KNOWLEDGE DISCOVERY IN GRAPH DATA

PROJECT SUBJECTS

M2-DS DATA SCIENCE, PARIS SACLAY UNIVERSITY

DETAILS ON: https://www.lri.fr/~sais/M2DS/projects/

Project submission at: https://ecampus.paris-saclay.fr/course/view.php?id=38304&section=5#section-5

1. LINK VALIDATION PROJECT

Input

Defense: 05/01/2021

- A set of identity links
- A OWL ontology, a set of mappings between properties, a set of functional properties

What is expected from you

- A ***program*** able to discover some ‘false’ sameAs links. The idea is only to look for functional datatype properties that have different values.
- An analysis of the results on a sample (erroneous data ?, …)

- Group of 3 students

2. DATA LINKING AND KEYS PROJECT

Input provided by the instructor

Defense: 05/01/2021

- An OAEI benchmark (e.g IIMB about cinema)
- A tool that is able to discover a set of composite key with exceptions from the dataset (SAKEY – ISWC14) (e.g name+firstname+birthdate is a key)

What is expected from you

- A ***program*** able to automatically select a subset of properties in order to eliminate unrelevant properties (long text, sameAS, typeOf) before applying SAKEY
- An analysis of the results of SAKEY with and without your preprocessing step.
- A set of strategies that can be used to merge keys with exceptions when they are discovered in different datasets (theoritically only)
- Group of 2 or 3 students

3. DATA LINKING AND KEYS PROJECT

Input provided by the instructor

Defense: 05/01/2021

- An OAEI benchmark (e.g IIMB about cinema)
- A tool that is able to discover a set of composite key with exceptions from the dataset (SAKEY – ISWC14) (e.g name+firstname+birthdate is a key)

What is expected from you

- A ***program*** that performes a post processing of Sakey.
- Generate key graphs for one class from the keys discovered for all classes (replace an object properties by the path that lead to keys detected for the range of the property)
- Give the algorithm and the program.
- Compare the number of keys obtained with local keys or graph keys as well as time execution.

- Group of 2 or 3 students