กรอบงานการตอบคำถามภาษาไทยโดยใช้แบบจำลองประเภททรานส์ฟอร์เมอร์พร้อมด้วยข้อมูลเสริม

นายภูริ ฟักมงคล

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2564
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

Thai Question Answering Framework Using Transformer Based Models With Augmented Data

Mr. Puri Phakmongkol

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Computer Science
Department of Computer Engineering
FACULTY OF ENGINEERING
Chulalongkorn University
Academic Year 2021

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	กรอบงานการตอบคำถามภาษาไทยโดยใช้แบบจำลองประเภททรานส์ฟอร์เมอร์พร้อมด้วยข้อมูลเสริม
โดย	นายภูริ ฟักมงคล
สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ผู้ช่วยศาสตราจารย์ ดร.พีรพล เวทีกูล

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

..... คณบดีคณะวิศวกรรมศาสตร์
(ศาสตราจารย์ ดร.สุพจน์ เตชวรสินสกุล)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(ศาสตราจารย์ ดร.บุญเสริม กิจศิริกุล)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(ผู้ช่วยศาสตราจารย์ ดร.พีรพล เวทีกูล)

..... กรรมการ
(ดร.ดวงดาว วิชาดากุล)

..... กรรมการภายนอกมหาวิทยาลัย
(ดร.ปรัชญา บุญขวัญ)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภูริ ฟักมงคล : กรอบงานการตอบคำถามภาษาไทยโดยใช้แบบจำลองประเภททรานส์ฟอร์เมอร์พร้อมด้วยข้อมูลเสริม. (Thai Question Answering Framework Using Transformer Based Models With Augmented Data) อ.ที่ปรึกษาหลัก : ผศ. ดร.พีรพล เวทีกูล

การตอบคำถาม (Question Answering) เป็นหนึ่งในงานด้านการประมวลผลภาษาธรรมชาติ ที่สามารถให้อุปกรณ์เข้าใจบทความและสามารถตอบคำถามที่ให้ได้ มีงานศึกษาทางด้านตอบคำถามในภาษาอังกฤษซึ่งมีจำนวนของชุดข้อมูลที่มีจำนวนมาก แต่ภาษาไทยเป็นภาษาที่มีจำนวนของชุดข้อมูลที่น้อยกว่า ทำให้ประสิทธิภาพของแบบจำลองการตอบคำถามภาษาไทยมีคะแนนเอฟวันเพียง 70% ในขณะที่แบบจำลองการตอบคำถามภาษาอังกฤษ สามารถมีคะแนนเอฟวันสูงถึง 90% ดังนั้น ในงานวิจัยชิ้นนี้ต้องการเพิ่มประสิทธิภาพของแบบจำลองการตอบคำถามภาษาไทย โดยการสร้างชุดข้อมูลเพิ่มด้วยแบบจำลอง Multilingual Text-to-Text Transfer Transformer (mT5) และการประมวลผลข้อมูลที่เหมาะสมกับภาษาไทย วิธีนี้ สามารถสร้างชุดข้อมูลการตอบคำถามได้มากกว่าหนึ่งแสนชุดจากบทความ Wikipedia ภาษาไทย นอกจากนี้ งานวิจัยชิ้นนี้ยังศึกษารูปแบบวิธีการฝึกสอนในรูปแบบต่าง ๆ โดยใช้ชุดข้อมูลที่สร้างขึ้นเพื่อเพิ่มประสิทธิภาพให้กับแบบจำลองการตอบคำถามให้ได้มากที่สุด อีกทั้ง ในงานวิจัยนี้ได้เสนอการวัดผลแบบจำลองการตอบคำถามภาษาไทยโดยใช้คะแนนเอฟวันระดับพยางค์ ซึ่งมีความเหมาะสมกับภาษาไทยมากกว่าการวัดผลด้วยวิธีการดั้งเดิม โดยใช้คะแนนความถูกต้องและคะแนนเอฟวันระดับคำ ผู้วิจัยได้ทำการทดลองบนชุดข้อมูลการตอบคำถามภาษาไทยทั้ง 2 ชุด ได้แก่ Thai Wiki QA และ iApp Wiki QA พบว่าผลการทดลองโดยใช้วิธีการสร้างชุดข้อมูลของงานวิจัยชิ้นนี้สามารถให้ประสิทธิภาพของแบบจำลองสูงขึ้นบนชุดข้อมูลทั้ง 2 ชุด โดยใช้แบบจำลองประเภททรานส์ฟอร์เมอร์ ได้แก่ Roberta และ mT5

สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์ ลายมือชื่อนิสิต .....

ปีการศึกษา 2564 ลายมือชื่อ อ.ที่ปรึกษาหลัก .....

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

6272068021 : MAJOR COMPUTER SCIENCE
KEYWORD: Natural Language Processing, Question Answering

Puri Phakmongkol : Thai Question Answering Framework Using Transformer Based Models With Augmented Data. Advisor: Asst. Prof. PEERAPON VATEEKUL, Ph.D.

Question Answering (QA) is a natural language processing task that enables the machine to understand a given context and answer a given question. There are several QA research trials containing high resources of the English language. However, Thai is one of the languages that have low availability of labeled corpora in QA studies. According to previous studies, while the English QA models could achieve more than 90% of F1 scores, Thai QA models could obtain only 70% in our baseline. In this study, we aim to improve the performance of Thai QA models by generating more question-answer pairs with Multilingual Text-to-Text Transfer Transformer (mT5) along with data preprocessing methods for Thai. With this method, the question-answer pairs can synthesize more than 100 thousand pairs from provided Thai Wikipedia articles. Utilizing our synthesized data, many fine-tuning strategies were investigated to achieve the highest model performance. Furthermore, we have presented that the syllable-level F1 is a more suitable evaluation measure than Exact Match (EM) and the word-level F1 for Thai QA corpora. The experiment was conducted on two Thai QA corpora: Thai Wiki QA and iApp Wiki QA. The results show that our augmented model is the winner on both datasets compared to other modern transformer models: Roberta and mT5.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Field of Study:	Computer Science	Student's Signature .....
Academic Year:	2021	Advisor's Signature .....

กิตติกรรมประกาศ

การที่วิทยานิพนธ์ฉบับนี้สามารถสำเร็จได้ด้วยดีนั้น เนื่องจากผู้วิจัยได้รับความช่วยเหลือและความเมตตาจากบุคคลท่านอื่น ๆ อย่างมากมาย ทั้งในรูปแบบของคำแนะนำก็ดี หรือกำลังใจก็ดี ดังนั้น ผู้วิจัยจึงขอแสดงความขอบคุณมา ณ ที่นี้

ขอขอบคุณอาจารย์ที่ปรึกษา ผศ. ดร. พีรพล เวทีกูล ผู้ที่คอยช่วยเหลือและให้คำแนะนำในการทำวิจัยชิ้นนี้ตั้งแต่เริ่มต้นจนเสร็จสมบูรณ์ นอกจากนี้ อาจารย์ยังคอยให้กำลังใจและแก้ไขปัญหาอย่างเต็มที่อย่างเสมอ

ขอขอบคุณคณะกรรมการสอบวิทยานิพนธ์ ซึ่งประกอบด้วย ศ. ดร. บุญเสริม กิจศิริกุล อ. ดร. ดวงดาว วิชาดากุล และ ดร. ปรัชญา บุญขวัญ ที่ให้เกียรติเป็นคณะกรรมการให้กับผู้วิจัย และให้ความรู้และข้อเสนอแนะในการทำวิจัย

ขอขอบคุณเพื่อน ๆ ที่เข้าศึกษารุ่นเดียวกัน ที่ช่วยสอนและให้คำแนะนำระหว่างการเรียนรายวิชาและสามารถนำมาปรับใช้ในงานวิจัยชิ้นนี้

สุดท้ายนี้ ขอขอบคุณครอบครัว ที่ให้กำลังใจและสนับสนุนผู้วิจัยอย่างเสมอจนสามารถสร้างงานวิจัยชิ้นนี้ขึ้นมาได้

ภูริ ฟักมงคล

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญ

หน้า

..... ค

บทคัดย่อภาษาไทย ..... ค

..... ง

บทคัดย่อภาษาอังกฤษ ..... ง

กิตติกรรมประกาศ ..... จ

สารบัญ ..... ฉ

สารบัญตาราง ..... ฌ

สารบัญรูปภาพ ..... ญ

บทที่ 1 บทนำ 1

1.1 ที่มาและความสำคัญ 1

1.2 วัตถุประสงค์ของงานวิจัย 2

1.3 ขอบเขตของงานวิจัย 2

1.4 ประโยชน์ที่ได้รับ 3

1.5 วิธีการดำเนินการวิจัย 3

1.6 ผลงานวิจัยที่ตีพิมพ์ 4

บทที่ 2 แนวคิดและทฤษฎีที่เกี่ยวข้อง 6

2.1 การตอบคำถามในการประมวลผลภาษาธรรมชาติ 6

2.2 กลไกความสนใจ (Attention Mechanism) 7

2.3 ทรานส์ฟอร์เมอร์ (Transformer) 8

2.4 RoBERTa 9

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2.5 Text-to-Text Transfer Transformer และ Multilingual Text-to-Text Transfer Transformer 10
2.6 การวัดผลแบบจำลองการตอบคำถาม 10
บทที่ 3 เอกสารและงานวิจัยที่เกี่ยวข้อง 12
3.1 งานวิจัยทางด้านการตอบคำถามในปัจจุบัน 12
3.2 งานวิจัยทางด้านการตอบคำถามในภาษาไทย 13
3.3 งานวิจัยทางด้านการสร้างชุดข้อมูล 13
3.4 WangchanBERTa 14
3.5 สรุปการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง 15
บทที่ 4 แนวคิดและวิธีการดำเนินงาน 16
4.1 ขั้นตอนการเตรียมข้อมูล 16
4.2 ขั้นตอนการฝึกสอนแบบจำลองสำหรับสร้างชุดข้อมูลเสริม 17
4.3 ขั้นตอนการฝึกสอนแบบจำลองตอบคำถาม โดยใช้ชุดข้อมูลเสริม 21
4.4 การวัดประสิทธิภาพของแบบจำลองการตอบคำถามระดับพยางค์ 23
บทที่ 5 การทดลองและผลการทดลอง 25
5.1 ชุดข้อมูลที่ใช้ 25
5.2 เครื่องมือและการตั้งค่าพารามิเตอร์ของแบบจำลอง 26
5.3 การทดสอบวัดประสิทธิภาพ 29
5.4 ผลการทดลองและสรุปผลการทดลอง 29
5.5 วิเคราะห์การเพิ่มประสิทธิภาพของแบบจำลอง 35
5.6 วิเคราะห์ประสิทธิภาพของแบบจำลองการตอบคำถามโดยประเภทของคำถาม 37
5.7 การวิเคราะห์ตัวเลือกตัวตัดคำ 40
บทที่ 6 สรุปการวิจัยและแนวทางการวิจัยในขั้นถัดไป 42
6.1 สรุปการวิจัย 42

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

6.2 ข้อเสนอแนะและข้อจำกัดในงานวิทยานิพนธ์นี้ 42
6.3 แนวทางวิจัยในขั้นต่อไป 43
ภาคผนวก 44
ตัวอย่างของชุดข้อมูลเสริมที่ถูกสร้าง 44
บรรณานุกรม 48
ประวัติผู้เขียน 50

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สารบัญตาราง

หน้า

ตาราง 1 แผนภูมิแกนต์แสดงแผนการวิจัย 5

ตาราง 2 ตัวอย่างของข้อมูลการตอบคำถามภาษาไทย ประเภทการสกัดชุดข้อความ 7

ตาราง 3 การเลือกแบบจำลองของวิทยานิพนธ์นี้เทียบกับงานวิจัยของ Raul Puri และคณะ ในขั้นตอนการสร้างชุดข้อมูลเสริม 17

ตาราง 4 ตัวอย่างของการแบ่งคำและพยางค์ในวิทยานิพนธ์นี้ 24

ตาราง 5 จำนวนชุดข้อมูลฝึกสอน ชุดข้อมูลตรวจสอบและชุดข้อมูลทดสอบ 26

ตาราง 6 การตั้งค่าของไฮเพอร์พารามิเตอร์ที่ใช้งานในแบบจำลองการตอบคำถาม 27

ตาราง 7 การตั้งค่าของไฮเพอร์พารามิเตอร์ที่ใช้งานในชุดของแบบจำลองสำหรับสร้าง ชุดคำถาม-คำตอบ 28

ตาราง 8 จำนวนข้อมูลสำหรับฝึกสอนในแต่ละชุดข้อมูล เมื่อรวมกับชุดข้อมูลเสริมแล้ว 30

ตาราง 9 ผลการทดลองการใช้ชุดข้อมูลเสริมโดยการเปรียบเทียบกับการไม่ใช้ชุดข้อมูลเสริม ในชุดข้อมูล Thai Wiki QA 32

ตาราง 10 ผลการทดลองการใช้ชุดข้อมูลเสริมโดยการเปรียบเทียบกับการไม่ใช้ชุดข้อมูลเสริม ในชุดข้อมูล iApp Wiki QA 33

ตาราง 11 ตัวอย่างของคำตอบจากแบบจำลองและผลเฉลย และผลการคำนวณคะแนนเอฟวันระดับพยางค์ โดยทุกตัวอย่างจะมีคะแนนเอฟวันระดับคำเป็น 0 35

ตาราง 12 ตัวอย่างของคำสำคัญที่ใช้แบ่งประเภทคำถาม 38

ตาราง 13 ประสิทธิภาพของแบบจำลองตามประเภทของคำถาม ของชุดข้อมูล Thai Wiki QA 39

ตาราง 14 ประสิทธิภาพของแบบจำลองตามประเภทของคำถาม ของชุดข้อมูล iApp Wiki QA 39

ตาราง 15 ผลของคะแนนเอฟวันระดับคำในตัวตัดคำที่ต่างกันของชุดข้อมูล Thai Wiki QA 40

ตาราง 16 ผลของคะแนนเอฟวันระดับคำในตัวตัดคำที่ต่างกันของชุดข้อมูล iApp Wiki QA 41

ตาราง 17 ตัวอย่างของชุดข้อมูลเสริมที่ถูกสร้าง 44

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญรูปภาพ

หน้า

รูป 1 โครงสร้างของแบบจำลองทรานส์ฟอร์เมอร์ ซึ่งประกอบด้วย Encoder และ Decoder [4] .. 8

รูป 2 โครงสร้างของแบบจำลอง BERT ที่ประกอบด้วย Encoder ของแบบจำลองทรานส์ฟอร์เมอร์ 9

รูป 3 ขั้นตอนการฝึกสอนและสร้างชุดข้อมูลเพิ่มโดยใช้ชุดข้อมูล SQuAD [17]..... 14

รูป 4 กรอบงานการตอบคำถามภาษาไทยที่นำเสนอในวิทยานิพนธ์นี้..... 16

รูป 5 แสดงข้อมูลที่ใช้ในการฝึกสอนแบบจำลองสำหรับเลือกคำตอบ..... 18

รูป 6 แสดงข้อมูลขาเข้าที่ใช้ในการฝึกสอนแบบจำลองสำหรับสร้างคำถาม 19

รูป 7 แสดงข้อมูลที่ใช้ในการฝึกสอนแบบจำลองสำหรับกรองคำถาม 20

รูป 8 แผนภาพแสดงขั้นตอนการสร้างคู่คำถาม-คำตอบสำหรับชุดข้อมูลเสริมโดยรวม 21

รูป 9 ขั้นตอนการฝึกสอนแบบตามลำดับ..... 22

รูป 10 ขั้นตอนการฝึกสอนแบบรวมข้อมูล..... 22

รูป 11 กราฟแสดงการกระจายตัวของคะแนนเอฟวันระดับคำ สำหรับชุดข้อมูลทดสอบของชุดข้อมูล Thai Wiki QA โดยที่ สีน้ำเงินคือการกระจายตัวของข้อมูลจากแบบจำลอง mT5-Base โดยที่ไม่ได้ใช้ชุดข้อมูลเสริม และ สีส้มคือการกระจายตัวจากแบบจำลองของการทดลอง mT5-SEQ-FLT..... 36

รูป 12 กราฟแสดงการกระจายตัวของคะแนนเอฟวันระดับคำ สำหรับชุดข้อมูลทดสอบของชุดข้อมูล iApp Wiki QA โดยที่ สีน้ำเงินคือการกระจายตัวของข้อมูลจากแบบจำลอง mT5-Base โดยที่ไม่ได้ใช้ชุดข้อมูลเสริม และ สีส้มคือการกระจายตัวจากแบบจำลองของการทดลอง mT5-SEQ-ALL 37

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญ

การตอบคำถาม (Question Answering) เป็นหนึ่งในการวิจัยทางด้านการประมวลผลภาษาธรรมชาติ (Natural Language Processing) โดยคอมพิวเตอร์จะต้องทำการตอบคำถาม โดยใช้ความรู้จากบทความที่ให้พร้อมกับคำถาม ลักษณะของคำตอบจะเป็นส่วนใดส่วนหนึ่งของบทความ ซึ่งสามารถเป็นคำเพียงหนึ่งคำหรือจะเป็นวลีก็ได้ การตอบคำถามที่แม่นยำจะเป็นพื้นฐานที่สำคัญให้สามารถนำไปใช้ประโยชน์ได้มากมาย ได้แก่ การสร้างระบบการตอบคำถามเพื่อดึงความรู้จากบทความ การสร้างระบบแสดงคำตอบใน Search Engine หรือ การสร้างระบบแชทบอท (Chatbot) เป็นต้น

เนื่องจากการพัฒนาการตอบคำถามเป็นงานที่ท้าทาย ทำให้มีนักวิจัยหลายคนพัฒนาระบบตอบคำถามในภาษาต่าง ๆ โดยภาษาอังกฤษเป็นหนึ่งในภาษาที่นักวิจัยให้ความสนใจ สำหรับการพัฒนาการตอบคำถามในภาษาอังกฤษนั้น มีผู้วิจัยจำนวนหนึ่งศึกษาวิจัยด้านการตอบคำถามโดยใช้การเรียนรู้ของเครื่อง (Machine Learning) และมีชุดข้อมูลสำหรับฝึกเป็นจำนวนมาก เช่น Minjoon Seo และคณะ [1] ได้เสนอแบบจำลองสำหรับตอบคำถามโดยใช้แบบจำลอง LSTM และกลไกความสนใจ งานวิจัยเหล่านี้ ส่งผลให้มีการพัฒนาทั้งในด้านของแบบจำลอง (Model) และด้านเทคนิคที่มีความหลากหลาย ทำให้มีประสิทธิภาพของการตอบคำถามสำหรับชุดข้อมูลภาษาอังกฤษในระดับสูง

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ที่ผ่านมามีนักวิจัยได้ทำการพัฒนาการตอบคำถามที่รองรับภาษาไทยอยู่จำนวนหนึ่ง เช่น Hatsanai Decha และ Karn Patanukhom [2] และ Theerit Lapchaicharoenkit [3] อย่างไรก็ตามผู้วิจัยพบว่าการพัฒนาการตอบคำถามในภาษาไทยยังมีข้อจำกัดและมีความท้าทายที่มากกว่าภาษาอังกฤษ เนื่องจากภาษาไทยมีความกำกวมด้านโครงสร้างไวยากรณ์ โดยเฉพาะอย่างยิ่ง การตัดแบ่งคำหรือประโยค นอกจากนี้ ระบบการตอบคำถามภาษาไทยยังอยู่ในระดับเริ่มต้น เครื่องมือทางด้านการประมวลผลภาษาธรรมชาติยังไม่ทำงานได้อย่างสมบูรณ์ และยังมีข้อจำกัดด้านจำนวนชุดข้อมูลภาษาไทยสำหรับการวิจัยด้วย จึงมีความจำเป็นที่จะต้องพัฒนาระบบการตอบคำถามภาษาไทยให้มีประสิทธิภาพที่ดีขึ้น

ปัจจุบัน การเรียนรู้ของเครื่อง มีการพัฒนาอย่างก้าวกระโดด อันเนื่องจากข้อมูลในปัจจุบันที่มีมากขึ้น ความสำเร็จของการเรียนรู้เชิงลึก (Deep Learning) ที่สามารถแก้ปัญหาได้อย่างหลากหลาย ทั้งในด้านของ คอมพิวเตอร์วิทัศน์ (Computer Vision) และการประมวลผลภาษาธรรมชาติ อีกทั้ง การพัฒนาของฮาร์ดแวร์ที่มีความสามารถที่รองรับการคำนวณและแก้ปัญหาที่ซับซ้อนได้เร็วยิ่งขึ้น ทำให้มีผู้วิจัยแบบจำลองต่าง ๆ ที่มีประสิทธิภาพสูงขึ้นมามากมาย เช่น ทรานส์ฟอร์เมอร์ (Transformer) [4] เป็นแบบจำลองที่ทำงานโดยเรียนรู้ข้อมูลที่เป็นลำดับโดยใช้เพียงกลไกความสนใจเท่านั้น ทำให้สามารถประมวลผลได้เร็ว และได้รับความนิยมอย่างยิ่งในการพัฒนาแบบจำลองด้านการประมวลผลภาษาธรรมชาติ ได้แก่ การสร้างแบบจำลองตัวแทนภาษา เช่น BERT [5] RoBERTa [6] และ T5 [7] ซึ่งได้ผลลัพธ์ที่ดีในการใช้งานคู่กับชุดข้อมูลภาษาอังกฤษ ดังนั้น ผู้วิจัยเห็นว่า แบบจำลองทรานส์ฟอร์เมอร์จะสามารถนำมาประยุกต์ใช้กับการตอบคำถามภาษาไทยได้

งานวิจัยนี้ จึงมุ่งเน้นที่จะเพิ่มประสิทธิภาพของแบบจำลองการตอบคำถามสำหรับชุดข้อมูลภาษาไทย โดยใช้แบบจำลองทรานส์ฟอร์เมอร์แบบต่าง ๆ อีกทั้ง เสนอขั้นตอนการประมวลผลข้อมูลภาษาไทย และ การสร้างชุดข้อมูลที่เป็นคู่คำถามและคำตอบให้มีจำนวนมากขึ้น โดยใช้แบบจำลองการเรียนรู้เชิงลึก โดยทางผู้วิจัยหวังเป็นอย่างยิ่งว่า งานวิจัยชิ้นนี้จะช่วยแก้ไขข้อจำกัดทางด้านจำนวนชุดข้อมูลและเป็นส่วนหนึ่งของการพัฒนาต่อยอดงานวิจัยทางด้านการตอบคำถามภาษาไทยต่อไป

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

1.2 วัตถุประสงค์ของงานวิจัย

งานวิจัยนี้เสนอการใช้งานแบบจำลองทรานส์ฟอร์เมอร์ประเภทต่าง ๆ กับการตอบคำถามภาษาไทยเพื่อเพิ่มประสิทธิภาพของแบบจำลองการตอบคำถามให้ดีขึ้น และมุ่งแก้ไขข้อจำกัดด้านจำนวนชุดข้อมูลภาษาไทย

1.3 ขอบเขตของงานวิจัย

- งานวิจัยนี้จะทำการแก้ไขข้อจำกัดด้านจำนวนชุดข้อมูลภาษาไทย
- งานวิจัยนี้จะทดสอบโดยใช้แบบจำลองทรานส์ฟอร์เมอร์ต่าง ๆ ที่รองรับภาษาไทย

- แบบจำลองการตอบคำถามภาษาไทย จะตอบคำถามโดยใช้บทความที่ให้พร้อมกับคำถามเท่านั้น
- ใช้ข้อมูลจากชุดข้อมูลการตอบคำถามภาษาไทย ได้แก่ Thai Wiki QA และ iApp Wiki QA เท่านั้น
- งานวิจัยนี้เสนอวิธีการวัดผลที่เหมาะสมกับงานตอบคำถามภาษาไทยโดยใช้คะแนนความถูกต้อง (Exact match) คะแนนเอฟวันระดับคำ (Word-level F1) และ คะแนนเอฟวันระดับพยางค์ (Syllable-level F1) เท่านั้น

1.4 ประโยชน์ที่ได้รับ

1. ได้ทดลองการใช้งานแบบจำลองทรานส์ฟอร์เมอร์สำหรับการตอบคำถามภาษาไทย
2. ได้เทคนิคที่ใช้กับการตอบคำถามภาษาไทยที่มีประสิทธิภาพสูงขึ้น
3. สามารถสร้างชุดข้อมูลภาษาไทยเพิ่มเติม เพื่อนำมาฝึกสอนและเพิ่มประสิทธิภาพให้กับแบบจำลองสำหรับตอบคำถามได้

1.5 วิธีการดำเนินการวิจัย

วิธีการดำเนินงานวิจัย สามารถแบ่งออกเป็นขั้นตอน ดังแสดงในตารางที่ 1 ดังนี้

1. เลือกหัวข้องานวิจัย
2. ศึกษาวรรณกรรมที่เกี่ยวข้องกับหัวข้องานวิจัย
3. ศึกษาชุดข้อมูลการตอบคำถามภาษาไทย พร้อมจัดเตรียมข้อมูลให้อยู่ในรูปแบบที่พร้อมใช้งาน
4. ออกแบบและพัฒนาวิธีที่นำเสนอ ได้แก่ การประมวลผลข้อมูลภาษาไทยก่อนใช้งาน ขั้นตอนการสร้างคู่คำถาม-คำตอบ รวมถึง ออกแบบการวัดผลที่เหมาะสมกับภาษาไทย
5. ประเมินและตรวจสอบผลการทดลอง
6. วิเคราะห์และสรุปผลการทดลอง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

7. เขียนรายงานวิจัย
8. ส่งรายงานวิจัย
9. จัดทำวิทยานิพนธ์
10. ส่งรายงานวิจัย

1.6 ผลงานวิจัยที่ตีพิมพ์

Enhance Text-to-Text Transfer Transformer with Generated Questions for Thai Question Answering โดย ภูริ ฟักมงคล และ พีรพล เวทีกูล ในวารสารระดับนานาชาติ Applied Sciences (ISSN 2076-3417) อยู่ในฐานข้อมูล **Scopus/SCImago Journal Rank (SJR): Q2/ Impact Factor (2020), 2.679**

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตาราง 1 แผนภูมิแกนต์แสดงแผนการวิจัย

งาน	2562		2563		2564	
	ภาคต้น	ภาคปลาย	ภาคต้น	ภาคปลาย	ภาคต้น	ภาคปลาย
เลือกหัวข้องานวิจัย		X				
ศึกษางานที่เกี่ยวข้อง		X	X			
ศึกษาและจัดเตรียมชุดข้อมูลการตอบแบบสอบถาม			X			
ออกแบบและพัฒนาวิธีที่นำเสนอ			X	X		
ประเมินผลตรวจสอบผล				X		
วิเคราะห์และสรุปผลการทดลอง				X	X	
เขียนรายงานวิจัย					X	
ส่งรายงานวิจัย					X	
จัดทำวิทยานิพนธ์					X	
เสนอผลงานวิจัย					X	

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 2
แนวคิดและทฤษฎีที่เกี่ยวข้อง

ในบทนี้นำเสนอทฤษฎีที่เกี่ยวข้องกับงานวิจัยนี้ ประกอบด้วย การตอบคำถามในการประมวลผลภาษาธรรมชาติ กลไกความสนใจ ทรานส์ฟอร์เมอร์ RoBERTa และ Text-to-Text Transfer Transformer

2.1 การตอบคำถามในการประมวลผลภาษาธรรมชาติ

การตอบคำถาม เป็นงานทางด้านการประมวลผลภาษาธรรมชาติ ที่ต้องการให้เครื่องคอมพิวเตอร์สามารถเข้าใจบทความและตอบคำถามที่เกี่ยวข้องกับบทความได้ การตอบคำถามโดยใช้บทความ บางครั้งถูกเรียกว่าความเข้าใจการอ่านของเครื่อง (Machine Reading Comprehension) หมายความว่าในการตอบคำถาม เครื่องคอมพิวเตอร์จะต้องสามารถสกัดคำตอบที่ถูกต้อง และสร้างคำตอบที่เกี่ยวข้องกับบทความที่ให้ได้ ซึ่ง Shanshan Liu และคณะ [8] ได้จำแนกกลุ่มของการตอบคำถามโดยใช้บทความทั้งหมด 4 ประเภท ตามลักษณะการตอบคำถามในชุดข้อมูล ได้แก่ การทดสอบแบบเติมคำ (Cloze test) การทดสอบปรนัย (Multiple Choices) การสกัดช่วงข้อความ (Span Extraction) และ การตอบอิสระ (Free Answering)

การทดสอบแบบเติมคำ จะมีลักษณะของคำถามเป็นประโยคที่มีบางคำถูกลบออก สิ่งที่คอมพิวเตอร์ต้องทำคือ หาคำที่เหมาะสมแล้วนำมาเติมในคำถามจากบทความที่ให้ ตัวอย่างของชุดข้อมูลที่มีงานเป็นการทดสอบแบบเติมคำคือ CNN & Daily Mail [9]

การทดสอบปรนัย เป็นงานที่คอมพิวเตอร์จะต้องตอบคำถามโดยการเข้าใจบทความที่ให้คู่กับคำถาม และตอบคำถามจากตัวเลือกที่ให้ไว้ งานการตอบคำถามประเภทนี้ได้แรงบันดาลใจมาจากการสอบวัดระดับทางภาษา ตัวอย่างของชุดข้อมูลคือ RACE [10]

การสกัดช่วงข้อความ เป็นงานที่คอมพิวเตอร์จะต้องทำการเลือกช่วงข้อความที่ติดกันอย่างถูกต้องจากในบทความเพื่อตอบคำถาม ทำให้คอมพิวเตอร์จะต้องสามารถเข้าใจภาษาธรรมชาติที่อยู่บนบทความ เพื่อที่จะสามารถเลือกช่วงข้อความที่เป็นคำตอบได้อย่างแม่นยำ ชุดข้อมูลที่นิยมใช้ในงานวิจัยของการตอบคำถามประเภทนี้คือ Stanford Question Answering Dataset (SQuAD) [11]

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

โดยที่การตอบคำถามในงานวิจัยนี้จัดอยู่ในประเภทการสกัดช่วงข้อความ ซึ่งมีลักษณะของข้อมูลตามตารางที่ 2

ตาราง 2 ตัวอย่างของข้อมูลการตอบคำถามภาษาไทย ประเภทการสกัดชุดข้อความ

คำถาม	สฺวี เปย์หง เป็นนักจิตรกรชาวอะไร
บทความ	สฺวี เปย์หง สฺวี เปย์หง (; 19 กรกฎาคม ค.ศ. 1895 – 26 กันยายน ค.ศ. 1953) เป็นนักจิตรกรชาว**จีน** เกิดที่มณฑลเจียงซู เขาเป็นที่รู้จักมากที่สุดในการวาดภาพม้าและนก นอกจากนี้เขายังได้รับการยกย่องว่าเป็นหนึ่งในคนแรกที่จะสร้างภาพวาดสีน้ำมันที่มีรูปแบบจีนมหากาพย์ - การแสดงความสามารถของเขาคือใช้เทคนิคของศิลปะตะวันตกแบบระดับสูงภาพวาดของเขา
คำตอบ	จีน

การตอบอิสระ เป็นงานที่คอมพิวเตอร์จะต้องตอบคำถามโดยใช้ความรู้จากบทความเช่นเดียวกับการสกัดช่วงข้อความ แต่งานนี้ คอมพิวเตอร์สามารถสร้างคำตอบได้อย่างอิสระโดยไม่จำเป็นต้องใช้ข้อความจากบทความ ทำให้คอมพิวเตอร์จะต้องนำความรู้มาสรุปแล้วสร้างคำตอบสำหรับตอบคำถาม ตัวอย่างของชุดข้อมูลคือ MS MARCO [12]

2.2 กลไกความสนใจ (Attention Mechanism)

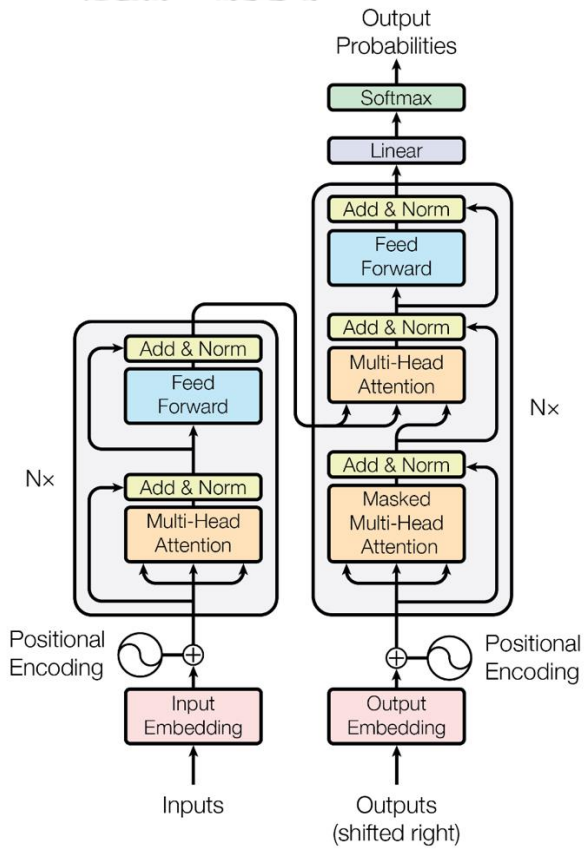
กลไกความสนใจ เป็นกลไกที่ช่วยให้แบบจำลองการเรียนรู้เชิงลึกแบบลำดับต่อลำดับ (Sequence-to-Sequence Model) สามารถกำหนดความสนใจกับข้อมูลได้ โดยใช้คะแนนความสนใจในการระบุว่าข้อมูลส่วนใดสำคัญที่ส่งผลต่อคำตอบ

กลไกความสนใจถูกเสนอขึ้นครั้งแรกโดย Dzmitry Bahdanau และ KyungHyun ChoYoshua Bengio [13] ซึ่งเป็นงานวิจัยทางด้านการแปลภาษาด้วยเครื่อง (Machine Translation) โดยที่กลไกความสนใจสามารถช่วยให้แบบจำลองสามารถเลือกคำที่สำคัญสำหรับใช้ในการแปลได้ นอกจากนี้ กลไกความสนใจยังถูกใช้งานอย่างแพร่หลายในแบบจำลองที่เกี่ยวข้องกับการประมวลผลภาษาธรรมชาติ เช่น Minjoon Seo และคณะ [1] ได้พัฒนาแบบจำลอง Bidirectional

Attention Flow (BiDAF) ได้ใช้กลไกความสนใจคู่กับแบบจำลอง LSTM เพื่อสร้างแบบจำลองสำหรับตอบคำถาม เป็นต้น

2.3 ทรานส์ฟอร์เมอร์ (Transformer)

ทรานส์ฟอร์เมอร์เป็นแบบจำลองที่สามารถเรียนรู้ข้อมูลในลักษณะเป็นลำดับได้ เช่นเดียวกับแบบจำลองประเภทนิวรอลเน็ตเวิร์กแบบวกกลับ (Recurrent Neural Network) แต่มีการใช้เพียงกลไกความสนใจเท่านั้น โดยถูกเสนอโดย Ashish Vaswani และคณะ [4]

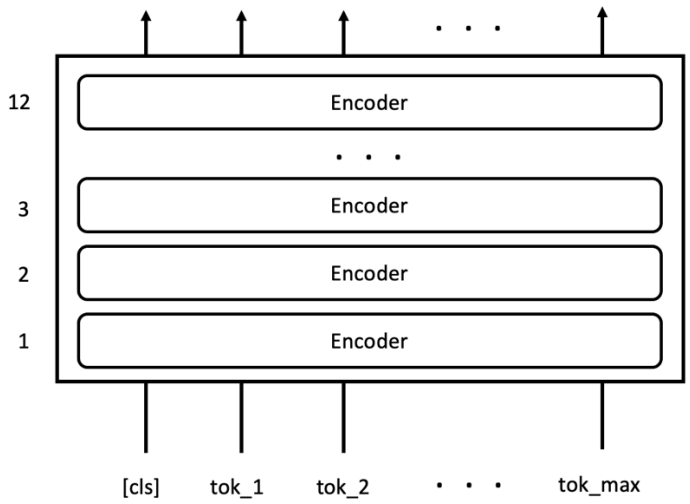


รูป 1 โครงสร้างของแบบจำลองทรานส์ฟอร์เมอร์ ซึ่งประกอบด้วย Encoder และ Decoder [4]

ทรานส์ฟอร์เมอร์เป็นแบบจำลองที่มีลักษณะเป็น Encoder-Decoder โดยมีกลไกความสนใจประเภท Self-attention ในชั้น Multi-Head Attention และ Fully-connected layer ประกอบทั้งส่วน Encoder สำหรับใช้รับลำดับของข้อมูลขาเข้า เพื่อแปลงข้อมูลให้อยู่ในลักษณะของตัวแทน (Representation) ก่อนที่จะถูกส่งไปยัง Decoder เพื่อพยากรณ์ผลลัพธ์ออกมาทีละค่า ดังรูปที่ 1

อีกทั้ง แบบจำลองทรานส์ฟอร์เมอร์ยังมีการเข้ารหัสตำแหน่ง (Positional Encoding) เพื่อระบุตำแหน่ง ต่าง ๆ ของชุดข้อมูลขาเข้าได้ ส่งผลให้แบบจำลองทรานส์ฟอร์เมอร์สามารถเรียนรู้ข้อมูลในลักษณะลำดับได้

ทรานส์ฟอร์เมอร์ได้รับความนิยมในการทำแบบจำลองทางด้านการประมวลผลภาษาธรรมชาติเป็นอย่างมาก โดยเฉพาะการทำแบบจำลองตัวแทนภาษา (Language Representation Model) เช่น แบบจำลอง Bidirectional Encoder Representations from Transformers (BERT) ที่ถูกเสนอโดย Jacob Devlin และคณะ [5] ได้สร้างแบบจำลองโดยนำโครงสร้างของทรานส์ฟอร์เมอร์ในส่วนของ Encoder ตามรูปที่ 2 เพื่อนำมาฝึกสอนให้แบบจำลองสามารถสร้างการฝังคำตามบริบท (Contextual Word Embedding) อีกทั้งยังสามารถใช้ในการถ่ายทอดการเรียนรู้ (Transfer Learning) เพื่อทำการฝึกสอนเพิ่มในงานเฉพาะด้านได้



รูป 2 โครงสร้างของแบบจำลอง BERT ที่ประกอบด้วย Encoder ของแบบจำลองทรานส์ฟอร์เมอร์

2.4 RoBERTa

Robustly optimized BERT approach (RoBERTa) [6] เป็นแบบจำลองตัวแทนภาษา โดยใช้โครงสร้างเป็นทรานส์ฟอร์เมอร์ส่วนของ Encoder เช่นเดียวกับ BERT แต่มีขั้นตอนการฝึกสอนก่อน (Pretraining) และตัวตัดคำ (Tokenizer) ที่ต่างกัน โดยที่ RoBERTa จะมีการเปลี่ยนงานระหว่างกันฝึกสอน ได้แก่ การเปลี่ยนการปกคลุม (Masking) ทุกรอบของการฝึกสอน, การเลือกที่จะ

ไม่ใช้การพยากรณ์ประโยคถัดไป (Next Sentence Prediction) และเปลี่ยนตัวตัดคำเป็น Byte-Pair Encoding (BPE)

2.5 Text-to-Text Transfer Transformer และ Multilingual Text-to-Text Transfer Transformer

แบบจำลอง Text-to-Text Transfer Transformer (T5) ถูกนำเสนอโดย Colin Raffel และคณะ [7] เป็นแบบจำลองที่นำส่วนประกอบของแบบจำลองทรานส์ฟอร์เมอร์มาใช้งานทั้งในส่วน Encoder และ Decoder เป็นส่วนประกอบ โดยจุดประสงค์ของแบบจำลอง T5 คือ ต้องการพัฒนาแบบจำลองที่สามารถทำงานกับโจทย์ทางด้านการประมวลผลภาษาธรรมาชาติได้หลากหลาย โดยการมองปัญหาเป็นลักษณะข้อความเป็นข้อความ (Text-to-Text) กล่าวคือ เป็นแบบจำลองที่รับข้อความเป็นข้อมูลขาเข้า เพื่อพยากรณ์ข้อความออกมา จึงสามารถนำแบบจำลองไปใช้กับโจทย์ที่หลากหลาย เช่น การตอบคำถาม การสรุปความ (Document Summarization) หรือ การจำแนกอารมณ์ (Sentiment Classification)

ต่อมามีการนำแบบจำลอง T5 มาทำการฝึกสอนก่อน [14] โดยใช้ข้อมูลที่มีจำนวนภาษาทั้งหมด 101 ภาษา โดยมีภาษาไทยประกอบ แบบจำลอง T5 แบบรองรับหลายภาษา ถูกเรียกว่า Multilingual Text-to-Text Transfer Transformer (mT5)

2.6 การวัดผลแบบจำลองการตอบคำถาม

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

การวัดผลแบบจำลองการตอบคำถามประเภทสกัดช่วงข้อความ ตามงานวิจัยจะวัดผลโดยใช้ 2 ตัวชี้วัด ได้แก่ คะแนนความถูกต้อง (Exact Match) และ คะแนนเอฟวัน (F1 Score) โดยมีรายละเอียดการคำนวณดังนี้

คะแนนความถูกต้อง จะใช้วัดว่าแบบจำลองสามารถตอบคำถามที่ตรงกับวลีผลเฉลยได้อย่างถูกต้องทุกประการเป็นสัดส่วนเท่าไร กำหนดให้ $ground_truth$ เป็นวลีผลเฉลย และ $prediction$ เป็นคำตอบที่ได้จากแบบจำลอง จะสามารถคิดคะแนนความถูกต้องได้ ดังนี้

$$Exact\ Match\ (EM) = \frac{\sum_{i=0}^{N} Match(ground_truth_i,\ prediction_i)}{N} \quad 1$$

คะแนนเอฟวัน เป็นตัววัดประสิทธิภาพโดยทั่วไปของงานด้านการจำแนก (Classification Task) โดยการวัดประสิทธิภาพของการตอบคำถามจะพิจารณาเป็นคำ (Word Token) ของวลีผลเฉลยและคำตอบที่ได้จากแบบจำลอง โดยคะแนนเอฟวันจะคำนวณได้จากค่าความเที่ยง (Precision) และค่าการเรียกคืน (Recall) ดังนี้

กำหนดให้

- TP (True Positive) คือจำนวนคำที่อยู่ในคำตอบจากแบบจำลองและวลีผลเฉลย
- FP (False Positive) คือจำนวนคำที่อยู่ในคำตอบจากแบบจำลองแต่ไม่อยู่ในวลีผลเฉลย
- FN (False Negative) คือจำนวนคำที่ไม่อยู่ในคำตอบจากแบบจำลองแต่อยู่ในวลีผลเฉลย

โดยที่

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ทั้งนี้ คะแนนเอฟวันที่ใช้วัดผลจะเป็นคะแนนเอฟวันแบบเฉลี่ยรวม (Macro F1) หมายความว่า เป็นการนำคะแนนเอฟวันของทุก ๆ คำตอบมาทำการเฉลี่ยกัน กำหนดให้ $F1_i$ เป็นคะแนนเอฟวันของคำถามข้อ i จะสามารถคำนวณคะแนนเอฟวันแบบเฉลี่ยรวมได้ดังนี้

$$F1 = \frac{\sum_{i=1}^{N} F1_i}{N} \quad (5)$$

บทที่ 3
เอกสารและงานวิจัยที่เกี่ยวข้อง

ในบทนี้นำเสนอเอกสารและงานวิจัยที่เกี่ยวข้อง ประกอบด้วย งานวิจัยทางด้านการตอบคำถามในปัจจุบัน งานวิจัยทางด้านการตอบคำถามในภาษาไทย งานวิจัยทางด้านการสร้างชุดข้อมูล และ WangchanBERTa

3.1 งานวิจัยทางด้านการตอบคำถามในปัจจุบัน

ในปัจจุบัน แบบจำลองทรานส์ฟอร์เมอร์ได้รับความนิยมในงานการประมวลผลภาษาธรรมชาติด้านการตอบคำถามอย่างแพร่หลาย เนื่องจากแบบจำลองนี้ได้สามารถทำประสิทธิภาพได้สูงขึ้นอย่างก้าวกระโดดเมื่อเทียบกับแบบจำลองที่เคยใช้

แบบจำลอง BERT ซึ่งถือว่าเป็นแบบจำลองตัวแทนภาษาแบบจำลองแรก ที่นำส่วนประกอบของแบบจำลองทรานส์ฟอร์เมอร์มาใช้งานและสามารถถ่ายทอดการเรียนรู้ได้ ได้รับคะแนนความถูกต้อง (Exact Match) และคะแนน F1 ที่ 80.8 และ 88.5 ตามลำดับโดยใช้ $BERT_{BASE}$ และสามารถได้คะแนน 84.1 และ 90.9 ตามลำดับโดยใช้ $BERT_{LARGE}$ [5] ในขณะที่ BiDAF ซึ่งเป็นแบบจำลองที่ใช้กลไกความสนใจคู่กับแบบจำลอง LSTM ได้คะแนนความถูกต้อง และคะแนน F1 ที่ 68.0 และ 77.3 ตามลำดับ [1] เมื่อทดสอบบนชุดข้อมูล SQuAD 1.1 [11] ทำให้เห็นว่าแบบจำลองประเภททรานส์ฟอร์เมอร์สามารถช่วยเพิ่มประสิทธิภาพของการตอบคำถามได้

ต่อมา Mandar Josh และคณะ [15] ได้แก้ไขแบบจำลอง BERT ให้สามารถพยากรณ์ช่วงข้อความได้ โดยใช้ชื่อว่า SpanBERT โดยการแก้ไขขั้นตอนการปกคลุม จากเดิมที่ BERT จะปกคลุมเป็นคำ แต่ SpanBERT จะทำการปกคลุมเป็นช่วงของข้อความ และเพิ่มการพยากรณ์จากขอบของช่วงข้อความทำให้แบบจำลองต้องเรียนรู้การพยากรณ์ช่วงข้อความโดยใช้คำที่อยู่บริเวณขอบของช่วงข้อความ จากการเปลี่ยนแปลงดังกล่าว ส่งผลให้แบบจำลอง SpanBERT สามารถพยากรณ์ช่วงข้อความได้ดีกว่า ซึ่งเป็นความสามารถที่เหมาะกับงานการตอบคำถาม ทั้งนี้ SpanBERT ได้แก้ไขแบบจำลองโดยใช้ $BERT_{LARGE}$ โดยที่ได้คะแนนความถูกต้อง และคะแนน F1 ที่ 88.8 และ 94.6 ตามลำดับ

3.2 งานวิจัยทางด้านการตอบคำถามในภาษาไทย

Hatsanai Decha และ Karn Patanukhom ได้สร้างระบบการตอบคำถามภาษาไทย [2] โดยใช้การสกัดคำสำคัญ (Keyword) จากคำถามและนำไปหาชุดของคำตอบจากบทความ จากนั้นจึงนำมาหาคำตอบจากชุดของคำตอบโดยใช้คะแนนที่ได้จากการจับคู่ของคำถามและประโยคที่มีคำตอบเรียกว่า Word order consistency ซึ่งหากมีคะแนนที่สูง หมายความว่าประโยคที่มีคำตอบนั้นมีความคล้ายกับคำถามนั้น ระบบนี้ได้รับการทดลองบนเอกสาร 50 บทความ และมีจำนวนคำถามทั้งหมด 210 คำถาม

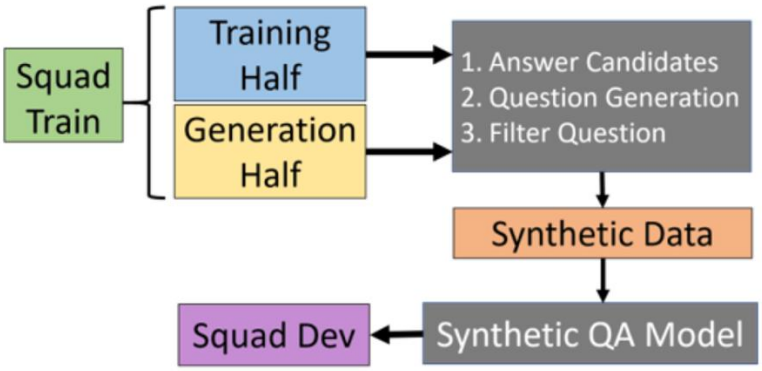
Theerit Lapchaicharoenkit ได้สร้างแบบจำลองการตอบคำถามภาษาไทย [3] โดยได้แก้ไขแบบจำลอง BiDAF ให้สามารถรองรับการตอบคำถามได้ 2 ประเภท คือ การสกัดช่วงข้อความและการตอบคำถามประเภทใช่หรือไม่ โดยการเพิ่มตัวจำแนกคำถามและตัวตอบคำถามประเภทใช่หรือไม่ไว้ในแบบจำลอง การตอบคำถามนั้น จะเลือกจากตัวจำแนกคำถาม ว่าจะเลือกตอบคำถามประเภทใช่หรือไม่ หรือจะเลือกตอบเป็นช่วงข้อความ อีกทั้งยังใช้การฝังคำตามบริบท โดยใช้แบบจำลอง BERT ที่มีการฝึกสอนก่อนโดยใช้ภาษาไทยเพียงอย่างเดียว ได้ทำการทดลองบนชุดข้อมูลสำหรับแข่งในงาน National Software Contest ปี พ.ศ. 2561-2562 ซึ่งประกอบด้วยคู่คำถาม-คำตอบทั้งหมด 17,000 คู่ ประกอบด้วยคำถามประเภทการสกัดช่วงข้อความทั้งหมด 15,000 คำถาม และคำถามประเภทใช่หรือไม่ทั้งหมด 2,000 คำถาม

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

3.3 งานวิจัยทางด้านการสร้างชุดข้อมูล

Bhuwan Dhingra และคณะ [16] ได้นำเสนอวิธีการเพิ่มประสิทธิภาพของแบบจำลองการตอบคำถาม โดยใช้วิธีเพิ่มจำนวนชุดข้อมูลให้มากขึ้น เนื่องจากปัญหาที่ชุดข้อมูลการตอบคำถามในบางขอบเขต (Domain) นั้นมีจำกัด วิธีการที่นำเสนอคือ ใช้การสร้างคำถามประเภทเติมคำโดยการสกัดคำถามและคำตอบโดยใช้โครงสร้างการเขียนบทความของภาษาอังกฤษที่มักจะมีการแนะนำถึงภาพรวมของบทความไว้ที่ส่วนต้นของบทความ วิธีนี้ช่วยเพิ่มประสิทธิภาพของแบบจำลองได้ ทางผู้วิจัยได้ทดลองโดยการนำมาฝึกสอนโดยใช้แบบจำลอง BiDAF พบว่าการฝึกสอนก่อนด้วยชุดข้อมูลที่สร้างขึ้น จากนั้นจึงทำการฝึกสอนด้วยชุดข้อมูลจริง สามารถเพิ่มประสิทธิภาพได้ 0.32% และ 0.11% สำหรับคะแนนความถูกต้อง และคะแนน F1 ตามลำดับ

Raul Puri และคณะ [17] ได้นำเสนอวิธีการสร้างชุดข้อมูลเพิ่มโดยใช้แบบจำลองการเรียนรู้เชิงลึก โดยใช้ 3 แบบจำลองในการสร้างชุดข้อมูล ได้แก่ แบบจำลองสำหรับเลือกคำตอบ (Answer Generation model) ซึ่งให้แบบจำลอง BERT เรียนรู้วิธีการเลือกคำตอบจากบทความ แบบจำลองการสร้างคำถาม (Question Generation model) ซึ่งให้แบบจำลอง GPT-2 เรียนรู้การสร้างคำถามจากบทความและคำตอบ และแบบจำลองสำหรับกรองคำถาม สำหรับใช้ในขั้นตอนกรองสองทิศทาง (Roundtrip Filtration) ซึ่งเป็นการนำแบบจำลอง BERT โดยการฝึกสอนให้ตอบคำถามโดยใช้ข้อมูลจริง มาพยากรณ์คำตอบจากบทความและคำถามที่สร้างขึ้น ถ้าสามารถพยากรณ์คำตอบได้ตรงกับคำตอบที่เลือกจากแบบจำลองสำหรับเลือกคำตอบ ก็ถือว่าข้อมูลที่สร้างมาสามารถยอมรับได้ อีกทั้งผู้วิจัยได้หลีกเลี่ยงการสูญเสียของข้อมูลที่สร้างเพิ่มโดยการสร้างคำถามจำนวน 2 คำถามต่อ 1 คู่ของบทความ-คำตอบ โดยขั้นตอนการสร้างชุดข้อมูลเสริม และการฝึกสอนได้สรุปในรูปที่ 3 ทั้งนี้ การทดลองของผู้วิจัยสามารถสร้างบทความเพิ่มโดยใช้บทความจาก Wikipedia ได้ถึง 19,925,130 คู่ และเมื่อนำมาฝึกสอนเพิ่มจากข้อมูล SQuAD 1.1 พบว่าสามารถเพิ่มประสิทธิภาพได้ 1.7% และ 1.2% สำหรับคะแนนความถูกต้อง และคะแนน F1 ตามลำดับ



รูป 3 ขั้นตอนการฝึกสอนและสร้างชุดข้อมูลเพิ่มโดยใช้ชุดข้อมูล SQuAD [17]

3.4 WangchanBERTa

Lalita Lowphansirikul และคณะ [18] ได้ทำการฝึกสอนแบบจำลอง RoBERTa โดยใช้ชุดข้อมูลภาษาไทยที่มีขนาดถึง 78 GB ซึ่งประกอบด้วยข้อมูลจากโซเชียลมีเดีย หัวข้อข่าว และข้อมูลภาษาไทยที่สามารถเข้าถึงได้อย่างสาธารณะอื่น ๆ โดยที่ได้ผ่านการประมวลผลข้อความ (Text-Processing) ด้วยวิธีการที่เหมาะสมกับภาษาไทย เช่น การเลือกใช้ตัวตัดคำ SentencePiece [19] ที่

ผ่านการฝึกสอนด้วยชุดข้อมูลภาษาไทย แบบจำลอง WangchanBERTa สามารถให้ผลที่ดีกว่าแบบจำลองที่รองรับภาษาไทยอื่น ๆ เช่น Multilingual BERT (mBERT) [5]

3.5 สรุปการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง

จากการศึกษาวรรณกรรมและงานวิจัยที่เกี่ยวข้องกับงานวิจัยนี้ จึงสามารถสรุปวิธีการและแบบจำลองที่สามารถนำมาใช้กับงานวิจัยนี้ได้ ดังนี้

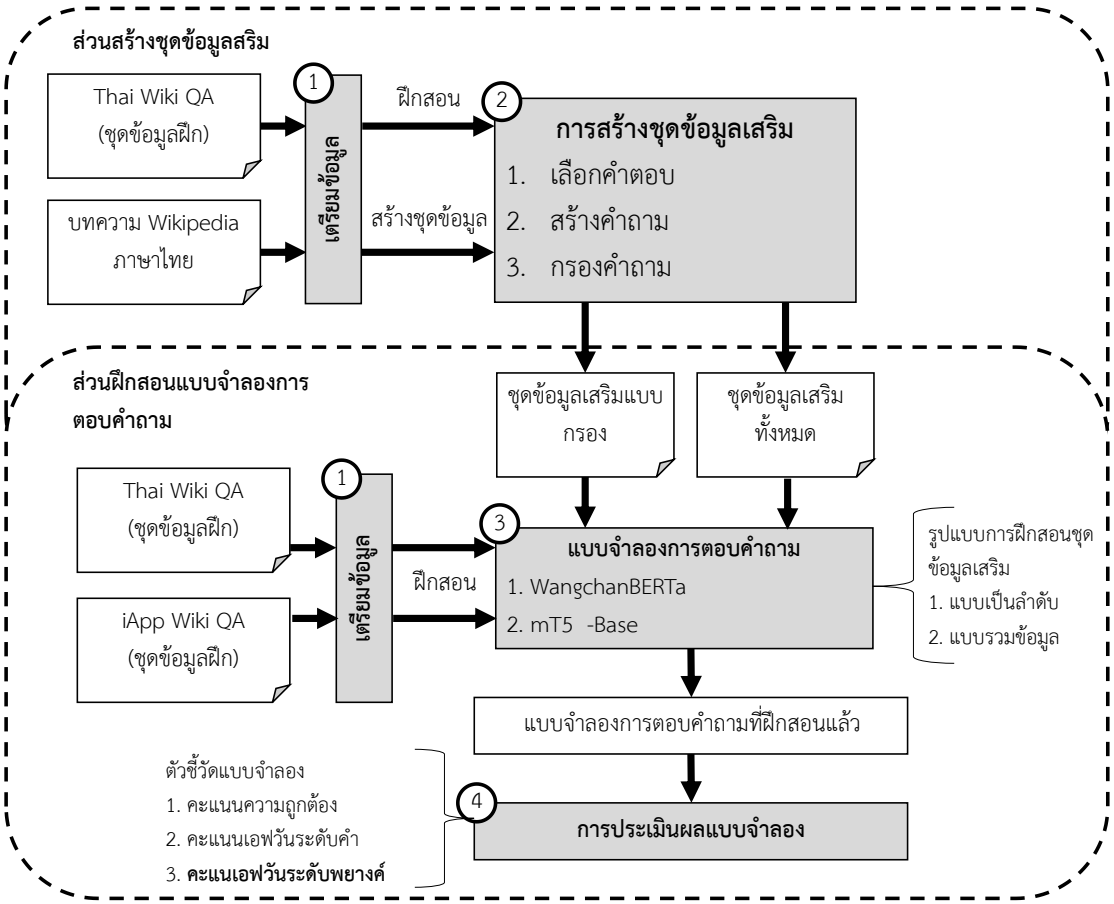
ในการเลือกแบบจำลองตัวแทนภาษา จำเป็นต้องเลือกแบบจำลองที่ฝึกสอนก่อนด้วยข้อมูลภาษาไทย เพื่อให้แบบจำลองสามารถใช้งานกับชุดข้อมูลภาษาไทยได้ ทางผู้วิจัยเลือกนำแบบจำลอง WangchanBERTa และแบบจำลอง mT5 มาใช้ในงานวิจัยชิ้นนี้ เนื่องจากแบบจำลอง WangchanBERTa ผ่านการฝึกสอนก่อนด้วยชุดข้อมูลภาษาไทยและการประมวลผลข้อความได้ถูกออกแบบให้เหมาะสมกับภาษาไทย และแบบจำลอง mT5 เป็นแบบจำลองที่รองรับหลากหลายภาษาโดยมีภาษาไทยเป็นหนึ่งในนั้น อีกทั้ง แบบจำลอง mT5 เป็นแบบจำลองที่สามารถทำงานได้หลากหลายรูปแบบซึ่งเหมาะสมกับงานวิจัยนี้

สำหรับวิธีการสร้างชุดข้อมูลคำถาม-คำตอบ ทางผู้วิจัยเลือกวิธีการของ Raul Puri และคณะ [17] เนื่องจากภาษาไทยเป็นภาษาที่ไม่มีการเขียนตามโครงสร้างที่ชัดเจนเหมือนภาษาอังกฤษ ส่งผลให้ไม่สามารถใช้งานวิธีการของ Bhuwan Dhingra และคณะ [16] ซึ่งอาศัยการสร้างชุดข้อมูลโดยการสกัดจากโครงสร้างการเขียนของบทความได้ อีกทั้ง ภาษาไทยเป็นภาษาที่มีความกำกวมสูง การใช้งานแบบจำลองการเรียนรู้เชิงลึกในการสร้างชุดข้อมูลเพิ่มจึงมีความเหมาะสมมากกว่า

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

บทที่ 4
แนวคิดและวิธีการดำเนินงาน

ในบทนี้จะอธิบายส่วนประกอบของกรอบงานการตอบคำถามภาษาไทยตามที่ได้เสนอ เช่น ขั้นตอนการเตรียมข้อมูลสำหรับข้อมูลภาษาไทย การสร้างชุดข้อมูลเสริม การฝึกสอนแบบจำลองการตอบคำถามภาษาไทยพร้อมกับข้อมูลเสริม และ การประเมินผลแบบจำลอง ขั้นตอนของกรอบงานและส่วนประกอบต่าง ๆ ได้อธิบายในรูปที่ 4



รูป 4 กรอบงานการตอบคำถามภาษาไทยที่นำเสนอในวิทยานิพนธ์นี้

4.1 ขั้นตอนการเตรียมข้อมูล

ในทุก ๆ การทดลอง ข้อมูลภาษาไทยจะต้องผ่านการประมวลผลข้อมูลภาษาไทยก่อนด้วยวิธีการที่เหมาะสมก่อนที่จะถูกนำไปใช้งาน ทั้งในส่วนของการฝึกสอนและการทดสอบแบบจำลอง ในขั้นตอนแรกของการประมวลผลคือ เปลี่ยนตัวอักษรภาษาอังกฤษในข้อมูลให้เป็นตัวพิมพ์เล็กก่อน

จากนั้นในขั้นตอนที่สอง คือการนำข้อมูลไป normalize ซึ่งขั้นตอนนี้จะทำให้ข้อความในข้อมูลถูกแก้ไขและทำให้อยู่ในรูปแบบเดียวกัน โดยการนำตัวอักษรที่ซ้ำกันออก และทำการเปลี่ยนการจัดเรียงตัวอักษรให้ถูกต้อง ในขั้นตอนนี้ สามารถที่จะช่วยลดจำนวนข้อความที่ผิดและสามารถช่วยให้แบบจำลองทำงานได้อย่างถูกต้อง ผู้เขียนได้ใช้วิธีการจาก PyThaiNLP ในการทำ normalize [20] ด้วยวิธีดังกล่าว

4.2 ขั้นตอนการฝึกสอนแบบจำลองสำหรับสร้างชุดข้อมูลเสริม

วิธีในการสร้างชุดข้อมูลเสริม ซึ่งประกอบด้วยคู่ของคำถามและคำตอบนั้น ผู้เขียนได้ใช้วิธีเดียวกับ Raul Puri และคณะ [17] โดยประกอบด้วย 3 ขั้นตอน ได้แก่ (1) การเลือกคำตอบ (2) การสร้างคำถาม และ (3) การกรองคำถาม วิธีนี้สามารถสร้างข้อมูลเสริมซึ่งประกอบด้วย บทความ c คำถาม q และ คำตอบ a โดยใช้เซ็ตของบทความ A โดยใช้ความน่าจะเป็น $p(q, a|c)$

ข้อแตกต่างระหว่างงานวิจัยของ Raul Puri และคณะ กับวิทยานิพนธ์นี้คือการเลือกแบบจำลองพื้นฐานในขั้นตอนการสร้างชุดข้อมูลเสริม ในวิทยานิพนธ์นี้ แบบจำลองต่าง ๆ จะถูกเลือกให้มีคุณสมบัติของแบบจำลองให้ใกล้เคียงกับงานของ Raul Puri และคณะ เช่น เลือกใช้แบบจำลอง WangchanBERTa แทนการใช้แบบจำลอง BERT หรือการเลือกใช้แบบจำลอง mT5 แทนการใช้แบบจำลอง GPT-2 ซึ่งมีความเหมาะสมกับการใช้งานกับภาษาไทยมากกว่า การใช้แบบจำลองในขั้นตอนการสร้างข้อมูลเสริมได้ถูกสรุปในตาราง 3 อย่างไรก็ตาม ในขั้นตอนการเลือกคำตอบ ผู้เขียนได้เลือกใช้แบบจำลอง mT5 แทนการใช้แบบจำลอง WangchanBERTa เนื่องจากทางผู้เขียนพบว่าแบบจำลอง mT5 สามารถสร้างคำตอบที่เหมาะสมกว่าแบบจำลอง WangchanBERTa เนื่องจากคำตอบที่ถูกเลือกจาก WangchanBERTa มีลักษณะเป็นวลีเป็นส่วนมาก ซึ่งไม่เหมาะกับการเป็นคำตอบสำหรับคำถาม

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

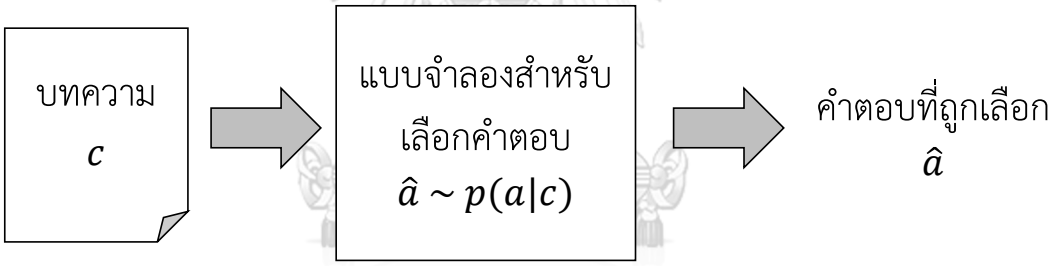
ตาราง 3 การเลือกแบบจำลองของวิทยานิพนธ์นี้เทียบกับงานวิจัยของ Raul Puri และคณะ ในขั้นตอนการสร้างชุดข้อมูลเสริม

งานวิจัย	การเลือกคำตอบ	การสร้างคำถาม	การกรองคำถาม
Raul Puri และคณะ	BERT	GPT-2	BERT
วิทยานิพนธ์นี้	mT5-Large	mT5-Large	WangchanBERTa

4.2.1 การฝึกสอนแบบจำลองสำหรับเลือกคำตอบ

เนื่องจากความยากและความกำกวมของภาษาไทย การหาคำตอบที่เหมาะสมกับบทความโดยใช้การทำตามกฎอาจไม่เพียงพอที่จะเลือกคำตอบที่มีคุณภาพได้ เนื่องจากเครื่องมือการประมวลผลภาษาธรรมชาติของภาษาไทยนั้นยังไม่มีประสิทธิภาพที่เพียงพอ ทำให้ไม่สามารถทำงานได้ถูกต้องกับทุก ๆ คำหรือประโยค ซึ่งส่งผลให้บางครั้งทำให้คุณสมบัติของข้อความถูกนำมาใช้อย่างไม่เหมาะสม

ในการแก้ปัญหานี้ ผู้เขียนจึงใช้แบบจำลองสำหรับเลือกคำตอบ $p(a|c)$ เพื่อเลือกคำตอบที่เหมาะสม $\hat{a}$ กับบทความ ในวิทยานิพนธ์นี้ ผู้เขียนได้เลือกใช้แบบจำลอง mT5-Large เป็นแบบจำลองพื้นฐาน โดยใช้บทความ c เป็นข้อมูลขาเข้า เพื่อให้แบบจำลองสามารถเรียนรู้การกระจายตัวของคำตอบในชุดข้อมูล ตามที่แสดงในรูปที่ 5 โดยที่คำตอบจะถูกเลือกจากคะแนนความน่าจะเป็นที่สูงที่สุด

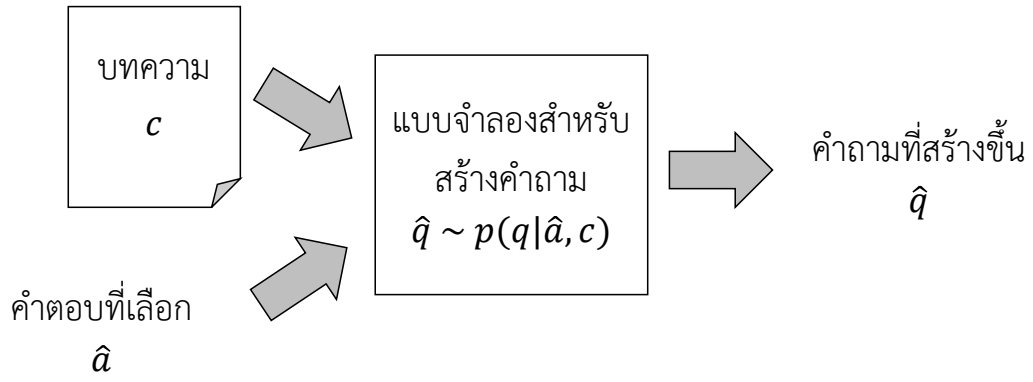


รูป 5 แสดงข้อมูลที่ใช้ในการฝึกสอนแบบจำลองสำหรับเลือกคำตอบ

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

4.2.2 การฝึกสอนแบบจำลองสำหรับสร้างคำถาม

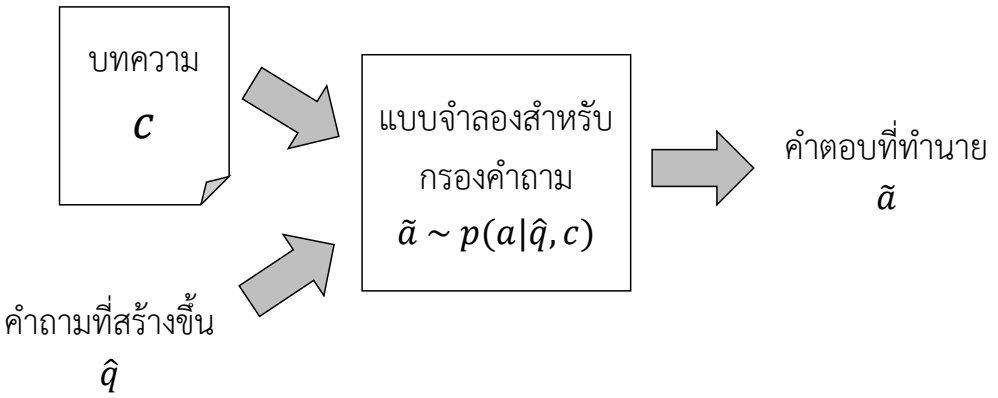
ในขั้นตอนการสร้างคำถามจะใช้ แบบจำลองสำหรับสร้างคำถาม $p(q|\hat{a}, c)$ ในการสร้างคำถามที่มีความเหมาะสมกับคำตอบและบทความ แบบจำลองนี้ใช้แบบจำลอง mT5-Large เป็นแบบจำลองพื้นฐานโดยใช้บทความ c และคำตอบที่เลือกโดยแบบจำลองสำหรับเลือกคำถาม $\hat{a}$ เป็นข้อมูลขาเข้าให้กับแบบจำลอง ตามที่แสดงในรูปที่ 6



รูป 6 แสดงข้อมูลขาเข้าที่ใช้ในการฝึกสอนแบบจำลองสำหรับสร้างคำถาม

4.2.3 การฝึกสอนแบบจำลองสำหรับกรองคำถาม

หลังจากได้คำถามที่สร้างขึ้นจากขั้นตอนการสร้างคำถาม $\hat{q}$ จากแบบจำลองการสร้างคำถาม และคำตอบที่เลือกจากขั้นตอนการเลือกคำตอบ $\hat{a}$ ซึ่งได้มาจากแบบจำลองสำหรับเลือกคำตอบ ก็จะทำให้ได้เป็นเซ็ตของข้อมูล 1 ชุดคือ $(c, \hat{q}, \hat{a})$ ที่สามารถนำไปใช้เป็นข้อมูลเสริมได้ทันที แต่ก่อนที่จำนำไปใช้งานจะเป็นจะต้องตรวจสอบก่อนว่าข้อมูลเสริมชุดนี้สามารถใช้งานและยอมรับได้ด้วยการใช้แบบจำลองสำหรับกรองคำถาม แบบจำลองนี้ ถูกฝึกสอนด้วยชุดข้อมูลการตอบคำถามโดยใช้ชุดข้อมูลจริงเป็นผลเฉลย หลังจากนั้น นำคำถามที่สร้างขึ้น $\hat{q}$ และบทความของข้อมูล c เข้าแบบจำลองสำหรับกรองคำถามเพื่อให้แบบจำลองทำนายคำตอบ $\tilde{a}$ ตามรูปที่ 7 จากนั้น ทำการเปรียบเทียบว่าคำตอบจากแบบจำลองสำหรับกรองคำถาม $\tilde{a}$ กับคำตอบจากข้อมูล $\hat{a}$ ที่ถูกเลือกโดยแบบจำลองสำหรับเลือกคำตอบ ตรงกันหรือไม่ ถ้าคำตอบทั้ง 2 ตรงกัน โดยวัดจากคะแนนเอฟวันจะเป็น 100 จะถือว่าข้อมูลนี้เป็นข้อมูลที่มีคุณภาพและสามารถยอมรับได้ ดังนั้นวิธีการในการสร้างชุดข้อมูลเสริมโดยรวมจะเป็นไปตามรูปที่ 8

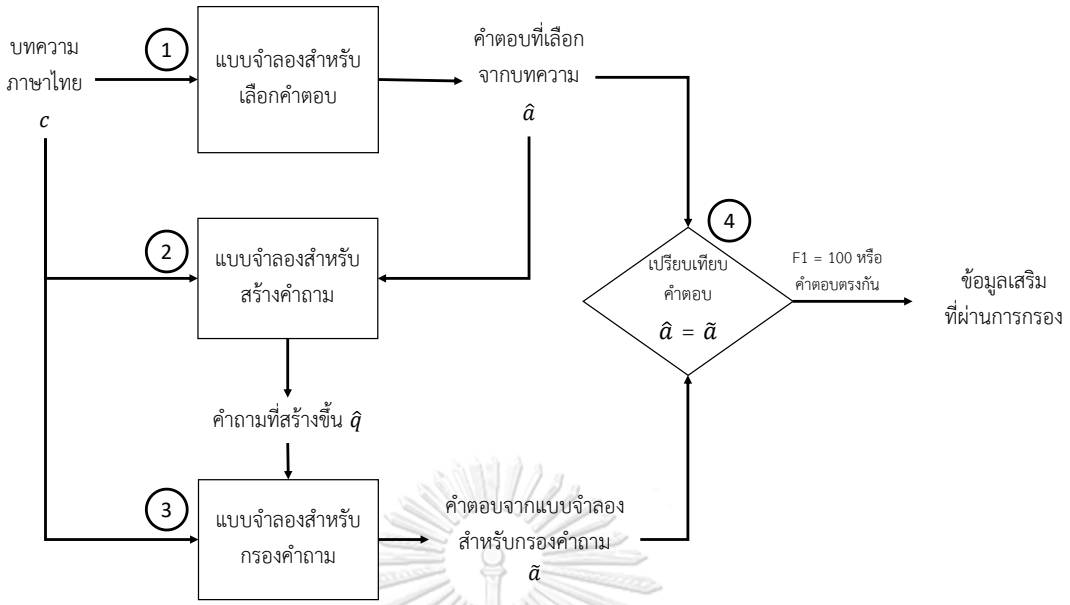


รูป 7 แสดงข้อมูลที่ใช้ในการฝึกสอนแบบจำลองสำหรับกรองคำถาม

ในขั้นตอนนี้ ผู้เขียนได้เลือกแบบจำลอง WangchanBERTa เป็นแบบจำลองพื้นฐาน เนื่องจากแบบจำลองนี้มีลักษณะใกล้เคียงกับงานของ Raul Puri และคณะ ที่เลือก BERT เป็นแบบจำลองพื้นฐานในขั้นตอนนี้ นอกจากนี้ แบบจำลอง WangchanBERTa ยังเหมาะสมกับภาษาไทยมากกว่า เนื่องจากเป็นแบบจำลองที่ผ่านการเรียนรู้ด้วยเอกสารภาษาไทยมาแล้ว โดยก่อนที่จะนำแบบจำลอง WangchanBERTa มาใช้สำหรับกรองคำถาม จะนำไปฝึกสอนด้วยงานการตอบคำถามภาษาไทยก่อน ด้วยชุดข้อมูลการตอบคำถามภาษาไทย

ในขั้นตอนการทดลอง ผู้เขียนได้ทดลองการใช้ชุดข้อมูลเสริมอยู่ 2 ชนิด ได้แก่ (1) ชุดข้อมูลเสริมแบบกรอง และ (2) ชุดข้อมูลเสริมทั้งหมด โดยที่ ชุดข้อมูลเสริมแบบกรอง จะเป็นเซ็ตของข้อมูล $(c, \hat{q}, \hat{a})$ ที่ผ่านขั้นตอนการกรองคำถาม ในขณะที่ ชุดข้อมูลเสริมทั้งหมด จะเป็นเซ็ตของข้อมูล $(c, \hat{q}, \hat{a})$ หลังจากผ่านขั้นตอนการสร้างคำถาม โดยที่จะผ่านขั้นตอนการกรองคำถามหรือไม่ก็ได้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



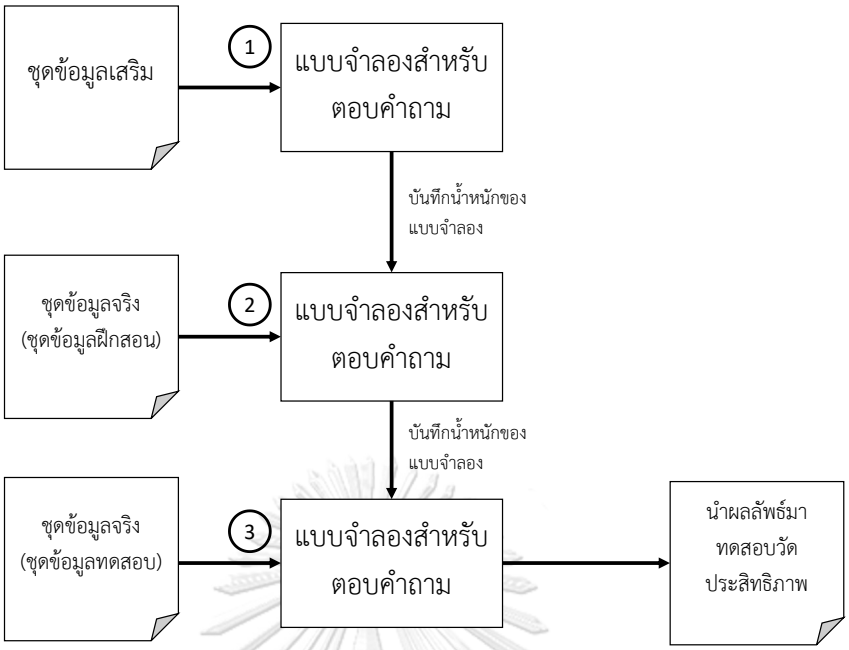
รูป 8 แผนภาพแสดงขั้นตอนการสร้างคู่คำถาม-คำตอบสำหรับชุดข้อมูลเสริมโดยรวม

4.3 ขั้นตอนการฝึกสอนแบบจำลองตอบคำถาม โดยใช้ชุดข้อมูลเสริม

ในการฝึกสอนแบบจำลองสำหรับตอบคำถามภาษาไทย จะใช้แบบจำลองประเภททรานส์ฟอร์เมอร์ (Transformer) ได้แก่ WangchanBERTa และ mT5-Base ซึ่งเป็นแบบจำลองที่สามารถใช้กับข้อมูลภาษาไทยได้ นอกจากนี้ ในวิทยานิพนธ์นี้ยังมีการศึกษารูปแบบการฝึกสอนแบบจำลองการตอบคำถามที่มีการใช้ชุดข้อมูลเสริมร่วมกับข้อมูลจริง (Training strategy) ได้แก่ การฝึกสอนแบบตามลำดับ และ การฝึกสอนแบบรวมข้อมูล

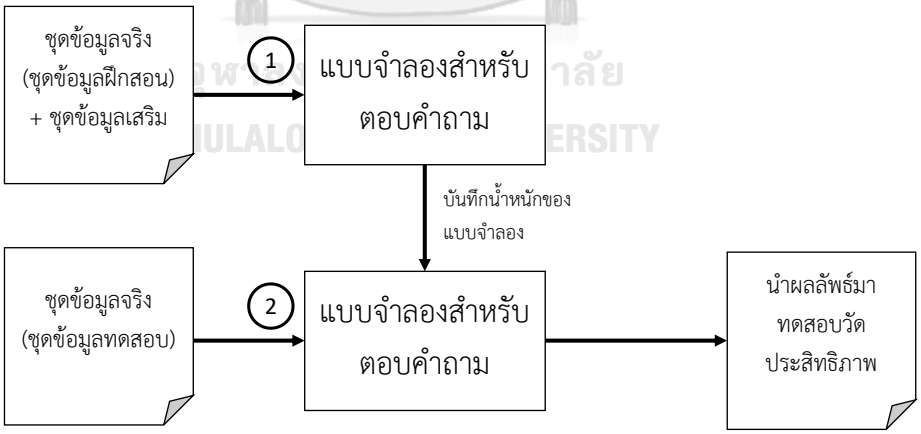
จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

รูปแบบการฝึกสอนแบบตามลำดับ (Sequence Strategy) หมายถึงการฝึกสอนด้วยชุดข้อมูลเสริมก่อน แล้วตามด้วยชุดข้อมูลจริง ตามลำดับ โดยขั้นตามตามฝึกสอนได้ระบุไว้ในรูปที่ 9 ดังนี้



รูป 9 ขั้นตอนการฝึกสอนแบบตามลำดับ

อีกรูปแบบการฝึกสอนคือ การฝึกสอนแบบรวมข้อมูล (Merge Strategy) หมายถึงการนำข้อมูลจากชุดข้อมูลเสริมรวมกับชุดข้อมูลฝึกจริง แล้วทำการฝึกสอนพร้อมกัน โดยขึ้นตอนการฝึกสอนได้ระบุไว้ในรูปที่ 10 ดังนี้



รูป 10 ขั้นตอนการฝึกสอนแบบรวมข้อมูล

4.4 การวัดประสิทธิภาพของแบบจำลองการตอบคำถามระดับพยางค์

เนื่องจากความไม่สมบูรณ์ของตัวตัดคำ (Word tokenizer) ขอภาษาไทย การวัดผลในระดับคำเหมือนการทดลองภาษาอังกฤษอาจไม่เพียงพอ เนื่องจาก ในภาษาอังกฤษจะมีช่องว่าง (space) สำหรับแบ่งระหว่างคำ ทำให้ภาษาอังกฤษสามารถแบ่งข้อความออกเป็นคำได้โดยง่าย ในขณะที่ภาษาไทยจะมีการแบ่งคำที่กำกวมมากกว่าเนื่องจากไม่มีช่องว่างระหว่างคำ ดังนั้น คะแนนเอฟวันระดับคำ ซึ่งใช้ตัวตัดคำเป็นส่วนประกอบในการคำนวณ จะขึ้นอยู่กับคุณภาพของตัวตัดคำที่เลือกใช้ในการแก้ไขปัญหานี้ ทางผู้เขียนจึงนำเสนอการคำนวณคะแนนเอฟวันระดับพยางค์ด้วย

คะแนนเอฟวันระดับพยางค์ (Syllable-level F1) คือคะแนนเอฟวันที่คำนวณจากการแบ่งพยางค์ ซึ่งเป็นตัวชี้วัดที่เหมาะสมมากกว่าคะแนนเอฟวันระดับคำ สำหรับภาษาที่มีการแบ่งคำที่กำกวมเช่นภาษาไทย เนื่องจากเหตุผลดังต่อไปนี้

1. การใช้ตัวตัดคำที่ต่างกันส่งผลให้คะแนนเอฟวันระดับคำต่างกัน ทำให้คะแนนของประสิทธิภาพไม่สามารถนำมาเปรียบเทียบกับงานวิจัยอื่น ๆ ที่ใช้ตัวตัดคำต่างกันได้
2. เนื่องจากความไม่สมบูรณ์ของตัวตัดคำ ทำให้การตัดคำมีการผิดพลาดเมื่อทำการตัดคำที่มีความคล้ายกัน
3. เนื่องจากความกำกวมของภาษาไทย บางคำของภาษาไทยสามารถตัดคำได้หลากหลายรูปแบบโดยเฉพาะคำนามเฉพาะ ในทางกลับกัน การแบ่งคำออกเป็นทีละพยางค์จะมีความกำกวมน้อยกว่า เนื่องจากการตัดพยางค์จะมีเพียงวิธีเดียวที่จะสามารถแบ่งคำออกเป็นพยางค์ที่ยังคงการออกเสียงของพยางค์นั้นได้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ในการทดลอง จะมีการใช้ตัวตัดคำ “newmm” ซึ่งเป็นหนึ่งในตัวตัดคำภาษาไทยที่ทำงานได้เร็วและมีความน่าเชื่อถือ อย่างไรก็ตาม ในตารางที่ 4 จะเห็นได้ว่า ตัวตัดคำ “newmm” ไม่สามารถตัดคำให้อยู่ในรูปแบบที่ถูกต้องได้ ในการแก้ปัญหานี้ จึงมีการวัดผลการตอบคำถามด้วยคะแนนเอฟวันระดับพยางค์คู่กับคะแนนเอฟวันระดับคำ เนื่องจากการใช้พยางค์ในการคำนวณคะแนนเอฟวันจะมีความแม่นยำเมื่อเทียบกับการแบ่งคำตามหลักภาษาศาสตร์มากกว่าการใช้คำจากตัวตัดคำ เพราะว่าการตัวแบ่งพยางค์สามารถแบ่งคำที่ซ้อนทับกันระหว่างคำตอบจากแบบจำลองและผลเฉลยได้ ในขณะที่ตัวแบ่งคำไม่สามารถทำได้

ตาราง 4 ตัวอย่างของการแบ่งคำและพยางค์ในวิทยานิพนธ์นี้

	คำโดยดั้งเดิม	การแบ่งคำที่ถูกต้อง	การแบ่งคำโดยใช้ newmm	การแบ่งพยางค์
วลีผลเฉลย	มลายู	มลายู	มลายู	ม | ลา | ยู
คำตอบจากแบบจำลอง	ภาษามลายู	ภาษา | มลายู	ภาษามลายู	ภา | ษา | ม | ลา | ยู
คะแนนเอฟวัน	-	66.6	0	74.99

ในหนึ่งคำของภาษาไทยสามารถมีได้หนึ่งหรือหลายพยางค์คล้ายกับภาษาอังกฤษ การแบ่งคำออกเป็นพยางค์หมายถึงการแบ่งคำออกเป็นหน่วยของการออกเสียงที่มีเพียง 1 การออกเสียง เช่น จากตารางที่ 4 คำว่า “มลายู” สามารถอ่านได้ว่า /mala:ju:/ ซึ่งประกอบด้วย 3 พยางค์ คือ “ม | ลา | ยู” ซึ่งแต่ละส่วนจะสามารถอ่านได้ว่า /ma/ /la:/ และ /ju:/ ตามลำดับ อีกหนึ่งตัวอย่างคือคำว่า “ภาษา” สามารถออกเสียงได้ว่า /pa:sa:/ ซึ่งประกอบด้วย 2 พยางค์ “ภา | ษา”ซึ่งแต่ละส่วนจะสามารถอ่านได้ว่า /pa:/ และ /sa:/ ตามลำดับ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 5

การทดลองและผลการทดลอง

ในบทนี้จะกล่าวถึงการทดลอง และผลการทดลองเบื้องต้น ประกอบด้วย ชุดข้อมูลที่ใช้ การฝึกสอนแบบจำลองการเรียนรู้เชิงลึก การทดสอบวัดประสิทธิภาพ และ ผลการทดลอง

5.1 ชุดข้อมูลที่ใช้

ในการวิจัยนี้ ได้ใช้ชุดข้อมูลการตอบคำถามภาษาไทยทั้งหมด 2 ชุด ได้แก่ ชุดข้อมูล Thai Wiki QA และ ชุดข้อมูล iApp Wiki QA โดยจำนวนข้อมูลในชุดข้อมูลทั้ง 2 ถูกสรุปในตารางที่ 5

ชุดข้อมูล Thai Wiki QA [21] เป็นชุดข้อมูลการตอบคำถามประเภท Open Domain ซึ่งถูกสร้างจากเอกสาร Wikipedia ภาษาไทย โดยที่ผู้สร้างคู่คำถามและคำตอบมาจากหลากหลายอาชีพและช่วงอายุ โดยชุดข้อมูลนี้ได้เคยถูกนำมาใช้ในการแข่งขัน National Software Contest (NSC) ในปี พ.ศ. 2561-2562 ลักษณะของคำถามของชุดข้อมูล Thai Wiki QA จะมีทั้งหมด 2 ประเภท ได้แก่ คำถามที่มีคำตอบเป็นข้อเท็จจริง ซึ่งจะมีลักษณะคำตอบเป็นคำหรือวลี และคำถามที่มีคำตอบเป็นการตอบรับหรือปฏิเสธ ซึ่งลักษณะคำตอบจะมีเพียง 2 คำตอบ คือ ใช่ หรือ ไม่ แต่ในการวิจัยครั้งนี้จะใช้เฉพาะคำถามที่มีคำตอบเป็นข้อเท็จจริงเท่านั้น ซึ่งมีชุดคำถาม-คำตอบทั้งหมด 15,000 คู่ โดยผู้เขียนได้แบ่งชุดข้อมูล เป็น ชุดข้อมูลฝึกสอน ชุดข้อมูลตรวจสอบ และ ชุดข้อมูลทดสอบ ดังที่สรุปในตาราง 3

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

นอกจากชุดข้อมูล Thai Wiki QA เป็นชุดข้อมูลที่มีชุดคำถาม-คำตอบแล้ว ทางผู้จัดทำชุดข้อมูล ยังได้เผยแพร่เอกสาร Wikipedia ภาษาไทยสำหรับใช้ในการตอบคำถามประเภท Open Domain จำนวน 125,302 บทความ โดยทางผู้เขียนได้นำบทความเหล่านี้มาใช้ในการสร้างคู่คำถาม-คำตอบโดยใช้ชุดของแบบจำลองสำหรับสร้างคำถาม

ชุดข้อมูล iApp Wiki QA[1] เป็นชุดข้อมูลการตอบคำถามที่มีประเภทคำถามเป็นคำถามที่มีคำตอบเป็นข้อเท็จจริง ซึ่งถูกสร้างจากเอกสาร Wikipedia ภาษาไทย พัฒนาโดย บจก. ไอแอพพ์ เทคโนโลยี โดยมีชุดข้อมูลคำถาม-คำตอบทั้งหมด 7,242 คู่

ตาราง 5 จำนวนชุดข้อมูลฝึกสอน ชุดข้อมูลตรวจสอบและชุดข้อมูลทดสอบ

ชุดข้อมูล	ชุดข้อมูลฝึกสอน	ชุดข้อมูลตรวจสอบ	ชุดข้อมูลทดสอบ
Thai Wiki QA	9,045	1,005	4,950
iApp Wiki QA	5,761	742	739

5.2 เครื่องมือและการตั้งค่าพารามิเตอร์ของแบบจำลอง

ทุก ๆ การทดลองในวิทยานิพนธ์นี้ ทั้งส่วนของการสร้างชุดข้อมูลเสริม และส่วนการฝึกสอนแบบจำลองการตอบคำถาม ผู้เขียนได้ใช้เครื่องมือ “Transformers” ของ HuggingFace ในการพัฒนาในส่วนของแบบจำลองและการฝึกสอนแบบจำลอง เนื่องจากเครื่องมี “Transformers” ได้มีโครงสร้างของแบบจำลองสำหรับฝึกสอนและนำแบบจำลองไปใช้งาน ตัวตัดคำสำหรับแบบจำลองเนื่องจากการใช้งานแบบจำลองประเภททรานส์ฟอร์เมอร์ จะต้องใช้ตัวตัดคำเดียวกันกับการฝึกสอนของแบบจำลอง โดยทั้ง WangchanBERTa และ mT5 ได้ใช้ตัวตัดคำ SentencePiece อีกทั้งเครื่องมือยังมีการตั้งค่าเบื้องต้นสำหรับการฝึกสอนที่พร้อมใช้งาน โดย ทุก ๆ แบบจำลองในวิทยานิพนธ์นี้จะใช้การตั้งค่าเริ่มต้นของเครื่องมือ ยกเว้น อัตราการเรียนรู้, Weight Decay, ขนาดรุ่นของการผลิต และ Epoch ตามรายละเอียดในตารางที่ 6 และ 7

ในการฝึกสอนแบบจำลองการตอบคำถาม ผู้เขียนได้ใช้ชุดของไฮเพอร์พารามิเตอร์ (Hyperparameter) แตกต่างกันระหว่างแบบจำลอง WanchanBERTa และ mT5 เนื่องจากจำนวนพารามิเตอร์ที่สามารถฝึกสอนได้ของแบบจำลองทั้งสองมีขนาดที่ต่างกันค่อนข้างมาก โดยมีรายละเอียดของไฮเพอร์พารามิเตอร์ตามตาราง 6 ดังนี้

[1] https://github.com/iapp-technology/iapp-wiki-qa-dataset

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตาราง 6 การตั้งค่าของไฮเพอร์พารามิเตอร์ที่ใช้งานในแบบจำลองการตอบคำถาม

พารามิเตอร์	WanchanBERTa	mT5
ตัวเพิ่มประสิทธิภาพ (Optimizer)	Adam	Adam
อัตราการเรียนรู้ (Learning rate)	5×10^{-6}	5×10^{-6}
ขนาดรุ่นของการผลิต (Batch size)	12	4
Weight Decay	0.01	0.01
Epochs	25	25

ในการฝึกสอนแบบจำลองสำหรับสร้างชุดข้อมูลเสริมทั้ง 3 แบบจำลอง ทางผู้เขียนได้ใช้งานชนิดของแบบจำลองและชุดของไฮเพอร์พารามิเตอร์ที่ต่างกันเช่นเดียวกับการฝึกสอนแบบจำลองสำหรับสร้างชุดข้อมูลเสริม โดยมีรายละเอียดของไฮเพอร์พารามิเตอร์ตามตารางที่ 7 ดังนี้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตาราง 7 การตั้งค่าของไฮเพอร์พารามิเตอร์ที่ใช้งานในชุดของแบบจำลองสำหรับสร้างชุดคำถาม-คำตอบ

พารามิเตอร์	แบบจำลองสำหรับเลือกคำตอบ	แบบจำลองสำหรับสร้างคำถาม	แบบจำลองสำหรับกรองคำถาม
ชนิดของแบบจำลอง	mT5-Large	mT5-Large	WangchanBERTa
ตัวเพิ่มประสิทธิภาพ (Optimizer)	Adam	Adam	Adam
อัตราการเรียนรู้ (Learning rate)	5×10^{-6}	5×10^{-6}	5×10^{-6}
ขนาดรุ่นของการผลิต (Batch size)	4	4	12
Weight Decay	0.01	0.01	0.01
Epochs	25	25	25

สาเหตุที่ทางผู้เขียนเลือกขนาดรุ่นของการผลิตที่ต่างกันในแต่ละประเภทของแบบจำลอง เนื่องจากเหตุผลทางเทคนิค เนื่องจากระบบที่ใช้ในการทดลอง คือ DGX A100 ซึ่งมี GPU รุ่น NVIDIA A100 สามารถใช้งานขนาดรุ่นของการผลิตที่ขนาด 4 เท่านั้น สำหรับแบบจำลอง mT5-Base และ mT5-Large ดังนั้น ในการฝึกสอนแบบจำลองประเภท WangchanBERTa จึงเลือกที่ 12 เพื่อให้แบบจำลองทั้ง 2 ประเภทสามารถเปรียบเทียบกันได้

นอกจากนี้ ในการทดลองของวิทยานิพนธ์นี้ยังใช้เครื่องมือ PyThaiNLP [20] ซึ่งเป็นเครื่องมือสำหรับใช้ประมวลผลภาษาธรรมชาติของภาษาไทย โดยที่เครื่องมือนี้จะใช้ในการประมวลผลก่อน อีกทั้ง เครื่องมือนี้ยังมีตัวตัดคำ “newmm” และตัวตัดพยางค์ที่ใช้ในการทดลองอีกด้วย

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

5.3 การทดสอบวัดประสิทธิภาพ

ในการวิจัยนี้ จะวัดประสิทธิภาพที่ชุดข้อมูลทดสอบ โดยจะวัดจากคำตอบจากแบบจำลองเทียบกับวลีผลเฉลย โดยใช้ตัววัดประสิทธิภาพทั้งหมด 3 อย่างดังที่กล่าวไว้ในหัวข้อ 2.6 และ 4.4 ดังนี้

1. คะแนนความถูกต้อง (Exact Match)
2. คะแนนเอฟวันระดับคำ (Word-level F1)
3. คะแนนเอฟวันระดับพยางค์ (Syllable-level F1)

5.4 ผลการทดลองและสรุปผลการทดลอง

ในการทดลองในวิทยานิพนธ์นี้ได้ถูกทดลองและเปรียบเทียบผลการทดลองระหว่างวิธีการฝึกสอนในรูปแบบต่าง ๆ ได้แก่ การใช้งานแบบจำลองพื้นฐาน รูปแบบการใช้งานชุดข้อมูลเสริม และกลุ่มของชุดข้อมูลที่ใช้งาน นอกจากนี้ ผู้เขียนยังมีการวิเคราะห์ประสิทธิภาพที่เพิ่มขึ้นของแบบจำลอง การวิเคราะห์ประสิทธิภาพของแบบจำลองตามประเภทของคำถาม และ วิเคราะห์คะแนนเอฟวันระดับคำที่มีการเปลี่ยน เมื่อเปลี่ยนตัวตัดคำ

5.4.1 ผลการสร้างชุดข้อมูลเสริม

ในการสร้างชุดข้อมูลเสริมจะใช้บทความที่ผู้จัดทำชุดข้อมูล Thai Wiki QA ได้เผยแพร่จำนวน 125,302 บทความ โดยใช้ชุดของแบบจำลองสำหรับสร้างชุดคำถาม-คำตอบสามารถสร้างชุดคำถาม-คำตอบได้ทั้งหมด 110,768 คู่ โดยที่ในชุดคำถาม-คำตอบที่สร้างขึ้นเหล่านี้ จะมี 53,575 ชุดที่ผ่านแบบจำลองสำหรับกรองคำถาม โดยเมื่อรวมกับจำนวนชุดข้อมูลฝึกของแต่ละชุดข้อมูลแล้ว จะได้จำนวนของชุดข้อมูลตามตาราง 8 และตัวอย่างของชุดข้อมูลเสริมที่ถูกสร้างจะอยู่ในตารางที่ 17 ในส่วนภาคผนวก

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตาราง 8 จำนวนข้อมูลสำหรับฝึกสอนในแต่ละชุดข้อมูล เมื่อรวมกับชุดข้อมูลเสริมแล้ว

ชุดข้อมูล	ชุดข้อมูลฝึก	ชุดข้อมูลฝึก + ชุดข้อมูลเสริมแบบกรอง	ชุดข้อมูลฝึก + ชุดข้อมูลเสริมทั้งหมด
Thai Wiki QA	9,045	62,610 (+592.2%)	119,813 (+1,224.6%)
iApp Wiki QA	5,761	59,326 (+929.8 %)	116,529 (+1,922.7%)

5.4.2 ผลการทดลองโดยรวม

ผู้เขียนได้ทำการทดลองเพิ่มประสิทธิภาพของแบบจำลองการตอบคำถามภาษาไทยบน 2 ชุดข้อมูล ได้แก่ Thai Wiki QA และ iApp Wiki QA และทำการศึกษาเปรียบเทียบผลการทดลองใน 3 มุมมอง ได้แก่ รูปแบบวิธีการฝึกสอน กลุ่มของชุดข้อมูลเสริม และ แบบจำลองประเภททรานส์ฟอร์เมอร์พื้นฐานที่เลือกใช้ (WangchanBERTa และ mT5-Base) โดยใช้ตัวชี้วัดที่เหมาะสมกับการวัดผลแบบจำลองการตอบคำถาม ได้แก่ คะแนนความถูกต้อง (Exact match) คะแนนเอฟวันระดับคำ (W-F1) และ คะแนนเอฟวันระดับพยางค์ (S-F1) ผลการทดลองจะอยู่ในตาราง 9 สำหรับชุดข้อมูล Thai Wiki QA และ ตาราง 10 สำหรับชุดข้อมูล iApp Wiki QA

ในตารางผลการทดลอง ผู้เขียนได้สร้างชื่อสำหรับระบุแบบจำลอง รูปแบบวิธีการฝึกสอน และกลุ่มของชุดข้อมูลเสริมที่ใช้เพื่อให้ง่ายสำหรับอ้างอิง โดยแบบจำลองจะเป็นไปได้ทั้งหมด 2 แบบจำลอง ได้แก่ WangchanBERTa (WBT) และ mT5-Base (mT5) รูปแบบวิธีการฝึกสอนจะเป็นไปได้ทั้งหมด 2 รูปแบบ ได้แก่ การฝึกสอนแบบเป็นลำดับ (SEQ หรือ Sequence Strategy) และ การฝึกสอนแบบรวมข้อมูล (MRG หรือ Merge Strategy) และกลุ่มของชุดข้อมูลเสริมที่ใช้จะเป็นไปได้ 2 กลุ่ม ได้แก่ กลุ่มข้อมูลเสริมแบบกรอง (FLT) และกลุ่มข้อมูลเสริมทั้งหมด (ALL) โดยที่ทั้ง 3 ส่วนจะประกอบเข้าด้วยกันแล้วแบ่งด้วยเครื่องหมายยัติภาค (“-“) ตัวอย่างเช่น mT5-SEQ-FLT หมายถึง การใช้แบบจำลอง mT5 พร้อมกับชุดข้อมูลเสริมแบบกรอง และฝึกสอนแบบเป็นลำดับ

จากผลการทดลองพบว่า การฝึกสอนด้วยชุดข้อมูลเสริมสามารถช่วยเพิ่มประสิทธิภาพของแบบจำลองการตอบคำถามได้ทุกกรณีและทุกตัวชี้วัด เมื่อเปรียบเทียบกับการฝึกสอนโดยไม่ใช้ข้อมูลเสริม สำหรับชุดข้อมูล Thai Wiki QA การทดลอง mT5-SEQ-FLT หรือการใช้งานแบบจำลอง mT5

โดยใช้ชุดข้อมูลเสริมแบบกรอง และฝึกสอนแบบเป็นลำดับ จะได้ประสิทธิภาพที่สูงที่สุด โดยมีคะแนนมากกว่าการใช้แบบจำลอง mT5 ที่ 6.28% 5.14% และ 4.94% สำหรับตัวชี้วัดคะแนนความถูกต้อง คะแนนเอฟวันระดับคำ และ คะแนนเอฟวันระดับพยางค์ตามลำดับ และสำหรับชุดข้อมูล iApp Wiki QA พบว่า การทดลอง mT5-SEQ-ALL หรือการใช้แบบจำลอง mT5-Base โดยใช้ชุดข้อมูลเสริมทั้งหมด และฝึกสอนแบบเป็นลำดับ จะได้ประสิทธิภาพสูงที่สุด โดยมีคะแนนมากกว่าการใช้แบบจำลอง mT5-Base อยู่ที่ 28.68% 18.51% และ 19.44% สำหรับตัวชี้วัดคะแนนความถูกต้อง คะแนนเอฟวันระดับคำ และ คะแนนเอฟวันระดับพยางค์ตามลำดับ จะสังเกตได้ว่า ในชุดข้อมูล iApp Wiki QA การใช้งานข้อมูลเสริมทั้งหมดสามารถให้ประสิทธิภาพที่สูงกว่าการใช้งานชุดข้อมูลเสริมแบบกรองเล็กน้อย เนื่องจากจำนวนของข้อมูลจริงของชุดข้อมูล iApp Wiki QA มีจำนวนน้อยกว่าชุดข้อมูล Thai Wiki QA ซึ่งต้องการจำนวนของข้อมูลในการฝึกสอนมากกว่า อย่างไรก็ตาม คะแนนของแบบจำลองที่ใช้งานชุดข้อมูลเสริมแบบกรอง และชุดข้อมูลเสริมทั้งหมดไม่ต่างกันมาก จึงสามารถสรุปได้ว่าการใช้ mT5-SEQ-FLT เป็นการทดลองที่สามารถให้ผลที่ดีที่สุดทั้ง 2 ชุดข้อมูล

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตาราง 9 ผลการทดลองการใช้ชุดข้อมูลเสริมโดยการเปรียบเทียบกับการไม่ใช้ชุดข้อมูลเสริม ในชุดข้อมูล Thai Wiki QA

รูปแบบการฝึกสอน	ไม่ใช้ชุดข้อมูลเสริม			+ ชุดข้อมูลเสริมแบบกรอง (FLT)			+ ชุดข้อมูลเสริมทั้งหมด (ALL)		
	EM	W-F1	S-F1	EM	W-F1	S-F1	EM	W-F1	S-F1
WangchanBERTa (WBT)									
การฝึกสอนแบบเป็นลำดับ (SEQ)	43.92	70.73	74.71	45.90	73.35	77.55	46.48	74.4	78.6
การฝึกสอนแบบรวมข้อมูล (MRG)				44.63	71.11	75.15	43.35	71.09	75.26
mT5-Base (mT5)									
การฝึกสอนแบบเป็นลำดับ (SEQ)	64.14	78.24	80.35	**70.42**	**83.64**	**85.29**	69.03	83.02	84.74
การฝึกสอนแบบรวมข้อมูล (MRG)				69.01	82.88	84.68	63.66	79.3	81.17

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตาราง 10 ผลการทดลองการใช้ชุดข้อมูลเสริมโดยการเปรียบเทียบกับการไม่ใช้ชุดข้อมูลเสริม ในชุดข้อมูล iApp Wiki QA

รูปแบบการฝึกสอน	ไม่ใช้ชุดข้อมูลเสริม			+ ชุดข้อมูลเสริมแบบกรอง (FLT)			+ ชุดข้อมูลเสริมทั้งหมด (ALL)		
	EM	W-F1	S-F1	EM	W-F1	S-F1	EM	W-F1	S-F1
WangchanBERTa (WBT)									
การฝึกสอนแบบเป็นลำดับ (SEQ)	30.58	69.68	73.71	32.88	71.81	74.42	33.15	72.98	75.14
การฝึกสอนแบบรวมข้อมูล (MRG)				29.77	69.59	71.97	31.66	70.37	72.82
mT5-Base (mT5)									
การฝึกสอนแบบเป็นลำดับ (SEQ)	29.36	63.46	63.71	56.02	81.31	82.58	**58.05**	**81.97**	**83.15**
การฝึกสอนแบบรวมข้อมูล (MRG)				57.10	81.42	82.57	53.45	79.53	80.81

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

5.4.3 การเปรียบเทียบระหว่างรูปแบบวิธีการฝึกสอน

ความแตกต่างระหว่างรูปแบบวิธีการฝึกสอน คือ ขั้นตอนในการฝึกสอนชุดข้อมูลเสริม ซึ่งมีอยู่ 2 รูปแบบ คือ การฝึกสอนแบบเป็นลำดับ หมายถึงเริ่มทำการฝึกสอนด้วยชุดข้อมูลเสริมก่อนแล้วตามด้วยชุดข้อมูลจริง และการฝึกสอนแบบรวมข้อมูล หมายถึง การนำข้อมูลเสริมและข้อมูลจริงมารวมกันแล้วฝึกสอนทีเดียว จากผลการทดลอง พบว่าการฝึกสอนแบบเป็นลำดับสามารถให้ผลการทดลองที่ดีกว่าการใช้การฝึกสอนแบบรวมข้อมูลอย่างชัดเจน ตามที่แสดงในตารางที่ 9 และ 10 ในคอลัมน์รูปแบบการฝึกสอน

5.4.4 การเปรียบเทียบระหว่างการใช้งานกลุ่มของชุดข้อมูลเสริม

ในการทดลอง การใช้งานชุดข้อมูลเสริมแบบกรองหรือการใช้งานชุดข้อมูลเสริมทั้งหมด ไม่มีความแตกต่างกันระหว่างจำนวนของการทดลองที่ให้ผลที่ดีกว่า อย่างไรก็ตาม หากพิจารณาเฉพาะการใช้รูปแบบการฝึกสอนแบบเป็นลำดับแล้วพบว่า การใช้ชุดข้อมูลเสริมทั้งหมดมีจำนวนการทดลองที่ดีกว่าการใช้ชุดข้อมูลเสริมแบบกรอง จึงสามารถสรุปได้ว่า การใช้งานชุดข้อมูลเสริมแบบกรองมีความเหมาะสมกว่าการใช้ชุดข้อมูลเสริมทั้งหมด

อย่างไรก็ตาม จากผลการทดลองของชุดข้อมูล iApp Wiki QA ตามตารางที่ 10 การทดลองของแบบจำลอง mT5-Base กับชุดข้อมูลเสริมทั้งหมด โดยใช้การฝึกสอนแบบเป็นลำดับ (mT5-SEQ-ALL) ให้ผลคะแนนที่ดีกว่าการใช้แบบจำลอง mT5 กับชุดข้อมูลเสริมแบบกรอง โดยใช้การฝึกสอนแบบเป็นลำดับ (mT5-SEQ-FLT) เล็กน้อย แต่จำนวนชุดข้อมูลฝึกต่างกันค่อนข้างมาก (พิจารณาจำนวนชุดข้อมูลฝึกจากตาราง 8) ในกรณีนี้สามารถสรุปได้ว่า การใช้งานชุดข้อมูลเสริมแบบกรองดีกว่าการใช้ชุดข้อมูลเสริมทั้งหมด ในแง่ของระยะเวลาที่ใช้ในการฝึกสอน

5.4.5 การเปรียบเทียบระหว่างแบบจำลองพื้นฐานที่ใช้งาน

จากการทดลองการเปรียบเทียบแบบจำลองพื้นฐานโดยการฝึกสอนเฉพาะข้อมูลจริงเพียงอย่างเดียว พบว่าผลการชนะของแบบจำลองขึ้นอยู่กับชุดข้อมูล อย่างไรก็ตาม หากพิจารณาการทดลองที่มีการใช้ชุดข้อมูลเสริมแล้ว การใช้แบบจำลอง mT5-Base สามารถให้ประสิทธิภาพที่สูงกว่าแบบจำลอง WangchanBERTa ทุกกรณี ตามตารางที่ 9 และ 10 ในคอลัมน์ “+ ชุดข้อมูลเสริมแบบกรอง” และ “+ ชุดข้อมูลเสริมทั้งหมด”

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

5.4.6 การเปรียบเทียบผลการทดลองด้วยคะแนน F1 ระดับพยางค์

จากผลการทดลองในตารางที่ 9 และ 10 จะสังเกตได้ว่า คะแนนเอฟวันระดับพยางค์ (S-F1) จะมีค่าที่สูงกว่าคะแนนเอฟวันระดับคำ (W-F1) ทุกการทดลอง โดยที่การทดลอง mT5-SEQ-FLT ที่ชุดข้อมูล Thai Wiki QA พบว่าคะแนนเอฟวันระดับพยางค์มีค่ามากกว่าคะแนนเอฟวันระดับคำอยู่ที่ 1.65% และการทดลอง mT5-SEQ-ALL ในชุดข้อมูล iApp Wiki QA คะแนนเอฟวันระดับพยางค์มีค่ามากกว่าคะแนนเอฟวันระดับคำอยู่ที่ 1.18% เนื่องจากในบางคำตอบจากแบบจำลองหรือผลเฉลยจะมีคำที่ตัวตัดคำ (word tokenizer) ที่ใช้ในการคำนวณคะแนนเอฟวันระดับคำ ไม่สามารถทำงานได้

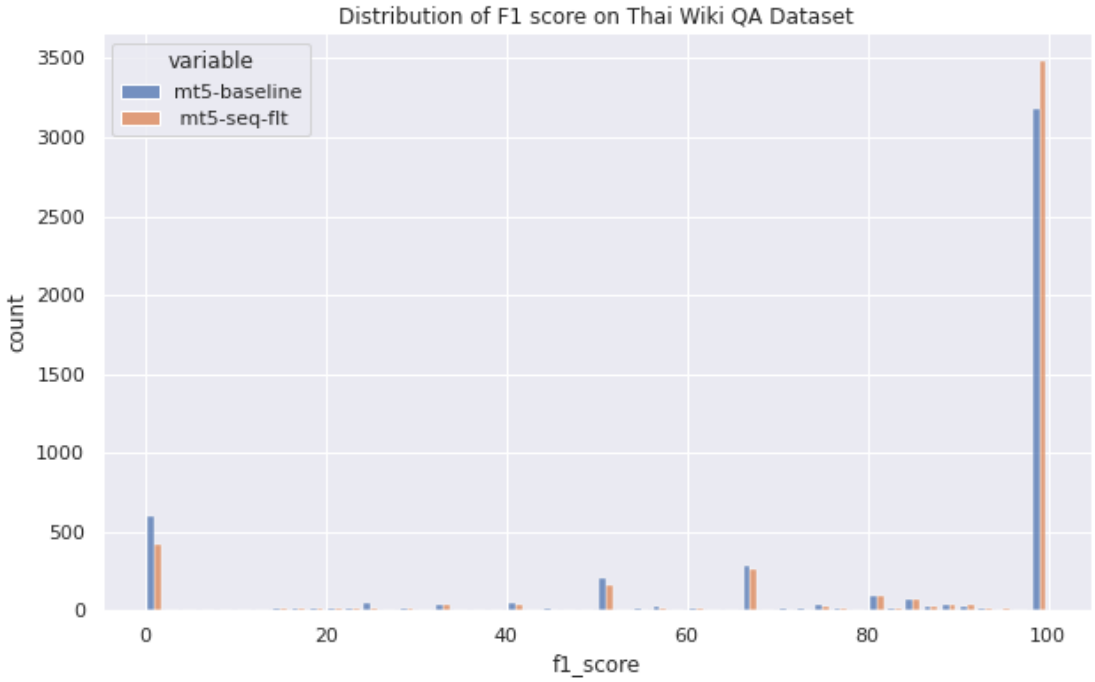
อย่างถูกต้อง เนื่องจากความไม่แม่นยำในการทำงานของตัวตัดคำเอง ส่งผลในคำที่มีอยู่ร่วมกันของทั้งคำตอบจากแบบจำลองและผลเฉลยไม่สามารถถูกนำไปคำนวณคะแนนเอฟวันอย่างถูกต้อง อย่างไรก็ตาม การคำนวณคะแนนเอฟวันระดับพยางค์จะไม่มีปัญหานี้ในการแบ่งคำเป็นระดับพยางค์ เนื่องจากการแบ่งคำเป็นพยางค์สามารถแบ่งได้คำที่มีร่วมกันระหว่างคำตอบจากแบบจำลองและผลเฉลยได้เพียงแบบเดียว ซึ่งมีความกำกวมน้อยกว่า การคำนวณคะแนนเอฟวันระดับพยางค์จึงสามารถทำให้คะแนนเอฟวันใกล้เคียงกับความเป็นจริงมากกว่า ดังตัวอย่างในตารางที่ 11

ตาราง 11 ตัวอย่างของคำตอบจากแบบจำลองและผลเฉลย และผลการคำนวณคะแนนเอฟวันระดับพยางค์ โดยทุกตัวอย่างจะมีคะแนนเอฟวันระดับคำเป็น 0

คำตอบจากแบบจำลอง	ผลเฉลย	คะแนนเอฟวันระดับพยางค์
มหาวิทยาลัยรามคำแหง	รามคำแหง	66.67
คณะนิติศาสตร์	นิติศาสตร์	85.71
คมนาคม	กระทรวงคมนาคม	75.00
ไม้ล้มลุกขนาดเล็ก	ล้มลุก	57.14

5.5 วิเคราะห์การเพิ่มประสิทธิภาพของแบบจำลอง

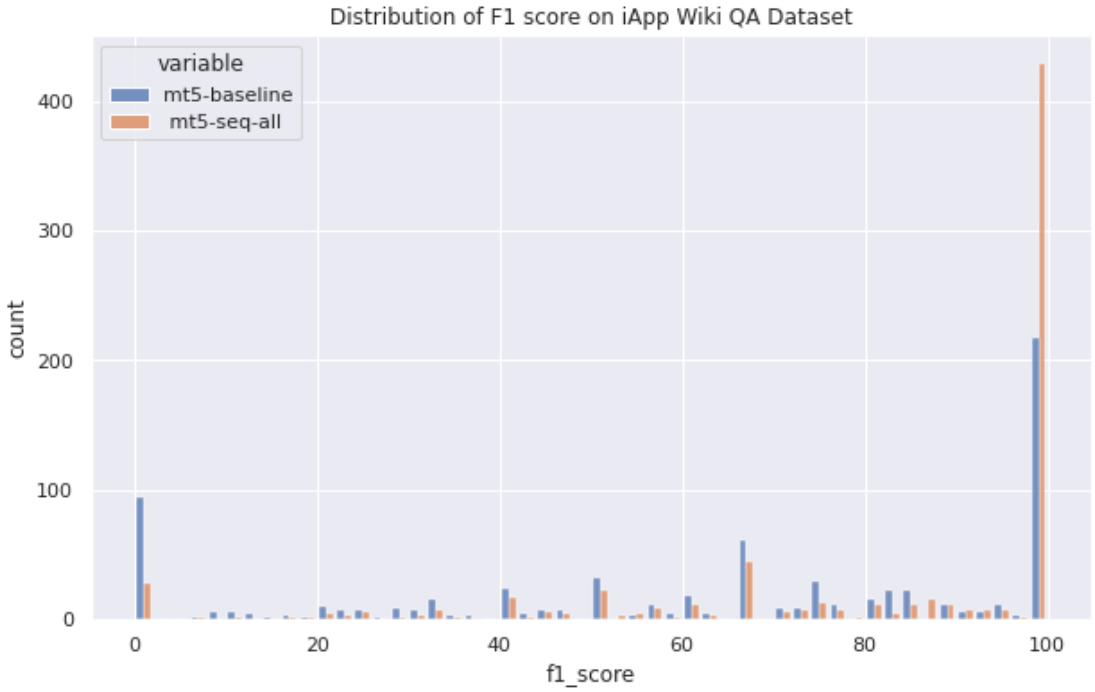
ในการวิเคราะห์ประสิทธิภาพที่เพิ่มขึ้นของแบบจำลองการตอบคำถาม ผู้เขียนได้ใช้กราฟที่แสดงการกระจายตัวของคะแนนเอฟวันระดับคำกับชุดข้อมูลทดสอบ เพื่อพิจารณาว่าการกระจายตัวของคะแนนเอฟวันมีความต่างกันอย่างไร โดยจะทำการเปรียบเทียบก่อนและหลังการใช้ชุดข้อมูลเสริม ผลของชุดข้อมูล Thai Wiki QA จะแสดงอยู่ในรูปที่ 11 จะเห็นได้ว่า ในกราฟสีน้ำเงินซึ่งเป็นกราฟที่แสดงการกระจายตัวของคะแนนเอฟวันระดับคำของแบบจำลอง mT5-Base โดยไม่ได้ใช้ชุดข้อมูลเสริม จะมีจำนวนข้อมูลที่ได้คะแนนเอฟวันระดับคำเป็น 0 มากกว่า 500 ข้อมูล ซึ่งเกิดจากผลการทำนายคำตอบของแบบจำลอง ไม่สามารถตอบคำถามได้ตรงกับผลเฉลยเลย แต่หลังจากใช้ชุดข้อมูลเสริม ตามการทดลอง mT5-SEQ-FLT ในกราฟสีส้มจะพบว่า จำนวนข้อมูลที่แบบจำลองสามารถตอบได้อย่างถูกต้อง โดยจะได้ข้อมูลที่มีคะแนนเอฟวันระดับคำเป็น 100 จะมีจำนวนเพิ่มขึ้นในขณะที่จำนวนข้อมูลที่ได้คะแนนเอฟวันระดับคำเป็น 0 ลดลง ซึ่งจะแสดงให้เห็นว่าวิธีการนี้สามารถช่วยเพิ่มประสิทธิภาพของแบบจำลองได้จริง



รูป 11 กราฟแสดงการกระจายตัวของคะแนนเอฟวันระดับคำ สำหรับชุดข้อมูลทดสอบของชุดข้อมูล Thai Wiki QA โดยที่ สีน้ำเงินคือการกระจายตัวของข้อมูลจากแบบจำลอง mT5-Base โดยที่ไม่ได้ใช้ชุดข้อมูลเสริม และ สีส้มคือการกระจายตัวจากแบบจำลองของการทดลอง mT5-SEQ-FLT

สำหรับผลของชุดข้อมูล iApp Wiki QA จะแสดงอยู่ในรูปที่ 12 ซึ่งจะมีลักษณะที่คล้ายกับผลของชุดข้อมูล Thai Wiki QA จะเห็นได้ว่า ในกราฟสีน้ำเงินซึ่งเป็นกราฟที่แสดงการกระจายตัวของคะแนนเอฟวันระดับคำของแบบจำลอง mT5 โดยไม่ได้ใช้ชุดข้อมูลเสริม จะมีจำนวนข้อมูลที่ได้คะแนนเอฟวันระดับคำเป็น 0 มากกว่า 50 ข้อมูล ซึ่งเกิดจากผลการทำนายคำตอบของแบบจำลองไม่สามารถตอบคำถามได้ตรงกับผลเฉลยเลย แต่หลังจากใช้ชุดข้อมูลเสริม ตามการทดลอง mT5-SEQ-ALL ในกราฟสีส้มจะเห็นได้ว่า จำนวนข้อมูลที่แบบจำลองสามารถตอบได้อย่างถูกต้อง โดยจะได้คะแนนเอฟวันระดับคำเป็น 100 จะมีจำนวนเพิ่มขึ้นมาก ในขณะที่จำนวนข้อมูลที่ได้คะแนนเอฟวันระดับคำเป็น 0 ลดลง ซึ่งจะแสดงให้เห็นว่าวิธีการนี้สามารถช่วยเพิ่มประสิทธิภาพของแบบจำลองได้ถึงแม้จะมีการเปลี่ยนชุดข้อมูล

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY



รูป 12 กราฟแสดงการกระจายตัวของคะแนนเอฟวันระดับคำ สำหรับชุดข้อมูลทดสอบของชุดข้อมูล iApp Wiki QA โดยที่ สีน้ำเงินคือการกระจายตัวของข้อมูลจากแบบจำลอง mT5-Base โดยที่ไม่ได้ใช้ชุดข้อมูลเสริม และ สีส้มคือการกระจายตัวจากแบบจำลองของการทดลอง mT5-SEQ-ALL

5.6 วิเคราะห์ประสิทธิภาพของแบบจำลองการตอบคำถามโดยประเภทของคำถาม

จากคำถามของชุดข้อมูลจะสามารถแบ่งชนิดของคำถามออกเป็น 6 ประเภทตามชนิดของคำตอบ โดยประเภทของคำถามจะมีดังนี้

- “ใคร”: กลุ่มของคำถามที่ต้องการคำตอบเป็นชื่อบุคคล
- “อะไร”: กลุ่มของคำถามที่ต้องการคำตอบเป็นชื่อของสิ่งของ
- “ที่ไหน”: กลุ่มของคำถามที่ต้องคำตอบเป็นชื่อสถานที่ เช่น ชื่อประเทศ ชื่อจังหวัด หรือชื่อรัฐ เป็นต้น
- “ปี”: กลุ่มของคำถามที่ต้องคำตอบเป็นเลขปี โดยจะเป็นปี ค.ศ. หรือปี พ.ศ. ก็ได้
- “วันที่”: กลุ่มของคำถามที่ต้องคำตอบเป็นวันที่
- “ตัวเลข”: กลุ่มของคำถามที่ต้องคำตอบเป็นตัวเลข

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ในการแบ่งประเภทของคำถามสามารถทำได้โดยการตรวจจับคำสำคัญ (Keyword detection) โดยที่คำสำคัญของคำถามแต่ละประเภทได้ระบุไว้ในตาราง 12 จากนั้น เมื่อทำการแบ่งชนิดของคำถามเรียบร้อยแล้ว ก็จะนำมาวัดผลในแต่ละประเภทของคำถามโดยที่ผลการทดลองจะแสดงในตารางที่ 13 และ 14 ผลการทดลองแสดงให้เห็นว่า การทดลองที่ดีที่สุดของแต่ละชุดข้อมูลสามารถเพิ่มประสิทธิภาพได้เมื่อเปรียบเทียบกับแบบจำลองพื้นฐานที่ไม่ใช้ข้อมูลเสริมในทุก ๆ ประเภทของคำถาม ยกเว้นประเภท “ปี” ในชุดข้อมูล iApp Wiki QA ที่มีประสิทธิภาพน้อยกว่าแบบจำลองพื้นฐานที่ไม่ใช้ข้อมูลเสริมเล็กน้อย หลังจากการตรวจสอบพบว่าข้อมูลที่จัดอยู่ในประเภท “ปี” มีอยู่เพียง 8 ข้อมูล เนื่องจากจำนวนข้อมูลที่น้อย ค่าเฉลี่ยของคะแนนเอฟวันจึงเปลี่ยนแปลงได้ง่าย อย่างไรก็ตาม คะแนนเอฟวันระดับพยางค์ของข้อมูลประเภทนี้เพิ่มขึ้นหลังจากการฝึกสอนด้วยข้อมูลเสริม ซึ่งแสดงให้เห็นว่าการทดลองที่ดีที่สุดของชุดข้อมูล iApp Wiki QA (mT5-SEQ-ALL) สามารถหาคำตอบที่แม่นยำกว่าเมื่อเทียบกับแบบจำลองที่ไม่ได้ใช้ชุดข้อมูลเสริม ซึ่งก็เป็นข้อพิสูจน์ว่าวิธีการสร้างชุดข้อมูลเสริมนี้สามารถเพิ่มประสิทธิภาพแบบจำลองการตอบคำถามภาษาไทยได้จริง

ตาราง 12 ตัวอย่างของคำสำคัญที่ใช้แบ่งประเภทคำถาม

ใคร	**อะไร**	**ที่ไหน**	**ปี**	**วันที่**	**จำนวน**
ใคร	อะไร	ที่ไหน	ปีใด	วันที่เท่าใด	เท่าไร
คนใด	...ใด	ที่ใด	ปีไหน	วันใด	เท่าใด
คนไหน	...ไหน	ประเทศใด	พ.ศ. ใด	เมื่อใด	กี่...
		จังหวัดใด	ค.ศ. ใด		

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตาราง 13 ประสิทธิภาพของแบบจำลองตามประเภทของคำถาม ของชุดข้อมูล Thai Wiki QA

ชนิดคำถาม	mT5-Base ไม่ใช้ข้อมูลเสริม			mT5-SEQ-FLT			ร้อยละที่เพิ่มขึ้น		
	EM	W-F1	S-F1	EM	W-F1	S-F1	EM	W-F1	S-F1
ประสิทธิภาพโดยรวม	**64.14**	**78.24**	**80.35**	**70.42**	**83.64**	**85.29**	**9.79**	**6.90**	**6.15**
ใคร	69.16	82.03	81.40	73.05	85.87	85.30	5.62	4.68	4.79
อะไร	62.90	77.45	80.48	68.82	82.89	85.24	9.41	7.02	5.91
ที่ไหน	67.54	83.23	83.02	76.61	88.82	88.93	13.42	6.72	7.12
ปี	76.04	83.06	85.03	83.83	88.75	89.53	10.24	6.85	5.29
วันที่	58.88	79.53	79.21	72.08	90.13	89.43	22.42	13.33	12.90
จำนวน	61.99	74.14	76.67	68.19	78.94	81.32	10.00	6.47	6.07

ตาราง 14 ประสิทธิภาพของแบบจำลองตามประเภทของคำถาม ของชุดข้อมูล iApp Wiki QA

ชนิดคำถาม	mT5-Base ไม่ใช้ข้อมูลเสริม			mT5-SEQ-ALL			ร้อยละที่เพิ่มขึ้น		
	EM	W-F1	S-F1	EM	W-F1	S-F1	EM	W-F1	S-F1
ประสิทธิภาพโดยรวม	**29.36**	**63.46**	**63.71**	**58.05**	**81.97**	**83.15**	**97.72**	**29.17**	**30.51**
ใคร	48.57	67.03	66.30	65.71	86.10	87.77	35.29	28.45	32.38
อะไร	32.18	65.00	66.56	56.02	81.03	82.64	74.08	24.66	24.16
ที่ไหน	33.33	68.22	71.85	55.56	79.72	85.17	66.69	16.85	18.54
ปี	62.50	84.17	79.17	62.50	82.08	81.25	0.00	−2.48	2.63
วันที่	9.68	63.91	58.22	62.90	87.97	86.43	549.79	37.64	48.45
จำนวน	13.83	55.49	53.70	60.64	80.90	80.76	338.47	45.79	50.39

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

5.7 การวิเคราะห์ตัวเลือกตัวตัดคำ

ในการทดลองในวิทยานิพนธ์นี้ได้ใช้ตัวตัดคำ “newmm” สำหรับใช้ในการคำนวณคะแนนเอฟวันระดับคำ อย่างไรก็ตาม ภาษาไทยก็มีตัวเลือกการใช้ตัวตัดคำอื่น ๆ ทางผู้เขียนจึงทำการทดลองเพิ่มเติมเพื่อเปรียบเทียบตัวตัดคำอื่น ๆ ทางผู้เขียนได้เลือก “AttaCut” ซึ่งเป็นตัวตัดคำที่พัฒนาจากแบบจำลองการเรียนรู้เชิงลึกสำหรับภาษาไทย ผลการเปรียบเทียบได้แสดงในตาราง 15 และ 16

จากผลการทดลองพบว่า คะแนนเอฟวันระดับคำจากตัวตัดคำ “newmm” จะมีค่ามากกว่าคะแนนเอฟวันจากตัวตัดคำ “AttaCut” ทุก ๆ กรณี ซึ่งหมายความว่า ตัวตัดคำ “AttaCut” มีการแบ่งคำที่ผิดพลาดมากกว่าตัวตัดคำ “newmm” ในคำที่มีความกำกวมซึ่งเป็นสาเหตุให้คะแนนเอฟวันระดับคำน้อยกว่า การทดลองนี้สามารถพิสูจน์ได้ว่าการเปลี่ยนตัวตัดคำจะส่งผลต่อคะแนนเอฟวันระดับคำ ในทางกลับกัน คะแนนเอฟวันระดับพยางค์จะสามารถแก้ปัญหาการตัดแบ่งคำได้ เนื่องจากการแบ่งคำเป็นพยางค์จะมีความกำกวมน้อยกว่า

ตาราง 15 ผลของคะแนนเอฟวันระดับคำในตัวตัดคำที่ต่างกันของชุดข้อมูล Thai Wiki QA

การทดลอง	“newmm” คะแนนเอฟวันระดับคำ	“AttaCut” คะแนนเอฟวันระดับคำ	ตัวแบ่งคำเป็นพยางค์ คะแนนเอฟวันระดับพยางค์
WangchanBERTa ไม่ใช้ข้อมูลเสริม	70.73	64.46	74.71
mT5-Base ไม่ใช้ข้อมูลเสริม	78.24	73.58	80.35
mT5-SEQ-FLT	83.64	78.98	85.29

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตาราง 16 ผลของคะแนนเอฟวันระดับคำในตัวตัดคำที่ต่างกันของชุดข้อมูล iApp Wiki QA

การทดลอง	“newmm” คะแนนเอฟวันระดับคำ	“AttaCut” คะแนนเอฟวันระดับคำ	ตัวแบ่งคำเป็นพยางค์ คะแนนเอฟวันระดับพยางค์
WangchanBERTa ไม่ใช้ข้อมูลเสริม	69.68	65.42	71.81
mT5-Base ไม่ใช้ข้อมูลเสริม	63.46	59.75	63.71
mT5-SEQ-ALL	81.87	78.87	83.15

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

บทที่ 6
สรุปการวิจัยและแนวทางการวิจัยในขั้นถัดไป

ในบทนี้นำเสนอแผนการดำเนินงานหลังจากการเสนอโครงร่างวิทยานิพนธ์

6.1 สรุปการวิจัย

วิทยานิพนธ์นี้นำเสนอการใช้งานแบบจำลองประเภททรานส์ฟอร์มเมอร์สำหรับงานการตอบคำถามภาษาไทย โดยมีจุดประสงค์ที่จะแก้ปัญหาข้อจำกัดด้านจำนวนชุดข้อมูลด้วยการสร้างชุดข้อมูลเสริมสำหรับฝึกสอน พร้อมกับการประมวลผลข้อมูลเบื้องต้น เพื่อเพิ่มประสิทธิภาพของแบบจำลองการตอบคำถามภาษาไทย ขั้นตอนการสร้างชุดข้อมูลเสริมประกอบด้วย 3 ขั้นตอน ได้แก่ การเลือกคำตอบ การสร้างคำถาม และการกรองคำถาม อีกทั้งในการทดลองได้ศึกษารูปแบบการฝึกสอนในหลายรูปแบบ ได้แก่ แบบจำลองการตอบคำถามพื้นฐาน รูปแบบการฝึกสอนชุดข้อมูล (การฝึกสอนแบบเป็นลำดับและการฝึกสอนแบบรวมข้อมูล) และกลุ่มของชุดข้อมูลเสริม นอกจากนี้ วิทยานิพนธ์นี้ได้นำเสนอรูปแบบของการวัดผลแบบจำลองเพิ่มเติม โดยใช้การประเมินผลระดับพยางค์ ซึ่งมีความเหมาะสมกับภาษาไทยมากกว่า เนื่องจากในการแบ่งพยงค์จะไม่มีความกำกวม ในการทดลองได้ทดลองบน 2 ชุดข้อมูลการตอบภาถามภาษาไทย ได้แก่ Thai Wiki QA และ iApp Wiki QA ผลการทดลองพบว่าด้วยวิธีการสร้างชุดข้อมูลเสริมที่เสนอ สามารถเพิ่มประสิทธิภาพของแบบจำลองได้ โดยมีประสิทธิภาพของคะแนน F1 ระดับคำจาก 78.24 เป็น 83.64 สำหรับชุดข้อมูล Thai Wiki QA และ 63.46 เป็น 81.97 สำหรับชุดข้อมูล iApp Wiki QA

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

6.2 ข้อเสนอแนะและข้อจำกัดในงานวิทยานิพนธ์นี้

อย่างไรก็ตาม วิธีการสร้างชุดข้อมูลเสริมที่นำเสนอในวิทยานิพนธ์นี้เหมาะสมกับงานด้านการตอบคำถามประเภทสกัดช่วงข้อความ (Span Extraction) เท่านั้น หมายความว่า คำตอบของคำถามต้องเป็นส่วนใดส่วนหนึ่งที่อยู่ติดกันในบทความเท่านั้น

6.3 แนวทางวิจัยในขั้นต่อไป

จากวิธีที่วิทยานิพนธ์นี้นำเสนอสามารถต่อยอดเพื่อทำงานวิจัยในขั้นถัดไปได้ ดังนี้

1. เนื่องจากวิธีที่นำเสนอเหมาะสำหรับงานด้านการตอบคำถามประเภทสกัดช่วงข้อความเท่านั้น โดยยังมีงานตอบคำถามอยู่อีกหลายประเภท ดังนั้น หากสามารถปรับปรุงขั้นตอนการสร้างชุดข้อมูลเสริมให้สามารถใช้งานกับงานตอบคำถามได้หลากหลายประเภทมากยิ่งขึ้น จะช่วยให้สามารถครอบคลุมการใช้งานที่หลากหลายรูปแบบมากกว่า
2. แบบจำลองประเภททรานส์ฟอร์เมอร์พื้นฐานที่ใช้ในวิทยานิพนธ์นี้ ได้แก่ แบบจำลอง WangchanBERTa และแบบจำลอง mT5 เป็นแบบจำลองที่ผ่านการฝึกสอนก่อนจากชุดข้อมูลที่ค่อนข้างหลากหลาย ในขณะที่ชุดข้อมูลที่นำมาทดสอบได้แก่ Thai Wiki QA และ iApp Wiki QA ถูกสร้างมาจากชุดข้อมูลที่สกัดมาจากบทความใน Wikipedia ซึ่งรูปแบบบทความอาจไม่ตรงกับสิ่งที่แบบจำลองได้เรียนรู้มาก่อน ดังนั้น หากสามารถนำแบบจำลองมาทำการฝึกสอนก่อนด้วยบทความที่มาจาก Wikipedia คาดว่าจะสามารถเพิ่มประสิทธิภาพได้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาคผนวก

ประกอบด้วยตัวอย่างของชุดข้อมูลเสริมที่ถูกสร้าง ดังนี้

ตัวอย่างของชุดข้อมูลเสริมที่ถูกสร้าง

ตาราง 17 ตัวอย่างของชุดข้อมูลเสริมที่ถูกสร้าง

1	บทความ	ประเทศฟิลิปปินส์ใน ค.ศ. 1979 เหตุการณ์ที่เกิดขึ้นใน ค.ศ. 1979 ในประเทศฟิลิปปินส์เหตุการณ์เหตุการณ์. - ประเทศฟิลิปปินส์ในโอลิมปิกซีเกมส์ 1979
	คำถาม	เหตุการณ์ที่เกิดขึ้นใน ค.ศ. 1979 ในประเทศใด
	คำตอบ	ฟิลิปปินส์
2	บทความ	เดอะคลิกไฟฟ์ เดอะคลิกไฟฟ์ () เป็นวงดนตรีสัญชาติอเมริกันจากบอสตัน แมสซาชูเซตส์ สหรัฐอเมริกาประวัติ ประวัติ. 4 คนจากสมาชิกดั้งเดิมจาก 5 คน เข้าศึกษาที่ berklee college of music ได้มีส่วนร่วมกับวงร็อกมาก่อน ก่อนที่จะรวมตัวกันในนามวง the click ซึ่งต่อมามีชื่อเป็น the click five สมาชิกในวงอยู่ด้วยกันในบอสตัน เช่าบ้านพักด้วยกันบนถนนอิมรี ซึ่งก็เป็นชื่ออัลบั้มแรกของพวกเขา พวกเขาแสดงในงานครั้งแรกให้กับคอนเสิร์ต kiss 108 fm hatchshell เมื่อกรกฎาคม 2004 หลังจากนั้นพวกเขาได้เซ็นสัญญากับค่ายใหญ่ ได้บินไปนิวยอร์กและลอสแอนเจลิส อยู่ใต้สังกัด epic records scout dave rocco พวกเขาได้เซ็นสัญญากับ lava records พวกเขาเริ่มอัดเพลงในอัลบั้มแรกที่ qdivision studios หลังจากเสร็จสิ้นการทำเพลง พวกเขาออกทัวร์โดยเริ่มจากงานเล็ก ๆ ในบอสตันไปจนที่ใหญ่ ๆ ในอเมริกาเหนือ มิวสิกวิดีโอตัวแรกของเขาคือ "just the girl" ออกในฤดูร้อนปี 2005 ทางช่องเอ็มทีวีและวีเอชวัน เขาออกอัลบั้มแรกที่ชื่อ greetings from imrie house เมื่อ 16 สิงหาคม ค.ศ. 2005 หลังจากออกทัวร์ในอเมริกายาวนาน 9 เดือนจึงทัวร์เอเชียในปี 2006 ในประเทศสิงคโปร์ มาเลเซีย และญี่ปุ่น ในปลายปี 2006 นักร้องนำอีริก ดิล ออกจากวงไป ต่อมาต้นปี 2008 ทางวงได้นักร้องนำคนใหม่ที่ชื่อ ไคล์ แพทริก นักเรียนจากเบิร์กลี อีกคน ทางวงได้บันทึกเสียงอัลบั้มที่ 2 ที่ชื่อ modern minds and pastimes วางขายเมื่อ 26 มิถุนายน ค.ศ. 2007 หลังจากทัวร์เล็ก ๆ ในอเมริกา ก็กลับมาในเอเชีย ต่อมา 2

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

		สิงหาคม ค.ศ. 2008 พวกเขาได้รับรางวัลเอ็มทีวี เอเชีย อวอร์ดส 2008 ที่มาเลเซีย พวกเขายังเป็นส่วนร่วมในโครงการให้ความรู้เกี่ยวกับการค้ามนุษย์ที่ชื่อโครงการว่า mtv exit โดยได้แสดงในหลายประเทศในเอเชีย รวมถึงประเทศไทย ที่ลานพาร์ก พารากอนผลงานผลงาน. - greetings from imrie house - modern minds and pastimes
	คำถาม	วงดนตรีสัญชาติอเมริกันจากบอสตัน แมสซาชูเซตส์ สหรัฐอเมริกา มีชื่อว่าอะไร
	คำตอบ	เดอะคลิกไฟฟ์
3	บทความ	ฟรันซ์ ดยุกแห่งโฮเฮนเบิร์ก ฟรันซ์ ดยุกแห่งโฮเฮนเบิร์ก ()พระประวัติ พระประวัติ. ฟรันซ์ ดยุกแห่งโฮเฮนเบิร์ก ประสูติเมื่อวันที่ 13 กันยายน พ.ศ. 2470 เป็นพระโอรสพระองค์ใหญ่ใน แม็กซิมิเลียน ดยุกแห่งโฮเฮนเบิร์ก และ เคาน์เตสเอลิซาเบธ วอลวัลเดอบูร์ก โดยทรงเป็นพระนัดดาใน อาร์ชดยุกฟรันซ์ แฟร์ดีนันด์แห่งออสเตรีย และ โซฟีอา ดัชเชสแห่งโฮเฮนเบิร์ก โดยเมื่อแรกประสูตินั้นทรงมีพระยศที่ ฮิสเซอรีนไฮเนส เจ้าชายฟรันซ์แห่งโฮเฮนเบิร์ก โดยในฐานะที่พระองค์นั้นเสกสมรสต่างฐานันดรทำให้ไม่มีสิทธิ์ในการสืบราชบัลลังก์ฮังการี หลังจากการล่มสลายของสถาบันพระมหากษัตริย์ในออสเตรีย พระอิสริยยศทั้งหมดจึงถูกยกเลิก และผู้ที่สืบเชื้อสายจากราชวงศ์นั้นจะต้องมีคำว่า ฟอน ขั้นระหว่างชื่อและราชสกุลนั้นๆ พระนามของพระองค์จึงเป็น ฟันซ์ ฟอน โฮเฮนเบิร์กเสกสมรส เสกสมรส. ฟรันซ์ ดยุกแห่งโฮเฮนเบิร์ก เสกสมรสกับ เจ้าหญิงเอลิซาเบธ ดัชเชสแห่งโฮเฮนเบิร์ก พระธิดาใน แกรนด์ดัชเชสชาร์ล็อตแห่งลักเซมเบิร์ก ณ เมืองลักเซมเบิร์ก เมื่อวันที่ 9 พฤษภาคม พ.ศ. 2499 โดยเจ้าหญิงเอลิซาเบธนั้น รับพระราชทานพระอิสริยยศที่ ดัชเชสแห่งโฮเฮนเบิร์ก โดยมีพระธิดาดังนี้1. เจ้าหญิงอนิตา 2. เจ้าหญิงโซฟี เนื่องจากพระองค์นั้นไม่มีพระโอรส ตำแหน่งดยุกแห่งโฮเฮนเบิร์กจึงตกไปเป็นของพระอนุชาคือ จอร์จ ดยุกแห่งโฮเฮนเบิร์ก
	คำถาม	ตำแหน่งดยุกแห่งโฮเฮนเบิร์กตกไปเป็นของพระอนุชาคือใคร
	คำตอบ	จอร์จ ดยุกแห่งโฮเฮนเบิร์ก
4	บทความ	มหาวิทยาลัยสุภานุวงศ์ มหาวิทยาลัยสุภานุวงศ์ (;) เป็นมหาวิทยาลัยรัฐในประเทศลาว ตั้งอยู่ที่หลวงพระบาง ชื่อของมหาวิทยาลัยตั้งตามเจ้าสุภานุวงศ์ ที่เป็น

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

		ประธานประเทศคนแรกของประเทศลาว ซึ่งเป็นวิศวกรที่ได้ศึกษาจบจากประเทศฝรั่งเศส มหาวิทยาลัยนี้ได้รับเงินสนับสนุนการก่อสร้างจากทางรัฐบาลประเทศเกาหลีใต้ อธิการบดีคนปัจจุบันคือ คำผาย ศรีสวัสดิ์การจัดการศึกษา การจัดการศึกษา. การจัดระบบการศึกษา หลักสูตรปริญญาตรี 5 ปี มีการเรียนการสอนวิชาวิทยาศาสตร์พื้นฐาน 1 ปี และวิชาเฉพาะแต่ละสาขา 4 ปี ปัจจุบันมีนักศึกษาประมาณ 3,900 คน ใน 5 คณะวิชา 14 สาขาวิชา ดังนี้1. คณะศึกษาศาสตร์ ประกอบด้วย สาขาการสอนคณิตศาสตร์ (ครูคณิต) สาขาภาษาลาวและวรรณคดีและสาขาภาษาอังกฤษ 2. คณะเศรษฐศาสตร์และการบริหาร ประกอบด้วย สาขาบริหารธุรกิจสากล(ธุรกิจระหว่างประเทศ) สาขาบริหารธุรกิจไอที สาขาบริหารธุรกิจการท่องเที่ยว สาขาเศรษฐศาสตร์และสาขาการบริหาร 3. คณะเกษตรศาสตร์ ประกอบด้วย สาขาการปลูกพืช สาขาการเลี้ยงสัตว์ และสาขาทรัพยากรป่าไม้ 4. คณะวิศวกรรมศาสตร์ ประกอบด้วย สาขาวิศวกรรมไฟฟ้า สาขาวิศวกรรมคอมพิวเตอร์ และสาขาวิศวกรรมโยธา 5. คณะสถาปัตยกรรมศาสตร์ มี 1 สาขา คือ สาขาสถาปัตยกรรมศาสตร์ 6. คณะอักษรศาสตร์ มี 2 สาขา คือ สาขาวิชาภาษาลาว และสาขาวิชาภาษาอังกฤษ
	คำถาม	ชื่อของมหาวิทยาลัยสุภานุวงศ์ ประเทศลาว ตั้งตามบุคคลใด
	คำตอบ	เจ้าสุภานุวงศ์
5	บทความ	ณภศศิ สุรวรรณ ณภศศิ สุรวรรณ เกิดเมื่อวันที่ 16 สิงหาคม พ.ศ. 2534 เป็นนักแสดงชาวไทย มายด์จบการศึกษาชั้นมัธยมศึกษาตอนต้นที่โรงเรียนสาธิต "พิบูลบำเพ็ญ" มหาวิทยาลัยบูรพา จังหวัดชลบุรี จบการศึกษาชั้นมัธยมศึกษาตอนปลายจากโรงเรียนเตรียมอุดมศึกษา ปัจจุบันจบการศึกษาจากคณะนิเทศศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย มายด์ผลงานในวงการบันเทิงมามากมาย โดยเฉพาะจากการแสดงโฆษณาต่างๆ เริ่มเข้าวงการจากการได้รับตำแหน่ง miss uthaitip freshy idol 2008 และมีผลงานแจ้งเกิดคือละครชุด อุบัติรักข้ามขอบฟ้า 2ประวัติประวัติ. มายด์มีพี่น้อง 2 คน โดยมีน้องสาวอีก 1 คน เรียนอยู่ที่โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ เดิมอาศัยอยู่ที่จังหวัดชลบุรี ต่อมาสอบได้ที่โรงเรียนเตรียมอุดมศึกษา สายศิลป์ คำนวณ (ต.อ.70) จึงย้ายมาอยู่ที่กรุงเทพมหานคร เมื่อ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

		เรียนอยู่ที่โรงเรียนเตรียมอุดมศึกษามายด์ทำกิจกรรมต่างๆมากมาย เช่น เป็นนางนพมาศ ปี 2550 ถือป้ายโรงเรียนในกิจกรรมกีฬาสี เชียร์ลีดเดอร์ในกีฬาประเพณีโรงเรียนเตรียมอุดมศึกษาและโรงเรียนเตรียมทหาร ครั้งที่ 25 และ 26 นอกจากนี้ยังอยู่ในโครงการความสามารถพิเศษทางภาษาไทย (gifted ไทย) ปัจจุบันศึกษาอยู่คณะนิเทศศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย เอกวารสารสนเทศ ผลงานในรั้วมหาวิทยาลัย เป็นผู้นำเชียร์ คณะนิเทศศาสตร์ ในงานกีฬาเฟรชชี่ และจุฬาคทากรผลงานละครโทรทัศน์ซีรีส์ภาพยนตร์พิธีกรผลงาน. พิธีกร. - พิธีกรรายการ honda live arcade by pgm-fl ออกอากาศทาง ช่อง bang channel - พิธีกรรายการ \'knock knock’ tv " @mind ตะลุย newzeland - พิธีกร รายการ bangster ออกอากาศทาง ช่อง bang channel - พิธีกร รายการ oishi kudasai ออกอากาศทาง ช่อง bang channel - รายการ pink riders บุษบาสารถี ช่อง 5 ทุกวันเสาร์ 16.05 (ร่วมกับ ส้ม - มารี เออเจนี เลอเลย์ และ ชาช่า - ริต์ตา รามณรงค์)นิตยสารนิตยสาร. - knock knock - centerpoint vol. 1 no. 24 january 2009 - เธอกับฉัน vol. 27 no. 603 june 2010โฆษณาโฆษณา. - กูลิโกะ ป๊อกกี้ - pretz รสแฮมและชีส - ดัชมิลล์ - ภาพนิ่ง พรีเซนเตอร์น้ำผลไม้ hapมิวสิกวิดีโอมิวสิกวิดีโอ. - เพลง darling ของ b.o.y. - เพลง ความเหงาโทรเข้ามา ของ เตชินท์ ชยุติรางวัลที่ได้รับรางวัลที่ได้รับ. - ชนะเลิศ การประกวด uthaitip freshy idol 2008ผลงานเพลงผลงานเพลง. - u tip freshy idol 4
	คำถาม	ผลงานแจ้งเกิดของณภศศิ สุรวรรณ คือละครชุดใด
	คำตอบ	อุบัติรักข้ามขอบฟ้า
6	บทความ	นาวน้ำ นาวน้ำ เป็นพืชในสกุลการเวก วงศ์ annonaceae เป็นไม้พุ่มรอเลื้อยเนื้อแข็ง ลำต้นและกิ่งมีหนามแข็งแหลมยาว ใบเดี่ยว ดอกเดี่ยว ออกตรงข้ามกับใบ สีเขียวหรือม่วงแดง กลิ่นแรงช่วงใกล้ค่ำ ผลแบบผลกลุ่ม ผลย่อยรูปหอกกลับ สุกเป็นสีเหลือง ออกดอกช่วงมกราคม – สิงหาคม พบครั้งแรกที่จังหวัดอุบลราชธานีโดยหมอคาร์เมื่อ 23 มกราคม พ.ศ. 2467
	คำถาม	นาวน้ำเป็นไม้พุ่มรอเลื้อยเนื้อแข็งพบครั้งแรกที่จังหวัดใด

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บรรณานุกรม

1. Seo, M., et al., *Bidirectional attention flow for machine comprehension.* arXiv preprint arXiv:1611.01603, 2016.
2. Decha, H. and K. Patanukhom. *Development of thai question answering system*. in *Proceedings of the 3rd International Conference on Communication and Information Processing*. 2017.
3. Lapchaicharoenkit, T. and P. Vateekul. *Machine Reading Comprehension on Multiclass Questions Using Bidirectional Attention Flow Models with Contextual Embeddings and Transfer Learning in Thai Corpus*. in *Proceedings of the 8th International Conference on Computer and Communications Management*. 2020.
4. Vaswani, A., et al. *Attention is all you need*. in *Advances in neural information processing systems*. 2017.
5. Devlin, J., et al., *Bert: Pre-training of deep bidirectional transformers for language understanding.* arXiv preprint arXiv:1810.04805, 2018.
6. Liu, Y., et al., *Roberta: A robustly optimized bert pretraining approach.* arXiv preprint arXiv:1907.11692, 2019.
7. Raffel, C., et al., *Exploring the limits of transfer learning with a unified text-to-text transformer.* arXiv preprint arXiv:1910.10683, 2019.
8. Liu, S., et al., *Neural machine reading comprehension: Methods and trends.* Applied Sciences, 2019. **9**(18): p. 3698.
9. Hermann, K.M., et al., *Teaching machines to read and comprehend.* Advances in neural information processing systems, 2015. **28**: p. 1693-1701.
10. Lai, G., et al., *Race: Large-scale reading comprehension dataset from examinations.* arXiv preprint arXiv:1704.04683, 2017.
11. Rajpurkar, P., et al., *Squad: 100,000+ questions for machine comprehension of text.* arXiv preprint arXiv:1606.05250, 2016.
12. Nguyen, T., et al. *MS MARCO: A human generated machine reading comprehension dataset*. in *CoCo@ NIPS*. 2016.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

13. Bahdanau, D., K. Cho, and Y. Bengio, *Neural machine translation by jointly learning to align and translate.* arXiv preprint arXiv:1409.0473, 2014.
14. Xue, L., et al., *mt5: A massively multilingual pre-trained text-to-text transformer.* arXiv preprint arXiv:2010.11934, 2020.
15. Joshi, M., et al., *Spanbert: Improving pre-training by representing and predicting spans.* Transactions of the Association for Computational Linguistics, 2020. **8**: p. 64-77.
16. Dhingra, B., D. Pruthi, and D. Rajagopal, *Simple and effective semi-supervised question answering.* arXiv preprint arXiv:1804.00720, 2018.
17. Puri, R., et al., *Training question answering models from synthetic data.* arXiv preprint arXiv:2002.09599, 2020.
18. Lowphansirikul, L., et al., *WangchanBERTa: Pretraining transformer-based Thai Language Models.* arXiv preprint arXiv:2101.09635, 2021.
19. Kudo, T. and J. Richardson, *Sentencepiece: A simple and language independent subword tokenizer and detokenizer for neural text processing.* arXiv preprint arXiv:1808.06226, 2018.
20. and, W.P., et al., *PyThaiNLP/pythainlp: PyThaiNLP v2.3.2 Release!* 2021.
21. Trakultaweekoon, K., et al. *The first wikipedia questions and factoid answers corpus in the thai language*. in *2019 14th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)*. 2019. IEEE.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ประวัติผู้เขียน

ชื่อ-สกุล	ภูริ ฟักมงคล
วัน เดือน ปี เกิด	14 กันยายน 2537
สถานที่เกิด	จังหวัดน่าน, ประเทศไทย
วุฒิการศึกษา	วิศวกรรมศาสตรบัณฑิต สาขาวิศวกรรมไฟฟ้าสื่อสารและอิเล็กทรอนิกส์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University