



Internship M2/ Final Year Engineering

Automatic Question Generation in the Context of Learning Environments

5-6 month internship in the LIRIS laboratory (TWEAK team)

Supervisor: Nadia Yacoubi Ayadi

This research internship is part of the TWEAK (LIRIS) team's research program, and aims to consolidate the team's work in the field of Computer Environments for Human Learning (CEHL). The main objective is to embed artificial intelligence in personalized and adaptive e-learning systems, in terms that pedagogical objectives, assessment methods and learning paths are adapted to the learner's unique profile. Setting up an adaptive e-learning system requires the collection and processing of different types of data: text (course material, bibliographic references, textbooks), knowledge bases (ConceptNet or Wikidata). This will require the use of automatic natural language processing, both for "understanding" textual content with NLU (Natural Language Understanding) algorithms/models, and for content generation (adapted course, questions) with NLG (Natural Language Generation) generative algorithms/models. It is within this framework that this internship aims to propose a question generation (QG) model [Mulla & Charpure, 2023] based on a given educational content and auxiliary knowledge sources.

For the year 2024, we plan to recruit a research intern, preferably at M2 or last year engineering level. During the internship, the student will develop, as a first step, an initial exploratory analysis of existing question/answer (QR) datasets in the literature, specific to the field of education (statistics, sources used, themes, question types, context, explanations) [Hadifar, et al, 2023, Hardalov et al., 2020] . In the second stage, we aim to study and experiment different pre-trained models such as BERT, BART and T5 for training AQ and GQ models [Hao et al., 2017]. Finally, we aim to propose an improvement of one (or more) pre-trained model(s) through the infusion of knowledge from open knowledge sources, such as ConceptNet or Wikidata to meet education requirements. For his experiments, the student

will need to draw on existing datasets in order to provide a concrete and comparative assessment of the proposed QR and QG models.

We plan to operationalize and evaluate the approach proposed in this internship by incorporating it into a Multiple Choice Questionnaire (MCQ) tool available in the OPALE application [Jean-Daubias and Quoc Do, 2019] used by students at Université Lyon 1. This embodiment will enable us, through user studies, to have a real assessment of the quality of the questions generated.

Prerequisites:

We are looking for a candidate with knowledge of machine learning (ideally also deep learning and Automatic Natural Language Processing) .

Apply now:

Send an email to Nadia YACOUBI AYADI nadia.yacoubi-ayadi@univ-lyon1.fr with your CV and a cover letter.

Bibliography :

[Hadifar, et al, 2023] A. Hadifar, S. K. Bitew, J. Deleu, C. Develder and T. Demeester, "EduQG: A Multi-Format Multiple-Choice Dataset for the Educational Domain," in *IEEE Access*, vol. 11, pp. 20885-20896, 2023, doi: 10.1109/ACCESS.2023.3248790.

[Hardalov et al., 2020] M. Hardalov, T. Mihaylov, D. Zlatkova, Y. Dinkov, I. Koychev, and P. Nakov. 2020. EXAMS: A Multi-subject High School Examinations Dataset for Cross-lingual and Multilingual Question Answering. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5427–5444, Online. Association for Computational Linguistics.

[Hao et al., 2017] Yanchao Hao, Yuanzhe Zhang, Kang Liu, Shizhu He, Zhanyi Liu, Hua Wu, and Jun Zhao. 2017. An End-to-End Model for Question Answering over Knowledge Base with Cross-Attention Combining Global Knowledge. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 221–231, Vancouver, Canada. Association for Computational Linguistics.

[Jean-Daubias et Quoc Do, 2019] Stéphanie Jean-Daubias & Tuong Phu Quoc Do (2019). « Des usages pédagogiques pour OPALE, appli mobile à destination des usagers du campus LyonTech-la Doua ». Colloque évolutions pédagogiques.

[Mulla & Charpure, 2023] Mulla, Nikahat & Gharpure, Prachi. (2023). Automatic question generation: a review of methodologies, datasets, evaluation metrics, and applications. *Progress in Artificial Intelligence*. 12. 10.1007/s13748-023-00295-9.