



รายงานการวิจัย

เรื่อง

ออนโทโลยีพื้พพิวเลชันจากเอกสาร สำหรับโดเมนการท่องเที่ยว
Ontology Population from Text for Tourism Domain

โดย

อรรรรณ อิมสมบัติ

มหาวิทยาลัยธุรกิจบัณฑิตย์

รายงานผลการวิจัยนี้ได้รับทุนอุดหนุนจากมหาวิทยาลัยธุรกิจบัณฑิตย์

พ.ศ. 2561

ชื่อเรื่อง : ออนโทโลยีเพื่อพหุวัจนจากเอกสาร สำหรับโดเมนการท่องเที่ยว
ผู้วิจัย : อรวรรณ อิมสมบัติ สถาบัน : มหาวิทยาลัยธุรกิจบัณฑิต
ปีที่พิมพ์ : พุทธศักราช 2561 สถานที่พิมพ์ : มหาวิทยาลัยธุรกิจบัณฑิต
แหล่งที่เก็บรายงานการวิจัยฉบับสมบูรณ์ : ศูนย์บริการวิจัย มหาวิทยาลัยธุรกิจบัณฑิต
จำนวนหน้าวิจัย : 49 หน้า ลิขสิทธิ์ : มหาวิทยาลัยธุรกิจบัณฑิต
คำสำคัญ : ออนโทโลยีเพื่อพหุวัจน, โดเมนการท่องเที่ยว, คอนดิชันนอลแรนดอมฟิลด์,
พีเจอร์เวจ เค-เนียร์เรสเนเบอร์

บทคัดย่อ

งานวิจัยนี้นำเสนอเทคนิคในการทำออนโทโลยีเพื่อพหุวัจนจากเอกสาร โดยประยุกต์ใช้เทคนิคการประมวลผลภาษาธรรมชาติ และเทคนิคการเรียนรู้ด้วยเครื่องจักร เพื่อเป็นการขยายความรู้ในออนโทโลยีในโดเมนที่สนใจให้มีปริมาณมากขึ้น เช่น โดเมนการท่องเที่ยว การแพทย์ และการเกษตร ซึ่งจะทำให้เกิดประสิทธิภาพในการนำออนโทโลยีไปประยุกต์ใช้ในงานต่างๆ เช่น ระบบค้นคืนสารสนเทศ ระบบให้คำแนะนำ และระบบถามตอบ เป็นต้น โดยในงานวิจัยนี้สนใจสร้างออนโทโลยีในโดเมนการท่องเที่ยว การทำออนโทโลยีเพื่อพหุวัจน คือ การเพิ่มอินสแตนซ์ให้กับคอนเซ็ปต์ในออนโทโลยี โดยอินสแตนซ์ส่วนใหญ่มักเป็นชื่อเฉพาะ เช่น ชื่อสถานที่ ดังนั้นจึงได้มีการประยุกต์ใช้เทคนิคการสกัดนิพจน์ระบุนามมาใช้ในการหาอินสแตนซ์ เริ่มต้นเอกสารนำเข้าจะถูกประมวลผลด้วยเทคนิคการประมวลผลภาษาธรรมชาติ คือ การตัดคำ และการกำกับหน้าที่ของคำ หลังจากนั้นจะถูกเรียนรู้ด้วยเทคนิค Conditional Random Fields (CRFs) เพื่อระบุขอบเขตของนิพจน์ระบุนาม ซึ่งได้แก่ ชื่อสถานที่ท่องเที่ยวต่าง ๆ และชื่อสถานที่ท่องเที่ยวเหล่านี้จะถูกจำแนกตามประเภทของคอนเซ็ปต์ หรือประเภทของสถานที่ท่องเที่ยวด้วยเทคนิค Feature-weight k-Nearest Neighbor นอกจากนี้ยังได้มีการประยุกต์ใช้เทคนิครูปแบบการเกิดร่วมกัน (Co-occurrence patterns) ในการสกัดความสัมพันธ์ระหว่างกิจกรรม กับสถานที่ท่องเที่ยวต่าง ๆ

จากการทดลองกับเอกสารภาษาไทยในโดเมนการท่องเที่ยวจำนวน 121,567 คำ ผลการทดลองพบว่าสามารถสกัดชื่อสถานที่ท่องเที่ยวได้จำนวน 857 แห่ง สามารถสกัดกิจกรรมการท่องเที่ยวได้จำนวน 84 กิจกรรม และสกัดความสัมพันธ์ได้จำนวน 941 ความสัมพันธ์ โดยขั้นตอน

การสกัดชื่อสถานที่ท่องเที่ยวมีค่าความแม่นยำ (precision) 74.38% และค่าระลึก (recall) 71.06%
และขั้นตอนสกัดความสัมพันธ์มีค่าความแม่นยำ (precision) 70.16% และค่าระลึก (recall) 69.07%

DRAFT

Title : Ontology Population from Text for Tourism Domain
Researcher : Aurawan Imsombut **Institution** : Dhurakij Pundit University
Year of Publication: 2018 **Publisher** : Dhurakij Pundit University
Sources : Dhurakij Pundit Research Service Center
Number of Pages : 49 Pages **Copyright** : Dhurakij Pundit University
Keywords : Ontology Population, Tourism Domain, Conditional Random Fields (CRFs), Feature-weighted k-Nearest Neighbor

Abstract

This research proposes the technique for populating ontology from text using natural language processing and machine learning techniques. The objective of the ontology population is to extend the knowledge in the ontology for covering the interested domain such as tourism, medication and architecture. This enhances the efficient use of ontologies for various applications e.g. information retrieval system, recommendation system and question answering system. In this research, It is focused on populating an ontology in the tourism domain. The ontology population task is adding instances to the concept in the ontology which most of the instances are unique names such as name of places. Therefore, the name entity recognition technique has been applied to find the instances. The first stage is input documents to be processed with natural language processing technique. Next, the Conditional Random Fields (CRFs), the supervised learning technique, is applied to identify the boundary of the name entity specifically name of places. These names related to their attractions are classified based to the concept or types of tourist attractions with Feature-weight k-Nearest Neighbor technique. Furthermore, co-occurrence patterns are used for extracting activity names and identified relationships between activities and attractions.

According to the experiment, 121,567 words in Thai travel domain documents, the results showed that the proposed techniques is able to extract 857 tourist attractions, 84 activity names and 941 relationships. The tourist attractions instances extracting step provided 74.38% and 71.06% of precision and recall measures respectively. In addition, relationships extracting step yielded 70.16% of precision and 69.07% of recall measures.



กิตติกรรมประกาศ

งานวิจัยฉบับนี้สำเร็จลุล่วงได้อย่างสมบูรณ์ โดยได้รับความอนุเคราะห์อย่างยิ่งจาก รองศาสตราจารย์ ดร. นุชรี เปรมชัยสวัสดิ์ ที่ปรึกษาวิทยาลัยครีเอทีฟดีไซน์ แอนด์ เอ็นเตอร์เทนเมนต์เทคโนโลยี และ ผู้ช่วยศาสตราจารย์ ดร. วรสิทธิ์ ชูชัยวัฒนา คณบดีวิทยาลัยครีเอทีฟดีไซน์ แอนด์ เอ็นเตอร์เทนเมนต์เทคโนโลยี ซึ่งได้เสียสละเวลาอันมีค่าให้คำแนะนำถึงแนวทางต่างๆ ในการวิจัย รวมทั้งชี้แนะแนวทางในการแก้ปัญหาอันเป็นประโยชน์เพื่อให้การวิจัยมีความสมบูรณ์มากยิ่งขึ้น

ขอขอบคุณ อาจารย์ เฉลิมพล ศิริกายน ผู้ช่วยศาสตราจารย์ ดร. วรภัทร ไพริเกรง และ ดร. เจษฎา ขจรฤทธิ์ ที่ช่วยให้คำแนะนำแนวทางในการวิเคราะห์ข้อมูล และการพัฒนาโปรแกรมประกอบการวิจัย

ขอขอบคุณมหาวิทยาลัยธุรกิจบัณฑิตที่ให้ทุนสนับสนุนในการทำวิจัยในครั้งนี้ นอกจากนี้ ขอขอบพระคุณครอบครัว และเพื่อนๆ ที่ให้กำลังใจในการทำวิจัยนี้ ผู้วิจัยหวังเป็นอย่างยิ่งว่า งานวิจัยนี้จะเป็นประโยชน์กับผู้ที่ต้องการศึกษาด้านการทำออนไลน์เพื่อพรีวิวเลชันจากเอกสาร ในโดเมนต่างๆ และหากมีข้อผิดพลาดประการใดในงานวิจัยฉบับนี้ ผู้วิจัยต้องกราบขอภัยเป็นอย่างสูงมา ณ ที่นี้ด้วย

อรรรรณ อิมสมบัติ
หัวหน้าโครงการ
มีนาคม 2561

สารบัญ

หน้าที่

บทคัดย่อภาษาไทย.....	ก
บทคัดย่อภาษาอังกฤษ.....	ข
กิตติกรรมประกาศ.....	ค
สารบัญ.....	ง
สารบัญตาราง.....	จ
สารบัญภาพ	ฉ
บทที่ 1 บทนำ	1
1.1 หลักการและเหตุผล.....	1
1.2 วัตถุประสงค์.....	4
1.3 ขอบเขตของงาน	4
1.4 ขั้นตอนและระยะเวลาการดำเนินงาน.....	6
1.5 ผลที่คาดว่าจะได้รับ	6
1.6 นิยามศัพท์.....	7
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....	8
2.1 ทฤษฎีที่เกี่ยวข้อง.....	8
2.2 งานวิจัยที่เกี่ยวข้อง	15
บทที่ 3 ระเบียบวิธีวิจัย.....	17
3.1 การดำเนินการวิจัย	17
3.2 การกำหนดประชากรและกลุ่มตัวอย่าง	23
3.3 การประเมินผล	23
บทที่ 4 ผลการวิจัย.....	25
4.1 ผลการวิจัย.....	25
4.2 การวิเคราะห์ผลการวิจัย	29

บทที่ 5 สรุปผลการวิจัย.....	32
5.1 สรุปผลการวิจัย.....	32
5.2 ข้อเสนอแนะ.....	33
บรรณานุกรม.....	34
ภาคผนวก ก.....	36
ภาคผนวก ข.....	37
ภาคผนวก ค.....	39
ประวัติผู้วิจัย.....	48

DR

สารบัญตาราง

หน้าที่

ตารางที่ 1.1	ขั้นตอนและระยะเวลาการดำเนินงาน.....	6
ตารางที่ 3.1	ตัวอย่าง Co-occurrence patterns สำหรับสร้างความสัมพันธ์ hasAttraction.....	22
ตารางที่ 3.2	ตัวอย่าง Co-occurrence patterns สำหรับสร้างความสัมพันธ์ hasActivity.....	22
ตารางที่ 4.1	ผลการประเมินประสิทธิภาพของระบบในการกำหนดขอบเขตของนิพจน์ระบุนาม ด้วยเทคนิค CRFs	26
ตารางที่ 4.2	ผลการประเมินประสิทธิภาพของระบบในการจำแนกประเภทสถานที่ท่องเที่ยวด้วย เทคนิค Feature weight k-NN.....	26
ตารางที่ 4.3	ตารางแสดงจำนวนอินสแตนซ์ของสถานที่ท่องเที่ยวที่น่าสนใจ (Attraction).....	27
ตารางที่ 4.4	ผลการประเมินประสิทธิภาพของระบบในการสกัดความสัมพันธ์	28
ตารางที่ 4.5	ตารางแสดงจำนวนอินสแตนซ์ของกิจกรรมในการท่องเที่ยว (Activity).....	29

สารบัญรูปภาพ

	หน้าที่
รูปที่ 1.1 ออนโทโลยีการท่องเที่ยว	3
รูปที่ 2.1 ตัวอย่าง Lightweight Ontology และการประยุกต์ใช้	9
รูปที่ 2.2 ตัวอย่าง UMLS Ontology	10
รูปที่ 2.3 ตัวอย่างออนโทโลยีการท่องเที่ยว.....	11
รูปที่ 3.1 การดำเนินการวิจัย	17
รูปที่ 3.2 ตัวอย่างการตัดคำ และการกำกับหน้าที่ของคำ	18
รูปที่ 3.3 ตัวอย่างการกำกับเอกสารในรูปแบบ IOB format.....	20

บทที่ 1

บทนำ

1.1 หลักการและเหตุผล

อุตสาหกรรมการท่องเที่ยวเป็นอุตสาหกรรมที่ทำรายได้จำนวนมากให้กับประเทศ เนื่องจากประเทศไทยเป็นประเทศที่มีทรัพยากรธรรมชาติที่สวยงาม มีแหล่งท่องเที่ยวทางวัฒนธรรมและประเพณีที่น่าสนใจที่สืบทอดกันมาอย่างยาวนาน ทำให้มีจำนวนนักท่องเที่ยวเพิ่มขึ้นในแต่ละปี และภาครัฐได้มีการรณรงค์ ส่งเสริมและประชาสัมพันธ์การท่องเที่ยวอย่างต่อเนื่องผ่านทางเว็บไซต์ เอกสารประชาสัมพันธ์ สื่อสิ่งพิมพ์ต่างๆ นอกจากนี้ข้อมูลเกี่ยวกับการท่องเที่ยวยังมีให้สืบค้นมากมายทางเว็บไซต์ต่างๆ ซึ่งอาจสร้างโดยหน่วยงานที่เกี่ยวข้องกับการท่องเที่ยว เช่น ส่วนราชการ หรือผู้ประกอบการ เป็นต้น อีกทั้งข้อมูลที่เผยแพร่โดยนักท่องเที่ยวที่ได้ไปเที่ยวตามสถานที่ต่างๆ ได้ทำการรีวิวผ่านสื่อสังคมออนไลน์ก็เป็นข้อมูลที่สำคัญและมีประโยชน์อย่างมากสำหรับนักท่องเที่ยวหน้าใหม่ที่จะสืบค้น และค้นคว้าหาข้อมูล เนื่องจากนักท่องเที่ยวที่จะไปเที่ยวยังสถานที่ต่างๆ จำเป็นจะต้องทราบข้อมูลเกี่ยวกับสถานที่นั้นๆ เช่น สถานที่ท่องเที่ยวที่น่าสนใจ ที่พัก เทศกาล และกิจกรรมที่มีในสถานที่ต่างๆ ซึ่งส่วนใหญ่นักท่องเที่ยวมักจะใช้การค้นหาผ่านเสิร์ชเอนจินต่างๆ แต่เนื่องจากเว็บไซต์มีการเติบโตอย่างรวดเร็ว ทำให้มีข้อมูลที่หลากหลายและมีปริมาณมาก ประกอบกับการสืบค้นข้อมูลของเสิร์ชเอนจินเป็นการสืบค้นแบบเปรียบเทียบคำค้นกับคำที่มีในเอกสารเท่านั้น ดังนั้นอาจจะให้ผลลัพธ์การสืบค้นไม่ครบถ้วนตรงตามความต้องการของผู้ใช้ อย่างไรก็ตามหากต้องการสืบค้นด้วยเทคนิคการสืบค้นขั้นสูง (Advanced Search) ซึ่งเป็นการสร้างคำค้นที่มีความซับซ้อนมากขึ้น เช่น ผู้ใช้ต้องการหาแหล่งท่องเที่ยวในจังหวัดเชียงราย ที่เป็นประเภทแหล่งท่องเที่ยวเชิงเกษตร ที่มีกิจกรรมให้อาหารสัตว์ จะสามารถทำได้หากในการสืบค้นนั้นมีการนำฐานความรู้เกี่ยวกับคำศัพท์หรือออนโทโลยีทางการท่องเที่ยวมาช่วยในการสืบค้นด้วย โดยผลลัพธ์ที่ได้จากการสืบค้น จะแสดงเป็น สิ่งพาร์ค ซึ่งเป็นสถานที่ท่องเที่ยวที่มีคุณสมบัติตรงตามที่ใช้สืบค้น โดยในเอกสารอาจจะไม่มีข้อความว่าให้อาหารสัตว์เลย แต่อาจจะมีข้อความให้อาหารยีราฟ และม้าลาย แต่หากในออนโทโลยีมี

ข้อมูลความสัมพันธ์ระหว่างสถานที่ห้างร้านและกิจกรรมการให้อาหารสัตว์จะช่วยให้สามารถแสดงผลการค้นหาที่ซับซ้อนนี้ได้ และจะทำให้ผู้ใช้ได้ข้อมูลที่ตรงกับความต้องการ มีความสะดวก และรวดเร็วในการค้นหามากยิ่งขึ้น

ออนโทโลยี (Ontology) เป็นคำที่ใช้กันอย่างแพร่หลายในสาขาที่เกี่ยวข้องกับ ปัญญาประดิษฐ์ (AI) และวิศวกรรมความรู้ (Knowledge Engineering) โดย Swartout ได้ให้นิยามว่า “ออนโทโลยี คือ กลุ่มของคำที่มีโครงสร้างแบบลำดับชั้นสำหรับอธิบายขอบเขตเนื้อหาที่สนใจ และสามารถใช้เป็นโครงร่างพื้นฐานสำหรับเป็นฐานความรู้ได้ (An ontology is a hierarchically structured set of terms for describing a domain that can be used as a skeletal foundation for a knowledge base.)” (Swartout et al., 1997) จึงกล่าวได้ว่า

ออนโทโลยีเป็นระบบคำศัพท์ที่มีความสัมพันธ์ในเชิงความหมาย เช่น Generic-Specific, Part-of และใช้เป็นโครงร่างพื้นฐานในการอธิบายความรู้เฉพาะด้าน เช่น การท่องเที่ยว การเกษตร เป็นต้น

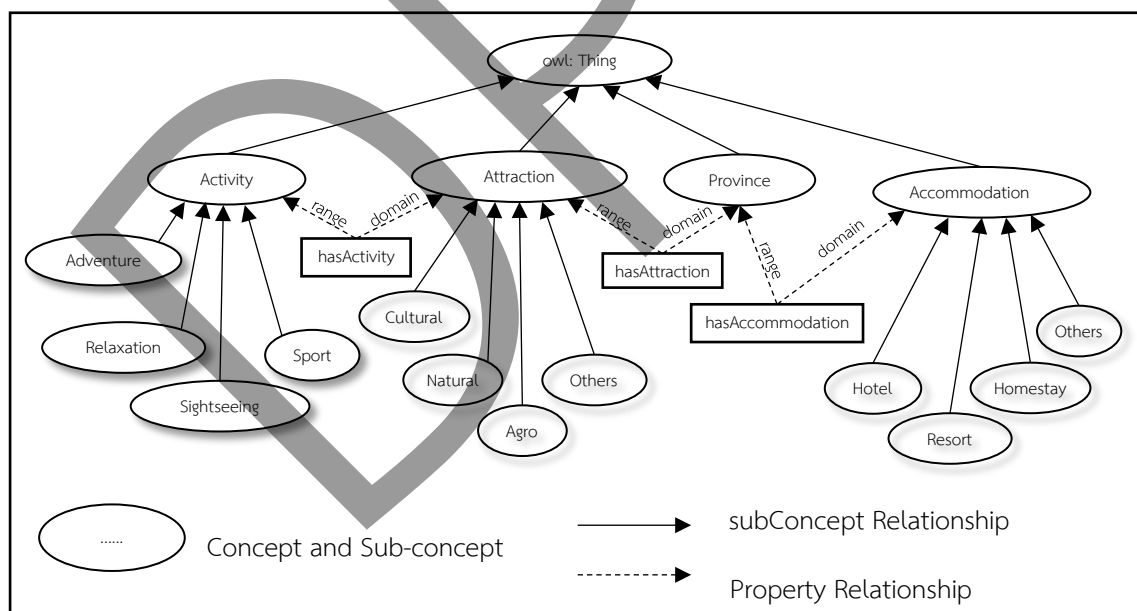
ออนโทโลยีสามารถนำมาประยุกต์ใช้ในการทำงานของระบบงานต่างๆ เช่น ช่วยในการทำงานของระบบสืบค้นข้อมูล โดยช่วยขยายคำค้น ทำให้สามารถค้นหาและเข้าถึงข้อมูลที่ผู้ใช้งานต้องการ ช่วยในระบบให้คำแนะนำ เนื่องจากข้อมูลที่จัดเก็บในออนโทโลยีถือเป็นฐานความรู้ในขอบเขตที่สนใจ ดังนั้นจึงสามารถนำความรู้ที่มีในออนโทโลยีมาช่วยในการทำงานของระบบให้คำแนะนำได้

องค์ประกอบของออนโทโลยี มีดังนี้

- 1) คอนเซ็ปต์หรือแนวคิด (Concepts) คือ ขอบเขตของความรู้ในเรื่องใดเรื่องหนึ่ง และสามารถทำการอธิบายรายละเอียดได้
- 2) คุณสมบัติ (Properties) คือ คุณสมบัติต่างๆ ที่นำมาใช้อธิบายรายละเอียดคอนเซ็ปต์
- 3) ความสัมพันธ์ (Relationships) คือ รูปแบบการแสดงความสัมพันธ์ระหว่างคอนเซ็ปต์
- 4) ข้อกำหนดการสร้างความสัมพันธ์ (Axioms) คือ เงื่อนไขหรือตรรกะในการแปลงความสัมพันธ์ระหว่างคอนเซ็ปต์กับคอนเซ็ปต์ หรือคุณสมบัติเพื่อการแปลงความหมายที่ถูกต้อง
- 5) ตัวอย่างข้อมูล (Instance) คือ คำศัพท์ที่มีการกำหนดความหมายไว้ในออนโทโลยีเรื่องนั้นๆ

การจัดเก็บออนโทโลยีจะมักจะใช้ภาษา RDFS หรือภาษา OWL เป็นต้น การทำออนโทโลยีพ็อพพิวเลชัน (Ontology Population) คือการเพิ่มตัวอย่างข้อมูล หรืออินสแตนซ์ (Instance) ลงในออนโทโลยีเพื่อเป็นการขยายความรู้ที่มีในออนโทโลยีเพื่อประยุกต์ใช้ในงานต่างๆได้หลากหลาย กว้างขวางมากขึ้น

งานวิจัยนี้นำเสนอการทำออนโทโลยีพ็อพพิวเลชันจากเอกสารภาษาไทย โดยเป็นการประยุกต์ใช้เทคนิคการประมวลผลภาษาธรรมชาติ (Natural Language Processing) และการเรียนรู้ด้วยเครื่องจักร (Machine Learning) เพื่อสกัดคำที่เป็นอินสแตนซ์ของคอนเซ็ปต์ และความสัมพันธ์ระหว่างอินสแตนซ์ต่างๆ โดยในงานวิจัยนี้ได้นำออนโทโลยีทางการท่องเที่ยว ได้แก่ OnTour Ontology (Prantner, 2004) ออนโทโลยีการท่องเที่ยว (Kongthong, 2010) และออนโทโลยีชนบท (หัตถยา, 2554) มาพิจารณา และคัดเลือกเฉพาะคอนเซ็ปต์ และความสัมพันธ์ที่เป็นข้อมูลที่น่าสนใจ (Attraction) ที่พัก (Accommodation) และกิจกรรมในการท่องเที่ยว (Activity) มาสร้างเป็นออนโทโลยีสำหรับใช้ในงานวิจัยนี้ โดยสามารถแสดงได้ดังรูปที่ 1.1



รูปที่ 1.1 ออนโทโลยีการท่องเที่ยว

ขั้นตอนในการวิจัยเริ่มจากการรวบรวมเอกสารการท่องเที่ยวจากเว็บไซต์ที่เกี่ยวข้อง หลังจากนั้นนำเอกสารที่ได้มาตัดคำ สกัดนิพจน์ระบุนาม (Name Entity Extraction) ซึ่งทำหน้าที่เป็นอินสแตนซ์ของแต่ละคอนเซ็ปต์ด้วยเทคนิคคอนดิชันนอลแรนดอมฟิลด์ (Conditional Random Field) ซึ่งนำคุณสมบัติของคำ และคำรอบข้างที่มักจะเกิดร่วมกัน มาใช้ในการพิจารณาเพื่อสกัดนิพจน์ระบุนามและจำแนกประเภทตามคอนเซ็ปต์ต่างๆทางการท่องเที่ยวด้วยเทคนิคฟีเจอร์เวท-เค-เนียร์เรสเนเบอร์ (Feature weight k-Nearest Neighbor) หลังจากนั้นจะนำอินสแตนซ์ต่างๆมาพิจารณาความสัมพันธ์ที่เกิดขึ้นระหว่างกัน โดยให้เครื่องเรียนรู้จากคำที่มักจะเกิดร่วมกัน เช่น ประโยค ปางอุ๋ง ตั้งอยู่ที่ ตำบลหมอกจำแป่ อำเภอเมือง จังหวัดแม่ฮ่องสอน ในที่นี้คำว่า ตั้งอยู่ที่ ถือเป็นคำที่บอกความสัมพันธ์ระหว่างคอนเซ็ปต์ สถานที่ท่องเที่ยวที่น่าสนใจ (Attraction) และจังหวัด (Province) ซึ่งสามารถสร้างเป็นแพทเทิร์นที่เรียกว่า Co-occurrence pattern เพื่อใช้ในการบอกความสัมพันธ์ต่างๆของคำที่เกิดขึ้นในเอกสารได้ โดยในเอกสารจะมีคำที่เป็นคำที่บอกความสัมพันธ์อยู่หลากหลาย ในงานวิจัยนี้จะเป็นการศึกษาคำที่สามารถใช้ในการกำหนดความสัมพันธ์ เพื่อสร้างความสัมพันธ์ระหว่างอินสแตนซ์ต่างๆเพื่อเป็นการเพิ่มความรู้อันมีในออนโทโลยีให้มากขึ้น

1.2 วัตถุประสงค์

- 1) เพื่อศึกษาคุณสมบัติของเอกสารภาษาไทยในการทำออนโทโลยีฟ็อฟฟิวเลชัน
- 2) เพื่อศึกษาวิธีการทำออนโทโลยีฟ็อฟฟิวเลชันจากเอกสารในโดเมนการท่องเที่ยว
- 3) เพื่อศึกษาคำที่ใช้กำหนดความสัมพันธ์ระหว่างอินสแตนซ์ของการท่องเที่ยว
- 4) เพื่อประเมินความถูกต้องของการทำออนโทโลยีฟ็อฟฟิวเลชันโดยพิจารณาจาก ความแม่นยำ ค่าระลึก และ อัตราการรู้จำ

1.3 ขอบเขตของงาน

1.3.1 หัวข้อการท่องเที่ยวที่ศึกษา

ข้อมูลการท่องเที่ยวที่นำมาศึกษาและวิเคราะห์จะเกี่ยวข้องกับหัวข้อดังต่อไปนี้

- สถานที่ท่องเที่ยวที่น่าสนใจ (Attraction) แบ่งเป็น

* สถานที่ท่องเที่ยวเชิงวัฒนธรรม (Cultural Tourism) เช่น วัด พระธาตุ อนุสาวรีย์ หอศิลป์ พิพิธภัณฑ์ อุทยานประวัติศาสตร์ อนุสรณ์สถาน

* สถานที่ท่องเที่ยวเชิงเกษตร (Agro Tourism) เช่น โครงการหลวง สวนผลไม้ สถานีเกษตร โครงการหลวง ศูนย์อนุรักษ์พันธุ์สัตว์ สวนพฤกษศาสตร์

* สถานที่ท่องเที่ยวเชิงนิเวศ (Natural Tourism) เช่น ถ้ำ ดอย ป่า หุบ น้ำตก ทะเลสาบ บ่อน้ำพุร้อน หุบดอกไม้ สวนสัตว์

* สถานที่ท่องเที่ยวอื่น ๆ (Others) เช่น ถนนคนเดิน ตลาด ตลาดย้อนยุค สวนน้ำ สวนสนุก ห้างสรรพสินค้า เป็นต้น

- กิจกรรมในการท่องเที่ยว (Activity) แบ่งเป็น

* กิจกรรมผจญภัย (Adventure) เช่น เดินป่า ไต่หน้าผา ไรด์ตัว โหนสลิง ล่องแก่ง บ้านจี้จัมพ์ พายเรือคายัค ขับรถไกลคาร์ท เพนท์บอล เช่น

* กิจกรรมสันทนาการ (Relaxation) เช่น พอกโคลน ล่องเรือ อาบน้ำแร่ ทำบุญตักบาตร เป็นต้น

* กิจกรรมชมธรรมชาติ (Sightseeing) เช่น ดูนก ชมวิว ชมเมือง ชมดอกไม้ เป็นต้น

* กิจกรรมกีฬา (Sport) เช่น ปั่นจักรยาน มวยไทย ดิกอล์ฟ ตกปลา เป็นต้น

1.3.2 ข้อมูลที่นำมาทดลอง

งานวิจัยนี้ทำการทดลองกับเอกสารเกี่ยวกับการท่องเที่ยวในภาคเหนือ จำนวน 5 จังหวัด ได้แก่ เชียงราย เชียงใหม่ น่าน แม่ฮ่องสอน และลำปาง เนื่องจากเป็นจังหวัดที่มีสถานที่ท่องเที่ยวจำนวนมาก ได้รับความนิยมจากนักท่องเที่ยว และเป็นภูมิภาคที่มีสถานที่ท่องเที่ยว (Attraction) และกิจกรรมในการท่องเที่ยว (Activity) ที่น่าสนใจ โดยผู้วิจัยได้ทำการรวบรวมเอกสารภาษาไทยจำนวน 250 หน้าเว็บเพจ จากเว็บไซต์เกี่ยวกับการท่องเที่ยวต่างๆ เช่น เว็บไซต์ <http://thai.tourismthailand.org/> www.เที่ยวเชียงใหม่.com และ www.paduykan.com

1.4 ขั้นตอนและระยะเวลาการดำเนินงาน

ตารางที่ 1.1 ขั้นตอนและระยะเวลาการดำเนินงาน

ขั้นตอนการดำเนินการวิจัย	เดือน											
	1	2	3	4	5	6	7	8	9	10	11	12
1. กำหนดปัญหาการวิจัย	↔											
2. ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง	↔											
3. กำหนดวัตถุประสงค์การวิจัยและประโยชน์ที่จะได้รับ			↔									
4. กำหนดสมมติฐานการวิจัย			↔									
5. รวบรวมข้อมูล และกำกับคำตอบ		↔	↔	↔	↔							
6. พัฒนาโปรแกรมการสกัดคุณลักษณะ และการเรียนรู้		↔	↔	↔	↔	↔	↔	↔	↔			
7. วิเคราะห์ผลลัพธ์และปรับปรุงระบบ								↔	↔	↔	↔	
8. สรุปและอภิปรายผล และจัดทำรายงาน										↔	↔	↔

1.5 ผลที่คาดว่าจะได้รับ

- 1) ได้ฐานความรู้ทางการท่องเที่ยว ซึ่งสามารถค้นคว้าได้โดยผ่าน entities หรือ คำที่เกี่ยวข้อง เช่น สถานที่ท่องเที่ยวประเภทต่าง ๆ และกิจกรรมในการท่องเที่ยว
- 2) เมื่อมีการนำออนโทโลยีไปใช้ในการค้นหาข้อมูลของเว็บไซต์การท่องเที่ยว เช่น การท่องเที่ยวแห่งประเทศไทย จะช่วยให้นักท่องเที่ยวสามารถค้นหาข้อมูลการท่องเที่ยวได้อย่างรวดเร็ว สามารถเลือกสถานที่ท่องเที่ยว และกิจกรรมได้ตรงตามที่ต้องการ และมีประสิทธิภาพมากยิ่งขึ้น
- 3) นักวิจัยที่สนใจสามารถศึกษาแนวทางในการพัฒนาออนโทโลยีทางการท่องเที่ยวจากเอกสารภาษาไทย

1.6 นิยามศัพท์

- ออนโทโลยี (Ontology) คือ กลุ่มของคำที่มีโครงสร้างแบบลำดับชั้นสำหรับอธิบายขอบเขตเนื้อหาที่สนใจ และสามารถใช้เป็นโครงร่างพื้นฐานสำหรับเป็นฐานความรู้ได้
- ออนโทโลยีพ็อพิวเลชัน (Ontology Population) คือ ขั้นตอนในการเพิ่มตัวอย่างข้อมูลหรืออินสแตนซ์ (Instance) ลงในออนโทโลยีเพื่อเป็นการขยายความรู้ที่มีในออนโทโลยีให้มากขึ้น
- อินสแตนซ์ (Instance) คือ คำศัพท์ที่มีการกำหนดความหมาย หรือคำศัพท์ที่มีความสัมพันธ์ is-a กับคอนเซ็ปต์หรือแนวคิด (Concepts) ที่ถูกกำหนดไว้ในออนโทโลยี
- การท่องเที่ยวเชิงวัฒนธรรม (Cultural Tourism) หมายถึง การท่องเที่ยวเพื่อชมงานประเพณีต่างๆ ได้รับความเพลิดเพลินในสุนทรียศิลป์ ได้รับความรู้ ความเข้าใจต่อสภาพสังคมและวัฒนธรรม
- การท่องเที่ยวเชิงเกษตร (Agro Tourism) หมายถึง การเดินทางท่องเที่ยวไปยังพื้นที่เกษตรกรรม สวนเกษตร ฟาร์มปศุสัตว์ เพื่อชื่นชมความสวยงาม ความสำเร็จ และเพลิดเพลินในแหล่งเกษตรกรรม
- การท่องเที่ยวเชิงนิเวศ (Natural Tourism) หมายถึง การท่องเที่ยวอย่างมีความรับผิดชอบในแหล่งธรรมชาติ ที่มีเอกลักษณ์เฉพาะถิ่นที่เกี่ยวกับระบบนิเวศ ภายใต้การจัดการสิ่งแวดล้อมและการท่องเที่ยวอย่างมีส่วนร่วมของท้องถิ่น เพื่อมุ่งเน้นให้เกิดจิตสำนึกต่อการรักษาระบบนิเวศอย่างยั่งยืน
- การท่องเที่ยวแบบอื่น ๆ (Others) เป็นการท่องเที่ยวประเภทอื่น ๆ ที่ไม่ใช่ 3 ประเภทที่กล่าวมาข้างต้น เช่น การท่องเที่ยวเพื่อสุขภาพ การท่องเที่ยวเพื่อนันทนาการ ซึ่งส่วนใหญ่จะเป็นแหล่งท่องเที่ยวที่มนุษย์สร้างขึ้น เพื่อการพักผ่อนและเสริมสร้างสุขภาพ ให้ความสนุกสนาน ความบันเทิง และการศึกษาหาความรู้ เช่น สวนสัตว์ สวนสนุก แหล่งช้อปปิ้ง

บทที่ 2

ทฤษฎี และงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงทฤษฎีที่เกี่ยวข้อง และงานวิจัยที่เกี่ยวข้องในการทำออนโทโลยีเพื่อพหุเลชันจากเอกสาร โดยมีทฤษฎีที่เกี่ยวข้อง ได้แก่ ออนโทโลยี การสกัดนิพจน์ระบุนาม และการเรียนรู้ของเครื่องจักร

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 ออนโทโลยี (Ontology)

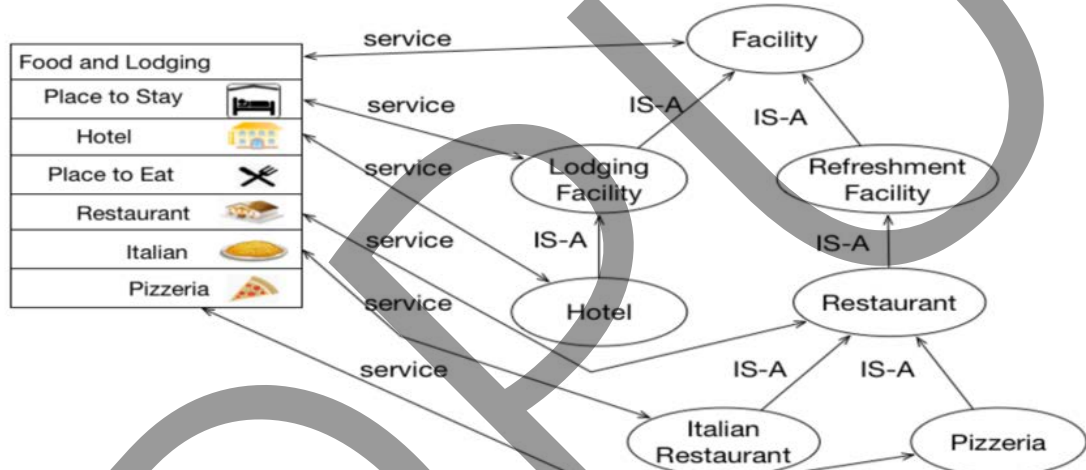
ออนโทโลยี (Ontology) เป็นคำที่ใช้กันอย่างแพร่หลายในสาขาที่เกี่ยวข้องกับปัญญาประดิษฐ์ (Artificial Intelligence: AI) และ วิศวกรรมความรู้ มีผู้ที่ให้คำนิยามของออนโทโลยีไว้หลายแบบ คำนิยามที่มีผู้นิยมใช้กันอย่างแพร่หลายคือคำนิยามของ Gruber [Gruber,1993] ที่นิยามว่า “ออนโทโลยีคือการให้รายละเอียดในเชิงมโนภาพ (An ontology is an explicit specification of a conceptualization.)” แต่อย่างไรก็ตามคำนิยามของ Gruber เป็นคำนิยามที่กว้างเกินไป เพื่อให้เข้าใจได้ง่ายขึ้นจึงขอลำถึงคำนิยามของ Searout [Searout et al., 1996] เพิ่มเติม คือ “ออนโทโลยี คือกลุ่มของคำที่มีโครงสร้างแบบลำดับชั้นสำหรับอธิบายขอบเขตเนื้อหาที่สนใจ และสามารถใช้เป็นโครงร่างพื้นฐานของความสัมพันธ์ของเทอมสำหรับใช้เป็นฐานความรู้ได้ (An ontology is a hierarchically structured set of terms for describing a domain that can be used as a skeletal foundation for a knowledge base.)” จึงกล่าวได้ว่า ออนโทโลยีเป็นระบบคำศัพท์ที่มีความสัมพันธ์ในเชิงความหมาย เช่น Generic-Specific, Part-of และใช้เป็นโครงร่างพื้นฐานในการอธิบายความรู้เฉพาะด้าน เช่น การเกษตร คอมพิวเตอร์ ออนโทโลยีสามารถนำมาประยุกต์ใช้ในการทำงานของระบบงานต่างๆ เช่น ช่วยในการทำงานของระบบสืบค้นข้อมูล โดยช่วยขยายคำค้น ทำให้ค้นหาและเข้าถึงข้อมูลที่ผู้ใช้งานต้องการจริงๆ ออนโทโลียังมีส่วนช่วยในการทำงานของระบบการประมวลผลสารสนเทศให้มีประสิทธิภาพมากขึ้น เช่น การจัดรวมสารสนเทศ

(Information Integration), การจำแนกเอกสาร (Document Classification), การทำความสะอาดข้อมูลในระบบฐานข้อมูล (Data Cleaning in Database System)

ออนโทโลยีแบ่งได้เป็นกลุ่มใหญ่ๆ 2 กลุ่ม ตามองค์ประกอบหลักของออนโทโลยี คือ

1) Lightweight ontology คือ ออนโทโลยีที่ประกอบด้วย concept, concept taxonomies, ความสัมพันธ์ระหว่างคอนเซ็ปต์ และคุณสมบัติของคอนเซ็ปต์ เช่น WordNet¹ จะเก็บข้อมูลคำศัพท์ ชนิดของคำ (Part of Speech) คำอธิบายคำศัพท์ (Glossary) และความสัมพันธ์ระหว่างคำศัพท์ เช่น Antonym, Hyponym, Meronym เป็นต้น

รูปที่ 2.1 แสดงตัวอย่าง Lightweight Ontology เกี่ยวกับโรงแรม ร้านอาหาร และการประยุกต์ใช้ในการสร้างทางเลือกให้กับผู้ใช้เกี่ยวกับที่พัก ร้านอาหาร และรายการอาหาร

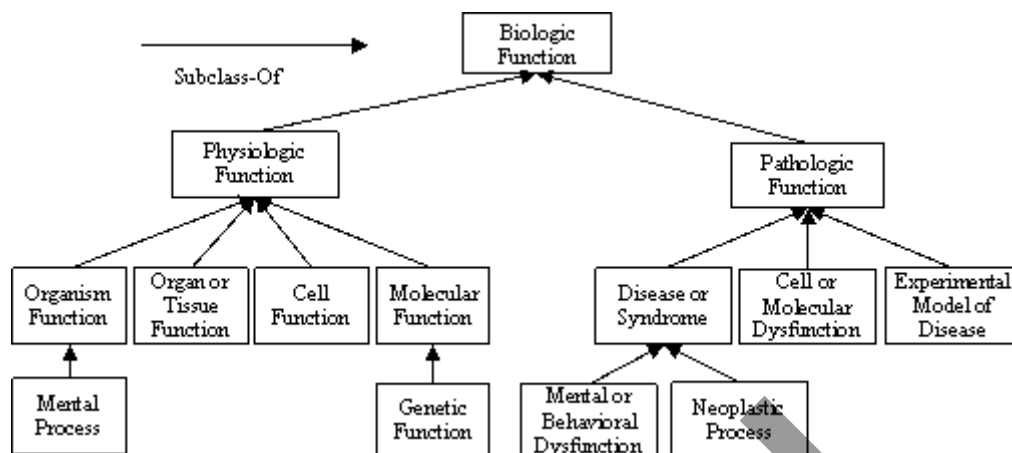


รูปที่ 2.1 ตัวอย่าง Lightweight Ontology และการประยุกต์ใช้ (Ojha, Sajan & Jovanovic, Mladjan & Giunchiglia, Fausto., 2015)

2) Heavyweight ontology คือ ออนโทโลยีที่มีข้อมูลเกี่ยวกับสัจพจน์ (axiom) หรือข้อกำหนดในการสร้างความสัมพันธ์ และข้อกำหนดในการนำค่าต่างๆมาเติมในคลาสและอินสแตนซ์ ซึ่งเพิ่มขึ้นจาก lightweight ontology จะช่วยให้เข้าใจความหมายของคอนเซ็ปต์มากยิ่งขึ้น ตัวอย่างเช่น UMLS² (Unified Medical Language System)

¹ <http://www.cogsci.princeton.edu/cgi-bin/webwn/>

² <http://umlsk.nim.nih.govh>



รูปที่ 2.2 ตัวอย่าง UMLS Ontology³

งานในการพัฒนาออนโทโลยีเป็นงานที่ต้องใช้เวลาและแรงงานของผู้เชี่ยวชาญเป็นอย่างมาก อีกทั้งยังมีการเกิดของคำใหม่เพิ่มขึ้นอยู่ตลอดเวลา จึงเป็นการยากในการที่จะให้ผู้เชี่ยวชาญเป็นผู้เพิ่มเติมข้อมูลอยู่เสมอ ดังนั้นการบำรุงรักษาออนโทโลยีแบบอัตโนมัติเป็นงานที่จำเป็นและสำคัญ

ในโดเมนการท่องเที่ยวมีงานวิจัยในการสร้างออนโทโลยีหลายงาน โดย Prantner et. al. (2007) ได้นำเสนอการทบทวนวรรณกรรมเกี่ยวกับการสร้างออนโทโลยีไว้ดังนี้

Harmonise Ontology ซึ่งพัฒนาในโครงการ EU Project Harmonise ซึ่งมีวัตถุประสงค์ในการสร้างเพื่อแก้ไขปัญหาการทำงานร่วมกันในด้านการท่องเที่ยว (e-tourism) เพื่อการสนับสนุนองค์กรการท่องเที่ยวให้สามารถแลกเปลี่ยนข้อมูลระหว่างกันได้โดยไม่ต้องเปลี่ยนโครงสร้างข้อมูลและระบบข้อมูลท้องถิ่นของตน ออนโทโลยีนี้ถูกจัดเก็บในรูปแบบภาษา RDF โดยมีคอนเซ็ปต์และคุณสมบัติส่วนใหญ่เกี่ยวข้องกับที่พักและเหตุการณ์ต่างๆ

Mondeca Tourism Ontology (www.mondeca.com) เป็นออนโทโลยีทางการท่องเที่ยวที่รวบรวมคอนเซ็ปต์ที่สำคัญทางการท่องเที่ยวซึ่งถูกกำหนดไว้ในอรรถาภิธาน (Thesaurus) ของ WTO (World Tourism Organization) (www.world-tourism.org) โดยในอรรถาภิธานจะมีข้อมูลและคำจำกัดความของหัวข้อการท่องเที่ยวและกิจกรรมสันทนาการ ข้อมูลที่บรรจุในออนโทโลยี Mondeca จะประกอบด้วย ข้อมูลการท่องเที่ยว ข้อมูลทางวัฒนธรรม แพคเกจการท่องเที่ยว และเนื้อหาที่มีเดียการท่องเที่ยว ซึ่งถูกเก็บในรูปแบบภาษา OWL และมีประมาณ 1000 คอนเซ็ปต์

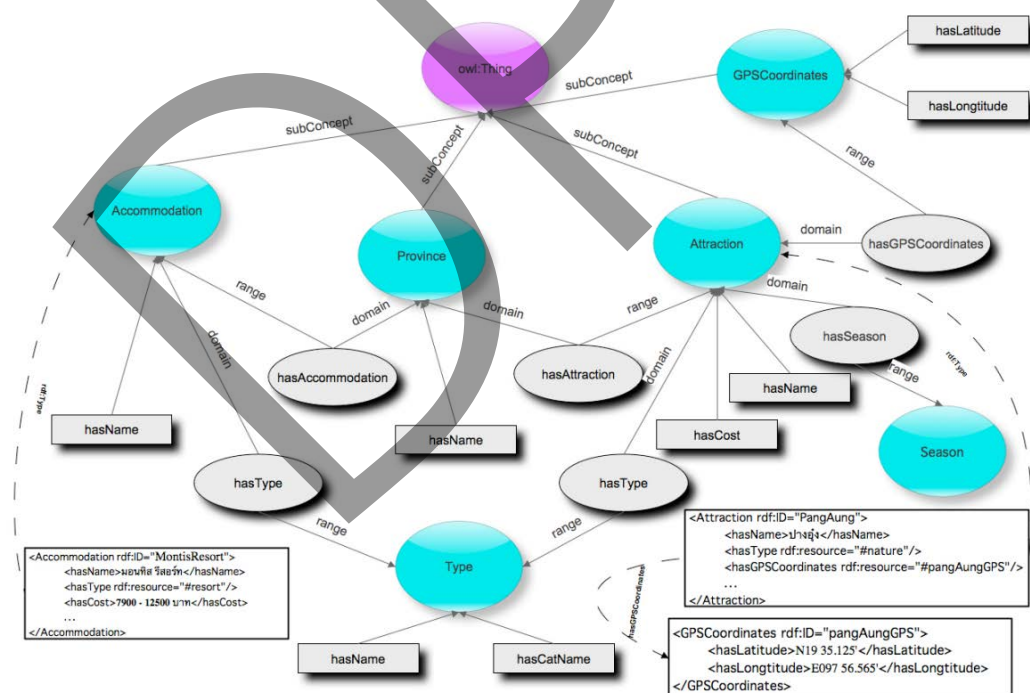
³ <https://www.nlm.nih.gov/research/umls/>

OnTour Ontology (Prantner, 2004) ถูกสร้างขึ้นโดย DERI (Digital Enterprise Research Institute) ประกอบด้วยคอนเซ็ปต์เกี่ยวกับที่ตั้ง (Location) ที่พัก (Accommodation) รวมถึงคอนเซ็ปต์ที่อธิบายถึงกิจกรรมสันทนาการและข้อมูลทางภูมิศาสตร์ ถูกเก็บในรูปแบบภาษา OWL-DL (<http://ontour.deri.org/ontology/ontour-02.owl>) เอกสารเกี่ยวกับออนโทโลยีนี้มีเผยแพร่ที่ <http://e-tourism.deri.at/ont/index.html>

OTA Specification ถูกสร้างโดย OTA (Open Travel Alliance) ซึ่งมีสมาชิกเป็นองค์กรที่เป็นตัวแทนทุกภาคส่วนของอุตสาหกรรมการท่องเที่ยว OTA Specification ถูกกำหนดในรูปแบบ XML ซึ่งเกี่ยวข้องกับเหตุการณ์และกิจกรรมต่าง ๆ ของการท่องเที่ยว

งานวิจัยเกี่ยวกับการสร้างและใช้งานออนโทโลยีทางการท่องเที่ยวภาษาไทย ได้แก่ ออนโทโลยีการท่องเที่ยวของ Kongthon (2010) และออนโทโลยีชนบทของหัตถยา (2554)

รูปที่ 2.3 แสดงตัวอย่างออนโทโลยีการท่องเที่ยวของ Kongthon (2010) โดยวิธีสี่เท่าแทนคอนเซ็ปต์ กรอบสี่เหลี่ยมเล็กที่มีเงาแทนคุณสมบัติ วิธีสี่ขาวแทนความสัมพันธ์ และกรอบสี่เหลี่ยมใหญ่แทนตัวอย่างข้อมูล ออนโทโลยีการท่องเที่ยวนี้ประกอบด้วยคอนเซ็ปต์เกี่ยวกับสถานที่ท่องเที่ยว ที่พัก จังหวัด ประเภท ที่พัก GPS และฤดูกาล



รูปที่ 2.3 ตัวอย่างออนโทโลยีการท่องเที่ยว

ที่มา: Kongthon, 2010

หัตยา (2554) ได้พัฒนาออนโทโลยีการท่องเที่ยวชนบทเพื่อใช้ในการสืบค้นข้อมูล ซึ่งเป็นออนโทโลยีที่สร้างโดยคนหรือผู้วิจัยเอง จากการศึกษาข้อมูลความต้องการของนักท่องเที่ยวจากเว็บทริปแอดไวเซอร์จำนวน 500 บทวิจารณ์ และจากการทบทวนงานวิจัยที่เกี่ยวข้อง และนำไปทดสอบโดยการประยุกต์ใช้กับการสืบค้นข้อมูลให้ตรงกับความต้องการของผู้ใช้

ออนโทโลยีที่ใช้ในงานวิจัยนี้ได้ปรับปรุงและพัฒนาจากคอนเซ็ปต์ที่มีในออนโทโลยีการท่องเที่ยวของ Prantner (2004) Kongthorn (2010) และหัตยา (2554) ดังแสดงในรูปที่ 1.1

2.1.2 การสกัดนิพจน์ระบุนาม (Name Entity Recognition)

นิพจน์ระบุนาม (Named Entity: NE) คือ นิพจน์ที่ทำหน้าที่ระบุชี้เฉพาะถึงสิ่งใดๆ เช่น ชื่อบุคคล ชื่อองค์กร ชื่อสถานที่ รวมไปถึงข้อความแสดงวันเวลา ปริมาณเงินและเปอร์เซ็นต์ เป็นต้น (Grishman and Sundheim, 1995) การสกัดนิพจน์ระบุนามจึงมีความสัมพันธ์กับการทำออนโทโลยีหรือพิวเลชัน เนื่องจากอินสแตนซ์ของแต่ละแนวคิด (คอนเซ็ปต์) ในออนโทโลยีมักจะเป็นชื่อเฉพาะ เช่น ชื่อสถานที่ ที่เป็นอินสแตนซ์ในคอนเซ็ปต์สถานที่ท่องเที่ยว ดังนั้นการสกัดนิพจน์ระบุนามจึงถือเป็นกระบวนการหนึ่งในการทำออนโทโลยีหรือพิวเลชัน

อย่างไรก็ตามการสกัดนิพจน์ระบุนามจะมีความยากเนื่องจากนิพจน์เหล่านี้เป็นคำไม่รู้จักเพราะไม่ปรากฏอยู่ในพจนานุกรม โดยเฉพาะในภาษาไทยจะมีความยุ่งยากยิ่งขึ้น เนื่องจากภาษาไทยมีคุณลักษณะเหล่านี้ (Chanlekha and Kawtrakul, 2004) ได้แก่

1. ภาษาไทยไม่มีข้อสันเทศจากตัวอักษร เช่น อักษรพิมพ์ใหญ่ที่ใช้ในภาษาอังกฤษ หรือชนิดของตัวอักษร เช่น ตัวอักษรคันจิ คาตากานะ ฮิรากานะ ของภาษาญี่ปุ่น ที่สามารถช่วยในการบ่งบอกนิพจน์ระบุนาม จึงทำให้การระบุตำแหน่งของนิพจน์ระบุนามในภาษาไทยยากกว่าภาษาอื่น
2. ภาษาไทยไม่มีการใช้ช่องว่าง หรืออักขระพิเศษเป็นตัวแยกขอบเขตระหว่างคำ ซึ่งทำให้เกิดปัญหาในการหาขอบเขตของนิพจน์ระบุนามในภาษาไทย
3. นิพจน์ระบุนามในภาษาไทยจำนวนมากมีโครงสร้างที่ไม่แน่นอน รวมทั้งนิพจน์ระบุนามในภาษาไทยยังสามารถสร้างขึ้นจากคำใดๆ ก็ได้ จึงทำให้เป็นการยากในการสร้างแบบจำลองหรือกฎที่จะสกัดนิพจน์ระบุนามที่พิจารณาจากโครงสร้างภายในของนิพจน์ระบุนาม

จากการทบทวนวรรณกรรมผู้วิจัยสามารถสรุปได้ว่าแนวทางการสกัดนิพจน์ระบุนามแบ่งได้เป็นกลุ่มหลักๆ 3 กลุ่ม ได้แก่ แนวทางการใช้กฎ หรือฐานความรู้ที่สร้างขึ้นโดยผู้เชี่ยวชาญ แนวทางแบบอัตโนมัติโดยใช้สถิติ หรือการเรียนรู้ด้วยเครื่องจักร และแนวทางแบบผสม

แนวทางการใช้กฎกฎฮิวริสติกในการสกัดนิพจน์ระบุนาม ซึ่งกฎเหล่านี้ถูกสร้างขึ้นโดยใช้พื้นฐานจากรูปแบบการใช้อักษรพิมพ์ใหญ่ เครื่องหมายวรรคตอน ข้อมูลพิเศษเกี่ยวกับชื่อ และลักษณะรูปคำ (Lexical Feature) ที่เกี่ยวข้อง ข้อมูลเหล่านี้ได้แก่ คำนำหน้าชื่อบุคคล คำระบุนัยขององค์กร และชื่อสถานที่ ข้อดีของแนวทางนี้คือมักให้ความถูกต้องสูง แต่มีข้อด้อยคือ ต้องมีการสร้างกฎและข้อมูลพิเศษ ลักษณะรูปแบบคำจำนวนมาก และมักทำงานได้ดีเฉพาะในโดเมนที่พัฒนาเท่านั้น

แนวทางการใช้สถิติ หรือการเรียนรู้ด้วยเครื่องจักร เป็นการฝึกฝนระบบให้เรียนรู้จากคลังเอกสารจริง เทคนิคที่นิยมใช้ เช่น แบบจำลองมาร์คอฟแบบซ่อนเร้น (Hidden Markov Model: HMM) แมกซ์เอนโทรปี (Maximum Entropy) ต้นไม้การตัดสินใจ (Decision Tree) Support Vector Machine Neural Network และ Conditional Random Fields (CRFs) แนวทางนี้มีข้อดีคือ สามารถปรับระบบให้ไปทำงานในโดเมนใหม่ได้สะดวกกว่าแนวทางแบบใช้กฎ แต่มีข้อด้อยคือ ต้องใช้คลังเอกสารจำนวนมากในการฝึกฝนระบบ และในการเรียนรู้ของระบบจะต้องใช้การประมวลผลคอมพิวเตอร์ที่มีความซับซ้อน

แนวทางแบบผสมเป็นวิธีการที่ผสมกันระหว่างการสร้างกฎโดยผู้เชี่ยวชาญ และการใช้หลักการทางสถิติ หรือการเรียนรู้ด้วยเครื่องจักร (Machine Learning) ซึ่งเป็นการนำกฎที่สร้างโดยผู้เชี่ยวชาญเป็นฐานความรู้ในระบบสำหรับช่วยในการเรียนรู้ของเครื่องจักร จึงเป็นการนำข้อดี และลดข้อด้อยของแต่ละเทคนิคลงและทำให้ได้ผลลัพธ์ที่มีความถูกต้องสูง

2.1.3 Conditional Random Fields (CRFs)

CRFs เป็นเทคนิคการเรียนรู้แบบมีผู้สอน (Supervised Learning) โดยจะทำการแปลงข้อมูลนำเข้า (Input Text) ให้เป็น Feature Vector และจะสร้างโหนด (Node) คำตอบที่ควรจะเป็นไปได้ทั้งหมด และหลังจากนั้นจะเลือกโหนด Node ที่มีความน่าจะเป็นมากที่สุดมาเป็นคำตอบ ซึ่งเทคนิค CRFs ช่วยแก้ปัญหาการเลือกคำตอบแบบลำเอียง (Labeled biased) เนื่องจาก CRFs เป็นเทคนิคแบบ Discriminative Models แบบจำลองของ CRFs มีลักษณะเป็นแบบจำลองกราฟไม่ระบุ

ทิศทาง (Undirected Graphical Model) และหาค่าความน่าจะเป็นของ Label ถัดไปโดยนำ Label ก่อนหน้าทั้งหมดที่มีลำดับเหตุการณ์เป็นเงื่อนไขมาคำนวณด้วยน้ำหนักของคุณสมบัติต่างๆ จากสถานะที่ต่างกัน จึงทำให้ไม่เอนเอียงไปสถานะใดสถานะหนึ่ง ถูกออกแบบมาใช้ในการเรียนรู้ข้อมูลที่มีลำดับอย่างเช่นข้อมูลภาษา จึงนิยมนำมาประยุกต์ใช้ในงานด้านการประมวลผลภาษาธรรมชาติ เช่น การกำกับหน้าที่ของคำ (POS Tagging) และการรู้จำนิพจน์ระบุนาม

ในการประยุกต์ใช้งาน CRFs ในงานการรู้จำนิพจน์ระบุนาม ข้อมูลนำเข้าคือลำดับของข้อความในเอกสาร จะเรียกว่า Observation Sequence เขียนแทนด้วย $s = \langle s_1, s_2, \dots, s_T \rangle$ และข้อมูลออกหรือการจำแนกนิพจน์ระบุนามจะเรียกว่า State Sequence เขียนแทนด้วย $o = \langle o_1, o_2, \dots, o_T \rangle$ โดยสามารถแสดงความสัมพันธ์ของทั้งสองตัวแปรได้โดยใช้ฟังก์ชันความน่าจะเป็นแบบมีเงื่อนไข $P(s|o)$ ดังนี้

$$P(s|o) = \frac{1}{Z_o} \exp(\sum_{t=1}^T \sum_{k=1}^K \lambda_k f_k(s_{t-1}, s_t, o, t)) \quad (1)$$

โดย $f_k(s_{t-1}, s_t, o, t)$ เป็นฟังก์ชันคุณลักษณะ (Feature) ต่าง ๆ λ_k เป็นค่าน้ำหนักของแต่ละคุณลักษณะที่ได้จากการเรียนรู้ Z_o เป็นค่า Normalization Factor ได้จากการคำนวณในการเรียนรู้ สามารถคำนวณได้ดังนี้

$$Z_o = \sum_s \exp(\sum_{t=1}^T \sum_{k=1}^K \lambda_k f_k(s_{t-1}, s_t, o, t)) \quad (2)$$

ค่า T คือ จำนวนลำดับของข้อมูลทั้งหมดในที่นี้คือจำนวนคำในเอกสาร ค่า t คือลำดับที่กำลังพิจารณา จะมีค่าตั้งแต่ 1 ถึง T อย่างไรก็ตามเทคนิค CRFs มีข้อจำกัดคือ หากมีจำนวนข้อมูลในการเรียนรู้เป็นจำนวนมาก จะทำให้มีการใช้หน่วยความจำมากขึ้นตามไปด้วย ทำให้ไม่เหมาะสมกับการทำงานกับข้อมูลจำนวนมาก โดยอาจจะมีการแก้ปัญหาด้วยการประยุกต์ใช้เทคนิคในการเลือกคุณลักษณะ หรือ Feature Selection มาช่วยเพื่อลดขนาดของข้อมูล

2.1.4 Feature-weighted k-Nearest Neighbor (kNN)

เทคนิค kNN เป็นเทคนิคการจำแนกกลุ่มแบบมีผู้สอน (Supervised Learning) ที่ทำงานได้รวดเร็ว การคำนวณไม่ซับซ้อน โดยจะจำแนกกลุ่มให้กับข้อมูลโดยพิจารณาจากเพื่อนบ้านที่ใกล้ที่สุด k ตัว หรือจำแนกกลุ่มตามข้อมูลที่มีค่าความคล้ายของคุณลักษณะต่างๆ ที่มากที่สุดดังสมการ (3) โดย c_i หมายถึงกลุ่ม i และ e คือ ข้อมูลที่ต้องการจำแนก หรือ testing data และ $Sim(e, e_j)$ เป็นค่า

ความคล้ายของข้อมูล e และ e_j โดยสามารถคำนวณได้ตามสมการ (4) ซึ่ง e_j คือข้อมูลตัวอย่างที่ได้รับการจำแนกกลุ่มไว้แล้ว ซึ่งจะมีคุณลักษณะ (k) ทั้งหมด n ตัว และ $\delta(e_j, c_i)$ จะมีค่าเป็น 1 ถ้า e_j อยู่ในกลุ่ม i ถ้าไม่ใช่จะมีค่าเป็น 0

$$score(e, c_i) = \sum_{e_j \in KNN(e)} Sim(e, e_j) \delta(e_j, c_i) \quad (3)$$

$$Sim(e, e_j) = \sqrt{\sum_{k=1}^n (e_k - e_{jk})} \quad (4)$$

$$\delta(e_j, c_i) = \begin{cases} 1 & e_j \in c_i \\ 0 & e_j \notin c_i \end{cases} \quad (5)$$

อย่างไรก็ตามในความเป็นจริงตามคุณลักษณะแต่ละตัวจะมีความสำคัญต่อการจำแนกกลุ่มไม่เท่ากัน ดังนั้นในการคำนวณค่าความคล้ายจึงมีการให้น้ำหนักกับคุณลักษณะที่มีผลในการจำแนกกลุ่มมากกว่าคุณลักษณะที่ไม่มีผลต่อการจำแนกกลุ่ม ซึ่งเรียกเทคนิคนี้ว่า Feature-weighted k-Nearest Neighbor (kNN) โดยแสดงได้ดังสมการ (6)

$$Sim(e, e_j) = \sqrt{\sum_{k=1}^n w_k (e_k - e_{jk})} \quad (6)$$

โดย W_k ค่าน้ำหนักของคุณลักษณะ k โดยค่าน้ำหนักที่ใช้ในงานวิจัยนี้ใช้ค่า Correlation coefficient ซึ่งมีค่าระหว่าง -1 ถึง +1 ดังสมการ (7)

$$Correlation\ coefficient = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{(n-1) \sqrt{S_x^2 S_y^2}} \quad (7)$$

โดย X และ Y คือคุณลักษณะที่พิจารณา $\bar{X}$ และ $\bar{Y}$ ค่าเฉลี่ยของแต่ละคุณลักษณะ และ S_x และ S_y ส่วนเบี่ยงเบนมาตรฐานของคุณลักษณะ x และ y ตามลำดับ

อย่างไรก็ตามหากข้อมูลที่นำมาใช้ในการเรียนรู้มีความลำเอียง (bias) หรือมีข้อผิดพลาด (error) จะส่งผลทำให้เกิดความผิดพลาดในการจำแนกข้อมูลได้

2.2 งานวิจัยที่เกี่ยวข้อง

มีงานวิจัยจำนวนมากที่ศึกษาเกี่ยวกับการทำออนโทโลยีพ็อบพิวเลชันซึ่งได้แก่การกำหนดและ จำแนกอินสแตนซ์ตามคอนเซ็ปต์ โดยส่วนใหญ่จะเป็นการประยุกต์เทคนิคการประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP) ร่วมกับเทคนิคการสกัดสารสนเทศ (Information Extraction) และการเรียนรู้ด้วยเครื่องจักร (Machine Learning)

Martinez et al. (2011) นำเสนอการประยุกต์ใช้เทคนิคการประมวลผลภาษาธรรมชาติ ร่วมกับเทคนิคการสกัดสารสนเทศ โดยใช้ GATE Tools เพื่อสกัดนิพจน์ระบุนามจากเอกสารเกี่ยวกับ ร้านอาหารและโรงแรม และใช้อัลกอริทึมฮิวริสติก (Heuristic Algorithm) ในการแก้ปัญหาความกำกวมต่างๆ ในการเพิ่มอินสแตนซ์ในออนโทโลยี

Faria et al. (2012) นำเสนอการประยุกต์ใช้การประมวลผลภาษาธรรมชาติ ร่วมกับเทคนิคการสกัดสารสนเทศ ในการสร้างกฎเพื่อทำออนโทโลยีพ็อบพิวเลชันจากเอกสารแบบอัตโนมัติ โดยได้ทำการทดลองกับเอกสารทางกฎหมาย และการท่องเที่ยว

Zhang et al. (2009) ได้ประยุกต์ใช้เทคนิคการประมวลผลภาษาธรรมชาติ และการเรียนรู้ด้วยเครื่องจักร โดยใช้ Maximum Entropy ในการสกัดความสัมพันธ์ระหว่างเอนทิตีต่างๆ ในโดเมนการท่องเที่ยว

Nanba et al. (2009) ใช้เทคนิคการประมวลผลภาษาธรรมชาติ และใช้คอนดิชันนัลแรนดอมฟิลด์ ในการวิเคราะห์ข้อมูลจากบล็อก (Blog) ทางการท่องเที่ยว เพื่อสกัดสารสนเทศเกี่ยวกับความสัมพันธ์ระหว่างชื่อสถานที่ และสินค้าท้องถิ่นของแต่ละสถานที่

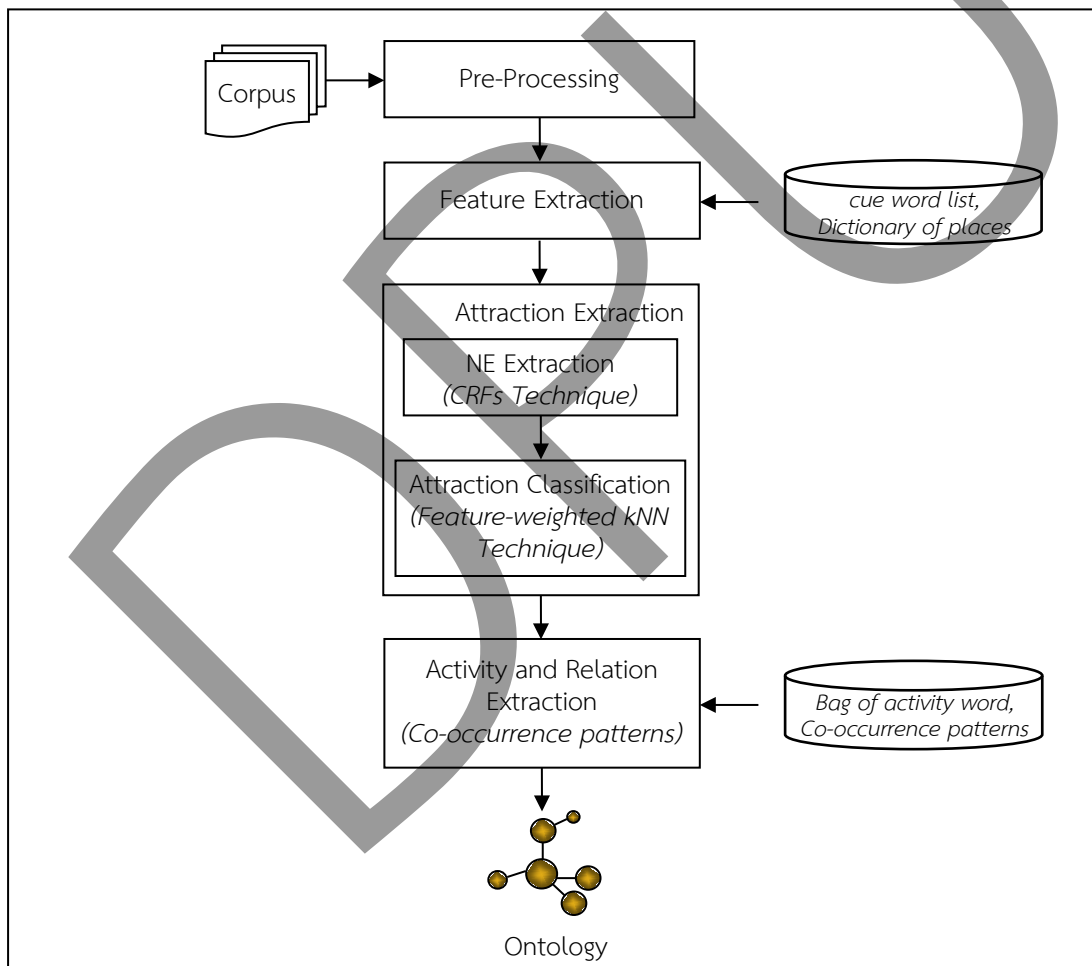
นอกจากนี้ยังมี Carlson et al. (2010), Giuliano and Gliozo (2008), Cimiano et al. (2005) และ Etzioni et al. (2005) ที่ประยุกต์ใช้เทคนิคการประมวลผลภาษาธรรมชาติ ร่วมกับเทคนิคการสกัดสารสนเทศ และการเรียนรู้ด้วยเครื่องจักร ในการทำออนโทโลยีพ็อบพิวเลชัน

บทที่ 3

ระเบียบวิธีวิจัย

ในส่วนนี้เป็นการนำเสนอการดำเนินการวิจัย การกำหนดประชากรและกลุ่มตัวอย่างในการทดลอง และการประเมินผล

3.1 การดำเนินการวิจัย



รูปที่ 3.1 การดำเนินการวิจัย

Corpus หรือคลังข้อความ เป็นเอกสารภาษาไทยจากเว็บไซต์เกี่ยวกับการท่องเที่ยวต่างๆ ในภาคเหนือ จำนวน 5 จังหวัด ได้แก่ เชียงราย เชียงใหม่ น่าน แม่ฮ่องสอน และลำปาง ที่เกี่ยวข้องกับสถานที่ท่องเที่ยวที่น่าสนใจ (Attraction) และกิจกรรมในการท่องเที่ยว (Activity)

3.1.1 ขั้นตอนก่อนการประมวลผล (Pre-Processing)

ขั้นตอนก่อนการประมวลผล เป็นขั้นตอนก่อนการทำงาน เนื่องจากภาษาไทยไม่มีการเว้นวรรคเพื่อแบ่งคำเหมือนในภาษาอังกฤษ ดังนั้นจึงต้องมีการนำเอกสารมาผ่านกระบวนการตัดคำเพื่อหาขอบเขตของคำก่อน เพื่อช่วยในการพิจารณาเกี่ยวกับความหมายและคุณสมบัติของคำต่างๆ โดยในงานวิจัยนี้ได้ใช้โปรแกรมตัดคำเล็กซ์โต (LexTo) ซึ่งเป็นโปรแกรมตัดคำที่ใช้เทคนิคการเลือกคำที่ยาวที่สุด (Longest Matching) โดยอ้างอิงจากพจนานุกรม (LexTo, 2559) หลังจากนั้นเอกสารจะถูกนำไปผ่านกระบวนการกำกับหน้าที่ของคำ หรือ POS Tagging ด้วยเทคนิค Hidden Markov Model (HMM) โดยเครื่องมือที่พัฒนาขึ้นมาเอง และอ้างอิงชุดหน้าที่ของคำจากงาน Sormlertlamvanich, Virach et. al., (1999) ของศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC) สามารถแสดงตัวอย่างได้ดังรูปที่ 3.2

รูปที่ 3.2 ตัวอย่างการตัดคำ และการกำกับหน้าที่ของคำ

ประโยคตัวอย่าง: ดอยแม่สลองเป็นดอยที่มีชื่อเสียงของจังหวัดเชียงราย		
ตัวอย่างการตัดคำ: ดอย แม่ สลอง เป็น ดอย ที่ มี ชื่อ เสียง ของ จังหวัด เชียงราย		
ตัวอย่างการกำกับหน้าที่ของคำ:		
	ดอย/ncn	แม่/ncn
	สลอง/unk	เป็น/vcs
	ดอย/ncn	ที่/prel
	มี/vex	ชื่อ/ncn
	เสียง/ncn	ของ/prep
	จังหวัด/ncn	เชียงราย/npn
หมายเหตุ	ncn	หมายถึง คำนามทั่วไป (Common Noun)
	unk	หมายถึง คำที่ไม่ทราบหน้าที่ของคำ (Unknown)
	vcs	หมายถึง คำกริยาแสดงสภาวะ (Complementary State Verb)
	prel	หมายถึง สรรพนามเชื่อมความ (Relative Pronoun)
	vex	หมายถึง คำกริยา มี (Existential Verb)
	prep	หมายถึง คำบุพบท (Preposition)
	npn	หมายถึง คำนามชี้เฉพาะ (Proper Noun)

3.1.2 ขั้นตอนการสกัดคุณลักษณะ (Feature Extraction)

ขั้นตอนการสกัดคุณลักษณะ เป็นการสกัดคุณลักษณะที่สำคัญเพื่อใช้ในการเรียนรู้สำหรับให้ระบบจำแนกขอบเขต และระบุประเภทของนิพจน์ระบุนาม โดยคุณลักษณะที่ใช้มีดังนี้

Lexical & POS features

- คำ และ POS ของคำที่พิจารณา
- คำและ POS ของคำที่เกิดขึ้นก่อนหน้าจำนวน 3 คำ
- คำและ POS ของคำที่เกิดขึ้นตามหลังจำนวน 3 คำ

Dictionary features

- คำที่พิจารณาเป็นคำที่อยู่ในรายการคำบอกนัย (cue word list) หรือไม่
- คำที่เกิดขึ้นก่อนหน้าเป็นคำที่อยู่ในรายการคำบอกนัย หรือไม่
- เป็นคำที่ไม่มีในพจนานุกรมหรือไม่ (หมายเหตุ ชื่อเฉพาะมักจะไม่มีปรากฏในพจนานุกรมคำทั่วไป)
- เป็นคำที่ปรากฏในพจนานุกรมชื่อจังหวัด อำเภอ และตำบล หรือไม่

Repeated occurrence

- คำที่พิจารณาเกิดขึ้นในเอกสารมากกว่า 3 ครั้งหรือไม่

หมายเหตุ

1. คำบอกนัย (cue word) เป็นคำศัพท์ซึ่งสามารถช่วยให้ทราบถึงการเกิดขึ้นของนิพจน์ระบุนาม เช่น คำนำหน้าชื่อ จะช่วยบอกได้ว่าข้อความที่ตามหลังมาจะเป็นนิพจน์ระบุนามที่เป็นชื่อคน โดยรายการคำที่บอกนัย (cue word list) ของงานวิจัยนี้ได้แสดงไว้ในภาคผนวก ข

2. จำนวนคำที่ใช้ในการทดลองเป็น 3 คำ เนื่องจากให้ผลการทดลองที่มีค่าความแม่นยำสูงที่สุด

3. คำที่พิจารณาเกิดขึ้นในเอกสารมากกว่า 3 ครั้ง ใช้จำนวน 3 ครั้ง เนื่องจากถ้าเป็นคำที่มีความสำคัญกับเอกสารจะมีการกล่าวถึงหลายๆ ครั้ง และจากการทบทวนวรรณกรรม พบว่าในงานวิจัยส่วนใหญ่ใช้จำนวนมากกว่า 3 ครั้งเพื่อแสดงว่าคำนั้นเป็นคำที่มีความสำคัญกับเอกสารนั้นๆ

3.1.3 ขั้นตอนการสกัดชื่อสถานที่ท่องเที่ยว (Attraction Extraction)

3.1.3.1 การสกัดนิพจน์ระบุนามด้วยเทคนิค Conditional Random Fields (CRFs)

ขั้นตอนการสกัดนิพจน์ระบุนาม ซึ่งนิพจน์ระบุนามถือเป็นอินสแตนท์ของคอนเซ็ปต์ต่างๆในออนโทโลยี ในงานวิจัยนี้จะใช้เทคนิค CRFs ในการสกัดนิพจน์ระบุนาม โดยเรียนรู้จากตัวอย่างที่มีการกำกับคลาสคำตอบไว้ โดยได้ประยุกต์ใช้ไลบรารีโปรแกรม CRF++ 0.58 ซึ่งพัฒนาโดย Taku Kudo⁴

โดยตัวอย่างแต่ละตัวอย่าง หรือคำแต่ละคำที่เกิดในเอกสารจะถูกกำกับคลาสคำตอบซึ่งแสดงขอบเขตของนิพจน์ระบุนามไว้ โดยใช้รูปแบบการกำกับคำตอบแบบ IOB format ซึ่งย่อมาจาก Inside-outside-beginning (tagging) ซึ่งถูกนำเสนอโดย Ramshaw and Marcus (1995) เพื่อใช้ในการกำกับเอกสารเพื่อการเรียนรู้ โดยมีความหมายดังนี้

- B: Begin คำเริ่มต้นของนิพจน์ระบุนาม
- I: Inside คำที่อยู่ภายในนิพจน์ระบุนาม
- O: Outside คำอื่น ๆ ที่ไม่ใช่ในนิพจน์ระบุนาม

รูปที่ 3.3 เป็นตัวอย่างการกำกับประโยค “ดอยแม่สลองเป็นดอยที่มีชื่อเสียงของจังหวัดเชียงราย” ในรูปแบบ IOB format

รูปที่ 3.3 ตัวอย่างการกำกับเอกสารในรูปแบบ IOB format

ดอย/B	แม่/I	สลอง/I	เป็น/O	ดอย/O	ที่/O	มี/O	ชื่อ/O	เสียง/O	ของ/O
จังหวัด/B	เชียงราย/I								

3.1.3.2 การจำแนกประเภทสถานที่ท่องเที่ยวด้วยเทคนิค Feature-weighted k-Nearest Neighbor (kNN)

หลังจากนั้นจะนำเข้าสู่กระบวนการเรียนรู้ด้วยเทคนิค Feature-weighted k-Nearest Neighbor (kNN) เพื่อจำแนกประเภทของสถานที่ท่องเที่ยว ซึ่งในงานวิจัยนี้แบ่งเป็น 4 กลุ่ม ได้แก่ สถานที่ท่องเที่ยวเชิงวัฒนธรรม (Culture) สถานที่ท่องเที่ยวเชิงเกษตร (Agriculture) สถานที่ท่องเที่ยวเชิงนิเวศ(Natural) และสถานที่ท่องเที่ยวอื่น ๆ (Others)

⁴ <https://taku910.github.io/crfpp/>

3.1.4 ขั้นตอนการสกัดความสัมพันธ์ (Relation Extraction)

ขั้นตอนการหาความสัมพันธ์ระหว่างกิจกรรมและสถานที่ โดยใช้เทคนิค Co-occurrence patterns ในการระบุความสัมพันธ์ โดย Co-occurrence patterns เป็นรูปแบบที่แสดงความสัมพันธ์ของคำที่มีการเกิดร่วมกันได้ เช่น pattern “<Attraction> ตั้งอยู่ที่ ... <Province>” สามารถแสดงความสัมพันธ์ hasAttraction ระหว่าง จังหวัด (Province) และสถานที่ท่องเที่ยวได้ (Attraction) เช่นประโยค “ปางอุ๋ง ตั้งอยู่ที่ ตำบลหมอกจำแป่ อำเภอเมือง จังหวัดแม่ฮ่องสอน” สามารถสร้างความสัมพันธ์ hasAttraction(แม่ฮ่องสอน, ปางอุ๋ง) และ pattern “กิจกรรมที่สำคัญ ... <Attraction> เช่น <Activity>...” สามารถแสดงความสัมพันธ์ hasActivity ระหว่าง Attraction และ Activity ได้ หรือ hasActivity(Attraction, Activity) ตัวอย่างเช่น ประโยค “กิจกรรมที่สำคัญ บริเวณลำน้ำกก เช่น ล่องแพ ตั้งเต็นท์ และเดินป่า” สามารถสกัดความสัมพันธ์ hasActivity(ลำน้ำกก, ล่องแพ) has Activity(ลำน้ำกก, ตั้งเต็นท์) และ hasActivity(ลำน้ำกก, เดินป่า) ได้ โดยเริ่มต้นจะมีการสร้างถุงคำ (Bag of words) ของกิจกรรมไว้ส่วนหนึ่ง เพื่อใช้เป็นการเริ่มต้นในการเปรียบเทียบการเกิดขึ้นของกิจกรรม ร่วมกับสถานที่ในเอกสารตามรูปแบบที่กำหนด ซึ่งในเอกสารอาจจะมีชื่อกิจกรรมเกิดมากกว่า 1 กิจกรรม ทำให้สามารถสกัดชื่อกิจกรรมได้เพิ่มขึ้น ตารางที่ 3.1 แสดงตัวอย่าง Co-occurrence patterns สำหรับสร้างความสัมพันธ์ hasAttraction ระหว่างอินสแตนซ์ในคอนเซ็ปต์จังหวัด และคอนเซ็ปต์สถานที่ท่องเที่ยว และตารางที่ 3.2 แสดงตัวอย่าง Co-occurrence patterns สำหรับสร้างความสัมพันธ์ hasActivity ระหว่างอินสแตนซ์สถานที่ท่องเที่ยว และกิจกรรม

ตารางที่ 3.1 ตัวอย่าง Co-occurrence patterns สำหรับสร้างความสัมพันธ์ hasAttraction

co-occurrence patterns	ตัวอย่างประโยค
<Attraction> ตั้งอยู่ที่ ... <Province>	ปางอุ๋ง ตั้งอยู่ที่ ตำบลหมอกจำแป่ อำเภอเมือง จังหวัดแม่ฮ่องสอน
<Attraction> อยู่ห่างจาก ... <Province>	น้ำตกสิริภูมิ อยู่ห่างจากตัวอำเภอเมือง เชียงใหม่ ประมาณ ...
การเดินทางไป <Attraction> โดย รถ... จาก ... <Province>	การเดินทางไป บ่อน้ำพุร้อนฝาง โดยรถส่วนตัว จาก ตัวเมือง เชียงใหม่ ไปตามทางหลวงหลวง
ติดต่อสอบถามเพิ่มเติม: ... <Attraction> ... <Province>	ติดต่อสอบถามเพิ่มเติม: วัดเจ็ดยอด ถ. เจ็ดยอด ต. เวียง อ.เมือง จ.เชียงราย
<Attraction> ที่ตั้ง: ... <Province>	อุทยานแห่งชาติสาละวิน ที่ตั้ง : อ. แม่สะเรียง จ. แม่ฮ่องสอน 58110

ตารางที่ 3.2 ตัวอย่าง Co-occurrence patterns สำหรับสร้างความสัมพันธ์ hasActivity

co-occurrence patterns	ตัวอย่างประโยค
นักท่องเที่ยวสามารถ <Activity> ได้ที่ <Attraction>	นักท่องเที่ยวสามารถมาทางเดินที่ได้ที่บริเวณสถานี เกษตรที่สูงขุนช่างเคี่ยน
ที่ <Attraction> นักท่องเที่ยว สามารถ <Activity>	ที่...สมอลล์ ฟาร์มนักท่องเที่ยวสามารถเพลิดเพลิน ไป กับกิจกรรมขี่ม้าสไตร์คาวบอย นั่งเกวียนชมทุ่ง ให้ อาหารสัตว์ ปลูกผัก และกิจกรรมอื่นๆ อีกมากมาย
นอกจาก <Activity> ที่ <Attraction> แล้ว นักท่องเที่ยวยัง สามารถ <Activity> ได้	นอกจากเดินชมและถ่ายภาพความสวยงามของแก รงค์แคนยอน เชียงใหม่แล้ว นักท่องเที่ยวยังสามารถ ลงไปเล่นน้ำ และกระโดดหน้าผาลงน้ำได้
สิ่งที่น่าสนใจในการมาเที่ยว <Attraction> คือ <Activity>	สิ่งที่น่าสนใจในการมาเที่ยวม่อนแจ่ม คือ ได้แคมป์ปิ้ง ชมวิว ชมแปลงพืชผักและผลไม้เมืองหนาว
<Attraction> ... เหมาะที่จะมา <Activity>	ทะเลสาบเชียงแสนมีบรรยากาศร่มรื่น เหมาะที่จะมา พายเรือแคนูเล่น
กิจกรรม <Activity> ที่ <Attraction> ...	กิจกรรมปั่นหน้าผาที่เขื่อนกิ่วลม จังหวัดลำปาง เป็นที่ นิยมมาก

3.2 การกำหนดประชากรและกลุ่มตัวอย่าง

3.2.1 ประชากร

การวิจัยนี้เป็นการประยุกต์ใช้เทคนิคการประมวลผลภาษาธรรมชาติและการเรียนรู้ด้วยเครื่องจักร ประกอบกับการทำออนโทโลยีทางการท่องเที่ยว เพื่อสร้างฐานความรู้ทางการท่องเที่ยวที่ช่วยให้การค้นหาข้อมูลทางการท่องเที่ยวมีประสิทธิภาพมากขึ้น ในที่นี้ผู้วิจัยจะจงเลือกศึกษาเกี่ยวกับการท่องเที่ยวในภาคเหนือ ซึ่งประกอบด้วย จังหวัดเชียงราย เชียงใหม่ น่าน แม่ฮ่องสอน และลำปาง เนื่องจากเป็นจังหวัดที่มีสถานที่ท่องเที่ยวจำนวนมาก และได้รับความนิยมจากนักท่องเที่ยว

ข้อมูลที่ใช้ในการทำออนโทโลยีทางการท่องเที่ยวดังกล่าวมาจากเอกสารการท่องเที่ยวซึ่งเป็นภาษาไทยที่ปรากฏในหน้าเว็บเพจต่างๆที่เกี่ยวข้อง ดังนั้นประชากรซึ่งเป็นเป้าหมายของการสุ่มตัวอย่าง คือ หน้าเว็บเพจภาษาไทยที่เกี่ยวข้องกับการท่องเที่ยวใน 5 จังหวัดที่ระบุข้างต้น

3.2.2 วิธีการเก็บรวบรวมข้อมูล

ผู้วิจัยเลือกเว็บเพจภาษาไทยที่เกี่ยวข้องกับการท่องเที่ยวใน 5 จังหวัดทางภาคเหนือ จำนวน 250 หน้าเว็บเพจจากเว็บไซต์เกี่ยวกับการท่องเที่ยวต่างๆ เช่น เว็บไซต์ <http://thai.tourismthailand.org/> <http://www.เที่ยวเชียงใหม่.com> <http://www.padaykan.com> รายละเอียดของเว็บไซต์ที่นำมาใช้แสดงอยู่ในภาคผนวก ค

3.3 การประเมินผล

ในการประเมินความถูกต้องของการทำงานของระบบจะพิจารณาจากค่าความแม่นยำ ค่าระลึก และ อัตราการรู้จำ ซึ่งมีรายละเอียดดังนี้

ค่าความแม่นยำ (Precision) คือ ค่าที่แสดงให้เห็นว่าระบบที่พัฒนาขึ้นมีความแม่นยำในการหาคำตอบที่ต้องการเพียงใด (Han, 2011) โดยคำนวณได้จาก

$$\text{ความแม่นยำ (Precision)} = \frac{\text{จำนวนคำตอบที่ถูกต้องที่ระบบค้นพบ}}{\text{จำนวนคำตอบทั้งหมดที่ระบบตอบมา}} \times 100 \quad (8)$$

ค่าระลึก (Recall) หรือค่าความครบถ้วน เป็นค่าที่แสดงให้เห็นว่าระบบได้ทำการดึงคำตอบที่ถูกต้องออกมาได้ครอบคลุมจากคำตอบที่มีทั้งหมดมากน้อยเพียงใด (Han, 2011) สามารถคำนวณได้ดังนี้

$$\text{ค่าระลึก (Recall)} = \frac{\text{จำนวนคำตอบที่ถูกต้องที่ระบบค้นพบ}}{\text{จำนวนคำตอบทั้งหมดที่มีในเอกสาร}} \times 100 \quad (9)$$

อัตราการเรียนรู้จำ (F-measure) คือ ค่าเฉลี่ยที่ให้ความสำคัญกับความแม่นยำ และความครบถ้วนเท่าๆ กัน (Han, 2011) โดยสามารถคำนวณได้จากสูตร

$$\text{อัตราการเรียนรู้จำ (F-measure)} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (10)$$

บทที่ 4

ผลการวิจัย

ในบทนี้จะเป็นการนำเสนอผลการวิจัยและการวิเคราะห์ผลการวิจัยในแต่ละขั้นตอนของการทำออนโทโลยีพร้อมพิวเลชัน ได้แก่ การกำหนดขอบเขตของนิพจน์ระบุนามด้วยเทคนิค CRFs การจำแนกประเภทสถานที่ท่องเที่ยวด้วยเทคนิค Feature weight k-NN และการสกัดความสัมพันธ์ของกิจกรรม กับสถานที่ท่องเที่ยว

4.1 ผลการวิจัย

จากการทดลองในการประมวลผลเอกสารจำนวน 250 เอกสาร ซึ่งเป็นเอกสารเผยแพร่เกี่ยวกับสถานที่ท่องเที่ยว และกิจกรรมที่น่าสนใจ จากเว็บไซต์เกี่ยวกับการท่องเที่ยว ซึ่งเอกสารจะถูกนำไปผ่านกระบวนการประมวลผลภาษาธรรมชาติ ด้วยโปรแกรมการตัดคำ และกำกับหน้าที่ของคำ โดยเอกสารที่นำเข้ามาทดลองประกอบด้วยคำจำนวนทั้งสิ้น 121,567 คำ โดยในการทดลอง เอกสารจะถูกแบ่งออกเป็น 2 กลุ่ม คือ กลุ่มสำหรับการเรียนรู้ (train) เป็นจำนวนร้อยละ 70 ของเอกสารทั้งหมด หรือประมาณ 175 เอกสาร หรือจำนวน 85,919 คำ และกลุ่มสำหรับทดสอบ (test) จำนวนร้อยละ 30 ของเอกสารทั้งหมด หรือประมาณ 75 เอกสาร หรือจำนวน 35,648 คำ ผลการประเมินประสิทธิภาพของระบบในการกำหนดขอบเขตของนิพจน์ระบุนามด้วยเทคนิค CRFs สามารถแสดงได้ดังตารางที่ 4.1 โดยเป็นการประเมินผลในการกำหนดขอบเขต โดยจำแนกคำเป็น 3 กลุ่ม คือ B (Begin) หมายถึง คำเริ่มต้นของนิพจน์ระบุนาม I (Inside) หมายถึง คำที่อยู่ภายในนิพจน์ระบุนาม และ O (Outside) หมายถึง คำอื่น ๆ ที่ไม่ใช่ในนิพจน์ระบุนาม

ตารางที่ 4.1 ผลการประเมินประสิทธิภาพของระบบในการกำหนดขอบเขตของนิพจน์ระบุนามด้วยเทคนิค CRFs

Class	Precision	Recall	F-measure
B (begin)	70.45%	76.02%	73.13%
I (Inside)	64.80%	83.67%	73.04%
O (outside)	99.51%	98.95%	99.23%
ค่าเฉลี่ยรวม	78.25%	86.21%	81.80%

ผลลัพธ์ที่ได้จากกระบวนการกำหนดขอบเขตของนิพจน์ระบุนามจะถูกตรวจสอบและแก้ไขให้ถูกต้องโดยผู้เชี่ยวชาญก่อนนำเข้าในกระบวนการถัดไป เพื่อเป็นการวัดผลการทดลองในขั้นตอนถัดไปอย่างแท้จริง ขั้นตอนถัดไปเป็นกระบวนการจำแนกประเภทของสถานที่ท่องเที่ยวด้วยเทคนิค Feature weight k-NN โดยจำแนกประเภทสถานที่ท่องเที่ยวออกเป็น 4 ประเภท คือ สถานที่ท่องเที่ยวเชิงวัฒนธรรม (Cultural Tourism) สถานที่ท่องเที่ยวเชิงเกษตร (Agro Tourism) สถานที่ท่องเที่ยวเชิงนิเวศ (Natural Tourism) และอื่นๆ (Others) สามารถแสดงผลการทดลองได้ดังตารางที่ 4.2

ตารางที่ 4.2 ผลการประเมินประสิทธิภาพของระบบในการจำแนกประเภทสถานที่ท่องเที่ยวด้วยเทคนิค Feature weight k-NN

Attraction	Precision	Recall	F-measure
สถานที่ท่องเที่ยวเชิงวัฒนธรรม (Cultural Tourism)	83.62%	89.87%	86.63%
สถานที่ท่องเที่ยวเชิงนิเวศ (Natural Tourism)	78.42%	75.69%	77.03%
สถานที่ท่องเที่ยวเชิงเกษตร (Agro Tourism)	73.56%	62.14%	67.37%
อื่นๆ (Others)	61.90%	56.52%	59.09%
ค่าเฉลี่ยรวม	74.38%	71.06%	72.53%

อินสแตนซ์ของสถานที่ท่องเที่ยวที่น่าสนใจ ที่ระบบสามารถสกัดได้ สามารถแสดงจำนวนจำแนกตามประเภท ได้ดังตารางที่ 4.3

ตารางที่ 4.3 ตารางแสดงจำนวนอินสแตนซ์ของสถานที่ท่องเที่ยวที่น่าสนใจ

ประเภทคอนเซ็ปต์ของสถานที่ท่องเที่ยวที่น่าสนใจ (Attraction)	จำนวน
สถานที่ท่องเที่ยวเชิงวัฒนธรรม (Cultural Tourism)	333
สถานที่ท่องเที่ยวเชิงนิเวศ (Natural Tourism)	332
สถานที่ท่องเที่ยวเชิงเกษตร (Agro Tourism)	107
อื่นๆ (Others)	85
รวม	857

ตัวอย่างชื่อสถานที่ท่องเที่ยวประเภทต่าง ๆ แสดงได้ ดังนี้

สถานที่ท่องเที่ยวเชิงวัฒนธรรม เช่น คุ่มเจ้าบุรีรัตน์ คุ่มเจ้าราชบุตร คุ่มหลวง บ้านเทอดไทย บ้านเบนนิวาส บ้านเสานัก บ้านเหลียวยังง่วน บ้านไถลื้อ บ้านรวมไทย บ้านรักไทย บ้านสันติชล บ้านสี่เสาหกละ บ้านหนองบัว พระตำหนักเอื้องเงิน พระตำหนักดอยตุง พระตำหนักปางตอง พระตำหนักภูพิงคราชนิเวศน์ พระธาตุจอมกิตติ พระธาตุจอมทอง พระธาตุจอมมอญ พระธาตุดอยเวา พระธาตุดอยกองมู พระธาตุดอยจอมทอง พิพิธภัณฑ์บ้านดำ พิพิธภัณฑ์สถานแห่งชาติเชียงใหม่ พิพิธภัณฑ์สถานแห่งชาติน่าน พิพิธภัณฑ์สถานแห่งชาติลำปาง พิพิธภัณฑ์อุบาคา วัดเชียงภูมิ วัดเชียงมั่น วัดเชียงมั่น วัดเม็งราย วัดเวียงคำ วัดจองกลาง วัดจองคำ วัดจอมแจ้ง วัดผาแดงหลวง วัดพญาภู วัดพญาวัด วัดสันติธรรม วัดหนองบัว วิหารพระเจ้าทองทิพย์ วิหารพระเจ้าทันใจ วิหารพระเจ้าพันองค์ ศูนย์ศิลปชีพแม่ต้า ศูนย์ศิลปชีพแม่ฮ่องสอน หอพิพิธภัณฑ์โบราณล้านนาวัดไหล่หิน หอคำเจ้าเมืองน่าน อนุสาวรีย์ครูบาศรีวิชัย อนุสาวรีย์พ่อขุนเม็งรายมหาราช

สถานที่ท่องเที่ยวเชิงนิเวศ เช่น เขตรักษาพันธุ์สัตว์ป่าเชียงดาว เขตรักษาพันธุ์สัตว์ป่าลุ่มน้ำปาย เขตรักษาพันธุ์สัตว์ป่าอมก๋อย แก่งเสือเต้นสอง แก่งสบห้วยผา แก่งสบห้วยหลวง แก่งสบห้วยอ้อ กัวแม่ปาน กัวลม จุดชมวิวยอดผาป่อง จุดชมวิวม่อนหมาก ดอยเขาน้อย ดอยเชียงเมี่ยง ดอยเชียงดาว ดอยเมี่ยง ดอยเสมอดาว ดอยเหนือ ดอยหลวงเชียงดาว ดอยหลวงเพียงดาว ดอยอินทนนท์ ถ้ำปลา ถ้ำผาสวรรค์ ถ้ำผีแมน ถ้ำพระนอน ถ้ำพระบาท น้ำตกแม่เย็น น้ำตกแม่เฒ่า น้ำตกแม่กลาง น้ำตกแม่ปาน น้ำตกมณฑาธาร น้ำตกม่อนหินไหล น้ำตกวชิรธาร ผาซ้อ ผาซู้ ผาน้ำหยด ผาหมอน ผาหมี ผาหัวสิงห์ ภูกระตัง ภูชี้ฟ้า ภูชี้ป่า ลำน้ำแ่ง ลำน้ำกก ลำน้ำว้า ลำน้ำว้า อุทยานแห่งชาติแจ้ซ้อน อุทยานแห่งชาติแม่เงา อุทยานแห่งชาติแม่เฒ่า อุทยานแห่งชาติแม่จรม อุทยานแห่งชาติแม่ตะไคร้ อุทยานแห่งชาติแม่วาว อุทยานแห่งชาติขุนแจ อุทยานแห่งชาติขุนแจ อุทยานแห่งชาติดอยผ้าห่มปก

สถานที่ท่องเที่ยวเชิงเกษตร เช่น โครงการหลวงแม่โถ โครงการหลวงแม่ปุนหลวง โครงการหลวงแม่ลาน้อย โครงการหลวงขุนแปะ โครงการหลวงชลประทานลำปาง โครงการหลวงตีนตก โครงการหลวงบ้านวัดจันทร์ โครงการหลวงม่อนเงาะ โครงการหลวงวัดจันทร์ โครงการหลวงหนองหอย โครงการหลวงห้วยน้ำลึก โครงการหลวงห้วยลึก โครงการอันเนื่องมาจากพระราชดำริดอยกิ่วลม โรงงานชาบ้านปงตอง ไร่ชาลู่เดช ไร่บุญรอด ไร่รุ่งเรืองจันทร์ ไร่อิงจันทร์ ศูนย์วิจัยเกษตรหลวงเชียงใหม่ ศูนย์วิจัยข้าวสะเมิง ศูนย์อนุรักษ์ช้างไทย สถานีเกษตรแม่จอนหลวง สถานีเกษตรขุนวาง สถานีเกษตรที่สูงขุนช่างเคี่ยน สถานีเกษตรที่สูงดอยม่อนล้าน สถานีเกษตรที่สูงสันป่าเกี๊ยะ สถานีเกษตรหลวงอ่างขาง สวนแม่ฟ้าหลวง สวนกุหลาบอังกฤษ สวนคำดอย สวนป่าทุ่งเกวียน สวนสัมมนา ธร หน่วยจัดการต้นน้ำแม่ตะมาน หน่วยจัดการต้นน้ำแม่สะลม หน่วยจัดการต้นน้ำแม่หยอด หน่วยจัดการต้นน้ำแม่อวม หน่วยจัดการต้นน้ำขุนแม่ยะ หน่วยจัดการต้นน้ำขุนน่าน หน่วยจัดการต้นน้ำดอยผาหลวง หน่วยจัดการต้นน้ำพร้าว

อื่นๆ เช่น ถนนคนเดินเชียงใหม่ ถนนคนเดินเมืองปาย ถนนคนเดินท่าแพ ถนนคนเดินลำปาง ถนนคนเดินวัวลาย ถนนคนม่วนเชียงราย กาดกองต้า กาดคำเที่ยง กาดมีโชค กาดตันพยอม กาดตันลำไย กาดด่อนกอง กาดหลวง เชียงราย กาดหลวง เชียงใหม่ กาดหน้ามอ ตลาดอัสวิน ตลาดแม่สาย ตลาดวโรรส ตลาดสายหยุด ฟาร์มแกะฮักยู ร้านกาแฟ Coffee in love ร้านกาแฟก่อนตะวันลับแนว เหลี่ยมภูผา เชียงราย ไนท์ บาร์ชาร์

ในการสกัดความสัมพันธ์ของกิจกรรม กับสถานที่ท่องเที่ยว (hasActivity) และความสัมพันธ์ระหว่างจังหวัดและสถานที่ท่องเที่ยว (hasAttraction) ด้วย Co-occurrence pattern สามารถแสดงผลการประเมินได้ดังตารางที่ 4.4

ตารางที่ 4.4 ผลการประเมินประสิทธิภาพของระบบในการสกัดความสัมพันธ์

ความสัมพันธ์	Precision	Recall	F-measure
hasAttraction	74.15%	79.70%	76.82%
hasActivity	66.17%	58.44%	62.07%
ค่าเฉลี่ยรวม	70.16%	69.07%	69.45%

กิจกรรมในการท่องเที่ยว (Activity) สามารถจำแนกตามประเภทกิจกรรมด้วยผู้เชี่ยวชาญ ได้ดังตารางที่ 4.5

ตารางที่ 4.5 ตารางแสดงจำนวนอินสแตนซ์ของกิจกรรมในการท่องเที่ยว

ประเภทคอนเซ็ปต์ของกิจกรรมในการท่องเที่ยว (Activity)	จำนวน
กิจกรรมผจญภัย (Adventure)	46
กิจกรรมสันทนาการ (Relaxation)	17
กิจกรรมชมธรรมชาติ (Sightseeing)	10
กิจกรรมกีฬา (Sport)	11
รวม	84

ตัวอย่างรายชื่อกิจกรรมในประเภทต่างๆ มีดังนี้

กิจกรรมผจญภัย เช่น เดินป่า ไรด์ตัว โหนสลิง ขึ้นกระเจตกเทศ ซี่ม้า บ้านจี้จัมพ์ ปีนบอลลูน ปีนเขา ปีนผา ปีนหน้าผา ขับรถเอทีวี ล่องแก่ง ล่องแพ ล่องคยัค ส่องจิ้งจก

กิจกรรมสันทนาการ ได้แก่ อาบน้ำแร่ แช่น้ำร้อน ทานขันโตก ทำบุญตักบาตร พอกโคลน นวดสปา เล่นน้ำตก ซ้อมของ

กิจกรรมชมธรรมชาติ ได้แก่ เดินศึกษาธรรมชาติ ชมวิว ชมเมือง ชมดอกบัวตอง ชมฝูงแกะ ชมการประกวดธิดา เทียบบ้านชาวเขา ดูนก

กิจกรรมกีฬา ได้แก่ ปั่นจักรยาน พายเรือแคนู ตีกอล์ฟ ตกปลา ตกปลาบึกยักษ์ ชกมวย

4.2 การวิเคราะห์ผลการวิจัย

4.2.1 การกำหนดขอบเขตของนิพจน์ระบุนามด้วยเทคนิค CRFs

ความผิดพลาดในการกำหนดขอบเขตของนิพจน์ระบุนามในส่วนการจำแนกคำเริ่มต้น หรือ คลาส B มักจะมีความผิดพลาดในการจำแนกชื่อสถานที่ที่เป็นคำเดียว ไม่มีคำขึ้นต้นที่อยู่ในรายการคำบอกนัย เช่น ปาย ทำให้ระบบจำแนกเป็นคลาส O แทน ในส่วนคลาส I มักจะมีความผิดพลาดในการจำแนกชื่อสถานที่ที่มีขนาดยาว ประกอบด้วยคำจำนวนมาก เช่น วัดพระสิงห์วรวิหาร ซึ่งส่วนใหญ่ในเอกสารมักจะมีชื่อสั้น คือ วัดพระสิงห์ ทำให้ระบบกำหนดขอบเขตของสถานที่แห่งนี้เพียง วัดพระสิงห์ ไม่รวมคำว่าวรวิหาร ทำให้การจำแนกคลาส I มีความผิดพลาดได้

4.2.2 การจำแนกประเภทสถานที่ท่องเที่ยวด้วยเทคนิค Feature weight k-NN

ผลลัพธ์ในการจำแนกสถานที่ท่องเที่ยวด้วยเทคนิค Feature weight k-NN พบว่า สถานที่ท่องเที่ยวเชิงวัฒนธรรม ให้ค่าความแม่นยำ ค่าระลึก และค่า F-measure สูงที่สุด ทั้งนี้เนื่องจากสถานที่ท่องเที่ยวในกลุ่มนี้มักจะเป็นวัด อนุสาวรีย์ และพิพิธภัณฑสถาน ซึ่งจะมีคำนำหน้าชื่อสถานที่ หรือคำบอกนัยที่ค่อนข้างชัดเจน ทำให้ระบบสามารถจำแนกสถานที่ท่องเที่ยวประเภทนี้ได้ค่อนข้างถูกต้อง

สถานที่ท่องเที่ยวเชิงนิเวศให้ค่าความแม่นยำในลำดับที่ 2 เนื่องจาก มีจำนวนตัวอย่างในการเรียนรู้ค่อนข้างเยอะ และคำขึ้นต้นที่ใช้เป็นคำบอกนัยในการจำแนกประเภทค่อนข้างชัดเจน ไม่มีความกำกวมกับสถานที่ท่องเที่ยวประเภทอื่น เช่น อุทยานแห่งชาติ ดอย ม่อน ป่า และน้ำตก เป็นต้น แต่มีคำนำหน้าชื่อสถานที่บางคำที่มีความกำกวม เช่น คำว่า “บ้าน” ตัวอย่างเช่น บ้านป่าบงเปียง ถูกจำแนกให้เป็นสถานที่ท่องเที่ยวเชิงนิเวศเนื่องจากมีนาขั้นบันได นักท่องเที่ยวสามารถไปพักโฮมสเตย์ และเรียนรู้การทำนาได้ ส่วนสถานที่ท่องเที่ยว บ้านไถลื้อ จะถูกจัดให้อยู่ในประเภทสถานที่ท่องเที่ยวเชิงวัฒนธรรม เนื่องจากบ้านไถลื้อเป็นหมู่บ้านที่สืบทอดวัฒนธรรมไถลื้อ ทำให้ระบบมีความกำกวมในการแยกประเภทสถานที่ท่องเที่ยวที่มีคำขึ้นต้นเหล่านี้

สถานที่ท่องเที่ยวเชิงเกษตร มีค่าความถูกต้องในการจำแนกในลำดับปานกลาง เนื่องจากชื่อสถานที่ท่องเที่ยวประเภทนี้หลายแห่ง จะประกอบด้วยคำทั่วไป หรือเป็นการนำชื่อตำบล หรือชื่ออำเภอมาประกอบในชื่อ ทำให้มีความกำกวมในการจำแนกเป็นชื่อทั่วไป เช่น สวนดอกคัตเตอร์ แม่ริม แม่สาออร์คิด เป็นต้น

ค่าความถูกต้องในการจำแนกสถานที่ประเภทอื่นๆ จะมีค่อนข้างต่ำ เนื่องจากตัวอย่างที่ใช้ในการเรียนรู้ของสถานที่ท่องเที่ยวประเภทนี้ค่อนข้างน้อย และชื่อสถานที่ที่มีความหลากหลาย ซึ่งส่งผลให้เกิดความกำกวมได้ เช่น ชื่อร้านค้า นอกจากนั้นคำขึ้นต้นบางคำมีความกำกวมกับสถานที่ท่องเที่ยวเชิงเกษตร เช่น คำว่า “ฟาร์ม” ตัวอย่างเช่น ฟาร์มงูแม่สา จัดเป็นสถานที่ท่องเที่ยวเชิงเกษตร เนื่องจากเป็นสถานที่รวบรวมพันธุ์งู ศึกษาด้านการขยายพันธุ์งู และมีการแสดงของงู ส่วนฟาร์มแกะฮักยู จัดเป็นสถานที่ท่องเที่ยวประเภทอื่นๆ เนื่องจากเป็นสถานที่สำหรับถ่ายภาพให้อาหารแกะ และขายผลิตภัณฑ์จากนมแกะ

4.2.3 การสกัดความสัมพันธ์ด้วย Co-occurrence pattern

ความสัมพันธ์ hasAttraction มีรูปคำในการเกิดความสัมพันธ์ที่ค่อนข้างชัดเจน เช่น คำว่า “ตั้งอยู่ที่” นอกจากนี้รายชื่อจังหวัด ซึ่งเป็นอินสแตนซ์ในคอนเซ็ปต์จังหวัดมีรายชื่อที่ครบถ้วน ชัดเจน ดังนั้นการสกัดความสัมพันธ์นี้ด้วย Co-occurrence pattern จึงสามารถสกัดความสัมพันธ์ได้ค่อนข้างถูกต้อง และครอบคลุม

ความสัมพันธ์ hasActivity มีรูปคำในการเกิดความสัมพันธ์ระหว่างสถานที่ท่องเที่ยว และกิจกรรมที่ค่อนข้างหลากหลาย และชื่อกิจกรรมส่วนใหญ่เป็นกริยาวิเศษณ์ ซึ่งในงานวิจัยนี้ไม่ได้มีการกำหนดวลี ทำให้การสกัดความสัมพันธ์ และการสกัดชื่อกิจกรรมเพิ่มเติมทำได้ค่อนข้างยาก ตัวอย่างเช่น ประโยค “ล่องแก่งสุตะระทิกพิชิตแม่น้ำปาย” “บริการทัวร์นั่งช้าง สอบถามได้ที่ ทัวร์เมิงไต 053-611-979 อ.ปาย” “ตกเย็น ห้ามพลาดการปั่นจักรยานเที่ยวรอบ ๆ บริเวณโครงการหลวงบ้าน วัดจันทร์” จะเห็นว่ามีความสัมพันธ์ที่ค่อนข้างหลากหลาย ทำให้การระลึก (recall) ของระบบในการสกัดความสัมพันธ์นี้ค่อนข้างต่ำ นอกจากนี้หากต้องการใช้เทคนิค Co-occurrence pattern ในโดเมนอื่น จะต้องมีการศึกษาค้นเอกสาร เพื่อสร้างรูปแบบการเกิดร่วมกัน ให้สอดคล้องกับการเกิดความสัมพันธ์ในรูปแบบต่าง ๆ ต่อไปอีก

บทที่ 5

สรุปผลการวิจัย

5.1 สรุปผลการวิจัย

งานวิจัยนี้นำเสนอการทำออนโทโลยีเพื่อพหุวัจนจากเอกสารภาษาไทยในโดเมนการท่องเที่ยว ซึ่งเป็นการขยายความรู้ในออนโทโลยีให้ครอบคลุมในโดเมนที่สนใจ เพื่อประโยชน์ในการนำออนโทโลยีไปใช้ในแอปพลิเคชันต่างๆ เช่น ระบบค้นคืนสารสนเทศ ระบบให้คำแนะนำ และระบบถามตอบ เป็นต้น โดยประยุกต์ใช้เทคนิคการประมวลผลภาษาธรรมชาติ และการเรียนรู้ด้วยเครื่องจักร เพื่อสกัดคำที่เป็นอินสแตนซ์ของคอนเซ็ปต์ และความสัมพันธ์ระหว่างอินสแตนซ์ต่างๆ

ขั้นตอนในการวิจัยเริ่มจากการรวบรวมเอกสารการท่องเที่ยวจากเว็บไซต์ที่เกี่ยวข้อง และนำเอกสารผ่านการประมวลผลภาษาธรรมชาติด้วยการตัดคำ กำกับหน้าที่ของคำ ถัดไปจะเป็นขั้นตอนการสกัดอินสแตนซ์ของคอนเซ็ปต์สถานที่ท่องเที่ยว ซึ่งอินสแตนซ์ส่วนใหญ่จะเป็นชื่อเฉพาะ หรือเรียกว่านิพจน์ระบุนาม ดังนั้นจึงมีการประยุกต์ใช้เทคนิคการสกัดนิพจน์ระบุนามมาใช้ในการสกัดอินสแตนซ์ โดยจะมีการสกัดคุณลักษณะ ที่สำคัญเพื่อใช้ในการเรียนรู้เพื่อจำแนกอินสแตนซ์ โดยคุณลักษณะที่ใช้จะเป็นคุณสมบัติของคำ และคำรอบข้างที่มักจะเกิดร่วมกัน ได้แก่ คุณลักษณะคำและหน้าที่ของคำ คุณลักษณะการเกิดในพจนานุกรม และคุณลักษณะการเกิดซ้ำของคำในเอกสาร และนำเอกสารที่มีการกำกับคำตอบการระบุขอบเขตของนิพจน์ระบุนามมาผ่านการเรียนรู้ด้วยเทคนิคคอนดิชันนอลแรนดอมฟิลด์ หลังจากนั้นจึงนำนิพจน์ระบุนามที่ได้มาจำแนกประเภทสถานที่ท่องเที่ยวด้วยเทคนิค พีเจอาร์เอจ เค-เนียร์เรสเนเบอร์ นั้นจะนำอินสแตนซ์ต่างๆมาพิจารณาความสัมพันธ์ที่เกิดขึ้นระหว่างกันโดยใช้รูปแบบการเกิดร่วมกัน เพื่อสร้างความสัมพันธ์ระหว่างคอนเซ็ปต์จังหวัดและสถานที่ท่องเที่ยว และความสัมพันธ์ระหว่างสถานที่ท่องเที่ยวกับกิจกรรม

จากการทดลองกับเอกสารภาษาไทยในโดเมนการท่องเที่ยวจำนวน 121,567 คำ ผลการทดลองพบว่าสามารถสกัดชื่อสถานที่ท่องเที่ยวได้จำนวน 857 แห่ง สามารถสกัดกิจกรรมการท่องเที่ยวได้จำนวน 84 กิจกรรม และสกัดความสัมพันธ์ได้จำนวน 941 ความสัมพันธ์ โดยขั้นตอนการสกัดชื่อสถานที่ท่องเที่ยวมีค่าความแม่นยำ 74.38% และค่าระลอก 71.06% และขั้นตอนการสกัดความสัมพันธ์มีค่าความแม่นยำ 70.16% และค่าระลอก 69.07%

5.2 ข้อเสนอแนะ

งานวิจัยนี้เป็นการศึกษาสถานที่ท่องเที่ยวในกลุ่มจังหวัดภาคเหนือ ซึ่งมีสถานที่ในกลุ่มการท่องเที่ยวเชิงนิเวศ ส่วนใหญ่เป็น ดอย อุทยานแห่งชาติ โครงการเกษตรที่สูง ดังนั้นการประยุกต์ใช้เทคนิคนี้กับสถานที่ท่องเที่ยวในภาคอื่น ๆ จะต้องมีการกำกับเอกสารสถานที่ท่องเที่ยวที่มีลักษณะเฉพาะในแต่ละพื้นที่เพิ่มเติม และกำหนดคำรายการคำบอกนัย ซึ่งเป็นคำขึ้นต้นสถานที่ท่องเที่ยวต่างๆ เพิ่มเติม เช่น ในภาคใต้ จะมีสถานที่ท่องเที่ยวทางทะเล เกาะ ทางภาคตะวันออกเฉียงเหนือจะเรียกชื่อภูเขา ขึ้นต้นด้วยคำว่า ภู แทนที่จะเป็นคำว่า ดอย เหมือนทางภาคเหนือ เป็นต้น แต่ในส่วนสถานที่ท่องเที่ยวประเภทอื่น ๆ เช่น สถานที่ท่องเที่ยวเชิงวัฒนธรรม สถานที่ท่องเที่ยวเชิงเกษตร จะสามารถประยุกต์ใช้เครื่องมือ และเทคนิคของงานวิจัยนี้ได้เลย เนื่องจากคำขึ้นต้นของสถานที่เหล่านี้ในภูมิภาคต่างๆ มักจะเรียกไม่ต่างกัน เช่น ขึ้นต้นด้วยคำว่า วัด พิชัยภัณฑ์ และตลาด เป็นต้น

นอกจากนี้ผู้วิจัยวางแผนการทำวิจัยด้วยการประยุกต์ใช้เทคนิคการเรียนรู้ของเครื่องจักรประเภทอื่นๆ เพื่อเพิ่มประสิทธิภาพการทำงานให้กับระบบ เช่น เทคนิค Deep Learning

บรรณานุกรม

- A. Kongthon, et. al. “Thailand’s Tourism Information Service Based on Semantic Search and Opinion Mining”, In Proceeding of International Technical Conference on Circuits/Systems, Computers and Communications, 2010.
- Borthwick, A., Sterling, J., Agichtein, E. and Grishman, R. NYU: Description of the MENE Named Entity System as Used in MUC-7. In Proc. of 7th Message Understanding Conference (MUC-7), Fairfax, Virginia, 1998.
- H. Chanlekha and A. Kawtrakul, Thai Named Entity Extraction by incorporating Maximum Entropy Model with Sample Heuristic Information, International Joint Conference of Natural Language Processing (IJCNLP-2004), 2004.
- J. Han, Micheline Kamber and Jian Pei, “Data Mining: Concepts and Techniques”, 3rd Edition, Morgan Kaufmann, 2011.
- J. Lafferty, A. McCallum, and F. Pereira, Conditional Random Fields: Probabilistic Models for Sequence Data, Proceedings of the Eighteenth International Conference on Machine Learning, 2001.
- K. Prantner, OnTour – The Ontology, DERI Innsbruck, 2004.
- N. Chinchor, OVERVIEW OF MUC-7/MET-2 [Online]. Retrieved from http://www.itl.nist.gov/iaui/894.05/related_projects/muc/proceedings/muc_7_proceedings/overview.html, Retrieved date 15 November 2017.
- N. Tirasaroj and W. Aroonmanakun, The Effect of Answer Patterns for Supervised Named Entity Recognition in Thai, in Proceeding of the 25th Pacific Asia Conference on Language, Information and Computation, 2011.
- Ojha, Sajan & Jovanovic, Mladjan & Giunchiglia, Fausto. Entity-Centric Visualization of Open Data, 2015. p.149-166.
- Prantner K, Ding Y, Luger M, Yan Z. Tourism ontology and semantic management system: state-of-the-arts analysis. In: Proceedings of IADIS (International Association for Development of the Information Society) international conference WWW/Internet 2007, 2007. p. 111–15.
- R. Vijaykrishan and L. Sobha, Domain Focused Named Entity Recognizer for Tamil Using Conditional Random Fields, in Proceedings of the Third International

- Joint Conference on Natural Language Processing (IJCNLP-08) Workshop on Name Entity Recognition for South and South East Asian Language, 2008.
- Ramshaw, L., and Marcus, M. P. (1995). Text Chunking Using Transformation-Based Learning. In Proceedings of the Third ACL Workshop on Very Large Corpora, Association for Computational Linguistics, 1995.
- Sajan Raj Ojha, Mladjan Jovanovic, and Fausto Giunchiglia, Entity-Centric Visualization of Open Data, in Proceedings of Human-Computer Interaction – INTERACT 2015: 15th IFIP TC 13 International Conference, Bamberg, Germany, 2015.
- Sornlertlamvanich, Virach & Takahashi, N & Isahara, H., Thai Part-of-speech Tagged Corpus: ORCHID, Journal of the Acoustical Society of Japan Volume 20 Issue 3, 1999, pp. 189-198.
- W. Swartout, R. Patil, K. Knight, and T. Russ. “Toward distributed use of large-scale ontologies.” In Spring Symposium Series on Ontological Engineering, pages 33–40, Stanford, 1997. AAAI Press.
- วิกิพีเดีย สารานุกรมเสรี, การท่องเที่ยว [Online]. สืบค้นจาก <https://th.wikipedia.org/wiki/การท่องเที่ยว>, วันที่สืบค้น 15 พฤศจิกายน 2559.
- ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ, โปรแกรมตัดคำสำหรับข้อความภาษาไทย LexTo (Thai Lexeme Tokenizer) [Online]. สืบค้นจาก <http://www.sansarn.com/lexto/>, วันที่สืบค้น 15 พฤศจิกายน 2559.
- หัตถยา คชรัตน์, การพัฒนาออนไลน์การท่องเที่ยวชนบท, วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีสุรนารี 2554.

ภาคผนวก ก

ส่วนหนึ่งของงานวิจัยนี้ได้รับการตีพิมพ์เผยแพร่ ดังนี้

- Imsoambut, A. and Kajornrit, J., A Comparison of Statistical and Data Mining Techniques for Enrichment Ontology with Instances, Journal of Reviews on Global Economics: Special Issue - Recent Topical Research on Global, Energy, Health & Medical, and Tourism Economics, and Global Software, p. 375-379, 2017.
- Imsoambut, A. and Paireekreng, W., “Extract Knowledge for Populating Thai Tourism Ontology from Texts Using Feature-weighted k-Nearest Neighbor”, In Proceeding of the 1st International Conference on Information Technology (InCIT2016), Thailand, p. 142-148, 2016.
- Imsoambut, A. and Sirikayon, C., “An Alternative Technique for Populating Thai Tourism Ontology from Texts Based on Machine Learning”, In Proceeding of the 15th IEEE/ACIS International Conference on Computer and Information Science (ICIS 2016), Okayama, Japan, p. 227-230, 2016.

ภาคผนวก ข

รายการคำบอกนัย (Cue Word List)

สถานที่ท่องเที่ยวเชิงวัฒนธรรม (Cultural Tourism)

วัด	กู๋	ศูนย์วัฒนธรรม
อนุสาวรีย์	ศูนย์หัตถกรรม	หอวัฒนธรรม
หอศิลป์	หอ	แหล่งโบราณคดี
พระบรมราชานุสาวรีย์	พระสถูปเจดีย์	พระธาตุ
อนุสรณ์สถาน	พิพิธภัณฑ	อุทยานประวัติศาสตร์
สำนักสงฆ์	พระตำหนัก	บ้าน
หมู่บ้าน	คุ้ม	เวียง
ย่าน	ชุมชน	

สถานที่ท่องเที่ยวเชิงนิเวศ (Natural Tourism)

ดอย	โป่ง	โป่งน้ำร้อน
น้ำพุร้อน	บ่อน้ำร้อน	บ่อน้ำพุร้อน
ป่า	ภู	ภูเขา
บึง	ผา	ป่า
หนอง	แก่ง	ถ้ำ
หล่ม	ทุ่ง	ไร่
ทะเลสาบ	ทะเลหมอก	จุดชมวิว
น้ำตก	อ่างเก็บน้ำ	เขื่อน
เส้นทางศึกษาธรรมชาติ	อุทยานแห่งชาติ	ม่อน
ก๊ว	หาด	ห้วย

สถานที่ท่องเที่ยวเชิงเกษตร (Agro Tourism)

สวน	ไร่	สถานีเกษตร
-----	-----	------------

สถานีประมง

โครงการหลวง

ศูนย์พัฒนาโครงการ

สถานีทดลองเกษตร

ศูนย์อนุรักษ์พันธุ์

เขตรักษาพันธุ์

โครงการ

สวนพฤกษศาสตร์

ศูนย์ศึกษาพัฒนา

สถานที่ท่องเที่ยวอื่น ๆ (Others)

ร้าน

ตลาด

ถนนคนเดิน

กาด

ภาคผนวก ค

รายชื่อหน้าเว็บไซต์ที่ถูกลำข้อมูลมาใช้ในการทำวิจัย

<http://www.reviewchiangmai.com/2435>
<http://www.paiduaykan.com/province/north/chiangmai/grandcanyon.html>
http://www.paiduaykan.com/76_province/north/chiangmai/maechaem.html
http://www.paiduaykan.com/76_province/north/chiangmai/maekalngluang.html
<http://www.paiduaykan.com/province/north/chiangmai/watchan.html>
<http://www.paiduaykan.com/province/north/chiangmai/giewmaepan.html>
http://www.paiduaykan.com/76_province/north/chiangmai/maeya.html
<http://www.teawteenai.com/ขุนช่างเคี่ยน>
http://www.paiduaykan.com/76_province/north/chiangmai/khunchangkain.html
<http://www.teawteenai.com/สถานที่ท่องเที่ยว/คุ้มเจ้าบุรีรัตน์-มหาอินทร์--3905>
<http://www.paiduaykan.com/province/north/chiangmai/monngo.html>
<http://www.lampang.go.th/travel/data/placetour/muangpan.htm>
<http://2017.sadoodta.com/info/ชุมชนวัดเกตุ>
http://www.paiduaykan.com/76_province/north/chiangmai/maetaman.html
<http://www.paiduaykan.com/province/north/chiangmai/doiphahompok.html>
<http://www.paiduaykan.com/province/north/chiangmai/doimonjong.html>
<http://www.paiduaykan.com/province/north/chiangmai/monlan.html>
<http://www.paiduaykan.com/province/north/chiangmai/doichiangdao.html>
http://www.paiduaykan.com/76_province/north/chiangmai/walkingstreet.html
<http://www.paiduaykan.com/province/north/chiangmai/thamchiangdao.html>
<http://www.paiduaykan.com/province/north/chiangmai/samoengriceresearch.html>
http://www.paiduaykan.com/76_province/north/chiangmai/maechaem.html
<http://www.paiduaykan.com/province/north/chiangmai/maeklangwaterfall.html>
<http://www.paiduaykan.com/province/north/chiangmai/maeyawaterfall.html>
<http://www.paiduaykan.com/province/north/chiangmai/phadokesiew.html>
<http://www.paiduaykan.com/province/north/chiangmai/siripoomwaterfall.html>
<http://www.paiduaykan.com/province/north/chiangmai/fanghotspring.html>
<http://www.mae-kampong.com/index.php?lay=show&ac=article&id=399648>

<http://www.paiduaykan.com/province/north/chiangmai/papongpieng.html>
<http://www.paiduaykan.com/province/north/chiangmai/banphamorn.html>
<http://www.paiduaykan.com/province/north/chiangmai/huayhom.html>
<http://www.paiduaykan.com/province/north/chiangmai/phachor.html>
<http://www.paiduaykan.com/province/north/chiangmai/darapirom.html>
<http://www.bhubingpalace.org/index2.php?ge=contact>
<http://www.paiduaykan.com/province/north/chiangmai/monjam.html>
<http://www.paiduaykan.com/province/north/chiangmai/watchediluang.html>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/วัดเจดีย์หลวงวรวิหาร--126>
<http://www.paiduaykan.com/province/north/chiangmai/watchiangman.html>
<http://www.dhammathai.org/watthai/north/watphrathatdoisuthep.php>
<http://www.paiduaykan.com/province/north/chiangmai/watprathatsrijomthong.html>
<http://www.paiduaykan.com/province/north/chiangmai/watphrasing.html>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/วัดพระสิงห์วรวิหาร--124>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/วัดสวนดอก--122>
<http://www.paiduaykan.com/province/north/chiangmai/maetho.html>
<http://www.paiduaykan.com/province/north/chiangmai/khunpae.html>
<http://www.paiduaykan.com/province/north/chiangmai/teentok.html>
<http://www.paiduaykan.com/province/north/chiangmai/monngoroyalproject.html>
<http://www.paiduaykan.com/province/north/chiangmai/huailuk.html>
<http://www.paiduaykan.com/province/north/chiangmai/majonlounge.html>
<http://www.paiduaykan.com/province/north/chiangmai/orchidagriculture.html>
http://www.paiduaykan.com/76_province/north/chiangmai/angkhang.html
<http://www.amazingthaitour.com/ภาคเหนือ/เที่ยวสมอลล์-ฟาร์ม-เชียงใหม่-small-farm-chiang-mai/>
<http://www.paiduaykan.com/province/north/chiangmai/queensrikitgarden.html>
<http://www.paiduaykan.com/province/north/chiangmai/suansomthanathon.html>
http://www.paiduaykan.com/76_province/north/chiangmai/chiangmaizoo.html
http://www.paiduaykan.com/76_province/north/maehongson/travelmaehongson.html
<http://www.emagtravel.com/archive/singha-park.html>

<http://www.chiangraifocus.com/travel/review/55/แซะ-ชม-ชิม-องุ่นไร่เม็ด--@ไร่อิงจันทร์>
<http://thai.tourismthailand.org/สถานที่ท่องเที่ยว/คูพระเจ้าเม็งราย--187>
<http://thai.tourismthailand.org/สถานที่ท่องเที่ยว/ดอยแม่สลอง--195>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/ดอยผาตั้ง--212>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/ดอยมด--4381>
<http://www.tourdoi.com/doi/langkaluang/general.htm>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/ดอยหัวแม่คำ--196>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/ทะเลสาบเชียงแสน--206>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/น้ำตกแม่เฒ--4380>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/น้ำตกขุนแจ--4376>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/บ้านเทอดไทย--4384>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/บ้านหาดบ้าย--4410>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/พระตำหนักดอยตุงและสวนแม่ฟ้าหลวง--4404>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/วัดพระธาตุจอมแจ้ง--4203>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/พระธาตุจอมแว่--4377>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/พระธาตุดอยเวา--4420>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/พระธาตุดอยตุง--4406>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/พระธาตุดอยปู่เข้า--209>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/พิพิธภัณฑสถานและศูนย์ศึกษาชาวเขา--4386>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/พิพิธภัณฑสถานบ้านดำ--6604>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/พิพิธภัณฑสถานพระประทับโกลด์แลนด์--4414>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/พิพิธภัณฑสถานแห่งชาติเชียงใหม่--199>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/พิพิธภัณฑสถานอุปลาคำ--190>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/รอยพระบาทแห่งสันติภาพ--4385>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/ลานทองอุทยานวัฒนธรรมลุ่มน้ำโขง--4415>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/วัดเจ็ดยอด--203>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/วัดกลางเวียง--4390>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/วัดถ้ำป่าอาชาทอง--4416>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/วัดถ้ำผาจอม--4379>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/วัดพระเจ้าล้านทอง--201>

<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/วัดพระเจ้าล้านทอง--201>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/วัดพระแก้ว--4391>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/วัดพระธาตุเจดีย์หลวง--200>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/วัดพระธาตุจอมกิติ--6063>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/วัดพระธาตุดอยจอมทอง--191>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/วัดพระธาตุผาเงา--6037>
<http://www.paiduaykan.com/province/north/chiangmai/watphrasing.html>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/วัดมิ่งเมือง--4392>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/วัดมุงเมือง--4393>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/วัดร่องขุน--189>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/วัดศรีเกิด--4394>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/วัดศรีบุญเรือง--3918>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/วัดสังฆาแก้วดอนพัน--205>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/ศูนย์ภาษาและวัฒนธรรมจีนสิรินธร--193>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/ศูนย์ส่งเสริมการเกษตรดอยผาหม่น--4411>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/อนุสาวรีย์พ่อขุนเม็งรายมหาราช--186>
<http://www.paiduaykan.com/province/north/maehongson/lovestrawberrypai.html>
http://www.paiduaykan.com/76_province/north/maehongson/pangung.html
<http://www.paiduaykan.com/province/north/maehongson/mealanoi.html>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/กองแลน--6053>
<http://www.paiduaykan.com/province/north/maehongson/worldviewpoint.html>
<http://www.paiduaykan.com/province/north/maehongson/doimeang.html>
<http://www.paiduaykan.com/province/north/maehongson/doiphuiko.html>
http://www.paiduaykan.com/76_province/north/maehongson/thamlod.html
http://www.paiduaykan.com/76_province/north/maehongson/maxican_sunflower.html
<http://travel.kapook.com/view96917.html>
http://www.paiduaykan.com/76_province/north/maehongson/banrakthai.html
<http://www.paiduaykan.com/province/north/maehongson/baanhuayhom.html>
<http://www.painaidii.com/diary/diary-detail/001470/lang/th/>
http://www.paiduaykan.com/76_province/north/maehongson/pai.html

http://www.paiduaykan.com/76_province/north/maehongson/pang-tong.html
http://www.paiduaykan.com/76_province/north/maehongson/phrathatdoikongmu.html
<http://www.paiduaykan.com/province/north/maehongson/pratadseejom.html>
http://www.paiduaykan.com/76_province/north/maehongson/watjongklang-jongkham.html
http://www.paiduaykan.com/76_province/north/maehongson/watjongklang-jongkham.html
http://www.paiduaykan.com/76_province/north/maehongson/LongNeckKaren.html
http://www.paiduaykan.com/76_province/north/maehongson/bannasantichon.html
<http://www.painaidii.com/diary/diary-detail/001470/lang/th/>
<http://travel.kapook.com/view96917.html>
http://www.paiduaykan.com/76_province/north/nan/saodinnanoi.html
<http://www.teeteawthai.com/เยือนไทลื้อมะเก่า>
<http://www.prapayneethai.com/กำแพงเมืองน่าน>
<http://www.lannatouring.com/Nan/Interesting-article/Khumchao-Nan.htm>
<http://www.lannatouring.com/Nan/Interesting-article/Wat-Prang-Dikduem.htm>
<http://www.lannatouring.com/Nan/Destination-guide/Muangnan5.htm>
<http://www.sadoodta.com/info/น้ำตกภูฟ้า>
www.thaitravelism.com/back/ภาคเหนือ/น่าน/185-น้ำตกศิลาเพชร-น่าน.html
www.thaitravelism.com/back/ภาคเหนือ/น่าน/259-บ่อเกลือ-น่าน.html
<http://www.lannatouring.com/Nan/Interesting-article/WatNongbua-Nan.htm>
<http://www.bansansuk.com/travel/Pha%20Chu/>
<http://www.lannatouring.com/Nan/Interesting-article/Nan-NationalMuseum.htm>
[lannatouring.com/Nan/Destination-guide/Muangnan2.htm](http://www.lannatouring.com/Nan/Destination-guide/Muangnan2.htm)
<http://www.paiduaykan.com/province/north/nan/watbantolaeng.html>
<http://www.bansansuk.com/travel/Wat%20Boon%20stand/>
http://www.uncledeng.com/portfolio-view/2015_wat_payapoo/
http://www.paiduaykan.com/76_province/north/nan/prathatkaonoi.html
<http://www.paiduaykan.com/province/north/nan/watphrathatpingsakad.html>
http://www.paiduaykan.com/76_province/north/nan/wadprathatchangkam.html
http://www.paiduaykan.com/76_province/north/nan/watphrathatchaehaeng.html
<http://www.paiduaykan.com/province/north/nan/watphuket.html>

http://www.paiduaykan.com/76_province/north/nan/watpumin.html
http://www.paiduaykan.com/76_province/north/nan/watmingmuang.html
<http://www.paiduaykan.com/province/north/nan/watrongngae.html>
<http://www.paiduaykan.com/province/north/nan/watsripanton.html>
www.thaitravelism.com/back/ภาคเหนือ/น่าน/184-วัดสวนตาล-น่าน.html
<http://www.xn--72cf2b2bcb7a9ayfd9ie7bxdk.com/topic/98>
<http://www.paiduaykan.com/province/north/nan/wathuakuang.html>
http://www.paiduaykan.com/76_province/north/nan/watnongbua.html
<http://www.thaitravelism.com/back/ภาคเหนือ/น่าน/180-สถานีพัฒนาการเกษตรที่สูง-น่าน.html>
www.thaitravelism.com/back/ภาคเหนือ/น่าน/186-หอศิลป์ริม่าน-น่าน.html
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/อนุสาวรีย์วีรกรรม-พลเรือน-ตำรวจ-ทหารและ-พิพิธภัณฑสถานทหารทุ่งช้าง--4310>
<http://ridceo.rid.go.th/petboon/kongen.php>
<https://www.touron thai.com/travelguide/placeinfo.php?id=1377>
<http://www.lampang.go.th/travel/data/placetour/maemo.htm>
http://www.paiduaykan.com/76_province/north/lampang/kadkongta.html
<http://www.lampang.go.th/travel/data/placetour/wangnua.htm>
http://www.paiduaykan.com/76_province/north/lampang/kadkongta.html
<http://www.thoenpost.com/ab.thoen/place/nampha%20pha-ngom%20cave/pha-ngom%20cave/cave/index.htm>
<https://www.touron thai.com/article/1338>
<http://www.lampang.go.th/travel/data/placetour/maung.htm>
<http://www.paiduaykan.com/province/north/lampang/hugyou.html>
http://www.biogang.net/ecotourism_view.php?uid=61772&id=129823
http://www.dmr.go.th/ewtadmin/ewt/dmr_web/main.php?filename=n06
<http://wangnuasweety.net/index.php>
<http://wattumsukkasem.blogspot.com/>
http://www.paiduaykan.com/76_province/north/lampang/wat-chedi-sao.html
www.amazingthaitour.com/ภาคเหนือ/วัดเฉลิมพระเกียรติพระจอมเกล้าราชานุสรณ์-วัดพระบาทปู่ผาแดง-อำเภอแจ้ห่ม-จังหวัดลำปาง/

<http://www.sadoodta.com/info/วัดเสลารัตนปัพพตาราม-วัดไหล่หินหลวงแก้วช้างยืน>
http://www.nfemaetha.com/showdownload2.php?no_pp=858
<http://pitchaya.net/วัดบ้านก่อ/>
<http://www.thaiticketmajor.com/variety/travel/1899/>
http://www.nfemaetha.com/showdownload2.php?no_pp=848
<https://tanaporn264.wordpress.com/สารน่ารู้เรื่องวัดป่า/>
<http://pitchaya.net/วัดพระเกิด/>
<http://www.lampang.go.th/travel/data/placetour/maung.htm>
http://www.thongteaw.com/Travel_tour_content_ลำปาง/วัดพระธาตุจอมปิง%20อ.เกาะคา%20จ.ลำปาง.html
<http://www.lampang.go.th/travel/data/placetour/kokha.htm>
<http://www.lampang.go.th/travel/data/placetour/maephrik.htm>
<http://www.tripchiangmai.com/chiangmaiboard/index.php?topic=7002.0#.W4-aac4zYdV>
<http://www.lampang.go.th/travel/data/placetour/maung.htm>
<http://www.banmuang.co.th/news/region/5071>
<https://www.touronthai.com/article/1358>
<http://www.lampang.go.th/travel/data/placetour/soemngam.htm>
<https://thai.tourismthailand.org/สถานที่ท่องเที่ยว/ศูนย์อนุรักษ์ช้างไทย-และ-สวนป่าทุ่งเกวียน--4132>
<http://www.handicrafttourism.com/ศักยภาพหมู่บ้าน/หมู่บ้านหลุก/>
<http://student.nu.ac.th/white/Thoen/travel/2.html>
<http://www.indraoutlet.com/aboutus.html>
<http://adventure.tourismthailand.org/thai/?p=265>
<https://www.ilovetogo.com/Article/71/3854/เริ่มแล้ว-Scenario-Paintball-แห่งเดียวในประเทศไทยที่เชียงใหม่ไนท์ซาฟารี>
<http://www.landingchiangmai.com/tours/trek-maetang/index.htm>
<http://www.sawadee.co.th/chiangmai/topactivities.html>
<http://adventure.tourismthailand.org/thai/?p=271>
<http://adventure.tourismthailand.org/thai/?p=241>
<http://www.touchchiangmai.com/th/Tours-detail/fishing-chiang-mai-holiday-tour/>
<http://adventure.tourismthailand.org/thai/?p=520>

<http://adventure.tourismthailand.org/thai/?p=269>
<http://adventure.tourismthailand.org/thai/?p=89>
<http://adventure.tourismthailand.org/thai/?p=188>
<http://adventure.tourismthailand.org/thai/?p=261>
<http://adventure.tourismthailand.org/thai/?p=164>
<http://www.cots.go.th/travelView.php?id=119&aid=1>
<http://www.sawadee.co.th/chiangrai/topactivities.html>
<http://www.cots.go.th/travelView.php?id=118&aid=1>
<http://www.chiangraifocus.com/2012/attractions/details.php?id=110&district=1>
<http://www.cots.go.th/travelView.php?id=117&aid=11>
<http://www.cots.go.th/travelView.php?id=120&aid=1>
<http://www.cots.go.th/travelView.php?id=121&aid=1>
<http://adventure.tourismthailand.org/thai/?p=263>
<http://www.maehongson4u.com/activities/bicycle.html>
<http://adventure.tourismthailand.org/thai/?p=186>
<http://www.maehongson4u.com/activities/elephant.html>
<http://www.maehongson4u.com/activities/rafting.html>
<http://adventure.tourismthailand.org/thai/?p=82>
<http://adventure.tourismthailand.org/thai/?p=216>
<http://www.maehongson4u.com/activities/bambooraft.html>
<http://adventure.tourismthailand.org/thai/?p=94>
<https://www.touronthai.com/article-6001003.html>
<http://www.lannatouring.com/Nan/Destination-guide/Muangnan5.htm>
http://tiaownan.blogspot.com/2012/02/blog-post_2499.html
http://www.nantouring.com/index.php?option=com_content&view=article&id=82%3A2012-08-30-12-21-19&catid=45%3A2012-06-07-02-55-34&Itemid=76&lang=th
<https://www.lovethailand.org/travel/th/3-ลำปาง/2218-ค่ายฝึกการรบพิเศษประตูผา.html>
<http://www.lampang.go.th/travel/data/placetour/maung.htm>
<http://adventure.tourismthailand.org/thai/?p=191>
<http://adventure.tourismthailand.org/thai/?p=167>

<http://adventure.tourismthailand.org/thai/?p=191>

http://www.thailovetrip.com/view.php?id_view=321

<http://www.lampang.go.th/travel/data/placetour/maung.htm>

<http://www.oceansmile.com/N/Nan/NANm3.htm>

<http://www.poobpubtour.com/2010/09/jaeson.html>

หมายเหตุ เข้าถึงเว็บไซต์เมื่อ กรกฎาคม - สิงหาคม 2559

DRU

ประวัติผู้วิจัย

ผู้ช่วยศาสตราจารย์ ดร. อรรวรรณ อิ่มสมบัติ

วศ.ด. (วิศวกรรมคอมพิวเตอร์) มหาวิทยาลัยเกษตรศาสตร์

วท.ม. (การจัดการระบบสารสนเทศ) สถาบันบัณฑิตพัฒนบริหารศาสตร์

วท.บ. (วิทยาการคอมพิวเตอร์) มหาวิทยาลัยเกษตรศาสตร์

ตำแหน่งปัจจุบัน

หัวหน้าหลักสูตรการออกแบบดิจิทัลเชิงสร้างสรรค์ วิทยาลัยครีเอทีฟดีไซน์ แอนด์
เอ็นเตอร์เทนเมนต์เทคโนโลยี มหาวิทยาลัยธุรกิจบัณฑิต

ผลงานทางวิชาการ

Imsombut, A. and Kajornrit, J., A Comparison of Statistical and Data Mining Techniques for Enrichment Ontology with Instances, Journal of Reviews on Global Economics: Special Issue - Recent Topical Research on Global, Energy, Health & Medical, and Tourism Economics, and Global Software, p. 375-379 (2017).

Imsombut, A. and Paireekreng, W., “Extract Knowledge for Populating Thai Tourism Ontology from Texts Using Feature-weighted k-Nearest Neighbor”, In Proceeding of the 1st International Conference on Information Technology (InCIT2016), Thailand (2016).

Imsombut, A. and Sirikayon, C., “An Alternative Technique for Populating Thai Tourism Ontology from Texts Based on Machine Learning”, 15th IEEE/ACIS International Conference on Computer and Information Science (ICIS 2016), Japan (2016).

อรรวรรณ อิ่มสมบัติ และ ฐากร พุกษวันประสุต, “การวิเคราะห์และการประเมินคุณภาพแบบทดสอบออนไลน์”, การประชุมวิชาการระดับชาติด้านเทคโนโลยีสารสนเทศ ครั้งที่ 6, นครราชสีมา, 2557.

ฉัตรมณี เพชรผึ้ง และ อรรวรรณ อิ่มสมบัติ, “ระบบสนับสนุนการตัดสินใจการประเมินการจัดการคุณภาพการปฏิบัติทางการเกษตรที่ดีสำหรับพืชอาหาร”, In Proceeding of the 28th

- National Graduate Research Conference, "Advancing Knowledge through Graduate Research", 2556.
- จุฬามาศ เทียนสอาด และ อรวรรณ อิมสมบัติ, “ระบบผู้เชี่ยวชาญเพื่อวินิจฉัยและให้คำแนะนำผู้ป่วยไตวายเรื้อรังโดยใช้ฐานความรู้ออนโทโลยี”, Journal of Information Science and Technology, Vol.3, No.2 , pp.22-30, 2555.
- จุฬามาศ เทียนสอาด และ อรวรรณ อิมสมบัติ, “การพัฒนาระบบผู้เชี่ยวชาญเพื่อวินิจฉัยและให้คำแนะนำในผู้ป่วยไตวายเรื้อรัง ที่ได้รับการฟอกเลือดด้วยเครื่องไตเทียม” Th 4th National conference on Information Technology (NCIT2012), Thailand (2012).
- อรวรรณ อิมสมบัติ และ มนต์ชัย พงษ์นุ่มกุล, “การศึกษาเปรียบเทียบตัวชี้วัดและเทคนิคการประเมินผลการจัดการความรู้”, งานประชุมวิชาการสมาคมสถาบันอุดมศึกษาเอกชนแห่งประเทศไทย (สสอท), กรุงเทพฯ, (2554).
- Imsombut, A., “A Statistical Approach for Semantic Relation Extraction”, In Proceeding of the Eighth Symposium on Natural Language Processing 2009 (SNLP 2009), Bangkok, (2009).
- อรวรรณ อิมสมบัติ, “การสร้างออนโทโลยีเฉพาะงานด้านการเกษตรจากเอกสารวิชาการ : ผลการวิจัยเบื้องต้น”, การประชุมวิชาการการบริหารและการจัดการ ครั้งที่ 4, กรุงเทพฯ, (2551).
- Imsombut, A., “Agricultural Task-oriented Ontology Construction from Technical Corpus: Preliminary Report”, In Proceeding of the 4th National Conference of Administrative and Management, Bangkok, Thailand (2008).
- Imsombut, A., “Non-Taxonomic Relation Extraction for Thai Ontology Learning”, In Proceeding of The Seventh Symposium on Natural Language Processing 2007 (SNLP 2007), Thailand (2007).