Extraction and completion of knowledge graph from historical sources.

June 19, 2025

keywords: NLP, Knowledge graph, Graph completion

Duration: 6 months

The Hakken team of Sony Research and Sony CSL, Paris are collaborating in a project about sustainable agriculture. The aim of the project is to look into old and new textual ressources about agriculture (especially market gardening) to make this knowledge accessible and potentially make original recommendations about farming practices.

The objective of the internship is to build a knowledge graph from the a corpus of texts from the 19th century [1] and evaluate an algorithm for graph completion. It includes preprocessing the data (translation,...), exploring methods to build knowledge graph and comparing it to the current template based method. The candidate should have good knowledge in NLP, especially about knowledge graphs. The intern will be supervised both by researchers in Sony Research's team Hakken (Alessandra Toniato, Pablo Sanchez, and Tarek Besold) and in Sony CSL Paris (David Colliaux, Pater Hanappe).

[1] Colliaux, David, and Remi van Trijp. "The discourse of the French method: making old knowledge on market gardening accessible to machines and humans." Computational Humanities Research 2024.

Contact: David Colliaux david.colliaux@sony.com