Universität Leipzig

Fakultät für Mathematik und Informatik
Institut für Informatik

Design and Implementation of DBpedia Archivo - an Augmented Ontology Archive

Bachelorarbeit

Leipzig, März 2021

vorgelegt von

Denis Streitmatter
Studiengang Bachelor Informatik

Betreuende Hochschullehrer:

Dr.-Ing. Sebastian Hellmann
Institut für Angewandte Informatik/KILT

Johannes Frey, M.Sc.
Institut für Angewandte Informatik/KILT

Dr. Eric Peukert
Abteilung Datenbanken Universität Leipzig

Streitmatter, Denis:
Design and Implementation of DBpedia Archivo - an Augmented Ontology Archive
Bachelorarbeit, Universität Leipzig
Leipzig, 2021.

Contents

Abstract

Over the last years, a huge amount of work has been done to improve the ability of machines to utilize data on the Web. One approach is the Semantic Web, using ontologies as a way to make the knowledge of a domain machine-usable. Even though many ontologies were developed and published, a unified system to handle those has not surfaced, leaving consumers as well as publishers to deal with many uncertainties and challenges.

This thesis presents DBpedia Archivo, an augmented ontology archive. It discovers, crawls, versions, and archives ontologies available on the Web. Each version of them is persisted on the DBpedia Databus. Additionally, Archivo augments the ontologies with different tests and features. The goals of Archivo are to provide a backup service for ontology-versions as well as to encourage publishers to follow best practices. For this Archivo rates the ontologies with a star system, making problems visible at a glance. A comparison to existing, similar systems is given.

1 Introduction and Motivation

1.1 Motivation

1.2 Structure of this Thesis

2

Preliminaries

3

Problem Analysis

4

Design

5
Evaluation

6

Conclusions and Future Work

6.1 Conclusions

6.2 Future Work

A Appendix

Common Abbreviations

Acronym	Meaning
NIR	Non-information Resource. An IRI that stands for a abstract entity on the web, in this thesis mainly ontologies.
VCS	Version Control System. Stands for a system to collaborative edit documents. Most noteable are git and the service GitHub.
URI	Uniform Resource Identifier. A string consisting of ASCII characters identifying a resource, mostly known for its usage in the World Wide Web.
IRI	The extended version of URIs, allowing not only ASCII characters but also most of Unicode characters.

Table A.1: Commonly used abbreviations in this thesis

Prefixes

Prefix	**URI**
ex	`http://example.org/resources/`
exo	`http://example.org/ontology/`
owl	`http://www.w3.org/2002/07/owl#`
rdfs	`http://www.w3.org/2000/01/rdf-schema#`
dct	`http://purl.org/dc/terms/`
dc	`http://purl.org/dc/elements/1.1/`
xhv	`http://www.w3.org/1999/xhtml/vocab#`
sh	`http://www.w3.org/ns/shacl#`
skos	`http://www.w3.org/2004/02/skos/core#`
dbo	`http://dbpedia.org/ontology/`
xsd	`http://www.w3.org/2001/XMLSchema#`

Table A.2: Prefixes and related URIs used in this thesis

B

Bibliography

Erklärung

Ich versichere, dass ich die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Quellen und Hilfsmittel angefertigt habe, insbesondere sind wörtliche oder sinngemäße Zitate als solche gekennzeichnet. Mir ist bekannt, dass Zuwiderhandlung auch nachträglich zur Aberkennung des Abschlusses führen kann. Ich versichere, dass das elektronische Exemplar mit den gedruckten Exemplaren übereinstimmt.

Datum: .....

(Unterschrift)