

Conversion of traditional on-line accessible dictionaries into Web of Data formats

What need was solved?

KDictionaries is an Israeli company dedicated to the creation and provision of cross-lingual lexical data. Over the years, the company has published several monolingual, bilingual and multilingual online dictionaries. KDictionaries wanted to adapt its lexical data so that language technology applications could interoperate with it. To achieve this objective, the Spanish dictionary of the “KDictionaries multilingual Global Series” dataset was modelled to be converted from its proprietary format into RDF for its later publication as linguistic linked data. The model was to be compatible with the English and German datasets of the same series. This would also allow the linking of KDictionaries datasets to other open external datasets.

What services were provided?

Several services were provided to achieve this goal:

- Methodological guidelines to transform the resource from its proprietary XML format to RDF, according to the lemon-ontolex model, a de-facto standard for representing lexical and linguistic information on the Web of Data.
- Definition of the model followed to convert the KDictionaries dataset into RDF based primarily on the lemon-ontolex model, with some extensions to meet the needs of the rich dictionary data.
- Definition of a suitable strategy to construct permanent Uniform Resource Identifiers (URIs) that identify the elements in the model.
- Identification of a set of mappings between the XML proprietary data and the lemon-ontolex model.

The relation with digitization

KDictionaries source datasets were already in a digital format, in XML. However, their dictionaries were not linked among each other in a way that maximised their reusability, and their proprietary format prevented the data from being linked to external linguistic resources. The modelling of their data as RDF was a first step to facilitate the exploitation of their content by linked-data aware technologies by abstracting from proprietary digital formats and moving into standardised representation mechanisms for digital content.

Customer

KDictionaries
<https://www.lexicala.com/>
 Israel