

Using DevOps Principles to Continuously Monitor RDF Data Quality

Roy Meissner
Institute for Applied Informatics
Hainstrasse 11
Leipzig, Germany
meissner@informatik.uni-leipzig.de

Kurt Junghanns
Institute for Applied Informatics
Hainstrasse 11
Leipzig, Germany
kjunghanns@informatik.uni-leipzig.de

ABSTRACT

One approach to continuously achieve a certain data quality level is to use an integration pipeline that continuously checks and monitors the quality of a data set according to defined metrics. This approach is inspired by Continuous Integration pipelines, that have been introduced in the area of software development and DevOps to perform continuous source code checks. By investigating in possible tools to use and discussing the specific requirements for RDF data sets, an integration pipeline is derived that joins current approaches of the areas of software-development and semantic-web as well as reuses existing tools. As these tools have not been built explicitly for CI usage, we evaluate their usability and propose possible workarounds and improvements. Furthermore, a real-world usage scenario is discussed, outlining the benefit of the usage of such a pipeline.

The full paper is included in the ACM Proceedings of the 12th International Conference on Semantic Systems, SEMANTICS 2016, ISBN 978-1-4503-4752-5.

DOI <http://dx.doi.org/10.1145/2993318.2993351>