VIETNAM NATIONAL UNIVERSITY HO CHI MINH CITY
HO CHI MINH CITY UNIVERSITY OF TECHNOLOGY
FACULTY OF COMPUTER SCIENCE AND ENGINEERING

REPORT

CAPSTONE PROJECT

DEVELOPING AI CHATBOT APPLICATION: QUESTION REFORMULATION PROBLEM

MAJOR: COMPUTER SCIENCE

COUNCIL	: **COMPUTER SCIENCE - 01 CLC**
INSTRUCTORS	: **QUẢN THÀNH THƠ**
	: **BÙI CÔNG TUẤN**
SECRETARY	: **NGUYỄN QUANG ĐỨC**
REVIEWER	: **NGUYỄN ĐỨC DŨNG**

—o0o—

STUDENT : **ĐẶNG CÔNG KHANH - 2053105**

HO CHI MINH CITY, MAY 2024

VIETNAM NATIONAL UNIVERSITY HO CHI MINH CITY
HO CHI MINH CITY UNIVERSITY OF TECHNOLOGY
FACULTY OF COMPUTER SCIENCE AND ENGINEERING

REPORT

CAPSTONE PROJECT

DEVELOPING AI CHATBOT APPLICATION: QUESTION REFORMULATION PROBLEM

MAJOR: COMPUTER SCIENCE

COUNCIL	: **COMPUTER SCIENCE - 01 CLC**
INSTRUCTORS	: **QUẢN THÀNH THƠ**
	: **BÙI CÔNG TUẤN**
SECRETARY	: **NGUYỄN QUANG ĐỨC**
REVIEWER	: **NGUYỄN ĐỨC DŨNG**

—o0o—

STUDENT : **ĐẶNG CÔNG KHANH - 2053105**

HO CHI MINH CITY, MAY 2024

Declaration of Authenticity

The author - Đặng Công Khanh, declares that this Capstone Project was composed and implemented entirely by myself under the guidance and supervision of Assoc. Prof. Quản Thành Thơ and Mr. Bùi Công Tuấn at the Faculty of Computer Science and Engineering, Vietnam National University - Ho Chi Minh City University of Technology.

In the process of researching and implementing this Specialized Project, the author had referenced multiple previous studies from other authors. All of them have been fully and clearly stated in the References part.

This Specialized Project has not been published in any form under any circumstances.

The author,
Đặng Công Khanh

Acknowledgement

First and foremost, the author would like to express the most sincere gratitude to Associate Professor, Ph.D. Quản Thành Thơ and Mr. Bùi Công Tuấn. Thank you for allocating your valuable time to guide, give advice, orientate, and encourage me throughout the outline and implementation phase of this Specialized Project.

The author would also like to thank Mr. Phạm Quốc Nguyên and Mr. Lê Minh Khôi for the advice and questions on the aspects which the author did not think about carefully or did not know that was needed. Your questions and advice were a great help.

Next, the author is grateful that the lecturers at the Faculty of Computer Science and Engineering, Vietnam National University - Ho Chi Minh City University of Technology have delivered us sufficient fundamental knowledge, which serves as the springboard so that the author can compose this specialized project and contribute to Vietnamese Computer Science field.

In addition, the author would like to cheer him selves for the efforts when full-time study and project composition happen simultaneously. Although making the first steps was not easy, the author did not give up. By constructing a sensible plan and putting in the best efforts, the author were able to keep the pace and finish the project just in time.

Last but not least, the author would like to thank his family for always supporting him on the journey at university. My parents' diligence has become the largest source of motivation for him to work hard. Furthermore, seniors at the authors company also conscientiously guide and train the author in a professional working environment when taking part in practical application projects, which really shapes the individual thoughts and behaviors to be better.

The author,

Đặng Công Khanh

Abstraction

Artificial Intelligence (AI) applications, particularly chatbots, have become integral in modern society, driven by the need for efficient customer service solutions. In the wake of successful introductions like ChatGPT, many companies are actively investing in AI applications to align with customer preferences and market trends.

During this 4.0 revolution, to extend their popularity to more customers, many organizations have gone online by both making their accounts on social networks and making their organization's website. These accounts and websites usually have admins and mods to help or answer customers' questions anytime they are needed. For these admins and mods to be able to answer the questions, they have to be trained with knowledge about the company's production and spend time practicing answering questions. Yet not every employee performs the same, and they cannot answer questions all the time, so they have to take shifts. To reduce the cost of training, hiring, and performance, chatbot is a solution.

However, the adoption of chatbots presents challenges, particularly in linguistic diversity and accuracy. While numerous models cater to English, there's a noticeable scarcity of effective solutions for languages like Vietnamese. Moreover, inherent ambiguities in user queries often lead to incorrect responses, undermining user satisfaction. To address these challenges, one promising approach is the utilization of question reformulation methods. By refining user queries, these methods enhance the chatbot's comprehension, thereby improving information extraction and response accuracy.

The objective is to find a good reformulation method to help the model better understand the question, therefore increasing the confidence when extracting information to answer and ensuring more satisfactory answers to the user.

Contents

List of Tables

List of Figures

1

Introduction

In Chapter 1, the author will present an overview of the urgency of the topic, objectives, and scope of the project's research. The report outline will also be presented.

Contents

1.1 Problem statement

The advancement of artificial intelligence (AI) technologies, particularly in the domain of natural language processing (NLP), has revolutionized human-computer interactions. AI-powered chatbots have emerged as strong tools in various domains, facilitating efficient communication between users and systems. However, despite their widespread adoption, challenges persist in achieving powerful understanding and interpretation of user queries. Addressing this challenge requires innovative approaches in NLP research to enhance chatbots' ability to analyze and respond accurately to user input.

Because people use language in many different ways, it can be hard for chatbots to understand questions, and they often end up giving answers that aren't quite right. This error can happens for many reasons (e.g. the user did not express their intent clearly, ...). One reason that this happens is people sometimes don't finish their sentences, leaving out words or phrases because they think the chatbot will understand what they mean without having to repeat everything. They do this to make the conversation faster and avoid saying the same things over and over again. For example, someone might ask, "How do I get to that market?" without saying which market they mean. To understand what they're asking, the chatbot has to remember what was said before, like if they previously asked about a market called "Ha Dong". Then, the chatbot can figure out what they mean by "that" market.

But even if the chatbot knows what was said before, it's still hard to understand the current question completely. The chatbot has to connect the old information to the new question and figure out which details to add to make sense of it. Choosing the right details to add isn't easy, because the chatbot has to make the question clearer and understands the intention of the user. So, dealing with unknown intent sentences in questions is a tough problem for chatbots. But by finding better ways to understand and respond to these kinds of questions will make chatbots easier to use and more helpful in conversations with people.

1.2 Goals

This project is a part of a bigger plan to make a question-answering chatbot that helps university students to communicate with in Vietnamese language. The author aims to make it easier for students to get the information they need quickly and easily.

The main focus is on making the chatbot better at understanding questions from students. This can be done by getting it to look at the questions before answering them. This way, the chatbot can understand what the student is asking more accurately.

Basically, the goal is to make sure that when students ask questions, the chatbot understands them well and gives helpful answers. This can be achieved by changing the questions a bit to make them clearer if needed, but the important parts will remain unchanged. This will make it easier for students to use the chatbot and get the information they're looking for about university rules in Vietnamese, making the question less ambiguous while preserving the semantics of the question.

1.3 Scope

The project will involve the design and implementation of a reformulation model that contributes to the field of question reformulation as follows:

- Question Reformulation Model: The primary focus will be on developing a Vietnamese language question reformulation model. This model will employ techniques such as conference and ellipsis resolution to enhance the clarity and precision of user inquiries, thus improving the effectiveness of the chatbot system.

- Preprocessing Agent: the preprocessing agent tasked with optimizing user questions before they enter the chatbot. This agent will streamline the question processing workflow, making inquiries more understandable and easier to answer. By optimizing the input data, the overall functionality and user experience of the chatbot system is aimed to be enhanced.

1.4 Thesis structure

There are six chapters in this project's proposal:

- Chapter 1 outlines the problems, goals, scope and thesis structure of the project.
- Chapter 2 focuses on providing a literature review relevant to the project, covering topics such as Natural Language Processing and Question Reformulation.
- Chapter 3 discusses related works and recent developments in Question Reformulation.
- Chapter 4 details the proposed approaches and the flow of the project.
- Chapter 5 goes into the experiment setups, the results and analysis.
- Chapter 6 provides a summary of the entire project and addresses potential improvements and challenges encountered by the author during the course of the thesis.

2

Background Knowledge

In this chapter, the author will discuss the essential knowledge for the project, including various approaches and the chosen solution. The author will delve into the advantages and disadvantages of each approach, ultimately highlighting the rationale behind the chosen solution. This chapter will provide relevant knowledge for Question Reformulation.

Contents

Question reformulation is a significant area of study within the fields of natural language processing (NLP). It involves modifying the phrasing of a question while preserving its original intent. This process can enhance the performance of various applications, including search engines, conversational agents, and question-answering systems. Understanding the theoretical foundations, methodologies, and applications of question reformulation is essential for developing effective and user-friendly systems.

2.1 Reformulation in Natural Language Processing (NLP)

In the context of Natural Language Processing (NLP), reformulation refers to the process of modifying or rephrasing text or queries to improve their clarity, precision, or relevance for downstream tasks.

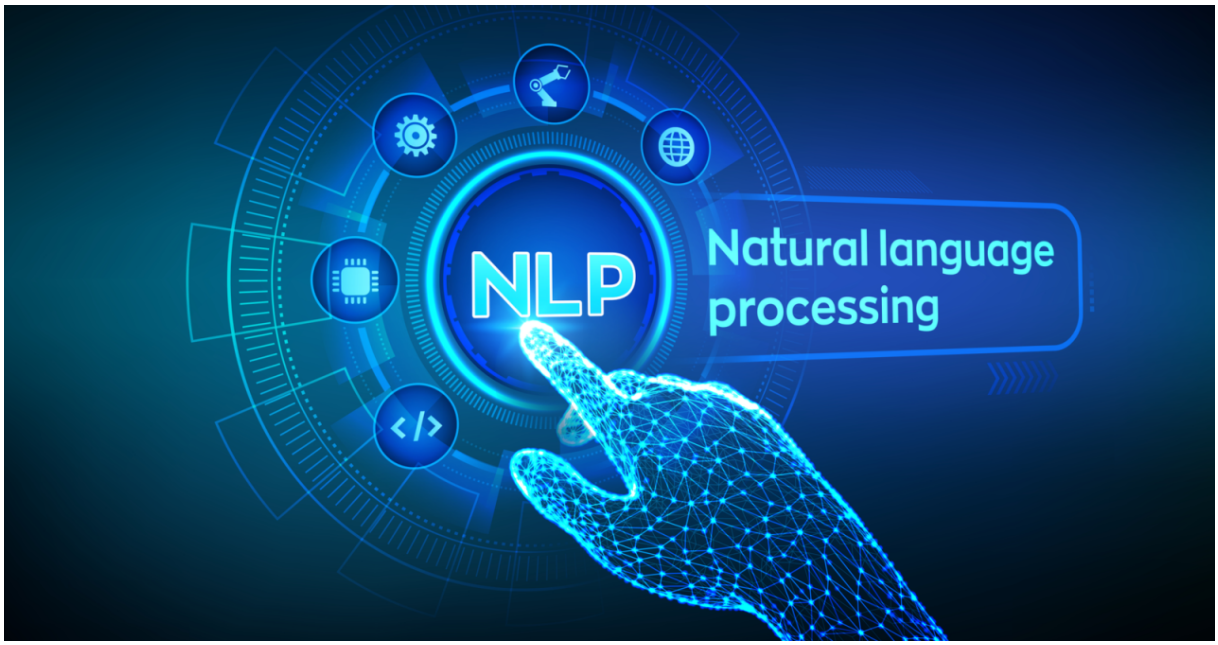

Figure 2.1: Natural language processing.

Recent research has seen the integration of reformulation techniques with other NLP tasks like paraphrase detection, semantic parsing, and context-aware question answering. These integrated approaches aim to develop more adaptable and powerful systems capable of effectively processing a wide range of user inputs. By combining multiple NLP tasks, these systems can achieve higher accuracy and better user experience.

Within the broader scope of reformulation, specific challenges such as coreference resolution and ellipsis resolution have garnered considerable attention. Coreference resolution involves identifying and linking pronouns and other referring expressions to their correct antecedents, which is vital for maintaining coherence in reformulated sentences. Without resolving coreferences, a reformulated query might still retain ambiguities that can confuse downstream processing. Ellipsis resolution, on the other hand, focuses on interpreting and filling in omitted information based on contextual clues. Both tasks are essential for producing complete and contextually accurate reformulations, which can significantly improve the effectiveness of NLP systems.

Reformulation techniques also play a crucial role in enhancing the performance of machine translation systems. By reformulating input text to be clearer and more grammatically correct, translation models can produce more accurate translations. This is particularly important in cases where the source language contains idiomatic expressions, ambiguous pronouns, or culturally specific references. Reformulation helps to standardize the input, making it easier for translation models to generate high-quality outputs that preserve the original meaning.

Overall, reformulation is a fundamental aspect of NLP that contributes to the adaptability of various language processing applications. By continuously evolving and integrating reformulation techniques with other NLP tasks, researchers and developers can create more sophisticated and user-friendly systems. These advancements help bridge the gap between human language variability and the rigid requirements of computational models, leading to more effective human-computer interactions.

2.2 Types of Reformulation

Reformulation techniques aim to enhance text or queries in various ways. Below are some methods of reformulation that the author had found.

2.2.1 Query Expansion

Query Expansion: adding additional terms or synonyms to a query to retrieve more relevant information.

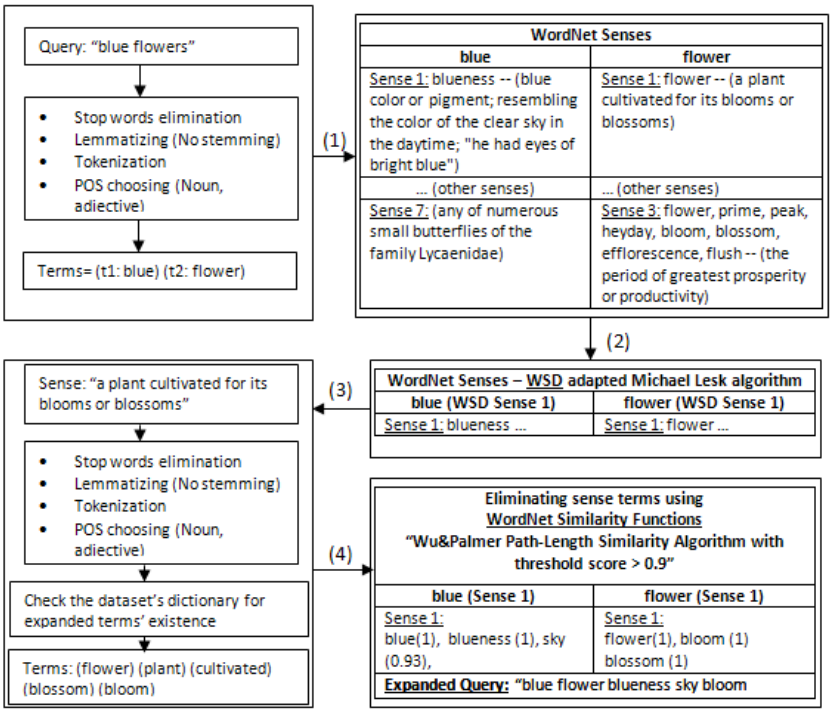

Figure 2.2: Query Expansion Example. (Source: [12])

The figure illustrates the process of query expansion, a technique used in question reformulation to improve search results. By taking an initial user query, the system performs several steps to broaden its scope and identify related terms. Stop words, like "the" or "is," are eliminated first.

Next, words are reduced to their base form and separated into individual units (tokenization). Part-of-speech tagging assigns a grammatical category (noun, verb, etc.) to each word.

Finally, Word Sense Disambiguation (WSD) tackles words with multiple meanings, ensuring the intended meaning is used. By consulting an expanded terms dictionary and employing a path length similarity algorithm, the system finds related terms that can enhance the search. This process effectively reformulates the user's question into a more

comprehensive search strategy, potentially retrieving a wider range of relevant information.

2.2.2 Sentence Simplification

Sentence simplification: restructuring complex sentences into simpler, easier-to-understanding forms without altering the original meaning. This aids in improving readability and comprehension.

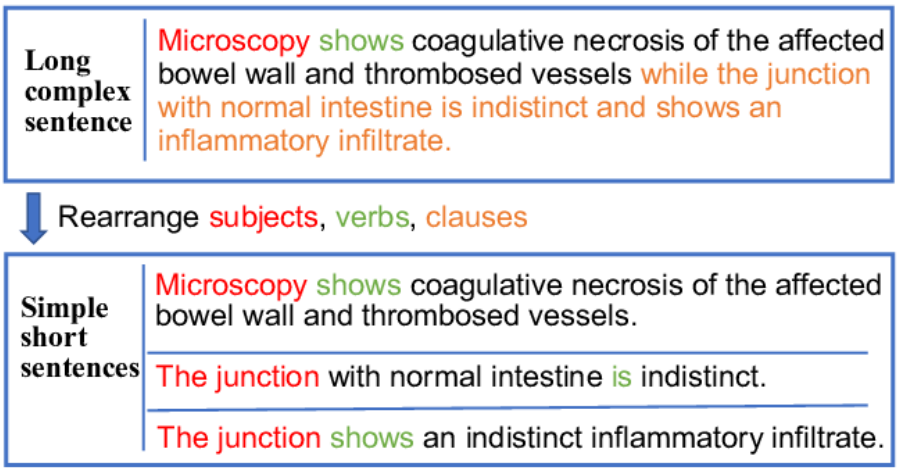

Figure 2.3: Sentence Simplification Example. (Source:[13])

Looking at the figure, we can see that the original sentence is a complex one, containing multiple clauses that describe the findings of a microscopic examination of a bowel tissue sample. This technique rearrange subjects, verbs and clauses in a proper way so that a long and complex sentence breaks the sentence down into three simplified version that is shorter and easier to understand.

Reading and trying to understand a long, complex sentence can be a nightmare, and this happens quite often. In this example, to make the explanation of microscopy clearer, we can divide the complex sentence into three simpler sentences: first, defining the concept of microscopy; second, stating that the junction with normal intestine is indistinct; and third, noting that the junction shows an indistinct inflammatory infiltrate. This approach creates short, simple sentences that are much easier to handle.

2.2.3 Paraphrasing

Paraphrasing: expressing the same meaning using different words or sentence structures. Paraphrasing can be a great help in text summarization, machine translation, and question answering.

Source text	Paraphrase
The need for investors to earn a commercial return may put upward pressure on prices	The need for profit is likely to *push up* prices

Figure 2.4: Paraphrasing Example. (Source: [14])

Paraphrasing is an essential skill that enhances clarity and understanding, particularly in complex or technical texts. Take, for instance, the source text: "the need for investors to earn a commercial return may put upward pressure on prices." This can be paraphrased as: "the need for profit is likely to push up prices." The original text uses more technical terms like "commercial return" and "put upward pressure," which may not be easily understood by everyone, including computers. By rephrasing it to "the need for profit" and "push up prices," the sentence becomes simpler and more direct.

This reformulation helps in breaking down complex ideas into more digestible pieces, making the information more accessible. Paraphrasing is particularly useful in fields like education, where the goal is to convey information clearly and effectively, and in computer processing, where simpler sentences are easier to analyze and process. Thus, paraphrasing not only aids in human comprehension but also enhances the efficiency of computational understanding and data processing.

2.2.4 Ellipsis Resolution

Ellipsis resolution: a linguistic phenomenon in which parts of a sentence are omitted, and have to be retrieved from discourse or real-world context.

Ellipsis is a form of anaphora (besides coreference) that often functions to reduce redundancy in language and improve discourse cohesion (Menzel 2017; Mitkov 1999). Languages provide various mechanisms to elide information, based on which different ellipses are defined in linguistics.

Ellipses are not very frequent in text[10] but for improving the accuracy of Natural Language Processing (NLP) systems that handle data with ellipses, they are important (Zhang et al. 2019; Dean, Cheung, and Precup 2016).[11]

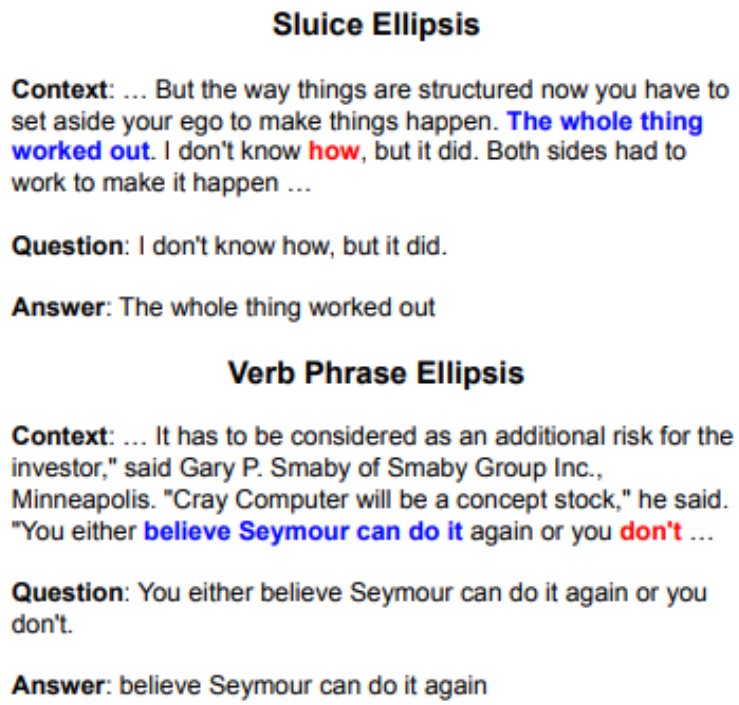

Figure 2.5: Ellipsis resolution example. (Source:[16])

The image illustrates two types of ellipsis: Sluice Ellipsis and Verb Phrase Ellipsis. In Sluice Ellipsis, a question like "I don't know how, but it did" is used, where the context clarifies the answer: "The whole thing worked out." The ellipsis omits the repeated information for brevity. In Verb Phrase Ellipsis, the sentence "You either believe Sey-

mour can do it again or you don’t" is used, and the context provides the answer: "believe Seymour can do it again." These examples demonstrate how ellipsis can streamline communication by omitting redundant information, relying on the context to fill in the gaps.

2.2.5 Coreference Resolution

Coreference resolution: a natural language processing task that identifies when different expressions in a text refer to the same entity. For example, in the sentences "Alice went to the store. She bought some milk," coreference resolution recognizes that "Alice" and "She" refer to the same person.

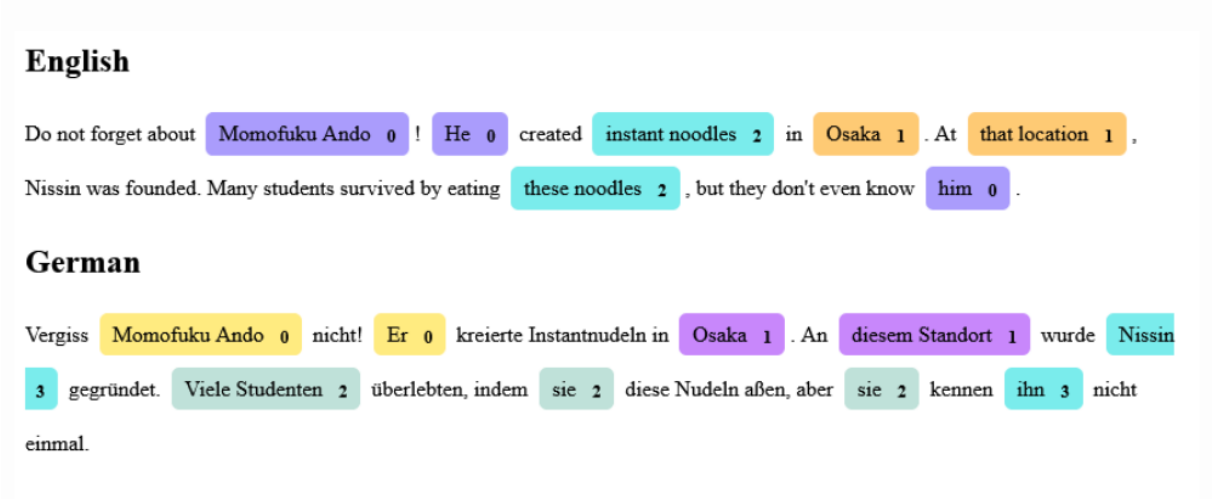

Figure 2.6: Coreference resolution example. (Source:[15])

The figure illustrates coreference resolution in both English and German texts, where the task is to identify and link expressions that refer to the same entity. For example, "Momofuku Ando" is mentioned initially, and subsequent references like "He" and "him" are linked back to this entity. Similarly, "instant noodles" and "these noodles" are identified as the same entity.

In the English text, "Momofuku Ando" is the antecedent for "He" and "him," while "instant noodles" corresponds to "these noodles." The locations "Osaka" and "that location" are also connected. The German text mirrors this structure, with "Momofuku

Ando" being linked to "Er" and "ihn," and "Instantnudeln" to "diese Nudeln."

Coreference resolution plays a crucial role in understanding and reformulating text. By identifying these links, it ensures coherence and consistency, making the text more comprehensible. When translating or summarizing, coreference resolution helps maintain the correct references, avoiding confusion about who or what is being discussed. For instance, without coreference resolution, it might be unclear that "He" refers to "Momofuku Ando" and not another male figure.

This process is very essential for natural language processing applications like chatbots and automated summarization tools. By understanding the relationships between different parts of a text, these systems can generate more accurate and coherent outputs, improving user interaction and information retrieval. In summary, coreference resolution enhances text clarity and coherence, crucial for both human understanding and computational text processing.

2.2.6 Question Reformulation

Question reformulation: the process of rephrasing a question to improve clarity, broaden its scope, narrow its focus, or make it easier to understand. This technique is often used to ensure that the question is interpreted correctly, to obtain more precise answers, or to address the question from different angles. It is commonly applied in search engines, natural language processing, and educational contexts to enhance the effectiveness of information retrieval and comprehension.

Question Reformulation technique is often used in information retrieval tasks, and is usually seen in question-answering models to:

- Refine the question to improve search results: by changing the formation to make the question easier to understand or easier to extract the information, therefore retrieving more relevant and precise results.

- Make better communication between user and chatbot by overcoming the language

ambiguity, complex sentence structure of the question, or handling synonyms, word variations, or different ways of user asking the same question.

- Assist in some other tasks by altering the wording or structure of the initial query.

Q: What did Gaston do after the world series?
History: …
Q: Where did he go in 2001?
A: In 2002, he was hired by the Jays as special assistant to president and chief executive officer Paul Godfrey.
Q: How long did he stay there? **R:** How long did Gaston stay at the Jays? **Ref:** How long did Gaston stay at the Jays?

Figure 2.7: Question reformulation example. (Source:[2])

The provided figure demonstrate question reformulation, a crucial technique in natural language processing and information retrieval. In the given scenario, the original question "What did Gaston do after the world series?" is followed by the question "Where did he go in 2001?" The answer specifies that "In 2002, he was hired by the Jays as special assistant to president and chief executive officer Paul Godfrey."

Following this, the next question, "How long did he stay there?" is ambiguous due to the unclear reference of "he" and "there." To resolve this, the question is reformulated as "How long did Gaston stay at the Jays?" which clarifies both the subject and the location being inquired about. The reformulated version removes ambiguity and ensures that the question directly refers to Gaston's tenure with the Jays.

Question reformulation is one of the best techniques for several reasons. Firstly, it enhances clarity by eliminating vague pronouns and unspecified locations, making the question straightforward. Secondly, it improves the precision of information retrieval by specifying the exact subject and context, which is crucial for accurate responses. Lastly, it ensures that the questions are more user-friendly and easier to understand, which is beneficial in both human communication and automated systems.

This technique is especially valuable in automated systems like chatbots, search engines, and question-answering systems where ensuring precise and relevant responses is paramount. By refining questions to be more explicit and clear, these systems can provide more accurate and useful answers, enhancing the overall user experience and the effectiveness of the information retrieval process.

3

Related works

In Chapter 3, we will explore previous research relevant to Question Reformulation, examining existing approaches and considering their strengths and limitations.

Contents

This project primarily builds upon the methodologies and insights presented in the paper "Action-Based Network for Conversational Question Reformulation" by Zheyu Ye, Jiangning Liu, Qian Yu, and Jianxun Ju [2]. By leveraging their innovative approach to question reformulation using action-based networks and reinforcement learning, this project try to further enhance the clarity and relevance of user queries in conversational AI systems. The foundational concepts and techniques from their work serve as the core framework for the developments and improvements of this.

In the realm of natural language processing and conversational AI, the task of question reformulation plays a crucial role in enhancing the clarity and relevance of user queries within interactive systems. The study "Action-Based Network for Conversational Question Reformulation" by Zheyu Ye, Jiangning Liu, Qian Yu, and Jianxun Ju [2] provides a notable contribution to this field. Their research introduces an innovative approach to refining questions in a conversational context, addressing common challenges such as ambiguity, redundancy, and context maintenance.

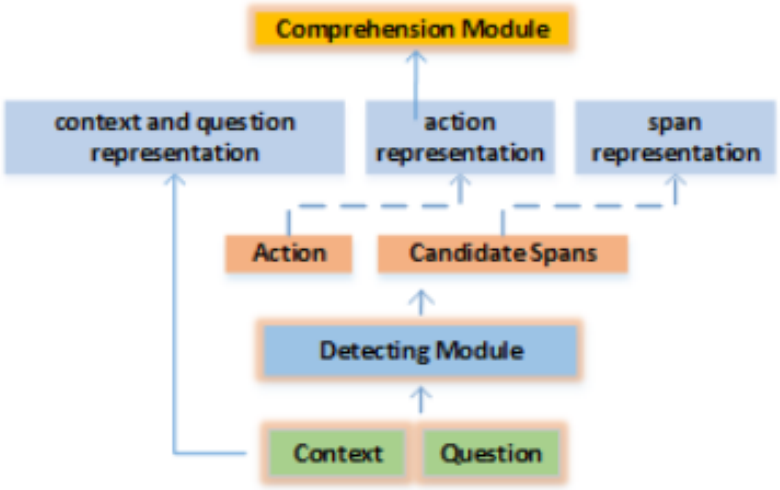

Figure 3.1: Overall architecture of proposed ActNet, consisting of the encoder component, the detecting component and the comprehension component

3.1 Objective

The central idea proposed by the authors revolves around the development of an action-based network that is called ActNet, which can dynamically reformulate questions to

improve the overall interaction quality between users and AI systems. The primary goal is to create a model that understands the context and intent behind user questions and can rephrase them in a way that enhances clarity and relevance, thereby making the responses from the AI more accurate and useful.

3.2 Methodology

To achieve this, the authors designed an action-based network that leverages deep learning techniques. The network consists of several key components:

3.2.1 Encoder

The authors stated that language model pre-training, such as BERT-base, has proven effective for various NLP tasks, including natural language inference and named entity recognition. And by concatenating questions and their contexts with separator tokens, the model can better understand the relationships between them.

The encoder captures the conversational history and relevant context surrounding the user's query. This is crucial for maintaining the continuity and coherence of the conversation, ensuring that the reformulated question is contextually appropriate.

3.2.2 Detecting Module

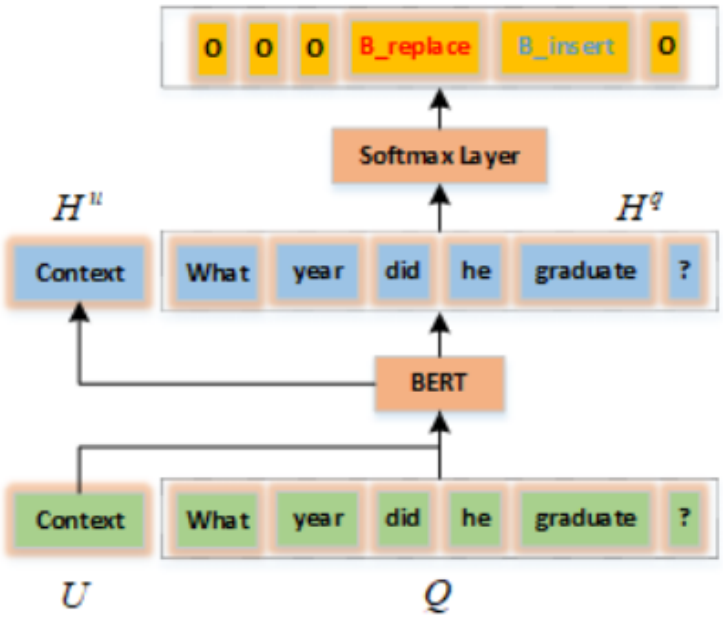

Figure 3.2: Architecture of encoder and detecting module. The green block is the text input, and the blue block is the representation output of that. The encoder uses a pre-trained language model BERT-base to encode the text. The question representations are injected into a softmax layer equipped with a greedy search to generate action tags

To handle references and omissions in a question, the authors had used a sequence tagging method to mark their positions. Each token gets a BIO tag: B for the beginning, I for inside, and O for no action needed. They refine B tags into B-insert for insertions and B-replace for replacements. These tags helped deciding the specific action needed between text segments.

A linear layer with softmax activation computes the probability of each tag. During inference, a greedy search decodes the tags, identifying spans starting with B tags as candidates for handling references and omissions.

This module identifies the type of reformulation action needed for a given question. Actions could include rephrasing, elaboration, simplification, or context addition. By recognizing the required action, the system can determine the most appropriate way to reformulate the question.

3.2.3 Comprehension Module

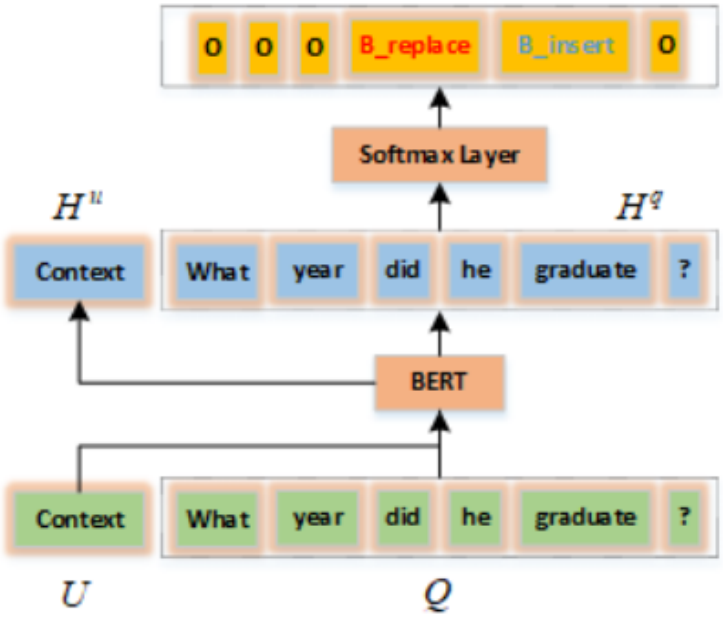

Figure 3.3: Architecture of comprehension module. The brown block is the probability of the start token of the answer span and the green block is the probability of the end token. The context representation and the candidate spans representation of the question is increased to an attention layer to generate answer span probability.

In this module, the authors used a machine reading comprehension (MRC) method to find information in the conversation history that is referenced or missing in the question. The process involves taking a query and context, predicting an answer as a text segment in the context. Each potential reference in the question is treated as a query, and the conversation history as the context. The goal is to estimate the likelihood of finding the correct segment. They simplify by using the start token of the span to represent the entire query's meaning and employ an attention neural network to determine the start and end positions of the answer span within the context.

3.3 Implementation

To optimize the performance of the network, the authors employed reinforcement learning techniques. This approach allows the model to learn from interactions and improve its reformulation strategies over time based on feedback and rewards, which are derived

from the quality of the responses generated.

The implementation of the action-based network involves training the model on a large dataset of conversational interactions. The dataset includes various types of questions and corresponding reformulations, providing a rich source of examples for the network to learn from. The training process incorporates supervised learning to initialize the network's parameters, followed by reinforcement learning to fine-tune the reformulation strategies.

During training, the model is evaluated on its ability to generate reformulations that improve the clarity and relevance of questions. Metrics such as BLEU score, ROUGE score, and human evaluation are used to assess the quality of the reformulated questions. Additionally, the impact of reformulated questions on the performance of downstream tasks, such as information retrieval and question answering, is also considered.

3.4 Results

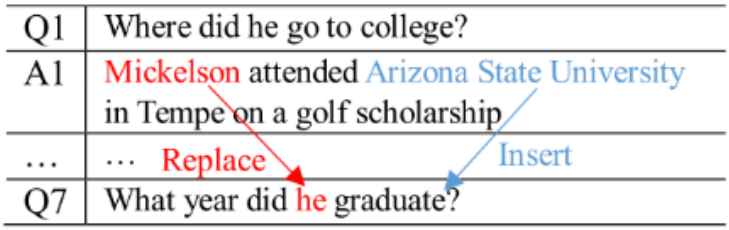

Figure 3.4: How the action-based network works. Red means coreference replacement and blue means omission supplement.

The results presented by the authors demonstrate significant improvements in the quality of conversational question reformulation. The action-based network was shown to effectively identify appropriate reformulation actions and generate questions that are clearer and more contextually relevant. Compared to baseline models, the proposed approach achieved higher scores on various evaluation metrics, indicating its effectiveness in enhancing conversational interactions.

Moreover, the study highlights the broader impact of improved question reformulation on conversational AI systems. By generating better reformulated questions, the system can provide more accurate and relevant responses, leading to a more satisfying user experience. This has implications for a wide range of applications, from customer service chatbots to virtual assistants and educational tools.

3.4.1 Advantages

The action-based network devised by Zheyu Ye, Jiangning Liu, Qian Yu, and Jianxun Ju [2] demonstrates remarkable prowess in dynamically reformulating questions based on conversational context. This ability results in clearer and more pertinent interactions. Employing reinforcement learning, the research facilitates continuous enhancement through feedback loops, a process bolstered by comprehensive evaluation metrics confirming its efficacy. This adaptability, coupled with its contextual comprehension, renders it exceptionally advantageous for applications such as customer service and virtual assistants. Consequently, it substantially elevates user satisfaction and communication efficiency.

The advantages of this approach are numerous. Firstly, its dynamic nature enables it to adapt swiftly to changing contexts, ensuring that interactions remain relevant and coherent. Secondly, the incorporation of reinforcement learning allows for continuous improvement, as the system learns from its interactions and refines its responses over time. Moreover, the comprehensive evaluation metrics utilized in the research had made considerable validation of the effectiveness of the approach.

3.4.2 Disadvantages

Nevertheless, the implementation of this sophisticated network is not without its challenges. Its complexity and resource-intensive nature present significant difficulty, particularly concerning the addition of extensive, high-quality conversational datasets and the requirement for substantial computational power. Additionally, concerns arise regarding the scalability of the approach to real-world applications, the potential for misinterpre-

tation of reformulation actions, and the inherent distinction involved in evaluating conversational systems.

The complexities associated with implementing such advanced systems underscore the need for consideration and further refinement. Addressing these challenges is essential to ensure the practical and effective deployment of the network across diverse domains.

4

Proposed methods

In Chapter 4, the author will describe the coreference resolution model that was utilized in this research. This section will include an exploration of key architecture that was adapted and its application in the developing of HCMUT chatbot.

Contents

4.1 Technical setup

Implementing coreference resolution to solve question reformulation problems can greatly enhance natural language understanding systems. Leveraging libraries such as crosslingual-coreference, Spacy models like en_core_web_sm, and pretrained multilingual models like facebook/mbart-large-50-many-to-many-mmt can provide a robust foundation for this task.

Overall, integrating these libraries and models into the question reformulation pipeline can greatly empower the system to accurately identify and resolve coreference mentions, leading to more coherent and contextually relevant reformulated questions across different languages and domains.

4.1.1 SpaCy models

SpaCy is a library for advanced Natural Language Processing in Python and Cython. It's built on the very latest research, and was designed from day one to be used in real products.[8]

Figure 4.1: SpaCy models

SpaCy comes with pretrained pipelines and currently supports tokenization and training for 70+ languages. It features state-of-the-art speed and neural network models for tagging, parsing, named entity recognition, text classification and more, multi-task learning with pretrained transformers like BERT, as well as a production-ready training sys-

tem and easy model packaging, deployment and workflow management. spaCy is commercial open-source software, released under the MIT license.[8]

Spacy models, particularly en_core_web_sm, provide efficient and accurate linguistic annotations, including part-of-speech tagging, dependency parsing, and named entity recognition. These annotations are essential for identifying and resolving coreference mentions within individual sentences, forming the basis for more complex coreference resolution algorithms.

SpaCy pipelines, models, and APIs are easy to use. Along with its speed and accuracy, many researchers, developers, and practitioners in NLP-related fields use SpaCy. SpaCy also has a rich ecosystem, good documentation, and active community support, which makes it beginners-friendly.

4.1.2 crosslingual-coreference library

crosslingual-coreference is a Python library that supports the coreferencing of documents between languages using SpaCy models with the assumption that cross-lingual embeddings should work for languages with similar sentence structures[7]. This library also supported SpaCy API to perform the coreference resolution operation

Figure 4.2: crosslingual-coreference library

The crosslingual-coreference library offers functionalities to resolve coreference mentions across languages, enabling seamless processing of multilingual text. By leveraging

this library, we can ensure that coreference resolution is not limited by language barriers, crucial for handling diverse datasets commonly encountered in question reformulation tasks.

4.1.3 en_core_web_sm Model

The en_core_web_sm model from SpaCy is renowned for versatility in handling various NLP tasks, making it an excellent choice for coreference resolution. As a lightweight yet powerful model, it includes pre-trained pipelines for tokenization, part-of-speech tagging, dependency parsing, named entity recognition, and more. These pipelines work together to analyze the structure and meaning of the text, which is crucial for accurately resolving coreferences.

One of the key strengths of the en_core_web_sm model lies in its ability to understand and process complex linguistic structures. This model has been trained on a diverse and extensive corpus, enabling it to capture a wide range of linguistic patterns and nuances. Its high accuracy in tagging and parsing helps in identifying the relationships between words and phrases, which is essential for pinpointing the correct antecedents for pronouns and other referring expressions.

Moreover, the model's efficiency and scalability make it suitable for real-time applications. Its optimized design allows for fast processing speeds, which is crucial for interactive systems like chatbots and virtual assistants. By quickly resolving coreferences in user queries, the model helps these systems provide timely and relevant responses. This responsiveness enhances user experience, maintaining a natural and coherent conversation flow.

However, the en_core_web_sm model by SpaCy primarily excels in English-language tasks due to its extensive training on English corpora. Its proficiency lies in understanding and processing intricate linguistic structures specific to the English language. With its accurate tagging and parsing capabilities, it effectively identifies relationships

between English words and phrases, facilitating precise coreference resolution. This model's efficiency further enhance its suitability for real-time applications in English-centric contexts, such as chatbots and virtual assistants, ensuring rapid and accurate responses to user queries in English.

4.1.4 facebook/mbart-large-50-many-to-many-mmt

A Facebook's multilingual translation model, created using the original mBART model and extended to add extra 25 languages to support multilingual machine translation models of 50 languages[9] because of its popularity and fast translation. In this project, the author adopted this model to replace the QA agent in answering questions with given documents.

Moreover, pretrained multilingual models like facebook/mbart-large-50-many-to-many-mmt offer extensive language coverage and capture crosslingual semantic representations. By fine-tuning these models on question reformulation datasets, we can leverage their multilingual capabilities to perform coreference resolution across various languages, further enhancing the versatility and effectiveness of the system.

4.2 Proposed methods

4.2.1 General pipeline overview

As stated before, the goal in this project is to try to use a coreference resolution model to de-contextualize the questions from the user to improve the performance of the QA model (in this case the LLM chatbot model of HCMUT Admission Office). The first step for the model to achieve this goal is to be able to coreference the questions given by the user based on the chat history or what referred as the context.

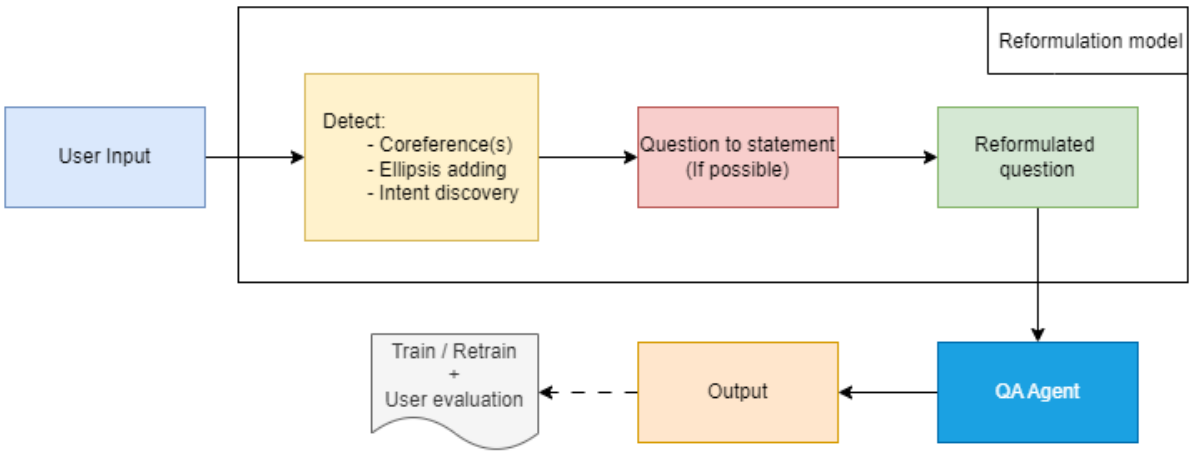

Figure 4.3: The general pipeline overview of the system.

The coreference agent sits before the chatbot agent doing the "preprocessing" job on the questions before feeding them to the chatbot. Assumed the chatbot is a black box and we do not have access to its internal component, assuming that the information generated from the chatbot is 100% correct but additional information might be required by the user.

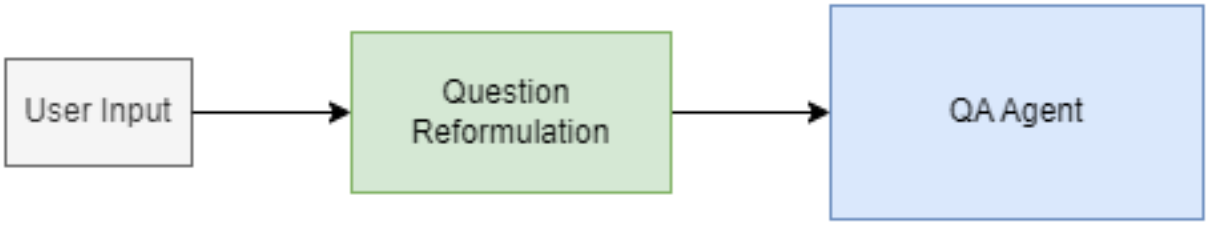

Figure 4.4: Another look at the general pipeline overview of the system.

The model consists of 2 layers. The first layer is the coreference resolution model. This model will take the query from the user, along with the chat history (the context), and try to replace the pronouns in the query (if any) with the correct entities extracted from the context. After that, the coreference agent will pass the reformulated question to the QA model to generate the answer.

The general pipeline of the system visually represents the workflow of the coreference resolution and question reformulation system designed to enhance the QA model's

performance. The process begins with user input, which could be a question or a statement requiring clarification. This input is fed into the coreference resolution model, the first critical component of the system. This model is tasked with detecting and resolving coreferences, adding any missing ellipses, and uncovering the user's intent. By addressing these linguistic challenges, the model transforms ambiguous user queries into clearer, contextually accurate questions.

Following this initial detection phase, the system attempts to rephrase questions as statements wherever applicable. This step aims to further de-contextualize and simplify the user's input, ensuring that the subsequent QA model receives straightforward and well-defined queries. The reformulated question is then generated and prepared for the QA agent. This agent, which operates as a black box, processes the refined input to generate precise and contextually relevant responses. The assumption is that with clearer input, the QA agent's output will be more accurate and useful to the user.

The final stage involves evaluating the system's output and incorporating user feedback to continually train and retrain the model. This iterative process is essential for refining the coreference resolution and question reformulation mechanisms. By leveraging user evaluations, the system can adapt to a wide range of conversational scenarios, ensuring its effectiveness in real-world applications. The ultimate objective is to create a seamless integration between the preprocessing layers and the QA model, leading to an enhanced user experience through more accurate and contextually appropriate answers.

4.2.2 Handling Coreference Resolution

The input of the model is a user's Vietnamese query. In the beginning, the input went through a translation model. In this research, *facebook/mbart-large-50-many-to-many-mmt*[9] is used to translate the query to English for a better coreference result as the *en_core_web_sm* model works best for English and similar structured languages.

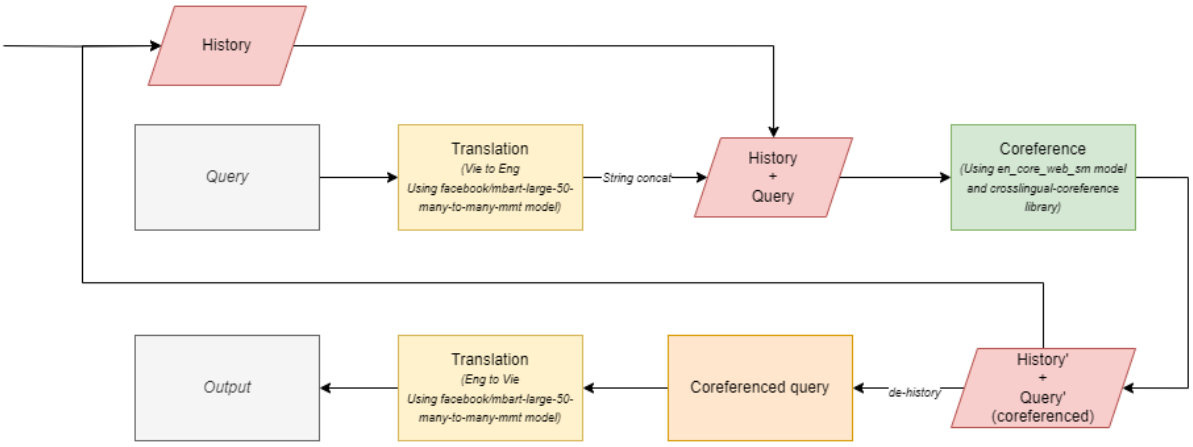

Figure 4.5: The coreference pipeline of the system.

After having translated the query, it is string-concatenated with the history to perform the coreference. In the beginning, the history is an empty string. The coreferenced string will then be separated into two parts: The English coreferenced query, which is copied from the string will go through the translation model to translate into Vietnamese to pass to the chatbot model. And the coreferenced string (the concatenated history and query) will then be concatenated with the answer received from the chatbot to form a new history. After the translation, the model will then return a Vietnamese answer, along with coreferenced query as an output.

Generally, this is the proposed method of how Coreference is solved. Starting with a Vietnamese query from the user, the system leverages the facebook/mbart-large-50-many-to-many-mmt model for translation from Vietnamese to English. This translation step is crucial because the en_core_web_sm model used for coreference resolution performs optimally with English input. By translating the query into English, the system ensures more accurate identification and resolution of coreferences, thus enhancing the overall coherence of the query.

Once translated, the English query is concatenated with any existing history of the conversation. Initially, this history is empty, but as the conversation progresses, it accumulates past interactions. This concatenated string, which now contains the entire context of the conversation, is then passed to the coreference resolution model. The

coreference resolution step is essential for replacing pronouns and ambiguous references in the query with the actual entities they refer to, based on the context provided by the history. This process results in a clear and contextually enriched query that can be more accurately understood by the QA model.

The output from the coreference resolution model is divided into two parts: the coreferenced query and the coreferenced string (which includes both the history and the current query). The coreferenced query is translated back into Vietnamese using the same translation model. This ensures that the query remains in the user's preferred language while benefiting from the enhanced clarity provided by the coreference resolution process. This translated query is then sent to the chatbot model to generate a response.

Simultaneously, the coreferenced string (the concatenated history and query) is updated by appending the answer received from the chatbot. This updated string forms the new history, which will be used in subsequent interactions to provide context. By maintaining a dynamic history that evolves with each user interaction, the system ensures that the context is always up-to-date, enabling more accurate and context-aware responses from the chatbot.

The final output of the system includes both the Vietnamese answer from the chatbot and the coreferenced query. The Vietnamese answer is presented to the QA agent, ensuring a seamless and natural conversational experience. Meanwhile, the coreferenced query can be used for further analysis and improvement of the system. This comprehensive pipeline demonstrates the integration of advanced translation and coreference resolution techniques to enhance the performance and user experience of QA systems.

4.2.3 Developing a Vietnamese Coreference Resolution model

The design of a Vietnamese coreference resolution model aimed to enhance the efficiency of the QA system by eliminating the dependency on the translation model. This approach focused on leveraging phoBERT, a powerful pre-trained model for Vietnamese,

to perform Named Entity Recognition (NER) tasks. By extracting entities from the conversation history and labeling them, the model could maintain context and accurately resolve coreferences within the user's queries. The system then employed NER to identify pronouns in the current query, which were subsequently matched with the labeled entities through unsupervised classification. This method aimed to replace ambiguous pronouns with specific entities, ensuring that the QA agent received clear and contextually precise questions.

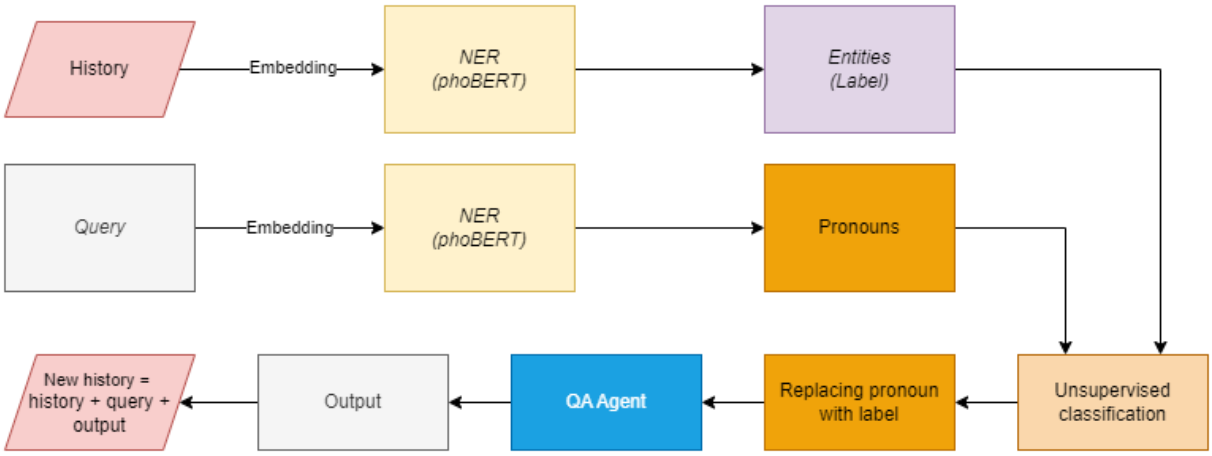

Figure 4.6: The Vietnamese model pipeline with a pseudo QA agent.

The figure illustrates a detailed pipeline designed for Vietnamese coreference resolution aimed at improving the efficiency and accuracy of QA systems. The process begins with the conversation history and the user's query, both of which are embedded using the phoBERT model to facilitate Named Entity Recognition (NER). By embedding these inputs, the system can effectively capture the semantic meaning and context of the text, which is crucial for accurately identifying entities and pronouns.

The first step involves running the NER task on the conversation history to extract all relevant entities. These entities serve as labels that represent the various subjects, objects, and other important references within the conversation. Extracting these entities accurately is essential for maintaining the context of the conversation, as they will be used to resolve any pronouns or ambiguous references in the subsequent queries.

Simultaneously, the NER task is also performed on the user's current query to identify pronouns that need to be resolved. The identified pronouns are then subject to an unsupervised classification process, where each pronoun is matched with the appropriate entity label extracted from the history. This matching process involves sophisticated algorithms that analyze the context and ensure that each pronoun is correctly replaced with the corresponding entity, thus resolving any ambiguities.

Once the pronouns are replaced with their respective entities, the coreferenced query is formed. This query, now clear and contextually enriched, is sent to the QA agent for processing. The QA agent, benefiting from the precise and unambiguous input, can generate accurate and relevant responses. This step is crucial as it bypasses the need for a translation model, thereby reducing the overall execution time of the system.

Finally, the output from the QA agent, along with the coreferenced query, is used to update the conversation history. The new history now includes the latest query and its response, ensuring that the context remains up-to-date for future interactions. This dynamic updating of history is vital for maintaining continuity in the conversation and enhancing the overall user experience by ensuring that each query is understood within the correct context.

However, the implementation of this model presented several challenges. One significant difficulty was the effective extraction of labels from the history. Coreference resolution often involves overlapping entities and references, making it complex to identify and categorize entities accurately. The overlap property of coreference resolution, where a single entity might have multiple references in different contexts, further complicated the task. Developing a powerful method to handle these overlaps was crucial but proved to be time-consuming and challenging.

Another major challenge was the unsupervised classification of pronouns. The goal was to determine the correct entity for each pronoun without relying on pre-labeled training data. This required the development of sophisticated algorithms capable of ac-

curately matching pronouns with their corresponding entities based on context. Finding an effective and reliable way to perform this classification was a significant difficulty, as it involved ensuring high precision and recall in identifying the correct entities, which is essential for maintaining the coherence and accuracy of the QA system's responses.

Due to these complexities and the limited time available, the improved model could not be fully implemented and tested. The author faced difficulties not only in the technical aspects of label extraction and classification but also in integrating these components into a cohesive system that could operate efficiently in real-time. Despite these challenges, the conceptual framework provided valuable insights into the potential of using native language models for coreference resolution and highlighted the areas that require further research and development.

5

Experiments

In this chapter, the author will demonstrate the result of the adapted coreference resolution model in a QA model.

Contents

5.1 BERTScore

BERTScore is an automatic evaluation metric for text generation. Analogously to common metrics, BERTScore computes a similarity score for each token in the candidate sentence with each token in the reference sentence.

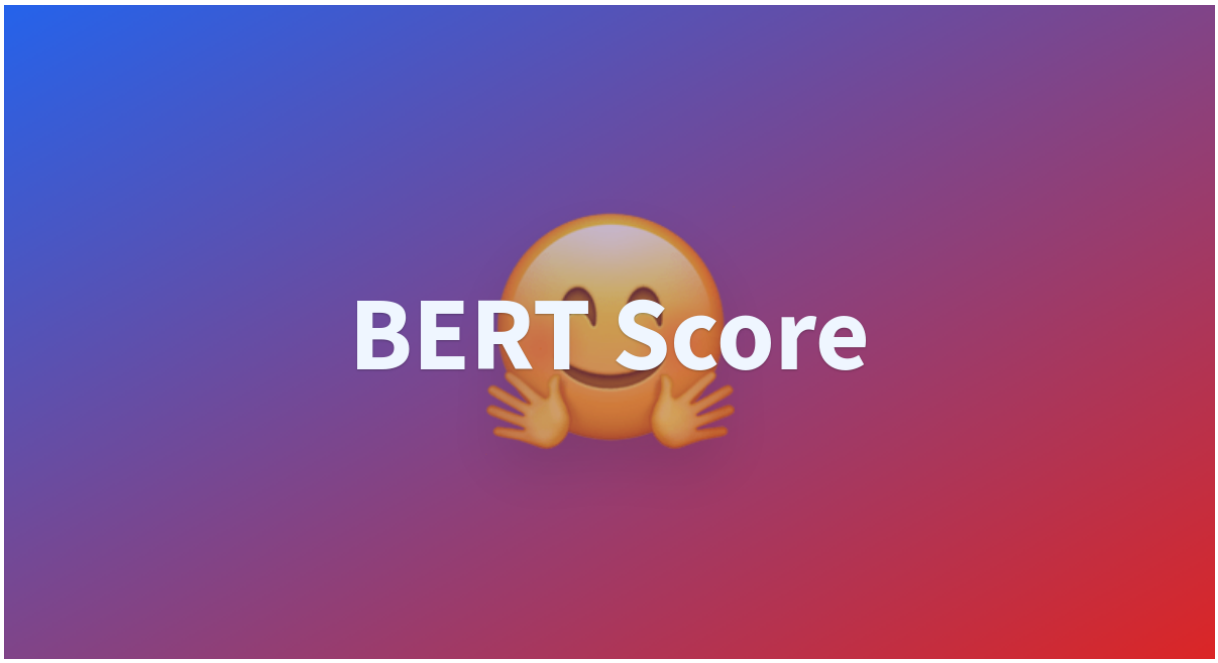

Figure 5.1: BERTScore - a Hugging Face Space by evaluate-metric.

However, instead of exact matches, it computes token similarity using contextual embedding. It evaluates using the outputs of 363 machine translation and image captioning systems. BERTScore correlates better with human judgments and provides stronger model selection performance than existing metrics. Finally, BERTScore is more robust to challenging examples when compared to existing metrics.[4]

In this project, BertScore is used as the main measure to evaluate the similarity of the answers output by the QA model to show that the co-reference model helps increase the confidence of the model and the information retrieved.

5.2 Result with Admission Dataset given by the Admission Office

The experimentation involved transforming and retesting the accuracy of the cross-lingual co-reference model using BertScore on specific data. The outcome demonstrated an improvement in sentence similarity by 0.6%, elevating it from 93.01% to 93.63%. However, further validation is essential due to potential data.

The primary challenge in testing this data-set comes from the data's apparent misalignment with the task requirements for co-reference. Co-reference needs historical conversational data where pronouns substitute nouns. However, the available data is a one-question-one-answer format, lacking the conversational context. Therefore, as a temporary measurement, the author self-handedly co-referenced and corefered the data as the context and concatenated string of history and question. After that, translating them into English. Thus, further experimenting with other data processing techniques or exploring alternative data-sets becomes crucial for certainty.

In conclusion, while the BERTScore resulted in promising in elevating sentence similarity, the need for further validation still exists. The mismatch between data characteristics and co-reference requirements underscores the necessity for diverse data sources or refined data processing methods to ensure the consistency of the model's performance.

5.3 Result with the CoQA Dataset and roberta-base-squad2 model of deepset from Hugging Face Hub

The final aim was to assess the model's functionality using the CoQA dataset since a Vietnamese dataset was unavailable. Therefore, the author also skipped the question translation step, instead, only translate the human input answer. The model employed for this evaluation acted as a QA system using deepset/roberta-base-squad2. The result was small but still positive, enhancing the confidence of the model's answers by over 0.7% (from 36.02% to 36.73%) under the condition that the data underwent co-reference

beforehand. This evaluation included of 10 dialogue segments and 141 questions for demonstration purpose and processing time estimation for later tests. The author hoped that with the help of Knowledge Graph (KG) the co-reference process for the data is no longer needed.

The confidence level in the model's answers witnessed an uplift, meeting the target increase of over 0.8%. The similarity between the original and extracted answers slightly increased to 82.59% from 81.80% according to BertScore. Manual inspection revealed that the reduced similarity in the answers stemmed from the QA system retrieving more information and the pronouns got replaced by the entities, regardless of co-reference application. Both co-referenced and non-co-referenced instances exhibited similar distinction. However, at some point, the model did not work as well as expected. This shows that the project also need other methods to achieve better reformulation result.

For example, under the same condition of the QA model have not been train for the CoQA dataset:

- The question: "Did they want Cotton to change the color of her fur?" resulted in the model without coreference (Un-Coref) returning the answer "I only wanted to be more like you", while the model with coreferenced model (Coref) returned the answer "5 other sisters wouldn't want that", which math the result of "no".

- The question: "What guided RJ home?" resulted in the model Un-Coref returning the answer "his father's flashlight", the Coref model returned the answer "Reginald Eppes's father's flashlight". Both answers are correct to the result "The flashlight", but the answer of Coref model has more information than the Uncoref model.

- Sometimes, both models can be wrong in retrieving the answer, like in the question "What kind of dishes does she bring?", the Un-coref model returns the answer "Ipad", while the Coref model returned the answer "Lucy can hardly do dishes in return". Both are wrong from the result "hot soup and a container with rice, vegetables and either chicken, meat or shrimp, sometimes with a kind of pancake". However, the Coref model seems to have helped the QA model extract more accu-

rate information.

Although there are also cases where the Coref model returned the wrong answer while the Un-coref model returned the right one, it is suspected that due to the incorrect coreference of the question or the searched document. This problem requires further research into different ways to reformulate the question and handle the document.

5.4 Analysis

In general, the assessment showed an increase in the model's answer confidence, meeting the specified threshold. However, a slight decline in the similarity between original and extracted answers was observed, primarily due to the QA system's tendency to retrieve more information, regardless of co-reference application. The model's ability to handle pronoun-embedded questions indicates its utility in extracting additional details. Additionally, the failure of the model in some occasions suggested the author to explore methods to eliminate context co-reference dependency, potentially leveraging knowledge graphs, finding other ways to handling questions and data, opening a pathway for future developments.

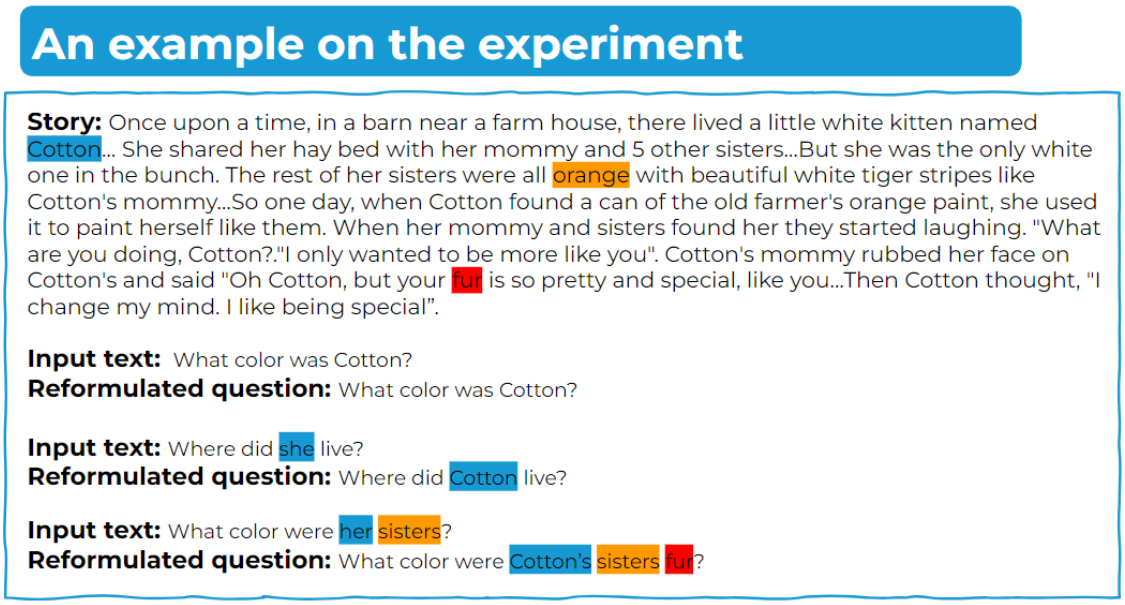

Figure 5.2: An example on the experiment

In the given story, "Once upon a time, in a barn near a farmhouse, there lived a little white kitten named Cotton," there encounter a series of scenarios where a question posed about the text needs to be reformulated for clarity and context. For instance, when asked "what color was Cotton," the question remains unchanged because it is already complete and clear. However, when the input text is "where did she live," the reformulated question correctly changes to "where did Cotton live." This transformation is intentional and demonstrates a specific rule or behavior in the reformulation process. Ideally, the reformulated question should capture the intended information while attaching to set patterns or rules, as shown in this example.

Another example is when the input text is "what color were her sisters," and the desired reformulated question should be "what color were Cotton's sisters' fur" to ensure clarity and specificity. This example highlights the challenges in maintaining contextual accuracy and specificity in question reformulation processes. The primary goal is to generate questions that are precise and reflective of the intended meaning within the story. By examining these scenarios, we can understand the complexities involved in creating an effective question reformulation system that not only follows specific rules but also enhances comprehension and retains the contextual integrity of the original text. This is especially important in educational and narrative contexts, where accurate and meaningful questions can significantly impact understanding and engagement.

Model advantages while testing with this dataset:

- Proficiency in handling questions embedded within passages containing pronouns.
- Enhanced extraction of additional information; nevertheless, a summarization model might still be necessary for users seeking more concise answers.
- Improved disambiguation capabilities: the model showcases a notable advantage in disambiguating ambiguous references within passages, especially those involving pronouns. Its proficiency in discerning context allows for more accurate comprehension, leading to better-informed responses.
- Potential exploration to obviate the need for context co-reference, such as leverag-

ing knowledge graphs (KG) to enrich the data pool.

6

Conclusion

In the following chapter, the author will engage in a comprehensive exploration of avenues for further improvement within the context of the thesis. This discussion will encompass a detailed examination of potential enhancements, refining aspects of the research to bolster its overall depth and effectiveness

Contents

During the length of the thesis, the author briefly mentioned some improvements when discussing many of the functionalities of the system. The author would like to revisit some of those improvements, as well as discussing new potential improvements.

6.1 Lacking real data

Although the author received a dataset from the Admission Office containing questions that students asked in real-time chats with university technicians, it was challenging to fully utilize this dataset. The dataset was very noisy, and the data is preprocessed - which make significant information removed, rendering it incomplete. Furthermore, there was no metric available for the author to evaluate the coreferenced questions accurately.

Integrating the CoQA dataset, initially designed for English, into Vietnamese research was a necessary step due to the lack of native data. Despite language differences, this adaptation was crucial in exploring question answering systems and computational linguistics. It laid a solid foundation for the project, though it came with its challenges. The author tackled these obstacles by using existing methodologies and adjusting them to fit their research. This approach helped bridge the gap between English and Vietnamese, advancing cross-linguistic computational research.

6.2 Vietnamese Coreference Model

The author attempted to enhance the efficiency of the model by implementing a coreference resolution model that operates on a Vietnamese corpus. This improvement would significantly reduce the execution time of the model, as it would eliminate the need for the time-consuming translation model.

However, due to time constraints and inefficiencies, the author was unable to implement the improved model on time. Additionally, challenges were faced in effectively extracting labels due to the overlapping nature of the coreference resolution problem and ongoing research into finding an effective method for the unsupervised classifica-

tion task.

So instead, the translation module is enhanced to make the model run slightly faster. Overall, it makes the model consistently runs query with less than 0.5 seconds per query. Which is quite acceptable in the authors' opinion.

6.3 Conclusion

Question Reformulation is not a new problem in the NLP field. However, due to the popularity of LLMs, in-depth research on the problem is not many. Reformulation refers to the process of modifying or rephrasing text or queries to improve their clarity, precision, or relevance for downstream tasks. The question reformulation problem is the task of rephrasing a question or removing the context from the question without changing its original meaning to help the computer better understand the user's intent. The solution to question reformulation problem promises a way to make the question-answering model in the chatbot to be lighter, more robust and better at understanding the user's question.

The assessment of the model across various datasets has provided valuable insights. The crosslingual-coreference model showcased a commendable 0.6% increase in sentence similarity, demonstrating its potential for improvement. However, the dataset is not perfectly optimal for testing, meaning that it needed to be check more to be sure the results are reliable.

Similarly, the evaluation using the CoQA data-set reflected promising outcomes in enhancing answer confidence by over 0.7%. Nevertheless, a slight reduction in the similarity between original and extracted answers indicates the need for a deeper investigation into the model's information retrieval mechanisms.

Positively, the model exhibit strengths in specific areas, such as handling pronoun-embedded questions and coreference resolution. Both assessments highlight the necessity for refining data processing techniques and exploring alternative data-sets to reinforce the mod-

els' reliability and effectiveness.

References

[1] Ben Kantor, Amir Globerson (2019), Coreference Resolution with Entity Equalization. Retrieved from https://aclanthology.org/P19-1066.pdf

[2] Zheyu Ye, Jiangning Liu, Qian Yu, Jianxun Ju (2021), Action based Network for Conversation Question Reformulation. Retrieved from https://arxiv.org/abs/2111.14445

[3] Yuqing Tang, Chau Tran, Xian Li, Peng-Jen Chen, Naman Goyal, Vishrav Chaudhary, Jiatao Gu, Angela Fan (2020), Multilingual Translation with Extensible Multilingual Pretraining and Finetuning. Retrieved from https://arxiv.org/pdf/2008.00401.pdf

[4] Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi (2020), BERTSCORE: EVALUATING TEXT GENERATION WITH BERT. Retrieved from https://arxiv.org/pdf/1904.09675.pdf

[5] Zheyu Ye, Jiangning Liu, Qian Yu, Jianxun Ju (2021), Action based Network for Conversation Question Reformulation. Retrieved from https://arxiv.org/pdf/2111.14445.pdf

[6] Sheng-Chieh Lin, Jheng-Hong Yang, Rodrigo Nogueira, Ming-Feng Tsai, Chuan-Ju Wang and Jimmy Lin (2020), Conversational Question Reformulation via Sequence-to-Sequence. Retrieved from Architectures and Pretrained Language Models https://arxiv.org/pdf/2004.01909.pdf

[7] Introduction of crosslingual-coreference library. Retrieved from https://pypi.org/project/crosslingual-coreference/

[8] Introduction of SpaCy library. Retrieved from https://pypi.org/project/spacy/

[9] The adapted translation model. Retrieved from https://huggingface.co/facebook/mbart-large-50

[10] The reported frequency of noun ellipses is 1.99% (Khullar, Majmundar, and Shrivastava 2020); and that of VPE along with related phenomenon is 1% (Bos and Spenader 2011).[11]

[11] Payal Khullar(2021), Are Ellipses Important for Machine Translation?. Retrieved from https://direct.mit.edu/coli/article/47/4/927/106771/Are-Ellipses-Impo

[12] Query Expansion Example. Retrieved from https://www.researchgate.net/publication/228966554_DEU_at_ImageCLEF_2009_

[13] Sentence Simplification Example. Retrieved from https://www.researchgate.net/publication/353485830_Towards_Visual_Questio

[14] Paraphrasing Example. Retrieved from https://www.futurelearn.com/info/courses/en

[15] Coreference Resolution example. Retrieved from https://pypi.org/project/crosslingual-coreference/

[16] Ellipsis Resolution example. Retrieved from https://aclanthology.org/2021.eacl-main.68.pdf

Appendix

Appendix A: Task Evaluation

Throughout the project implementation, the author effectively employed an agile methodology, managed to finish the project through iterative sprints, each lasting approximately two weeks. At the conclusion of each sprint, the author conducted a comprehensive meeting to review the progress and strategize for the upcoming sprint.

The author communication with Assoc. Prof. Quan Thanh Tho and Mr. Bui Cong Tuan played a pivotal role. Their guidance and support had assisted in overcoming challenges, and they facilitated connections with stakeholders who supplied valuable real-world data and feedback. This collaboration enhanced the authors' understanding of the project, enable the author to align with stakeholder expectations.

Adopting an iterative and incremental approach, the author developed the project, focusing on delivering a minimum viable product with essential features and functionalities. Regular testing and evaluation were integral to maintaining the project's quality and performance. This experience provided valuable lessons, including the utilization of diverse technologies, effective teamwork, communication skills, and adept time and resource management. The author take pride in the project and aspire to contribute positively to the coreference field.

Appendix B: Task Description

Task	Description
Define the problem	Meetings with the instructors to comprehend the issue at hand about its significance, and the underlying reasons for the project necessity.
Researching	Involved an in-depth exploration of information related to the project's problem, drawing insights from diverse resources like academic papers and relevant data sources. To lay the groundwork for effective problem-understanding
Design workflow	Utilizing the identified problem, mapping out the comprehensive project pipeline, and outlining the project's current position within the larger project framework.
Explore potential solutions	Engaging in discussions with the instructors to receive guidance on project direction and technology choices for the overarching project.

Table 6.1: Table of task description on defining the problem and exploring potential solutions

Task	Description
Selecting the library	Engaging in meetings with the instructors, reviewing academic papers to make an informed decision on selecting an optimal library.
Coding and Running Small Tests	Library experimental by coding and conducting small-scale tests to ensure the selected library's suitability for the project requirements.
Data preprocessing	Generating expected questions from the data-set's existing questions to provide further evaluation and insight into the model's performance.
Complete the code and run the data-set	Finalizing the coding process and executing it on the data-set for comprehensive analysis and evaluation of the model's functionality.
Evaluate result	Utilizing BertScore for assessing the model's efficacy, precision, and confidence.

Table 6.2: Table of task description on Developing the model