

Link Maintenance in the Semantic Web

Andre Gomes Regino¹, Julio Cesar dos Reis¹

¹Institute of Computing – University of Campinas (Unicamp)
Campinas – SP – Brazil

{andre.regino,jreis}@ic.unicamp.br

Abstract. *Connections among data elements represent the core of Semantic Web. The connections are built with semi-automatic linking algorithms using a variety of similarity calculus. The data interconnected by these algorithms demands automatic methods and tools to maintain its consistency. Even though the constant update of RDF connections is considered an important process for the evolution of these structured datasets, such changing operations can influence the well-formed links, which turns difficult the consistency of the connections over time. In this work, we aim to investigate new methods responsible for fixing and updating links among ontologies in the Linked Open Data context.*

1. Context

Links between LOD datasets are at the heart of the Web of Data. Although the implementation of change operations in LOD datasets is essential to assure structured data evolution these operations can affect established links, which might turn them invalid or inconsistent. These links are maintained sporadically and manually [Bizer et al. 2009]. Also, ontologies, vocabularies and data schemas can change the definition and structure of RDF data. The manual maintenance remains hardly accomplishable due to the overwhelming number of links available.

2. Goal

We aim to investigate, formalize and implement semi automatic link maintenance actions in order to recognize affected links and turn them up-to-date. Figure 1 shows an evolution of a removal of a given triple and the absence of a link removal associated to that triple.

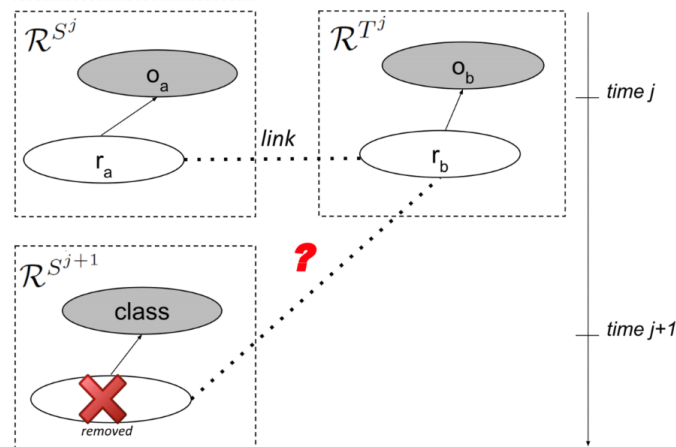


Figure 1. Problem Characterization

3. Methodology

In order to keep to links up-to-date, we are building a framework composed by three main steps, listed as:

- **Step A:** Detection of changes, given two versions of the same dataset as input, the framework maps every simple or complex change through these versions;
- **Step B:** Recognition of affected links, given the changes mapped at Step A, discover which of the links became structurally or semantically broken;
- **Step C:** Application of maintenance action, given the list of broken links as output of Step B, select based on a given number of actions which one is appropriate to make the links consistent.

4. Initial Results

Table 4 shows the results we collected in a study [Regino et al. 2019] that interrelates changes in triples (lines) with changes in links (columns) performed in a life sciences dataset named Agrovoc¹. The first table shows that Agrovoc dataset applies the concept of Linked Data, linking 99% of their newly added triples to an external dataset. In second table, however, 96.15% of identified removed cases shows that if an internal triple is removed, the connected link remained untouched, generating cases of structurally broken links. The third table, regarding modification, shows that the fourth sub-case concerns the most frequent one, in which the modification of triples led to unchanged links. This case needs additional studies to further observe to which extend these unchanged links remained semantically inconsistent due to the modifications of the associated RDF triples.

Triples / Links	Add	No Add
Add	98.84%	0.41%

Triples / Links	Remove	No Remove
Remove	3.85%	96.15%

Triples / Links	Add	Remove	Modify	No Change
Modify	0%	0.04%	4.41%	95.55%

Table 1. Add, Remove and Modify Actions

5. Next Steps

We are now focusing on developing novel strategies to address the challenges on identifying broken links and maintaining them (Steps B and C of Section 3).

References

- Bizer, C., Heath, T., and Berners-Lee, T. (2009). Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems*, 5(3):1–22.
- Regino, A. G., dos Reis, J. C., Matsoui, J., Bonacin, R., Morshed, A., and Sellis, T. (2019). Understanding link changes in lod via the evolution of life science datasets. In *Workshop Semantic Web solutions for large-scale biomedical data analytics (SeWeBMeDA-2019) co-located at the 18th International Semantic Web Conference (ISWC'19), Auckland, New Zeland. (accepted for publication)*.

¹<http://aims.fao.org/standards/agrovoc/linked-data>