รหัสโครงการ SUT2-204-57-12-04

รายงานการวิจัย

ระบบถามตอบเชิงความหมายออนไลน์สำหรับผลไม้เศรษฐกิจของไทย (An Online Semantic Question Answering System for Thai Economic Fruits)

ได้รับทุนอุดหนุนการวิจัยจาก
มหาวิทยาลัยเทคโนโลยีสุรนารี

ผลงานวิจัยเป็นความรับผิดชอบของหัวหน้าโครงการวิจัยแต่เพียงผู้เดียว

รหัสโครงการ SUT2-204-57-12-04

รายงานการวิจัย

ระบบถามตอบเชิงความหมายออนไลน์สำหรับผลไม้เศรษฐกิจของไทย (An Online Semantic Question Answering System for Thai Economic Fruits)

คณะผู้วิจัย

หัวหน้าโครงการ

ผู้ช่วยศาสตราจารย์ ดร.ศุภกฤษฎิ์ นิวัฒนากูล

สาขาวิชาเทคโนโลยีสารสนเทศ สำนักวิชาเทคโนโลยีสังคม

มหาวิทยาลัยเทคโนโลยีสุรนารี

ผู้ร่วมวิจัย

อาจารย์ ดร.ฉัตรภัสร์ ฐิติอัคราวงศ์

ได้รับทุนอุดหนุนการวิจัยจากมหาวิทยาลัยเทคโนโลยีสุรนารี ประจำงบประมาณ พ.ศ. 2557

มหาวิทยาลัยเทคโนโลยีสุรนารี

ผลงานวิจัยเป็นความรับผิดชอบของหัวหน้าโครงการวิจัยแต่เพียงผู้เดียว

ธันวาคม 2560

กิตติกรรมประกาศ

การวิจัยครั้งนี้ได้รับทุนอุดหนุนการวิจัยจากมหาวิทยาลัยเทคโนโลยีสุรนารี ประจำปีงบประมาณ พ.ศ. 2557 และงานวิจัยฉบับนี้สำเร็จลุล่วงได้ด้วยดี ด้วยเพราะได้รับการสนับสนุนอันดีจากสาขาวิชาเทคโนโลยีสารสนเทศ สถานวิจัย สำนักวิชาเทคโนโลยีสังคม มหาวิทยาลัยเทคโนโลยีสุรนารี ที่อำนวยความสะดวกด้านการดำเนินการต่าง ๆ ที่เกี่ยวข้องกับงานวิจัย ขอขอบคุณผู้ช่วยนักวิจัย คุณเจษฎา สิงห์ทองชัย คุณลดาวัลย์ แย้มครวญ รวมถึงผู้เชี่ยวชาญ ผู้ทดสอบการใช้งานระบบที่มีส่วนช่วยในขั้นตอนต่าง ๆ ของการพัฒนาและประเมินระบบต้นแบบ นอกจากนี้ ยังได้รับความกรุณาอย่างดียิ่งจากผู้ทรงคุณวุฒิที่ได้เสียสละเวลาในการตรวจข้อเสนอโครงการวิจัย รวมถึงรายงานการวิจัยฉบับสมบูรณ์ และในโอกาสนี้ ผู้วิจัยขอขอบคุณทุก ๆ ท่านที่มิได้ทั้งหมด ที่มีส่วนช่วยเหลือ สนับสนุน และให้กำลังใจจนกระทั่งงานวิจัยนี้ได้สำเร็จลุล่วงไปได้ด้วยดี

คุณค่าและประโยชน์อันจะพึงมีจากงานวิจัยฉบับนี้ขอมอบแด่บิดามารดา ครูอาจารย์ ครอบครัว และเพื่อนร่วมงานในสำนักวิชาเทคโนโลยีสังคม มหาวิทยาลัยเทคโนโลยีสุรนารี ทุกท่านที่ให้กำลังใจและช่วยเหลือตลอดมา

คณะผู้วิจัย

มหาวิทยาลัยเทคโนโลยีสุรนารี

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์หลักเพื่อ 1) พัฒนา และประเมินกระบวนการถามตอบเชิงความหมายออนไลน์สำหรับผลไม้เศรษฐกิจของไทย และ 2) เพื่อออกแบบ พัฒนา และประเมินระบบถามตอบเชิงความหมายออนไลน์ เพื่อช่วยให้ผู้ใช้ได้เข้าถึงความรู้เกี่ยวกับผลไม้เศรษฐกิจของไทยที่อยู่กระจัดกระจายบนระบบเครือข่ายอินเตอร์เน็ต ในการวิจัยดังกล่าวได้มีการใช้ออนโทโลยีอธิบายความหมาย ความสัมพันธ์และโครงสร้างของความรู้ที่จัดเก็บผ่านภาษาโอดับเบิ้ลยูแอล (Web Ontology language : OWL) และใช้การวิเคราะห์พฤติกรรมการตั้งคำถามจากผู้ใช้เพื่อนำมาประกอบในการพัฒนากระบวนการถามตอบเชิงความหมายดังกล่าว ผลการวิจัยได้พัฒนากระบวนการถามตอบเชิงความหมายสำหรับผลไม้เศรษฐกิจของไทย ประกอบด้วย 4 ขั้นตอน ได้แก่ 1) การจัดความรู้ให้เป็นระบบ 2) การประมวลผลคำถามเชิงความหมาย 3) การจับคู่คำถามและคำตอบ และ 4) การสร้างชุดคำตอบ ได้มีการพัฒนาระบบถามตอบเชิงความหมายออนไลน์โดยพัฒนาส่วนต่อประสานผู้ใช้ให้มีปฏิสัมพันธ์ระหว่างผู้ใช้กับระบบเป็นแบบถามตอบ ผลลัพธ์จากการถามตอบแสดงผ่านเว็บเบราว์เซอร์ประกอบด้วยข้อมูลที่เกี่ยวข้องรวมถึงลิงก์เชื่อมโยงไปยังเอกสาร ผลการประเมินระบบถามตอบเชิงความหมาย พบว่าระบบสามารถให้คำตอบได้ตรงตามความต้องการครบถ้วน โดยมีค่าความแม่นยำร้อยละ 87 ค่าความระลึกร้อยละ 85 และค่าอัตราการรู้จำร้อยละ 86 ผู้ใช้มีความพึงพอใจต่อระบบถามตอบเชิงความหมายสำหรับผลไม้เศรษฐกิจของไทยโดยรวมอยู่ในระดับมาก ทั้งในรูปแบบการระบุข้อคำถามเหมาะสม สะดวกต่อการใช้งาน การค้นคืนความรู้ตรงกับความต้องการ และการเชื่อมโยงไปยังแหล่งความรู้ที่พบ

คำสำคัญ: กระบวนการถามตอบ ระบบถามตอบเชิงความหมายออนไลน์ ผลไม้เศรษฐกิจไทย

มหาวิทยาลัยเทคโนโลยีสุรนารี

Abstract

This research primarily aimed to 1) develop and evaluate an online semantic question answering process for Thai economic fruits and 2) design, develop, and evaluate an online semantic question answering system to support users to access scattered knowledge about Thai economic fruits that is available on the Internet. This research employed ontology to describe meanings, relationship, and knowledge structure through Web Ontology Language (OWL) and behaviour analysis in questioning from users for deploying in this question answering process development. The research results presented the question answering process development for Thai economic fruits that consisted of 4 processes: 1) managing knowledge systematically 2) processing semantic questions 3) matching questions and answers, and 4) generating a set of answers. In addition, this research paper developed an online semantic question answering system by developing a user interface that interacts between users and the system as question answering format. The results from question answering showed via web browser that contained relevant information, including links to documents. The results from this system displayed that it was able to provide answers that met the full requirements by gaining 87 percent for precision rate, 85 percent for recall rate, and 86 percent for F-measure rate. Furthermore, the overall satisfaction from users with this system for Thai economic fruits were in a high level in both a proper question identification, easy to use, knowledge retrieval that met the requirements, and links to the source of knowledge found.

มหาวิทยาลัยเทคโนโลยีสุรนารี

Keywords: Question answering process; an online semantic question answering system; Thai economic fruits

สารบัญ

หน้า

กิตติกรรมประกาศ ..... ก
บทคัดย่อภาษาไทย ..... ข
บทคัดย่อภาษาอังกฤษ ..... ค
สารบัญ ..... ง
สารบัญตาราง ..... ช
สารบัญภาพ ..... ช
บทที่ 1 บทนำ 1
1.1 ความสำคัญและที่มาของปัญหาการวิจัย 1
1.2 วัตถุประสงค์ของการวิจัย 4
1.3 สมมติฐานการวิจัย 4
1.4 ขอบเขตของการวิจัย 4
1.5 ข้อตกลงเบื้องต้น 4
1.6 ประโยชน์ที่ได้รับจากการวิจัย 5
1.7 คำอธิบายศัพท์ 5
บทที่ 2 ปริทัศน์วรรณกรรมและงานวิจัยที่เกี่ยวข้อง 6
2.1 ระบบถามตอบผ่านเว็บ (Web-Based Question Answering System) 6
2.1.1 ระบบถามตอบ 8
2.1.2 กระบวนการประมวลผลคำถาม 9
2.2 การสกัดข้อมูล (Data Extraction) 11
2.3 การจัดหมวดหมู่ 13
2.4 การค้นคืนสารสนเทศ (Information Retrieval) 13
2.4.1 กระบวนการค้นคืนสารสนเทศ 15

มหาวิทยาลัยเทคโนโลยีสุรนารี

2.4.2 รูปแบบการค้นคืนสารสนเทศ 16
2.4.3 การวัดประสิทธิภาพการค้นคืนสารสนเทศ 18
2.5 การประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP) 19
2.5.1 ระดับของการประมวลผลภาษาธรรมชาติ 20
2.5.2 ขั้นตอนการประมวลผลภาษาธรรมชาติ 20
2.5.3 องค์ประกอบของการประมวลผลภาษาธรรมชาติ 21
2.5.4 เทคนิคการประมวลผลภาษาธรรมชาติ 22
2.6 การจัดการฐานความรู้ด้วยออนโทโลยี 22
2.6.1 ประเภทของออนโทโลยี 24
2.6.2 ส่วนประกอบหลักของออนโทโลยี 25
2.6.3 ประโยชน์ของออนโทโลยี 26
2.6.4 ภาษาและเครื่องมือที่ใช้ในการจัดการฐานความรู้ออนโทโลยี 26
2.7 ผลไม้เศรษฐกิจของไทย 36
2.8 เอกสารและงานวิจัยที่เกี่ยวข้อง 37
2.9 กรอบแนวคิดการวิจัย 42
บทที่ 3 วิธีการดำเนินการวิจัย 46
3.1 การศึกษาข้อมูลเบื้องต้น 46
3.1.1 การวิเคราะห์พฤติกรรมการตั้งคำถามของผู้ใช้ 46
3.1.2 โครงสร้างของความรู้ทางด้านผลไม้เศรษฐกิจของไทย 52
3.2 การวิเคราะห์และออกแบบระบบ 53
3.3 การพัฒนาระบบ 57
3.4 การประเมินผลระบบ 57
3.5 การสร้างและหาประสิทธิภาพของเครื่องมือ 58
บทที่ 4 ผลการวิจัยและอภิปรายผล 61

4.1 ผลการพัฒนากระบวนการถามตอบเชิงความหมายออนไลน์.....61
4.1.1 การประมวลผลความรู้.....61
4.1.2 การประมวลผลคำถามเชิงความหมาย.....64
4.1.3 การจับคู่คำถาม-คำตอบ.....65
4.1.4 การสร้างชุดคำตอบ.....66
4.2 ผลการประเมินระบบ.....67
4.2.1 ผลการประเมินความถูกต้องและความสามารถของระบบ.....67
4.2.2 ผลการวิเคราะห์ความพึงพอใจ.....68
4.3 การอภิปรายผล.....70
บทที่ 5 สรุปผลการวิจัยและอภิปรายผล.....72
5.1 สรุปผลการวิจัย.....72
5.2 ข้อเสนอแนะ.....73
บรรณานุกรม.....74
ภาคผนวก ก.....83
ภาคผนวก ข.....88
ภาคผนวก ค.....103
ประวัติผู้วิจัย.....106

มหาวิทยาลัยเทคโนโลยีสุรนารี

สารบัญตาราง

ตารางที่ 2.1 ตัวกระทำทางบูลีน....17
ตารางที่ 2.2 อธิบายตัวอย่างคลาสและคุณสมบัติของรถยนต์....26
ตารางที่ 2.3 การประกาศถึงคุณลักษณะเฉพาะของสมาชิกภายในคลาส....34
ตารางที่ 2.4 ชนิดข้อมูลที่อ้างอิงตาม XML schema datatypes....35
ตารางที่ 2.5 ประเภทข้อกำหนดคุณสมบัติของ OWL....35
ตารางที่ 2.6 สรุปเปรียบเทียบงานวิจัยที่เกี่ยวข้องระบบถามตอบ....42
ตารางที่ 3.1 รายชื่อเว็บไซต์ที่แสดงถึงพฤติกรรมการตั้งคำถามที่เกี่ยวข้องผลไม้เศรษฐกิจ....47
ตารางที่ 3.2 การวิเคราะห์พฤติกรรมการตั้งคำถาม....50
ตารางที่ 3.3 แสดงจำนวนประชากรและกลุ่มตัวอย่างของข้อคำถาม จำแนกตามกลุ่มการตั้งคำถาม และการจัดหมวดหมู่คำถาม....51
ตารางที่ 3.4 วิเคราะห์คำและจัดหมวดหมู่คำถาม....56
ตารางที่ 3.5 ผลการประเมินความสอดคล้องของข้อคำถามโดยผู้เชี่ยวชาญ....60
ตารางที่ 4.1 ผลการประเมินความถูกต้องและความสามารถของระบบ....67
ตารางที่ 4.2 ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของความพึงพอใจของผู้ใช้ที่มีต่อระบบถามตอบ....70

มหาวิทยาลัยเทคโนโลยีสุรนารี

สารบัญภาพ

ภาพที่ 2.1 ขั้นตอนการทำงานระบบถามตอบ 7
ภาพที่ 2.2 สถาปัตยกรรมของระบบถามตอบ 8
ภาพที่ 2.3 กระบวนการสกัดข้อมูล 12
ภาพที่ 2.4 กระบวนการค้นคืนสารสนเทศ 14
ภาพที่ 2.5 ลำดับขั้นของการวิเคราะห์ทางภาษา (Linguistic Analysis) 20
ภาพที่ 2.6 ขั้นตอนของการประมวลผลภาษาธรรมชาติ 21
ภาพที่ 2.7 ประเภทของออนโทโลยี 24
ภาพที่ 2.8 ตัวอย่างส่วนประกอบของออนโทโลยี 25
ภาพที่ 2.9 ตัวอย่างข้อมูลอินสแตนส์ในการอธิบายรายละเอียดของข้อมูล 25
ภาพที่ 2.10 ตัวอย่างองค์ประกอบการแทนแนวคิดของรถยนต์ 27
ภาพที่ 2.11 โครงสร้างของเมทาดาทาของเอกสาร RDF (ที่มา W3C, 1999) 30
ภาพที่ 2.12 การอธิบายความสัมพันธ์ของ Resource ในรูปแบบ RDFS (ที่มา W3C, 1999) 30
ภาพที่ 2.13 ระบบถามตอบเชิงความหมายออนไลน์สำหรับผลไม้เศรษฐกิจของไทย 45
ภาพที่ 3.1 แสดงขั้นตอนการดำเนินการวิจัย 46
ภาพที่ 3.2 การกำหนดออฟเจคพร็อพเพอร์ตี้เพิ่มเติมในออนโทโลยีผลไม้เศรษฐกิจ 53
ภาพที่ 3.3 ตัวอย่างการจัดกลุ่มเอกสาร 53
ภาพที่ 3.4 กรอบการทำงานของระบบถามตอบเชิงความหมายออนไลน์ 54
ภาพที่ 3.5 แสดงตัวอย่างการตัดคำของข้อคำถาม 55
ภาพที่ 4.1 กระบวนการประมวลผลความรู้ 62
ภาพที่ 4.2 กระบวนการประมวลผลคำถามเชิงความหมาย 64
ภาพที่ 4.3 กระบวนการจับคู่คำถาม-คำตอบ 65
ภาพที่ 4.4 กระบวนการสร้างชุดคำตอบ 67
ภาพที่ 4.5 ตัวอย่างผลลัพธ์ของคำถาม “วิธีขยายพันธุ์ลำไย” 68
ภาพที่ 4.6 ตัวอย่างผลลัพธ์ของคำถาม “โรคและแมลงมังคุด” 69

มหาวิทยาลัยเทคโนโลยีสุรนารี

บทที่ 1 บทนำ

1.1 ความสำคัญและที่มาของปัญหาการวิจัย

การค้นหาข้อมูลในปัจจุบันโดยส่วนใหญ่ จะค้นหาจากข้อมูลที่จัดเก็บอยู่บนเครือข่ายอินเตอร์เน็ต เนื่องจากเป็นแหล่งที่มีข้อมูลมากมายมหาศาลและหลากหลาย มีความสะดวกต่อผู้ใช้ในการค้นหา สามารถค้นหาได้ทุกที่ทุกเวลาเพียงแค่มีอุปกรณ์ที่สามารถติดต่อเครือข่ายอินเตอร์เน็ตได้ การค้นหาข้อมูลจากเครือข่ายอินเตอร์เน็ตจะใช้บริการจากโปรแกรมค้นหา (Search engine) ที่อำนวยความสะดวกให้กับผู้ใช้ค้นหาข้อมูลได้สะดวกสบายมากขึ้น เพียงแต่ผู้ใช้ระบุคำสำคัญหรือคำค้นที่ต้องการ โปรแกรมค้นหาจะนำรายการข้อมูลหรือเอกสารมาแสดงเพื่อให้ผู้ใช้ได้เลือกข้อมูลหรือเอกสารที่ตรงกับความต้องการของผู้ใช้ นอกจากนี้ โปรแกรมค้นหายังอาจมีความสามารถที่ให้ผู้ใช้ระบุขอบเขตของข้อมูลที่ต้องการค้นหาได้ เช่น การระบุช่วงเวลา การระบุภาษา การกำหนดตามขอบเขตพื้นที่ การใช้คำเชื่อมต่าง ๆ เป็นต้น ตัวอย่างของโปรแกรมค้นหาที่ให้บริการในปัจจุบัน เช่น Google, Yahoo, Baidu, Bing เป็นต้น

ข้อมูลที่กระจายอยู่บนระบบเครือข่ายอินเทอร์เน็ตมีหลากหลายรูปแบบ เช่น ข้อมูลที่เป็นตัวอักษร (Textual Information) รูปภาพ (Image) วีดีโอ (Video) เป็นต้น การมีข้อมูลในปริมาณที่มากและหลากหลายรูปแบบ ถึงแม้จะมีผลดีที่ทำให้ผู้ใช้สามารถค้นหาข้อมูลได้จากเว็บไซต์ต่าง ๆ ที่หลากหลายมากขึ้น อีกทั้งสามารถค้นหาข้อมูลได้หลากหลายรูปแบบ แต่การที่ผู้ใช้สามารถเข้าถึงข้อมูลที่หลากหลายรูปแบบและมีที่มาที่หลากหลายนั้น ได้นำมาสู่ปัญหาในการค้นคืนสารสนเทศสำหรับผู้ใช้เอง ได้แก่ ผลลัพธ์ที่ได้จากการค้นหาอาจไม่ตรงตามความต้องการของผู้ใช้ ผลลัพธ์ที่ได้จากการค้นหามีทั้งที่เกี่ยวข้องและไม่เกี่ยวข้องกับสิ่งที่ผู้ใช้ต้องการ หรือผลลัพธ์ที่ได้มีปริมาณมากเกินไปจนกระทั่งไม่สามารถเข้าไปดูได้ครบทั้งหมด ผลลัพธ์ที่ได้จะแสดงตามคำค้น (Keywords) ที่ตรงกับผู้ใช้กำหนดสำหรับการค้นหา ทำให้เอกสารทั้งหมดที่อยู่ในฐานข้อมูลไม่ถูกนำมาแสดง และเอกสารบางส่วนที่นำมาแสดงก็ไม่ตรงกับความต้องการอันเนื่องมาจากความหมายของคำที่อาจมีได้หลายความหมาย ผู้ใช้ไม่สามารถระบุแน่ชัดว่าเอกสารที่แสดงนั้นตรงกับความต้องการหรือไม่ ต้องใช้วิธีการสุ่มเข้าไปอ่านรายละเอียดจากเอกสารซึ่งทำให้เสียเวลาเป็นอย่างยิ่ง นอกจากนี้ เมื่อพบเอกสารที่ต้องการแล้วก็ไม่สามารถแสดงการเชื่อมโยงไปยังเอกสารอื่น ๆ ที่เกี่ยวข้องได้

ในการค้นหาข้อมูลของผู้ใช้นั้น ผู้ใช้ต้องการค้นหาข้อมูลที่เกี่ยวข้องกับเรื่องที่ตรงกับที่ตนเองสนใจเท่านั้น มีผู้ใช้จำนวนไม่น้อยที่ขาดความชำนาญในการใช้คำค้น ไม่สามารถกำหนดคำหลักหรือคำที่เฉพาะเจาะจงสำหรับการค้นหาที่เหมาะสมได้ ทำให้ผลของการค้นคืนมีปริมาณมากทั้งที่ตรงกับความ

ต้องการและไม่ตรงกับความต้องการ และการที่ได้ผลลัพธ์จากการค้นหาที่มีปริมาณมากเกินไป ทำให้ผู้ใช้เสียเวลาในการคัดเลือกเอกสารที่ตรงกับความต้องการอย่างแท้จริง ซึ่งมีความเป็นได้น้อยที่ผู้ใช้จะดูเอกสารจากผลการค้นคืนทั้งหมด ทำให้เอกสารบางส่วนหรือจำนวนมากไม่ถูกเปิดอ่าน ซึ่งอาจเป็นการเสียโอกาสสำหรับผู้ใช้ เนื่องจากอาจมีเอกสารที่ตรงกับความต้องการของผู้ใช้จากเอกสารเหล่านั้น เนื่องจากเอกสารที่ค้นคืนมาได้มีจำนวนมากเกินไป

ปัจจุบันมีการเสนอแนวทางการเพิ่มประสิทธิภาพการค้นหาด้วยวิธีการค้นหาเชิงความหมาย (Semantic search) เพื่อที่จะให้ผู้ใช้ได้ผลลัพธ์จากการค้นหาตรงกับความต้องการมากที่สุด โดยจะมีการจัดเก็บข้อมูลของเอกสารพร้อมทั้งสร้างความสัมพันธ์เชื่อมโยงระหว่างเอกสาร หรือเชื่อมโยงระหว่างคำค้น รวมถึงการจัดทำเป็นหมวดหมู่เพื่อให้ง่ายและสะดวกต่อการเข้าถึงข้อมูล ในการค้นหาเชิงความหมายจะประกอบไปด้วยสองส่วนที่สำคัญ ประการแรก ได้แก่การสร้างฐานความรู้เพื่อช่วยในการค้นหา ซึ่งในปัจจุบันได้มีการจัดทำฐานความรู้ด้วยออนโทโลยี (Ontology) เพื่อนิยามความหมายของข้อมูล ประการที่สอง การหาขั้นตอนวิธีที่เหมาะสมสำหรับการค้นหาข้อมูลจากฐานความรู้ที่ได้จัดทำขึ้น

ออนโทโลยี (Ontology) เป็นการนิยามหรือกำหนดรูปแบบโครงสร้าง (model) ของความสัมพันธ์ระหว่างข้อมูล ภายในขอบเขตขององค์ความรู้เพื่ออธิบายสิ่งที่เราสนใจ (domain) มีการกำหนดขอบเขตความคิดไว้อย่างชัดเจนและมีการยอมรับเพื่อใช้งานร่วมกัน ลักษณะของออนโทโลยี เป็นโครงสร้างลำดับชั้น (Hierarchy) ที่ใช้ในการจำแนกกลุ่มของวัตถุหรือสิ่งที่กำลังสนใจ รองรับมาตรฐานของเอ็กซ์เอ็มแอล (eXtensible Markup Language : XML) ที่ออกแบบมาสำหรับการจัดเก็บข้อมูลบนเว็บที่มีวัตถุประสงค์ในการแลกเปลี่ยนข้อมูลระหว่างกัน ออนโทโลยีจึงเป็นแนวทางหนึ่งที่ช่วยในการจัดการฐานความรู้และถูกนำมาประยุกต์ใช้กับงานระบบต่าง ๆ ไม่ว่าจะเป็นระบบข้อมูลทางด้านชีววิทยา ระบบพาณิชย์อิเล็กทรอนิกส์ รวมไปถึงระบบงานต่าง ๆ ที่พัฒนาขึ้นบนเว็บเชิงความหมายเพื่อช่วยในการจัดเก็บและค้นคืน การแลกเปลี่ยน และการนำกลับมาใช้ใหม่ของข้อมูล สารสนเทศ หรือความรู้เหล่านั้น

เนื่องจากประเทศไทยเป็นประเทศที่มีภูมิประเทศและภูมิอากาศที่เหมาะแก่การเพาะปลูก จึงทำให้ประเทศไทยเป็นแหล่งปลูกผลไม้และพืชผลทางการเกษตรหลายชนิด ในการผลิตผลไม้ถือได้ว่าประเทศไทยเป็นประเทศที่มีสภาพพื้นที่และภูมิอากาศที่เหมาะสมต่อการผลิตผลไม้เมืองร้อนหลากหลายชนิดตั้งแต่ภาคเหนือจนถึงภาคใต้ ส่วนฤดูกาลให้ผลผลิตผลไม้แต่ละชนิดก็ยังแตกต่างกันไปในแต่ละสภาพพื้นที่ของแต่ละภาค จึงเป็นข้อดีประการหนึ่งที่ส่งผลให้ไทยมีผลไม้หลากหลายชนิดหมุนเวียนออกสู่ตลาดตลอดทั้งปี ทำให้มีความได้เปรียบในเรื่องของความยืดหยุ่นในการส่งออก ซึ่งผลไม้นับเป็นพืชเศรษฐกิจที่สำคัญประเภทหนึ่งของประเทศไทย ที่นำรายได้เข้าประเทศปีละหลายพันล้านบาท สร้างรายได้ให้กับเกษตรกรผู้เพาะปลูกและผู้ส่งออกของไทยได้เป็นอย่างดี และผลไม้ไทยยังเป็นที่นิยมบริโภคกันทั่วไปทั้งในประเทศและต่างประเทศ นอกจากนี้ ความต้องการบริโภคผลไม้นับวันจะเพิ่มสูงขึ้น

มหาวิทยาลัยเทคโนโลยีสุรนารี

เนื่องมาจากสาเหตุหลักคือจำนวนประชากรที่เพิ่มมากขึ้นและความสนใจในสุขภาพก็มีมากขึ้นด้วย ดังนั้นการเพาะปลูกจึงไม่ได้มุ่งเพื่อการบริโภคอย่างเดียว แต่มุ่งเพื่อส่งออกต่างประเทศอีกด้วย

ประเทศไทยมีความได้เปรียบโดยเปรียบเทียบจากการส่งออกผลไม้และเป็นประเทศผู้นำในการส่งออกผลไม้ในตลาดโลกและตลาดอาเซียน เนื่องจากผลไม้ประเทศไทยเป็นที่ยอมรับของตลาด โดยมีการศึกษาศักยภาพการผลิตและการตลาดผลไม้ของประเทศไทย พบว่า ประเทศไทยมีความได้เปรียบโดยเปรียบเทียบจากการแข่งขันส่งออกผลไม้ไปทั้งในตลาดโลกและตลาดอาเซียนในช่วง 7 ปี (2547 – 2554) และยังคงมีศักยภาพในการส่งออกอยู่ในระดับสูงและมากกว่าประเทศอื่น ๆ ได้แก่ ประเทศเวียดนาม ซึ่งสะท้อนให้เห็นว่าประเทศไทยเป็นผู้นำในการส่งออกผลไม้ทั้งในตลาดโลกและตลาดอาเซียน ซึ่งในปี 2554 ประเทศไทยมีมูลค่าการส่งออกสินค้าทั้งหมด 6,882,642 ล้านบาท เพิ่มขึ้นจาก 6,176,170 ล้านบาทของปี 2553 หรือเพิ่มขึ้นร้อยละ 11.44 เป็นมูลค่าสินค้าเกษตรและผลิตภัณฑ์ 1,447,716 ล้านบาท เพิ่มขึ้นจาก 1,135,750 ล้านบาทของปี2553 หรือร้อยละ 27.47 เป็นผลไม้จำนวน 81,513 ล้านบาท (สำนักงานเศรษฐกิจการเกษตร, 2555) และในปี 2556 ที่ผ่านมา มีการส่งออกผลไม้ไปยังต่างประเทศ ได้แก่ ลำไย เป็นมูลค่า 8,503 ล้านบาท ทุเรียน 7,344 ล้านบาท และมังคุด 4,296 ล้านบาท (สำนักงานเศรษฐกิจการเกษตร, 2556) ซึ่งลำไย ทุเรียน และมังคุด ถือเป็นผลไม้เศรษฐกิจที่ทำรายได้ให้กับประเทศไทยเป็นอย่างมาก

จากภาพรวมของรายได้ที่ได้จากการส่งออกผลไม้ไทย พบว่าประเทศไทยมีความได้เปรียบในเรื่องของการแข่งขันที่จะส่งผลไม้ไทยออกสู่ตลาดโลก ซึ่งทำรายได้ให้กับประเทศอย่างมากมาย เนื่องจากมีสภาพภูมิอากาศและพื้นที่ที่เหมาะสมแก่การเพาะปลูกพืชผลทางการเกษตร และยังทำให้เกษตรกรผู้เพาะปลูกมีรายได้จากการส่งผลไม้ออกสู่ตลาดอีกด้วย ดังนั้นจะเห็นได้ว่าผลไม้มีความสำคัญต่อเศรษฐกิจของไทย ในการค้นหาความรู้เกี่ยวกับผลไม้เศรษฐกิจของไทยสามารถค้นหาได้จากระบบเครือข่ายอินเตอร์เน็ต แต่ปัญหาที่สำคัญ คือการที่จะเข้าถึงความรู้ที่เป็นประโยชน์เหล่านั้นได้อย่างรวดเร็วและถูกต้องได้อย่างไร ดังนั้น การที่มีระบบที่สามารถสนับสนุนการค้นหาความรู้ได้ง่าย สะดวก รวดเร็ว และตรงต่อความต้องการของผู้ใช้ จึงเป็นสิ่งจำเป็น ด้วยเหตุผลที่ว่า จะสนับสนุนให้สามารถนำความรู้เกี่ยวกับผลไม้เศรษฐกิจของไทยมาใช้ประโยชน์ในภาคเกษตรกรรมซึ่งเป็นภาคที่มีความสำคัญต่อการพัฒนาประเทศไทยไปสู่ความยั่งยืนมากขึ้น

ในการวิจัยนี้ มุ่งแก้ปัญหาการค้นหาความรู้เกี่ยวกับผลไม้เศรษฐกิจของไทยที่อยู่กระจัดกระจายบนระบบเครือข่ายอินเตอร์เน็ต เน้นกระบวนการวิจัยเกี่ยวกับการถามตอบเชิงความหมายสำหรับผลไม้เศรษฐกิจของไทย โดยที่ผู้ใช้สามารถป้อนคำถามที่ต้องการ แล้วระบบจะทำการวิเคราะห์คำถามเพื่อค้นหาคำตอบ หากพบว่าคำถามยังไม่ชัดเจน ก็จะมีการถามกลับไปยังผู้ถามเพื่อให้ได้คำตอบที่ชัดเจนแล้วจึงนำเสนอคำตอบออกมา ระบบนี้จะสนับสนุนให้ผู้ใช้สามารถได้รับคำตอบที่ตรงหรือใกล้เคียงกับความต้องการของผู้ถามมากขึ้น อันจะเป็นประโยชน์ต่อนักวิจัย เกษตรกร หรือผู้ที่สนใจ

1.2 วัตถุประสงค์ของการวิจัย

1.2.1 เพื่อออกแบบกระบวนการถามตอบเชิงความหมายออนไลน์สำหรับผลไม้เศรษฐกิจของไทย

1.2.2 เพื่อพัฒนา และประเมินระบบถามตอบเชิงความหมายออนไลน์สำหรับผลไม้เศรษฐกิจของไทย

1.3 สมมติฐานการวิจัย

1.3.1 ผลการประเมินประสิทธภาพของระบบถามตอบเชิงความหมายออนไลน์สำหรับความรู้ด้านไม้ผลเศษฐกิจของประเทศไทย มีค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และอัตราการรู้จำ (F-measure) ไม่น้อยกว่าร้อยละ 80

1.3.2 การประเมินความพึงพอใจในการใช้งานระบบถามตอบเชิงความหมายออนไลน์สำหรับความรู้ด้านไม้ผลเศรษฐกิจของประเทศไทยโดยผู้ใช้ มีผลประเมินในภาพรวมระดับมากขึ้นไป

1.4 ขอบเขตของการวิจัย

ในการวิจัยครั้งนี้จะทำการออกแบบ พัฒนา และประเมินระบบถามตอบเชิงความหมายออนไลน์สำหรับผลไม้เศรษฐกิจของไทย โดยพัฒนาฐานความรู้ผลไม้เศรษฐกิจของไทยจากเอกสาร หนังสือ และเว็บไซต์ที่เกี่ยวข้อง

สำหรับกลุ่มเป้าหมายในการเข้าถึงระบบนี้ ได้แก่ เกษตรกรไทยผู้ทำการเพาะปลูกผลไม้เศรษฐกิจของไทย ชุมชน นักวิชาการเกษตร รวมถึงประชาชนผู้สนใจทั่วไป ที่สามารถเข้าถึงระบบเครือข่ายอินเตอร์เน็ตและสามารถใช้งานระบบผ่านเว็บเบราว์เซอร์ได้

1.5 ข้อตกลงเบื้องต้น

ในงานวิจัยนี้ผู้วิจัยได้กำหนดข้อตกลงเบื้องต้นของการวิจัย ดังนี้

1.5.1 เอกสารความรู้ผลไม้เศรษฐกิจที่นำมาเป็นตัวอย่างจะค้นหาจากระบบเครือข่ายอินเตอร์เน็ตด้วยเสิร์ชเอ็นจิน แล้วเลือกรายการความรู้ที่มีการเผยแพร่ตั้งแต่ปี พ.ศ. 2550 เป็นต้นมา โดยเลือกรายการที่มีคำอธิบายครบถ้วนตามความต้องการ

1.5.2 ฐานความรู้ผลไม้เศรษฐกิจที่พัฒนาขึ้น จะจัดเก็บเฉพาะคำอธิบายเอกสารความรู้ และลิงก์เชื่อมโยงไปยังเอกสารความรู้เท่านั้น

1.5.3 ระบบถามตอบเชิงความหมายออนไลน์ที่พัฒนาขึ้นจะจัดทำส่วนต่อประสานผู้ใช้เฉพาะส่วนการถามตอบเอกสารความรู้ในฐานความรู้ที่พัฒนาขึ้นมาเท่านั้น โดยผลลัพธ์ที่แสดงผ่านเว็บ

เบราว์เซอร์ ที่ประกอบด้วย ชื่อเอกสารความรู้ ชื่อผู้เขียน ลิงก์เชื่อมโยงไปยังเอกสาร รายละเอียดโดยย่อ และจัดลำดับแสดงรายการผลลัพธ์

1.6 ประโยชน์ที่ได้รับจากการวิจัย

1.6.1 ได้ฐานความรู้ผลไม้เศรษฐกิจของไทย ที่เป็นประโยชน์ต่อผู้ใช้ในทุกกลุ่มที่เกี่ยวข้องที่สามารถเข้าถึงความรู้ต่าง ๆ ได้อย่างสะดวกรวดเร็ว และตรงกับความต้องการ

1.6.2 ได้กระบวนการที่เหมาะสมสำหรับการค้นหาความรู้ผลไม้เศรษฐกิจของไทย

1.6.3 ได้ระบบถามตอบเชิงความหมายออนไลน์สำหรับผลไม้เศรษฐกิจของไทยเพื่อเป็นประโยชน์ต่อประชาชน ชุมชน ที่จะนำความรู้ที่ได้นำไปประยุกต์ใช้

1.7 คำอธิบายศัพท์

1.7.1 ระบบถามตอบเชิงความหมายออนไลน์ (Online semantic question answering system) หมายถึง ระบบที่มีจุดประสงค์เพื่ออำนวยความสะดวกให้ผู้ใช้ได้เข้าถึงสารสนเทศที่เป็นข้อความที่อยู่ในรูปแบบอิเล็กทรอนิกส์ โดยจะมีการพัฒนาให้มีการปฏิสัมพันธ์ระหว่างผู้ใช้กับระบบแบบเป็นการถามตอบเพื่อให้ได้มาซึ่งเอกสารความรู้ที่ต้องการ

1.7.2 ผลไม้เศรษฐกิจของไทย (Thai economic fruits) หมายถึง ผลไม้ที่ทำรายได้เข้าประเทศในสามลำดับแรกในปี 2556 ได้แก่ ลำไย ทุเรียน และมังคุด

มหาวิทยาลัยเทคโนโลยีสุรนารี

บทที่ 2 ปริทัศน์วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

การวิจัยในครั้งนี้เน้นการออกแบบ พัฒนา ทดสอบ และประเมินกระบวนการถามตอบเชิงความหมายออนไลน์สำหรับผลไม้เศรษฐกิจของไทย โดยผู้วิจัยได้ศึกษาเอกสาร และงานวิจัยที่เกี่ยวข้องตามลำดับ ดังนี้

2.1 ระบบถามตอบผ่านเว็บ

2.2 การสกัดข้อมูล

2.3 การจัดหมวดหมู่

2.4 การค้นคืนสารสนเทศ

2.5 การประมวลผลภาษาธรรมชาติ

2.6 การจัดการฐานความรู้ด้วยออนโทโลยี

2.7 ผลไม้เศรษฐกิจของไทย

2.8 เอกสารและงานวิจัยที่เกี่ยวข้อง

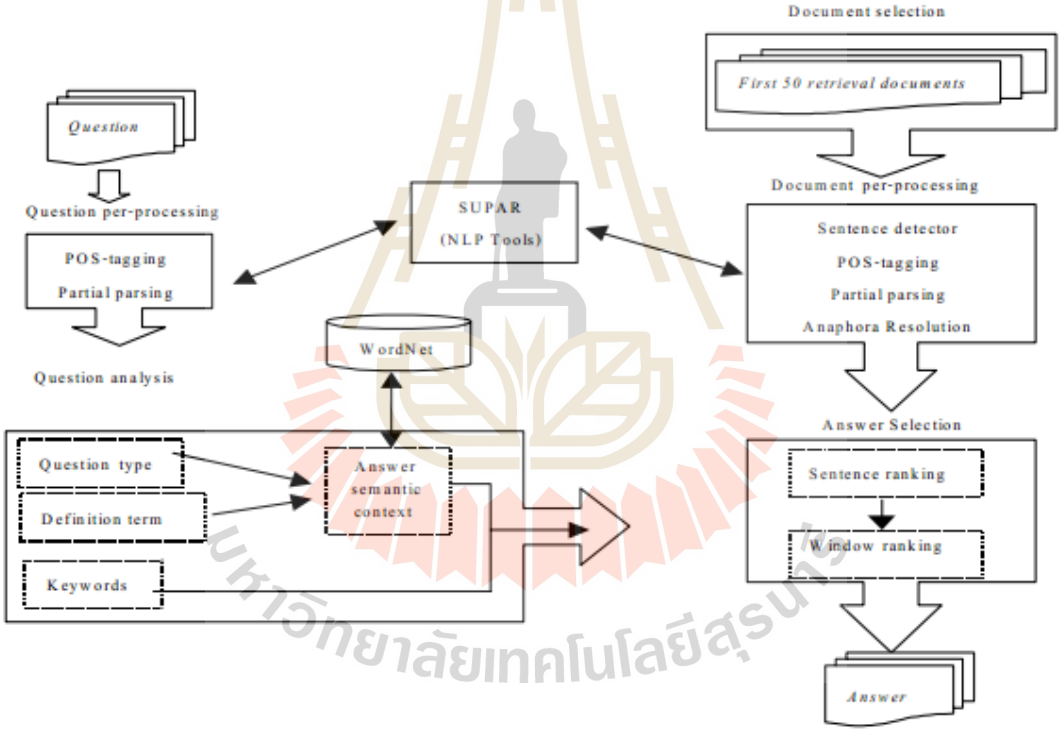
2.9 กรอบแนวคิดการวิจัย

2.1 ระบบถามตอบผ่านเว็บ (Web-Based Question Answering System)

InfoLab Group (1993) ได้อธิบายเกี่ยวกับระบบถามตอบผ่านเว็บเป็นระบบแรกของโลก ได้แก่ START พัฒนาขึ้นในปี 1993 โดยบอริสเททซ์ และคณะ ที่มหาวิทยาลัย Massachusetts Institute of Technology (MIT) ประเทศสหรัฐอเมริกา ซึ่งระบบนี้จะยอมให้ผู้ใช้ป้อนคำถามที่เป็นภาษาอังกฤษ แล้วจะค้นหาคำตอบมาให้ผู้ใช้ตามความรู้ หรือข้อมูลที่มีอยู่ จากการทดสอบการสอบถามข้อมูลพบว่าระบบนี้ยังมีการปรับปรุงข้อมูลให้เป็นปัจจุบัน และในปี 2000 งานวิจัยของ Steven and et al.) ได้นำระบบการประมวลผลภาษาธรรมชาติเข้ามาทำงานร่วมกับระบบการค้นคืนสารสนเทศเพื่อพัฒนาระบบถามตอบ โดยทำการสกัดลักษณะนาม (Entity extraction) ที่ต้องการจากข้อความสั้นที่ถูกดึงมาจากเอกสาร ซึ่งมีการแบ่งประเภทลักษณะนาม (Entity classification) แล้วจัดหมวดหมู่ประเภทของคำนาม และเลือกคำตอบ (Entity ranking) จากหมวดหมู่ของลักษณะนามที่เข้าคู่กับหมวดหมู่ของคำถามที่ต้องการค้นหา ซึ่งการจับเข้าคู่กันใช้เทคนิคแบบบูลีน (Boolean) เช่น ถ้าการจัดหมวดหมู่ลักษณะนามในคำถามเป็นประเภทวันที่ และมีข้อคำถามอย่างน้อย 1 ข้อคำถามที่เป็นลักษณะนามเกี่ยวกับวันที่ที่อยู่ในเอกสาร ดังนั้นหมวดหมู่คำนามในเอกสารที่เกี่ยวกับวันที่จะถูกเลือก

และการเลือกคำตอบจะวิเคราะห์จากความถี่ของคำและตำแหน่งของลักษณะนามที่ปรากฏในข้อความสั้นที่ถูกค้นคืน

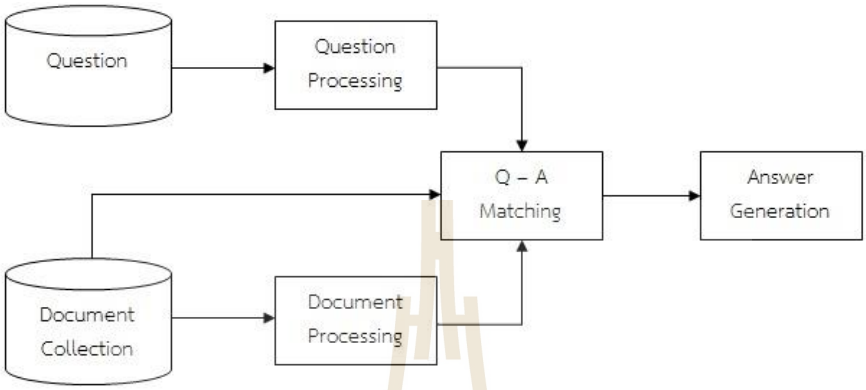
ดังภาพที่ 2.1 Luis and Antonio (2000) ได้นำเสนอการสร้างตัวแทนคำถามจากคำสำคัญ (Keywords) ที่สร้างจากการวิเคราะห์ความหมายที่แตกต่างกันของคำตามลักษณะเฉพาะของคำตอบที่ต้องการ และจัดลำดับคำตอบจากการวัดความคล้ายคลึงระหว่างตัวแทนจากประโยคคำถามกับประโยคที่ได้จากเอกสาร มาใช้กับระบบการประมวลผลภาษาธรรมชาติในระบบถามตอบ กล่าวคือระบบจะสร้างคำที่เป็นตัวแทนการวิเคราะห์ความหมายที่แตกต่างกันของคำ (Semantic context representation) ของคำตอบ ซึ่งมีตัวแทนจากการวิเคราะห์ความหมายที่แตกต่างกันของคำช่วยในการค้นหาคำตอบ กรณีการวิเคราะห์คำที่เป็นคำถาม (Wh-term) ไม่สามารถค้นหาคำตอบได้ และจะช่วยระบบในการวิเคราะห์คำตอบที่แตกต่างกันระหว่างคำตอบที่เป็นไปได้โดยการเปรียบเทียคำตอบที่ได้จากผลลัพธ์ของระบบกับตัวแทนนำมาคำนวณค่าความคล้ายคลึงระหว่างประโยคเพื่อให้ได้คำตอบ



ภาพที่ 2.1 ขั้นตอนการทำงานระบบถามตอบ

นอกจากนี้แล้ว Attardi et al. (2001) ได้นำเสนอระบบถามตอบ PiQASso ซึ่งได้รวบรวมข้อมูลจากหลาย ๆ ห้องสมุด แล้วใช้การเข้าถึงข้อมูลเหล่านั้นด้วยระบบคำถาม-คำตอบ ซึ่งจากผลการดำเนินการ พบว่ามีคำถามที่ไม่สามารถตอบได้มากถึง 49 % และได้นำเสนอให้มีการปรับปรุงการวิเคราะห์คำถาม ด้วยการปรับแต่งอนุกรมวิธาน (Taxonomy) และการใช้การสกัด และขยายคำสำคัญเพื่อช่วยการจำแนกความหมายของคำ

ระบบถามตอบ เป็นระบบที่มีจุดประสงค์ เพื่ออำนวยความสะดวกให้ผู้ใช้ได้เข้าถึงสารสนเทศที่เป็นข้อความที่อยู่ในรูปแบบอิเล็กทรอนิกส์ ซึ่งคำถาม และคำตอบจะเป็นภาษาธรรมชาติ หรือภาษาที่มนุษย์ใช้สื่อสารกัน ซึ่งผลลัพธ์ที่ได้จะเป็นผลลัพธ์ที่กระชับ ที่อาจเป็นคำตอบโดยตรงของคำถามหรือเป็นข้อความสั้น ๆ ที่ถูกดึงออกมาจากเอกสารแต่ไม่ใช่ข้อความทั้งเอกสาร ต่อมา Radev et al. (2005) ได้นำเสนอสถาปัตยกรรมของระบบถามตอบที่ประกอบด้วย 4 ส่วน ดังภาพที่ 2.2



ภาพที่ 2.2 สถาปัตยกรรมของระบบถามตอบ

2.1.1 ระบบถามตอบ

โดยทั่วไประบบถามตอบ มักถูกมองว่าเป็นระบบเดียวกันกับระบบค้นคืนสารสนเทศ และระบบสกัดสารสนเทศ แต่ในความเป็นจริงแล้วแต่ละระบบมีความแตกต่างกันตรงที่ทั้งคำถาม และคำตอบของระบบคำตอบ จะเป็นภาษาธรรมชาติหรือเป็นภาษาที่มนุษย์ใช้ในการสื่อสาร ซึ่งจากสถาปัตยกรรมของระบบคำตอบ ที่ประกอบไปด้วยขั้นตอนต่าง ๆ ที่มีรายละเอียดดังต่อไปนี้

1) การประมวลผลคำถาม (Question Processing)

เป็นขั้นตอนในการวิเคราะห์คำถาม โดยจะมีการสร้างตัวแทนจากการแบ่งประเภทคำถาม ซึ่งประเภทของคำถามอาจได้มาจากการวิเคราะห์คำสำคัญในประโยคของคำถาม การวิเคราะห์จากประเภทของคำตอบที่ต้องการ ความสัมพันธ์ของคำในคำถาม และกฎสำหรับการเลือกคำตอบที่เป็นไปได้ โดยดูจากคำสำคัญที่อยู่ในคำถาม ซึ่งในการแบ่งประเภทคำถาม อาจแบ่งออกเป็นคำถามเกี่ยวกับบุคคล คำถามเกี่ยวกับสถานที่ หรือคำถามเกี่ยวกับเวลา เป็นต้น

2) การประมวลผลเอกสาร (Document Processing)

ในการประมวลผลเอกสารนั้น เป็นการเปลี่ยนแปลงเอกสารที่อยู่ในคลังเอกสารให้อยู่ในรูปแบบที่ทำให้การสืบค้นมีประสิทธิภาพมากขึ้น โดยทั่วไประบบส่วนใหญ่มักจะใช้การสร้างดัชนี (Indexing) เหมือนกับระบบค้นคืนสารสนเทศ

3) การเปรียบเทียบคำถาม-คำตอบ (Question-Answering Matching)

เป็นการเปรียบเทียบระหว่างคำถามที่ผู้ใช้ถาม กับคำถามที่อยู่ในเอกสาร เพื่อให้ได้ชุดคำตอบที่เป็นไปได้สำหรับคำถาม ก่อนที่จะทำการวิเคราะห์คำตอบที่แท้จริงอย่างละเอียด และนำคำตอบที่เป็นไปได้มาตรวจสอบความถูกต้องอีกครั้ง เพื่อความถูกต้องและตรงกับคำถามมากที่สุด

4) การสร้างคำตอบ (Answer Generation)

เมื่อได้คำตอบที่ดีที่สุดสำหรับคำถามแล้ว ซึ่งจะเป็นข้อความสั้น ๆ ที่มีคำตอบที่แท้จริงอยู่ในข้อความนั้น ๆ ทำการสลัดคำตอบเหล่านั้นออกมา และส่งผลลัพธ์ให้กับผู้ใช้

2.1.2 กระบวนการประมวลผลคำถาม

ระบบถามตอบ เป็นระบบที่มีการรับข้อมูล หรือคำถามในรูปแบบภาษาธรรมชาติ ดังนั้นจึงมีความเกี่ยวข้องกับการประมวลผลภาษาธรรมชาติ ซึ่งจะช่วยให้แสดงผลลัพธ์ที่ถูกต้อง และตรงตามความต้องการของผู้ใช้มากที่สุด (Rožanc, 2013) ซึ่งการค้นคืนด้วยภาษาธรรมชาติจะมีประสิทธิภาพได้นั้น จำเป็นต้องมีกระบวนการในการประมวลผลคำถามหรือประโยคที่ผู้ใช้ป้อนเข้ามาในระบบ ซึ่งโดยทั่วไปมักจะใช้กับข้อความ หรือประโยคภาษาไทย ซึ่งส่วนเกี่ยวข้องกับการประมวลผลภาษาธรรมชาติ ได้แก่ การตัดคำภาษาไทย และการแบ่งประโยคภาษาไทย (เกศสุดา ตุริยากรณ์ และชุลีรัตน์ จรัสกุลชัย, 2546) ซึ่งมีรายละเอียดดังต่อไปนี้

1) การตัดคำภาษาไทย

เนื่องจากในภาษาไทย ไม่มีสัญลักษณ์ที่บ่งบอกถึงขอบเขตของคำ ซึ่งแตกต่างจากภาษาอังกฤษที่ใช้การเว้นวรรคในการแบ่งคำแต่ละคำออกจากกัน ซึ่งระบบถามตอบ จะมีการประมวลผลคำทั้งในส่วนคำถาม และคำตอบ ที่ต้องมีการอาศัยเครื่องมือในการตัดคำ โดยในการตัดคำสามารถแบ่งออกตามลักษณะของฐานข้อมูลได้ 3 วิธี ได้แก่ การตัดคำโดยใช้กฎ (Rule-Bases Approach) การตัดคำโดยใช้พจนานุกรม (Dictionary-Based Approach) และการตัดคำโดยใช้คลังข้อมูล (Corpus-Based Approach) (กานดา รุณนะพงศา และปโยธร อุราธรรมกุล, 2549) ซึ่งแต่ละวิธีมีรายละเอียดดังต่อไปนี้

1.1) การตัดคำโดยใช้กฎ (Rule-Bases Approach) การตัดคำโดยใช้กฎ เป็นการตัดคำโดยการตรวจสอบจากกฎเกณฑ์ทางอักขรวิธีที่สร้างขึ้นมา โดยอาศัยหลักไวยากรณ์ภาษาไทยเริ่มจากการพัฒนาการตัดพยางค์ก่อน เนื่องจากพยางค์มีรูปแบบ และกฎเกณฑ์ที่แน่นอนมากกว่าคำ จากนั้นจึงนำมาเป็นเกณฑ์ในการกำหนดขอบเขตของคำ วิธีการนี้จะมีข้อจำกัด คือผลของการตัดคำอาจได้เป็นกลุ่มคำที่สามารถตัดคำได้อีก ความถูกต้องของการตัดคำที่ได้จะมีค่าต่ำ เนื่องจากภาษาไทยมีกฎเกณฑ์ที่ไม่แน่นอน แต่มีข้อดี คือมีการทำงานที่รวดเร็วเพราะใช้ทรัพยากรในการประมวลผลน้อย

1.2) การตัดคำโดยใช้พจนานุกรม (Dictionary-Based Approach) การตัดคำโดยใช้พจนานุกรม เป็นการตัดคำโดยใช้การเปรียบเทียบคำกับคำที่จัดเก็บอยู่ในพจนานุกรมร่วมกับการใช้กฎในการตัดคำด้วย ซึ่งเป็นการแก้ปัญหาการตัดคำของการใช้กฎ เนื่องจากการใช้กฎเพียงอย่างเดียว

มหาวิทยาลัยเทคโนโลยีสุรนารี

ไม่สามารถหาขอบเขตของคำได้ถูกต้องทั้งหมด โดยวิธีการนี้จะต้องมีการจัดเก็บคำทั้งหมดไว้ในพจนานุกรม แล้วนำข้อความที่ป้อนเข้ามาไปค้นหา และเปรียบเทียบสายอักขระกับคำในพจนานุกรม ปัญหาที่พบในวิธีนี้ คือไม่สามารถจัดเก็บคำทั้งหมดลงในพจนานุกรมได้ เนื่องจากมีคำใหม่ ๆ เกิดขึ้นอยู่ตลอดเวลา ดังนั้นความถูกต้องของวิธีนี้จึงขึ้นอยู่กับปริมาณของคำในพจนานุกรม และต้องสูญเสียพื้นที่ในการจัดเก็บคำตามปริมาณคำศัพท์ในพจนานุกรมด้วย วิธีการตัดคำโดยพจนานุกรมนี้ สามารถแบ่งออกได้อีก 2 วิธี คือ

1.2.1) วิธีการเทียบคำที่ยาวที่สุด (Longest Matching) โดยวิธีการเทียบคำที่ยาวที่สุดนี้ จะทำการเปรียบเทียบคำที่ป้อนเข้ามากับคำที่อยู่ในพจนานุกรม ถ้าไม่พบคำที่สามารถเทียบกับคำที่อยู่ในพจนานุกรมได้ ระบบจะทำการลดความยาวของข้อความลงทีละตัวอักษร ตามหลักทางอักขรวิธี จนสามารถที่จะเทียบคำกับคำในพจนานุกรมได้ ซึ่งวิธีการนี้การตัดคำมีความถูกต้องประมาณ 80% แต่มีข้อด้อย คือการเทียบกับคำแรกที่มีความยาวมากเกินไป ทำให้การตัดคำที่มีตามมามีความผิดพลาดได้ เนื่องจากคำในภาษาไทยบางคำเป็นคำประสมที่เกิดจากหลายคำรวมกัน

1.2.2) การตัดคำให้ได้จำนวนคำและคำที่ไม่พบในพจนานุกรมน้อยที่สุด (Maximal Matching) เป็นวิธีที่ใช้ในการแก้ปัญหาจากการตัดคำ ด้วยวิธีการเทียบคำที่ยาวที่สุด ที่เกิดความผิดพลาดจากการเทียบกับคำที่มีความยาวมากเกินไป วิธีการนี้จะใช้การตัดคำโดยวิธีการเทียบคำที่ยาวที่สุดก่อนจากนั้นจะทำการย้อนกลับ (Backtracking) เพื่อทำการตัดคำที่สามารถเป็นไปได้อีกครั้ง โดยใช้วิธีพิจารณาทางเลือกของการตัดคำที่เป็นไปได้ทั้งหมดก่อนที่จะเลือกการตัดคำ และผลลัพธ์ของการตัดคำจะเลือกจากการตัดคำที่มีต้นทุนน้อยที่สุด และได้คำที่ไม่พบในพจนานุกรมให้น้อยที่สุด

1.3) การตัดคำโดยใช้คลังข้อมูล (Corpus-Based Approach) การตัดคำโดยใช้คลังข้อมูล เป็นการตัดคำที่นำวิธีการทางสถิติ หรือวิธีการกลไกการเรียนรู้เข้ามาใช้ในกระบวนการประมวลผลทางภาษา โดยใช้คลังข้อมูลทางภาษาที่ใช้แรงงานคน ในการเตรียมข้อมูลมาเป็นฐานความรู้ที่จะใช้สำหรับการตัดคำ ซึ่งการตัดคำโดยใช้คลังข้อมูลสามารถแบ่งได้ 2 วิธี คือ

1.3.1) วิธีการตัดคำที่อาศัยค่าความน่าจะเป็น เป็นการตัดคำโดยใช้แบบจำลองไตรแกรมกำกับหน้าที่ของคำ (Part-Of-Speech Trigram Model) ในการหารูปแบบของการตัดคำ และลำดับหมวดคำ (Tag Sequence) ที่เป็นไปได้มากที่สุด ซึ่งวิธีการนี้จะต้องมีการใช้คลังข้อมูลที่มีการตัดคำ และการกำกับหน้าที่ของคำเตรียมไว้ก่อนแล้ว ผลลัพธ์ของวิธีการนี้จะเป็นการเลือกรูปแบบการตัดคำที่มีค่าความน่าจะเป็นมากที่สุด

1.3.2) วิธีการตัดคำที่อาศัยคุณลักษณะของคำ เป็นวิธีการที่แก้ไขข้อผิดพลาดของการตัดคำที่อาศัยค่าความน่าจะเป็นของการจัดประเภทของคำ ที่จะนำมาเป็นแบบจำลองในการตัดคำ ซึ่งวิธีการนี้เป็นวิธีการแบบผสม (Hybrid Approach) โดยนำคุณลักษณะของคำที่มีความกำกวมในการตัดคำ มาช่วยเลือกการตัดคำที่ถูกต้อง หลังจากที่มีการตัดคำจากวิธีการตัดคำ

ให้ได้จำนวนคำ และคำที่ไม่มีในพจนานุกรมน้อยที่สุด โดยการสร้างแบบจำลองอาจจะใช้กลไกการเรียนรู้

2) การแบ่งประโยคภาษาไทย

รูปแบบการเขียนของภาษาไทยนั้น นอกจากจะไม่มีสัญลักษณ์บ่งบอกขอบเขตของแต่ละคำแล้ว ก็ยังไม่มีสัญลักษณ์ที่สามารถบ่งบอกถึงขอบเขตของประโยคได้อย่างชัดเจน ซึ่งวิธีการแบ่งประโยคภาษาไทย ได้แก่ การใช้เทคนิคกลไกการเรียนรู้ เรียกว่า วินโดว์ (Pisan & Virah) โดยวิธีการนี้จะใช้ประโยชน์ของช่องว่างในการระบุจุดสิ้นสุดของแต่ละประโยค ซึ่งมีวิธีการคือ จะทำการสกัดประโยคจากข้อความย่อหน้า โดยการตรวจสอบช่องว่างที่มีความเหมาะสมที่จะเป็นตัวแบ่งประโยคด้วยการพิจารณาบริบทรอบ ๆ ช่องว่างนั้น เพื่อระบุว่าช่องว่างนั้น เป็นช่องว่างที่เป็นตัวแบ่งประโยคหรือไม่ ซึ่งคุณลักษณะที่ใช้สำหรับการเรียนรู้นั้นจะใช้ทั้งคำ และหน้าที่ของคำที่อยู่รอบข้างทางด้านซ้ายของช่องว่างจำนวน 2 คำ และทางด้านขวาของช่องว่างจำนวน 2 คำ นอกจากนั้นยังใช้จำนวนคำที่อยู่ทางซ้าย และทางขวาของช่องว่าง รวมทั้งหมดจะใช้คุณลักษณะจำนวน 10 คุณลักษณะด้วยกัน

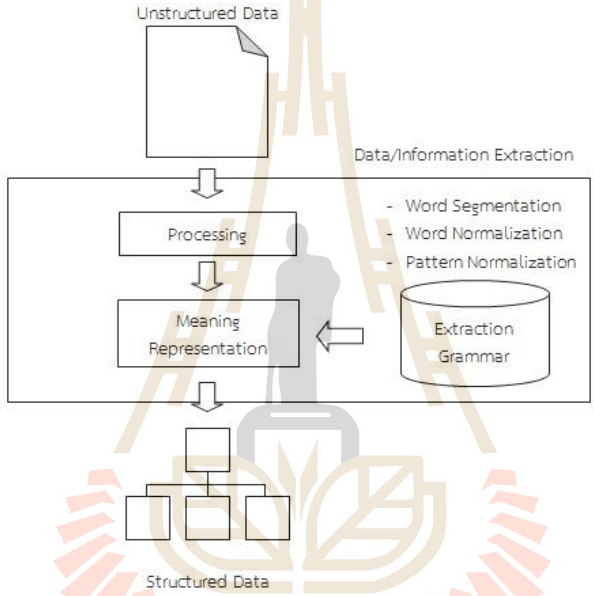
2.2 การสกัดข้อมูล (Data Extraction)

YongYue and HuoSong (2009) ได้อธิบายเกี่ยวกับการสกัดข้อมูลเป็นการสกัดความหมายออกมาจากเอกสารที่ไม่มีโครงสร้าง ดังเช่น ข้อความภาษาธรรมชาติ เพื่อนำข้อมูลดังกล่าวนำไปประมวลผลต่อไป วิธีในการสกัดข้อมูลอาจเป็นการสกัดเพื่อให้ได้ซึ่ง ชื่อ นามวลี บทบาทเชิงความหมาย ความสัมพันธ์ หรือลำดับเวลา ฯลฯ สอดคล้องกับงานวิจัยของ Maedche and Staab (2000) ที่ได้ศึกษาการสกัดข้อมูลหรือสารสนเทศเน้นการวิจัยไปในทิศทางของสกัดข้อมูลแบบอัตโนมัติ ดังเช่น TEXT-Concept Map-Ontology (TEXCOMON) ที่สกัดข้อมูลออกมา เพื่อสร้างโดเมนออนโทโลยีทางด้านหน่วยการเรียนรู้ หรือ TEXT-TO-ONTO ที่สกัดข้อมูลสำหรับสภาวะแวดล้อมเกี่ยวกับการเรียน สำหรับในงานวิจัยนี้จะใช้เทคนิคการสกัดข้อมูล เพื่อวิเคราะห์คำถามจากผู้ใช้ โดยทั่วไปแหล่งที่มาของข้อมูลที่ไม่มีโครงสร้าง ได้แก่ เว็บไซต์ อีเมล์ เอกสาร และไฟล์ PDF

การสกัดข้อมูลนั้น รวมไปถึงการสกัดความรู้ (Knowledge Extraction) ด้วย ซึ่งเป็นการค้นหาความรู้ที่เฉพาะเจาะจงจากคลังข้อความภาษาธรรมชาติ หรือคลังข้อความที่ไม่มีโครงสร้าง โดยการวิเคราะห์ข้อมูล เพื่อหาความสัมพันธ์ของข้อมูล เพื่อให้ได้ความรู้ในเรื่องนั้น ๆ และนำความรู้ที่สกัดได้ไปใช้ประโยชน์ การสกัดข้อมูลเป็นกระบวนการที่ซับซ้อน ที่ต้องมีการวิเคราะห์และดึงข้อมูลที่เกี่ยวข้องจากแหล่งข้อมูล เช่น ฐานข้อมูล ในรูปแบบที่เฉพาะเจาะจง ส่วนใหญ่ของการสกัดข้อมูลมักมาจากแหล่งข้อมูลที่ไม่มีโครงสร้าง และรูปแบบข้อมูลที่แตกต่างกัน ข้อมูลที่ไม่มีโครงสร้างนี้จะอยู่รูปแบบต่าง ๆ เช่น ตาราง ดัชนี และข้อมูลที่ใช้เพื่อการการวิเคราะห์ เป็นต้น

YongYue and HuoSong (2009) ได้ศึกษาถึงวิธีการสกัดข้อมูลโดยมีการแบ่งหมวดหมู่ และโครงสร้างของข้อมูลที่มีเนื้อหาเฉพาะเจาะจงเปลี่ยนรูปแบบจากแหล่งข้อมูลที่ไม่มีโครงสร้าง เช่น

ข้อความภาษาธรรมชาติ ให้เป็นข้อมูลที่มีโครงสร้างที่สามารถนำไปใช้ในการประมวลผลด้านข้อมูลต่อไป กระบวนการในการสกัดข้อมูล เริ่มจากการจัดการมูลที่ไม่มีโครงสร้าง โดยการตัดคำ (Word Segmentation) การแปลงรูปแบบคำ (Word Normalization) และการแปลงรูปแบบประโยคให้อยู่ในรูปแบบมาตรฐาน (Pattern Normalization) ซึ่งกระบวนการดังกล่าวจะมีความเกี่ยวข้องกับกระบวนการค้นคืนสารสนเทศ (Information Retrieval) ในส่วนของการตัดคำ ซึ่งจะมีการอธิบายในหัวข้อถัดไป เมื่อมีการจัดการกับข้อมูลที่ไม่มีโครงสร้างแล้ว จะนำผลลัพธ์ที่ได้ไปสกัด โดยใช้ไวยากรณ์สำหรับการแยกข้อมูลออกจากประโยค (Extraction Grammar) เพื่อให้ได้ข้อมูลที่มีโครงสร้างตามรูปแบบที่ต้องการ ทั้งนี้ เพื่อให้ผู้ใช้สามารถค้นคืนเอกสารที่ต้องการได้อีกด้วย (Surendran, 2013) โดยกระบวนการในการรสกัดข้อมูลแสดงดังภาพที่ 2.3



ภาพที่ 2.3 กระบวนการสกัดข้อมูล

Eckerstorfer (2011) ได้อธิบายเกี่ยวกับการสกัดข้อมูลบนเว็บ (Web data extraction) ไว้ว่า ปัจจุบันข้อมูลได้กระจัดกระจายอยู่ในแหล่งข้อมูลที่แตกต่างกัน ซึ่งข้อมูลส่วนใหญ่ถูกจัดเก็บบนเว็บไซต์ต่าง ๆ ซึ่งผู้ใช้สามารถนำไปใช้งานได้เลย เพียงแค่ค้นหาข้อมูลจากคำค้นผ่านการค้นหาข้อมูลบนเว็บไซต์ แต่การที่ผู้ใช้จะค้นคืนเอกสาร หรือข้อมูลที่ต้องการได้นั้นยังมี ข้อกำจัด เนื่องจากผู้ใช้มีความเชี่ยวชาญหรือประสบการณ์ในการค้นคืนเอกสารแตกต่างกัน จึงทำให้ผู้จัดเก็บข้อมูลหรือเอกสารต่าง ๆ ต้องจัดเก็บข้อมูลให้อยู่ในรูปแบบที่จะทำให้ผู้ใช้สามารถค้นคืนเอกสารที่ต้องการได้ง่าย แต่ก็เป็นเรื่องยาก เนื่องจากเอกสารส่วนใหญ่เป็นเอกสารที่ไม่มีโครงสร้าง ดังนั้นจึงต้องมีการสกัดข้อมูล เพื่อให้ผู้ใช้สารมารถดึงข้อมูลที่ต้องการจากเว็บไซต์ได้ เว็บไซต์ที่แตกต่างกัน ข้อมูลจะถูกนำเสนอแตกต่างกัน และโครงสร้างของข้อมูลจะมีการเปลี่ยนแปลงเมื่อถ่ายโอนข้อมูลจากเว็บไซต์หนึ่งไปยังเว็บไซต์อื่น ๆ นอกจากนี้เว็บไซต์ที่ใช้การเข้ารหัสที่แตกต่างกัน และองค์ประกอบ HTML ที่แตกต่างกัน ในการแสดง

เนื้อหา ซึ่งเมื่อมีการสกัดข้อมูลจะทำให้มีความสามารถในการดึงข้อมูลจากแหล่งข้อมูลต่าง ๆ ทั้งแบบมีโครงสร้าง และแบบไม่มีโครงสร้าง

2.3 การจัดหมวดหมู่

การจัดหมวดหมู่การเรียนรู้แบบมีผลเฉลย แบ่งออกเป็น 2 ขั้นตอน คือ การเรียนรู้เพื่อสร้างกลุ่มเอกสารต้นแบบ และแยกหมวดหมู่ของเอกสารที่สนใจ โดยการตรวจสอบหาความคล้ายคลึงกันกับกลุ่มเอกสารต้นแบบ ซึ่งปัจจุบันมีอัลกอลิทึมสำหรับการจัดหมวดที่หลากหลาย เช่น

1) ต้นไม้ตัดสินใจ (Decision Tree) ต้นไม้จะประกอบด้วยโหนดแทนคุณลักษณะ และโหนดล่างสุดแทนหมวดหมู่ การสร้างต้นไม้ตัดสินใจ C4.5 จะใช้ค่าอัตราเกน (Gain Ratio) เพื่อเลือกคุณลักษณะที่จะเป็นรากหรือโหนด ถ้าให้ชุดข้อมูล M ประกอบด้วยค่าที่เป็นไปได้ $\{m_1, m_2,..., m_n\}$ และให้ความน่าจะเป็นที่จะเกิดค่า m_1 มีค่าเท่ากับ $P(m_1)$ จะได้ค่าเกนของ M เขียนแทนด้วย I(M) (Quinlan, 1993)

2) ซัพพอร์ทเวกเตอร์แมชชีน (Support Vector Machine : SVM) ซึ่งนิเวศ จิระวิชิตชัย, ปริญญา สงวนสัตย์ และพยุง มีสัจ. (2553) อธิบายไว้ว่าเป็นแนวคิดที่แบ่งข้อมูลออกเป็น 2 ส่วน โดยสร้างเส้นแบ่งระหว่างกึ่งกลางกลุ่มให้มีระยะห่างจากขอบเขตของสองกลุ่มมากที่สุด ซึ่งจะสร้างฟังก์ชันวัดความคล้ายคลึงที่เรียกว่า เคอร์เนลฟังก์ชัน (Kernel Function) ให้เหมาะสมกับข้อมูลที่มีมิติสูง เช่น กำหนดให้ $(x_i, y_i),...,(x_n,y_n)$ เป็นตัวอย่างที่ใช้สำหรับการสอน n คือ จำนวนข้อมูลตัวอย่าง m คือ จำนวนมิติข้อมูลเข้า และ y คือผลลัพธ์มีค่า +1 หรือ -1 (Joachims, 1998) ดังสมการ

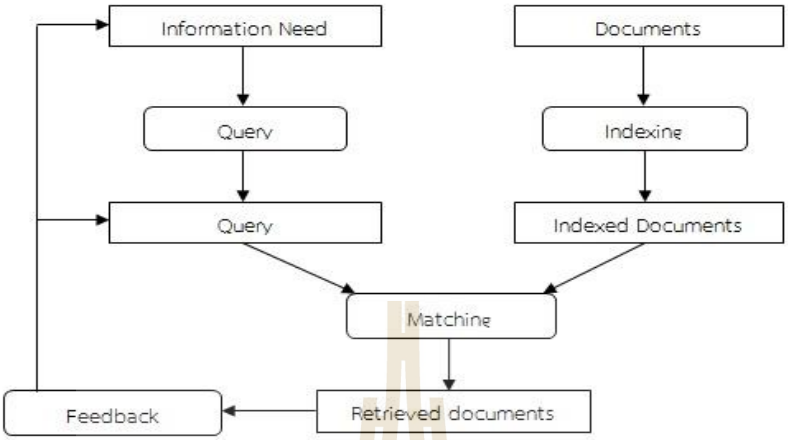
$$(x_i, y_i),...,(x_n,y_n) \text{ เมื่อ } x \in Rm, y \in \{+1, -1\}$$

สำหรับปัญหาเชิงเส้นมิติข้อมูลขนาดสูงได้ถูกแบ่งเป็น 2 กลุ่ม สามารถคำนวณได้จากสมการ $(w*x)+b=0$ เมื่อ w คือค่าน้ำหนัก และ b คือค่า bias สมการใช้สำหรับจำแนกประเภทของข้อมูล คือ $(w*x)+b>0$ ถ้า $y_i = +1$ และ $(w*x)+b<0$ ถ้า $y_i = -1$ ซึ่ง SVM มีเคอร์เนลฟังก์ชัน (Kernel Function) เพื่อให้สามารถประยุกต์การแก้ปัญหา เช่น Linear, Polynomial, Radial Basis Function เป็นต้น

2.4 การค้นคืนสารสนเทศ (Information Retrieval)

Song et al (2011) ได้อธิบายถึงการค้นคืนสารสนเทศ (Information Retrieval) ว่าเป็นกระบวนการที่เกี่ยวข้องกับการค้นหาและดึงสารสนเทศจากแหล่งสารสนเทศ เพื่อให้ผู้ใช้ได้รับสารสนเทศที่ตรงกับความต้องการได้อย่างถูกต้อง รวดเร็ว ซึ่งระบบค้นคืนสารสนเทศในปัจจุบันเป็นที่รู้จัก และใช้กันอย่างแพร่หลาย คือ Search Engine เช่น Google หรือ Yahoo เป็นต้น เป็นระบบค้นคืนสารสนเทศจากคำค้น (Query) ของผู้ใช้ และแสดงเป็นรายการเอกสารที่สัมพันธ์กับคำค้นของผู้ใช้

โดยทั่วไปการค้นคืนสารสนเทศมีกระบวนการพื้นฐานต่าง ๆ แสดงดังภาพที่ 2.4 (ไกรศักดิ์ เกสร, 2554) ดังนี้



ภาพที่ 2.4 กระบวนการค้นคืนสารสนเทศ

ระบบค้นคืนสารสนเทศ คือโปรแกรมที่จัดเก็บ และจัดการสารสนเทศ หรือเอกสารต่าง ๆ ซึ่งมีสารสนเทศที่เป็นทั้งตัวอักษร และสารสนเทศมัลติมีเดีย โดยกระบวนการค้นคืนสารสนเทศ จะเริ่มต้นจากการที่ผู้ใช้เกิดความต้องการสารสนเทศ (Information Need) โดยผู้ใช้จะคิดคำค้นที่จะใช้ในการค้นคืนสารสนเทศที่ต้องการ เป็นกระบวนการในการกำหนดคำค้น หรือข้อคำถาม (Query formulation) ผลลัพธ์ที่ได้จากกระบวนการนี้ คือ ข้อคำถาม (Query) หรือคำค้นที่จะใช้ในการค้นคืนสารสนเทศ เมื่อผู้ใช้นำเอาข้อคำถามไปค้นจากระบบ ระบบจะทำการประมวลผลโดยนำเอาคำสำคัญ (Keyword) ที่อยู่ในคำค้นของผู้ใช้มาเปรียบเทียบกับคำสำคัญ ที่อยู่ในดัชนี (Index) มาเปรียบเทียบและจับคู่คำสำคัญนั้น ๆ เรียกว่ากระบวนการการจับคู่ (Matching) โดยคำสำคัญที่อยู่ในดัชนีนั้นนั้น เกิดจากการนำเอกสารต่าง ๆ (Documents) มาทำดัชนี (Indexing) หรือเรียกว่าเป็นตัวแทนเอกสาร (Representation Document) (Cambridge University, 2008) ซึ่งกระบวนการในการทำดัชนีจะกล่าวถึงในหัวข้อถัดไป โดยหลังจากการทำดัชนีแล้ว จะได้คำสำคัญที่จะถูกใช้เปรียบเทียบกับคำค้นของผู้ใช้ เมื่อมีการเปรียบเทียบคำค้นจากผู้ใช้กับคำสำคัญในดัชนี จะได้เอกสารที่มีความสัมพันธ์กับคำค้นของผู้ใช้ และจะแสดงผลการค้นหาให้กับผู้ใช้ (Cambridge University, 2008) ซึ่งผลการค้นหานั้นอาจเป็นเอกสารที่ผู้ใช้ต้องการ เป็นประโยชน์ต่อผู้ใช้ เรียกเอกสารเหล่านั้นว่า เอกสารที่เกี่ยวข้องกับความต้องการ (Relevant documents) แต่หากผลลัพธ์ที่ได้ไม่ตรงกับความต้องการของผู้ใช้ อาจส่งผลต่อความต้องการสารสนเทศของผู้ใช้ หรือผู้ใช้อาจมีการเปลี่ยนแปลงคำค้น เพื่อให้ตนเองได้ผลลัพธ์ที่ตรงกับความต้องการมากขึ้น ซึ่งระบบค้นคืนที่ดี จะต้องมีการตรวจสอบคำค้น และมีการประมวลผล เพื่อให้

ผู้ใช้ได้สารสนเทศที่เกี่ยวข้อง และตรงกับความต้องการของผู้ใช้มากที่สุด (Apple Computer, Inc., 2005)

2.4.1 กระบวนการค้นคืนสารสนเทศ

การค้นคืนสารสนเทศนั้นมีกระบวนการหลัก ๆ 2 กระบวนการดังที่ได้กล่าวมาแล้ว คือ การทำดัชนี (Indexing) และการเปรียบเทียบ (Matching) ซึ่งมีรายละเอียดดังต่อไปนี้

1) การทำดัชนี (Indexing) คือการสร้างตัวแทนเอกสารต่าง ๆ ในฐานข้อมูล โดยจัดหมวดหมู่อย่างเป็นระบบเรียกว่า ดัชนี (Index) การทำดัชนีข้อมูลเป็นกระบวนการที่มีความสำคัญอย่างยิ่งต่อประสิทธิภาพในการค้นคืนสารสนเทศ เนื่องจากเป็นการเลือกคำ หรือประโยคที่มีความสำคัญของเอกสารนั้น เพื่อแทนข้อมูลทั้งหมดของเอกสาร เป็นการสร้างดัชนี เพื่อใช้สำหรับการสืบค้นนั่นเอง ซึ่งการทำดัชนีข้อมูล มีขั้นตอนดังต่อไปนี้

1.1) การตัดคำ (Tokenizing) คือการตัดคำที่อยู่ในเอกสารออกเป็นคำ ๆ ถ้าเป็นในภาษาอังกฤษจะถือว่าเว้นวรรคเป็นช่องว่างระหว่างคำ ถือว่าเป็นขั้นตอนที่ง่าย แต่ถ้าเป็นภาษาอื่น ๆ เช่นภาษาไทย ที่ไม่มีช่องว่างระหว่างคำ จึงถือว่าเป็นขั้นตอนที่ยากสำหรับภาษาไทย เนื่องจากคำบางคำในภาษาไทยอาจเขียนแบบเดียวกัน แต่อ่านออกเสียงไม่เหมือนกัน หรืออาจมีความหมายต่างกับ เช่น ตาก-ลม หรือ ตา-กลม ดังนั้นหากต้องตัดคำเหล่านี้ จำเป็นต้องดูบริบทของประโยคอื่น ๆ เพื่อจะสร้างดัชนีได้อย่างถูกต้อง ซึ่งแนวทางที่จะนำมาแก้ปัญหานี้ คือการประมวลผลภาษาธรรมชาติ (Natural Language Processing) มาช่วยในการวิเคราะห์คำ

1.2) การตัดคำที่ไม่สำคัญออกไป (Removal of Stop Words) คำที่ไม่สำคัญหรือคำหยุด (Stop Words) คือคำที่มีความถี่สูงในเอกสาร หรือปรากฏอยู่ในเอกสารเป็นจำนวนมาก แต่ไม่มีความหมาย หรือไม่จำเป็นสำหรับการค้นคืนเอกสาร สามารถตัดออกได้โดยที่ไม่ทำให้ประสิทธิภาพในการค้นคืนต่ำลง ได้แก่ คำสรรพนาม คำสันธาน คำบุพบท เช่น ถ้าเป็นคำในภาษาอังกฤษ ได้แก่ a, an, the, in, on

1.3) การเปลี่ยนคำให้อยู่ในรูปแบบดั้งเดิม (Stemming) ถ้าเป็นคำภาษาอังกฤษจะมีการเปลี่ยนจากพหูพจน์ให้เป็นเอกพจน์ เปลี่ยนตัวพิมพ์ใหญ่ให้เป็นพิมพ์เล็ก หรือตัดคำท้ายออก เช่น ing, ed, s, es ออกจากคำ เช่น เปลี่ยนจาก playing เป็น play หรือเปลี่ยนจากรูปอดีตที่เติม ed เป็นรูปปัจจุบันที่ไม่เติม ed เช่น จาก played เป็น play เป็นต้น

ดัชนีเป็นสิ่งจำเป็นสำหรับการค้นคืนสารสนเทศ เนื่องจากดัชนีเป็นตัวชี้ไปยังตัวข้อมูล และแหล่งข้อมูล จะทำให้ผู้ใช้สามารถค้นหาสารสนเทศได้ตรงประเด็น และชี้แหล่งสารสนเทศได้อย่างถูกต้อง รวดเร็ว ประหยัดเวลาและเพิ่มประสิทธิภาพในการค้นคืน (Smith, 2004) ซึ่งการทำดัชนีสามารถใช้เทคนิคต่าง ๆ เช่น การทำดัชนีแบบอินเวิร์ส ขั้นตอนสำคัญของการทำดัชนีแบบอินเวิร์ท คือ การเรียงลำดับข้อมูลซึ่งโดยทั่วไปจะเรียงลำดับตามตัวอักษร คำที่ปรากฏในเอกสารต่างกัน แต่เป็นคำ

เดียวกันจะถือเป็นหนึ่งคำ และมีการนับความถี่ของคำที่ปรากฏอยู่ในทุกเอกสารมารวมกัน โดยขั้นตอนสำคัญของการทำดัชนีแบบอินเวิร์ทนั้น คือการเรียงลำดับตามตัวอักษร คำเดียวที่ปรากฏอยู่ในเอกสารต่างกันจะถือว่าเป็นหนึ่งคำและนำความถี่ของคำที่ปรากฏในเอกสารมารวมกัน ซึ่งการทำดัชนีแบบอินเวิร์ทประกอบด้วยส่วนสำคัญ 2 ส่วน คือ 1) พจนานุกรม (Dictionary) คือส่วนที่ทำหน้าที่ในการบันทึกข้อมูลสถิติ เช่น จำนวนคำที่ปรากฏอยู่ในเอกสาร เรียกว่า ความถี่เอกสาร (Document Frequency) ซึ่งข้อมูลดังกล่าวจะนำไปใช้ในการเรียงลำดับข้อมูลผลลัพธ์ และ 2) การประกาศ (Posting) คือการแสดงรายการเอกสารทั้งหมดที่มีคำที่ผู้ใช้ค้นหาปรากฏอยู่ โดยแสดงข้อมูลในรูปแบบของหมายเลขเอกสาร (Document ID)

การทำดัชนีแบบอินเวิร์ท ถือเป็นวิธีที่มีประสิทธิภาพ สำหรับเอกสารที่เป็นตัวอักษร นอกจากการทำดัชนีแบบอินเวิร์ทแล้วยังมีวิธีการทำดัชนีอื่น ๆ ที่มีประสิทธิภาพ ได้แก่ การทำดัชนีตำแหน่งคำ (Positional Indexing) เป็นต้น แต่อย่างไรก็ตาม ความมีประสิทธิภาพในการค้นคืนสารสนเทศจะเกิดขึ้นได้ก็ต่อเมื่อ ผู้ทำดัชนีเลือกใช้คำที่สอดคล้องกับคำที่ผู้ใช้ค้นมากที่สุด และมีรายการโยงไปยังสารสนเทศที่มีความเกี่ยวข้องสัมพันธ์กัน (Apple Computer, Inc., 2005)

2) การเปรียบเทียบ (Matching)

Apple Computer, Inc (2005) ได้อธิบายถึงการเปรียบเทียบ คือการนำคำที่อยู่ในดัชนีมาเปรียบเทียบกับคำที่อยู่ในคำค้น (Query) ของผู้ใช้ หากคำค้นที่ผู้ใช้ค้นหาตรงกับคำสำคัญที่อยู่ในดัชนี จะมีการจับคู่คำเหล่านั้น มีการเรียงลำดับความคล้ายคลึงระหว่างเอกสาร และผลลัพธ์ และแสดงผลการค้นหาตามคำที่ผู้ใช้ค้นหา

2.4.2 รูปแบบการค้นคืนสารสนเทศ

ปัจจุบันมีสารสนเทศมากมายที่น่าสนใจบนอินเทอร์เน็ต ทำให้ต้องมีการใช้เทคนิคการค้นคืนข้อมูลต่าง ๆ เพื่อความสะดวกรวดเร็วในการค้นหาข้อมูลสำคัญ ๆ ที่ผู้ใช้ต้องการ โดยในการค้นคืนสารสนเทศนั้นมีรูปแบบในการค้นคืนที่แตกต่างกัน วัตถุประสงค์เพื่อให้ผู้ใช้ได้สารสนเทศที่ต้องการอย่างตรงประเด็น และรวดเร็ว ซึ่งในงานวิจัยนี้ได้ยกตัวอย่างรูปแบบการค้นคืนสารสนเทศ ได้แก่ Boolean Model และ Vector Space Model ซึ่งมีรายละเอียดดังต่อไปนี้

1) แบบจำลองบูลีน (Boolean Model)

การค้นคืนโดยใช้วิธีการบูลีน คือการค้นหาโดยใช้เงื่อนไขจากนิพจน์ในทางตรรกศาสตร์ เป็นรูปแบบการค้นคืนเอกสารที่ยังคงนิยมใช้ในปัจจุบัน (ภรศิษฐ์ กุหลาบทิพย์, 2550) มีวัตถุประสงค์เพื่อค้นหาประเด็น และหาเอกสารที่ตรงกับนิพจน์ (Boolean Expression) โดยถือนิพจน์ทางตรรกศาสตร์นั้นคือข้อคำถาม หรือคำค้น (Query) จากผู้ใช้ โดยเอกสารจะถูกแทนด้วยเซตของคำ (Keywords) ซึ่งในการค้นคืนจะประกอบด้วย ข้อคำถามหรือคำค้น (Query) ที่เป็น Boolean Expression ที่เชื่อมโยงด้วยตัวกระทำทางตรรกะ AND OR และ NOT โดยเอกสารที่มีคำสำคัญใน

Query ปรากฏอยู่ จะถูกเลือกมาพิจารณาเป็นผลลัพธ์ในการค้นคืน โดยจะพิจารณาจากเอกสารทั้งหมด และเลือกเอาเอกสารที่ตรงกับนิพจน์ของข้อคำถามเท่านั้น โดยพิจารณาตามลำดับของตัวกระทำ ซึ่งผลของการเปรียบเทียบได้ผลลัพธ์เป็น TRUE หรือ FALSE อย่างใดอย่างหนึ่ง ซึ่งเอกสารที่เป็นผลลัพธ์จะมีค่าเป็นจริง (TRUE) เท่านั้น โดยผู้ใช้จะต้องใช้คำค้นที่ประกอบไปด้วยคำที่ต้องการค้น และตัวกระทำทางบูลีน ได้แก่ AND OR และ NOT ซึ่งตัวกระทำแต่ละตัวมีความหมายดังตารางที่ 2.1

ตารางที่ 2.1 ตัวกระทำทางบูลีน

Boolean	**เครื่องหมาย**	**ความหมาย**
AND	+	AND operator ใช้สัญลักษณ์ในการสืบค้นด้วยเครื่องหมาย + เมื่อใช้ and เป็นตัวเชื่อมในการค้นหา ระบบจะค้นหาเอกสารที่มีคำหรือข้อความทั้งสองปรากฏอยู่ เช่น ทุเรียน + หมอนทอง หมายถึงให้ค้นหาเอกสารที่มีคำว่า “ทุเรียน” และ “หมอนทอง” ปรากฏอยู่ทั้งสองคำ แต่จะปรากฏเพียงคำใดคำหนึ่งไม่ได้
OR	&	OR Operator ใช้สัญลักษณ์ในการสืบค้นด้วยเครื่องหมาย & เมื่อใช้ or เป็นตัวเชื่อในการค้นหา ระบบจะค้นหาเอกสารที่มีคำใดคำหนึ่งหรือข้อความใดข้อความหนึ่งปรากฏอยู่ เช่น ทุเรียน & หมอนทอง หมายถึงให้ค้นหาเอกสารที่มีคำว่า “ทุเรียน” หรือ “หมอนทอง” คำใดคำหนึ่งหรือทั้งสองคำก็ได้
NOT	-	NOT Operator ใช้สัญลักษณ์ในการสืบค้นด้วยเครื่องหมาย - เมื่อใช้ NOT เป็นตัวเชื่อมในการค้นหา ระบบจะค้นหาเอกสารที่เอกสารที่มีคำหรือข้อความแรกปรากฏอยู่เท่านั้น โดยต้องไม่มีคำหรือข้อความที่สองปรากฏอยู่ในเอกสารใด ๆ เช่น ทุเรียน – หมอนทอง หมายถึง ให้ค้นหาเอกสารที่มีคำว่า “ทุเรียน” เพียงคำเดียวเท่านั้น ซึ่งต้องไม่มีคำว่า “หมอนทอง” ปรากฏอยู่

การค้นคืนโดยใช้ Boolean Expression นั้น มีข้อดีคือ มีความแม่นยำหากผู้ใช้มีความรู้หรือมีความชำนาญในการกำหนดคำค้นภายใต้รูปแบบของ Boolean ซึ่งจะทำให้ได้ ผลลัพธ์ที่ตรงกับความต้องการของผู้ใช้ เนื่องจากเอกสารที่ถูกค้นคืน จะแสดงเฉพาะเอกสารที่ตรงกับเงื่อนไขใน Query เท่านั้น แต่อย่างไรก็ตามก็ยังมีข้อจำกัดสำหรับผู้ใช้ที่ไม่ชำนาญหรือไม่สามารถกำหนดคำค้นในแบบ Boolean Expression ได้ จะทำให้ได้ผลลัพธ์ที่ไม่ถูกต้อง ซึ่งเป็นการยากสำหรับผู้ใช้ส่วนใหญ่ นอกจากนี้ เอกสารที่ค้นคืนออกมาได้อาจมีปริมาณที่น้อยเกินไป ทำให้ไม่เพียงพอต่อความต้องการของ

ผู้ใช้ ไม่มีการเรียงลำดับของเอกสาร อีกทั้งยังไม่มีการแสดงรายการเอกสารที่คล้ายกันอีกด้วย ดังนั้นผู้ที่จะใช้รูปแบบการค้นคืนแบบ Boolean จะต้องเป็นผู้ที่มีความรู้ในการกำหนดคำค้นโดยระบบเงื่อนไขทางตรรกศาสตร์ได้เป็นอย่างดี จึงจะสามารถค้นคืนเอกสารที่ต้องการได้

2) แบบจำลองเวกเตอร์สเปซ (Vector Space Model : VSM)

การแทนเอกสารด้วยแบบจำลองเวกเตอร์สเปซ มีการใช้งานอย่างกว้างขวาง เนื่องจากมีความยืดหยุ่น และมีประสิทธิภาพ VSM เป็นการกำหนดให้เอกสารแต่ละฉบับเปรียบเหมือนเวกเตอร์ของคำ โดยขนาดของเวกเตอร์ขึ้นอยู่กับจำนวนของคำ ที่ปรากฏอยู่ในเอกสารฉบับนั้น ๆ ใช้การคำนวณหาค่าความเหมือน หรือเปรียบเทียบข้อมูล โดยใช้วิธีการคำนวณแบบ cosine (ณรงค์พิสิษฐ์ อึ้งรังสี, 2554) ซึ่งผลลัพธ์ที่ได้ จะมีค่าจาก 0 ถึง 1 เมื่อผลลัพธ์ที่ได้มีค่าเข้าใกล้ 0 หมายถึง ข้อมูลที่เปรียบเทียบกันมีความเหมือนกันมาก VSM มีการให้น้ำหนักของคำ (Term Weighting) ซึ่งน้ำหนักของเวกเตอร์ขึ้นอยู่กับจำนวนของคำที่ปรากฏในเอกสารนั้น ๆ (Term Frequency: TF) ถ้าค่า TF มาก หมายถึงมีคำนั้น ๆ ปรากฏในเอกสารมาก และสัดส่วนของจำนวนเอกสารทั้งหมดกับจำนวนเอกสารที่มีคำปรากฏอยู่ในเอกสาร (Inverse Document Frequency: IDF) หรือเรียกว่าค่า TF-IDF (บารมี โอสธีรกุล และธวัชชัย งามสันติวงศ์, 2555)

อย่างไรก็ตาม ถึงแม้ว่าจะมีการใช้แบบจำลองเวกเตอร์สเปซอย่างกว้างขวาง แต่การค้นคืนเอกสารด้วยแบบจำลองเวกเตอร์สเปซก็ยังมีข้อจำกัดอยู่ เนื่องจากว่าในการค้นคืนนั้นมีการคิดแยกคำ ทำให้ไม่สามารถค้นคืนแบบกลุ่มคำหรือลำดับของคำได้ และไม่ได้สนับสนุนการค้นหาแบบการใช้คำที่มีความหมายคล้ายกัน (Synonym) แต่จะสามารถค้นหาความคล้ายกันของเอกสารได้ก็ต่อเมื่อ เอกสารนั้นใช้คำคำเดียวกันในหลาย ๆ เอกสารที่มีความคล้ายกัน แต่การใช้คำค้นต่างกันแบบจำลองเวกเตอร์สเปซจะไม่สามารถหาความคล้ายกันของเอกสารได้ (ณรงค์พิสิษฐ์ อึ้งรังสี, 2554)

2.4.3 การวัดประสิทธิภาพการค้นคืนสารสนเทศ

เป็นการประเมินระบบค้นคืนสารสนเทศ ว่ามีความสามารถในการค้นคืนสารสนเทศได้ตรงตามความต้องการของผู้ใช้ มีความรวดเร็ว ถูกต้องแม่นยำ และครบถ้วนสมบูรณ์หรือไม่ การวัดประสิทธิภาพการค้นคืนมีหลายวิธี ซึ่งวิธีที่นิยมใช้ในการเปรียบเทียบ คือ การวัดค่าความระลึก (Recall) ค่าความแม่นยำ (Precision) และอัตราการรู้จำ (F-Measure) (ภรศิษฐ์ กุหลาบทิพย์, 2550)

1) ค่าความระลึก (Recall)

ค่าความระลึก (Recall) คือค่าที่แสดงให้เห็นว่าผลลัพธ์ที่ค้นออกมามีความถูกต้องเพียงใด โดยคำนวณจากอัตราส่วนของเอกสารทั้งหมด ที่เกี่ยวข้องกับเอกสารทั้งหมดที่มีอยู่ในฐานข้อมูล ดังสมการ

$$Recall = \frac{|Ra|}{|R|}$$

จากสมการ Ra แทนเอกสารทั้งหมดที่เกี่ยวข้อง และ R แทนเอกสารทั้งหมดที่มีอยู่ในฐานข้อมูล โดยค่าความระลึกเป็นการวัดความสามารถของระบบสืบค้น ในการค้นคืนเอกสารทั้งหมดหรือข้อมูลที่เกี่ยวข้องกับคำค้น (Relevance) หากค่า Recall เท่ากับ 1 หรือค่า Ra = R หมายถึง เอกสาร หรือผลลัพธ์ที่ค้นคืนมาได้ทั้งหมดสอดคล้องกับคำค้น หรือข้อคำถาม แต่ถ้าค่า Recall เป็น 0 หรือ Ra = 0 หมายถึง ไม่มีเอกสารหรือผลลัพธ์ที่ถูกค้นคืนออกมาไม่ตรงกับคำค้นหรือข้อคำถาม

2) ค่าความแม่นยำ (Precision)

ค่าความแม่นยำ (Precision) คือค่าที่แสดงถึงการค้นคืนสารสนเทศได้ตรงกับความต้องมากน้อยเพียงใด คำนวณจากอัตราส่วนของเอกสารทั้งหมด ที่เกี่ยวข้องกับจำนวนของเอกสารทั้งหมดที่ถูกค้นออกมาดังสมการ

$$Precision = \frac{|Ra|}{|A|}$$

จากสมการ Ra แทนเอกสารทั้งหมดที่เกี่ยวข้อง และ A แทนเอกสารทั้งหมดที่ถูกค้นออกมา โดยค่าความแม่นยำเป็นการวัดความสามารถของระบบสืบค้น ในการค้นคืนเอกสารหรือข้อมูลได้สอดคล้องกับคำค้นหรือข้อคำถามให้ได้มากที่สุด หากค่า Precision เป็น 1 หรือ Ra = A หมายถึง เอกสารหรือข้อมูลทุกรายการมีความเกี่ยวข้องกัน แต่ถ้าค่า Precision เป็น 0 หรือ Ra = 0 หมายถึง ไม่มีเอกสารหรือข้อมูลที่มีความเกี่ยวข้องกัน

3) อัตราการรู้จำ (F-Measure)

อัตราการรู้จำ (F-Measure) เป็นค่าเฉลี่ยที่ให้ความสำคัญกับความแม่นยำและความระลึกเท่า ๆ กัน โดยรวมค่าความแม่นยำและค่าความระลึกเป็นค่าเดียวกัน โดยคำนวณได้ดังสมการ

$$F\text{-}Measure = \frac{2 \times Recall \times Precision|}{Precision + Recall}$$

2.5 การประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP)

การประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP) เป็นการประมวลผลข้อความที่มาจากภาษาของมนุษย์ให้เป็นภาษาที่เข้าใจง่าย ไม่ซับซ้อน กล่าวคือการทำให้คอมพิวเตอร์เข้าใจภาษาของมนุษย์ มีการนำไปประยุกต์ใช้กับระบบค้นคืนสารสนเทศ (Information Retrieval) ซึ่งการประมวลผลภาษาธรรมชาติจัดอยู่ในระดับต่ำ (Low Level) ของการวิเคราะห์ทางภาษา (Linguistic Analysis) ดังภาพที่ 2.5 (Chopra, Prashar และ Sain, 2013)

ภาพที่ 2.5 ลำดับขั้นของการวิเคราะห์ทางภาษา (Linguistic Analysis)

2.5.1 ระดับของการประมวลผลภาษาธรรมชาติ

ระดับของการประมวลผลภาษาธรรมชาติ สามารถแบ่งออกเป็น 5 ระดับ (Chopra, Prashar และ Sain, 2013) ดังต่อไปนี้

1) Morphological Analysis คือการวิเคราะห์คำในหน่วยคำว่าสามารถแยกออกมาเป็นหน่วยย่อย ๆ อะไรบ้าง โดยระบุและอธิบายโครงสร้างของคำได้ เช่น คำว่า “การทำงาน” สามารถแบ่งเป็นคำที่มีหน่วยย่อย ๆ ได้ คือ “การ”+ “ทำงาน” เป็นต้น

2) Syntactic Analysis คือการวิเคราะห์หลักไวยากรณ์ เป็นการวิเคราะห์เกี่ยวกับการสร้างประโยค เพื่อดูว่าคำที่อยู่ภายในประโยคมีหน้าที่อะไร เช่น คำใดทำหน้าที่เป็นประธาน คำใดทำหน้าที่เป็นกริยา เป็นต้น โดยจะวิเคราะห์ลำดับของคำในรูปแบบของโครงสร้าง เช่น ใช้โครงสร้างต้นไม้เพื่อบอกความสัมพันธ์ของคำต่าง ๆ เป็นต้น ซึ่งในการประมวลผลภาษาธรรมชาติจะใช้หลักการโครงสร้างประโยคดังกล่าวในการประมวลผลแต่ละประโยค จึงจะทำให้คอมพิวเตอร์สามารถตีความและเข้าใจความหมายได้ดียิ่งขึ้น

3) Semantic Analysis คือการวิเคราะห์ทางความหมาย เป็นการวิเคราะห์ความหมายของคำว่าหมายถึงสิ่งใด และหาความหมายของคำเมื่ออยู่ในบริบทที่ต่างกัน เพราะคำบางคำจะมีความหมายต่างกันเมื่ออยู่ในประโยคที่มีบริบทต่าง ๆ กัน หรือเป็นคำที่มีความหมายเดียวกัน เป็นการแปลความหมายของคำโดยพิจารณาจากบริบท

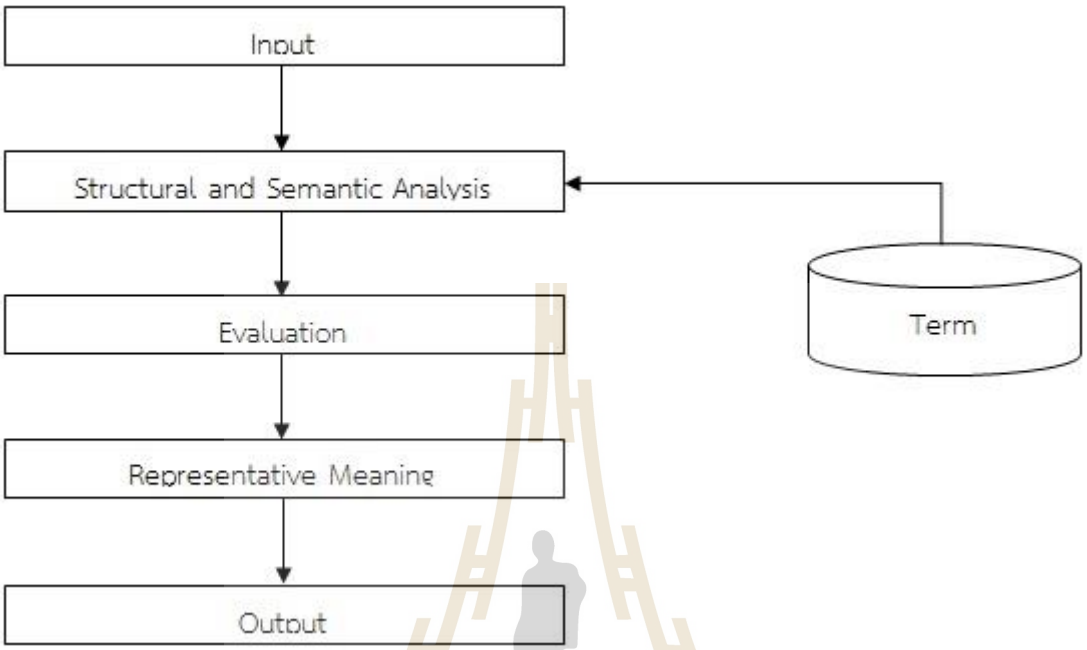
4) Discourse Integration คือการพิจารณาโครงสร้างของประโยคว่าควรจัดอยู่ในประเภทใด และวิเคราะห์โครงสร้างของเอกสาร เช่น การแยกแยะว่าส่วนคือส่วนที่แสดงความคิดเห็น ส่วนใดที่เป็นข้อเท็จจริง เป็นต้น

5) Pragmatic Analysis คือ เป็นการวิเคราะห์และแปลความหมายของประโยคอีกครั้งหนึ่ง เพื่อพิจารณาว่าประโยคหรือคำมีความหมายตรงกับความต้องการในการสื่อความหมายหรือไม่

2.5.2 ขั้นตอนการประมวลผลภาษาธรรมชาติ

ระบบประมวลผลภาษาธรรมชาติประกอบด้วย 2 ส่วน คือ การทำความเข้าใจภาษา (Natural Language Understanding: NLU) และการสร้างภาษา (Natural Language Generation: NLG) ซึ่ง

ทั้ง 2 ส่วนมีลักษณะสมมาตรกัน (Symmetry) แต่กลับทิศทางกัน (Reversibility) (Chopra, Prashar และ Sain, 2013) ซึ่งขั้นตอนของการประมวลผลภาษาธรรมชาติแสดงดังภาพที่ 2.6



ภาพที่ 2.6 ขั้นตอนของการประมวลผลภาษาธรรมชาติ

2.5.3 องค์ประกอบของการประมวลผลภาษาธรรมชาติ

องค์ประกอบของการประมวลผลภาษาธรรมชาติ ประกอบด้วย 5 องค์ประกอบ (Chopra, Prashar และ Sain, 2013) ดังนี้

1) Parser เป็นการทำหน้าที่ในการวิเคราะห์ไวยากรณ์ของประโยคที่ถูกป้อนเข้าสู่ระบบ ซึ่งแบ่งประโยคออกเป็นคำ โดยแต่ละคำจะถูกนำไปจับคู่กับโครงสร้างที่ใช้ในการวิเคราะห์คำที่อยู่ในลักษณะของต้นไม้เรียกว่า พาร์เซอร์ ทรี (Parser Tree)

2) Lexicon เป็นพจนานุกรมที่ใช้ในการวิเคราะห์ความหมายที่มีการสะกดคำที่ถูกต้อง ความหมายและหน้าที่คำนั้น ๆ ในประโยค ในกรณีที่มีมากกว่าหนึ่งหมาย Lexicon จะแสดงความหมายทั้งหมดออกมา

3) Understander เป็นส่วนของการทำความเข้าใจที่ทำงานร่วมกับฐานความรู้ เพื่อกำหนดความหมายของประโยค โดยใช้พาร์เซอร์ ทรี (Parser Tree) ในการค้นหาและอ้างอิงข้อมูลในฐานความรู้

4) Knowledge Base เป็นฐานความรู้ที่เป็นที่จัดเก็บความรู้ต่าง ๆ ที่มีความเกี่ยวข้องกับภาษาธรรมชาติที่มนุษย์ใช้ในการสื่อสาร ฐานความรู้คือสิ่งที่จะช่วยในเรื่องของการกำหนดความหมายของประโยคที่ผู้ใช้ป้อนเข้าสู่ระบบ

5) Generator เป็นการสร้างโครงสร้างเฉพาะของประโยค หลังจากได้ทำการค้นหาและทำความเข้าใจของประโยคแล้ว ซึ่งเป็นโครงสร้างเฉพาะของประโยคที่อยู่ในรูปแบบที่ระบบสามารถใช้แทนความหมายและความเข้าใจของประโยคได้ และถูกจัดเก็บไว้ในหน่วยความจำของระบบโครงสร้างของประโยคที่สร้างขึ้นมาใหม่

2.5.4 เทคนิคการประมวลผลภาษาธรรมชาติ

เทคนิคการประมวลผลภาษาธรรมชาติ ปัจจุบันที่นิยมใช้มีอยู่ด้วยกัน 2 เทคนิคคือ เทคนิคการวิเคราะห์คำหลัก (Keyword Analysis) และเทคนิคการวิเคราะห์ไวยากรณ์ (Syntax Analysis) (Lame, 2005) ซึ่งมีรายละเอียดดังต่อไปนี้

1) เทคนิคการวิเคราะห์คำหลัก (Keyword Analysis) เป็นการวิเคราะห์คำจากประโยคว่าคำใดเป็นคำหลัก โดยเปรียบเทียบกับคำที่จัดเก็บไว้ในฐานความรู้ แต่ถ้าคำหลักที่มีอยู่ไม่ตรงกับความต้องการของผู้ใช้นั้น ผู้ใช้สามารถเพิ่มคำหลักลงในฐานความข้อมูลได้ เพื่อช่วยให้ระบบสามารถวิเคราะห์ได้ตรงกับผู้ใช้มากที่สุด โดยขั้นตอนพื้นฐานของกระบวนการวิเคราะห์โดยใช้คำหลักในการประมวลผลภาษาธรรมชาติเกี่ยวข้องกับขั้นตอนการแบ่งคำแต่ละคำในประโยค(Parsing) และขั้นตอนการจับคู่รูปแบบของคำ (Pattern Matching)

2) เทคนิคการวิเคราะห์ไวยากรณ์ (Syntax Analysis) เป็นการวิเคราะห์ ไวยากรณ์ความหมายและความสัมพันธ์ระหว่างคำในประโยคที่ผู้ใช้ป้อนเข้าสู่ระบบอย่างละเอียด เพื่อนำไปประมวลผลหาผลลัพธ์มาแสดงให้กับผู้ใช้ ซึ่งเทคนิคนี้มีการวิเคราะห์รูปประโยควิเคราะห์ความหมายของประโยค หาความหมายของคำในประโยค วิเคราะห์หาความหมายที่แท้จริงของประโยค และทดลองเชื่อมประโยคหลายประโยคเข้าด้วยกัน

2.6 การจัดการฐานความรู้ด้วยออนโทโลยี

ออนโทโลยี (Ontology) คือการอธิบายรูปแบบโครงสร้างความสัมพันธ์ระหว่างข้อมูลในขอบเขตที่สนใจ (Domain) ให้ได้ใจความและมีความถูกต้องมากที่สุด มีการระบุแนวคิดที่เป็นทางการอย่างชัดเจน ภายใต้การยอมรับร่วมกันของบุคคลที่เกี่ยวข้อง (Expert Consensus) ออนโทโลยี เป็นเทคโนโลยีทางด้านการพัฒนาภาษาเชิงความหมายของภาษาสมัยใหม่ โดยเป็นภาษาที่เครื่องคอมพิวเตอร์สามารถเข้าใจ ตีความ และทำตามคำสั่งได้ ซึ่งมีผู้ให้คำนิยามของออนโทโลยีไว้หลายแบบดังนี้

ออนโทโลยี คือรูปแบบอย่างเป็นทางการที่ชัดเจน และมีการแบ่งปันคุณสมบัติของแนวความคิดที่ใช้ร่วมกัน (Gruber, 1993) เป็นชุดของสิ่งที่มีโครงสร้างเป็นลำดับขั้นของข้อตกลงที่จะอธิบายโดเมนที่สามารถนำมาใช้เป็นรากฐานโครงร่างสำหรับฐานความรู้ เป็นแนวคิดที่ว่าด้วยเรื่องคุณสมบัติที่เป็นที่รู้จักและยอมรับร่วมกัน ในขณะที่มีการใช้คำศัพท์ที่เฉพาะเจาะจงในการอธิบายเอ็นทิตี้ (Entity) คลาส (Class) คุณสมบัติ (Property) และความสัมพันธ์ (Relationship) ที่เกี่ยวข้องกับโดเมนที่ให้ความสนใจ ซึ่งจำเป็นต้องมีความเฉพาะเจาะจงของคำที่ใช้ (คำศัพท์) และข้อตกลงที่จะกำหนดความหมายของคำเหล่านั้นพร้อมความสัมพันธ์ระหว่างคำ

นอกจากนี้ Fensel (2004) ได้ให้คำนิยามของออนโทโลยีว่ามี 4 แนวคิดหลัก ๆ ดังนี้

1. แนวความคิด คือรูปแบบนามธรรมของปรากฏการณ์
2. เป็นทางการ คือคำอธิบายทางคณิตศาสตร์ที่แม่นยำ
3. ชัดเจน คือการกำหนดขอบเขตและความสัมพันธ์ของสิ่งต่าง ๆ ได้อย่างชัดเจน
4. ร่วมกัน คือการยอมรับร่วมกันของบุคคลต่าง ๆ ที่เกี่ยวข้องกับกำหนดขอบเขต และความสัมพันธ์ของสิ่งต่าง ๆ

ออนโทโลยี คือการจัดการองค์ความรู้ หรือฐานความรู้ให้อยู่ในรูปแบบรงสร้างลำดับชั้น (Hierarchy) คล้ายกับโครงสร้างต้นไม้ (Tree) เป็นโครงสร้างที่อำนวยความสะดวกให้กับกระบวนการค้นหาสารสนเทศเชิงความหมายให้มีประสิทธิภาพมากขึ้น โดยอธิบายได้ดังนี้

1. แนวคิด (Concept) คือขอบเขตของความรู้หรือเรื่องใดเรื่องหนึ่ง ที่สามารถอธิบายรายละเอียดได้ เป็นพื้นฐานสำคัญในการสร้างฐานความรู้

2. คุณสมบัติ (Property) คือคุณสมบัติที่ใช้ในการอธิบายแนวคิดหรือคุณลักษณะของคลาส

3. ความสัมพันธ์ (Relationship) คือความสัมพันธ์ระหว่างแนวคิด โดยประกอบไปด้วยความสัมพันธ์ ดังนี้

3.1 is-a คือความสัมพันธ์ที่มีคุณสมบัติการถ่ายทอดจากคลาสแม่ไปยังคลาสลูก เช่น PC is-a Computer เป็นต้น ซึ่งอธิบายได้ว่า PC มีคุณสมบัติเป็น Computer

3.2 has-a หรือ part-of คือความสัมพันธ์ที่คลาสหนึ่งมีคลาสอื่นเป็นส่วนประกอบ เช่น คลาส Computer มีคลาส Keyboard เป็นส่วนประกอบ

3.3 syn-of คือความสัมพันธ์ของคลาสที่มีความหมายเหมือนกัน เช่น Data processor syn-of Computer หมายความว่า Computer มีความหมายเดียวกับ Data processor สามารถใช้แทนกันได้

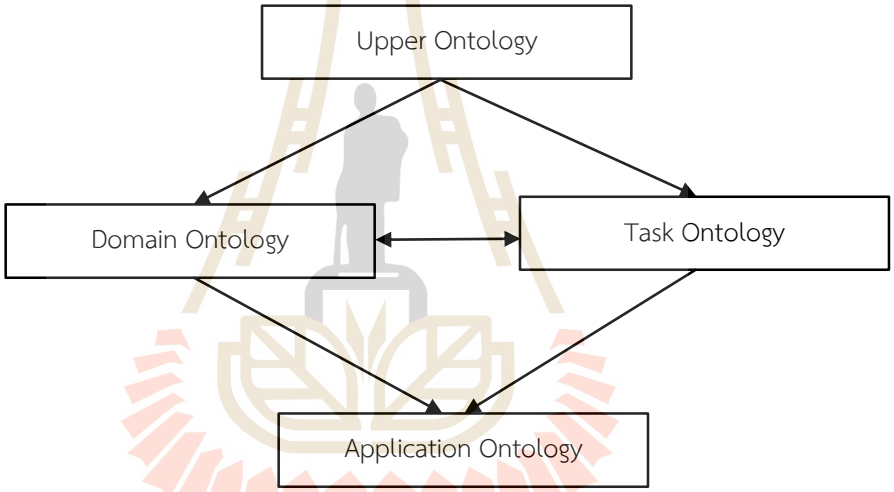
3.4 instance-of คือความสัมพันธ์ที่แสดงถึงการเป็นตัวแทน หรือสมาชิกของคลาส เช่น HP instance-of Computer หมายความว่า HP เป็นยี่ห้อคอมพิวเตอร์ยี่ห้อหนึ่งซึ่งถือเป็นคอมพิวเตอร์ประเภทหนึ่ง

4. ข้อกำหนดในการสร้างความสัมพันธ์ (Axioms) คือเงื่อนไขหรือตรรกะในการแปลงความสัมพันธ์ระหว่างแนวคิด หรือระหว่างคุณสมบัติ เพื่อการแปลงความหมายที่ถูกต้อง เป็นกลไกสำคัญสำหรับการอนุมานความรู้ เพื่อสร้างความรู้ใหม่

5. ข้อมูล (Instance) คือคำศัพท์ที่มีการกำหนดความหมายไว้ในออนโทโลยีในเรื่องนั้นๆ

2.6.1 ประเภทของออนโทโลยี

การกำหนดประเภทของออนโทโลยี สามารถกำหนดได้หลากหลายประเภท ขึ้นอยู่กับวัตถุประสงค์และความเหมาะสมในการอธิบายข้อมูลแสดงดัง ภาพที่ 2.7 (Guarino, 1998)



ภาพที่ 2.7 ประเภทของออนโทโลยี

1) ออนโทโลยีระดับบน (Upper ontology) เป็นออนโทโลยีที่อธิบายแนวคิดทั่วไป เช่น พื้นที่ เวลา วัตถุ เหตุการณ์หรือการกระทำ หรืออะไรก็ตามที่ไม่ได้ขึ้นอยู่กับปัญหาหรือโดเมนเฉพาะ

2) ออนโทโลยีโดเมน (Domain Ontology) เป็นออนโทโลยีที่อธิบายคำศัพท์สำหรับโดเมนทั่วไป (Generic Domain) โดยอาศัยการถ่ายทอดคุณลักษณะเฉพาะของกิจกรรมจากออนโทโลยีระดับบน

3) ออนโทโลยีกิจกรรม (Task Ontology) เป็นออนโทโลยีที่เหมาะสำหรับงานที่เฉพาะเจาะจง เช่นการประกอบชิ้นส่วนเข้าด้วยกัน ซึ่งเป็นงานหรือกิจกรรม โดยใช้วิธีการเฉพาะ โดยอาศัยการถ่ายทอดคุณลักษณะเฉพาะของกิจกรรมจากออนโทโลยีระดับบน

4) ออนโทโลโลยีโลคอล (Application Ontology or Local Ontology) เป็นออนโทโลยีที่อธิบายการประยุกต์ ใช้แนวคิดซึ่งขึ้นอยู่กับโดเมนเฉพาะและงานหรือกิจกรรม เป็นแนวความคิดที่ตอบสนองต่อบทบาทของโดเมนที่มีความเฉพาะเจาะจง

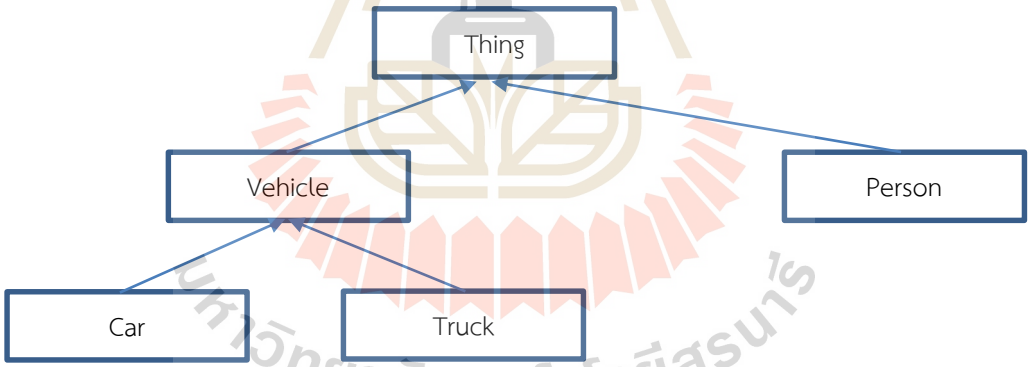
2.6.2 ส่วนประกอบหลักของออนโทโลยี

1) การกำหนดข้อมูลโครงสร้างสำหรับการอธิบายข้อมูลเชิงความหมาย คือการกำหนดออนโทโลยีระดับบน เพื่อเป็นโมเดลแสดงโครงสร้างการอธิบายข้อมูลเชิงความหมาย ซึ่งต้องมีการกำหนดคลาส (Class) คุณสมบัติ (Property) และเงื่อนไข (Restriction) ดังตัวอย่างแสดงดังภาพที่ 2.8 ซึ่งในการอธิบายข้อมูลมีส่วนประกอบ ดังนี้

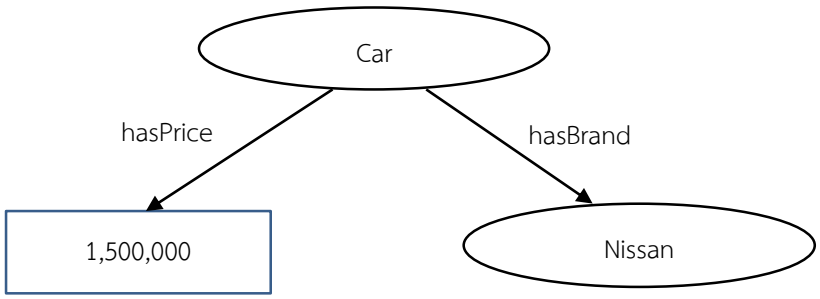
1.1) คลาส (Class) หรือ แนวคิด (Concept) คือการที่แสดงถึงสิ่งที่เราสนใจ และอธิบายได้ว่าคลาสต่าง ๆ ประกอบไปด้วยอะไรบ้างภายในโดเมน

1.2) คุณสมบัติ (Property) คือการแสดงถึงการกำหนดความสัมพันธ์ (Relation) หรือคุณลักษณะของคลาส เพื่อเชื่อมโยงระหว่างคลาสด้วยการระบุค่าคุณสมบัติ

2) ข้อมูลอินสแตนส์ (Instance Data) คือการอธิบายรายละเอียดของข้อมูล ซึ่งใช้ข้อมูลโครงสร้างในการอธิบายแสดงตัวอย่างดังภาพที่ 2.9



ภาพที่ 2.8 ตัวอย่างส่วนประกอบของออนโทโลยี



ภาพที่ 2.9 ตัวอย่างข้อมูลอินสแตนส์ในการอธิบายรายละเอียดของข้อมูล

มหาวิทยาลัยเทคโนโลยีสุรนารี

ตารางที่ 2.2 อธิบายตัวอย่างคลาสและคุณสมบัติของรถยนต์

เอ็นทิตี้	ประเภทของข้อมูล	ความหมาย
Vehicle	Class	ยานพาหนะ
Car	Class	รถยนต์
Truck	Class	รถบรรทุก
Person	Class	บุคคล
hasPrice	Properties	ระบุราคาของรถยนต์
hasBrand	Properties	ระบุยี่ห้อของรถยนต์

2.6.3 ประโยชน์ของออนโทโลยี

ออนโทโลยี (Ontology) สามารถนำมาใช้เพื่อสนับสนุนความหลากหลายที่ดีของงานวิจัย ในพื้นที่ที่หลากหลาย (Wache และคณะ, 2001) เช่น เป็นตัวแทนของความรู้ การประมวลผลภาษาธรรมชาติ การค้นคืนสารสนเทศ ฐานข้อมูล การจัดการความรู้ การบูรณาการฐานข้อมูลออนไลน์ ห้องสมุดดิจิทัล ระบบสารสนเทศทางภูมิศาสตร์ การค้นคืนสารสนเทศด้วยภาพหรือระบบค้นคืนหลากหลายรูปแบบ เป็นต้น ซึ่งการที่จะนำออนโทโลยีไปประยุกต์ในกับงานด้านใดนั้น ขึ้นอยู่กับวัตถุประสงค์ของการใช้งาน (Fan และคณะ, 2012) โดยสามารถแบ่งประโยชน์ของออนโทโลยีได้ดังต่อไปนี้

1) ระบบสืบค้นข้อมูล การนำออนโทโลยีไปใช้กับระบบสืบค้นข้อมูลนั้น ทำให้ค้นหาและเข้าถึงข้อมูลที่ต้องการของผู้ใช้และได้ข้อมูลที่ถูกต้องมากยิ่งขึ้น

2) ระบบงานที่ต้องการผนวกความรู้ (Knowledge Integration) เพื่อสร้างความเข้าใจเบื้องต้นของความรู้ระหว่างโดเมน

3) ระบบงานที่มีความต้องการนำข้อมูลกลับมาใช้ใหม่ (Reuse)

2.6.4 ภาษาและเครื่องมือที่ใช้ในการจัดการฐานความรู้ออนโทโลยี

ภาษาที่ใช้พัฒนาออนโทโลยีแบ่งตามรูปแบบภาษา (Guarino, 1998) ได้ดังต่อไปนี้

1) Logic based คือรูปแบบภาษาที่ใช้อธิบายออนโทโลยี เป็นการใช้ตรรกะสำหรับอนุมานความรู้ โดยมีส่วนประกอบคือ ภาคแสดง (Predicate) อาร์กิวเมนต์ (Argument) และตัวบ่งปริมาณ (Ouantifier) คุณสมบัติของภาษาที่พัฒนาแบบลอจิกเบส (Logic based) คือ วากยสัมพันธ์ (Syntax) ซึ่งมีความชัดเจนและเป็นทางการ เช่น ภาษา CYCL, CLASSIC, LOOm เป็นต้น

2) Frame based คือรูปแบบภาษาที่ใช้อธิบายออนโทโลยีด้วยหลักการของการคิดเชิงปฏิสัมพันธ์ของมนุษย์ เมื่อนึกถึงวัตถุหรือสิ่งใดจะเชื่อมโยงกับคุณลักษณะเด่นของวัตถุหรือสิ่งนั้นด้วย ซึ่งประกอบไปด้วย เซตของคุณสมบัติ (Attribute) หรือสล็อท (Slot) และแฟ็คเซ็ต (Facets) ภาษาที่พัฒนาแบบเฟรมเบส (Frame based) เช่น OKBC, F-logic เป็นต้น

(Car <Model modelname> <Color color> <Price price>)	(Honda <Model SUV> <Color White> <Price 1,500,000 Bath>)

ภาพที่ 2.10 ตัวอย่างองค์ประกอบการแทนแนวคิดของรถยนต์

จากภาพที่ 2.10 เป็นการอธิบายองค์ประกอบการแทนแนวคิดของรถยนต์ (Car) ซึ่งประกอบด้วยคุณสมบัติ 3 คุณสมบัติ ได้แก่ รุ่น (Model), สี (Color) และราคา (Price) และการแทนค่าคุณสมบัติของตัวอย่างที่เป็นสมาชิกของแนวคิด (Instance) คือ Honda

3) Web Based คือภาษาที่อธิบายทรัพยากรบนเว็บ โดยอธิบายออนโทโลยีโดยใช้พื้นฐานของลอจิคเบสและเฟรมเบส เป็นภาษาที่มีหลักการและแบบแผนอ้างอิงได้ และอยู่ในรูปแบบที่มนุษย์สามารถเข้าใจได้ง่าย ตัวอย่างภาษาแบบเว็บเบส ได้แก่ XML (Extensible Markup Language), RDF (Resource Description Framework), OWL (Web Ontology Language) เป็นต้น

3.1) XML (Extensible Markup Language) เป็นภาษาที่ออกแบบมาสำหรับการแลกเปลี่ยนระหว่างเครื่องคอมพิวเตอร์ที่แตกต่างกัน และเน้นการแลกเปลี่ยนข้อมูลผ่านอินเทอร์เน็ต นอกจากนี้ยังมีจุดประสงค์เพื่อติดต่อกันในระบบที่มีความแตกต่างกัน เช่น มีระบบปฏิบัติการที่แตกต่างกันหรือโปรแกรมประยุกต์แตกต่างกัน พัฒนาโดย World Wide Web Consortium หรือ W3C (2015)

XML เป็นภาษาที่ใช้เน้นส่วนที่เป็นข้อมูล โดยสามารถกำหนดชื่อแท็ก (Element) และชื่อแอตทิบิวต์ได้ตามความต้องการของผู้สร้างเอกสาร โดยเอกสารนั้นจะต้องมีความเป็น Well-formed ซึ่งเป็นไวยากรณ์พื้นฐานของเอกสาร XML เช่น เอกสาร XML ต้องเริ่มต้นด้วย <?xml version=

"1.0" ?> เอกสาร XML 1 เอกสาร จะต้องมีแท็กรูทเพียงแท็กเดียว หมายความว่า แท็กและข้อมูลต่าง ๆ จะต้องอยู่ภายในแท็กแรกสุดเพียงแท็กเดียว การเปิดและปิดแท็กจะต้องไม่มีการคร่อมกัน ส่วน DTD และ Schema จะมีหรือไม่มีก็ได้ ขึ้นอยู่กับว่ามีผู้ใช้เอกสารนั้นมากน้อยแค่ไหน เอกสาร XML จึงเป็นแค่เท็กซ์ไฟล์ชนิดหนึ่งที่มีแท็กเปิดและแท็กปิดครอบข้อมูลไว้ตรงกลางเท่านั้น ทำให้เอกสาร XML ถูกใช้ในการติดต่อกับระบบที่ต่างกัน เนื่องจากความง่ายในการสร้างเอกสาร การนำเอกสาร XML ไปใช้งาน จะสนใจแต่ข้อมูลที่ถูกกำกับด้วยแท็กมากกว่า

เนื่องจากเอกสาร XML สามารถกำหนดชื่อแท็ก และชื่อแอตทิบิวต์ได้ตามความต้องการของผู้สร้างเอกสาร ทำให้ในการเน้นข้อมูลใดข้อมูลหนึ่ง สามารถมีเอกสาร XML หลายรูปแบบ ผู้เขียนอาจใช้ชื่อแท็กต่างกัน ทั้งที่สื่อความหมายไปที่สิ่งเดียวกัน หากว่าเอกสาร XML นั้น ถูกนำไปใช้ติดต่อกับระบบอื่นๆ อาจทำให้สื่อความหมายไม่ตรงกัน ดังนั้นจึงต้องมีการกำหนดรูปแบบที่เป็นมาตรฐานขึ้น (ตกลงรูปแบบระหว่างกัน) โดย DTD และ Schema จะเป็นตัวกำหนดว่าเอกสาร XML นั้น จะต้องมีแท็กอะไรบ้าง ภายในแท็กนั้นจะมีแท็ก แอตทิบิวต์ หรือข้อมูลอะไรได้บ้าง โดย DTD จะต่างกับ Schema ตรงที่ Schema เป็นเอกสาร xml ด้วย

เอกสาร XML ประกอบด้วย 2 ส่วนหลัก ๆ คือ Prolog และ Root Element (Frey, n.d.) ซึ่งมีรายละเอียดดังนี้

- Prolog หรือส่วนหัวของเอกสาร XML ใช้อธิบายเอกสาร XML โดยแสดงเวอร์ชั่นของเอกสาร และตัวอักษรที่ใช้ในเอกสาร

```
<?xml version = “1.0” encoding = “UTF-8”?>
```

- Root Element เป็นส่วนเนื้อหาทั้งหมดของเอกสาร ประกอบด้วย แท็กเปิดและแท็กปิด

```
<staff>Somchai</staff>
```

จากข้อความข้างต้นอธิบายได้ว่า <staff> คือแท็กเปิด ส่วน Somchai คือข้อมูล และ </staff> คือแท็กปิด แท็กปิดจะต้องมีชื่อเหมือนแท็กเปิด โดยตามหลังเครื่องหมาย “/” เนื่องจากภาษา XML เป็นภาษาที่มีผลกับอักษรตัวพิมพ์ใหญ่และตัวพิมพ์เล็ก (case sensitive) ดังนั้นการใส่แท็กเปิดและแท็กปิดไม่ตรงกัน จะทำให้โปรแกรมไม่ทำงาน เช่น ใส่แท็กเปิดเป็น <staff> แต่ใส่แท็กปิดเป็น </Staff> จะถือว่าเป็นคนละแท็กกัน ดังนั้นแท็กเปิดและแท็กปิดไม่เป็นแท็กเดียวกัน เป็นต้น

มหาวิทยาลัยเทคโนโลยีสุรนารี

นอกจากนี้แท็กของ XML ยังสามารถอยู่ในแท็กอื่น ๆ ได้อีกด้วย

```
<staff>
    <name>Somchai</name>
        <id>P001</id>
</staff>
```

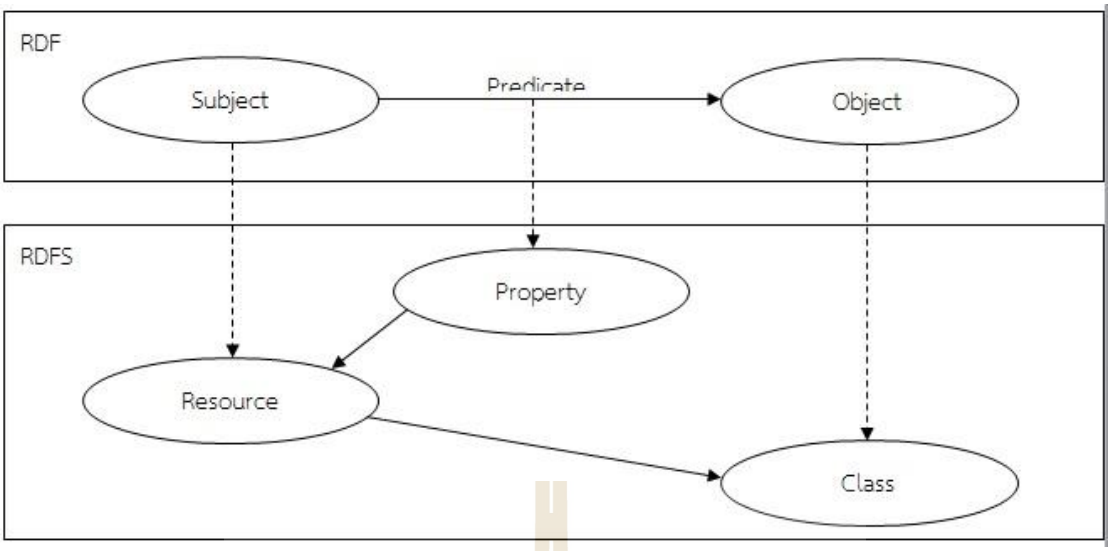
3.2) RDF (Resource Description Framework) พัฒนาขึ้นโดย W3C (1999) เป็นกรอบการทำงานที่เป็นตัวแทนของข้อมูลในเว็บ เป็นภาษามาตรฐานที่อธิบายลักษณะและความสัมพันธ์ของข้อมูล ออกแบบมาเพื่อให้คอมพิวเตอร์สามารถเข้าใจข้อมูลร่วมกัน โดยมีโครงสร้างพื้นฐานที่ช่วยในการแลกเปลี่ยน การเข้ารหัส และการนำกลับมาใช้ใหม่ของเมทาดาทา (Metadata) ช่วยเพิ่มการค้นพบและการเข้าถึงข้อมูลดังกล่าว มีส่วนช่วยในการอธิบายข้อมูลสารสนเทศที่จัดเก็บจากแหล่งข้อมูลต่าง ๆ (W3C, 1999) RDF ใช้ XML เป็นโครงสร้างและรูปแบบสำหรับการแลกเปลี่ยนและการประมวลผลของข้อมูล (Frey, n.d.) โดยโครงสร้างของ RDF ประกอบด้วย 3 ส่วน เรียกว่า Triple ได้แก่ ทรัพยากร (Subject/Resource) คุณสมบัติของทรัพยากร (Predicate) และค่าของคุณสมบัติ (Object)

- Subject หรือ Resource คือส่วนของทรัพยากรที่แทนด้วย URI (Uniform Resource Identifier) โดย URI เป็นการเข้ารหัสของที่อยู่กับวัตถุบนอินเทอร์เน็ต ซึ่งอาจอยู่ในรูปแบบของ URL ที่จะใช้ในการเข้าถึงทรัพยากร

- Predicate หรือ Property คือส่วนคุณสมบัติของทรัพยากร ซึ่งเป็นคุณสมบัติเฉพาะของทรัพยากร ซึ่งอาจแทนด้วยรูปแบบของ Element

- Object คือส่วนของค่าของคุณสมบัติ เรียกว่า Property Value ซึ่งอาจเป็น Object ซึ่งได้แก่แท็กอื่น ๆ ที่อยู่ภายในแท็กนั้น หรืออาจเป็นค่าอักขระซึ่งจะเรียกว่า Literal

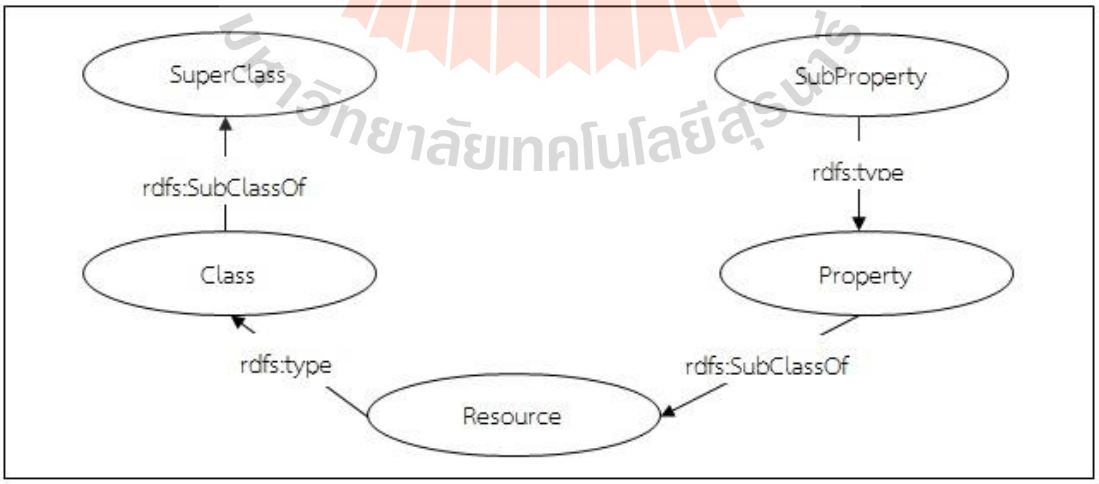
3.3) RDFS (RDF Schema) เป็นส่วนต่อขยายจาก RDF เป็นภาษาเพิ่มความสามารถในการอธิบายข้อมูลมากขึ้น ใช้ในการอธิบายโครงสร้างของเมทาดาทา (Metadata) ของทรัพยากรบนเว็บ โดยกำหนดคำศัพท์ (Vocabulary) เพื่ออธิบายโครงสร้างของเมทาดาทา (W3C, 1999) ซึ่งประกอบไปด้วยคุณสมบัติและค่าของคุณสมบัติที่อธิบายให้อยู่ในรูปแบบของ Class และ Property โดยสามารถสืบทอดเป็น SubClass และ SubProperty ได้ เมื่อเอกสาร RDF มี RDF Schema กำกับจะช่วยให้เครื่องอนุมานข้อมูลได้ เป็นการกำหนดโครงสร้างของเมทาดาทาซึ่งใช้อธิบายความสัมพันธ์ของแหล่งทรัพยากร



ภาพที่ 2.11 โครงสร้างของเมทาดาทาของเอกสาร RDF (ที่มา W3C, 1999)

จากภาพที่ 2.11 เป็นการอธิบายโครงสร้างเมทาดาทา ซึ่งประกอบไปด้วยคุณสมบัติและค่าของคุณสมบัติ มากำหนดให้เป็นโหนดเพื่ออธิบาย Resource โดยในส่วนของ Predicate ซึ่งเป็นคุณสมบัติ จะถูกกำหนดเป็นโหนด Property และส่วนของ Object ซึ่งเป็นค่าของคุณสมบัติจะถูกกำหนดให้เป็นโหนด Class โดยทั้งโหนด Property และโหนด Class จะสามารถสืบทอดเป็น SubProperty และ SubClass ได้ตามลำดับ

เมื่อทำการอธิบายส่วนของเมทาดาทาให้เป็นโหนด Property และโหนด Class รวมถึงมีการสืบทอดเป็น SubProperty และ SubClass จะมีการจัดวางความสัมพันธ์ ระหว่างโหนดในรูปแบบของ RDF ซึ่งมีความสัมพันธ์ดังภาพที่ 2.12



ภาพที่ 2.12 การอธิบายความสัมพันธ์ของ Resource ในรูปแบบ RDFS (ที่มา W3C, 1999)

จากภาพที่ 2.20 เป็นการจัดวางความสัมพันธ์ของแต่ละโหนดในรูปแบบของ RDFS โดยโหนด Resource กับโหนด Class มีความสัมพันธ์กัน เรียกความสัมพันธ์ดังกล่าวว่าเป็น rdfs:SubClassOf, โหนด Class กับ SubClass มีความสัมพันธ์กัน ที่เรียกว่า rdfs:SubClassOf และโหนด Property กับ SubProperty มีความสัมพันธ์กันที่เรียกว่า rdfs:type ดังนั้น RDFS สามารถอธิบายกลุ่มของทรัพยากรบนเว็บที่มีความเกี่ยวข้องกันและสามารถเชื่อมต่อความสัมพันธ์ระหว่างทรัพยากรบนเว็บ (ภรศิษฐ์ กุหลาบทิพย์, 2550) เช่น “Pupil” เป็นกลุ่มประเภทของ “Student” และ “Student” ก็เป็นคลาสย่อยของ “People” เป็นต้น

ในการสร้าง RDFS จะกำหนดคำศัพท์ (Vocabulary) เพิ่มเข้าไปเพื่ออธิบายโครงสร้างของเมทาดาทาที่อธิบายในโครงสร้าง RDF เพื่อให้ข้อมูลมีความหมายมากขึ้น เช่น การบรรยายคลาส การบรรยายการสืบทอดคลาส ข้อจำกัดของ RDFS ประกอบด้วยคลาส และคุณสมบัติของคลาส การแสดงคุณสมบัติของคลาสถูกกำหนดโดย rdf:Property RDFS มี 4 คุณสมบัติ คือ rdfs:subClassOf, rdfs:type, rdfs:range, rdfs:domain แต่คุณสมบัติเหล่านี้ไม่เพียงพอต่อการบรรยายข้อมูลให้มีความหมายเชิงตรรกะได้ ดังนั้นจึงได้มีการกำหนดภาษาที่ใช้บรรยายโครงสร้างความสัมพันธ์ และความหมายเชิงตรรกะได้ โดยภาษาดังกล่าวเรียกว่า ภาษา OWL (Web Ontology Language)

3.4) Web Ontology Language: OWL คือมาตรฐานสากลสำหรับการเข้ารหัส และแลกเปลี่ยนออนโทโลยี ออกแบบมาเพื่อสนับสนุนเว็บเชิงความหมาย (W3C, 2004) ดังนั้นคอมพิวเตอร์จะสามารถจัดการกระบวนการได้อย่างฉลาดมากยิ่งขึ้น ซึ่งส่วนประกอบที่สำคัญของเว็บเชิงความหมายคือการระบุและการใช้ออนโทโลยี OWL คือภาษาหนึ่งในส่วนของออนโทโลยีสำหรับเว็บ ถูกสร้างโดย W3C เมื่อกุมภาพันธ์ 2004 ภาษา OWL ถูกออกแบบมาเพื่อให้ทำงานร่วมกับภาษา XML (eXtensible Markup Language) โดยใช้ในการนิยามคลาส คุณสมบัติ และอธิบายถึงความสัมพันธ์ระหว่างคลาสและคุณสมบัติ โดยถูกพัฒนาต่อจากภาษา RDF (Resource Description Framework) (Rajapaksha และ Kodagoda, 2008) ช่วยให้เครื่องคอมพิวเตอร์สามารถประมวลผลแบบเชิงความหมายได้ สามารถใช้ OWL ในการแสดงความหมายของคำ และความสัมพันธ์ระหว่างคำ กล่าวคือ สามารถแทนความหมายของเนื้อหาของเอกสารบนเว็บ โดยที่เครื่องคอมพิวเตอร์สามารถแปลความหมายได้ (W3C, 2004) ซึ่งภาษาย่อยของ OWL ดังนี้

3.4.1) OWL Lite เหมาะกับการเก็บข้อมูลความสัมพันธ์แบบระดับชั้น (Hierarchy) และมีข้อกำหนดอย่างง่าย สนับสนุนการใช้งานเบื้องต้น เหมาะกับความต้องการของผู้ใช้ที่ต้องการเพียงการจัดลำดับ และการแบ่งประเภทโดยใช้ลักษณะหรือเกณฑ์ในการแบ่งแบบง่าย ๆ มีการกำหนดโครงสร้างในรูปแบบลำดับชั้น และมีการบังคับใช้คุณสมบัติพื้นฐานในการกำหนดโครงสร้างข้อมูล ง่ายต่อการพัฒนา

3.4.2) OWL DL เป็นภาษาที่รองรับทุกคำสั่งของ OWL แต่จะต้องเขียนโปรแกรมภายใต้ข้อกำหนด เช่น คลาสสามารถเป็นซับคลาสของคลาสอื่น ๆ ได้มากกว่าหนึ่งคลาส DL ย่อมาจาก Description logic มีข้อบังคับในการใช้ RDF การกำหนดคลาส และค่าของข้อมูล โดยจัดให้มีคุณสมบัติที่เหมาะกับการใช้งานด้านฐานข้อมูล และแทนความรู้ที่อยู่บนพื้นฐานของการอธิบายด้วยตรรกะ สามารถบรรยายข้อมูลและโครงสร้างข้อมูล ในรูปแบบโครงสร้างภาษา OWL ด้วยข้อจำกัดของคลาสและคุณสมบัติได้

3.4.3) OWL Full เป็นภาษาที่มีความซับซ้อนมากที่สุด โดยมีลักษณะพิเศษ เช่น Syntax อิสระจาก RDF หรือคลาสสามารถเป็นอินสแตนท์ได้ เป็นต้น สนับสนุนผู้ใช้งานที่ต้องการความครบถ้วนและมีโครงสร้างที่สมบูรณ์แบบ OWL Full มีการผสมผสานระหว่าง OWL และ RDFS โดยไม่มีการบังคับในส่วนของการแบ่งคลาส การกำหนดคุณสมบัติและค่าของข้อมูล ผู้ใช้สามารถอธิบายข้อมูลในรูปแบบ RDFS ได้อย่างอิสระ

โครงสร้างของภาษา OWL มีการบรรยายข้อมูลแบบผสมผสานกันระหว่างการใช้ไวยากรณ์ของ RDF, RDFS และ XML ซึ่งแบ่งตามประเภทของการใช้งาน ผลลัพธ์ที่ได้จะอยู่ภายใต้รูปแบบของ RDF Triples (W3C, 2004) เอกสารของภาษา OWL ประกอบไปด้วยส่วนประกอบดังต่อไปนี้

- Namespace การกำหนด Namespace จะประกาศไว้ที่ส่วนเริ่มต้นของเอกสารเพื่อเป็นการกำหนดกลุ่มในการอ้างอิงข้อมูล เอกสาร OWL ที่ถูกสร้างจะขึ้นอยู่กับโครงสร้างที่ถูกนิยามด้วย RDF, RDFS และชนิดข้อมูลของ RDFS การเขียน Namespace จะประกาศไว้ภายใต้คำสั่งของ rdf:RDF syntax ดังตัวอย่าง

```
rdf:RDF xmlns:rdf=http://www.w3.org/1999/02/22-rdf-syntax-ns#”
xmlns:rdfs=”http://www.w3.org/2000/01/rdf-schema#”
xmlns:owl=http://www.w3.org/2002/07/owl#”
xmlns:xsd=http://www.w3.org/2000/10/XMLSchema#”
```

- Ontology Headers แสดงการอธิบายรายละเอียดเบื้องต้นของออนโทโลยี ว่าเป็นโครงสร้างข้อมูลเกี่ยวกับอะไร ภายใต้อิลิเมนต์ <owl:Ontology rdf:about=””> ประกอบด้วยคำสั่ง <owl:versionInfo> ใช้แสดงรุ่นของข้อมูลที่สร้าง คำสั่ง <rdf:comment> ใช้แสดงการอ้างอิงเกี่ยวกับแหล่งข้อมูลว่ามาจากที่ใด ดังตัวอย่าง

```
<owl:Ontology rdf: about=””>
<owl:versionInfo>V 1.17 2003/02/26</owl:versionInfo>
<rdfs:comment>An example ontology</rdfs:comment>
<owl:imports rdf:resource=”http://www.example.com/owl”/></owl:Ontology>
```

- อิลิเมนต์พื้นฐาน (Basic element) ของ OWL มีดังต่อไปนี้

- คลาส และคลาสย่อย คลาสคือเค้าโครงที่ใช้ในการกำหนดกลุ่มของข้อมูลหรือวัตถุ (Object) โดยทุก ๆ OWL คลาสจะมีความสัมพันธ์กับข้อมูลที่มีลักษณะเฉพาะหรือที่เรียกว่า (Class Extension) ซึ่งเป็นลักษณะของการสืบทอดคลาส ซึ่งมีลักษณะเฉพาะ (Individual) ที่อยู่ภายใน Class Extension เรียกว่าตัวแทนข้อมูล โดยคลาสจะมีลักษณะเป็นลำดับชั้น มีการสืบทอดคุณสมบัติของคลาส ในโครงสร้างของภาษา OWL จะประกอบด้วยคลาส

1. owl:Thing ทุก ๆ สิ่งที่มีลักษณะเฉพาะจะเป็นสมาชิกของคลาส owl:Thing โดยกำหนดให้ owl:Thing เป็นคลาสใหญ่ที่สามารถครอบคลุมทุกคลาสข้อมูลได้

2. owl:Nothing ใช้กำหนดคลาสที่ไม่มีสมาชิกอยู่ภายในคลาสนั้น ๆ หรือเป็นคลาสว่าง

ตัวย่างในการสร้างคลาสของระบบลงทะเบียน ที่ประกอบด้วยคลาส 2 คลาส ที่มีความสัมพันธ์กัน คือคลาสหลักสูตร (Course) และคลาสนักศึกษา (Student) โดย 2 คลาสมีความสัมพันธ์ที่มีชื่อมีว่าลงทะเบียน (Enrolled) ในคลาสประกอบไปด้วยคุณสมบัติของคลาส โดยคลาสหลักสูตร คุณสมบัติของคลาส ได้แก่ รหัสหลักสูตร (Code) คำอธิบายหลักสูตร (Description) และจำนวนผู้ลงทะเบียน (NumEnrolled) ส่วนคลาสนักศึกษา ประกอบด้วยคุณสมบัติ รหัสนักศึกษา (ID) และชื่อนักศึกษา (Name) ดังนั้นในการประกาศชื่อคลาสสามารถทำได้ดังนี้

```
<owl:Class rdf:ID=”Course”/>
<owl:Class rdf:ID=”Student”/>
```

จากตัวอย่าง ในการประกาศชื่อคลาส จะมีคลาสเริ่มต้นคือ owl:Class โดยกำหนดให้ owl:Class เป็นคลาสใหญ่ที่สามารถครอบคลุม ทุกคลาสข้อมูลได้ ซึ่ง owl:Class สามารถประกอบไปด้วยซับคลาสได้ การกำหนดชื่อคลาสอยู่ภายใต้ syntax rdf:ID=”Course” ภายในเครื่องหมาย “ ” จะแทนชื่อคลาสที่ถูกกำหนดขึ้น

การแทนชื่อคลาสในภาษา OWL สามารถเขียนได้หลายรูปแบบ เช่น #Course เขียนแทนด้วย rdf:resource=“#Course” โครงสร้างที่เป็นพื้นฐานของคลาสคือ rdf:subClassOf โดยจะมีความสัมพันธ์ ดังนี้ ถ้า X เป็นคลาสย่อย (Subclass) ของ Y ดังนั้น ทุก ๆ instance ของ X จะเป็น instance ของ Y ด้วย ซึ่ง rdf:subClassOf เป็นความสัมพันธ์แบบ transitive ถ้า X เป็นคลาสย่อยของ Y และ Y เป็นคลาสย่อยของ Z ดังนั้น X เป็นคลาสย่อยของ Z ด้วย ดังตัวอย่างการประกาศคลาสที่มีคลาสย่อยอยู่ภายใน ดังนี้

```
<owl:Class rdf:ID=”Subject”>
<rdfs:subClassOf rdf:resource=”#Course”/>
</owl:Class>
```

จากตัวอย่างเป็นกำหนดคลาส “Subject” เป็นคลาสย่อย (Subclass) ของคลาส “Course” ความสัมพันธ์แบบคลาสย่อยนี้ทำให้เกิดคลาสทั่วไป (Generic class) และคลาสจำเพาะเจาะจง (Specific class)

การนิยามคลาสจะประกอบด้วย 2 ส่วน ได้แก่ ชื่อคลาส ที่ใช้แทนการอ้างอิงการใช้งานคลาสนั้น และการแสดงรายละเอียดหรือข้อจำกัดต่าง ๆ ส่วนของการอธิบายสมาชิกภายในคลาสนั้น ทำได้โดยการประกาศถึงคุณลักษณะเฉพาะของสมาชิกภายในคลาส (W3C, 2004) ดังแสดงในตารางที่ 2.3

ตารางที่ 2.3 การประกาศถึงคุณลักษณะเฉพาะของสมาชิกภายในคลาส

Classifier	Code	Description	NumEnrolled
Course	IT204999	Semantic Web	0

จากตารางที่ 2.3 แสดงข้อมูลสมาชิกของคลาส Course ข้อมูลที่แสดงถึงความเฉพาะเจาะจง คือ Code เขียนในรูปแบบภาษา OWL ได้ดังนี้ <owl:Thing tdf:ID=”IT204999”/>

คุณสมบัติของคลาส ในภาษา OWL การกำหนดคุณสมบัติของคลาสออกแบ่งเป็น 2 ประเภท ได้แก่ Object Properties และ Datatype Properties

- Object Properties เป็นการแสดงความสัมพันธ์ระหว่าง instance ของคลาสที่เกี่ยวข้องกัน 2 คลาส

- Datatype Properties เป็นการกำหนดคุณสมบัติ เพื่อแสดงความสัมพันธ์ระหว่าง instance ของคลาสที่เป็นค่าชนิดพื้นฐานและ RDF literals รวมถึงการแสดงชนิดข้อมูล XML Schema datatypes

คุณสมบัติและชนิดข้อมูล การเขียนโครงสร้างของภาษา OWL ประกอบด้วยชนิดของข้อมูลที่ต้องการใช้งาน ซึ่งการประกาศใช้งานจะประกาศผ่านคำสั่ง XML Schema datatypes โดยจะอ้างอิงตามชนิดข้อมูลที่ใช้งานจริง และมีการประกาศการใช้งาน URI ที่มีการอ้างอิงการใช้งานชนิดข้อมูลนั้น ๆ ขนิดข้อมูลที่อ้างอิงตาม XML schema datatypes แสดงดังตารางที่ 2.4 (W3C, 2004)

ตารางที่ 2.4 ชนิดข้อมูลที่อ้างอิงตาม XML schema datatypes

ชนิดข้อมูล	คำสั่ง
String	xsd: string, xsd: normalizedString, xsd: token, xsd: language, xsd: NMTOKEN, xsd: Name, xsd: NCName, xsd: hexBinary, xsd:base64Binary, xsd:anyURI
Boolean	xsd:Boolean
Numerical	xsd: decimal, xsd: float, xsd: double, xsd: integer, xsd: positiveInteger, xsd: nonPositiveInteger, xsd: negativeInteger, xsd: nonNegativeInteger, xsd: long, xsd: int, xsd: short, xsd: byte, xsd: unsignedLong, xsd: unsignedInt, xsd: unsignedShort, xsd:unsignedByte
Time	xsd: dateTime, xsd: time, xsd: date, xsd: gYearMonth, xsd: gYear, xsd:gMonth, xsd:gDay, xsd:gMonth

ข้อกำหนดของคุณสมบัติ ภาษา OWL มีข้อกำหนดในการใช้งานคุณสมบัติ และลักษณะพิเศษของการบรรยายคลาส โดยอธิบายคลาสในรูปแบบ anonymous class รวมถึงทุกคลาสที่เป็น individual class ได้ถูกกำหนดให้เป็นข้อจำกัดในการใช้งานคุณสมบัติ โดยโครงสร้างภาษา OWL แบ่งข้อกำหนดของคุณสมบัติเป็น 2 ประเภท ดังตารางที่ 2.5

ตารางที่ 2.5 ประเภทข้อกำหนดคุณสมบัติของ OWL

ประเภท	คำสั่ง
1. Value constraints	Owl:allValuesFrom, owl:someValuesFrom, owl:hasValue
2. Cardinality constraints	Owl:maxCardinality, owl:minCardinality, owl:cardinality

ในการแทนค่าข้อกำหนดคุณสมบัติ ยกตัวอย่างดังต่อไปนี้

- การแทนค่าข้อกำหนดของคุณสมบัติโดยคำสั่ง owl:allValuesFrom เป็นการจำกัดความต้องการสำหรับทุก ๆ instance ของคลาสให้มีคุณสมบัติที่เป็นลักษณะเฉพาะเท่านั้น ซึ่งค่าของคุณสมบัตินี้ทุก ๆ สมาชิกภายในคลาสจะถูกกำหนดภายใต้คำสั่ง owl:allValuesFrom เช่น การนิยามว่ามนุษย์ทุกคนจะต้องมีผู้ให้กำเนิด ดังตัวอย่าง

```
<owl:Restriction>
<owl:onProperty rdf:resource="#hasParent"/>
<owl:allValuesFrom rdf:resource="#Human"/>
</owl:Restriction>
```

- การแทนค่าข้อกำหนดของคุณสมบัติโดยคำสั่ง owl:someValuesFrom เป็นการแสดงความหมายว่าอย่างน้อย 1 อย่างของคุณสมบัติจะต้องเกี่ยวข้องกับ instance ของการบรรยายคลาส หรือค่าข้อมูลภายในขอบเขตของข้อมูล เช่น การนิยามว่าโดยทั่วไปผู้ให้กำเนิดอาจมีชีวิตอยู่ทั้ง 2 คน หรืออาจมีชีวิตเพียงบิดาหรือมารดา ดังตัวอย่าง

```
<owl:Restriction>
<owl:onProperty rdf:resource="#hasParent"/>
<owl:someValuesFrom rdf:resource="#Physician"/>
</owl:Restriction>
```

2.7 ผลไม้เศรษฐกิจของไทย

ผลไม้ หมายถึง ผลที่เกิดจากการขยายพันธุ์โดยอาศัยเพศของพืชบางชนิด ซึ่งมนุษย์สามารถรับประทานได้ เช่น แอปเปิ้ล มะม่วง และ ทุเรียน เป็นต้น ทั้งนี้สำนักงานเศรษฐกิจการเกษตร (2555) ได้อธิบายเกี่ยวกับผลไม้ว่าเป็นผลผลิตที่เกิดขึ้นจากสิ่งมีชีวิตจำพวกพืช โดยมีรูปทรงคล้ายทรงกลมหรือทรงรี มีเปลือกหรือมีสิ่งที่ห่อหุ้มเนื้อที่อยู่ข้างใน ซึ่งมักจะถูกนำไปเป็นอาหารโดยมนุษย์หรือสัตว์ และอาจมีความแตกต่างกันตามสายพันธุ์

การขยายพันธุ์ สามารถขยายพันธุ์ได้โดยดอก เมล็ด หรือ อื่นๆ ซึ่งผลไม้ที่ออกมานี้ตอนแรกจะมีขนาดเล็ก และมักจะไม่ค่อยถูกนำมารับประทานโดยมนุษย์ แต่เมื่อเติบโตจนสุกงอม จะมีลักษณะที่แตกต่างไปจากเดิม เช่น เปลี่ยนจากสีเขียวเป็นสีเหลือง มีกลิ่นหอม และรสหวาน เป็นต้น สามารถนำมา

มหาวิทยาลัยเทคโนโลยีสุรนารี

รับประทาน หรือประกอบอาหาร ส่วนมากมักจะนำมาทำเป็นอาหารหวาน ถ้าผลไม้สุกงอมเต็มที่จะมีลักษณะที่ก่อให้เกิดประโยชน์ได้น้อยลง เช่น เน่าเสีย บูด ขึ้นรา เป็นต้น และจะหลุดร่วงจากต้นลงสู่พื้นดินหรือพื้นน้ำ กลายเป็นอาหารให้แก่ห่วงโซ่อาหารลำดับถัดไป เช่น แบคทีเรีย จุรินทรีย์ จนกลายเป็นอินทรียธาตุ หรืออนินทรียธาตุ หมุนเวียนเป็นวัฏจักรต่อไป

การที่จะบอกได้ว่าเป็นผลไม้อะไรนั้น จำเป็นต้องมีสิ่งบ่งชี้อื่นๆ ประกอบหลายอย่าง เช่น เปลือกมีลักษณะเป็นหนาม และแข็ง เนื้อข้างในสีเหลือง หมายถึง ทุเรียน เป็นต้น ผลไม้ในความหมายทั่วไป หมายถึง ผลไม้ที่สามารถรับประทาน โดยไม่ต้องนำไปปรุงในครัวก่อนแต่มีรสชาติที่ดี ซึ่งอาจจะต้องปอกเปลือกก่อนรับประทาน ดังนั้นอาหารหลายชนิดจึงเป็นผลไม้ในเชิงพฤกษศาสตร์แต่กลับถูกจัดว่าเป็นผักในเชิงการทำครัว อันได้แก่ ผลของพืชจำพวกฟัก (ฟักทอง แฟง และ แตงกวา) มะเขือเทศ ถั่วลันเตา ถั่วฝักยาว ข้าวโพด พริกหยวก และเครื่องเทศ

ผลไม้เศรษฐกิจ คือผลไม้ที่ทำรายได้เข้าประเทศ เป็นผลไม้ที่ส่งออกไปขายยังต่างประเทศ ซึ่งผลไม้นับเป็นพืชเศรษฐกิจที่สำคัญประเภทหนึ่งของประเทศไทย โดยผลไม้สามารถทำรายได้เข้าประเทศปีละหลายพันล้านบาท และผลไม้ไทยยังเป็นที่นิยมบริโภคกันทั่วไปทั้งในประเทศและต่างประเทศ (สำนักงานเศรษฐกิจการเกษตร, 2555) นอกจากนี้ ความต้องการบริโภคผลไม้นับวันจะเพิ่มสูงขึ้น เนื่องมาจากสาเหตุหลักคือจำนวนประชากรที่เพิ่มมากขึ้นและความสนใจในสุขภาพก็มีมากขึ้นด้วย ในการผลิตผลไม้ถือได้ว่าไทยเป็นประเทศที่มีสภาพพื้นที่และภูมิอากาศที่เหมาะสมต่อการผลิตผลไม้หลายชนิด ส่วนฤดูกาลให้ผลผลิตผลไม้แต่ละชนิดก็ยังแตกต่างกันไปในแต่ละสภาพพื้นที่ของแต่ละภาค จึงเป็นข้อดีประการหนึ่งที่ส่งผลให้ไทยมีผลไม้หลากหลายชนิดหมุนเวียนออกสู่ตลาดตลอดทั้งปี

2.8 เอกสารและงานวิจัยที่เกี่ยวข้อง

การจัดการฐานความรู้ และการค้นคืนความรู้ มีงานวิจัยที่เกี่ยวข้องเพื่อให้การจัดการฐานความรู้เป็นไปอย่างมีประสิทธิภาพ และสามารถค้นคืนความรู้ได้อย่างรวดเร็ว และถูกต้อง ตรงตามความต้องการของผู้ใช้ สำหรับงานวิจัยที่เกี่ยวข้อง แสดงตามงานวิจัยของผู้วิจัยแต่ละท่าน ดังนี้

เกศสุดา ตุริยกรณ์ และชุลีรัตน์ จรัสกุลชัย (2546) ได้ทำการวิจัย ระบบคำถาม-คำตอบประยุกต์ใช้กับข้อความภาษาไทย โดยการเก็บรวบรวมเอกสารภาษาไทย และค้นคืนให้ได้คำตอบที่ต้องการ และได้นำความน่าจะเป็นเชิงสถิติเข้ามาประยุกต์ใช้ในระบบคำถาม-คำตอบ เช่น การจัดหมวดหมู่เอกสาร (Text Classification) เพื่อลดความยุ่งยากในการวิเคราะห์เชิงภาษาศาสตร์ และเพื่อความสะดวกในการจัดเอกสารให้ตรงกับประเภทของคำตอบที่ต้องการ ซึ่งในการทดลองเป็นคำถามที่เกี่ยวกับบุคคล (Who Question) คำถามเกี่ยวกับสถานที่ (Where Question) และคำถามเกี่ยวกับเวลา (When Question) และจะนำเอกสารที่มีหมวดหมู่ตรงกับคำถามมาเลือกคำตอบ (Answer Selection) โดยใช้อัตราเปรียบเทียบการเกิดร่วมกัน (Co-occurrence ratio) เพื่อให้ได้เอกสารที่มีความถูกต้อง และใกล้เคียงกับคำถามมากที่สุด ซึ่งในงานวิจัยนี้ใช้วิธีการวัดประสิทธิภาพของระบบโดย

ใช้ค่า MRAR (Mean Reciprocal Answer Rank) และใช้ผลลัพธ์ที่ได้จากรายการคำตอบใน 5 อันดับเฉลี่ยด้วยจำนวนคำถามทั้งหมด 60 คำถามที่ใช้ในการทดสอบ แบ่งเป็นคำถามเกี่ยวกับบุคคล คำถามเกี่ยวกับสถานที่ และคำถามเกี่ยวกับเวลา อย่างละ 20 คำถาม ถ้าคำตอบไม่ได้อยู่ภายใน 5 อันดับหรือไม่มีคำตอบที่ถูกต้องจะได้คะแนนเป็นศูนย์ จากการศึกษาพบว่า เปอร์เซ็นต์ของการได้รับคำตอบที่ตรงกับคำถามเป็น 63.3 % ส่วนการวัดประสิทธิภาพได้ค่าของการวัดประสิทธิภาพเป็น 0.518 ซึ่งถือเป็นค่าที่น่าพอใจ

เพ็ญพรรณ อัศวนพเกียรติ, อรินทิพย์ ธรรมชัยพิเนต และกฤษณะ ไวยมัย (2546) ได้ทำการวิจัย ออนโทโลยีชีวภาพ: ระบบสำหรับสืบค้น และวิเคราะห์ข้อมูลทางด้านชีววิทยา โดยนำจุดเด่นของภาษาสืบค้น และการใช้ภาพเพื่อสื่อความเข้าใจมารวมกัน และมีการพิจารณาเส้นทางที่ดีที่สุดด้วยการพิจารณาคุณภาพของแผนการสืบค้น ซึ่งได้พัฒนา 3 ส่วน ได้แก่ ส่วนการพัฒนาฐานความรู้ โดยการใช้ออนโทโลยีจากระบบ BACIIS มาพัฒนา และเปลี่ยนแปลงให้มีความเข้าใจมากขึ้น ส่วนการรวบรวมข้อมูล ได้เพิ่มประสิทธิภาพของการสืบค้นด้วยการพิจารณาคุณภาพของแผนการสืบค้น และความสามารถในการสืบค้นของแต่ละฐานข้อมูล และส่วนการพัฒนารูปแบบการสืบค้นที่เหมาะสม ได้มีการนำข้อดีของการสืบค้น (query language) และการใช้ภาพเพื่อสื่อความเข้าใจ (graphic user interface) มารวมกัน จากการศึกษาพบว่า เมื่อเปรียบเทียบระบบ TAMBIS และ BACIIS กับระบบดังกล่าว ผลลัพธ์ที่ได้ คือระบบสามารถสืบค้นข้อมูลที่ซับซ้อนมากกว่า และใช้งานง่ายเหมาะสำหรับผู้ใช้งานเริ่มต้น และนักชีววิทยา เนื่องจากระบบ TAMBIS ใช้งานยาก และช้า เพราะการออกแบบออนโทโลยีค่อนข้างซับซ้อน และยังไม่สามารถระบุเงื่อนไขของฐานข้อมูลที่จะค้นหาได้

อานนท์ กรเสวกวิสัย (2552) ได้ทำการวิจัยระบบสืบค้นรูปภาพบนอินเทอร์เน็ต โดยใช้หลักการเว็บเชิงความหมาย ซึ่งงานวิจัยนี้ได้ออกแบบโครงสร้างในการอธิบายรายละเอียดการสืบค้นด้วยรูปภาพสถานที่ท่องเที่ยวในประเทศไทย โดยจำลองการสืบค้นข้อมูลด้วยรูปภาพ ให้อยู่บนพื้นฐานของแบบจำลองการสืบค้นสารสนเทศ ประกอบด้วยกระบวนการสำหรับสืบค้นที่สำคัญ 2 กระบวนการ คือ (1) การสืบค้นรูปภาพบนเว็บ และ (2) การจัดเตรียมข้อมูลสำหรับการสืบค้น ในการพัฒนาต้นแบบดังกล่าว มีการออกแบบออนโทโลยีออกเป็น 4 ระดับ ประกอบด้วย ระดับประเภทของรูปภาพ ระดับประเภทสถานที่ท่องเที่ยว ระดับประเภทหมวดย่อยของสถานที่ท่องเที่ยว และคุณสมบัติของสถานที่ท่องเที่ยว ส่วนในการเตรียมข้อมูลสำหรับการสืบค้นนั้น มีการออกแบบขั้นตอนการนำเข้าข้อมูล และขั้นตอนการประมวลผล เพื่อเตรียมข้อมูลรูปภาพให้อยู่ในรูปแบบที่เหมาะสมสำหรับการสืบค้น เป็นการประมวลแบบ Batch Processing โดยมีเป้าหมายเพื่อทำหน้าที่เป็นตัวแทนเชิงความหมายของแต่ละรูปภาพ เพื่อเป็นข้อมูลสำหรับการเปรียบเทียบในการสืบค้น ด้วยการสร้าง Lexical Chain จากคำบรรยายภาพ และการบรรยายภาพด้วยภาษา RDF และการทำดัชนี จากการศึกษาพบว่า ข้อมูลที่มีความหมายตรงกับข้อมูลสืบค้นหรือมีความหมายเหมือนกับข้อมูลที่สืบค้น และในกรณีที่ในหน้าเว็บไซต์ประกอบด้วยรูปภาพ 1 ภาพ ผลลัพธ์ที่ได้จะตรงกับความหมายที่แท้จริงของรูปภาพมากกว่า กรณีที่

เว็บไซต์มีรูปภาพตั้งแต่ 2 ภาพขึ้นไป เนื่องจากยังไม่สามารถระบุคำอธิบายภาพว่าส่วนใดคือคำอธิบายของแต่ละภาพ นอกจากนี้ ได้มีการประเมินประสิทธิภาพการสืบค้นด้วยค่า Precision และ Recall โดยผลที่ได้อยู่ในระดับปานกลาง

นฤพนธ์ พนาวงศ์ และ จักรกฤษณ์ เสน่ห์ (2553) ได้ทำการวิจัย ระบบค้นหาสถานที่ท่องเที่ยวในประเทศไทย ด้วยหลักการออนโทโลยีแลเนมแมทชิ่ง ซึ่งงานวิจัยนี้ได้พัฒนาระบบค้นหาสถานที่ท่องเที่ยวในประเทศไทย โดยใช้หลักการออนโทโลยีในการออกแบบข้อมูลท่องเที่ยวต่าง ๆ สำหรับการค้นหาเพื่อให้ได้ผลลัพธ์ที่ตรงกับความต้องการ และครอบคลุมข้อมูลท่องเที่ยวมากที่สุด และใช้ทฤษฎีเนมแมทชิ่ง เพื่อแก้ปัญหาคำที่มีการเขียนผิดหรือไม่ถูกต้องตามหลักภาษาไทย และภาษาอังกฤษที่ปรากฏอยู่บนเว็บไซต์การท่องเที่ยวต่าง ๆ เพื่อให้ระบบทำการค้นหาสถานที่ท่องเที่ยวในประเทศไทยได้อย่างมีประสิทธิภาพ นอกจากนั้นได้มีการใช้วิธีการผสม (Hybrid Approaches) คือการนำอัลกอริทึมหลาย ๆ แบบมารวมกัน เพื่อให้ครอบคลุมปัญหาทั้งหมดของการเกิดคำพ้องรูป และพ้องเสียง ใช้การรวมกันจากกฎพื้นฐานของตัวสะกด และกฎพื้นฐานของการออกเสียงในการสร้างคำที่สามารถเขียนได้หลากหลายทั้งในภาษาไทย และภาษาอังกฤษ โดยใช้หลักการจากการที่ผู้ใช้ป้อนคำค้น ระบบจะทำการสร้างคำต่าง ๆ ที่สามารถเขียนได้หลายรูปแบบ และทำการดึงข้อมูลเหล่านั้นจากกูเกิ้ลแล้วแสดงผลให้กับผู้ใช้ ระบบดังกล่าวมีความสามารถในการนำข้อมูลต่าง ๆ ไปใช้ในการวางแผนการท่องเที่ยว และจัดแพ็คเกจการท่องเที่ยว เพื่อตอบสนองความต้องการต่อผู้ใช้ให้ได้รับความสะดวกรวดเร็วในการวางแผนการท่องเที่ยวอีกด้วย โดยการประเมินประสิทธิภาพของระบบ ประเมินจากการค้นคืนข้อมูลต่าง ๆ ที่ได้มีความครบถ้วนมากที่สุด โดยการแสดงความสัมพันธ์ของสถานที่ต่าง ๆ ได้อย่างครบถ้วน จากผลการศึกษาพบว่า การค้นหาข้อมูลสถานที่ท่องเที่ยวที่มีชื่อที่คล้ายกัน ทั้งภาษาไทย และภาษาอังกฤษ ทำให้เกิดประโยชน์กับนักท่องเที่ยวชาวต่างชาติมากขึ้น ส่วนในเรื่องของการวางแผนการท่องเที่ยวนั้น ได้แสดงผลในรูปแบบแผนที่ และมีการแนะนำเส้นทางโดยแสดงลูกศรลำดับการเดินทาง ทำให้นักท่องเที่ยวตัดสินใจได้รวดเร็ว และสะดวกสบาย

นฤพนธ์ พนาวงศ์ และจักรกฤษณ์ เสน่ห์ (2554) ได้ทำการวิจัยการพัฒนาระบบสืบค้นข้อมูลออนโทโลยีท่องเที่ยวด้วยภาษา SPARQL และจัดเก็บข้อมูลท่องเที่ยวทั้งหมด มีการจัดกลุ่มข้อมูลตามความสัมพันธ์ไว้ด้วยกันด้วยหลักการออนโทโลยี โดยแสดงตัวอย่างการเก็บข้อมูลในไฟล์ OWL มีการใช้ RDF API หรือเรียกว่า RAP สำหรับภาษา PHP เป็นเครื่องมือในการสืบค้นข้อมูลด้วยเทคโนโลยีเว็บเชิงความหมาย ซึ่งเมื่อป้อนชื่อจังหวัด หรือแหล่งท่องเที่ยวที่ต้องการค้นหา ระบบจะเชื่อมต่อไปยังฐานข้อมูลออนโทโลยีท่องเที่ยวผ่านทาง Ontology Access API ผ่านภาษา SPARQL แล้วส่งต่อไปยัง Ontology Mediation Layer เพื่อทำการประมวลผล แล้วนำไปตรวจสอบว่าถูกต้องตามรูปแบบของ SPARQL หรือไม่ แล้วแสดงข้อมูลพื้นฐานของจังหวัดนั้นต่อผู้ใช้ จากการศึกษาพบว่า เมื่อนักท่องเที่ยวป้อนคำค้นหาสถานที่ท่องเที่ยว หรือจังหวัดที่ต้องการรู้ข้อมูลการท่องเที่ยว ระบบสามารถแสดงข้อมูลได้อย่างถูกต้อง และมีความสัมพันธ์กันระหว่างจังหวัดนั้น ๆ แต่ก็พบว่าหากนักท่องเที่ยวป้อนคำค้นหาที่ไม่

ถูกต้องตามหลักภาษาไทย หรือภาษาอังกฤษระบบจะไม่สามารถค้นพบข้อมูล เนื่องจากไม่มีส่วนการตรวจสอบความคล้ายคลึงกันของคำค้นหานั้น ๆ

วิชุดา โชติรัตน์, ผุสดี บุญรอด และศรีมาจ ณ วิเชียร (2554) ได้ทำการวิจัยการพัฒนาฐานความรู้ออนโทโลยีสำหรับวิเคราะห์ข่าวออนไลน์อัตโนมัติ โดยศึกษาปัจจัยที่มีผลต่อการวิเคราะห์เนื้อหาข่าวออนไลน์ เพื่อการออกแบบ และพัฒนาฐานความรู้ออนโทโลยีสำหรับอธิบายสิ่งต่าง ๆ ที่เป็นองค์ประกอบที่สำคัญในเนื้อหาข่าว ทำให้ระบบสามารถวิเคราะห์เนื้อหา และอนุมาข้อมูลต่าง ๆ ที่เกี่ยวข้องในเนื้อหาข่าวได้อย่างมีประสิทธิภาพ โดยมีการพัฒนาฐานความรู้ ออนโทโลยีโดเมนข่าวออนไลน์ ภายใต้ขอบเขตข่าวที่เกี่ยวข้องกับพื้นที่ 5 จังหวัดชายแดนภาคใต้ ซึ่งประกอบด้วยคลาสเวลาที่เกิดเหตุการณ์ บุคคลที่เกี่ยวข้อง รูปแบบเหตุการณ์ อาวุธ และผลกระทบ และได้ทำการส่งออกข้อมูลให้อยู่ในรูปแบบของภาษา OWL แล้วได้ทดลองกับข่าวจำนวน 200 ข่าว และทดสอบฐานความรู้ออนโทโลยีด้วยการแบ่งประเด็นการทดสอบออกเป็น 4 ด้าน ได้แก่ 1) การสื่อความหมาย (Interpretability: I) 2) ความถูกต้อง (Accuracy: A) 3) ความชัดเจน (Clarity: C) และ 4) ความครอบคลุม (Comprehensiveness: O) จากการทดสอบพบว่า ประเด็นทั้ง 4 ด้าน ผ่านการพิจารณาจากการประเมินผลจากผู้เชี่ยวชาญ ซึ่งอยู่ในระดับดี ทั้ง 4 ด้าน และจากผลการวิจัยสามารถนำฐานความรู้ออนโทโลยีที่พัฒนา ไปประยุกต์ใช้ร่วมกับการวิเคราะห์ข้อมูลด้วยเทคนิคการทำเหมืองข้อมูล หรือเทคนิคอื่น ๆ เพื่อประโยชน์ในการวิเคราะห์ข้อมูลที่รวดเร็ว และการศึกษาความสัมพันธ์ของสิ่งต่าง ๆ ที่สนใจจากเนื้อหาข่าวออนไลน์ภาษาไทยได้ในอนาคต

จุฑาวรรณ สิทธิโชคสถาพร (2555) ได้ทำการวิจัยต้นแบบออนโทโลยี เพื่อการค้นคืนสารสนเทศเชิงความหมาย สำหรับงานสารบรรณอิเล็กทรอนิกส์ กรณีศึกษางานบริหารธุรกิจ คณะแพทยศาสตร์ มหาวิทยาลัยสงขลานครินทร์ ซึ่งวิทยานิพนธ์นี้ได้นำแนวคิดออนโทโลยีมาพัฒนาการค้นคืนสารสนเทศเชิงความหมายสำหรับงานสารบรรณอิเล็กทรอนิกส์ และกำหนดการพัฒนาข้อมูลเชิงความหมายเฉพาะข้อมูลชนิดที่มีโครงสร้าง ให้สอดคล้องกับฐานข้อมูลเชิงความสัมพันธ์ในระบบงานสารบรรณอิเล็กทรอนิกส์ในปัจจุบัน ซึ่งผู้ใช้เป็นผู้สร้างข้อมูลเชิงความหมายนั้น โดยระบุข้อมูลจากรูปแบบที่ถูกออกแบบจากการพิจารณาตามแนวคิดออนโทโลยี ซึ่งในการประเมินผลระบบนั้น โดยการสร้างแบบสอบถามสำหรับออนโทโลยี ที่เกี่ยวข้องกับคลาส (Class) คุณสมบัติ (Slot) ตัวดำเนินการ (Operator) และค่าของข้อมูล (Value) จากการศึกษาพบว่า การสืบค้นจากต้นแบบออนโทโลยีช่วยขยายคำค้นที่นอกเหนือจากที่ผู้ใช้ระบุ ทำให้สามารถค้นคืนสารสนเทศได้ครอบคลุมตามความหมายที่ผู้ใช้ต้องการมากยิ่งขึ้น เมื่อเปรียบเทียบกับการค้นคืนจากระบบเดิมของงานสารบรรณ

พยุง มีสัจ, วาทินี นุ้ยเพียร และผุสดี บุญรอด (2556) ได้ทำการวิจัยระบบค้นคืนเชิงความหมายจากข้อมูลบรรณานุกรมโดเมน Information System โดยพัฒนาระบบการค้นคืนเชิงความหมาย ที่เป็นการประยุกต์ใช้งาน เพื่อให้ผู้ใช้สามารถค้นคืนข้อมูลงานวิจัยโดเมน Information System ได้ตรงตามความต้องการของผู้ใช้มากที่สุด ซึ่งในการพัฒนาระบบได้ใช้ฐานข้อมูล MySQL ร่วมกับโครงสร้างออนโทโลยี ที่

ออกแบบโดยใช้มาตรฐานของภาษา XML เพื่อใช้เก็บข้อมูลรูปแบบเอกสารที่มีความแตกต่างกัน ให้สามารถใช้งานร่วมกันได้ ซึ่งเป็นการใช้ออนโทโลยีแบบ Taxonomic ที่มีความสัมพันธ์แบบ Is-a และความสัมพันธ์แบบ Part-of หลังจากนั้นทำการตรวจสอบความสัมพันธ์ของโครงสร้าง เพื่อให้ได้ผลลัพธ์ที่ถูกต้อง มีการนำภาษา RDF มาใช้อธิบายรูปแบบความสัมพันธ์ของข้อมูล และมีการใช้ OWL อธิบายโครงสร้างข้อมูล โดยใช้การประเมินประสิทธิภาพด้วยการวัดค่าความถูกต้อง ค่าความแม่นยำ และการวัดประสิทธิภาพโดยรวม จากผู้เชี่ยวชาญจำนวน 3 ท่าน โดยทำการทดสอบการค้นคืนจากระบบที่พัฒนาขึ้นด้วยจำนวนคำค้น 15 คำค้น และให้ผู้เชี่ยวชาญพิจารณาผลลัพธ์ของคำที่ได้ กับชื่อบทความวิจัย โดยการให้คะแนนถูก-ผิด แล้วนำคะแนนจากทั้ง 3 ท่าน มาหาค่าเฉลี่ย ซึ่งหากคะแนนที่ได้จากการค้นคืนตรงอย่างน้อย 2 ท่าน ถือว่าถูกต้อง เนื่องจากมีคะแนนเกินครึ่ง ซึ่งผลที่ได้จากการวัดค่าความแม่นยำ คือ 97.09% ถือว่าอยู่ในระดับที่น่าพอใจ

เสกสรรค์ ศิวิลัย และจักรกฤษณ์ เสน่ห์ นมะหุต (2556) ได้ทำการวิจัยการพัฒนาระบบคำถาม-คำตอบ ในการแนะนำอาหารที่เหมาะสมสำหรับผู้ป่วย โดยเน้นไปที่การแสดงวิธีการรวบรวมและวิธีการแปลงประโยคคำถามภาษาไทยเกี่ยวกับอาหารที่เหมาะสมสำหรับผู้ป่วยที่เป็นภาษาธรรมชาติไปเป็นภาษาที่ใช้ในการสอบถามฐานข้อมูลความรู้ออนโทโลยี (SPARQL) เริ่มจากผู้ใช้ทำการป้อนคำถามด้วยภาษาธรรมชาติ จากนั้นระบบจะทำการตรวจสอบ และปรับปรุงคำถามให้อยู่ในรูปแบบที่ถูกต้อง โดยการแก้ไขคำผิด หรือคำที่มีความหมายเหมือนกัน จากนั้นทำการตัดคำในประโยคร่วมกับคำที่ได้จากการรวบรวมคำถามจากเว็บไซต์สนทนา แล้วทำการวิเคราะห์คำถาม และจัดรูปแบบ ก่อนทำการแปลงให้เป็นภาษาที่ใช้ในการสอบถามฐานข้อมูลความรู้ ออนโทโลยี จากนั้นทำการสืบค้นข้อมูลจากออนโทโลยีอาหารซึ่งทำงานร่วมกับกลไกอนุมาน จากการศึกษาพบว่า ระบบดังกล่าวช่วยให้ผู้ป่วยได้รับคำแนะนำในการบริโภคอาหารได้ถูกต้อง และเหมาะสมกับผู้ป่วยได้อย่างมีประสิทธิภาพ และสะดวกรวดเร็วมากขึ้น แต่อย่างไรก็ตาม ยังมีประโยคคำถามในบางกรณีที่ยากต่อการสกัด ซึ่งมีผลต่อการแนะนำอาหารที่เหมาะสมสำหรับผู้ป่วย เช่น คำถามที่มีความยาว และคลุมเครือ ยังใช้ไม่ได้กับระบบนี้ หรือใช้กับระบบนี้แล้วได้ผลลัพธ์ที่ไม่มีประสิทธิภาพเพียงพอ

คิม และคณะ (Kim et al., 2001) ได้ทำการวิจัยระบบคำถามตอบข้อความภาษาเกาหลี โดยใช้วิธีการประมวลผลภาษาธรรมชาติ และการค้นคืนสารสนเทศ มีการแบ่งประเภทคำ, การสกัดลักษณะนาม, การสร้างรูปแบบของการเข้าคู่กัน และใช้การค้นคืนสารสนเทศในการหาน้ำหนักของคำด้วยความถี่ของคำและค่าส่วนกลับของความถี่เอกสาร ซึ่งมีกระบวนการ 2 ขั้นตอน ได้แก่ 1) การสร้างดรรชนี ในการค้นหาคำตอบโดยการสกัดประโยคที่เป็นคำนามจากเอกสารเทียบกับคำในพจนานุกรมที่คาดว่าน่าจะเป็นคำตอบที่เป็นไปได้จากกลุ่มเอกสาร และให้ค่าน้ำหนักของคำในแต่ละประโยคไว้ในฐานข้อมูล ซึ่งมีการแบ่งประเภทตามหมวดหมู่ของการวิเคราะห์ความหมายของคำที่แตกต่างกัน และ 2) การค้นหา จะเป็นกระบวนการแบ่งกำหนดข้อคำถามจากผู้ใช้ โดยเลือกดรรชนีจากฐานข้อมูลที่ประกอบด้วยประโยคที่คาดว่าจะเป็นคำตอบกับข้อคำถามจากผู้ใช้นำมาวัดความคล้ายคลึงระหว่างข้อคำถามกับประโยคคำตอบ และทำการจัดลำดับผลลัพธ์คำตอบ

แฮมโม และคณะ (Hammo and et al., 2002) ได้ทำการวิจัยระบบคำถามตอบในภาษาอารบิค โดยการประมวลผลภาษาธรรมชาติทำงานร่วมกับการค้นคืนสารสนเทศ ซึ่งการประมวลภาษาธรรมชาติวิเคราะห์คำในไวยากรณ์ของคำถามและวิเคราะห์เอกสารเพื่อนำไปวิเคราะห์หาประโยคคำตอบให้กับข้อคำถาม และมีการค้นคืนสารสนเทศจากการวิเคราะห์คำถามและสร้างเป็นข้อคำถามสำหรับการค้นคืน

2.9 กรอบแนวคิดการวิจัย

กรอบแนวคิดการวิจัยผู้วิจัยนำเสนอในรูปแบบการเปรียบเทียบงานวิจัยที่เกี่ยวข้อง ซึ่งสามารถจำแนกได้ 3 ประเด็น คือ ฐานความรู้ การถามตอบ และการสืบค้น แสดงดังตารางที่ 2.6

ตารางที่ 2.6 สรุปเปรียบเทียบงานวิจัยที่เกี่ยวข้องระบบถามตอบ

ผู้วิจัย	ฐานความรู้		การถามตอบ			การสืบค้น
	ออนโทโลยี	การจัดกลุ่มความรู้	การวิเคราะห์คำถาม	การเปรียบเทียบคำถามตอบ	การสร้างคำตอบ	
เกศสุดา ตุริยกรณ์และชุลีรัตน์ จรัสกุลชัย (2546)		√	ตัดคำให้อยู่ในรูปเวกเตอร์และแบ่งประเภทคำถาม	แบ่งเอกสารเป็นข้อความสั้น และจัดหมวดหมู่เอกสารตามประเภทคำตอบ	หาอัตราเปรียบเทียบการเกิดร่วมกัน	การบีบอัดเมตริกซ์
ยุทธศักดิ์ วุฒิกร, และคณะ (2553)			ตัดคำและวิเคราะห์คำในประโยคด้วย Stanford Parser	สร้างตัวแทนเชิงความหมายและพิจารณาจากความสัมพันธ์ของกราฟมโนภาพ	ชุดความสัมพันธ์ของกราฟเทียบกับกฎ	
นฤพนธ์ พนาวงศ์และจักรกฤษณ์ เสน่ห์ (2554)	√	√				วัดความคล้ายคลึงกับคำค้น

มหาวิทยาลัยเทคโนโลยีสุรนารี

ตารางที่ 2.6 สรุปเปรียบเทียบงานวิจัยที่เกี่ยวข้องระบบถามตอบ

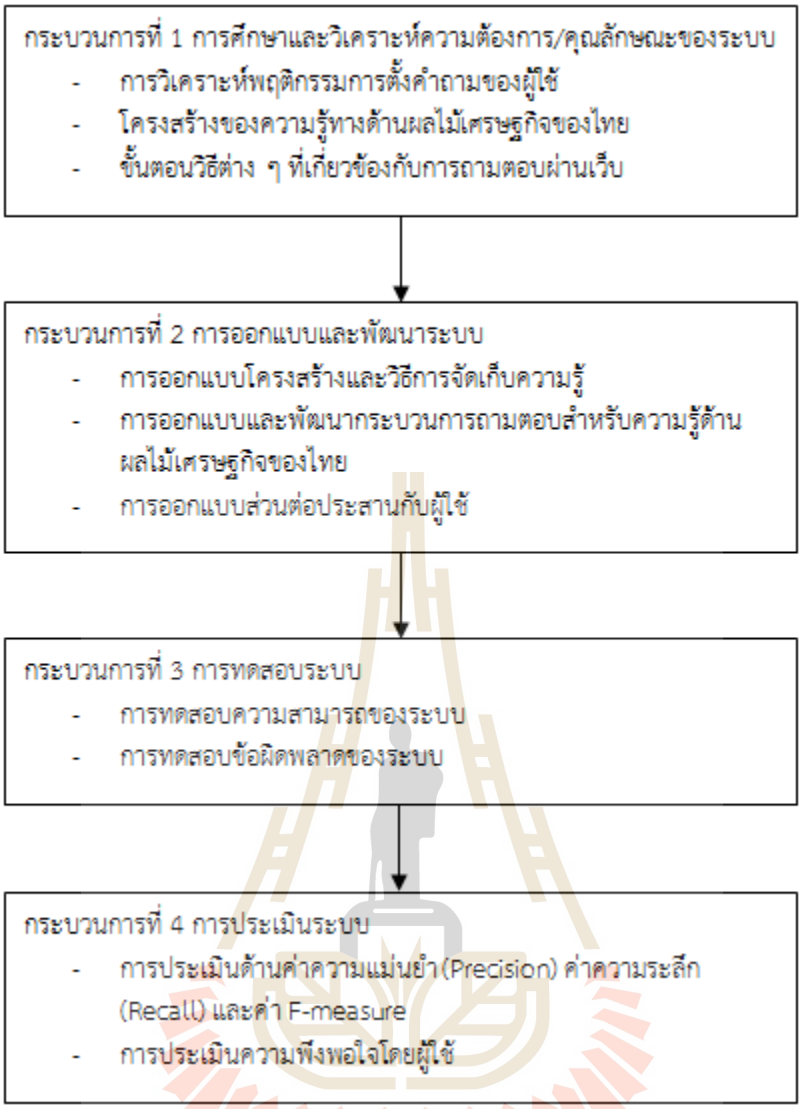
ผู้วิจัย	ฐานความรู้		การถามตอบ			การสืบค้น
	ออนโทโลยี	การจัดกลุ่มความรู้	การวิเคราะห์คำถาม	การเปรียบเทียบคำถามตอบ	การสร้างคำตอบ	
สิงห์ทัย สุขสว่าง โรจน์ และระพีพรรณ พิริยะกุล (2558)			ตัดคำโดยใช้พจนานุกรม และวิเคราะห์คำจัดหมวดหมู่	การแจงประโยคด้วยกฎ	แบ่งกลุ่มด้วย K-Mean หารูปแบบคำตอบ	
Luis and Antonio, 2000		√	สร้างตัวแทนคำถามจากคำสำคัญ	ใช้ตัวแทนวิเคราะห์ความหมายที่แตกต่างของคำตามลักษณะคำตอบ	วัดความคล้ายคลึงระหว่างตัวแทนประโยคคำถามกับประโยคที่ได้จากเอกสาร	
Kim et al. (2001)		√	แบ่งประเภทคำและสกัดลักษณะนาม	เปรียบเทียบคำนามกับคำในพจนานุกรม		วัดความคล้ายคลึงระหว่างข้อคำถามกับประโยค
Hammo and et al. (2002)	√	√	ตัดคำ และวิเคราะห์ไวยกรณ์ของคำถาม	เปรียบเทียบประโยคคำถามกับคำตอบ		
Radev and et al.(2005)			สร้างตัวแทนจากการแบ่งประเภทคำถาม	เปรียบเทียบประโยคคำถาม กับคำตอบที่	เลือกข้อความสั้น	

มหาวิทยาลัยเทคโนโลยีสุรนารี

ตารางที่ 2.6 สรุปเปรียบเทียบงานวิจัยที่เกี่ยวข้องระบบถามตอบ

ผู้วิจัย	ฐานความรู้		การถามตอบ			การสืบค้น
	ออนโทโลยี	การจัดกลุ่มความรู้	การวิเคราะห์คำถาม	การเปรียบเทียบคำถามตอบ	การสร้างคำตอบ	
				ปรากฏในเอกสาร		
ผู้วิจัย	√	√	ตัดคำ โดยใช้พจนานุกรมและวิเคราะห์คำจัดหมวดหมู่	เปรียบเทียบระหว่างคำถามที่ผู้ใช้ถาม กับคำตอบที่อยู่ในเอกสารด้วยแจ็กการ์ด	เปรียบเทียบไวยกรณ์และเลือกข้อความสั้น	คำนวณหาค่าน้ำหนักของคำค้นจากเส้นทางที่สั้นที่สุดและวัดความคล้ายคลึงระหว่างข้อคำถามกับประโยค

จากตารางที่ 2.6 ผู้วิจัยนำเสนอในรูปแบบการเปรียบเทียบงานวิจัยที่เกี่ยวข้องกับงานวิจัยนี้ ซึ่งมีการนำเอาขั้นตอนวิธีที่หลากหลายมาใช้สำหรับการออกแบบและพัฒนาระบบถามตอบเชิงความหมายออนไลน์สำหรับผลไม้เศรษฐกิจของไทย ซึ่งสามารถสรุปเป็นกรอบแนวคิดแสดงดังภาพที่ 2.13 ดังนี้

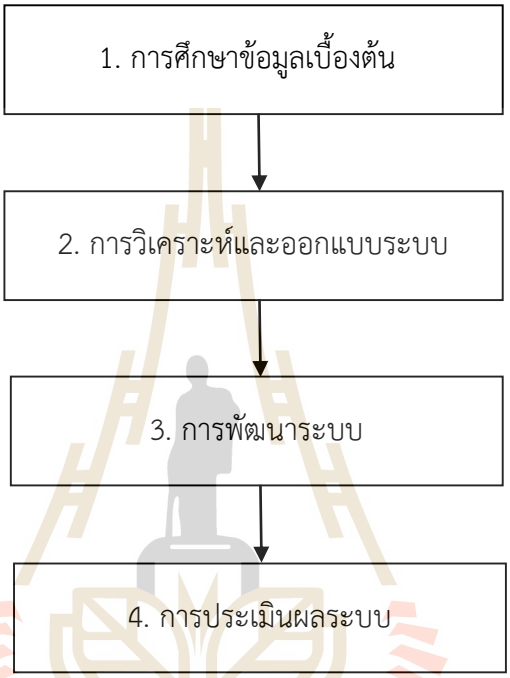


ภาพที่ 2.13 ระบบถามตอบเชิงความหมายออนไลน์สำหรับผลไม้เศรษฐกิจของไทย

มหาวิทยาลัยเทคโนโลยีสุรนารี

บทที่ 3 วิธีการดำเนินการวิจัย

การดำเนินการวิจัยในครั้งนี้ เป็นการวิจัยและพัฒนา โดยมีการประยุกต์ใช้ทฤษฎี หลักการและเทคนิคที่เกี่ยวข้องกับการถามตอบ การจัดการฐานความรู้ด้วยออนโทโลยี และการสกัดความรู้ (Knowledge Extraction) มาประยุกต์ใช้ให้เหมาะสมต่อการพัฒนาระบบถามตอบเชิงความหมายสำหรับผลไม้เศรษฐกิจของไทยควบคู่กับการพัฒนาระบบตามกระบวนการของวงจรการพัฒนาระบบ System Development Life Cycle (SDLC) โดยมีลำดับขั้นตอนในการวิจัยดังภาพที่ 3.1 ดังนี้



ภาพที่ 3.1 แสดงขั้นตอนการดำเนินการวิจัย

3.1 การศึกษาข้อมูลเบื้องต้น

การศึกษาข้อมูลเบื้องต้นจากเอกสารที่เกี่ยวข้องจากหนังสือ เอกสาร และข้อมูลต่าง ๆ บนระบบเครือข่ายอินเทอร์เน็ต ซึ่งประกอบไปด้วยที่สำคัญ ได้แก่ ข้อมูลระบบถามตอบเชิงความหมาย ข้อมูลผลไม้เศรษฐกิจของไทย และข้อมูลการวิจัยที่เกี่ยวข้อง เพื่อนำข้อมูลไปใช้ในการวิเคราะห์และออกแบบระบบ มีกระบวนการศึกษา วิเคราะห์ความต้องการและคุณลักษณะของระบบ ดังนี้

3.1.1 การวิเคราะห์พฤติกรรมการตั้งคำถามของผู้ใช้

จากการศึกษาถึงพฤติกรรมการตั้งคำถามของผู้ใช้ที่เกี่ยวข้องผลไม้เศรษฐกิจ ผู้วิจัยได้ทำการสำรวจพฤติกรรมการตั้งคำถามโดยเก็บรวบรวมข้อมูลดังนี้

1) ข้อมูลแบบปฐมภูมิ (Primary Data) โดยเป็นข้อมูลที่ผู้วิจัยได้เป็นผู้เก็บรวบรวมข้อมูลชุดคำถามที่เกี่ยวข้องกับผลไม้เศรษฐกิจของประเทศไทย จากการสอบถามเกษตรกรทั่วไป กรณีนึกถึงผลไม้ทุเรียน มังคุด และลำไย จะนึกถึงอะไร จำนวน 15 คน

2) ข้อมูลแบบทุติยภูมิ (Second Data) โดยเป็นข้อมูลที่ผู้วิจัยได้สืบค้นเกี่ยวกับคำถามจากผู้ใช้ที่นิยมสอบถามเกี่ยวกับผลไม้ที่เป็นทุเรียน มังคุด และลำไย จาก 30 เว็บไซต์ ดังตารางที่ 3.1

ตารางที่ 3.1 รายชื่อเว็บไซต์ที่แสดงถึงพฤติกรรมการตั้งคำถามที่เกี่ยวข้องผลไม้เศรษฐกิจ

ที่	ตำแหน่งที่อยู่เว็บไซต์	รายละเอียดคำถามเกี่ยวกับ
1	http://www.thaibizchina.com/thaibizchina/th/faqs/?SECTION_ID=640	การตลาดและการส่งออกผลไม้เศรษฐกิจ
2	http://www.oknation.net/blog/horti-asia/2013/04/24/entry-1	พันธุ์ การปลูกดูแลรักษา โรคและแมลง สารเคมีที่ใช้ป้องกันโรคและแมลง การแปรรูปเครื่องจักรหรือเครื่องมือการผลิด การแปรรูป การตลาดและส่งออกของมังคุด
3	http://www.oknation.net/blog/horti-asia/2013/04/30/entry-2	การปลูก การดูแลรักษา โรคและแมลง การแปลรูปของมังคุด
4	http://www.eto.ku.ac.th/neweto/e-book/plant/tree_fruit/mangoste.pdf	การปลูกมังคุด
5	http://health.kapook.com/view93601.html	ประโยชน์ของมังคุด
6	http://medplant.mahidol.ac.th/user/qa.asp	โรคและแมลง การสกัด และสารเคมีที่ใช้ป้องกันโรคและแมลง
7	http://knowledge.kasetbay.com/	การปลูก ดูแลรักษาทุเรียน มังคุด ลำไย
8	http://www.thaidurianloverclub.com/th/store/faqs	พันธุ์ การปลูก ดูแลรักษาทุเรียน
9	http://www.soursopherbs.com/webboard/viewtopic/1	ทุเรียน
10	http://www.soursopcancer.com/article/25/15-คำถามที่พบบ่อย-สำหรับผลิตภัณฑ์ทุเรียนเทศรักษามะเร็ง	การแปรรูปและผลิตภัณฑ์ทุเรียน

มหาวิทยาลัยเทคโนโลยีสุรนารี

ตารางที่ 3.1 รายชื่อเว็บไซต์ที่แสดงถึงพฤติกรรมการตั้งคำถามที่เกี่ยวข้องผลไม้เศรษฐกิจ

ที่	ตำแหน่งที่อยู่เว็บไซต์	รายละเอียดคำถามเกี่ยวกับ
11	https://th-th.facebook.com/เทคนิคสวนทุเรียน-784218128294625/	ถามตอบเกี่ยวกับทุเรียน
12	http://www.kasetloongkim.com/modules.php?name=Forums&file=viewtopic&t=4558	รวบรวมปัญหาการถามตอบปัญหาเกษตรกร ทางรายการวิทยุ 12 AUG AM 594 เวลา 08.10-09.00 & 20.05-20.30 ทุกวัน และ FM 91.0 (07.00-08.00 / วันอาทิตย์)
13	http://www.kasetporpeang.com/forums/index.php?topic=100995.0;wap2	ถามตอบปัญหาเกษตรกร
14	http://www.myquestionth.com/question/5035	ถามตอบปัญหาเกษตรกร
15	http://www.thalingchan.go.th/webboard2/webboard.php?Category=board	แลกเปลี่ยนความรู้เกี่ยวกับทุเรียน
16	https://www.moac.go.th/faq_detail.php?fa_id=1081	กระทู้ถามตอบเกี่ยวกับทุเรียนของกระทรวงเกษตรและสหกรณ์
17	https://www.doctor.or.th/article/detail/7518	มหัศจรรย์ทุเรียน ราชาของผลไม้
18	https://www.gotoknow.org/posts/514467	GotoKnow เกี่ยวกับลำไย
19	https://web.facebook.com/groups/www.longankipqew/?_rdr	ปัญหาถามตอบเกี่ยวกับลำไยผ่านสังคมออนไลน์
20	http://bengaltube.com/watch/7oGqJERhFR4	คำถามที่พบบ่อยเกี่ยวกับลำไยของชาวบ้าน
21	http://m.naewna.com/view/agriculture/165941	ปัญหาเกษตรที่นี่มีคำตอบ : ลำไยกับการปรับปรุงบำรุงดิน
22	https://www.moac.go.th/faq_list.php?offset=1040&f_id=&f_sub_id=2&filename=index&BID= https: / / sites. google. com/ site/ webhlakhlay/ baeb-thdsxb-reuxng-khun-ru-khae-hin-keiyw-kab-lai	แบบข้อคำถามเกี่ยวกับพันธุ์ การปลูกดูแลรักษา โรคและแมลง สารเคมีที่ใช้ป้องกันโรคและแมลง การแปรรูป เครื่องจักรหรือเครื่องมือการ

มหาวิทยาลัยเทคโนโลยีสุรนารี

ตารางที่ 3.1 รายชื่อเว็บไซต์ที่แสดงถึงพฤติกรรมการตั้งคำถามที่เกี่ยวข้องผลไม้เศรษฐกิจ

ที่	ตำแหน่งที่อยู่เว็บไซต์	รายละเอียดคำถามเกี่ยวกับ
		ผลิต การแปรรูป การตลาดและส่งออกของลำไย
23	http://www.opsmoac.go.th/faq_detail.php?fa_id=11	ถอดองค์ความรู้คำถามเกี่ยวกับลำไย
24	http://www.geotubepk.com/nBMaJo71O0E/.html	ถอดองค์ความรู้คำถามเกี่ยวกับดินกับการปลูกลำไย
25	http://m.naewna.com/view/agriculture/157593	คำถามปัญหาเกี่ยวกับการบำรุงดินของมังคุด
26	http://www.durianchan.com/durian/index.php?board=7.0	กลุ่มถามตอบทุเรียน มังคุด และลำไย เครือข่ายจันทบุรี
27	http://guru.sanook.com/16786/	ถามตอบเกี่ยวกับมังคุด
28	http://www.chemtrack.org/Board-Detail.asp?TID=0&ID=469	ถามตอบสารเคมีที่ใช้ป้องกันโรคและแมลง
29	https://th-th.facebook.com/เทคนิคสวนมังคุด-1049372465089389/	ถามตอบเกี่ยวกับมังคุด
30	http://www.kasetporpeang.com/forums/index.php?topic=110254.0	กระทู้ถามตอบเกี่ยวกับทุเรียน มังคุด ลำไย และเรื่องเกษตรอื่นๆ

จากการเก็บรวบรวมข้อมูลข้างต้นผู้วิจัยได้ประยุกต์ใช้แนวคิดของ วิลสัน(Wilson, 1997) ที่ได้อธิบายว่าคุณลักษณะเฉพาะบุคคล ลักษณะแหล่งสารสนเทศ เช่นการเข้าถึงและ ความน่าเชื่อถือของแหล่งสารสนเทศ มีผลต่อการแสวงหาสารสนเทศที่ต้องการ

จากการเก็บรวบรวมข้อมูลแบบปฐมภูมิ (Primary Data) โดยสอบถามเกษตรกรทั่วไป กรณีความต้องการสอบถามเกี่ยวกับผลไม้ทุเรียน มังคุด และลำไย จำนวน 15 คน และการเก็บข้อมูลแบบทุติยภูมิ (Second Data) จากข้อมูลใน 30 เว็บไซต์ ซึ่งผู้วิจัยได้วิเคราะห์การตั้งคำถามที่เกี่ยวข้องใน 7 กลุ่ม ได้แก่ พันธุ์ การปลูกและดูแลรักษา โรคและแมลง สารเคมีที่ใช้ป้องกันโรคและแมลง การแปรรูป เครื่องจักรหรือเครื่องมือการผลิต การตลาดและการส่งออก โดยจัดหมวดหมู่คำถามจากการวิเคราะห์การตั้งคำถามไว้ใน 5 หมวด ได้แก่ ใคร (Who) อะไร (What) ทำไม (Why) เมื่อไร (When) ที่ไหน (Where) และอย่างไร (How) ซึ่งได้จำนวนข้อคำถามและจัดหมวดหมู่ ดังตารางที่ 3.2

ตารางที่ 3.2 การวิเคราะห์พฤติกรรมการตั้งคำถาม

การวิเคราะห์ข้อคำถามผลไม้เศรษฐกิจที่เกี่ยวข้องกับทุเรียน มังคุด และลำไย	จำนวนข้อคำถามในการจัดหมวดหมู่						รวม
	Who	What	Why	When	Where	How	
1. พันธุ์	10	49	43	34	40	22	198
2. การปลูกและดูแลรักษา	10	54	33	34	40	5	176
3. โรคและแมลง	7	23	75	54	32	2	193
4. สารเคมีที่ใช้ป้องกันโรคและแมลง	0	25	38	21	13	6	103
5. การแปรรูป	0	28	20	12	15	5	80
6. เครื่องจักรหรือเครื่องมือการผลิต	18	30	10	18	8	8	92
7. การตลาดและการส่งออก	8	34	25	21	10	10	108
รวมข้อคำถามทั้งหมด	53	243	244	194	158	58	950

เนื่องจากข้อคำถามที่จะนำมาทดสอบสำหรับการวิจัยนี้มีจำนวนมาก ดังนั้นผู้วิจัยได้คัดเลือกข้อคำถามเพื่อนำมาใช้สำหรับการถามตอบในระบบถามตอบเชิงความหมายสำหรับผลไม้เศรษฐกิจของไทยครั้งนี้กำหนดขนาดตัวอย่าง (จำนวนข้อคำถาม) ใช้วิธีเปิดตาราง Yamane (Taro Yamane. 1967 : 886) เพื่อหาจำนวนขนาดกลุ่มตัวอย่างจากกลุ่มการตั้งคำถาม และการจัดหมวดหมู่คำถามโดยได้กำหนดความคลาดเคลื่อนของการสุ่มตัวอย่างที่ยอมให้เกิดระหว่างค่าจริงและค่าประมาณร้อยละ 0.05 ตามสูตร

จากสูตร $n = \frac{N}{1+N(e)^2}$

เมื่อ n คือ จำนวนตัวอย่างข้อคำถาม หรือขนาดของกลุ่มตัวอย่าง

N คือ จำนวนข้อคำถามทั้งหมด

e คือ ความคาดเคลื่อนในการสุ่มตัวอย่าง (Sampling error) ในที่นี้จะกำหนดเท่ากับ +/-0.05 ภายใต้ความเชื่อมั่น 95%

แทนค่าสูตร $n = \frac{950}{1+950(0.05)^2}$

$n = 399.57$ หรือ 400 ข้อคำถาม

อย่างไรก็ตาม เพื่อให้ผลการวิจัยที่ได้รับจากกลุ่มตัวอย่างที่สุ่มมาสามารถใช้อธิบายกลุ่ม ข้อคำถามเป้าหมายได้อย่างถูกต้อง และเพื่อเป็นการเพิ่มความเที่ยงตรงของข้อสรุปที่จะได้รับจากตัวอย่าง จึงใช้กลุ่มตัวอย่างจำนวน 400 ข้อคำถาม โดยมีการสุ่มตัวอย่างแบบอาศัยความน่าจะเป็น (Probability

sampling) เพื่อจะได้เป็นการกระจายกลุ่มตัวอย่างให้ครอบคลุมกลุ่มการตั้งคำถาม และการจัดหมวดหมู่คำถามและไม่ให้มีการกระจุกตัวอยู่ในกลุ่มการตั้งคำถาม และการจัดหมวดหมู่คำถามใดคำถามหนึ่งมากเกินไป

การสุ่มตัวอย่างในการวิจัยนี้ใช้วิธีการสุ่มตัวอย่างแบบกลุ่ม (Cluster Random Sampling) โดยแบ่งกลุ่มการตั้งคำถาม และการจัดหมวดหมู่คำถาม อย่างเป็นสัดส่วน (Proportional Random Sampling) ตามสูตร (มานิตย์ สิงห์ทองชัย, 2557: 40)

$$n_h = \frac{N_h * n}{N}$$

เมื่อ n_h แทนจำนวนข้อคำถามในแต่ละหมวดหมู่

N_h แทนจำนวนข้อคำถามทั้งหมดในแต่ละหมวดหมู่

n แทนจำนวนตัวอย่างข้อคำถามที่ทำการศึกษา(n=400)

N แทนจำนวนข้อคำถาม

เพื่อให้ได้จำนวนกลุ่มตัวอย่างจากข้อคำถาม รวมเป็น 400 ข้อคำถาม จะได้ตัวอย่างที่ต้องการดังตารางที่ 3.3

ตารางที่ 3.3 แสดงจำนวนประชากรและกลุ่มตัวอย่างของข้อคำถาม จำแนกตามกลุ่มการตั้งคำถามและการจัดหมวดหมู่คำถาม

การวิเคราะห์ข้อคำถามผลไม้เศรษฐกิจที่เกี่ยวข้องกับทุเรียน มังคุด และลำไย	**จำนวนการเลือกข้อคำถามจากข้อคำถามทั้งหมด**						**รวม**
	Who	What	Why	When	Where	How	
1. พันธุ์	10***(4)***	49***(21)***	43***(18)***	34***(14)***	40***(17)***	22***(9)***	198***(83)***
2. การปลูกและดูแลรักษา	10***(4)***	54***(23)***	33***(14)***	34***(14)***	40***(17)***	5***(2)***	176***(74)***
3. โรคและแมลง	7***(3)***	23***(10)***	75***(32)***	54***(23)***	32***(13)***	2***(1)***	193***(81)***
4. สารเคมีที่ใช้ป้องกันโรคและแมลง	0***(0)***	25***(11)***	38***(16)***	21***(9)***	13***(5)***	6***(3)***	103***(43)***
5. การแปรรูป	0***(0)***	28***(12)***	20***(8)***	12***(5)***	15***(6)***	5***(2)***	80***(34)***

ตารางที่ 3.3 แสดงจำนวนประชากรและกลุ่มตัวอย่างของข้อคำถาม จำแนกตามกลุ่มการตั้งคำถามและการจัดหมวดหมู่คำถาม

การวิเคราะห์ข้อคำถามผลไม้เศรษฐกิจที่เกี่ยวข้องกับทุเรียน มังคุด และลำไย	**จำนวนการเลือกข้อคำถามจากข้อคำถามทั้งหมด**						**รวม**
	Who	**What**	**Why**	**When**	**Where**	**How**	
6. เครื่องจักรหรือเครื่องมือการผลิต	18***(8)***	30***(13)***	10***(4)***	18***(8)***	8***(3)***	8***(3)***	92***(39)***
7. การตลาดและการส่งออก	8***(3)***	34***(14)***	25***(11)***	21***(9)***	10***(4)***	10***(4)***	108***(45)***
รวม	53***(22)***	243***(102)***	244***(103)***	194***(82)***	158***(67)***	58***(24)***	950***(400)***

จากตารางที่ 3.3 พบว่าข้อคำถามที่ผู้วิจัยนำมาใช้สำหรับการถามตอบในระบบถามตอบเชิงความหมายสำหรับผลไม้เศรษฐกิจของไทยที่เกี่ยวข้องกับพันธุ์, การปลูกและดูแลรักษา, โรคและแมลง, สารเคมีที่ใช้ป้องกันโรคและแมลง, การแปรรูป, เครื่องจักรหรือเครื่องมือการผลิด, การตลาดและการส่งออก ของทุเรียน มังคุด และลำไย จำนวน 83, 74, 81, 43, 34, 39 และ 45 ข้อคำถามตามลำดับ

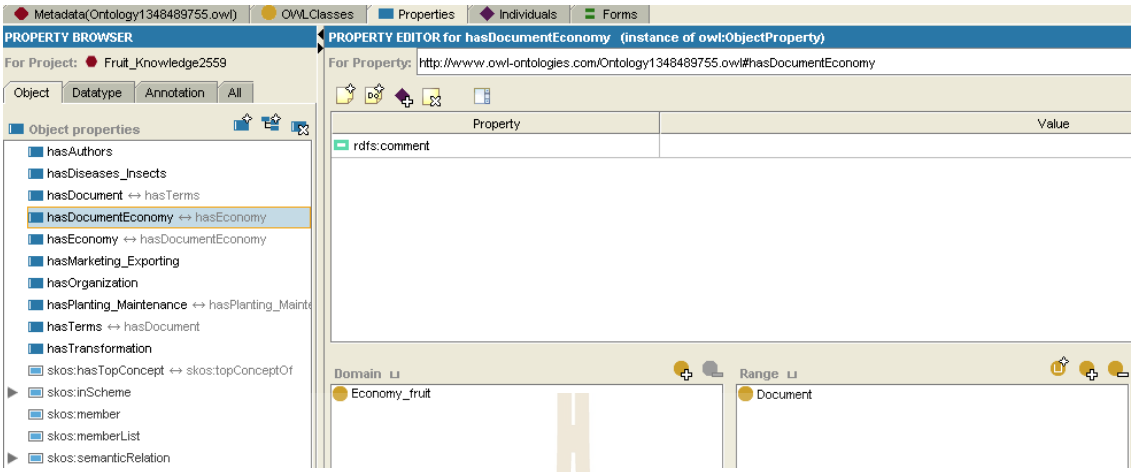
3.1.2 โครงสร้างของความรู้ทางด้านผลไม้เศรษฐกิจของไทย

การศึกษาความรู้ทางด้านผลไม้เศรษฐกิจของไทยสำหรับงานวิจัยนี้ ได้ใช้ออนโทโลยีด้านผลไม้เศรษฐกิจที่วิจัยและพัฒนาขึ้นโดย ศุภกฤษฎิ์ นิวัฒนากูล (2556) มาประยุกต์ใช้ในการออกแบบฐานความรู้ที่อยู่ในรูปแบบภาษาโอดับเบิ้ลยูแอล ซึ่งมีโครงสร้างของฐานความรู้ที่ประกอบไปด้วยคลาสผลไม้เศรษฐกิจ (Economy_Fruit) การปลูกและบำรุงรักษา (Planting_and_Maintenance) โรคและแมลง (Diseases_and_Insects) สารเคมี (Chemicals) การแปรรูป (Transformation) เครื่องจักรหรือเครื่องมือ (Machines_and_Equipments) การตลาดและการส่งออก (Marketing_and_Exporting) และเอกสารความรู้ผลไม้เศรษฐกิจ (Documents)

ซึ่งกระบวนปรับปรุงฐานความรู้ที่นำมาใช้นั้นผู้วิจัยได้ปรับปรุงออนโทโลยีผลไม้เศรษฐกิจของไทย ดังนี้

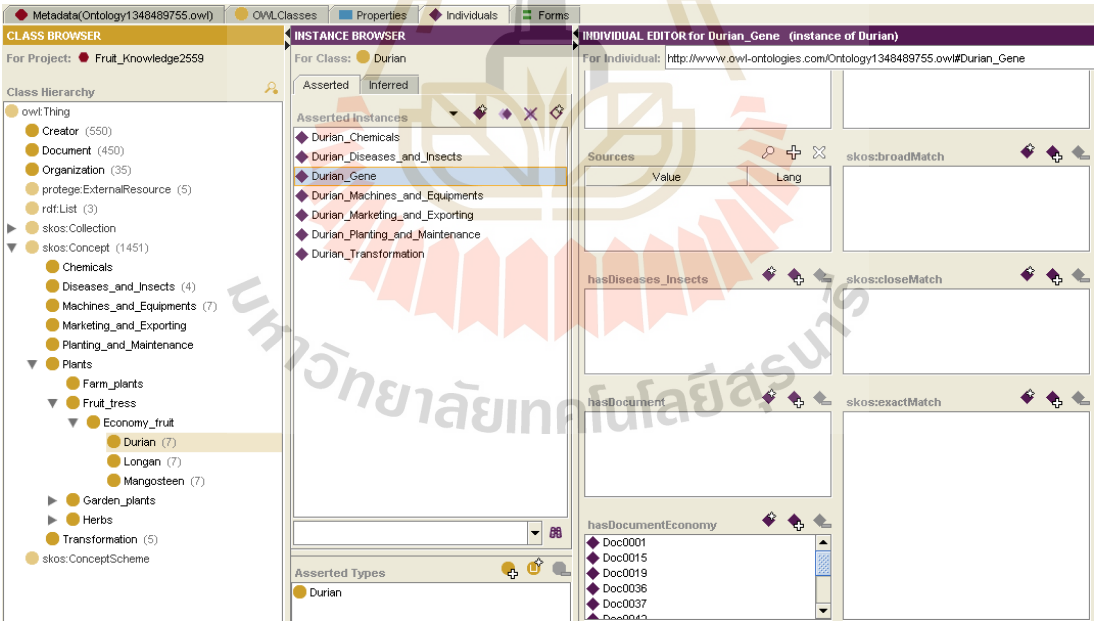
1) กำหนดออฟเจคพร็อพเพอร์ตี้ในฐานความรู้ผลไม้เศรษฐกิจเพิ่มเติม ได้แก่ hasDocumentEconomy เป็นพร็อพเพอร์ตี้ที่เชื่อมโยงความสัมพันธ์ระหว่างคลาส Economy_fruit

และ Document ซึ่ง Inverse กับออฟเจคพร็อพเพอร์ตี้ hasEconomy ซึ่งเป็นพร็อพเพอร์ตี้ที่เชื่อมโยงความสัมพันธ์ระหว่างคลาส Document และ Economy_fruit แสดงดังภาพที่ 3.2



ภาพที่ 3.2 การกำหนดออฟเจคพร็อพเพอร์ตี้เพิ่มเติมในออนโทโลยีผลไม้เศรษฐกิจ

1.2) จัดกลุ่มเอกสารในออกเป็น 7 กลุ่มตามการวิเคราะห์พฤติกรรมการตั้งคำถามของผู้ใช้ ได้แก่ พันธุ์ การปลูกและบำรุงรักษา โรคและแมลง สารเคมี การแปรรูป เครื่องจักรหรือเครื่องมือ การตลาดและการส่งออก สำหรับผลไม้เศรษฐกิจ 3 ชนิด คือ ทุเรียน มังคุด และลำไย แสดงดังภาพที่ 3.3

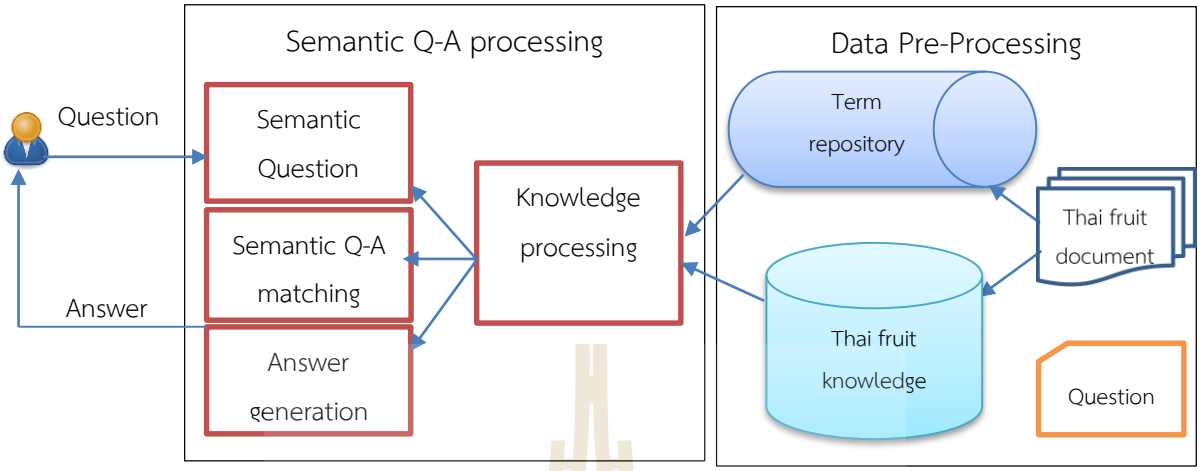


ภาพที่ 3.3 ตัวอย่างการจัดกลุ่มเอกสาร

3.2 การวิเคราะห์และออกแบบระบบ

เป็นการวิเคราะห์ระบบ การออกแบบโครงสร้างฐานความรู้ การออกแบบหน้าจอนำเข้าข้อมูลจากผู้ใช้ หน้าจอผลลัพธ์ของระบบ รวมถึงการออกแบบขั้นตอนวิธี (Algorithm) ของการทำงานในส่วน

ต่าง ๆ ของระบบถามตอบเชิงความหมายผ่านเว็บ ซึ่งมีกรอบการทำงานของระบบถามตอบเชิงความหมายงานวิจัยนี้ มีส่วนประกอบสำคัญ ดังภาพที่ 3.4



ภาพที่ 3.4 กรอบการทำงานของระบบถามตอบเชิงความหมายออนไลน์

จากภาพที่ 3.4 การทำงานของระบบถามตอบเชิงความหมายออนไลน์สำหรับผลไม้เศรษฐกิจของไทย ประกอบด้วย 2 ส่วน ดังนี้

1. การเตรียมข้อมูล (Data Pre-Processing) เป็นการเตรียมข้อมูลเบื้องต้นสำหรับการทำงานของระบบถามตอบเชิงความหมายออนไลน์ ดังนี้

1.1 เอกสารที่เป็นความรู้ผลไม้เศรษฐกิจของไทย (Thai fruit documents) ได้แก่ ทุเรียน มังคุด และลำไย ที่นำมาเป็นตัวอย่างจะค้นหาจากระบบเครือข่ายอินเตอร์เน็ตด้วยเสิร์ชเอ็นจิน แล้วเลือกรายการความรู้ที่มีการเผยแพร่ตั้งแต่ปี พ.ศ. 2550 เป็นต้นมา โดยเลือกรายการที่มีคำอธิบายครบถ้วนตามความต้องการ

1.2 คลังคำ (Term repository) ได้ใช้ตัวแบบ Simple knowledge organization system (SKOS) สำหรับสร้างคลังคำ โดยมีการนำคำศัพท์มาจากอรรถาภิธานศัพท์เกษตรไทย

1.3 ฐานความรู้ผลไม้เศรษฐกิจ (Thai fruit knowledge) ได้นำออนโทโลยีผลไม้เศรษฐกิจของ ศุภกฤษฎิ์ นิวัฒนากูล (2556) มาประยุกต์ใช้เป็นฐานความรู้ ซึ่งจัดเก็บเฉพาะชื่อเรื่อง คำสำคัญ คำอธิบายความรู้ และแหล่งอ้างอิงต้นฉบับไปยังเอกสารความรู้ผลไม้เศรษฐกิจเท่านั้น

1.4 ข้อคำถาม (Question) เป็นการเตรียมข้อสอบถาม จำนวน 400 คำถามที่เกี่ยวข้องกับผลไม้เศรษฐกิจของไทย ซึ่งตัวอย่างคำถามแสดงในภาคผนวก ก

2. กระบวนการถามตอบเชิงความหมาย (Semantic Q-A processing) ประกอบด้วย 4 ขั้นตอน ดังนี้

2.1 การประมวลผลความรู้ (Knowledge Processcing) เป็นกระบวนการสกัดคุณลักษณะเพื่อสร้างรายการคำ และนำคุณลักษณะที่ได้มาสร้างดัชนีเพื่อเลือกคุณลักษณะความน่าจะ

เป็นด้วยการเลือก 20% ของอันดับแรกของผลรวมค่าดัชนี และทำการจัดกลุ่มโดยใช้ซัพพอร์ทเวกเตอร์แมชชีน

2.2 การประมวลผลคำถามเชิงความหมาย (Semantic question processing) เป็นขั้นตอนการวิเคราะห์และประมวลผลข้อคำถาม โดยประยุกต์แนวคิดของ ธนารักษ์ ธีระมั่นคง และ ศศิพร อุษณวศิน (Theeramunkong and Usanavasin, 2001) ที่ใช้การสร้างตัวแทนจากการวิเคราะห์คำสำคัญในประโยคคำถามมาทำการตัดคำโดยใช้หลักการตัดคำจากคลังข้อมูล (Corpus based approach) โดยใช้กลไกการเรียนรู้ ซึ่งมีการตัดคำให้เป็นคำหรือพยางค์เดียว (Monosyllables) ตามกฎวิธีการผันคำในภาษาไทย และใช้คลังข้อมูลในการสร้างเครือข่ายคำ พร้อมสถิติการเกิดของคำเป็นตัวแปรเพื่อใช้ในการจำแนกหาขอบเขตของคำ (Word boundaries)

งานวิจัยนี้ใช้โปรแกรมตัดคำ THSplitlib (สุวิชา เผือกอิ่ม, 2556) แสดงดังภาพที่ 3.5 ซึ่งเป็นโปรแกรม Open Source สำหรับการตัดคำภาษาไทย ใช้หลักการตัดคำจากพจนานุกรม (Dictionary-Based Approach) ในการเปรียบเทียบการตัดคำกับคำที่จัดเก็บในพจนานุกรม ซึ่งผู้วิจัยได้นำข้อคำถามจำนวน 400 คำถาม จากขั้นตอนการเตรียมข้อมูลมาผ่านกระบวนการของโปรแกรมตัดคำเพื่อจำแนกคำออกจากประโยค เช่น

Q1 “การปลูกลำไยทำอย่างไรครับ”

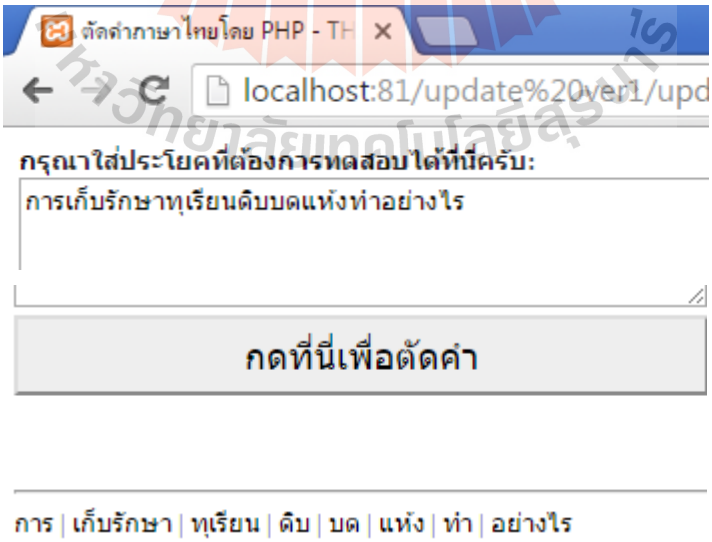
“การ | ปลูก | ลำไย | ทำ | อย่างไร | ครับ”

Q2 “อยากทราบการแปรรูปทุเรียนทำอย่างไรครับ”

“อยาก | ทราบ | การ | แปรรูป | ทุเรียน | ทำ | ยังไง | ครับ”

Q3 “ที่ไหนรับซื้อมังคุดบ้างครับ”

“ที่ | ไหน | รับ | ซื้อ | มังคุด | บ้าง | ครับ”



ภาพที่ 3.5 แสดงตัวอย่างการตัดคำของข้อคำถาม

จากนั้นผู้วิจัยจะทำการวิเคราะห์คำในประโยคที่ผ่านกระบวนการจำแนกคำมาแล้ว เช่นการตีความหมายของประโยคคำถามที่มีรูปประโยคไม่เหมือนกันแต่มีความหมายเหมือนกัน และคำกำกวม เป็นต้น ซึ่งขั้นตอนนี้จะดำเนินการด้วยผู้เชี่ยวชาญ (สิงห์ทัย สุขสว่างโรจน์ และระพีพรรณ พิริยะกุล, 2558) ดังนี้

1. จำแนกกลุ่มประโยค

“การ | ปลูก | ลำไย | ทำ | อย่างไร | ครับ”

“อยาก | ทราบ | การ | แปรรูป | ทุเรียน | ทำ | ยังไง | ครับ”

“ที่ | ไหน | รับ | ซื้อ | มังคุด | บ้าง | ครับ”

2. กำจัดคำที่ไม่มีสาระสำคัญในประโยคออกไป

“~~การ~~ | ปลูก | ลำไย | ทำ | อย่างไร |~~ครับ~~”

“~~อยาก~~ | ~~ทราบ~~ | การ | แปรรูป | ทุเรียน | ทำ | ยังไง | ~~ครับ~~”

“ที่ | ไหน | รับ | ซื้อ | มังคุด | ~~บ้าง~~ | ~~ครับ~~”

3. วิเคราะห์คำและจัดหมวดหมู่คำถาม ได้แก่ ใคร (Who) อะไร (What) อย่างไร (Why) เมื่อไร (When) ที่ไหน (Where) และอย่างไร (How) แสดงดังตารางที่ 3.4

ตารางที่ 3.4 วิเคราะห์คำและจัดหมวดหมู่คำถาม

ภาคประธาน	ภาคแสดง	คีย์เวิร์ด	หมวดหมู่คำถาม
ผู้ถาม	ปลูกลำไย	ทำอย่างไร	How
ผู้ถาม	การแปรรูปทุเรียน	ทำยังไง	How
ผู้ถาม	รับซื้อมังคุด	ที่ไหน	Where

4. สร้างรูปแบบคำถามบนพื้นฐานของการแจงประโยค

ทำอย่างไร | ปลูก | v | ลำไย | Noun

ทำยังไง | การแปรรูป | v | ทุเรียน | Noun

ที่ไหน | รับซื้อ | v | มังคุด | Noun

2.3 การจับคู่คำถามตอบ (Semantic Q-A matching) โดยกระบวนการนี้ทำงานหลังจากที่ได้รายการคำที่เกี่ยวข้องจากขั้นตอนการวิเคราะห์คำถามที่นำเข้าจากผู้ใช้นำไปคำนวณหา TF-IDF เพื่อนำไปค้นหาคำตอบจากจากฐานความรู้ โดยจะค้นหาความรู้ หรือเอกสารที่เกี่ยวข้องทั้งหมด และทำการเปรียบเทียบระหว่างข้อคำถาม กับคำตอบที่อยู่ในเอกสาร ซึ่งใช้แจ็คการ์ด สำหรับวัดความคล้ายคลึงระหว่างข้อคำถามและข้อความที่ถูกเลือกเป็นคำตอบ สอดคล้องกับการศึกษา

ของศุภกฤษฎ์ นิวัฒนากูล (2556) ที่ได้ใช้แจ๊คการ์ดหาค่าความเกี่ยวข้องระหว่างคำสองคำ ข้อสังเกตพบว่า การวัดประสิทธิภาพความคล้ายด้วยสัมประสิทธิ์แจ๊คการ์ด สามารถทำนายค่าได้ถูกต้องดี และมีเสถียรภาพในการทำงานสูงมาก

2.4 การสร้างชุดคำตอบ (Answer generation) โดยกระบวนการนี้จะนำรายการคำจากส่วนของ Semantic question processing และรายการความรู้ที่ได้จาก Semantic Q-A matching มาทำการเปรียบเทียบความคล้ายคลึงและสร้างเป็นคำตอบเพื่อเสนอต่อผู้ใช้ต่อไป

3.3 การพัฒนาระบบ

การพัฒนาระบบในงานวิจัยนี้ได้พัฒนาระบบการค้นคืนสารสนเทศ ที่ได้จากการสกัดสารสนเทศ โดยประมวลผลภาษาธรรมชาติเพื่อนำมาใช้สำหรับการค้นหาคำถามที่เกี่ยวข้องกับคำตอบ ซึ่งใช้เครื่องมือ ได้แก่ 1) Protégé สำหรับเป็นเครื่องมือในการสร้างออนโทโลยี 2) SPARQL Query Language สำหรับเป็นภาษาสำหรับการค้นหาความรู้ และ 3) RAP API for PHP สำหรับการสร้างส่วนต่อประสานผู้ใช้ (User Interface)

การนำซอฟต์แวร์มาใช้ในการพัฒนาระบบ จะเน้นการนำซอฟต์แวร์ที่เป็น Open Source หรือ Freeware เป็นหลักในการพัฒนาระบบ และใช้วิธีการเช่าเครื่องแม่ข่ายสำหรับเก็บข้อมูล

3.4 การประเมินผลระบบ

การประเมินระบบ จะดำเนินการเป็น 2 ส่วน ดังนี้

ส่วนที่ 1 เป็นการประเมินความถูกต้องและความสามารถของระบบถามตอบเชิงความหมายออนไลน์ โดยการกำหนดข้อมูลที่ต้องการค้นหา หลังจากนั้นจะทำการค้นหาโดยระบบและค้นหาโดยผู้เชี่ยวชาญ เพื่อนำข้อมูลมาเปรียบเทียบกัน ในส่วนนี้จะเป็นการวิเคราะห์โดยใช้การคำนวณค่าความระลึก (Recall Value), ค่าความแม่นยำ (Precision Value) และค่า F-Measure

ค่าความแม่นยำ (Precision) คือค่าที่แสดงให้เห็นว่าระบบที่พัฒนาขึ้นมีความแม่นยำเพียงใด ดังสมการที่ 3.2

$$Precision = \frac{TP}{TP+FP} \times 100\%$$

สมการที่ 3.2 ค่าความแม่นยำ

จากสมการ อธิบายได้ดังนี้

TP (True Position) แทน เอกสารที่เลือก และถูกเลือกโดยผู้เชี่ยวชาญ

FP (False Position) แทน เอกสารที่เลือก แต่ไม่ถูกเลือกโดยผู้เชี่ยวชาญ

ค่าความระลึก (Recall) คือค่าที่แสดงให้เห็นว่าระบบได้ทำการดึงคำตอบออกมาแล้วมีความเกี่ยวข้องกับคำถามที่ผู้เชี่ยวชาญค้นหามากเพียงใด ดังสมการที่ 3.3

มหาวิทยาลัยเทคโนโลยีสุรนารี

$$Recall = \frac{TP}{TP+FN} \times 100\%$$

สมการที่ 3.3 ค่าความระลึก

จากสมการ อธิบายได้ดังนี้

TP (True Position) แทน เอกสารที่เลือก และถูกเลือกโดยผู้เชี่ยวชาญ

FN (False Negative) แทน เอกสารที่ไม่เลือก แต่ถูกเลือกโดยผู้เชี่ยวชาญ

อัตราการรู้จำ (F-measure) คือค่าเฉลี่ยที่ให้ความสำคัญกับความแม่นยำและค่าความระลึก คำนวณได้จากสมการที่ 3.4

$$F\text{-}Measure = \frac{2 \times Recall \times Precision|}{Precision + Recall}$$

สมการที่ 3.4 อัตราการรู้จำ

ส่วนที่ 2 เป็นการทดสอบโดยใช้วิธีประเมินความพึงพอใจของผู้ใช้ระบบถามตอบที่ผู้วิจัยได้พัฒนาขึ้นครั้งนี้เป็นแบบสอบถามเกี่ยวกับความยากง่ายในการใช้งานระบบและความพึงพอใจต่อผลลัพธ์ที่ได้ ซึ่งสอบถามจากผู้ใช้จำนวนไม่น้อยกว่า 30 คน วิเคราะห์ข้อมูลใช้สถิติเบื้องต้น เช่น ร้อยละ ค่าเฉลี่ย และค่าเบี่ยงเบนมาตรฐาน

3.5 การสร้างและหาประสิทธิภาพของเครื่องมือ

ในการวิจัยครั้งนี้ใช้การสร้างและหาประสิทธิภาพของเครื่องมือ ซึ่งได้แก่แบบสอบถามความพึงพอใจ ดังนี้

1) สร้างแบบสอบถามเครื่องมือประเมินผลระบบ เป็นความคิดเห็นเกี่ยวกับความพึงพอใจต่อการใช้งานระบบถามตอบเชิงความหมายสำหรับผลไม้เศรษฐกิจของไทย ลักษณะแบบสอบถามเป็นแบบมาตราส่วนประมาณค่า (Rating Scale) จำนวน 5 ข้อ โดยมีเกณฑ์การให้คะแนนตามวิธีการของลิเคอร์ท (Likert) มี 5 ระดับ ดังนี้

พึงพอใจมากที่สุด	ให้	5	คะแนน
พึงพอใจมาก	ให้	4	คะแนน
พึงพอใจปานกลาง	ให้	3	คะแนน
พึงพอใจน้อย	ให้	2	คะแนน
พึงพอใจน้อยที่สุด	ให้	1	คะแนน

ผู้วิจัยได้ทำการวิเคราะห์ข้อมูลในการศึกษาโดยนำแบบประเมินความพึงพอใจของผู้ใช้ที่มีต่อระบบถามตอบเชิงความหมายออนไลน์ผลไม้เศรษฐกิจของประเทศไทยมาวิเคราะห์ระดับความเหมาะสมโดยใช้สถิติค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน และเกณฑ์การให้ระดับคะแนนประเมินแบบสอบถามความพึงพอใจ ดังนี้ (มานิตย์ สิงห์ทองชัย. 2558: 43)

ระดับค่าเฉลี่ย = (คะแนนสูงสุด – คะแนนต่ำสุด) / จำนวนชั้น

= (5-1) /5 =0.8

ค่าเฉลี่ย 4.24 - 5.00 แปลความว่า พึงพอใจมากที่สุด

ค่าเฉลี่ย 3.43 – 4.23 แปลความว่า พึงพอใจมาก

ค่าเฉลี่ย 2.62 – 3.42 แปลความว่า พึงพอใจปานกลาง

ค่าเฉลี่ย 1.81 – 2.61 แปลความว่า พึงพอใจน้อย

ค่าเฉลี่ย 1.00 – 1.80 แปลความว่า พึงพอใจน้อยที่สุด

2) นำแบบประเมินที่สร้างขึ้นเสนอต่อผู้เชี่ยวชาญ จำนวน 5 ท่านสำหรับตรวจสอบระหว่างข้อคำถามกับจุดประสงค์ของการประเมิน โดยประเมินค่าดัชนีความสอดคล้อง (Item-Object Congruence Index: IOC) สำหรับการประเมินค่าดัชนีความสอดคล้องข้อคำถามกับจุดประสงค์การประเมิน (IOC) มีเกณฑ์การให้คะแนน ดังนี้

คะแนน +1 เมื่อแน่ใจว่าข้อคำถามนั้นประเมินตามจุดประสงค์

คะแนน 0 เมื่อไม่แน่ใจว่าข้อคำถามนั้นประเมินตามจุดประสงค์

คะแนน -1 เมื่อแน่ใจว่าคำถามนั้นไม่สอดคล้องกันระหว่างคำถามกับจุดประสงค์

จากนั้นจะทำการวิเคราะห์ข้อมูลการหาค่าดัชนี ความสอดคล้องระหว่างข้อคำถามของแบบประเมินกับจุดประสงค์ในการประเมิน ซึ่งในการคำนวณค่า IOC ใช้สูตรดังสมการ 3.1 ดังนี้ (กรมวิชาการ. 2545 : 65)

$$IOC = \frac{\sum R}{N}$$

สมการที่ 3.1 สูตรคำนวณค่า IOC

IOC คือ ดัชนีความสอดคล้องระหว่างคำถามกับจุดประสงค์

R คือ คะแนนของผู้เชี่ยวชาญ

∑R คือ ผลรวมของคะแนนผู้เชี่ยวชาญแต่ละคน

N คือ จำนวนผู้เชี่ยวชาญ

เกณฑ์การเลือกค่า IOC โดยข้อคำถามแต่ละข้อต้องได้ค่าตั้งแต่ 0.60 ขึ้นไป ผลการประเมินความถูกต้องของระบบพบว่า IOC มีค่าเฉลี่ยเท่ากับ 0.92 และข้อคำถามทุกข้อมีผลการประเมินสูงกว่า 0.60 ทุกข้อคำถาม รายละเอียดแสดงดังตารางที่ 3.5

ตารางที่ 3.5 ผลการประเมินความสอดคล้องของข้อคำถามโดยผู้เชี่ยวชาญ

รายการ	ค่า IOC
1. รูปแบบการระบุข้อคำถามเหมาะสม สะดวกต่อการใช้งาน	0.80
2. การค้นคืนความรู้ตรงกับความต้องการ	1.00
3. ผลลัพธ์ที่ได้จากการค้นคืนตรงกับความต้องการ	1.00
4. สามารถเข้าถึงแหล่งความรู้ผลไม้เศรษฐกิจของไทย	0.80
5. การเชื่อมโยงไปยังแหล่งความรู้ที่พบ	1.00
รวม	0.92

3) นำแบบสอบถามที่ผ่านการประเมินค่า IOC แล้วไปให้ผู้ใช้จำนวน 30 ท่าน โดยผู้ใช้นี้จะทำการเลือกกลุ่มตัวอย่างแบบเฉพาะเจาะจง โดยต้องเป็นผู้ที่มีความสามารถในการใช้งานคอมพิวเตอร์ และมีความรู้เกี่ยวกับทางด้านการเกษตร โดยให้ทำการทดลองสืบค้นข้อมูลจากระบบถามตอบเชิงความหมายออนไลน์สำหรับผลไม้เศรษฐกิจของไทย หลังจากนั้นจะประเมินความสามารถในการทำงานของระบบสืบค้นตามข้อคำถามที่ได้ออกแบบไว้ และนำข้อมูลที่ได้มาวิเคราะห์และสรุปผลต่อไป

บทที่ 4 ผลการวิจัยและอภิปรายผล

การดำเนินการวิจัยครั้งนี้ ผู้วิจัยเสนอผลการวิจัยการพัฒนาระบบถามตอบเชิงความหมายออนไลน์สำหรับผลไม้เศรษฐกิจของไทย ดังนี้

4.1 ผลการพัฒนากระบวนการถามตอบเชิงความหมายออนไลน์

4.1.1 การประมวลผลความรู้

4.1.2 การประมวลผลคำถามเชิงความหมาย

4.1.3 การจับคู่คำถาม-คำตอบ

4.1.4 การสร้างชุดคำตอบ

4.2 ผลการประเมินประสิทธิภาพ

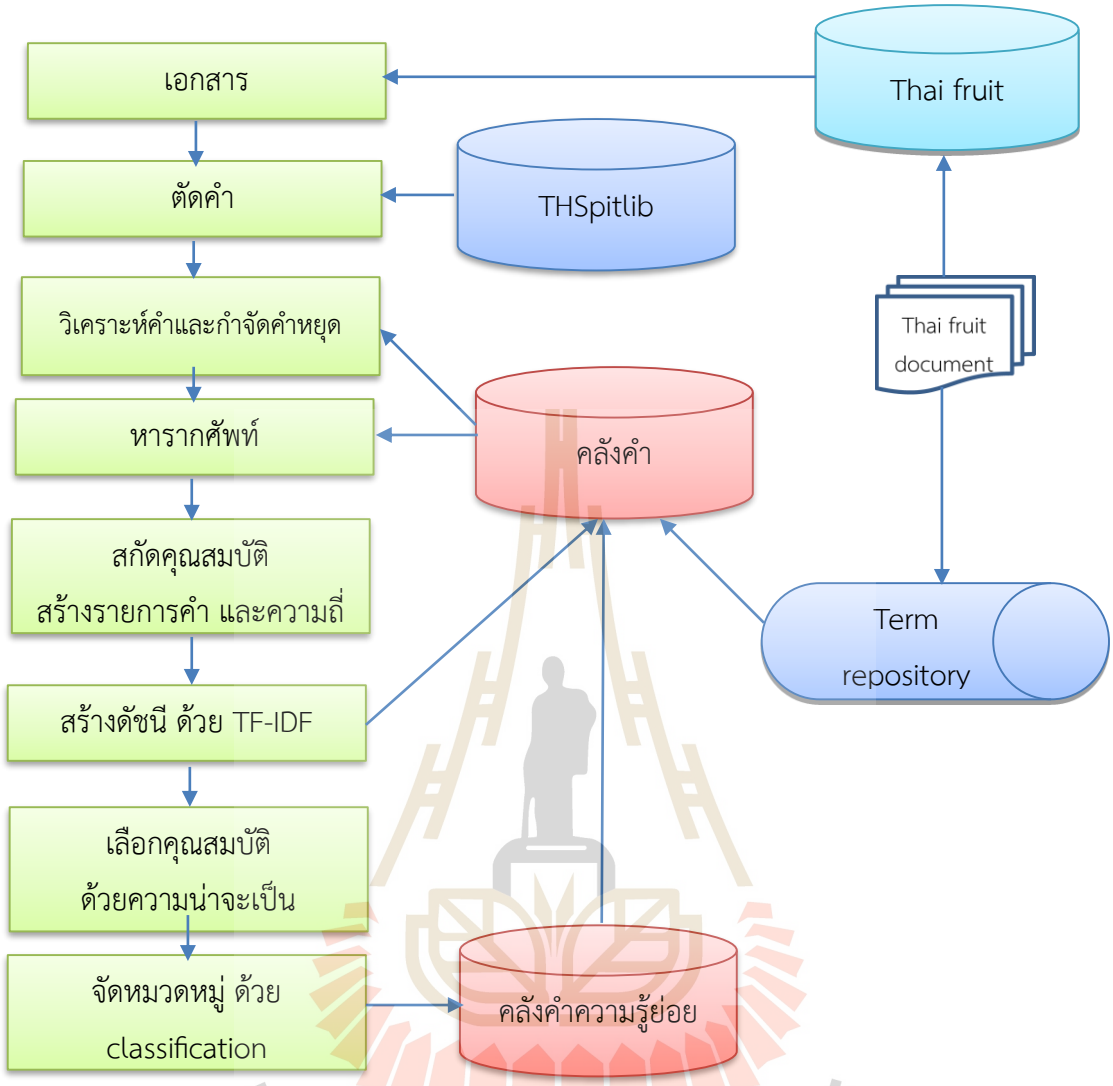
4.2.1 ผลการประเมินความถูกต้องและความสามารถของระบบ

4.2.2 ผลการวิเคราะห์ความพึงพอใจ

4.1 ผลการพัฒนากระบวนการถามตอบเชิงความหมายออนไลน์

4.1.1 การประมวลผลความรู้

การประมวลผลความรู้จะเริ่มจากกระบวนการตัดคำจากเอกสารในฐานความรู้ที่จัดเก็บในฐานความรู้ผลไม้เศรษฐกิจ ระบบจะทำการตัดคำด้วยโปรแกรมตัดคำ THSplitlib (สุวิชา เผือกอิ่ม, 2556) ซึ่งเป็นโปรแกรม Open Source สำหรับการตัดคำภาษาไทย ใช้หลักการตัดคำจากพจนานุกรม (Dictionary-Based Approach) ในการเปรียบเทียบการตัดคำกับคำที่จัดเก็บในพจนานุกรม ซึ่งมีการกำจัดคำหยุดมักเป็นคำที่ปรากฏขึ้นบ่อยครั้งในความรู้และไม่มีความหมายต่อความรู้ เช่นคำบุพบทที่ใช้เชื่อมคำหรือกลุ่มคำให้สัมพันธ์กัน คำสันธานทำหน้าที่เชื่อมคำกับคำ เป็นต้น ซึ่งการกำจัดคำหยุดจะช่วยให้ลดขนาดของดัชนีลงเพื่อช่วยในการประมวลผลให้รวดเร็วขึ้น (Jaruskulchai, 1998) กระบวนการประมวลผลความรู้แสดงรายละเอียดดังภาพที่ 4.1



ภาพที่ 4.1 กระบวนการประมวลผลความรู้

จากนั้นจะทำการหารากศัพท์จากรูปเดิมของคำ หรือหาคำที่มีความหมายคล้ายกันเพื่อปรับรวมให้เป็นคำเดียวซึ่งวิธีการนี้จะใช้ผู้เชี่ยวชาญรวบรวมคำศัพท์ที่มีความหมายคล้ายคลึงกัน จัดเก็บในคลังคำ จากนั้นทำการสกัดคุณลักษณะสร้างรายการคำโดยใช้คำวามถี่ของคำเดี่ยว นำคุณลักษณะที่ได้มาสร้างดัชนี ด้วย TF-IDF แล้วเลือกคุณลักษณะด้วยความน่าจะเป็นโดยเลือก 20% อันดับแรกของผลรวมค่าดัชนีของคุณลักษณะเพื่อลดการคำนวณโดยอาศัยหลักการ Pareto principle (80-20 rule) และทำการจัดกลุ่มด้วย classification โดยใช้ซัพพอร์ทเวกเตอร์แมชชีน ซึ่งเหมาะกับข้อมูลที่มีมิติของข้อมูลสูง (นิเวศ จิระวิชิตชัย,ปริญญา สงวนสัตย์ และพยุง มีสัจ, 2553)

ตัวอย่างตารางคุณลักษณะ

	w_1	w_2	...	w_x
d_1	f_{11}	f_{12}	...	f_{1x}
d_2	f_{21}	f_{22}	...	f_{2x}
...	...	...	...	...
d_y	f_{y1}	f_{y2}	...	f_{yx}

สร้างดัชนีจากคุณลักษณะด้วย TF-IDF

$$TFIDF(t,d,D)=TF(t,d)\times IDF(t,D)$$

$$TF(t,d)=0.5+\frac{0.5\times f_{t,d}}{\max\{f_{t',d} : t' \in d\}}$$

$$IDF(t,D)=\log\frac{N}{|\{d \in : t \in d\}|}$$

เมื่อ

$TFIDF(t,d,D)$ = ค่า TF-IDF

$TF(t,d)$ = ค่าความถี่คำ

$IDF(t,D)$ = ค่าส่วนกลับความถี่เอกสาร

$f_{t,d}$ = จำนวนคำ t ที่ปรากฏบนเอกสาร d

$\max\{f_{t',d} : t' \in d\}$ = จำนวนคำ t' ใด ๆ ที่นับได้สูงสุดบนเอกสาร D

N = จำนวนเอกสารทั้งหมด

$|\{d \in D : t \in d\}|$ = จำนวนเอกสารที่มีคำ t ปรากฏอยู่

ตัวอย่างตารางดัชนีคำ

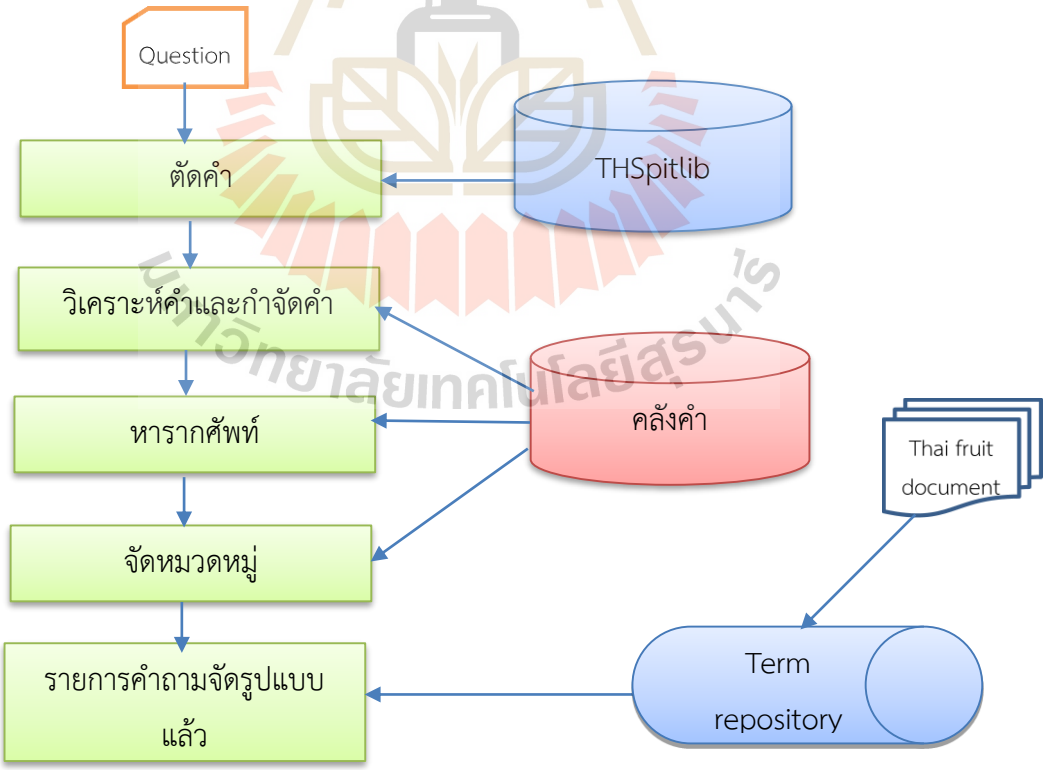
	w_1	w_2	...	w_x
d_1	$tfidf_{11}$	$tfidf_{12}$	...	$tfidf_{1x}$
d_2	$tfidf_{21}$	$tfidf_{22}$	...	$tfidf_{2x}$
...	...	...	...	...
d_y	$tfidf_{y1}$	$tfidf_{y2}$	...	$tfidf_{yx}$

มหาวิทยาลัยเทคโนโลยีสุรนารี

4.1.2 การประมวลผลคำถามเชิงความหมาย

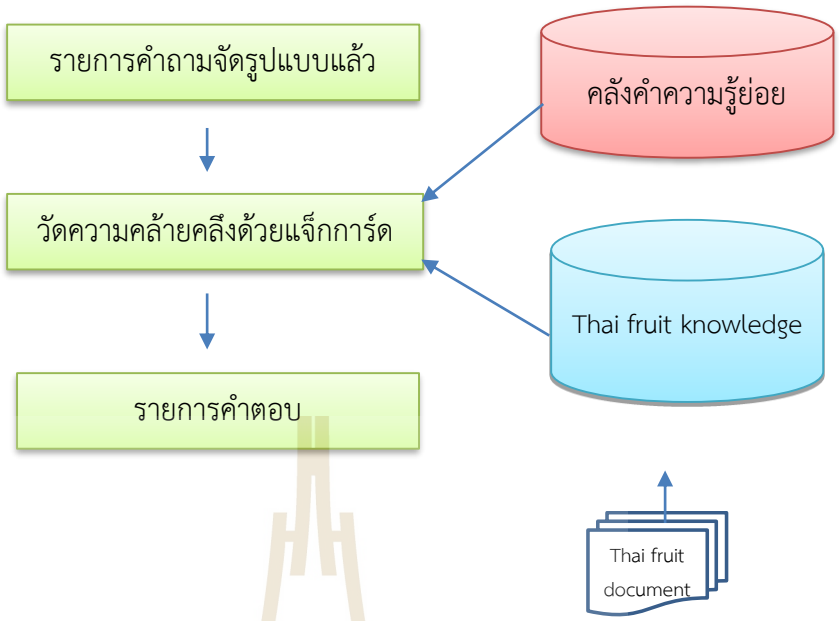
เมื่อผู้ใช้กรอกคำถามเข้ามา ระบบจะทำการตัดคำด้วยโปรแกรมตัดคำ THSplitlib (สุวิชา เผือกอิ่ม, 2556) ใช้หลักการเปรียบเทียบการตัดคำกับคำที่จัดเก็บในพจนานุกรม เช่น กรอกคำถามว่า “การเก็บรักษาทุเรียนดิบบดแห้งทำอย่างไร” จะได้ผลลัพธ์ว่า “การ/เก็บรักษา/ทุเรียน/ดิบ/บด/แห้ง/ทำ/อย่างไร” ส่วนวิเคราะห์และกำจัดคำต้องทำการกำจัดคำหยุดในที่นี้รวมถึงวิเศษณ์ ซึ่งมีหน้าที่ขยายคำนาม สรรมนาม กริยา และวิเศษณ์ เช่น จะ ที่ ความ การ แดง ดำ เปียก แห้ง เป็นต้น เนื่องจากคำเหล่านี้ไม่ใช่ใจความสำคัญของการสืบค้น จะสร้างความสับสนในการหาคำตอบ

ดังนั้นจาก รายการคำทั้งหมดจะเหลือเพียง “เก็บรักษา/ทุเรียน/บด/อย่างไร” จากนั้นจะเป็นการหารากศัพท์ และการจัดหมวดหมู่ซึ่งมีด้วยกันสองส่วน ส่วนแรกคือการจัดหมวดหมู่เพื่อให้รู้ว่าคำถามนี้อยู่ในหมวดหมู่ความรู้ใด เช่น “เก็บรักษา”-การปลูกและบำรุงรักษา (Planting_and_Maintenance) และ “บด”-การแปรรูป (Transformation) เป็นต้น ส่วนที่สองคือการจัดหมวดหมู่เพื่อให้รู้ว่าคำถามนี้ถามเรื่องอะไร เช่น “อย่างไร”-ขั้นตอนหรือวิธีการ เป็นต้น สุดท้ายการสร้างรูปแบบคำถามในรูปแบบที่สามารถสืบค้นได้ด้วยคอมพิวเตอร์คือการเปรียบคำถามให้อยู่ในรูปแบบรายการคำค้นที่มีความครอบคลุมคำถามที่ผู้ระบุเข้ามานั้นเอง เช่น “การเก็บรักษาทุเรียนดิบบดแห้งทำอย่างไร”->“ขั้นตอน/วิธีการ/เก็บรักษา/ทุเรียน/บด”แสดงดังภาพที่ 4.2



ภาพที่ 4.2 กระบวนการประมวลผลคำถามเชิงความหมาย

4.1.3 การจับคู่คำถาม-คำตอบ



ภาพที่ 4.3 กระบวนการจับคู่คำถาม-คำตอบ

ภาพที่ 4.3 เป็นการแสดงกระบวนการจับคู่คำถาม-คำตอบ โดยเมื่อได้รูปแบบคำจากขั้นตอนที่ผ่านมาแล้ว จะมีการนำรูปแบบคำถามไปทำการเปรียบเทียบเอกสารความรู้ที่เก็บอยู่ในฐานความรู้ จากนั้นทำการเปรียบเทียบด้วยการวัดความคล้ายคลึงกันระหว่างข้อคำถาม กับคำตอบเฉพาะกับเอกสารที่อยู่ในหมวดหมู่ที่เกี่ยวข้องกับคำถามเท่านั้น เช่น “เก็บรักษา”-การปลูกและบำรุงรักษา (Planting_and_Maintenance) และ “บด”-การแปรรูป (Transformation) เป็นต้น

การวัดความคล้ายคลึงระหว่างคำถามที่จัดรูปแบบแล้วกับคำตอบด้วยแจ็กการ์ด ดังสมการ

$$\square(\square,\square) = \frac{|\square \cap \square|}{|\square \cup \square|} = \frac{\square \cdot \square}{|\square|^{\square} + |\square|^{\square} - \square \cdot \square}$$

เมื่อ

$J(A,B)$ = ความคล้ายคลึง

A = เวกเตอร์คำค้น

B = เวกเตอร์เอกสาร

4.1.4 การสร้างชุดคำตอบ

เมื่อได้ข้อความที่จะมาเป็นคำตอบแล้ว ซึ่งข้อคำถามนั้นอาจมีความยาวเกินไป หรือจัดรูปแบบข้อความแล้วไม่ได้ใจความ ระบบต้องปรับปรุงข้อคำตอบให้สามารถเข้าใจได้โดยง่ายโดยมีกระบวนการแสดงดังภาพที่ 4.4 เริ่มต้นจากการนำข้อความจากเอกสารที่ได้รับการวิเคราะห์ในส่วนการจับคู่คำถามตอบ เปรียบเทียบไวยากรณ์โดยการอาศัยคลังคำออนไลน์ เนื่องจากคำถามหลักไวยากรณ์นั้นมีความละเอียดอ่อนต้องอาศัยผู้เชี่ยวชาญและแหล่งข้อมูลอ้างอิงเพื่อสร้างฐานความรู้ขึ้นมา เช่น “กวน”เป็นคำนาม แปลตามภาษาถิ่นอีสาน หมายถึงตำแหน่งขุนนางในภาคอีสานสมัยโบราณ “กวน”เป็นคำนามแปลว่าผลไม้ที่เคี่ยวกับน้ำตาลแล้วคนให้เข้ากัน และ”กวน”เป็นคำกริยา แปลว่าคนให้เข้ากัน, ทำให้รำคาญ หรือชวนทะเลาะ เป็นต้น ด้วยวิธีนี้ระบบจึงทราบได้ว่าคำไหนมีไวยากรณ์ทางภาษาประเภทใดและทำการเลือกข้อความสั้น คือการเลือกข้อความสำคัญที่ว่าสามารถถอดไวยากรณ์ของข้อความนั้นมาได้ถูกต้องแม่นยำแค่ไหน เนื่องจากระบบมีการใส่ความรู้จำทางไวยากรณ์ที่ง่าย ๆ ไว้ คือ ประโยคต้องประกอบด้วย ประธาน+กริยา+(กรรม) ดังนั้นวิธีการเลือกจึงมองหาคำหลักและองค์ประกอบที่สมบูรณ์ คือ นาม+กริยา+(นาม) เช่น “ทุเรียนบดสามารถเก็บไว้ได้นาน 1-2 ปี ในภาชนะบรรจุปิดสนิทความชื้นต่ำ” สามารถถอดไวยากรณ์ได้ดังนี้

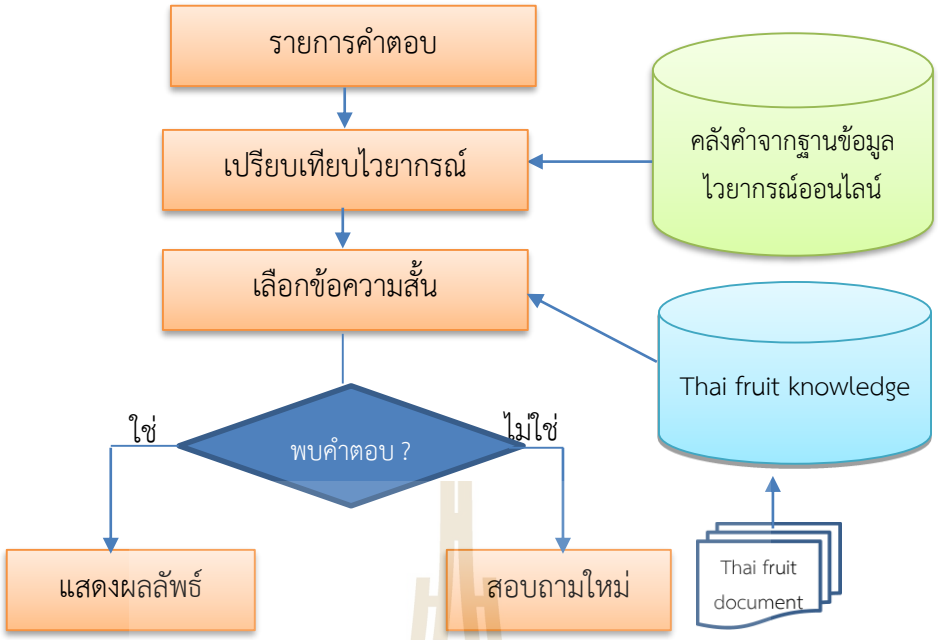
ทุเรียน-น.	บด-ก./น	สามารถ-ว.	เก็บ-ก./น.
ไว้-ก.	ได้-ก.	นาน-ว.	1-2-มย.
ปี-น.	ใน-บ.	ภาชนะ-น.	บรรจุ-ก.
ปิด-ก.	สนิท-ว.	ความชื้น-น.	ต่ำ-ว./ก.

เมื่อ น. คือ นาม ก. คือ กริยา ว. คือ วิเศษ บ. คือ บุพบท มย. คือ ไม่ได้นิยาม

หากเรียงกันเฉพาะไวยากรณ์จะได้ว่า “น. ก./น. ว. ก./น. ก. ก. ว. มย. น. บ. น. ก. ก. ว. น. ว./ก.” สังเกตว่าคำที่เน้นตัวหนานั้นคือ คำที่ผู้ใช้ใส่เข้ามา จึงทำให้ประโยคที่จะถูกนำมาแสดงมีแค่ประโยคเดียว คือ “น. ก./น. ว. ก./น. ก. ก. ว. มย. น.” ซึ่งตรงกับข้อความ “ทุเรียนบดสามารถเก็บไว้ได้นาน 1-2 ปี” มาถึงตรงนี้ผู้วิจัยพบว่าแม้ประโยคที่ได้จะไม่ได้ตอบคำถามอย่างตรงไปตรงมา แต่ก็เพียงพอที่จะบอกผู้ใช้ได้ว่าเอกสารนั้นอาจจะมีข้อความที่เกี่ยวข้องกับคำถามของผู้ใช้มากน้อยเพียงใด ซึ่งการสร้างชุดคำตอบนี้ ได้แบ่ง 2 กรณี ดังนี้

1. กรณีวิเคราะห์ข้อคำถามและสร้างชุดคำตอบได้ ระบบถามตอบจะทำการสกัดความรู้ในฐานความรู้ที่ตรงกับหมวดหมู่มาแสดง โดยแสดงชุดคำตอบ พร้อมลิงค์ชื่อของเอกสาร และผู้เขียน

2. กรณีวิเคราะห์ข้อคำถามที่ไม่สามารถสร้างชุดคำตอบได้ หรือไม่พบคำตอบในฐานความรู้ ระบบถามตอบจะส่งข้อความให้ผู้ค้นหาทราบและให้ระบุคำถามใหม่อีกครั้งเพื่อจะใช้ในการค้นหาข้อมูลที่เกี่ยวข้องกับคำค้นที่ผู้ใช้ต้องการต่อไป



ภาพที่ 4.4 กระบวนการสร้างชุดคำตอบ

4.2 ผลการประเมินระบบ

4.2.1 ผลการประเมินความถูกต้องและความสามารถของระบบ

การประเมินความถูกต้องและความสามารถของระบบในงานวิจัยนี้ได้ทดลองให้ผู้ป้อนข้อคำถาม โดยมีข้อคำถามที่ผู้ใช้ใช้ในการสอบถามรวมทั้งสิ้นจำนวน 400 คำถามที่เกี่ยวข้องกับผลไม้เศรษฐกิจของไทย ได้แก่ ทุเรียน มังคุด และลำไย โดยทำการประมวลผลตามกรอบการทำงานของระบบถามตอบเชิงความหมายออนไลน์ จากนั้นนำผลลัพธ์ที่ได้นำมาตรวจสอบความถูกต้องและประเมิน โดยเปรียบเทียบกับการกำหนดความสำคัญของเอกสารต่อคำค้นของผู้เชี่ยวชาญ ซึ่งคำตอบที่ได้จากการประมวลผลผู้ใช้สามารถมองเห็นได้ชัดเจนว่าเป็นคำตอบที่ถูกต้องหรือไม่ถูกต้อง ซึ่งบางกรณีที่คำตอบไม่ชัดเจน ให้ถือว่าคำตอบเหล่านั้นเป็นคำตอบที่ไม่ถูกต้อง สรุปผลการประเมินความถูกต้องและความสามารถของระบบการถามตอบเชิงความหมายออนไลน์สำหรับผลไม้เศรษฐกิจของไทยแสดงดังตารางที่ 4.1 รายละเอียดการคำนวณแสดงภาคผนวก ข

ตารางที่ 4.1 ผลการประเมินความถูกต้องและความสามารถของระบบ

ข้อคำถาม (Query)	การวัดความถูกต้องและความสามารถของระบบ		
	Precision	Recall	F-measure
400	0.87	0.85	0.86

ที่มา : การคำนวณภาคผนวก ข

จากตารางที่ 4.1 พบว่าระบบสามารถให้คำตอบได้ตรงตามความต้องการครบถ้วน โดยมีค่าความแม่นยำ (Precision) และค่าความระลึก (Recall) เท่ากับ 0.87, 0.85 ตามลำดับ และเมื่อพิจารณาค่าอัตราการรู้จำ (F-measure) เป็นการหาค่าเฉลี่ยที่ให้ความสำคัญกับความแม่นยำและความระลึกครบถ้วนเท่า ๆ กันพบว่าระบบสามารถให้คำตอบได้อย่างถูกต้อง เท่ากับ 0.86 นั้นเป็นเพราะระบบถามตอบที่พัฒนาขึ้นสามารถค้นหาหลัก คำพ้อง คำกว้างกว่า (Broader Term-BT) คำแคบกว่า (Narrower Term-NT) และ คำเกี่ยวข้อง (Related Term-RT) ได้

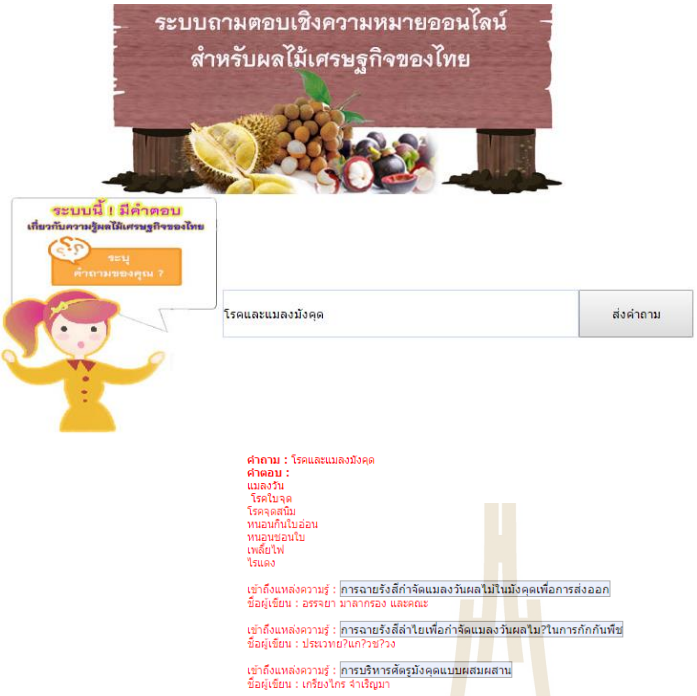
4.2.2 ผลการวิเคราะห์ความพึงพอใจ

การทดสอบระบบถามตอบเชิงความหมายผ่านส่วนติดต่อกับผู้ใช้งานบนเว็บบราวเซอร์จากฐานความรู้ที่สร้างขึ้นผ่านเครือข่ายอินเทอร์เน็ต โดยมีผลการประเมินความพึงพอใจจากผู้ใช้ระบบ ดังนี้

1) ตัวอย่างผลการถามตอบจากกลุ่มตัวอย่างของผู้ใช้งานระบบ

ภาพที่ 4.5 ตัวอย่างผลลัพธ์ของคำถาม “วิธีขยายพันธุ์ลำไย”

มหาวิทยาลัยเทคโนโลยีสุรนารี



ภาพที่ 4.6 ตัวอย่างผลลัพธ์ของคำถาม “โรคและแมลงมังคุด”

จากภาพที่ 4.5 และภาพที่ 4.6 แสดงตัวอย่างผลลัพธ์จากการระบุคำถาม “วิธีขยายพันธุ์ลำไย” และโรคและแมลงมังคุด จากกลุ่มตัวอย่างผู้ใช้งาน พบว่าระบบสามารถค้นหาคำตอบจากฐานความรู้ได้โดยอัตโนมัติ และแสดงรายละเอียดเอกสารที่มีความเกี่ยวข้องกับคำถาม และแหล่งเข้าถึงความรู้ได้อย่างครบถ้วนแม่นยำ

ข้อสังเกตพบว่า สำหรับการถามตอบเชิงความหมายออนไลน์สำหรับผลไม้เศรษฐกิจของไทยระบบสามารถเข้าใจการประมวลผลข้อคำถามของผู้ใช้ที่ระบุเป็นประโยคเข้ามาสอบถาม โดยใช้เทคนิคการประมวลผลภาษาธรรมชาติ สำหรับการวิเคราะห์กลุ่มคำ ชนิดกลุ่มคำ และความหมายของประโยคตามกระบวนการถามตอบเชิงความหมายออนไลน์ที่ผู้วิจัยได้นำเสนอ ซึ่งผลลัพธ์ที่ได้จากการถามตอบระบบมีข้อผิดพลาดในส่วนของยังมีบางคำตอบที่มีความถูกต้องแต่ไม่ได้ถูกดึงขึ้นมาแสดงผล เช่นจากการระบุข้อคำถามโรคและแมลงมังคุด คำตอบที่ถูกต้องอีกหนึ่งคำตอบ ได้แก่ โรคใบแห้งและขอบใบแห้ง และบางครั้งการเลือกเอกสารที่เกี่ยวข้องจะให้คำตอบการเลือกเอกสารที่ไม่เกี่ยวข้องติดมาด้วย ในอีกประเด็นคือการใช้คำถามที่เป็นคำค้นเดียวจะให้ผลคำตอบที่ถูกต้องครบถ้วน และส่งผลให้ค่าคะแนนเอกสารที่เกี่ยวข้องสูงไปด้วย

2) ผลการวิเคราะห์ความพึงพอใจของผู้ใช้ที่มีต่อระบบถามตอบเชิงความหมายสำหรับผลไม้เศรษฐกิจของไทย แสดงดังตารางที่ 4.2 สำหรับรายละเอียดแสดงดังภาคผนวก ค

ตารางที่ 4.2 ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของความพึงพอใจของผู้ใช้ที่มีต่อระบบถามตอบ

ข้อคำถาม	$\bar{x}$	S.D.	ระดับความพึงพอใจ
1. รูปแบบการระบุข้อคำถามเหมาะสม สะดวกต่อการใช้งาน	4.73	0.45	มากที่สุด
2. การค้นคืนความรู้ตรงกับความต้องการ	4.17	0.46	มาก
3. ผลลัพธ์ที่ได้จากการค้นคืนตรงกับความต้องการ	4.03	0.76	มาก
4. สามารถเข้าถึงแหล่งความรู้ผลไม้เศรษฐกิจของไทย	4.10	0.40	มาก
5. การเชื่อมโยงไปยังแหล่งความรู้ที่พบ	4.13	0.57	มาก
รวม	4.23	0.24	มาก

ที่มา : จากการคำนวณ (ภาคผนวก ค)

จากตารางที่ 4.2 พบว่าผู้ใช้มีความพึงพอใจต่อระบบถามตอบเชิงความหมายสำหรับผลไม้เศรษฐกิจของไทย โดยรวมอยู่ในระดับมาก ($\bar{x}$=4.23) เมื่อพิจารณาเรียงลำดับเป็นรายข้อตามค่าเฉลี่ยจากมากไปหาน้อย 3 ตามลำดับแรก ได้ดังนี้ รูปแบบการระบุข้อคำถามเหมาะสม สะดวกต่อการใช้งาน ($\bar{x}$=4.73) การค้นคืนความรู้ตรงกับความต้องการ ($\bar{x}$=4.17) และการเชื่อมโยงไปยังแหล่งความรู้ที่พบ ($\bar{x}$=4.13)

4.3 การอภิปรายผล

กระบวนการถามตอบเชิงความหมายออนไลน์ที่ได้นำเสนอและพัฒนาระบบขึ้นสามารถค้นหาคำตอบ และค้นหาเอกสารความรู้ที่เกี่ยวข้อง มีค่าความแม่นยำเท่ากับ 87% ค่าระลึกเท่ากับ 85% ค่าอัตราการรู้จำ เท่ากับ 86% และจากผลการประเมินความพึงพอใจจากกลุ่มตัวอย่างผู้ใช้งานระบบพบว่าผู้ใช้มีความพึงพอใจต่อระบบถามตอบเชิงความหมายสำหรับผลไม้เศรษฐกิจของไทย โดยรวมอยู่ในระดับมากในรูปแบบการระบุข้อคำถามเหมาะสม สะดวกต่อการใช้งาน การค้นคืนความรู้ตรงกับความต้องการ และการเชื่อมโยงไปยังแหล่งความรู้ที่พบ เหตุที่เป็นเช่นนี้อาจเนื่องมาจากผู้วิจัยศึกษาหลักการ ทฤษฎีกระบวนการพัฒนาระบบถามตอบที่เกี่ยวข้องกับการ การสกัดข้อมูล การจัดการฐานความรู้ที่สามารถตอบสนองการค้นคืนเชิงความหมายด้วยออนโทโลยี และพฤติกรรมการค้นคืนสารสนเทศ ตลอดจนกระบวนการวิจัยได้ประยุกต์แนวทางการพัฒนาจากวงจรการพัฒนาระบบ SDLC (System Development Life Cycle) และในมุมมองของผู้ใช้งานระบบได้เอื้ออำนวยให้ผู้ใช้ระบุข้อคำถามเป็นประโยค ซึ่งได้นำแนวคิดประมวลผลภาษาธรรมชาติมาใช้กับกระบวนการถามตอบที่พัฒนาขึ้น หากเกษตรกรหรือผู้ที่สนใจต้องการสอบถามเกี่ยวกับความรู้ผลไม้เศรษฐกิจ ระบบนี้ก็นับว่าเป็นระบบที่สามารถช่วยให้คำตอบแก่ผู้ใช้ได้ แต่การให้คำตอบนั้นยังมีข้อจำกัดในส่วนของความรู้เฉพาะ พันธุ์ การปลูกและบำรุงรักษา โรคและแมลง สารเคมี การแปรรูป เครื่องจักรหรือเครื่องมือ

การตลาดและการส่งออก สำหรับผลไม้เศรษฐกิจ 3 ชนิด คือทุเรียน มังคุด และลำไย เท่านั้น หากต้องการให้ระบบมีความรู้และให้คำตอบได้นอกเหนือจากนี้ก็สามารถเพิ่มความรู้ที่เกี่ยวข้องกับผลไม้เศรษฐกิจของประเทศไทยในฐานความรู้ได้ และที่สำคัญกระบวนการถามตอบที่ได้นำเสนอนี้เป็นเพียงแนวทางสำหรับการพัฒนาระบบถามตอบในเรื่อง หรือโดเมนอื่น ๆ ได้ และควรมีการวิจัยหรือปรับปรุงเทคนิควิธีการประมวลผลคำถามเชิงความหมาย การจับคู่คำถามตอบ และการสร้างชุดคำตอบ อันจะส่งผลต่อการให้คำตอบที่ถูกต้อง และตรงกับความต้องการของผู้ใช้ให้มากที่สุด ซึ่งสอดคล้องกับการวิจัยของเกศสุดา ตุริยกรณ์ และชุลีรัตน์ จรัสกุลชัย (2546) ที่ได้เสนอแนวคิดกระบวนการเกี่ยวกับระบบคำถาม-คำตอบประยุกต์ใช้กับข้อความภาษาไทย โดยการจัดหมวดหมู่เอกสาร (Text Classification) ใช้คำถามที่เกี่ยวกับบุคคล (Who Question) คำถามเกี่ยวกับสถานที่ (Where Question) และคำถามเกี่ยวกับเวลา (When Question) และจะนำเอกสารที่มีหมวดหมู่ตรงกับคำถามมาเลือกคำตอบ (Answer Selection) ซึ่งเปรียบเทียบอัตราการเกิดร่วมกัน (Co-occurrence ratio) เพื่อให้ได้เอกสารที่มีความถูกต้อง และใกล้เคียงกับคำถามมากที่สุด ซึ่งระบบให้คำตอบตรงกับข้อคำถาม 63.3 % และสอดคล้องกับการวิจัยของเสกสรรค์ ศิวิลัย และจักรกฤษณ์ เสน่ห์ นมะหุต (2556) ได้เสนอแนวคิดกระบวนการเกี่ยวกับระบบคำถาม-คำตอบสำหรับการแนะนำอาหารที่เหมาะสมสำหรับผู้ป่วย โดยเน้นไปที่การแสดงวิธีการรวบรวม และวิธีการแปลงประโยคคำถามภาษาไทยเกี่ยวกับอาหารที่เหมาะสมสำหรับผู้ป่วยที่เป็นภาษาธรรมชาติ ใช้ฐานข้อมูลความรู้ออนโทโลยี แต่จากการศึกษานี้ทำให้พบว่าคำถามที่มีความยาว และคลุมเครือ ยังใช้ไม่ได้กับระบบนี้ หรือใช้กับระบบนี้แล้วได้ผลลัพธ์ที่ไม่มีประสิทธิภาพเพียงพอ และสอดคล้องกับการวิจัยของคิม และคณะ (Kim et al., 2001) ได้เสนอแนวคิดกระบวนการเกี่ยวกับระบบคำถาม-คำตอบข้อความภาษาเกาหลี โดยมีการสกัดลักษณะนามสร้างรูปแบบของการเข้าคู่ข้อคำถามกับคำตอบ และการค้นคืนความรู้ใช้วิธีการหาน้ำหนักคำด้วย TF-IDF นอกจากนี้ เรเดฟ และคณะ (Radev and et al., 2005) ยังได้เสนอแนวคิดเกี่ยวกับระบบคำถาม-คำตอบ นั้นจะประกอบด้วยกระบวนการประมวลผลคำถาม (Question Processing) การประมวลผลเอกสาร (Document Processing) การเปรียบเทียบคำถาม-คำตอบ (Question-Answering Matching) และการสร้างคำตอบ (Answer Generation)

บทที่ 5 สรุปผลการวิจัยและอภิปรายผล

จากการศึกษาวิจัยและพัฒนาระบบถามตอบเชิงความหมายสำหรับผลไม้เศรษฐกิจของไทยครั้งนี้มีผลการวิจัยและข้อเสนอแนะมีดังนี้

5.1 สรุปผลการวิจัย

งานวิจัยนี้ได้ศึกษาและวิเคราะห์กระบวนการและออกแบบพัฒนาระบบถามตอบเชิงความหมายออนไลน์สำหรับผลไม้เศรษฐกิจของไทย ซึ่งประยุกต์ใช้ออนโทโลยีผลไม้เศรษฐกิจ ThaiFo (ศุภกฤษฎิ์ นิวัฒนากูล, 2556) มาเป็นฐานความรู้ โดยได้ออกแบบระบบถามตอบเชิงความหมายออนไลน์สำหรับผลไม้เศรษฐกิจของไทยออกเป็นกระบวนการที่สำคัญ 4 ขั้นตอน ได้แก่

1) กระบวนการสร้างฐานความรู้ย่อยใช้กระบวนการตัดคำเอกสารที่จัดเก็บไว้ในฐานความรู้เปรียบเทียบการตัดคำกับคำที่จัดเก็บในพจนานุกรม มีการกำจัดคำหยุดเพื่อลดขนาดของดัชนี โดยหารากศัพท์จากความหมายของคำโดยผู้เชี่ยวชาญที่ทำหน้าที่รวบรวมคำศัพท์ แล้วทำการการสกัดคุณลักษณะสร้างดัชนี ด้วย TF-IDF เลือกคุณสมบัติด้วยความน่าจะเป็น และทำการจัดกลุ่ม

2) การประมวลผลคำถามเชิงความหมาย (Semantic question processing) ใช้ตัวแทนสำหรับการวิเคราะห์คำในประโยคข้อคำถาม และตัดคำให้เป็นพยางค์เดียวตามกฎวิธีการผันคำในภาษาไทยจากคลังข้อมูล (Corpus based approach)

3) การจับคู่คำถามตอบ (Semantic Q-A matching) จะทำการวัดความคล้ายคลึงด้วยเทคนิควิธีของแจ็กการ์ด (Jaccard) โดยจะค้นหาความรู้ หรือเอกสารที่เกี่ยวข้องทั้งหมด และทำการเปรียบเทียบระหว่างข้อคำถาม กับคำตอบที่อยู่ในฐานความรู้ และ

4) การสร้างชุดคำตอบ (Answer generation) มีการวิเคราะห์การจับคู่คำถามตอบเปรียบเทียบไวยกรณ์และทำการเลือกข้อความสั้นมานำเสนอเป็นผลลัพธ์

ผลการประเมินระบบในส่วนของการประเมินความแม่นยำและความสามารถ พบว่า ระบบที่นำเสนอมีค่าความแม่นยำเท่ากับ 87% ค่าความระลึกเท่ากับ 85% และค่าอัตราการรู้จำเท่ากับ 86% การประเมินผลความพึงพอใจโดยผู้เชี่ยวชาญพบว่า มีความพึงพอใจต่อระบบถามตอบเชิงความหมายสำหรับผลไม้เศรษฐกิจของไทยโดยรวมอยู่ในระดับมากในรูปแบบการระบุข้อคำถามเหมาะสม สะดวกต่อการใช้งาน การค้นคืนความรู้ตรงกับความต้องการ และการเชื่อมโยงไปยังแหล่งความรู้ที่พบ

5.2 ข้อเสนอแนะ

สำหรับกระบวนการการออกแบบกระบวนการถามตอบเชิงความหมายออนไลน์สำหรับผลไม้เศรษฐกิจของไทยสามารถเพิ่มความแม่นยำของระบบได้ในประเด็นดังต่อไปนี้

1) การพัฒนาโปรแกรมตัดคำภาษาไทยที่มีความถูกต้อง ซึ่งจากการวิจัยพบว่าการตัดคำยังมีข้อผิดพลาด ดังนั้น การเพิ่มความถูกต้องของการตัดคำภาษาไทยให้มีความถูกต้องมากยิ่งขึ้น จะทำให้มีผลต่อการเพิ่มขึ้นของความแม่นยำโดยตรง เนื่องจากเป็นกระบวนการแรกของระบบ หากมีการตัดคำที่ไม่ถูกต้องย่อมส่งผลต่อการดำเนินการในขั้นตอนต่อไป การตัดคำสำหรับคำถามจากผู้ใช้ในกระบวนการประมวลผลคำถามเชิงความหมาย ยังพบความผิดพลาดของการตัดคำ ดังน้น อาจมีการปรับการใช้เทคนิควิธีการตัดคำโดยอาจมีการเปรียบเทียบการตัดคำโดยใช้พจนานุกรม (Dictionary-Based Approach) กับการตัดคำโดยใช้คลังข้อมูล (Corpus-Based Approach) ซึ่งเป็นการตัดคำที่นำวิธีการทางสถิติ หรือวิธีการกลไกการเรียนรู้เข้ามาใช้ในกระบวนการประมวลผลทางภาษา

2) กระบวนการจับคู่คำถามตอบ ซึ่งวัดความคล้ายคลึงกันระหว่างรายการคำถามที่ผ่านการจัดรูปแบบกับคำตอบด้วยเทคนิควิธีแจ็กการ์ด สามารถนำวิธีการวัดความคล้ายคลึงด้วยวิธีแจ็กการ์ดมาใช้ในการตรวจสอบการมีอยู่ของตัวอักษรที่ซ้ำกันได้ ตัวอย่างเช่น คำว่า “ผลไม้เศรษฐกิจ” พิมพ์ตกเป็น “ผลไม้เศรษฐกจ” และ พิมพ์เกินหรือพิมพ์ซ้ำเป็น “ไม้ผลเศรรษฐกิจ” เพื่อที่จะกำหนดเป็นคำที่ถูกต้องหรือเลือกใช้คำที่มีอยู่ในคลังคำสำหรับใช้ในกระบวนการจับคู่คำถามตอบได้

3) การสร้างชุดคำตอบของข้อความที่เป็นคำตอบยังมีข้อจำกัดในการทำงาน ซึ่งข้อความภาษาไทยจะไม่มีรูปแบบในการเขียนที่แน่นอนการเขียนสามารถเขียนได้หลายรูปแบบ ส่วนกระบวนการสร้างชุดคำตอบควรปรับเปลี่ยนวิธีการเปรียบเทียบไวยากรณ์ให้มีความเหมาะสมอันส่งผลให้ระบบถามตอบสามารถตอบสนองต่อความต้องการของผู้ใช้ให้มากที่สุด โดยมีประสิทธิภาพของการให้คำตอบที่ตรงกับประเด็นคำถามมากขึ้น

ข้อเสนอแนะสำหรับการประยุกต์ใช้ระบบการถามตอบเชิงความหมายสำหรับการนำไปประยุกต์ใช้โดยการเพิ่มเนื้อหาความรู้ของผลไม้เศรษฐกิจอื่น ๆ เข้าไป เช่น ลองกอง เงาะ เป็นต้น จะทำให้ฐานความรู้มีความสมบูรณ์มากยิ่งขึ้น และนอกจากนี้สามารถนำกระบวนการในถามตอบไปประยุกต์ใช้กับโดเมนอื่น ๆ ที่คล้ายคลึงได้ เช่น พืชไร่ ผัก เป็นต้น

บรรณานุกรม

กานดา รุณนะพงศา และ ปโยธร อุราธรรมกุล. (2549). การตัดคำภาษาไทยโดยการปรับปรุงกฎและพจนานุกรมแบบใหม่. รายงานวิจัย คณะวิศวกรรมศาสตร์ มหาวิทยาลัยขอนแก่น.

เกศสุดา ตุริยากรณ์ และ ชุลีรัตน์ จรัสกุลชัย. (2546). ระบบคำถาม-คำตอบประยุกต์ใช้กับข้อความภาษาไทย. เรื่องเต็มการประชุมทางวิชาการของมหาวิทยาลัยเกษตรศาสตร์ ครั้งที่ 41: สาขาวิทยาศาสตร์ สาขาการจัดการทรัพยากรและสิ่งแวดล้อม, หน้า 76-83.

ไกรศักดิ์ เกสร. (2554). การค้นหาข้อมูลเชิงความหมาย: แนวคิดใหม่ของโปรแกรมการค้นหา (Search Engine) และแนวทางการพัฒนาในอนาคต. วารสารวไลยอลงกรณ์ปริทัศน์ ปีที่ 1, ฉบับที่ 2, หน้า 1-16.

ไกรศักดิ์ เกสร. (2555). การแสดงผลลัพธ์การค้นหาข้อมูล: จุดเล็กๆ ที่สำคัญและไม่ควรมองข้ามสำหรับเสริชเอ็นจิน. วารสารเทคโนโลยีสารสนเทศ. Vol. 8, No. 1, หน้า : 62-68.

จุฑาวรรณ สิทธิโชคสถาพร. (2555). ต้นแบบออนโทโลยีเพื่อการค้นคืนสารสนเทศเชิงความหมายสำหรับงานสารบรรณอิเล็กทรอนิกส์ กรณีศึกษา งานบริหารและธุรการ คณะแพทยศาสตร์ มหาวิทยาลัยสงขลานครินทร์. วิทยานิพนธ์ วิทยาศาสตร์มหาบัณฑิต สาขาวิชาการจัดการเทคโนโลยีสารสนเทศ มหาวิทยาลัยสงขลานครินทร์.

ณรงค์พิสิษฐ์ อึ้งรังษี. (2554). ระบบถามตอบปัญหาคอมพิวเตอร์ด้วยการค้นคืนรายกรณี. การศึกษาตามหลักสูตร วิทยาศาสตร์มหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ ภาควิชาเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

ธีรศักดิ์ สังข์ศรี. (2551). การใช้ฐานความรู้เวิร์ดเน็ตในการปรับปรุงการจัดกลุ่มเอกสารบนเว็บเชิงความหมาย. รายงานการศึกษาอิสระปริญญาวิทยาศาสตร์มหาบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น.

นภัสกร กรวยสวัสดิ์. (2555). Folksonomy กับการเพิ่มประสิทธิภาพในการค้นคืนสารสนเทศ. วารสารสารสนเทศศาสตร์, Volume 29, Issue 3, หน้า : 65-80.

นฤพนธ์ พนาวงศ์ และ จักรกฤษณ์ เสน่ห์. (2553). ระบบค้นหาสถานที่ท่องเที่ยวในประเทศไทยด้วยหลักการออนโทโลจีและเนมแมทชิ่ง. Journal of Information Science and Technology, Vol.1, Issue.2, pp.60-69.

นฤพนธ์ พนาวงศ์ และ จักรกฤษณ์ เสน่ห์. (2554). การพัฒนาระบบสืบค้นข้อมูลออนโทโลจีท่องเที่ยวด้วยภาษา SPARQL. เอกสารประกอบการประชุมวิชาการ “นเรศวรวิจัย” ครั้งที่ 7 กลุ่มวิทยาศาสตร์และเทคโนโลยี. หน้า 195-204.

นายพิชยา จึงสมานญาติ และ มณเฑียร รัตนศิริวงศ์วุฒิ. (2554). ระบบสารสนเทศเพื่อการจัดการงานบริการไฟฟ้าด้วยเทคนิคฐานกฎร่วมกับออนโทโลยี. เอกสารประกอบการประชุมเสนอผลงานวิจัยระดับบัณฑิตศึกษา มสธ. ครั้งที่ 1.

นิเวศ จิระวิชิตชัย,ปริญญา สงวนสัตย์ และพยุง มีสัจ. (2553). การจัดหมวดหมู่เอกสารภาษาไทยแบบอัตโนมัติด้วยซัพพอร์ตเวกเตอร์แมชชีน. การประชุมวิชาการระดับชาติ NCCIT2010, หน้า 87-82.

บารมี โอสธีรกุล และ ธวัชชัย งามสันติวงศ์. (2555). ระบบเทียบโอนรายวิชาโดยใช้แบบจำลองเวกเตอร์สเปซ. การประชุมเสนอผลงานวิจัยระดับบัณฑิตศึกษามหาวิทยาลัยสุโขทัย ธรรมาธิราช ครั้งที่ 2.

ปาทิตา สุขสมบูรณ์ การ์เซีย, อัจฉรา หลีระพงษ์ และ นันทิยา อริยะพิชัย. (2553). การประยุกต์ใช้เทคโนโลยีออนโทโลยีและ Semantic Web สำหรับระบบสืบค้นสารสนเทศการท่องเที่ยว. รายงานการวิจัย ทุนอุดหนุนวิจัย มหาวิทยาลัยสงขานครินทร์ วิทยาเขตตรัง.

พนิดา ตันศิริ. (มปป.). เว็บเชิงความหมายของเว็บ3.0. [ออนไลน์]. เข้าถึงได้จาก: http://www.bu.ac.th/knowledgecenter/executive_journal/oct_dec_10/pdf/aw6.pdf.

พยุง มีสัจ, วาทินี นุ้ยเพียร และ ผุสดี บุญรอด. (2556). ระบบการค้นคืนเชิงความหมายจากข้อมูลบรรณานุกรมโดเมน Information System. Journal of Information Science and Technology, Vol.4, No.1, pp.11-20.

พิชญสินี กิจวัฒนาถาวร. (2552). การพัฒนาระบบสกัดความรู้อัตโนมัติจากบทวิจารณ์ออนไลน์ของลูกค้า. วิทยานิพนธ์ ปริญญาวิทยาการสารสนเทศมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีสุรนารี.

เพ็ญพรรณ อัศวนพเกียรติ, อรินทิพย์ ธรรมชัยพิเนต และ กฤษณะ ไวยมัย. (2546). ออนโทโลยีชีวภาพ: ระบบสำหรับสืบค้นและวิเคราะห์ข้อมูลทางด้านชีววิทยา. เอกสารประกอบการประชุมทางวิชาการของมหาวิทยาลัยเกษตรศาสตร์ ครั้งที่ 41: สาขาวิศวกรรมศาสตร์และสาขาสถาปัตยกรรมศาสตร์, หน้า 277-285.

ภรศิษฐ์ กุหลาบทิพย์. (2550). การค้นหาข้อมูลบนเว็บเชิงความหมายโดยใช้ฐานความรู้ Wordnet. การศึกษาอิสระปริญญาวิทยาศาสตร์มหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น.

มานิตย์ สิงห์ทองชัย. (2557). การวิเคราะห์ตัวแบบเชิงปริมาณเพื่อการตัดสินใจ. พิมพ์ครั้งที่ 1 นครสวรรค์ : คณะวิทยาการจัดการ มหาวิทยาลัยราชภัฏนครสวรรค์.

_____. (2558). ผลการจัดการเรียนแบบใฝ่รู้ที่มีต่อพฤติกรรมการเรียนรู้ ผลสัมฤทธิ์ทางการเรียนและความพึงพอใจในรายวิชาการเงินธุรกิจ (3251101) ของนักศึกษาระดับปริญญาตรี. รายงานการวิจัย นครสวรรค์ : มหาวิทยาลัยราชภัฏนครสวรรค์.

เยาวลักษณ์ จิตรดารงค์ และ กฤดาภัทร สีหารี. (2556). การบูรณาการข้อมูลสารสนเทศขององค์กรชุมชนด้วยออนโทโลยี. เอกสารประกอบการประชุมวิชาการ 2013 International Computer Science and Engineering. หน้า : 460-465.

วิชุดา โชติรัตน์, ผุสดี บุญรอด และ ศจีมาจ ณ วิเชียร. (2554). การพัฒนาฐานความรู้ออนโทโลยีสำหรับวิเคราะห์ข่าวออนไลน์โดยอัตโนมัติ. วารสารเทคโนโลยีสารสนเทศ, Vol. 7, No. 14, หน้า : 13-18.

วิไลพร เลิศมหาเกียรติ และ อนิราช มิ่งขวัญ. (2553). นวัตกรรมการค้นคืนสารสนเทศแบบสหความสัมพันธ์. วารสารวิชาการพระจอมเกล้าพระนครเหนือ ปีที่ 20, ฉบับที่ 3, หน้า 514-523.

วิศิษฏ์ วรรณภูมิ และ ศิฬาณี นุชิตประสิทธิชัย. (2552). การสืบค้นข้อมูลการให้บริการเว็บเซอร์วิสเชิงความหมาย. เอกสารประกอบการประชุมวิชาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศครั้งที่ 5, หน้า : 158-163.

วิศิษฏ์ วรรณภูมิ. (2551). ระบบการสืบค้น ข้อมูลการให้บริการเว็บเซอร์วิสเชิงความหมาย. การศึกษาตางหลักสูตร วิทยาศาสตร์มหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

ศุภกฤษฎิ์ นิวัฒนากูล และคณะ. (2556). การเข้าถึงความรู้ทางการเกษตรด้วยเทคโนโลยีเว็บเชิงความหมาย. รายงานการวิจัย สาขาวิชาเทคโนโลยีสารสนเทศ สำนักวิชาเทคโนโลยีสังคม มหาวิทยาลัยเทคโนโลยีสุรนารี.

สมชาย ปราการเจริญ. (2548). ออนโทโลยี: ทางเลือกของการพัฒนาฐานความรู้ในรูปแบบเชิงเนื้อหา. เอกสารประกอบการประชุมทางวิชาการระดับชาติด้านคอมพิวเตอร์ และเทคโนโลยีสารสนเทศ ครั้งที่ 1 (NCCIT05), หน้า 92-99.

สมาน ลอยฟ้า. (2554). การค้นคืนสารสนเทศด้วยระบบ Federated Search. วารสารสารสนเทศศาสตร์, Volume 28, Issue 2, หน้า : 73-84.

สลิล โทไวยะ และ สมชาย ปราการเจริญ. (2553). การพัฒนาระบบสืบค้นฐานข้อมูลความรู้ในการแก้ไขปัญหาของระบบเครือข่ายแลนไร้สายด้วยเทคโนโลยีออนโทโลยี. เอกสารประกอบการประชุมวิชาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ครั้งที่ 6, หน้า : 928-933.

สำนักงานนโยบาย และยุทธศาสตร์การค้า. (2556). แนวโน้มการส่งออก 2556 และการทำธุรกิจ E-Commerce. วารสาร สนย., ปีที่ 3 ฉบับที่ 19.

สำนักงานเศรษฐกิจการเกษตร. (2555). รายงานประจำปี 2555. [ออนไลน์]. เข้าถึงได้จาก: http://www.oae.go.th/download/journal/AnnualOAE2555.pdf.

สิรินาฏ พรศิริประทาน. (มปป.). การส่งออกผักและผลไม้สดไทยไปสหภาพยุโรป. [ออนไลน์]. เข้าถึงได้จาก: www.itd.or.th/articles?download=74%3Aar40.

เสกสรรค์ ศิวิลัย และ จักรกฤษณ์ เสน่ห์ นมะหุต. (2556). การพัฒนาระบบคำถาม-คำตอบในการแนะนำอาหารที่เหมาะสมสำหรับผู้ป่วย. เรื่องเต็มประกอบการประชุมวิชาการระดับประเทศด้านเทคโนโลยีสารสนเทศ (National Conference on Information Technology: NCIT) ครั้งที่ 5, หน้า 167-172.

อรนุช บุญวัน และ ผุสดี บุญรอด. (2553). ระบบค้นคืนตราไปรษณีย์ยากรไทยโดยใช้การประมวลภาษาธรรมชาติ. เอกสารประกอบการประชุมวิชาการระดับชาติ ด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ครั้งที่ 6 (The 6th National Conference on Computing and Information Technology), หน้า 754-759.

อรวรรณ อุไรเรืองพันธุ์ และ สมจิตร อาจอินทร์. (2552). การสรุปเอกสารเชิงความหมายโดยใช้ออนโทโลยี. เอกสารประกอบการประชุมวิชาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ครั้งที่ 5, หน้า : 294-299.

อานนท์ ไกรเสวกวิสัย. (2552). ระบบสืบค้นรูปภาพบนอินเตอร์เน็ตโดยใช้หลักการเว็บเชิงความหมาย. การศึกษาเฉพาะบุคคล. วิทยาศาสตร์มหาบัณฑิต สาขาเทคโนโลยีสารสนเทศและการจัดการ มหาวิทยาลัยกรุงเทพ.

อิศเรศ ชาติวิริยะกุล. (2552). ระบบสืบค้นงานวิจัยด้านเทคโนโลยีสารสนเทศโดยวิธีเชิงความหมาย. การศึกษาตามหลักสูตร วิทยาศาสตร์มหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลันเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

มหาวิทยาลัยเทคโนโลยีสุรนารี

Maedche, A., and Staab, S. (2000). The text-to-onto ontology learning environment. In ICCS-2000 - Eight International Conference on Conceptual Structures, Software Demonstration, 2000.

Zouaq, A., and Nkambou, R. (2008). Building domain ontologies from text for educational purposes. IEEE Transaction on Learning Technologies. 1(1): 49-62.

Zouaq, A., and Nkambou, R. (2009). Enhancing learning objects with an ontology-based memory. IEEE Transactions on Knowledge and Data Engineering. 12(6): 881-893.

Chopra, A., Prashar, A., and Sain, C. (2013). Natural language processing. International Journal of Technology Enhancemrnt and Emerging Engineering Reserch. 1(4): 131-134.

Gouveia, A., and Cardoso, J. (2007). OWL – web ontology language. In Encyclopedia of Information Science and Technology, Information Science Reference, Pennsylvania, USA.

Apple Computer, Inc., (2005). Search Basics. [Online]. Available: https://developer.apple.com/library/mac/documentation/UserExperience/Conceptual/SearchKitConcepts/searchKit_basics/searchKit_basics.html.

Barry Smith. (2004). Beyond Concepts: Ontology as Reality Representation. Proceedings of FOIS International Conference on Formal Ontology and Information Systems.

Cambridge University. (2008). A first take at building an inverted index. [Online]. Available: http://nlp.stanford.edu/IR-book/html/htmledition/a-first-take-at-building-an-inverted-index-1.html.

Chen YongYue and Xia HuoSong. (2009). Research on Knowledge Extraction and Visualization in Knowledge Retrieve. Proceedings of the 2009 International Conference on Intelligent Human-Machine Systems and Cybernetics. pp : 66-69.

Dan Shen and Mirella Lapata. (2007). Using Semantic Roles to Improve Question Answering. Proceedings of the 2007 Joint Conference on Empirical Methods

in Natural Language Processing and Computational Natural Language Learning, pp : 12–21.

Deborah L. McGuinness and Frank van Harmelen. (2004). OWL Web Ontology Language Overview. [Online]. Available: http://www.w3.org/TR/owl-features/.

Devika K and Subu Surendran. (2013). An Overview of Web Data Extraction Techniques. International Journal of Scientific Engineering and Technology, Volume 2, Issue 4, pp : 278-287.

Deyou Tang and Yang Xu. (2012). An Ontology Framework for Evolution of Plain Text. Proceedings of the 2012 3rd International Conference on System Science, Engineering Design and Manufacturing Informatization, pp : 5-8.

Dieter Fensel. (2004). Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce. IEEE Inteligent System, 16(1), 8-14.

Dov Winer. (2014). Review of Ontology Based Storytelling Devices. Lecture Notes in Computer Science Volume 8002, pp 394-405.

Dragomir Radev, Weiguo Fan Hong Qi, Harris Wu and Amardeep Grewal. (2005). Probabilistic Question Answering on the Web. JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY, 56(6), pp :571–583.

Emilio Ferraraa, Pasquale De Meob, Giacomo Fiumarac and Robert Baumgartner. (2014). Web Data Extraction, Applications and Techniques: A Survey. Preprint submitted to Knowledge-based systems January 23, 2014.

Eric Miller. (1998). An Introduction to the Resource Description Framework. Bulletin of the American Society for Information Science, October/November.

Florian Eckerstorfer. (2011). Web Data Extraction - Overview and Comparison of selected State-of-the-Art Tools. [Online]. Available: http://storage.braincrafted.com /Eckerstorfer2011WebDataExtraction.pdf.

G. Attardi, A. Cisternino, F. Formica, M. Simi and A. Tommasi. (2001). PIQASso: Pisa question answering system. In TREC 2001. Available: http://citeseerx.ist.psu.edu/viewdoc /summary?doi=10.1.1.75.8404.

มหาวิทยาลัยเทคโนโลยีสุรนารี

Guiraude Lame. (2005). Using NLP Techniques to Identify Legal Ontology Components: Concepts and Relations. Lecture Notes in Artificial Intelligence, Volume 3369, pp : 169-184.

H. Wache, T. Vögele, U. Visser, H. Stuckenschmidt, G. Schuster, H. Neumann and S. Hübner . (2001). Ontology-based Integration of Information - A Survey of Existing Approaches. Proceedings of IJCAI-01 Workshop: Ontologies and Information Sharing, Seattle, WA, USA, pp. 108-117.

Hyun Namgoong, Moonyoung Chung, Kyung-il Kim, HyeonSung Cho and Yunku Chung. (2006). Effective Semantic Web Services Discovery using Usability. Proceedings of the 8th International Conference on Advanced Communication Technology, pp : 2199-2203.

I. Rožanc. (2013). Framework for Web Application Domain Knowledge Extraction. Proceedings of the 2013 36th International Convention on the Information & Communication Technology Electronics & Microelectronics (MIPRO) , pp : 705-710.

InfoLab Group. (1993). START Natural Language Question Answer System. [Online]. Available: http://start.csail.mit.edu/.

Jaruskulchai, C. (1998). An Automatic Indexing for Thai Text Retrieval. PhD Thesis, George Washington University. USA.

J. Fan, A. Kalyanpur, D. C. Gondek and D. A. Ferrucci. (2012). Automatic knowledge extraction from documents. IBM J. RES. & DEV., VOL. 56, NO. 3/4.

Jeff Heflin. (2004). OWL Web Ontology Language Use Cases and Requirements. [Online]. Available: http://www.w3.org/TR/2004/REC-webont-req-20040210/.

Joachims. (1998). Text categorization with support vector machines: Learning with mary relevant features. Proc. of the 10th European Conf. on Machine Learning, pp. 137-142.

Kanjana Charoenwattanachaikul. (2007). Cross-Language Question-Answering System. A Special Project Studies The Degree of Master of Engineering (Computer

Engineering) Faculty of Engineering King Mongkut's University Of Technology Thonburi.

M. F. Moens. (2006). Information Extraction: Algorithms and Prospects in a Retrieval Context. Springer : Netherlands, p 246.

Maire Casey and Claus Pahl. (2003). Web Components and the Semantic Web. Electronic Notes in Theoretical Computer Science 82, No. 5, pp : 156-163.

Mike Frey. (n.d.). Extensible Markup Language XML- Based Component Architecture Model.[Online].Available: http://www.liquidhub.com/docs/viewpoints/ Horizons_XML.pdf.

Nicola Guarino. (1998). Formal Ontology in Information System. 1st International conference On Formal Ontology in Information System (FOIS'98).

Quinlan, J. R. (1993). C4.5 : Programs for Machine Learning. Morgan Kaufmann Publishers.

Ruihua Song, Liqian Yu, Ji-Rong Wen and Hsiao-Wuen Hon. (2011). A Proximity Probabilistic Model for Information Retrieval. [Online]. Available: http://research.microsoft.com/ pubs/144542/ppm.pdf.

Samantha K. Rajapaksha and Nuwan Kodagod. (2008). Internal Structure and Semantic Web Link Structure Based Ontology Ranking. Proceedings of the 4th International Conference on Information and Automation for Sustainability, pp : 86-90.

Saowalak Sangkae. (2002). Question-Answering In Thai Language Using Syntactic-Semantic Analysis. A Special Project Studies The Degree of Master of Engineering (Computer Engineering) Faculty of Engineering King Mongkut's University Of Technology Thonburi.

Scott Farrar and Terry Langendoen (2003). A linguistic ontology for the semantic web. Glot International Vol. 7, No. 3.

Steffen Lamparter, Marc Ehrig and Christoph Tempich. (2013). Knowledge Extraction from Classification Schemas. Lecture Notes in Computer Science Volume 3290, 2004, pp 618-636.

Susan Dumais, Michele Banko, Eric Brill, Jimmy Lin and Andrew Ng. (2002). Web Question Answering: Is More Always Better?. Microsoft Research.

Thomas Krichel. (2002). The Semantic Web and an Introduction to RDF. [Online]. Available: http://openlib.org/home/krichel/papers/anhalter.letter.pdf.

Tom Gruber. (1993). Toward Principle for the Design of Ontology Used for Knowledge Sharing. [Online]. Available: http://tomgruber.org/writing/onto-design.html.

Tony Lawson. (2004). A Conception of Ontology. [Online]. Available: http://www .csog.group. cam. ac.uk /A_Conception_of_Ontology.pdf.

W3C. (1999). Resource Description Framework (RDF) Model and Syntax Specification. [Online]. Available: http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/.

W3C. (2015). Extensible Markup Language (XML). [Online]. Available: http:// www.w3.org/XML/.

WANG Yong-gui and JIA Zhen. (2010). Research on Semantic Web Mining. Proceedings of the 2010 International Conference On Computer Design And Appliations (ICCDA 2010), pp :67-70.

Xiao Hang Wang, Tao Gu, Da Qing Zhang and Hung Keng Pung. (2004). Ontology Based Context Modeling and Reasoning using OWL. Proceedings of the Second IEEE Annual Conference on Pervasive Computing and Communications Workshops, pp : 18-22.

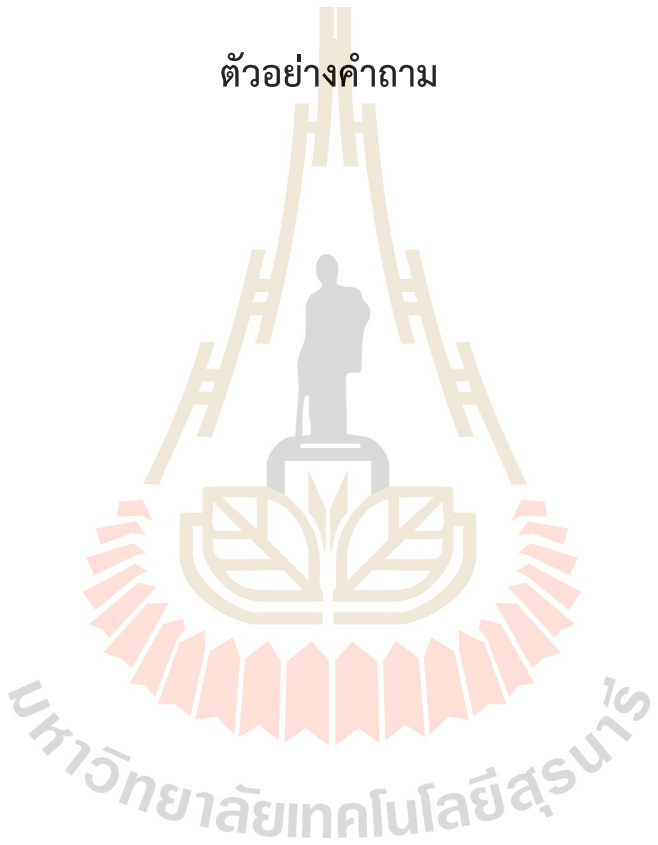
Ya Gao, Fang Yuan and Ming Zhang. (2010). Data Extraction Based on Index Path in Web. Proceedings of the 2010 Second International Workshop on Education Technology and Computer Science, pp : 157-160.

Yamane, Taro. (1967). Statistics : An introductory analysis. New York : Harper and Row.

Yinsheng Li, Yi Liao, Anting Zhang and Yueting Chai. (2011). Multi-Dimension Semantic Dictionary for Online Intelligence. Proceedings of the 2011 15th International Conference on Computer Supported Cooperative Work in Design, pp : 445-450.

ภาคผนวก ก

ตัวอย่างคำถาม



ตัวอย่างคำถาม จำนวน 100 คำถาม จากทั้งหมด 400 คำถาม

1. การปลูกทุเรียนมีวิธีการอย่างไรบ้าง
2. ทุเรียนมีการปลูกมากในจังหวัดใด
3. โรคในทุเรียนมีอะไรบ้าง
4. การไม่ตัดรากแก้วในทุเรียนเพราะอะไร
5. ทุเรียนชอบอากาศแบบใด
6. การขยายพันธุ์ทุเรียนทำได้อย่างไรบ้าง
7. การให้น้ำทุเรียนทำอย่างไร
8. ทุเรียนเริ่มออกดอกให้ผลเมื่อมีอายุเท่าไหร่
9. การดูแลทุเรียนอายุ 8-10 เดือนทำได้อย่างไร
10. การใส่ปุ๋ยทุเรียนในปีที่ 1 ทำได้อย่างไร
11. การใส่ปุ๋ยทุเรียนในปีที่ 2 ทำอย่างไร
12. การใส่ปุ๋ยทุเรียนในปีที่ 3 มีวิธีการอย่างไร
13. จะให้ปุ๋ยกับทุเรียนอายุ 4 ปีอย่างไร
14. การดูแลรักษามังคุดทำอย่างไร
15. ทุเรียนมีพันธุ์ที่สำคัญอะไรบ้าง
16. หนอนกินใบมังคุดจะแก้ไขอย่างไร
17. จะกำจัดหนอนกินใบมังคุดได้อย่างไร
18. ต้นมังคุดมีรูปร่างหน้าตาเป็นอย่างไร
19. เราสามารถปลูกมังคุดได้ทั่วไปทุกแห่งหรือไม่
20. มังคุดชอบอากาศแบบใด
21. มังคุดมีพันธุ์อะไรบ้าง
22. ที่ไหนรับซื้อมังคุดบ้าง
23. จะส่งออกมังคุดไปจำหน่ายที่ประเทศใดได้บ้าง
24. ทุเรียนสามารถแปรรูปเป็นผลิตภัณฑ์ได้บ้าง
25. ทุเรียนกวนทำอย่างไร
26. ลำไยจะเก็บผลผลิตได้เมื่อปลูกได้กี่ปี
27. โรคและแมลงทุเรียนมีอะไรบ้าง
28. วิธีการกำจัดโรคและแมลงในทุเรียน
29. การควบคุมการสุกของผลทุเรียน
30. การเก็บรักษาทุเรียนดิบบดแห้งทำอย่างไร
31. การป้องกันศัตรูทุเรียนที่สำคัญทำอย่างไร

มหาวิทยาลัยเทคโนโลยีสุรนารี

32. ส่งออกทุเรียนไปประเทศอะไรได้บ้าง
33. การบรรจุภัณฑ์ทุเรียนแปรรูป
34. มังคุดมีประโยชน์อย่างไร
35. โรคและแมลงมังคุด
36. การเก็บรักษาลำไย
37. การปลูกลำไยทำอย่างไร
38. การกำจัดด้วงในทุเรียนมีวิธีการใดบ้าง
39. การแปรรูปผลิตภัณฑ์น้ำมังคุดทำอย่างไรบ้าง
40. การป้องกันกำจัดเพลี้ยในมังคุด
41. โรคที่สำคัญของลำไยมีอะไรบ้าง
42. การขยายพันธ์ลำไยมีอยู่กี่วิธี
43. การปลูกมังคุดได้ทั่วไปทุกแหงเลยหรือไม
44. ทุเรียนไส้ซึม เกิดจากอะไร
45. จะป้องกันไรแดงในทุเรียนได้อย่างไร
46. จะใช้สารเคมีอะไรที่กำจัดไรแดง
47. เพลี้ยไก่แจ้ในทุเรียนมักเกิดในช่วงเวลาใด
48. วิธีการป้องกันเพลี้ยไก่แจ้ทำอย่างไร
49. สารเคมีอะไรที่ใช้ป้องกันโรครากเน่าและโคนเน่า
50. ลำไยชอบอากาศแบบไหนมากที่สุด
51. ทุเรียนพันธุ์ชนะมีลักษณะอย่างไร
52. ลำไยป่ามีลักษณะอย่างไร
53. ลักษณะของลำไยพื้นเมืองเป็นแบบใด
54. ลำไยที่ปลูกในประเทศไทยแบ่งออกเป็นกี่กลุ่ม อะไรบ้าง
55. ลำไยพันธุ์การค้าที่นิยมปลูกคือพันธุ์อะไร
56. ลำไยที่ให้ผลขนาดใหญ่ เนื้อหนา มีรสหวาน เมล็ดมีขนาดเล็ก ได้แก่พันธุ์อะไรบ้าง
57. ลำไยที่แบ่งตามลักษณะระยะการออกดอกได้แก่พันธุ์อะไร
58. อุณหภูมิที่เหมาะสมต่อการปลูกลำไยคือเท่าใด
59. พื้นที่เหมาะสมกับการปลูกลำไยควรมีลักษณะอย่างไร
60. ลำไยแทงช่อแล้วมีวิธีการดูแลอย่างไร
61. วิธีการดูแลลำไยดอกบาน
62. การให้ปุ๋ยลำไยมีวิธีการให้ปุ๋ยอย่างไร
63. การดูแลรักษาลำไยช่วงติดผลและเก็บเกี่ยว
64. แมลงชอบระบาดในลำไยมากที่สุดช่วงใด

มหาวิทยาลัยเทคโนโลยีสุรนารี

65. การป้องกันและกำจัดหนอนคืบลำไยทำอย่างไร
66. สารที่ใช้ฉีดพ่นหนอนคืบกินใบ และแมลงค่อมทองในลำไย
67. การป้องกันมวนลำไยควรฉีดพ่นด้วยสารชนิดใด
68. การป้องกันโรคหลังการเก็บเกี่ยวมีวิธีการป้องกันอย่างไร
69. ลำไยมีสรรพคุณทางยารักษาโรคอะไรได้บ้าง
70. รากลำไยใช้รักษาโรคอะไร
71. ชื่อวิทยาศาสตร์ของทุเรียนมีชื่อว่าอะไร
72. มังคุดมีชื่อวิทยาศาสตร์อะไร
73. ลำไยจำแกนได้กี่สายพันธุ์
74. การแตกใบครั้งที่หนึ่งของลำไยควรใช้ปุ๋ยสูตรอะไร
75. การป้องกันกำจัดโรคใบติดในทุเรียนทำอย่างไร
76. โรคราสีชมพูในทุเรียนเกิดจากอะไร
77. การป้องกันกำจัดโรคราสีชมพูในทุเรียนทำได้อย่างไร
78. โรคจุดสนิมในทุเรียนมีวิธีการป้องกันกำจัดอย่างไร
79. โรคผลเน่าในทุเรียนจะมีวิธีป้องกันกำจัดอย่างไร
80. ทุเรียนสามารถมีสรรพคุณรักษาโรคอะไรได้บ้าง
81. มังคุดอยู่ในวงศ์อะไร
82. ชื่อวิทยาศาสตร์ของมังคุดมีชื่อว่าอะไร
83. มังคุดมีชื่อภาษาอังกฤษว่าอะไร
84. ผลมังคุดเป็นแบบใด
85. การให้ปุ๋ยต้นมังคุดหลังการเก็บเกี่ยวอย่างไร
86. ช่วงมังคุดออกดอกควรให้ปุ๋ยอย่างไร
87. วิธีการให้ปุ๋ยหลังผลมังคุดติด 3-4 สัปดาห์
88. จะจัดการเพิ่มผลผลิตให้มังคุดมีคุณภาพได้อย่างไร
89. การเก็บเกี่ยวมังคุดทำอย่างไร
90. การปฏิบัติหลังการเก็บเกี่ยวมังคุดทำอย่างไร
91. การป้องกันกำจัดหนอนกินใบมังคุดทำอย่างไร
92. โรคพืชในมังคุดมีอะไรบ้าง
93. มังคุดมีศัตรูพืชอะไร
94. จะฉีดพ่นสารเคมีใดในการกำจัดโรคใบจุด
95. หนอนชอนใบมังคุดจะกำจัดด้วยสารเคมีใด
96. เพลี้ยไฟในมังคุดจะฉีดพ่นสารใด
97. จะกำจัดไรแดงที่เกิดในมังคุดได้อย่างไร

มหาวิทยาลัยเทคโนโลยีสุรนารี

98. พื้นที่ไหนที่เกษตรกรนิยมปลูกมังคุด
99. ดินแบบใดที่มังคุดจะเจริญเติบโตได้ดี
100. มังคุดมีสรรพคุณทางยาแก้โรคอะไร

ภาคผนวก ข

ผลการประเมินประสิทธิภาพจากการค้นคืนความรู้

มหาวิทยาลัยเทคโนโลยีสุรนารี

คำถาม	จำนวนคำตอบ		Classification context				Precision	Recall	F-measure
	เกี่ยวข้อง	ค้นพบ	TP	TN	FP	FN			
Q1	7	7	7	293	0	0	1.00	1.00	1.00
Q2	10	9	9	290	0	1	1.00	0.90	0.95
Q3	3	3	3	297	0	0	1.00	1.00	1.00
Q4	12	12	12	288	0	0	1.00	1.00	1.00
Q5	5	5	5	295	0	0	1.00	1.00	1.00
Q6	9	9	9	291	0	0	1.00	1.00	1.00
Q7	21	21	21	279	0	0	1.00	1.00	1.00
Q8	29	25	21	267	4	8	0.84	0.72	0.78
Q9	17	17	15	281	2	2	0.88	0.88	0.88
Q10	7	7	7	293	0	0	1.00	1.00	1.00
Q11	63	52	47	232	5	16	0.90	0.75	0.82
Q12	4	4	3	295	1	1	0.75	0.75	0.75
Q13	18	16	15	281	1	3	0.94	0.83	0.88
Q14	6	6	6	294	0	0	1.00	1.00	1.00
Q15	27	24	23	272	1	4	0.96	0.85	0.90
Q16	48	41	36	247	5	12	0.88	0.75	0.81
Q17	29	23	19	267	4	10	0.83	0.66	0.73
Q18	15	14	14	285	0	1	1.00	0.93	0.97
Q19	54	49	46	243	3	8	0.94	0.85	0.89
Q20	20	19	18	279	1	2	0.95	0.90	0.92
Q21	8	8	7	291	1	1	0.88	0.88	0.88
Q22	5	5	4	294	1	1	0.80	0.80	0.80
Q23	22	17	16	277	1	6	0.94	0.73	0.82
Q24	42	40	34	252	6	8	0.85	0.81	0.83
Q25	8	7	6	291	1	2	0.86	0.75	0.80
Q26	16	7	7	284	0	9	1.00	0.44	0.61
Q27	34	33	33	266	0	1	1.00	0.97	0.99
Q28	28	25	23	270	2	5	0.92	0.82	0.87

มหาวิทยาลัยเทคโนโลยีสุรนารี

คำถาม	จำนวนคำตอบ		Classification context				Precision	Recall	F-measure
	เกี่ยวข้อง	ค้นพบ	TP	TN	FP	FN			
Q29	31	28	24	265	4	7	0.86	0.77	0.81
Q30	36	27	26	263	1	10	0.96	0.72	0.83
Q31	22	25	22	275	3	0	0.88	1.00	0.94
Q32	33	31	29	265	2	4	0.94	0.88	0.91
Q33	24	23	22	275	1	2	0.96	0.92	0.94
Q34	4	3	3	296	0	1	1.00	0.75	0.86
Q35	20	20	19	279	1	1	0.95	0.95	0.95
Q36	8	7	7	292	0	1	1.00	0.88	0.93
Q37	32	31	29	266	2	3	0.94	0.91	0.92
Q38	41	39	32	252	7	9	0.82	0.78	0.80
Q39	19	21	18	278	3	1	0.86	0.95	0.90
Q40	6	4	4	294	0	2	1.00	0.67	0.80
Q41	39	39	35	257	4	4	0.90	0.90	0.90
Q42	46	48	42	248	6	4	0.88	0.91	0.89
Q43	38	42	35	255	7	3	0.83	0.92	0.88
Q44	42	40	37	255	3	5	0.93	0.88	0.90
Q45	26	23	19	270	4	7	0.83	0.73	0.78
Q46	27	29	23	267	6	4	0.79	0.85	0.82
Q47	29	34	26	263	8	3	0.76	0.90	0.83
Q48	10	13	9	286	4	1	0.69	0.90	0.78
Q49	8	8	7	291	1	1	0.88	0.88	0.88
Q50	33	34	31	264	3	2	0.91	0.94	0.93
Q51	26	25	25	274	0	1	1.00	0.96	0.98
Q52	34	35	28	259	7	6	0.80	0.82	0.81
Q53	49	48	43	246	5	6	0.90	0.88	0.89
Q54	13	12	12	287	0	1	1.00	0.92	0.96
Q55	49	44	36	243	8	13	0.82	0.73	0.77
Q56	14	10	10	286	0	4	1.00	0.71	0.83
Q57	9	13	9	287	4	0	0.69	1.00	0.82
Q58	35	36	34	263	2	1	0.94	0.97	0.96

มหาวิทยาลัยเทคโนโลยีสุรนารี

คำถาม	จำนวนคำตอบ		Classification context				Precision	Recall	F-measure
	เกี่ยวข้อง	ค้นพบ	TP	TN	FP	FN			
Q59	24	26	23	273	3	1	0.88	0.96	0.92
Q60	18	23	18	277	5	0	0.78	1.00	0.88
Q61	2	4	2	296	2	0	0.50	1.00	0.67
Q62	12	7	7	288	0	5	1.00	0.58	0.74
Q63	44	42	41	255	1	3	0.98	0.93	0.95
Q64	18	18	15	279	3	3	0.83	0.83	0.83
Q65	7	11	6	288	5	1	0.55	0.86	0.67
Q66	42	42	36	252	6	6	0.86	0.86	0.86
Q67	25	26	21	270	5	4	0.81	0.84	0.82
Q68	32	32	31	267	1	1	0.97	0.97	0.97
Q69	20	20	18	278	2	2	0.90	0.90	0.90
Q70	7	3	3	293	0	4	1.00	0.43	0.60
Q71	39	39	34	256	5	5	0.87	0.87	0.87
Q72	12	9	8	287	1	4	0.89	0.67	0.76
Q73	44	46	40	250	6	4	0.87	0.91	0.89
Q74	31	29	28	268	1	3	0.97	0.90	0.93
Q75	25	23	23	275	0	2	1.00	0.92	0.96
Q76	35	39	34	260	5	1	0.87	0.97	0.92
Q77	2	1	1	298	0	1	1.00	0.50	0.67
Q78	41	45	35	249	10	6	0.78	0.85	0.81
Q79	34	31	30	265	1	4	0.97	0.88	0.92
Q80	15	16	13	282	3	2	0.81	0.87	0.84
Q81	18	14	13	281	1	5	0.93	0.72	0.81
Q82	47	2	2	253	0	45	1.00	0.04	0.08
Q83	38	39	35	258	4	3	0.90	0.92	0.91
Q84	40	43	38	255	5	2	0.88	0.95	0.92
Q85	24	24	20	272	4	4	0.83	0.83	0.83
Q86	49	54	44	241	10	5	0.81	0.90	0.85
Q87	43	42	37	252	5	6	0.88	0.86	0.87
Q88	24	25	20	271	5	4	0.80	0.83	0.82

มหาวิทยาลัยเทคโนโลยีสุรนารี

คำถาม	จำนวนคำตอบ		Classification context				Precision	Recall	F-measure
	เกี่ยวข้อง	ค้นพบ	TP	TN	FP	FN			
Q89	21	18	18	279	0	3	1.00	0.86	0.92
Q90	18	16	15	281	1	3	0.94	0.83	0.88
Q91	19	19	17	279	2	2	0.89	0.89	0.89
Q92	14	15	14	285	1	0	0.93	1.00	0.97
Q93	8	8	7	291	1	1	0.88	0.88	0.88
Q94	38	35	33	260	2	5	0.94	0.87	0.90
Q95	6	5	5	294	0	1	1.00	0.83	0.91
Q96	8	7	7	292	0	1	1.00	0.88	0.93
Q97	23	27	22	272	5	1	0.81	0.96	0.88
Q98	48	50	47	249	3	1	0.94	0.98	0.96
Q99	13	14	11	284	3	2	0.79	0.85	0.81
Q100	48	51	40	241	11	8	0.78	0.83	0.81
Q101	3	3	3	297	0	0	1.00	1.00	1.00
Q102	42	45	37	250	8	5	0.82	0.88	0.85
Q103	44	47	37	246	10	7	0.79	0.84	0.81
Q104	24	28	21	269	7	3	0.75	0.88	0.81
Q105	36	30	26	260	4	10	0.87	0.72	0.79
Q106	12	12	12	288	0	0	1.00	1.00	1.00
Q107	20	19	18	279	1	2	0.95	0.90	0.92
Q108	49	41	34	244	7	15	0.83	0.69	0.76
Q109	50	53	42	239	11	8	0.79	0.84	0.82
Q110	44	52	37	241	15	7	0.71	0.84	0.77
Q111	31	25	21	265	4	10	0.84	0.68	0.75
Q112	29	26	24	269	2	5	0.92	0.83	0.87
Q113	9	9	7	289	2	2	0.78	0.78	0.78
Q114	23	23	19	273	4	4	0.83	0.83	0.83
Q115	43	35	33	255	2	10	0.94	0.77	0.85
Q116	42	34	31	255	3	11	0.91	0.74	0.82
Q117	4	4	3	295	1	1	0.75	0.75	0.75
Q118	12	10	9	287	1	3	0.90	0.75	0.82

มหาวิทยาลัยเทคโนโลยีสุรนารี

คำถาม	จำนวนคำตอบ		Classification context				Precision	Recall	F-measure
	เกี่ยวข้อง	ค้นพบ	TP	TN	FP	FN			
Q119	46	53	37	238	16	9	0.70	0.80	0.75
Q120	10	9	8	289	1	2	0.89	0.80	0.84
Q121	1	1	1	299	0	0	1.00	1.00	1.00
Q122	33	30	29	266	1	4	0.97	0.88	0.92
Q123	3	3	3	297	0	0	1.00	1.00	1.00
Q124	6	6	6	294	0	0	1.00	1.00	1.00
Q125	11	9	8	288	1	3	0.89	0.73	0.80
Q126	38	31	29	260	2	9	0.94	0.76	0.84
Q127	49	48	47	250	1	2	0.98	0.96	0.97
Q128	37	35	30	258	5	7	0.86	0.81	0.83
Q129	23	21	20	276	1	3	0.95	0.87	0.91
Q130	44	44	38	250	6	6	0.86	0.86	0.86
Q131	24	21	19	274	2	5	0.90	0.79	0.84
Q132	9	9	8	290	1	1	0.89	0.89	0.89
Q133	36	32	28	260	4	8	0.88	0.78	0.82
Q134	39	38	33	256	5	6	0.87	0.85	0.86
Q135	27	22	19	270	3	8	0.86	0.70	0.78
Q136	38	32	26	256	6	12	0.81	0.68	0.74
Q137	42	38	36	256	2	6	0.95	0.86	0.90
Q138	47	53	40	240	13	7	0.75	0.85	0.80
Q139	7	8	7	292	1	0	0.88	1.00	0.93
Q140	19	21	18	278	3	1	0.86	0.95	0.90
Q141	12	12	11	287	1	1	0.92	0.92	0.92
Q142	14	14	14	286	0	0	1.00	1.00	1.00
Q143	3	3	3	297	0	0	1.00	1.00	1.00
Q144	15	12	11	284	1	4	0.92	0.73	0.81
Q145	19	22	18	277	4	1	0.82	0.95	0.88
Q146	4	4	4	296	0	0	1.00	1.00	1.00
Q147	47	42	39	250	3	8	0.93	0.83	0.88
Q148	40	44	36	252	8	4	0.82	0.90	0.86

มหาวิทยาลัยเทคโนโลยีสุรนารี

คำถาม	จำนวนคำตอบ		Classification context				Precision	Recall	F-measure
	เกี่ยวข้อง	ค้นพบ	TP	TN	FP	FN			
Q149	28	30	25	267	5	3	0.83	0.89	0.86
Q150	21	24	21	276	3	0	0.88	1.00	0.93
Q151	43	45	41	253	4	2	0.91	0.95	0.93
Q152	31	25	22	266	3	9	0.88	0.71	0.79
Q153	23	23	21	275	2	2	0.91	0.91	0.91
Q154	32	28	24	264	4	8	0.86	0.75	0.80
Q155	20	17	16	279	1	4	0.94	0.80	0.86
Q156	25	28	21	268	7	4	0.75	0.84	0.79
Q157	36	37	32	259	5	4	0.86	0.89	0.88
Q158	7	7	6	292	1	1	0.86	0.86	0.86
Q159	29	24	21	268	3	8	0.88	0.72	0.79
Q160	4	4	4	296	0	0	1.00	1.00	1.00
Q161	15	18	14	281	4	1	0.78	0.93	0.85
Q162	15	13	12	284	1	3	0.92	0.80	0.86
Q163	7	7	6	292	1	1	0.86	0.86	0.86
Q164	30	24	20	266	4	10	0.83	0.67	0.74
Q165	48	52	45	245	7	3	0.87	0.94	0.90
Q166	14	16	13	283	3	1	0.81	0.93	0.87
Q167	38	39	32	255	7	6	0.82	0.84	0.83
Q168	14	12	11	285	1	3	0.92	0.79	0.85
Q169	34	37	34	263	3	0	0.92	1.00	0.96
Q170	26	28	23	269	5	3	0.82	0.88	0.85
Q171	32	37	27	258	10	5	0.73	0.84	0.78
Q172	37	35	31	259	4	6	0.89	0.84	0.86
Q173	12	13	10	285	3	2	0.77	0.83	0.80
Q174	19	18	17	280	1	2	0.94	0.89	0.92
Q175	50	46	40	244	6	10	0.87	0.80	0.83
Q176	4	4	4	296	0	0	1.00	1.00	1.00
Q177	35	38	34	261	4	1	0.89	0.97	0.93
Q178	1	1	1	299	0	0	1.00	1.00	1.00

มหาวิทยาลัยเทคโนโลยีสุรนารี

คำถาม	จำนวนคำตอบ		Classification context				Precision	Recall	F-measure
	เกี่ยวข้อง	ค้นพบ	TP	TN	FP	FN			
Q179	4	4	3	295	1	1	0.75	0.75	0.75
Q180	35	30	29	264	1	6	0.97	0.83	0.89
Q181	34	32	28	262	4	6	0.88	0.82	0.85
Q182	31	27	23	265	4	8	0.85	0.74	0.79
Q183	23	20	19	276	1	4	0.95	0.83	0.88
Q184	28	31	23	264	8	5	0.74	0.82	0.78
Q185	41	40	34	253	6	7	0.85	0.83	0.84
Q186	2	2	2	298	0	0	1.00	1.00	1.00
Q187	16	16	13	281	3	3	0.81	0.81	0.81
Q188	43	47	41	251	6	2	0.87	0.95	0.91
Q189	14	14	13	285	1	1	0.93	0.93	0.93
Q190	37	31	29	261	2	8	0.94	0.78	0.85
Q191	3	3	3	297	0	0	1.00	1.00	1.00
Q192	6	6	5	293	1	1	0.83	0.83	0.83
Q193	48	51	40	241	11	8	0.78	0.83	0.81
Q194	26	30	26	270	4	0	0.87	1.00	0.93
Q195	37	41	33	255	8	4	0.80	0.89	0.85
Q196	42	37	33	254	4	9	0.89	0.79	0.84
Q197	44	38	37	255	1	7	0.97	0.84	0.90
Q198	4	4	4	296	0	0	1.00	1.00	1.00
Q199	31	28	25	266	3	6	0.89	0.81	0.85
Q200	9	9	9	291	0	0	1.00	1.00	1.00
Q201	7	7	6	292	1	1	0.86	0.86	0.86
Q202	34	35	29	260	6	5	0.83	0.85	0.84
Q203	49	54	46	243	8	3	0.85	0.94	0.89
Q204	33	35	27	259	8	6	0.77	0.82	0.79
Q205	13	11	11	287	0	2	1.00	0.85	0.92
Q206	5	4	4	295	0	1	1.00	0.80	0.89
Q207	39	46	35	250	11	4	0.76	0.90	0.82
Q208	13	14	12	285	2	1	0.86	0.92	0.89

มหาวิทยาลัยเทคโนโลยีสุรนารี

คำถาม	จำนวนคำตอบ		Classification context				Precision	Recall	F-measure
	เกี่ยวข้อง	ค้นพบ	TP	TN	FP	FN			
Q209	31	26	21	264	5	10	0.81	0.68	0.74
Q210	44	42	37	251	5	7	0.88	0.84	0.86
Q211	0	0	0	300	0	0	0.00	0.00	0.00
Q212	33	33	31	265	2	2	0.94	0.94	0.94
Q213	15	17	14	282	3	1	0.82	0.93	0.88
Q214	4	4	4	296	0	0	1.00	1.00	1.00
Q215	42	47	36	247	11	6	0.77	0.86	0.81
Q216	2	2	2	298	0	0	1.00	1.00	1.00
Q217	50	47	39	242	8	11	0.83	0.78	0.80
Q218	46	48	41	247	7	5	0.85	0.89	0.87
Q219	25	25	24	274	1	1	0.96	0.96	0.96
Q220	43	43	41	255	2	2	0.95	0.95	0.95
Q221	6	6	6	294	0	0	1.00	1.00	1.00
Q222	7	7	7	293	0	0	1.00	1.00	1.00
Q223	4	4	4	296	0	0	1.00	1.00	1.00
Q224	4	4	3	295	1	1	0.75	0.75	0.75
Q225	39	39	32	254	7	7	0.82	0.82	0.82
Q226	5	6	4	293	2	1	0.67	0.80	0.73
Q227	49	53	48	246	5	1	0.91	0.98	0.94
Q228	47	39	37	251	2	10	0.95	0.79	0.86
Q229	15	16	15	284	1	0	0.94	1.00	0.97
Q230	4	4	4	296	0	0	1.00	1.00	1.00
Q231	26	31	22	265	9	4	0.71	0.85	0.77
Q232	34	36	30	260	6	4	0.83	0.88	0.86
Q233	25	27	21	269	6	4	0.78	0.84	0.81
Q234	47	52	44	245	8	3	0.85	0.94	0.89
Q235	28	27	26	271	1	2	0.96	0.93	0.95
Q236	13	13	13	287	0	0	1.00	1.00	1.00
Q237	29	27	26	270	1	3	0.96	0.90	0.93
Q238	25	21	18	272	3	7	0.86	0.72	0.78

มหาวิทยาลัยเทคโนโลยีสุรนารี

คำถาม	จำนวนคำตอบ		Classification context				Precision	Recall	F-measure
	เกี่ยวข้อง	ค้นพบ	TP	TN	FP	FN			
Q239	24	28	20	268	8	4	0.71	0.83	0.77
Q240	7	6	5	292	1	2	0.83	0.71	0.77
Q241	14	16	12	282	4	2	0.75	0.86	0.80
Q242	27	25	22	270	3	5	0.88	0.81	0.85
Q243	36	40	35	259	5	1	0.88	0.97	0.92
Q244	33	37	32	262	5	1	0.86	0.97	0.91
Q245	41	48	34	245	14	7	0.71	0.83	0.76
Q246	38	37	33	258	4	5	0.89	0.87	0.88
Q247	33	27	26	266	1	7	0.96	0.79	0.87
Q248	2	2	2	298	0	0	1.00	1.00	1.00
Q249	50	60	49	239	11	1	0.82	0.98	0.89
Q250	23	26	22	273	4	1	0.85	0.96	0.90
Q251	4	4	4	296	0	0	1.00	1.00	1.00
Q252	12	14	11	285	3	1	0.79	0.92	0.85
Q253	22	21	19	276	2	3	0.90	0.86	0.88
Q254	43	36	32	253	4	11	0.89	0.74	0.81
Q255	31	30	26	265	4	5	0.87	0.84	0.85
Q256	22	21	18	275	3	4	0.86	0.82	0.84
Q257	0	0	0	300	0	0	0.00	0.00	0.00
Q258	5	4	4	295	0	1	1.00	0.80	0.89
Q259	17	16	14	281	2	3	0.88	0.82	0.85
Q260	0	0	0	300	0	0	0.00	0.00	0.00
Q261	39	39	38	260	1	1	0.97	0.97	0.97
Q262	20	18	16	278	2	4	0.89	0.80	0.84
Q263	44	48	44	252	4	0	0.92	1.00	0.96
Q264	1	1	1	299	0	0	1.00	1.00	1.00
Q265	44	42	39	253	3	5	0.93	0.89	0.91
Q266	32	33	32	267	1	0	0.97	1.00	0.98
Q267	15	14	13	284	1	2	0.93	0.87	0.90
Q268	25	20	17	272	3	8	0.85	0.68	0.76

มหาวิทยาลัยเทคโนโลยีสุรนารี

คำถาม	จำนวนคำตอบ		Classification context				Precision	Recall	F-measure
	เกี่ยวข้อง	ค้นพบ	TP	TN	FP	FN			
Q269	42	49	41	250	8	1	0.84	0.98	0.90
Q270	19	22	17	276	5	2	0.77	0.89	0.83
Q271	42	34	31	255	3	11	0.91	0.74	0.82
Q272	14	16	12	282	4	2	0.75	0.86	0.80
Q273	40	41	38	257	3	2	0.93	0.95	0.94
Q274	28	29	26	269	3	2	0.90	0.93	0.91
Q275	6	7	5	292	2	1	0.71	0.83	0.77
Q276	49	57	46	240	11	3	0.81	0.94	0.87
Q277	9	8	7	290	1	2	0.88	0.78	0.82
Q278	5	5	5	295	0	0	1.00	1.00	1.00
Q279	25	25	22	272	3	3	0.88	0.88	0.88
Q280	15	13	12	284	1	3	0.92	0.80	0.86
Q281	48	52	46	246	6	2	0.88	0.96	0.92
Q282	32	30	24	262	6	8	0.80	0.75	0.77
Q283	39	36	34	259	2	5	0.94	0.87	0.91
Q284	14	14	13	285	1	1	0.93	0.93	0.93
Q285	3	3	3	297	0	0	1.00	1.00	1.00
Q286	38	33	32	261	1	6	0.97	0.84	0.90
Q287	8	9	7	290	2	1	0.78	0.88	0.82
Q288	35	30	27	262	3	8	0.90	0.77	0.83
Q289	43	43	41	255	2	2	0.95	0.95	0.95
Q290	46	53	43	244	10	3	0.81	0.93	0.87
Q291	15	14	13	284	1	2	0.93	0.87	0.90
Q292	7	8	7	292	1	0	0.88	1.00	0.93
Q293	15	18	12	279	6	3	0.67	0.80	0.73
Q294	47	39	37	251	2	10	0.95	0.79	0.86
Q295	39	39	33	255	6	6	0.85	0.85	0.85
Q296	39	34	28	255	6	11	0.82	0.72	0.77
Q297	49	46	46	251	0	3	1.00	0.94	0.97
Q298	40	39	38	259	1	2	0.97	0.95	0.96

มหาวิทยาลัยเทคโนโลยีสุรนารี

คำถาม	จำนวนคำตอบ		Classification context				Precision	Recall	F-measure
	เกี่ยวข้อง	ค้นพบ	TP	TN	FP	FN			
Q299	11	12	10	287	2	1	0.83	0.91	0.87
Q300	26	31	25	268	6	1	0.81	0.96	0.88
Q301	35	35	30	260	5	5	0.86	0.86	0.86
Q302	17	17	15	281	2	2	0.88	0.88	0.88
Q303	31	34	31	266	3	0	0.91	1.00	0.95
Q304	12	14	10	284	4	2	0.71	0.83	0.77
Q305	22	18	15	275	3	7	0.83	0.68	0.75
Q306	46	45	40	249	5	6	0.89	0.87	0.88
Q307	22	26	20	272	6	2	0.77	0.91	0.83
Q308	29	28	27	270	1	2	0.96	0.93	0.95
Q309	40	35	33	258	2	7	0.94	0.83	0.88
Q310	21	22	18	275	4	3	0.82	0.86	0.84
Q311	35	29	27	263	2	8	0.93	0.77	0.84
Q312	18	21	17	278	4	1	0.81	0.94	0.87
Q313	38	41	32	253	9	6	0.78	0.84	0.81
Q314	16	18	13	279	5	3	0.72	0.81	0.76
Q315	14	15	13	284	2	1	0.87	0.93	0.90
Q316	10	8	7	289	1	3	0.88	0.70	0.78
Q317	22	26	20	272	6	2	0.77	0.91	0.83
Q318	49	58	41	234	17	8	0.71	0.84	0.77
Q319	43	43	40	254	3	3	0.93	0.93	0.93
Q320	33	32	30	265	2	3	0.94	0.91	0.92
Q321	49	46	45	250	1	4	0.98	0.92	0.95
Q322	22	22	21	277	1	1	0.95	0.95	0.95
Q323	37	44	34	253	10	3	0.77	0.92	0.84
Q324	12	10	8	286	2	4	0.80	0.67	0.73
Q325	23	27	20	270	7	3	0.74	0.87	0.80
Q326	27	28	23	268	5	4	0.82	0.85	0.84
Q327	15	17	14	282	3	1	0.82	0.93	0.88
Q328	43	39	38	256	1	5	0.97	0.88	0.93

มหาวิทยาลัยเทคโนโลยีสุรนารี

คำถาม	จำนวนคำตอบ		Classification context				Precision	Recall	F-measure
	เกี่ยวข้อง	ค้นพบ	TP	TN	FP	FN			
Q329	44	42	41	255	1	3	0.98	0.93	0.95
Q330	39	40	35	256	5	4	0.88	0.90	0.89
Q331	11	12	9	286	3	2	0.75	0.82	0.78
Q332	25	22	22	275	0	3	1.00	0.88	0.94
Q333	28	33	25	264	8	3	0.76	0.89	0.82
Q334	27	30	24	267	6	3	0.80	0.89	0.84
Q335	20	16	13	277	3	7	0.81	0.65	0.72
Q336	0	0	0	300	0	0	0.00	0.00	0.00
Q337	24	24	20	272	4	4	0.83	0.83	0.83
Q338	44	42	41	255	1	3	0.98	0.93	0.95
Q339	18	15	14	281	1	4	0.93	0.78	0.85
Q340	25	26	21	270	5	4	0.81	0.84	0.82
Q341	46	50	43	247	7	3	0.86	0.93	0.90
Q342	2	2	2	298	0	0	1.00	1.00	1.00
Q343	45	36	34	253	2	11	0.94	0.76	0.84
Q344	13	13	11	285	2	2	0.85	0.85	0.85
Q345	17	17	15	281	2	2	0.88	0.88	0.88
Q346	14	12	10	284	2	4	0.83	0.71	0.77
Q347	2	2	2	298	0	0	1.00	1.00	1.00
Q348	42	36	32	254	4	10	0.89	0.76	0.82
Q349	2	2	2	298	0	0	1.00	1.00	1.00
Q350	48	48	40	244	8	8	0.83	0.83	0.83
Q351	28	33	24	263	9	4	0.73	0.86	0.79
Q352	39	44	39	256	5	0	0.89	1.00	0.94
Q353	24	27	22	271	5	2	0.81	0.92	0.86
Q354	0	0	0	300	0	0	0.00	0.00	0.00
Q355	35	29	29	265	0	6	1.00	0.83	0.91
Q356	6	7	6	293	1	0	0.86	1.00	0.92
Q357	5	4	4	295	0	1	1.00	0.80	0.89
Q358	38	32	30	260	2	8	0.94	0.79	0.86

มหาวิทยาลัยเทคโนโลยีสุรนารี

คำถาม	จำนวนคำตอบ		Classification context				Precision	Recall	F-measure
	เกี่ยวข้อง	ค้นพบ	TP	TN	FP	FN			
Q359	8	7	6	291	1	2	0.86	0.75	0.80
Q360	14	13	11	284	2	3	0.85	0.79	0.81
Q361	16	15	13	282	2	3	0.87	0.81	0.84
Q362	39	32	29	258	3	10	0.91	0.74	0.82
Q363	43	44	39	252	5	4	0.89	0.91	0.90
Q364	23	26	22	273	4	1	0.85	0.96	0.90
Q365	6	5	5	294	0	1	1.00	0.83	0.91
Q366	43	47	40	250	7	3	0.85	0.93	0.89
Q367	27	29	26	270	3	1	0.90	0.96	0.93
Q368	11	9	9	289	0	2	1.00	0.82	0.90
Q369	17	14	11	280	3	6	0.79	0.65	0.71
Q370	19	17	15	279	2	4	0.88	0.79	0.83
Q371	9	8	8	291	0	1	1.00	0.89	0.94
Q372	16	17	15	282	2	1	0.88	0.94	0.91
Q373	47	52	41	242	11	6	0.79	0.87	0.83
Q374	38	35	34	261	1	4	0.97	0.89	0.93
Q375	10	9	9	290	0	1	1.00	0.90	0.95
Q376	31	37	28	260	9	3	0.76	0.90	0.82
Q377	33	32	32	267	0	1	1.00	0.97	0.98
Q378	13	14	12	285	2	1	0.86	0.92	0.89
Q379	32	26	23	265	3	9	0.88	0.72	0.79
Q380	17	20	16	279	4	1	0.80	0.94	0.86
Q381	31	29	24	264	5	7	0.83	0.77	0.80
Q382	37	31	29	261	2	8	0.94	0.78	0.85
Q383	35	31	30	264	1	5	0.97	0.86	0.91
Q384	50	56	46	240	10	4	0.82	0.92	0.87
Q385	18	19	16	279	3	2	0.84	0.89	0.86
Q386	18	20	15	277	5	3	0.75	0.83	0.79
Q387	49	44	39	246	5	10	0.89	0.80	0.84
Q388	49	44	37	244	7	12	0.84	0.76	0.80

มหาวิทยาลัยเทคโนโลยีสุรนารี

คำถาม	จำนวนคำตอบ		Classification context				Precision	Recall	F-measure
	เกี่ยวข้อง	ค้นพบ	TP	TN	FP	FN			
Q389	27	22	18	269	4	9	0.82	0.67	0.73
Q390	43	42	40	255	2	3	0.95	0.93	0.94
Q391	0	0	0	300	0	0	0.00	0.00	0.00
Q392	5	4	3	294	1	2	0.75	0.60	0.67
Q393	26	25	22	271	3	4	0.88	0.85	0.86
Q394	35	37	33	261	4	2	0.89	0.94	0.92
Q395	40	40	34	254	6	6	0.85	0.85	0.85
Q396	16	13	11	282	2	5	0.85	0.69	0.76
Q397	16	18	14	280	4	2	0.78	0.88	0.82
Q398	39	40	35	256	5	4	0.88	0.90	0.89
Q399	32	26	25	267	1	7	0.96	0.78	0.86
Q400	7	6	6	293	0	1	1.00	0.86	0.92
เฉลี่ย							0.87	0.85	0.86

มหาวิทยาลัยเทคโนโลยีสุรนารี

ภาคผนวก ค

การประเมินความพึงพอใจ

มหาวิทยาลัยเทคโนโลยีสุรนารี

การประเมินแบบสอบถามความพึงพอใจ

คะแนนความพึงพอใจของผู้ใช้ที่มีต่อระบบถามตอบเชิงความหมายออนไลน์สำหรับผลไม้เศรษฐกิจของไทย

คนที่	Q1	Q2	Q3	Q4	Q5
1	4	4	3	4	4
2	5	4	5	5	4
3	5	4	3	5	3
4	5	4	3	4	3
5	4	5	4	5	4
6	4	5	3	5	4
7	4	4	4	4	4
8	5	4	5	4	4
9	5	4	4	4	5
10	5	4	4	4	5
11	5	5	4	4	5
12	5	4	4	4	5
13	5	5	3	4	5
14	5	4	4	4	4
15	5	4	5	4	4
16	5	4	5	4	4
17	5	4	5	4	4
18	5	4	4	4	4
19	5	5	4	4	4
20	5	4	5	4	4
21	4	3	4	4	4
22	5	4	5	4	4
23	4	4	4	4	4
24	5	4	5	4	4
25	5	4	3	3	5
26	4	4	3	4	5
27	4	4	3	4	4

มหาวิทยาลัยเทคโนโลยีสุรนารี

คนที่	Q1	Q2	Q3	Q4	Q5
28	5	5	4	4	4
29	5	4	4	4	4
30	5	4	5	4	3
$\overline{x}$	4.73	4.17	4.03	4.10	4.13
S.D.	0.45	0.61	0.76	0.40	0.57

หมายเหตุ

Q1 รูปแบบการระบุข้อคำถามเหมาะสม สะดวกต่อการใช้งาน

Q2 การค้นคืนความรู้ตรงกับความต้องการ

Q3 ผลลัพธ์ที่ได้จากการค้นคืนตรงกับความต้องการ

Q4 สามารถเข้าถึงแหล่งความรู้ผลไม้เศรษฐกิจของไทย

Q5 การเชื่อมโยงไปยังแหล่งความรู้ที่พบ

ผลลัพธ์การวิเคราะห์ด้วยโปรแกรมสำเร็จรูปทางสถิติเพื่อการวิจัยทางสังคมศาสตร์บนวินโดวส์ ดังนี้

Descriptive Statistics

	N	Mean	Std. Deviation
q1	30	4.73	.450
q2	30	4.17	.461
q3	30	4.03	.765
q4	30	4.10	.403
q5	30	4.13	.571
sum	30	4.23	.241
Valid N (listwise)	30		

ค่าความเชื่อมั่นของแบบสอบถามความพึงพอใจ

Reliability Statistics

Cronbach's Alpha	N of Items
.762	10

มหาวิทยาลัยเทคโนโลยีสุรนารี

ประวัติผู้วิจัย

ชื่อ นายศุภกฤษฎ์ นิวัฒนากูล
Mr.Suphakit Niwattanakul

ตำแหน่ง
ผู้ช่วยศาสตราจารย์ ประจำสาขาวิชาเทคโนโลยีสารสนเทศ สำนักวิชาเทคโนโลยีสังคม

ที่อยู่
สาขาวิชาเทคโนโลยีสารสนเทศ สำนักวิชาเทคโนโลยีสังคม
มหาวิทยาลัยเทคโนโลยีสุรนารี ต.สุรนารี อ.เมือง จ.นครราชสีมา 30000
โทรศัพท์ 044-224560 โทรสาร 044-224205
E-Mail suphakit@ccs.sut.ac.th

ประวัติการศึกษา

- Ph.D. (Computer Science) University of La Rochelle, France (2008)
- วท.ม. (เทคโนโลยีการจัดการระบบสารสนเทศ) มหาวิทยาลัยมหิดล (2542)
- ทษ.บ. (บริหารธุรกิจการเกษตร) มหาวิทยาลัยแม่โจ้ เชียงใหม่ (2533)

งานวิจัย

- ศุภกฤษฎ์ นิวัฒนากูล. (2555). การเข้าถึงความรู้ทางการเกษตรด้วยเทคโนโลยีเว็บเชิงความหมาย.
- ศุภกฤษฎ์ นิวัฒนากูล และนิศาชล จำนงศรี. (2555). การพัฒนาระบบการจัดการความรู้ในชุมชนเศรษฐกิจฐานรากในเขตอีสานใต้ ประเทศไทย.
- การเข้าถึงความรู้ด้วยตัวแบบออนโทโลยี (Access to Knowledge based on an Ontology Model). (2551). วิทยานิพนธ์ระดับปริญญาเอก : มหาวิทยาลัยลาโรแชล ประเทศฝรั่งเศส
- การพัฒนาต้นแบบสำหรับการสร้างแบบทดสอบผ่านระบบเครือข่ายอินเตอร์เน็ต (Web Based Testing Model Via Internet). (2546).
- ระบบการสอนเสริมที่ชาญฉลาดสำหรับการเรียนรู้โครงสร้างข้อมูล (Intelligence Tutoring System for Data Structure Learning). (2542). วิทยานิพนธ์ระดับปริญญาโท : มหาวิทยาลัยมหิดล

ชื่อ ฉัตรภัสร์ ฐิติอัคราวงศ์

Chatphat Titiakarawongse

ตำแหน่ง

อาจารย์ประจำสาขาวิชาเทคโนโลยีสารสนเทศ สำนักวิชาเทคโนโลยีสังคม

ที่อยู่

สาขาวิชาเทคโนโลยีสารสนเทศ สำนักวิชาเทคโนโลยีสังคม

มหาวิทยาลัยเทคโนโลยีสุรนารี ต.สุรนารี อ.เมือง จ.นครราชสีมา 30000

โทรศัพท์ 044-224514 โทรสาร 044-224205

E-Mail : chatphat@sut.ac.th

ประวัติการศึกษา

- ปริญญาเอก : Doctor of Information Technology Edith Cowan University (2555)
- ปริญญาโท : Master of Business Systems Monash University (2544)
- ปริญญาตรี : สังคมสงเคราะห์ศาสตร์บัณฑิต มหาวิทยาลัยธรรมศาสตร์ (2541)

งานวิจัย

- Subsorn, P., & Singh, K. (2007). DSS applications as a business enhancement strategy.
  Proceedings of the Third Transforming Information and Learning Conference (TILC 2007). Perth: ECU.
- Subsorn, P., Xiao, J., & Singh, K. (2008a). A web-based application of group decision
  making in a fuzzy environment. Proceedings of the Fifth International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ETICON 2008). Krabi, Thailand.
- Subsorn, P. (2008b). Enhancing rubber forecasting: The case of the Thai rubber industry. Poster session presented at the Ninth Postgraduate Electrical Engineering and Computing Symposium (PEECS 2008), Perth, Western Australia.

- Subsorn, P., & Xiao, J. (2009a). Developing forecasts for Thai rubber latex prices: Non-
neural network training and neural network training approaches. Proceedings of the Tenth Postgraduate Electrical Engineering and Computing Symposium (PEECS 2009). Perth: ECU.
- Subsorn, P., & Xiao, J. (2009b). Price forecasting in the public agricultural rubber industry in Thailand: A preliminary investigation. Proceedings of the Third International Conference on Artificial Intelligence in Science and Technology (AISAT 2009). Hobart, Tasmania.
- Subsorn, P., Xiao, J., & Clayden, J. (2010). Forecasting rubber production using intelligent time series analysis to support decision makers. In G. Devlin (Ed.), Advances in Decision Support Systems. Vienna, Austria: IN-TECH Education and Publishing.
- Subsorn, P., & Limwiriyakul, S. (2011a). A comparative analysis of the security of internet banking in Australia: A customer perspective. Proceedings of the 2nd International Cyber Resilience Conference (ICR2011), Perth, Western Australia.
- Subsorn, P. & Limwiriyakul, S. (2011b). A comparative analysis of internet banking security in Thailand: A customer perspective. Proceedings of the 3rd International Social Science, Engineering and Energy Conference 2011 (I-SEEC2011), Nakhon Pathom, Thailand.
- Xiao, J. & Subsorn, P. (2012a). A new latex price forecasting model to reduce the risk
of rubber overproduction in Thailand. In Jie Lu, Lakhmi Jain, Guangquan Zhang (Ed.). Risk Management in Decision Making: Intelligent Methodologies and Application, Springer.
- Subsorn, P. & Limwiriyakul, S. (2012b). An analysis of internet banking security of foreign subsidiary banks in Australia: A customer perspective. International Journal of Computer Science Issues. IJCSI, 9 (2).
- Limwiriyakul S. & Subsorn, P. (2012c). A customer perspective investigation on internet banking security of licensed banks in Hong Kong," Proceeding of the 2012 International Conference on Security and Management (SAM'12), Las Vegas, USA.