

Title

Using Machine Learning Tools in the Cloud: Experience Gained from the Ask4Summary research project

Abstract

Ask4Summary (<https://ask4summary.vipresearch.ca/>) is a research that has a system periodically running backend services to process text-based content (e.g., a course's learning materials and the CORD-19 dataset). The CORD-19 dataset includes a growing number of academic articles regarding Coronaviruses; at present, there are more than 717,000 full text articles in the CORD-19 dataset. When Ask4Summary processes these text-based content, it uses Natural Language Processing (NLP) techniques that include tokenization, n-grams extraction, and part-of-speech (PoS) tagging. It then identifies the keywords from a user's question and uses cosine similarity to summarize the associated content and present to the user. When the Ask4Summary is fed with course materials like PDFs, Word and Powerpoint documents, it can provide summaries for students' questions related to the course. Currently Ask4Summary can serve users via three ways: web system, Moodle plugin, and chatbot. In this talk, I would also like to share some of our experiences and insights from using Amazon's Comprehend Keyphrase Extraction and Syntax Analysis APIs in the Ask4Summary research project.