Everything for the Users, Nothing by the Users: Lessons Learnt from a Heterogeneous Data Mapping Languages User Study

Herminio García-González

IT and Communications Service, University of Oviedo, Asturias, Spain
garciaherminio@uniovi.es

Abstract. Usability has been seen as a new requirement in the Semantic Web community. Heterogeneous data mapping languages enable users to create knowledge graphs from legacy datasets. Even though some of these tools claim to be user friendly, this is not empirically demonstrated. In this paper, we revisit our previous usability experiment with these languages and from its outcomes we envisage next actions and problems that should be tackled in the topic. Covering them should lead to a better adoption among users.

Keywords: Data mapping languages · Data integration · Semantic Web · Usability

1 Introduction

Recent achievements in the data mapping topic enable users to define heterogeneous data sources integration in a declarative fashion instead of using *ad-hoc* solutions which redounds in a higher flexible and faster process [10]. This fact allows users to invest less time and resources while constructing a knowledge graph. Thus, the final goal is to ease users' workflows. In addition, some of the proposed languages claim to be user friendly (i.e., YARRRML [8] and ShExML [5]) or easy to learn by semantic web experts (i.e., SPARQL-Generate [9]). However, this quality should be quantified in order to establish proper comparisons.

Recent trends in the semantic web community have seen the necessity to understand users and put them in the center of our solutions, improving their productivity and taking care of their needs[1]. Moreover, this has also been highlighted in the Knowledge Graph Construction W3C Community Group[2]. Users

[1] http://www.juansequeda.com/blog/2020/11/16/international-semantic-web-conference-iswc-2020-trip-report/

[2] https://w3id.org/kg-construct/tpac/#report

or usability studies allow to understand users' problems as well as their difficulties, needs and perceptions. They are, therefore, a huge analysis tool when deciding future actions on the topic.

However, to the best of our knowledge, only our recent study [5] has tackled the topic of usability in heterogeneous data mapping languages. In this paper, we briefly summarise our previous experiment, explaining the followed methodology and its outcomes. Then, from these results we build our argumentation on actions that should be taken in the community to better understand and address users' problems.

2 Brief experiment description

Currently, there are various languages that allow to integrate heterogeneous data sources into a single Knowledge Graph. First comers based their syntax in the W3C Recommendation R2RML[3], like RML [3] and xR2RML [11], offering an inherited RDF-based syntax. Conversely, more recent approaches based their syntax in other languages like SPARQL-Generate (based on SPARQL), YARRRML (based on YAML) or ShExML (based on ShEx); looking for a better user friendliness.

While having an RDF-based syntax could seem a good thing for machine readability, inference and processability, it is also true that it makes the final language verbose (even when using the Turtle syntax). In this regard, we could consider these languages more like middle representation languages rather than languages intended to be used directly by users. Following this idea, including them into a user study would not be fair, as users would need more time and keystrokes to achieve the same solution. Therefore, it would be a bias from the beginning, as these users would not be in the same conditions, in comparison with other users assigned with much less verbose languages. Thus, we selected those languages that claim to be user friendly and that offer a similar syntax in terms of verbosity.

The experiment was designed as a mixed-method approach, that is to say, involving a quantitative and a qualitative design and analysis [2]. On one hand, the quantitive part (involving capture of behavioural and performance metrics) allows for an objective and direct evaluation of the users interaction with the tools and their tasks achievements. On the other hand, the qualitative part (involving a questionnaire for subjective variables) enable to gather the users' perceptions. Using a mixed-method approach we can correlate both sets of measures to have a better understanding on how the users interact with and perceive the tool.

The sample consisted of 20 students pursuing a MSc in Web Engineering (first course out of two). The experiment was hosted the final day of the semantic web subject in which the students were introduced to semantic technologies (RDF, SPARQL, Shape Expressions, among others). Therefore, we can categorise the sample as first-time users with some background knowledge. So, our results and conclusions will be in line with this described profile.

[3] https://www.w3.org/TR/r2rml/

The sample was randomly distributed in three groups, one per language, so previous knowledge background bias could be mitigated. The experiment was divided in two tasks. The first one consisted in creating a set of mapping rules from a given input and the expected output. The second one was to perform a small modification to the mapping rules created in the former task. Therefore, first task measured global usability whereas the second one measured modifiability of the mapping rules developed with the assigned language.

3 Results & Highlights

In the first task, 17 students (out of 20) submitted results (7 for ShExML, 4 for YARRRML, and 4 for SPARQL-Generate) and in the second task only 7 students did so (6 for ShExML and 1 for YARRRML). These total results reveal that the SPARQL-Generate users had problems when reaching a working mapping and that the YARRRML users found difficult to modify an existing set of mapping rules and/or that they invested too much time in the first task. We performed a statistical analysis (cf. [5] to see the full statistical report in detail) per variable for the three languages as well as a pair-wise comparison to see in which variables and among which languages there were differences.

Task 1: In quantitative analysis significant differences were found in elapsed seconds (particularly between ShExML and YARRRML), completeness percentage (between ShExML and SPARQL-Generate) and precision (between ShExML and SPARQL-Generate). This comes to corroborate that the YARRRML users invested much more time than the ShExML users when finding working solutions. The difference in completeness percentage and precision between ShExML and SPARQL-Generate reveals that the SPARQL-Generate users were not able to find working solutions.

It is also worth to mention that there were no significant differences in the rest of quantitative variables which come to support that there are no differences in languages verbosity (i.e., no significant differences in keystrokes) and in the offered web playground, which was used by the users to develop the mapping rules (i.e., left button clicks, right button clicks, mouse wheel scroll and meters travelled by the mouse). This is of quite importance as, like we mentioned in Section 2, a difference in languages verbosity would imply much more time invested by users to reach a working solution. In the same way, a significant difference in user interfaces could lead to the similar biases when solely analysing the languages usability.

In qualitative analysis significant differences were found in general satisfaction (between ShExML and YARRRML), learnability (between ShExML and other both languages), mapping definition easiness (between ShExML and both other languages) and easiness of use (between ShExML and YARRRML). These qualitative results come to corroborate and complement the quantitative ones, so difficulties in finding working solutions by the SPARQL-Generate users are translated to a worse learnability and mapping definitions easiness impression. More time consumed to find working solutions by the YARRRML users is trans-

lated to a worse impression in the four variables causing a descent in general usability indicators (i.e., general satisfaction level and easiness of use).

Task 2: In task 2 no significant differences could be established due to the very low sample sizes. Only one YARRRML user was able to submit a non-working solution whereas 6 ShExML users submitted a solution for this task. This could be caused by the extra time needed by the YARRRML users wrt the ShExML users. The ShExML modifiability variable was rated with 5 points by 83% of the ShExML users and with 3 points by the YARRRML single user. The SPARQL-Generate users were unable to reach this task as they had problems finishing the first one.

These differences reveal that the design of SPARQL-Generate is having a bad effect on first-time users which found it hard to operate and learn. However, it would be interesting to discern which parts of the languages are causing the differences between ShExML and YARRRML. As an hypothesis, we can explain them due to their different syntaxes because ShExML uses keywords which can make the language more self-explanatory and offers modularity in its iterators, which reminds the, well-known by developers, object-oriented paradigm.

Bad results in some qualitative variables for the three languages reveal another interesting picture. They perceive that the languages design lead to commit some errors (error proneness), that the error reporting system was not useful to solve their errors (error reporting system) and that they do not see much applicability to these tools (applicability). As we analyse further in the following section, these three aspects should be handled urgently by the community.

4 Actions to take

In the light of the previously commented results it is important how new features are added and designed so they do not have a bad impact on usability and learnability. In the semantic web community we care a lot about new features and technical improvements but we tend to care less about users, we should involve them more and develop more user-centric approaches.

Following the previous argument, the users from our experiment told us that the three languages lead them to commit some errors (the languages are not designed taking users mental models into account), that the error reporting system was not useful (it is another point that reveals that users are not involved in the development process) and that they do not see much applicability in these tools. This last perception reveals something that could be extended to the whole semantic web community. With these languages and tools we are producing knowledge graphs, so in the end they do not see applicability to neither of them. In addition, applicability (and related variables like learnability) on first-time users reveal a derivate and correlated one: adoption. If first-time users do not see much applicability, and technologies are hard to learn, they are not going to adopt them. Therefore, these points are urgent ones that should be addressed by the community.

From a methodological point of view, we have to adopt stronger methods [12] which support our hypothesis and claims. Focusing in this community, Heyvaert et al. [7] compared user performance and perception from expert and non-experts users while using RMLEditor. In addition, they established a comparison with RML on users which had previous knowledge of it. However, the comparisons are merely established using percentages. Lefrançois et al. [9] carried out a performance evaluation between the RML and SPARQL-Generate main implementations. Again, this comparison is established based only on mean times. These weak comparisons could drive to some erroneous conclusions as they lack the power of a statistical test [4]. In plain words, we want to corroborate that our findings were not obtained just by chance. Besides, from a mathematical point of view, we want to take into account the variability of our sample (i.e., the variance) so we are not assuming bad conclusions for the negligence of not analysing it. Moreover, we are losing the evidence strength measure of these conclusions (see effect size [1]). Thus, we have to learn from experimental[4] and social sciences on how observational experiments are performed, analysed and reported[5].

As we mentioned, our study only covered first-time users with some background knowledge so it is necessary to run these kinds of experiments with other profiles to have a whole perspective on the topic [12]. Doing so, we will be able to cover requirements from all users types and, thus, increase the overall adoption. In addition, a comparison between visual and non-visual approaches along different users profiles should be carried out to discern users preferences and the target type of user to which each tool should be mainly addressed.

Finally, in order to contrast our hypothesis about differences between ShExML and YARRRML it would involve running more complex experiments which could come closer to the users' mental model processes, so we can understand which language constructions and syntax are better. One possibility is to use cognitive models and frameworks [6] which could deliver explanations to our prior empirical study.

5 Conclusions

The Semantic Web community recent trend to focus on users, understand them, and take care of their needs has opened another perspective in the semantic technological stack which advocates to put the user in the center of our thoughts. Although some heterogeneous data mapping languages claimed its user friendliness, it was not empirically supported. Thus, in this paper we have briefly summarised our previous usability experiment in the field of heterogeneous data mapping languages where ShExML demonstrated a better usability on first-time users.

We have also claimed actions that should be addressed to solve the elucidated common problems, further experiments that cover more users' profiles,

[4] https://slideshare.net/miriamfs/vision-track-october2020fernandezv5

[5] https://slideshare.net/tammavalentina/the-tao-of-knowledge-the-journey-vs-the-goal

more complex experiments that could explain users' mental model processes and the use of stronger methodological instruments and metrics. Dealing with the exposed points, we envisage a promising future for the Knowledge Graph Construction community, decreasing their technologies complexity barriers, having more users being attracted, and in short, improving their adoption.

References

1. Coe, R.: It's the effect size, stupid: What effect size is and why it is important (2002)
2. Creswell, J.W.: Mixed-method research: Introduction and application. In: Handbook of educational policy, pp. 455–472. Elsevier (1999)
3. Dimou, A., Sande, M.V., Colpaert, P., Verborgh, R., Mannens, E., de Walle, R.V.: RML: A Generic Language for Integrated RDF Mappings of Heterogeneous Data. In: Proceedings of the Workshop on Linked Data on the Web co-located with WWW 2014, Seoul, Korea, April 8, 2014. (2014)
4. Gallo, A.: A refresher on statistical significance. Harvard Business Review **16** (2016)
5. García-González, H., Boneva, I., Staworko, S., Labra-Gayo, J.E., Lovelle, J.M.C.: ShExML: improving the usability of heterogeneous data mapping languages for first-time users. PeerJ Computer Science **6**, e318 (2020)
6. Hansen, M.E., Lumsdaine, A., Goldstone, R.L.: Cognitive architectures: a way forward for the psychology of programming. In: ACM Symposium on New Ideas in Programming and Reflections on Software, Onward! 2012, part of SPLASH '12, Tucson, AZ, USA, October 21-26, 2012. pp. 27–38. ACM (2012)
7. Heyvaert, P., Dimou, A., Herregodts, A., Verborgh, R., Schuurman, D., Mannens, E., de Walle, R.V.: RMLEditor: A Graph-Based Mapping Editor for Linked Data Mappings. In: The Semantic Web. Latest Advances and New Domains - ESWC 2016, Heraklion, Crete, Greece, May 29 - June 2, 2016, Proceedings. Lecture Notes in Computer Science, vol. 9678, pp. 709–723. Springer (2016)
8. Heyvaert, P., Meester, B.D., Dimou, A., Verborgh, R.: Declarative Rules for Linked Data Generation at Your Fingertips! In: ESWC 2018 Satellite Events, Heraklion, Crete, Greece, June 3-7, 2018, Revised Selected Papers. pp. 213–217 (2018)
9. Lefrançois, M., Zimmermann, A., Bakerally, N.: A SPARQL Extension for Generating RDF from Heterogeneous Formats. In: The Semantic Web - 14th International Conference, ESWC 2017, Portorož, Slovenia, May 28 - June 1, 2017, Proceedings, Part I. Lecture Notes in Computer Science, vol. 10249, pp. 35–50 (2017)
10. Meester, B.D., Heyvaert, P., Verborgh, R., Dimou, A.: Mapping Languages: Analysis of Comparative Characteristics. In: Joint Proceedings of the 1st International Workshop on Knowledge Graph Building and 1st International Workshop on Large Scale RDF Analytics co-located with ESWC 2019, Portorož, Slovenia, June 3, 2019. CEUR Workshop Proceedings, vol. 2489, pp. 37–45. CEUR-WS.org (2019)
11. Michel, F., Djimenou, L., Faron-Zucker, C., Montagnat, J.: Translation of Relational and Non-relational Databases into RDF with xR2RML. In: WEBIST 2015 - Proceedings of the 11th International Conference on Web Information Systems and Technologies, Lisbon, Portugal, 20-22 May, 2015. pp. 443–454. SciTePress (2015)
12. Pesquita, C., Ivanova, V., Lohmann, S., Lambrix, P.: A Framework to Conduct and Report on Empirical User Studies in Semantic Web Contexts. In: EKAW 2018, Nancy, France, November 12-16, 2018, Proceedings. Lecture Notes in Computer Science, vol. 11313, pp. 567–583. Springer (2018)