



Updating Semantic Annotations in RDF Format

Mariana T. Bisca*, Júlio C. dos Reis.

Abstract

Semantic annotations are sets of labels that help structure the information available in web documents, making it easier for machines to interpret this information automatically. While previous works have explored the (semi-) automatic maintenance of these semantic annotations in relation to ontology-level changes, this work will focus on the effects of instance-level changes, specifically regarding annotations in RDF format. Given a set of annotations affected by an instance-level change, a classification-based approach will be used in order to determine which out of all possible update operations is most appropriate for each annotation, thus keeping the information system consistent and up-to-date.

Key words:

Semantic Annotations, Web Semantics, RDF

Introduction

Though there is a nearly infinitely large amount of useful information spread around the world wide web, the unstructured nature of web data makes it hard to access this information directly. Semantic annotations provide a way to add structure to web documents, describing the different entities present in a given document as well as the underlying relationships between these entities. This work focuses specifically in semantic annotations that use the Resource Description Framework (RDF), a data model consisting mainly of triples formed by a “subject” and an “object” that relate to each other by a given “predicate”. In this context, the “subject” usually refers to a document or entity, while the “object” can be a concept, an instance or perhaps even a real value.

The full set of instances, concepts and relationships that can be used to describe the domain in which an information system is set is given by its underlying ontology. As most domains are constantly evolving throughout time, it's important that the ontologies used to represent them evolve in the same manner, keeping the information system.

This work will study the way that instance-level changes to the ontology affect previously obtained semantic annotations and seek to develop a way to keep these annotations consistent and up-to-date.

Results and Discussion

The main results produced by this work will be the technique designed to (partially) automate the maintenance of semantic annotations.

One possible approach to build this technique is to frame it as a classification problem. Through the study of the inconsistencies caused by the instance level changes, it is possible to define a finite set of potential operations that can be used to update that annotation in order restore overall consistency.

Important details about methodology can be described in this item.

The technique is therefore only a matter of determining which of the potential operations is most appropriate for each inconsistent annotation. This kind of classifier can be constructed either by writing explicit rules or by training a machine learning model. The details of the solution will be part of the continuation of this work.

Conclusions

Semantic annotations play an important part in leveraging the potential of initially unstructured information, and keeping these annotations up-to-date manually can be time-consuming and lead to errors. This work outlines the way to construct a technique that can at least partially automate this process, and will implement and test this technique in the future, thus making it easier to maintain the overall consistency of an information system so that it can be used to build powerful applications.

Acknowledgement

We gratefully acknowledge FAEPEX/UNICAMP for funding this research. The authors also thank the Institute of Computing (IC) community for their support.

¹ Cardoso, S. D.; Chantal, R. D.; Silveira, M.; Pruski, C.; Combining rules, background knowledge and change patterns to maintain semantic annotations. American Medical Informatics Association. 2017. 505.

² Christen, V.; Lin, Y. C.; Groß, A.; Cardoso, S. D.; Pruski, C.; Silveira, M.; Rahm, E.; A learning-based approach to combine medical annotation results. International Conference on Data Integration in the Life Sciences. 2018. 135-143.