Arabic Ontology Data

Overview

This directory contains a copy of the Arabic ontology extracted on 09-06-2024.

The Arabic ontology is stored as two CSV files: **concepts** and **relations**. Below is a description of the content of these two files:

Concepts File

The concepts file contains a list of concepts (nodes). Each concept is described by the following columns:

- **Concept ID:** The identifier for the concept in the ontology.
- **Arabic Synsets:** A list of Arabic synonyms for the concept, separated by "|".
- **English Synsets:** A list of English synonyms for the concept, separated by "|". Note that only a few concepts have English terms.
- **Gloss:** The ontological definition of the concept.
- **Example:** Contextual examples showing how the Arabic synonyms can be used.
- **Data Source Id:** An internal number that will be explained in a separate note later.

IMPORTANT REMARK: We would like to highlight that the Concept IDs with a source id = 200 are well-designed and well-studied. However, concept IDs in other dataSources are draft concepts under construction. These draft concepts, along with their synonyms and relations, should be carefully used.

Relations File

The relations file represents the ontology tree. Each row in the file contains the following columns:

- **Concept ID:** The identifier for the concept in the ontology.
- **subTypeOf:** Indicates the parent-child relationship between concepts. For example, if concept ID 1 is a subtype of concept ID 2, in this case the concept ID 1 is the child, and the number in the subTypeOf column is the parent.

- **partOf:** Used similarly to subTypeOf.
- **instanceOf:** Used similarly to subTypeOf.

Copyright

All data are copyrighted by Birzeit University.

Citation

Users and researchers utilizing our data and ideas must acknowledge and cite the following articles in a proper and clear manner:

[1] Mustafa Jarrar: The Arabic Ontology - An Arabic Wordnet with Ontologically Clean Content. Applied Ontology Journal, 16:1, 1-26. IOS Press. 2021

[2] Mustafa Jarrar, Hamzeh Amayreh: An Arabic-Multilingual Database with a Lexicographic Search Engine. The 24th International Conference on Applications of Natural Language to Information Systems (NLDB 2019). Pages(234-246). LNCS 11608, Springer. 2019

License

This data cannot be merged, mixed, or distributed, either fully or partially.