Ruben Taelman

Using an existing website as a queryable low-cost LOD publishing interface

Demo

Brecht Van de Vyvere , Ruben Taelman , Pieter Colpaert , Ruben Verborgh
In Proceedings of the 16th Extended Semantic Web Conference: Posters and Demos (2019)

Maintaining an Open Dataset comes at an extra recurring cost when it is published in a dedicated Web interface. As there is not often a direct financial return from publishing a dataset publicly, these extra costs need to be minimized. Therefore we want to explore reusing existing infrastructure by enriching existing websites with Linked Data. In this demonstrator, we advised the data owner to annotate a digital heritage website with JSON-LD snippets, resulting in a dataset of more than three million triples that is now available and officially maintained. The website itself is paged, and thus hydra partial collection view controls were added in the snippets. We then extended the modular query engine Comunica to support following page controls and extracting data from HTML documents while querying. This way, a SPARQL or GraphQL query over multiple heterogeneous data sources can power automated data reuse. While the query performance on such an interface is visibly poor, it becomes easy to create composite data dumps. As a result of implementing these building blocks in Comunica, any paged collection and enriched HTML page now becomes queryable by the query engine. This enables heterogenous data interfaces to share functionality and become technically interoperable.

Article   BibTeX