Ordia: A Web application for Wikidata lexemes

Finn Årup Nielsen
Cognitive Systems, DTU Compute, Technical University of Denmark, Kongens Lyngby

Summary

Ordia is a web application for Wikidata lexemes querying the Wikidata Query Service on-the-fly when the user navigates the Ordia interface. A user may view individual lexemes, their forms and senses as well as their compound and derivation graph. Other webpages on Ordia aggregates lexemes for language, lexical category, grammatical feature, properties or references. Ordia has also special features: text-to-lexemes and integration of knowledge graph embedding similarity results.

WikiData lexemes

WikiData has as of May 2019 over 46,000 lexemes. Ordia has a dedicated page that shows statistics:

Statistics

The text-to-lexemes facility in Ordia enables a user to enter a text in the web interface and let Ordia extracts synonyms and words. Extracted words are sent for matching to the Wikidata Query Service and Ordia displays the result of the matching in a table.

Text-to-lexemes

A number of so-called aspects show specific information related to: language, lexical categories, grammatical features, properties or references. Each of these aspects has a dynamically updated index page (e.g., /language for language) and dynamically constructed pages for individual Wikidata items (e.g., /language/Q24905 for the Danish language). Here is the ontology graph that is part of the lexical category aspect for verb (Q24905)

Aspects

The user should easily be able to perform powerful SPARQL queries by navigating the Ordia Interface URLs should be predicatable, e.g., /lexical-category/Q24905 shows the Wikidata item for verb (Q24905) as a lexical category. /property/P31/value/Q639512 is the aspect for instance of (P31) cranberry morpheme (Q639512). Each page should link to other pages and let the user discover new lexicographic relations. The interface should use graphics whenever possible, e.g., graphs for leemce and concept relations and for displaying images associated with senses of lexemes.

Conceptual choices

The Javascript DataTable package allows the user to filter results displayed in tables in Ordia on the client side.

Search results

Copy-and-paste of a text into Ordia’s text-to-lexemes facility to see if all words match Wikidata lexemes, and if not enable the user to quickly create the missing lexemes. Get an overview of the Wikidata linguistic ontology and its usage, from, e.g., the lexical category aspect. Searching for lexemes View statistics, e.g., number of lexemes, number of forms, lexemes per language and lexemes per lexical category.

Text-to-lexemes?
text=Blue+cars,+green+bikes+and+red+motorcycles+must+stop+at+the+crossing.
text-language=en

Aspects

https://tools.wmflabs.org/ordia/
text-to-lexemes?text=blue+cars,_green+bikes+and+red+motorcycles+must+stop+at+the+crossing.&text-language=en

Future

Current use

This work is funded by the Innovation Fund Denmark through the projects DAnish Center for Big Data Analytics driven Innovation (DABAI) and Teaching platform for developing and automatically tracking early stage literacy skills (ATEL).

Acknowledgment

References


Current annotation for Wikidata lexemes is too sparse for Ordia to make sense as a cross-language dictionary, etymological dictionary or synonym dictionary. Possible future extensions could include internalization of the user interface and Wikidata entry directly in Ordia.

Acknowledgment