How Redundant Is It? - An Empirical Analysis on Linked Datasets

Honghan Wu[1], Boris Villazon-Terrazas[2], Jeff Z. Pan[1], and Jose Manuel Gomez-Perez[2]

[1] Department of Computing Science, University of Aberdeen, UK
[2] iSOCO, Intelligent Software Components S.A., Spain

Abstract. While there are some popular vocabularies widely used across linked open data, many linked data sets do not have T-Box axioms at all. How does such fact, i.e., the usage patterns on T-Boxes, affect the linked data consumption? This might be an interesting question to be asked by linked data consumers. In this paper, we analysis one particular aspect of this question i.e., redundancy analysis, on several popular datasets and vocabularies in web of data. We start with the analysis on semantic redundancies of datasets given their T-Boxes. Then, we propose useful t-box axioms which are helpful for consumption but are absent in the dataset in question. Finally, we reveal how the linkages of the web of data, both in concept-level and instance-level, affects data set redundancies.

1 Introduction

@Boris

2 Graph Pattern Based Redundancy Identification

3 A-Box Redundancy

4 T-Box Rules VS. A-Box Redundancy

5 Redundancy deviation by vocabulary Linkage

6 Graph Pattern Based Redundancy Identification

7 Related Work

@Boris

8 Conclusion

References