

SPARK: Exploiting LLMs and Knowledge Graphs to Build Structured and Personalized Learning Paths

Intended audience: Master 2 students

Duration: 6 months (proposed dates: 01/02/2026 - 01/08/2026)

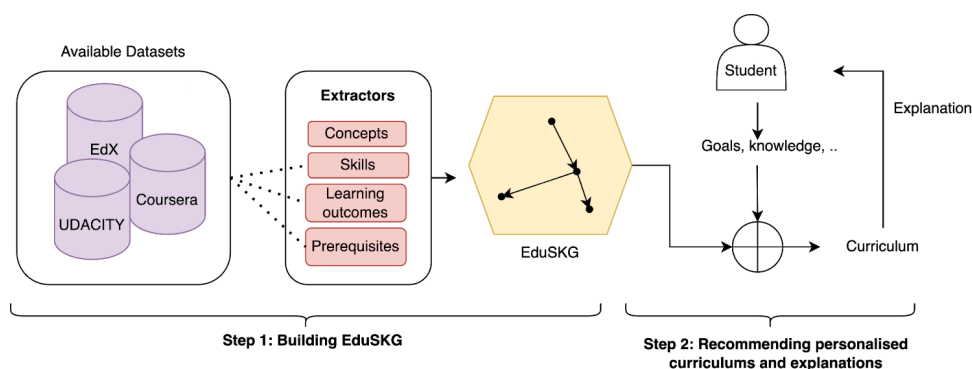
Supervisors: Nadia Yacoubi Ayadi (nadia.yacoubi-ayadi@univ-lyon1.fr), Bruno Yun (bruno.yun@univ-lyon1.fr)

Keywords: Artificial Intelligence, Education, Large Language Models

Gratification: Around 640€/month (approximately 3200€ for 5 months)

Scientific context

This project investigates the use of Large Language Models (LLMs) and knowledge graphs to advance the design of next-generation intelligent tutoring systems (ITS). The focus is on supporting personalised learning, effort regulation, and goal setting by enabling learners to engage more actively in the selection of their learning materials and pathways. Indeed, the rapid expansion of online educational platforms, offering an extensive range of courses and learning pathways, poses a significant challenge for learners in identifying the most relevant and suitable options. While multiple courses already exist, effectively exploiting them for personalised learning requires careful processing and structuring. In this project, there are two main steps.



First, we plan to rely on LLMs and NLP parsing libraries to automatically identify key pedagogical elements from course materials. Specifically, we aim to extract: (i) course learning outcomes (learning objectives that can be categorized according to Bloom's Taxonomy), (ii) prerequisites (essential prior knowledge or skills required before course enrollment), (iii) topics (categorization of the course content based on IEEE taxonomy), and (iv) skills (specific skills that learners are expected to acquire upon completion). These elements will be structured into an Education Structured Knowledge Graph (EdSKG), which will serve as a powerful framework for structuring and leveraging educational resources. For this step, there are multiple challenges to overcome:

1- Heterogeneity of course materials - Course resources differ significantly in modality (e.g., videos, images, HTML), granularity, and pedagogical design. Developing a unified system capable of ingesting and processing such diverse inputs is a non-trivial task.

2. Dataset alignment - Multiple useful datasets available (e.g., UDACITY for learning outcomes, prerequisites, and skills; EDX and COURSERA for course descriptions), yet they are not directly interoperable. Aligning these heterogeneous sources into a unified training framework requires careful mapping and normalization.

3- Lack of annotated datasets for concept extraction - Currently, there is no annotated dataset that directly links pedagogical resources to their relevant concepts. While several general-purpose approaches for key concepts and phrases extraction exist (e.g., KBIR), their effectiveness in this context remains uncertain and requires empirical evaluation.

To address these challenges, we propose to explore several strategies. For the first one, we investigate multimodal feature extraction to encode different modalities into embeddings and normalize them into a shared embedding space or alternatively, transform all material into a common intermediate representation (e.g. text and metadata) using OCR techniques, multimodal LLMs, parsers and automatic speech-recognition, etc.. For the second challenge, we plan to use transfer learning to address dataset alignment by projecting heterogeneous sources into a shared semantic space and imputing missing elements. This will ensure consistent representations across formats and styles, enabling automatic generation of outcomes, skills, or prerequisite links where they are absent. For the last challenge, we will evaluate the extracted concepts by different methods using either human evaluations or LLM juries.

The second step of this project will consist in exploiting EduSKG to automatically connect learners with courses that best match their interests and previous learning experiences. Ultimately, EduSKG will serve as a foundation for our system to recommend curriculums (learning paths), ensure that students receive adaptive guidance aligned with their goals, prior knowledge, and skill mastery. To achieve this, we will draw on automated planning techniques, which allow us to treat curriculum generation as a sequence of courses leading from a learner's current state (skills, knowledge) to a desired goal state (target competencies).

Another objective of this step is to build on recent findings in explainable AI by incorporating visual explanations (e.g., "what-if" scenarios) to help learners understand the impact of different courses on their learning path. By combining learner control with transparent

recommendations, the ITS will encourage students not only to address weaknesses but also to consolidate strengths, fostering more balanced and self-regulated learning trajectories.

References

- Daniel Reales et al: Core Concept Identification in Educational Resources via Knowledge Graphs and Large Language Models. SN Comput. Sci. 5(8): 1029 (2024)
- Gaganpreet Jhajj et al: Educational Knowledge Graph Creation and Augmentation via LLMs. In Proceedings of Intelligent Tutoring Systems (ITS'2024) (2): 292-304.
- Abdessamad Chanaa & Nour-Eddine El Faddouli: Prerequisites-based course recommendation: recommending learning objects using concept prerequisites and metadata matching. Smart Learn. Environ. 11(1): 16 (2024)
- Abu-Rasheed et al: Llm-assisted knowledge graph completion for curriculum and domain modelling in personalized higher education recommendations. In Proceedings of IEEE Global Engineering Education Conference (EDUCON'2025) (pp. 1-5). IEEE.
- Silan Hu & Xiaoning Wang: FOKE: A Personalized and Explainable Education Framework Integrating Foundation Models, Knowledge Graphs, and Prompt Engineering. CoRR abs/2405.03734 (2024)
- Sonal Jain & Jyoti Pareek: Automatic extraction of prerequisites and learning outcome from learning material. Int. J. Metadata Semant. Ontologies 8(2): 145-154 (2013)
- Boubaker Nadia, Kodia Zahra, and Yacoubi Ayadi, Nadia. A Hybrid Approach to Post-hoc Explanations in Recommender Systems: Embedding Meets Semantics. In Proceedings of the International Conference on Cooperative Information Systems (CoopIS'2025) – To Appear.

Who we are looking for

We are looking for a motivated Master's level (Master 2) or final-year engineering student in Computer Science with the following qualifications:

- ✓ French and English proficiency at B2 level
- ✓ Strong skills in imperative programming, Python, and the Git version control system
- ✓ Knowledge in Artificial Intelligence, in particular Reinforcement Learning
- ◆ Knowledge in abstract argumentation is a plus, but not strictly required
- ◆ Proficiency in LaTeX is also a plus

How to apply

Submit your application by email to the supervisors (see above) with the following documents:

- a CV detailing your training experience, as well as personal and/or school projects;
- your grade report for the last 2 years, including temporary grades if the official report is not available;
- a cover letter (non-specific cover letters will not be considered).

Application date: as soon as possible, ideally before end of November 2025

Other details

The internship will take place either at the LIRIS laboratory, located in the Nautibus building (22 avenue Pierre de Coubertin, 69622 Villeurbanne, France). It will be co-supervised by Dr. Nadia Ayadi Ayadi (LIRIS, TWEAK) and Dr. Bruno Yun (LIRIS, SyCoSMA). The student will join the project and have the opportunity to participate in scientific presentations and meetings with other AI researchers from both research teams.