

Marie Skłodowska-Curie Actions (MSCA)
Innovative Training Networks (ITN)
H2020-MSCA-ITN-2014

WDAqua
“Answering Questions Using Web Data”
ETN, Grant no. 642795

<http://wdaqua.informatik.uni-bonn.de>

Dissemination level	Confidential, only for members of the consortium (including the Commission Services)
Type of document	Other
Contractual date of delivery	M16
Actual date of delivery	2015-05-02
Deliverable number	D7.1
Deliverable name	OpenQA Reference Implementation Document
Deliverable leader	FRAUNHOFER
Work package	WP1
Tasks	T7.2
Coordinator (name, affiliation, contact)	Christoph Lange (FRAUNHOFER; math.semantic.web@gmail.com)
EC project officer	Matias Pandolfo

1. Overview

The objective of the Cloud-based open question answering architecture (OpenQA) for the second year of the WDAqua project is to design a distributed, scalable open architecture that supports pluggable integration of many different approaches for individual stages in a QA pipeline. This architecture will be matured subsequently; meanwhile, components for it will be implemented by several researchers in the project.

We aim towards a message driven interoperable question answering (QA) system. To achieve this, we have collected requirements from different data models used in state-of-the-art QA systems to cover the need of the existing QA systems. To cover these needs, we have further developed the *qa* vocabulary, a generalised vocabulary, using the W3C-Web Annotation Data Model¹. This vocabulary has a machine-readable, extensible representation to address the requirements collected. A conceptual view of QA systems has been set up using this vocabulary. We have further developed Qanary – A Methodology for Vocabulary-driven Open Question Answering Systems. Using Qanary, we have integrated several independent QA components into a QA pipeline according to the OpenQA specification; the initial batch of these components constitutes our reference implementation.

We consider it reasonable to devote *one* self-contained document to the high-level description of the Architecture to avoid redundancy. We chose to deliver this description in D4.1 “OpenQA specification and reference implementation”. The implementation corresponding to the current Deliverable D7.1 is available from the public source repository of the WDAqua project:

- <https://github.com/WDAqua/QAOntology> hosts the *qa* vocabulary. Its documentation is embedded as machine-readable annotations into its implementation; however, we also link to a human-friendly rendering of the documentation, which has been generated from this source.
- <https://github.com/WDAqua/Qanary> hosts the generic framework for implementing QA pipelines, as well as the implementations of several components for Entity Linking (i.e. Named Entity Recognition and Named Entity Disambiguation). Both together serve as the initial reference implementation. Their documentation is, once more, part of the Java source code, but has also been exported to a human-friendly view.

[1]: <https://www.w3.org/TR/annotation-model/>