

# Knowledge Graph Construction from Text For Enhanced Retrieval Augmented Generation

6-month internship @ CEA List

## Internship context

Based in Saclay (Essonne), the LIST is one of the two institutes of CEA Tech, the Technological Research Division of the CEA. Dedicated to intelligent digital systems, its mission is to carry out technological developments of excellence on behalf of industrial partners, in order to create value.

Within the LIST, the Laboratory of Textual and Visual Semantic Analysis (LASTI) conducts its research in the field of natural language processing and computer vision to extract, classify and generate information. The laboratory's research themes include learning with few data, trustworthiness and multimodality.

## Missions

The emerging GraphRAG paradigm combines large language models (LLMs) with structured knowledge graphs (KGs) to enhance factual grounding, multi-hop reasoning, and generation interpretability. This approach was notably formalized in the 2024 Microsoft *GraphRAG* paper, which shows the superiority of graph-guided generation over classic dense retrieval in long-context scenarios. However, a major bottleneck in deploying GraphRAG systems is the construction of KGs. Traditional KGs (e.g., DBpedia, Wikidata) are poorly adapted to domain-specific texts. Simultaneously, most knowledge still resides in unstructured natural language (scientific articles, documentation, etc.). To address this, recent works such as Microsoft GraphRAG (2024) and KGEN (2025) have explored using LLMs as general-purpose extractors of graphs from text.

While the community is aware of the quality gap between knowledge graphs built with expert-in-the-loop (domain-specific ontology) versus the use of LLMs, the domain lacks a measurement of how this gap affects the downstream task of graphRAG. This internship aims to address that question and can follow these steps:

- Reproduce and compare LLM-based graph construction pipelines
  - Implement/reuse components from Microsoft GraphRAG and KGEN
  - Apply both methods to domain-specific corpora (e.g., Natural disaster articles).
- Design ontology-guided generation mechanisms
  - Define or reuse a lightweight domain ontology (set of domain entities and relations)
  - Explore : Prompt templating with ontology constraints, post-filtering for ontology compliance, etc.
- Evaluate and compare graph construction quality (extraction metrics, graph metrics, semantic fidelity metrics, etc.)
- Assess downstream utility
  - Use both graphs in a GraphRAG QA pipeline (retrieving subgraphs to guide LLM generation).
  - Compare factual accuracy, latency, and interpretability of outputs with and without ontology guidance.

This internship may be seen as an introduction to research. It may lead to the publishing a scientific paper if the results are convincing. Successful outcomes may also lead to the extension of this work through a PhD thesis, within the broader scope of the LASTI team's AI research.

## Qualifications

- M2 or final-year engineering student
- Knowledge in Natural Language Processing
- Familiarity with prompt engineering, information extraction, and knowledge graphs
- Machine learning skills (deep learning, generative AI...)
- Python proficiency in a deep learning framework (especially PyTorch or Tensorflow)

## Job-related benefits

Joining the CEA List and the LASTI as an intern means:

- Working in one of the most innovative research organizations in the world, addressing societal challenges to build the world of tomorrow
- Discovering a rich ecosystem: privileged connections between the industrial and academic sectors
- Conducting research autonomously and creatively: encouragement to promote results (scientific articles, patents, open-source codes...)
- Benefit from an internal computing infrastructure with more than 300 state-of-the-art GPUs
- Receive a stipend between €1300 and €1400 per month
- Have the opportunity to continue with a PhD or as a research engineer after the internship
- Have the possibility of remote work, receive a 75% reimbursement on public transportation costs, and benefit from the "mobili-jeune" aid to reduce rent costs...

To apply, please send your CV, a cover letter, and the title of the internship to:

[lastirecrute@cea.fr](mailto:lastirecrute@cea.fr)

If you are interested in more than one internship, please indicate your order of preference.