THE NATION THAILAND

TECH

WabiQA: A Wikipedia-Based Thai Question-Answering System

Mahidol University has made it a priority to conduct high-quality research to meet the needs of a rapidly changing world and respond to both Thailands and the worlds problems. Artificial Intelligence (AI) technology has become more prevalent in everyday life than ever.

As a significant application, AI technology is often used to facilitate the discovery of knowledge in large databases such as "Wikipedia", a website-based free multilingual encyclopedia, hosted by the non-profit Wikimedia Foundation. Such a collaborative encyclopedia houses over 35 million articles, from which the AI technology can assist in automatically retrieving answers to natural-language questions. With such capability, users no longer have to waste time searching for the answers on their own.

WabiQA is a novel system for automatically answering questions in the Thai language utilizing the Thai Wikipedia articles as the knowledge source. The system was developed by a research team at the Faculty of Information and Communication Technology, Mahidol University (ICT Mahidol). WabiQA takes questions in the natural Thai language, such as "When was the Faculty of ICT, Mahidol University established?" as input, then displays the answer, e.g., "20 May 2009," along with the specific document in which the answer is found. However, all answers must be in the (Thai language) Wikipedia database only.

WabiQA implements the BM25F-based retriever to identify articles on Wikipedia that are most likely to contain the answer or are related to a given question. Then, a Bi-Directional Long-Short Term Memory (BiLSTM) model is applied to read the Thai article and locate candidate answers. Lastly, the Attention layers are used as general answer predictors to quantify the confidence that the target text is the answer to the input question. The test result showed that the WabiQA was able to reduce search time by 97.81 percent. Furthermore, the research team also developed a prototype mobile application that aims to facilitate Thai users with visual impairments using voice-to-speech technology and an intelligent question-answer categorization.

Dr. Thanapon Noraset, the project advisor, said, "This research is a collaboration among the Faculty of ICT's research team, students, and partners, that aims to develop artificial intelligence research and innovations, driven by the technological capabilities of the Thai people, to address difficult real-world problems. Personally, I am impressed by the ability of Mahidol students to conduct such world-class research. In this study, we applied advanced artificial intelligence technology to learn and understand the Thai context. Furthermore, this technology can also be used to perform an automated analysis of large and heterogeneous data composed in Thai, such as searching and summarizing social media opinions on products, events, or policies."

This work was published in the Information Processing & Management Journal, a leading international academic journal ranked Quartile 1 (Q1) in Information Systems. Moreover, the WabiQA also won the first prize award from Thailand's 21st National Software Contest 2019 (NSC) under the category "Question-Answering Program from Thai Wikipedia." Currently, the research team is working to improve the WabiQA so that it is more effective and suitable for usage.

This research demonstrated the potential of advanced artificial intelligence in understanding questions and finding answers in the Thai language and revealed several potential future research directions and real-world applications. In particular, the research team would like to explore how to apply a similar approach to create an AI system that can answer questions for a specific domain, such as legal documents, official announcements, and insurance policies. Additionally, the research team wishes to investigate and develop unique strategies for utilizing data from multilingual sources. This should improve the system's accuracy while also expanding its coverage in additional languages.

Published : December 31, 2021

By : THE NATION