

# Knowledge Graph Enrichment using Textual Documents

## General Information

- Supervisors: Mehwish Alam and Yves Roquet
- Starting date: February, 2024
- Internship duration: 6 months
- Location: La Banque Postale Asset Management (LBPAM).

**Keywords:** Knowledge Graphs, Knowledge Graph Enrichment, Human Interaction

## Context

A key player in multi-specialist conviction management and sustainable finance, LBP AM capitalizes on its expertise and that of its subsidiaries, Tocqueville Finance and La Financière de l'Echiquier (LFDE), to offer a wide range of open-ended funds, dedicated solutions and mandates to its customers, institutional investors, insurers, mutual insurers, large corporations and external distributors. LBP AM has five investment divisions: real & private assets, multi-asset & absolute performance, quantitative solutions, equity management via Tocqueville Finance, and, finally, collective & private management via LFDE. On June 30, 2023, LBP AM and its subsidiaries had consolidated assets under management and distribution of €70 billion. With a 100 % SRI approach, LBP AM deploys a model that systematically integrates extra-financial analysis into investment decisions, in order to respond to major societal and environmental challenges.

Within the IT teams, the digital lab department carries out digital projects and works in close collaboration with the business units. To do this, we rely on a variety of technical solutions, including a complex Data platform (MarkLogic), a Robot Process Automation tool (UiPath), data visualization tools, and a Python development environment. Natural Language Processing (NLP) topics are handled via our MarkLogic platform and our Python ecosystem.

## Context of the Internship

The growing number of documents every day poses a challenge for users while exploring and retrieving relevant documents. Often, an overwhelming amount of results is shown

without further knowledge about the context or background of the intended subject or the hidden information about the documents leading to an information overload. Knowledge Graphs (KGs) can act as a potential solution by providing means to structure those documents and allow navigation [1]. KGs represent information in the form of entities and relations, e.g.,  $\langle Paris, is\_capital\_of, France \rangle$ , where *Paris* and *France* represent entities and *is\_capital\_of* represent a relation. However, KGs are not always static, i.e., new information is published every day leading to the necessity of allowing updates to the KG.

This internship focuses on building an interactive platform for supporting evolving KGs based on the verification module allowing an interaction with the human (expert). The existing KG used by LBPAM already provides a mechanism for exploration, however, it does not support the accommodation of new facts which are never seen by the KG. The goal is to allow the updates of this existing KG with the help of facts from the new documents without duplications. The goal is to perform automated detection of the keywords in the textual document and perform matching against the labels of the entities within a KG with the help of textual similarity based on neural language models such as Word2vec [2]. In case of the absence of the entity in the KG the newly extracted keyword will become the candidate for the KG enrichment process. However, this automated extraction approach can lead to irrelevant keywords. A basic interface will be developed to enable human, in this case, expert interaction. This interface will allow the expert to accept and reject the suggestions made by the automated system. The accepted changes will be incorporated within the existing KG.

## Profile Required

- Student of Master 1 or Master 2;
- Ability to work independently and as part of a team;
- Excellent communication and writing skills;
- Good command of French (required) and English.

## Contact Information

Please send your CVs and transcripts to Mehwish Alam (mehwish.alam@telecom-paris.fr) and Yves Roquet (yves.roquet@lbpam.com).

## References

- [1] Silvana Castano, Mattia Falduti, Alfio Ferrara, and Stefano Montanelli. A knowledge-centered framework for exploration and retrieval of legal documents. *Inf. Syst.*, 106:101842, 2022.
- [2] Tomás Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality. In *27th Annual Conference on Neural Information Processing Systems.*, pages 3111–3119, 2013.