# Introduction

Over the past couple of decades there has been an exponential increase in the amount of data that is generated by the web and more recently the internet of things (IOT). This exponential increase has not always been accompanied by a similar increase in the extraction of information and knowledge from this data. The lag in information gain is mainly due to the large volume and unknown structure of the data that makes their analysis quite a difficult and complex task [1]. To overcome this issue, it is essential to develop the necessary tools that will help understand and manage the content of the data. Those tools will allow the correct classification of the data and will incorporate data transformation, cleansing and standardisation as part of pre-processing steps.

On top of data mining and knowledge retrieval there are many cases where different services offered by different applications need to be integrated or data need to be transferred from one system to another (e.g. system migrations). In those cases, it is important to have an interface agreement whereby data from interacting systems need to be well curated and mapped. Oftentimes this is a cumbersome and manual task especially when no additional documentation or system specs are provided.

Within the space described above, the research question we will try to answer as part of this project is the following:

*How can we enhance a set of data given as input (e.g. tabular data) with semantic meaning using existing knowledge graphs (e.g. DBpedia, WikiData) as reference?*

This project is inspired by the Sem Tab challenge that has been organised annually since 2019. The scope of the challenge is organised in 3 separate but overlapping tasks listed below:

* CTA Task: Assign a class from a KG to an entire column of a table
* CEA Task: Assign an individual entity of a KG to each specific cell
* CPA Task: Assign the relationship (i.e. object property) between 2 table columns

The above tasks are reflected in Figure 1

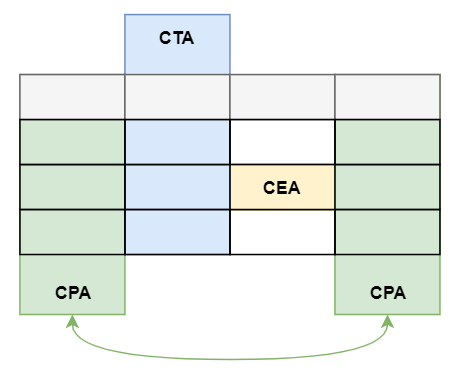


Figure 1. Matching tabular data to classes (CTA), entities (CEA) and properties (CPA)

# Critical Context

Dfdf dsf sd f sd f sdf sd f sdf sd f dsf

# Approaches: Methods & Tools for Design, Analysis & Evaluation

# Work Plan

# Risks

# Ref

[1] A. Lausch, A. Schmidt, L. Tischendorf, “Data mining and linked open data – New perspectives for data analysis in environmental research”, Ecological Modelling 295, 2015, p. 5-17 [https://doi.org/10.1016/j.ecolmodel.2014.09.018](https://0-doi-org.wam.city.ac.uk/10.1016/j.ecolmodel.2014.09.018)  
  
  
  
F. Kalloubi, E. H Nfaoui, O. El Beqqali, “Micro blog semantic context retrieval system based on linked open data and graph-based theory”, Expert Systems With applications 53, 2016, p. 138-148

[10.1109/SITA.2014.6847286](https://0-doi-org.wam.city.ac.uk/10.1109/SITA.2014.6847286)

T. Di Noia, V. C. Ostuni, J. Rosati, P. Tomeo, E. Di Sciascio, R. Mirizzi, C. Bartolini, “Building a relatedness graph from Linked Open Data: A case study in the IT domain”, Expert Systems With Applications 44, 2016, p. 354-366

[https://doi.org/10.1016/j.eswa.2015.08.038](https://0-doi-org.wam.city.ac.uk/10.1016/j.eswa.2015.08.038)

The diagram in Figure 1 illustrates the steps taken for the first 3 tasks of the coursework (i.e. to create an ontology based of the give data, load the data to a KG and extract useful insights).

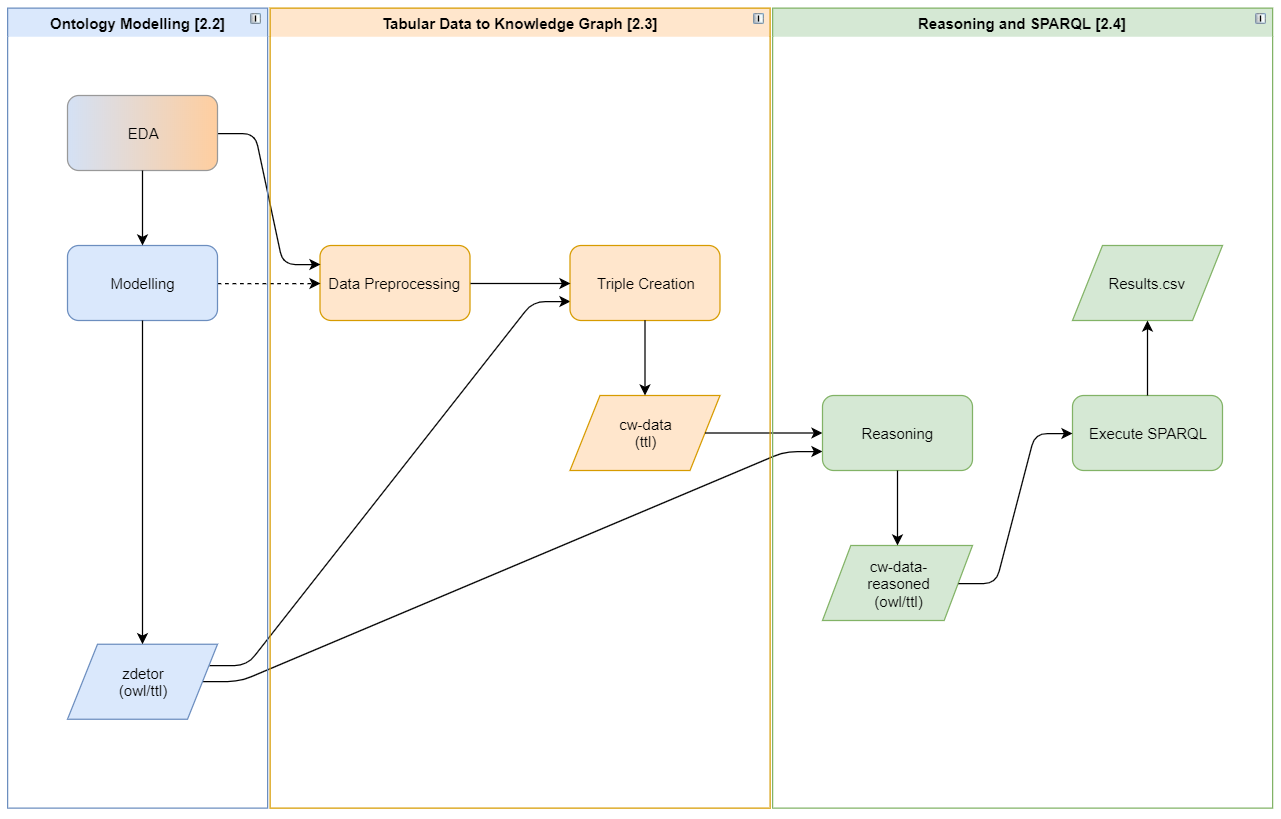


Figure 1. Process of modelling and converting tabular data to a KG