Semester Project Definition

The main goal of the project is to build a knowledge graph by applying the methodology being introduced in the lecture. The following presents the details about the workpackages and submission of the project.

1 Workpackages

WP1 – Knowledge Creation

For this workpackage , you need to decide on a domain and identify 2 different data sources that can be used as a basis for knowledge creation. Ideally, you should select sources with different characteristics, for example, one unstructured (e.g., text) and one (semi-)structured (e.g., a JSON, CSV, XML, relational database). Then, use an appropriate knowledge creation method to create RDF data from the data sources you found.

Deliverables

D1.1. Domain Selection and Identification of sources (06.12.2023)

D1.2. RDF data created from the sources (13.12.2023)

WP2 – Knowledge Hosting

The created data should be stored in a triple store. Upload the generated RDF data to the triple store of your choice. Do not forget to attach provenance information. Use named graphs to host triples coming from different sources. Attach the at least the following provenance information to each named graph:

- Creation date
- Name of the data source
- Number of triples in the named graph

Deliverables

D2.1. Knowledge graph stored in a triple store (20.12.2023)

WP3 – Knowledge Assessment

Assess the quality of your knowledge graph for two dimensions as introduced in the lecture, namely correctness and completeness. Calculate an aggregated quality score for your knowledge graph. You can use one of the tools presented in the lecture or implement your own custom assessment tool.

Deliverables

D3.1. Calculation of the quality scores for the correctness and completeness dimensions as well as the calculation of an aggregated quality score (03.01.2027)

WP4 – Knowledge Enrichment

Find another source different than the initial two. This source could an existing knowledge graph. Enrich your knowledge graph following the methodology presented in the lecture. Apply duplicate detection. You may use one of the tools presented in the lecture or implement your own solution.

Deliverables

D4.1. Enriched knowledge graph with linked duplicate instances (17.01.2023)

WP5 – Knowledge Cleaning

Apply error detection on your knowledge graph and identify errors. Use one of the tools presented in the lecture or your own custom implementation.

Deliverables

D5.1. Error detection implementation and a report of the found errors (31.01.2023)

2 Submission

There is no formal template or instructions for the presentation of the deliverables. However, a GitHub repository must be maintained for each group and each deliverable must be documented in a Readme file and/or Wiki. This repository will be the final submission of the project. The deadline for the final submission is **14.02.2024.**