Building a retrieval system using a large-scale language model

Background

The emergence of large-scale language models (LLMs) has led to significant developments in natural language processing. In particular, advances in understanding questions and fluently generating answers with LLMs have been remarkable. However, interactive sessions with LLMs in business are not easy, as it requires judging the truth or falsehood of seemingly correct answers. It is also known that LLMs are not good at answering questions regarding local information that is not generally known to the public. LLMs can be trained with local information but costs are high. Also, it is difficult to dramatically improve LLM performance from unstructured information retrieval. Therefore, a system configuration that obtains answers by supplementing information to an LLM from a database of non-public, local information, is effective.

Goal

We will build a system that receives questions in natural text, searches for appropriate data in a local database, and generates answers using an LLM. Considering methods for evaluating and ranking of search results and prompts for the LLM, the system will improve on current solutions to generate relevant answers.

Subjects

1.1. Retrieval processing using LLMs.
1.2. Method for creating search queries from natural text necessary for retrieving supplemental local information for feeding into LLMs.
1.3. Methods for evaluating and ranking search results.
2.1. Interactive, web-based interface development.
2.2. Design of prompts for combining search results and the LLM.
2.3. Display search results.

Qualifications

- Python, Linux, and Web app development.
- Knowledge of python libraries, especially Transformers and LangChain.
- Good communication skills in English and, preferably, also basic Japanese language skills.

Contact information

This project is hosted by SEL. Students are invited to apply for a scholarship from SEL through the Sweden-Japan Foundation from 2024-03-15. For more information, please visit the SEL web site or contact the SEL public relations team at info@sel.co.jp.