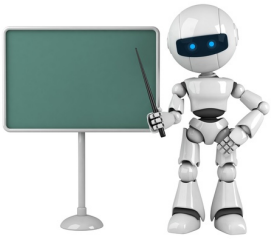
ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Distributed Computing

Prof. R. Wattenhofer

Automated Exercise Generation



Introduction

With the rapid pace of discoveries and the large diversity of topics, it is not always possible to find an up-to-date course about a subject of interest. Even if a course exists it is unlikely personalized to students existing knowledge and interest. Some known concepts might be repeated while other interesting ones might be missing. The large number of man-hours needed to build and keep a course up to date is a major bottleneck for personalized teaching. Solving this bottleneck is only going to become more relevant as automated teaching assistants are more widely adopted in education.

In this thesis, we try to develop a prototype for automated, personalized exercise generation. Based on recent advancements in Deep Learning and Reinforcement Learning, we aim to build an algorithm capable of automating the creation of exercises and keeping them up-to-date by scraping the web and aggregating the relevant information into a personalized course.

Project Description

The project aims at designing a pipeline that, given a topic and user interest, can output exercises without the need to specify information sources. As a result of autonomously searching for relevant information, the exercises can be generated and updated without human intervention. The exercises are presented in the form of flashcards containing question and answer pairs. It will extend a previous project titled “Automated Exercises Coache” which provides the environment in which generated flashcards can be studied. The goal of the project is to provide students with personalized exercises about a subject of interest and updating those automatically from publicly accessible information on the web.

Detailed Project Outline

The following primary tasks are seen as essential to realize our ambition. Each task is given a rough estimate for the expected time allocation on the right. We are using the generic term "Knowledge Concept" to refer to a unit of information or skill.[1] The terminology is taken from S. Pandey and G. Karypis [1]

- **Web-scraper:** Build a web-scraper capable of searching for websites about a given topic and extract the relevant text. Relevant text is to be understood as text that is not otherwise used for navigation or advertisement. (⋆⋆)
- **Aggregation & Filtering:** Develop a method to aggregate the scraped texts and filter out false positives. The resulting representation should be designed to allow for the discovery of relationships between Knowledge Concepts. (⋆⋆⋆)
- **Extraction:** Using the representation from task 2, devise a method to select the Knowledge Concepts that should be processed into flashcards. As the selection criterion may be very subjective, an interface is to be created to allow the user to overwrite the selection if necessary. (⋆⋆⋆)
- **Question generation:** Process the selected Knowledge Concepts into flashcards. This requires the generation of suitable Question and Answer pairs. Using the representation from task 2, further enhance each flashcard with information about the necessary context to understand the question. (⋆⋆)
- **Testing:** Using the existing flashcard environment, test the pipeline in a user study. (⋆⋆⋆⋆)

Extensions

- **Automatic course construction:** Using the relationship structure between flashcards, create a partial order in which the cards should be studied. If context information is needed provide a short text or a link to a website where the user can read more about the question. (⋆⋆)

References

[1] S. Pandey and G. Karypis, "A self-attentive model for knowledge tracing," 2019.

[1] An example for a unit of information is: "The ETH Zurich was established in 1855". On the other hand, the knowledge on how to subtract two numbers is considered a skill as it also requires applying the knowledge to a new situation.