Analyzing textbooks using dynamic graphs

Supervisors: Zoltan Miklos (zoltan.miklos@irisa.fr), Mickael Foursov

Keywords: concept hierarchies, evolving graphs, knowledge extraction

Internship of ca. 6 months, from February/Mars 2021.

Research lab IRISA, Rennes, DRUID research team. Given the sanitary context, it is likely that that the internship will be (partially) in tele-work

Reference: **DRUID_2021_01**

Context

Networks are used to model a wide-range of phenomena, in a wide-range of scientific domains, from physic to biology or social sciences. For example, in the case of text, one can build term co-occurrence graphs to represent the co-utilization of words in a document collection. While one often focuses on static graphs, these graphs are often dynamic in real setting.

By analyzing the co-occurrence graph of the terms of a textbook, we hope to get some insights about the structure of the textbook and also the relationships between the concepts presented in the book. In particular we would like to construct these co-occurrence graphs for each section or chapter, and compare them. By analyzing the evolution of these graphs, we would like to understand the partial order of concepts (with respect to specificity or pre-requisites) presented in the book. This understanding could help to enrich textbooks with structured annotations and to help to combine different learning resources.

Internship

A number of techniques have been proposed to extract structured knowledge and to construct knowledge graphs from unstructured text or from semi-structured data [1]. Researchers have tried to reconstruct concept graphs also for educational resources [2]. These graphs often exhibit a hierarchical structure [3,4]. If we analyze these concept graphs for smaller fragments of a document (or a collection of documents) we can gain some understanding in which order the concepts are introduced and how they depend on each other [5,6].

The goal of the internship:

- Extract co-occurrence graphs from fragments (per sections/chapters)
- Analyze the evolution of these graphs
- Identify partial orders between the identified concepts, construct the hierarchical structure of the concepts or to relate the concepts to an existing hierarchy

- Document the work and develop scientific publications

Possibility to continue the work, as a PhD thesis. (If the internship results are satisfactory and if the requested funding is secured.)

Skills

- Basic knowledge of machine learning techniques and deep learning
- Good Python programming skills
- Interest in research, scientific curiosity
- English (a good command, both in reading and in writing)

Bibliography

[1] Xin Luna Dong et al. Multi-modal Information Extraction from Text, Semi-structured, and Tabular Data on the Web. KDD'2020 tutorial
[2] Liu et al. Concept Graph Learning from Educational Data. Journal of Artificial Intelligence Recherch 55 (2016) 1059-1090
[3] Jean-Claude Falmagne and Jean-Paul Doignon. Learning Spaces. Springer, 2010.
[4] Jean-Paul Doignon and Jean-Claude Falmagne. Knowledge Spaces. Springer, 1999.
[5] Valls-Vargas et al. Towards Automatically Extracting Story Graphs from Natural Language Strories. AAAI-17 Workshop
[6] Roy et al. Inferring Concept Prerequisite Relations from Online Educational Resources. IAAI-2019

Internship application

Please send your application documents (CV, motivation letter, results transcripts, etc.) to zoltan.miklos@irisa.fr.   Please put the reference DRUID_2021_01 in the subject line.