

การสร้างคำถามภาษาไทยโดยใช้ MT5



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2564
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

Question Generation in the Thai Language Using MT5



Miss Nutthanit Wiwatbutsiri

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	การสร้างคำถามภาษาไทยโดยใช้ MT5
โดย	น.ส.ณัฐนิช วิวัฒน์บุตรศิริ
สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	รองศาสตราจารย์ ดร.อติวงศ์ สุชาโต
อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม	รองศาสตราจารย์ ดร.โปรดปราน บุญยพุกกณะ ดร.เนืองวงศ์ ทวยเจริญ

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้บัณฑิตวิทยาลัยรับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของ
ของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

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

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(อาจารย์ ดร.เอกพล ช่างสูวนิช)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(รองศาสตราจารย์ ดร.อติวงศ์ สุชาโต)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม
(รองศาสตราจารย์ ดร.โปรดปราน บุญยพุกกณะ)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม
(ดร.เนืองวงศ์ ทวยเจริญ)

..... กรรมการภายนอกมหาวิทยาลัย
(ดร.อนันต์ลดา โชติมงคล)

ณัฐนิช วิวัฒน์บุตตรสิริ : การสร้างคำถามภาษาไทยโดยใช้ MT5. (Question Generation in the Thai Language Using MT5) อ.ที่ปรึกษาหลัก : รศ. ดร.อดิวงค์ สุชาติ, อ.ที่ปรึกษาร่วม : รศ. ดร.โปรดปราน บุญยพุกกณะ,ดร.เนืองวงศ์ ทวยเจริญ

มีงานวิจัยเกี่ยวกับเรื่องการสร้างคำถามอยู่จำนวนมากในขอบเขตภาษาอังกฤษแต่แทบไม่มีงานวิจัยเรื่องการสร้างคำถามในภาษาไทย มีชุดข้อมูลคำถาม-คำตอบในขอบเขตของภาษาอังกฤษมากกว่า 1 ล้านคู่คำถาม-คำตอบซึ่งมีจำนวนมากเมื่อเปรียบเทียบกับในขอบเขตของภาษาไทยที่มีอยู่เพียงประมาณ 12,000 คู่ งานวิจัยนี้ขอนำเสนอวิธีพัฒนาการสร้างคำถามอัตโนมัติจากบทความ โดยไม่ต้องมีคำตอบในการสร้างคำถาม ภายใต้เงื่อนไขการฝึกสอนจากชุดข้อมูลที่มีอยู่อย่างจำกัด โดยแบบจำลองการสร้างคำถามอัตโนมัติซึ่งฝึกสอนโดยแบบจำลองที่ผ่านการเรียนรู้มาก่อน MT5 จากชุดข้อมูลที่มนุษย์สร้างขึ้น สามารถสร้างคำถามจากชุดข้อมูลภาษาไทยที่เมื่อประเมินอัตโนมัติโดยวัดจากคะแนน BLEU-1 ได้คะแนน 56.19 เราจึงนำเสนอวิธีการเพิ่มประสิทธิภาพการสร้างคำถามจากการสังเคราะห์ข้อมูลและกลไกที่นำเสนอเพิ่มเติมโดยยังคงใช้เพียงแบบจำลองที่ผ่านการเรียนรู้มาก่อน MT5 ซึ่งแบบจำลองที่ผ่านการพัฒนาแล้วมีคะแนน BLEU-1 ถึง 59.03 มากกว่าแบบจำลองที่ผ่านมา นอกจากนี้ผลการประเมินประสิทธิภาพของคำถามโดยมนุษย์ยังแสดงคะแนนด้านความไพเราะ 4.40 คะแนน, ด้านความเกี่ยวข้องกับบทความ 4.65 คะแนนและด้านการตอบคำถามได้จากบทความ 4.7 คะแนนจากทั้งหมด 5 คะแนน

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์
ปีการศึกษา 2564

ลายมือชื่อนิสิต
ลายมือชื่อ อ.ที่ปรึกษาหลัก
ลายมือชื่อ อ.ที่ปรึกษาร่วม
ลายมือชื่อ อ.ที่ปรึกษาร่วม

6270075021 : MAJOR COMPUTER SCIENCE

KEYWORD: Automatic question generation, Thai language, Synthetic datasets

Nutthanit Wiwatbutsiri : Question Generation in the Thai Language Using MT5. Advisor: Assoc. Prof. ATIWONG SUCHATO, Ph.D. Co-advisor: Assoc. Prof. Proadpran Punyabukkana, Ph.D., Prof. NUENGWONG TUAYCHAROEN

There are numerous publications of Question Generation (QG) in English but less in Thai. More than million question-answer pairs are available in the English language, compared with only around 12,000 question-answer pairs in the Thai language. This paper presents a method to improve automatic Thai QG from given passages without an answer. Under a dataset of insufficient size. Our evaluation showed that a QG model which was trained by the pre-trained model MT5 from a Thai dataset achieved a BLEU-1 score of 56.19. We proposed a method to generate synthetic data and an additional mechanism by using a single pre-trained model. Our best model outperformed the previous model by achieving a BLEU-1 score of 59.03. The results and from human evaluation in fluency score was 4.40, the relevance score is 4.65, and the answer-ability score is 4.7 from 5.0.



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Field of Study: Computer Science

Academic Year: 2021

Student's Signature

Advisor's Signature

Co-advisor's Signature

Co-advisor's Signature

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สามารถสำเร็จลุล่วงได้ด้วยดี เนื่องจากอาจารย์ที่ปรึกษาทุกท่านที่สละเวลาให้คำแนะนำ คำปรึกษา และความเมตตาและตรวจแก้ไข จากรองศาสตราจารย์ ดร.อดิวงค์ สุชาติ อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก รองศาสตราจารย์ ดร.โปรดปราน บุญยพุกกณะและดร. เนื่องวงศ์ ทวยเจริญ อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม ผู้วิจัยจึงขอกราบขอบพระคุณอย่างสูงไว้ ณ ที่นี้

เพื่อปรับปรุงข้อบกพร่องต่างๆ ส่งผลให้วิทยานิพนธ์ฉบับนี้ถูกต้อง และมีความสมบูรณ์มากยิ่งขึ้น

ขอกราบขอบพระคุณ คณาจารย์คณะวิศวกรรมศาสตร์ ภาควิชาคอมพิวเตอร์ทุกท่านที่ให้ความรู้และคำแนะนำ

ข้อคิดเห็นต่างๆที่ดีเสมอมา

ขอกราบขอบพระคุณผู้ร่วมประเมินงานวิจัยนี้ทุกท่านที่ให้ความร่วมมือในการทำวิจัยเป็นอย่างดีมาโดยตลอด และขอขอบคุณเพื่อน ๆ พี่ ๆ น้อง ๆ ที่ให้คำแนะนำ ความช่วยเหลือ รับฟังและเป็นที่กำลังใจให้เสมอมา

ขอกราบขอบพระคุณทุกคนในครอบครัวที่ให้การสนับสนุนในเรื่องการศึกษา ให้คำแนะนำอบรม สั่งสอน

ตลอดจนเป็นแบบอย่างในการดำรงชีวิต และเป็นกำลังใจที่สำคัญยิ่ง ส่งผลให้ประสบความสำเร็จในการศึกษา

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ณัฐนิช วิวัฒน์บุตรสิริ

สารบัญ

	หน้า
.....	ค
บทคัดย่อภาษาไทย.....	ค
.....	ง
บทคัดย่อภาษาอังกฤษ.....	ง
กิตติกรรมประกาศ.....	จ
สารบัญ.....	ฉ
สารบัญตาราง.....	ช
สารบัญรูปภาพ.....	ฌ
บทที่ 1 บทนำ	1
1.1. ที่มาและความสำคัญ.....	1
1.2. วัตถุประสงค์ของงานวิจัย.....	2
1.3. ขอบเขตการวิจัย	2
1.4. ประโยชน์ที่คาดว่าจะได้รับ.....	2
บทที่ 2 ทฤษฎีที่เกี่ยวข้อง	3
2.1. การตัดคำ.....	3
2.2. การเข้ารหัส	4
2.3. การถอดรหัส.....	4
2.4. ทราจิสฟอรัมเมอร์	5
2.5. แบบจำลองที่ผ่านการฝึกสอนมาแล้วและการถ่ายทอดความรู้.....	6
2.6. ชุดข้อมูลฝึกสอนบทความและคู่คำถาม-คำตอบภาษาไทย.....	6
2.7. เอ็น-แกรม	11

2.8. คะแนนเบลอ	13
2.9. การตั้งคำถาม.....	13
บทที่ 3 งานวิจัยที่เกี่ยวข้อง	14
3.1. งานวิจัยเกี่ยวกับการสร้างชุดคำถาม	14
3.2. งานวิจัยเกี่ยวกับการสังเคราะห์ข้อมูล	16
บทที่ 4 แนวคิดและวิธีการดำเนินงาน.....	19
4.1. แบบจำลองการสร้างคำถาม.....	20
4.2. แบบจำลองการสร้างคำถามจากข้อมูลสังเคราะห์	24
บทที่ 5 การทดลองและผลการทดลอง	25
5.1. การเตรียมข้อมูล	25
5.2. การนำข้อมูลเข้าเพื่อฝึกสอน	28
5.3. การสร้างคำถามจากแบบจำลองการสร้างชุดคำถาม	29
5.4. การคัดกรองคำถาม	31
5.5. การสังเคราะห์ข้อมูล	34
5.6. การประเมินผลการทดลอง	34
บทที่ 6 สรุปผลการทดลอง ข้อเสนอแนะ และงานวิจัยในอนาคต	40
6.1. สรุปผลการทดลอง.....	40
6.2. ข้อเสนอแนะ	41
6.3. งานวิจัยในอนาคต	41
ภาคผนวก.....	1
บรรณานุกรม.....	5
ประวัติผู้เขียน.....	8

สารบัญตาราง

	หน้า
ตารางที่ 1 ชื่อเรียกและผลลัพธ์ที่แตกต่างตามจำนวนเอ็น.....	12
ตารางที่ 2 ป้ายการสกัดนิพจน์เฉพาะหรือชื่อเฉพาะที่ใช้ใน PyThaiNLP จาก https://pythainlp.github.io/dev-docs/api/tag.html	20
ตารางที่ 3 ตัวอย่างบทความรูปแบบนำเข้าปกติและรูปแบบนำเข้า NER	27
ตารางที่ 4 จำนวนข้อมูลในแต่ละชุดข้อมูล รูปแบบนำเข้าและประเภท	28
ตารางที่ 5 ตารางแสดงค่าพารามิเตอร์หลักที่ใช้ในการฝึกสอน	28
ตารางที่ 6 ตารางแสดงค่าพารามิเตอร์หลักที่ใช้ในการถอดรหัสเพื่อสร้างคำถาม	29
ตารางที่ 7 ตัวอย่างบทความและคำถามจากแต่ละแบบจำลอง	29
ตารางที่ 8 ผลการวัดความสามารถของการคัดกรองคำถามในรูปแบบของค่าความแม่นยำ(Precision) ค่าความถูกต้อง(Recall) ค่าความเที่ยงตรง(Accuracy) และค่าเฉลี่ยระหว่างความแม่นยำและความ ถูกต้อง(F1).....	33
ตารางที่ 9 สัดส่วนข้อมูลทดสอบเพื่อประเมินตามกลุ่มความยาวของบทความ	34
ตารางที่ 10 ผลการประเมินอัตโนมัติจากคะแนนเบลอเปรียบเทียบระหว่างแบบจำลองและรูปแบบ ข้อมูลนำเข้า	36
ตารางที่ 11 ผลการประเมินอัตโนมัติจากคะแนนเบลอเปรียบเทียบระหว่างแบบจำลองที่นำเสนอกับ งานวิจัยที่ใกล้เคียง	36
ตารางที่ 12 ตัวอย่างบทความ คำถามและผลการประเมินโดยมนุษย์.....	37
ตารางที่ 13 ผลประเมินคำถามจากผู้ประเมินแต่ละท่านตามเกณฑ์ที่กำหนด	38
ตารางที่ 14 ผลลัพธ์เปรียบเทียบการประเมินโดยมนุษย์ระหว่างงานวิจัยก่อนหน้าและงานวิจัยที่ นำเสนอโดยมีเกณฑ์การประเมินเดียวกัน.....	39

สารบัญรูปร่างภาพ

	หน้า
รูปที่ 1 การเตรียมข้อมูลนำเข้าในแต่ละรูปแบบ.....	15
รูปที่ 2 โครงสร้างแบบจำลองจากกลไกการตัดลอกจากสูญเสียบและทำนายประเภทคำถาม.....	16
รูปที่ 3 กระบวนการฝึกสอนและสังเคราะห์ข้อมูล.....	17
รูปที่ 4 ผลการใช้ข้อมูลสังเคราะห์ในแบบจำลองการตอบคำถาม.....	18
รูปที่ 5 เปรียบเทียบผลลัพธ์ของการไม่กรองข้อมูลสังเคราะห์ การกรองข้อมูลสังเคราะห์แบบไปกลับ และการสร้างคำถามเพิ่มขึ้นร่วมกับการกรองข้อมูลแบบไปกลับ จึงเป็นแนวทางในการนำไปพัฒนางานการสร้างคำถามจากการสังเคราะห์ชุดข้อมูล โดยจะใช้แบบจำลองการสร้างคำถามที่ได้จากการฝึกสอนในชุดข้อมูลที่มนุษย์สร้างขึ้น กำหนดให้สร้างชุดคำถามจากบทความมาจำนวน 5-10 ประโยค จากนั้นนำผลลัพธ์ที่ได้เข้าขั้นตอนคัดกรองประโยคคำถาม เพื่อนำไปใช้เป็นชุดข้อมูลฝึกสำหรับการสร้างคำถามในรอบต่อไป โดยมีความเป็นไปได้จากผลงานวิจัยข้างต้นว่าจะช่วยให้ผลการประเมินแบบจำลองการสร้างคำถามดียิ่งขึ้นจากการฝึกสอนจากข้อมูลสังเคราะห์.....	18
รูปที่ 6 แผนภาพการสร้างแบบจำลองการสร้างคำถาม.....	19
รูปที่ 7 ตัวอย่างข้อมูลบทความภาษาไทยจาก NSC ก่อนทำความสะอาดข้อมูล.....	25
รูปที่ 8 ตัวอย่างข้อมูลบทความภาษาไทยจาก NSC หลังทำความสะอาดข้อมูล.....	26
รูปที่ 9 ตัวอย่างการแปลงข้อมูลนำเข้าปกติเป็นข้อมูลนำเข้า NER.....	26

บทที่ 1

บทนำ

1.1. ที่มาและความสำคัญ

การศึกษาในปัจจุบันมีรูปแบบที่เปลี่ยนแปลงไปจากเดิม ผู้เรียนสามารถเข้าถึงแหล่งเรียนรู้จำนวนมากได้ง่ายขึ้นจากอินเทอร์เน็ต แบบฝึกหัดตอบคำถามเป็นเครื่องมือที่ช่วยพัฒนาผู้เรียนให้เกิดความชำนาญและเสริมสร้างความเข้าใจในเนื้อหาที่ศึกษา ที่ผ่านมาการสร้างแบบฝึกหัดต้องใช้ความรู้ เวลาและแรงกายของมนุษย์ในการสร้างแบบฝึกหัดตอบคำถาม แต่เนื่องด้วยเวลาและแรงกายของมนุษย์มีอยู่อย่างจำกัด ทำให้การสร้างแบบฝึกหัดอาจให้เพียงพอต่อแหล่งความรู้ที่มีอยู่อย่างมหาศาลในโลกของอินเทอร์เน็ต จึงนำการเรียนรู้ของเครื่องมาพัฒนาแบบจำลองที่ช่วยสร้างชุดคำถามอัตโนมัติจากข้อเท็จจริงที่ปรากฏในบทความที่น่าสนใจ ซึ่งจะทำได้แบบฝึกหัดช่วยในเรื่องของความจำอย่างไม่จำกัดตามแต่ที่ผู้เรียนต้องการ

ที่ผ่านมาได้มีการพัฒนาแบบจำลองการสร้างชุดคำถามอัตโนมัติด้วยวิธีต่าง ๆ โดยใช้ข้อมูลนำเข้าเป็นข้อความ เช่น การใช้กฎ (rule-based) สร้างรูปแบบของประโยคคำถาม [1, 2] การสร้างกราฟความรู้ (knowledge graph) [3] และการใช้แบบจำลองแบบต่อเนื่อง (sequence-to-sequence model) [4] ซึ่งปัจจุบันได้นำเอาองค์ความรู้ในเรื่องของการเรียนรู้เชิงลึก (deep learning) มาใช้ในการพัฒนาเพื่อให้ได้ผลลัพธ์เป็นประโยคคำถามที่มีคุณภาพมากขึ้นในการวัดผลคือมีความเป็นภาษามนุษย์ สามารถสร้างประโยคคำถามที่มีความหมายและเกี่ยวข้องกับบทความ ซึ่งการพัฒนาโดยใช้อุปกรณ์ความรู้ในเรื่องการเรียนรู้เชิงลึกจำเป็นต้องใช้ชุดข้อมูลคำถาม-คำตอบ ที่ประกอบไปด้วยบทความและชุดคำถามจำนวนมากเป็นตัวอย่างในการฝึกสอน งานวิจัยที่ผ่านมาได้มีการนำข้อมูลมาจากหลากหลายแหล่ง เช่น ชุดคำถาม-คำตอบของสแตนฟอร์ด (Stanford Question Answering Dataset หรือ SQuAD) ชุดคำถาม-คำตอบทริเวีย (TriviaQA) เป็นต้น ซึ่งเป็นชุดข้อมูลภาษาอังกฤษและมีจำนวนข้อมูลประมาณ 100,000 คู่คำถาม-คำตอบและ 95,000 คู่คำถาม-คำตอบตามลำดับ

สำหรับภาษาไทยชุดข้อมูลคำถาม-คำตอบที่มีการเผยแพร่ให้ผู้ที่สนใจสามารถนำไปใช้คือข้อมูลจากศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (National Electronics and Computer Technology Center หรือ NECTEC) ที่ใช้ในโครงการแข่งขันพัฒนาโปรแกรมคอมพิวเตอร์แห่งประเทศไทย (National Software Contest: NSC) ในหัวข้อการแข่งขันพัฒนาโปรแกรมถามตอบจากคลังข้อมูลวิกิพีเดียภาษาไทย (Question answering program from Thai Wikipedia) โดยมีชุดคำถาม 4,000 คู่คำถาม-คำตอบ ชุดข้อมูลคำถาม-คำตอบจากบริษัทไอแอปเทคโนโลยี (iApp Technology Co., Ltd.) โดยมีชุดคำถาม 9,170 คู่คำถาม-คำตอบและชุดข้อมูลภาษาไทยจากการแปลชุดคำถามของสแตนฟอร์ดเวอร์ชัน 1.1 ที่ได้รับการแปลโดยผู้เชี่ยวชาญด้านภาษา (Cross-lingual Question Answering Dataset หรือ XQuAD) มี 1,190 คำถาม จากงานวิจัย [5] ซึ่งให้เห็นว่าจำนวนข้อมูลที่ใช้ในการฝึกสอนมีผลต่อประสิทธิภาพของแบบจำลอง โดยการเรียนรู้เชิงลึกต้องการข้อมูลที่มีขนาดใหญ่สำหรับการฝึกสอนแบบจำลองให้มีประสิทธิภาพ เมื่อเปรียบเทียบระหว่างจำนวนข้อมูลชุดคำถาม-คำตอบภาษาไทยกับจำนวนข้อมูลชุดคำถาม-คำตอบที่เป็นภาษาอังกฤษพบว่ามีความแตกต่างกันมาก เมื่อนำชุดข้อมูล

ภาษาไทยดังกล่าวฝึกสอนแบบจำลองโดยการเรียนรู้เชิงลึกเพื่อสร้างคำถามภาษาไทย ทำให้ได้ผลลัพธ์ที่ไม่มีคุณภาพ เช่น ไม่ใช่ประโยคคำถาม ประโยคไม่เป็นภาษาธรรมชาติ เป็นคำถามที่ไม่ปรากฏคำตอบในบทความ

ดังนั้นจึงมีแนวคิดที่จะนำเสนออัลกอริทึมพัฒนาแบบจำลองการสร้างชุดคำถามภาษาไทยที่มีรูปแบบและโครงสร้างทางภาษาที่ซับซ้อนและแตกต่างจากภาษาอื่น ๆ ให้ได้ผลลัพธ์ที่มีคุณภาพมากขึ้นจากชุดข้อมูลที่มีจำกัด โดยนำแบบจำลองที่ผ่านการฝึกสอนมาแล้ว (pre-trained model) และการถ่ายทอดความรู้ (transfer learning) มาใช้ร่วมกับการสังเคราะห์ชุดข้อมูลฝึกสอนเพื่อให้ได้แบบจำลองการสร้างคำถามที่สามารถสร้างคำถามจากบทความที่กำหนดให้ ซึ่งจากงานวิจัยที่ผ่านมายังไม่มีผู้เสนอแบบจำลองการสร้างคำถามภาษาไทยเพื่อสร้างคำถามจากบทความที่กำหนดโดยไม่ใช้คำตอบในการฝึกสอนแบบจำลองสร้างคำถาม (answer-agnostic) และยังไม่มีการนำเสนอเรื่องการสังเคราะห์ชุดข้อมูลฝึกสอนเพื่อใช้ในการสร้างคำถามในภาษาไทยมาก่อน จากอัลกอริทึมที่นำเสนอ จะช่วยให้แบบจำลองสร้างคำถามได้อย่างอิสระภายใต้บทความที่กำหนด

1.2. วัตถุประสงค์ของงานวิจัย

นำเสนอแบบจำลองการสร้างคำถามภาษาไทยจากข้อเท็จจริงที่ปรากฏในบทความโดยไม่ใช้คำตอบในการฝึกสอนแบบจำลองสร้างคำถาม (answer-agnostic) ให้สามารถสร้างคำถามที่มีคุณภาพมากขึ้นจากชุดข้อมูลคำถาม-คำตอบภาษาไทยที่มีจำนวนจำกัด โดยนำเสนออัลกอริทึมเพื่อให้แบบจำลองเรียนรู้ข้อมูลฝึกสอนที่สังเคราะห์ขึ้นใหม่จากผลลัพธ์ของการสร้างชุดคำถามและกระบวนการคัดกรอง

1.3. ขอบเขตการวิจัย

- 1.3.1. แบบจำลองฝึกสอนจากชุดข้อมูลภาษาไทยซึ่งเป็นข้อมูลที่นำมาจากวิกิพีเดียภาษาไทย และสร้างคำถาม คำตอบโดยผู้เชี่ยวชาญทางภาษา
- 1.3.2. คำถามที่สร้างจากแบบจำลองเป็นคำถามข้อเท็จจริงจากบทความ ซึ่งมีคำตอบคำถามได้แก่ ใคร อะไร ไหน ที่ไหน เมื่อไร ใด ก็เท่าไร
- 1.3.3. ไม่รองรับคำที่พิมพ์ผิดจากชุดข้อมูลที่สร้างโดยมนุษย์
- 1.3.4. ความยาวของบทความไม่น้อยกว่า 100 ตัวอักษรและไม่เกิน 3,000 ตัวอักษร

1.4. ประโยชน์ที่คาดว่าจะได้รับ

- 1.4.1. แบบจำลองการสร้างคำถามภาษาไทยที่ฝึกสอนมาแล้ว
- 1.4.2. ชุดข้อมูลบทความและคำถามภาษาไทย
- 1.4.3. เกณฑ์พื้นฐานในงานการสร้างคำถามภาษาไทย

บทที่ 2

ทฤษฎีที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงทฤษฎีที่สำคัญที่ใช้ในการสร้างและพัฒนาแบบจำลองสำหรับงานวิจัยนี้ โดยงานวิจัยนี้ประกอบด้วยแบบจำลอง 2 ส่วนคือแบบจำลองการสร้างชุดคำถามและแบบจำลองคัดกรองคำถาม ทั้ง 2 แบบจำลองเป็นการประมวลผลภาษาธรรมชาติซึ่งจะต้องมีการแปลงข้อมูลที่เป็นประโยคภาษาธรรมชาติให้เป็นข้อมูลที่เครื่องเข้าใจและประมวลผล เริ่มจากการการตัดคำ (tokenization) เพื่อแยกประโยคให้เป็นคำ จากนั้นนำคำเข้ารหัสโดยการเข้ารหัส (encoding) เพื่อให้สามารถประมวลผลผ่านการเรียนรู้ของเครื่อง (machine learning) ได้แบบจำลองที่มีการฝึกสอนในงานที่ต้องการ

สำหรับแบบจำลองการสร้างชุดคำถามได้นำแบบจำลองที่ผ่านการฝึกสอนมาแล้ว (pre-trained model) และการถ่ายทอดความรู้ (transfer learning) มาใช้โดยการปรับค่าพารามิเตอร์และฝึกสอนโดยชุดข้อมูลที่กำหนด (fine-tuning) แบบจำลองนี้มีสถาปัตยกรรมทรานส์ฟอร์มเมอร์ (transformer) ซึ่งมีความสามารถในด้านการประมวลผลภาษาธรรมชาติ แบบจำลองการสร้างคำถามจะถูกนำเข้าบทความเพื่อประมวลผลและถอดรหัสโดยการถอดรหัส (decoding) ได้ผลลัพธ์เป็นชุดคำถาม

2.1. การตัดคำ

การตัดคำคือการแบ่งคำในข้อความเพื่อเข้ารหัสแต่ละคำแทนข้อความทั้งประโยค สามารถทำได้หลายรูปแบบ เช่น การแบ่งข้อความเป็นตัวอักษร การแบ่งข้อความเป็นวลี การแบ่งข้อความเป็นคำ ซึ่งแต่ละวิธีมีข้อดีข้อเสียแตกต่างกันออกไปขึ้นอยู่กับนำไปใช้ หากเลือกตัวตัดคำ (tokenizer) ที่เหมาะสมจะช่วยให้ได้คำที่มีความหมายถูกต้องตามบริบทของประโยคซึ่งมีผลต่อประสิทธิภาพในงานเนื่องจากมีผลต่อความหมายของคำ สำหรับภาษาไทยที่มีลักษณะการเขียนในประโยคติดกันไม่มีการเว้นวรรคระหว่างคำทำให้ความซับซ้อนในการตัดคำมากขึ้น มีผู้สร้างและเผยแพร่ตัวตัดคำสำหรับภาษาไทยเพื่อให้ผู้ใช้งานนำไปใช้ในงานด้านการประมวลผลทางภาษา เช่น ตีปัด (deepcut) [6] อดตาคัด (attacut) [7] ฯลฯ การตัดคำจะถูกนำไปใช้ทั้งส่วนของบทความและชุดคำถาม

ตัวอย่าง “ผิดกะเพรานั้นจะเกิดขึ้นราวสมัยรัชกาลที่ 7 เพราะคนจีนนำเอามาขายในร้านอาหารตามสั่ง อย่งไรก็ดี ผัดกะเพรานั้นจะเพ็งนิยมราว พ.ศ. 2500”

การตัดคำโดยอดตาคัด: ['ผิด', 'กะเพรา', 'น่า', 'จะ', 'เกิด', 'ขึ้น', 'ราว', 'สมัย', 'รัชกาล', 'ที่', '7', 'เพราะ', 'คน', 'จีน', 'นำ', 'เอา', 'มา', 'ขาย', 'ใน', 'ร้าน', 'อาหาร', 'ตาม', 'สั่ง', 'อย่าง', 'ไร', 'ก็', 'ดี', 'ผิด', 'กะเพรา', 'น่า', 'จะ', 'เพ็ง', 'นิยม', 'ราว', 'พ.ศ.', '2500']

การตัดคำโดยตีปัด: ['ผิด', 'กะเพรา', 'น่า', 'จะ', 'เกิด', 'ขึ้น', 'ราว', 'สมัย', 'รัชกาล', 'ที่', '7', 'เพราะ', 'คน', 'จีน', 'นำ', 'เอา', 'มา', 'ขาย', 'ใน', 'ร้าน', 'อาหาร', 'ตาม', 'สั่ง', 'อย่าง', 'ไร', 'ก็', 'ดี', 'ผิด', 'กะ', 'เพรา', 'น่า', 'จะ', 'เพ็ง', 'นิยม', 'ราว', 'พ.ศ.', '2500']

2.2. การเข้ารหัส

การเข้ารหัสคือการแปลงข้อมูลรูปภาพหรือข้อความให้อยู่ในรูปแบบเวกเตอร์ตัวเลขเพื่อให้เครื่องสามารถประมวลผลจากข้อมูลได้ ในด้านภาษาแล้วเราแสดงข้อความให้อยู่ในรูปแบบตัวเลขได้หลายวิธี เช่น การเข้ารหัสจากดัชนี (index-based encoding) คำฝังตัว (word embedding) โดยตัวเลขมีนัยสำคัญในการแทนที่คำนั้น ๆ ขึ้นอยู่กับแต่ละวิธี

2.2.1. การเข้ารหัสจากดัชนีคือการแปลงข้อความให้อยู่ในรูปแบบเวกเตอร์ตัวเลข โดยเป็นการจับคู่ระหว่างคำและตัวเลขดัชนีของคำซึ่งอ้างอิงจากพจนานุกรมคำศัพท์ ในการเข้ารหัสด้วยวิธีนี้จะมีการกำหนดขนาดสูงสุดของเวกเตอร์ ในกรณีที่ข้อมูลไม่ถึงตามที่กำหนดจะมีการแทนค่าด้วยตัวเลข “0” ให้ครบตามจำนวนข้อมูลในเวกเตอร์

ตัวอย่าง “เขาสวมเสื้อสีเขียว” จากตัวอย่างพจนานุกรมคำศัพท์

พจนานุกรมคำศัพท์

ดัชนี	...	259	...	278	...	595	596	...	610	...
คำศัพท์	...	เขา	...	เขียว	...	สวม	เสื้อ	...	สี	...

$$\text{รูปแบบเวกเตอร์ตัวเลขดังนี้} \begin{bmatrix} 259 & 595 & 596 & 610 \\ 278 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

2.2.2. คำฝังตัวคือการแปลงข้อความให้อยู่ในรูปแบบเวกเตอร์ตัวเลขที่มีค่าจากการเรียนรู้ความสัมพันธ์ของคำ หากคำที่มีความใกล้เคียงกันจะมีการเข้ารหัสที่คล้ายกัน ลดความเบาบางของข้อมูลทำให้การเข้ารหัสมีประสิทธิภาพซึ่งเป็นตัวเลขทศนิยมที่เป็นได้ทั้งค่าบวกและลบ ค่าดังกล่าวเป็นค่าที่ได้จากการคำนวณจากบริบทของคำที่ใกล้เคียงค่าดังกล่าวจากพจนานุกรมที่กำหนด โดยผู้สร้างสามารถกำหนดค่าพารามิเตอร์ เช่น ขนาดของเวกเตอร์ที่ต้องการ จำนวนสูงสุดของคำที่ใช้ในการฝึกฝน จำนวนรอบที่ฝึกฝน การสร้างเวกเตอร์ตัวเลขเพื่อแทนคำที่นิยมใช้กัน เช่น เวิร์ดทูเวก (Word2Vec) [8] และโกลฟ (GloVe) [9]

ตัวอย่าง “เขาสวมเสื้อสีเขียว” โดยใช้เวิร์ดทูเวกของชุดข้อมูลพจนานุกรมคำศัพท์ตัวอย่าง

$$\text{รูปแบบเวกเตอร์ตัวเลขดังนี้} \begin{bmatrix} 1.2 & -0.1 & 4.3 & 3.2 \\ 0.8 & 0.4 & 0.1 & 0.5 \\ 2.7 & 1.1 & -0.9 & 0.4 \\ 1 & 0.3 & 2.5 & -0.7 \\ -0.6 & 2.1 & 3.8 & 2.1 \end{bmatrix}$$

2.3. การถอดรหัส

การถอดรหัสคือการแปลงข้อมูลรูปแบบเวกเตอร์ตัวเลขให้อยู่ในรูปแบบภาษาธรรมชาติที่มนุษย์สามารถเข้าใจได้ โดยใช้ความน่าจะเป็นในการเลือกคำที่เหมาะสมในภาษาธรรมชาติจากพจนานุกรม เมื่อเริ่มต้นจะมีการกำหนดค่าเริ่มต้นให้ การถอดรหัสสามารถทำได้หลายวิธีโดยมีอัลกอริทึมที่ใช้เลือกคำจากค่าความน่าจะเป็นต่างกันไปดังนี้

- การค้นหาแบบละโมภ (greedy search) คือ การเลือกคำที่มีความน่าจะเป็นสูงที่สุดให้เป็นคำต่อไป โดยพิจารณาจากคำก่อนหน้าไปเรื่อย ๆ จนจบการทำงานเมื่อพบคำเฉพาะจบประโยคหรือครบจำนวนคำตามความยาวที่กำหนด โดยมีสูตรการคำนวณดังนี้

$$w_t = \operatorname{argmax}_w P(w|w_{1:t-1}) \quad (2.1)$$

- การค้นหาแบบลำแสง (beam search) คือ พิจารณาคำที่มีความน่าจะเป็นสูงสุดตามจำนวนบีมที่กำหนด เพื่อเป็นทางเลือกในการเลือกคำต่อไปที่อาจให้ความน่าจะเป็นที่ดีกว่า ในแต่ละทางเลือกจะจบการทำงานเมื่อพบคำเฉพาะจบประโยคหรือครบจำนวนคำตามความยาวที่กำหนด ทำจนกระทั่งจบการทำงานทุกทางเลือก ให้เลือกเส้นทางที่มีค่าความน่าจะเป็นสูงที่สุดแล้วสร้างประโยคจากคำที่เส้นทางนั้นเลือกทั้งหมด
- การสุ่มตัวอย่าง k ตัวแรก (top-k sampling) คือ การเลือกคำที่มีความน่าจะเป็นสูงที่สุด k ตัวแรก (top-k) หากกำหนดให้ k เป็น 1 จะเป็นการค้นหาแบบละโมภและสุ่มเลือกข้อมูลที่มีความน่าจะเป็นจาก k ตัวสามารถใช้ร่วมกับความน่าจะเป็นที่ p (top-p) ที่จะพิจารณาค่าความน่าจะเป็นเฉพาะที่มีผลรวมตามค่า p ที่กำหนด กรณีที่ใช้การสุ่ม k ตัวแรกร่วมกับความน่าจะเป็นที่ p อัลกอริทึมจะสุ่ม k ตัวแรกก่อนจึงเลือกคำที่มีความน่าจะเป็นที่ p มาสุ่มเลือกในการสร้างประโยค นอกจากนี้มีการนำค่าอุณหภูมิ (temperature) มาช่วยเลือกคำที่เหมาะสมขึ้นโดยการเพิ่มความน่าจะเป็นไปได้ของคำที่มีความน่าจะเป็นสูงและลดความน่าจะเป็นไปได้ของคำที่มีความน่าจะเป็นน้อยตัวอย่างดังสูตรที่ 2.2 และ 2.3 คือ

$$\sum_{w \in V_{top-k}} P(w| \text{"ผัด"}) \quad (2.2)$$

$$\sum_{w \in V_{top-k}} P(w| \text{"ผัด"}, \text{"กะเพรา"}) \quad (2.3)$$

ตัวอย่าง การถอดรหัสเพื่อสร้างคำถามจากบทความนำเข้า

การค้นหาแบบละโมภ	การค้นหาแบบลำแสง	การสุ่มตัวอย่าง k ตัวแรก
ข้าวผัดกะเพรามีต้นกำเนิดมาจากประเทศใด	ผัดกะเพราเป็นอาหารที่ใช้เนื้อสัตว์ชนิดใดเป็นหลัก	ผัดกะเพรามีชื่อเป็นภาษาอังกฤษว่าอะไร

2.4. ทรานส์ฟอร์มเมอร์

ทรานส์ฟอร์มเมอร์ (Transformer) เป็นสถาปัตยกรรมออกแบบโครงสร้างมาเพื่อแก้ปัญหาด้านภาษาในงานที่เป็นแบบต่อเนื่อง โดยนำกลไกแอตเทนชันประเภทเซลฟ์แอตเทนชัน (self-attention) มาช่วยแทนการใช้โครงข่ายประสาทเทียมแบบคอลโลลูชัน (Convolutional Neural Network หรือ CNN) หรือ โครงข่ายประสาทเทียมแบบวนกลับ (Recurrent Neural Network หรือ RNN) ซึ่งช่วยแก้ปัญหากรณีข้อความมีขนาดยาวและความเข้าใจในบริบทของภาษามากขึ้น กลไกของเซลฟ์แอตเทนชันคือการให้แบบจำลองเรียนรู้ด้วยตนเองโดยนำคำทุกคำมา

เปรียบเทียบกันเองในประโยคเพื่อเลือกค่าที่เหมาะสมและแสดงออกมาเป็นค่าน้ำหนัก (weighted values) นำไปใช้กับข้อมูลนำเข้าตัวถัดไป แต่ปัญหาที่พบคือกรณีที่มีความสัมพันธ์ของค่ามีความหลากหลาย จึงได้นำเทคนิคหลาย ๆ อย่างเข้ามาช่วยให้การทำงานมีประสิทธิภาพมากขึ้น เช่น การนำเทคนิคแอดเทนชันแบบหลายหัว (multi-head attention) เป็นเทคนิคที่นำเซลล์แอดเทนชันหลาย ๆ ชุดมาหาความสัมพันธ์แล้วจึงประมวลผลรวมเป็นค่าเดียวและเลือกค่าที่เหมาะสมที่สุด ทราנסฟอร์มเมอร์จึงเป็นที่นิยมในการสร้างแบบจำลองที่แก้ไขปัญหาของงานการประมวลผลภาษาธรรมชาติโดยมีประเภทแตกต่างกันจากการเรียนรู้ของแบบจำลอง เช่น ทีโฟฟ (T5) เอ็มทีโฟฟ (MT5) เบิร์ต (Bidirectional Encoder Representations from Transformers หรือ BERT) เอ็กแอลเอ็ม (XLM)

เนื่องจากประสิทธิภาพในการทำงานของทราנסฟอร์มเมอร์ที่กล่าวไว้ข้างต้น ในงานวิจัยนี้จึงนำแบบจำลองที่มีสถาปัตยกรรมทราנסฟอร์มเมอร์ประเภทข้อความถึงข้อความ (text-to-text transformer) ซึ่งจะมีฟังก์ชันการทำงานที่รองรับการนำเข้าข้อมูลประเภทข้อความและการสร้างข้อมูลนำออกเป็นข้อความ โดยผู้ใช้งานสามารถปรับแต่งค่าพารามิเตอร์ของแต่ละฟังก์ชันได้อย่างอิสระ จากขอบเขตในเอกสารที่ระบุรายละเอียดฟังก์ชันและคุณสมบัติของพารามิเตอร์เพื่อให้แบบจำลองมีความเหมาะสมในการเรียนรู้งานที่ต้องการจากชุดข้อมูลที่ฝึกสอน

2.5. แบบจำลองที่ผ่านการฝึกสอนมาแล้วและการถ่ายทอดความรู้

ในอดีตการสร้างแบบจำลองเพื่อใช้ในการแก้ปัญหาต่าง ๆ มักจะเริ่มต้นศูนย์ ทำให้ต้องใช้เวลาและข้อมูลจำนวนมากในการฝึกแบบจำลองเพื่อให้ได้ผลลัพธ์ตามวัตถุประสงค์ที่ต้องการ จึงได้มีการสร้างแบบจำลองให้ฝึกสอนในเรื่องพื้นฐานจากคลังข้อมูลขนาดใหญ่ก่อน เราเรียกแบบจำลองนี้ว่าแบบจำลองที่ผ่านการฝึกสอนมาแล้ว (pre-trained model) ได้น้ำหนักจากการฝึกสอนมาแล้ว (pre-trained weights) และนำค่าน้ำหนักนี้ไปใช้เป็นตัวตั้งต้นสำหรับการแก้ปัญหาในเรื่องที่เราสนใจ

การถ่ายทอดความรู้ (Transfer learning) คือแก้ปัญหาที่สนใจโดยการนำแบบจำลองที่ผ่านการฝึกสอนมาแล้วมาจากข้อมูลขนาดใหญ่เป็นแบบจำลองตั้งต้น จากนั้นปรับค่าพารามิเตอร์ (fine-tuning) ให้สอดคล้องกับชุดข้อมูลที่เรากำหนด เพื่อนำไปฝึกสอนให้ได้ผลลัพธ์ตามวัตถุประสงค์ ซึ่งช่วยลดเวลาในการฝึกแบบจำลองและเพิ่มประสิทธิภาพแบบจำลองได้ดียิ่งขึ้น มีความเหมาะสมกับงานวิจัยนี้ที่มีชุดข้อมูลฝึกสอนภาษาไทยน้อยเมื่อเปรียบเทียบกับชุดข้อมูลฝึกสอนลักษณะเดียวกันในภาษาอังกฤษ การใช้อรรถความรู้ในเรื่องการถ่ายทอดความรู้จึงมีความเป็นไปได้สูงที่จะช่วยให้งานด้านการสร้างคำถามอัตโนมัติในภาษาไทยมีประสิทธิภาพมากกว่าการสร้างแบบจำลองการสร้างคำถามอัตโนมัติในภาษาไทยตั้งแต่เริ่มต้นในชุดข้อมูลที่มีจำนวนน้อย

2.6. ชุดข้อมูลฝึกสอนบทความและคู่คำถาม-คำตอบภาษาไทย

การเรียนรู้ของเครื่องจำเป็นต้องใช้ชุดข้อมูลเพื่อฝึกสอนแบบจำลองในเรียนรู้ข้อมูลนำเข้าเพื่อให้ได้ผลลัพธ์ตามที่ต้องการ งานวิจัยนี้ใช้การเรียนรู้ของเครื่องในการพัฒนาแบบจำลองการสร้างชุดคำถามและแบบจำลองคัดกรองคำถาม ข้อมูลฝึกสอนจึงเป็นส่วนสำคัญในการพัฒนางานวิจัยนี้ ซึ่งได้ใช้ชุดข้อมูลบทความและคู่คำถาม-คำตอบจากหลายแหล่งที่มาดังนี้

- ข้อมูลบทความและคำถาม-คำตอบภาษาไทยจาก NECTEC [10] เป็นข้อมูลที่รวบรวมจากคลังข้อมูลวิกิพีเดียภาษาไทย ที่มีการสร้างชุดคำถาม-คำตอบโดยผู้ใช้งานทั่วไปจากบทความบนวิกิพีเดียภาษาไทยที่มีเนื้อหาหลากหลาย เช่น ด้านวิทยาศาสตร์ การท่องเที่ยว กีฬา ฯลฯ มีวัตถุประสงค์สำหรับใช้ในการตอบคำถาม ดาวน์โหลดเมื่อวันที่ 17 ธันวาคม พ.ศ. 2560 มีการคัดเลือกคำถามโดยจัดหมวดหมู่ของคำถามได้ 7 ประเภท ได้แก่ อะไร ใคร ไหน เมื่อไหร่ ไต ก็ เท่าไร โดยไม่นับรวมคำถามที่มีความซับซ้อน หลังจากคัดเลือกแล้วทำให้ได้ชุดข้อมูลที่มีจำนวนบทความ 2,289 บทความและคู่คำถาม-คำตอบจำนวน 4,000 คู่ จัดเก็บในรูปแบบเจสัน (JavaScript Object Notation หรือ JSON) เผยแพร่ให้ผู้ที่สนใจสามารถดาวน์โหลดเพื่อนำไปใช้ได้ที่ <https://aiforthai.in.th/corpus.php>



ตัวอย่าง

```
<doc id="115035" url="https://th.wikipedia.org/wiki?curid=115035" title="เบนจี้">เบนจี้ เบนจี้
() เป็นชื่อตัวละครหมาพันธุ์พันธุ์ที่ปรากฏอยู่ในภาพยนตร์หลายเรื่องที่เขียนบท และกำกับโดย โจ
แคมป์ ในช่วงทศวรรษ 1970 ถึง 1980 ภาพยนตร์เรื่องแรกในชุด ใช้ชื่อเรื่องว่า เบนจี้ ... Benji: Off the
Leash! (ภาพยนตร์) - พ.ศ. 2550, Benji: The Barkening (ภาพยนตร์)</doc>
```

```
[
  {
    "question_id":1,
    "question_type":1,
    "question":"เอเลี่ยนส์เป็นภาพยนตร์แนวไหน",
    "answer":"นิยายวิทยาศาสตร์แอคชั่น",
    "answer_begin_position ":149,
    "answer_end_position":172,
    "article_id":473065
  },
  {
    "question_id":2,
    "question_type":1,
    "question":"พระเจ้าอาฟองโซ่ที่ 1 แห่งโปรตุเกส เป็นผู้สืบเชื้อสายจากราชวงศ์ใด",
    "answer":"ราชวงศ์กาเปเตียง",
    "answer_begin_position ":218,
    "answer_end_position":234,
    "article_id":289367
  }
]
```

- ข้อมูลบทความและคำถาม-คำตอบภาษาไทยจากบริษัทไอแอปเทคโนโลยีจำกัด เป็นชุดข้อมูลจากวิกิพีเดียภาษาไทยจากบทความ 1,961 บทความและคู่คำถาม-คำตอบ 9,170 คู่ จัดเก็บในรูปแบบเจสัน (JavaScript Object Notation หรือ JSON) เผยแพร่ให้ผู้ที่สนใจสามารถดาวน์โหลดเพื่อนำไปใช้ได้ที่ <https://github.com/iapp-technology/iapp-wiki-qa-dataset>

ตัวอย่าง

```
{
  'article_id': '01KZTrxgvC5mOovXFMPJ',
  'question_id': '01KZTrxgvC5mOovXFMPJ_000',
  'context': 'พัทธ์ธีรา ศรุติพงษ์โกคิน (เกิด 3 ธันวาคม พ.ศ. 2533) หรือชื่อเล่นว่า อร เป็น
นักแสดงหญิงชาวไทย สำเร็จมัธยมศึกษาจากCatholic Cathedral College ประเทศนิวซีแลนด์ และ
ปริญญาตรีจากRaffles International College สาขา Business Marketing\๓\๓เข้าสู่วงการตั้งแต่อายุ 6
ขวบ ... เป็นพิธีกรให้กับช่อง ทูรู มิวสิก',
  'question': 'พัทธ์ธีรา ศรุติพงษ์โกคิน เกิดวันที่เท่าไร',
  'answers': {
    'text': ['3 ธันวาคม พ.ศ. 2533'],
    'answer_start': [31],
    'answer_end': [50]
  },
  'title': 'พัทธ์ธีรา ศรุติพงษ์โกคิน',
  'created_by': 'gmnjGRF0y0g7QRZDd9Qgz3AgiHJ3',
  'created_on': '2019-08-07 14:00:38.778000+00:00',
  'is_pay': {
    'status': True,
    'total': 2.5,
    'date': '2019-08-13 10:47:28.095000+00:00'
  }
}
```

- ข้อมูลบทความและคำถาม-คำตอบภาษาไทยจากการแปลชุดคำถามของสแตนด์ฟอร์เวิร์ด 1.1 ที่ได้รับการแปลโดยผู้เชี่ยวชาญด้านภาษา (Cross-lingual Question Answering Dataset หรือ XQuAD) [32] แปลเป็น 10 ภาษาได้แก่ ภาษาสเปน ภาษาเยอรมัน ภาษากรีก ภาษารัสเซีย ภาษาตุรกี ภาษาอารบิก ภาษาเวียดนาม ภาษาจีน ภาษาฮินดีและภาษาไทย มีวัตถุประสงค์สำหรับใช้ในงานการตอบคำถามข้ามภาษา ได้ชุดข้อมูลบทความ 240 บทความและคู่คำถาม-คำตอบ 1,190 คู่ จัดเก็บในรูปแบบเจสัน (JavaScript Object Notation หรือ JSON) เผยแพร่ให้ผู้ที่สนใจสามารถดาวน์โหลดเพื่อนำไปใช้ได้ที่ <https://github.com/deepmind/xquad>

ตัวอย่าง

```
{
  "context": "ทีมรับของแพนเธอร์สถอดใจที่คะแนน 308 ได้อันดับที่หกของลีก ในขณะที่เป็นผู้นำในเอ็นเอฟแอลด้วยการอินเตอร์เซป 24 ครั้งและได้รับเลือกให้เล่นในโปรโบว์ล สี่ ครั้ง ... ผู้เล่นในตำแหน่งคอร์เนอร์แบ็กซึ่งกลายมาเป็นซัดตาวนคอร์เนอร์ระหว่างฤดูกาล และทำอินเตอร์เซปไปสี่ครั้ง ซึ่งสองครั้งในนั้นกลายเป็นทัซตาวน",
  "qas": [
    {
      "answers": [
        {
          "answer_start": 33,
          "text": "308"
        }
      ],
      "id": "56beb4343aeaaa14008c925b",
      "question": "ทีมรับของแพนเธอร์สยอมแพ้ที่คะแนนเท่าไร"
    }
  ]
}
```

- ข้อมูลบทความจากวิกิพีเดียภาษาไทย เป็นบทความจากวิกิพีเดียภาษาไทยที่ได้จากโครงการแข่งขันพัฒนาโปรแกรมคอมพิวเตอร์แห่งประเทศไทย (National Software Contest หรือ NSC) ในการแข่งขันพัฒนาโปรแกรมถามตอบจากคลังข้อมูลวิกิพีเดียภาษาไทย (Question answering program from Thai Wikipedia) มีจำนวนบทความทั้งหมดประมาณ 125,000 บทความ สำหรับงานวิจัยนี้บทความจะถูกนำไปใช้ในการสังเคราะห์ข้อมูลฝึกสอน โดยบทความทั้งหมดเป็นบทความที่ไม่เคยถูกนำไปใช้ในระหว่างขั้นตอนการฝึกสอนก่อนหน้านี้ เผยแพร่ให้ผู้ที่สนใจสามารถดาวน์โหลดเพื่อนำไปใช้ได้ที่ <http://copycatch.in.th/thai-qa-task.html>

ตัวอย่าง

<doc id="665" url="https://th.wikipedia.org/wiki?curid=665" title="ประเทศเยอรมนี">ประเทศเยอรมนี ประเทศเยอรมนี (; ดอยชฺลันท) หรือชื่ออย่างเป็นทางการคือ สหพันธ์สาธารณรัฐเยอรมนี (;) เป็นสหพันธ์สาธารณรัฐแบบรัฐสภาในยุโรปกลาง มีรัฐองค์ประกอบ 16 รัฐ มีพื้นที่ 357,021 ตารางกิโลเมตร และมีภูมิอากาศตามฤดูกาลแบบอบอุ่นเป็นส่วนใหญ่ มีประชากรประมาณ 82 ล้านคน ซึ่งเป็นประเทศที่มีประชากรมากที่สุดในสหภาพยุโรป ประเทศเยอรมนีเป็นจุดหมายการเข้าเมืองยอดนิยมอันดับสองในโลกรองจากสหรัฐอเมริกา เมืองหลวงและมหานครใหญ่สุดของประเทศคือ กรุงเบอร์ลิน ขณะที่เขตเมืองขยายใหญ่สุดคือ รุร์ โดยมีศูนย์กลางหลักคอร์ทมุนด์และเอสเซิน นครหลักอื่นของประเทศ ได้แก่ ฮัมบวร์ค มิวนิก โคโลญ แฟรงก์เฟิร์ต ชตุทท์การ์ท ดิสเซิลดอร์ฟ ไลพ์ซิจ เบร์ลิน เดรสเดิน ฮันโนเฟอร์และเนือร์นแบร์ก ประเทศนี้มีระบอบการปกครองแบบประชาธิปไตยเชิงเสรีภาพและรัฐสวัสดิการ พรมแดนทางทิศเหนือติดทะเลเหนือ เดนมาร์ก และทะเลบอลติก ทิศตะวันออกติดโปแลนด์และเช็กเกีย ทิศใต้ติดออสเตรีย และสวิตเซอร์แลนด์ ทิศตะวันตกติดฝรั่งเศส ลักเซมเบิร์ก เบลเยียม และเนเธอร์แลนด์ มีเมืองหลวงและเมืองใหญ่ของประเทศคือเบอร์ลิน เยอรมนีมีประชากรประมาณ 80 ล้านคนและเป็นประเทศที่มีความหนาแน่นประชากรสูงสุดแห่งหนึ่ง อีกทั้งยังเป็นประเทศที่มีคนย้ายถิ่นมากที่สุดเป็นอันดับสามของโลก หลังจากที่สหรัฐอเมริกาเยอรมนีเป็นปลายทางการย้ายถิ่นที่สองได้รับความนิยมมากที่สุดในโลก เยอรมนีเป็นหนึ่งในสมาชิกผู้ก่อตั้งสหภาพยุโรปและยังก่อตั้งสหภาพการเงินกับสมาชิกในสหภาพยุโรปอีก 17 ประเทศ โดยใช้ชื่อว่ายูโรโซน เยอรมนีเป็นสมาชิกของกลุ่ม ... ชนิดของเบียร์ในเยอรมันได้แก่ Alt, Bock, Dunkel, Kölsch, เลเกอร์, Malzbier, Pils และ Weizenbier จากการสำรวจ 18 ประเทศตะวันตกที่บริโภคเครื่องดื่มเป็นต่อหัวมากที่สุด เยอรมนีอยู่ในอันดับ 14 สำหรับเครื่องดื่ม

โดยหลักของงานวิจัยนี้เป็นเรื่องการประมวลผลภาษาธรรมชาติที่จะต้องมีการจัดการข้อมูลจากภาษาธรรมชาติให้แสดงอยู่ในรูปแบบข้อมูลที่เครื่องสามารถนำไปเรียนรู้ได้ ดังนั้นในทฤษฎีที่เกี่ยวข้องจึงนำเสนอในด้านของการจัดการข้อมูลที่เป็นข้อความ เพื่อให้ผู้อ่านสามารถเข้าใจกระบวนการเรียนรู้ตั้งแต่การแปลงข้อความเข้าจนถึงกระบวนการที่ได้ผลลัพธ์ของข้อความในรูปแบบของประโยคคำถาม นอกจากนี้จึงขอนำเสนอองค์ความรู้ที่มีประสิทธิภาพสูงและได้รับความนิยมในปัจจุบันของงานด้านการประมวลผลภาษาธรรมชาติและถูกนำมาใช้เพื่อทำการวิจัยนี้

2.7. เอ็น-แกรม

เอ็น-แกรม (N-gram) คือการประมวลผลร่วมกันทางด้านภาษาและความน่าจะเป็น โดยเอ็นหมายถึงจำนวนคำวลี หรือตัวอักษรที่ต่อเนื่องกันในข้อความ ในงานด้านภาษามักนำเอ็นแกรมมาใช้ในการสกัดข้อมูลเพื่อประมวลผลสำหรับงานที่ต้องการ เช่น การประเมินประสิทธิภาพของแบบจำลอง โดยจำนวนเอ็นที่เพิ่มขึ้นจะมีชื่อเรียกและผลลัพธ์ที่แตกต่างกัน กรณีที่มีจำนวนเอ็นตั้งแต่ 3 ขึ้นไป จะใช้การอ่านเลขเป็นชื่อเรียกในแต่ละแกรมดังตารางที่ 1

ตัวอย่าง “โจรสลัด คือบุคคลที่ปล้นหรือโจรกรรมในทะเล หรือบางครั้งตามชายฝั่งหรือท่าเรือต่างๆ”

ตารางที่ 1 ชื่อเรียกและผลลัพธ์ที่แตกต่างตามจำนวนเอ็น

จำนวนเอ็น	ชื่อเรียก	ตัวอย่าง
1	Unigram (ยู-นิ-แกรม)	(‘โจรสลัด’), (‘คือ’), (‘บุคคล’), (‘ที่’), (‘ปล้น’), (‘หรือ’), (‘โจรกรรม’), (‘ใน’), (‘ทะเล’), (‘ ’), (‘หรือ’), (‘บาง’), (‘ครั้ง’), (‘ตาม’), (‘ชาย’), (‘ฝั่ง’), (‘หรือ’), (‘ท่า’), (‘เรือ’), (‘ต่างๆ’)
2	Bigram (ไบ-แกรม)	(‘โจรสลัด’, ‘คือ’), (‘คือ’, ‘บุคคล’), (‘บุคคล’, ‘ที่’), (‘ที่’, ‘ปล้น’), (‘ปล้น’, ‘หรือ’), (‘หรือ’, ‘โจรกรรม’), (‘โจรกรรม’, ‘ใน’), (‘ใน’, ‘ทะเล’), (‘ทะเล’, ‘ ’), (‘ ’, ‘หรือ’), (‘หรือ’, ‘บาง’), (‘บาง’, ‘ครั้ง’), (‘ครั้ง’, ‘ตาม’), (‘ตาม’, ‘ชาย’), (‘ชาย’, ‘ฝั่ง’), (‘ฝั่ง’, ‘หรือ’), (‘หรือ’, ‘ท่า’), (‘ท่า’, ‘เรือ’), (‘เรือ’, ‘ต่างๆ’)
3	Trigram (ไตร-แกรม)	(‘โจรสลัด’, ‘คือ’, ‘บุคคล’), (‘คือ’, ‘บุคคล’, ‘ที่’), (‘บุคคล’, ‘ที่’, ‘ปล้น’), (‘ที่’, ‘ปล้น’, ‘หรือ’), (‘ปล้น’, ‘หรือ’, ‘โจรกรรม’), (‘หรือ’, ‘โจรกรรม’, ‘ใน’), (‘โจรกรรม’, ‘ใน’, ‘ทะเล’), (‘ใน’, ‘ทะเล’, ‘ ’), (‘ทะเล’, ‘ ’, ‘หรือ’), (‘ ’, ‘หรือ’, ‘บาง’), (‘หรือ’, ‘บาง’, ‘ครั้ง’), (‘บาง’, ‘ครั้ง’, ‘ตาม’), (‘ครั้ง’, ‘ตาม’, ‘ชาย’), (‘ตาม’, ‘ชาย’, ‘ฝั่ง’), (‘ชาย’, ‘ฝั่ง’, ‘หรือ’), (‘ฝั่ง’, ‘หรือ’, ‘ท่า’), (‘หรือ’, ‘ท่า’, ‘เรือ’), (‘ท่า’, ‘เรือ’, ‘ต่างๆ’)
4	Four-gram (โฟร์-แกรม)	(‘โจรสลัด’, ‘คือ’, ‘บุคคล’, ‘ที่’), (‘คือ’, ‘บุคคล’, ‘ที่’, ‘ปล้น’), (‘บุคคล’, ‘ที่’, ‘ปล้น’, ‘หรือ’), (‘ที่’, ‘ปล้น’, ‘หรือ’, ‘โจรกรรม’), (‘ปล้น’, ‘หรือ’, ‘โจรกรรม’, ‘ใน’), (‘หรือ’, ‘โจรกรรม’, ‘ใน’, ‘ทะเล’), (‘โจรกรรม’, ‘ใน’, ‘ทะเล’, ‘ ’), (‘ใน’, ‘ทะเล’, ‘ ’, ‘หรือ’), (‘ทะเล’, ‘ ’, ‘หรือ’, ‘บาง’), (‘ ’, ‘หรือ’, ‘บาง’, ‘ครั้ง’), (‘หรือ’, ‘บาง’, ‘ครั้ง’, ‘ตาม’), (‘บาง’, ‘ครั้ง’, ‘ตาม’, ‘ชาย’), (‘ครั้ง’, ‘ตาม’, ‘ชาย’, ‘ฝั่ง’), (‘ตาม’, ‘ชาย’, ‘ฝั่ง’, ‘หรือ’), (‘ชาย’, ‘ฝั่ง’, ‘หรือ’, ‘ท่า’), (‘ฝั่ง’, ‘หรือ’, ‘ท่า’, ‘เรือ’), (‘หรือ’, ‘ท่า’, ‘เรือ’, ‘ต่างๆ’)

2.8. คะแนนเบลอ

คะแนนเบลอ (BLEU หรือ Bilingual Evaluation Understudy) คือมาตรวัดหนึ่งที่ใช้ในการประเมินผลงานด้านการสร้างข้อความโดยอัตโนมัติ เป็นเกณฑ์ที่นิยมใช้ในการประเมินประสิทธิภาพของแบบจำลองในด้านภาษา ซึ่งวัดความคล้ายกันของคำในประโยคระหว่างประโยคอ้างอิงจากมนุษย์และประโยคที่แบบจำลองสร้างขึ้นจากแหล่งข้อมูลเดียวกัน แสดงผลในรูปแบบคะแนนระหว่าง 0-1 โดยคะแนนที่เข้าใกล้ 1 คือแบบจำลองสามารถสร้างประโยคที่มีความใกล้เคียงกับประโยคที่มนุษย์สร้างขึ้น กลไกการประมวลผลเพื่อประเมินคะแนนเบลอจะใช้เอ็น-แกรมร่วมด้วย เพื่อเทียบความใกล้เคียงระหว่างประโยค ซึ่งมักจะระบุจำนวนแกรมที่ใช้ในการประเมินคะแนนเบลอ เช่น BLEU-1 หมายถึงใช้แกรมในการประเมินค่าคะแนนเบลอและ BLEU-2 หมายถึงใช้ไบแกรมในการประเมินค่าคะแนนเบลอ เป็นต้น

2.9. การตั้งคำถาม

จากทฤษฎีการตั้งคำถาม [11] สามารถจำแนกประเภทคำถามที่พัฒนาในด้านสติปัญญาออกเป็น 6 ระดับ ได้แก่ ความจำ ความเข้าใจ การประยุกต์ การวิเคราะห์ การประเมินและสร้างสรรค์ ซึ่งแบบจำลองที่นำเสนอสามารถสร้างคำถามจากบทความอยู่ในระดับความจำและความเข้าใจ โดยถามในเรื่องข้อมูลและข้อเท็จจริงจากบทความที่กำหนดให้ คำถามในกลุ่มนี้จะใช้คำถามประเภท ใคร, อะไร, อย่างไร, ที่ไหน, เมื่อไร ซึ่งเป็นคำถามที่ไม่มีคำตอบที่แน่นอน ผู้เรียนจะใช้ทักษะการอ่านภาษาไทย ความจำในเนื้อหาบทความซึ่งเป็นทักษะพื้นฐานที่ใช้ในการเรียนรู้



บทที่ 3

งานวิจัยที่เกี่ยวข้อง

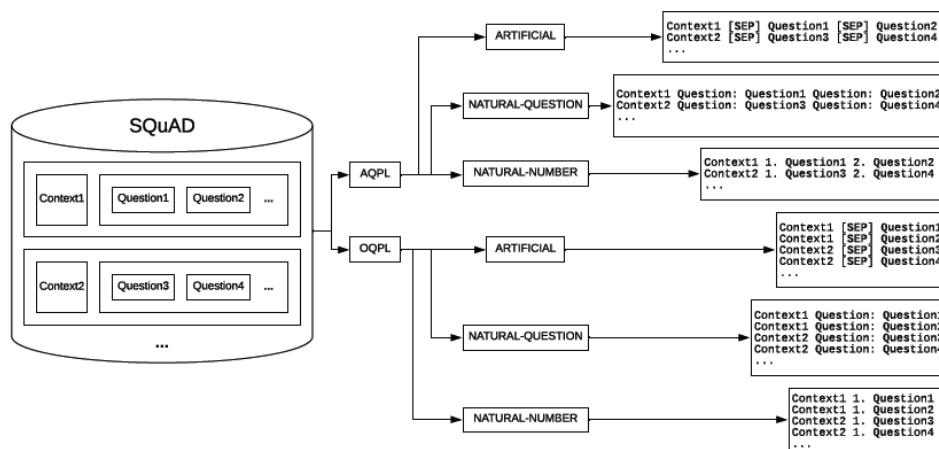
ในงานวิจัยนี้ต้องการนำเสนอการสร้างชุดคำถามอัตโนมัติโดยใช้การเรียนรู้เชิงลึก จากข้อมูลฝึกสอนจำนวนน้อย โดยมีงานวิจัยที่เกี่ยวข้องดังนี้

3.1. งานวิจัยเกี่ยวกับการสร้างชุดคำถาม

มีหลายงานวิจัยที่การแก้ปัญหาในเรื่องการสร้างชุดคำถามอัตโนมัติมาตั้งแต่ปีค.ศ. 2003 มีผู้วิจัยการสร้างคำถามอัตโนมัติโดยการใช้การแปลงกฎและฐานข้อมูลคำศัพท์เพื่อสร้างคำถามประเภทคำถามตัวเลือก [12] ต่อมาในปีคริสต์ศักราช 2010 ได้มีการวิจัยเพิ่มโดยการใช้กฎ (rule-based) ควบคู่กับการใช้แบบจำลองการถดถอยโลจิสติก (logistic regression) ใช้ในการจัดอันดับคำถาม [13] และได้เริ่มเข้าสู่ช่วงแพร่หลายของการเรียนรู้แบบโครงข่ายประสาทเทียม (neural network) จึงเกิดงานวิจัย [14, 15] ที่นำโครงข่ายประสาทเทียมแบบวนกลับ (Recurrent Neural Network หรือ RNN) หรือ หน่วยความจำระยะสั้นแบบยาว (Long Short Term Memory หรือ LSTM) มาใช้ในการสร้างชุดคำถาม หลังจากนั้นในปีค.ศ. 2018 ได้มีการเผยแพร่เรื่องกลไกแอตเทนชัน จึงเกิดงานวิจัยรูปแบบใหม่ในการสร้างคำถามอัตโนมัติ ที่นำกลไกแอตเทนชันมาใช้แทนโครงข่ายประสาทเทียมแบบวนกลับ ในรูปแบบของทรานส์ฟอร์มเมอร์ เช่น BERT [16, 17] T5 [18] GPT2 [19] เป็นต้น ในงานวิจัยที่นำเสนอเป็นการสร้างแบบจำลองที่ไม่จำเป็นต้องมีคำตอบเป็นข้อมูลนำเข้าในการฝึกสอนแบบจำลอง ซึ่งช่วยลดเวลา ความซับซ้อนและขนาดของแบบจำลองโดยมีงานวิจัยการสร้างคำถามที่ใกล้เคียงกันดังนี้

3.1.1. งานวิจัยของ Luis Enrico Lopez และคณะ [18] ในปีค.ศ. 2020 ใช้เทคนิคทรานส์ฟอร์มเมอร์ผ่านแบบจำลองที่ผ่านการฝึกสอนมาแล้วจีพีที-2 (GPT-2) ในการสร้างชุดคำถาม โดยศึกษาปัจจัยที่มีผลกระทบต่อประสิทธิภาพของแบบจำลอง แสดงรูปแบบข้อมูลนำเข้าดังรูปที่ 1 มีรายละเอียดแต่ละรูปแบบดังนี้

- การใช้ตัวค้นประโยชน์สำหรับข้อมูลนำเข้ามี 3 รูปแบบ ได้แก่
 - 1) คำประดิษฐ์ (artificial) ใช้คำพิเศษ “[SEP]” ในการแบ่งประโยคของข้อความ โดยสร้างข้อมูลนำเข้าให้อยู่ในรูปแบบ
 - 2) ภาษาธรรมชาติ (natural question) ใช้คำถามภาษาธรรมชาติ “Question:” ในการแบ่งประโยคของข้อความ โดยสร้างข้อมูลนำเข้าให้อยู่ในรูปแบบ
 - 3) ตัวเลข (natural-number) ใช้ตัวเลขจำนวนจริง ในการแบ่งประโยคของข้อความ โดยสร้างข้อมูลนำเข้าให้อยู่ในรูปแบบ
- รูปแบบของข้อมูลนำเข้าแบ่งเป็น 2 รูปแบบ ได้แก่
 - 1) รวมทุกคำถามของบทความในข้อความเดียว
 - 2) หนึ่งคำถามต่อหนึ่งข้อความ

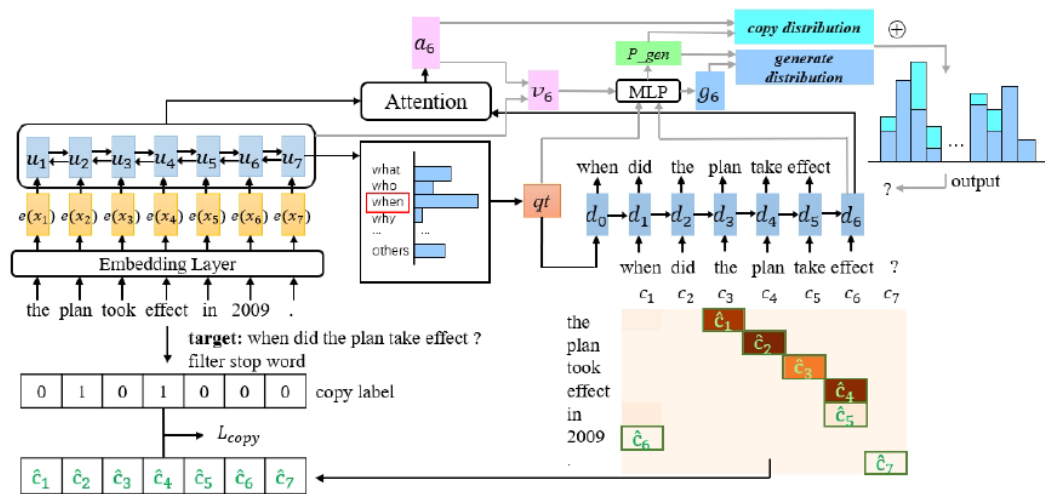


รูปที่ 1 การเตรียมข้อมูลนำเข้าในแต่ละรูปแบบ

ผลการทดลองของงานวิจัยนี้พบว่าแบบจำลองที่มีประสิทธิภาพโดยรวมที่ดีที่สุดคือการจัดรูปแบบข้อมูลนำเข้าโดยจำกัดข้อมูลนำเข้าหนึ่งคำถามต่อหนึ่งข้อความ (OQPL-Artificial) มีการเชื่อมต่อกำถามและข้อความโดยใช้คำประดิษฐ์ จากการประเมินด้วยค่า BLEU, METEOR และ ROUGE L ต่อมาได้ทำการเปรียบเทียบระหว่างประโยคคำถามที่มีคำตอบและไม่มีคำตอบโดยใช้คำประดิษฐ์ [ANSS] แทนหน้าข้อความคำตอบ และใช้คำประดิษฐ์ [ANSE] แทนท้ายข้อความคำตอบที่ปรากฏในบทความ พบว่าค่าการประเมินที่ได้แบบไม่มีคำตอบมีประสิทธิภาพดีกว่า

3.1.2. งานวิจัยของ Wu, X., N. Jiang, และ Y. Wu [20] นำเสนอแบบจำลองการสร้างคำถามอัตโนมัติที่ไม่ต้องใช้คำตอบในการสร้างคำถาม โดยเสนอ 2 วิธีที่ใช้ในการพัฒนาดังรูปที่ 2 โดยนำเสนอโลกดังนี้

- 1) ทำนายประเภทของคำถาม เป็นการทำนายประเภทของคำถามที่ควรถาม เพื่อให้แบบจำลองสร้างคำถามหลากหลายประเภทจากบทความเดียวกันและพัฒนาประสิทธิภาพการสร้างคำถาม ประเภทคำถามที่ปรากฏในชุดข้อมูลทั้งหมด 7 ประเภทได้แก่ อะไร (what), ใคร (who), อย่างไร (how), ที่ไหน (where), เมื่อไร (when), ใช่/ไม่ใช่ (yes/no) และ อื่น ๆ ทำนายประเภทคำถามโดยใช้เพอร์เซ็ปตรอนหลายชั้น เลือกจำนวนที่มีค่าความน่าจะเป็นสูงที่สุดในแต่ละประเภทคำถามตามจำนวนที่กำหนด
- 2) กลไกการคัดลอกคำสูญเสีย ในงานวิจัยนี้มีข้อสังเกตจากการสร้างคำถามของมนุษย์ พบว่าคำถามส่วนใหญ่มักนำส่วนหนึ่งของบทความหรือคำสำคัญที่ปรากฏในบทความมาใช้ในการตั้งคำถาม จึงเป็นที่มาของการพิจารณาการคัดลอกคำสูญเสียของคำเพื่อมาใช้ในการสร้างประโยค เปรียบเทียบแต่ละคำระหว่างบทความและการถอดรหัสเพื่อสร้างคำถาม



รูปที่ 2 โครงสร้างแบบจำลองจากกลไกการคัดลอกคำสูญเสียและทำนายประเภทคำถาม

งานวิจัยนี้ใช้หน่วยความจำระยะสั้นแบบยาวไปกลับ (bi-LSTM) 2 ชั้นในการสร้างแบบจำลองการเข้ารหัส และหน่วยความจำระยะสั้นแบบยาวไปกลับ 1 ชั้นในการถอดรหัสเพื่อสร้างคำถาม โดยฝึกสอนแบบจำลองเอง ผ่านกลไกทั้ง 2 ที่ได้นำเสนอข้างต้น ผลลัพธ์ของแบบจำลองคือคำถามที่มีการประเมิน 2 แบบคือประเมินโดยอัตโนมัติได้แก่ เบลู 1-4 (BLEU1-4) มีเทีย (METEOR) รูแอล (ROUGE-L) และประเมินโดยมนุษย์ซึ่งมีเกณฑ์คือความไพเราะ ความเกี่ยวข้องและสามารถหาคำตอบได้

จากงานวิจัยที่กล่าวมาข้างต้นจึงมีแนวทางในการพัฒนาแบบจำลองการสร้างคำถามโดยไม่ต้องใช้คำตอบในการสร้าง เป็นแบบจำลองที่ไม่ซับซ้อนโดยนำแบบจำลองที่ฝึกสอนมาแล้ว ปรับค่าตัวแปร (fine-tuning) และจัดรูปแบบข้อมูลการนำเข้าข้อมูลที่แตกต่างกันและนำเสนอแนวทางในการพัฒนาเพื่อเพิ่มประสิทธิภาพแบบจำลองในการสร้างคำถาม คัดกรองคำถามที่มีความใกล้เคียงกับคำถามที่มนุษย์สร้างจากการพิจารณาคำถามที่มีความใกล้เคียงกับคำสำคัญในบทความ ใช้การประเมินโดยอัตโนมัติและเกณฑ์การประเมินโดยมนุษย์ตามงานวิจัยที่ได้กล่าวไว้ข้างต้นเพื่อเป็นเกณฑ์พื้นฐานในการเปรียบเทียบผลลัพธ์ระหว่างแบบจำลอง

3.2. งานวิจัยเกี่ยวกับการสังเคราะห์ข้อมูล

มีงานวิจัยที่สังเคราะห์ข้อมูลเพื่อพัฒนางานตอบคำถาม (question answering) [21] ซึ่งจากงานวิจัยสามารถเพิ่มค่าความแม่นยำ (accuracy) จากการสังเคราะห์คำถามและคำตอบได้มากกว่าการใช้ชุดข้อมูลฝึกสอนจากมนุษย์เท่านั้น โดยไม่มีมนุษย์อยู่ในระหว่างกระบวนการมีเพียงการใช้แบบจำลองแต่ละแบบร่วมกัน นำเสนออัลกอริทึมการสังเคราะห์ข้อมูลดังรูปที่ 3 ในการฝึกสอนมีการแบ่งชุดข้อมูลออกเป็น 2 ส่วนเท่า ๆ กัน ส่วนแรกใช้ฝึกสอนการสกัดคำตอบ การสร้างคำถามและการกรองคำถาม ส่วนที่สองใช้ในการสร้างข้อมูลสังเคราะห์เพื่อพัฒนาแบบจำลองการตอบคำถาม ซึ่งกระบวนการดังกล่าวแบ่งออกเป็น 3 ขั้นตอนดังนี้

- 1) การสกัดคำตอบจากบทความ โดยใช้ทรานส์ฟอร์มเมอร์เบิร์ดในการทำนายตำแหน่งเริ่มต้นและตำแหน่งสิ้นสุดของช่วงคำเพื่อให้ได้ประโยคคำตอบ ดังสมการนี้

$$\hat{a} \sim p(a|c) \quad (3.1)$$

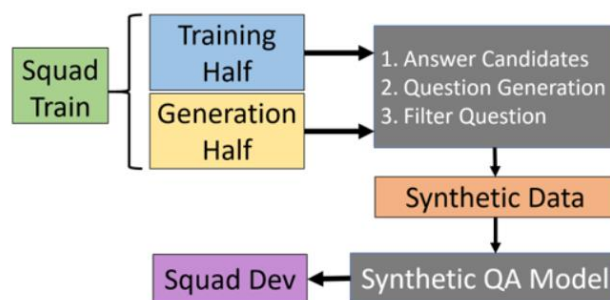
2) การสร้างคำถาม ใช้แบบจำลองที่ผ่านการฝึกสอนมาแล้วจีพีที-2 (GPT-2) โดยข้อมูลนำเข้ามีรูปแบบเป็นประโยคที่นำ บทความ คำตอบและคำถามรวมเข้าด้วยกันและคั่นแต่ละส่วนด้วยตัวจบประโยค นอกจากนี้ยังมีการเน้นส่วนที่เป็นช่วงคำตอบในบทความ ดังสมการนี้

$$\hat{q} \sim p(q|\hat{a}, c) \quad (3.2)$$

3) การกรองแบบไปและกลับ ใช้แบบจำลองการตอบคำถามที่ฝึกสอนจากชุดข้อมูลที่มีอยู่ ในการกรองคำถามและคำตอบที่สังเคราะห์ขึ้นใหม่ โดยเปรียบเทียบคำตอบจากแบบจำลองการตอบคำถามกับคำถามที่สังเคราะห์ (a^*) เทียบกับคำตอบที่สังเคราะห์ (\hat{a}) ในบทความเดียวกัน หากมีคำตอบที่ตรงกัน นับว่าคำถามที่สังเคราะห์นี้สามารถยอมรับได้

$$\hat{a} \stackrel{?}{=} a^* \sim p(a|c, \hat{q}) \quad (3.3)$$

จากผลงานวิจัยดังรูปที่ 4 พบว่าการสังเคราะห์ข้อมูลเพื่อแก้ปัญหาจำนวนข้อมูลฝึกสอนที่มนุษย์สร้างขึ้นอย่างจำกัด มาเป็นข้อมูลที่ใช้ในการฝึกสอนแบบจำลองสามารถเพิ่มประสิทธิภาพของงานตอบคำถามได้ โดยพิจารณาจากคะแนนอีเอ็ม (EM-score) และคะแนนเอฟวัน (F1-score) มีค่าเพิ่มขึ้นเมื่อเปรียบเทียบกับงานก่อนหน้า โดยการคัดกรองข้อมูลการฝึกสอนเป็นส่วนสำคัญสำหรับงานวิจัยนี้ พิจารณาได้จากการเปรียบเทียบระหว่างชุดข้อมูลสังเคราะห์ที่ไม่ผ่านขั้นตอนการกรองกับชุดข้อมูลสังเคราะห์ที่ผ่านการกรองแบบไปและกลับ (roundtrip) ดังรูปที่ 5 พบว่าการกรองแบบไปและกลับช่วยเพิ่มคะแนนอีเอ็มให้ดีขึ้นและยังมีประสิทธิภาพมากขึ้นเมื่อกำหนดให้สร้างมากกว่า 1 คำถามจากคำตอบชุดเดียวกันร่วมกับการกรองแบบไปกลับ (Overgenerate & RT) ซึ่งทำให้ได้จำนวนชุดข้อมูลฝึกสอนที่มีคุณภาพยอมรับได้มากขึ้น



รูปที่ 3 กระบวนการฝึกสอนและสังเคราะห์ข้อมูล

Model Size				# Questions	EM	F1
Answer	Question	Filter	QA			
345M	345M	345M	345M	116721	85.3	92.0
1.2B	1.2B	1.2B	345M	184992	87.1	93.2
Human Generated Data			345M	42472	86.3	93.2

รูปที่ 4 ผลการใช้ข้อมูลสังเคราะห์ในแบบจำลองการตอบคำถาม

Filter Model	# Questions	345M QA		Large QA	
		EM	F1	EM	F1
Synthetic Questions + Real Answers					
None	42472	83.4	90.9	79.0	87.0
Roundtrip (RT)	24310	84.0	91.3	76.5	84.4
Overgenerate & RT	47772	85.6	92.4	81.7	88.7
Synthetic Questions + Synthetic Answers					
None	229297	79.1	87.9	78.2	86.8
Roundtrip (RT)	93866	86.3	92.7	84.1	90.5
Overgenerate & RT	184992	87.1	93.2	85.2	91.5
Human Generated Data	42472	86.3	93.2	82.4	89.7

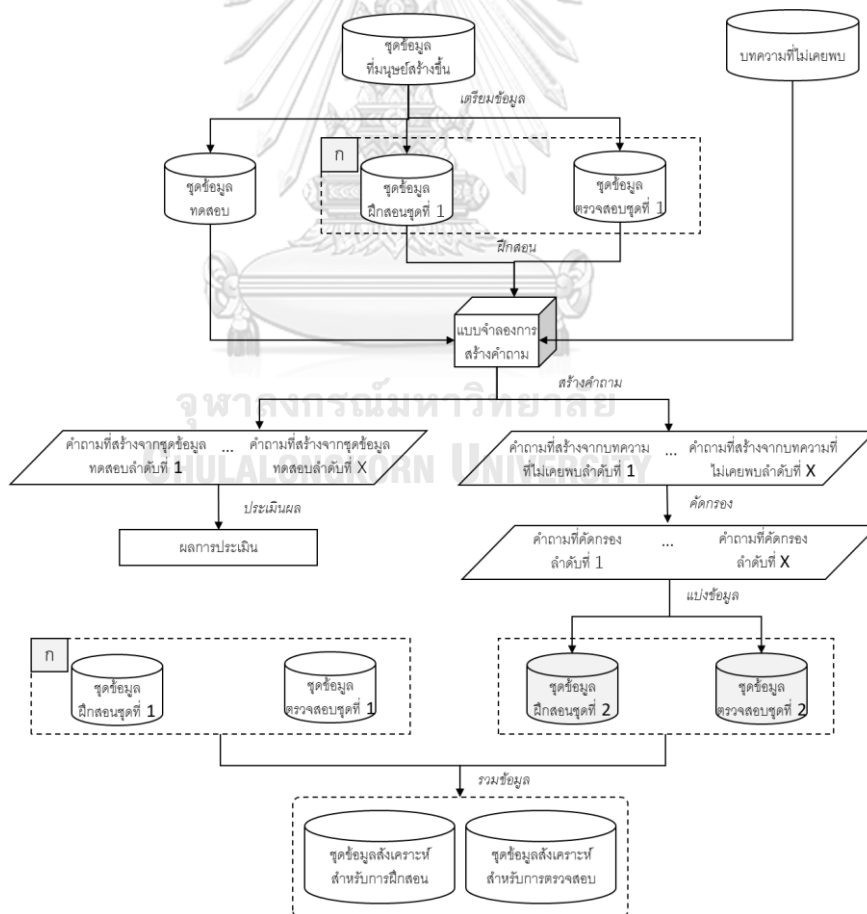
รูปที่ 5 เปรียบเทียบผลลัพธ์ของการไม่กรองข้อมูลสังเคราะห์ การกรองข้อมูลสังเคราะห์แบบไปกลับและการสร้างคำถามเพิ่มขึ้นพร้อมกับกรกรองข้อมูลแบบไปกลับ จึงเป็นแนวทางในการนำไปพัฒนาการสร้างคำถามจากการสังเคราะห์ชุดข้อมูล โดยจะใช้แบบจำลองการสร้างคำถามที่ได้จากการฝึกสอนในชุดข้อมูลที่มนุษย์สร้างขึ้น กำหนดให้สร้างชุดคำถามจากบทความมาจำนวน 5-10 ประโยค จากนั้นนำผลลัพธ์ที่ได้เข้าขั้นตอนคัดกรองประโยคคำถาม เพื่อนำไปใช้เป็นชุดข้อมูลฝึกสำหรับการสร้างคำถามในรอบต่อไป โดยมีความเป็นไปได้จากผลงานวิจัยข้างต้นว่าจะช่วยให้ผลการประเมินแบบจำลองการสร้างคำถามดีขึ้นจากการฝึกสอนจากข้อมูลสังเคราะห์

บทที่ 4

แนวคิดและวิธีการดำเนินงาน

งานวิจัยนี้เป็นการนำเสนอการพัฒนาแบบจำลองจากการสังเคราะห์ข้อมูลฝึกสอน เนื่องจากชุดข้อมูลฝึกสอนภาษาไทยที่เป็นบทความและคำถามมีจำนวนน้อยเมื่อเปรียบเทียบกับภาษาอื่น โดยนำแนวคิดจากงานวิจัยที่ได้ศึกษา มาเรื่องการสังเคราะห์ข้อมูลและการใช้แบบจำลองที่ผ่านการฝึกสอนมาแล้วในการพัฒนาคุณภาพการสร้างคำถามจากบทความภาษาไทย

จากงานวิจัยที่มีการศึกษาเรื่องการสร้างคำถาม [18, 21] โดยใช้แบบจำลองที่ผ่านการฝึกสอนการสร้างข้อความ เป็นแบบจำลองตั้งต้นและปรับค่าพารามิเตอร์ของแบบจำลองเพื่อให้เหมาะสมกับชุดข้อมูลภาษาไทย จากงานวิจัยที่กล่าวถึงในข้างต้นเป็นการใช้แบบจำลองที่ไฟฟ์ (T5) และจีพีที-2 (GPT-2) ตามลำดับซึ่งไม่รองรับการสร้างข้อความในภาษาไทย งานวิจัยนี้จึงใช้แบบจำลองที่ผ่านการฝึกสอนมาแล้วเอ็มทีไฟฟ์ (mT5) โดยมีโครงสร้างของแบบจำลองมาจากทีไฟฟ์ที่ผ่านการฝึกสอนภาษาไทยมาแล้ว เพื่อใช้ในการสร้างคำถามจากบทความดังรูปที่ 6 ซึ่งมีรายละเอียดในแต่ละขั้นตอนดังนี้



รูปที่ 6 แผนภาพการสร้างแบบจำลองการสร้างคำถาม

4.1. แบบจำลองการสร้างคำถาม

1) การเตรียมข้อมูล

ชุดข้อมูลที่ใช้ในงานวิจัยนี้แบ่งเป็น 2 ส่วนดังนี้

ก. ชุดข้อมูลที่ประกอบด้วยบทความและคำถาม-คำตอบภาษาไทยที่สร้างโดยนักภาษาศาสตร์เพื่อใช้ในการพัฒนางานด้านการสร้างคำถามและการตอบคำถามจาก 3 แหล่งข้อมูลได้แก่

- สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ (สวทช.) ในโครงการแข่งขันพัฒนาโปรแกรมคอมพิวเตอร์แห่งประเทศไทย (NSC)
- บริษัทไอแอฟเทคโนโลยีจำกัด
- การแปลชุดคำถามของสแตนฟอร์ดเวอร์ชัน 1.1

ข. ชุดข้อมูลบทความที่ไม่เคยพบ ในเรื่องทั่วไปจากวิกิพีเดียภาษาไทย ซึ่งมาจาก สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ (สวทช.) ในโครงการแข่งขันพัฒนาโปรแกรมคอมพิวเตอร์แห่งประเทศไทย

ข้อมูลจากแหล่งข้อมูลข้างต้นมีโครงสร้างการจัดเก็บข้อมูลที่แตกต่างกัน จึงผ่านกระบวนการทำความสะอาดข้อมูลโดยลบข้อมูลที่ไม่จำเป็น เช่น ตัวอักษรพิเศษ เอชทีเอ็มแอลแท็ก จากนั้นจัดรวมข้อมูลให้อยู่ในรูปแบบเดียวกัน จากงานวิจัยเรื่องของการแปลงข้อมูลนำเข้ามีผลต่อความสามารถแบบจำลองการสร้างคำถาม รวมถึงงานวิจัยเรื่องการทำนิพจน์เฉพาะหรือชื่อเฉพาะมาใช้ในการตอบคำถาม (Question Answering) ทำให้ได้ผลลัพธ์ที่ดีขึ้น แต่ยังไม่มีการทดลองกับงานด้านการสร้างคำถาม ในงานวิจัยนี้จึงจัดรูปแบบการนำเข้าข้อมูลเพื่อทดลองประสิทธิภาพแบบจำลองจากการฝึกสอนในชุดข้อมูลฝึกสอนที่แตกต่างกันโดยเสนอรูปแบบนำเข้าข้อมูล 2 รูปแบบดังนี้

- รูปแบบนำเข้าปกติ เป็นรูปแบบนำเข้าที่ไม่มีขั้นตอนใดเพิ่มเติม ตัวอย่างบทความและคำถาม
- รูปแบบนำเข้า NER เป็นรูปแบบนำเข้าที่เพิ่มป้ายการสกัดนิพจน์เฉพาะหรือชื่อเฉพาะ (Named Entity tag) ในประโยคเข้าไปในบทความนำเข้า เพื่อเน้นข้อมูลที่สำคัญในประโยคโดยใช้คลังโปรแกรม (library) ของ PyThaiNLP เนื่องจากเป็นที่รู้จักอย่างแพร่หลายในการทำงานด้านภาษาที่เป็นภาษาไทย มีเอกสารแนะนำการใช้งานจึงนำมาใช้สำหรับการทดลองในงานวิจัยนี้ โดยมีป้ายกำกับดังตารางที่ 2

ตารางที่ 2 ป้ายการสกัดนิพจน์เฉพาะหรือชื่อเฉพาะที่ใช้ใน PyThaiNLP จาก <https://pythainlp.github.io/dev-docs/api/tag.html>

ป้ายการสกัดนิพจน์เฉพาะหรือชื่อเฉพาะ	ตัวอย่าง
DATE	2/21/2004, 16 ก.พ., จันทร์
TIME	16.30 น., 5 วัน, 1-3 ปี
EMAIL	info@nrpsc.ac.th
LEN	30 กิโลเมตร, 5 กม.

LOCATION	ไทย, จ.ปราจีนบุรี, กำแพงเพชร
ORGANIZATION	กรมวิทยาศาสตร์การแพทย์, อย.
PERSON	น.พ.จรัล, นางประนอม ทองจันทร์
PHONE	1200, 0 2670 8888
URL	http://www.bangkokhealth.com/
ZIP	10400, 11130
Money	2.7 ล้านบาท, 2,000 บาท
LAW	พ.ร.บ.โรคระบาด พ.ศ.2499, รัฐธรรมนูญ

แบ่งข้อมูลทั้ง 2 รูปแบบตามประเภทชุดข้อมูลดังนี้

- ชุดข้อมูล ก. สำหรับการฝึกสอนแบบจำลองในการสร้างคำถาม แบ่งเป็นชุดฝึกสอนและชุดตรวจสอบ และชุดข้อมูลทดสอบด้วยอัตราส่วนที่ 80:16:4 ตามลำดับ
- ชุดข้อมูล ข. สำหรับสร้างคำถามเพื่อใช้ในการสังเคราะห์ชุดข้อมูล แบ่งเป็นชุดละ 10,000 บทความ

2) การฝึกสอน

เป็นการใช้แบบจำลองที่ผ่านการฝึกสอนมาแล้วของเอ็มทีไฟฟ์ โดยกำหนดรูปแบบข้อมูลนำเข้าคือบทความ และข้อมูลนำออกคือประโยคคำถาม ทำการตัดคำและเข้ารหัส โดยเลือกใช้ตัวตัดคำที่ผ่านการฝึกสอนมาแล้วของเอ็มทีไฟฟ์ เนื่องจากแบบจำลองที่ฝึกสอนมาแล้วที่นำมาใช้ในการสร้างแบบจำลองการสร้างคำถามคือเอ็มทีไฟฟ์ การใช้แบบจำลองที่ฝึกสอนมาแล้วเหมือนกันจะช่วยให้ประสิทธิภาพดีกว่าการนำตัวตัดคำอื่นมาใช้ ในขั้นการฝึกสอนแต่ละรอบมีการกำหนดค่าพารามิเตอร์ต่าง ๆ เช่น ความเร็วในการเรียนรู้ (learning rate) จำนวนรอบการฝึกสอน (epoch) ขนาดแบตช์ (batch size) ซึ่งมีผลต่อการเรียนรู้ของแบบจำลองในการฝึกสอน ในแต่ละรอบที่ฝึกสอนจะมีการเก็บสถิติไว้และบันทึกแบบจำลองรอบที่มีผลดีที่สุดไว้ โดยพิจารณาจากค่าสูญเสีย (loss) ที่มีค่าน้อยที่สุดในชุดข้อมูลตรวจสอบ

3) การสร้างคำถาม

นำแบบจำลองที่ได้จากขั้นตอนที่ 3 มาสร้างคำถามจากชุดข้อมูลทดสอบและชุดข้อมูลที่ไม่เคยพบ โดยมีข้อมูลนำเข้าคือบทความ ผ่านแบบจำลองและถอดรหัสในขั้นตอนนี้จะมีการปรับค่าพารามิเตอร์ซึ่งมีผลต่อการสร้างประโยคคำถามเพื่อให้ได้คำถามที่คาดว่าจะมีประสิทธิภาพมากที่สุดจากการประเมินโดยอัตโนมัติและค่าที่จะระบุพารามิเตอร์ที่จะทำให้ได้ผลลัพธ์เดิมทุกครั้งที่มีการสร้างคำถามจากแบบจำลองการสร้างคำถาม โดยอัลกอริทึมที่เลือกใช้คือการสุ่มตัวอย่าง k และ p ตัวแรกได้ผลลัพธ์ดังนี้

- คำถามจากชุดข้อมูลทดสอบ คือ มีข้อมูลนำเข้าเป็นชุดข้อมูลทดสอบ ใช้ในการประเมินประสิทธิภาพของแบบจำลองการสร้างคำถาม
- คำถามจากบทความที่ไม่เคยพบ คือ มีข้อมูลนำเข้าเป็นบทความที่ไม่เคยถูกนำไปใช้ในระหว่างขั้นตอนการฝึกสอนและการตรวจสอบ ใช้ในการสังเคราะห์ชุดข้อมูลฝึกสอนเพื่อพัฒนาประสิทธิภาพของแบบจำลองการสร้างคำถาม

4) การคัดกรองคำถาม

กระบวนการคัดกรองคำถามเพื่อนำมาใช้ในการฝึกสอนแบบจำลองให้มีประสิทธิภาพมากขึ้น โดยกระบวนการคัดกรองมี 4 ขั้นตอนดังนี้

4.1) ลบคำถามที่เหมือนกันจากบทความเดียวกัน

ในการสร้างคำถามจากขั้นตอนที่ 4 มีการกำหนดจำนวนคำถามที่ต้องการตอบบทความ ซึ่งกลไกการสร้างคำถามจากความน่าจะเป็นมีโอกาสที่ทำให้ได้คำถามที่เหมือนกันทั้งหมดหรือแตกต่างกันเพียงช่องว่าง ซึ่งส่งผลต่อการนำคำถามไปใช้ฝึกสอน จึงมีแนวทางในการลบคำถามที่เหมือนกันจากบทความเดียวกันตามขั้นตอนดังนี้

4.1.1) สร้างชุดข้อมูล 2 ชุดเพื่อเก็บคำถามต้นแบบและคำถามที่ลบช่องว่าง

4.1.2) ลบช่องว่างที่ปรากฏในคำถามที่พิจารณาออก

4.1.3) ตรวจสอบคำถามจากกระบวนการที่ 2 กับชุดข้อมูลที่เก็บคำถามที่ลบช่องว่าง กรณีที่คำถามที่พิจารณาปรากฏในชุดข้อมูลที่เก็บคำถามที่ลบช่องว่างจะพิจารณาคำถามต่อไป แต่หากไม่พบคำถามดังกล่าวในชุดข้อมูลที่เก็บคำถามที่ลบช่องว่าง จะเก็บคำถามต้นแบบในชุดข้อมูลที่เก็บคำถามต้นแบบ

4.2) ลบคำถามที่มีตัวเลขที่ไม่ปรากฏในบทความ

ผลการสร้างคำถามที่มีจำนวนตัวเลข แบบจำลองมักสร้างคำถามโดยขาดความแม่นยำของตัวเลขที่ไม่ปรากฏในบทความ ทำให้ได้คำถามที่ไม่มีประสิทธิภาพตัวอย่างเช่น บทความกล่าวถึงวิกฤตต้มยำกุ้งในปีพ.ศ. 2540 แต่คำถามที่สร้างระบุว่าวิกฤตต้มยำกุ้งในปีพ.ศ. 2546 เพื่อหลีกเลี่ยงคำถามที่ไม่มีคุณภาพซึ่งไม่ตรงตามเท็จจริง จึงมีกระบวนการตรวจสอบตัวเลขที่ปรากฏในคำถามจะต้องเป็นตัวเลขที่ปรากฏในบทความโดยมีขั้นตอนดังนี้

4.2.1) สร้างชุดเก็บข้อมูลตัวเลขที่ปรากฏในบทความ

4.2.2) สกัดตัวเลขที่ปรากฏในบทความลงในชุดจากข้อ 1)

4.2.3) สกัดตัวเลขที่ปรากฏในคำถาม

4.2.4) พิจารณาค่าตัวเลขที่ปรากฏในคำถามเพื่อตรวจสอบว่ามีค่าอยู่ในชุดเก็บข้อมูลตัวเลขที่ปรากฏในบทความหรือไม่ กรณีที่พบจะทำตัวถัดไปจนจบ หากไม่พบจะถือว่าคำถามนั้นไม่ผ่านเงื่อนไขนี้

4.3) คัดกรองคำถามที่มีรูปแบบประโยคคำถามตามที่กำหนด

ในขั้นตอนนี้ใช้แนวทางการกำหนดเงื่อนไข หรือกฎให้ครอบคลุมรูปแบบประโยคคำถามภาษาไทย โดยใช้นิพจน์ปรกติ (Regular expression) พิจารณาประโยคที่ปรากฏนิพจน์ของคำถาม โดยการใช้กฎในการคัดกรองช่วยคัดกรองประโยคคำถามได้อย่างมีประสิทธิภาพ

4.4) คัดกรองคำถามที่มีความเกี่ยวข้องกับบทความ

การสร้างคำถามของแบบจำลองมักจะมีเกี่ยวข้องกับบทความ จึงนำเรื่องความน่าจะเป็น เพื่อช่วยในการจำแนกคำถามที่จะได้รับการยอมรับคือมีความเกี่ยวข้องกับบทความ โดยใช้เอ็น-แกรมและ สูตรคำนวณหาความน่าจะเป็น (1) หากคำถามที่มีค่าความน่าจะเป็นผ่านเกณฑ์ที่กำหนดไว้คือมากกว่าหรือ เท่ากับ 0.75 จะผ่านการคัดกรอง โดยมีขั้นตอนดังนี้

- 4.4.1) สร้างชุดข้อมูลเอ็น-แกรมเพื่อเก็บจำนวนข้อมูลเอ็นแกรมของบทความและคำถาม
- 4.4.2) นับจำนวนข้อมูลของค่าที่ตรงกันระหว่างบทความและคำถาม
- 4.4.3) คำนวณสัดส่วนของจำนวนตามสูตร (1)
- 4.4.4) พิจารณาเลือกคำถามที่ผ่านเกณฑ์ที่กำหนด

$$\text{ความน่าจะเป็น} = \frac{\text{จำนวนแกรมที่ตรงกันระหว่าง} \times \text{ประโยคอ้างอิงและประโยคที่ระบบสร้าง}}{\text{จำนวนคำทั้งหมดที่ระบบสร้าง}} \quad (4.1)$$

5) การสังเคราะห์ข้อมูล

นำข้อมูลที่ได้จากการคัดกรองแบ่งออกเป็น 2 ชุดคือข้อมูลฝึกสอนและข้อมูลตรวจสอบในอัตราส่วน 80 ต่อ 20 ตามลำดับ รวมข้อมูลที่ใช้ฝึกสอนกับข้อมูลผลลัพธ์ที่ได้จากการคัดกรองคำถามตามประเภทชุดข้อมูล บันทึกเป็นชุดข้อมูลสังเคราะห์เพื่อใช้ในการฝึกสอน

6) การประเมินผลการทดลอง

ประเมินประสิทธิภาพของแบบจำลองจากประโยคคำถามที่สร้างจากชุดข้อมูลทดสอบโดยสุ่มเลือก 100 บทความและคู่คำถาม-คำตอบ โดยแบ่งการประเมินออกเป็น 2 รูปแบบดังนี้

- 6.1) ประเมินโดยอัตโนมัติ ในงานวิจัยนี้ใช้คะแนนเบลอในการประเมินความใกล้เคียงกันระหว่างคำถามที่มนุษย์สร้างและคำถามที่แบบจำลองสร้าง โดยใช้จำนวนเอ็น-แกรมตั้งแต่ 1-4
- 6.2) ประเมินโดยมนุษย์ กำหนดให้มีผู้ประเมินจำนวน 3 ท่าน โดยจะต้องเป็นผู้มีความรู้ทางด้าน การอ่าน จับใจความโดยมีคุณสมบัติคือ จบการศึกษาระดับปริญญาตรีในคณะอักษรศาสตร์ มนุษยศาสตร์ ครุศาสตร์ สังคมศาสตร์สาขาภาษาไทย ในแต่ละหัวข้อข้อมูลประเมินจะต้องให้คะแนน 0 หมายถึง ประโยคไม่ผ่านเกณฑ์และไม่สามารถยอมรับได้ หรือ 1 หมายถึงประโยคผ่านเกณฑ์และยอมรับได้ กำหนดเกณฑ์การประเมินดังนี้
 - ความไพเราะ คือ ประโยคใช้คำถามถูกต้องตามหลักไวยากรณ์ มีความเป็นภาษาธรรมชาติที่มนุษย์อ่านแล้วเข้าใจได้
 - มีความหมายและบริบทที่เกี่ยวข้องเหมาะสมกับบทความ คือ ประโยคมีความหมายและบริบทที่เกี่ยวข้องกับบทความ บริบทของคำมีความเป็นธรรมชาติของภาษาและเหมาะสม
 - สามารถหาคำตอบได้จากในบทความ คือ สามารถตอบคำถามได้จากข้อมูลของบทความที่กำหนด (พร้อมระบุคำตอบถ้ามี)

4.2. แบบจำลองการสร้างคำถามจากข้อมูลสังเคราะห์

จากงานวิจัยเกี่ยวกับการสังเคราะห์ข้อมูลเพื่อพัฒนาประสิทธิภาพแบบจำลองการตอบคำถาม [21] มีการคัดกรองข้อมูลที่ยอมรับได้โดยใช้แบบจำลองการตอบคำถามคัดกรองแบบไปและกลับโดยพิจารณาจากคำตอบ แต่ในงานวิจัยนี้เราพิจารณาถึงการสร้างคำถามที่มีความเกี่ยวข้องกับบทความ ดังนั้นจึงมีแนวทางที่จะสร้างแบบจำลองใหม่จากงานวิจัยเกี่ยวกับการจำแนกข้อความ [1, 22] เพื่อคัดกรองคำถามที่ดีจากสมมติฐานว่าประโยคคำถามที่ได้จากการสร้างคำถามอัตโนมัติที่ดีจะต้องมีรูปแบบประโยคที่ใกล้เคียงกับประโยคคำถามที่มนุษย์สร้างขึ้นและต้องมีความเกี่ยวข้องกับบทความเพื่อนำไปใช้เป็นชุดข้อมูลฝึกสอน โดยใช้ผลลัพธ์จากการรวมข้อมูลฝึกสอนและข้อมูลตรวจสอบในขั้นตอนการสังเคราะห์ข้อมูลเป็นข้อมูลฝึกสอนแบบจำลองใหม่ ได้เป็นแบบจำลองการสร้างคำถามจากข้อมูลสังเคราะห์

แบบจำลองการสร้างคำถามจากข้อมูลสังเคราะห์จะถูกนำไปสร้างคำถามจากชุดข้อมูลทดสอบและชุดข้อมูลบทความที่ไม่เคยพบเพื่อประเมินประสิทธิภาพของแบบจำลองนั้น ๆ และเพื่อสร้างชุดข้อมูลสังเคราะห์ใหม่



บทที่ 5

การทดลองและผลการทดลอง

การทดลองนี้แบ่งออกเป็น 6 ขั้นตอนหลักได้แก่ การเตรียมข้อมูล การนำเข้าข้อมูลเพื่อฝึกสอน การสร้างคำถาม จากแบบจำลองการสร้างคำถาม การคัดกรองคำถาม การคัดกรองคำถาม การสังเคราะห์ข้อมูลและการประเมินผล การทดลอง โดยรายละเอียดในแต่ละขั้นตอนมีดังต่อไปนี้

5.1. การเตรียมข้อมูล

5.1.1. ทำความสะอาดข้อมูล โดยการลบข้อมูลที่ไม่จำเป็นในบทความ คำถามและคำตอบ เช่น ตัวอักษรพิเศษ เอชทีเอ็มแอลแท็กตัวอย่างดังรูปที่ 7 และ รูปที่ 8

```
<doc id="115035" url="https://th.wikipedia.org/wiki?curid=115035" title="เบนจี้">เบนจี้ เบนจี้ () เป็นชื่อตัวละครหมาพันธุ์ขนสั้นที่ปรากฏอยู่ในภาพยนตร์หลายเรื่องที่เขียนบท และกำกับโดย โจ แคมป์ ในช่วงทศวรรษ 1970 ถึง 1980 ภาพยนตร์เรื่องแรกในชุด ใช้ชื่อเรื่องว่า เบนจี้ เช่นเดียวกับตัวละคร ถ่ายทำที่เมืองดัลลัส รัฐเท็กซัส ฉายครั้งแรกในปี พ.ศ. 2517 ภาพยนตร์ได้รับการเสนอชื่อเข้าชิงรางวัลออสการ์ และได้รางวัลลูกโลกทองคำ สาขาเพลงประกอบยอดเยี่ยม จากเพลง Benji's Theme (I Feel Love) ร้องโดย ชาร์ลี ริช หมาที่แสดงเป็นเบนจี้ตัวแรก ชื่อว่า ฮิกกินส์ (พ.ศ. 2502 - พ.ศ. 2518) มีอายุถึง 15 ปี แล้วในขณะแสดง หลังจากภาพยนตร์ออกฉายได้ไม่นาน มันก็ตายในปี พ.ศ. 2518 เบนจี้ในภาพยนตร์เบนจี้ ในภาพยนตร์. - พ.ศ. 2517, Benji (ภาพยนตร์) - พ.ศ. 2520, For the Love of Benji (ภาพยนตร์) - พ.ศ. 2521, Benji's Very Own Christmas Story (ภาพยนตร์โทรทัศน์) - พ.ศ. 2523, Oh Heavenly Dog (ภาพยนตร์) - พ.ศ. 2523, Benji at Work (ภาพยนตร์โทรทัศน์) - พ.ศ. 2524, Benji Takes a Dive at Marineland (ภาพยนตร์โทรทัศน์) - พ.ศ. 2526, Benji, Zax & the Alien Prince (ภาพยนตร์ซีรีส์) - พ.ศ. 2530, Benji the Hunted (ภาพยนตร์) - พ.ศ. 2547, Benji: Off the Leash! (ภาพยนตร์) - พ.ศ. 2550, Benji: The Barkening (ภาพยนตร์)</doc>
```

รูปที่ 7 ตัวอย่างข้อมูลบทความภาษาไทยจาก NSC ก่อนทำความสะอาดข้อมูล

เบนจี้ เบนจี้ เป็นชื่อตัวละครหมาพันธุ์พันธุ์ที่ปรากฏอยู่ในภาพยนตร์หลายเรื่องที่เกี่ยวข้องกับ โดย โจ แคมป์ ในช่วงทศวรรษ 1970 ถึง 1980 ภาพยนตร์เรื่องแรกในชุด ใช้ชื่อเรื่องว่า เบนจี้ เช่นเดียวกับ ตัวละคร ถ่ายทำที่เมืองดัลลัส รัฐเท็กซัส ฉายครั้งแรกในปี พ.ศ. 2517 ภาพยนตร์ได้รับการเสนอชื่อเข้าชิง รางวัลออสการ์ และได้รางวัลลูกโลกทองคำ สาขาเพลงประกอบยอดเยี่ยม จากเพลง Benji's Theme ร้องโดย ชาร์ลี ริช หมาที่แสดงเป็นเบนจี้ตัวแรก ชื่อว่า อิกกินส์ มีอายุถึง 15 ปีแล้วในขณะแสดง ภาพยนตร์ออกฉายได้ไม่นาน มันก็ตายในปี พ.ศ. 2518 เบนจี้ในภาพยนตร์เบนจี้ในภาพยนตร์. พ.ศ. 2517 Benji พ.ศ. 2520 For the Love of Benji พ.ศ. 2521 Benji's Very Own Christmas Story พ.ศ. 2523 Oh Heavenly Dog พ.ศ. 2523 Benji at Work พ.ศ. 2524 Benji Takes a Dive at Marineland พ.ศ. 2526 Benji Zax the Alien Prince พ.ศ. 2530 Benji the Hunted พ.ศ. 2547 Benji Off the Leash พ.ศ. 2550 Benji The Barkening

รูปที่ 8 ตัวอย่างข้อมูลบทความภาษาไทยจาก NSC หลังทำความสะอาดข้อมูล

5.1.2. จัดรูปแบบการนำเข้าข้อมูล ในการทดลองนี้จัดรูปแบบการนำเข้าข้อมูลออกเป็น 2 รูปแบบคือ รูปแบบนำเข้าปกติและรูปแบบนำเข้า NER โดยมีตัวอย่างข้อมูลนำเข้าดังตารางที่ 3

- รูปแบบนำเข้าปกติ บทความและคำถามที่จากต้นฉบับที่ผ่านการทำความสะอาดและไม่มีขั้นตอนใดเพิ่มเติม
- รูปแบบนำเข้า NER แปลงบทความจากต้นฉบับโดยเพิ่มป้ายกำกับการสัคตินิพจน์เฉพาะหรือชื่อเฉพาะในประโยคเข้าไปในบทความนำเข้าโดยคลังโปรแกรม (library) PyThaiNLP ตัวอย่างการเพิ่มป้ายกำกับการสัคตินิพจน์เฉพาะหรือชื่อเฉพาะดังรูปที่ 9

การรุกรานแบกแดด ชามาร์คินด์ เออร์เกนซ์ เคียฟ วลาดีเมียร์และอีกหลายแห่งก่อให้เกิด การสังหารหมู่ อย่างเช่นที่หลายส่วนของชุมชนสถานใต้ได้ถูกทำลายลงจนสิ้นเชิง ฮูลากู ข่าน ทายาทของเขาทำลายส่วนใหญ่ของอิหร่านทางตอนเหนือ แล้วปล้นสะดมแบกแดด...

NER Tagging

<PERSON>การรุกรานแบกแดด ชามาร์คินด์ เออร์เกนซ์ เคียฟ วลาดีเมียร์</PERSON> และอีกหลายแห่งก่อให้เกิดการสังหารหมู่ อย่างเช่นที่หลายส่วนของ<LOCATION>ชุมชนสถานใต้</LOCATION>ได้ถูกทำลายลงจนสิ้นเชิง <PERSON>ฮูลากู ข่าน</PERSON> ทายาทของเขาทำลายส่วนใหญ่ของอิหร่านทางตอนเหนือ แล้วปล้นสะดมแบกแดด...

รูปที่ 9 ตัวอย่างการแปลงข้อมูลนำเข้าปกติเป็นข้อมูลนำเข้า NER

ตารางที่ 3 ตัวอย่างบทความรูปแบบนำเข้าปกติและรูปแบบนำเข้า NER

บทความปกติ	บทความ NER
<p>โยโกะ โอะโนะ โยโกะ โอะโนะ เป็นนักร้องชาวญี่ปุ่น และเป็นภรรยาคนที่สองของจอห์น เลนนอน นักร้องนำของวงเดอะบีเทิลส์ โอะโนะเกิดที่กรุงโตเกียว ในปี พ.ศ. 2476 ได้แต่งงานครั้งแรกกับ โทะชิ อิชियะนะงิ พ.ศ. 2499 ครั้งที่สองกับแอนโทนี ค็อกซ์ พ.ศ. 2499 โดยมีบุตรสาวคนแรกชื่อ เคียวโกะ ชาน คอกซ์ และแต่งงานครั้งที่สามกับนักร้องชาวอังกฤษ จอห์น เลนนอน เมื่อวันที่ 20 มีนาคม พ.ศ. 2512 โดยมีลูกชายชื่อ ฌอน เลนนอน หรือ ฌอน ทะโร โอะโนะ เลนนอน เกิดวันเดียวกับบิดาของเขาในวันที่จอห์น เลนนอนอายุครบ 35 ปี</p>	<p><PERSON>โยโกะ โอะโนะ โยโกะ โอะโนะ </PERSON> เป็นนักร้องชาว<LOCATION>ญี่ปุ่น </LOCATION> และเป็นภรรยาคนที่สองของจอห์น เลนนอน นักร้องนำของวง<ORGANIZATION>เดอะบีเทิลส์</ORGANIZATION> โอะโนะเกิดที่<LOCATION>กรุงโตเกียว</LOCATION> ในปี <DATE>พ.ศ. 2476</DATE> ได้แต่งงานครั้งแรกกับ <PERSON>โทะชิ อิชियะนะงิ</PERSON> พ.ศ. 2499 ครั้งที่สองกับ<PERSON>แอนโทนี ค็อกซ์</PERSON> พ.ศ. 2499 โดยมีบุตรสาวคนแรกชื่อ <PERSON>เคียวโกะ ชาน คอกซ์</PERSON> และแต่งงานครั้งที่สามกับนักร้องชาว<LOCATION>อังกฤษ</LOCATION> จอห์น เลนนอน เมื่อวันที่ <DATE>20 มีนาคม พ.ศ. 2512</DATE> โดยมีลูกชายชื่อ ฌอน เลนนอน หรือ <PERSON>ฌอน ทะโร โอะโนะ เลนนอน</PERSON> เกิดวันเดียวกับบิดาของเขาในวันที่จอห์น เลนนอนอายุครบ <TIME>35 ปี</TIME></p>
<p>นิยาม 5 นิยาม 5 หมายถึง กำหนดอันแน่นอน ความ เป็นไปอันมีระเบียบแน่นอนของธรรมชาติ กฎธรรมชาติ อันครอบคลุมทุกสิ่งทุกอย่าง ทุกสิ่งทุกอย่างดำเนินไป ตามกฎทั้ง 5 ประการนี้ ได้แก่ 1. อุดุนิยาม คือ กฎธรรมชาติเกี่ยวกับปรากฏการณ์ธรรมชาติ เช่น อุณหภูมิ ดินฟ้าอากาศ สิ่งแวดล้อม 2. พีชนิยาม คือ กฎธรรมชาติที่เกี่ยวกับพันธุกรรม กระบวนการถ่ายทอด ข้อมูลของสิ่งมีชีวิตทั้งพืช สัตว์ เชื้อโรค ผ่านการสืบพันธุ์ 3. จิตนิยาม คือ กฎธรรมชาติเกี่ยวกับการทำงานของจิต เจตสิก 4. กรรมนิยาม คือ กฎแห่งกรรม คือกฎแห่งการกระทำและผลของการกระทำ 5. ธรรมนิยาม อันได้แก่ กฎไตรลักษณ์ คือ อนิจจัง ทุกขัง อนัตตา</p>	<p>นิยาม <TIME>5 นิยาม</TIME> 5 <PERSON> หมายถึง กำหนดอันแน่นอน ความ เป็นไปอันมีระเบียบแน่นอนของธรรมชาติ กฎธรรมชาติ อันครอบคลุมทุกสิ่งทุกอย่าง ทุกสิ่งทุกอย่างดำเนิน</PERSON>ไปตามกฎทั้ง 5 ประการนี้ ได้แก่<MONEY>1.</MONEY> อุดุนิยาม คือ <PERSON>กฎธรรมชาติเกี่ยวกับปรากฏการณ์ธรรมชาติ เช่น อุณหภูมิ ดินฟ้าอากาศ สิ่งแวดล้อม</PERSON> <TIME>2.</TIME> พีชนิยาม คือ กฎธรรมชาติที่เกี่ยวกับพันธุกรรม กระบวนการถ่ายทอดข้อมูลของสิ่งมีชีวิตทั้งพืช สัตว์ เชื้อโรค ผ่านการสืบพันธุ์ <TIME>3.</TIME> จิตนิยาม คือ กฎธรรมชาติเกี่ยวกับการทำงานของจิต เจตสิก <TIME>4.</TIME> กรรมนิยาม คือ กฎแห่งกรรม คือ <ORGANIZATION>กฎแห่งการกระทำและผลของการกระทำ 5.</ORGANIZATION> ธรรมนิยาม อันได้แก่ กฎไตรลักษณ์ คือ อนิจจัง ทุกขัง อนัตตา</p>

ต้นแกะแห่งทาทารี ต้นแกะแห่งทาทารี คือต้นไม้มิในตำนานจากเอเชียกลางที่เชื่อกันว่ามีผลเป็นแกะ ตัวแกะเชื่อมติดกับต้นไม้มด้วยสายสะดือและเล็มหญ้ารอบต้น เมื่อต้นไม้มหายไ้หมดทั้งต้นไม้มและแกะก็จะตาย แต่ในความเป็นจริงแล้วต้นไม้มชนิดนี้คือว่านลูกไก่ทอง ซึ่งเป็นเฟิร์นชนิดหนึ่ง	<ORGANIZATION>ต้นแกะแห่งทาทารี ต้นแกะแห่งทาทารี</ORGANIZATION> คือต้นไม้มิในตำนานจากเอเชียกลาง</LOCATION>ที่เชื่อกันว่ามีผลเป็นแกะ ตัวแกะเชื่อมติดกับต้นไม้มด้วยสายสะดือและเล็มหญ้ารอบต้น เมื่อต้นไม้มหายไ้หมดทั้งต้นไม้มและแกะก็จะตาย แต่ในความเป็นจริงแล้วต้นไม้มชนิดนี้คือว่านลูกไก่ทอง</PERSON> ซึ่งเป็นเฟิร์นชนิดหนึ่ง
--	---

5.1.3. แบ่งข้อมูล แบ่งข้อมูลแต่ละรูปแบบที่จะใช้ในการฝึกออกเป็นส่วนคือชุดข้อมูลฝึกสอน ชุดข้อมูลตรวจสอบและชุดข้อมูลทดสอบด้วยอัตราส่วน 80:16:4 ตามลำดับ จำนวนข้อมูลแต่ละชุดข้อมูลดังตารางที่ 4

ตารางที่ 4 จำนวนข้อมูลในแต่ละชุดข้อมูล รูปแบบนำเข้าและประเภท

	ประเภท	ข้อมูลนำเข้าปกติ		ข้อมูลนำเข้า NER	
		บทความ	คำถาม	บทความ	คำถาม
ชุดข้อมูลที่สร้างโดยมนุษย์	ข้อมูลฝึกสอน	3,582	8,610	3,582	8,610
	ข้อมูลตรวจสอบ	1,321	3,041	1,321	3,041
ชุดข้อมูลที่สังเคราะห์รอบที่ 1	ข้อมูลฝึกสอน	10,571	33,717	10,486	32,274
	ข้อมูลตรวจสอบ	3,096	9,223	3,073	8,958
ชุดข้อมูลที่สังเคราะห์รอบที่ 2	ข้อมูลฝึกสอน	18,047	73,990	17,612	57,496
	ข้อมูลตรวจสอบ	4,992	19,364	4,881	15,253

5.2. การนำข้อมูลเข้าเพื่อฝึกสอน

ในการวิจัยนี้ใช้แบบจำลองที่ผ่านการฝึกสอนมาแล้วเอ็มทีไฟฟ์ (mT5) โดยใช้ตัวตัดคำ (Tokenizer) และเข้ารหัสคือ “google/mt5-base” และ ฝึกสอนด้วย “google/mt5-base” โดยข้อมูลนำเข้าคือผลลัพธ์ในข้อ 5.1 กำหนดให้มีการฝึกสอนจำนวน 5 และ 10 รอบ บันทึกน้ำหนักในรอบของการฝึกสอนที่ดีที่สุดในรอบการทำงาน โดยพิจารณาจากค่าสูญเสียในชุดข้อมูลตรวจสอบ (Validation loss) ที่น้อยที่สุด จากการทดสอบพบว่าแบบจำลองที่มีค่าสูญเสียและมีผลลัพธ์ที่ดีมีการกำหนดค่าพารามิเตอร์หลักดังตารางที่ 5

ตารางที่ 5 ตารางแสดงค่าพารามิเตอร์หลักที่ใช้ในการฝึกสอน

ชื่อพารามิเตอร์	ค่าพารามิเตอร์
num_train_epochs	5
model_name_or_path	google/mt5-base
tokenizer_name_or_path	google/mt5-base
max_seq_length	256

learning_rate	1e-4
seed	42

5.3. การสร้างคำถามจากแบบจำลองการสร้างชุดคำถาม

ใช้แบบจำลองที่ได้จากหัวข้อที่ 5.2 สร้างคำถามจากบทความในชุดข้อมูลทดสอบและชุดข้อมูลที่ไม่เคยพบมาก่อน วิธีที่เลือกนำมาทดสอบการสร้างคำถามวิธีการค้นหาแบบลำแสง (Beam Search) โดยกำหนดจำนวนคำถามที่ต้องการให้สร้างบทความละ 10 คำถาม กำหนดค่าความน่าจะเป็นของคำ (Top-P) และจำนวนคำที่น่าจะเป็น K ตัวแรก (Top-K sampling) ซึ่งมีการกำหนดค่าพารามิเตอร์หลักดังตารางที่ 6 ได้ผลลัพธ์ของการสร้างคำถามโดยใช้แบบจำลองคือประโยคคำถามและคะแนนความน่าจะเป็นที่สร้างคำถามนั้น ตัวอย่างบทความและคำถามที่ได้จากแต่ละแบบจำลองดังตารางที่ 7

ตารางที่ 6 ตารางแสดงค่าพารามิเตอร์หลักที่ใช้ในการถอดรหัสเพื่อสร้างคำถาม

ชื่อพารามิเตอร์	ค่าพารามิเตอร์
num_return_sequences	10
min_length	3
max_length	30
output_scores	True
top_p	0.9
top_k	80

ตารางที่ 7 ตัวอย่างบทความและคำถามจากแต่ละแบบจำลอง

บทความที่ 1	
เดอะฟอกออฟวอร์ อีเลเฟเวนเลสเซ็นส์ฟรอมเดอะไลฟ์ออฟโรเบิร์ต แม็กนามารา คือภาพยนตร์สารคดีจากสหรัฐอเมริกาที่กำกับโดย เออร์โรล มอร์ริส เกี่ยวกับชีวิตและช่วงเวลาของโรเบิร์ต แม็กนามารา อดีตรัฐมนตรีว่าการกระทรวงกลาโหมแห่งสหรัฐอเมริกา เดอะ ฟอก ออฟ วอร์นำเสนอถึงความไม่แน่นอนในสนามรบระหว่างสงคราม ภาพยนตร์เรื่องนี้ได้รับรางวัลออสการ์ ประเภทภาพยนตร์สารคดียอดเยี่ยม และ Independent Spirit Award for Best Documentary Feature	
แบบจำลอง	คำถาม
แบบจำลองที่ฝึกสอนจากชุดข้อมูลที่สร้างโดยมนุษย์และมีรูปแบบข้อมูลนำเข้าปกติ	เดอะฟอกออฟวอร์ กำกับโดยใคร
แบบจำลองที่ฝึกสอนจากชุดข้อมูลที่สร้างโดยมนุษย์และมีรูปแบบข้อมูลนำเข้า NER	เดอะฟอกออฟวอร์ กำกับโดยใคร
แบบจำลองที่ฝึกสอนจากชุดข้อมูลสังเคราะห์รอบที่ 1 และมีรูปแบบข้อมูลนำเข้าปกติ	ภาพยนตร์สารคดีจากสหรัฐอเมริกาเรื่องเดอะฟอกออฟวอร์ กำกับโดยใคร

แบบจำลองที่ฝึกสอนจากชุดข้อมูลสังเคราะห์ รอบที่ 1 และมีรูปแบบข้อมูลนำเข้า NER	เดอะฟ็อกออฟวอร์ เป็นภาพยนตร์สารคดีจากสหรัฐอเมริกาที่กำกับโดยใคร
แบบจำลองที่ฝึกสอนจากชุดข้อมูลสังเคราะห์ รอบที่ 2 และมีรูปแบบข้อมูลนำเข้าปกติ	ภาพยนตร์สารคดีจากสหรัฐอเมริกาเรื่องเดอะฟ็อกออฟวอร์ ได้รับรางวัลออสการ์ประเภทภาพยนตร์อะไร
แบบจำลองที่ฝึกสอนจากชุดข้อมูลสังเคราะห์ รอบที่ 2 และมีรูปแบบข้อมูลนำเข้า NER	ภาพยนตร์สารคดีเรื่องเดอะฟ็อกออฟวอร์ เป็นภาพยนตร์แนวอะไร

บทความที่ 2

ภาษาพาเซนต์ ภาษาพาเซนต์ เป็นภาษาที่พัฒนามาจากภาษาเปอร์เซียกลาง โดยตัดคำยืมจากภาษากลุ่มเซมิติก เช่น ภาษาอราเมอิกออกไป เริ่มแรกพบในข้อความศักดิ์สิทธิ์ของศาสนาโซโรอัสเตอร์ ภาษานี้เป็นภาษาเขียนหลังจากสิ้นสุดยุคซัสซานีเยน และลดความสำคัญลงเมื่อศาสนาอิสลามแพร่เข้ามาถึง จนกลายเป็นภาษาตายในที่สุด

แบบจำลอง	คำถาม
แบบจำลองที่ฝึกสอนจากชุดข้อมูลที่สร้างโดยมนุษย์และมีรูปแบบข้อมูลนำเข้าปกติ	ภาษาที่พัฒนามาจากภาษาเปอร์เซียกลางคือภาษาใด
แบบจำลองที่ฝึกสอนจากชุดข้อมูลที่สร้างโดยมนุษย์และมีรูปแบบข้อมูลนำเข้า NER	ภาษาพาเซนต์เป็นภาษาพัฒนามาจากภาษาใด
แบบจำลองที่ฝึกสอนจากชุดข้อมูลสังเคราะห์ รอบที่ 1 และมีรูปแบบข้อมูลนำเข้าปกติ	ภาษาพาเซนต์ เริ่มแรกพบในข้อความศักดิ์สิทธิ์ของศาสนาโซโรอัสเตอร์เรียกว่าอะไร
แบบจำลองที่ฝึกสอนจากชุดข้อมูลสังเคราะห์ รอบที่ 1 และมีรูปแบบข้อมูลนำเข้า NER	ภาษาพาเซนต์ เป็นภาษาที่พัฒนามาจากภาษาเปอร์เซียกลาง โดยตัดคำยืมจากภาษากลุ่มเซมิติกเช่น
แบบจำลองที่ฝึกสอนจากชุดข้อมูลสังเคราะห์ รอบที่ 2 และมีรูปแบบข้อมูลนำเข้าปกติ	ภาษาพาเซนต์ เป็นภาษาที่พัฒนามาจากภาษาเปอร์เซียกลาง โดยตัดคำยืมจากภาษากลุ่มเซมิติกเช่นอะไร
แบบจำลองที่ฝึกสอนจากชุดข้อมูลสังเคราะห์ รอบที่ 2 และมีรูปแบบข้อมูลนำเข้า NER	ภาษาพาเซนต์ เป็นภาษาที่พัฒนามาจากภาษาเปอร์เซียกลาง โดยตัดคำยืมจากภาษากลุ่มอะไรออกไป

บทความที่ 3

นายสวัสดิ์ วัฒนายากร เกิดที่จังหวัดปัตตานี เป็นบุตรคนที่ 7 ในจำนวน 10 คนของขุนธำรงพันธุ์ภักดี กับนางสร้อยทอง วัฒนายากร สมรสกับคุณหญิงปานจิตต์ วัฒนายากร อดีตผู้ช่วยปลัดกระทรวงเกษตรและสหกรณ์ มีบุตรชาย 1 คน คือ ทศนพงษ์ วัฒนายากรนายสวัสดิ์ วัฒนายากร ถึงแก่อสัญกรรมเมื่อวันที่ 13 มีนาคม พ.ศ. 2555 ที่โรงพยาบาลรามธิบดี ด้วยโรคมะเร็งในหลอดลม รวมอายุ 77 ปี 5 เดือน มีพิธีพระราชทานน้ำหลวงอาบศพ เมื่อวันที่ 14 มีนาคม ณ ศาลาบำเพ็ญกุศล วัดเบญจมบพิตรดุสิตวนาราม โดยพระบาทสมเด็จพระเจ้าอยู่หัวฯ ทรงพระกรุณาโปรดเกล้าฯ ให้ พลเอกเปรม ติณสูลานนท์ ประธานองคมนตรีและรัฐบุรุษ เป็นผู้แทนพระองค์ ในการนี้ ทรงพระกรุณาโปรดเกล้าฯ พระราชทานโกศมณฑป พร้อมเครื่องประกอบเกียรติยศศพ พระพิธีธรรมสวดพระอภิธรรมศพ 7 คืน กับพระราชทานบำเพ็ญกุศลงาน 7 วัน และทรงรับไว้ในพระบรมราชานุเคราะห์นายสวัสดิ์ วัฒนา

<p>ยากร มีศักดิ์เป็นน้องชายของนายมงคล วัฒนายากร ซึ่งเป็นบิดาของ รองศาสตราจารย์ ดร.ปณิธาน วัฒนายากร อดีตรองเลขาธิการนายกรัฐมนตรีฝ่ายการเมือง</p>	
แบบจำลอง	คำถาม
แบบจำลองที่ฝึกสอนจากชุดข้อมูลที่สร้างโดย มนุษย์และมีรูปแบบข้อมูลนำเข้าปกติ	นายสวัสดิ์ วัฒนายากร มีศักดิ์เป็นน้องชายของใคร
แบบจำลองที่ฝึกสอนจากชุดข้อมูลที่สร้างโดย มนุษย์และมีรูปแบบข้อมูลนำเข้า NER	สวัสดิ์ วัฒนายากร เป็นบุตรคนที่เท่าไรของขุนธำรงพันธุ์ภักดี
แบบจำลองที่ฝึกสอนจากชุดข้อมูลสังเคราะห์ รอบที่ 1 และมีรูปแบบข้อมูลนำเข้าปกติ	นายสวัสดิ์ วัฒนายากร ถึงแก่อสัญกรรมเมื่อวันที่เท่าไร
แบบจำลองที่ฝึกสอนจากชุดข้อมูลสังเคราะห์ รอบที่ 1 และมีรูปแบบข้อมูลนำเข้า NER	สวัสดิ์ วัฒนายากร ได้เสียชีวิตจากโรคมะเร็งในหลอดลม รวมอายุกี่ปี
แบบจำลองที่ฝึกสอนจากชุดข้อมูลสังเคราะห์ รอบที่ 2 และมีรูปแบบข้อมูลนำเข้าปกติ	บุตรคนที่ 7 ในจำนวน 10 คนของขุนธำรงพันธุ์ภักดีกับนาง สร้อยทอง วัฒนายากร คือ
แบบจำลองที่ฝึกสอนจากชุดข้อมูลสังเคราะห์ รอบที่ 2 และมีรูปแบบข้อมูลนำเข้า NER	เจ้าหญิงสาวคนใดได้รับรางวัลพระราชทานจาก คณะกรรมการนโยบายองค์กรเกษตรและสหกรณ์ในปี พ.ศ. 2558

จากผลลัพธ์การสร้างคำถามของแบบจำลองที่ต่างกันแสดงให้เห็นว่าแบบจำลองมีการเรียนรู้ที่จะสร้างคำถามที่มีความยาวและซับซ้อนมากขึ้นจากบทความเดียวกัน ซึ่งมีบางประโยคที่ไม่มีความแตกต่างในด้านความหมายหรือรูปประโยค แต่บางประโยคไม่สมบูรณ์หรือไม่สามารถหาคำตอบได้จากบทความ

5.4. การคัดกรองคำถาม

ในขั้นตอนนี้จะคัดกรองคำถามจากแต่ละบทความโดยยกตัวอย่างคำถามที่สร้างจากแบบจำลองจากตารางที่ 4 ในแต่ละขั้นตอนดังนี้

5.4.1. ลบคำถามที่เหมือนกันจากบทความเดียวกัน

ลบคำถามที่เหมือนกันจากบทความเดียวตามขั้นตอนดังนี้

- 1) สร้างชุดข้อมูล 2 ชุดเพื่อเก็บคำถามต้นแบบและคำถามที่ลบช่องว่างได้ผลลัพธ์ดังนี้

ชุดข้อมูลที่เก็บคำถามต้นแบบ: {}

ชุดข้อมูลที่เก็บคำถามที่ลบช่องว่าง: {}

- 2) ลบช่องว่างที่ปรากฏในคำถามที่พิจารณาออกได้ผลลัพธ์ดังนี้

ตัวอย่างคำถามต้นแบบ: “เหมย หลันฟิง เป็นนักแสดงชายผู้รับบทอะไร”

ตัวอย่างคำถามที่ลบช่องว่าง: “เหมยหลันฟิงเป็นนักแสดงชายผู้รับบทอะไร”

3) ตรวจสอบคำถามจากกระบวนการที่ 2 กับชุดข้อมูลที่เก็บคำถามที่ลบช่องว่าง กรณีที่คำถามที่พิจารณาปรากฏในชุดข้อมูลที่เก็บคำถามที่ลบช่องว่างจะพิจารณาคำถามต่อไป แต่หากไม่พบคำถามดังกล่าวในชุดข้อมูลที่เก็บคำถามที่ลบช่องว่าง จะเก็บคำถามต้นแบบในชุดข้อมูลที่เก็บคำถามต้นแบบในกรณีตัวอย่างจะได้ผลลัพธ์ดังนี้

ชุดข้อมูลที่เก็บคำถามต้นแบบ: {“เหมย หลันฟิง เป็นนักแสดงชายผู้รับบทอะไร”}

ชุดข้อมูลที่เก็บคำถามที่ลบช่องว่าง: {“เหมยหลันฟิงเป็นนักแสดงชายผู้รับบทอะไร”}

5.4.2. ลบคำถามที่มีตัวเลขที่ไม่ปรากฏในบทความ

ลบคำถามที่มีตัวเลขที่ไม่ปรากฏในบทความตามขั้นตอนดังนี้

1) สร้างชุดเก็บข้อมูลตัวเลขที่ปรากฏในบทความ จากตัวอย่างบทความตารางที่ 4 จะได้ผลลัพธ์ดังนี้

ชุดข้อมูลตัวเลขที่ปรากฏในบทความ: {}

2) สกัดตัวเลขที่ปรากฏในบทความลงในชุดจากข้อ 1)

ชุดข้อมูลตัวเลขที่ปรากฏในบทความ: {8, 11, 1, 4, 1920, 2480, 2488, 1920, 1960, 1919, 1930, 1934, 1935, 1952, 2551}

3) สกัดตัวเลขที่ปรากฏในคำถาม พิจารณาคำถามที่ปรากฏในตารางที่ 4 ในแต่ละคำถามจะพบคำถามว่า “ผลงานการแสดงของเหมย หลันฟิงในปี ค.ศ. 1929 เป็นภาพยนตร์ที่กำกับโดยใคร” ซึ่งเมื่อสกัดตัวเลขจะได้เลข “1929”

4) พิจารณาตัวเลขที่ปรากฏในคำถามทุกตัวเพื่อตรวจสอบว่ามีค่าอยู่ในชุดเก็บข้อมูลตัวเลขที่ปรากฏในบทความหรือไม่ กรณีที่พบจะทำตัวถัดไปจนครบ หากไม่พบจะถือว่าคำถามนั้นไม่ผ่านเงื่อนไขนี้ ซึ่งจากผลลัพธ์ในข้อ 3) ตัวเลข “1929” ไม่ปรากฏในชุดเก็บข้อมูลตัวเลขที่ปรากฏในบทความ ดังนั้นคำถามนี้จึงไม่ผ่านเงื่อนไขที่กำหนดในข้อนี้

5.4.3. คัดกรองคำถามที่มีรูปแบบประโยคคำถามตามที่กำหนด

เลือกประโยคคำถามโดยใช้กฎการสร้างคำถามในประโยคภาษาไทย ซึ่งจะต้องปรากฏวลีที่เป็นคำถามดังต่อไปนี้ในประโยค “ใคร,ทำไม,กี่,อะไร,เท่าไร,ที่ไหน,อย่างไร,ไหน,ใด,เมื่อไร,เท่าไร,เมื่อไหร่” โดยกำหนดนิพจน์ปกติดังนี้ “.*ใคร.*|.*ทำไม.*|.*กี่.*|.*อะไร.*|.*เท่าไร.*|.*ที่ไหน.*|.*อย่างไร.*|.*ไหน.*|.*ใด.*|.*เมื่อไร.*|.*เมื่อไหร่.*|.*เท่าไร.*|.*เมื่อไหร่.*|.*ใด.*|.*[^\s]*ที่ร\$”

5.4.4. คำถามที่มีความเกี่ยวข้องกับบทความ

ในการพิจารณาคำถามที่เกี่ยวข้องกับบทความโดยจะต้องมีรูปแบบประโยคที่มีค่าปรากฏในบทความ จากการทดลองกับชุดข้อมูลที่มีผลเฉลยจากมนุษย์ได้ผลลัพธ์ดังตารางที่ 8 เนื่องจากสร้างชุดข้อมูลสังเคราะห์ควรเป็นคำถามที่มีคุณภาพ การพิจารณาที่ค่าความแม่นยำจึงเหมาะสมในการพิจารณาเกณฑ์เปอร์เซ็นต์ความน่าจะเป็นเพื่อให้ชุดข้อมูลสังเคราะห์มีคุณภาพ จึงกำหนดเกณฑ์ที่ในความน่าจะเป็นที่คำถามมีความเกี่ยวข้องกับบทความมากกว่า 75 เปอร์เซ็นต์ ซึ่งหากกำหนดเกณฑ์ที่มากกว่านี้จะทำให้ชุดข้อมูลสังเคราะห์มีข้อมูลลดลง

ตารางที่ 8 ผลการวัดความสามารถของการคัดกรองคำถามในรูปแบบของค่าความแม่นยำ(Precision) ค่าความถูกต้อง(Recall) ค่าความเที่ยงตรง(Accuracy) และค่าเฉลี่ยระหว่างความแม่นยำและความถูกต้อง(F1)

เกณฑ์เปอร์เซ็นต์ความน่าจะเป็น	ความแม่นยำ	ความถูกต้อง	ความเที่ยงตรง	ค่าเฉลี่ยระหว่างความแม่นยำและความถูกต้อง
>= 60	0.763	0.607	0.605	0.676
>= 65	0.785	0.538	0.587	0.639
>= 70	0.808	0.470	0.565	0.594
>= 75	0.823	0.419	0.544	0.555

แสดงตัวอย่างบทความและคำถามที่ผ่านการคัดกรองความเกี่ยวข้องกับบทความดังต่อไปนี้

บทความที่ 1 แพคิเซอฟาโลซอร์ส แพคิเซอฟาโลซอร์ส เป็นไดโนเสาร์หัวแข็ง หรือแพคิเซอฟาโลซอร์ชนิดหนึ่ง ไดโนเสาร์ชนิดนี้มีหัวหนาถึง 25 เซนติเมตร ซึ่งน่าจะมีไว้ต่อสู้หรือป้องกันตัว ฟอสซิลของค้นพบที่รัฐไวโอมิง ในปี ค.ศ. 1931 โดยนักล่าฟอสซิลชื่อกิลมอร์ แพคิเซอฟาโลซอร์อาศัยอยู่ในยุคครีเทเชียสตอนปลายเมื่อประมาณ 7065 ล้านปีก่อน มีญาติอย่างพรีโนเซฟาลี และดราโกเร็กซ์หัวประหลาด หัวประหลาด แพคิเซอฟาโลซอร์สมีกระดูกส่วนหัวหนาถึง 25 เซนติเมตร หรือ 12 นิ้ว ส่วนคอต่อกับส่วนล่างของกะโหลก ไม่ใช่หลังหัวเหมือนกับสัตว์เลื้อยคลานทั่วไป เมื่อกัมหัวคอจะตรงเป็นแนวเดียวกับกระดูกสันหลัง ซึ่งเหมาะสำหรับรับแรงกระแทก ผู้เชี่ยวชาญคาดว่ามันน่าจะมีไว้ใช้เพื่อต่อสู้แย่งตัว หรือมีไว้ใช้เพื่อป้องกันตัวจากนักล่าเช่นไทแรนโนซอร์ส แต่ บ็อบ เบ็กเกอร์ผู้เชี่ยวชาญด้านไดโนเสาร์โต้แย้งว่า กะโหลกของแพคิเซอฟาโลซอร์สไม่แข็งพอที่จะใช้ต่อสู้ แต่ยังไม่มีการศึกษาของเขา เนื่องจากสมองของแพคิเซอฟาโลซอร์สและไดโนเสาร์หัวแข็งมีขนาดเล็ก และถ้าสมองเล็กก็หมายถึงสมองจะได้รับการกระทบเหือนน้อยกว่าสมองใหญ่

คำถาม

- 1) ไดโนเสาร์ชนิดใดมีหัวหนาถึง 25 เซนติเมตร
- 2) แพคิเซอฟาโลซอร์สมีกระดูกส่วนหัวหนาถึงเท่าไร
- 3) ไดโนเสาร์ชนิดนี้มีหัวหนาถึงเท่าไร

บทความที่ 2 อีพีซี อีพีซี ย่อมาจากคำว่าวิศวกรรม การจัดซื้อ และการก่อสร้าง สัญญาอีพีซีเป็นรูปแบบทั่วไปของการจัดทำสัญญาในอุตสาหกรรมก่อสร้าง ภายใต้สัญญาอีพีซี ผู้รับเหมาจะออกแบบ จัดหาอุปกรณ์ที่จำเป็นและก่อสร้าง หรือ ติดตั้ง ทั้งโดยตนเองหรือผู้รับเหมาช่วง ผู้รับเหมาจะดำเนินโครงการด้วยความเสี่ยงเองเพื่อกำหนดค่างานเป็นงบประมาณในการก่อสร้าง ซึ่งขึ้นอยู่กับขอบเขตงานที่ตกลงกัน

คำถาม

- 1) สัญญาอีพีซีเป็นรูปแบบทั่วไปของการจัดทำสัญญาในอุตสาหกรรมก่อสร้าง ภายใต้สัญญาอะไร
- 2) สัญญาอีพีซี เป็นรูปแบบทั่วไปของการจัดทำสัญญาในอุตสาหกรรมก่อสร้าง ภายใต้สัญญาใด

บทความที่ 3 ซิซึ ซิซึ หรือในชื่อภาษาจีน จีจี เป็นแพนด้ายักษ์เพศเมียที่มีชื่อเสียงจากสวนสัตว์ลอนดอน ในประเทศอังกฤษ แม้ว่า ซิซึ จะมีได้เป็นแพนด้ายักษ์ตัวแรกของสวนสัตว์ลอนดอน แต่มันก็เป็นที่ตั้งใจดูผู้คนในฐานะสัตว์ที่มีความน่ารักมากที่สุดของสวนสัตว์แห่งนี้ ซิซึ เกิดในปีค.ศ. 1957 และนำตัวมาจากมณฑลเสฉวนในเดือนธันวาคม ค.ศ. 1957 ก่อนที่จะส่งไปยังสวนสัตว์ปักกิ่งช่วงเดือนมกราคม ค.ศ. 1958 ตัวแพนจัดหาสัตว์ชาวออสเตรเลียซึ่งมีชื่อว่า เฮย์นีย์ เดมเมอร์ ได้รับซิซึมาจากการแลกเปลี่ยนสถาบันแอนทริกกันฮูฟสตอกเพื่อกระซับมิตรในเดือนพฤษภาคม ค.ศ.1958 และนำไปอยู่ที่สวนสัตว์มอสโก

คำถาม

- 1) ซิซึ เป็นแพนด้ายักษ์เพศเมียที่มีชื่อเสียงจากสวนสัตว์ลอนดอนในประเทศอังกฤษ เกิดวันที่เท่าไร
- 2) ซิซึ เป็นแพนด้ายักษ์เพศเมียที่มีชื่อเสียงจากสวนสัตว์ลอนดอน ในประเทศใด
- 3) ซิซึเป็นแพนด้ายักษ์เพศเมียที่มีชื่อเสียงจากสวนสัตว์ลอนดอนในประเทศอังกฤษ ซึ่งมีชื่อว่าอะไร
- 4) ซิซึ เป็นแพนด้ายักษ์เพศเมียที่มีชื่อเสียงจากสวนสัตว์ลอนดอนในประเทศอังกฤษมีชื่อภาษาจีนว่าอะไร
- 5) ซิซึ เป็นแพนด้ายักษ์เพศเมียที่มีชื่อเสียงจากสวนสัตว์ลอนดอนในประเทศอะไร
- 6) ซิซึ หรือในชื่อภาษาจีนเรียกว่าอะไร

5.5. การสังเคราะห์ข้อมูล

รวมชุดข้อมูลที่ได้จากการคัดกรองเข้ากับชุดข้อมูลตั้งต้นที่ใช้ในการฝึกแบบจำลองการสร้างคำถาม โดยผลลัพธ์ที่ได้จากการรวมคือชุดข้อมูลสังเคราะห์ดังตารางที่ 4 โดยในชุดข้อมูลสังเคราะห์ของรอบที่ 1 เกิดจากการรวมกันของชุดข้อมูลที่สร้างโดยมนุษย์กับคำถามที่คัดกรองจากการสร้างคำถามและชุดข้อมูลสังเคราะห์รอบที่ 2 เกิดจากการรวมกันของชุดข้อมูลสังเคราะห์รอบที่ 1 กับคำถามที่คัดกรองจากการสร้างคำถามของชุดข้อมูลสังเคราะห์รอบที่ 1 ข้อมูลที่ได้จากการสังเคราะห์จะถูกใช้ในการฝึกสอนแบบจำลองโดยมีขั้นตอนการดำเนินการเช่นเดียวกับที่ได้กล่าวไว้ในข้างต้น

5.6. การประเมินผลการทดลอง

สุ่มเลือกข้อมูลทดสอบจากความยาวของบทความโดยแบ่งเป็นกลุ่มความยาวและจำนวนตามสัดส่วนของกลุ่มความยาวทั้งหมดที่ปรากฏในชุดข้อมูลทดสอบดังตารางที่ 9 ชุดข้อมูลทดสอบทั้งหมด 100 บทความและคำถามที่มีคะแนนสูงที่สุดของแต่ละบทความเพื่อใช้ในการประเมินการประเมินประสิทธิภาพของแบบจำลองการสร้าง

ตารางที่ 9 สัดส่วนข้อมูลทดสอบเพื่อประเมินตามกลุ่มความยาวของบทความ

ความยาวของบทความ	สัดส่วนต่อบทความทั้งหมด	จำนวนที่ปรากฏในชุดข้อมูลทดสอบ	จำนวนบทความที่สุ่มคัดเลือกเพื่อประเมิน
น้อยกว่า 500	26.04	63	26
500 – 1000	29.34	71	30
1000 – 1500	13.23	32	13

1500 – 2000	9.92	24	10
2000 – 2500	6.20	15	6
2500 – 3000	15.29	37	15
รวม	100.00	242	100

การประเมินแบ่งเป็น 2 ส่วนดังนี้

- 1) การประเมินโดยอัตโนมัติ ในขั้นตอนนี้ใช้มาตรวัดคะแนนเบลอโดยกำหนดพารามิเตอร์ที่พิจารณาคะแนนเบลอในแต่ละจำนวนแกรมและเพิ่มกระบวนการเกี่ยวข้องข้อมูลให้เป็นมาตรฐานกรณีที่มีความยาวประโยคสั้น ซึ่งส่งผลต่อคะแนนเบลอในแต่ละแกรม กำหนดให้คำถามที่สร้างโดยมนุษย์เป็นประโยคอ้างอิงและคำถามที่มีคะแนนสูงที่สุดในบทความเป็นประโยค

ผลลัพธ์ของการเปรียบเทียบระหว่างแบบจำลองที่ฝึกสอนจากชุดที่ต่างกันและรูปแบบข้อมูลนำเข้าที่ต่างกันดังตารางที่ 10 จากการทดลองสังเกตได้ว่าแบบจำลองที่มีการฝึกสอนจากชุดข้อมูลสังเคราะห์ให้ผลลัพธ์ที่ดีกว่าแบบจำลองที่ฝึกสอนจากชุดข้อมูลที่สร้างโดยมนุษย์ในทุกคะแนนเบลอ เมื่อเปรียบเทียบระหว่างแบบจำลองที่ฝึกสอนจากชุดข้อมูลสังเคราะห์ พบว่าแบบจำลองที่ฝึกสอนจากชุดข้อมูลสังเคราะห์รอบที่ 1 ได้คะแนนเบลอ 1 และคะแนนเบลอ 2 สูงที่สุดสมมติฐานว่าชุดข้อมูลสังเคราะห์จากรอบที่ 2 อาจมีคำถามที่ไม่ตรงตามเกณฑ์ที่กำหนดแต่ผ่านการคัดกรองจึงทำให้ประสิทธิภาพของแบบจำลองลดลง แต่คะแนนเบลอ 3 และคะแนนเบลอ 4 จากแบบจำลองที่ฝึกสอนจากชุดข้อมูลสังเคราะห์รอบที่ 2 สูงกว่า สมมติฐานว่าแบบจำลองเรียนรู้จากชุดข้อมูลที่ใช้ในการฝึกสอนเพิ่มขึ้น ทำให้แบบจำลองพยายามสร้างคำถามที่มีความยาวและความซับซ้อนมากขึ้น ซึ่งจากการวัดผลด้วยคะแนนเบลอโดยการเปรียบเทียบชุดคำถามที่มนุษย์สร้างกับคำถามที่แบบจำลองสร้างจากบทความเดียวกันมีการใช้ฟังก์ชันการเกลี่ย (smoothing function) กรณีที่คำถามเป็นประโยคสั้นสำหรับคะแนนเบลอที่จำนวนเอ็นสูง จึงอาจส่งผลให้คะแนนเบลอ 3 และคะแนนเบลอ 4 ของแบบจำลองที่ฝึกสอนจากชุดข้อมูลสังเคราะห์รอบที่ 2 สูงกว่า

การทดลองรูปแบบข้อมูลนำเข้าแบบปกติและแบบ NER พบว่าโดยรวมแบบปกติให้ผลคะแนนที่ดีกว่า สมมติฐานว่ากระบวนการในการจัดเตรียมข้อมูลนำเข้าแบบ NER มีการติดป้ายกำกับการสกัดนิพจน์เฉพาะหรือชื่อเฉพาะไม่เหมาะสมอาจจะด้วยข้อจำกัดทางด้านประสิทธิภาพของคลังโปรแกรม PyThaiNLP หรือการทำรูปแบบนำเข้าแบบ NER ไม่เหมาะสมต่อแบบจำลองการสร้างคำถาม จึงเป็นแนวทางในการปรับเปลี่ยนการนำเข้าข้อมูลหรือใช้คลังโปรแกรมอื่นเพื่อติดป้ายนิพจน์คำนามในอนาคต

ตารางที่ 10 ผลการประเมินอัตโนมัติจากคะแนนเบลอเปรียบเทียบระหว่างแบบจำลองและรูปแบบข้อมูลนำเข้า

	ประเภท ข้อมูลนำเข้า	BLEU-1	BLEU-2	BLEU-3	BLEU-4
แบบจำลองที่ฝึกสอน จากชุดข้อมูลที่สร้าง โดยมนุษย์	ข้อมูลนำเข้าปกติ	56.19	43.93	36.61	31.88
	ข้อมูลนำเข้า NER	56.68	43.46	36.71	31.70
แบบจำลองที่ฝึกสอน จากชุดข้อมูลสังเคราะห์ รอบที่ 1	ข้อมูลนำเข้าปกติ	59.03	44.49	37.49	32.82
	ข้อมูลนำเข้า NER	56.65	43.13	36.45	32.46
แบบจำลองที่ฝึกสอน จากชุดข้อมูล สังเคราะห์รอบที่ 2	ข้อมูลนำเข้าปกติ	57.30	44.00	38.65	34.00
	ข้อมูลนำเข้า NER	55.19	41.37	36.02	32.44

การเปรียบเทียบระหว่างแบบจำลองที่มีคะแนนเบลอมากที่สุดกับแบบจำลองในงานวิจัยที่ใกล้เคียง โดยงานวิจัยที่นำมาเปรียบเทียบคืองานวิจัยที่สร้างคำถามภาษาอังกฤษเนื่องจากให้ภาษาไทยยังไม่ม้งานวิจัยที่ได้ผลลัพธ์ในเรื่องนี้มาก่อน ผลลัพธ์ของแบบจำลองการสร้างคำถามภาษาไทยมีคะแนนที่มากกว่างานวิจัยที่ใกล้เคียงอาจด้วยเรื่องของภาษาที่มีความแตกต่างกัน ดังตารางที่ 11

ตารางที่ 11 ผลการประเมินอัตโนมัติจากคะแนนเบลอเปรียบเทียบระหว่างแบบจำลองที่นำเสนอกับงานวิจัยที่ใกล้เคียง

แบบจำลอง	BLEU-1	BLEU-2	BLEU-3	BLUE-4
OQPL Standard [18]	55.60	31.03	16.56	7.89
CopyLossEnhanced [20]	45.08	27.98	19.38	13.90
แบบจำลอง การสร้างคำถามภาษาไทย	59.03	44.49	37.49	32.82

- 2) การประเมินโดยมนุษย์ จากเกณฑ์และรูปแบบการประเมินจากบทความและคำถามจำนวน 100 ข้อ ข้อละ 1 คะแนน คะแนนเต็มคือ 100 คะแนนผ่านเอกสารเอกซ์เซล ได้ผลลัพธ์ของผู้ประเมินแต่ละท่านดังตารางที่ 12 และตารางที่ 13 สรุปผลคะแนนประเมินโดยเทียบกับงานวิจัยก่อนหน้าดังตารางที่ 14 โดยในแต่ละเกณฑ์วิเคราะห์ได้ดังนี้

- ความไพเราะของประโยค คือมีองค์ประกอบของประโยคที่ครบถ้วนตามโครงสร้างของประโยค คำถามได้คะแนนเฉลี่ย 88% เนื่องด้วยประโยคคำถามมีคำเกินหรือไม่สมบูรณ์เช่น “ตัวละครตัวละครของเมก้า XLR คือใคร”, “ขอทราบชื่อสินค้าหน่อย”

- มีความหมายของประโยคและบริบทที่เกี่ยวข้องเหมาะสมกับบทความ คือ คำถามมีความเกี่ยวข้องกับบทความและเหมาะสมกับบริบทได้คะแนนเฉลี่ย 93% คำถามที่ผู้ประเมินให้ไม่ผ่านในเกณฑ์นี้เช่น “เมืองหลวงของจังหวัดสมุทรสงครามมีชื่อว่าอะไร”
- สามารถหาคำตอบได้จากบทความ คือถามในเรื่องที่ปรากฏข้อมูลในบทความได้คะแนนเฉลี่ย 94% ในเกณฑ์นี้เนื่องด้วยคำถามไม่มีคำตอบในบทความจริงหรือคำถามกำกวมทำให้ไม่สามารถตอบได้เช่น “จังหวัดพะเยาเป็นจังหวัดในกลุ่มใด”, “โจรสลัดที่มีชื่อเสียงคือใคร”

ตารางที่ 12 ตัวอย่างบทความ คำถามและผลการประเมินโดยมนุษย์

บทความที่ 1				
ภาษาพาเซนต์ ภาษาพาเซนต์ เป็นภาษาที่พัฒนามาจากภาษาเปอร์เซียกลาง โดยตัดคำยืมจากภาษากลุ่มเซมิติกเช่น ภาษาอาหรับออกไป เริ่มแรกพบในข้อความศักดิ์สิทธิ์ของศาสนาโซโรอัสเตอร์ ภาษานี้เป็นภาษาเขียนหลังจากสิ้นสุดยุคชัสซาเนียน และลดความสำคัญลงเมื่อศาสนาอิสลามแพร่เข้ามาถึง จนกลายเป็นภาษาตายในที่สุด				
คำถามที่ 1				
ภาษาพาเซนต์ เริ่มแรกพบในข้อความศักดิ์สิทธิ์ของศาสนาอะไร				
ผู้ประเมิน	ความพึงพอใจ	ความเกี่ยวข้องเหมาะสม	สามารถหาคำตอบได้	คำตอบ (ถ้ามี)
ผู้ประเมิน 1	ไม่ยอมรับ	ยอมรับ	ยอมรับ	ศาสนาโซโรอัสเตอร์
ผู้ประเมิน 2	ยอมรับ	ยอมรับ	ยอมรับ	โซโรอัสเตอร์
ผู้ประเมิน 3	ยอมรับ	ยอมรับ	ยอมรับ	โซโรอัสเตอร์
บทความที่ 2				
ค่อม ชวนชื่น ค่อม ชวนชื่น ชื่อจริง อาคม ปรีดากุล เกิดเมื่อวันที่ 1 มกราคม 2500 ที่กรุงเทพมหานคร เป็นนักแสดงตลกชาวไทย อีกทั้งเป็นอดีตสมาชิกตลกคณะชวนชื่นประวัติ ประวัติ. ค่อม ชวนชื่น ชื่อจริง อาคม ปรีดากุล เกิดเมื่อวันที่ 1 มกราคม 2500 ที่กรุงเทพมหานคร เป็นนักแสดงตลกชาวไทย และเป็นอดีตสมาชิกตลกคณะชวนชื่น มักจะเรียกติดปากว่า น้ำเหิน ด้วยความที่ไม่รู้หนังสือเลย แต่น้ำค่อมกลับเป็นคนที่มีความสามารถพิเศษเรื่องการจำบทได้อย่างแม่นยำ ในระยะประมาณ 10 ปีที่ผ่านมาค่อมมีผลงานการแสดงไม่ต่ำกว่า 60 เรื่อง...				
คำถามที่ 2				
นักแสดงตลกชาวไทย คือใคร				
ผู้ประเมิน	ความพึงพอใจ	ความเกี่ยวข้องเหมาะสม	สามารถหาคำตอบได้	คำตอบ (ถ้ามี)
ผู้ประเมิน 1	ยอมรับ	ไม่ยอมรับ	ไม่ยอมรับ	
ผู้ประเมิน 2	ยอมรับ	ไม่ยอมรับ	ยอมรับ	ค่อม ชวนชื่น
ผู้ประเมิน 3	ยอมรับ	ยอมรับ	ยอมรับ	ค่อม ชวนชื่น

บทความที่ 3				
<p>เร็กนาร์อ็อกออนไลน์ เร็กนาร์อ็อกออนไลน์ เป็นเกม MMORPG มุมมองบุคคลที่สามจากด้านเฉียงบน ระบบกึ่งสามมิติ ที่ถูกพัฒนาขึ้นโดยบริษัท กราวิตี้ คอร์ปอเรชั่น ประเทศเกาหลีใต้ เริ่มทดสอบระบบเซิร์ฟเวอร์วันที่ 1 พฤศจิกายน 2544 เป็นระยะเวลา 9 เดือน ก่อนเปิดให้บริการเป็นครั้งแรกในเกาหลีใต้เมื่อ 31 สิงหาคม พ.ศ. 2545 โดยเนื้อหาส่วนใหญ่มาจากการ์ตูนเรื่อง RAGNAROK ภูตเทพวิบัติ ซึ่งประพันธ์โดยอี มย็องจิน สำหรับในประเทศไทย เปิดให้บริการวันที่ 16 กันยายน 2545 แบบโคลสเบต้าจำนวน 3 เซิร์ฟเวอร์ได้แก่ เคออส โลกิ และไอริส วันที่ 25 ตุลาคม 2545 เปิดบริการแบบโอเพ่นเบต้าทั้งหมด 6 เซิร์ฟเวอร์ได้แก่ เคออส โลกิ ไอริส ลิเดีย เฟนรีร์ และซาราร์ห์ โดยบริษัทเอเชียซอฟท์ คอร์ปอเรชั่น จำกัด...</p>				
คำถามที่ 3				
เนื้อหาส่วนใหญ่มาจากการ์ตูนเรื่องใด				
ผู้ประเมิน	ความพึงพอใจ	ความเกี่ยวข้องเหมาะสม	สามารถหาคำตอบได้	คำตอบ (ถ้ามี)
ผู้ประเมิน 1	ไม่ยอมรับ	ยอมรับ	ยอมรับ	RAGNAROK ภูตเทพวิบัติ
ผู้ประเมิน 2	ยอมรับ	ไม่ยอมรับ	ยอมรับ	เรื่อง RAGNAROK ภูตเทพวิบัติ
ผู้ประเมิน 3	ยอมรับ	ยอมรับ	ยอมรับ	RAGNAROK ภูตเทพวิบัติ

ตารางที่ 13 ผลประเมินคำถามจากผู้ประเมินแต่ละท่านตามเกณฑ์ที่กำหนด

ลำดับผู้ประเมิน	ความพึงพอใจของ ประโยค	มีความหมายและบริบทที่เกี่ยวข้อง เหมาะสมกับบทความ	สามารถหาคำตอบได้ จากในบทความ
1	76	91	89
2	90	80	90
3	93	99	94
เฉลี่ยรวม	88	93	94

ผลลัพธ์ของการประเมินในแต่ละเกณฑ์จากตารางที่ 13 พบว่าคำถามส่วนใหญ่สามารถหาคำตอบได้จากในบทความ ความแตกต่างของผู้ประเมิน 1 และ ผู้ประเมิน 2 ที่ให้ความเห็นในเรื่องของเกณฑ์ความพึงพอใจของประโยคและเกณฑ์มีความหมายและบริบทที่เกี่ยวข้องเหมาะสมกับบทความแม้ว่าจะมีการกำหนดเกณฑ์ร่วมกันแล้ว ตัวอย่างจากบทความที่ 3 ผู้ประเมิน 1 ให้ความเห็นกรณีไม่ยอมรับในเกณฑ์เรื่องความพึงพอใจเนื่องจากรูปประโยคคำถามไม่มีประธานแต่ยังมีความเกี่ยวข้องกับบทความ ส่วนผู้ประเมิน 2 ให้ความเห็นว่าประโยคใช้คำถามได้ถูกต้องแต่คำถามไม่มีการชี้เฉพาะเจาะจงทำให้ไม่สามารถระบุได้ว่ามีความเกี่ยวข้องกับบทความหรือไม่ ดังนั้นการประเมินโดยมนุษย์อาจต้องใช้ผู้ประเมินเพิ่มขึ้นหรือมีการกำหนดเกณฑ์การประเมินที่แน่ชัดร่วมกันมากกว่าในงานวิจัยนี้

งานวิจัยใกล้เคียงที่ใช้เกณฑ์การประเมินเดียวกันโดยเป็นการสร้างคำถามภาษาอังกฤษซึ่งมีคะแนนเต็มข้อละ 5 คะแนน จำนวน 100 ข้อ คะแนนเต็ม 500 คะแนน ได้ปรับการคิดคะแนนให้อยู่ใน

มาตรฐานเดียวกัน โดยใช้สูตรที่ (5.1) ผลลัพธ์ดังตารางที่ 10 ได้คะแนนในแต่ละเกณฑ์สูงกว่างานวิจัยที่เปรียบเทียบ แม้ว่าบริบททางภาษาและผู้ประเมินที่แตกต่างกัน

$$\text{คะแนนมาตรฐาน} = \frac{\text{คะแนนเฉลี่ยการประเมินจากแบบจำลองที่นำเสนอ} \times 5}{100} \quad (5.1)$$

ตารางที่ 14 ผลลัพธ์เปรียบเทียบการประเมินโดยมนุษย์ระหว่างงานวิจัยก่อนหน้าและงานวิจัยที่นำเสนอโดยมีเกณฑ์การประเมินเดียวกัน

แบบจำลอง	ความพึงพอใจ ของประโยค	มีความหมายและบริบทที่ เกี่ยวข้องเหมาะสมกับบทความ	สามารถหาคำตอบ ได้จากในบทความ
CopyLossEnhanced [20]	4.23	4.23	4.20
แบบจำลอง การสร้างคำถามภาษาไทย	4.40	4.65	4.70

บทที่ 6

สรุปผลการทดลอง ข้อเสนอแนะ และงานวิจัยในอนาคต

ในงานวิจัยนี้นำเสนอขั้นตอนการพัฒนาประสิทธิภาพของแบบจำลองการสร้างคำถามเพื่อให้ได้คำถามภาษาไทยที่มีคุณภาพภายใต้จำนวนข้อมูลฝึกสอนที่จำกัด จากการค้นคว้างานวิจัยที่เกี่ยวข้อง นำเสนอแนวคิดและขั้นตอนการดำเนินงานจนกระทั่งทดลอง พิสูจน์ให้เห็นว่าแนวคิดที่นำเสนอสามารถพัฒนาประสิทธิภาพแบบจำลองได้จากผลการทดลองดังนี้

6.1. สรุปผลการทดลอง

การประเมินแบบอัตโนมัติโดยคะแนนเบลอ แบบจำลองที่ฝึกสอนจากชุดข้อมูลสังเคราะห์รอบที่ 1 โดยมีรูปแบบข้อมูลนำเข้าปกติให้ผลลัพธ์ที่ดีกว่าแบบจำลองอื่น ๆ โดยวิเคราะห์ในแต่ละส่วนได้เป็นดังนี้

- เปรียบเทียบระหว่างชุดข้อมูลฝึกสอน จากผลการทดลองเปรียบเทียบระหว่างแบบจำลองที่ฝึกสอนจากชุดข้อมูลที่สร้างโดยมนุษย์กับแบบจำลองที่ฝึกสอนจากชุดข้อมูลสังเคราะห์พบว่าการที่ให้แบบจำลองเรียนรู้จากชุดข้อมูลที่มากขึ้นช่วยให้แบบจำลองมีประสิทธิภาพที่ดีขึ้น แต่เมื่อเทียบระหว่างแบบจำลองที่ฝึกสอนจากชุดข้อมูลสังเคราะห์รอบที่ 1 และ 2 จากชุดข้อมูลที่มากขึ้นแต่กลับได้ค่าคะแนนลดลง สามารถตั้งข้อสังเกตได้ว่าบทความที่ใช้ในการสร้างชุดข้อมูลสังเคราะห์และกระบวนการคัดกรองมีผลต่อประสิทธิภาพของแบบจำลองการสร้างคำถาม
- เปรียบเทียบระหว่างรูปแบบข้อมูลนำเข้า จากงานวิจัยก่อนหน้าที่มีการปรับรูปแบบข้อมูลนำเข้าเพื่อหาแบบนำเข้าข้อมูลที่ช่วยให้แบบจำลองมีประสิทธิภาพที่ดีที่สุด จึงทดสอบการนำ NER มาใช้ในงานวิจัยซึ่งได้คะแนนลดลงจากรูปแบบนำเข้าปกติ จึงตั้งข้อสังเกตได้ว่า NER ไม่ควรนำมาพร้อมใช้ในรูปแบบข้อมูลนำเข้าหรือควรมีขั้นตอนการนำมาใช้เพิ่มขึ้นจากที่งานวิจัยนี้ได้ทดลองเช่น การฝึกสอน NER จากชุดคำตอบ นอกจากนี้อาจเกิดจากกระบวนการทำป้ายกำกับพจนานุกรมหรือชื่อเฉพาะไม่สามารถสกัดช่วงคำที่ถูกต้องจากคลังโปรแกรม(library) ที่ใช้ในงานวิจัยนี้
- เปรียบเทียบคะแนนเบลอ จากผลการทดลองเปรียบเทียบคะแนนเบลอ ระหว่างแบบจำลองที่ฝึกสอนจากชุดข้อมูลสังเคราะห์รอบที่ 1 และ 2 พบว่าค่าของแบบจำลองที่ฝึกสอนจากชุดข้อมูลสังเคราะห์รอบที่ 1 มีคะแนนเบลอ 1 และ 2 สูงกว่า แต่แบบจำลองที่ฝึกสอนจากชุดข้อมูลสังเคราะห์รอบที่ 2 มีคะแนนเบลอ 3 และ 4 สูงกว่า อาจเกิดจากแบบจำลองที่ฝึกสอนจากชุดข้อมูลสังเคราะห์รอบที่ 2 มีการเรียนรู้จากชุดข้อมูลที่มากขึ้น ทำให้มีบางประโยคคำถามที่แบบจำลองพยายามสร้างประโยคที่ยาวและซับซ้อนขึ้น ซึ่งอัลกอริทึมที่ใช้คำนวณคะแนนเบลอมีการใช้ฟังก์ชันการเกลี่ย (smoothing function) สำหรับพิจารณาประโยคที่สั้นในคะแนนเบลอ 3 และ 4 ซึ่งเมื่อเทียบแล้วจึงอาจส่งผลให้แบบจำลองที่ฝึกสอนจากชุดข้อมูลสังเคราะห์รอบที่ 2 ให้ผลคะแนนเบลอ 3 และ 4 สูงกว่า

การประเมินผลโดยมนุษย์ได้กำหนดเกณฑ์ด้านความไม่ไพเราะ มีความหมายและบริบทที่เกี่ยวข้องเหมาะสมกับบทความและสามารถหาคำตอบได้จากบทความโดยใช้แบบจำลองที่มีคะแนน BLEU โดยรวมสูงที่สุดมาใช้ในการ

ประเมิน พบว่าโดยรวมแล้วแบบจำลองสามารถสร้างคำถามที่ทำให้มนุษย์เข้าใจได้สามารถหาคำตอบได้จากบทความ และมีความเกี่ยวข้องกับบทความ แต่แบบจำลองยังสร้างคำถามได้ไม่เพราะตามรูปประโยคคำถามที่มนุษย์ใช้กันใน ภาษาธรรมชาติ นอกจากนี้เมื่อพิจารณาผลการประเมินในแต่ละผู้ประเมินพบว่าในคำถามเดียวกันมีจุดที่พิจารณา และให้ความสำคัญต่างกัน อาจด้วยบริบทของบทความและดุลยพินิจของผู้ประเมินแม้ว่าจะมีการกำหนดเกณฑ์การ ประเมิน ขอบเขตและตัวอย่างแล้ว

6.2. ข้อเสนอแนะ

ในการพัฒนาประสิทธิภาพแบบจำลองการสร้างคำถามทำได้หลายวิธีจากงานวิจัยที่มีการเผยแพร่ โดย ในงานวิจัยนี้เน้นการพัฒนาแบบจำลองการสร้างคำถามโดยใช้แบบจำลองที่ไม่ซับซ้อนและนำเสนอขั้นตอนการ พัฒนาประสิทธิภาพแบบจำลองการสร้างคำถามจากการฝึกสอนที่มีจำนวนข้อมูลฝึกสอนอยู่อย่างจำกัดเมื่อ เปรียบเทียบกับชุดข้อมูลฝึกสอนในภาษาอื่น ผู้พัฒนาแบบจำลองการสร้างคำถามภาษาไทยอาจพิจารณาแนวทางอื่น ในการพัฒนาคุณภาพแบบจำลองการสร้างคำถามหรือพัฒนาเรื่องกระบวนการคัดกรองคำถาม นอกจากนี้ผู้พัฒนา พิจารณาและให้ความสำคัญกับขั้นตอนการทำความสะอาดข้อมูลเรื่องคำที่เขียนผิดหรือไม่สมบูรณ์ซึ่งส่งผลให้ แบบจำลองเรียนรู้และสร้างคำถามที่มีคำผิดเนื่องจากชุดข้อมูลฝึกสอนซึ่งมีผลต่อการสร้างคำถาม ในด้านการ ประเมินสามารถเพิ่มการประเมินอัตโนมัติในรูปแบบอื่นเพื่อเปรียบเทียบประสิทธิภาพแบบจำลองได้ชัดเจนยิ่งขึ้น สำหรับการประเมินโดยมนุษย์สามารถเพิ่มจำนวนผู้ประเมินและขอบเขตคุณวุฒิผู้ประเมินให้เหมาะสมต่อเกณฑ์ คุณภาพคำถามและจัดอบรมการประเมินให้เป็นมาตรฐานเดียวกัน

6.3. งานวิจัยในอนาคต

จากผลการทดลองที่เปรียบเทียบกับงานวิจัยที่ใกล้เคียงกันในภาษาอื่นเนื่องจากไม่มีงานวิจัยที่สร้างคำถามและ ประเมินอย่างแพร่หลายในภาษาไทย จึงส่งผลให้ไม่มีตัวเปรียบเทียบที่แน่ชัดในภาษาเดียวกัน อย่างไรก็ตามใน งานวิจัยนี้มีผลการเปรียบเทียบระหว่างแบบจำลองที่สร้างจากการฝึกสอนของชุดข้อมูลฝึกสอนโดยมนุษย์และชุด ข้อมูลสังเคราะห์ที่นำเสนอ ถือเป็นเกณฑ์พื้นฐานให้งานวิจัยเรื่องการสร้างคำถามในภาษาไทยสำหรับงานวิจัยใน อนาคต โดยผู้วิจัยสามารถนำเสนอแบบจำลองการสร้างคำถามโดยใช้การสร้างแบบจำลองด้วยวิธีอื่นหรือปรับปรุง วิธีการคัดกรองคำถามที่มีคุณภาพเพื่อให้ชุดข้อมูลสังเคราะห์มีคำถามที่มีคุณภาพมากขึ้น เปลี่ยนคลังโปรแกรมที่ใช้ ในสัปดาห์เฉพาะหรือชื่อเฉพาะ เพิ่มขั้นตอนการทำความสะอาดข้อมูลให้ละเอียดยิ่งขึ้น



ภาคผนวก

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY



ลำดับ	บทความ	คำถาม
1	<p>เจ้าหญิงนาเดจดาแห่งบัลแกเรีย 30 มกราคม ค.ศ. 1899 – 15 กุมภาพันธ์ ค.ศ. 1958 ทรงเป็นสมาชิกของราชวงศ์บัลแกเรียพระราชประวัติเจ้าหญิงนาเดจดาทรงมีพระชนมายุ 8 พรรษา ค.ศ. 1907พระนางทรงประสูติในเมืองไซเพีย ที่ทรงเป็นพระธิดาองค์สุดท้ายของพระเจ้าซาร์เฟอร์ดินานด์ที่ 1 แห่งบัลแกเรียและพระมเหสีองค์แรก เจ้าหญิงมารี หลุยส์แห่งปาร์มา ซึ่งได้สิ้นพระชนม์หลังจากทรงให้ประสูติกาลแก่พระนาง ร่วมกับพระเชษฐภคินี เจ้าหญิงยูโดเซียแห่งบัลแกเรีย พระนางได้รับการศึกษาภายใต้การดูแลของพระวิมาดาเจ้าหญิงเอลเนอร์ รอสแห่งคอสมทริกส์เจ้าหญิงนาเดจดาทรงได้อภิเษกสมรส เมื่อวันที่ 24 มกราคม ค.ศ. 1924 ที่ Bad Mergentheim เยอรมนี กับดยุก อัลเบิร์ต อ็อยเกนแห่งเวือร์ทเทิมแบร์ค พระราชโอรสองค์ที่สองของอัลเบิร์ต ดยุกแห่งเวือร์ทเทิมแบร์ค ทั้งสองพระองค์ทรงได้มีพระราชโอรสและพระธิดาอยู่ 5 พระองค์ ได้แก่ดยุก แฟร์ดินันท์ อ็อยเกนแห่งเวือร์ทเทิมแบร์ค ดัชเชส มากาเรตตา หลุยส์แห่งเวือร์ทเทิมแบร์ค อภิเษกสมรสในปี ค.ศ. 1970 กับ François Luce de Cheigny ดยุก อ็อยเกน อีเบอร์ฮาร์ดแห่งเวือร์ทเทิมแบร์ค อภิเษกสมรสในปี ค.ศ. 1962 กับอาร์ชดัชเชสเล็คซานดราแห่งออสเตรีย พระธิดาของเจ้าหญิงอิลีเยนาแห่งโรมาเนียดยุกอเล็กซานเดอร์ อ็อยเกน แห่งเวือร์ทเทิมแบร์ค ดัชเชส โซเฟีย โซเฟีย แห่งเวือร์ทเทิมแบร์ค อภิเษกสมรสในปี ค.ศ. 1969 กับ Antonio Manuel Róxo de RamosBandeira เจ้าหญิงนาเดจดาได้สิ้นพระชนม์ ซึ่งมีพระชนมายุได้เพียง 59 พรรษาในอุบัติเหตุที่การท เยอรมนี.</p>	<p>เจ้าหญิงนาเดจดาแห่งบัลแกเรียมีพระชนมายุกี่พรรษา</p>
2	<p>อิติมการศึกษาวัดภูที่มีขนาดเล็กมาก เช่น วัดภูที่อยู่ไกล วัตถุประสงค์อยู่สูง เป็นต้น ใช้เพียงแว่นขยายและเลนส์อื่นเดี่ยวสองดู เช่นเดียวกับการใช้แว่นขยายส่องดูลายมือ ช่วงปี พ.ศ. 2133 แซคารีเยส เจนเซน ช่างทำแว่นชาวดัตช์ ประดิษฐ์กล้องจุลทรรศน์ชนิดเลนส์ประกอบประกอบด้วยแว่นขยายสองอัน ต่อมา กาลิเลโอ กาลิเลอีได้สร้างแว่นขยายส่องดูสิ่งมีชีวิตเล็ก ๆ ในปี พ.ศ. 2208 รอเบิร์ต ฮุกได้ประดิษฐ์กล้องจุลทรรศน์ชนิดเลนส์ประกอบที่มีลำกล้องรูปร่างสวยงาม ป้องกันการรบกวนจากแสงภายนอกได้ และไม่ต้องถือเลนส์ให้ซ้อนกัน เขาส่องดูไม้คอร์กที่ฝานบาง ๆ แล้วพบช่องเล็ก ๆ มากมาย เขาเรียกช่องเหล่านั้นว่าเซลล์ ซึ่งหมายถึงห้องว่าง ๆ หรือห้องซึ่งเซลล์ที่ฮุกเห็นเป็นเซลล์ที่ตายแล้ว เหลือแต่ผนังเซลล์ของพืชซึ่งแข็งแรงกว่าเยื่อหุ้มเซลล์ในสัตว์ จึงทำให้คงรูปร่างอยู่ได้ ฮุกจึงได้ชื่อว่าเป็นผู้ที่ตั้งชื่อเซลล์ในปี พ.ศ. 2215 อันโตนิ ฟัน เลเวนฮุก ชาวดัตช์ สร้างกล้องจุลทรรศน์ชนิดเลนส์เดี่ยวจากแว่นขยายที่เขาตนเอง แวะนขยายบางอันขยายได้ถึง 270 เท่า เขาใช้กล้องจุลทรรศน์ตรวจดูหยดน้ำจากบึงและแม่น้ำ และจากน้ำที่กรองไว้หม้อ เห็นสิ่งมีชีวิตเล็ก ๆ มากมาย นอกจากนี้ เขายังส่องดูสิ่งมีชีวิตต่าง ๆ เช่น เม็ดเลือดแดง กล้ามเนื้อ เป็นต้น เมื่อเขาพบสิ่งเหล่านี้ เขารายงานไปยังราช</p>	<p>นักวิทยาศาสตร์ชาวดัตช์ชื่อใคร</p>

	<p>สมาคมแห่งกรุงลอนดอน จึงได้รับการยกย่องว่าเป็นผู้ประดิษฐ์กล้องจุลทรรศน์ พ.ศ. 2367 ดูโรเจอร์ นักพฤกษศาสตร์ชาวฝรั่งเศสศึกษาเนื้อเยื่อพืช และสัตว์พบว่าประกอบด้วยเซลล์ พ.ศ. 2376 โรเบิร์ต บราวน์ นักพฤกษศาสตร์ชาวอังกฤษ เป็นคนแรกที่พบว่าเซลล์และพืชมีนิวเคลียสเป็นก่อนกลุมๆ อยู่ภายในเซลล์ พ.ศ. 2378 นูก นะดีอาร์ท นักสัตวศาสตร์ชาวฝรั่งเศส ศึกษาจุลินทรีย์และสิ่งมีชีวิตอื่น ๆ พบว่าภายในประกอบด้วยของเหลวใส ๆ จึงเรียกว่า ซาร์โคเด ซึ่งเป็นภาษาฝรั่งเศสมาจากศัพท์กรีกว่า ซาร์ค ซึ่งแปลว่าเนื้อ พ.ศ. 2381 มัททิวส์ ชไลเดน นักพฤกษศาสตร์ชาวเยอรมัน ศึกษาเนื้อเยื่อพืชชนิดต่าง ๆ พบว่าพืชทุกชนิดประกอบด้วยเซลล์ พ.ศ. 2382 ชไลเดน และทีโอดอร์ ขวาน จึงร่วมกันตั้งทฤษฎีเซลล์ ซึ่งมีใจความสรุปได้ว่า สิ่งมีชีวิตทุกชนิดประกอบด้วยเซลล์และผลิตภัณฑ์จากเซลล์ พ.ศ. 2382 ปูร์กีญ นักสัตววิทยาชาวเชโกสโลวาเกีย ศึกษาไข่และตัวอ่อนของสัตว์ชนิดต่าง ๆ พบว่าภายในมีของเหลวใส เหนียว อ่อนนุ่มเป็นวุ้น เรียกว่าโปรทอปลาซึม ต่อจากนั้นมีนักวิทยาศาสตร์อีกมากมายทำการศึกษาร่วมกับเซลล์ด้วยกล้องจุลทรรศน์ชนิดเลนส์ประกอบ และได้พัฒนาให้ดียิ่งขึ้น จนกระทั่งปี พ.ศ. 2475 นักวิทยาศาสตร์ชาวเยอรมันคือแอนสท์ รัสกาและมิกซ์ นอลล์ ได้เปลี่ยนแปลงกระบวนการของกล้องจุลทรรศน์ที่ใช้แสงและเลนส์มาใช้ลำอิเล็กตรอน ทำให้เกิดกล้องจุลทรรศน์อิเล็กตรอนขึ้นในระยะต่อ ๆ มา ปัจจุบันมีกำลังขยายกว่า 5 แสนเท่า</p>	<p>เมืองหลวงของประเทศไทย ชื่ออะไร</p>
3	<p>ไปรอภา ไปรอภา เป็นเมืองหลวงและเมืองที่ใหญ่ที่สุดของประเทศไทยตามประวัติศาสตร์ ทางตะวันตกของประเทศไทย เช่นกัน ตั้งอยู่ชายหาดทางใต้ของเกาะชังเตียกู มีทิวทัศน์ที่สวยงามเป็นแห่งชาติ 4 แห่ง เป็นเมืองศูนย์กลางการค้า ส่งออกกาแกพอ้อย ผลไม้เมืองร้อน มีอุตสาหกรรมการประมง</p>	<p>จตุญ ธรรมศิลป์ เกิดวันที่เท่าไร</p>
4	<p>จตุญ ธรรมศิลป์ เกิดวันที่ 1 เมษายน พ.ศ. 2499 เป็นผู้กำกับ ค่ายผลิตละคร ดาราวิดีโอ ช่อง 7 สี และมีผลงานเรื่องแรกคือเรื่อง เจ้าหญิงแดงอ่อน และได้สร้างผลงานละครหลังข่าวไว้อีกมากมาย เช่น รากนครา ฟ้าใหม่ หลงเงาจันทร์ จิ้งจอกป่าก็แรง อากัง เสียวีตลกลอยแสงด้วยโรคมะเร็ง เมื่อช่วงเช้าวันที่ 11 พฤศจิกายน 2558 ที่โรงพยาบาลจุฬาลงกรณ์ ทรานส์คริปต์ ผลงานละครโทรทัศน์. สถานีโทรทัศน์สีทองที่พบของ 7 ปราสาทมืด ท้องทุ่งหงส์ทอง มณีรั้ว ตะวันชิงพลบ เมืองใต้พลั ภาพอาคารพจน์ แก้วสารพัดนึก ภูตแม่น้ำโขง นางทาส ศิระขมมาร ทวิภพ วิมานะพรวัว กระสือ ใ้อ...มาดา ภูตพิพิศาศ สายโลหิต กำกับร่วม แม่นาคพระนคร ญาติกา รัตนโกสินทร์ ปอบผีฟ้า การะเกด กิ่งไม้ พลังรัก ป่วงหงส์ แม่นาค หนุ่มทิพย์ ม่านบังใจ รากนครา เงาใจ ปิ่นไพร บอดีการ์ดสาว สาวน้อย รุ่งทิพย์ นະหน้าทอง มหาสง สาวน้อยในตะเกียงแก้ว ภาค 2 ฟ้าใหม่ นางสารสัมพันธ์ บ้านร้อยดอกไม้ เหมราช หลงเงาจันทร์ จิ้งจอกป่าก็แรง ยายหนูลูกพ่อ ปุโลมแผ้ว</p>	<p>จตุญ ธรรมศิลป์ เกิดวันที่เท่าไร</p>

	<p>ทรัพย์ สายยศวิฑิต คุป่วนอลวน มือนาง ตะวันยอแสง สวรรค์สร้าง บ้านเดดอกรัก อสุรน้อยในตะเกียงแก้ว อายูกรัก คือหัตถาครองพิภพ</p>	
5	<p>จังหวัดพะเยา เป็นจังหวัดในภาคเหนือตอนบน บริเวณที่ตั้งของตัวเมืองพะเยาในปัจจุบันอยู่ติดกับกว๊านพะเยา เดิมเป็นที่ตั้งของเมือง ภูมามยาว หรือ พะยาว ที่ก่อตั้งขึ้นเมื่อพุทธศตวรรษที่ 16 โดยกษัตริย์องค์แรกคือ พญาจอมธรรม ซึ่งเป็นราชบุตรองค์หนึ่งจากเมืองหริภุญไชยในเชียงใหม่ และพะยาวอีกหลายองค์ เช่น พญาเจือง วิจารณ์แห่งเผ่าไหลวโนพื้นที่ลุ่มแม่น้ำโขง และพญาจำเมืองซึ่งได้กระทำการสถาปนาเป็นไมตรีต่อกันกับพญามังรายแห่งนครพิงค์เชียงใหม่ และพญาร่วงรวมค่าแห่งสุโขทัย ภายหลังมีการเปลี่ยนแปลงอำนาจ และตกอยู่ภายใต้อิทธิพลของอาณาจักรล้านนา ในสมัยพญาคำฟู เมื่อถึงสมัยกรุงรัตนโกสินทร์ ใน พ.ศ. 2386 ตรงกับสมัยรัชกาลที่ 3 เมืองพะเยาถูกตั้งขึ้นใหม่พร้อมเมืองเชียงรายและเมืองงาวเพื่อเป็นเมืองหน้าด่านต่อสู้กับกองทัพพม่าที่ตั้งอยู่ที่เมืองเชียงแสน โดยให้เมืองพะเยาขึ้นตรงต่อ นครลำปาง และในท้ายที่สุดก่อนที่พะเยาจะถูกลกฐานะขึ้นเป็นจังหวัด พะเยาอยู่ใต้การปกครองของจังหวัดเชียงรายในฐานะ อำเภอพะเยา และเมื่อวันที่ 28 สิงหาคม พ.ศ. 2520 อำเภอพะเยาได้ยกฐานะขึ้นเป็น จังหวัดพะเยา นับเป็นจังหวัดที่ 72 ของประเทศไทยประวัติศาสตร์ ประวัติศาสตร์ พะเยาเป็นจังหวัดในกลุ่มล้านนาตะวันออกร่วมกับเชียงราย มีสำเนียงภาษาถิ่นเหนือเป็นเอกลักษณ์ต่างจากภาษาถิ่นอื่นในจังหวัดเชียงใหม่ นอกจากนี้ ความสัมพันธ์กับเชียงรายในแง่ของสำเนียงภาษาและวัฒนธรรมแล้ว พะเยาและเชียงรายยังมีความสัมพันธ์กันในลักษณะบ้านพี่เมืองน้องฉันทน์ศรีอยุธยา เนื่องจากในอดีต ราชวงศ์ที่ปกครองทั้งสองเมืองต่างสืบเชื้อสายมาจากราชวงศ์ลาวจาก หิริภุญไชย เชียงแสน ซึ่งได้แตกสายออกมาเป็นราชวงศ์สำคัญในดินแดนพะเยา เชียงแสนและล้านนาในยุคต่อมา เช่น ราชวงศ์เชียงแสนที่ปกครองเมืองเชียงราย ราชวงศ์เชียงแสนสายพญาเงี้ยวเมือง และราชวงศ์มังรายซึ่งปกครองเชียงใหม่เดิมจากบรรพบุรุษร่วมกับภูมิศาสตร์อาณาเขต ภูมิภาคเขต. จังหวัดพะเยามีเนื้อที่ประมาณ 6335 ตารางกิโลเมตร ตอนเหนือมีพื้นที่ติดกับจังหวัดเชียงราย ทิศตะวันออกเฉียงใต้ติดกับประเทศลาวและจังหวัดน่าน ส่วนทางตะวันตกติดกับจังหวัดลำปาง และทางด้านใต้ติดกับจังหวัดแพร่การแบ่งเขตการปกครองตั้งอยู่ลักษณะบริเวณการปกครอง. สัญลักษณ์ประจำจังหวัด. คำขวัญประจำจังหวัด กว๊านพะเยาแหล่งชีวิต ศักดิ์สิทธิ์พระเจ้าตนหลวง บวงสรวงพ่อขุนงำเมือง งามลือเลื่องดอยบุษราคัม ทรายพระจำจังหวัด รูปพระเจ้าตนหลวงวัดศรีโคมคำ ต้นไม้ประจำจังหวัด ต้นสารภี ดอกไม้ประจำจังหวัด ดอกสารภีรายนามผู้ว่าราชการจังหวัดหน่วยการปกครอง หน่วยงานปกครอง. การปกครองแบ่งออกเป็น 9 อำเภอ 68 ตำบล 779 หมู่บ้าน 2 เทศบาลเมือง 33 เทศบาลตำบล 36 องค์การบริหารส่วนตำบลการ</p>	<p>ใครคือผู้ก่อตั้งจังหวัดพะเยา</p>

	<p>ปกครองส่วนภูมิภาค การปกครองส่วนภูมิภาค. การปกครองส่วนท้องถิ่นการขนส่งทางถนน การขนส่ง. ทางถนน. โดยรถยนต์สามารถเดินทางได้หลายเส้นทาง เช่น 1. ใช้ทางหลวงหมายเลข 32 ผ่านจังหวัดพระนครศรีอยุธยา อำเภอ สิงห์บุรี นครสวรรค์ จากนั้นใช้ทางหลวงหมายเลข 1 ผ่านกำแพงเพชร ตาก ลำปาง แล้วเข้าสู่ตัวเมืองพะเยา ระยะทาง 969 กิโลเมตร 2. ใช้ทางหลวงหมายเลข 32 ผ่านจังหวัดพระนครศรีอยุธยา อำเภอ สิงห์บุรี นครสวรรค์ จากนั้นเลี้ยวขวาเข้าสู่ทางหลวงหมายเลข 117 จนถึงพิษณุโลก แยกซ้ายเข้าสู่ทางหลวงหมายเลข 12 ไปจนถึงสุโขทัย เลี้ยวขวาไปตามทางหลวงหมายเลข 101 ผ่านอำเภอเด่นชัย จังหวัดแพร่...</p>
--	---





ลำดับ	บทความ	คำถาม
1	<p>เจ้าหญิงนาเดจดาแห่งบัลแกเรีย 30 มกราคม ค.ศ. 1899 – 15 กุมภาพันธ์ ค.ศ. 1958 ทรงเป็นสมาชิกของราชวงศ์บัลแกเรียพระราชประวัติเจ้าหญิงนาเดจดาทรงมีพระชนมายุ 8 พรรษา ค.ศ. 1907พระนางทรงประสูติในเมืองโซเฟีย ที่ทรงเป็นพระธิดาองค์สุดท้ายของพระเจ้าซาร์เฟอร์ดินานด์ที่ 1 แห่งบัลแกเรียและพระมเหสีองค์แรก เจ้าหญิงมารี หลุยส์แห่งปาร์มา ซึ่งได้สิ้นพระชนม์หลังจากทรงให้ประสูติกาลแก่พระนาง ร่วมกับพระเชษฐภคินี เจ้าหญิงยูโดเซียแห่งบัลแกเรีย พระนางได้รับการศึกษาภายใต้การดูแลของพระวิมาดาเจ้าหญิงเอลเนอร์ รอสแห่งคอสมทริกส์เจ้าหญิงนาเดจดาทรงได้อภิเษกสมรส เมื่อวันที่ 24 มกราคม ค.ศ. 1924 ที่ Bad Mergentheim เยอรมนี กับดยุก อัลเบิร์ต อ็อยเกนแห่งเวือร์ทเทิมแบร์ค พระราชาเออร์สองค์ที่สองของอัลเบิร์ต ดยุกแห่งเวือร์ทเทิมแบร์ค ทั้งสองพระองค์ทรงได้มีพระราชโอรสและพระธิดาอยู่ 5 พระองค์ ได้แก่ดยุก แฟร์ดินันท์ อ็อยเกนแห่งเวือร์ทเทิมแบร์ค ดัชเชส มากาเรตตา หลุยส์แห่งเวือร์ทเทิมแบร์ค อภิเษกสมรสในปี ค.ศ. 1970 กับ François Luce de Cheigny ดยุก อ็อยเกน อีเบอร์ฮาร์ดแห่งเวือร์ทเทิมแบร์ค อภิเษกสมรสในปี ค.ศ. 1962 กับอาร์ชดัชเชสเล็กซานดราแห่งออสเตรีย พระธิดาของเจ้าหญิงอีเลินาแห่งโรมาเนียดยุกอเล็กซานเดอร์ อ็อยเกน แห่งเวือร์ทเทิมแบร์ค ดัชเชส โซเฟีย โซเฟีย แห่งเวือร์ทเทิมแบร์ค อภิเษกสมรสในปี ค.ศ. 1969 กับ Antonio Manuel Róxo de RamosBandeira เจ้าหญิงนาเดจดาได้สิ้นพระชนม์ ซึ่งมีพระชนมายุได้เพียง 59 พรรษาในอุบัติเหตุที่การท เยอรมนี.</p>	<p>เจ้าหญิงนาเดจดาแห่งบัลแกเรีย มีพระชนมายุกี่พรรษา</p>
2	<p>อิติมการศึกษาวัดุมุที่มีขนาดเล็กมาก เช่น วัดุมุที่อยู่ไกล วัดุมุที่อยู่สูง เป็นต้น ใช้เพียงแว่นขยายและเลนส์อั้นเดียวส่องดู เช่นเดียวกับการใช้แว่นขยายส่องดูลายมือ ช่วงปี พ.ศ. 2133 แซคาเรียส เจนเซน ชาวทักแวนชวาทต์ซ์ ประดิษฐ์กล้องจุลทรรศน์ชนิดเลนส์ประกอบประกอบด้วยแว่นขยายสองอัน ต่อมา กาลิเลโอ กาลิเลอีได้สร้างแว่นขยายส่องดูสิ่งมีชีวิตเล็ก ๆ ในปี พ.ศ. 2208 รอเบิร์ต ฮุกได้ประดิษฐ์กล้องจุลทรรศน์ชนิดเลนส์ประกอบที่มีลำกล้องรูปร่างสวยงาม ป้องกันการรบกวนจากแสงภายนอกได้ และไม่ต้องถือเลนส์ให้ซ้อนกัน เขาส่องดูไม้คอร์กที่ฝานบาง ๆ แล้วพบช่องเล็ก ๆ มากมาย เขาเรียกช่องเหล่านั้นว่าเซลล์ ซึ่งหมายถึงห้องว่าง ๆ หรือห้องซึ่งเซลล์ที่ฮุกเห็นเป็นเซลล์ที่ตายแล้ว เหลือแต่ผนังเซลล์ของพืชซึ่งแข็งแรงกว่าเยื่อหุ้มเซลล์ในสัตว์ จึงทำให้คงรูปร่างอยู่ได้ ฮุกจึงได้ชื่อว่าเป็นผู้ที่ตั้งชื่อเซลล์ในปี พ.ศ. 2215 อันโตนิ ฟัน เลเวนฮุก ชาวดัตช์ สร้างกล้องจุลทรรศน์ชนิดเลนส์เดียวจากแว่นขยายที่เขาตนเอง แวะนขยายบางอันขยายได้ถึง 270 เท่า เขาใช้กล้องจุลทรรศน์ตรวจดูหยดน้ำจากบึงและแม่น้ำ และจากน้ำฝนที่รองไว้ในหม้อ เห็นสิ่งมีชีวิตเล็ก ๆ มากมาย นอกจากนี้ เขายังส่องดูสิ่งมีชีวิตต่าง ๆ เช่น เม็ดเลือดแดง กล้ามเนื้อ เป็นต้น เมื่อเขาพบสิ่งเหล่านี้ เขารายงานไปยังราช</p>	<p>ใครคือผู้ประดิษฐ์กล้องจุลทรรศน์ชนิดเลนส์ประกอบ</p>

	<p>สมาคมแห่งกรุงลอนดอน จึงได้รับการยกย่องว่าเป็นผู้ประดิษฐ์กล้องจุลทรรศน์ พ.ศ. 2367 ดูโรเจอร์ นักพฤกษศาสตร์ชาวฝรั่งเศสศึกษาเนื้อเยื่อพืช และสัตว์พบว่าประกอบด้วยเซลล์ พ.ศ. 2376 โรเบิร์ต บราวน์ นักพฤกษศาสตร์ชาวอังกฤษ เป็นคนแรกที่พบว่าเซลล์และพืชมีนิวเคลียสเป็นก่อนกลมๆ อยู่ภายในเซลล์ พ.ศ. 2378 นูก นะดีอาร์ทเจอร์ นักสัตวศาสตร์ชาวฝรั่งเศส ศึกษาจุลินทรีย์และสิ่งมีชีวิตอื่น ๆ พบว่าภายในประกอบด้วยของเหลวใส ๆ จึงเรียกว่า ซาร์โคโคต ซึ่งเป็นภาษาฝรั่งเศสมาจากศัพท์กรีกว่า ซาร์ค ซึ่งแปลว่าเนื้อ พ.ศ. 2381 มัททิวส์ ชไลเดน นักพฤกษศาสตร์ชาวเยอรมัน ศึกษาเนื้อเยื่อพืชชนิดต่าง ๆ พบว่าพืชทุกชนิดประกอบด้วยเซลล์ พ.ศ. 2382 ชไลเดน และทีโอดอร์ ชวาน จึงร่วมกันตั้งทฤษฎีเซลล์ ซึ่งมีใจความสรุปได้ว่า สิ่งมีชีวิตทุกชนิดประกอบด้วยเซลล์และผลิตภัณฑ์จากเซลล์ พ.ศ. 2382 ปูร์กีญ นักสัตววิทยาชาวเชโกสโลวาเกีย ศึกษาไข่และตัวอ่อนของสัตว์ชนิดต่าง ๆ พบว่าภายในมีของเหลวใส เหนียว อ่อนนุ่มเป็นวุ้น เรียกว่าโปรทอปลาซึ่งต่อมาในมหาวิทยาลัยอิมมัญญูเรียศึกษาเกี่ยวกับเซลล์ด้วยกล้องจุลทรรศน์ชนิดเลนส์ประกอบ และได้พัฒนาให้ดียิ่งขึ้น จนกระทั่งปี พ.ศ. 2475 นักวิทยาศาสตร์ชาวเยอรมันคือเอนสท์ รัสกาและมิกซ์ นอลล์ ได้เปลี่ยนแปลงกระบวนการของกล้องจุลทรรศน์ที่ใช้แสงและเลนส์มาใช้ลำอิเล็กตรอน ทำให้เกิดกล้องจุลทรรศน์อิเล็กตรอนขึ้นในระยะต่อ ๆ มา ปัจจุบันมีกำลังขยายกว่า 5 แสนเท่า</p>	<p>เมืองหลวงของประเทศคาบูเวร์ดีมีชื่อว่าอะไร</p> <p>จตุรรมศิลป์ เกิดวันที่เท่าไร</p>
3	<p>ไปรอกา ไปรอกา เป็นเมืองหลวงและเมืองที่ใหญ่ที่สุดของประเทศคาบูเวร์ดี ประเทศในมหาสมุทรแอตแลนติก ทางตะวันตกของประเทศ เซเนกัล ตั้งอยู่ชายหาดทางใต้ของเกาะเซเนกัล มีท่าเรือเฟอร์รีและมีสนามบินแห่งชาติ 4 แห่ง เป็นเมืองศูนย์กลางการค้า ส่งออกกาแฟ อ้อย ผลไม้เมืองร้อน มีอุตสาหกรรมการประมง</p>	<p>จตุรรมศิลป์ เกิดวันที่เท่าไร</p>
4	<p>จตุรรมศิลป์ จตุรรมศิลป์ เกิดวันที่ 1 เมษายน พ.ศ. 2499 เป็นผู้กำกับ ค่ายผลิตละคร ดาราทีวีไอ ช่อง 7 สี และมีผลงานเรื่องแรกคือเรื่อง เจ้าหญิงแดงอ่อน และได้สร้างผลงานละครหลังข่าวไว้อีกมากมาย เช่น รากนครา ฟ้าใหม่ หลงเงาจันทร์ จิ้งจอกป่าที่แรง อากัง เสียวีตลกลอยแสงด้วยโรคมะเร็ง เมื่อช่วงเช้าวันที่ 11 พฤศจิกายน 2558 ที่โรงพยาบาลจุฬาลงกรณ์และศิริราช ศาลา ๓ อาคารโทรทัศน์. สถานีโทรทัศน์สีกองทัพบกช่อง 7 ปราสาทมิต ท้องทุ่งน พงษ์ทอง มณีรั้ว ตะวันชิงพลบ เมืองใต้พลี ภาพอาคารพจน์ แก้วสารพัดนึก ภูตแม่น้ำโขง นางทาส ศีระขมมาร ทวิภพ วิมานะพรวัว กระสือ ใ้อ...มาดา ภูตพิพิศาศ สายโลหิต กำกับร่วม แม่มาดพระนคร ญาติกา รัตนโกสินทร์ ปอบผีฟ้า การะเกด กิ่งไม้ พลังรัก ปวงหงส์ แม่นาค หม่อมทิพย์ ม่านบังใจ รากนครา เงามใจ ปิ่นไพร บอดีการ์ดสาว สาวน้อย รุ่งทิพย์ นะหน้าทอง มหาหงส์ สาวน้อยในตะเกียง แก้ว ภาค 2 ฟ้าใหม่ นางสารสัมพันธ์ บ้านร้อยดอกไม้ เหมราช หลงเงาจันทร์ จิ้งจอกป่าที่แรง ยายหนูลูกพ่อ ปุ้เสมผ้า</p>	<p>จตุรรมศิลป์ เกิดวันที่เท่าไร</p>

	<p>ปกครองส่วนภูมิภาค การปกครองส่วนภูมิภาค. การปกครองส่วนท้องถิ่นการขนส่งทางถนน การขนส่ง. ทางถนน. โดยรถยนต์สามารถเดินทางได้หลายเส้นทาง เช่น 1. ใช้ทางหลวงหมายเลข 32 ผ่านจังหวัดพระนครศรีอยุธยา อำเภอ สิงห์บุรี นครสวรรค์ จากนั้นใช้ทางหลวงหมายเลข 1 ผ่านกำแพงเพชร ตาก ลำปาง แล้วเข้าสู่ตัวเมืองพะเยา ระยะทาง 969 กิโลเมตร 2. ใช้ทางหลวงหมายเลข 32 ผ่านจังหวัดพระนครศรีอยุธยา อำเภอ สิงห์บุรี นครสวรรค์ จากนั้นเลี้ยวขวาเข้าสู่ทางหลวงหมายเลข 117 จนถึงพิษณุโลก แยกซ้ายเข้าสู่ทางหลวงหมายเลข 12 ไปจนถึงสุโขทัย เลี้ยวขวาไปตามทางหลวงหมายเลข 101 ผ่านอำเภอเด่นชัย จังหวัดแพร่...</p>
--	---





ภาคผนวก ค.

ตัวอย่างผลลัพธ์การสร้างคำถามจาก

แบบจำลองเชิงสัญลักษณ์จากชุดข้อมูลสังเคราะห์รอบที่ 1 โดยมีรูปแบบข้อมูลนำเข้าปกติ

ลำดับ	บทความ	คำถาม
1	<p>เจ้าหญิงนาเดจดาแห่งบัลแกเรีย 30 มกราคม ค.ศ. 1899 – 15 กุมภาพันธ์ ค.ศ. 1958 ทรงเป็นสมาชิกของราชวงศ์บัลแกเรียพระราชประวัติเจ้าหญิงนาเดจดาทรงมีพระชนมายุ 8 พรรษา ค.ศ. 1907พระนางทรงประสูติในเมืองไซเพีย ที่ทรงเป็นพระธิดาองค์สุดท้ายของพระเจ้าซาร์เฟอร์ดินานด์ที่ 1 แห่งบัลแกเรียและพระมเหสีองค์แรก เจ้าหญิงมารี หลุยส์แห่งปาร์มา ซึ่งได้สิ้นพระชนม์หลังจากทรงให้ประสูติกาลแก่พระนาง ร่วมกับพระเชษฐภคินี เจ้าหญิงยูโดเซียแห่งบัลแกเรีย พระนางได้รับการศึกษาภายใต้การดูแลของพระวิมาดาเจ้าหญิงเอลเนอร์ รอสแห่งคอสมทริกส์เจ้าหญิงนาเดจดาทรงได้อภิเษกสมรส เมื่อวันที่ 24 มกราคม ค.ศ. 1924 ที่ Bad Mergentheim เยอรมนี กับดยุก อัลเบิร์ต อ็อยเกนแห่งเวือร์ทเทิมแบร์ค พระราชาเออร์สองค์ที่สองของอัลเบิร์ต ดยุกแห่งเวือร์ทเทิมแบร์ค ทั้งสองพระองค์ทรงได้มีพระราชโอรสและพระธิดาอยู่ 5 พระองค์ ได้แก่ดยุก แฟร์ดินันท์ อ็อยเกนแห่งเวือร์ทเทิมแบร์ค ดัชเชส มาการิตตา หลุยส์แห่งเวือร์ทเทิมแบร์ค อภิเษกสมรสในปี ค.ศ. 1970 กับ François Luce de Cheigny ดยุก อ็อยเกน อีเบอร์ฮาร์ดแห่งเวือร์ทเทิมแบร์ค อภิเษกสมรสในปี ค.ศ. 1962 กับอาร์ชดัชเชสเล็คซานดราแห่งออสเตรีย พระธิดาของเจ้าหญิงอิลีเยนาแห่งโรมาเนียดยุกอเล็กซานเดอร์ อ็อยเกน แห่งเวือร์ทเทิมแบร์ค ดัชเชส โซเฟีย แห่งเวือร์ทเทิมแบร์ค อภิเษกสมรสในปี ค.ศ. 1969 กับ Antonio Manuel Róxo de RamosBandeira เจ้าหญิงนาเดจดาได้สิ้นพระชนม์ ซึ่งมีพระชนมายุได้เพียง 59 พรรษาในอุบัติเหตุที่การท เยอรมนี.</p>	<p>เจ้าหญิงนาเดจดาแห่งบัลแกเรียทรงได้รับการศึกษาภายใต้การดูแลของใคร</p>
2	<p>อิติมการศึกษาวัดภูที่มีขนาดเล็กมาก เช่น วัดภูที่อยู่ไกล วัตถุประสงค์อยู่สูง เป็นต้น ใช้เพียงแว่นขยายและเลนส์อื่นเดี่ยวสองดู เช่นเดียวกับการใช้แว่นขยายส่องดูลายมือ ช่วงปี พ.ศ. 2133 แซคาเรียส แจนเซน ช่างทำแว่นชาวดัตช์ ประดิษฐ์กล้องจุลทรรศน์ชนิดเลนส์ประกอบประกอบด้วยแว่นขยายสองอัน ต่อมา กาลิเลโอ กาลิเลอีได้สร้างแว่นขยายส่องดูสิ่งมีชีวิตเล็ก ๆ ในปี พ.ศ. 2208 รอเบิร์ต ฮุกได้ประดิษฐ์กล้องจุลทรรศน์ชนิดเลนส์ประกอบที่มีลำกล้องรูปร่างสวยงาม ป้องกันการรบกวนจากแสงภายนอกได้ และไม่ต้องถือเลนส์ให้ซ้อนกัน เขาส่องดูไม้คอร์กที่ฝานบาง ๆ แล้วพบช่องเล็ก ๆ มากมาย เขาเรียกช่องเหล่านั้นว่าเซลล์ ซึ่งหมายถึงห้องว่าง ๆ หรือห้องซึ่งเซลล์ที่ฮุกเห็นเป็นเซลล์ที่ตายแล้ว เหลือแต่ผนังเซลล์ของพืชซึ่งแข็งแรงกว่าเยื่อหุ้มเซลล์ในสัตว์ จึงทำให้คงรูปร่างอยู่ได้ ฮุกจึงได้ชื่อว่าเป็นผู้ที่ตีั้งชื่อเซลล์ในปี พ.ศ. 2215 อันโตนิ ฟัน เลเวนฮุก ชาวดัตช์ สร้างกล้องจุลทรรศน์ชนิดเลนส์เดี่ยวจากแว่นขยายที่เขาตนเอง แวะนขยายบางอันขยายได้ถึง 270 เท่า เขาใช้กล้องจุลทรรศน์ตรวจดูหยดน้ำจากบึงและแม่น้ำ และจากน้ำฝนที่รองไว้ในหม้อ เห็นสิ่งมีชีวิตเล็ก ๆ มากมาย นอกจากนี้ เขายังส่องดูสิ่งมีชีวิตต่าง ๆ เช่น เม็ดเลือดแดง กล้ามเนื้อ เป็นต้น เมื่อเขาพบสิ่งเหล่านี้ เขารายงานไปยังราช</p>	<p>กล้องจุลทรรศน์ชนิดเลนส์ประกอบมีขนาดเท่าไร</p>

	<p>สมาคมแห่งกรุงลอนดอน จึงได้รับการยกย่องว่าเป็นผู้ประดิษฐ์กล้องจุลทรรศน์ พ.ศ. 2367 ดูโรเจอร์ นักพฤกษศาสตร์ชาวฝรั่งเศสศึกษาเนื้อเยื่อพืช และสัตว์พบว่าประกอบด้วยเซลล์ พ.ศ. 2376 โรเบิร์ต บราวน์ นักพฤกษศาสตร์ชาวอังกฤษ เป็นคนแรกที่พบว่าเซลล์และพืชมีนิวเคลียสเป็นก่อนกลมๆ อยู่ภายในเซลล์ พ.ศ. 2378 นูก นะดีอาร์ท นักสัตวศาสตร์ชาวฝรั่งเศส ศึกษาจุลินทรีย์และสิ่งมีชีวิตอื่น ๆ พบว่าภายในประกอบด้วยของเหลวใด ๆ จึงเรียกว่า ซาร์โคเด ซึ่งเป็นภาษาฝรั่งเศสมาจากศัพท์กรีกว่า ซาร์ค ซึ่งแปลว่าเนื้อ พ.ศ. 2381 มัททิวส์ ชไลเดน นักพฤกษศาสตร์ชาวเยอรมัน ศึกษาเนื้อเยื่อพืชชนิดต่าง ๆ พบว่าพืชทุกชนิดประกอบด้วยเซลล์ พ.ศ. 2382 ชไลเดน และทีโอดอร์ ขวาน จึงร่วมกันตั้งทฤษฎีเซลล์ ซึ่งมีใจความสรุปได้ว่า สิ่งมีชีวิตทุกชนิดประกอบด้วยเซลล์และผลิตภัณฑ์จากเซลล์ พ.ศ. 2382 ปูร์กีญ นักสัตววิทยาชาวเชโกสโลวาเกีย ศึกษาไข่และตัวอ่อนของสัตว์ชนิดต่าง ๆ พบว่าภายในมีของเหลวใส เหนียว อ่อนนุ่มเป็นวุ้น เรียกว่าโปรทอปลาซึ่งต่อมาในมีนิกวิทยาศาสตร์อีกมากมายทำการศึกษาร่วมกับเซลล์ด้วยกล้องจุลทรรศน์ชนิดเลนส์ประกอบ และได้พัฒนาให้ดียิ่งขึ้น จนกระทั่งปี พ.ศ. 2475 นักวิทยาศาสตร์ชาวเยอรมันคือแอนสท์ รัสกาและมิกซ์ นอลล์ ได้เปลี่ยนแปลงกระบวนการของกล้องจุลทรรศน์ที่ใช้แสงและเลนส์มาใช้ลำเล็กลึกลง ทำให้เกิดกล้องจุลทรรศน์อิเล็กตรอนขึ้นในระยะต่อ ๆ มา ปัจจุบันมีกำลังขยายกว่า 5 แสนเท่า</p>	<p>เมืองหลวงของประเทศกาบูเวร์ดี ประเทศโต</p>
3	<p>ไปรอกา ไปรอกา เป็นเมืองหลวงและเมืองที่ใหญ่ที่สุดของประเทศกาบูเวร์ดี ประเทศในมหาสมุทรแอตแลนติก ทางตะวันตกของประเทศ เซเนกัล ตั้งอยู่ชายหาดทางใต้ของเกาะเซเนดิยู มีท่าเรือเฟอร์รี่และมีสนามบินแห่งชาติ 4 แห่ง เป็นเมืองศูนย์กลางการค้า ส่งออกกาแฟ อ้อย ผลไม้เมืองร้อน มีอุตสาหกรรมการประมง</p>	<p>เมืองหลวงของประเทศกาบูเวร์ดี ประเทศโต</p>
4	<p>จตุญ ธรรมศิลป์ จตุญ ธรรมศิลป์ เกิดวันที่ 1 เมษายน พ.ศ. 2499 เป็นผู้กำกับ ค่ายผลิตละคร ดาราวิดีโอ ช่อง 7 สี และมีผลงานเรื่องแรกคือเรื่อง เจ้าหญิงแดงอ่อน และได้สร้างผลงานละครหลังข่าวไว้อีกมากมาย เช่น รากนครา ฟ้าใหม่ หลงเงาจันทร์ จิ้งจอกป่าก็แรง อาแกง เสียชีวิตลงอย่างสงบด้วยโรคมะเร็ง เมื่อช่วงเช้าวันที่ 11 พฤศจิกายน 2558 ที่โรงพยาบาลจุฬาลงกรณ์และศิริราช กรุงเทพมหานคร สถานที่โทรทัศน์ สภาวัฒนธรรมจังหวัดภูเก็ต หอประชุม หงส์ทอง มณีรัว ตะวันชิงพลบ เมืองใต้พลี ภาพอาคารพจน์ แก้วสารพัดนึก ภูตแม่น้ำโจง นางทาส ศิระขะมาร ทวิภพ วิมานะพรวัว กระสือ ไร่...มาดา ภูตพิพิศาส สายโลหิต กำกับร่วม แม่มาดพระนคร ญาติกา รัตนโกสินทร์ ปอบผีฟ้า การะเกด กิ่งไม้ พลึงรัก ป่างหงส์ แม่นาค หน่มทิพย์ ม่านบังใจ รากนครา เงาใจ ปิ่นไพร บอดีการ์ดสาว สาวน้อย รุ่งทิพย์ นะหน้าทอง มหาหงส์ สาวน้อยในตะเกียง แก้ว ภาค 2 ฟ้าใหม่ นางสารสัมพันธ์ บ้านร้อยดอกไม้ เหมราช หลงเงาจันทร์ จิ้งจอกป่าก็แรง ยายหนูลูกพ่อ ปุโลมแผ้ว</p>	<p>จตุญ ธรรมศิลป์ มีผลงานเรื่องแรกคือเรื่องอะไร</p>

	<p>ปกครองส่วนภูมิภาค การปกครองส่วนภูมิภาค. การปกครองส่วนท้องถิ่นการขนส่งทางถนน การขนส่ง. ทางถนน. โดยรถยนต์สามารถเดินทางได้หลายเส้นทาง เช่น 1. ใช้ทางหลวงหมายเลข 32 ผ่านจังหวัดพระนครศรีอยุธยา อำเภอ สิงห์บุรี นครสวรรค์ จากนั้นใช้ทางหลวงหมายเลข 1 ผ่านกำแพงเพชร ตาก ลำปาง แล้วเข้าสู่ตัวเมืองพะเยา ระยะทาง 969 กิโลเมตร 2. ใช้ทางหลวงหมายเลข 32 ผ่านจังหวัดพระนครศรีอยุธยา อำเภอ สิงห์บุรี นครสวรรค์ จากนั้นเลี้ยวขวาเข้าสู่ทางหลวงหมายเลข 117 จนถึงพิษณุโลก แยกซ้ายเข้าสู่ทางหลวงหมายเลข 12 ไปจนถึงสุโขทัย เลี้ยวขวาไปตามทางหลวงหมายเลข 101 ผ่านอำเภอเด่นชัย จังหวัดแพร่...</p>
--	---





ภาคผนวก ง.

ตัวอย่างผลลัพธ์การสร้างคำถามจาก

แบบจำลองฝึกสอนจากชุดข้อมูลสังเคราะห์รอบที่ 1 โดยมีรูปแบบข้อมูลนำเข้า NER

ลำดับ	บทความ	คำถาม
1	<p>เจ้าหญิงนาเดจดาแห่งบัลแกเรีย 30 มกราคม ค.ศ. 1899 – 15 กุมภาพันธ์ ค.ศ. 1958 ทรงเป็นสมาชิกของราชวงศ์บัลแกเรียพระราชประวัติเจ้าหญิงนาเดจดาทรงมีพระชนมายุ 8 พรรษา ค.ศ. 1907พระนางทรงประสูติในเมืองโซเฟีย ที่ทรงเป็นพระธิดาองค์สุดท้ายของพระเจ้าซาร์เฟอร์ดินานด์ที่ 1 แห่งบัลแกเรียและพระมเหสีองค์แรก เจ้าหญิงมารี หลุยส์แห่งปาร์มา ซึ่งได้สิ้นพระชนม์หลังจากทรงให้ประสูติกาลแก่พระนาง ร่วมกับพระเชษฐภคินี เจ้าหญิงยูโดเซียแห่งบัลแกเรีย พระนางได้รับการศึกษาภายใต้การดูแลของพระวิมาดาเจ้าหญิงเอลเนอร์ รอสแห่งคอสมทริกส์เจ้าหญิงนาเดจดาทรงได้อภิเษกสมรส เมื่อวันที่ 24 มกราคม ค.ศ. 1924 ที่ Bad Mergentheim เยอรมนี กับดยุก อัลเบิร์ต อ็อยเกนแห่งเวือร์ทเทิมแบร์ค พระราชาโอรสองค์ที่สองของอัลเบิร์ต ดยุกแห่งเวือร์ทเทิมแบร์ค ทั้งสองพระองค์ทรงได้มีพระราชโอรสและพระธิดาอยู่ 5 พระองค์ ได้แก่ดยุก แฟร์ดินันท์ อ็อยเกนแห่งเวือร์ทเทิมแบร์ค ดัชเชส มากาเรตตา หลุยส์แห่งเวือร์ทเทิมแบร์ค อภิเษกสมรสในปี ค.ศ. 1970 กับ François Luce de Cheigny ดยุก อ็อยเกน อีเบอร์ฮาร์ดแห่งเวือร์ทเทิมแบร์ค อภิเษกสมรสในปี ค.ศ. 1962 กับอาร์ชดัชเชสเล็กลีซานดราแห่งออสเตรีย พระธิดาของเจ้าหญิงอีเลินาแห่งโรมาเนียดยุกอเล็กซานเดอร์ อ็อยเกน แห่งเวือร์ทเทิมแบร์ค ดัชเชส โซเฟีย แห่งเวือร์ทเทิมแบร์ค อภิเษกสมรสในปี ค.ศ. 1969 กับ Antonio Manuel Róxo de RamosBandeira เจ้าหญิงนาเดจดาได้สิ้นพระชนม์ ซึ่งมีพระชนมายุได้เพียง 59 พรรษาในอุบัติเหตุที่การท เยอรมนี.</p>	เจ้าหญิงนาเดจดาได้อภิเษกสมรสกับใคร
2	<p>อดีตการศึกษาวิศวะที่มีขนาดเล็กมาก เช่น วัตถุที่อยู่ไกล วัตถุที่อยู่สูง เป็นต้น ใช้เพียงแว่นขยายและเลนส์อื่นเดี่ยวสองดู เช่นเดียวกับการใช้แว่นขยายส่องดูลายมือ ช่วงปี พ.ศ. 2133 แซคาเรียส แจนเซน ชาวที่กแว่นชาวดัตช์ ประดิษฐ์กล้องจุลทรรศน์ชนิดเลนส์ประกอบประกอบด้วยแว่นขยายสองอัน ต่อมา กาลิเลโอ กาลิเลอีได้สร้างแว่นขยายส่องดูสิ่งมีชีวิตเล็ก ๆ ในปี พ.ศ. 2208 รอเบิร์ต ฮุกได้ประดิษฐ์กล้องจุลทรรศน์ชนิดเลนส์ประกอบที่มีลำกล้องรูปร่างสวยงาม ป้องกันการรบกวนจากแสงภายนอกได้ และไม่ต้องถือเลนส์ให้ซ้อนกัน เขาส่องดูไม้คอร์กที่ฝานบาง ๆ แล้วพบช่องเล็ก ๆ มากมาย เขาเรียกช่องเหล่านั้นว่าเซลล์ ซึ่งหมายถึงห้องว่าง ๆ หรือห้องซึ่งเซลล์ที่ฮุกเห็นเป็นเซลล์ที่ตายแล้ว เหลือแต่ผนังเซลล์ของพืชซึ่งแข็งแรงกว่าเยื่อหุ้มเซลล์ในสัตว์ จึงทำให้คงรูปร่างอยู่ได้ ฮุกจึงได้ชื่อว่าเป็นผู้ที่ตั้งชื่อเซลล์ในปี พ.ศ. 2215 อันโตนิ ฟัน เลเวนฮุก ชาวดัตช์ สร้างกล้องจุลทรรศน์ชนิดเลนส์เดี่ยวจากแว่นขยายที่เขาตนเอง แวะนขยายบางอันขยายได้ถึง 270 เท่า เขาใช้กล้องจุลทรรศน์ตรวจดูหยดน้ำจากบึงและแม่น้ำ และจากน้ำฝนที่รองไว้ในหม้อ เห็นสิ่งมีชีวิตเล็ก ๆ มากมาย นอกจากนี้ เขายังส่องดูสิ่งมีชีวิตต่าง ๆ เช่น เม็ดเลือดแดง กล้ามเนื้อ เป็นต้น เมื่อเขาพบสิ่งเหล่านี้ เขารายงานไปยังราช</p>	นักวิทยาศาสตร์ชาวเยอรมันคือใคร

	<p>สมาคมแห่งกรุงลอนดอน จึงได้รับการยกย่องว่าเป็นผู้ประดิษฐ์กล้องจุลทรรศน์ พ.ศ. 2367 ดูโรเจอร์ นักพฤกษศาสตร์ชาวฝรั่งเศสศึกษาเนื้อเยื่อพืช และสัตว์พบว่าประกอบด้วยเซลล์ พ.ศ. 2376 โรเบิร์ต บราวน์ นักพฤกษศาสตร์ชาวอังกฤษ เป็นคนแรกที่พบว่าเซลล์และพืชมีนิวเคลียสเป็นก่อนกลุมๆ อยู่ภายในเซลล์ พ.ศ. 2378 นูก นะดีอาร์ท นักสัตวศาสตร์ชาวฝรั่งเศส ศึกษาจุลินทรีย์และสิ่งมีชีวิตอื่น ๆ พบว่าภายในประกอบด้วยของเหลวใส ๆ จึงเรียกว่า ซาร์โคเด ซึ่งเป็นภาษาฝรั่งเศสมาจากศัพท์กรีกว่า ซาร์ค ซึ่งแปลว่าเนื้อ พ.ศ. 2381 มัททิวส์ ชไลเดน นักพฤกษศาสตร์ชาวเยอรมัน ศึกษาเนื้อเยื่อพืชชนิดต่าง ๆ พบว่าพืชทุกชนิดประกอบด้วยเซลล์ พ.ศ. 2382 ชไลเดน และทีโอดอร์ ชวาน จึงร่วมกันตั้งทฤษฎีเซลล์ ซึ่งมีใจความสรุปได้ว่า สิ่งมีชีวิตทุกชนิดประกอบด้วยเซลล์และผลิตภัณฑ์จากเซลล์ พ.ศ. 2382 ปูร์กีญ นักสัตววิทยาชาวเชโกสโลวาเกีย ศึกษาไข่และตัวอ่อนของสัตว์ชนิดต่าง ๆ พบว่าภายในมีของเหลวใส เหนียว อ่อนนุ่มเป็นวุ้น เรียกว่าโปรทอปลาซึมต่อจากนั้นมีนักวิทยาศาสตร์อีกมากมายทำการศึกษาร่วมกันเกี่ยวกับเซลล์ด้วยกล้องจุลทรรศน์ชนิดเลนส์ประกอบ และได้พัฒนาให้ดียิ่งขึ้น จนกระทั่งปี พ.ศ. 2475 นักวิทยาศาสตร์ชาวเยอรมันคือแอนสท์ รัสกาและมิกซ์ นอลล์ ได้เปลี่ยนแปลงกระบวนการของกล้องจุลทรรศน์ที่ใช้แสงและเลนส์มาใช้ลำอิเล็กตรอน ทำให้เกิดกล้องจุลทรรศน์อิเล็กตรอนขึ้นในระยะต่อ ๆ มา ปัจจุบันมีกำลังขยายกว่า 5 แสนเท่า</p>	เมืองหลวงของประเทศคาบูเวร์ดี มีชื่อว่า อะไร
3	ไปรอกา ไปรอกา เป็นเมืองหลวงและเมืองที่ใหญ่ที่สุดของประเทศคาบูเวร์ดี ประเทศในมหาสมุทรแอตแลนติก ทางตะวันตกของประเทศ เซเนกัล ตั้งอยู่ชายหาดทางใต้ของเกาะเซเนกัล มีท่าเรือเฟอร์รี่และมีสนามบินแห่งชาติ 4 แห่ง เป็นเมืองศูนย์กลางการค้า ส่งออกกาแฟ อ้อย ผลไม้เมืองร้อน มีอุตสาหกรรมการประมง	
4	<p>จตุญจ ธรรมศิลป์ จตุญจ ธรรมศิลป์ เกิดวันที่ 1 เมษายน พ.ศ. 2499 เป็นผู้กำกับ ค่ายผลิตละคร ดาราวิดีโอ ช่อง 7 สี และมีผลงานเรื่องแรกคือเรื่อง เจ้าหญิงแดงอ่อน และได้สร้างผลงานละครหลังข่าวไว้อีกมากมาย เช่น รากนครา ฟ้าใหม่ หลงเงาจันทร์ จิ้งจอกป่าแกร่ง อาแกง เสียดชีวิตลงอย่างสวยงามด้วยโรคมะเร็ง เมื่อช่วงเช้าวันที่ 11 พฤศจิกายน 2558 ที่โรงพยาบาลจุฬาลงกรณ์และศิริราช กรุงเทพมหานคร</p> <p>ละครโทรทัศน์: สถานีโทรทัศน์สีกองทัพบกช่อง 7 ปราสาทมิด ท้อพูน พงษ์ทอง มณีรั้ว ตะวันชิงพลบ เมืองโศกโศก ภาพอาคารพจน์ แก้วสารพัดนึก ภูตแม่น้ำโขง นางพาส ศีระชะมาร ทวิภพ วิมานมะพร้าว กระสือ ไร่...มาดา ภูตพิศواس สายโลหิต กำกับร่วม แม่มาดพระนคร ญาติกา รัตนโกสินทร์ ปอบผีฟ้า การะเกด กิ่งไม้ พลังรัก ปวงหงส์ แม่นาค หนุ่มทิพย์ ม่านบังใจ รากนครา เงามใจ ปิ่นไพร บอดีการ์ดสาว สาวน้อย รุ่งทิพย์ นะหน้าทอง มหาหงส์ สาวน้อยในตะเกียง แก้ว ภาค 2 ฟ้าใหม่ นางสารสัมพันธ์ บ้านร้อยดอกไม้ เหมราช หลงเงาจันทร์ จิ้งจอกป่าแกร่ง ยายหนูลูกพ่อ ปุโลมแผ้ว</p>	ละครโทรทัศน์เรื่องแรกคือเรื่องอะไร

	<p>ทรัพย์ สายยศวิฑิต เธอคือชีวิต คู่ป่วนอลวน มือนาง ตะวันยอแสง สวรรค์สร้าง บ้านเดดอกรัก อสุรน้อยในตะเกียงแก้ว อายูกรัก คือหัตถ์ถาครองพิภพ</p>	<p>ทางหลวงหมายเลข 32 ผ่านจังหวัดอะไร</p>
<p>5</p>	<p>จังหวัดพะเยา เป็นจังหวัดในภาคเหนือตอนบน บริเวณที่ตั้งของตัวเมืองพะเยาในปัจจุบันอยู่ติดกับกว๊านพะเยา เดิมเป็นที่ตั้งของเมือง ภูกามยาว หรือ พะยาว ที่ก่อตั้งขึ้นเมื่อพุทธศตวรรษที่ 16 โดยกษัตริย์องค์แรกคือ พญาจอมธรรม ซึ่งเป็นราชบุตรองค์หนึ่งจากเมืองหริภุญไชยเมืองเชียงแสน และเป็นบรรพบุรุษของกษัตริย์เมืองพะยาวอีกหลายองค์ เช่น พญาเจือง วิชาบุรุษแห่งเผ่าไหลวโนพันที่ลุ่มแม่น้ำโขง และพญางำเมืองซึ่งได้กระทำการยึดสาบานเป็นไมตรีต่อกันกับพญามังรายแห่งนครพิงค์เชียงใหม่ และพญาร่วงรามคำแหงแห่งสุโขทัย ภายหลังมีการเปลี่ยนแปลงอำนาจ และตกอยู่ภายใต้อิทธิพลของอาณาจักรล้านนา ในสมัยพญาคำฟู เมื่อถึงสมัยกรุงรัตนโกสินทร์ ใน พ.ศ. 2386 ตรงกับสมัยรัชกาลที่ 3 เมืองพะเยาถูกตั้งขึ้นใหม่พร้อมเมืองเชียงรายและเมืองงาวเพื่อเป็นเมืองหน้าด่านต่อสู้กับกองทัพพม่าที่ตั้งอยู่ที่เมืองเชียงแสน โดยให้เมืองพะเยาขึ้นตรงต่อ นครลำปาง และในท้ายที่สุดก่อนที่พะเยาจะถูกลกฐานะขึ้นเป็นจังหวัด พะเยาอยู่ใต้การปกครองของจังหวัดเชียงรายในฐานะ อำเภอพะเยา และเมื่อวันที่ 28 สิงหาคม พ.ศ. 2520 อำเภอพะเยาได้ยกฐานะขึ้นเป็น จังหวัดพะเยา นับเป็นจังหวัดที่ 72 ของประเทศไทยประวัติศาสตร์ ประวัติศาสตร์. พะเยาเป็นจังหวัดในกลุ่มล้านนาตะวันออกร่วมกับเชียงราย มีสำเนียงภาษาถิ่นเหนือเป็นเอกลักษณ์ต่างจากภาษาถิ่นเหนือในจังหวัดเชียงใหม่ นอกจากความสัมพันธ์กับเชียงรายในแง่ของสำเนียงภาษาและวัฒนธรรมแล้ว พะเยาและเชียงรายยังมีความสัมพันธ์กันในลักษณะบ้านพี่เมืองน้องฉันทน์ศรีอยุธยา เนื่องจากในอดีต ราชวงศ์ที่ปกครองทั้งสองเมืองต่างสืบเชื้อสายมาจากราชวงศ์ลาวจาก หิริภุญไชยเมืองเชียงแสน ซึ่งได้แตกสายออกมาเป็นราชวงศ์สำคัญในดินแดนพยาว เชียงแสนและล้านนาในยุคต่อมา เช่น ราชวงศ์เชียงแสนที่ปกครองเมืองเชียงราย ราชวงศ์เชียงแสนสายพญาเงี้ยวเมือง และราชวงศ์มังรายซึ่งปกครองเชียงใหม่ตามจากบรรพบุรุษร่วมกับภูมิศาสตร์อาณาเขต วิชาประวัติศาสตร์. อาณาเขต. จังหวัดพะเยามีเนื้อที่ประมาณ 6335 ตารางกิโลเมตร ตอนเหนือมีพื้นที่ติดกับจังหวัดเชียงราย ทิศตะวันออกติดกับประเทศลาวและจังหวัดน่าน ส่วนทางตะวันตกติดกับจังหวัดลำปาง และทางด้านใต้ติดกับจังหวัดแพร่การเมืองการปกครองตั้งอยู่ลักษณะประจำจังหวัดการปกครอง. สัญลักษณ์ประจำจังหวัด. คำขวัญประจำจังหวัด กว๊านพะเยาแหล่งชีวิต ศักดิ์สิทธิ์พระเจ้าตนหลวง บวงสรวงพ่อขุนงำเมือง งามลือเลื่องดอยบุษราคัม ทรายประจำจังหวัด รูปพระเจ้าตนหลวงวัดศรีโคมคำ ต้นไม้ประจำจังหวัด ต้นสารภี ดอกไม้ประจำจังหวัด ดอกสารภีการานามผู้ว่าราชการจังหวัดหน่วยการปกครอง หน่วยงานปกครอง. การปกครองแบ่งออกเป็น 9 อำเภอ 68 ตำบล 779 หมู่บ้าน 2 เทศบาลเมือง 33 เทศบาลตำบล 36 องค์การบริหารส่วนตำบลการ</p>	

	<p>ปกครองส่วนภูมิภาค การปกครองส่วนภูมิภาค. การปกครองส่วนท้องถิ่นการขนส่งทางถนน การขนส่ง. ทางถนน. โดยรถยนต์สามารถเดินทางได้หลายเส้นทาง เช่น 1. ใช้ทางหลวงหมายเลข 32 ผ่านจังหวัดพระนครศรีอยุธยา อำเภอ สิงห์บุรี นครสวรรค์ จากนั้นใช้ทางหลวงหมายเลข 1 ผ่านกำแพงเพชร ตาก ลำปาง แล้วเข้าสู่ตัวเมืองพะเยา ระยะทาง 969 กิโลเมตร 2. ใช้ทางหลวงหมายเลข 32 ผ่านจังหวัดพระนครศรีอยุธยา อำเภอ สิงห์บุรี นครสวรรค์ จากนั้นเลี้ยวขวาเข้าสู่ทางหลวงหมายเลข 117 จนถึงพิษณุโลก แยกซ้ายเข้าสู่ทางหลวงหมายเลข 12 ไปจนถึงสุโขทัย เลี้ยวขวาไปตามทางหลวงหมายเลข 101 ผ่านอำเภอเด่นชัย จังหวัดแพร่...</p>
--	---





ลำดับ	บทความ	คำถาม
1	<p>เจ้าหญิงนาเดจดาแห่งบัลแกเรีย 30 มกราคม ค.ศ. 1899 – 15 กุมภาพันธ์ ค.ศ. 1958 ทรงเป็นสมาชิกของราชวงศ์บัลแกเรียพระราชประวัติเจ้าหญิงนาเดจดาทรงมีพระชนมายุ 8 พรรษา ค.ศ. 1907พระนางทรงประสูติในเมืองไซเพีย ที่ทรงเป็นพระธิดาองค์สุดท้ายของพระเจ้าซาร์เฟอร์ดินานด์ที่ 1 แห่งบัลแกเรียและพระมเหสีองค์แรก เจ้าหญิงมารี หลุยส์แห่งปาร์มา ซึ่งได้สิ้นพระชนม์หลังจากทรงให้ประสูติกาลแก่พระนาง ร่วมกับพระเชษฐภคินี เจ้าหญิงยูโดเซียแห่งบัลแกเรีย พระนางได้รับการศึกษาภายใต้การดูแลของพระวิมาดาเจ้าหญิงเอลเนอร์ รอสแห่งคอสมทริกส์เจ้าหญิงนาเดจดาทรงได้อภิเษกสมรส เมื่อวันที่ 24 มกราคม ค.ศ. 1924 ที่ Bad Mergentheim เยอรมนี กับดยุก อัลเบิร์ต อ็อยเกนแห่งเวือร์ทเทิมแบร์ค พระราชาโอรสองค์ที่สองของอัลเบิร์ต ดยุกแห่งเวือร์ทเทิมแบร์ค ทั้งสองพระองค์ทรงได้มีพระราชโอรสและพระธิดาอยู่ 5 พระองค์ ได้แก่ดยุก แฟร์ดินันท์ อ็อยเกนแห่งเวือร์ทเทิมแบร์ค ดัชเชส มากาเรตตา หลุยส์แห่งเวือร์ทเทิมแบร์ค อภิเษกสมรสในปี ค.ศ. 1970 กับ François Luce de Cheigny ดยุก อ็อยเกน อีเบอร์ฮาร์ดแห่งเวือร์ทเทิมแบร์ค อภิเษกสมรสในปี ค.ศ. 1962 กับอาร์ชดัชเชสเล็คซานดราแห่งออสเตรีย พระธิดาของเจ้าหญิงอีเลินาแห่งโรมาเนียดยุกอเล็กซานเดอร์ อ็อยเกน แห่งเวือร์ทเทิมแบร์ค ดัชเชส โซเฟีย แห่งเวือร์ทเทิมแบร์ค อภิเษกสมรสในปี ค.ศ. 1969 กับ Antonio Manuel Róxo de RamosBandeira เจ้าหญิงนาเดจดาได้สิ้นพระชนม์ ซึ่งมีพระชนมายุได้เพียง 59 พรรษาในอุบัติเหตุที่การท เยอรมนี.</p>	<p>เจ้าหญิงนาเดจดาทรงได้รับการศึกษาภายใต้การดูแลของใคร</p>
2	<p>อิติมการศึกษาวัดภูที่มีขนาดเล็กมาก เช่น วัดภูที่อยู่ไกล วัตถุประสงค์อยู่สูง เป็นต้น ใช้เพียงแว่นขยายและเลนส์อื่นเดี่ยวสองดู เช่นเดียวกับการใช้แว่นขยายส่องดูลายมือ ช่วงปี พ.ศ. 2133 แซคาเรียส เจนเซน ช่างทำแว่นชาวดัตช์ ประดิษฐ์กล้องจุลทรรศน์ชนิดเลนส์ประกอบประกอบด้วยแว่นขยายสองอัน ต่อมา กาลิเลโอ กาลิเลอีได้สร้างแว่นขยายส่องดูสิ่งมีชีวิตเล็ก ๆ ในปี พ.ศ. 2208 รอเบิร์ต ฮุกได้ประดิษฐ์กล้องจุลทรรศน์ชนิดเลนส์ประกอบที่มีลำกล้องรูปร่างสวยงาม ป้องกันการรบกวนจากแสงภายนอกได้ และไม่ต้องถือเลนส์ให้ซ้อนกัน เขาส่องดูไม้คอร์กที่ผานบาง ๆ แล้วพบช่องเล็ก ๆ มากมาย เขาเรียกช่องเหล่านั้นว่าเซลล์ ซึ่งหมายถึงห้องว่าง ๆ หรือห้องซึ่งเซลล์ที่อุกเห็นเป็นเซลล์ที่ตายแล้ว เหลือแต่ผนังเซลล์ของพืชซึ่งแข็งแรงกว่าเยื่อหุ้มเซลล์ในสัตว์ จึงทำให้คงรูปร่างอยู่ได้ ฮุกจึงได้ชื่อว่าเป็นผู้ที่ตั้งชื่อเซลล์ในปี พ.ศ. 2215 อันโตนิ ฟัน เลเวนฮุก ชาวดัตช์ สร้างกล้องจุลทรรศน์ชนิดเลนส์เดี่ยวจากแว่นขยายที่เขาตนเอง แวะนขยายบางอันขยายได้ถึง 270 เท่า เขาใช้กล้องจุลทรรศน์ตรวจดูหยดน้ำจากบึงและแม่น้ำ และจากน้ำฝนที่รองไว้ในหม้อ เห็นสิ่งมีชีวิตเล็ก ๆ มากมาย นอกจากนี้ เขายังส่องดูสิ่งมีชีวิตต่าง ๆ เช่น เม็ดเลือดแดง กล้ามเนื้อ เป็นต้น เมื่อเขาพบสิ่งเหล่านี้ เขารายงานไปยังราช</p>	<p>กล้องจุลทรรศน์ชนิดใดที่ใหญ่ที่สุดในโลก</p>

	<p>สมาคมแห่งกรุงลอนดอน จึงได้รับการยกย่องว่าเป็นผู้ประดิษฐ์กล้องจุลทรรศน์ พ.ศ. 2367 ดูโรเจอร์ นักพฤกษศาสตร์ชาวฝรั่งเศสศึกษาเนื้อเยื่อพืช และสัตว์พบว่าประกอบด้วยเซลล์ พ.ศ. 2376 โรเบิร์ต บราวน์ นักพฤกษศาสตร์ชาวอังกฤษ เป็นคนแรกที่พบว่าเซลล์และพืชมีนิวเคลียสเป็นก่อนกลมๆ อยู่ภายในเซลล์ พ.ศ. 2378 นูก นะดีอาร์ท นักสัตวศาสตร์ชาวฝรั่งเศส ศึกษาจุลินทรีย์และสิ่งมีชีวิตอื่น ๆ พบว่าภายในประกอบด้วยของเหลวใด ๆ จึงเรียกว่า ซาร์โคเด ซึ่งเป็นภาษาฝรั่งเศสมาจากศัพท์กรีกว่า ซาร์ค ซึ่งแปลว่าเนื้อ พ.ศ. 2381 มัททิวส์ ชไลเดน นักพฤกษศาสตร์ชาวเยอรมัน ศึกษาเนื้อเยื่อพืชชนิดต่าง ๆ พบว่าพืชทุกชนิดประกอบด้วยเซลล์ พ.ศ. 2382 ชไลเดน และทีโอดอร์ ขวาน จึงร่วมกันตั้งทฤษฎีเซลล์ ซึ่งมีใจความสรุปได้ว่า สิ่งมีชีวิตทุกชนิดประกอบด้วยเซลล์และผลิตภัณฑ์จากเซลล์ พ.ศ. 2382 ปูร์กีเย นักสัตววิทยาชาวเฮลโลแลนศึกษาไข่และตัวอ่อนของสัตว์ชนิดต่าง ๆ พบว่าภายในมีของเหลวใส เหนียว อ่อนนุ่มเป็นวุ้น เรียกว่าโปรทอปลาซึ่งต่อมาได้มีการศึกษาเกี่ยวกับเซลล์ด้วยกล้องจุลทรรศน์ชนิดเลนส์ประกอบ และได้พัฒนาให้ดียิ่งขึ้น จนกระทั่งปี พ.ศ. 2475 นักวิทยาศาสตร์ชาวเยอรมันคือเอนสท์ รัสกาและมิกซ์ นอลล์ ได้เปลี่ยนแปลงกระบวนการของกล้องจุลทรรศน์ที่ใช้แสงและเลนส์มาใช้ลำอิเล็กตรอน ทำให้เกิดกล้องจุลทรรศน์อิเล็กตรอนขึ้นในระยะต่อ ๆ มา ปัจจุบันมีกำลังขยายกว่า 5 แสนเท่า</p>
3	<p>ไปรอกา เป็นเมืองหลวงและเมืองที่ใหญ่ที่สุดของประเทศคาบูเวร์ดี ประเทศในมหาสมุทรแอตแลนติก ทางตะวันตกของประเทศ เซเนกัล ตั้งอยู่ชายหาดทางใต้ของเกาะเซ็งเตียกู มีท่าเรือเฟอร์รี่และมีสนามบินแห่งชาติ 4 แห่ง เป็นเมืองศูนย์กลางการค้า ส่งออกกาแฟ อ้อย ผลไม้เมืองร้อน มีอุตสาหกรรมการประมง</p>
4	<p>จตุญ ธรรมศิลป์ จตุญ ธรรมศิลป์ เกิดวันที่ 1 เมษายน พ.ศ. 2499 เป็นผู้กำกับ ค่ายผลิตละคร ดาราวิดีโอ ช่อง 7 สี และมีผลงานเรื่องแรกคือเรื่อง เจ้าหญิงแดงอ่อน และได้สร้างผลงานละครหลังข่าวไว้อีกมากมาย เช่น รากนครา ฟ้าใหม่ หลงเงาจันทร์ จิ้งจอกป่าก็แรง อากัง เสียวีตล่องอย่างสวยโรคมะเร็ง เมื่อช่วงเช้าวันที่ 11 พฤศจิกายน 2558 ที่โรงพยาบาลจุฬาลงกรณ์และศิริราชรักษาตัว ภายละครโทรทัศน์. สถานีโทรทัศน์สีกองทัพบกช่อง 7 ปราสาทมืด ท้อพูน หงส์ทอง มณีรั้ว ตะวันชิงพลบ เมืองโหดเพล้ ภาพอาคารพจน์ แก้วสารพัดนึก ภูตแม่น้ำโขง นางทาส ศีระขมมาร ทวิภพ วิมานะพรวัว กระสือ ไร่...มาดา ภูตพิศวาส สายโลหิต กำกักร่วม แม่ขนาดพระนคร ญาติกา รัตนโกสินทร์ ปอบผีฟ้า การะเกด กิ่งไม้ พลั้งรัก ป่วงหงส์ แม่นาค หม่อมทิพย์ ม่านบังใจ รากนครา เงาใจ ปิ่นไพร บอดีการ์ดสาว สาวน้อย รุ่งทิพย์ นะหน้าทอง มหาเฮง สาวน้อยในตะเกียง แก้ว ภาค 2 ฟ้าใหม่ นางสารสัมพันธ์ บ้านร้อยดอกไม้ เหมราช หลงเงาจันทร์ จิ้งจอกป่าก็แรง ยายหนูลูกพ่อ ปู่โสมเฒ่า</p>

ไปรอกา เป็นเมืองหลวงของประเทศอะไร

เจ้าหญิงแดงอ่อน คือใคร

	<p>ปกครองส่วนภูมิภาค การปกครองส่วนภูมิภาค. การปกครองส่วนท้องถิ่นการขนส่งทางถนน การขนส่ง. ทางถนน. โดยรถยนต์สามารถเดินทางได้หลายเส้นทาง เช่น 1. ใช้ทางหลวงหมายเลข 32 ผ่านจังหวัดพระนครศรีอยุธยา อำเภอ สิงห์บุรี นครสวรรค์ จากนั้นใช้ทางหลวงหมายเลข 1 ผ่านกำแพงเพชร ตาก ลำปาง แล้วเข้าสู่ตัวเมืองพะเยา ระยะทาง 969 กิโลเมตร 2. ใช้ทางหลวงหมายเลข 32 ผ่านจังหวัดพระนครศรีอยุธยา อำเภอ สิงห์บุรี นครสวรรค์ จากนั้นเลี้ยวขวาเข้าสู่ทางหลวงหมายเลข 117 จนถึงพิษณุโลก แยกซ้ายเข้าสู่ทางหลวงหมายเลข 12 ไปจนถึงสุโขทัย เลี้ยวขวาไปตามทางหลวงหมายเลข 101 ผ่านอำเภอเด่นชัย จังหวัดแพร่...</p>
--	---





ภาคผนวก ข.

ตัวอย่างผลลัพธ์การสร้างคำถามจาก

แบบจำลองฝึกสอนจากชุดข้อมูลสังเคราะห์รอบที่ 2 โดยมีรูปแบบข้อมูลนำเข้า NER

จุฬาลงกรณ์มหาวิทยาลัย

CHULALONGKORN UNIVERSITY

ลำดับ	บทความ	คำถาม
1	<p>เจ้าหญิงนาเดจดาแห่งบัลแกเรีย 30 มกราคม ค.ศ. 1899 – 15 กุมภาพันธ์ ค.ศ. 1958 ทรงเป็นสมาชิกของราชวงศ์บัลแกเรียพระราชประวัติเจ้าหญิงนาเดจดาทรงมีพระชนมายุ 8 พรรษา ค.ศ. 1907พระนางทรงประสูติในเมืองโซเฟีย ที่ทรงเป็นพระธิดาองค์สุดท้ายของพระเจ้าซาร์เฟอร์ดินานด์ที่ 1 แห่งบัลแกเรียและพระมเหสีองค์แรก เจ้าหญิงมารี หลุยส์แห่งปาร์มา ซึ่งได้สิ้นพระชนม์หลังจากทรงให้ประสูติกาลแก่พระนาง ร่วมกับพระเชษฐภคินี เจ้าหญิงยูโดเซียแห่งบัลแกเรีย พระนางได้รับการศึกษาภายใต้การดูแลของพระวิมาดาเจ้าหญิงเอลเนอร์ รอสแห่งคอสมทริกส์เจ้าหญิงนาเดจดาทรงได้อภิเษกสมรส เมื่อวันที่ 24 มกราคม ค.ศ. 1924 ที่ Bad Mergentheim เยอรมนี กับดยุก อัลเบิร์ต อ็อยเกนแห่งเวือร์ทเทิมแบร์ค พระราชาเออร์สองค์ที่สองของอัลเบิร์ต ดยุกแห่งเวือร์ทเทิมแบร์ค ทั้งสองพระองค์ทรงได้มีพระราชโอรสและพระธิดาอยู่ 5 พระองค์ ได้แก่ดยุก แฟร์ดินันท์ อ็อยเกนแห่งเวือร์ทเทิมแบร์ค ดัชเชส มากาเรตตา หลุยส์แห่งเวือร์ทเทิมแบร์ค อภิเษกสมรสในปี ค.ศ. 1970 กับ François Luce de Cheigny ดยุก อ็อยเกน อีเบอร์ฮาร์ดแห่งเวือร์ทเทิมแบร์ค อภิเษกสมรสในปี ค.ศ. 1962 กับอาร์ชดัชเชสเล็กซานดราแห่งออสเตรีย พระธิดาของเจ้าหญิงอีเลินาแห่งโรมาเนียดยุกอเล็กซานเดอร์ อ็อยเกน แห่งเวือร์ทเทิมแบร์ค ดัชเชส โซเฟีย แห่งเวือร์ทเทิมแบร์ค อภิเษกสมรสในปี ค.ศ. 1969 กับ Antonio Manuel Róxo de RamosBandeira เจ้าหญิงนาเดจดาได้สิ้นพระชนม์ ซึ่งมีพระชนมายุได้เพียง 59 พรรษาในอุบัติเหตุที่การท เยอรมนี.</p>	เจ้าหญิงนาเดจดาทรงอภิเษกสมรสกับใคร
2	<p>อิติมการศึกษาวัดุมุที่มีขนาดเล็กมาก เช่น วัดุมุที่อยู่ไกล วัดุมุที่อยู่สูง เป็นต้น ใช้เพียงแว่นขยายและเลนส์อีนเดียวส่องดู เช่นเดียวกับการใช้แว่นขยายส่องดูลายมือ ช่วงปี พ.ศ. 2133 แซคาเรียส เจนเซน ช่างทำแว่นชาวดัตช์ ประดิษฐ์กล้องจุลทรรศน์ชนิดเลนส์ประกอบประกอบด้วยแว่นขยายสองอัน ต่อมา กาลิเลโอ กาลิเลอีได้สร้างแว่นขยายส่องดูสิ่งมีชีวิตเล็ก ๆ ในปี พ.ศ. 2208 รอเบิร์ต ฮุกได้ประดิษฐ์กล้องจุลทรรศน์ชนิดเลนส์ประกอบที่มีลำกล้องรูปร่างสวยงาม ป้องกันการรบกวนจากแสงภายนอกได้ และไม่ต้องถือเลนส์ให้ซ้อนกัน เขาส่องดูไม้คอร์กที่ผานบาง ๆ แล้วพบช่องเล็ก ๆ มากมาย เขาเรียกช่องเหล่านั้นว่าเซลล์ ซึ่งหมายถึงห้องว่าง ๆ หรือห้องซึ่งเซลล์ที่ฮุกเห็นเป็นเซลล์ที่ตายแล้ว เหลือแต่ผนังเซลล์ของพืชซึ่งแข็งแรงกว่าเยื่อหุ้มเซลล์ในสัตว์ จึงทำให้คงรูปร่างอยู่ได้ ฮุกจึงได้ชื่อว่าเป็นผู้ที่ตีตั้งชื่อเซลล์ในปี พ.ศ. 2215 อันโตนิ ฟัน เลเวนฮุก ชาวดัตช์ สร้างกล้องจุลทรรศน์ชนิดเลนส์เดียวจากแว่นขยายที่เขาตนเอง แวะนขยายบางอันขยายได้ถึง 270 เท่า เขาใช้กล้องจุลทรรศน์ตรวจดูหยดน้ำจากบึงและแม่น้ำ และจากน้ำฝนที่รองไว้ในหม้อ เห็นสิ่งมีชีวิตเล็ก ๆ มากมาย นอกจากนี้ เขายังส่องดูสิ่งมีชีวิตต่าง ๆ เช่น เม็ดเลือดแดง กล้ามเนื้อ เป็นต้น เมื่อเขาพบสิ่งเหล่านี้ เขารายงานไปยังราช</p>	ในปี พ.ศ. ไต

	<p>สมาคมแห่งกรุงลอนดอน จึงได้รับการยกย่องว่าเป็นผู้ประดิษฐ์กล้องจุลทรรศน์ พ.ศ. 2367 ดูโรเจอร์ นักพฤกษศาสตร์ชาวฝรั่งเศสศึกษาเนื้อเยื่อพืช และสัตว์พบว่าประกอบด้วยเซลล์ พ.ศ. 2376 โรเบิร์ต บราวน์ นักพฤกษศาสตร์ชาวอังกฤษ เป็นคนแรกที่พบว่าเซลล์และพืชมีนิวเคลียสเป็นก่อนกลมๆ อยู่ภายในเซลล์ พ.ศ. 2378 นูก นะดีอาร์ทเจอร์ นักสัตวศาสตร์ชาวฝรั่งเศส ศึกษาจุลินทรีย์และสิ่งมีชีวิตอื่น ๆ พบว่าภายในประกอบด้วยของเหลวใส ๆ จึงเรียกว่า ซาร์โคโด ซึ่งเป็นภาษาฝรั่งเศสมาจากศัพท์กรีกว่า ซาร์ค ซึ่งแปลว่าเนื้อ พ.ศ. 2381 มัททิวส์ ชไลเดน นักพฤกษศาสตร์ชาวเยอรมัน ศึกษาเนื้อเยื่อพืชชนิดต่าง ๆ พบว่าพืชทุกชนิดประกอบด้วยเซลล์ พ.ศ. 2382 ชไลเดน และทีโอดอร์ ชวาน จึงร่วมกันตั้งทฤษฎีเซลล์ ซึ่งมีใจความสรุปได้ว่า สิ่งมีชีวิตทุกชนิดประกอบด้วยเซลล์และผลิตภัณฑ์จากเซลล์ พ.ศ. 2382 ปูร์กีเนีย นักสัตววิทยาชาวเชโกสโลวาเกีย ศึกษาไข่และตัวอ่อนของสัตว์ชนิดต่าง ๆ พบว่าภายในมีของเหลวใส เหนียว อ่อนนุ่มเป็นวุ้น เรียกว่าโปรทอปลาซึมต่อจากนั้นมีนักวิทยาศาสตร์อีกมากมายทำการศึกษาร่วมกันเกี่ยวกับเซลล์ด้วยกล้องจุลทรรศน์ชนิดเลนส์ประกอบ และได้พัฒนาให้ดียิ่งขึ้น จนกระทั่งปี พ.ศ. 2475 นักวิทยาศาสตร์ชาวเยอรมันคือแอนสท์ รัสกาและมิกซ์ นอลล์ ได้เปลี่ยนแปลงกระบวนการของกล้องจุลทรรศน์ที่ใช้แสงและเลนส์มาใช้ลำเล็กลึกลง ทำให้เกิดกล้องจุลทรรศน์อิเล็กตรอนขึ้นในระยะต่อ ๆ มา ปัจจุบันมีกำลังขยายกว่า 5 แสนเท่า</p>	เมืองหลวงของประเทศคาบูเวร์ดี มีท่าเรือเพอร์รีและมีสนามบินแห่งชาติที่แห่ง
3	<p>โปรอา ไปรออา เป็นเมืองหลวงและเมืองที่ใหญ่ที่สุดของประเทศคาบูเวร์ดี ประเทศในมหาสมุทรแอตแลนติก ทางตะวันตกของประเทศเซเนกัล ตั้งอยู่ชายหาดทางใต้ของเกาะซังเตียกู มีท่าเรือเพอร์รีและมีสนามบินแห่งชาติ 4 แห่ง เป็นเมืองศูนย์กลางการค้า ส่งออกกาแฟ อ้อย ผลไม้เมืองร้อน มีอุตสาหกรรมการประมง</p>	ละครเรื่องแรกคือเรื่องอะไร
4	<p>จตุญ ธรรมศิลป์ จตุญ ธรรมศิลป์ เกิดวันที่ 1 เมษายน พ.ศ. 2499 เป็นผู้กำกับ ค่ายผลิตละคร ดาราวิดีโอ ช่อง 7 สี และมีผลงานเรื่องแรกคือเรื่อง เจ้าหญิงแดงอ่อน และได้สร้างผลงานละครหลังข่าวไว้อีกมากมาย เช่น รากนครา ฟ้าใหม่ หลงเงาจันทร์ จิ้งจอกป่าที่แรง อาแกง เสียชีวิตลงอย่างสงบด้วยโรคมะเร็ง เมื่อช่วงเช้าวันที่ 11 พฤศจิกายน 2558 ที่โรงพยาบาลจุฬารัตน 5 วัฒนา กรุงเทพมหานคร. สถานที่โทรทัศน์กองทัพบกช่อง 7 ปราสาทหินดีด หองหนุ่ย หงส์ทอง มณีรั้ว ตะวันชิงพลบ เมืองโศกโศก ภาพอาคารพจน์ แก้วสารพัดนึก ภูตแม่น้ำโขง นางทาส ศีระสมาร ทวิภพ วิมานะพรวัว กระสือ ไร่...มาดา ภูตพิพิศาศ สายโลหิต กำกับร่วม แม่มาดพระนคร ญาติกา รัตนโกสินทร์ ปอบผีฟ้า การะเกด กิ่งไม้ พลั้งรัก ป่วงหงส์ แม่มาด หนุ่มทิพย์ ม่านบังใจ รากนครา เงาใจ ปิ่นไพร บอดีการ์ดสาว สาวน้อย รุ่งทิพย์ นะหน้าทอง มหาเฮง สาวน้อยในตะเกียง แก้ว ภาค 2 ฟ้าใหม่ นางสารสัมพันธ์ บ้านร้อยดอกไม้ เหมราช หลงเงาจันทร์ จิ้งจอกป่าที่แรง ยายหนูลูกพ่อ ปุโลมแผ้ว</p>	

	<p>ทรัพย์ สายยศวิฑิต คุป่วนอลวน มือนาง ตะวันยอแสง สวรรค์สร้าง บ้านเดดอกรัก อสุรน้อยในตะเกียงแก้ว อาญากรัก คือหัตถาครองพิภพ</p>	<p>ในปี พ.ศ. ใด</p>
5	<p>จังหวัดพะเยา จังหวัดพะเยา ะญาว เป็นจังหวัดในภาคเหนือตอนบน บริเวณที่ตั้งของตัวเมืองพะเยาในปัจจุบันอยู่ติดกับกว๊านพะเยา เดิมเป็นที่ตั้งของเมือง ภูกามยาว หรือ พะยาว ที่ก่อตั้งขึ้นเมื่อพุทธศตวรรษที่ 16 โดยกษัตริย์องค์แรกคือ พญาจอมธรรม ซึ่งเป็นคราษัตริย์องค์หนึ่งจากเมืองหริภุญไชย เชียงใหม่ และพะเยา ซึ่งได้กระทำการทำศึกต่อกันกับพญามังรายแห่งนครพิงค์เชียงใหม่ วิชาเจือง วิชาบุรุษแห่งเผ่าไทลาวในพื่นที่ลุ่มแม่น้ำโขง และพญาเงี้ยว ซึ่งได้กระทำการทำศึกต่อกันกับพญามังรายแห่งนครพิงค์เชียงใหม่ และพญา ร่วงรามคำแหงแห่งสุโขทัย ภายหลังมีการเปลี่ยนแปลงอำนาจ และตกอยู่ภายใต้อิทธิพลของอาณาจักรล้านนา ในสมัยพญาคำฟู เมื่อถึงสมัยกรุงรัตนโกสินทร์ ใน พ.ศ. 2386 ตรงกับสมัยรัชกาลที่ 3 เมืองพะเยาถูกตั้งขึ้นใหม่พร้อมเมืองเชียงรายและเมืองงาวเพื่อเป็นเมืองหน้าด่านต่อสู้กับกองทัพพม่าที่ตั้งอยู่ที่เมืองเชียงแสน โดยให้เมืองพะเยาขึ้นตรงต่อ นครลำปาง และในท้ายที่สุดก่อนที่พะเยาจะถูยกฐานะขึ้นเป็นจังหวัด พะเยาอยู่ใต้การปกครองของจังหวัดเชียงรายในฐานะ อำเภอพะเยา และเมื่อวันที่ 28 สิงหาคม พ.ศ. 2520 อำเภอพะเยาได้ยกฐานะขึ้นเป็น จังหวัดพะเยา นับเป็นจังหวัดที่ 72 ของประเทศไทยประวัติศาสตร์ ประวัติศาสตร์. พะเยาเป็นจังหวัดในกลุ่มล้านนาตะวันออกร่วมกับเชียงราย มีสำเนียงภาษาถิ่นเหนือเป็นเอกลักษณ์ต่างจากภาษาถิ่นอื่นในจังหวัดเชียงใหม่ นอกจากนี้ ความสัมพันธ์กับเชียงรายในแง่ของสำเนียงภาษาและวัฒนธรรมแล้ว พะเยาและเชียงรายยังมีความสัมพันธ์กันในลักษณะบ้านพี่เมืองน้องฉันทน์ศรีอยุธยา เนื่องจากในอดีต ราชวงศ์ที่ปกครองฝั่งสองเมืองต่างสืบเชื้อสายมาจากราชวงศ์ลาวจาก หิริภุญไชย เชียงแสน ซึ่งได้แตกสายออกมาเป็นราชวงศ์สำคัญในดินแดนพะเยา เชียงแสนและล้านนาในยุคต่อมา เช่น ราชวงศ์เชียงแสนที่ปกครองเมือง เชียงราย ราชวงศ์เชียงแสนสายพญาเงี้ยวเมือง และราชวงศ์มืองซึ่งปกครองเชียงใหม่เดิมก้าเกิดมาจากบรรพบุรุษร่วมกับภูมิศาสตร์อาณาเขต ภูมิศาสตร์. อาณาเขต. จังหวัดพะเยามีเนื้อที่ประมาณ 6335 ตารางกิโลเมตร ตอนเหนือมีพื้นที่ติดกับจังหวัดเชียงราย ทิศตะวันออกเฉียงใต้กับประเทศลาวและจังหวัดน่าน ส่วนทางตะวันตกติดกับจังหวัดลำปาง และทางด้านใต้ติดกับจังหวัดแพร่การเมืองการ ปกครองสัญลักษณ์ประจำจังหวัดการปกครอง. สัญลักษณ์ประจำจังหวัด. คำขวัญประจำจังหวัด กว๊านพะเยาแหล่งชีวิต ศักดิ์สิทธิ์พระเจ้าตนหลวง บวงสรวงพ่อขุนงำเมือง งามลือเลื่องดอยบุษราคัม ทรายประจำจังหวัด รูปพระเจ้าตนหลวงวัดศรีโคมคำ ต้นไม้ประจำจังหวัด ต้นสารภี ดอกไม้ประจำจังหวัด ดอกสารภีการัญญาการจังหวัดหน่วยการปกครอง หน่วยการปกครอง. การปกครองแบ่งออกเป็น 9 อำเภอ 68 ตำบล 779 หมู่บ้าน 2 เทศบาลเมือง 33 เทศบาลตำบล 36 องค์การบริหารส่วนตำบลการ</p>	

	<p>ปกครองส่วนภูมิภาค การปกครองส่วนภูมิภาค. การปกครองส่วนท้องถิ่นการขนส่งทางถนน การขนส่ง. ทางถนน. โดยรถยนต์สามารถเดินทางได้หลายเส้นทาง เช่น 1. ใช้ทางหลวงหมายเลข 32 ผ่านจังหวัดพระนครศรีอยุธยา อำเภอ สิงห์บุรี นครสวรรค์ จากนั้นใช้ทางหลวงหมายเลข 1 ผ่านกำแพงเพชร ตาก ลำปาง แล้วเข้าสู่ตัวเมืองพะเยา ระยะทาง 969 กิโลเมตร 2. ใช้ทางหลวงหมายเลข 32 ผ่านจังหวัดพระนครศรีอยุธยา อำเภอ สิงห์บุรี นครสวรรค์ จากนั้นเลี้ยวขวาเข้าสู่ทางหลวงหมายเลข 117 จนถึงพิษณุโลก แยกซ้ายเข้าสู่ทางหลวงหมายเลข 12 ไปจนถึงสุโขทัย เลี้ยวขวาไปตามทางหลวงหมายเลข 101 ผ่านอำเภอเด่นชัย จังหวัดแพร่...</p>
--	---



บรรณานุกรม

1. Kwankajornkiet, C., A. Suchato, and P. Punyabukkana. *Automatic multiple-choice question generation from Thai text*. in 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE). 2016. IEEE.
2. Le, N.-T., T. Kojiri, and N. Pinkwart, *Automatic question generation for educational applications—the state of art*, in *Advanced computational methods for knowledge engineering*. 2014, Springer. p. 325-338.
3. Zhao, J., X. Deng, and H. Sun, *Easy-to-Hard: Leveraging Simple Questions for Complex Question Generation*. arXiv preprint arXiv:1912.02367, 2019.
4. Liu, B. *Neural Question Generation based on Seq2Seq*. in *Proceedings of the 2020 5th International Conference on Mathematics and Artificial Intelligence*. 2020.
5. Linjordet, T. and K. Balog. *Impact of Training Dataset Size on Neural Answer Selection Models*. in *European Conference on Information Retrieval*. 2019. Springer.
6. Kittinaradorn, R., et al., *A Thai word tokenization library using deep neural network*. 2018.
7. Chormai, P., P. Prasertsom, and A. Rutherford, *AttaCut: A Fast and Accurate Neural Thai Word Segmenter*. arXiv preprint arXiv:1911.07056, 2019.
8. Mikolov, T., et al., *Distributed representations of words and phrases and their compositionality*. arXiv preprint arXiv:1310.4546, 2013.
9. Pennington, J., R. Socher, and C.D. Manning. *Glove: Global vectors for word representation*. in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. 2014.
10. 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.
11. Bloom, B.S., *Taxonomy of educational objectives: The classification of educational goals*. Cognitive domain, 1956.

12. Mitkov, R. *Computer-aided generation of multiple-choice tests*. in *Proceedings of the HLT-NAACL 03 workshop on Building educational applications using natural language processing*. 2003.
13. Heilman, M. and N.A. Smith. *Good question! statistical ranking for question generation*. in *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. 2010.
14. Serban, I.V., et al., *Generating factoid questions with recurrent neural networks: The 30m factoid question-answer corpus*. arXiv preprint arXiv:1603.06807, 2016.
15. Du, X., J. Shao, and C. Cardie, *Learning to ask: Neural question generation for reading comprehension*. arXiv preprint arXiv:1705.00106, 2017.
16. Kriangchaivech, K. and A. Wangperawong, *Question generation by transformers*. arXiv preprint arXiv:1909.05017, 2019.
17. Chan, Y.-H. and Y.-C. Fan. *A recurrent BERT-based model for question generation*. in *Proceedings of the 2nd Workshop on Machine Reading for Question Answering*. 2019.
18. Lopez, L.E., et al., *Transformer-based End-to-End Question Generation*. arXiv preprint arXiv:2005.01107, 2020.
19. Liu, B., et al. *Asking questions the human way: Scalable question-answer generation from text corpus*. in *Proceedings of The Web Conference 2020*. 2020.
20. Wu, X., N. Jiang, and Y. Wu, *A question type driven and copy loss enhanced framework for answer-agnostic neural question generation*. arXiv preprint arXiv:2005.11665, 2020.
21. Puri, R., et al., *Training question answering models from synthetic data*. arXiv preprint arXiv:2002.09599, 2020.
22. Yan, D. and S. Guo, *Leveraging contextual sentences for text classification by using a neural attention model*. *Computational intelligence and neuroscience*, 2019. 2019.



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ประวัติผู้เขียน

ชื่อ-สกุล ณัฐนิช วิวัฒน์บุตรสิริ
วัน เดือน ปี เกิด 17 มีนาคม 2536
สถานที่เกิด กรุงเทพฯ



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY