

From Spreadsheet to Linked Open Data via Ontology

Parita Amin^[0000–0003–3475–7223] and Daryl H. Hepting^[0000–0002–3138–3521]

Department of Computer Science, University of Regina, Regina, SK S4S 0A2, Canada
`{pda324,heptingd}@uregina.ca`

Abstract. The Semantic Web, over the years, has developed significantly in terms of study and research but still has numerous unanswered questions. Linked Open Vocabulary (LOV) for feedback survey is one of those topics which lack enough research content. This research focuses on development of Linked Open Data (LOD) using Protégé¹ with an aim to convert data that is stored in tabular format (e.g. Spreadsheet, Comma Separated Values (CSV)) into LOD.

Keywords: Linked Open Data · Survey · Protégé · Likert · Open-ended.

1 Motivation

Since 2012, Hepting has been collecting feedback² from students in the classes he teaches. Most of the data collected has been on paper, but some recent course offerings have collected this feedback data using moodle’s X feature. The feedback instrument has 9 Likert items using a 5 point Likert scale (Never, Rarely, Sometimes, Often, Always) and 3 open-ended questions. More detail on the feedback instrument can be found on his website². A goal of making the dataset available as linked open data is to enable more robust analysis along with embracing open data principles [7]. Advice about publishing LOD can be found in works by Borgesius, Gray, and van Eechoud [2] and W3C [8].

Presented here is the process to take this student feedback data from spreadsheet to fully linked open data, with the possibility of generalizing this effort to any sort of data that a citizen would like to publish.

The steps begin with searching for a Linked Open Vocabulary [3] that meets the citizen’s needs. If one is found, then that LOV can be used to map data from the spreadsheet into RDF. If one is not found, an ontology/vocabulary needs to be developed to allow the mapping to occur.

There are LOV’s for surveys (see Scandolari et al. [6] and Fox and Katsumi [4]) but they don’t deal with Likert items.

The vocabulary need not be perfect the first time. The goal is to model the feedback instrument (responses). What is the most concise and most straightforward way to do this modelling? In SWftWO [1], the authors propose this test...

¹ https://protegewiki.stanford.edu/wiki/Main_Page

² <http://www2.cs.uregina.ca/~hepting/teaching/evaluation.html>

2 Development of Linked Open Dataset for Feedback Survey Using Protégé

There exist many ontology editors for the development of LOVs, of which Protégé is one of the most used. “Protégé is a free, open-source platform that provides a growing user community with a suite of tools to construct domain models and knowledge-based applications with ontologies” [5]. Here, Protégé has been used to describe new vocabulary for student feedback survey. The vocabulary for feedback survey can be described in different ways. The vocabulary to describe the survey instrument (the form) would be different from the vocabulary used to describe the responses. This paper focuses on describing the vocabulary for the survey instrument which may consist of the “Type of Survey”, “Survey Details”, and “Survey Questions”. These concepts form the basis for the survey instrument.

2.1 Classes and Class Hierarchy

The root node of the class hierarchy in Protégé is “owl:Thing”. “Survey”, “SurveyQuestions”, “SurveyAnswers”, “Respondent”, and “ResponseDate” are the subclasses of “owl:Thing”. The class hierarchy can be further built using Tools > Create class hierarchy... The “Survey” class focuses on the type of the survey. The “Respondent” class signifies the respondents of the survey and the “ResponseDate” is for the date on which the respondent responded to their survey. The “SurveyQuestions” class represents the question section whereas the “SurveyAnswers” class represents the answer section of the survey. “FeedbackSurvey” is the only subclass of the “Survey” class whereas the “SurveyQuestions” and “SurveyAnswers” have two subclasses each, corresponding to likert and open-ended type question-answers. The “SubClass Of” field of Class Description view allows describing the parent class(es)

2.2 Object Properties and Object Property Hierarchy

The next step in ontology development in Protégé is describing the object properties. Object properties are used to describe the relationship between two entities. For instance, “Survey” and “FeedbackSurvey” classes are the entities that can be connected to each other using an object property (relationship). “Survey” is connected to “FeedbackSurvey” by “hasType” object property and “FeedbackSurvey” is connected to “Survey” by “isTypeOf”. In such cases, “hasType” is said to be an inverse of “isTypeOf”. Also, object properties have domains and ranges. Here, “FeedbackSurvey” is the range and “Survey” is the domain for “isTypeOf”. Similarly, all the object properties in survey ontology can have inverse, domain and range

2.3 Data Properties

Data Properties are used to restrict data entry for particular fields of the tabular data. For survey ontology, data restriction properties restrict the datatype

of values of answers for all 9 likert scale questions using Description view of data property hierarchy. The domain and range for these data properties are “Respondent” and “xsd:positiveInteger” respectively.

2.4 Individuals

Individuals are the instances of the classes described for survey ontology. For instance, “Respondent” class has instances, namely, “Respondent1”, “Respondent2”, “Respondent3”, ...

3 Conclusion

The research has potential for further development where the LOD (with 5-star rating³) can be created as a discoverable resource on the web which would enable the citizens to generalize, access and publish their data on the web.

References

1. Allemang, D., Hendler, J., Gandon, F.: Semantic Web for the Working Ontologist: Effective Modeling for Linked Data, RDFS, and OWL. Association for Computing Machinery, New York, NY, USA, 3 edn. (2020)
2. Borgesius, F.Z., Gray, J., van Eechoud, M.: Open Data, Privacy, and Fair Information Principles: Towards a Balancing Framework. *Berkeley Technology Law Journal* **30**(3), 2073–2131 (2015)
3. Dumontier, M., Vandenbussche, P.Y., Ateazing, G.A., Poveda-Villalón, M., Vatan, B.: Linked Open Vocabularies (LOV): A Gateway to Reusable Semantic Vocabularies on the Web. *Semantic Web* **8**(3), 437–452 (Jan 2017). <https://doi.org/10.3233/SW-160213>
4. Fox, M.S., Katsumi, M.: An Ontology for Surveys. In: Macer, T., et al. (eds.) *Proceedings of the Association for Survey Computing*. vol. 7, pp. 1–24. Association for Survey Computing (2016)
5. Musen, M.A.: The Protégé Project: A Look Back and a Look Forward. *AI Matters* **1**(4), 4–12 (Jun 2015). <https://doi.org/10.1145/2757001.2757003>
6. Scandolari, D., Re Calegari, G., Celino, I., Scrocca, M.: The Survey Ontology. Revision: V 1.1 (2021), <https://cefriel.github.io/survey-ontology/docs/index.html>
7. Tauberer, J.: *Open Government Data Definition: The Book*. Sebastopol, California, 2 edn. (2014), <https://opengovdata.io/2014/8-principles/>
8. World Wide Web Consortium: Best Practices for Publishing Linked Data (Jan 2014), <http://www.w3.org/TR/ld-bp/>

³ https://www.w3.org/2011/gld/wiki/5_Star_Linked_Data