Terminology and Ontology
Project

Mathieu Constant

2020/2021

Project

- Subject : develop a term identification system for a specific domain
- Presentation : November 25, 2020
- Deposit on Arche : before presentation
- By groups of two complementary people

Instructions

- Collect a corpus of articles in a specific domain
- Extract automatically the terms from this corpus
- Filter manually the list of terms
- Develop (or get) a rule-based identification system relying on this list
- Apply it on your corpus to obtain an annotated corpus
- Train a sequence tagger of your choice from this annotated corpus in a supervised way
- Given a new text of your targeted domain, apply your two identification systems, and evaluate/compare the results
- Optional : combine the two systems in one.