Internship Proposal: Advancing FLORA

Supervisors: Fabian M. Suchanek, Thomas Bonald, Yiwen Peng
Location: Télécom Paris (19 Place Marguerite Perey, 91120 Palaiseau)
Duration: 5 to 6 months
Starting: March or April 2026

The goal of the project is to advance the FLORA entity-alignment system [1]. This system can align two knowledge bases without the need for training data, by using a Fuzzy-Logic based formalism and algorithm. The system is the state of the art, and won the best paper award of the ISWC 2025 conference.

At the same time, several avenues remain to be explored: Most importantly, the initial phase of literal matching is currently too inefficient and too slow (it runs in quadratic time). This phase would have to be improved, both algorithmically and engineering-wise. Second, partly due to the first limitation, the system cannot currently run on large knowledge bases such as YAGO [2]. The system would thus have to be improved to be able to ingest such large datasets. Possibly, new datasets can be thought up, if weaknesses in existing datasets become apparent.

Depending on the depth of the problems and the depth of their solutions, this work has the potential to lead to a research publication. This project is suitable for students who want to go for a PhD afterwards.

From the student, this internship requires familiarity with (or the readiness to become familiar with) Python, knowledge bases, language models, and large-scale data handling. If the internship goes well, it can lead to a PhD thesis (for which funding would have to be secured, e.g., via Hi Paris).

References

[1] Yiwen Peng, Thomas Bonald, and Fabian Suchanek. Flora: Unsupervised knowledge graph alignment by fuzzy logic. In *International Semantic Web Conference (ISWC)*, 2025.

[2] Fabian M. Suchanek, Mehwish Alam, Thomas Bonald, Lihu Chen, Pierre-Henri Paris, and Jules Soria. YAGO 4.5: A Large and Clean Knowledge Base with a Rich Taxonomy. In *SIGIR*, 2024.