# Extracting and evaluating educational concept dependencies implicit in Large Language Models

**Dominik Glandorf**
Matrikelnummer 6007407
dominik.glandorf@student.uni-tuebingen.de

**Anastasiia Alekseeva**
Matrikelnummer 5994775
anastasiia.alekseeva@student.uni-tuebingen.de

GitHub repository: https://github.com/mlcolab/learning-dependencies

## Abstract

## 1 Introduction

Large Language Models are trained on immense corpora of text and have proven to capture included factual information by building on it when performing well in downstream tasks such as question answering. This observation led to the research line of Knowledge Extraction from LLMs (Cohen, 2023). In this work, we focus on the extraction of educational knowledge that is implicitly captured in LLMs and poses a subfield that has rather stagnated in the previous years.

Effective and efficient instruction does not only incorporate what to teach but also how to teach. Guidelines for *instructional sequencing* emphasize the order of instruction. More precisely, prerequisites of educational content should be either known to the student or taught first (**?**). Within educational content, Merill (1983) differentiated facts, concepts, principles, rules, procedures, interpersonal skills, attitudes, and their sole recall from their application. To simplify, we will focus on concepts and their relations. If one concept is a prerequisite of another, we call this relation a concept dependency. For example, to understand the concept of a derivative, having knowledge about the concept of a function will facilitate or even enable learning. When the dependencies are thought of as directed edges between nodes that represent concepts, a concept dependency graph emerges which is a special type of a knowledge graph (**?**). This graph is also called *concept map* in the field of Learning Sciences.

Prerequisites can be inferred from learner behavior by testing their performance after being presented different instructional sequences (Pavlik et al., 2008, Vuong et al., 2011). However, this has the disadvantage of disengaging users with too difficult concepts before teaching easier or necessary ones. Experts usually dispose of the required knowledge about concepts to create concept maps. The high cost of expert knowledge motivates the automated extraction of concept dependencies from appropriate sources.

### 1.1 Related work

**?** defined the prerequisite relation in terms of the consumption of information about concepts. Vuong (2011) if a better graduation rate given prerequisite knowledge is fulfilled. Concepts are often equated with Wikipedia articles (Talukdar and Cohen, 2012; Wang, 2015).

Project Report for *QDS Research Project* 2022/23

## 2 Method

### 2.1 Information sources

For preprocessing the textbooks we used Wikifier (**?**).

### 2.2 Structured information extraction

## 3 Results

## 4 Discussion