Challenge Outcome and Conclusion

Martin Gaedke[1] and Florian Daniel[2]

[1] Technische Universität Chemnitz, Str. der Nationen 62, 09111 Chemnitz, Germany
martin.gaedke@informatik.tu-chemnitz.de
[2] Politecnico di Milano, Via Ponzio 34/5, 20133 Milano, Italy
florian.daniel@polimi.it

Abstract. In this final chapter, we report on the outcome of the ICWE 2016 Rapid Mashup Challenge (RMC), describe the voting system used, and draw some conclusions regarding the presented works.

Keywords: Mashups, Challenge, Benchmarking

1 Challenge Organization

We recall that every tool participating in the challenge was allocated 20 minutes that the authors could split freely into a presentation of the demonstrated approach and the demonstration itself. The goal of the presentation was to introduce the approach and/or tool, to illustrate its design and to enumerate its most important features, so that the audience would be prepared for the actual demonstration. The live demonstration, in fact, aimed at showcasing the on-the-fly development of a mashup chosen by the authors in front of the audience. The starting point for all demonstrations was an empty workspace or a code editor in which the components/resources to be reused in the mashup had been pre-registered and pre-defined, but not yet assembled. Some authors chose to follow an iterative process, whereby the mashup was grown incrementally, piece by piece. Others also included a more general overview of the mashup tool capabilities, which was useful to demonstrate the expressive power of the tool, but did not necessarily help them build the most impressive mashup during the allocated time frame.

Each time a mashup was completed and the time for the demonstration had expired, the audience had the opportunity to ask questions to the authors. This interactive session was very useful to provide the mashup authors with valuable feedback and the audience with clarifications regarding what it had just seen during the demonstration. At the end of the interactive session, the audience could provide its assessment of the approach via simple ratings collected through the ASQ system. The results were aggregated on the fly and the overall ranking updated and shown to the audience and the tool authors.

2 The ASQ Voting System

The challenge evaluation phase was supported by the ASQ system [1]. ASQ (a permutation over Slides-Questions-Answers) allows anyone with a Web browser

to follow a slideshow presentation and interact with the content by answering questions embedded in the slides. It was originally developed at the USI Faculty of Informatics to support in-classroom teaching activities by taking advantage of the fact that every student comes with his/her laptop to follow the lectures. Students not only can better read the content broadcast to their devices, but teachers can get real-time feedback about their level of understanding and thus adapt their pace and explanation depth during the lecture.

As such ASQ is a general tool and can be used also for any interactive presentation. In particular for the 2016 edition of the RMC, we reused the ASQ system already extended for the prior edition of the challenge. The extensions required were: (i) a special question type to gather ratings, over a 5-star scale, with the possibility to award also half stars; and (ii) a count-down timer activated at the beginning of each presentation/demo session to ensure every participant demonstrates his/her tool during the same amount of time.

The use of ASQ during the RMC enabled the easy involvement of the whole audience in the assessment of the presented approaches (including the authors themselves, who did however not vote in their own turn). A secondary benefit was that ASQ allows one to automate and increase the efficiency of the scoring process, where the answers are aggregated and the final ranking is recomputed after every participant is evaluated. The slides showing the metadata about the current participants (name of approach/tool, authors, abstract) were interleaved with the questions to evaluate them. This helped focusing the audience's attention and build a shared awareness of the proceedings of the challenge and to manage the time without introducing unnecessary delays.

3 Evaluation Criteria

In line with the call for participation of the RMC, every demonstration was evaluated according to five different criteria:

1. *Expressive power.* Each approach presented during the Challenge was accompanied by a filled feature checklist (introduced in the first chapter of this volume), which provided insight into the respective expressive power as declared by the authors. This year, we used the feature checklists only to select candidates before the Challenge, while during the Challenge we asked the audience to provide a subjective impression of the expressive power perceived. The more composition features an approach supports, the higher its expressive power.
2. *Flexibility.* This criterion aimed to assess the extensibility and adaptability of the approaches to different, possibly novel requirements. Although by now it is almost a decade that researchers have been working on mashups, every day new requirements and/or technologies pop up, and mashup tools are perhaps the instruments that are exposed most to this evolution. Flexibility is thus paramount. The lower the effort needed to extend/adapt an approach to novel requirements, the higher the flexibility.

3. *Maturity.* On the other hand, given a development instrument, it is important to understand its level of maturity, that is, the stability and readiness for production of the instrument. This is perhaps the criterion that varied most among the presented approaches in the 2016 edition of the RMC, and it was important to capture the difference of maturity among the approaches in order to enable the reader to properly interpret the presented results. The closer an instrument is to a production-ready system, the higher its maturity.
4. *Intuitiveness.* This criterion explored the end-user perspective of mashup development with the presented approaches and tried to quantify how the audience perceived the respective ease of use in practical settings. Of course, the more graphical and interactive an approach, the fewer skills are needed to operate it, and the higher its intuitiveness.
5. *Demo effectiveness.* The last criterion aimed to assess the effectiveness of the showcased mashup scenario and demonstration in convincing the audience of the power of the presented approach/tool. Partly, this criterion therefore also includes the "performance" of the presenter. So, the more the audience understood the demonstration, the higher the demo effectiveness.

These criteria are different from those used in the 2015 edition of the Challenge, as the overall setting of the comparison was different (less restricted demonstrations and, hence, harder to compare) and the presented approaches were more heterogeneous among each other (again, harder to compare).

4 Results

Going straight to the point, Table 1 summarizes the feedback obtained from the audience by each of the tools participating in the challenge and orders them in descending order based on the average vote achieved. We are thus glad to proclaim the winner of the 2016 edition of the Rapid Mashup Challenge: *Smart-Composition* by Michael Krug, Fabian Wiedemann, Markus Ast and Martin Gaedke. Congratulations!

Of course, the ranking does not only communicate the winner of the Challenge but also some relative positioning of all the presented approached. Before going to fast interpretations or comparisons with the ranking of last year, we would however like to point out again what we already explained earlier in this volume: this year we accepted proposals of very different maturity (see the respective column in the table), in order to provide an as rich as possible picture of the ongoing activities of the community. But attention, the maturity of a presented approach significantly impacts also the other criteria of the evaluation, as an approach or tool that is not yet at the level of development the authors envision themselves, of course, cannot be expected to score high in those criteria that are still under development. In some cases, the presented tool was even still at the level of a proof of concept prototype; this is, for instance, the case of CAMUS and Toolet, two approaches that were still in a very early stage of development but that we nevertheless felt deserved some space in the Challenge

Table 1: Final ranking of the 2015 RMC based on the feedback gathered from the audience during the challenge (the smallest vote possible was 0 stars, the highest 5 stars)

Rank	Tool	Expressive power	Flexibility	Maturity	Intuitiveness	Demo effectiveness	Average
1	SmartComposition	3.73	3.77	3.55	3.45	3.68	3.64
2	FlexMash 2.0	3.42	3.15	3.77	3.42	3.35	3.42
3	EFESTO	2.50	2.44	3.00	3.38	2.50	2.76
4	Search-based Mashups	2.67	2.56	3.72	2.33	2.44	2.74
5	Uduvudu Editor	2.88	2.71	2.33	2.42	2.75	2.62
6	Linked Widgets	2.32	2.46	2.82	1.93	2.18	2.34
7	CAMUS	1.85	1.60	1.95	1.60	1.90	1.78
8	Toolet	1.50	1.61	1.17	1.50	1.33	1.42

in order to allow the authors to explain their ideas and to obtain constructive feedback from the audience. Therefore, the ranking should be read as a comparison of screenshots of approaches or instruments taken at a given instant of time during their development, some already in a rather mature phase, others still in the conception phase.

This being said, it is interesting to note that the winner of the Challenge is not the approach that scored best in terms of maturity. Instead that audience particularly appreciated its flexibility and the effectiveness of the demo, next to the expressive power of the approach. The approach that was assessed as the most mature, instead, was FlexMash 2.0, where the "2.0" already hints at the evolution the tool has undergone over the last years. The hard work by the authors has payed off. Also search-based mashups presented by Eduard Daoud (not included in this volume for time restrictions) were considered very mature, in line with the nature of this industrial contribution to the Challenge.

Looking at all votes together, it seems that the presented approaches can be grouped into three clusters: proof-of-concept prototypes (ranks 7-8), advanced prototypes (ranks 3-6) and production-ready instruments (ranks 1-2). Search-based mashups are an exception, which is an instrument that is actually used in production in industrial contexts; while this is acknowledged by the audience with a high maturity score, the other criteria lower the position of the approach in the overall ranking.

Last year, the distribution of the votes was rather narrow. This year, we notice a significant difference between the votes of the different approaches. This is fully in line with the observation that last year the compared tools were very similar in terms of their maturity, while this year there is much more variety in the maturity.

5 Limitations

As with all rankings based on subjective feedback, the key issue is participation. And the ranking provided in Table 1 is no exception. In average, the audience during the Challenge consisted of approximately 20-25 people, most of which also participated in the voting process. More specifically, all votes reported in the ranking are based on individual feedbacks collected from 8-14 participants. The authors participating in the Challenge were asked not to participate in the voting process for their own presentation; no issue regarding this rule was reported, and we trust in the correct, ethical conduct of all participants.

As for the comparison of mashup approaches themselves, our comment of last year is still valid: Given the wide variety of approaches to mashup tool design, both from research and industry, and the lack of standard or commonly accepted benchmarks to assess development tools, it remains difficult to give a fair comparison of mashup development approaches/tools. But we do not only re-confirm this statement and even must add that this year the comparison was even harder, given the different levels of maturity of the presented works.

During the challenge, tools were demonstrated by their own authors, something that may invalidate any claim of usability or accessibility, especially by end-user programmers, usually associated with mashup tools. However, since every tool was used by the corresponding authors, the fairness of the comparison is not affected. In the future, it could be an interesting option to allow also the audience to try the proposed instruments, at least in very simple design scenarios. This would allow the audience to obtain a better feeling especially of the intuitiveness criterion, but also of the maturity criterion.

6 Conclusion

Summing up, we consider this second edition of the Rapid Mashup Challenge a success, similar to the one of the first edition. The quality of both the presented works and the presenters was high, and the topics brought forward by the presenters as well as by the audience were stimulating. Compared to the first edition, which was characterized by a set of mashup approaches at a comparable level of maturity, this year the approaches selected for presentation span all phases of the development life cycle, from early prototypes to production-ready tools. While this on the one hand hindered to some extent the comparison of the approaches, on the other hand it however conveys a very positive message: research on mashups and Web-based composition technologies is a topic of continuing interest and strategic value to both industry and academia. Although mashups, that is, the integration of all kinds of Web resources, have percolated into common software engineering practice, there is a continuous need for cutting-edge research that investigates that potential, opportunities and pitfalls of new technologies as they emerge and that aims to conceptualize and abstract the respective underlying principles to facilitate their use.

The approaches presented in the context of the RMC do exactly this, some of which even with the goal of enabling non-programmers to take part in the

development of composition-based applications. This volume represents tangible evidence of this effort and of the need to invest even further efforts into the directions outlined throughout this volume – directions we hope will be explored in the future editions of the Challenge.

After this second edition of the Challenge, we see that the challenge for the future of the Challenge is understanding how to reliably compare approaches that are as diverse as the ones that characterize the domain of assisted mashup development in an environment that is constantly changing and evolving. While in the first edition this was less evident, this year the problem emerged prominently. One key ingredient toward a more objective benchmarking of assistive mashup development techniques seems to be, as identified this year, a clear differentiation of the maturity levels of the competing approaches. Being the RMC a challenge that aims to provide a final ranking of approaches, also this year we proposed a possible (subjective) ranking; yet, as we pointed out, it is important to acknowledge that inside this ranking not all approaches are actually comparable. How to enable a fair comparison, if possible at all, is a question we leave to the Challenges to come.

Acknowledgment We would like to thank all participants for their enthusiasm and the audience for their active help with the evaluation of the presented approaches. We would also like to thank Vasileios Triglianos for his help and support with the ASQ tool used to organize the voting process.

References

1. Triglianos, V., Pautasso, C.: Interactive scalable lectures with ASQ. In: Proc. of the 14th International Conference on Web Engineering (ICWE 2014), Toulouse, France, Springer (July 2014) 515–518