WIKIPEDIA
The Free Encyclopedia

Linguistic Linked Open Data

In natural language processing, linguistics, and neighboring fields, **Linguistic Linked Open Data (LLOD)** describes a method and an interdisciplinary community concerned with creating, sharing, and (re-)using language resources in accordance with Linked Data principles. The **Linguistic Linked Open Data Cloud** was conceived and is being maintained by the Open Linguistics Working Group (OWLG) of the Open Knowledge Foundation, but has been a point of focal activity for several W3C community groups, research projects, and infrastructure efforts since then.

Definition and Development

Linguistic Linked Open Data describes the publication of data for linguistics and natural language processing using the following principles:[1]

LLOD Cloud (2016-05-24)

- Data should be openly licensed using licenses such as the Creative Commons licenses.
- The elements in a dataset should be uniquely identified by means of a URI.
- The URI should resolve, so users can access more information using web browsers.
- Resolving an LLOD resource should return results using web standards such as the Resource Description Framework (RDF).
- Links to other resources should be included to help users discover new resources and provide semantics.

The primary benefits of LLOD have been identified as:[2]

- Representation: Linked graphs are a more flexible representation format for linguistic data.
- Interoperability: Common RDF models can easily be integrated.
- Federation: Data from multiple sources can trivially be combined.
- Ecosystem: Tools for RDF and linked data are widely available under open source licenses.
- Expressivity: Existing vocabularies help express linguistic resources.
- Semantics: Common links express what you mean.
- Dynamicity: Web data can be continuously improved.

The home of the LLOD cloud diagram is under linguistic-lod.org[3]

LLOD vocabularies

Aside from gathering metadata and generating the LLOD cloud diagram, the LLOD community is driving the development of community standards with respect to vocabularies, metadata and best practice recommendations.

According to the state-of-the-art overview by Cimiano et al. (2020),[4] these include:

- for modelling lexical resources