PhD: « Sustainability and Explainability through Learning on Large Knowledge Graphs » – 36-months contract

Context:

The Fayol institute of Mines Saint-Étienne and the Laboratory of Informatics, Modeling and System Optimization (LIMOS, UMR 6158) is opening a PhD position in knowledge representation and machine learning, to work at the intersection of digital technologies and sustainable development.

The PhD proposal aims to find alternatives to Large Language Models (LLMs), characterized by a high number of trainable parameters and/or a high number of tokens in their training corpus. Using LLMs thus comes with high energy costs, in both training and inference phases, and a lack of transparency on the generated content. The objective of the PhD is to show that Knowledge Graphs such as Dbpedia, BabelNet or ConceptNet can be a solution to both problems. They have already been used in question answering tasks, despite their notable incompleteness on naive physics (basic spatial-temporal reasoning). The incompleteness of a Knowledge Graph can however be addressed by learning vector representations of the main concepts of the graph (its foundational ontology), whose geometric properties remain semantically interpretable.

Mission:

The objective of the PhD is to develop a method to train vector representations of concepts at low cost and produce a pre-trained language model from DBpedia (or similar). The pre-trained model would then be used for spatial-temporal reasoning in an industrial application, to prove the usefulness of such representations.

Activities:

Research will be conducted through bibliography management, experiment design and execution (on compute clusters), writing and presentation of results. During the PhD, the candidate will have to manage large volumes of data, use compute resources on dedicated CPUs/GPUs and develop code using machine learning libraries such as PyTorch. They will also have to apply lifecycle assessment methods to measure the environmental impact of a computation.

The PhD candidate may also help organize doctoral workshops or summer schools (with invited researchers) and attend the numerous scientific events being organized among members the Fayol institute or LIMOS.

Profile:

Technical skills and curriculum:

- Master's degree or equivalent, in the domains of computer science, data science or applied mathematics
- Prior knowledge in:
    - machine learning and/or natual langage processing

- formal logics and/org Semantic Web
- (large) relational databases and/or graph databases

Other skills:

- written and spoken English (writing of technical reports and oral presentations)
- pratical problem solving
- ability to generalize and formalize (mathematically)
- autonomy, initiative and intellectual curiosity

Hiring conditions:

- contract as per French public sector rules (36 months)
- expected start date: **October 1st, 2024**
- full-time
- based in Saint-Étienne, France

Application:

Applications (resumé, cover letter) must be submitted via RECRUITEE **before May 5th, 2024**.

Further questions:

Any further questions on the position may be addressed to:

Victor Charpenay (co-supervisor), victor.charpenay@emse.fr

Antoine Zimmermann (supervisor), antoine.zimmermann@emse.fr

Julie Jaffre (administrative staff), julie.jaffre@emse.fr, +33 4 77 42 00 17