

Identifying Relevant Dimensions for the Quality of Web Mashups: An Empirical Study

Tihomir Orehovački¹(✉), Cinzia Cappiello², and Maristella Matera²

¹ Department of Information and Communication Technologies,
Juraj Dobrila University of Pula, Zagrebačka 30, 52100 Pula, Croatia
tihomir.orehovacki@unipu.hr

² Dipartimento di Elettronica, Informazione e Bioingegneria,
Politecnico di Milano, Via Ponzio 34/5, 20133 Milan, Italy
{cinzia.cappiello,maristella.matera}@polimi.it

Abstract. Web mashups available online today are very often characterized by a poor quality. Several researchers justify this aspect by considering the situational, short-living nature of these applications. We instead believe such a low quality is also due to the lack of suitable quality models. This paper presents a quality model that tries to capture the nature of Web mashups by focusing on their component-based nature and the added value that they are required to introduce with respect to their single constituents. A finite set of indicators and attributes was first determined by reviewing the literature. An analysis of data collected from domain experts revealed a relevance of performance variables at different levels of granularity. An empirical study was then carried out to assess which dimensions are the most relevant with respect to the mashup quality as perceived by users.

Keywords: Web mashups · Quality evaluation · Model validation · Relevance identification · Empirical study

1 Introduction

Web mashups are composite applications that integrate reusable data, application logic and/or user interfaces typically - but not mandatorily, sourced from the Web [11]. After several years of research and development experiences on this class of applications, it is still difficult to find high-quality, useful mashups on the Web. On the one hand, it is true that still there are not stable development practices and tools. On the other hand, Web mashups are meant to satisfy situational, short-living needs and in this scenario quality might not be a primary concern. Moreover, suitable quality models, able to capture mashup peculiarities, are still lacking. If adequate models would be available, developers and final users as well would reach an increased awareness of how mashups could (and should) be.

We first observed a lack of adequate quality models when we conducted a study on the mashups published on programmableweb.com [5], the reference Web site for the community of API and Web mashup developers. To evaluate a subset of about 100

mashups available on that site we adopted quality metrics generally valid for Web applications. We compared the objective results achieved by computing such metrics with the findings of a heuristic evaluation conducted by a pool of independent evaluators acquainted with Web technologies and mashup development. The result was a sharp discrepancy between the two assessments, highlighting that mashup quality requires revised models able to capture the specifics of such applications. Certainly, traditional quality principles for Web applications must not be neglected but, even for simple Web mashups, generic Web models need to be repurposed.

Given the previous observations, this paper investigates the relevance of traditional quality factors with respect to the nature of mashups and what additional factors can be further considered to focus especially on the added value that such composite applications can introduce with respect to exploiting their single components. Drawing on an extensive literature review, we determined a finite set of indicators and attributes that contribute to the quality of mashups and employed them to design a conceptual model in the form of a quality requirements tree. An analysis of data collected from domain experts revealed a relevance of performance variables at different levels of granularity in a quality requirements tree. With the objective to examine the validity of the introduced conceptual model, an empirical study was then carried out. During the study, participants accomplished predefined scenarios of interaction with a representative sample of mashups and assessed them by exploiting the quality model. Study findings helped us determine which dimensions, among those identified in the quality requirements tree, are considered as relevant with respect to the mashup perceived quality.

The paper is organized as follows. Section 2 reviews the most relevant related works discussing the quality of mashups. Section 3 then illustrates the model that we defined. Section 4 describes how we validated the model and the main implications deriving from the conducted empirical study. Section 5 finally draws our conclusions.

2 Related Work

A quality model consists of a selection of quality characteristics that are relevant for a given class of software applications and/or for a given assessment process [13]. In the Web scenario, the first quality models focused on static Web sites [17, 39], then some authors started addressing more complex Web applications [25, 29]. Recently, quality models for Web 2.0 applications have been proposed [35, 38, 41]. In the more restricted mashup context, the quality dimensions suggested by all these works, as well as the one proposed in Software Engineering [15, 19] and Web Engineering [4, 27] may be partly appropriate to measure the internal quality of a mashup (e.g., code readability), as well as its external quality in-use (e.g., usability). However, Rio and Brito e Abreu [39] showed that, in order to effectively support system development and evaluation, quality model must be domain-dependent, as there is a strong impact of the application domain on the usefulness of quality dimensions. In line with these findings, this paper aims to assess the relevance of some quality dimensions with respect to the mashup peculiarities. This research is motivated by some past experience of some authors of this paper that showed that quality models that generally work well for

traditional Web applications do not help identify even sever problems when applied to the evaluation of mashups [5].

Other works also tried to explore diverse aspects of the mashup quality and usability. Drawing on recent standards [16] and usability guidelines for Web design, Insfran et al. [14] proposed the Mashup Usability Model that decomposes usability into appropriateness, recognisability, learnability, operability, user error protection, user interface aesthetics, and accessibility. Koschmider et al. [20] then emphasized that the selection of quality metrics also depends on the type of mashups. For example, UI mashup should be evaluated especially in terms of consistent graphical representation, while data and function mashups are required to fulfill criteria that address more the integration and the orchestration of the involved resources.

Quality of the mashup can be also observed from the perspective of heterogeneous components that constitute it [6] as well as from the aspect of the final composition [43]. In that respect, Cappiello et al. [5] developed a model which addresses: data quality (accuracy, timeliness, completeness, availability, and consistency), presentation quality (usability and accessibility), and composition quality (added value, component suitability, component usage, consistency, and availability). Nevertheless, results of a systematic mapping study [9] suggest that there is still a need for empirical research addressing the assessment of the proposed quality dimensions.

In this paper we capitalize on all the previous works. The quality dimensions that we investigated were selected by carefully reviewing all such works. In addition, as illustrated in the next sections, we experimentally show how some dimensions are more relevant than others. We believe this effort to validate the model is original and introduces a valuable contribution in the Web mashup domain.

3 A Quality Model for Web Mashups

Drawing on a comprehensive literature review that included prior studies on the assessment of mashups [5, 9, 11], mashup components [6], mashup tools [37], and Web 2.0 applications designed for collaborative writing [30, 31, 33], mind mapping [33, 35, 36], and diagramming [31, 35], an initial pool of 165 items meant for measuring diverse facets of quality in the context of mashups was designed. To ensure content validity at all levels of granularity in the model, the relevance of items was examined by two independent mashup experts on a three-point scale (1 - mandatory, 2 - desired, 3 - not relevant). Data collected from experts were examined with two criteria: content validity ratio (CVR) and average value of assigned relevance (\bar{x}). A total of 62 items which have not met the cut-off values of the aforementioned criteria (CVR = 0.99, $\bar{x} \geq 2.00$) [22, 23] were omitted from further analysis. This procedure resulted in a quality model for Web mashups which consists of 6 categories and 24 attributes.

System quality is composed of four quality attributes: efficiency, effectiveness, response time, and compatibility. *Efficiency* refers to the degree to which the employment of a Web mashup saves resources in a specified context. In that respect, Web mashups should be implemented in a way so that users can complete intended tasks in the shortest time possible and with minimal number of steps. *Effectiveness* is the extent to which users can, by means of a Web mashup, realize intended tasks

completely and accurately. Perceived effectiveness can be evaluated with two objective metrics: proportion of tasks completed and proportion of tasks that are completed correctly [15]. Keeping that in mind, the interface of a Web mashup should provide all functionalities needed for task completion. *Response time* reflects the degree to which a Web mashup efficiently reacts to users' actions. Considering that users have very low tolerance threshold related to the response time, this quality attribute plays an important role in a success of every Web application, including mashups [31]. Therefore, a Web mashup and its components need to be quickly loaded in a Web browser whereas execution of selected interface functionalities should take very little time. *Compatibility* represents a level to which a Web mashup operates properly with different types of devices and among different environments. Taking into account that Web mashups are commonly referred to as one of the representatives of Web 2.0 applications [34], they also have to meet the aforementioned compatibility criteria.

Service quality deals with attributes measuring the quality of interaction between a Web mashup and users. It includes three quality attributes: availability, reliability, and feedback. *Availability* denotes the extent to which a Web mashup and its components can be accessed at any time. *Reliability* refers to the degree to which a Web mashup is dependable, stable, and bug-free. Since this attribute belongs to a set of essential predictors of the quality and satisfaction of users [45], a Web mashup has to perform as intended, without errors or operational interruptions. *Feedback* is related to the extent to which a Web mashup returns appropriate messages and notifies users about its status or progress of displaying the content in its components. It originates from Nielsen's ten usability heuristics [26], and according to Seffah et al. [42] it can be used as a metric which indicates to what level a piece of software congruously responds to users' actions by supplying them with convenient messages. In that respect, a Web mashup should timely inform users with messages that are clear, understandable, precise, and useful.

Content Quality refers to the perception or the assessment of the suitability of the content that the mashup provides for a specific goal in a defined context. In this category, we consider five dimensions that analyze the provided data from different perspectives: content accuracy, content completeness, content credibility, content timeliness, and content added value. *Content accuracy* refers to the correctness of the content that is displayed as output in the mashup. Correctness is usually assessed as the similarity between the considered value and the correct one [40]. *Content completeness* refers to the ability of mashup components to produce all expected data values [7]. Usually completeness is evaluated on the basis of the query submitted by the user. In fact, it can be defined as the ratio between the number of values obtained and the expected one. *Content credibility* is related to the trustworthiness of the mashup components and thus to the related data sources. Trustworthiness is an important aspect to consider especially when accuracy is not precisely assessable: if data are gathered from a certificated source, their correctness is guaranteed. *Content timeliness* refers to the fact that data have to be accessed at the right time. Data should be temporally valid. This dimension can be assessed as the ratio between currency (the data's "age" from the time of component creation or last update) and volatility (the average period of data validity in a specific context) [3]. *Content added value* refers to the possibility to gather more information and thus value from integration of different components. It refers the

possibility to perform additional queries due to the fact that data are integrated or to acquire more knowledge simply for the fact that data are visualized together (and not necessarily integrated [5]).

Composition Quality aims to evaluate the orchestration among components and the way in which the mashup provides the desired features [5]. This category includes three dimensions: component suitability, composition added value and effectiveness of integrated visualization. *Component suitability* refers to the suitability of the offered component functionalities and data with respect to the output that the mashup is supposed to provide. *Composition added value* refers to the functionalities and data offered by the mashup. In particular, the added value is generated by the new functionalities enabled by the integration of components. In fact, effective mashups exploit and combine the functionalities offered by the components in order to provide new and advanced operations. *Effectiveness of integrated visualization* refers to the cohesiveness of the visualization that is the opportunity to visualize in the same screen data coming from different sources [8]. Users benefit from the fact that heterogeneous data are aggregated into a unified visualization.

Effort stands for attributes, which measure the effortlessness of the Web mashup use. The following five attributes constitute this dimension: minimal memory load, accessibility, ease of use, learnability, and understandability. *Minimal memory load* refers to the amount of mental and perceptive activity needed for completing an intended task by means of the Web mashup. It is commonly employed for measuring the amount of information user needs to memorize to complete a particular task [42]. *Accessibility* denotes the extent to which the Web mashup is usable to people with the widest range of characteristics and capabilities. In order to achieve this goal, Web mashups have to comply with as many guidelines suggested in [46]. For instance, both interface functionalities and content returned by the Web mashup should be of sufficient size to be readable to visually impaired people. *Ease of use* presents the degree to which the use of the Web mashup is free of effort. Considering that ease of use significantly contributes to the perceived usefulness, and users' satisfaction [32], the Web mashup should be easy to operate in a way that users have no need to seek assistance of any kind when using it. *Learnability* refers to the extent to which it is easy to learn how to use the Web mashup. *Understandability* is the degree to which functionalities of the mashup interface are clear and unambiguous to users.

User experience concerns the quality attributes such as usefulness, playfulness, satisfaction, and loyalty which directly contribute to the adoption of the Web mashup by users. *Usefulness* refers to the extent to which the employment of the Web mashup enhances users' performance in completing intended tasks. Findings of prior study indicate that usefulness has significant influence on users' satisfaction and loyalty [32]. Taking this into account, features provided by the Web mashup should be advantageous compared to those offered by any other alternative. *Playfulness* is the degree to which the use of the Web mashup successfully holds users' attention. *Satisfaction* is the level to which users like to have an interaction with the Web mashup. It is of a great importance that Web mashup meet users' expectations. *Loyalty* refers to the extent to which users are willing to continue to use the Web mashups and recommend it to others. The Web mashup should be able to turn occasional visitors into regular users that are willing to spread a good word among their families, friends, and colleagues.

4 Quality Model Validation

4.1 Research Design

We conducted an empirical study adopting a within-subjects research design contrasting four Web mashups (“Gaiagi 3D Driver” (<http://www.gaiagi.com/driving-simulator>), “Health Map” (<http://www.healthmap.org>), “Leeds Travel Info” (<http://www.leedstravel.info>), and “This is Now” (<http://now.jit.su>)) that are heterogeneous with respect to their purpose. At the beginning of the study, details on the architecture of mashups, their taxonomy, and practical usefulness were presented to participants. In the next step, predefined scenarios with representative steps of interaction with mashups were given to each student. After finishing scenarios with all four mashups, students were asked to complete an online post-use questionnaire. It was composed of 6 items related to participants’ demography and 103 items meant for measuring 24 diverse facets of mashups’ quality as deriving from the model illustrated in the previous section. Responses to the questionnaire items were modulated on a four-point Likert scale (1 – strongly agree, 4 – strongly disagree). Each attribute was measured with between two and seven items. For the purpose of data analysis, attributes and categories were operationalized as composite subjective measures. Values for quality attributes were estimated as a sum of responses to items that are assigned to them. The same holds for quality categories and overall perceived quality of evaluated mashups.

The assumption that data was sampled from a Gaussian distribution was examined with the Shapiro-Wilk Test. Considering that in all comparisons the Shapiro-Wilk statistic for at least one variable significantly deviated from a normal distribution ($p < .05$), the analysis of collected data was conducted with non-parametric tests. With the goal to explore differences among evaluated mashups, the Friedman’s ANOVA expressed as chi-square (χ^2) value was applied as the non-parametric counterpart to the one-way ANOVA with repeated measures. By employing separate Wilcoxon Signed-Rank Test (Z) on all possible pairs of evaluated mashups, actual differences existing among them were identified. In order to avoid a Type I error and declare results of pairwise comparisons significant, a Bonferroni correction was applied to the results of Wilcoxon Signed-Rank Tests. It was calculated by dividing the significance level of .05 by number of comparisons. The effect size (r) is an objective measure that reflects the relevance of difference between a pair of evaluated mashups. It represents a quotient of Z -value and a square root of the number of observations. Values of .10, .30, or .50 for the effect size can be, as a rule of thumb, interpreted as small, medium, or large, respectively [10].

As regards participants, 43 subjects took part in the empirical study. They ranged in age from 19 to 45 years ($M = 20.93$, $SD = 3.900$). The sample was composed of 83.72 % male and 16.28 % female students. At the time when the study took place, they were all in the second year of an undergraduate programme in Information Systems. Up to the implementation of the study, 51.16 % of participants had never used mashups before. Remaining 48.84 % of students are using mashups at least once a week where majority of them (85.71 %) is spending less than an hour on interaction with mashups. On the other hand, study participants are loyal users of popular Web 2.0 applications such as Facebook, Twitter and Instagram. Majority of students (53.49 %) are using those

social Web applications three or more times a day. When frequency in use expressed in hours is considered, majority of study participants (69.76 %) are spending between four and ten hours a week on interaction with the aforementioned Web 2.0 applications.

4.2 Findings

Since another instrument for measuring the perceived quality of mashups does not exist in literature, it was not possible to conduct benchmarking thus obtain a quantitative measure of validity. As an alternative, Lewis [23] proposes the assessment of sensitivity which affects the validity of the measuring instrument. The sensitivity of the introduced model and employed post-use questionnaire was examined by exploring differences among evaluated mashups.

Friedman’s ANOVA revealed a significant difference ($\chi^2(3) = 19.866, p = .000$) among the four mashups in the overall quality perceived by study participants. Drawing on this finding, a post-hoc analysis with the significance level set at $p < .0125$ was applied. It was discovered that significant difference in perceived quality exists between This is Now and Health Map ($Z = -4.302, p = .000, r = -.46$), This is Now and Gaiagi 3D Driver ($Z = -2.839, p = .005, r = -.31$), Health Map and Gaiagi 3D Driver ($Z = -2.494, p = .013, r = -.27$), and between Leeds Travel Info and Health Map ($Z = -2.476, p = .013, r = -.27$).

When categories of quality are considered, *composition quality* is associated with the highest level of relevance. This resulted from two large in size (.53 and .51) and two medium in size (both .39) differences among evaluated mashups. It is followed by quality categories *user experience* that uncovered three medium in size (in range from .47 to .40) differences, *effort* which demonstrated one large (.50) and three medium in size (in range from .48 to .37) differences, and *content quality* which has shown one large (.50), one medium (.37), and one small (.27) in size difference among four mashups. Finally, it appeared that *system quality* and *service quality* have the lowest degree of relevance among identified categories meant for measuring quality of mashups. Namely, system quality revealed one medium in size (.34) difference whereas service quality uncovered two small in size (.28 and .26) differences among mashups that took part in the study. The aforementioned findings are summarized in Table 1.

Table 1. Taxonomy of proposed categories with respect to their relevance in evaluating the quality of Web mashups

| Quality categories for Web mashups | χ^2 | Effects in size (r) | | |
|------------------------------------|----------|-------------------------------------|--|-------------------------------------|
| | | Large | Medium | Small |
| Composition quality | 37.899 | .53 ^c , .51 ^e | .39 ^d , .39 ^b | - |
| User experience | 33.393 | - | .47 ^e , .45 ^c , .40 ^f | - |
| Effort | 28.520 | .50 ^f | .48 ^c , .38 ^a , .37 ^d | - |
| Content quality | 22.584 | .50 ^e | .37 ^c | .27 ^d |
| System quality | 8.198 | - | .34 ^a | - |
| Service quality | 7.950 | - | - | .28 ^e , .26 ^d |

^a Gaiagi 3D Driver vs. Health Map, ^b Gaiagi 3D Driver vs. Leeds Travel Info,

^c Gaiagi 3D Driver vs. This is Now, ^d Health Map vs. Leeds Travel Info,

^e Health Map vs. This is Now, ^f Leeds Travel Info vs. This is Now

Table 2. Taxonomy of proposed attributes with respect to their relevance in evaluating the quality of Web mashups

| Quality attributes for Web mashups | χ^2 | Effects in size (r) | | |
|------------------------------------|----------|-------------------------------------|---|-------------------------------------|
| | | Large | Medium | Small |
| <i>Mandatory</i> | | | | |
| Content added value | 51.618 | .57 ^e , .50 ^c | .49 ^f | - |
| Response time | 43.793 | - | .49 ^e , .41 ^b , .38 ^f , .37 ^e | .29 ^a |
| Ease of use | 43.387 | - | .48 ^e , .44 ^c , .37 ^f , .32 ^d | - |
| Playfulness | 42.135 | - | .46 ^b , .44 ^d , .38 ^c , .38 ^e | - |
| Effect. of Int. Visualizations | 40.958 | .52 ^c , .52 ^e | .42 ^f , .36 ^b , .34 ^d | - |
| Satisfaction | 40.500 | .54 ^f | .44 ^b , .42 ^e , .37 ^c | - |
| <i>Sufficient</i> | | | | |
| Compatibility | 37.358 | - | .40 ^a , .40 ^f , .38 ^b | - |
| Loyalty | 36.951 | - | .42 ^b , .41 ^d , .40 ^e , .35 ^c | - |
| Usefulness | 31.190 | - | .44 ^e , .37 ^c | .27 ^d |
| Effectiveness | 30.404 | - | .48 ^e , .38 ^c , .34 ^d | - |
| Content Timeliness | 28.657 | - | .46 ^c , .43 ^a , .38 ^b | - |
| Composition Added Value | 24.467 | - | .39 ^e , .38 ^d , .35 ^e , .34 ^b | - |
| Content Credibility | 24.076 | - | .41 ^c , .40 ^e , .31 ^b | - |
| <i>Desired</i> | | | | |
| Reliability | 22.941 | - | .38 ^c , .37 ^b | .28 ^a |
| Learnability | 22.106 | - | .39 ^f , .33 ^c | .28 ^d |
| Minimal Memory Load | 21.995 | - | .37 ^f , .33 ^c | .27 ^e |
| Feedback | 20.610 | - | .40 ^e , .39 ^e | .28 ^f |
| Content Completeness | 20.334 | - | .48 ^e , .37 ^c , .33 ^d , .32 ^b | - |
| Composition Suitability | 20.213 | - | .40 ^e , .34 ^c | - |
| <i>Optional</i> | | | | |
| Accessibility | 12.895 | - | .33 ^f , .31 ^a | .29 ^c , .28 ^d |
| Content Accuracy | 8.385 | - | .34 ^d | - |
| Efficiency | 8.075 | - | .34 ^e | - |
| Availability | 8.006 | - | .30 ^f | - |
| <i>Not relevant</i> | | | | |
| Understandability | 5.140 | - | - | - |

^a Gaiagi 3D Driver vs. Health Map, ^b Gaiagi 3D Driver vs. Leeds Travel Info,

^c Gaiagi 3D Driver vs. This is Now, ^d Health Map vs. Leeds Travel Info,

^e Health Map vs. This is Now, ^f Leeds Travel Info vs. This is Now

Taking into account results related to attributes, they can be classified into five groups (mandatory, sufficient, desired, optional, and not relevant) of relevance with respect to the quality of mashups, as presented in Table 2. If the Web mashup does not comply with requirements specified by mandatory attributes its quality will be significantly reduced. Sufficient attributes are also very important, but failing to meet requirements defined by them will affect the overall perceived quality to a lesser extent

than in the case of mandatory attributes. If requirements that constitute desired attributes are not satisfied, the overall perceived quality will be penalized to some extent but users will not reject the Web mashup. Optional attributes have similar role as desired attributes, but their impact on overall perceived quality is lower than the impact of desired attributes. Not relevant attributes are the ones that might be important in the case of other types of software, but not in the context of Web mashups.

4.3 Discussion

The reported findings indicate several contributions and implications for academic scholars and practitioners. First, the concept of quality that is introduced in recent international standard on quality of software [15], was reworked and adapted to the context of mashups. More specifically, the proposed model is composed of attributes which originate from theories related to the acceptance of technology [2, 24, 43, 44], success of information systems [12], models and guidelines aimed for evaluating quality [4, 6, 15, 16, 28, 35, 38, 41], user experience [18, 21, 26], and usability [1, 42] thus reflecting pragmatic and hedonic facets of quality. Next, the relevance of the dimensions considered in the proposed model was empirically identified. Considering the results of data analysis, all dimensions of the model, except the understandability, demonstrated the significant differences among mashups and can be consequently used for the quality evaluation purposes in this context. All quality attributes together with the composite measure of overall quality that have met the criteria of sensitivity have shown between small (.26 in the case of the category which deals with the assessment of service quality) and large (.57 in the case of an attribute which measures the extent to which content returned by mashups adds value to users) effects in size (as specified by Cohen [10]) which additionally confirms their suitability in assessing the quality of mashups.

Third, for all attributes and categories that have proven to fulfill the criteria of sensitivity, the practical relevance for evaluating the quality of mashups was determined. The set forth relevance is based on the amount of differences that were discovered as well as on the effect size of each difference. It should be noted that when two or more attributes had similar number of differences identified, their level of relevance in evaluating the quality of Web mashups sequence was determined by the value of their overall effect size.

Given that introduced quality model and measuring instrument add to the extant body of knowledge, academic scholars can use them as a foundation for future advances in the field. Practitioners can employ the post-use questionnaire in order to examine quality of existing mashups. In addition, the reported findings can be used by practitioners as guidelines for the development of novel mashups.

As with majority of empirical studies, the work presented in this paper has several limitations. The first one is related to the homogeneity of study participants since heterogeneous sample of users may have different attitude towards facets of quality in the context of mashups. Keeping that in mind, results of the conducted study should be interpreted carefully. The second limitation concerns the sample of mashups that were involved in the study. Although reported findings have shown significant difference among heterogeneous mashups, it would be worth to investigate if introduced framework would yield significant differences among mashups that have similar purpose. Since each

type of mashups has its specific features which may affect dimensions of perceived quality, the last limitation indicates that reported findings cannot be generalized to all types of mashups. Taking the aforementioned into account, further studies should be carried out in order to draw sound conclusions and examine the robustness of study results.

5 Conclusions

This paper represents one of the few attempts to define a quality model capturing the peculiarities of mashups. Driven by the results of some past studies, we took a better look at the characteristics of mashups and identified some relevant dimensions that reflect the quality of such applications as perceived by the end users. Differently from other contributions, we experimentally assessed the relevance of the considered quality dimensions. Taking into account the overall difference that was found among evaluated Web mashups together with the number and strength of differences in pairwise comparisons, the proposed set of attributes was classified into five different groups. We in particular found that attributes such as content added value, response time, ease of use, playfulness, effectiveness of integrated visualizations and satisfaction strongly contribute to the overall perceived quality of Web mashups. On the other hand, it appeared that attributes like accessibility, content accuracy, efficiency, and availability, which have proven to be important for assessing other breeds of software, are less important in the context of evaluating the quality of Web mashups.

We recognize that many of the quality dimensions introduced in this paper are not easy to turn into operative metrics and to be automatically assessed, yet we also recognize that quality assessment to a large degree will always be a qualitative process. Our future work will be however devoted to the development of a measuring instrument as an extension of a quality-aware composition paradigm already implemented in mashup platform [8].

References

1. Alonso-Ríos, D., Vázquez-García, A., Mosqueira-Rey, E., Moret-Bonillo, V.: Usability: a critical analysis and a taxonomy. *Int. J. Hum. Comput. Interact.* **26**(1), 53–74 (2010)
2. Bhattacharjee, A.: Understanding information systems continuance: an expectation-confirmation model. *MIS Quart.* **25**(3), 351–370 (2001)
3. Bovee, M., Srivastava, R., Mak, B.: A conceptual framework and belief-function approach to assessing overall information quality. *Int. J. Intel. Syst.* **18**(1), 51–74 (2001)
4. Calero, C., Ruiz, J., Piattini, M.: Classifying Web metrics using the Web quality model. *Online Inf. Rev.* **29**(3), 227–248 (2005)
5. Cappiello, C., Daniel, F., Koschmider, A., Matera, M., Picozzi, M.: A quality model for mashups. In: Auer, S., Díaz, O., Papadopoulos, G.A. (eds.) *ICWE 2011*. LNCS, vol. 6757, pp. 137–151. Springer, Heidelberg (2011)
6. 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)

7. Capiello, C., Daniel, F., Matera, M., Pautasso, C.: Information quality in mashups. *IEEE Internet Comput.* **14**(4), 14–22 (2010)
8. Capiello, C., Matera, M., Picozzi, M., Daniel, F., Fernandez, A.: Quality-aware mashup composition: issues, techniques and tools. In: *Proceedings of the 8th International Conference on the Quality of Information and Communications Technology*, pp. 10–19. IEEE, Lisbon (2012)
9. Cedillo, P., Fernandez, A., Insfran, E., Abrahão, S.: Quality of Web mashups: a systematic mapping study. In: Sheng, Q.Z., Kjeldskov, J. (eds.) *ICWE Workshops 2013. LNCS*, vol. 8295, pp. 66–78. Springer, Heidelberg (2013)
10. Cohen, J.: A power primer. *Psychol. Bull.* **112**(1), 155–159 (1992)
11. Daniel, F., Matera, M.: *Mashups: Concepts, Models, and Architectures. Data-Centric Systems and Applications*. Springer, Heidelberg (2014)
12. DeLone, W.H., McLean, E.R.: The DeLone and McLean model of information systems success: a ten-year update. *J. Manag. Inf. Syst.* **19**(4), 9–30 (2003)
13. Fenton, N.E., Peeger, S.L.: *Software Metrics: A Rigorous and Practical Approach*. PWS Publishing, Boston (1997)
14. Insfran, E., Cedillo, P., Fernández, A., Abrahão, S., Matera, M.: Evaluating the usability of mashups applications. In: *Proceedings of the 8th International Conference on the Quality of Information and Communications Technology*, pp. 323–326. IEEE, Lisbon (2012)
15. ISO/IEC 25010: *Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models* (2011)
16. ISO/IEC 25012: *Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) – Data quality model* (2008)
17. Ivory, M.Y., Megraw, R.: Evolution of Web site design patterns. *ACM Trans. Inf. Syst.* **23**, 463–497 (2005)
18. Hassenzahl, M., Tractinsky, N.: User experience - a research agenda. *Behav. Inf. Technol.* **25**(2), 91–97 (2006)
19. Kan, S.H.: *Metrics and Models in Software Quality Engineering*. Addison-Wesley Longman Publishing Co., Boston (2002)
20. Koschmider, A., Hoyer, V., Giessmann, A.: Quality metrics for mashups. In: *Proceedings of the Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists*, pp. 376–380. ACM, Bela-Bela (2010)
21. Law, E.L.-C., Van Schaik, P.: Modelling user experience – an agenda for research and practice. *Interact. Comput.* **22**(5), 313–322 (2010)
22. Lawshe, C.H.: A quantitative approach to content validity. *Pers. Psychol.* **28**(4), 563–575 (1975)
23. Lewis, J.R.: IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *Int. J. Hum. Comput. Inter.* **7**(1), 57–78 (1995)
24. Liao, C., Palvia, P., Chen, J.-L.: Information technology adoption behavior life cycle: toward a Technology Continuance Theory (TCT). *Int. J. Inf. Manage.* **29**(4), 309–320 (2009)
25. Mavromoustakos, S., Andreou, A.S.: WAQE: a Web Application Quality Evaluation model. *Int. J. Web Eng. Technol.* **3**, 96–120 (2007)
26. Nielsen, J., Mack, R.L.: *Usability Inspection Methods*. Wiley, New York (1994)
27. Olsina, L., Covella, G., Rossi, G.: Web quality. In: Mendes, E., Mosley, N. (eds.) *Web Engineering: Theory and Practice of Metrics and Measurement for Web Development*, pp. 109–142. Springer, Heidelberg (2005)
28. Olsina, L., Lew, P., Dieser, A., Rivera, B.: Updating quality models for evaluating new generation Web applications. *J. Web Eng.* **11**(3), 209–246 (2012)
29. Olsina, L., Rossi, G.: Measuring Web application quality with WebQEM. *IEEE Multimedia* **9**, 20–29 (2002)

30. 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)
31. Orehovački, T.: Proposal for a set of quality attributes relevant for Web 2.0 application success. In: *Proceedings of the 32nd International Conference on Information Technology Interfaces*, pp. 319–326. IEEE Press, Cavtat (2010)
32. Orehovački, T., Babić, S.: Predicting students' continuance intention related to the use of collaborative Web 2.0 applications. In: *Proceedings of the 23rd International Conference on Information Systems Development*, pp. 112–122. Faculty of Organization and Informatics, Varaždin (2014)
33. Orehovački, T., Babić, S., Jadrić, M.: Exploring the validity of an instrument to measure the perceived quality in use of Web 2.0 applications with educational potential. In: Zaphiris, P., Ioannou, A. (eds.) *LCT 2014, Part I. LNCS*, vol. 8523, pp. 192–203. Springer, Heidelberg (2014)
34. Orehovački, T., Bubaš, G., Kovačić, A.: Taxonomy of Web 2.0 applications with educational potential. In: Cheal, C., Coughlin, J., Moore, S. (eds.) *Transformation in Teaching: Social Media Strategies in Higher Education*, pp. 43–72. Informing Science Press, Santa Rosa (2012)
35. Orehovački, T., Granić, A., Kermek, D.: Evaluating the perceived and estimated quality in use of Web 2.0 applications. *J. Syst. Softw.* **86**(12), 3039–3059 (2013)
36. Orehovački, T., Granić, A., Kermek, D.: Exploring the quality in use of Web 2.0 applications: the case of mind mapping services. In: Harth, A., Koch, N. (eds.) *ICWE 2011. LNCS*, vol. 7059, pp. 266–277. Springer, Heidelberg (2012)
37. Orehovački, T., Granollers, T.: Subjective and objective assessment of mashup tools. In: Marcus, A. (ed.) *DUXU 2014, Part I. LNCS*, vol. 8517, pp. 340–351. Springer, Heidelberg (2014)
38. Pang, M., Suh, W., Hong, J., Kim, J., Lee, H.: A new Web site quality assessment model for the Web 2.0 Era. In: Murugesan, S. (ed.) *Handbook of Research on Web 2.0, 3.0, and X.0: Technologies, Business, and Social Applications*, pp. 387–410. IGI Global, Hershey (2010)
39. Rio, A., Brito e Abreu, F.: Websites quality: Does it depend on the application domain? In: *Proceedings of the 7th International Conference on the Quality of Information and Communications Technology*, pp. 493–498. IEEE, Porto (2010)
40. Redman, T.C.: *Data Quality for the Information Age*. Artech House, Norwood (1996)
41. Sassano, R., Olsina, L., Mich, L.: Modeling content quality for the Web 2.0 and follow-on applications. In: Murugesan, S. (ed.) *Handbook of Research on Web 2.0, 3.0, and X.0: Technologies, Business, and Social Applications*, pp. 371–386. IGI Global, Hershey (2010)
42. Seffah, A., Donyae, M., Kline, R.B., Padda, H.K.: Usability measurement and metrics: a consolidated model. *Software Qual. J.* **14**(2), 159–178 (2006)
43. Venkatesh, V., Bala, H.: Technology acceptance model 3 and a research agenda on interventions. *Decis. Sci.* **39**(2), 273–315 (2008)
44. Venkatesh, V., Thong, J.Y.L., Xu, X.: Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. *MIS Q.* **36**(1), 157–178 (2012)
45. Webb, H.W., Webb, L.A.: SiteQual: an integrated measure of Web site quality. *J. Enterp. Inf. Manage.* **17**(6), 430–440 (2004)
46. World Wide Web Consortium: *Web Content Accessibility Guidelines (WCAG) 2.0* (2008). <http://www.w3.org/TR/WCAG20/>