Internship position at CNAM Paris on Knowledge graph completion leveraging pre-trained language models and GNNs (Level: Master / 3rd year engineering school)

Host ISID and Vertigo teams at Centre d'études et de recherche en informatique et communications (CEDRIC), CNAM Paris

Starting date As soon as possible

Duration 6 months

Gratification Internship gratification according to public organization's grid

Candidate profile
As a minimum requirement, the successful candidate should have:
- A master degree in one or more of the following areas: machine learning, natural language processing, symbolic AI, semantic web.
- Excellent programming skills (Java or Python)
- Excellent command of English
- Experience with machine learning and graphs

How to apply
The application should be formatted as **a single pdf file** and should include:
- A complete and detailed curriculum vitae
- A cover letter
- a report you wrote (ideally an internship or project report)
- The content of M1 and M2 courses and the corresponding grades
- The contact of one referee and a recommandation letter if possible

The pdf file should be sent to nada.mimouni@cnam.fr and fournier@cnam.fr

Keywords machine learning, knowledge graphs, semantic web

Supervision Nada Mimouni (Cnam), Raphaël Fournier-S'niehotta (Cnam), Jean-Claude Moissinac (Télécom Paris)

Context
In the field of cultural heritage, and painting in particular, the management of large collections has become increasingly complex over the years. Heritage data, including aspects such as names, creators, representations and images, have posed significant challenges for curators and researchers alike.
Semantic knowledge graphs have emerged as a promising approach to representing cultural heritage datasets. They provide a structured framework for integrating heterogeneous data sources, enabling comprehensive exploration and reasoning about cultural artworks and their relationships. However, existing knowledge graphs are far from complete in this domain, and creating and populating semantic models for heritage data is a resource-intensive undertaking, requiring substantial human expertise. Knowledge graph completion (KGC) approaches have been proposed to enhance knowledge graphs by completing their

missing connections. In this work, we aim to extend knowledge-graph completion techniques to this specific data domain, by leveraging both pre-trained language models and Graph Neural Networks (GNNs) to facilitate the efficient creation and extension of knowledge graphs.

The work will be organized as follows:
- Related work: analysis of existing approaches on the use of pre-trained language models and GNNs to improve knowledge graph completion (KGC).
- Data: collection and creation of benchmarks to evaluate the models.
- Methodology: definition of a methodology for data preparation and knowledge graph enrichment.
- Interpretation and evaluation: carry out a quantitative assessment of the proposed methods for KGC, based on the created benchmarks, in order to establish their effectiveness in this context. An effort towards explaining these results should be made.

References

- Nada Mimouni, Jean-Claude Moissinac, Anh Tuan Vu. Knowledge Base Completion With Analogical Inference on Context Graphs. Semapro 2019, Sep 2019, Porto, France.
- Nada Mimouni, Jean-Claude Moissinac, and Anh Tuan Vu. "Domain specific knowledge graph embedding for analogical link discovery". International Journal On Advances in Intelligent Systems 13 (2020): 140-150.
- Li, Luoqiu, et al. "Text-guided legal knowledge graph reasoning." Knowledge Graph and Semantic Computing: Knowledge Graph Empowers New Infrastructure Construction: 6th China Conference, CCKS 2021, Guangzhou, China, November 4-7, 2021, Proceedings 6. Springer Singapore, 2021.
- Kolyvakis, Prodromos, Alexandros Kalousis, and Dimitris Kiritsis. "Hyperbolic knowledge graph embeddings for knowledge base completion." The Semantic Web: 17th International Conference, ESWC 2020, Heraklion, Crete, Greece, May 31–June 4, 2020, Proceedings 17. Springer International Publishing, 2020.
- Safavi, Tara, and Danai Koutra. "CoDEx: A Comprehensive Knowledge Graph Completion Benchmark." Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2020.
- Hamann, Felix, et al. "Open-World Knowledge Graph Completion Benchmarks for Knowledge Discovery." Advances and Trends in Artificial Intelligence. From Theory to Practice: 34th International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, IEA/AIE 2021, Kuala Lumpur, Malaysia, July 26–29, 2021, Proceedings, Part II 34. Springer International Publishing, 2021.
- Petroni, Fabio, et al. "Language models as knowledge bases?." arXiv preprint arXiv:1909.01066 (2019).
- Shen, T., Zhang, F., Cheng, J.: A comprehensive overview of knowledge graph completion. Know.-Based Syst. 255(C) (2022). https://doi.org/10.1016/j.knosys.2022.109597.

- Hongcai Xu, Junpeng Bao, and Wenbo Liu. 2023. Double-Branch Multi-Attention based Graph Neural Network for Knowledge Graph Completion. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15257–15271, Toronto, Canada. Association for Computational Linguistics.
- Z. Ye, Y. J. Kumar, G. O. Sing, F. Song and J. Wang, "A Comprehensive Survey of Graph Neural Networks for Knowledge Graphs," in *IEEE Access*, vol. 10, pp. 75729-75741, 2022, doi: 10.1109/ACCESS.2022.3191784. https://ieeexplore.ieee.org/document/9831453