

A System to Assist Semantic Web Data Editing Through the Use of Case-Based Reasoning

Nicolas Lasolle^{1,2}[0000-0002-1253-649X], Olivier Bruneau¹, Jean Lieber²,
Emmanuel Nauer², and Siyana Pavlova²

¹ Université de Lorraine, CNRS, Université de Strasbourg, AHP-PreST, F-54000
Nancy, France

² Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France
nicolas.lasolle@univ-lorraine.fr

Abstract. RDF (*Resource Description Framework*) is the Semantic Web core model which is used to represent data thanks to the definition of triples. RDF data editing is often a manual task which may seem tedious for the concerned users, and which includes an important risk of errors. A Web editing tool along with a suggestion system have been created to assist the editing of the Henri Poincaré correspondence corpus. The tool uses RDFS (*RDF Schema*) entailment and case-based reasoning to provide relevant suggestions when editing a resource. Four versions are proposed and are compared to each other through a double evaluation. While this system has been created for and evaluated with the Henri Poincaré correspondence corpus, the proposed methods could be applied to other corpora.

Keywords: Digital humanities · Semantic Web data editing · Case-based reasoning · Historical corpora · SPARQL query transformation

1 Introduction

Manual RDF data editing is the task which consist in creating triples through a form-based interface or directly by writing them to a file using an adapted RDF serialization format (e.g. RDF/XML, Turtle). The Henri Poincaré correspondence corpus is centered around 2100 sent and received letters which constitute scientific, administrative and private exchanges. The RDF model has been chosen to represent corpus data, the RDFS language to represent ontology knowledge and the SPARQL language to query data. The manual editing of corpus data has been identified as a tedious task which requires the constant attention of the user to avoid different kinds of mistakes. A Web editing tool along with a suggestion system have been developed to assist users in this editing process. This system has already been presented in the context of the *International Semantic Web Conference* [2].

Copyright © 2021 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2 Presentation of the system

The goal of the suggestion system is to provide a ranked list of potential values associated to a field (i.e. subject, predicate or object) when editing a triple. Four versions of the suggestion system have been designed. The *basic version* orders the suggestions using the alphabetical order. The *deductive version* benefits from the use of the ontology knowledge when ranking the potential values. In particular, it takes advantage of the knowledge about the **domain** and **range** of the ontology properties. The *case-based version* uses case-based reasoning to exploit knowledge about similar edited resources. As an example, when one is editing a letter and wants to set its recipient, it may be relevant to search for letters having the same topics discussed, quoting the same individuals or institutions, or written within a relatively close temporal interval. A system based on SPARQL query transformation rules is reused to find resources similar to the resource which is currently being edited. This system, named SQTRL (SPARQL Query Transformation Rule Language), has proven useful in various contexts [1]. The *combination version* combines the methods used in the two previous versions.

Associated with the suggestion system, a Web user interface has been developed to visualize and to update RDF databases. This interface proposes an autocomplete mechanism that uses the suggestion system for providing values when editing a resource. A demonstration video of this tool in use for the Henri Poincaré correspondence corpus is available online.

Two evaluations have been conducted for this system. The first evaluation is human-based through a user who edits around 30 letters of the Henri Poincaré correspondence corpus using the Web editor. At the end of the editing, he/she is invited to complete a survey to give his/her feedback about the four suggestion system versions. A second evaluation is managed through a dedicated program which provides objective measures by simulating editing situations based on already edited content of the corpus. Both evaluations results suggest that the version combining RDFS entailment and case-based reasoning is the most efficient one and could significantly assist users in the editing process.

Acknowledgement. This work is supported partly by the French PIA project “Lorraine Université d’Excellence”, reference ANR-15-IDEX-04-LUE.

References

1. Bruneau, O., Gaillard, E., Lasolle, N., Lieber, J., Nauer, E., Reynaud, J.: A SPARQL Query Transformation Rule Language — Application to Retrieval and Adaptation in Case-Based Reasoning. In: Aha, D., Lieber, J. (eds.) *Case-Based Reasoning Research and Development*. ICCBR 2017. pp. 76–91. Springer, Cham (2017)
2. Lasolle, N., Bruneau, O., Lieber, J., Nauer, E., Pavlova, S.: Assisting the RDF Annotation of a Digital Humanities Corpus Using Case-Based Reasoning. In: Pan, J.Z., Tamma, V., d’Amato, C., Janowicz, K., Fu, B., Polleres, A., Seneviratne, O., Kagal, L. (eds.) *The Semantic Web - ISWC 2020*. pp. 617–633. Springer, Cham (2020)