



รายงานวิจัยฉบับสมบูรณ์

โครงการ การพัฒนาต้นแบบระบบสอบถามข้อมูลการท่องเที่ยวอัจฉริยะ

โดย

ดร.ไกรศักดิ์ เกษร คณะวิทยาศาสตร์

สำนักหอสมุด มหาวิทยาลัยนครสวรรค์

วันลงทะเบียน ๙ - เม.ย. ๒๕๕๘

เลขทะเบียน 1.6๗๓1216

เลขเรียกหนังสือ ๑ T58

๒๕  
๓๑๑๔๖  
1956

12 กุมภาพันธ์ 2557

สัญญาเลขที่ R2556B011

## รายงานวิจัยฉบับสมบูรณ์

โครงการ การพัฒนาต้นแบบระบบสอบถามข้อมูลการท่องเที่ยวอัจฉริยะ

คณะผู้วิจัย

พ.น.ดร.ไกรศักดิ์ เกษร สังกัด คณะวิทยาศาสตร์

มหาวิทยาลัยนเรศวร

สนับสนุนโดยงบประมาณแผ่นดิน มหาวิทยาลัยนเรศวร

ปีงบประมาณ 2556

## สารบัญ

บทคัดย่อ.....	iv
Abstract .....	iv
บทสรุปสำหรับผู้บริหาร.....	v
บทที่ 1 บทนำ .....	7
1.1 ที่มาและความสำคัญของปัญหา.....	7
1.2 จุดมุ่งหมายของการศึกษา.....	8
1.3 ขอบเขตของงานวิจัย .....	8
1.4 ประโยชน์ที่คาดว่าจะได้รับ .....	9
บทที่ 2 ทบทวนวรรณกรรม .....	10
2.1. ไม่ใช่การวิเคราะห์เชิงความหมาย (Non-semantic base technology).....	11
2.2. การวิเคราะห์เชิงความหมาย (Semantic based technology) .....	12
2.3. Query Recommendation (การจัดการคำถามที่ไม่รู้จัก) .....	12
2.4. ตัวอย่างระบบถามตอบที่มีในปัจจุบัน .....	15
บทที่ 3 ทฤษฎีที่เกี่ยวข้อง .....	20
3.1. กฎทางอักขระ .....	21
3.2. การตัดคำแบบ Longest Matching .....	22
3.3. การตัดคำด้วยกระบวนการย้อนกลับ (Backtracking).....	23
3.4. การตัดคำด้วยเทคนิคผสม ( Hybrid Approach ).....	24
บทที่ 4 ระเบียบวิธีวิจัย .....	26
4.1. ระบบตัดคำ (Word tokenizing).....	26
4.2. การวิเคราะห์คำถาม (Question analysis).....	27
4.3. การค้นหาคำตอบและการเรียงลำดับ (Answer matching) .....	29
4.4. การจัดการคำถามที่ไม่รู้จัก (Unknown question handling).....	30
4.5. เครื่องมือที่ใช้ในการพัฒนา .....	30
4.6. การออกแบบส่วนแสดงผล (Output Design).....	35
บทที่ 5 ผลการวิจัย.....	37
5.1. ความถูกต้องของการตอบคำถาม เช่น อะไร ที่ไหน.....	37
5.2. คำถามที่กำกวม.....	43
5.3. เปรียบเทียบประสิทธิภาพกับ Google .....	43

## สารบัญ (ต่อ)

บทที่ 6 บทสรุปและข้อเสนอแนะ .....	45
6.1. ความสำเร็จของงานวิจัยและแนวคิดใหม่ .....	45
6.2. ข้อจำกัดของงานวิจัย .....	46
บรรณานุกรม .....	47
ภาคผนวก ก รายงานตัวชี้วัด .....	49
ภาคผนวก ข จดหมายตอบรับ .....	50
ภาคผนวก ข บทความฉบับสมบูรณ์ .....	51





## สารบัญญภาพ

รูปที่ 1 ลักษณะการให้บริการของระบบอัตโนมัติในปัจจุบัน .....	15
รูปที่ 2 การถาม ตอบ สถานที่ท่องเที่ยว .....	16
รูปที่ 3 การถาม ตอบ ชื่อสถานที่ท่องเที่ยว .....	16
รูปที่ 4 การถาม ตอบ ร้านอาหาร .....	17
รูปที่ 5 การถาม ตอบ โรงแรมและที่พัก .....	18
รูปที่ 6 การสนทนาของชิมชิมิ .....	19
รูปที่ 7 ส่วนประกอบของระบบ ATQ.....	27
รูปที่ 8 ตัวอย่างการตัดคำโดยโปรแกรมแบ่งคำภาษาไทยด้วยเทคนิคไฮบริด .....	27
รูปที่ 9 ผังงานของระบบ .....	32
รูปที่ 10 ตำแหน่งของคำถาม .....	33
รูปที่ 11 หน้าจอหลัก .....	35
รูปที่ 12 หน้าจอแรกของระบบ .....	37
รูปที่ 13 หน้าจอรระบบ ATQ .....	38
รูปที่ 14 ข้อมูลการสนทนาเป็นข้อมูล .....	39
รูปที่ 15 ข้อมูลแหล่งที่ท่องเที่ยว .....	39
รูปที่ 16 ตำแหน่งบนแผนที่ .....	40
รูปที่ 17 แผนที่แบบปิกมุดหลายๆ ที่ .....	41
รูปที่ 18 แผนที่เฉพาะสถานที่ .....	42
รูปที่ 19 ข้อมูลของสถานที่ .....	42
รูปที่ 20 คำกำกวม .....	43
รูปที่ 21 ค่าความแม่นยำในการค้นหาข้อมูลที่ตรงกับคำถามผู้ใช้ .....	44

## บทคัดย่อ

อุตสาหกรรมการท่องเที่ยวจัดได้ว่าเป็นธุรกิจที่มีความเติบโตอย่างรวดเร็วและเป็นอุตสาหกรรมที่สร้างรายได้ให้กับประเทศอย่างสูงในแต่ละปี ภาครัฐจึงมีนโยบายให้ทุกฝ่ายช่วยกันพัฒนาส่งเสริมคุณภาพและมาตรฐานของการบริการและกระตุ้นการท่องเที่ยวโดยเพิ่มความหลากหลายของรูปแบบการให้บริการและการท่องเที่ยว ดังนั้นบทความนี้จึงนำเสนอระบบการให้บริการข้อมูลการท่องเที่ยวโดยอาศัยคอมพิวเตอร์เป็นเครื่องมือหลักในการให้ข้อมูลจากการสอบถามของผู้ใช้โดยอัตโนมัติ หัวใจสำคัญของระบบนี้คือ 1) เทคนิคการวิเคราะห์คำถาม เพื่อให้เข้าใจความหมายที่แท้จริงของคำถามนั้นๆ เนื่องจากคำถามอาจมีความกำกวม ซึ่งวิธีการที่นำเสนอนี้จะใช้บทบาทหรือหน้าที่ของคำต่างๆ มาช่วยในการวิเคราะห์คำถาม 2) นอกจากนี้ระบบที่นำเสนอยังเก็บข้อมูลการท่องเที่ยวในรูปแบบขององค์ความรู้โดยใช้ออนโทโลยี ทำให้สามารถจัดการคำถามที่มีความกำกวมได้อย่างมีประสิทธิภาพมากขึ้น ผลการทดลองชี้ให้เห็นว่าระบบเข้าใจคำถามและสามารถหาคำตอบที่สอดคล้องกับคำถามเพิ่มสูงขึ้นนอกจากนี้ระบบยังให้ข้อมูลแก่นักท่องเที่ยวที่ตรงประเด็นมากกว่าเมื่อเปรียบเทียบกับระบบค้นหาข้อมูลทั่วไป เช่น Google

## Abstract

With the rapid growth of the tourism industry, it has become a major business and generates a huge profits for the country. Therefore, the government tries to offer various services to tourists in order to enhance the quality of services and persuade travelers to visit Thailand. This paper presents a novel framework for a question-answer agent system that can automatically provide tourism information for tourists. The key ideas of this research are: 1) a novel technique for question analysis using word roles and 2) knowledge-based design using an ontology for handling ambiguous questions. The experimental results show that the proposed framework can efficiently find the correct information to a user and improve the precision value compared to Google.



### บทสรุปสำหรับผู้บริหาร

ประเทศไทยเป็นประเทศหนึ่งที่นักท่องเที่ยวสนใจเดินทางมาเยี่ยมชม เนื่องจากความโดดเด่นในเรื่องของทรัพยากรการท่องเที่ยวโดยมีความหลากหลายและมีเอกลักษณ์เฉพาะของสถานที่ท่องเที่ยวทางธรรมชาติและทางวัฒนธรรม แต่อย่างไรก็ตามจากผลสำรวจของกรมการท่องเที่ยว พบว่าประเทศไทยยังให้ข้อมูลเกี่ยวกับแหล่งท่องเที่ยวไม่เพียงพอหรืออุตสาหกรรมการท่องเที่ยวของไทยเองไม่สามารถสร้างสื่อประชาสัมพันธ์ที่เข้าถึงกลุ่มตลาดเป้าหมายได้ นักท่องเที่ยวอาจจะไม่มีความสะดวกในการเข้าไปที่สำนักงานการท่องเที่ยวแห่งประเทศไทยเพื่อสอบถามข้อมูลการท่องเที่ยวที่ตนเองสนใจ ทั้งนี้การเพิ่มศักยภาพทางการท่องเที่ยวเพื่อแก้ไขปัญหาดังกล่าวสามารถทำได้โดยพัฒนาระบบฐานข้อมูลและศูนย์รวมข้อมูลของการท่องเที่ยว ซึ่งสอดคล้องกับนโยบายของกระทรวงการท่องเที่ยวและกีฬาที่ต้องการพัฒนาการตลาดเชิงอิเล็กทรอนิกส์และระบบให้ข้อมูลการท่องเที่ยวในประเทศไทย การพัฒนาระบบสารสนเทศเพื่อให้ข้อมูลการท่องเที่ยวในประเทศไทยนี้มีวัตถุประสงค์เพื่ออำนวยความสะดวกในการให้ข้อมูลกับนักท่องเที่ยว อีกทั้งยังช่วยลดต้นทุนในการจ้างมนุษย์ในการทำหน้าที่ดังกล่าวได้อีกด้วย

การให้ข้อมูลด้านการท่องเที่ยวของประเทศไทยผ่านทางเว็บไซต์ต่างๆ ที่มีอยู่ในปัจจุบันยังมีประสิทธิภาพไม่เพียงพอที่จะนำเสนอข้อมูลให้ตรงกับความต้องการที่แท้จริงของนักท่องเที่ยว และระบบสนับสนุนการสืบค้นข้อมูลด้านการท่องเที่ยวยังไม่ดีเท่าที่ควร ตัวอย่างเช่น นักท่องเที่ยวต้องการค้นหาข้อมูลเกี่ยวกับโบราณสถานในประเทศไทยโดยใช้คำสำคัญในการค้นหาว่า "อารามหลวง" ระบบค้นหาข้อมูลในปัจจุบัน เช่น Google จะค้นพบเฉพาะเอกสารที่ปรากฏคำสำคัญนี้เท่านั้น เอกสารอื่นๆ ที่มีเนื้อหา (Concept) เดียวกันแต่ไม่มีคำสำคัญเหล่านี้ปรากฏอยู่ในเอกสาร ระบบค้นหาข้อมูลก็จะไม่สามารถค้นพบและนำเสนอต่อนักท่องเที่ยวได้ จากปัญหาดังกล่าวผู้วิจัยจึงมีแนวคิดที่จะพัฒนาระบบถาม-ตอบอัตโนมัติสำหรับให้ข้อมูลการท่องเที่ยวในเขตจังหวัดพิษณุโลก โดยต้นแบบของระบบนี้จะเน้นการให้ข้อมูลเพื่อตอบคำถามประเภทข้อมูลทั่วไป (What) และที่อยู่ของสถานที่ท่องเที่ยว (Where) เป็นหลักก่อนและจะขยายการทำงานไปสู่คำถามประเภทอื่นๆ เช่น คำถามเกี่ยวกับเวลา ต่อไปในอนาคต

กรอบแนวคิดที่นำเสนอในงานวิจัยนี้เรียกว่า "ระบบสอบถามข้อมูลการท่องเที่ยวอัจฉริยะ (Automatic Tourism Question-Answer System-ATQ)" แบ่งออกเป็น 2 ส่วนหลักๆ โดยที่แต่ละส่วนจะให้ความสำคัญกับงานที่ต่างกัน ทั้งนี้เพื่อให้ระบบคอมพิวเตอร์สามารถเข้าใจสิ่งที่ผู้ใช้หรือนักท่องเที่ยวต้องการสื่อสารกับระบบ โดยที่ส่วนที่ 1 จะเน้นไปที่การวิเคราะห์คำถาม (Question analysis) เพื่อนำไปสู่การตีความประโยค (Interpretation/Understanding) เพื่อให้เกิดความใจความหมายของคำถามต่างๆที่นักท่องเที่ยวสื่อสารกับระบบโดยการพิมพ์ผ่านหน้าจอ และส่วนที่ 2 คือการจัดการกับคำถามที่ระบบไม่รู้จัก เนื่องจากนักท่องเที่ยวแต่ละคนอาจจะเขียนคำถามที่ใช้คำพูดไม่เหมือนกันดังนั้นจึงเป็นไปได้

ยากที่ระบบจะเข้าใจคำถามทุกคำถามของนักท่องเที่ยว โดยขอบเขตของระบบต้นแบบอยู่ที่การให้ข้อมูลสถานที่ท่องเที่ยวในจังหวัดพิษณุโลกก่อน

ผลการทดลองปรากฏว่าวิธีการที่นำเสนอในงานวิจัยนี้ ระบบ ATQ สามารถปรับปรุงประสิทธิภาพการค้นหาข้อมูลและตอบคำถามโดยใช้ภาษาไทยได้สูงกว่า Google ที่มีความแม่นยำเฉลี่ยอยู่ที่ 85% ในขณะที่ Google มีความแม่นยำเฉลี่ยที่ 74%





# บทที่ 1

## บทนำ

### 1.1 ที่มาและความสำคัญของปัญหา

ประเทศไทยเป็นประเทศหนึ่งที่นักท่องเที่ยวสนใจเดินทางมาเยี่ยมชม เนื่องจากความโดดเด่นในเรื่องของทรัพยากรการท่องเที่ยวโดยมีความหลากหลายและมีเอกลักษณ์เฉพาะของสถานที่ท่องเที่ยวทางธรรมชาติและทางวัฒนธรรม แต่อย่างไรก็ตามจากผลสำรวจของกรมการท่องเที่ยว (2554) พบว่านักท่องเที่ยวส่วนใหญ่เดินทางเข้ามาในประเทศไทยโดยมีจุดประสงค์หลักคือ ต้องการเยี่ยมชมแหล่งท่องเที่ยวทางธรรมชาติและวัฒนธรรม แต่ไม่ได้รับบริการด้านข้อมูลเกี่ยวกับแหล่งท่องเที่ยวทางอย่างเพียงพอหรืออุตสาหกรรมการท่องเที่ยวของไทยเองไม่สามารถสร้างสื่อประชาสัมพันธ์ที่เข้าถึงกลุ่มตลาดเป้าหมายได้ นักท่องเที่ยวอาจจะไม่มีความสะดวกในการเข้าไปที่สำนักงานการท่องเที่ยวแห่งประเทศไทยเพื่อสอบถามข้อมูลการท่องเที่ยวที่ตนเองสนใจ ซึ่งปัญหาดังกล่าวนี้นับได้ว่าเป็นปัญหาสำคัญของประเทศอันหนึ่งเนื่องจากประเทศไทยเองมีทรัพยากรทางวัฒนธรรมที่โดดเด่น แต่เมื่อแหล่งท่องเที่ยวทางวัฒนธรรมไม่เป็นที่นิยมของนักท่องเที่ยวโดยเฉพาะนักท่องเที่ยวชาวไทย จึงทำให้การท่องเที่ยวของไทยเสียโอกาสทั้งในด้านเศรษฐกิจและทั้งในด้านการพัฒนาคุณภาพชีวิตของประชาชน

ทั้งนี้การเพิ่มศักยภาพทางการท่องเที่ยวเพื่อแก้ไขปัญหาดังกล่าวสามารถทำได้โดยการพัฒนาฐานข้อมูลและศูนย์รวมข้อมูลของการท่องเที่ยว (เทิดชาย ช่วยบำรุง, 2548) ซึ่งสอดคล้องกับนโยบายของกระทรวงการท่องเที่ยวและกีฬาที่ต้องการพัฒนาการตลาดเชิงอิเล็กทรอนิกส์และระบบให้ข้อมูลการท่องเที่ยวในประเทศไทย การพัฒนาระบบสารสนเทศเพื่อให้ข้อมูลการท่องเที่ยวในประเทศไทยนี้มีวัตถุประสงค์เพื่อประโยชน์ในการสร้างมูลค่าให้แก่ธุรกิจการท่องเที่ยวและนักท่องเที่ยว เนื่องจากระบบทางอิเล็กทรอนิกส์สามารถสนับสนุนการให้ข้อมูลแก่นักท่องเที่ยวโดยอัตโนมัติ โดยการตอบคำถามของนักท่องเที่ยวที่ใส่เข้ามาในระบบ โดยที่นักท่องเที่ยวไม่จำเป็นต้องไปค้นหาข้อมูลผ่านระบบค้นหาข้อมูล เช่น Google เนื่องจากบางครั้งระบบค้นหาข้อมูลดังกล่าวให้ผลลัพธ์การค้นหาที่ไม่ตรงกับสิ่งที่นักท่องเที่ยวต้องการหรือให้ผลลัพธ์การค้นหาที่มากเกินไป ดังนั้นนักท่องเที่ยวต้องเสียเวลาเลือกดูข้อมูลเหล่านั้นว่าข้อมูลใดเป็นข้อมูลที่ตนเองสนใจ และข้อเสียอีกประการหนึ่งคือนักท่องเที่ยวไม่รู้จะใช้คำศัพท์ใดในการค้นหาจึงจะทำให้ได้ข้อมูลที่ตนเองต้องการ เนื่องระบบค้นหาข้อมูลหรือ search engine จะให้ใส่คีย์เวิร์ด (keyword) เท่านั้น ส่วนการให้ข้อมูลด้านการท่องเที่ยวของประเทศไทยผ่านทางเว็บไซต์ต่างๆ ที่มีอยู่ในปัจจุบันยังมีประสิทธิภาพไม่เพียงพอที่จะนำเสนอข้อมูลให้ตรงกับความสนใจที่แท้จริงของนักท่องเที่ยว และระบบสนับสนุนการสืบค้นข้อมูลด้านการท่องเที่ยว

ยังไม่ได้เท่าที่ควร ตัวอย่างเช่น นักท่องเที่ยวต้องการค้นหาข้อมูลเกี่ยวกับโบราณสถานในประเทศไทยโดยใช้คำสำคัญในการค้นหาว่า “อารามหลวง” ระบบค้นหาข้อมูลในปัจจุบันเช่น Google จะค้นพบเฉพาะเอกสารที่ปรากฏคำสำคัญนี้เท่านั้น เอกสารอื่นๆ ที่มีเนื้อหา (concept) เดียวกันแต่ไม่มีคำสำคัญเหล่านี้ปรากฏอยู่ในเอกสาร ระบบค้นหาข้อมูลก็จะไม่สามารถค้นพบและนำเสนอต่อนักท่องเที่ยวได้ มีการพัฒนาวิธีการขึ้นมาหลายวิธีแต่ก็ยังไม่สามารถให้ผลลัพธ์ที่ตรงความต้องการเนื่องจากระบบยังมีปัญหาอย่างเช่น ระบบที่ยังไม่สามารถวิเคราะห์คำถามเชิงความหมายได้ (Non-Semantic) เป็นระบบที่ใช้วิธีการเทียบตัวอักษรกับฐานข้อมูลโดยตรง คำที่ค้นหาต้องเป็นคำที่มีอยู่ในฐานข้อมูลและต้องสะกดคำอย่างถูกต้องเท่านั้นระบบถึงจะแสดงคำตอบออกมาได้

ดังนั้นเพื่อเป็นการสร้างข้อได้เปรียบและทำให้การท่องเที่ยวของไทยเกิดการพัฒนาและมีการขยายตัวนั้นจำเป็นต้องมีอย่างอื่นที่ต้องมีการดำเนินนโยบายการพัฒนาประเทศไทยที่มีจุดมุ่งหมายเพื่อให้นักท่องเที่ยวได้รับข้อมูลเกี่ยวกับการท่องเที่ยวอย่างเพียงพอและถูกต้องโดยอาศัยคอมพิวเตอร์เป็นเครื่องมือหลักในการให้ข้อมูลการสอบถามจากผู้ใช้โดยอัตโนมัติ และช่วยประหยัดค่าใช้จ่ายในการจ้างแรงงานมนุษย์ในการทำหน้าที่ดังกล่าวอีกด้วย ซึ่งในการศึกษาครั้งนี้มุ่งเน้นจังหวัดพิษณุโลก เนื่องจากจังหวัดพิษณุโลกนี้มีความหลากหลายทางด้านสถานที่ท่องเที่ยว ศิลปวัฒนธรรมและวิถีชีวิตของคนในจังหวัดนี้นับได้ว่ามีความสำคัญและน่าสนใจเป็นอย่างยิ่งที่จะพัฒนาให้เป็นแหล่งท่องเที่ยวเชิงวัฒนธรรมได้ ดังปรากฏให้เห็นเป็นเอกลักษณ์ในแต่ละพื้นที่เช่นในเรื่องของความเชื่อ พิธีกรรม ศิลปะการแสดง วิถีชีวิตและผู้คนที่หลากหลายของจังหวัดพิษณุโลก ผู้วิจัยจึงได้เลือกศึกษาการพัฒนาแบบสอบถามข้อมูลการท่องเที่ยวอัจฉริยะ โดยเน้นการให้ข้อมูลแหล่งท่องเที่ยวในจังหวัดพิษณุโลก ซึ่งการศึกษาดังกล่าวนี้น่าจะสามารถนำมาใช้เป็นกรอบในการพัฒนาและเป็นต้นแบบในการพัฒนาระบบสอบถามข้อมูลการท่องเที่ยวด้านอื่นๆ ทั่วประเทศต่อไปได้

## 1.2 จุดมุ่งหมายของการศึกษา

- 1) พัฒนาด้านแบบระบบสอบถามข้อมูลอัตโนมัติที่ผู้ใช้สามารถใช้ภาษาธรรมชาติภาษาไทย (Natural Language) ในการสอบถามข้อมูลการท่องเที่ยวเชิงวัฒนธรรมได้
- 2) พัฒนาเทคนิควิธีการวิเคราะห์คำถาม และคำตอบโดยใช้วิธีการกำหนดรูปแบบเชิงความหมาย (Semantic Pattern)

## 1.3 ขอบเขตของงานวิจัย

### ขอบเขตด้านระบบ

- 1) สามารถรับคิวรีโดยใช้ภาษาธรรมชาติได้
- 2) สามารถตัดคำภาษาไทยได้ถูกต้องตามบริบท (Context) ของประโยค



- 3) สามารถเข้าใจความหมายของประโยคและติดต่อสื่อสารหรือโต้ตอบกับผู้ใช้ได้อย่างมี
- 4) ประสิทธิภาพ

#### ขอบเขตด้านข้อมูล

- 1) การศึกษาครั้งนี้มุ่งเน้นข้อมูลสถานที่ท่องเที่ยวในจังหวัดพิษณุโลก

#### 1.4 ประโยชน์ที่คาดว่าจะได้รับ

- 1) ได้ระบบตอบคำถามสถานที่ท่องเที่ยวในจังหวัดพิษณุโลกที่ให้บริการได้ตลอด 24 ชั่วโมง
- 2) นักท่องเที่ยวสามารถใช้ระบบให้การค้นหาข้อมูลท่องเที่ยวในจังหวัดพิษณุโลกได้ตลอดเวลา
- 3) สามารถเผยแพร่สถานที่ท่องเที่ยวในจังหวัดพิษณุโลกให้กับผู้สนใจที่จะเดินทางมาท่องเที่ยวได้
- 4) ลดต้นทุนในการจ้างงานเพื่อให้ข้อมูลกับนักท่องเที่ยว
- 5) สามารถนำระบบตอบคำถามสถานที่ท่องเที่ยวในจังหวัดพิษณุโลกไปประยุกต์ใช้ในจังหวัดอื่นๆ ได้



## บทที่ 2

### ทบทวนวรรณกรรม

ความต้องการที่จะให้มีแหล่งข้อมูลที่ถูกต้องและเหมาะสมเกี่ยวกับการท่องเที่ยวเชิงวัฒนธรรมเพื่อใช้ในเผยแพร่ความรู้และข้อมูลต่างๆ แก่นักท่องเที่ยวที่นั่นทำให้ ระบบสอบถามข้อมูลอัตโนมัติเป็นวิธีการหนึ่งที่เหมาะสมถูกนำมาใช้ในการสนับสนุนแนวคิดดังกล่าว การพัฒนาระบบระบบสอบถามข้อมูลโดยเฉพาะเกี่ยวกับการท่องเที่ยวเชิงวัฒนธรรมมีความสำคัญอย่างยิ่ง ปัจจุบันนักท่องเที่ยวนิยมหาข้อมูลเกี่ยวกับการท่องเที่ยวผ่านระบบคอมพิวเตอร์ (Rayman-Bacchus & A Molina, 2001) พบว่า คนร้อยละ 50 ใช้อินเทอร์เน็ตในที่ทำงานหาข้อมูลการท่องเที่ยว (Robinson, 1997) ยังได้ศึกษาเรื่องทางเลือกสำหรับนักท่องเที่ยวและให้ความเห็นว่าธุรกิจท่องเที่ยวเป็นองค์กรที่ต้องการระบบสารสนเทศที่ดีและมีคุณภาพเพื่อเพิ่มการติดต่อสื่อสาร และให้ข้อมูลเกี่ยวกับแหล่งท่องเที่ยวของตนแก่นักท่องเที่ยว นอกจากนี้จากการสำรวจของ ATLAS (2004) พบว่า แหล่งข้อมูลสำคัญของการท่องเที่ยวเชิงวัฒนธรรมมาจากอินเทอร์เน็ต มีผลสำรวจพบว่ามีองค์กรหลายแห่งที่ใช้ระบบสารสนเทศในการให้บริการข้อมูลและแนะนำเส้นทางท่องเที่ยวแก่นักท่องเที่ยว เช่น มีการใช้ตู้สาขาอิเล็กทรอนิกส์เพื่อให้ข่าวสารแก่นักท่องเที่ยวและอนุญาตให้นักท่องเที่ยวจองห้องพัก ร้านอาหาร เช่ารถ และจองการท่องเที่ยว การศึกษาเกี่ยวกับกลยุทธ์ในการนำเทคโนโลยีสารสนเทศมาใช้ในอุตสาหกรรมการท่องเที่ยวของ (Buhalis, 1998) ได้ชี้ให้เห็นถึงความสำคัญของการนำระบบสารสนเทศมาเป็นเครื่องมือในการดำเนินธุรกิจที่ทำให้เกิดความได้เปรียบในการแข่งขันทั้งในด้านช่องทางการเข้าถึงกลุ่มเป้าหมายและการสร้างความแตกต่างให้กับสินค้าและบริการต่างๆ เพื่อตอบสนองความพึงพอใจของนักท่องเที่ยว ดังนั้นธุรกิจการท่องเที่ยวใดที่ใช้ระบบสารสนเทศจะสามารถเก็บรวบรวมข้อมูลจากนักท่องเที่ยวและพฤติกรรมการท่องเที่ยวของนักท่องเที่ยวได้ จะเห็นได้ว่าการพัฒนาระบบสอบถามข้อมูลท่องเที่ยวอัจฉริยะจะสามารถช่วยกระตุ้นให้เกิดการท่องเที่ยวเชิงวัฒนธรรมมากยิ่งขึ้น

งานวิจัยที่เกี่ยวข้องกับในระบบตอบคำถามนั้นมีอยู่หลายฉบับ แต่ละฉบับก็จะบอกวิธีการใช้ภาษาธรรมชาติในการประมวลผลด้วยคอมพิวเตอร์ที่แตกต่างกันไป ซึ่งแต่ละระบบก็มีทั้งแบบการวิเคราะห์เชิงความหมาย (Semantic based technology) และไม่ใช้การวิเคราะห์เชิงความหมาย (Non-Semantic based technology) ตัวอย่างระบบที่มีการวิเคราะห์คำถาม



## 2.1. ไม่ใช่การวิเคราะห์เชิงความหมาย (Non-semantic base technology)

อรุณี (2548) ได้ทำระบบตอบคำถามโดยการนำประโยคที่เป็นภาษาธรรมชาติมาเปรียบเทียบกับกลุ่มคำกับข้อมูลที่เก็บในรูปแบบ XML ซึ่งเป็นระบบที่จะเน้นการตอบคำถามที่เกี่ยวกับเวลา เช่น เมื่อไร เวลาไหน วันไหน เดือนไหน ซึ่งยังไม่สามารถตอบคำถามประเภทอื่นๆ ได้ เช่น สถานที่ อย่างไร อะไร นอกจากนี้ระบบยังไม่สามารถวิเคราะห์คำถามเชิงความหมายได้ แต่การทำงานเป็นในลักษณะของการเปรียบเทียบอักขระ (String matching) (ณรงค์, 2552) ได้สร้างระบบแชตโรบอท (Chat robot) ภาษาไทยโดยการรับคำถามจากผู้ใช้เข้าสู่ระบบ โดยเริ่มจากระบบการตัดคำ และนำแต่ละคำที่ได้ ไปค้นหาจากองค์ความรู้ (Knowledge base) เรื่องอัญมณี อย่างไรก็ตามระบบนี้ยังไม่มีกลไกที่จะจัดการ (Handle) คำถามที่ไม่เคยพบมาก่อน เนื่องจากระบบจะตอบคำถามเฉพาะคำถามที่อยู่ในฐานข้อมูลเท่านั้น ดังนั้นระบบนี้จึงมีข้อจำกัดหากผู้ใช้ใส่คำถามที่ไม่ปรากฏในฐานข้อมูลของระบบ (วลัยลักษณ์, 2553) พัฒนาระบบผู้เชี่ยวชาญในการตอบคำถามอัตโนมัติเพื่อช่วย Help desk ตอบคำถามผู้ใช้ที่มีปัญหาทางด้าน ซอฟต์แวร์ ฮาร์ดแวร์ เน็ตเวิร์ค ซึ่งจะจำแนกประเภทปัญหาใหม่และค้นคืนในกลุ่มปัญหานั้นโดยวิธีวัดความคล้ายคลึงของเอกสารด้วยวิธีการ Cosine Similarity เนื่องจากข้อมูลที่เป็นคำหลักมีจำนวนมาก ควรเพิ่มคำที่มีความหมายคล้ายคลึงกันและทำการลดขนาดจำนวนคำด้วยเทคนิคการเลือกคุณลักษณะ (Feature Selection) ก่อนการจำแนกข้อมูลปัญหา (ทศพล, 2552) พัฒนาระบบตอบคำถามที่เกี่ยวกับเนื้อหาที่อยู่เอกสารต่างๆ ในฐานข้อมูลของระบบ นักวิจัยได้นำวิธีการ Rule-based (เงื่อนไข If-then) มาใช้ในการวิเคราะห์คำถามและค้นหาคำตอบให้กับผู้ใช้ นอกจากนี้ยังนำ WordNet (คือฐานข้อมูลคำศัพท์ออนไลน์ภาษาอังกฤษที่ออกแบบตามทฤษฎีทางภาษาศาสตร์) มาช่วยในการแบ่งประเภทของคำตอบและนำเทคนิค (Bag of Words Bow) มาใช้ในการคำนวณหาความสัมพันธ์ระหว่างประโยคคำถามและคำตอบ อย่างไรก็ตามวิธีการ Rule-based ที่ใช้ในงานวิจัยนี้ยังไม่รองรับการวิเคราะห์คำถามเชิงความหมายเช่นกัน

ชัชวาลและคณะ (2551) ได้พัฒนาระบบอับดุล หรือ ระบบบริการข้อมูลเชิงบูรณาการผ่านระบบสนทนาออนไลน์ ที่สามารถตอบคำถามผู้ใช้ได้เสมือนมนุษย์ ระบบจะให้บริการข้อมูลสาธารณะทั่วไปแก่ผู้ใช้ ซึ่งมีประโยชน์ที่ผู้ใช้ไม่ต้องเข้าถึงแหล่งข้อมูลโดยตรง หัวใจหลักของงานวิจัยนี้คือกระบวนการตัดคำ (Word Segmentation) และแปลงรูปแบบภาษา (Normalization) ให้อยู่ในรูปแบบที่คอมพิวเตอร์สามารถเข้าใจได้ และใช้ AIML Engine ในการพิจารณาคำตอบ แต่เนื่องจากระบบอับดุลนี้ใช้เพื่อตอบคำถามทั่วไป ดังนั้นจึงไม่สามารถตอบคำถามเชิงลึกได้โดยเฉพาะเรื่องของการท่องเที่ยว (Ravichandran & Hovy, 2002) ได้ พัฒนาการนำวิธีการเรียนรู้รูปแบบของคำถามและคำตอบโดยอัตโนมัติเป็นภาษาอังกฤษ ซึ่งจะมีฐานข้อมูลที่ระบุคำที่ต้องการจากอินเทอร์เน็ตด้วยมือ และจะสร้างรูปแบบคำถามและคำตอบด้วยเพิ่มข้อมูลและทำให้เป็นมาตรฐาน มีการคำนวณเปอร์เซ็นต์ของแต่ละรูปแบบและค่าเฉลี่ยของประเภทคำถาม รูปแบบที่ได้จะนำไปหาคำตอบ วิธีการนี้จะทำได้ดีในประเภทคำถาม BIRTHDATE หรือ LOCATION แต่ทำได้ไม่ดีในคำถาม DEFINITION ซึ่งเราควรนำไปพัฒนาต่อให้สามารถตอบคำถามได้ครอบคลุมและตรงความต้องการ

จากระบบต่างๆ ที่กล่าวมาข้างต้นนั้นเป็นระบบที่เป็น Non-semantic base technology ซึ่งทำให้การเข้าใจความหมายของคำถามนั้นมีข้อจำกัดอยู่มาก เช่น ผู้ใช้ใช้คำถามที่ไม่เหมือนกับคำถามในฐานข้อมูลระบบจะไม่สามารถหาคำตอบให้ผู้ใช้ได้ หรือหากผู้ใช้สะกดคำผิด ระบบก็ไม่สามารถหาคำตอบได้จากฐานข้อมูลเช่นกัน ดังนั้นนักวิจัยหลายท่านจึงพยายามพัฒนาเทคโนโลยีการวิเคราะห์คำถามเชิงความหมายขึ้น เพื่อลดข้อจำกัดดังกล่าว

## 2.2. การวิเคราะห์เชิงความหมาย (Semantic based technology)

ฉัตรชัยและคณะ (2551) ได้ทำระบบตอบคำถามโดยการนำคำที่ผ่านการตัดคำมาแล้วนำไปสร้างดัชนี จากนั้นจึงนำมาเข้ากระบวนการ Vector Space Model หาคำนำหนักของคำ และนำ Term ที่ได้ในรูปแบบของเมตริกมาเข้ากระบวนการ Latent Semantic Indexing (LSI) โดยผ่านกระบวนการ Semi-Discrete Decomposition (SDD) ซึ่งเป็นกระบวนการลดขนาดเมตริกซ์ (Matrix) และนำผลที่ได้มาเข้ากระบวนการเปรียบเทียบความคล้ายคลึงของเอกสารระหว่างคำถามและคำตอบ อย่างไรก็ตามงานวิจัยนี้เป็นการวิเคราะห์ความหมายจากการคำนวณทางสถิติ การค้นหาคำถามนั้นยังไม่ค่อยมีประสิทธิภาพมากนัก เนื่องจากความไม่ละเอียดของข้อมูลคำถาม, คำเหมือนคำคล้ายและคำรากศัพท์เมื่อเปรียบเทียบกับ การวิเคราะห์ด้วยเทคนิคอื่นๆ เช่น ออนโทโลยี ธนากร (2550) ได้ทำระบบตอบปัญหาโดยระบบอนุญาตให้ผู้ใช้ใส่คำถามเป็นคำสำคัญเท่านั้น (ไม่เป็นประโยค) นำคำสำคัญไปหาข้อมูลที่เก็บในออนโทโลยี และได้คำตอบที่มีความสัมพันธ์กับคำถามเชิงความหมาย (Semantic related) ข้อจำกัดของระบบนี้คือไม่สามารถที่จะรองรับคำถามที่อยู่ในรูปแบบประโยคได้ (Guo & Zhang, 2008) ได้นำ Semantic web และ Ontology มาใช้เป็นเครื่องมือที่สำคัญในการสร้างคลังความรู้ด้านคำศัพท์ด้วยออนโทโลยี เพื่อนำไปใช้ในการตอบคำถาม ข้อดีของงานวิจัยนี้คือ ระบบสามารถเรียนรู้คำถามที่ไม่อยู่ในฐานข้อมูลและสามารถขยายฐานความรู้ได้ โดยอาศัยความสัมพันธ์ของคำต่างๆ ในดัชนี แต่ระบบนี้สนับสนุนคำถามเฉพาะภาษาจีนเท่านั้น (Hao & Wenyin, 2008) ระบบได้ตอบสำหรับคนและคอมพิวเตอร์ ด้วยการให้ Semantic Pattern ในการทำให้คอมพิวเตอร์เข้าใจคำถามซึ่งจะนำคำถามที่ได้มาสร้างรูปแบบ (Pattern) ของตัวเองแล้วก็ไปเทียบกับคลังรูปแบบ ที่สร้างไว้ เพื่อวิเคราะห์ความหมายของคำถาม และดึงคำตอบที่เกี่ยวข้องออกมาแสดง (Shen, 2007) นำบทบาทของความหมายมาใช้เพื่อทำให้ระบบถามตอบมีประสิทธิภาพยิ่งขึ้น โดยการสร้างเฟรมเวิร์กสำหรับแยกคำตอบซึ่งดูจากพฤติกรรมทางความหมาย และมีการหาค่าความหมายของปัญหาและการแยกคำตอบโดยใช้กราฟแมชชีน งานวิจัยนี้ยังไม่สามารถทำให้ระบบสามารถเรียนรู้คำได้ด้วยตัวเอง ในอนาคตสามารถพัฒนาให้สามารถเรียนรู้คำที่เคยใช้ได้

## 2.3. Query Recommendation (การจัดการคำถามที่ไม่รู้จัก)



Pan et al. (2007) ได้พัฒนาระบบแนะนำคำถามให้กับผู้ใช้ เมื่อระบบไม่เข้าใจคำถามบางคำในคำถาม ระบบจะถามข้อมูลเพิ่มเติม ระบบนี้ได้นำวิธีการ Hybrid Query Recommendation ซึ่งเป็นการผสมผสานระหว่าง วิธีการเทียบคำจากฐานข้อมูลถ้าเจอความสัมพันธ์ของคำถามแต่ถ้าไม่เจอระบบจะสร้างคำแนะนำด้วยระบบเอง แต่งานวิจัยนี้ก็ยังมียุทธศาสตร์ค้นหาคำแนะนำจากคำศัพท์และโครงสร้างประโยค งานวิจัยที่สามารถขยายองค์ความรู้ของระบบได้อัตโนมัติ เช่น Hao et al. (2008) พัฒนารูปแบบการขยายฐานความรู้ด้วย Semantic Pattern Generation Guide (SPG-Guide) โดยการให้ผู้ใช้เลือกรูปแบบคำถามจากตัวเลือกที่มีอยู่ และ Admin จะเป็นคนตัดสินใจว่าถูกหรือผิดแล้วจึงนำไปเข้า Question Target category (QT-Category) เพื่อนำไปใช้ค้นหาคำตอบต่อไป ข้อเสียก็คือระบบไม่ได้มีส่วนวิเคราะห์คำถาม เพราะเป็นเพียงการตัดสินใจของมนุษย์ ดังนั้นจึงควรพัฒนาให้ระบบสามารถวิเคราะห์คำถามเองได้ Guo et al. (2008) นำองค์ความรู้ที่สร้างด้วยออนโทโลยีและเทคโนโลยีของ Semantic Web มาสร้างเป็นฐานข้อมูลคำตอบ และใช้วิธีการของออนโทโลยีมาใช้เพื่อขยายฐานความรู้ตามหลักการความสัมพันธ์ของออนโทโลยี ซึ่งวิธีทำให้ระบบสามารถตอบคำถามที่ซ้ำๆ กันได้อย่างรวดเร็ว แต่วิธีการนี้ใช้เพียงความสัมพันธ์ของคำถามแต่ไม่ได้เพิ่มลงในฐานข้อมูล ยุทธศักดิ์ (2553) พัฒนาระบบขยายฐานความรู้อัตโนมัติโดยใช้ความสัมพันธ์ของคำในเอกสารที่มีในฐานข้อมูล นำมาสร้างเป็น Syntax Tree จากนั้น แปลงให้อยู่ในรูปแบบของ Conceptual Graph แล้วนำมาสร้างความสัมพันธ์เพื่อวิเคราะห์สร้างคำถามและคำตอบโดยอัตโนมัติ งานวิจัยนี้เป็นการวิเคราะห์ข้อมูลจากเอกสารที่มีอยู่แล้วไม่ใช่การโต้ตอบกันกับระบบ จากงานวิจัยข้างต้นเป็นการขยายฐานความรู้หลากหลายรูปแบบที่น่าสนใจ และวิธีการพัฒนาระบบสนทนาที่ยังมีข้อจำกัดอีกอย่างคือระบบไม่เข้าใจคำถาม หรือไม่พบคำศัพท์นั้นๆ (Google, 2006) ได้นำ retrieval-based query recommendation มาใช้เพื่อช่วยสะกดคำที่ผิดของ Chrome Browser แต่ feature นี้ ช่วยแนะนำคำที่ถูกต้อง เพื่อตัดสินใจว่าจะใช้คำนั้นหรือไม่ ดังนั้นจึงควรนำไปพัฒนาให้ใช้กับประโยคต่อไป Dupret et al. (2006) ได้เสนอวิธีการแนะนำคำถามด้วยวิธีการ Click Through data ซึ่งเป็นเซตของค่าการระบุความคล้ายกันของคำถาม โดยงานวิจัยนี้มุ่งเน้นไปในทางการปรับปรุงคำถามให้ดีขึ้นมากกว่าความสัมพันธ์ของคำถาม แต่งานวิจัยนี้จะมุ่งเน้นคำถามเฉพาะทางมากกว่าคำถามทั่วไปและเป็นภาษาอังกฤษ เราจึงความพัฒนาต่อให้สามารถแนะนำคำถามทั่วไปในภาษาไทย Baeza et al. (2004) พัฒนารูปแบบการแนะนำคำถามจากคำถามเดิมที่เคยใช้งานมาแล้วเพื่อนำมาเป็นตัวเลือกให้กับผู้ใช้ด้วยการใช้ความสัมพันธ์ของคำถาม และกระบวนการจัดกลุ่ม (Clustering) โดยใช้เนื้อหาของคำถามเดิมที่เก็บไว้ใน Query log ซึ่งระบบนี้ไม่ได้ค้นหาเฉพาะความสัมพันธ์ของคำถามเพียงเท่านั้น แต่ยังสามารถจัดอันดับและบันทึกไว้ได้อีกด้วย

จากงานวิจัยที่เกี่ยวข้อง สามารถเปรียบเทียบงานวิจัยและสรุปได้ดัง

ตารางที่ 1 สรุปงานวิจัยที่เกี่ยวข้องกับระบบตอบคำถาม

Author	Technique	Non-Semantic base	Semantic base	Semantic Pattern Process	Query Recommendation
(อรุณี, 2548)	String Matching	/	X	X	X
(ทศพล, 2552)	Rule Base	/	X	X	X
(ชัชวาล สังกิต ตระการ et al., 2551)	AIML Engine	/	X	X	X
(จักรชัย, 2551)	Latent Semantic Indexing	X	P	X	X
(Guo & Zhang, 2008)	Ontology	X	/	X	X
(Hao & Wenyin, 2008)	Ontology	P	/	/	X
(Pan & Shaw, 2007)	Hybrid Query reccomment	P	/	X	/

/=มี X=ไม่มี P=รองรับการทำงานบางส่วน

จากการศึกษาวิจัยที่เกี่ยวข้องกับระบบตอบคำถามพบว่าระบบที่มีอยู่ในปัจจุบันยังไม่สามารถรองรับความต้องการของผู้ใช้ได้ทั้งหมด จึงสรุปได้ดังนี้

1) ระบบที่มีอยู่โดยส่วนมากใช้วิธีเทียบตัวอักษรเมื่อระบบเจอคำที่ไม่มีในฐานข้อมูล ระบบจะไม่สามารถประมวลผลคำนั้น และไม่สามารถหาคำตอบได้

2) ระบบส่วนใหญ่ไม่สามารถแนะนำคำถามให้แก่ผู้ใช้เมื่อผู้ใช้พิมพ์ผิด หรือระบบหาคำตอบไม่เจอ ดังนั้นเราจึงควรพัฒนาเพื่อให้ระบบมีประสิทธิภาพมากยิ่งขึ้น

ด้วยเหตุผลดังกล่าวผู้วิจัยจึงเกิดแนวคิดที่จะพัฒนาระบบตอบคำถามที่สามารถรองรับความแม่นยำของคำตอบ การผิดพลาดของคำถาม โดยใช้วิธีการเชิงความหมาย (Semantic Base) โดยใช้ฐานความรู้ที่เป็น

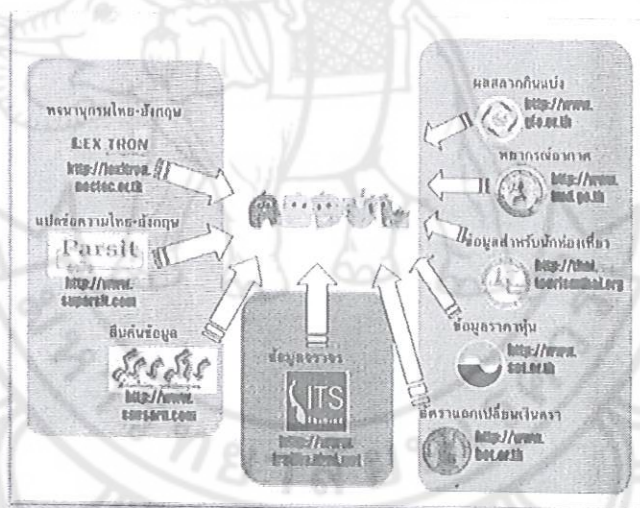


ออนโทโลยีเป็นเครื่องมือช่วยในการหาความหมายของคำและใช้วิธีการแนะนำคำถามมาใช้ในการแก้ไขข้อผิดพลาดของคำถาม

## 2.4. ตัวอย่างระบบถามตอบที่มีในปัจจุบัน

### 2.4.1. อับดุล (Abdul)

ระบบอับดุล (NECTEC, 2008) นั้นเป็นระบบบริการให้ข้อมูลแก่ผู้ใช้ในการพัฒนาระบบอับดุลจะมีขั้นตอนการตัดคำ (Word Segmentation) กระบวนการในการตัดคำของอับดุลนั้น ถือเป็นขั้นตอนสำคัญในการที่จะทำให้ระบบอับดุลเข้าใจความหมายของประโยคที่ทางผู้ใช้สนทนาเข้าได้อย่างถูกต้องและหัวใจหลักของระบบอับดุล คือ ด้านเทคโนโลยีถามตอบ (Question Answering) ถือว่าเป็นหัวใจหลักที่สำคัญจะทำให้ระบบสามารถค้นหาคำตอบจากองค์ความรู้ต่างๆที่มี เพื่อให้ในการตอบคำถามมีความถูกต้องมากที่สุด ระบบอับดุลให้บริการข้อมูลในหลายๆ ด้าน เช่น ราคาทอง ราคาน้ำมัน สภาพอากาศ ดวง หุ้น ผลออกสลากกินแบ่งรัฐบาล ข้อมูลที่พัก ร้านอาหาร เป็นต้น แสดงดังรูปที่ 1 (NECTEC, 2008)



รูปที่ 1 ลักษณะการให้บริการของระบบอับดุลในปัจจุบัน

แต่ระบบอับดุลนั้นไม่สามารถให้ข้อมูลบางส่วนได้อย่างละเอียดเพราะส่วนใหญ่ระบบอับดุลจะให้ข้อมูลในมุมมองที่กว้างออกไป ตัวอย่างเช่น สถานที่ท่องเที่ยว

จากรูปที่ 2 จะเห็นได้ว่าระบบอับดุลนั้นไม่สามารถในคำตอบกับผู้ใช้ได้ตามที่ผู้ใช้ได้ถามกับระบบอับดุล ดังรูปที่ 2 แต่ถ้าผู้ใช้มีการถามเฉพาะสถานที่ท่องเที่ยวที่ผู้ใช้ต้องการจะไปเที่ยว เช่น น้ำตก หรือ วัด ระบบอับดุลจะให้ข้อมูลดังรูปที่ 3



[Text input field] Say

สถานที่ท่องเที่ยว



การท่องเที่ยว



สถานที่ท่องเที่ยวในประเทศไทย



ใบทรานสคริปต์



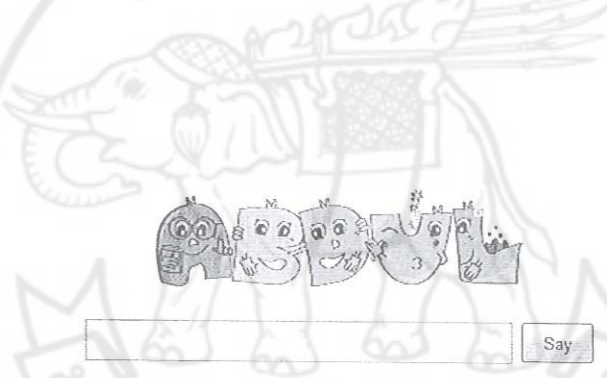
หาคำตอบไม่เจอ ช่วยเพิ่มให้หน่อยครับ



หาคำตอบไม่เจอ ช่วยเพิ่มให้หน่อยครับ



### รูปที่ 2 การถาม ตอบ สถานที่ท่องเที่ยว



[Text input field] Say

วัดใหญ่



วัด

ส่วนหนึ่งของ

ใบรู้เหมือนกันครับ สองถามใหม่



ข้อมูลจาก : วิกิพีเดีย

ในจังหวัดพิษณุโลกมีน้ำตกอะไรบ้าง



น้ำตก

น้ำตก หรือ โคน ใน ภาษาใต้ เป็นลักษณะทาง ภูมิศาสตร์ ที่เกิดจาก น้ำป่า ไหลลงมาจากจุดที่สูงกว่า ทำให้เป็นลักษณะทัศนียภาพของ น้ำตก น้ำตกที่พบได้ตาม อุทยาน ขึ้น

อันนี้ตอบไม่ได้อะครับ ช่วยใส่คำตอบให้ผมหน่อย



ข้อมูลจาก : วิกิพีเดีย

### รูปที่ 3 การถาม ตอบ ชื่อสถานที่ท่องเที่ยว

การให้ข้อมูลในการถามถึงสถานที่ท่องเที่ยวแบบระบุชื่อสถานที่ท่องเที่ยวไป เช่น น้ำตก หรือ วัด เป็นต้น ระบบอัจฉริยะจะให้ข้อมูลเกี่ยวกับความหมายของประโยคนั้นๆ แต่ถ้าผู้ใช้ถามข้อมูลที่แคบลงไปอีก เช่น "ในจังหวัดพิษณุโลกมีน้ำตกอะไรบ้าง" จะได้คำตอบ ดังรูปที่ 3 ระบบอัจฉริยะก็จะไม่สามารถให้ข้อมูลนั้นได้

The screenshot shows a chatbot interface with a search bar at the top containing the word "อาหาร". Below the search bar, there are two search results displayed in a list format. Each result includes a small icon of a person, a title, and a paragraph of text. The first result is titled "อาหาร" and the second is titled "ร้านอาหาร". To the right of each result, there is a button labeled "Say" and a small icon of a person. The background of the chatbot interface features a large, faint watermark of a Thai elephant.

#### รูปที่ 4 การถาม ตอบ ร้านอาหาร

จากรูปที่ 4 การให้ข้อมูลเกี่ยวกับร้านอาหารหรืออาหาร ระบบอัจฉริยะนั้นจะให้ข้อมูลแบบความหมายว่า "ร้านอาหารคืออะไร" และถ้ามีการถามถึงร้านอาหารที่แคบลงไปอีก เช่น "ร้านอาหารในจังหวัดพิษณุโลก" ระบบอัจฉริยะก็ไม่สามารถให้ข้อมูลเพิ่มเติมกับผู้ใช้ได้อีกดังรูปที่ 4 ส่วนการให้ข้อมูลแบบโรงแรมหรือที่พักของระบบอัจฉริยะจะได้ข้อมูลดังรูปที่ 5




 Say

โรงแรมในจังหวัดภูเก็ต



โรงแรม

โรงแรม หมายถึง สถานที่ประกอบการเชิงการค้าที่บริการผู้เดินทาง เพื่อบริการผู้เดินทางในเรื่องของที่พักอาศัย อาหาร และบริการอื่น ๆ ที่เกี่ยวข้องกับการพักอาศัยและเดินทาง หรืออาคารที่มีห้องนอนหลายห้อง ติดต่อเรียกรายกันอาคารหนึ่งหลังหรือหลายหลัง ซึ่งมีบริการต่าง ๆ เพื่อความสะดวกของผู้ที่มาพัก ซึ่งเรียกว่า "แขก" (guest)

ข้อมูลจาก : วิกิพีเดีย

หาศาลอมใจเจอ ช่วยเพิ่มให้หน่อยครับ



ที่พักในจังหวัดขอนแก่น



ที่พัก



ไม่รู้จักครับ



น่าจะ ไม่รู้จักครับ เพิ่มศาลอมใจให้หน่อย



## รูปที่ 5 การถาม ตอบ โรงแรมและที่พัก

### 2.4.2. ซิมซิมิ (Simsimi)

โปรแกรมซิมซิมิ (Simsimi) เป็นโปรแกรมเสริมเพื่อความสนุก ซิมซิมิพัฒนาโดยบริษัท ISMAKER INC. ตั้งแต่ พ.ศ.2545 และมีการเปิดตัวให้คนไทยรู้จักเมื่อ พ.ศ. 2554 โดยแท้จริงแล้วคือ Robot Chatting ส่วนใหญ่ผู้ใช้ที่เข้าไปสนทนากับโปรแกรมซิมซิมิเพื่อไปหาเพื่อนคุยในการผ่อนคลายในการสนทนากับโปรแกรม นั้นจะสนทนาไปในทางที่สนุก กวนๆ ขำๆ แต่ในการสนทนานั้นอาจจะใช้คำหรือประโยคที่ไม่สมบูรณ์ในการสนทนาและโปรแกรมสามารถเข้าใจและตอบกลับได้อย่างเสมือนจริง แต่โปรแกรมซิมซิมิจะไม่สามารถตอบข้อมูลที่เป็นการได้เพราะเป็นโปรแกรมที่พัฒนามาความสนุกเพลิดเพลิน ดังนั้นการเก็บรวบรวมข้อมูลเพื่อนำมาใช้เป็นคำตอบที่ใช้ในการสนทนากับผู้ใช้ได้มาจากการสุ่มกลุ่มวัยรุ่นที่ใช้คำในการโพสต์หรือแชท ทางโลกออนไลน์ เช่น Facebook Twitter Instagram และ Google จึงทำให้โปรแกรมซิมซิมิจะรู้ว่าคำตอบในฐานข้อมูลของโปรแกรมซิมซิมินั้นควรมีอะไรบ้าง ดังนั้นโปรแกรมซิมซิมิจึงได้มีวิเคราะห์คำถามมาเพื่อที่จะเก็บคำตอบที่จะใช้ในการสนทนากับผู้ใช้ที่อยู่ในฐานข้อมูลของโปรแกรมซิมซิมิ ในการที่ผู้ใช้เข้ามาใช้โปรแกรมซิมซิมินั้นก็ยังมีแนวโน้มที่จะใช้คำหรือประโยคที่ไม่สมบูรณ์แต่ปัจจุบันโปรแกรมซิมซิมิการพัฒนาอยู่เรื่อยๆ ในการสนทนา เช่น การสนทนาถามถึงสถานที่ท่องเที่ยว น้ำตก ร้านอาหาร โรงแรม หรือที่พัก ดังรูปที่ 6

SimSimi™ Talk Teach Quotes

Hi~!

แนะนำสถานที่ท่องเที่ยวในจังหวัดพิษณุโลก

ได้ครับ... อยากไปไหนครับ?

ป่าตอง

แซมหลายเค้อ

hey!

SimSimi learned 22,304,466 sentences.

Thai

Send

SimSimi™ Talk Teach Quotes

Hi~!

แนะนำร้านอาหารในจังหวัดพิษณุโลก

ร้านหมี่คนแซ่ลี อร่อยเว้อ อยู่ปากซอยทองหล่อ55ข้างร้านทอง

แนะนำโรงแรมในจังหวัดพิษณุโลก

ลองไปเช็คราคาที [www.checkroomrate.com](http://www.checkroomrate.com)

SimSimi learned 22,304,470 sentences.

Thai

Send

รูปที่ 6 การสนทนาของซิมซิมิ



### บทที่ 3

#### ทฤษฎีที่เกี่ยวข้อง

จากการศึกษางานวิจัยที่เกี่ยวข้อง ผู้วิจัยจึงเกิดแนวคิดที่จะพัฒนาระบบค้นหาข้ามภาษาสำหรับภาษาไทยและภาษาอังกฤษโดยใช้ออนโทโลยีเป็นฐานความรู้ ซึ่งได้ศึกษาถึงทฤษฎีต่างๆ ที่จะนำมาใช้ในงานวิจัยนี้ ซึ่งทฤษฎีที่สำคัญมีดังต่อไปนี้

- 1) กฎทางอักขระ
- 2) การตัดคำแบบ Longest matching
- 3) การตัดคำด้วยกระบวนการย้อนกลับ
- 4) การตัดคำด้วยเทคนิคผสม

หัชทัย (2547) ในการสร้างระบบสอบถามข้อมูลอัตโนมัติ นั้น แนวคิดสำคัญอันหนึ่งของระบบคือการสร้างดัชนีที่มีประสิทธิภาพ โดยทำการดัชนีคำที่สำคัญๆ ของเอกสารไว้ ไม่เพียงแต่ทำดัชนีคำครบถ้วนเท่านั้น แต่ยังต้องเน้นที่การทำดัชนีที่ถูกต้องอีกด้วย คำในภาษาอังกฤษนั้น สามารถที่จะทำดัชนีได้ง่ายกว่าภาษาไทย เนื่องจาก สามารถตัดคำได้ง่ายเพราะแต่ละคำมีช่องว่างระหว่างคำ ทำให้ระบบตัดคำสามารถตัดคำได้ง่ายกว่าภาษาไทย ปัญหาหลักของภาษาไทยคือเป็นภาษาที่แต่ละคำไม่ได้ถูกแยกกันโดยใช้ช่องว่าง ตัวอย่างเช่น ตากลม (ตาก-ลม) และ ตากลม (ตา-ลม) ซึ่งเขียนเหมือนกันทุกประการ ปัญหาหลักที่เป็นสาเหตุให้ทำดัชนีไม่ถูกต้องคือ ลักษณะของภาษาไทยเป็นภาษาที่ไม่มีช่องว่างระหว่างคำ จึงทำให้การตัดคำ (Tokenizing) เพื่อทำดัชนีข้อมูลสำหรับการสืบค้นข้อมูลสารสนเทศทำได้ยาก ส่งผลให้ผลลัพธ์จากการสืบค้นข้อมูลสารสนเทศมีประสิทธิภาพต่ำ นอกจากนั้นลักษณะการตัดคำยังสามารถทำได้หลายแบบ และแต่ละรูปแบบทำให้ได้ความหมายของคำที่แตกต่างกันไป นอกจากนี้คำศัพท์ที่ไม่ปรากฏอยู่ในพจนานุกรมก็เป็นปัญหาที่ทำให้การตัดคำไม่สามารถทำได้ถูกต้อง ซึ่งนอกจากจะตัดคำที่ไม่มีในพจนานุกรมผิดแล้ว อาจส่งผลทำให้คำรอบๆ ข้างมีการตัดคำผิดไปด้วย ดังนั้นการสร้างระบบสอบถามข้อมูลอัตโนมัติ เพื่อให้เข้าใจความหมายที่แท้จริงของคำถามระบบจำเป็นต้องตัดคำในประโยคออกเป็นคำๆ เพื่อพิจารณาความหมายของคำตามบริบท (Context)

สมมติฐานที่สำคัญในการตัดคำภาษาไทยโดยใช้การสกัดคำเฉพาะ (Name entity) ร่วมกับพจนานุกรมคือ การตัดคำภาษาไทยโดยใช้การสกัดการสกัดคำเฉพาะร่วมกับพจนานุกรม จะทำให้เพิ่มความแม่นยำในการค้นหาสถานที่มากขึ้น เมื่อเปรียบเทียบกับการทำตัดคำโดยวิธีการแบบเดิมๆ ข้อดีของการสกัดคำเฉพาะ เป็นแนวทาง



ที่เหมาะสม สำหรับนำมาใช้ในการตัดสินใจแยกแยะประเภทคำ รวมไปถึงการเลือกคำตอบที่ดีที่สุดที่ได้จากแนวทางการใช้กฎและฐานความรู้ โดยความถูกต้องและความเหมาะสมของข้อมูลที่ใช้ในการเรียนรู้ การเตรียมฐานความรู้ที่เหมาะสม รวมไปถึงการเลือกคุณสมบัติ และขอบเขตของบริบทที่ใช้พิจารณาให้เหมาะสม จะส่งผลให้ระบบสามารถตัดสินใจได้อย่างถูกต้อง รวมทั้งยังสามารถโน้มนำเอาสิ่งที่เรียนรู้จากคลังเอกสารฝึกฝนระบบ มาใช้กับปัญหาที่ไม่เคยพบในการเรียนรู้มาก่อนได้

ระบบตอบคำถามจำเป็นต้องมีการตัดคำที่มีอยู่ในประโยคให้แยกจากกันเพื่อหาความหมายจากคำเหล่านั้น เพราะถ้าไม่ตัดคำแล้วระบบก็ไม่สามารถค้นคืนคำตอบได้ ในภาษาอังกฤษการตัดคำนั้นสามารถทำได้ง่ายกว่าภาษาไทยเพราะภาษาอังกฤษมีการเว้นวรรคคำอยู่แล้ว แต่ในภาษาไทยจะเขียนคำติดกันจนเป็นประโยคโดยไม่มีการเว้นวรรคใดๆมาเป็นตัวแบ่งแยกระหว่างหน่วยคำ ดังนั้นจึงทำให้การตัดคำภาษาไทยเป็นเรื่องที่ยุ่งยาก

อรุณี (2548) วิธีการตัดคำภาษาไทยด้วยคอมพิวเตอร์ก็คือการหาขอบเขตของคำแล้วเปรียบเทียบแต่ละคำกับพจนานุกรมซึ่งจะต้องมีคำศัพท์เก็บไว้เป็นจำนวนที่มากพอและต้องสามารถเพิ่มศัพท์ใหม่ลงในพจนานุกรมได้ด้วยจึงจะสามารถตอบสนองความต้องการในการใช้งานได้ เพราะเมื่อเวลาผ่านไปก็จะมีคำศัพท์ใหม่เกิดขึ้นมาตามยุคตามสมัย ปัญหาอีกประการหนึ่งของการตัดคำคือเอกสารข้อความต่าง ๆ มีการใช้คำเฉพาะเช่นชื่อคน ชื่อสถานที่ คำผสมต่าง ๆ ทำให้ในทางปฏิบัติไม่สามารถบรรจุคำศัพท์ที่จะใช้ประกอบการตัดคำลงไปในพจนานุกรมได้ทั้งหมด ฉะนั้นจำเป็นที่จะต้องมียุทธวิธีอื่น ๆ มาประกอบในการตัดคำ เช่น การตัดคำแบบกฎทางอักขระ วิธี Longest Matching และการตัดคำด้วยกระบวนการย้อนกลับ

### 3.1. กฎทางอักขระ

วิรัช (2536) อธิบายว่าการเขียนประโยคภาษาไทยนั้นไม่มีการเว้นวรรคระหว่างคำ ไม่มีการใช้เครื่องหมายวรรคตอนที่ชัดเจน ไม่มีตัวบ่งหน้าที่ทางไวยากรณ์ ไม่มีการใช้ตัวอักษรตัวเล็ก-ใหญ่ และไม่มีเครื่องหมายระบุจุดจบของประโยค แต่ภาษาไทยนั้นมีกฎทางอักขระของการประสมอักษร 3 ลักษณะคือ การประสมอักษร การเว้นวรรค และการขึ้นย่อหน้า ซึ่งทั้ง 3 ลักษณะนี้เป็นตัวพิจารณาในการเปรียบเทียบของคำในพจนานุกรม การขึ้นย่อหน้าใหม่เป็นตัวบ่งชี้ให้รู้ว่าเป็นการสิ้นสุดของข้อความก่อนหน้านั้น การเว้นวรรคเป็นตัวบ่งชี้ความเป็นไปได้ของการสิ้นสุดของคำหรือประโยค และกฎทางอักขระวิธีนี้เป็นตัวบ่งชี้ถึงความเป็นไปได้ในการที่จะแยกสายอักขระเพื่อพิจารณาคำ ดังนั้นมีการแบ่งอักขระใหม่เป็น 5 กลุ่ม ดังนี้

กลุ่มที่ 1 อักขระที่ไม่มีผลกระทบของการใช้พื้นที่ได้แก่ ั ำ ๅ ๆ ็ ่ ้ ๐

กลุ่มที่ 2 อักขระที่ต้องมีพยัญชนะตามเสมอ ได้แก่ ๑ , ๒ , ๓ , ๔ , ๕

กลุ่มที่ 3 อักขระที่จำเป็นจะต้องมีรูปพยัญชนะอยู่หน้าเสมอได้แก่ ๖ , ๗ , ๘

กลุ่มที่ 4 อักขระที่มีตัวการ์นต์ เมื่อเข้ารูขของกลุ่มคำเป็นพยัญชนะสุดท้ายที่ไม่มีการออกเสียง จึงไม่มีการนำมาเป็นอักษรแรกของคำ

กลุ่มที่ 5 กลุ่มอักขระที่เหลือทั้งหมด

จากการแบ่งอักขระเป็นกลุ่ม ๆ ดังกล่าวมาแล้วสามารถสรุปได้ว่า อักขระในกลุ่มที่ 1 ถึง 4 จะเป็นตัวที่ช่วยได้ลดปริมาณของการตัดคำที่ต้องไปตรวจสอบกับพจนานุกรมลงได้ ทำให้ระบบการตัดคำสามารถทำงานได้เร็วยิ่งขึ้น

### 3.2. การตัดคำแบบ Longest Matching

กระบวนการตัดคำแบบ Longest Matching มีวิธีการคือเมื่อนำมาใช้กับข้อความ “ความก้าวหน้าทางด้านวิทยาศาสตร์มีบทบาทสำคัญ” จะได้ผลดังนี้

- 1) ตรวจสอบข้อความทั้งหมดโดยถือว่าเป็นคำศัพท์เดียวกับคำในพจนานุกรม ซึ่งก็จะไม่พบคำศัพท์ดังกล่าว
- 2) ตัดอักขระชุดสุดท้ายออกเหลือข้อความว่า “ความก้าวหน้าทางด้านวิทยาศาสตร์มีบทบาทสำคัญ”
- 3) นำข้อความที่เหลือไปตรวจสอบกับพจนานุกรม ซึ่งก็ไม่พบคำศัพท์ดังกล่าว
- 4) ตัดอักขระชุดสุดท้ายออก จะเหลือข้อความ “ความก้าวหน้าทางด้านวิทยาศาสตร์มีบทบาทสำคัญ” ตัด “คั” ออกตามกฎของอักขระวิธี
- 5) นำข้อความที่เหลืออยู่ไปตรวจสอบกับพจนานุกรมถ้ามี
- 6) ข้อความดังกล่าวแสดงว่าได้ผลของการตัดคำ 1 คำ โดยวิธีดังกล่าวจะได้ผลลัพธ์ดังตารางที่ 2

ตารางที่ 2 อธิบายผลลัพธ์การตัดคำ Longest Matching ในแต่ละขั้นตอนของข้อความ “ความก้าวหน้าทางด้านวิทยาศาสตร์มีบทบาทสำคัญ”

ลำดับที่	ส่วนของคำที่ยาวที่สุด	ส่วนของข้อความที่ต้องตรวจสอบต่อ
1	ความก้าวหน้า	ทางด้านวิทยาศาสตร์มีบทบาทสำคัญ
2	ทาง	ด้านวิทยาศาสตร์มีบทบาทสำคัญ
3	ด้าน	วิทยาศาสตร์มีบทบาทสำคัญ
4	วิทยาศาสตร์	มีบทบาทสำคัญ
5	มี	บทบาทสำคัญ
6	บทบาท	สำคัญ
7	สำคัญ	



### 3.3. การตัดคำด้วยกระบวนการย้อนกลับ (Backtracking)

จากทฤษฎีการตัดคำแบบ Longest Matching ให้ความถูกต้องเพียง 80 % เท่านั้น ข้อผิดพลาดของระบบนี้คือ การเลือกขอบเขตของคำที่ยาวเกินไปตั้งแต่แรก ให้ความหมายของข้อความที่ผ่านกระบวนการตัดคำแบบ Longest Matching ผิดเพี้ยนไป ดังนั้นจึงได้มีการเสนอทฤษฎีการตัดคำด้วยกระบวนการย้อนกลับ โดยเน้นการตัดคำให้ได้จำนวนคำและคำที่ไม่มีในพจนานุกรมน้อยที่สุด ในระบบเหล่านี้ต้องการการประมวลผลที่มีความเร็วสูง โดยหลักการตัดคำเหล่านี้สามารถที่จะให้ผลการตัดคำถูกต้องถึง 100% (วิรัช, 2536) ได้อธิบายถึงวิธีการตัดคำด้วยกระบวนการย้อนกลับว่าจะต้องทำงานต่อเนื่องจากวิธีการ Longest Matching โดยหลังจากได้ผลการตัดคำมาแล้วจะนำคำที่ได้มาเข้ากระบวนการหาความเป็นไปได้ของคำทั้งหมดด้วยกระบวนการย้อนกลับ จากซ้ายไปขวา ตัวอย่างเช่นประโยค “กีฬาเป็นการออกกำลังกายอย่างหนึ่ง” หลังจากใช้กระบวนการ Longest Matching จะได้คำดังนี้ “กีฬาเป็นการ/ออกกำลังกาย/อย่างหนึ่ง” ในกระบวนการย้อนกลับจะกระทำที่คำว่า “กีฬา” ซึ่งเป็นคำที่ไม่สามารถแบ่งได้อีกจึงจบกระบวนการของคำนี้และกระทำที่คำถัดไปคือ คำว่า “เป็นการ” เมื่อทำการตรวจสอบสามารถแบ่งได้คือคำว่า “เป็น” และส่วนที่เหลือ “การออกกำลังกายอย่างหนึ่ง” มาทำการตรวจสอบต่อได้คำว่า “การออกกำลังกาย/อย่างหนึ่ง” เมื่อสิ้นการหาความเป็นไปได้ของการตัดคำทั้งหมดก็นำคำต่าง ๆ ที่ได้มาจัดเรียงเป็นประโยคพร้อมทั้งคำนวณค่าจำนวนครั้งในการตัดคำ (Cost) ของแต่ละเส้นทาง เรียงประโยคทั้งหมดตามจำนวนครั้งในการตัดคำ (Cost) ที่คำนวณได้จากตัวอย่างข้างต้นสามารถแสดงผลการตัดคำได้ดังตารางที่ 3 โดยจำนวนครั้งในการตัดคำ (Cost) สามารถคำนวณได้จาก จำนวนคำที่ตัดได้ถูกต้อง (x) บวกกับจำนวนคำที่ตัดแล้วผิดพลาด (y) คุณด้วยหก สามารถแสดงได้ดังสมการดังนี้  $Cost = x + 6y$  ค่าจำนวนครั้งในการตัดคำที่สูงแสดงถึงผลการตัดคำที่ไม่ดี ค่าจำนวนครั้งให้น้อยให้ผลในการตัดคำที่ดี

ตารางที่ 3 ผลการตัดคำและค่า Cost

ผลการตัดคำ	Cost
“กีฬา/เป็น/การออกกำลังกาย/อย่างหนึ่ง”	4
“กีฬา/เป็นการ/ออกกำลังกาย/อย่างหนึ่ง”	5
“กีฬา/เป็นการ/ออก/กำลังกาย/อย่างหนึ่ง”	5
“กีฬา/เป็นการ/ออกกำลัง/กาย/อย่าง/หนึ่ง”	6
“กีฬา/เป็นการ/ออกกำลัง/กา/ย/อย่าง/หนึ่ง”	7
“กีฬา/เป็นการ/ออ/ก/กำลังกาย/อย่างหนึ่ง”	11
“กีฬา/เป็นการ/ออกกำลัง/กาย/อย่าง/หนึ่ง”	12
“กีฬา/เป็นการ/ออ/ก/กำลังกาย/อย่าง/หนึ่ง”	12
“กีฬา/เป็นการ/ออ/ก/กำลัง/กาย/อย่าง/หนึ่ง”	18



### 3.4. การตัดคำด้วยเทคนิคผสม ( Hybrid Approach )

วิธีการตัดคำนี้เป็นการผสมผสานวิธีการตัดคำด้วยวิธี เทียบกฎทางอักขระกับพจนานุกรม การตัดคำแบบ Longest Matching และการตัดคำด้วยกระบวนการย้อนกลับ (Backtracking) ซึ่งสามารถเพิ่มประสิทธิภาพการแบ่งคำได้แม่นยำยิ่งขึ้น (วรศักดิ์ และคณะ, 2552) พัฒนาระบบตัดคำภาษาไทยด้วยเทคนิคไฮบริดซึ่งระบบนี้มีความสามารถหลายอย่างได้แก่

- 1) ตรวจสอบคำเฉพาะ( NameEntity ) ในฐานข้อมูลได้
- 2) ตรวจสอบคำย่อได้
- 3) สามารถเพิ่มคำเฉพาะลงในฐานข้อมูลได้

เนื่องจากโปรแกรมนี้อมีประสิทธิภาพในระดับที่ดี ดังนั้นผู้วิจัยจึงได้นำโปรแกรมนี้อมาช่วยในการตัดคำในระบบคำถาม เพราะโปรแกรมนี้อสามารถตัดคำที่เป็นคำเฉพาะได้ และไฟล์มีขนาดเล็กจึงใช้เวลาในการประมวลผลน้อยมาก แต่โปรแกรมนี้อยังมีข้อจำกัดเช่น ไม่สามารถแยกคำที่มาจากภาษาต่างประเทศได้และไม่สามารถแยกชื่อที่มีนามสกุลต่อท้าย เช่น

นายสุชาติ หวังดีต่อนางสุจิตตรา จะตัดได้ นายสุชาติ/ /หวัง/ดี/ต่อ/นางสุ/จิต/ตรา



สำนักหอสมุด

๓ - เม.ย. ๒๕๕๘

ว. ๖๖๔๑๒๑๖

๑ ๙๕๘  
๑๖๔  
๙๙๗๗๓  
๒๕๕๖



## บทที่ 4

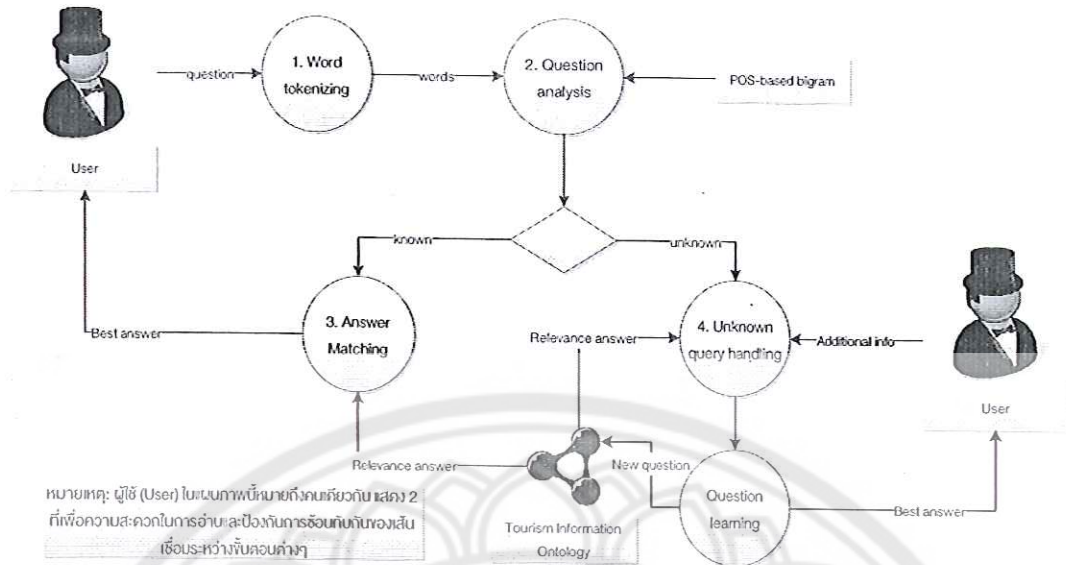
### ระเบียบวิธีวิจัย

ระบบสอบถามข้อมูลการท่องเที่ยวอัจฉริยะ (Automatic Tourism Question-Answer System-ATQ) คือระบบที่ให้ข้อมูลแก่นักท่องเที่ยวโดยอัตโนมัติโดยที่ผู้ใช้สามารถสอบถามข้อมูลการท่องเที่ยวกับระบบได้เหมือนกับคุยกับพนักงานคนหนึ่งที่สามารถให้ข้อมูลการท่องเที่ยวแก่ผู้ใช้ได้ โดยต้นแบบระบบที่จะพัฒนาขึ้นในการศึกษาวิจัยนี้จะจำกัดข้อมูล (Domain) อยู่ที่ข้อมูลการท่องเที่ยวในเขตพื้นที่จังหวัดพิษณุโลก หัวใจสำคัญของระบบนี้คือการสร้างระบบที่สามารถติดต่อสื่อสารกับผู้ใช้ได้อย่างมีประสิทธิภาพ เปรียบเสมือนเป็นตัวแทน (Agent) ของพนักงานในบริษัทท่องเที่ยวที่คอยให้ข้อมูลกับนักท่องเที่ยวได้ตลอดเวลาโดยจุดประสงค์หลักของโครงการวิจัยนี้อีกอย่างหนึ่งคือพยายามพัฒนาเทคนิคที่ทำให้ระบบคอมพิวเตอร์เข้าใจภาษาไทยและนำไปประยุกต์ใช้กับระบบสอบถามข้อมูลการท่องเที่ยวโดยกรอบแนวคิด (Framework) ของการศึกษาวิจัยนี้แบ่งออกเป็น 2 ส่วนหลักๆ โดยที่แต่ละส่วนจะให้ความสำคัญกับงานที่ต่างกัน ทั้งนี้เพื่อให้ระบบคอมพิวเตอร์สามารถเข้าใจสิ่งที่ผู้ใช้หรือนักท่องเที่ยวต้องการสื่อสารกับระบบ โดยที่ส่วนที่ 1 จะเน้นไปที่การการวิเคราะห์คำถาม (Question analysis) เพื่อนำไปสู่การตีความประโยค (Interpretation/Understanding) เพื่อให้เกิดความใจความหมายของคำถามต่างๆที่นักท่องเที่ยวสื่อสารกับระบบโดยการพิมพ์ผ่านหน้าจอ และส่วนที่ 2 คือการจัดการกับคำถามที่ระบบไม่รู้จัก เนื่องจากนักท่องเที่ยวแต่ละคนอาจจะเขียนคำถามที่ใช้คำพูดไม่เหมือนกัน ดังนั้นจึงเป็นไปได้ยากที่ระบบจะเข้าใจคำถามทุกคำถามของนักท่องเที่ยว ดังนั้นส่วนการจัดการคำถามที่รู้จักนี้ ระบบจะพยายามที่จะเรียนรู้คำถามใหม่ๆ ที่เข้ามาในระบบและเพิ่มเข้าไปในองค์ความรู้ของระบบโดยอัตโนมัติ ข้อมูลโดยสรุปของกรอบแนวคิดของระบบ ATQ (รูปที่ 7) สามารถอธิบายได้ดังนี้

#### 4.1. ระบบตัดคำ (Word tokenizing)

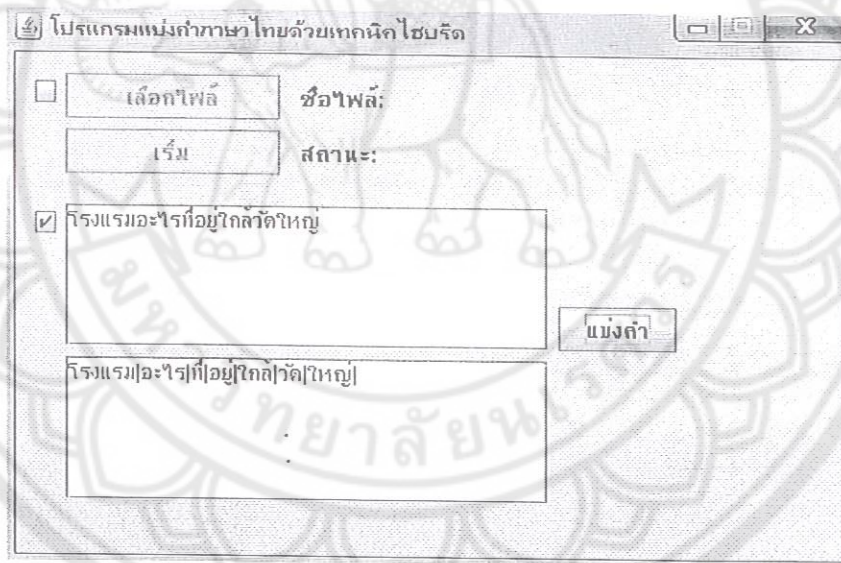
ขั้นตอนนี้เป็นกระบวนการนำประโยคคำถามที่ได้จากผู้ใช้มาแบ่งออกเป็นหน่วยของคำย่อยๆ เพื่อนำไปวิเคราะห์หาความหมายของคำถามต่อไป เนื่องจากเทคนิคการตัดคำไม่ใช่จุดประสงค์หลักของงานวิจัยนี้ ดังนั้นนักวิจัยจึงใช้ซอฟต์แวร์ตัดคำที่ถูกพัฒนาไว้เรียบร้อยแล้วที่ชื่อว่า “โปรแกรมตัดคำด้วยเทคนิคไฮบริด” (วรศักดิ์ และคณะ, 2552) ข้อดีของโปรแกรมนี้อีกพัฒนาโดยภาษาจาวาที่สามารถนำมาเชื่อมต่อกับระบบใหม่ที่นำเสนอได้โดยง่ายและสามารถตรวจสอบคำย่อจากข้อมูล เช่น จ. พ.ศ. ตรวจสอบชื่อเฉพาะในฐานข้อมูล และตรวจสอบคำที่คิดว่าน่าจะเป็นคำเฉพาะจากฐานข้อมูลคำขึ้นต้น





รูปที่ 7 ส่วนประกอบของระบบ ATQ

ตัวอย่างประโยคคำถามเช่น โรงแรมอะไรที่อยู่ใกล้วัดใหญ่ ผลที่ได้จากการตัดคำจากโปรแกรมนี้คือ โรงแรม/อะไร/ที่/อยู่/ใกล้/วัด/ใหญ่ แสดงรูปที่ 8



รูปที่ 8 ตัวอย่างการตัดคำโดยโปรแกรมแบ่งคำภาษาไทยด้วยเทคนิคไฮบริด

#### 4.2. การวิเคราะห์คำถาม (Question analysis)

เมื่อได้รับหน่วยคำจากกระบวนการตัดคำในขั้นตอนที่ 1 แล้ว ขั้นตอนถัดไปคือนำคำต่างๆ มาวิเคราะห์เพื่อแปลความหมายของคำถาม ในขั้นตอนนี้ประกอบด้วย 2 ขั้นตอนย่อยๆ คือ

#### 4.2.1. การวิเคราะห์โครงสร้างของประโยค (Question's structure analysis)

กระบวนการนี้จะนำคำที่ได้มาหาประเภทของคำ เช่น คำนามคำเฉพาะและหาประเภทของคำถามโดยเทียบจากฐานความรู้และนำมาเปลี่ยนโครงสร้างของประโยคใหม่โดยใส่ข้อมูลเพิ่มเติมให้กับคำต่างๆ เพื่อให้ง่ายต่อการวิเคราะห์ความหมายมากขึ้นซึ่งในขั้นแรกจะต้องหาหน้าที่ของคำก่อน โดยผลการหาหน้าที่ของคำจะแสดงในรูปแบบ <หน้าที่ของคำ, pos> โดยที่ pos (position) หมายถึงลำดับของคำในประโยคเพื่อประโยชน์ในการเข้าใจคำถามที่ถูกต้องมากยิ่งขึ้น ส่วน “หน้าที่ของคำ” จะประกอบด้วยคำประเภทต่างๆ ดังต่อไปนี้

1. <antecedent, pos:>หมายถึงคำที่บอกถึงคำนำหน้าชื่อเฉพาะของคำ เช่น วัดน้ำตกโรงแรมมหาวิทยาลัยร้านอาหารร้านอุทยานอุทยานแห่งชาติ เป็นต้น
2. <noun, pos:>หมายถึงชื่อที่บ่งบอกถึงสถานที่นั้นๆ เช่น วัดใหญ่น้ำตกแก่งของพลาโลมา เป็นต้น
3. <preposition, pos: >หมายถึง คำบุพบท ที่ปรากฏในประโยคคำถาม เช่น อยู่ใกล้ระหว่างใกล้ติดกับ เป็นต้น
4. <adjective, pos: >หมายถึง คำคุณศัพท์ ที่ปรากฏในประโยคคำถาม เช่น สวย ใหญ่ เล็ก เป็นต้น
5. <qtype, pos: >หมายถึงคำที่ใช้ตอบปัญหา เช่น อะไรที่ไหนอย่างไรตรงไหน เป็นต้น โดยที่หากคำเหล่านี้ปรากฏใกล้คำไหนจะถือว่าเป็นส่วนคำถามหลักของประโยค
6. <verb, pos>หมายถึง คำกริยา ที่ปรากฏในประโยคคำถาม เช่น มี ไหว้ ไป
7. <stop word, pos>หมายถึง คำที่ไม่มีความสำคัญในการนำมาใช้ประโยชน์เพื่อการวิเคราะห์คำถาม ซึ่งผู้วิจัยจะไม่นำมาใช้ในกระบวนการ เช่น จ๊ะหรือที่ เป็นต้น

ตัวอย่างเช่น จากขั้นตอนการตัดคำในขั้นตอนที่ 1 จะได้ “โรงแรม/อะไร/ที่/อยู่/ใกล้/วัด/ใหญ่” เมื่อผ่านขั้นตอนการวิเคราะห์โครงสร้างของประโยค ผลลัพธ์ที่ได้แสดงดังต่อไปนี้

โครงสร้างประโยค: “<antecedent, pos:โรงแรม, 1><qtype, pos: อะไร, 2><stop word, pos: ที่, 3><stopword, pos: อยู่, 4><preposition, pos: ใกล้, 5><antecedent, pos:วัด, 6><adjective, pos:ใหญ่, 7>”

โครงสร้างของประโยคที่ได้นี้จะไปประมวลผลต่อเพื่อลดความซ้ำซ้อนของคำ รวมคำที่ทำหน้าที่เหมือนกัน และตัดคำที่ไม่สำคัญทิ้ง ซึ่งจะเรียกขั้นตอนดังกล่าวว่า “การทำความสะอาดประโยค”



#### 4.2.2. การทำความสะอาดคำในประโยค (Sentence cleansing)

ขั้นตอนนี้จะช่วยลดคำที่ไม่จำเป็นในประโยคเพื่อการประมวลผลที่ถูกต้องและรวดเร็วยิ่งขึ้น ดังนั้นจะมีการรวมคำและตัดคำที่ไม่สำคัญทิ้งไป ในการกระทำดังกล่าวจะใช้หลักการของ Part-of-Speech (POS) bigram (Rajagopal et al. 2013) โดยมีหลักการดังต่อไปนี้

- 1) Antecedent+Noun: จะนำคำมารวมกัน แสดงว่าเป็นชื่อเฉพาะ
- 2) Antecedent+Adjective: จะนำคำมารวมกัน แสดงว่าเป็นชื่อเฉพาะ
- 3) Noun+Adjective: ระบบจะนำ 2 คำมารวมกัน
- 4) Noun+Noun: ทั้ง 2 คำจะถูกรวมกันเป็นคำเดียว ระบบจะคิดว่าทั้ง 2 คำถูกใช้ร่วมกันเพื่อแทนชื่อสถานที่หรือบางสิ่ง
- 5) Stopword: ระบบจะตัดคำที่ไม่สำคัญทิ้ง
- 6) สำหรับกรณีอื่นๆ ที่อยู่นอกเหนือจาก 5 ข้อนี้อันจะยังคงเดิม จะไม่ถูกตัดทิ้งและจะไม่ถูกนำมา รวมกัน เช่น คำกริยา(Verb)ต่างๆจะไม่ถูกนำไปรวมกับคำอื่นๆ

จากหลักการข้างต้นจะทำให้ได้ผลรับจากตัวอย่างในหัวข้อ 3.2.1 ดังต่อไปนี้

“<antecedent, pos: โรงแรม, 1><qtype, pos:อะไร, 2><preposition, pos: ใกล้, 5><antecedent, pos:วัดใหญ่, 6>”

ต่อจากนั้นระบบจะนำองค์ประกอบของคำถามดังกล่าวไปค้นหาข้อมูลในออนโทโลยีต่อไปโดยใช้ SPARQL และนำผลลัพธ์มาจัดลำดับความคล้ายคลึง (Similarity) โดยการคำนวณแบบโคไซน์ (Cosine)

#### 4.3. การค้นหาคำตอบและการเรียงลำดับ (Answer matching)

เมื่อได้คำสำคัญที่เหลือทั้งหมดจากขั้นตอนก่อนหน้า ระบบจะนำคำสำคัญเหล่านี้เพื่อไปค้นหาข้อมูลในออนโทโลยีซึ่งเก็บข้อมูลด้านการท่องเที่ยวเอาไว้ โครงสร้างของออนโทโลยีจะช่วยวิเคราะห์คำถามจากคำสำคัญที่เหลือและหาข้อมูลที่สอดคล้องกับคำถามในเชิงของความหมายหรือคอนเซปต์ (Concept matching) ซึ่งจะให้ได้คำตอบที่เกี่ยวข้องกับคำถาม มากกว่าวิธีการเปรียบเทียบตัวอักษร (String matching) จากตัวอย่างในขั้นตอนที่แล้ว ระบบทราบว่าคำถามจัดอยู่ในประเภท “อะไร” และคำถามที่มีลักษณะเป็นความสัมพันธ์กันระหว่างข้อมูล 2 ข้อมูลหรือเรียกว่า “Tuple” ซึ่งโครงสร้างของ Tuple จะประกอบด้วยประธาน-<ความสัมพันธ์>-กรรม ดังนั้นในตัวอย่างนี้จะได้ Tuple คือ “โรงแรม-<ใกล้>-วัดใหญ่” จากคำถามและความสัมพันธ์ของคำนามทั้ง 2 คำ ระบบจะทำการค้นหาคำตอบที่เกี่ยวข้องโดยใช้โครงสร้างของออนโทโล-



จะช่วยในการสกัดคำตอบที่เกี่ยวข้องออกมาทั้งหมดและแสดงให้กับผู้ใช้ทราบต่อไป “โรงแรม” เป็นคำสำคัญที่บอกให้ระบบทราบว่าผู้ใช้ต้องการทราบข้อมูลเกี่ยวกับโรงแรมที่อยู่ใกล้วัดใหญ่ ไม่ใช่ร้านอาหารหรือห้างสรรพสินค้าและความสัมพันธ์ “ใกล้” นี้จะต้องถูกระบุไว้ในอย่างชัดเจนในออนโทโลยี ระบบจะไม่ได้พิจารณาจากระยะทางระหว่างวัดใหญ่กับโรงแรม

#### 4.4. การจัดการคำถามที่ไม่รู้จัก (Unknown question handling)

ในบางคำถามเมื่อผ่านกระบวนการทำความเข้าใจแล้ว จะเหลือเพียงคำสำคัญไม่กี่คำ ซึ่งมีผลทำให้ไม่สามารถประกอบคำเป็น Tuple เพื่อไปเปรียบเทียบกับโครงสร้างของออนโทโลยีได้ ตัวอย่างเช่น “มีห้างสรรพสินค้าตรงไหนบ้าง” เมื่อผ่านกระบวนการตัดคำ วิเคราะห์โครงสร้างประโยคและการทำความเข้าใจแล้วจะได้

<verb, pose: มี, 1><antecedent, pos: ห้างสรรพสินค้า, 2> <qtype, pos: ตรงไหน, 3>

จากตัวอย่างดังกล่าวประเภทของคำถามคือหาตำแหน่งของสถานที่ (ตรงไหน) และคำสำคัญที่เหลือไม่สามารถประกอบเป็น Tuple ได้อย่างสมบูรณ์ อย่างไรก็ตามในกรณีเช่นนี้ระบบจะพยายามคาดเดาความหมายของคำถาม ระบบจะพยายามแปลงคำสำคัญเหล่านี้ให้อยู่ในรูปแบบ Tuple และเติมสิ่งที่หายไปด้วยตัวแปร (X) ดังนั้นเราจะได้ “X-<มี>-ห้างสรรพสินค้า” ดังนั้นระบบสามารถแปลความหมายของคำถามได้ว่า “ตำแหน่งใดๆ (X) ที่มีห้างสรรพสินค้า” และระบบจะค้นหาข้อมูลจาก Tuple ดังกล่าวโดยแทนที่ X ด้วยพิกัดบนแผนที่ของห้างสรรพสินค้าทุกแห่งในจังหวัดพิษณุโลก

#### 4.5. เครื่องมือที่ใช้ในการพัฒนา

เครื่องมือที่ใช้ในการพัฒนาระบบถามตอบแหล่งท่องเที่ยวจังหวัดพิษณุโลกแบบอัตโนมัติประกอบไปด้วยส่วนที่เป็น Software ดังนี้

##### 4.5.1. เครื่องมือและการพัฒนาเครื่องมือ

ซอฟต์แวร์

-ระบบปฏิบัติการ Windows 7 Ultimate

เครื่องมือที่ใช้พัฒนา

- ภาษา PHP

- โปรแกรม Adobe Dreamweaver CS5 เป็นโปรแกรมในการพัฒนาระบบ

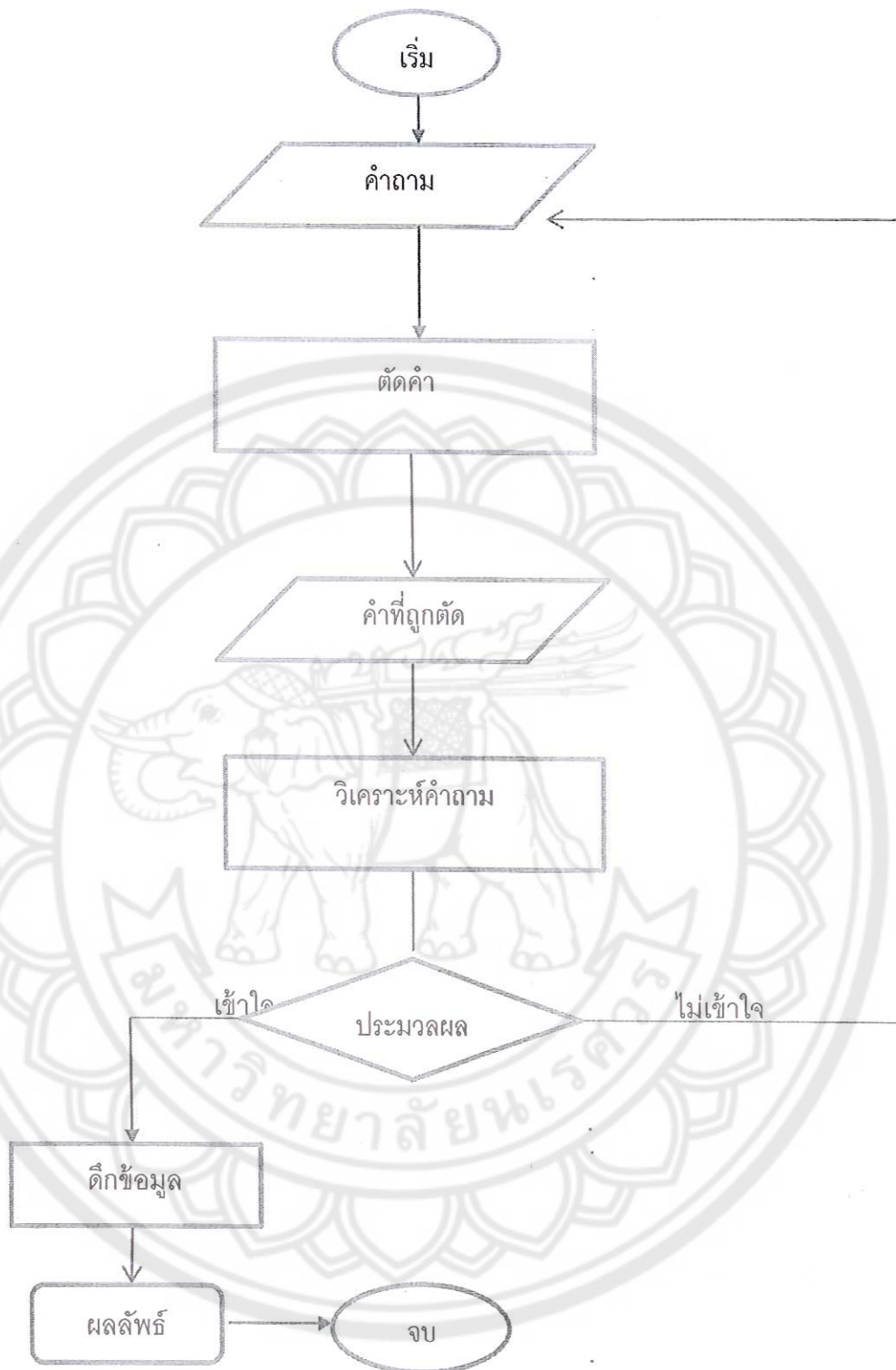
#### 4.5.2. วิธีการดำเนินการศึกษา

- 1) ศึกษาข้อมูลในการทำเว็บไซต์
- 2) ศึกษาหาข้อมูลเกี่ยวกับสถานที่ท่องเที่ยวในจังหวัดพิษณุโลก
- 3) วางแผนออกแบบระบบให้สอดคล้องกับผู้ใช้
- 4) ออกแบบหน้าเว็บไซต์
- 5) วิเคราะห์ความต้องการของระบบ
- 6) ดำเนินการเขียน พัฒนาเว็บไซต์
- 7) สร้างและพัฒนาโปรแกรม
- 8) ทดลอง ปรับปรุงและแก้ไขข้อผิดพลาดของโปรแกรม
- 9) จัดทำรูปเล่มและรายงาน

#### 4.5.3. ผังการทำงานของระบบ

ผู้วิจัยได้ออกแบบระบบในขั้นเริ่มต้นและปรับปรุงอย่างต่อเนื่อง จากการเก็บรวบรวมข้อมูล ความต้องการของผู้ใช้ จนได้การออกแบบในส่วนต่างๆ จากรูปที่ 9 สามารถอธิบายการทำงานของระบบได้ดังต่อไปนี้

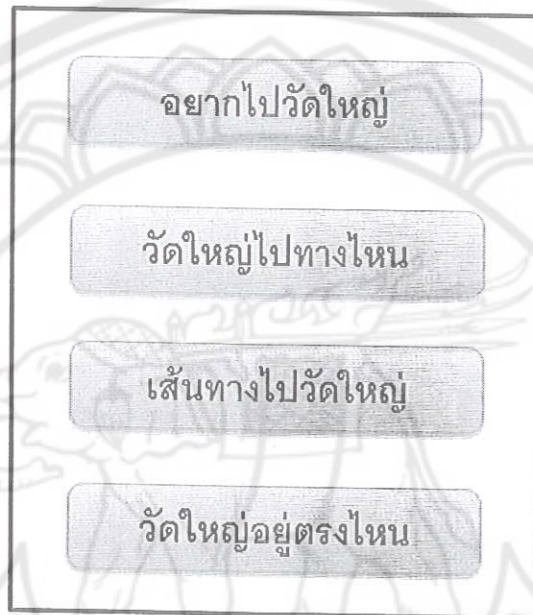
1. เมื่อระบบได้รับคำถามหรือประโยคมาจากผู้ใช้ระบบก็จะนำคำถามไปเข้าสู่กระบวนการตัดคำ โดยกระบวนการตัดคำของระบบนั้นจะใช้โปรแกรมตัดคำด้วยเทคนิคไฮบริดซึ่งจะได้มาจากคำถามหรือประโยคที่ผู้ใช้ใส่เข้ามาในระบบแล้วระบบจะนำคำถามหรือประโยคนั้นไปตรวจสอบในพจนานุกรมว่าคำถามหรือประโยคนั้นมีคำที่สามารถจะตัดออกมาได้ในลักษณะใดตามหลักของพจนานุกรม หลังจากการตรวจสอบแล้วโปรแกรมก็จะมีการตัดคำออกมา เช่น “วัดใหญ่ไปทางไหน” โปรแกรมจะตัดออกมาได้คำว่า “วัด | ใหญ่ | ไป | ทาง | ไหน” เมื่อระบบได้ตัดคำออกมาแล้วนั้นจะนำคำถามที่ตัดได้มาเก็บไว้



รูปที่ 9 ผังงานของระบบ



2. นำคำถามหรือประโยคที่ถูกเก็บไว้แต่ละคำไปค้นหาในฐานข้อมูลและการค้นหาในฐานข้อมูลนั้น จะใช้คำถามหรือประโยคที่ได้ผ่านกระบวนการตัดคำมาแล้วในการค้นหา โดยที่ประโยคหรือคำถามที่ถูกตัดออกมา จะมีความหมายเดียวกันแต่คำถามหรือประโยคนั้นอาจมีหลากหลายรูปแบบ เช่น อยากไปวัดใหญ่ วัดใหญ่ไปทางไหน เส้นทางไปวัดใหญ่ วัดใหญ่อยู่ตรงไหน คำถามหรือประโยคที่กล่าวมานี้มีความหมายเดียวกันไม่ว่า ตำแหน่งของคำถามหรือประโยคที่ถูกตัดมาถึงจะไม่ได้อยู่ในตำแหน่งเดียวกันก็ตามระบบก็สามารถเข้าใจใน ความหมายของคำถามหรือประโยคนั้นๆ ได้ว่าคำถามหรือประโยคนั้นหมายถึงอะไร ตัวอย่างดังรูปที่ 10



รูปที่ 10 ตำแหน่งของคำถาม

3. คำถามหรือประโยคที่ถูกตัดออกมา เช่น วัดใหญ่ น้ำตก ห้างสรรพสินค้า ระบบจะนำคำถามหรือ ประโยคที่เป็นคำหลักๆ ในการค้นหาข้อมูล เช่น คำว่า วัด | ใหญ่ | น้ำตก | ห้าง คำถามหรือประโยคเหล่านี้ ระบบก็จะตอบเป็นข้อมูลของสถานที่นั้นๆ แต่ถ้ามีการขยายความของคำถามหรือประโยค เช่น อยากไปวัด ใหญ่ วัดใหญ่ไปทางไหน เส้นทางไปวัดใหญ่ วัดใหญ่อยู่ตรงไหน ระบบสามารถเข้าใจความหมายของคำถามนี้ ว่าจะตอบคำถามเป็นแผนที่และแสดงตำแหน่งของสถานที่นั้นๆ

4. ถ้าคำถามหรือประโยคที่ไม่สามารถเข้าใจคำถามก็จะมี การส่งคำถามกับไปยังผู้ใช้เพื่อต้องการขอ ข้อมูลเพิ่มเติม

#### 4.5.4. กระบวนการแนะนำคำถาม

กระบวนการนี้จะรับเซตคำถามที่ไม่สามารถจับคู่กับรูปแบบในฐานความรู้รูปแบบได้เนื่องจาก

1. ผู้ใช้สะกดคำผิด
2. กระบวนการวิเคราะห์คำถามไม่สามารถหาชนิดคำถามในฐานความรู้
3. ชื่อคำเฉพาะไม่มีในฐานข้อมูลและนำมาเพื่อประมวลผลใหม่เพื่อด้วยกระบวนการต่อไปนี้

##### 1) สกัดคำที่ไม่รู้จัก

เมื่อได้รับชุดคำถามจากกระบวนการวิเคราะห์คำถามแล้ว ขั้นตอนนี้จะเลือกคำที่ไม่รู้จักออกมา แล้วนำไปค้นหาในฐานข้อมูลความรู้ เช่น ถามว่า “วัดพระแก้วอยู่ตรงไหน”

“[คำนำหน้าคำเฉพาะ=วัด] [ชื่อคำเฉพาะ=พระแก้ว] <คำชี้แนะ: อยู่> <ชนิดคำถาม: ตรงไหน>

จะเห็นว่าคำที่ฐานข้อมูลไม่รู้จักคือ พระแก้ว เพราะพระแก้วไม่ได้อยู่ในข้อมูลจังหวัดพิษณุโลกดังนั้นระบบจะทำการประมวลผลต่อในขั้นตอนถัดไป

##### 2) การแบ่งหมวดหมู่

กระบวนการนี้จะเลือกคำที่ผิดพลาดแล้วนำมาหาว่าคำนี้อยู่ในหมวดหมู่ใด ถ้าไม่เจอระบบจะเลือกโดยใช้ความสัมพันธ์ของออนโทโลยี จากกระบวนการข้างต้นระบบสามารถทราบอยู่แล้วว่าคำว่า “พระแก้ว” อยู่ในหมวดของวัดแต่ในฐานความรู้ของจังหวัดพิษณุโลกไม่มีคำว่า “พระแก้ว” ดังนั้นระบบจะตอบกลับผู้ใช้งานว่า “ไม่พบข้อมูล” เป็นต้น อย่างไรก็ตามระบบจะพยายามข้อมูลอื่นๆ ที่ใกล้เคียงโดยมีกระบวนการดังขั้นตอนถัดไป

##### 3) การสร้างประโยค

ขั้นตอนนี้จะนำ เซตคำถามเดิมมาสร้างคำถามใหม่ เช่น คุณหมายถึง

“วัดใหญ่อยู่ตรงไหน”

“วัดคู้วารีย์อยู่ตรงไหน”

“วัดจุนางอยู่ตรงไหน”

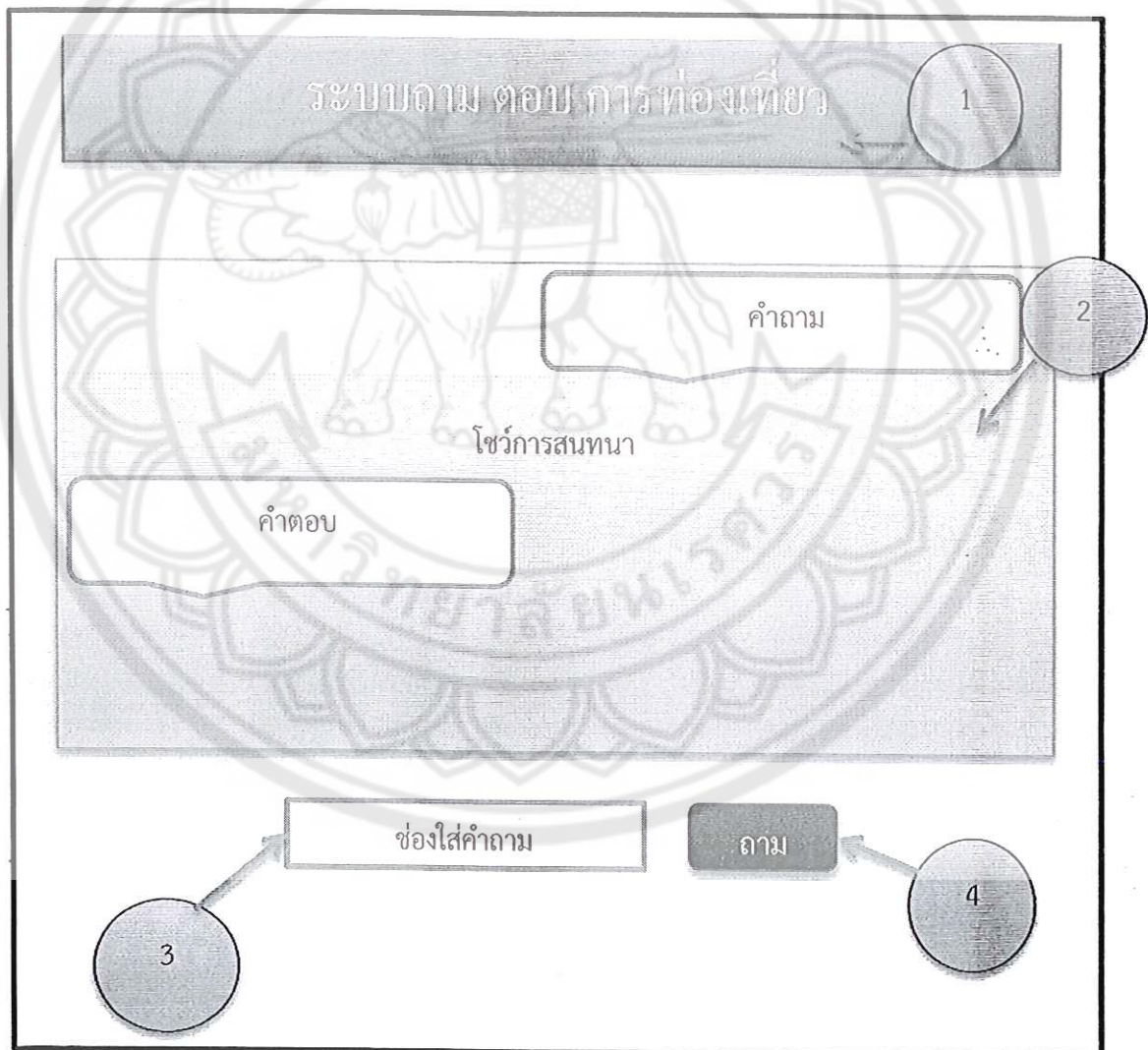
“วัดวังหินอยู่ตรงไหน”

“วัดวังทองอยู่ตรงไหน” เป็นต้น

เมื่อจบขั้นตอนนี้ ระบบจะส่งคำถามที่สร้างขึ้นให้กับผู้ใช้ เพื่อให้ผู้ใช้เลือกคำถามที่ต้องการ เมื่อผู้ใช้เลือกคำถามแล้วคำถามจะถูกส่งไปที่กระบวนการค้นหาคำตอบต่อไป

#### 4.6. การออกแบบส่วนแสดงผล (Output Design)

ในการออกแบบหน้าจอสำหรับระบบ ATQ นั้นจะเน้นการใช้งานที่ง่ายโดยจะแบ่งออกเป็น 4 ส่วนหลักดังรูปที่ 11



รูปที่ 11 หน้าจอหลัก



(1) ส่วนหัวของเว็บไซต์

(2) ส่วนของการแสดงข้อมูลในการสนทนาระหว่างผู้ใช้กับระบบ โดยคำถามจะแสดงบริเวณหน้าจอ ด้านขวามือ และคำตอบแสดงบริเวณหน้าจอด้านซ้ายมือ

(3) ส่วนของการรับคำถามจากผู้ใช้ที่ต้องการหาข้อมูลเกี่ยวกับสถานที่ท่องเที่ยวหรืออื่นๆ

(4) กดปุ่มเพื่อส่งคำถามเข้าไปประมวลผล

ตัวอย่างการทำงานของระบบสามารถอ่านรายละเอียดได้ในบทถัดไป



## บทที่ 5

### ผลการวิจัย

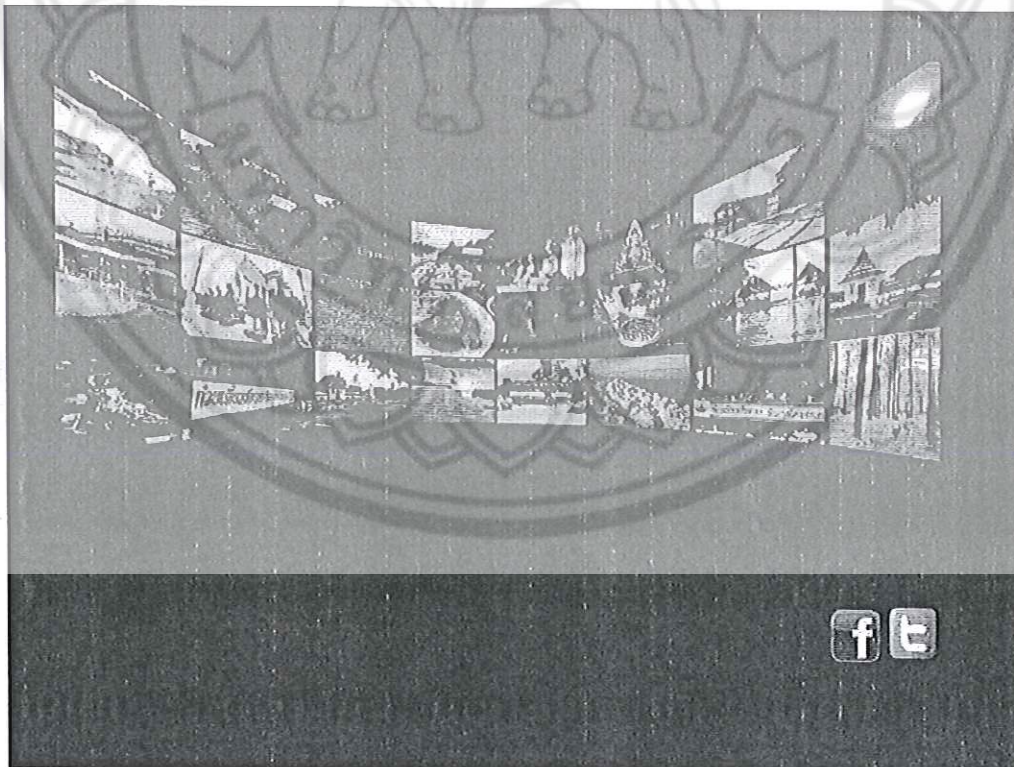
จากแนวคิดที่นำเสนอในหัวข้อที่ผ่าน ผู้วิจัยได้ทำการทดลองเพื่อทดสอบประสิทธิภาพของระบบ ATQ ดังต่อไปนี้

- 1) ความถูกต้องของการตอบคำถาม เช่น อะไร ที่ไหน
- 2) คำถามที่กำกวม

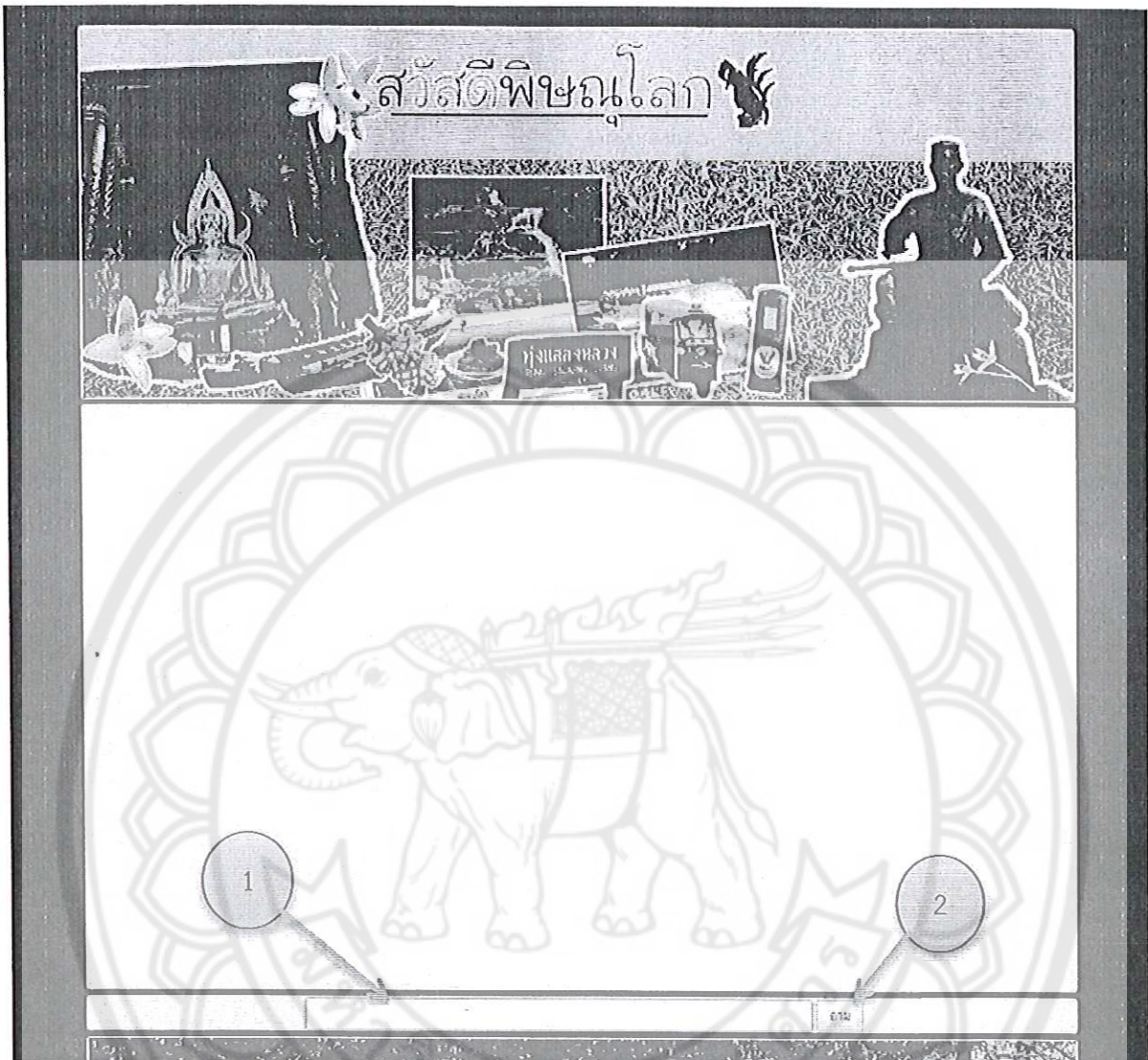
#### 5.1. ความถูกต้องของการตอบคำถาม เช่น อะไร ที่ไหน

##### 5.1.1. วิธีการทดลองคำถามประเภทข้อมูลทั่วไป

การทดลองนี้เป็นการทดลองใช้งานระบบตามที่ออกแบบไว้ เมื่อผู้ใช้เข้ามาใช้ระบบจะพบหน้าจอแรก แสดงดังรูปที่ 12



รูปที่ 12 หน้าจอแรกของระบบ

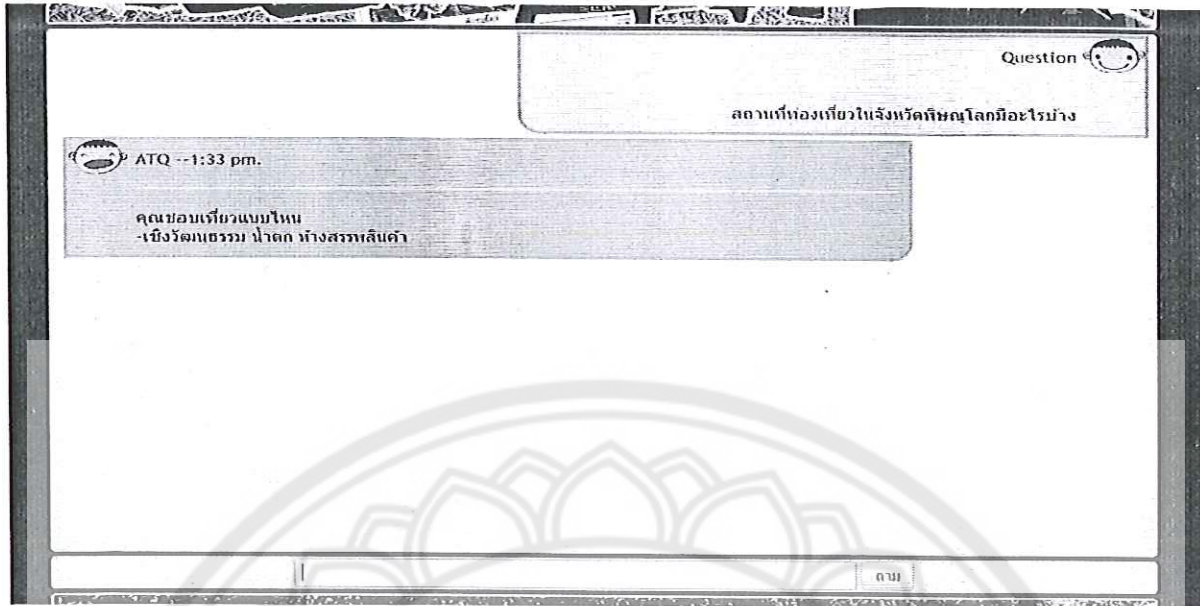


รูปที่ 13 หน้าจอระบบ ATQ

เมื่อกดหน้าแรกเข้ามาจะโชว์หน้าของ”ระบบถามตอบแหล่งท่องเที่ยวในจังหวัดพิษณุโลกแบบอัตโนมัติ” ดังรูปที่ 13 หน้าจอหลัก

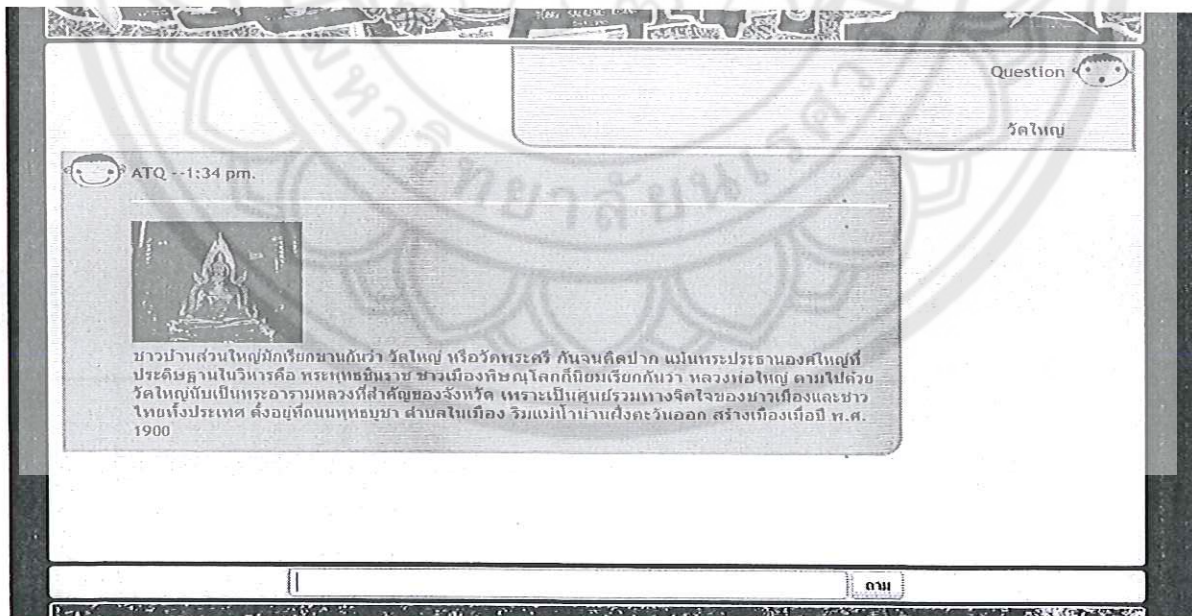
1. ให้ผู้ใช้ระบบใส่คำถามลงไปช่อง
2. เมื่อผู้ใช้ใส่คำถามเสร็จให้ผู้กดปุ่ม”ถาม”





รูปที่ 14 ข้อมูลการสนทนาเป็นข้อมูล

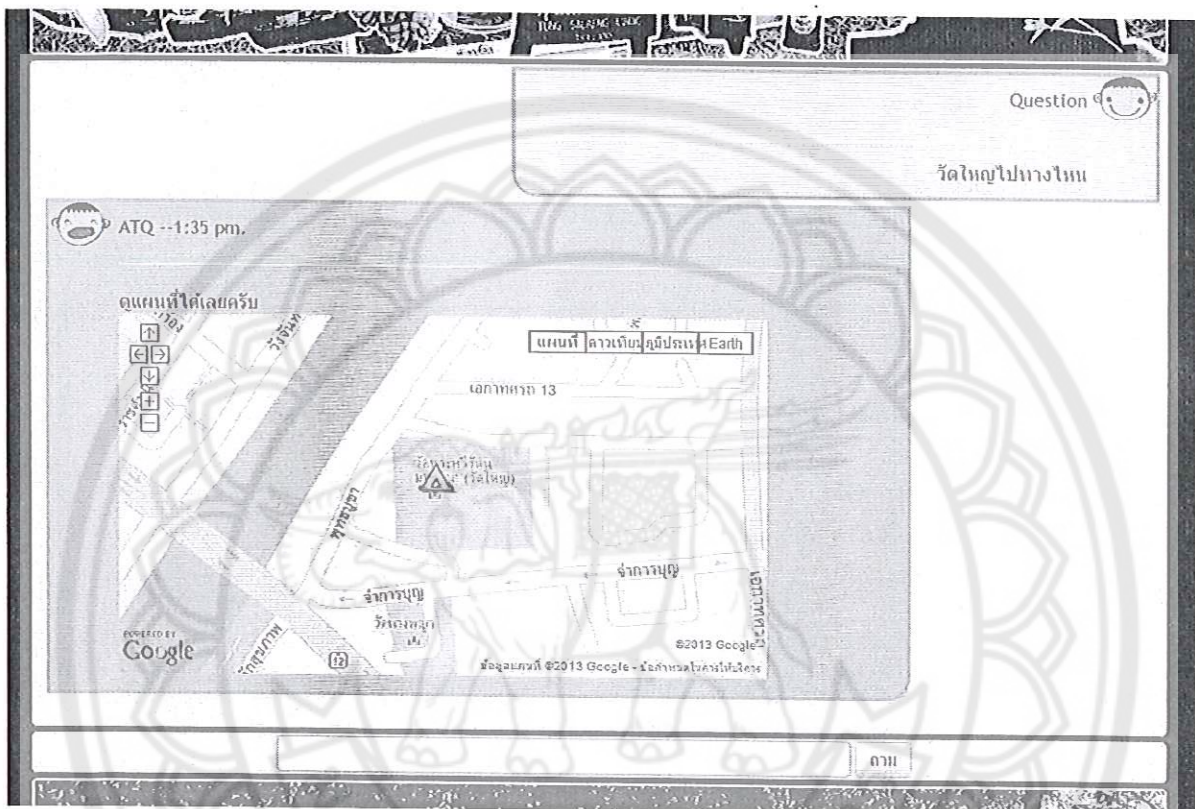
หลังจากที่ผู้ใช้ใส่คำถามไปในระบบตาม ดังรูปที่ 14 การที่ผู้ใช้ใส่คำถามเข้าไป เช่น “สถานที่ท่องเที่ยวในจังหวัดพิษณุโลกมีอะไรบ้าง” การให้คำตอบของระบบจะให้มาเป็นข้อมูลที่ตามเพิ่มเติม คุณชอบเที่ยวแบบไหน เช่น เชิงวัฒนธรรม น้ำตก หรือห้างสรรพสินค้า ถ้าผู้ใช้มีการถามเฉพาะสถานที่นั้น เช่น วัดใหญ่ จะแสดงข้อมูล ดังรูปที่ 15 แสดงข้อมูลเกี่ยวกับสถานที่ท่องเที่ยววันั้นๆ ที่ผู้ใช้ใส่คำถามเข้ามา เช่น วัดใหญ่ ระบบจะส่งคำตอบเป็นข้อมูลประวัติของวัดใหญ่มารับกับผู้ใช้



รูปที่ 15 ข้อมูลแหล่งที่ท่องเที่ยว

### 5.1.2. วิธีการทดลองคำถามประเภทตำแหน่งและที่ตั้งของสถานที่ต่างๆ

ถ้ามีการใส่คำถาม เช่น วัดใหญ่ไปทางไหน ทางไปวัดใหญ่ วัดใหญ่อยู่ตรงไหน คำถามประเภทนี้ระบบจะตอบมาเป็นแผนที่แบบปักหมุด ดังรูปที่ 16

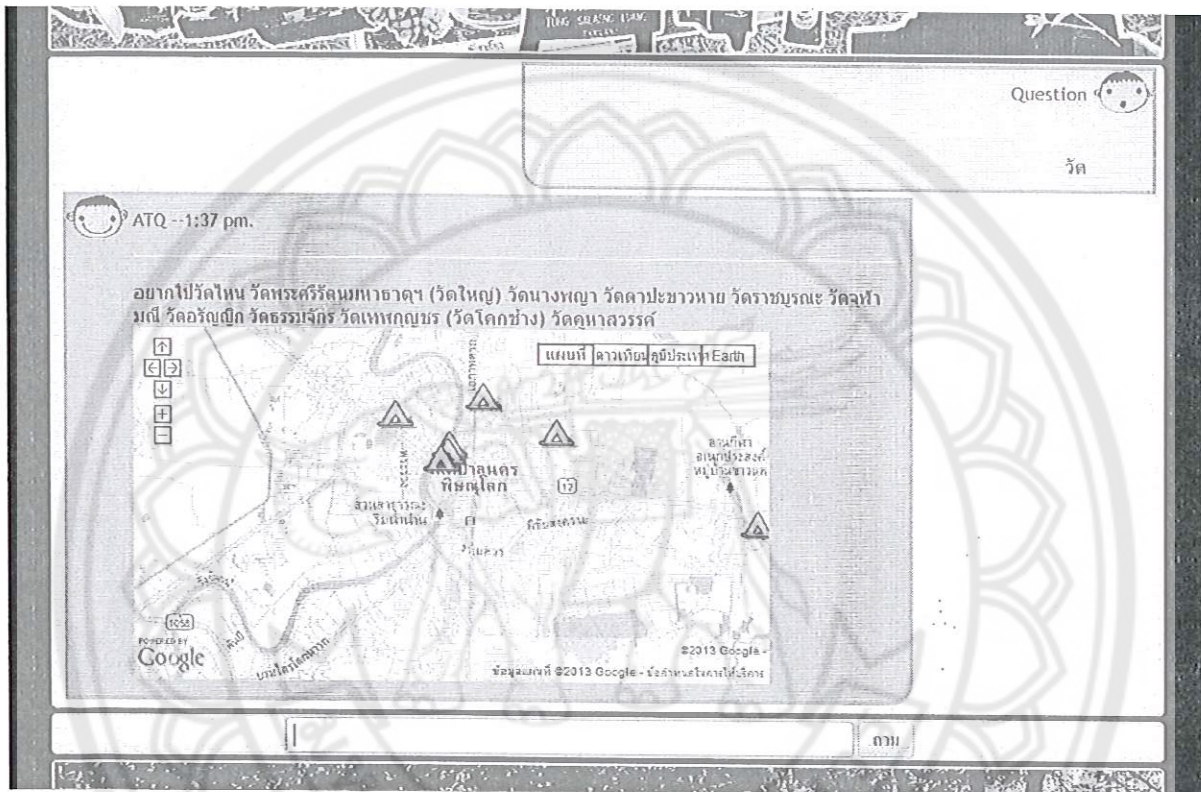


รูปที่ 16 ตำแหน่งบนแผนที่



การแสดงผลภาพแผนที่ที่จะแบ่งออกเป็น 2 แบบ

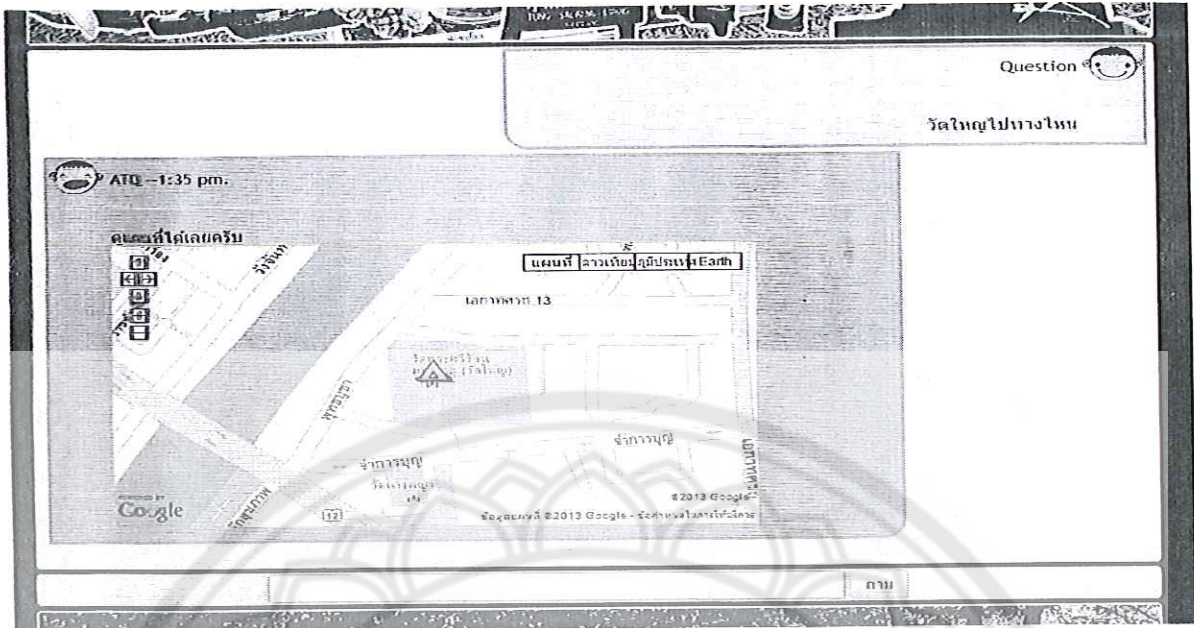
1. แสดงแผ่นแบบปักหมุดหลายๆที่ เช่น วัด จะแสดงดังรูปที่ 17



รูปที่ 17 แผ่นที่แบบปักหมุดหลายๆที่

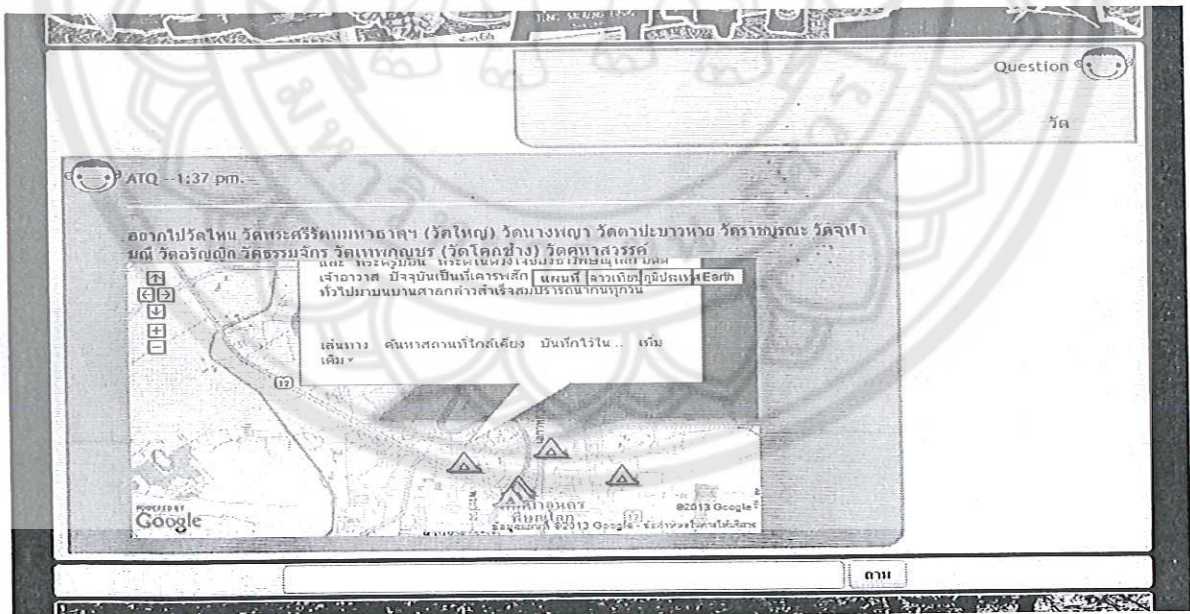
2. แสดงแผ่นที่แบบปักหมุดเฉพาะสถานที่ เช่น วัดใหญ่อยู่ตรงไหน ทางไปวัดใหญ่ ดังรูปที่ 18





รูปที่ 18 แผนที่เฉพาะสถานที่

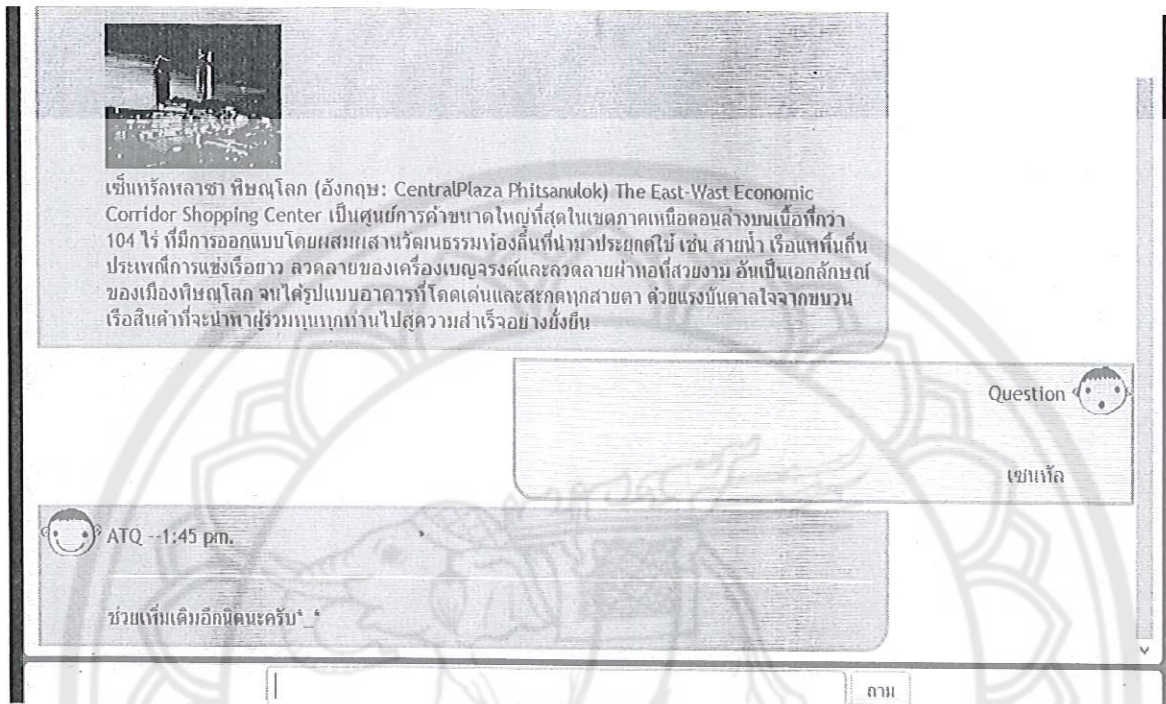
การตอบถามแบบแผนที่นั้นผู้ใช้สามารถกดไปที่มุมที่ปักอยู่ในแผนที่ที่จะแสดงข้อมูลเกี่ยวสถานที่นั้นๆ เช่น ประวัติของสถานที่ ดังรูปที่ 19



รูปที่ 19 ข้อมูลของสถานที่

## 5.2. คำถามที่กำกวม

คำถามที่กำกวมนั้นอาจจะมี ความหมายเดียวกันถ้าผู้ใช้ที่ใส่คำถามที่ผิดหลักของไวยากรณ์ เช่น เซนทรัล คำถามประเภทนี้ระบบไปจะสามารถเข้าใจได้ ดังรูปที่ 20



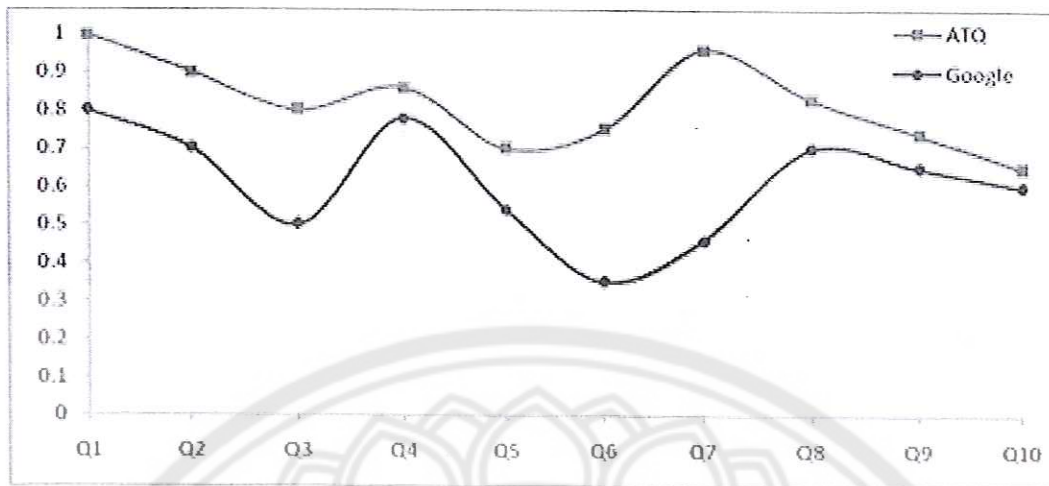
รูปที่ 20 คำกำกวม

ในขณะที่ระบบยังไม่สามารถแก้ไขคำผิดให้กับผู้ใช้ได้ ดังนั้นถ้าผู้ใช้จะต้องพิมพ์ชื่อของสถานที่ที่ผู้ใช้ต้องการได้อย่างถูกต้องซึ่งยังเป็นข้อจำกัดของระบบที่จะต้องปรับปรุงพัฒนาในเวอร์ชันต่อไป

## 5.3. เปรียบเทียบประสิทธิภาพกับ Google

เปรียบเทียบประสิทธิภาพการให้ข้อมูลกับ Google: ผู้วิจัยได้ทดลองทำการใส่คำถามจำนวน 10 คำถามเกี่ยวกับสถานที่ท่องเที่ยวในเขตจังหวัดพิษณุโลก จากชุดคำถามทั้งหมด 84 คำถาม ซึ่งแบ่งเป็นคำถามประเภทเส้นทาง ข้อมูลสถานที่ท่องเที่ยว โรงแรมและร้านอาหาร ซึ่งมีความซับซ้อนต่างกัน และใช้คำถามเดียวกันกับ Google และประเมินประสิทธิภาพของระบบทั้งสองจากค่าความแม่นยำ (Precision) ผลการทดลองแสดงดังรูปที่ 21 ซึ่งให้เห็นว่าระบบ ATQ สามารถปรับปรุงประสิทธิภาพการค้นหาข้อมูลและตอบคำถามโดยใช้ภาษาไทยได้สูงกว่า Google ที่มีความความแม่นยำเฉลี่ยอยู่ที่ 85% ในขณะที่ Google มีค่าความแม่นยำเฉลี่ยที่ 74%





รูปที่ 21 ค่าความแม่นยำในการค้นหาข้อมูลโดยตรงกับคำถามผู้ใช้





## บทที่ 6

### บทสรุปและข้อเสนอแนะ

ในบทนี้เป็นการสรุปการทำวิจัยในเรื่องการพัฒนาระบบสอบถามข้อมูลการท่องเที่ยวอัจฉริยะ ซึ่งในบทนี้จะเป็นการสรุปงานวิจัยทั้งหมดและสิ่งที่งานวิจัยนี้จะทำต่อไปในอนาคต เนื้อหาในบทนี้จะเป็นการสรุปงานวิจัย ซึ่งมีข้อหาดังต่อไปนี้

- 1) ความสำเร็จของงานวิจัยและแนวคิดใหม่
- 2) สรุปผลการทดลอง
- 3) ข้อจำกัดของงานวิจัย
- 4) สิ่งทีงานวิจัยนี้จะทำต่อไปในอนาคต

#### 6.1. ความสำเร็จของงานวิจัยและแนวคิดใหม่

##### 6.1.1. ความสำเร็จของงานวิจัย

งานวิจัยนี้ได้พัฒนาเทคนิคในการค้นหาข้อมูลที่เกี่ยวข้องกับข้อความในรูปแบบของระบบถาม-ตอบอัตโนมัติ (ATQ) ซึ่งสามารถที่จะเรียนรู้คำถามและจัดการกับคำถามที่ไม่รู้จักได้อย่างมีประสิทธิภาพ องค์ประกอบของระบบที่สำคัญมี 4 ส่วนได้แก่ 1) ส่วนการตัดคำ 2) ส่วนของการวิเคราะห์คำถาม 3) การค้นหาคำตอบที่เกี่ยวข้อง และ 4) การจัดการคำถามที่ไม่รู้จัก ซึ่งแนวคิดดังกล่าวได้ยอมรับให้นำเสนอในงานประชุมวิชาการ 1 งาน ดังแสดงได้ดังนี้

ไกรศักดิ์ เกษร. (2014). ระบบสอบถามข้อมูลอัจฉริยะสำหรับข้อมูลการท่องเที่ยวในประเทศไทย. นำเสนอในงานประชุมวิชาการระดับชาติวิทยาศาสตร์วิจัยครั้งที่ 6 (Science Research 6<sup>th</sup>), มหาวิทยาลัยบูรพา จังหวัดชลบุรี 21-23 มีนาคม 2557.

### 6.1.2. สรุปแนวคิดใหม่ในงานวิจัย

แนวคิดใหม่ในงานวิจัยนี้คือการพัฒนาหลักการในการวิเคราะห์ประโยค โดยการนำหลักการวิเคราะห์โครงสร้างประโยคมาวิเคราะห์ซึ่งจะใช้หน้าที่ของคำและตำแหน่งของคำมาวิเคราะห์และหลักการของ Part-of-Speech มาร่วมด้วยเพื่อให้การวิเคราะห์มีประสิทธิภาพมากขึ้น (อ่านรายละเอียดในหัวข้อที่ 4.2)

### 6.2. ข้อจำกัดของงานวิจัย

ในงานวิจัยนี้ได้พัฒนาระบบถามตอบอัตโนมัตินี้ยังมีข้อจำกัดบางประการในงานวิจัย ซึ่งสามารถทำการพัฒนาต่อยอดในอนาคตต่อไป

- 1) ระบบยังไม่สามารถตรวจสอบตำแหน่งปัจจุบันของผู้ใช้ได้ผ่านเครื่องคอมพิวเตอร์ ดังนั้นจึงทำให้มีข้อจำกัดในการกำหนดเส้นทางเพื่อไปยังสถานที่ท่องเที่ยวดังกล่าว
- 2) ข้อคำถามบางคำถามที่ยาวและซับซ้อน ระบบจะยังไม่สามารถตอบคำถามได้อย่างถูกต้อง หรือไม่เข้าใจคำถามนั่นเอง ดังนั้นอาจจะต้องพัฒนาวิธีการวิเคราะห์คำถามที่มีโครงสร้างซับซ้อนสูงขึ้น
- 3) ระบบยังไม่สามารถตอบคำถามที่เกี่ยวข้องกับเวลาได้ ดังนั้นในอนาคตผู้วิจัยจะทำการพัฒนาการทำงานในส่วนนี้ต่อไป

## บรรณานุกรม

- Baeza-Yates, R., Hurtado, C., & Mendoza, M. (2004). Query recommendation using query logs in search engines. *Proceedings of the 2004 international conference on Current Trends in Database Technology* (pp. 588–596). Berlin, Heidelberg: Springer-Verlag.
- Dupret, G., & Mendoza, M. (2006). Automatic Query Recommendation using Click-Through Data.
- google, google. (2006). Google spell check. <http://www.google.com/features.html#spell>.
- Guo, Q., & Zhang, M. (2008). Question Answering System Based on Ontology and Semantic Web. In G. Wang, T. Li, J. W. Grzymala-Busse, D. Miao, A. Skowron, & Y. Yao (Eds.), *Rough Sets and Knowledge Technology* (pp. 652–659).
- Hao, T., & Wenyin, L. (2008). Automatic Question Translation Based on Semantic Pattern. *Proceeding of the 4th International Conference on Semantics, Knowledge and Grid*.
- Pan, S., & Shaw, J. (2007). Natural Language Query Recommendation in Conversation Systems. *Proceedings of the 20th International Joint Conference on Artificial Intelligence*.
- Ravichandran, D., & Hovy, E. (2002). Learning surface text patterns for a Question Answering system. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics* (pp. 41–47). Stroudsburg, PA, USA: Association for Computational Linguistics.
- Rajagopal, D., Cambria, E., Olsher, D., & Kwok, K. (2013). A Graph-Based Approach to Commonsense Concept Extraction and Semantic Similarity Detection. In *Proceedings of the 22nd International Conference on World Wide Web Companion* (pp. 565–570).
- Rayman-Bacchus, & A Molina. (2001). Internet-based Tourism Services: Business Issue and Trend, 589–605.
- Shen, D. (2007). Using Semantic Role to Improve Question Answering. *Proceedings of EMNLP 2007*.



Wikipedia, W. (2013). Artificial Linguistic Internet Computer Entity. In *Wikipedia, the free encyclopedia*.

ฉัตรชัย สรณเจริญพงศ์. (2551). ระบบถามตอบ. มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

ัชชวาล สังคีตตระการ, ศราวุธ คงยัง, มารุต บุรณรัช, ชูชาติ หลุไชยะศักดิ์, & อลิสา คงทน. (2551). *อับดุล: ระบบบริการข้อมูลเชิงบูรณาการผ่านระบบสนทนาออนไลน์*.

ทศพล วงศ์ศรีไพศาล. (2552). การใช้กฎในระบบตอบคำถามภาษาไทย สำหรับทดสอบการอ่านเพื่อความเข้าใจ. มหาวิทยาลัยธรรมศาสตร์.

ธนากร สุวรรณวัฒน์. (2550). บริการช่วยเหลือเพื่อตอบปัญหาการติดตั้งและการใช้งานอุปกรณ์ระบบ Call Center โดยใช้อินเทอร์เน็ต. มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

นายณรงค์ โคตรสกิจ. (2552). การพัฒนาต้นแบบแชทบอทภาษาไทย กรณีศึกษา เรื่องอัญมณี. มหาวิทยาลัยเชียงใหม่.

ทศพล วงศ์ศรีไพศาล. (2552). การใช้กฎในระบบตอบคำถามภาษาไทย สำหรับทดสอบการอ่านเพื่อความเข้าใจ. มหาวิทยาลัยธรรมศาสตร์.

ธนากร สุวรรณวัฒน์. (2550). *บริการช่วยเหลือเพื่อตอบปัญหาการติดตั้งและการใช้งานอุปกรณ์ระบบ Call Center โดยใช้อินเทอร์เน็ต (วิทยานิพนธ์/Thesis)*. มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

วรศักดิ์ ตั้งกุลทวีทรัพย์, พีระศักดิ์ รัตนมณี, & ธนพล จินดาพิทักษ์. (2552). การแบ่งคำภาษาไทยด้วยเทคนิคไฮบริด (การแข่งขันสุดยอดซอฟต์แวร์แบ่งคำภาษาไทย). มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตหาดใหญ่.

ยุทธศักดิ์ วุฒิกกร. (2553). *การขยายฐานความรู้แบบอัตโนมัติสำหรับระบบถาม-ตอบ โดยใช้ความหมายของประโยค (วิทยานิพนธ์/Thesis)*. มหาวิทยาลัยธรรมศาสตร์.

วลัยลักษณ์ สุขสมบูรณ์. (2553). *ระบบถามตอบอัตโนมัติสำหรับ Helpdesk*. มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

อรุณี โอสานนท์. (2548). *ระบบตอบคำถามเป็นภาษาธรรมชาติ (ภาษาไทย) (วิทยานิพนธ์/Thesis)*. มหาวิทยาลัยวลัยลักษณ์.

## ภาคผนวก ก

## รายงานตัวชี้วัด

## 1. นำเสนอสนงการประชุมหรือสัมมนาระดับชาติ

ลำดับ ที่	ชื่อนักวิจัยและชื่อผลงาน	เป็นส่วนหนึ่งของงานวิจัยเรื่อง	ปีงบประมาณ	แหล่งทุน	ตีพิมพ์เป็น Proceedings
					Oral
1	ไกรศักดิ์ เกษร, “ระบบสอบถามข้อมูลอัจฉริยะสำหรับข้อมูลการท่องเที่ยวในประเทศไทย” วิทยาศาสตร์วิจัยครั้งที่ 6, มหาวิทยาลัยบูรพา จังหวัดชลบุรี	การพัฒนาต้นแบบระบบสอบถามข้อมูลการท่องเที่ยวอัจฉริยะ	2556	วช. (แผ่นดิน)	1

ภาคผนวก ข

จดหมายตอบรับ





From: ดร. กฤษณะ ชินสาร

Sent: Monday, February 10, 2014 9:40:38 AM (UTC+07:00) Bangkok, Hanoi, Jakarta

To: kraisak kesorn

Subject: [src6] แจ้งผลการพิจารณาผลงานวิจัยงานประชุมวิชาการ วิทยาศาสตร์วิจัยครั้งที่ ๖

เรียน Kraisak Kesorn:

หลังจากผู้ทรงคุณวุฒิได้ทำการพิจารณาบทความวิจัยในหัวข้อ

"Intelligent Question-Answer Agent System for Tourism Information in Thailand" แล้ว ทางฝ่ายวิชาการของงานประชุมฯ มีความยินดีที่จะแจ้งให้ผู้ส่งบทความทราบว่า ผลงานวิจัยของท่านได้รับการ "ตอบรับ" เพื่อนำเสนอผลงานในงานประชุมวิชาการ การประชุมวิชาการระดับชาติ "วิทยาศาสตร์วิจัยครั้งที่ 6" หลังจากที่ท่านได้นำการ "แก้ไข" ตามข้อเสนอแนะและส่งกลับมาให้ผู้ทรงคุณวุฒิได้พิจารณาอีกครั้ง

ขอขอบคุณที่ส่งผลงานเข้าร่วมนำเสนอในงานประชุมฯ

ดร. กฤษณะ ชินสาร  
Faculty of Informatics,  
Burapha University  
[krisana@buu.ac.th](mailto:krisana@buu.ac.th)



ภาคผนวก ข

บทความฉบับสมบูรณ์



# ระบบสอบถามข้อมูลอัจฉริยะสำหรับข้อมูลการท่องเที่ยวในประเทศไทย

## Intelligent Question-Answer Agent System for Tourism Information in Thailand

ไกรศักดิ์ เกษร

Kraisak Kesorn

ภาควิชาวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยนเรศวร

### บทคัดย่อ

อุตสาหกรรมการท่องเที่ยวจัดได้ว่าเป็นธุรกิจที่มีความเติบโตอย่างรวดเร็วและเป็นอุตสาหกรรมที่สร้างรายได้ให้กับประเทศอย่างสูงในแต่ละปี ภาครัฐจึงมีนโยบายให้ทุกฝ่ายช่วยกันพัฒนาส่งเสริมคุณภาพและมาตรฐานของการบริการและกระตุ้นการท่องเที่ยวโดยเพิ่มความหลากหลายของรูปแบบการให้บริการและการท่องเที่ยว ดังนั้นบทความนี้จึงนำเสนอระบบการให้บริการข้อมูลการท่องเที่ยวโดยอาศัยคอมพิวเตอร์เป็นเครื่องมือหลักในการให้ข้อมูลจากการสอบถามของผู้ใช้โดยอัตโนมัติ หัวใจสำคัญของระบบนี้คือ 1) เทคนิคการวิเคราะห์คำถาม เพื่อให้เข้าใจความหมายที่แท้จริงของคำถามนั้นๆ เนื่องจากคำถามอาจมีความกำกวม ซึ่งวิธีการที่นำเสนอนี้จะใช้บทบาทหรือหน้าที่ของคำต่างๆ มาช่วยในการวิเคราะห์คำถาม 2) นอกจากนั้นระบบที่นำเสนอยังเก็บข้อมูลการท่องเที่ยวในรูปแบบขององค์ความรู้โดยใช้ออนโทโลยี ทำให้สามารถจัดการคำถามที่มีความกำกวมได้อย่างมีประสิทธิภาพมากขึ้น ผลการทดลองชี้ให้เห็นว่าระบบเข้าใจคำถามและสามารถหาคำตอบที่สอดคล้องกับคำถามเพิ่มสูงขึ้นนอกจากนี้ระบบยังให้ข้อมูลแก่นักท่องเที่ยวที่ตรงประเด็นมากกว่าเมื่อเปรียบเทียบกับระบบค้นหาข้อมูลทั่วไป เช่น Google

คำสำคัญ : ออนโทโลยี, ระบบถาม-ตอบอัตโนมัติ, การวิเคราะห์ประโยค

### Abstract

With the rapid growth of the tourism industry, it has become a major business and generates a huge profits for the country. Therefore, the government tries to offer various services to tourists in order to enhance the quality of services and persuade travelers to visit Thailand. This paper presents a novel framework for a question-answer agent system that can automatically provide tourism information for tourists. The key ideas of this research are: 1) a novel technique for question analysis using word roles and 2) knowledge-based design using an ontology for handling ambiguous questions. The experimental results show that the proposed framework can efficiently find the correct information to a user and improve the precision value compared to Google.

Keywords : ontology, chat robot, sentence analysis



## 1. บทนำ

ประเทศไทยเป็นประเทศหนึ่งที่นักท่องเที่ยวสนใจเดินทางมาเยี่ยมเยือน เนื่องจากความโดดเด่นในเรื่องของทรัพยากรการท่องเที่ยว โดยมีความหลากหลายและมีเอกลักษณ์เฉพาะของสถานที่ท่องเที่ยวทางธรรมชาติและทางวัฒนธรรม แต่อย่างไรก็ตามจากผลสำรวจของกรมการท่องเที่ยว พบว่าประเทศไทยยังให้ข้อมูลเกี่ยวกับแหล่งท่องเที่ยวไม่เพียงพอหรืออุตสาหกรรมการท่องเที่ยวของไทยเองไม่สามารถสร้างสื่อประชาสัมพันธ์ที่เข้าถึงกลุ่มตลาดเป้าหมายได้อย่างมีประสิทธิภาพ นักท่องเที่ยวอาจจะไม่มีความสะดวกในการเข้าไปที่สำนักงานการท่องเที่ยวแห่งประเทศไทยเพื่อสอบถามข้อมูลการท่องเที่ยวที่ตนเองสนใจ ทั้งนี้การเพิ่มศักยภาพทางการท่องเที่ยวเพื่อแก้ไขปัญหาดังกล่าวสามารถทำได้โดยการพัฒนากระบวนการข้อมูลและศูนย์รวมข้อมูลของการท่องเที่ยว ซึ่งสอดคล้องกับนโยบายของกระทรวงการท่องเที่ยวและกีฬาที่ต้องการพัฒนาการตลาดเชิงอิเล็กทรอนิกส์และระบบให้ข้อมูลการท่องเที่ยวในประเทศไทย การพัฒนาระบบสารสนเทศเพื่อให้ข้อมูลการท่องเที่ยวในประเทศไทยนี้มีวัตถุประสงค์เพื่ออำนวยความสะดวกในการให้ข้อมูลกับนักท่องเที่ยว อีกทั้งยังช่วยลดต้นทุนในการจ้างมนุษย์ในการทำหน้าที่ดังกล่าวได้อีกด้วย

การให้ข้อมูลด้านการท่องเที่ยวของประเทศไทยผ่านทางเว็บไซต์ต่างๆ ที่มีอยู่ในปัจจุบันยังมีประสิทธิภาพไม่เพียงพอที่จะนำเสนอข้อมูลให้ตรงกับความต้องการที่แท้จริงของนักท่องเที่ยว และระบบสนับสนุนการค้นหาข้อมูลด้านการท่องเที่ยวยังไม่ดีเท่าที่ควร ตัวอย่างเช่น นักท่องเที่ยวต้องการค้นหาข้อมูลเกี่ยวกับโบราณสถานในประเทศไทยโดยใช้คำสำคัญในการค้นหาว่า "อารามหลวง" ระบบค้นหาข้อมูลในปัจจุบัน เช่น Google จะค้นพบเฉพาะเอกสารที่ปรากฏคำสำคัญนี้เท่านั้น เอกสารอื่นๆ ที่มีเนื้อหา (Concept) เดียวกันแต่ไม่มีคำสำคัญเหล่านี้ปรากฏอยู่ในเอกสาร ระบบค้นหาข้อมูลก็จะไม่สามารถค้นพบและนำเสนอต่อนักท่องเที่ยวได้มีการพัฒนาวิธีการขึ้นมาหลายวิธีแต่ก็ยังไม่สามารถให้ผลลัพธ์ที่ตรงความต้องการเนื่องจากระบบยังมีปัญหาอย่าง เช่น ระบบที่ยังไม่สามารถวิเคราะห์คำถามเชิงความหมายได้ (Non-Semantic): เป็นระบบที่ใช้วิธีการเทียบตัวอักษรกับฐานข้อมูลโดยตรง คำที่ค้นหาต้องเป็นคำที่มีอยู่ในฐานข้อมูลและต้องสะกดคำอย่างถูกต้องเท่านั้นระบบก็จะแสดงคำตอบออกมาได้

จากปัญหาดังกล่าวผู้วิจัยจึงมีแนวคิดที่จะพัฒนาระบบถาม-ตอบอัตโนมัติสำหรับให้ข้อมูลการท่องเที่ยวในเขตจังหวัดพิษณุโลก โดยต้นแบบของระบบนี้จะเน้นการให้ข้อมูลเพื่อตอบคำถามประเภทข้อมูลทั่วไป (What) และที่อยู่ของสถานที่ท่องเที่ยว (Where) เป็นหลักก่อนและจะขยายการทำงานไปสู่คำถามประเภทอื่นๆ ต่อไปในอนาคต

## 2. ทบทวนวรรณกรรม

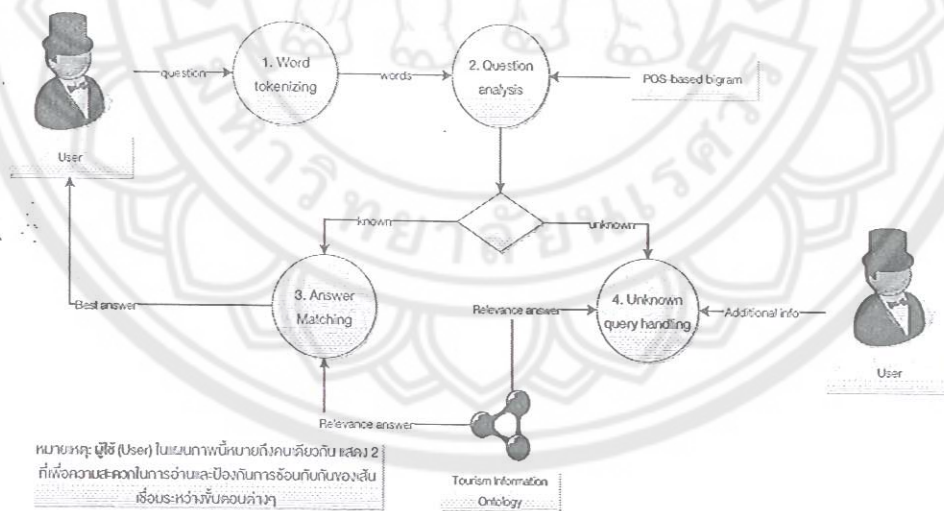
ความต้องการที่จะให้มีแหล่งข้อมูลที่ถูกต้องและเหมาะสมเกี่ยวกับการท่องเที่ยวเพื่อใช้ในการเผยแพร่ข้อมูลต่างๆ แก่นักท่องเที่ยวทำให้ระบบสอบถามข้อมูลอัตโนมัติเป็นวิธีการหนึ่งที่ถูกนำมาใช้ในการสนับสนุนแนวคิดดังกล่าว การสร้างระบบถามตอบอัตโนมัติได้รับความสนใจจากนักวิจัยจำนวนมาก

จตุณี (2548) ได้ทำระบบตอบถาม-ตอบโดยการนำประโยคที่เป็นภาษาธรรมชาติมาเปรียบเทียบกับกลุ่มคำกับข้อมูลที่เก็บในรูปแบบ XML ซึ่งเป็นระบบที่จะเน้นการตอบคำถามที่เกี่ยวกับเวลา (When) เท่านั้น เช่นเมื่อไรเวลาไหนวันไหนเดือนไหนซึ่งยังไม่สามารถตอบคำถามประเภทอื่นๆ ได้เช่นอะไร (What) สถานที่ (Where) อย่างไร (How) นอกจากนี้ระบบไม่สามารถวิเคราะห์คำถามเชิงความหมายได้ เนื่องจากการทำงานเป็นในลักษณะของการเปรียบเทียบอักขระ (String matching) ชีวาลและคณะ (2551) ได้พัฒนาระบบที่ชื่อว่า อับดุล (Abdul) หรือระบบบริการข้อมูลเชิงบูรณาการผ่านระบบสหราชอาณาจักรที่สามารถตอบคำถามผู้ใช้ได้เสมือนมนุษย์ระบบจะให้บริการข้อมูลสาธารณะทั่วไปแก่ผู้ใช้ซึ่งมีประโยชน์ที่ผู้ใช้ไม่ต้องเข้าถึงแหล่งข้อมูลโดยตรงหัวใจหลักของงานวิจัยนี้คือกระบวนการตัดคำ (Word Segmentation) และแปลงรูปแบบภาษาให้อยู่ในรูปแบบที่คอมพิวเตอร์สามารถเข้าใจได้และใช้ AIML Engine ในการพิจารณาคำตอบแต่เนื่องจากระบบอับดุลนี้ใช้เพื่อตอบคำถามทั่วไป ดังนั้นจึงไม่สามารถตอบคำถามเรื่องของการท่องเที่ยวได้อย่างมีประสิทธิภาพ ทศพล (2552) พัฒนาระบบตอบคำถามที่เกี่ยวกับเนื้อหาที่อยู่เอกสารต่างๆ ในฐานข้อมูลของงานวิจัย โดยใช้วิธีการ Rule-based (เงื่อนไข If-then) มาใช้ในการวิเคราะห์คำถามและค้นหาคำตอบให้กับผู้ใช้ นอกจากนี้ยังนำ WordNet (คือฐานข้อมูลคำศัพท์ออนไลน์ภาษาอังกฤษที่ออกแบบตามทฤษฎีทางภาษาศาสตร์) มาช่วยในการแบ่งประเภทของคำตอบและนาเทคนิค (Bag of Words) มาใช้ในการคำนวณหาความสัมพันธ์ระหว่างประโยคคำถามและคำตอบอย่างไรก็ตามวิธีการ Rule-based ที่ใช้ในงานวิจัยนี้ยังไม่รองรับการวิเคราะห์คำถามเชิงความหมายเช่นกัน Guo et al. (2008) ได้นำเว็บเชิงความหมาย (Semantic web) และออนโทโลยีมาใช้เป็นเครื่องมือที่สำคัญในการสร้างคลังความรู้ด้านคำศัพท์ด้วยออนโทโลยี เพื่อนำไปใช้ในการตอบคำถาม ข้อดีของงานวิจัยนี้คือ ระบบสามารถเรียนรู้คำถามที่ไม่อยู่ในฐานข้อมูลและสามารถขยายฐานความรู้ได้โดยอาศัยความสัมพันธ์ของคำต่างๆ ในดัชนี แต่ระบบนี้สนับสนุนคำถามเฉพาะภาษาจีนเท่านั้น Hao et al. (2008) ระบบได้ตอบสำหรับคน

และคอมพิวเตอร์ด้วยการใช้ Semantic Pattern ในการทำให้คอมพิวเตอร์เข้าใจคำถามซึ่งจะนำคำถามที่ได้มาสร้างรูปแบบ (Pattern) ของตัวเองแล้วก็ไปเทียบกับคลังรูปแบบ ที่สร้างไว้ เพื่อวิเคราะห์ความหมายของคำถาม และดึงคำตอบที่เกี่ยวข้องออกมาแสดง Shen (2007) นำบทบาทของความหมายมาใช้เพื่อทำให้ระบบถามตอบมีประสิทธิภาพยิ่งขึ้น โดยการสร้างเฟรมเวิร์กสำหรับแยกคำตอบซึ่งดูจากพฤติกรรมทางความหมายและมีการหาคำความหมายของปัญหาและการแยกคำตอบโดยใช้กราฟแมชชีน จากการศึกษาระบบงานเดิมพบข้อจำกัดคือ ระบบที่มีอยู่โดยส่วนมากใช้วิธีเทียบตัวอักษรเมื่อระบบเจอคำที่ไม่มีในฐานข้อมูล ระบบจะไม่สามารถประมวลผลคำนั้นและไม่สามารถหาคำตอบได้ด้วยเหตุผลดังกล่าวผู้วิจัยจึงเกิดแนวคิดที่จะพัฒนาระบบถาม-ตอบที่สามารถเข้าใจคำถามของผู้ใช้เชิงความหมายและสามารถจัดการคำถามที่ไม่รู้จักได้อย่างมีประสิทธิภาพมากขึ้น

3. ระบบถาม-ตอบคำถามการท่องเที่ยวอัตโนมัติ (Automatic Tourism Question-Answer System)

ระบบสอบถามข้อมูลการท่องเที่ยวอัตโนมัติ (Automatic Tourism Question-Answer System-ATQ) คือระบบที่ให้ข้อมูลแก่นักท่องเที่ยวโดยอัตโนมัติโดยที่ผู้ใช้สามารถสอบถามข้อมูลการท่องเที่ยวกับระบบได้เหมือนกับคุยกับพนักงานคนหนึ่งที่สามารถให้ข้อมูลการท่องเที่ยวแก่ผู้ใช้ได้ โดยต้นแบบระบบที่จะพัฒนาขึ้นในการศึกษาวิจัยนี้จะจำกัดข้อมูล (Domain) อยู่ที่ข้อมูลการท่องเที่ยวในเขตพื้นที่จังหวัดพิษณุโลก หัวใจสำคัญของระบบนี้คือการสร้างระบบที่สามารถติดต่อสื่อสารกับผู้ใช้ได้อย่างมีประสิทธิภาพ เปรียบเสมือนเป็นตัวแทน (Agent) ของพนักงานในบริษัทท่องเที่ยวที่คอยให้ข้อมูลกับนักท่องเที่ยวได้ตลอดเวลาโดยจุดประสงค์หลักของโครงการวิจัยนี้ก็คือพยายามพัฒนาเทคนิคที่ทำให้ระบบคอมพิวเตอร์เข้าใจภาษาไทยและนำไปประยุกต์ใช้กับระบบสอบถามข้อมูลการท่องเที่ยวโดยกรอบแนวคิด (Framework) ของการศึกษานี้แบ่งออกเป็น 2 ส่วนหลักๆ โดยที่แต่ละส่วนจะให้ความสำคัญกับงานที่ต่างกัน ทั้งนี้เพื่อให้ระบบคอมพิวเตอร์สามารถเข้าใจสิ่งที่ผู้ใช้หรือนักท่องเที่ยวต้องการสื่อสารกับระบบ โดยที่ส่วนที่ 1 จะเน้นไปที่การวิเคราะห์คำถาม (Question analysis) เพื่อนำไปสู่การตีความประโยค (Interpretation/Understanding) เพื่อให้ได้คิดความใจความหมายของคำถามต่างๆที่นักท่องเที่ยวสื่อสารกับระบบโดยการพิมพ์ผ่านหน้าจอ และส่วนที่ 2 คือการจัดการกับคำถามที่ระบบไม่รู้จัก เนื่องจากนักท่องเที่ยวแต่ละคนอาจจะเขียนคำถามที่ใช้คำพูดไม่เหมือนกันดังนั้นจึงเป็นไปได้ยากที่ระบบจะเข้าใจคำถามทุกคำถามของนักท่องเที่ยว ดังนั้นส่วนการจัดการคำถามที่ไม่รู้จักนี้ ระบบจะพยายามที่จะเรียนรู้คำถามใหม่ๆ ที่เข้ามาในระบบและเพิ่มเข้าไปในองค์ความรู้ของระบบโดยอัตโนมัติ ข้อมูลโดยสรุปของกรอบแนวคิดของระบบ ATQ สามารถอธิบายได้ดังนี้



รูปที่ 1 ส่วนประกอบของระบบ ATQ

3.1 ระบบตัดคำ (Word tokenizing)

ขั้นตอนนี้เป็นกระบวนการนำประโยคคำถามที่ได้จากผู้ใช้มาแบ่งออกเป็นหน่วยของคำย่อยๆ เพื่อนำไปวิเคราะห์หาความหมายของคำถามต่อไป เนื่องจากเทคนิคการตัดคำไม่ใช่จุดประสงค์หลักของงานวิจัยนี้ ดังนั้นนักวิจัยจึงใช้ซอฟต์แวร์ตัดคำที่ถูกพัฒนาไว้เรียบร้อยแล้วที่ชื่อว่า "โปรแกรมตัดคำด้วยเทคนิคไฮบริด" (วรศักดิ์ และคณะ, 2552) ข้อดีของโปรแกรมนี้คือพัฒนาโดยภาษาจาวาที่สามารถนำมาเชื่อมต่อกับระบบใหม่ที่น่าเสนอได้โดยง่ายและสามารถตรวจสอบคำย่อยจากข้อมูล เช่น จ. พ.ศ. ตรวจสอบชื่อเฉพาะในฐานข้อมูลและ



ตรวจสอบคำที่คิดว่าน่าจะเป็นคำเฉพาะจากฐานข้อมูลคำขึ้นต้น ตัวอย่างประโยคคำถามเช่น โรงแรมอะไรที่อยู่ใกล้วัดใหญ่ ผลที่ได้จากการตัดคำจากโปรแกรมนี้คือ โรงแรม/อะไร/ที่อยู่/ใกล้/วัด/ใหญ่

### 3.2 การวิเคราะห์คำถาม (Question analysis)

เมื่อได้รับหน่วยคำจากกระบวนการตัดคำในขั้นตอนที่ 1 แล้ว ขั้นตอนถัดไปคือนำคำต่างๆ มาวิเคราะห์เพื่อแปลความหมายของคำถามในขั้นตอนนี้ประกอบด้วย 2 ขั้นตอนย่อยๆ คือ

#### 3.2.1 การวิเคราะห์โครงสร้างของประโยค (Question's structure analysis)

กระบวนการนี้จะนำคำที่ได้มาหาประเภทของคำ เช่น คำนามคำเฉพาะและหาประเภทของคำถามโดยเทียบจากฐานความรู้และนำมาเปลี่ยนโครงสร้างของประโยคใหม่โดยใช้ข้อมูลเพิ่มเติมให้กับคำต่างๆ เพื่อให้ง่ายต่อการวิเคราะห์ความหมายมากขึ้นซึ่งในขั้นแรกจะต้องหาหน้าที่ของคำก่อน โดยผลการหาหน้าที่ของคำจะแสดงในรูปแบบ <หน้าที่ของคำ, pos> โดยที่ pos (position) หมายถึงลำดับของคำในประโยคเพื่อประโยชน์ในการเข้าใจคำถามที่ถูกต้องมากยิ่งขึ้น ส่วน "หน้าที่ของคำ" จะประกอบด้วยคำประเภทต่างๆ ดังต่อไปนี้

1. <antecedent, pos:>หมายถึงคำที่บอกลำนำหน้าชื่อเฉพาะของคำ เช่น วัดน้ำตก โรงแรมมหาวิทยาลัยร้านอาหารร้านอุทยานอุทยานแห่งชาติ เป็นต้น
2. <noun, pos:>หมายถึงชื่อที่บ่งบอกถึงสถานที่นั้นๆ เช่น วัดใหญ่ น้ำตกแห่งของพลาโลมา เป็นต้น
3. <preposition, pos: >หมายถึง คำบุพบท ที่ปรากฏในประโยคคำถาม เช่น อยู่ใกล้ระหว่างใกล้ติดกับ เป็นต้น
4. <adjective, pos: >หมายถึง คำคุณศัพท์ ที่ปรากฏในประโยคคำถาม เช่น สวย ใหญ่ เล็ก เป็นต้น
5. <qtype, pos: >หมายถึงคำที่ใช้ตอบปัญหา เช่น อะไรที่ใหนอย่างไรตรงใหน เป็นต้น โดยที่หากคำเหล่านี้ปรากฏใกล้คำถามจะถือว่าเป็นส่วนคำถามหลักของประโยค
6. <verb, pos>หมายถึง คำกริยา ที่ปรากฏในประโยคคำถาม เช่น มี ไหว้ ไป
7. <stop word, pos>หมายถึง คำที่ไม่มีความสำคัญในการนำมาใช้ประโยชน์เพื่อการวิเคราะห์คำถาม ซึ่งผู้วิจัยจะไม่นำมาใช้ในกระบวนการ เช่น จะหรือที่ เป็นต้น

ตัวอย่างเช่น จากขั้นตอนการตัดคำในขั้นตอนที่ 1 จะได้ "โรงแรม/อะไร/ที่อยู่/ใกล้/วัด/ใหญ่" เมื่อผ่านขั้นตอนการวิเคราะห์โครงสร้างของประโยค ผลลัพธ์ที่ได้แสดงดังต่อไปนี้

โครงสร้างประโยค: "<antecedent, pos:โรงแรม, 1><qtype, pos: อะไร, 2><stop word, pos: ที่, 3><stopword, pos: อยู่, 4><preposition, pos: ใกล้, 5><antecedent, pos:วัด, 6><adjective, pos:ใหญ่, 7>"

โครงสร้างของประโยคที่ได้นี้จะไปประมวลผลต่อเพื่อลดความซ้ำซ้อนของคำ รวมคำที่ทำหน้าที่เหมือนกัน และตัดคำที่ไม่สำคัญทิ้ง ซึ่งจะเรียกขั้นตอนดังกล่าวว่า "การทำคามสะอาดประโยค"

#### 3.2.2 การทำความสะอาดคำในประโยค (Sentence cleansing)

ขั้นตอนนี้จะช่วยลดคำที่ไม่จำเป็นในประโยคเพื่อการประมวลผลที่ถูกต้องและรวดเร็วยิ่งขึ้น ดังนั้นจะมีการรวมคำและตัดคำที่สำคัญทิ้งไป ในการกระทำดังกล่าวจะใช้หลักการของ Part-of-Speech (POS) bigram (Rajagopal, Cambria, Olsheř, & Kwok, 2013) โดยมีหลักการดังต่อไปนี้

- 1) Antecedent+Noun: จะนำคำมารวมกัน แสดงว่าเป็นชื่อเฉพาะ
- 2) Antecedent+Adjective: จะนำคำมารวมกัน แสดงว่าเป็นชื่อเฉพาะ
- 3) Noun+Adjective: ระบบจะนำ 2 คำมารวมกัน
- 4) Noun+Noun: ทั้ง 2 คำจะถูกรวมกันเป็นคำเดียว ระบบจะคิดว่าทั้ง 2 คำถูกใช้ร่วมกันเพื่อแทนชื่อสถานที่หรือบางสิ่ง
- 5) Stopword: ระบบจะตัดคำที่ไม่สำคัญทิ้ง



- 6) สำหรับกรณีอื่นๆ ที่อยู่นอกเหนือจาก 5 ข้อนี้ คำเหล่านี้จะยังคงเดิม จะไม่ถูกตัดทิ้งและจะไม่ถูกนำมารวมกัน เช่น คำกริยา (Verb) ต่างๆ จะไม่ถูกนำไปรวมกับคำอื่นๆ

จากหลักการข้างต้นจะทำให้ได้ผลลัพธ์จากตัวอย่างในหัวข้อ 3.2.1 ดังต่อไปนี้

"<antecedent, pos: โรงแรม, 1><qtype, pos: อะไร, 2><preposition, pos: ใกล้, 5><antecedent, pos: วัดใหญ่, 6>"

ต่อจากนั้นระบบจะนำองค์ประกอบของคำถามดังกล่าวไปค้นหาข้อมูลในออนโทโลยีต่อไปโดยใช้ SPARQL และนำผลลัพธ์มาจัดลำดับความคล้ายคลึง (Similarity) โดยการคำนวณแบบโคไซน์ (Cosine)

### 3.2.3 การจัดการคำถามที่ไม่รู้จัก (Unknown question handling)

ในบางคำถามเมื่อผ่านกระบวนการทำความเข้าใจความสะอาดคำในประโยคแล้ว จะเหลือเพียงคำสำคัญไม่กี่คำ ซึ่งมีผลทำให้ไม่สามารถประกอบคำเป็นทUPLE เพื่อไปเปรียบเทียบกับโครงสร้างของออนโทโลยีได้ ตัวอย่างเช่น "มีห้างสรรพสินค้าตรงไหนบ้าง" เมื่อผ่านกระบวนการตัดคำ วิเคราะห์โครงสร้างประโยคและการทำความเข้าใจความสะอาดคำในประโยคแล้วจะได้

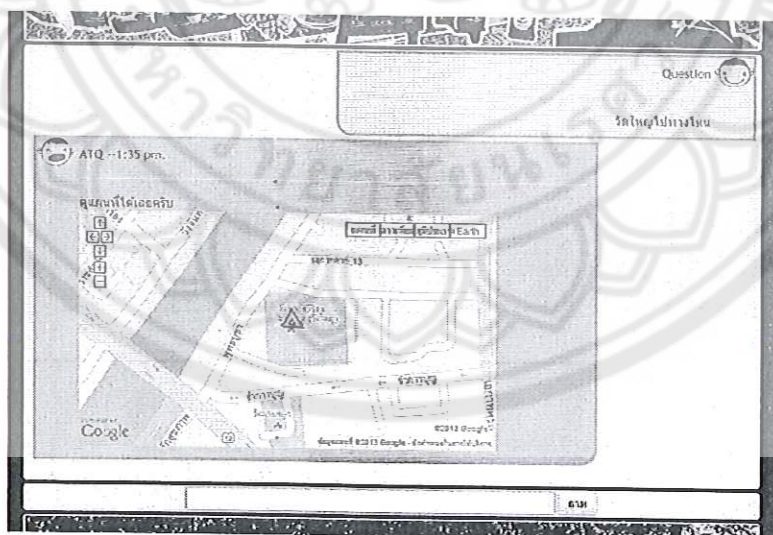
<verb, pose: มี, 1><antecedent, pos: ห้างสรรพสินค้า, 2> <qtype, pos: ตรงไหน, 3>

จากตัวอย่างดังกล่าวประเภทของคำถามคือหาตำแหน่งของสถานที่ (ตรงไหน) และคำสำคัญที่เหลือไม่สามารถประกอบเป็น TUPLE ได้อย่างสมบูรณ์ อย่างไรก็ตามในกรณีเช่นนี้ระบบจะพยายามคาดเดาความหมายของคำถาม ระบบจะพยายามแปลงคำสำคัญเหล่านี้ให้อยู่ในรูปแบบทUPLE และเติมสิ่งที่หายไปด้วยตัวแปร (X) ดังนั้นเราจะได้ "X-<มี>-ห้างสรรพสินค้า" ดังนั้นระบบสามารถแปลความหมายของคำถามได้ว่า "ตำแหน่งใดๆ (X) ที่มีห้างสรรพสินค้า" และระบบจะค้นหาข้อมูลจาก TUPLE ดังกล่าวโดยแทนที่ X ด้วยพิกัดบนแผนที่ของห้างสรรพสินค้าทุกแห่งในจังหวัดพิษณุโลก

## 4. ผลการทดลองและการอภิปราย

จากแนวคิดที่นำเสนอในหัวข้อที่ผ่าน ผู้วิจัยได้ทำการทดลองเพื่อทดสอบประสิทธิภาพของระบบ ATQ ดังต่อไปนี้

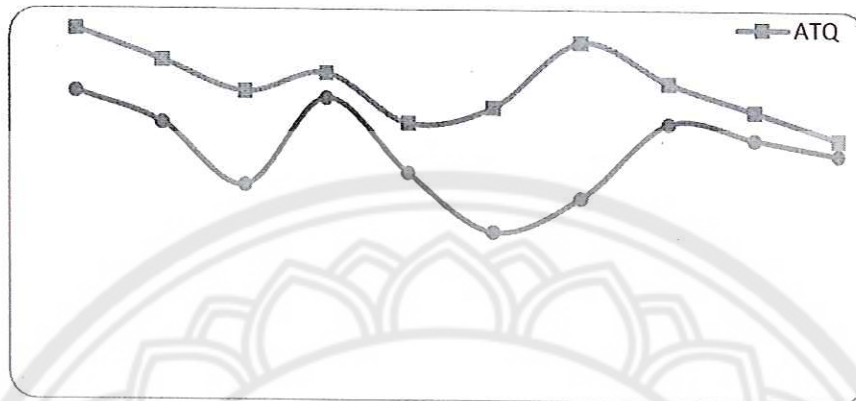
- 1) การวิเคราะห์คำถามเชิงความหมาย: เมื่อผู้ใช้ใส่คำถามที่ต่างกัน แต่มีความหมายเหมือนกัน ด้วยเทคนิคการวิเคราะห์คำถามเชิงความหมายที่นำเสนอ ระบบจะสามารถให้คำตอบที่เกี่ยวข้องกับคำถามได้อย่างถูกต้อง แม่นยำ เช่น วัดใหญ่อยู่ตรงไหน ทางไปวัดใหญ่ ที่ตั้งวัดใหญ่ ระบบจะให้คำตอบเป็นแผนที่ตำแหน่งของสถานที่ที่เกี่ยวข้องดังกล่าว แสดงดังรูปที่ 2



รูปที่ 2 การให้คำตอบประเภทตำแหน่งของสถานที่ที่เกี่ยวข้อง

- 2) เปรียบเทียบประสิทธิภาพการค้นหาข้อมูลกับ Google: ผู้วิจัยได้ทดลองทำการใส่คำถามจำนวน 10 คำถามที่สุ่มมาจากชุดคำถามที่ได้มาจากการสอบถามนับห้องที่เกี่ยวข้องจากจำนวนทั้งหมด 84 คำถาม เกี่ยวกับสถานที่ที่เกี่ยวข้องในเขตจังหวัดพิษณุโลก เช่น ตำแหน่งที่ตั้ง ประวัตินั้น และใช้คำถามเดียวกันกับ Google และประเมินประสิทธิภาพของระบบทั้งสองจากค่าความแม่นยำ (Precision) ซึ่งหมายถึงข้อมูลที่ถูกต้อง

เปรียบเทียบกับข้อมูลที่ค้นหาได้ทั้งหมดใน 10 อันดับแรก ผลการทดลองแสดงดังรูปที่ 3 ซึ่งเห็นว่าระบบ ATQ สามารถปรับปรุงประสิทธิภาพการค้นหาข้อมูลและตอบคำถามโดยใช้ภาษาไทยได้สูงกว่า Google ที่มีความแม่นยำเฉลี่ยอยู่ที่ 85% ในขณะที่ Google มีความแม่นยำเฉลี่ยที่ 74% สำหรับคำถามที่ระบบไม่สามารถหาคำตอบได้คือ คำถามที่มีโครงสร้างที่ซับซ้อน เช่น “โรงแรมใกล้เซ็นทรัลและร้านอาหารอีสาน” เมื่อวิเคราะห์ประโยคออกมาทำให้ระบบไม่สามารถหาข้อมูลในออนไลน์ที่ถูกต้องได้



รูปที่ 3 ค่าความแม่นยำในการค้นหาข้อมูลที่ตรงกับคำถามผู้ใช้

## 5. บทสรุป

แนวคิดของระบบถามตอบที่นำเสนอในบทความนี้สามารถปรับปรุงประสิทธิภาพการให้ข้อมูลหรือค้นหาข้อมูลการท่องเที่ยวให้สูงขึ้นโดยใช้หลักการของการวิเคราะห์ประโยคคำถาม โดยใช้หน้าที่ของคำวิเคราะห์เพื่อหาความหมายของคำถามนั้นๆ อย่างไรก็ตามวิธีการที่นำเสนอนี้ ยังมีข้อจำกัดคือใช้ได้กับระบบการท่องเที่ยวเท่านั้น เนื่องจากข้อมูลที่ใช้สอนระบบ (Train) และการวิเคราะห์ประโยคคำถามอยู่บนพื้นฐานของโดเมน (Domain) การท่องเที่ยวเท่านั้น ดังนั้นหากจะนำไปประยุกต์ใช้ถามตอบในเรื่องอื่นๆ จะต้องทำการสอนระบบใหม่เช่นกัน และระบบดังกล่าวยังไม่สามารถตอบคำถามที่เกี่ยวข้องกับเวลาได้ ซึ่งจะเป็นงานวิจัยที่พัฒนาต่อไปในอนาคต

## 6. กิตติกรรมประกาศ

งานวิจัยนี้ได้รับทุนสนับสนุนจากสำนักงานคณะกรรมการการวิจัยแห่งชาติ (วช.) ประจำปีงบประมาณ 2556 รหัสโครงการ R2556B011

## 7. เอกสารอ้างอิง

- Guo, Q., and Zhang, M. (2008). Question Answering System Based on Ontology and Semantic Web. In G. Wang, T. Li, J. W. Grzymala-Busse, D. Miao, A. Skowron, & Y. Yao (Eds.), *Rough Sets and Knowledge Technology*, 652–659.
- Hao, T., and Wenyin, L. (2008). Automatic Question Translation Based on Semantic Pattern. In *Fourth International Conference on Semantics, Knowledge and Grid*, 372–375.
- Rajagopal, D., Cambria, E., Olsher, D., and Kwok, K. (2013). A Graph-Based Approach to Commonsense Concept Extraction and Semantic Similarity Detection. In *Proceedings of the 22nd International Conference on World Wide Web Companion* 565–570.
- Shen, D. (2007). Using Semantic Role to Improve Question Answering. In *Proceedings of EMNLP 2007*.
- ชัชวาล สังคีตตระการ, ศราวุธ คงยัง, มารุต บุรณรัช, ชูชาติ นฤไชยะศักดิ์, และ อลิสา คงทน. (2551). อັบลู:ระบบบริการข้อมูลเชิงบูรณาการผ่านระบบสนทนาออนไลน์, *NECTEC Technical Journal*, 1-6.
- ทศพล วงศ์ศิริไพศาล. (2552). การใช้กฎในระบบตอบคำถามภาษาไทย สำหรับทดสอบการอ่านเพื่อความเข้าใจ. มหาวิทยาลัยธรรมศาสตร์.
- วรงค์ศักดิ์ ตังกุลทวีทรัพย์, พีระศักดิ์ รัตนมณี, และ ธนพล จินดาพิทักษ์. (2552). การแบ่งคำภาษาไทยด้วยเทคนิคไฮบริด, การแข่งขันสุดยอดซอฟต์แวร์แบ่งคำภาษาไทย, มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตหาดใหญ่.
- อรุณี โฉหารานนท์. (2548). ระบบตอบคำถามเป็นภาษารวมชาติ (ภาษาไทย) (วิทยานิพนธ์). มหาวิทยาลัยวลัยลักษณ์.