Exploring the Quality in Use of Web 2.0 Applications: The Case of Mind Mapping Services

Tihomir Orehovački[1], Andrina Granić[2], and Dragutin Kermek[1]

[1] University of Zagreb, Faculty of Organization and Informatics
Pavlinska 2, 42000 Varaždin, Croatia
`{tihomir.orehovacki,dragutin.kermek}@foi.hr`
[2] University of Split, Faculty of Science
Nikole Tesle 12, 21000 Split, Croatia
`andrina.granic@pmfst.hr`

Abstract. Research in Web quality has addressed quality in use as the most important factor affecting a wide acceptance of software applications. It can be conceived as comprising two complementary concepts, that is, usability and user experience, which accounts for the employment of more user-centred evaluations. Nevertheless, in the context of Web 2.0 applications, this topic has still not attracted sufficient attention from the HCI community. This paper addresses the quality in use of Web 2.0 applications on the case of mind mapping services. The evaluation methodology brings together three complementary methods. The estimated quality in use is measured by means of the logging actual use method, while the perceived quality in use is evaluated by means of the retrospective thinking aloud (RTA) method and a questionnaire. The contribution of our work is twofold. Firstly, we provide empirical evidence that the proposed methodology in conjunction with the model, set of attributes, and measuring instruments is appropriate for evaluating quality in use of Web 2.0 applications. Secondly, the analysis of qualitative data reveals that performance and effort based attributes considerably contribute to mind mapping services success.

Keywords: Web 2.0, Quality in Use, Evaluation Methodology, Study Results.

1   Introduction

Usability evaluation plays an essential role in the human-centred design process of interactive software applications. *Usability*, as a quality of use in context [4], is related to ease-of-use and ease-of-learning. More recently, a concept of *user experience* (UX) [9] has been gaining popularity, leading to a switch in the research focus from product-centred evaluation to more user-oriented one. Furthermore, due to the emergence of Web 2.0 applications, the role of user experience in the assessment process has become even more important. As a result of these developments, latest research in Web quality has been addressing *quality in use* that is considered to be one of the most important factors affecting a wide acceptance of software applications in general.

1.1 Research Background

In recent research in the field of usability, user experience and quality in use, e.g. [3], [6], [11], [13] along with the latest quality standard [10], no agreement has so far been reached on attributes which reflect the 'real quality' of a software application. Moreover, it is not clear how the concept of quality in use should be defined in the context of Web 2.0 applications. According to the ISO standard on quality models [10], usability (along with flexibility and safety) is a characteristic of quality in use, with effectiveness, efficiency and satisfaction as its sub-characteristics. In accordance with ISO 25010, Bevan perceived usability as performance in use and satisfaction in terms of its relation with user experience [3]. To encompass the overall user experience, satisfaction needs to be concerned with both pragmatic and hedonic user goals. Lew *et al.* [11] proposed extending the ISO 25010 standard to incorporate new characteristics, in particular information quality and learnability. They also argued for including usability and user experience concepts into the modelling framework. Taking the above into consideration, quality in use could be seen as comprising two complementary concepts: (i) usability, which refers to the product-centred evaluation of pragmatic attributes through the use of both subjective and objective measuring instruments as well as (ii) user experience (UX), which concerns the use of subjective measuring instruments for the assessment of hedonic attributes.

HCI literature offers a lot of different models, methods and standards aimed at evaluating the quality and usability of software applications. However, research related to the evaluation of Web 2.0 applications in general has been deficient. Recent studies suggested that the reason for that might be the inappropriateness of current approaches for evaluation of those applications. A research into usability assessment carried out by Hart *et al*. [8] revealed that the popular social networking site Facebook complies with only two of ten heuristics originally proposed by Nielsen [12]. They also reported that the attributes such as ease of use, usefulness and playfulness have a major impact on users' loyal behaviour. When subjected to conventional usability evaluation, YouTube appears to score badly as well, meeting only two traditional heuristics [17]. Thompson and Kemp [18] argued that one of the main reasons why Web 2.0 applications such as Flickr, Wikipedia and YouTube have a large number of active users is their focus on user experience. Moreover, they extended and modified a set of Nielsen's traditional heuristics with an objective to evaluate the usability of Web 2.0 applications. However, the validity of the proposed set of heuristics has so far not been empirically confirmed. In addition, current research is usually focused on the development of methods and models aimed for the evaluation of particular quality aspects (e.g. information quality [1]) or types of Web 2.0 applications (e.g. mashups [5]). All the afore-mentioned findings motivated us to initiate our research into the design of a methodology that would enable the evaluation of the quality in use of Web 2.0 applications, regardless of their type and the context in which they are used [14].

1.2 Proposed Classification of Quality in Use Attributes

Our analysis of relevant recent research in the field of Web quality and usability assessment resulted in a set of attributes that may have a significant role in the evaluation of the quality in use of Web 2.0 applications [16]. The developed

conceptual model shown in Figure 1 classifies quality in use attributes into six basic categories: system quality (SYQ), service quality (SEQ), information quality (INQ), performance (PFM), effort (EFO), and acceptability (ACP).
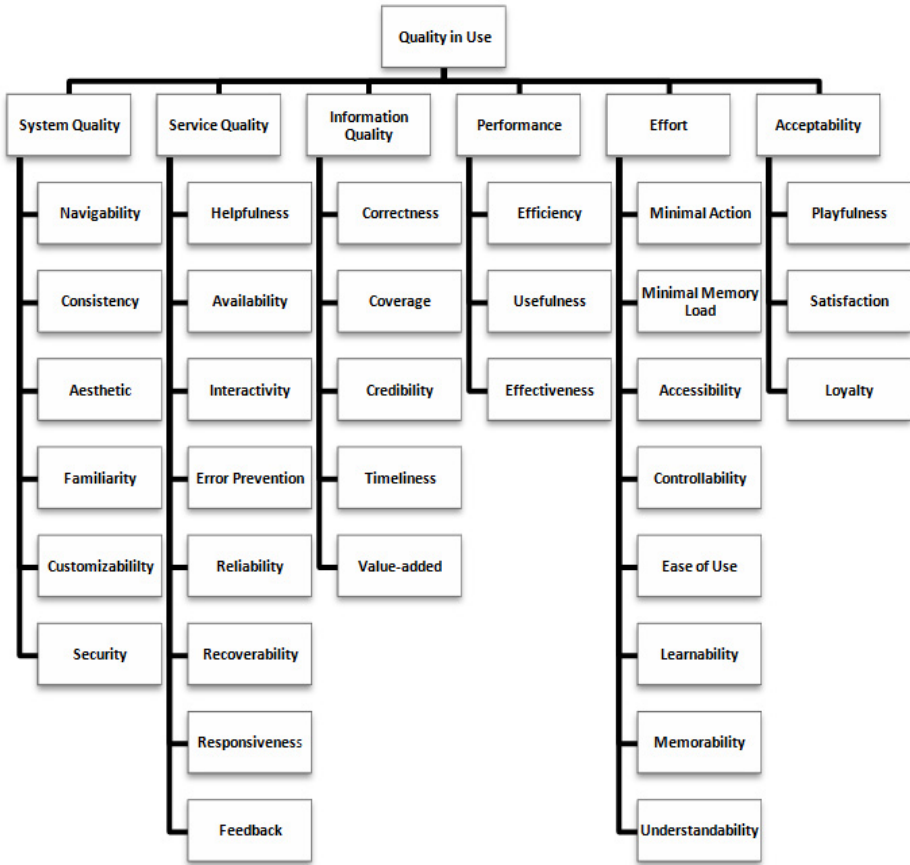

Fig. 1. Proposed model for evaluating Quality in Use of Web 2.0 applications [16]

System Quality (SYQ) measures quality in use of Web 2.0 application at the level of its interface features. It is comprised of six attributes: *navigability* (NVG, degree to which interface elements are well organized and alternative navigation mechanisms are provided), *consistency* (CNS, degree to which the same structure, design, terminology and components are used throughout a Web 2.0 application), *aesthetic* (AES, degree of visual attractiveness of a Web interface), *familiarity* (FML, degree to which a Web interface is similar to previously used applications), *customizability* (CUS, degree to which interface elements can be adapted to the characteristics of the task or user), and *security* (SCR, extent to which personal data and files are protected from unauthorized access). *Service quality* (SEQ) is the extent of quality of interaction between the user and a Web 2.0 application. This category is further decomposed into eight attributes: *helpfulness* (HLP, degree to which a Web application contains modules for user's assistance), *availability* (AVL, degree to which interface elements are continuously

available), *interactivity* (ITR, degree to which a Web 2.0 application creates the feeling of use of a desktop application), *error prevention* (ERP, degree to which a Web application prevents the occurrence of errors), *reliability* (RLB, degree to which a Web 2.0 application works without errors or interruptions), *recoverability* (RCV, the extent to which a Web 2.0 application can recover from errors and interruptions in its running), *responsiveness* (RSP, extent of the speed of a Web 2.0 application's response to users' requests and actions), and *feedback* (FDB, extent to which a Web 2.0 application displays its status or progress at any time). *Information Quality* (INQ) captures the quality of the content which proceeds out of using a Web 2.0 application. This category contains five different attributes: *correctness* (CRC, degree to which information content is free of errors), *coverag*e (CRG, degree to which information content is appropriate, complete and compactly represented), *credibility* (CDB, degree to which information content is unbiased, trustworthy, and verifiable), *timeliness* (TLS, degree to which information content is up to date), and *value-added* (VAD, degree to which information content is advantageous). *Performance* (PFM) refers to the quality of performing assignments by means of a Web 2.0 application interface functionalities. This category includes three attributes: *effectiveness* (EFE, degree to which an assignment can be achieved with accuracy, and completeness), *usefulness* (UFL, degree to which the user perceives a Web 2.0 application as the most appropriate solution for performing the assignment), and *efficiency* (EFI, degree to which a goal can be achieved with minimal consumption of resources). *Effort* (EFO) is the extent of perceived and estimated mental and physical energy when executing a task with Web 2.0 applications. This category is subdivided into eight attributes: *minimal action* (MAC, degree to which an assignment solution can be achieved in a minimum number of steps), *minimal memory load* (MEL, amount of information the user needs to remember when carrying out tasks), *accessibility* (ACS, extent to which a Web 2.0 application can be used by people with a widest range of disabilities), *controllability* (CTR, level of user's freedom while completing the task), *ease of use* (EOU, degree to which a Web 2.0 application can be used without help), *learnability* (LRN, which measures how easily the user can learn to use a Web interface functionalities), *memorability* (MRB, which measures how easy it is to memorize and remember how to use a Web 2.0 application), and *understandability* (UND, extent to which interface elements are clear and unambiguous to the user). To facilitate data collection in this study, two theoretically separated attributes, that is, minimal action and minimal memory load, are logically combined into a single attribute that is named physical and mental effort (PME). *Acceptability* (ACP) consists of attributes that directly contribute to the success of a Web 2.0 application, including *playfulness* (PLY, extent to which using a Web 2.0 application is fun and stimulating), *satisfaction* (STF, extent to which a Web 2.0 application can meet user's expectations) and *loyalty* (LOY, the users' intention to continue to use a Web application or to recommend it to their colleagues). The main aim of this paper is to investigate to what extent the proposed model and associated measuring instruments are appropriate for evaluating the quality in use of Web 2.0 applications, particularly mind mapping services.

2   Methodology

Participants. A total of 86 respondents (70.9% male, 29.1% female), aged 20.31 years (SD = 1.868) on average, participated in the survey. Participants were students

of Information Science from the University of Zagreb. All of them had been using popular Web 2.0 applications (Facebook and YouTube) on a regular basis (71% and 77.9%, respectively, did that twice a day or more often). The study was conducted within the Data Structures course. It should be noted that students had not participated in similar studies before.

Procedure and apparatus. The study adopted a within-subjects design contrasting four different Web 2.0 applications for mind mapping. During one semester, students had to solve four different programming tasks. In addition to writing the programming code, an integral part of each task was to graphically display an algorithm by means of a mind map. All the tasks were of equal complexity. Before the experiment started, we had defined which Web 2.0 application must be used when performing a particular task. Web 2.0 applications that were involved in the study are presented in Figure 2.
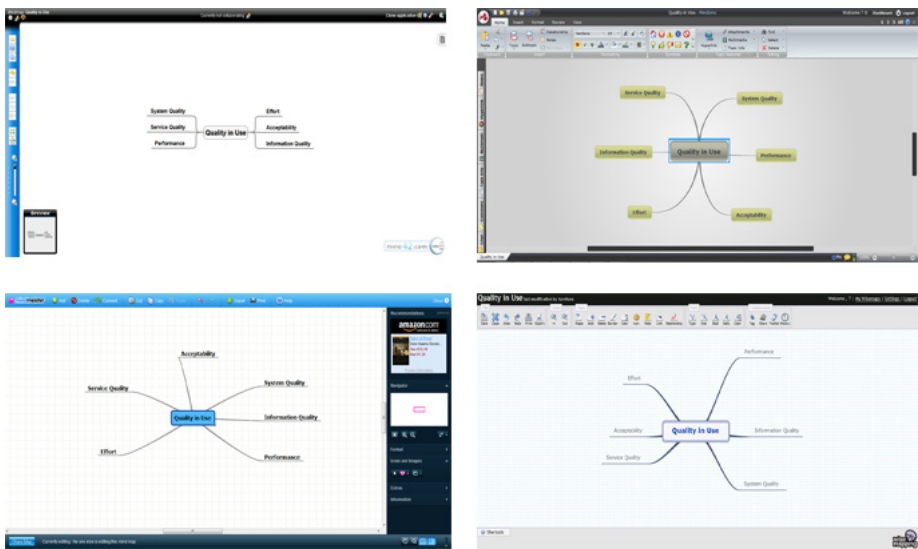

Fig. 2. Screenshots of evaluated Web 2.0 applications (clockwise, starting from top left: Mind 42, Mindomo, Wise Mapping, and Mindmeister)

Data were gathered by both objective and subjective means. Objective quality in use attributes (efficiency and effort) were measured using a Mousotron [2], which kept track of timing, mouse movements, mouse clicks, and keystrokes. To ensure maximum data accuracy, students were given detailed written and oral instructions at the beginning of each assignment. In the first step, students had to create an account and log in. Once the working environment had been loaded, it was necessary to run Mousotron and start task execution. Immediately after the task was completed, students needed to stop the data collecting process. The results gathered by means of Mousotron and the solutions for the assignment were supposed to be uploaded to the e-learning system. In order to obtain as much data about the advantages and disadvantages of the used mind mapping Web 2.0 applications as possible, the retrospective thinking aloud (RTA) method was employed. RTA allowed students to complete the assignment first, and then describe their experiences of working with a

Web 2.0 application. Immediately after they had completed the educational activity, students had to provide a critical review of the mind mapping application and deliver it in the form of a written report. One of the authors conducted a two-phase analysis of the data collected with the RTA method. Firstly, positive and negative comments related to the particular Web 2.0 applications were extracted from reports. Subsequently, each comment was attached to quality in use attributes whose definition they fit most closely. At the end of the semester, the perceived quality in use was evaluated by means of a post-use questionnaire.

Measures. Effort was measured by means of three metrics: distance traversed, mouse clicks, and keystrokes. *Distance* refers to the number of millimeters traversed by the user while moving the mouse between two points. *Mouse clicks* are the sum of all the mouse clicks (left, right, middle) that were made while reaching the task solution. Double clicks were not specifically analyzed, but were recorded as left clicks. *Keystrokes* denote the total number of keys on the keyboard that the user clicked while reaching the task solution. Other parameters that potentially affect the amount of physical effort, such as the number of the mouse wheel scrolls, are beyond the scope of this study. The sum of previous three metrics is referred to as *estimated effort*. Apart from physical effort, efficiency in use was also measured. *Time* is the amount of time expressed in seconds required to complete the task. *Mouse moving speed (MMS)* is the ratio of distance and time expressed in millimetres per minute (mm/min). *Keyboard typing speed (KTS)* is the ratio of keystrokes and time expressed in the number of keystrokes per minute (ks/min). *Mouse clicking speed (MCS)* is the ratio of mouse clicks and time expressed in the number of mouse clicks per minute (mc/min). The sum of keyboard typing speed, mouse moving speed and mouse clicking speed is referred to as *estimated efficiency*. Given that the execution of assignments was not time limited, objective assessment of the effectiveness was beyond the scope of this study. The post-use questionnaire was used for gathering data about perceived quality in use of the evaluated Web 2.0 applications. Each quality in use attribute was measured with between three and eight items. Responses were modulated on a five point Likert scale (1-strongly agree, 5-strongly disagree). In addition, overall preferences were assessed directly by a 4-point semantic differential item (1-best, 4-worst) in which users were asked to rank the quality in use of mind mapping Web 2.0 applications. The Cronbach's alpha values (presented in Table 1) ranged from .701 to .896, thus indicating a high reliability of the scale. Combining three different methods (logging actual use, questionnaire and RTA), complementary data of the estimated and perceived quality in use of the mind mapping applications was gathered. In this way, the process of detection of problems in Web 2.0 applications usage was accelerated, while the identification of key quality in use attributes was facilitated. The research results are presented in more detail in the following section.

Table 1. Internal reliability of scale

	Mind 42	Mindmeister	Mindomo	Wise Mapping
System Quality	.779	.767	.798	.794
Service Quality	.720	.701	.732	.741
Information Quality	.811	.789	.769	.788
Performance	.816	.741	.774	.796
Effort	.896	.868	.884	.885
Acceptability	.855	.850	.863	.831

3   Results

Given that data were not normally distributed (K-S 1,2,3,4 > 0.05), the analysis was conducted by means of non-parametric tests. All the reported results are expressed as the median value.

3.1   Estimated Quality in Use

Effort. The analysis of the data revealed statistically significant differences among the four Web 2.0 interfaces in terms of mouse movements during task execution ($\chi2(3) = 23.255$, $p < .001$). The Wilcoxon Signed-Rank Tests with Bonferroni pairwise comparisons revealed that participants needed to make significantly fewer mouse movements to solve the task using the Mind 42 than either Mindmeister ($Z = -3.585$, $p = .000$, $r = -.27$) or Mindomo ($Z = -4.433$, $p = .000$, $r = -.34$). Both effects were medium in size with the significance level set at $p < .008$. No significant differences were observed in other pairwise comparisons.

The type of the Web 2.0 application used significantly affected the amount of mouse clicks made by users during the mind map design ($\chi2(3) = 11.102$, $p < .05$). A comparison of the number of mouse clicks among all four Web 2.0 applications revealed that users made much fewer mouse clicks using Mind 42 or Mindmeister than using Mindomo or Wise Mapping. Therefore, the number of required comparisons in the post-hoc analysis was reduced and the significance level set at $p < .0125$. A significant difference in the number of mouse clicks was found between Mindmeister and Wise Mapping ($Z = -2.995$, $p = .003$, $r = -.23$), Mind 42 and Wise Mapping ($Z = -2.824$, $p = .005$, $r = -.22$), and Mindmeister and Mindomo ($Z = -2.556$, $p = .011$, $r = -.19$), respectively, while the difference between Mindmeister and Mindomo was not significant ($Z = -1.955$, $p = .051$). All the effects were small in size. No significant difference was found among all four Web 2.0 applications in terms of the amount of keystrokes made when reaching the task solution ($\chi2(3) = 1.806$, $p = .614$).

Table 2. Results of objective measure effort for four selected mind mapping Web 2.0 applications (note that a lower score means a better result)

	Mind 42	Mindmeister	Mindomo	Wise Mapping
Distance (mm)	30759	38049	41313	39411
Keystrokes	554	536	612	561
Mouse clicks	292	286	353	361
Effort	**31610**	**38993**	**42693**	**40248**

There was a significant difference in the amount of estimated effort among all four Web 2.0 applications ($\chi2(3) = 22.858$, $p < .001$). Pairwise comparison revealed a significant difference between Mind 42 and Mindomo ($Z = -4.407$, $p = .000$,

r = - .37), and between Mind 42 and Mindmeister (Z = -3.563, p = .000, r = - .27). The difference between Mind 42 and Wise Mapping was on the verge of the significance level (p < .0125). According to the summary of the results presented in Table 2, students experienced less effort using Mind 42 than any of the three remaining Web 2.0 applications.

Efficiency. Friedman's ANOVA revealed a significant difference among four different Web 2.0 applications in the speed of moving the mouse during task solving ($\chi2(3) = 40.083$, p < .001). As a follow up for this finding, a post-hoc analysis with the significance level set at p < 0.125 was applied. A significant difference in the speed of mouse movements was found between Mind 42 and Mindmeister (Z = - 4.567, p = .000, r = - .35), Mind 42 and Mindomo (Z = -4.825, p = .000, r = - .37), Wise Mapping and Mindmeister (Z = -4.192, p = .000, r = - .32), and between Wise Mapping and Mindomo (Z = -3.718, p = .000, r = - .28), respectively. Neither the keyboard typing ($\chi2(3) = 1.806$, p = .614) nor mouse clicking ($\chi2(3) = 6.402$, p = .094) speeds were significantly different among all four evaluated Web 2.0 applications. A significant difference in the estimated efficiency of evaluated Web 2.0 applications was discovered ($\chi2(3) = 41.829$, p < .001). A post hoc analysis showed a significant difference in the overall efficiency between Mindomo and Mind 42 (Z = - 4.851, p = .000, r = - .37), Mindmeister and Mind 42 (Z = -4.549, p = .000, r = - .35), Wise Mapping and Mindmeister (Z = -4.231, p = .000, r = - .35), and Wise Mapping and Mindomo (Z = -3.757, p = .000, r = - .29). According to the results presented in Table 3, of all four evaluated Web 2.0 applications, Mindmeister was the most efficient in accomplishing the assignments.

Table 3. Results of objective measure efficiency for four selected mind mapping Web 2.0 applications (note that a higher score means a better result)

	Mind 42	**Mindmeister**	**Mindomo**	**Wise Mapping**
MMS (mm/min)	1952	2573	2550	2074
KTS (ks/min)	35	34	35	33
MCS (mc/min)	18	19	19	20
Efficiency	**2013**	**2635**	**2604**	**2122**

3.2 Perceived Quality in Use

Rank. A significant value of chi square ($\chi2(3) = 37.381$, p < .001) indicates the existence of differences in the subjective ranking measure among the evaluated Web 2.0 applications. A post-hoc procedure with the significance level set at p < .0167 revealed differences between Wise Mapping and Mindmeister (Z = -4.800, p = .000, r = - .37), Wise Mapping and Mindomo (Z = -4.668, p = .000, r = - .36), Mindomo and Mind 42 (Z = -2.864, p = .004, r = - .22), Mindmeister and Mind 42 (Z = -2.671,

p = .008, r = - .20), and Wise Mapping and Mind 42 (Z = -2.605, p = .009, r = - .20). The results of overall subjective preferences presented in Table 4 indicate that Mindomo has the highest rank of perceived quality in use.

Table 4. Results of subjective measure rank for four selected mind mapping Web 2.0 applications (note that a lower score means a better result)

	M Rank	χ^2	df	p
Mind 42	2.65	37.381	3	< .001
Mindmeister	2.19			
Mindomo	2.04			
Wise Mapping	3.12			

Questionnaire. The results show that the scores of the four applications differ significantly ($\chi^2(3)$ = 27.599, p < .001). Wilcoxon Signed-Rank Tests with a Bonferroni correction were used to follow up on this finding. Significant differences were found between Wise Mapping and Mindomo (Z = -4.394, p = .000, r = - .34), Mindomo and Mind 42 (Z = -4.073, p = .000, r = - .31), Wise Mapping and Mindmeister (Z = -3.926, p = .000, r = - .30), and between Mindmeister and Mind 42 (Z = -2.915, p = .004, r = - .22). All the effects were medium in size. The summary of the results obtained from the post-use questionnaire is presented in Table 5.

Table 5. Results of overall perceived quality in use of four selected mind mapping Web 2.0 applications (note that a lower score means a better result)

	Median	SD	χ^2	p
Mind 42	334.00	50.471	27.599	< .001
Mindmeister	328.00	46.658		
Mindomo	321.50	47.000		
Wise Mapping	338.50	46.962		

Retrospective Thinking Aloud. Data analysis revealed that students had generated a total of 1711 comments related to the advantages (63.18%) and disadvantages (36.82%) of the used Web 2.0 applications. In general, effort and performance based attributes were reported in most cases during RTA sessions, while the attributes related to the information quality were mentioned rather rarely. In particular, the most important quality in use attributes reported by students are ease of use, effectiveness, controllability and interactivity, while in the context of Web 2.0 applications used for mind mapping the attributes such as consistency, feedback, accessibility and memorability are of little importance. Based on overall results presented in Table 6, the best ratio of reported advantages and disadvantages in use belongs to Mindomo, followed by Mindmeister, Mind 42, and Wise Mapping.

Table 6. Results of Retrospective Thinking Aloud (RTA) method

	Mind 42		Mindmeister		Mindomo		Wise Mapping	
	Pros	Cons	Pros	Cons	Pros	Cons	Pros	Cons
NVG	27	5	16	7	17	5	9	14
CNS								
AES	12	3	14	6	12		5	7
FML	5	1	5		22	1	4	5
CUS	1	13	15	15	21	5	5	17
SCR								1
HLP	1		3	3	7		7	1
AVL	1	1	6	8	12	5	5	1
ITR	24	1	24		27	5	15	18
ERP	10	3	8	4	5	5	2	9
RLB	7	1	10	6	14	2	7	13
RCV					1		1	1
RSP			1		3	5	1	4
FDB								
CRG	4	7	3	12	13	3	4	20
EFE	24	37	28	24	52	8	16	38
UFL	14	2	17	1	11	2	3	10
EFI	15	1	11	3	14	5	8	16
PME					2			1
ACS								
CTR		40	7	20	17	10	4	26
EOU	64	2	53	5	71	3	42	14
LRN	3	3	7	4	12	1	5	5
MRB								
UND	16	1	6	5	15	4	6	13
PLY	1		1		1			
STF	34	14	26	9	39	7	15	20
LOY	1	4		1	1			1
CMP*			3	1		3		23

* Attribute was not included in model proposed in [16]

4 Discussion and Concluding Remarks

The objective of the research described in this paper was the design of a methodology for evaluating quality in use of Web 2.0 applications [14]. In order to accomplish the research goal, the quality in use of mind mapping applications was evaluated with three different methods: logging actual use, questionnaire, and retrospective thinking aloud.

The purpose of the experiment presented in this paper was twofold. Firstly, we aimed to determine to what extent the conceptual model and the corresponding measuring instrument we developed would be suitable for the evaluation of Web 2.0 applications. The analysis of the data gathered by means of the logging actual use method revealed that in the analysis and comparison of Web 2.0 applications the following can be used: distance traversed, number of mouse clicks and mouse moving speed. Namely, through the use of these objective metrics a statistically significant

difference between the evaluated Web 2.0 applications can be determined. However, the measures of numbers of keystrokes, keyboard typing and mouse clicking speeds did not show discriminant validity. We believe that this occurred because the experiment was not time-limited. Another possible reason may be a narrow specialization of evaluated applications. Accordingly, the results obtained from the post-use questionnaire showed statistically significant differences among all four evaluated Web 2.0 applications. This suggests that Web 2.0 applications can be ranked by mean values.

Secondly, we aimed to identify the importance that users attach to certain attributes of quality in use and to detect whether the set of the most important attributes depends on the type of Web 2.0 application. The results gathered by means of the retrospective thinking aloud (RTA) method revealed the importance of effort (28.99%) and performance (21.43%) based attributes of quality in use. In particular, participants felt highly satisfied and comfortable working with Web 2.0 applications meeting the following quality in use attributes: ease of use, effectiveness, controllability, interactivity, navigability, customizability, efficiency, information content coverage, understandability, and reliability. In addition, data analysis showed that some attributes (consistency, feedback, accessibility, and memorability) were not mentioned at all during the RTA session. Possible reasons may be: majority of users not having any kind of disability; the ability to evaluate memorability when the application is re-used with a time lag, etc. On the other hand, compatibility, i.e. the degree to which a Web 2.0 application works equally well within different browsers, operating systems, or devices, which was not included in the proposed model, has proven to be an important indicator of problems in use of Web 2.0 applications. The results were similar to the findings presented in [15] which suggest that: (i) there is a general set of attributes that needs to be measured independently of the type of Web 2.0 applications; (ii) the weight of an attribute depends on the type of the evaluated Web 2.0 application; and (iii) there is a set of attributes aimed for measuring the quality in use of specific types of Web 2.0 applications. In our research, the results of the estimated and perceived quality in use do not match. Such findings are in accordance with those presented in e.g. [7]), indicating that quality in use should be measured with both subjective and objective instruments since they are aimed for evaluating different aspects of Web 2.0 applications. In addition, we must emphasize that a homogeneous set of four evaluated applications is a fairly modest sample on the basis of which generalizable sound conclusions on the importance of each category on the quality in use of Web 2.0 applications can not be drawn. Therefore, our future work will be focused on: (i) applying the proposed model to evaluate the quality in use of various Web 2.0 applications in a different context of use; (ii) revision of attributes, model and measuring instruments; (iii) improvement of a proposed methodology with an aim to facilitate analysis and comparison of the evaluated Web 2.0 applications.

References

1. Almeida, J.M., Gonçalves, M.A., Figueiredo, F., Pinto, H., Belém, F.: On the Quality of Information for Web 2.0 Services. IEEE Internet Computing 14(6), 47–55 (2010)
2. Blacksun Software (2011),
   `http://www.blacksunsoftware.com/mousotron.html`

3. Bevan, N.: Extending Quality in Use to Provide a Framework for Usability Measurement. In: Kurosu, M. (ed.) HCD 2009. LNCS, vol. 5619, pp. 13–22. Springer, Heidelberg (2009)

4. Bevan, N., Macleod, M.: Usability measurement in context. Behaviour & Information Technology 13, 132–145 (1994)

5. Cappiello, C., Daniel, F., Matera, M.: A Quality Model for Mashup Components. In: Gaedke, M., Grossniklaus, M., Díaz, O. (eds.) ICWE 2009. LNCS, vol. 5648, pp. 236–250. Springer, Heidelberg (2009)

6. Chiou, W.-C., Lin, C.-C., Perng, C.: A strategic framework for website evaluation based on a review of the literature from 1995-2006. Information & Management 47, 282–290 (2010)

7. Frøkjær, E., Hertzum, M., Hornbæk, K.: Measuring usability: Are effectiveness, efficiency, and satisfaction really correlated? In: Proceedings of the ACM CHI Conference on Human Factors in Computing Systems, pp. 345–352. ACM, New York (2000)

8. Hart, J., Ridley, C., Taher, F., Sas, C., Dix, A.: Exploring the Facebook Experience: A New Approach to Usability. In: 5th Nordic Conference on Human-Computer Interaction: Building Bridges, pp. 471–474. ACM, Lund (2008)

9. Hassenzahl, M., Tractinsky, N.: User experience - a research agenda. Behaviour & Information Technology 25(2), 91–97 (2006)

10. ISO/IEC 25010:2011. Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models (2011)

11. Lew, P., Olsina, L., Zhang, L.: Quality, Quality in Use, Actual Usability and User Experience as Key Drivers for Web Application Evaluation. In: Benatallah, B., Casati, F., Kappel, G., Rossi, G. (eds.) ICWE 2010. LNCS, vol. 6189, pp. 218–232. Springer, Heidelberg (2010)

12. Nielsen, J.: Heuristic evaluation. In: Nielsen, J., Mack, R.L. (eds.) Usability Inspection Methods, John Wiley & Sons, New York (1994)

13. Olsina, L., Sassano, R., Mich, L.: Specifying Quality Requirements for the Web 2.0 Applications. In: Proceedings of 7th International Workshop on Web-oriented Software Technology (IWWOST 2008), pp. 56–62. CEUR, Bratislava (2008)

14. Orehovački, T.: Development of a Methodology for Evaluating the Quality in Use of Web 2.0 Applications. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part IV. LNCS, vol. 6949, pp. 382–385. Springer, Heidelberg (2011)

15. Orehovački, T.: Perceived Quality of Cloud Based Applications for Collaborative Writing. In: Pokorny, J., et al. (eds.) Information Systems Development – Business Systems and Services: Modeling and Development, pp. 575–586. Springer, Heidelberg (2011)

16. Orehovački, T.: Proposal for a Set of Quality Attributes Relevant for Web 2.0 Application Success. In: 32nd International Conference on Information Technology Interfaces, pp. 319–326. IEEE Press, Cavtat (2010)

17. Silva, P.A., Dix, A.: Usability – Not as we know it! In: 21st British HCI Group Annual Conference on HCI 2007: People and Computers XXI: HCI...But not as We Know It, vol. 2, pp. 103–106. ACM, University of Lancaster (2007)

18. Thompson, A.-J., Kemp, E.A.: Web 2.0: extending the framework for heuristic evaluation. In: 10th International Conference NZ Chapter of the ACM's Special Interest Group on Human-Computer Interaction, pp. 29–36. ACM, New Zealand (2009)