The State of Linked Data

An Extensible Framework to Asses and Build Dataset Profiles

Ahmad Assaf[1,2], Aline Senart[2] and Raphaël Troncy[1]

[1] EURECOM, Sophia Antipolis, France. `<firstName.lastName@eurecom.fr>`
[2] SAP Labs France. `<firstName.lastName@sap.com>`

Abstract. Linked Open Data (LOD) has emerged as one of the largest collection of interlinked datasets on the web. Benefiting from this mine of data requires the existence of descriptive information about each dataset in the accompanying metadata. Such meta information is currently very limited to few data portals where they are provided manually, provide little or bad quality insights or not available at all. Several automatic profiling techniques are needed to generate descriptive and informative meta information. To address this issue, we propose a scalable automatic approach for extracting linked dataset meta information. This approach apply several techniques to check the validity of the attached metadata of a certain dataset as well as a whole Data Portal. Furthermore, we apply our framework on prominent Data Portals in order to analyze the general state of Linked Open Data.

Keywords: Linked Data, Dataset Profile, Metadata, Data Quality

1   Introduction
References