THESIS INFORMATION

Title: **MACHINE READING COMPREHENSION FOR VIETNAMESE TEXTS**

Major Code: **Computer Science**

Code: **9480101**

PhD Student: **NGUYEN VAN KIET**

Advisors: **Assoc. Prof. PhD. NGUYEN LUU THUY NGAN**
and PhD. NGUYEN GIA TUAN ANH

University: **UNIVERSITY OF INFORMATION TECHNOLOGY, VNU-HCM**

ABSTRACT

Recently, Machine Reading Comprehension is a challenging task of natural language processing, attracting many researchers in Artificial Intelligence. This PhD dissertation focuses on building high-quality corpora, developing machine reading comprehension models which are integrated into retriever-reader question aswering models for Vietnamese texts. This PhD dissertation is characterized by three main contributions, which are expounded upon in the subsequent sections.

1) *Creating corpora and evaluating machine reading comprehension models on Vietnamese texts*: In the first contribution, the PhD dissertation focuses on building and developing corpora for Vietnamese (as a low-resource language for natural language processing and artificial intelligence). Vietnamese machine reading comprehension corpora on Wikipedia-based and healthcare news-based texts are created to promote the development and evaluation of machine reading comprehension models and retriever-reader question answering models based on deep learning architectures and transformer-based language models. The state-of-the-art methods are evaluated on these datasets. From the experience of automatic reading comprehension on Vietnamese corpora, a challenging Vietnamese dataset with unanswerable questions was introduced at VLSP 2021 for evaluating machine reading comprehension models.

2) *Proposing machine reading comprehension models integrated with evidence extraction on Vietnamese texts*: In the second contribution, inheriting from the first evaluation results obtained on our machine reading comprehension corpora in the first contribution, the PhD dissertation presents automatic reading comprehension models integrated with evidence extraction and modern language models using transformer architecture. In addition, he researches and expands the evaluation and comparison with many advanced reading comprehension models to understand the proposed reading comprehension model on many different datasets.

3) *Proposing retriever-reader question answering models integrated with machine reading comprehension techniques for Vietnamese texts*: In third contribution, inheriting from the first evaluation results on our machine reading comprehension corpora in the first contribution and the findings on our machine reading comprehension model in second contribution, the PhD dissertation presented new Vietnamese retriever-reader question answering systems using evidence extraction and modern language models using transformer architecture.

During the completion of the thesis, the PhD student published seven scientific articles or papers, of which: four articles were published in prestigious journals (three SCIE-indexed journals and a national journal) and three papers were published at Scopus-indexed international conferences in computer science.

<div align="center">

Advisors **PhD Student**

NGUYEN LUU THUY NGAN   NGUYEN GIA TUAN ANH   NGUYEN VAN KIET

</div>