Information Integration and Analytic Data Processing

Individual Challenge

Challenges

Each member of the group must pick a different one

Delivery 25/05/2025

One page report detailing your work submitted in moodle + code

Project final submission should include all individuals works.

Choose one Challenge

New Question & Datasource

Add a new question/problem that requires at least one additional datasource

   a.   Characterize the new data, update the integrated schema, analyse and present the result

Data Cleaning Expert

Duplicate one of your datasets, manipulate the records to have many types of new significant data issues (See T5 slides)

   b.   Address the created data issues created with data cleaning procedures. Make sure you are dealing with multiple complex issues.

Extracting Knowledge

Create an workflow with LLM to extract knowledge from free form text to create a new datasource for your project

   c.   Characterize the new data, update the integrated schema

   d.   *the free form text can come from an existing datasource

Choose a Challenge

Quality Assessment

Create quality assessment metrics for the different steps of the project.

    a. Create a *gold standard* sample dataset, evaluate your entity matching, blocking etc.

Rich Data Characterization & Strategic Planning

Create a detailed characterization of your original datasets (see slides T5 data profiling)

    b. Find and discuss insights about your data. Discuss strategies of what should be done to address the flaws in your data (e.g. if missing data how could it be filled)

Rich Data Visualizations & Insights

Create rich visualizations and dashboards to explore your project data beyond the initial questions.

    a. Detail what we can learn from the data.