รหัสโครงการ SUT2-204-59-12-04

รายงานการวิจัย

ตัวแบบการค้นคืนเชิงความหมายสำหรับความรู้ผลไม้เศรษฐกิจไทยบนฐานออนโทโลยี

(Semantic Retrieval Model for Thai Economic Fruit Knowledge Based on Ontology)

ได้รับทุนอุดหนุนการวิจัยจาก

มหาวิทยาลัยเทคโนโลยีสุรนารี

ผลงานวิจัยเป็นความรับผิดชอบของหัวหน้าโครงการวิจัยแต่เพียงผู้เดียว

รหัสโครงการ SUT2-204-59-12-04

รายงานการวิจัย

ตัวแบบการค้นคืนเชิงความหมายสำหรับความรู้ผลไม้เศรษฐกิจไทยบนฐานออนโทโลยี

(Semantic Retrieval Model for Thai Economic Fruit Knowledge Based on Ontology)

คณะผู้วิจัย

หัวหน้าโครงการ

ผู้ช่วยศาสตราจารย์ ดร.ศุภกฤษฎิ์ นิวัฒนากูล

สาขาวิชาเทคโนโลยีสารสนเทศ สำนักวิชาเทคโนโลยีสังคม

มหาวิทยาลัยเทคโนโลยีสุรนารี

ได้รับทุนอุดหนุนการวิจัยจากมหาวิทยาลัยเทคโนโลยีสุรนารี ประจำงบประมาณ พ.ศ. 2559

มหาวิทยาลัยเทคโนโลยีสุรนารี

ผลงานวิจัยเป็นความรับผิดชอบของหัวหน้าโครงการวิจัยแต่เพียงผู้เดียว

เมษายน 2565

กิตติกรรมประกาศ

การวิจัยครั้งนี้ได้รับทุนอุดหนุนการวิจัยจากมหาวิทยาลัยเทคโนโลยีสุรนารี ประจำปีงบประมาณ พ.ศ. 2559 และงานวิจัยฉบับนี้สำเร็จลุล่วงได้ด้วยดี ด้วยเพราะได้รับการสนับสนุนอันดีจากสาขาวิชาเทคโนโลยีสารสนเทศ สถานวิจัย สำนักวิชาเทคโนโลยีสังคม มหาวิทยาลัยเทคโนโลยีสุรนารี ที่อำนวยความสะดวกด้านการดำเนินการต่าง ๆ ที่เกี่ยวข้องกับงานวิจัย ขอขอบคุณผู้มีส่วนเกี่ยวข้องทุกท่าน รวมถึงผู้เชี่ยวชาญ ผู้ทดสอบการใช้งานระบบที่มีส่วนช่วยในขั้นตอนต่าง ๆ ของการพัฒนาและประเมินระบบต้นแบบ นอกจากนี้ ยังได้รับความกรุณาอย่างดียิ่งจากผู้ทรงคุณวุฒิที่ได้เสียสละเวลาในการตรวจข้อเสนอโครงการวิจัย รวมถึงรายงานการวิจัยฉบับสมบูรณ์ และในโอกาสนี้ ผู้วิจัยขอขอบคุณทุก ๆ ท่าน ที่มิได้กล่าวถึงทั้งหมด ที่มีส่วนช่วยเหลือ สนับสนุน และให้กำลังใจจนกระทั่งงานวิจัยนี้ได้สำเร็จลุล่วงไปได้ด้วยดี

คุณค่าและประโยชน์อันจะพึงมีจากงานวิจัยฉบับนี้ขอมอบแด่บิดามารดา ครูอาจารย์ ครอบครัว และเพื่อนร่วมงานในสำนักวิชาเทคโนโลยีสังคม มหาวิทยาลัยเทคโนโลยีสุรนารี ทุกท่านที่ให้กำลังใจและช่วยเหลือตลอดมา

ผู้วิจัย

มหาวิทยาลัยเทคโนโลยีสุรนารี

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อ 1) ออกแบบและพัฒนาคลังคำด้วยการจัดทำดัชนีเชิงความหมายสำหรับผลไม้เศรษฐกิจไทย และ 2) เพื่อออกแบบ พัฒนา และประเมินตัวแบบค้นคืนเชิงความหมายสำหรับผลไม้เศรษฐกิจไทย เพื่อช่วยให้ผู้ใช้ได้เข้าถึงความรู้เกี่ยวกับผลไม้เศรษฐกิจไทยได้อย่างสะดวกและรวดเร็ว ผลการวิจัยได้พัฒนาคลังคำโดยใช้แนวทางการจัดทำดัชนีเชิงความหมายโดยนำภาษาเอสเคโอเอสมาประยุกต์ใช้ ซึ่งคลังคำที่พัฒนาขึ้นจำแนกตามหมวดหมู่ได้ 10 คอลเล็คชัน ประกอบด้วยคอนเซ็ปต์จำนวน 101 คอนเซ็ปต์ การออกแบบออนโทโลยีตามแนวทางกระบวนการวิศวกรรมออนโทโลยีประกอบด้วยคลาสหลักจำนวน 4 คลาส และคลาสย่อยจำนวน 12 คลาส การออกแบบออนโทโลยีโดยสร้างคุณสมบัติประเภท Datatype properties จำนวน 1 คุณสมบัติ คุณสมบัติประเภท Object properties จำนวน 12 คุณสมบัติสำหรับจัดเก็บรายการความรู้ เชื่อมโยงรายการความรู้ภายในฐานความรู้ และเชื่อมโยงรายการความรู้กับคำสำคัญในคลังคำ ใช้ภาษา SWRL สำหรับการสร้างกฎเพื่อเพิ่มเติมความรู้ในฐานความรู้ ตัวแบบการค้นคืนเชิงความหมายประกอบด้วย 3 ขั้นตอน 1) การประมวลคำสำคัญ โดยมีการประยุกต์ใช้ค่าสัมประสิทธิ์ความคล้ายคลึงของแจ็กการ์ด เพื่อระบุคำค้น 2) การประมวลผลคำค้น เพื่อสร้างคำสั่งสำหรับการค้นหาด้วยภาษา SPARQL สำหรับการค้นหาแบบ Concept-based และ Keyword-based และ 3) การแสดงผลลัพธ์ที่มีการคำนวณ ค่าความคล้ายคลึงเชิงมุมโคไซน์เพื่อหาคล้ายคลึงของแต่ละรายการความรู้ ผลการประเมินตัวแบบโดยการทดลองการค้นหา 30 ชุดคำค้นที่แตกต่างกัน แล้วนำมาวิเคราะห์ค่าความแม่นยำได้ร้อยละ 91.39 ค่าความระลึกร้อยละ 83.44 และค่าอัตราการรู้จำร้อยละ 87.23 ผลการประเมินระบบการค้นคืนเชิงความหมายที่พัฒนาขึ้นมาโดยทดสอบการค้นทั้งหมด 400 รายการ ผลการประเมินระบบในภาพรวมได้ค่าความแม่นยำร้อยละ 63.31 ค่าความระลึกร้อยละ 95.74 และค่าอัตราการรู้จำร้อยละ 76.22

คำสำคัญ: ตัวแบบการค้นคืนเชิงความหมาย, ดัชนีเชิงความหมาย, การวัดค่าความคล้ายคลึง, ผลไม้เศรษฐกิจไทย

มหาวิทยาลัยเทคโนโลยีสุรนารี

Abstract

This research aimed to 1) design and develop a word repository by using semantic indexation for Thai economic fruits and 2) design, develop, and evaluate a semantic search model to support users to quickly and conveniently access Thai economic fruits knowledge. The research results were to develop word repository using a semantic indexing approach by applying the SKOS language. The developed word repository can be classified into 10 collections, consisiting of 101 concepts. Designing an ontology based on an ontology engineering approach. It consists of 4 main classes and 12 subclasses. The ontology was designed by 1 dataype property and 12 object properties for describing knowledge and associating knowedge within their knowledge base and between the word repository. The SWRL language was used for defining rules to build additional knowledge in the knowledge base. The sementic retrieval model consists of 3 steps: 1) keyword processing 2) query processing and 3) display of results. The Jaccard similarity coefficient was applied to identify keywords to search knowledge in the knowledge base. The SPARQL language was used for querying based on concept-based and keyword-based search. The Cosine Similrity Measurement was applied to identiy the similarity of knowledge for displaying the results to users. The evaluation of semantic retrieval model by searching experiments on 30 different search terms showed a Precision of 91.39%, Recall of 83.44% and F-measure of 87.23%. The evaluation of semantic retrieval system by using experiments on 400 different search terms showed a Precision of 63.31%, Recall of 95.74% and F-measure of 76.22%.

มหาวิทยาลัยเทคโนโลยีสุรนารี

Keywords: Semantic retrieval model, Semantic indexation, Similarity measure, Thai economic fruits

สารบัญ

หน้า

กิตติกรรมประกาศ ..... ก
บทคัดย่อภาษาไทย ..... ข
บทคัดย่อภาษาอังกฤษ ..... ค
สารบัญ ..... ง
สารบัญตาราง ..... ช
สารบัญภาพ ..... ซ
บทที่ 1 บทนำ 1
1.1 ความสำคัญและที่มาของปัญหาการวิจัย 1
1.2 วัตถุประสงค์ของการวิจัย 3
1.3 สมมติฐานการวิจัย 3
1.4 ขอบเขตของการวิจัย 3
1.5 ข้อตกลงเบื้องต้น 4
1.6 ประโยชน์ที่ได้รับจากการวิจัย 4
1.7 คำอธิบายศัพท์ 5
บทที่ 2 ปริทัศน์วรรณกรรมและงานวิจัยที่เกี่ยวข้อง 6
2.1 การค้นคืนเชิงความหมาย (Semantic Retrieval) 6
2.2 โปรแกรมค้นหาเชิงความหมาย 8
2.2.1 ความหมายและคุณลักษณะของโปรแกรมค้นหาเชิงความหมาย 8
2.2.2 ประเภทของโปรแกรมค้นหาเชิงความหมาย 11
2.2.3 องค์ประกอบของโปรแกรมค้นหาเชิงความหมาย 12
2.2.4 การทำงานของโปรแกรมค้นหาเชิงความหมาย 14
2.2.5 ตัวอย่างโปรแกรมค้นหาเชิงความหมาย 14
2.3 การจัดเก็บความรู้และการจัดความรู้ให้เป็นระบบ 19

มหาวิทยาลัยเทคโนโลยีสุรนารี

2.3.1 ความรู้และการจัดการความรู้ 19
2.3.2 ฐานความรู้ด้วยตัวแบบออนโทโลยี 21
2.3.3 การออกแบบออนโทโลยี 25
2.3.4 การจัดความรู้ให้เป็นระบบ 26
2.4 การค้นคืนจากฐานความรู้ในออนโทโลยี 28
2.4.1 ภาษาสำหรับการค้นคืนจากฐานความรู้ในออนโทโลยี 28
2.4.2 ภาษากฎสำหรับฐานความรู้ในออนโทโลยี 28
2.5 การวัดค่าความคล้ายคลึง 31
2.5.1 การวัดค่าความคล้ายคลึงเชิงมุมโคไซน์ 31
2.5.2 การวัดค่าสัมประสิทธิ์ความคล้ายคลึงของแจ็กการ์ด 34
2.6 การประเมินประสิทธิภาพของการค้นหา 36
2.7 เอกสารและงานวิจัยที่เกี่ยวข้อง 38
2.7.1 การพัฒนาฐานความรู้ 38
2.7.2 การค้นคืนเชิงความหมาย 38
2.7.3 การประเมินประสิทธิภาพการค้นคืน 39
2.8 กรอบแนวคิดการวิจัย 40
บทที่ 3 วิธีการดำเนินการวิจัย 42
3.1 การศึกษาข้อมูลเบื้องต้น 42
3.2 การวิเคราะห์และออกแบบตัวแบบ 43
3.3 การพัฒนาตัวแบบ 44
3.4 การประเมินตัวแบบ 45
บทที่ 4 ผลการวิจัยและอภิปรายผล 46
4.1 ผลการออกแบบฐานความรู้สำหรับผลไม้เศรษฐกิจไทย 46
4.1.1 คลังคำ 46

มหาวิทยาลัยเทคโนโลยีสุรนารี

4.1.2 ออนโทโลยีผลไม้เศรษฐกิจไทย 48
4.2 ผลการพัฒนาตัวแบบการค้นคืนเชิงความหมาย 55
4.2.1 การประมวลผลคำสำคัญ 55
4.2.2 การประมวลผลคำค้น 58
4.2.3 การแสดงผลลัพธ์ 62
4.2.4 การประเมินผลตัวแบบ 65
4.3 ผลการพัฒนาระบบการค้นคืนเชิงความหมาย 65
4.4 การประเมินผลระบบและการอภิปรายผล 66
บทที่ 5 สรุปผลการวิจัยและอภิปรายผล 70
5.1 สรุปผลการวิจัย 70
5.2 ข้อเสนอแนะ 73
บรรณานุกรม 74
ภาคผนวก ก ผลการประเมินตัวแบบการค้นคืนเชิงความหมาย 79
ภาคผนวก ข รายละเอียดโปรแกรม 82
ภาคผนวก ค ผลการประเมินระบบค้นคืนเชิงความหมาย 109
ประวัติผู้วิจัย 119

มหาวิทยาลัยเทคโนโลยีสุรนารี

สารบัญตาราง

ตารางที่ 2.1 ความแตกต่างระหว่างโปรแกรมค้นหาคำสำคัญและโปรแกรมค้นหาเชิงความหมาย 9
ตารางที่ 2.2 การเปรียบเทียบโปรแกรมค้นหาเชิงความหมาย 18
ตารางที่ 2.3 เทคโนโลยีเว็บและการจัดการความรู้ 21
ตารางที่ 2.4 ตัวอย่างข้อมูลสำหรับคำนวณค่าความคล้ายคลึงมุมโคไซน์ 33
ตารางที่ 2.5 Confusion Matrix 36
ตารางที่ 2.6 สรุปเปรียบเทียบงานวิจัยที่เกี่ยวข้องตัวแบบค้นคืนเชิงความหมาย 40
ตารางที่ 4.1 ขอบเขตข้อมูลที่จัดเก็บของออนโทโลยีผลไม้เศรษฐกิจไทย 49
ตารางที่ 4.2 รายละเอียดของ Data type properties ในฐานความรู้ 51
ตารางที่ 4.3 รายละเอียดของ Object properties ในฐานความรู้ 51
ตารางที่ 4.4 คุณสมบัติเพื่อเชื่อมโยงความสัมพันธ์ระหว่างฐานความรู้กับคลังคำ 52
ตารางที่ 4.5 กลุ่มคำจากผู้ใช้และกลุ่มคำสำคัญในคลังคำ 56
ตารางที่ 4.6 ตัวอย่างการคำนวนค่าความคล้ายคลึงของคำสำคัญจากผู้ใช้กับคำในคลังคำ 57
ตารางที่ 4.7 ตัวอย่างการจำแนกกลุ่มของรายการคำค้น 59
ตารางที่ 4.8 ตัวอย่างผลลัพธ์ที่ได้จากการค้นหาในฐานความรู้ตามรายการคำค้นที่กำหนด 61
ตารางที่ 4.9 ตัวอย่างผลลัพธ์ที่ได้จากการค้นหาในฐานความรู้โดยใช้คำสำคัญจากผู้ใช้ 62
ตารางที่ 4.10 ข้อมูลสำหรับคำนวณค่าความคล้ายคลึงเชิงมุมโคไซน์ของรายการความรู้ 64
ตารางที่ 4.11 ผลการประเมินตัวแบบ 65
ตารางที่ 4.12 ผลการประเมินระบบการค้นคืนเชิงความหมาย 67

มหาวิทยาลัยเทคโนโลยีสุรนารี

สารบัญภาพ

ภาพที่ 2.1 องค์ประกอบของโปรแกรมค้นหาเชิงความหมาย 12
ภาพที่ 2.2 ตัวอย่างภาพหน้าจอของโปรแกรมค้นหาเชิงความหมาย DuckDuckGo 15
ภาพที่ 2.3 ตัวอย่างภาพหน้าจอโปรแกรมค้นหาเชิงความหมาย Sensebot 16
ภาพที่ 2.4 ตัวอย่างภาพหน้าจอโปรแกรมค้นหาเชิงความหมาย Swoogle 17
ภาพที่ 2.5 ความสัมพันธ์ระหว่างข้อมูล สารสนเทศ และความรู้ 20
ภาพที่ 2.6 สถาปัตยกรรมของเว็บเชิงความหมาย 22
ภาพที่ 2.7 ตัวอย่างความสัมพันธ์ของคำตามแบบโครงสร้างของเอสเคโอเอส 27
ภาพที่ 2.8 รูปแบบของภาษาเอ็กซ์เอ็มแอลของการสร้างกฎ 30
ภาพที่ 2.9 หน้าจอการสร้างกฎด้วยภาษาเอสดับบิวอาร์แอลด้วยโปรแกรมโปรเทเจ 31
ภาพที่ 2.10 การวัดค่าความคล้ายคลึงเชิงมุมโคไซน์ 32
ภาพที่ 2.11 ตัวอย่างความคล้ายคลึงของสองเอกสาร 36
ภาพที่ 4.1 ตัวอย่างความสัมพันธ์ของคำในคลังคำตามตัวแบบเอสเคโอเอส 48
ภาพที่ 4.2 ลำดับชั้นของคลาสหลักและคลาสย่อยของออนโทโลยีผลไม้เศรษฐกิจไทย 49
ภาพที่ 4.3 ตัวอย่างการระบุรายละเอียดข้อมูลในโปรแกรมโปรเทเจ 54
ภาพที่ 4.4 ขั้นตอนการประมวลผลคำสำคัญ 55
ภาพที่ 4.5 ขั้นตอนการประมวลผลคำค้น 58
ภาพที่ 4.6 ขั้นตอนการประมวลผลลัพธ์ 62
ภาพที่ 4.7 ตัวอย่างผลลัพธ์ที่ได้จากการค้นหาด้วยคำค้นจากผู้ใช้ “มังคุด ปลูกระยะ ดิน” 64
ภาพที่ 4.8 ตัวอย่างหน้าจอการค้นหาเชิงความหมายสำหรับผลไม้เศรษฐกิจไทย 66

มหาวิทยาลัยเทคโนโลยีสุรนารี

บทที่ 1 บทนำ

1.1 ความสำคัญและที่มาของปัญหาการวิจัย

ประเทศไทยตั้งอยู่บนพื้นฐานความเป็นสังคมเกษตร ผลผลิตและรายได้จากการเกษตรมีความสำคัญต่อเศรษฐกิจของประเทศไทยเป็นอย่างมาก ดังจะเห็นได้จากมูลค่าการส่งออกสินค้าทั้งหมดในปี 2563 จำนวน 7,173,166 ล้านบาท เป็นสินค้าเกษตรและผลิตภัณฑ์มีมูลค่ามากถึง 1,288,818 ล้านบาท หรือคิดเป็นร้อยละ 17.97 ของมูลค่าการส่งออกสินค้าทั้งหมดของประเทศ และพบว่า ผลไม้และผลิตภัณฑ์เป็นสินค้าที่มีมูลค่าการส่งออกมากที่สุดในกลุ่มสินค้าเกษตรและผลิตภัณฑ์ ซึ่งมีมูลค่ามากถึง 182,351 ล้านบาท หรือคิดเป็นร้อยละ 14.15 ของมูลค่าสินค้าเกษตรและผลิตภัณฑ์ สำหรับสินค้าในกลุ่มผลไม้และผลิตภัณฑ์ที่มีการส่งออกในปริมาณที่มาก ได้แก่ ทุเรียนและผลิตภัณฑ์ มีมูลค่าสินค้า จำนวน 72,459 ล้านบาท ลำไยและผลิตภัณฑ์มีมูลค่าสินค้า จำนวน 24,709 ล้านบาท (สำนักงานเศรษฐกิจการเกษตร, 2564) จะเห็นได้ว่าการเกษตรของไทยได้มีการกระจายการผลิตไปสู่พืชและผลไม้เศรษฐกิจใหม่ ๆ เพิ่มขึ้นหลายชนิด เช่น ข้าว ทุเรียน ลำไย และมังคุด เป็นต้น ซึ่งในแต่ละปีมีสินค้าทางการเกษตรที่ส่งออกจำหน่ายไปต่างประเทศเป็นจำนวนมาก สร้างรายได้ให้กับประเทศในปริมาณที่มากเช่นเดียวกัน

การพัฒนาด้านการเกษตรของไทยในระยะที่ผ่านมามีผลผลิตด้านการเกษตรขยายตัวในอัตราที่สูง นอกจากนั้นการเติบโตของสาขาเกษตรเป็นปัจจัยสำคัญต่อการเพิ่มพูนฐานะและรายได้ให้แก่เกษตรกรและบุคคลที่เกี่ยวข้อง และมีส่วนสำคัญต่อการแก้ปัญหาความยากจนของประชากรทั้งประเทศ แต่ในปัจจุบันความเหลื่อมล้ำในฐานะรายได้ระหว่างภาคเกษตรและภาคเศรษฐกิจอื่น ๆ และระหว่างภูมิภาคส่วนต่าง ๆ ของประเทศยังมีปัญหาอยู่อีกมากโดยเฉพาะการเข้าถึงการศึกษาหรือความรู้เฉพาะทางด้านการเกษตรโดยมีเทคโนโลยีสารสนเทศมาเป็นปัจจัยสำคัญที่จะมาช่วยสำหรับการเพิ่มผลผลิต ลดต้นทุน หรือการแก้ไขปัญหาจากการผลิตให้มีความสอดคล้องกับความต้องการต่อเศรษฐกิจการเกษตรของประเทศที่มีการเจริญเติบโตอย่างต่อเนื่อง ฉะนั้นผู้ผลิตหรือเกษตรกรจำเป็นอย่างยิ่งที่จะต้องมีความรู้เพื่อที่จะใช้ความรู้ในการแก้ไขปัญหาได้ทันท่วงที ถึงแม้ในปัจจุบันจะมีระบบอินเทอร์เน็ตที่ใช้โปรแกรมค้นหา ดังเช่น กูเกิ้ล ที่ให้บริการการค้นหาเอกสารหรือความรู้ที่มีความสะดวกและรวดเร็ว แต่เนื่องจากโปรแกรมค้นหาที่มีอยู่ในปัจจุบันไม่ได้ออกแบบมาเพื่อรองรับการค้นหาเอกสารหรือความรู้ทางการเกษตรเป็นการเฉพาะเจาะจง และระบบการค้นหายังเป็นการสืบค้นจากคำสำคัญเป็นหลักโดยไม่ได้พิจารณาจากความหมายของคำสำคัญ จึงทำให้ผลลัพธ์ที่ได้ไม่ตรงกับความต้องการ หรือยังมีเอกสารที่เกี่ยวข้องอยู่ในฐานข้อมูลหรือฐานความรู้ที่ไม่ได้ถูกค้นหาดึงมาใช้ประโยชน์

จากความต้องการของเกษตรกรและผู้เกี่ยวข้องที่ต้องการเข้าถึงความรู้ทางการเกษตรที่อยู่กระจัดกระจายบนอินเทอร์เน็ตโดยใช้บริการของโปรแกรมค้นหา แต่ปัญหาสำคัญของการใช้โปรแกรมค้นหาในปัจจุบันได้แก่ จำนวนเอกสารความรู้ที่ค้นคืนมาได้มีปริมาณมากเกินไปจนกระทั่งไม่สามารถจะเข้าถึงได้ทั้งหมด การค้นหาแบบใช้คำหลัก (Keywords) ตามที่ผู้ใช้ระบุทำให้เอกสารที่เกี่ยวข้องอีกจำนวนมากที่อยู่ในฐานความรู้อื่น ไม่ได้ถูกดึงขึ้นมาเนื่องจากไม่มีคำค้นตามที่ผู้ใช้ระบุ ซึ่งเป็นปัญหาของคำที่เขียนต่างกันแต่มีความหมายเดียวกัน นอกจากนี้การจำแนก การจัดทำดัชนี การกำหนดความสัมพันธ์ระหว่างความรู้ต่าง ๆ มีความสำคัญต่อการค้นหาเพื่อให้มีความแม่นยำ ตรงตามความต้องการของผู้ใช้มากขึ้น จึงเป็นเรื่องสำคัญที่จะต้องนำมาประยุกต์ใช้สำหรับการจัดการเอกสารความรู้ทางการเกษตรที่อยู่กระจัดกระจายบนเครือข่ายอินเตอร์เน็ต

ระบบค้นคืนสารสนเทศ (Information Retrieval System: IR) มีวัตถุประสงค์หลักเพื่อค้นคืนเอกสารหรือสารสนเทศตามคำสำคัญจากข้อคำถามของผู้ใช้ โดยจะมีหน้าที่หลักในการประมวลผลเอกสาร สร้างตัวแทนเอกสารหรือดัชนี ประมวลผลคำสำคัญ สร้างตัวแทนคำสำคัญ และสืบค้นเอกสารหรือความรู้โดยการเปรียบเทียบความเหมือนของตัวแทนคำสำคัญกับเอกสาร จากนั้นจะให้ผลลัพธ์ ซึ่งได้แก่ รายการของเอกสารเรียงตามลำดับความเหมือนของเอกสารกับคำค้น (Salton, 1989) ดังนั้นกระบวนการพื้นฐานของการค้นคืนสารสนเทศประกอบด้วยการทำดัชนีข้อมูล การเปรียบเทียบ และการเรียงลำดับ

ในการทำดัชนีข้อมูลสามารถดำเนินการได้โดยการประยุกต์ใช้วิธีการให้หมายเหตุ (Annotation) หรือ การให้ป้ายระบุ (Tagging) ซึ่งเป็นการให้ชื่อ คุณลักษณะ ความเห็น คำอธิบาย และอื่น ๆ กับเอกสารหรือข้อความที่ถูกเลือกขึ้นมา ซึ่งเป็นการเพิ่มสารสนเทศสำหรับบางส่วนของข้อมูลที่มีอยู่ในเอกสารเหล่านั้น ในปัจจุบันมีการพัฒนาการให้หมายเหตุเชิงความหมาย (Semantic annotation) ซึ่งจะนำข้อมูลที่ไม่มีโครงสร้างหรือกึ่งโครงสร้างที่อยู่ในเนื้อหานั้นไปเชื่อมโยงกับความรู้ที่มีโครงสร้างภายใต้โดเมนหนึ่ง ๆ นอกจากนี้แล้ว การให้หมายเหตุเชิงความหมาย สามารถแสดงผลลัพธ์ที่ได้จากการค้นหาโดยไม่ต้องมีการเชื่อมต่อที่ชัดเจนกับการค้นหาเริ่มต้นได้ ดังนั้น ถ้าการให้ป้ายระบุมีวัตถุประสงค์เพื่อช่วยในการค้นหาผลลัพธ์ที่เกี่ยวข้องทั้งหมด การให้หมายเหตุเชิงความหมายสามารถช่วยเพิ่มความแตกต่างและหลากหลายของกระบวนการเหล่านั้นได้ (Ontotext, 2013) อีกทั้งการให้หมายเหตุเชิงความหมายยังทำหน้าที่คล้ายเป็นสะพานเชื่อมต่อความกำกวมในการอธิบายสิ่งต่าง ๆ ของภาษาธรรมชาติ กับการแสดงค่าทางคอมพิวเตอร์ในภาษาที่มีรูปแบบเป็นทางการ (formal language) โดยการให้หมายเหตุเชิงความหมายจะแสดงให้คอมพิวเตอร์เข้าใจว่าข้อมูลจะเชื่อมต่อกันอย่างไรและความสัมพันธ์จะถูกประเมินแบบอัตโนมัติได้อย่างไร ซึ่งหลักการดังกล่าวอาจสามารถช่วยในการกรองข้อมูลที่ซับซ้อนและทำการค้นหาข้อมูลที่ต้องการได้ (Ontotext, 2013)

การเปรียบเทียบเอกสาร เป็นการวัดความคล้ายคลึงระหว่างเอกสาร โดยมีแนวคิดของการนำความรู้พื้นฐานเกี่ยวกับความสัมพันธ์ระหว่างคำในแง่ความหมายมาใช้ ซึ่งคำแต่ละคำจะมีความหมายใกล้เคียงกับคำอื่น ๆ ไม่เท่ากัน ดังนั้น เอกสารที่คล้ายคลึงกันจึงควรมีคำที่มีเอกสารความหมายใกล้เคียงกันปรากฏอยู่ โดยสามารถใช้วิธีการวัดได้หลากวิธี ดังเช่น วิธีระยะทางที่สั้นที่สุดระหว่างคำจากการนำเสนองานวิจัยของพิลาวัณย์ พลับรู้การ และกฤษณะ ไวยมัย (2544) หรืออัลกอริทึมที่คำนวนเส้นทางตามโครงสร้างที่สั้นที่สุด (Shortest-Path Algorithm) โดยจะค้นหาชุดของเส้นเชื่อม (Edge) ระหว่างสองโหนดหรือสองคำค้นที่มีผลรวมน้ำหนักน้อยที่สุด (McConnell, 2001)

การเรียงลำดับจะมีความสำคัญเป็นอย่างยิ่งในกรณีที่มีการสืบค้นความรู้หรือเอกสารในปริมาณมาก ๆ เพื่อให้ผู้ใช้สามารถเข้าถึงความรู้หรือเอกสารได้ตรงตามความต้องการมากที่สุด ในการเรียงลำดับสามารถใช้วิธีการคำนวณที่หลากหลาย ดังเช่น การวัดความคล้ายคลึงเชิงมุมโคไซน์ (Cosine Similarity) แบบแจ็คการ์ด (Jaccard) หรือแบบฟัซซี่ (Fuzzy)

ดังนั้น ในการวิจัยนี้ จึงนำหลักการดังที่กล่าวมาข้างต้นมาประยุกต์ใช้ในการพัฒนาตัวแบบการค้นคืนความรู้ โดยจะใช้ฐานความรู้ผลไม้เศรษฐกิจไทยเป็นฐานความรู้ตัวอย่าง ซึ่งในกระบวนการวิจัยจะมุ่งเน้นที่การค้นคืนความรู้เชิงความหมาย โดยที่ผู้ใช้สามารถป้อนข้อมูลที่ต้องการค้นหา แล้วระบบ จะทำการวิเคราะห์ คัดกรอง และจัดประเภทของข้อมูล เพื่อให้ผู้ใช้สามารถได้รับข้อมูลที่ตรง หรือใกล้เคียงกับความต้องการของผู้ใช้มากขึ้นได้ ซึ่งตัวแบบที่นำเสนอนี้มุ่งเน้นจะหาวิธีการที่จะทำให้การค้นคืนความรู้มีความตรงต่อความต้องการของผู้ใช้

1.2 วัตถุประสงค์ของการวิจัย

1.2.1 เพื่อออกแบบและพัฒนาการจัดทำดัชนีเชิงความหมายสำหรับผลไม้เศรษฐกิจไทย

1.2.2 เพื่อออกแบบ พัฒนา และประเมินตัวแบบค้นคืนเชิงความหมายสำหรับผลไม้เศรษฐกิจไทย

มหาวิทยาลัยเทคโนโลยีสุรนารี

1.3 สมมติฐานการวิจัย

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

1.4 ขอบเขตของการวิจัย

ในการวิจัยนี้จะทำการออกแบบฐานความรู้โดยใช้ตัวแบบออนโทโลยี จัดทำดัชนีเชิงความหมายจากเอกสารทางวิชาการ แล้วนำมาจัดเก็บในฐานความรู้ไม่น้อยกว่า 500 รายการ จากนั้นจะทำการ

ออกแบบ พัฒนา ทดสอบ และประเมินตัวแบบการค้นคืนความรู้เชิงความหมายสำหรับผลไม้เศรษฐกิจไทย

สำหรับกลุ่มเป้าหมายในการเข้าถึงระบบนี้ ได้แก่ เกษตรกรไทยผู้ทำการผลิตผลไม้เศรษฐกิจไทย ชุมชนนักวิชาการเกษตร รวมถึงประชาชนผู้สนใจทั่วไป ที่สามารถเข้าถึงระบบเครือข่ายอินเทอร์เน็ต และต้องการศึกษาและนำความรู้เกี่ยวกับการปลูก ดูแลรักษา การป้องกันโรคและศัตรูพืช และเทคโนโลยีการผลิตผลไม้เศรษฐกิจไทยไปใช้ให้เกิดประโยชน์ต่อการผลิตผลไม้เศรษฐกิจไทยอย่างยั่งยืน

1.5 ข้อตกลงเบื้องต้น

ในงานวิจัยนี้ผู้วิจัยได้กำหนดข้อตกลงเบื้องต้นของการวิจัย ดังนี้

1.5.1 ความรู้ผลไม้เศรษฐกิจที่นำมาเป็นตัวอย่างจะค้นหาจากระบบเครือข่ายอินเทอร์เน็ตด้วยโปรแกรมค้นหา จากเอกสารความรู้ทางวิชาการที่เผยแพร่จากแหล่งต่าง ๆ แล้วนำมาวิเคราะห์และเลือกมาเฉพาะข้อมูลที่เกี่ยวข้องกับเรื่องที่ต้องการจัดเก้บในฐานความรู้ ความรู้ที่จัดเก็บจะต้องเป็นความรู้ที่สามารถเผยแพร่ได้โดยไม่มีลิขสิทธิ์เท่านั้น

1.5.2 ฐานความรู้ผลไม้เศรษฐกิจที่พัฒนาขึ้น จะจัดเก็บความรู้ในผลไม้ 3 ชนิด ได้แก่ ทุเรียน ลำไย และมังคุด ซึ่งเป็นผลไม้ที่ส่งออกและมีมูลค่าทางเศรษฐกิจสูง

1.5.3 ตัวแบบค้นคืนเชิงความหมายสำหรับผลไม้เศรษฐกิจไทยที่พัฒนาขึ้นจะจัดทำส่วนต่อประสานผู้ใช้เฉพาะส่วนการค้นคืนความรู้ในฐานความรู้ที่พัฒนาขึ้นมาเท่านั้น คำสำคัญที่ใช้ในการค้นหาจะต้องพิมพ์ถูกต้อง และมีการเว้นวรรคระหว่างคำอย่างถูกต้อง สำหรับผลลัพธ์ที่แสดงผ่านบราวเซอร์ประกอบด้วย รายละเอียดความรู้ และการจัดลำดับแสดงรายการผลลัพธ์

1.6 ประโยชน์ที่ได้รับจากการวิจัย

1.6.1 เพื่อเป็นองค์ความรู้ในการวิจัยเกี่ยวกับการค้นคืนเชิงความหมายสำหรับผลไม้เศรษฐกิจไทย

1.6.2 เพื่อเป็นประโยชน์ต่อประชาชน ชุมชน ที่สามารถใช้ศึกษาเกี่ยวกับผลไม้เศรษฐกิจไทยที่ต้องการได้สะดวก รวดเร็ว และตรงต่อความต้องการมากขึ้น

1.6.3 เพื่อเป็นประโยชน์ต่อประชาชน ชุมชน ที่จะนำความรู้ที่ได้จากระบบนี้ นำไปประยุกต์ใช้ให้เข้ากับสถานการณ์ในขณะนั้น ได้สะดวกและรวดเร็วขึ้น

มหาวิทยาลัยเทคโนโลยีสุรนารี

1.7 คำอธิบายศัพท์

1.7.1 ตัวแบบค้นคืนเชิงความหมายสำหรับผลไม้เศรษฐกิจไทย (Semantic retrieval model for Thai economic fruits) หมายถึง ตัวแบบที่สำหรับให้ผู้ใช้ได้เข้าถึงความรู้ที่จัดเก็บอยู่ในฐานความรู้ที่มีการสร้างความสัมพันธ์ของเนื้อหาต่าง ๆ เพื่อช่วยให้การค้นคืนได้ถูกต้องและตรงตามความต้องการ

1.7.2 ผลไม้เศรษฐกิจไทย (Thai economic fruits) หมายถึง ทุเรียน ลำไย และมังคุด

มหาวิทยาลัยเทคโนโลยีสุรนารี

บทที่ 2 ปริทัศน์วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

การวิจัยในครั้งนี้เน้นการออกแบบ พัฒนา ทดสอบ และประเมินตัวแบบการค้นคืนเชิงความหมายสำหรับผลไม้เศรษฐกิจไทย โดยผู้วิจัยได้ศึกษาเอกสาร และงานวิจัยที่เกี่ยวข้องตามลำดับดังนี้

2.1 การค้นคืนเชิงความหมาย

2.2 โปรแกรมค้นหาเชิงความหมาย

2.3 ฐานความรู้ด้วยตัวแบบออนโทโลยี

2.4 การจัดทำดัชนี

2.5 การเปรียบเทียบความคล้ายคลึง

2.6 การเรียงลำดับ

2.7 ผลไม้เศรษฐกิจไทย

2.8 เอกสารและงานวิจัยที่เกี่ยวข้อง

2.9 กรอบแนวคิดการวิจัย

2.1 การค้นคืนเชิงความหมาย (Semantic Retrieval)

การค้นคืนเชิงความหมาย (Sematic Retrieval) หากพิจารณาจากที่มาของคำแล้ว จะมาจาก 2 คำ ได้แก่ “การค้นคืน (Retrieval)” และ “การให้ความหมาย (Semantic)” โดยความหมายของการค้นคืนที่มีการอธิบายไว้ จะเป็นกระบวนการที่เกี่ยวข้องกับการรวบรวม ค้นหา และดึงข้อมูลจากแหล่งที่เกี่ยวข้อง เพื่อให้ผู้ใช้ได้รับข้อมูลที่ตรงกับความต้องการของผู้ใช้ได้อย่างถูกต้อง รวดเร็ว และทันเวลา (Song et al, 2011) ดังนั้น ความหมายของการค้นคืนเชิงความหมาย จึงหมายถึงกระบวนการในการรวบรวม ค้นหา และดึงข้อมูลจากแหล่งที่เกี่ยวข้องตามความหมายที่ผู้ใช้ต้องการเพื่อให้ได้รับข้อมูลที่ตรงกับความต้องการของผู้ใช้ให้มากที่สุด

ระบบค้นคืนเชิงความหมาย (Semantic Retrieval System) เป็นกระบวนการในการค้นคืนข้อมูลตามความหมายที่ผู้ใช้ต้องการ ซึ่งประกอบด้วย 6 ส่วนประกอบที่สำคัญ ดังต่อไปนี้ (Storey et all, 2003)

1) ส่วนต่อประสานผู้ใช้ (Interface)

ส่วนต่อประสานผู้ใช้ (Interface) เป็นส่วนในการรับคำค้นจากผู้ใช้แล้วนำไปประมวลผลเพื่อค้นหาสิ่งที่ต้องการให้กับผู้ใช้ต่อไป คำค้นที่ผู้ใช้กำหนดอยู่ได้ในหลากหลายรูปแบบ เช่น คำค้น วลีสำคัญ ข้อความ เป็นต้น

2) เครื่องอนุมาน (Inference engine)

เครื่องอนุมาน (Interface engine) จะนำคำค้นจากผู้ใช้มาสกัดหาคำที่เกี่ยวข้องเพิ่มเติม โดยอาจมีการใช้แหล่งความรู้ระยะไกลมาช่วยในการสกัดหรือกำหนดคำค้นที่เกี่ยวข้อง เมื่อได้กลุ่มคำค้นที่เกี่ยวข้องกับที่ผู้ใช้ต้องการค้นหาแล้ว เครื่องอนุมานจะนำคำค้นเหล่านั้นไปค้นหาในฐานความรู้ต่อไป

3) แหล่งความรู้ระยะไกล (Remote knowledge sources)

แหล่งความรู้ระยะไกล (Remote knowledge soruces) เป็นแหล่งความรู้ที่อยู่ในหลากหลายรูปแบบบนเครือข่ายอินเทอร์เน็ต แหล่งความรู้ระยะไกลดังกล่าวนี้จะมีการจัดเก็บข้อมูลหรือความรู้ต่าง ๆ แล้วมีการจัดระบบความรู้เหล่านั้นเพื่อแสดงถึงความสัมพันธ์ ความเชื่อมโยงความหมาย คำเหมือน คำตรงข้าม และข้อมูลต่าง ๆ เพื่อให้สะดวกต่อการค้นหาและการนำข้อมูลหรือความรู้เหล่านั้นไปใช้ในการค้นหาได้อย่างเหมาะสมและถูกต้อง ตัวอย่างของแหล่งความรู้ระยะไกลดังเช่น WordNet, DARPA Ontology Library เป็นต้น

4) ฐานความรู้ (Local knowledge base)

เป็นการรวบรวมความรู้ต่าง ๆ ไว้ภายในหน่วยงาน หรือองค์กร ซึ่งเป็นความรู้ที่จำเป็นสำหรับองค์กรในการนำไปใช้ประโยชน์ต่อไป โดยในการจัดเก็บความรู้ดังกล่าวนี้ จะมีการออกแบบโครงสร้างของความรู้สำหรับการจัดเก็บ และมีการเชื่อมโยงความสัมพันธ์ของความรู้ต่าง ๆ เพื่อช่วยในการค้นคืนความรู้ได้ตรงกับความต้องการของผู้ใช้ต่อไป

5) การสร้างคำค้น (Query constructor)

เป็นกระบวนการในการสร้างคำค้นตามรูปแบบของภาษาหรือตัวแบบที่ใช้ เพื่อที่จะในการค้นหาความรู้ในฐานความรู้ ซึ่งอาจมีการสร้างเป็นคำค้นเดี่ยว ๆ หรืออาจเป็นกลุ่มคำค้นที่ได้มาจากการขยายคำค้นที่ผู้ใช้ระบุมาก็ได้

6) โปรแกรมค้นหา (Search engine)

โปรแกรมค้นหาจะทำหน้าที่นำคำค้นที่ได้จากการสร้างคำค้นไปค้นหาความรู้จากฐานความรู้ หรือแหล่งความรู้ระยะไกล และเมื่อได้ผลลัพธ์จากการค้นหามาแล้ว จะนำมาแสดงให้กับผู้ใช้ที่ต้องการค้นหา โปรแกรมค้นหาสามารถจัดรูปแบบการแสดงผลลัพธ์ รวมถึงการเรียงลำดับของผลลัพธ์ในรูปแบบต่าง ๆ ได้

ในการค้นหาจะมีการใช้โปรแกรมค้นหา (Search Engine) เช่น Google, Yahoo Bing เป็นต้น เป็นโปรแกรมที่ออกแบบเพื่อการค้นหาแหล่งจัดเก็บเอกสารออนไลน์ โดยให้ผลลัพธ์เป็นเอกสารอิเล็กทรอนิกส์แสดงผลในหน้าเว็บเพจที่ตรงความต้องการตามคำสำคัญ (Keyword) ที่ผู้ใช้ใช้ในการสืบค้นเอกสาร แบ่งเป็นรูปแบบที่ยึดเอกสารเป็นศูนย์กลาง (Document-centric model) อาจไม่เหมาะสำหรับงานสารสนเทศที่มีความซับซ้อนมาก ๆ ซึ่งต้องการสารสนเทศเชิงโครงสร้างที่บูรณาการจากแหล่งที่หลากหลาย จึงพัฒนามาเป็นรูปแบบที่ยึดเอนทิตี้เป็นศูนย์กลาง (Entity-centric

model) จะบรรยายเอนทิตี้ที่เป็นบุคคล สถานที่ หรือรายละเอียดข่าวสารต่าง ๆ มากกว่าเอกสาร (Harth, Hogan, Umbrich, & Decker, 2008) โปรแกรมค้นหาจะได้อธิบายรายละเอียดในหัวข้อ 2.2 ต่อไป

การค้นคืนเชิงความหมายในงานวิจัยนี้ เป็นกระบวนการในการรวบรวม และค้นหาข้อมูลจากฐานความรู้ผลไม้เศรษฐกิจไทยที่จัดทำขึ้น เพื่อดึงความรู้ที่เกี่ยวข้องตามความหมายที่ผู้ใช้ต้องการ โดยกระบวนการในการดำเนินนการจะเริ่มจากการสร้างส่วนต่อประสานผู้ใช้เพื่อรับข้อมูลคำค้นจากผู้ใช้ แล้วนำมาผ่านกระบวนการอนุมานเพื่อให้ได้กลุ่มคำค้นที่จะใช้นำไปค้นหา โดยมีการเชื่อมต่อกับคลังคำที่ได้พัฒนาขึ้นมา แล้วนำคำค้นที่ได้ไปดึงข้อมูลจากฐานความรู้เพื่อมาแสดงผลให้กับผู้ใช้ต่อไป และมีการแสดงผลตามความคล้ายคลึงของคำค้นจากผู้ใช้ ตัวแบบที่พัฒนาขึ้นจะทำหน้าที่เป็นเสมือนหนึ่งโปรแกรมค้นหาเชิงความหมาย

2.2 โปรแกรมค้นหาเชิงความหมาย

2.2.1 ความหมายและคุณลักษณะของโปรแกรมค้นหาเชิงความหมาย

สำหรับคำศัพท์ที่เกี่ยวข้องกับคำว่า “โปรแกรมค้นหาเชิงความหมาย” จาก การสำรวจวรรณกรรมที่เกี่ยวข้องในเรื่องนี้ พบว่า มีคำศัพท์จำนวนสองคำที่ใช้ในความหมายของโปรแกรมค้นหาเชิงความหมาย ได้แก่ คำว่า โปรแกรมค้นหาเชิงความหมาย (Semantic search engines) หรือ เอสเอสอี (SSE) และคำว่า โปรแกรมค้นหาเชิงความหมาย (Semantic web search engines) หรือ เอสดับเบิ้ลยูเอสอี (SWSE) ซึ่งมีนักวิชาการและองค์กรที่เกี่ยวข้องได้อธิบายความหมายไว้ดังนี้

- World Wide Web Consortium หรือ W3C (2010) อธิบายว่า หมายถึง แอปพลิเคชันสำหรับการค้นหาออนโทโลยี (Ontologies) จากการสืบค้นโดยใช้คำสำคัญที่เป็นภาษาธรรมชาติและการจัดอันดับผลลัพธ์ที่ค้นได้

- d'Aquin, Ding & Motta (2011) ให้คำจำกัดความว่า หมายถึง ระบบหนึ่ง ๆ ที่รวบรวม จัดทำดรรชนี และวิเคราะห์เอกสารเว็บเชิงความหมาย (Semantic web documents) ที่จัดเตรียมการสืบค้นและกลไกการสอบถาม ซึ่งเอกสารเว็บเชิงความหมายเป็นเอกสารที่ประกอบด้วยสารสนเทศที่เข้ารหัสโดยใช้มาตรฐานภาษาเว็บเชิงความหมาย (Standard semantic web languages) เช่น อาร์ดีเอฟ (RDF) อาร์ดีเอฟเอส (RDFS) และโอดับเบิ้ล ยูแอล (OWL) เป็นต้น

- Shah, Shah, & Deulkar (2015) อธิบายว่า หมายถึง เว็บเชิงความหมายหนึ่ง ๆ ที่สามารถตอบคำถามต่าง ๆ ตามการสื่อความหมายของข้อคำถามของผู้ใช้ทรัพยากรในคลังสารสนเทศและความหมายเฉพาะหรือรูปแบบความรู้

- Maitaouthong (2019) ให้ความหมายของโปรแกรมค้นหาเชิงความหมายว่าเป็นระบบที่ทำหน้าที่รวบรวม ทำดรรชนีและวิเคราะห์เอกสารเว็บเชิงความหมาย เพื่อการสืบค้นและ

การตอบคำถามที่สื่อความหมายตรงตามความต้องการสารสนเทศของผู้ใช้ ซึ่งเอกสารเว็บเชิงความหมายเป็นเอกสารที่เข้ารหัสโดยใช้มาตรฐานภาษาเว็บเชิงความหมาย เช่น อาร์ดีเอฟ (RDF) อาร์ดีเอฟเอส (RDFS) และโอดับเบิ้ลยูแอล (OWL) เป็นต้น

จากความหมายของโปรแกรมค้นหาเชิงความหมายข้างต้น ในงานวิจัยนี้จะดำเนินการตามความหมายที่เน้นถึงการทำเป็นระบบที่ทำหน้าที่รวบรวม ทำดรรชนี และวิเคราะห์ข้อมูลที่รวบรวมไว้ เพื่อให้การสืบค้นข้อมูลได้ตรงตามความหมายและความต้องการของผู้ใช้มากที่สุด

การค้นหาข้อมูลบนเว็บในระยะเริ่มแรกจะใช้โปรแกรมค้นหาคำสำคัญ (Keyword search engine) ซึ่งโปรแกรมดังกล่าวอาจมีข้อจำกัดค่อนข้างมาก ไม่ว่าจะเป็นผลการค้นที่ได้มีจำนวนมากและมักไม่ตรงความต้องการของผู้ใช้ การจัดเก็บข้อมูลบนเว็บเป็นฐานข้อมูลสากลที่ขาดโครงสร้างเชิงความหมาย และเป็นเรื่องยากสำหรับเครื่องจักรที่จะเข้าใจสารสนเทศที่ผู้ใช้ใช้สืบค้นในรูปแบบที่เป็นอักขระ ผลการค้นหาจึงให้ผลลัพธ์ที่คลุมเครือ (Ambiguous results) ดังนั้น จึงมีการพัฒนาเว็บเชิงความหมาย (Semantic web) ขึ้นมาแก้ปัญหาข้อจำกัดของเว็บในปัจจุบัน (Madhu, Govardhan, & Rajinikanth, 2011) ที่ทำให้มีการเชื่อมโยงข้อมูลของเว็บผู้พัฒนาและเว็บของแหล่งข้อมูลอื่นที่สัมพันธ์กัน ทำให้เกิดระบบสืบค้นที่มีประสิทธิภาพสามารถสืบค้นข้อมูลได้อย่างรวดเร็วและตรงประเด็น ภายใต้ความสัมพันธ์ของคำที่มีความหมายต่อกันและสามารถเชื่อมโยงไปยังข้อมูลที่ต้องการอย่างแท้จริง ส่งผลให้เกิดนวัตกรรมการสืบค้นข้อมูลผ่านฐานข้อมูลขนาดใหญ่ที่มีการเชื่อมโยงความสัมพันธ์ของข้อมูล และได้มีการพัฒนาเครื่องมือสำหรับการสืบค้นเว็บเชิงความหมายที่เรียกว่าโปรแกรมค้นหาเชิงความหมาย (Semantic search engines) สำหรับการสืบค้นเอกสารเว็บเชิงความหมายให้ได้ผลการค้นที่มีความแม่นยำและตรงความต้องการของผู้ใช้มากที่สุด

อย่างไรก็ตาม โปรแกรมค้นหาคำสำคัญและโปรแกรมค้นหาเชิงความหมายมีความแตกต่างกันในหลายประเด็น จากการศึกษาของ Malve & Chawan (2015) ที่ศึกษาความแตกต่างระหว่าง โปรแกรมค้นหาคำสำคัญกับโปรแกรมค้นหาเชิงความหมาย เมื่อนำผลการศึกษาดังกล่าว มาวิเคราะห์ จึงสามารถสรุปได้เป็น 3 ประเด็น คือ ลักษณะของโปรแกรม คำที่ใช้ค้น และผลการค้นหา ดังแสดงในตารางที่ 2.1

ตารางที่ 2.1 ความแตกต่างระหว่างโปรแกรมค้นหาคำสำคัญและโปรแกรมค้นหาเชิงความหมาย

ประเด็น	โปรแกรมค้นหาคำสำคัญ	โปรแกรมค้นหาเชิงความหมาย
1) ลักษณะของโปรแกรม	• เป็นโปรแกรมค้นหาแบบเดิม (Traditional search engines) ที่ให้ผลการค้นในบริบทที่เกี่ยวข้องกับข้อคำถาม	• เป็นโปรแกรมค้นหาที่ทำงานด้วยวิธีการเชิงความหมาย (Semantic based approach) ให้สารสนเทศที่ถูกต้อง แม่นยำและตรงความต้องการที่เกี่ยวข้องกับข้อคำถาม

ตารางที่ 2.1 ความแตกต่างระหว่างโปรแกรมค้นหาคำสำคัญและโปรแกรมค้นหาเชิงความหมาย

ประเด็น	โปรแกรมค้นหาคำสำคัญ	โปรแกรมค้นหาเชิงความหมาย
	• ใช้คำสำคัญ (Keywords) เพื่อขยายข้อคำถาม • ใช้ภาษาเอชทีเอ็มแอล (HTML) และ ภาษาเอกซ์เอ็มแอล (XML) เพื่อสร้างเมทาดาทา	• ใช้ออนโทโลยี (Ontology) ที่เชื่อมโยงความสัมพันธ์ระหว่างคำสำคัญต่าง ๆ • ใช้ภาษาเว็บเชิงความหมาย (Semantic web languages) และใช้อาร์ดีเอฟ (RDF) และโอดับเบิ้ลยูแอล (OWL) เพื่อสร้างเมทาดาทา
2) คำที่ใช้ค้น	• ไม่เน้นคำหยุด (Stop words) เนื่องจากไม่ให้ผลการค้นที่ถูกต้องแม่นยำตามที่ผู้ใช้ต้องการ (คำหยุด เช่น บุพบท สันธาน) • ไม่เน้นคำหรือวลีที่ใช้ตอบคำถามให้ ได้ผลการค้นที่ถูกต้องแม่นยำ	• เน้นคำหยุดและเครื่องหมายวรรคตอน เนื่องจากโปรแกรมค้นหาจะคำนวณอักขระทุกตัวที่ส่งผลต่อผลการสืบค้น • เน้นข้อความหรือคำที่ให้คำตอบตรงตามข้อคำถามของผู้ใช้
3) ผลการค้น	• สารสนเทศที่ค้นได้ขึ้นอยู่กับคำสำคัญและวิธีการจัดอันดับความสำคัญเว็บเพจ (Page rank algorithm) ซึ่งให้ผลการค้นที่เป็นข้อความโฆษณาที่ไม่พึงประสงค์ (Spam results) • แสดงเว็บเพจทั้งหมดที่อาจจะสอดคล้องหรืออาจจะไม่สอดคล้องกับข้อคำถามของผู้ใช้ และเลือกเว็บเพจที่ตรงความต้องการจากเว็บเพจจำนวนมาก	• สารสนเทศที่ค้นได้ไม่ได้ขึ้นอยู่กับคำ สำคัญและวิธีการจัดอันดับความสำคัญ เว็บเพจ โดยให้ผลการค้นที่ถูกต้อง แม่นยำ (Exact results) มากกว่าผล การค้นที่ไม่ตรงความต้องการ • แสดงเฉพาะผลการค้นที่ตอบข้อคำถามของผู้ใช้เท่านั้น

มหาวิทยาลัยเทคโนโลยีสุรนารี

2.2.2 ประเภทของโปรแกรมค้นหาเชิงความหมาย

การจำแนกประเภทของโปรแกรมค้นหาเชิงความหมาย สามารถแบ่งได้เป็น 2 ประเภท คือ โปรแกรมค้นหาออนโทโลยี (Ontology search engines) และ โปรแกรมค้นหาเชิงความหมาย (Semantic search engines)(Madhu, Govardhan, & Rajinikanth, 2011) โดย Esmaili & Abolhassani (2006) อธิบายว่า โปรแกรมค้นหาออนโทโลยีเป็นโปรแกรมที่พยายามปรับปรุงผลจากการสืบค้น ซึ่งใช้มาตรฐานและภาษาเว็บเชิงความหมาย และการใช้สารสนเทศเชิงบริบทที่ถูกแทนด้วยออนโทโลยีและเมทาดาทา ขณะที่โปรแกรมค้นหาเชิงความหมายเป็นโปรแกรมเฉพาะที่ใช้ในการค้นหาเอกสารเว็บเชิงความหมายเท่านั้น ซึ่งเป็นเอกสารที่ถูกแทนในภาษาเฉพาะของเว็บเชิงความหมาย โดยจำแนกได้ดังนี้

1. โปรแกรมค้นหาออนโทโลยี (Ontology search engines)

1.1 โปรแกรมค้นหาข้อมูลออนโทโลยี (Ontology meta search engines) เป็นโปรแกรมค้นหาที่ใช้สืบค้นชื่อของแฟ้มข้อมูลและใช้ทางเลือกต่าง ๆ ที่เป็นประเภทของแฟ้มข้อมูล เช่น อาร์เอสเอส (RSS) อาร์ดีเอฟ (RDF) โอดับเบิ้ลยูแอล (OWL) เป็นต้น หรือสืบค้นโดยใช้แท็กเชิงความหมาย (Semantic tags) ซึ่งโปรแกรมค้นหานี้มีชุดโปรแกรม การแสดงผลที่ดีสำหรับการเรียกดูและการควบคุมทิศทางออนโทโลยีที่ค้นหา

1.2 โปรแกรมค้นหาออนโทโลยีโดยใช้การรวบรวมข้อมูลเป็นฐาน (Crawler based ontology search engines) เป็นโปรแกรมที่ใช้การรวบรวมข้อมูลเฉพาะ เพื่อการค้นหาเอกสารเว็บเชิงความหมายบนเว็บ ดรรชนีของเอกสารเว็บเชิงความหมาย และการสกัดเมทาดาทาเกี่ยวกับเอกสารเว็บเชิงความหมาย ซึ่งการใช้โปรแกรมค้นหานี้ ผู้ใช้สามารถสืบค้นคลาสหรือคุณลักษณะเฉพาะ (Special class or property) และข้อมูลตัวอย่าง ทั้งนี้โปรแกรมค้นหาสามารถใช้ค้นหาโครงสร้างกราฟของเอกสารเว็บเชิง ความหมายบนเว็บได้

2. โปรแกรมค้นหาเชิงความหมาย (Semantic search engines)

2.1 โปรแกรมค้นหาโดยใช้บริบทเป็นฐาน (Context based search engines) เป็นโปรแกรมที่มีจุดมุ่งหมายในการเพิ่มสมรรถนะของโปรแกรมค้นหาแบบเดิมโดยเฉพาะ ค่าความแม่นยำ (Precision) และค่าความระลึก (Recall) เป็นความเข้าใจบริบทของเอกสารและคำถาม ซึ่งโปรแกรมนี้มีส่วนสำคัญ คือ เครื่องมือคำอธิบาย (Annotator) ที่มีหน้าที่ในการสร้างเมทาดาทาสำหรับหน้าเว็บที่ถูกรวบรวมข้อมูล โดยจำเป็นต้องสร้างเมทาดาทา สำหรับคำถามของผู้ใช้เพื่อที่จะค้นหาบริบทของคำถามนั้นๆ

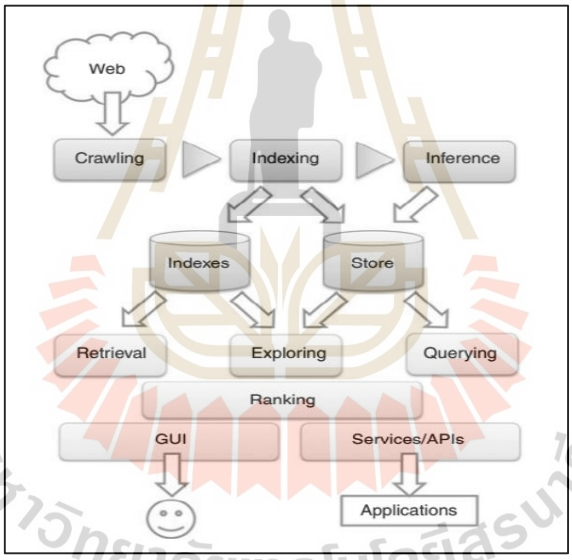
2.2 โปรแกรมค้นหาเชิงวิวัฒนาการ (Evolutionary search engines) เป็นโปรแกรมที่ค้นหาคำตอบจากปัญหาที่รู้จักกันดี ซึ่งเป็นการรวบรวมสารสนเทศในเรื่องใดเรื่องหนึ่งโดยอัตโนมัติ โปรแกรมนี้ใช้เมทาดาทาภายนอก (External metadata) ใช้โปรแกรมค้นหาปรกติและแสดง

สารสนเทศเพิ่มเติมที่ใกล้เคียงผลการค้นที่เป็นต้นฉบับ โดยโปรแกรมค้นหานี้เป็นเหมือนโปรแกรมค้นหาโดยใช้บริบทเป็นฐานที่หลากหลายสำหรับการค้นหาเว็บ

2.3 โปรแกรมค้นหาความสัมพันธ์เชิงความหมาย (Semantic association discovery engines) เป็นโปรแกรมเฉพาะของเว็บเชิงความหมายที่มีสมรรถนะในการสืบค้น โดยมีเป้าหมายคือการค้นหาความสัมพันธ์เชิงความหมายที่หลากหลายระหว่างคำค้นที่ป้อนลงไปซึ่งมักจะมี 2 คำ จากนั้นจะเป็นการจัดอันดับผลการค้นที่ขึ้นอยู่ กับการวัดระยะห่างระหว่างแนวคิดเชิงความหมาย (Semantic distance metrics) โปรแกรมค้นหานี้ทำงานได้ดีในบริบทของฐานความรู้

2.2.3 องค์ประกอบของโปรแกรมค้นหาเชิงความหมาย

สำหรับกิจกรรมโดยทั่วไปของโปรแกรมค้นหาเชิงความหมายประกอบด้วย องค์ประกอบต่าง ๆ (Components) ซึ่งโปรแกรมค้นหาแต่ละโปรแกรมอาจมีองค์ประกอบที่แตกต่างกัน โดยองค์ประกอบที่เกี่ยวข้องในกิจกรรมต่าง ๆ แสดงดังภาพที่ 2.1 (d'Aquin, Ding, & Motta, 2011)



ภาพที่ 2.1 องค์ประกอบของโปรแกรมค้นหาเชิงความหมาย

จากภาพที่ 2.1 แสดงให้เห็นถึงกิจกรรมที่เกิดขึ้นในการทำงานของ โปรแกรมค้นหาเชิงความหมาย ซึ่งเริ่มต้นจากการรวบรวมข้อมูลของเว็บเพจต่าง ๆ เพื่อนำมาจัดทำดรรชนีและสรุป/อนุมานข้อมูลที่ได้รวบรวมไว้ เมื่อผู้ใช้ทำการค้นคืน หรือสำรวจ หรือสอบถามเพื่อให้ระบบช่วยสืบค้นสารสนเทศที่ต้องการ ระบบจะดำเนินการประมวลผลข้อมูลโดยแสดงผลเว็บเพจที่ตรงความต้องการสูงสุดก่อนเป็นลำดับแรกที่หน้าจอสืบค้นของผู้ใช้

1. การรวบรวมข้อมูล (Crawling) เป็นกระบวนการสำคัญสำหรับระบบต่าง ๆ ที่มีเป้าหมายในการจัดเตรียมการเข้าถึงชุดของเอกสารเชิงความหมายที่เข้าถึงได้บนเว็บ ซึ่งเป็น

การรวบรวมเอกสารเว็บโดยใช้โปรแกรมเก็บข้อมูลจากเว็บเพจ (Crawler) ที่สามารถแสวงหาแหล่งที่แตกต่างกันของสารสนเทศเพื่อระบุที่อยู่ของเอกสารต่าง ๆ

2. การทำดรรชนี (Indexing) เป็นกระบวนการเชื่อมโยงเอกสารเชิงความหมายด้วยชุดของคำศัพท์ โดยระบบส่วนใหญ่ได้ปรับปรุงดรรชนีเพื่อการสืบค้นเอกสารฉบับเต็ม ด้วยข้อมูลเพิ่มเติม เช่น เมทาดาทาที่เกี่ยวข้องกับเอกสารแต่ละรายการหรือดรรชนีเนื้อหาของเอกสาร ซึ่งเป็นความสัมพันธ์ระหว่างเอนทิตี้ เพื่อให้การสอบถามและกลไกการค้นหา มีประสิทธิภาพ

3. การอนุมาน (Inference) เป็นกระบวนการที่ใช้ในการปรับปรุงชุดข้อมูลที่ได้รวบรวมและรวมถึงสารสนเทศที่เป็นข้อสรุป ซึ่งกลไกการทำงานที่มีเหตุผลมากอาจใช้เวลา ในการจัดทำดรรชนี ขณะที่กลไกการทำงานที่มีเหตุผลน้อยกว่าอาจต้องคำนึงถึงเวลาใน การสอบถาม

4. การค้นคืน (Retrieval) เป็นกระบวนการค้นคืนข้อมูลในระบบที่แตกต่างกัน ข้อมูลที่นำเข้าในระบบจะลำดับจากการสืบค้นด้วยคำสำคัญสู่การสอบถามที่เป็นทางการ ผลลัพธ์ที่ได้คือ ตัวระบุแหล่งทรัพยากรสากล (Uniform Resource Identifier: URI) ของเอกสารเว็บเชิงความหมาย คำศัพท์เว็บเชิงความหมาย (เช่น คลาส และ คุณสมบัติ) และ/หรือวัตถุต่าง ๆ ซึ่งผลลัพธ์อาจถูกนำเสนอด้วยเมทาดาทาที่เกี่ยวข้องเพิ่มเติมและ สามารถเรียกดูได้หลายวิธี

5. การควบคุมทิศทาง/การสำรวจ (Exploring) เป็นกระบวนการใน การจัดเตรียมผลลัพธ์ที่เรียกดูได้ โดยควบคุมทิศทางไปสู่เอกสารที่ค้นหาผ่านวัตถุที่มีการเชื่อมโยงความสัมพันธ์กันเพื่อตรวจสอบสารสนเทศในตัวเอกสารหรือเพื่อปรับปรุง ข้อคำถามผ่านกลไกการขยายคำถาม

6. การสอบถาม (Querying) เป็นกระบวนการจัดเตรียมวิธีการที่เป็นทางการในการสืบค้นคอลเล็กชั่นของเอกสารที่จัดเก็บอยู่ในระบบ เพราะลักษณะของการสืบค้นอยู่ ในรูปแบบของการสืบค้นด้วยคำสำคัญ เช่น การใช้ภาษา SPARQL ในการเข้าถึงเนื้อหา ของเอกสารได้โดยตรง เป็นต้น

7. การจัดอันดับ (Ranking) เป็นกระบวนการที่มีเป้าหมายในการอำนวยความสะดวกสำหรับการเลือกสารสนเทศที่ตรงความต้องการมากที่สุด ซึ่งความเข้าใจเกี่ยวกับ การตรงความต้องการสำหรับข้อมูลเชิงความหมายอาจเป็นความคลุมเครือและขึ้นกับ บริบท ดังนั้น ระบบที่แตกต่างกันจะใช้วิธีการที่แตกต่างกันในปัญหาของการจัดอันดับ

8. ส่วนต่อประสานการสืบค้น (Search interface) เป็นส่วนที่ระบบจัดเตรียมบริการสำหรับการเข้าถึงเมทาดาทาและผลลัพธ์การสืบค้น ซึ่งเป็นส่วนต่อประสานผู้ใช้กราฟิก (Graphical user Interface: GUI) สำหรับผู้ใช้ที่เป็นมนุษย์ โดยเทคโนโลยีที่ต่างกันอาจใช้นำเสนอส่วนต่อประสานได้มากมายและระดับของลักษณะเฉพาะโดยผ่านบริการต่าง ๆ อาจเกิดจากกลไกการสืบค้นอย่างง่ายไปสู่ส่วนต่อประสานโปรแกรมประยุกต์ (Application programming interface: API) เพื่อสำรวจและค้นหาเนื้อหาเชิงความหมายออนไลน์

2.2.4 การทำงานของโปรแกรมค้นหาเชิงความหมาย

การทำงานของโปรแกรมค้นหาเชิงความหมายจัดแบ่งรูปแบบของ การทำงานได้เป็น 4 วิธี โดยโปรแกรมค้นหาเชิงความหมายแต่ละโปรแกรมอาจใช้หนึ่งวิธี หรือใช้การผสมผสานกันมากกว่าหนึ่งวิธีในลักษณะของการทำงานที่แตกต่างกัน ซึ่งเป้าหมายของการสืบค้นเชิงความหมายคือเพื่อใช้ความหมาย (Meaning) ในการปรับปรุงประสบการณ์การสืบค้นของผู้ใช้ (Rashid, & Nisar, 2016, Sudeepthi, Anuradha, & Babu, 2012)

1. การวิเคราะห์เชิงบริบท (Contextual analysis) เป็นการวิเคราะห์บริบท ที่ช่วยให้เกิดความกระจ่างในคำถามที่สืบค้นและทำให้คำถามมีความหมายเดียว เช่น คำว่า “Goal” ในคำถามที่มีความหมายว่า “จุดมุ่งหมาย” หรือ “ประตูฟุตบอล” หรือสิ่งอื่น เป็นต้น

2. กลไกที่มีเหตุผล (Reasoning engines) เป็นชุดของข้อเท็จจริงต่าง ๆ ที่ถูกนำเสนอในระบบ โดยข้อเท็จจริงเพิ่มเติมจะได้รับการสรุปจากชุดของข้อเท็จจริง เช่น กรณี ที่ระบบรู้ว่าใครคือลูกของพอล (Paul's children) ซึ่งระบบจะรู้ว่าลูกแต่ละคนของพอลคือ ใครบ้าง ดังนั้น กลไกที่มีเหตุผลจะสามารถสรุปได้ว่าใครคือลูกคนโตของพอล เป็นต้น

3. ความเข้าใจภาษาธรรมชาติ (Natural language understanding) เป็นกลไกต่างๆ ในการประมวลผลเนื้อหาที่มีการจัดทำดรรชนี และประมวลคำถามที่ผู้ใช้ต้องการ สืบค้น โดยพยายามที่จะระบุขอบเขตของสารสนเทศ ซึ่งกลไกจะใช้ความสัมพันธ์ระหว่าง ถ้อยคำในประโยค (Syntax) และกฎเกณฑ์ต่างๆ (Rules) ในการระบุว่าเป็นบุคคล สถานที่ องค์กร และสิ่งต่าง ๆ

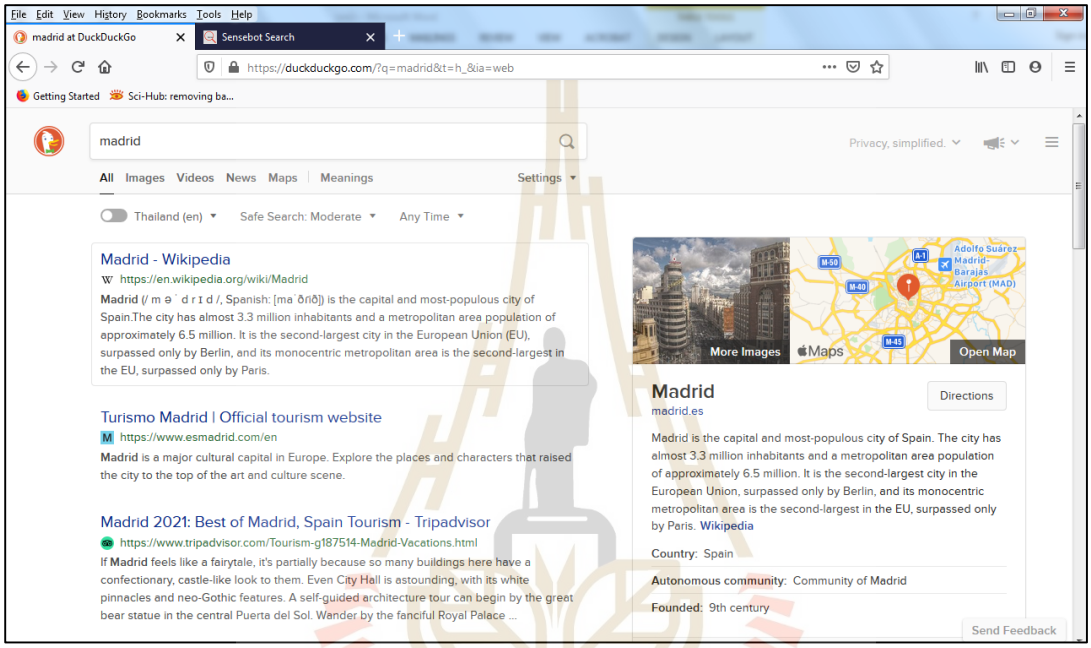
4. การสืบค้นและขยายการสืบค้นออนโทโลยี (Ontology search/expansion) เป็นการใช้ออนโทโลยีที่เป็นตัวแทนความรู้ในเรื่องใดเรื่องหนึ่งและการอธิบายคำถามต่าง ๆ เช่น กรณีที่ผู้ใช้พิมพ์คำถามโดยใช้คำว่า “Truck” ซึ่งหมายถึงรถบรรทุก ระบบจะเพิ่ม คำศัพท์จากออนโทโลยีของระบบ เช่น คำว่า “Vehicle” ที่หมายถึงยานพาหนะ เป็นต้น เพื่อให้การสืบค้นมีความเฉพาะเจาะจงและครอบคลุมกว้างขวางมากกว่า ซึ่งโปรแกรม ค้นหาเชิงความหมายจำนวนมากใช้วิธีการนี้

2.2.5 ตัวอย่างโปรแกรมค้นหาเชิงความหมาย

โปรแกรมค้นหาเชิงความหมายที่มีการพัฒนาและให้บริการอยู่ในปัจจุบันมี ด้วยกันหลายโปรแกรม เช่น DuckDuckGo, Factbites, senseBot และ Swoogle ซึ่ง โปรแกรมค้นหาแต่ละโปรแกรมต่างก็มีลักษณะการทำงานที่แตกต่างกัน หากแต่มีเป้าหมาย เหมือนกันคือการอำนวยความสะดวกในการค้นหาเอกสารต่าง ๆ บนเว็บที่ตรงความ ต้องการของผู้ใช้และเป็นเอกสารที่เกี่ยวข้องในเรื่องที่ผู้ใช้ต้องการค้นหามากที่สุด มีรายละเอียดดังนี้

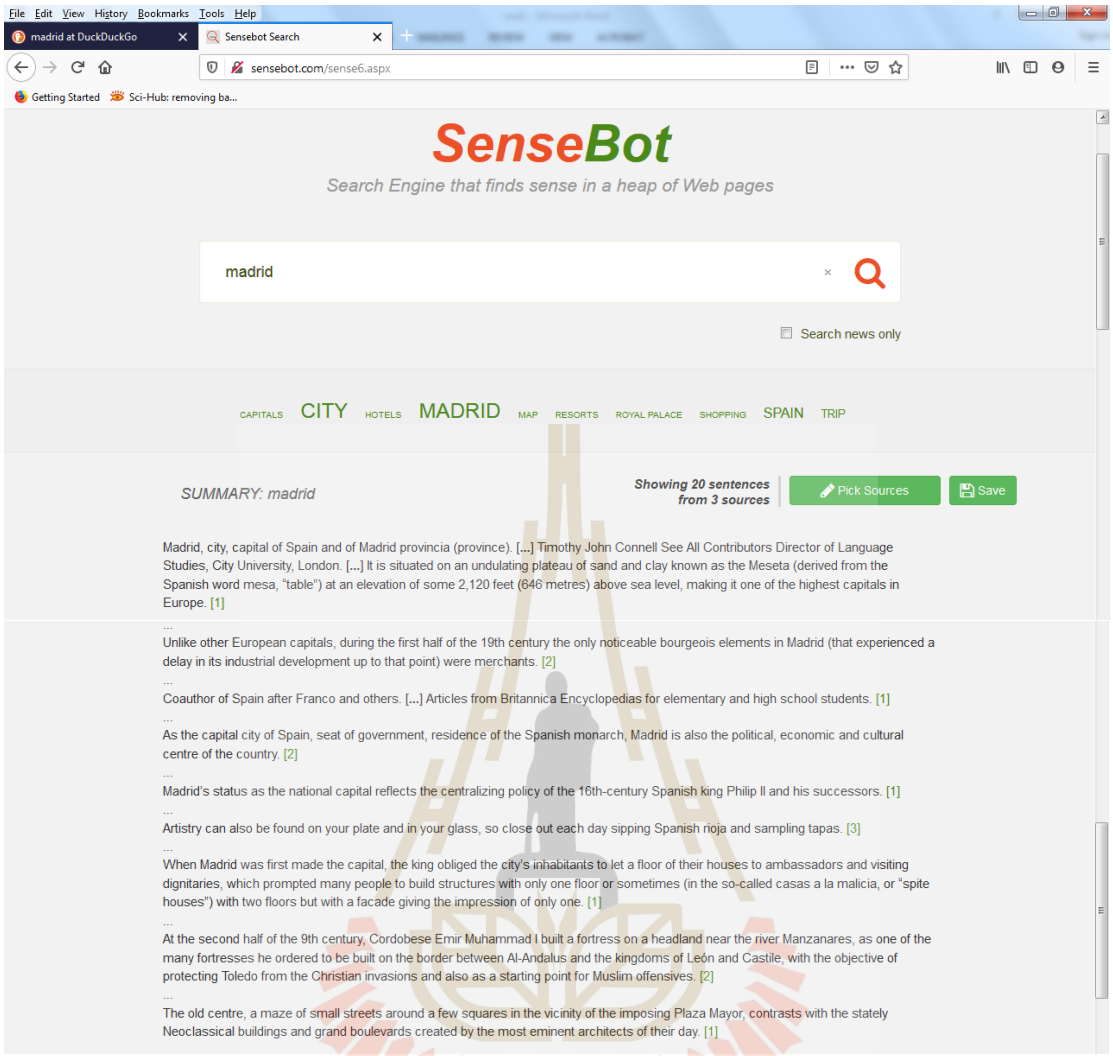
1. DuckDuckGo เป็นโปรแกรมค้นหาที่พัฒนาโดยบริษัท DuckDuckGo ซึ่ง เป็นบริษัทดำเนินงานด้านความปลอดภัยในการใช้อินเทอร์เน็ตหรือความเป็นส่วนตัวบน อินเทอร์เน็ต โดยมุ่งเน้นความเป็นส่วนตัวของผู้ใช้ (Privacy) และไม่จัดเก็บผลการสืบค้น ของผู้ใช้ โปรแกรมค้นหานี้ให้

สารสนเทศจากแหล่งที่ดีที่สุดและตรงความต้องการมากที่สุด ซึ่งผู้ใช้สามารถเลือกดูผลการสืบค้นได้หลายประเภท ได้แก่ เว็บเพจ รูปภาพ วีดิทัศน์ ข่าว และความหมายของคำ ผู้ใช้สามารถสืบค้นได้ที่เว็บไซต์ http://duckduckgo.com ดังตัวอย่างแสดงดังภาพที่ 2.2 ซึ่งหน้าจอสืบค้นต่อไปนี้เป็นตัวอย่างของการสืบค้นโดยใช้คำว่า “madrid” ได้ผลการค้น เป็นเว็บเพจที่นำเสนอสารสนเทศที่ต้องการ โดยผู้ใช้สามารถเชื่อมโยงไปดูเอกสารฉบับเต็มได้จากหัวข้อเรื่องที่ปรากฏอยู่เหนือข้อความที่เป็นรายละเอียดคร่าว ๆ ของเว็บเพจ หรือเชื่อมโยงจากเว็บไซต์ที่ปรากฏอยู่ใต้ข้อความของแหล่งข้อมูลแต่ละรายการ



ภาพที่ 2.2 ตัวอย่างภาพหน้าจอของโปรแกรมค้นหาเชิงความหมาย DuckDuckGo

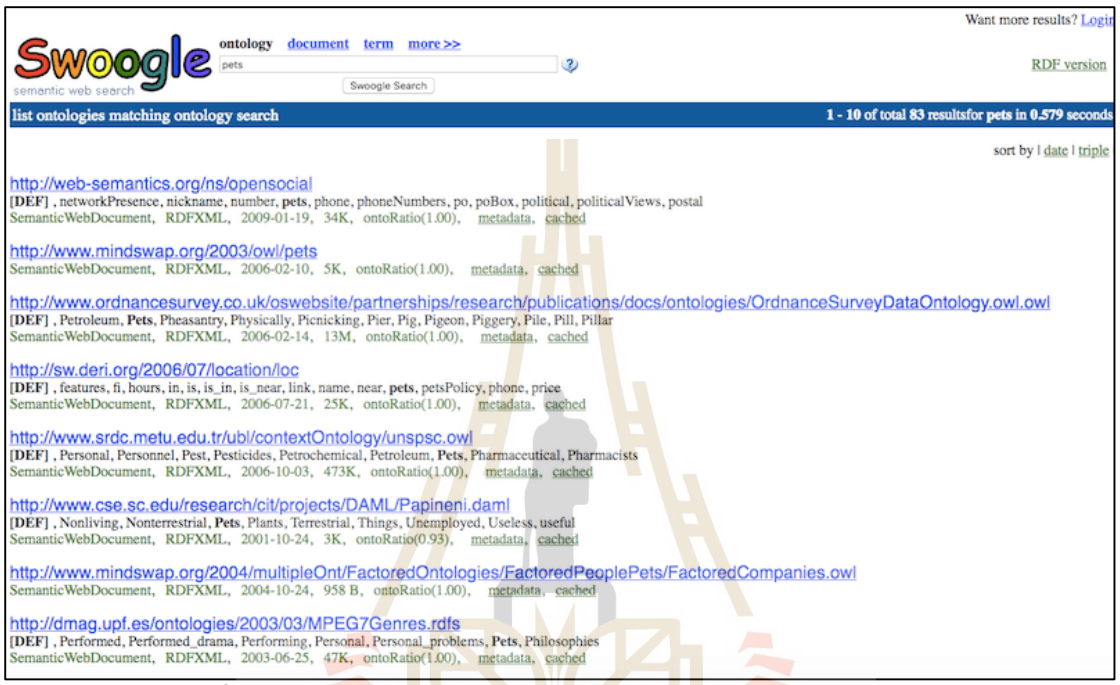
2. SenseBot เป็นโปรแกรมค้นหาที่พัฒนาโดยบริษัท Semantic Engines LLC โดยเป็นโปรแกรมค้นหาที่ให้บทสรุปของข้อความในการตอบข้อคำถามของผู้ใช้ ซึ่งเป็น คอลเล็กชั่นของจุดเชื่อมโยง (Links) ไปสู่เว็บเพจต่าง ๆ เป็นการวิเคราะห์ผลการค้นจาก เว็บที่ตรงความต้องการมากที่สุดและให้บทสรุปเนื้อหาที่นำเสนอในเว็บเพจ บทสรุปจะช่วย ในการประมวลหัวข้อที่เกี่ยวข้องกับข้อคำถามที่ผสมผสานกันของประเด็นสำคัญและ ตรงความต้องการมากที่สุด โปรแกรมค้นหานี้จะใช้การทำเหมืองข้อความ (Text mining) เพื่อวิเคราะห์คำในเว็บเพจต่าง ๆ และกำหนดเป็นแนวคิดหลัก จากนั้นจะดำเนินการสรุป เนื้อหาของเอกสารหลาย ๆ รายการเพื่อสร้างบทสรุปที่สอดคล้องกัน ผู้ใช้สามารถสืบค้น ได้ที่เว็บไซต์ http://sensebot.com แสดงดังภาพที่ 2.3 ซึ่งหน้าจอสืบค้นต่อไปนี้เป็นตัวอย่างของการสืบค้น โดยใช้คำว่า “madrid” ได้ผลการค้นเป็นเว็บเพจที่นำเสนอสารสนเทศที่ต้องการ โดย ผู้ใช้สามารถเชื่อมโยงไปดูเอกสารฉบับเต็มได้จากข้อความที่ปรากฏอยู่หลังคำว่า “SOURCE”



ภาพที่ 2.3 ตัวอย่างภาพหน้าจอโปรแกรมค้นหาเชิงความหมาย Sensebot

3. Swoogle เป็นโปรแกรมค้นหาที่พัฒนาขึ้นโดยกลุ่มวิจัยที่มีชื่อว่า UMBC ebiquity Research Group (Finin et al., 2014) ซึ่งประกอบด้วยคณาจารย์และนักศึกษาจากภาควิชาวิทยาการ คอมพิวเตอร์และวิศวกรรมไฟฟ้า และภาควิชาระบบสารสนเทศของ University of Maryland, Baltimore County โปรแกรมค้นหานี้ใช้สืบค้นเอกสารเว็บเชิงความหมายโดยเฉพาะออนโทโลยีเว็บเชิงความหมาย ซึ่งเป็นการวิเคราะห์คำในประโยคและทำดรรชนีให้กับคำที่อธิบายความหมายหลัก ๆ ของเอกสาร เช่น ตัวระบุแหล่งทรัพยากรสากลของเอกสาร คำศัพท์ที่ใช้และนิยามในเอกสาร คำอธิบายเกี่ยวกับเอกสาร และเนมสเปสที่ใช้ในเอกสาร โดยมุ่งเน้นให้ผู้ใช้สามารถใส่คำค้นเพื่อค้นหาเอกสารออนโทโลยีทั้งรูปแบบของอาร์ดีเอฟ (RDF) และโอดับเบิ้ลยูแอล (OWL) ผู้ใช้สามารถสืบค้นโปรแกรมค้นหานี้ได้ที่เว็บไซต์ http://swoogle.umbc.edu แสดงดังภาพที่ 2.4 ซึ่งการสืบค้นคำค้นเพียงคำเดียวสามารถระบุคำค้นตามหลังเมทาดาต้า พื้นฐาน เช่น เมื่อต้องการค้นหาคำว่า “madrid” ที่ปรากฏเป็นส่วนหนึ่งของตัวระบุแหล่ง ทรัพยากรสากลของเอกสาร ให้พิมพ์คำค้นว่า

url:madrid หรือการค้นโดยใช้คำที่ปรากฏในรายละเอียดของเอกสาร ให้พิมพ์คำค้นว่า descrmadrid หรือการค้นโดยใช้คำที่ ปรากฏเป็นชื่อของคลาสและคุณลักษณะที่อธิบายไว้ในเอกสาร ให้พิมพ์คำค้นว่า defmadrid เป็นต้น ซึ่งหน้าจอสืบค้นต่อไปนี้เป็นตัวอย่างของการสืบค้นโดยใช้คำว่า “url:madrid” ได้ผลการค้นเป็นเอกสารออนโทโลยีที่ต้องการ โดยผู้ใช้สามารถเชื่อมโยง ไปดูเอกสารออนโทโลยีได้จากเว็บไซต์ที่ปรากฏในแต่ละรายการ



ภาพที่ 2.4 ตัวอย่างภาพหน้าจอโปรแกรมค้นหาเชิงความหมาย Swoogle

จากตัวอย่างการสืบค้นสารสนเทศโดยใช้โปรแกรมค้นหาเชิงความหมายข้างต้น จะเห็นได้ว่าการพัฒนาโปรแกรมค้นหาเชิงความหมายเพื่อเป็นทางเลือกสำหรับผู้ใช้ในการสืบค้นสารสนเทศที่ต้องการบนเว็บและเป็นโปรแกรมค้นหาที่มาแทนที่โปรแกรม ค้นหาแบบเดิมนั้น มีหน่วยงานทั้งภาครัฐและภาคเอกชนที่ให้ความสำคัญในการพัฒนา โปรแกรมค้นหาเชิงความหมาย ไม่ว่าจะเป็นหน่วยงานวิจัยในมหาวิทยาลัยหรือบริษัทเอกชนที่มุ่งเน้นการพัฒนาโปรแกรมโดยใช้เทคโนโลยีเว็บเชิงความหมาย ซึ่งโปรแกรม ค้นหาเชิงความหมายจะให้แหล่งสารสนเทศที่หลากหลายโดยเฉพาะแหล่งสารสนเทศที่อธิบายความหมายของคำศัพท์ที่เป็นคำค้น แหล่งสารสนเทศที่ค้นได้มีความสอดคล้องกับคำค้นและตรงความต้องการของผู้ใช้ โดยโปรแกรมค้นหาจะวิเคราะห์คำในเอกสารแล้ว สร้างเป็นบทสรุป (Summary) หรือนับจำนวนคำของคำค้นที่ปรากฏในเอกสาร เพื่อให้ ผู้ใช้ได้ใช้พิจารณาเลือกแหล่งสารสนเทศที่ค้นหาได้อย่างมีประสิทธิภาพ นอกจากนี้ โปรแกรมค้นหาเชิงความหมายยังให้จุดเชื่อมโยงไปยังเอกสารเว็บเชิงความหมาย หรือ เว็บไซต์ที่เชื่อมโยงไปยังเอกสารออนโทโลยีอีกด้วย

เมื่อเปรียบเทียบโปรแกรมค้นหาเชิงความหมายที่ยกเป็นตัวอย่างทั้ง 3 โปรแกรมในประเด็นต่าง ๆ ได้แก่ วิธีการ เทคนิคที่ใช้ ประเภทของผลการค้น จุดเด่นและ ข้อจำกัด สามารถสรุปข้อมูลได้ตามตารางที่ 2.2 (Shah, Shah, & Deulkar, 2015)

ตารางที่ 2.2 การเปรียบเทียบโปรแกรมค้นหาเชิงความหมาย

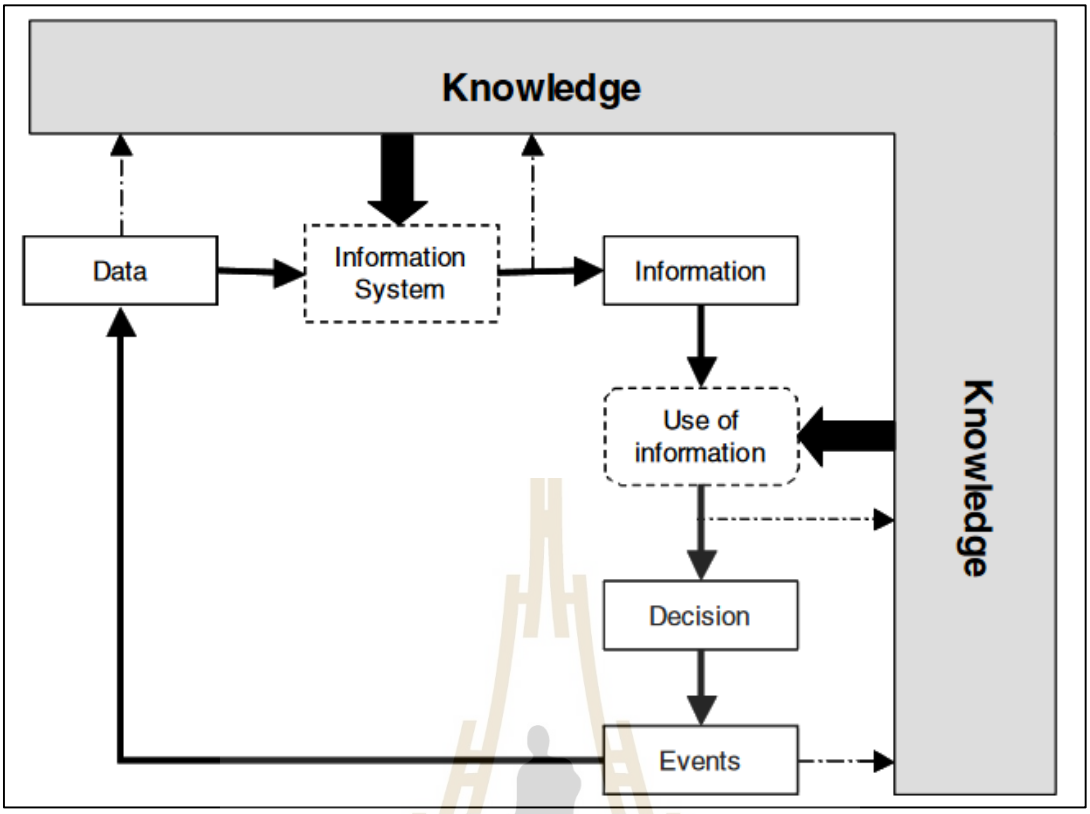
ประเด็นการเปรียบเทียบ	**โปรแกรมค้นหาเชิงความหมาย**		
	DuckDuckGo	**SenseBot**	**Swoogle**
วิธีการ	วิธีการจัดกลุ่มและการประมวลผลภาษาธรรมชาติ	วิธีการสืบค้นเชิงแนวคิด	วิธีการเชิงเนื้อหา
เทคนิคที่ใช้	การตอบคำถามทันที ซึ่งรวบรวมจากส่วนต่อประสานโปรแกรมประยุกต์หรือแหล่งข้อมูลคงที่ต่าง ๆ	การระบุแนวคิดเชิงความหมายหลักจากข้อคำถามของผู้ใช้โดยใช้วิธีการทำเหมืองข้อความที่เกิดจากการวิเคราะห์คำในเว็บเพจซึ่งใช้ในการแสดงบทสรุปที่สอดคล้องกัน	โปรแกรมค้นหาเชิงความหมายที่ทำดรรชนีโดยดึงข้อมูลจากเว็บซึ่งใช้ในการสืบค้นออนโทโลยีและข้อมูลตัวอย่าง
ประเภทของผลการค้น	บทสรุป	บทสรุป	อาร์ดีเอฟ(RDF) และโอดับเบิ้ลยูแอล (OWL)
จุดเด่น	- ให้บทสรุปของหัวข้อเรื่องและเรื่องที่เกี่ยวข้อง ซึ่งแสดงในกล่องปรากฏอยู่เหนือผลการค้น - เน้นความเป็นส่วนตัวและไม่บันทึกข้อมูลผู้ใช้ - ให้ผลการค้นจากแหล่งต่างๆ และดึงข้อมูลจากเว็บ	สรุปเอกสารที่หลากหลาย	- ค้นหาออนโทโลยีที่เหมาะสม - ค้นหาโครงสร้างข้อมูลตัวอย่างของเว็บเชิงความหมายที่เหมาะสม
ข้อจำกัด	ขาดลักษณะสำคัญในการค้นหารูปภาพและวีดิทัศน์	ทำงานด้วยโปรแกรมไฟร์ฟอกซ์(Firefox) และโปรแกรมค้นหา Googleเพื่อแสดงผลการค้น	การขยาย Swoogle เพื่อทำดรรชนีและใช้ค้นหาข้อมูลตัวอย่างจำนวนมาก

โปรแกรมค้นหาเชิงความหมาย นับเป็นเครื่องมือตัวใหม่ที่ใช้ค้นหาสารสนเทศบนเว็บ ซึ่งอาจมาแทนที่โปรแกรมค้นหาแบบเดิม เนื่องจากเป็นเครื่องมือที่ให้สารสนเทศ เป็นเอกสารเว็บเชิงความหมายและเอกสารออนโทโลยีที่คำนึงถึงความหมายที่แฝงอยู่ในเอกสารต่าง ๆ โดยให้ผลการค้นที่แม่นยำและตรงความต้องการของผู้ใช้มากกว่าโปรแกรมค้นหาแบบเดิม อนึ่งในการพัฒนาโปรแกรมค้นหาเชิงความหมายของหน่วยงานต่าง ๆ นั้น มีโครงการที่หน่วยงานทั้งภาครัฐและภาคเอกชนริเริ่มและดำเนินการอย่างเป็นรูปธรรมเกิดขึ้นเป็นจำนวนมาก บางโปรแกรมได้รับการพัฒนาและใช้สืบค้นได้ในปัจจุบัน แต่บางโปรแกรมอาจไม่ได้รับความนิยมหรือไม่สามารถเข้าถึงได้แล้วในปัจจุบัน ดังนั้นผู้ใช้ที่ต้องการค้นหาสารสนเทศโดยใช้โปรแกรมค้นหาเชิงความหมายจึงควรศึกษาลักษณะของโปรแกรมค้นหาและวิธีการใช้งาน เพื่อให้เกิดความเข้าใจว่าจะค้นหาสารสนเทศที่ ต้องการอย่างไรและจะได้ผลลัพธ์อะไรบ้างจากการใช้โปรแกรมค้นหานั้น ๆ ดังนั้น ในงานวิจัยนี้จะมีการพัฒนาตัวแบบการค้นคืน ที่ทำงานเสมือนหนึ่งเป็นโปรแกรมค้นหาเชิงความหมายสำหรับผลไม้เศรษฐกิจไทย

2.3 การจัดเก็บความรู้และการจัดความรู้ให้เป็นระบบ

2.3.1 ความรู้และการจัดการความรู้

ความรู้ มีคำที่เกี่ยวข้องสัมพันธ์กันอยู่หลายคำ ประกอบด้วยข้อมูล สารสนเทศ ความรู้และปัญญา โดยข้อมูล (Data) หมายถึง ข้อเท็จจริงต่าง ๆ ที่ใช้ในชีวิตประจำวัน อาจเป็นข้อมูลที่เกิดจากการปฏิบัติการ หรือการสังเกตการณ์ การกระทำ หรือลักษณะต่าง ๆ ของวัตถุ คน สัตว์ สิ่งของ มีลักษณะเป็นข้อมูลดิบที่ยังไม่ผ่านการกลั่นกรองหรือประมวลผล ไม่สามารถนำมาใช้ประโยชน์ได้ทันที มักปรากฏในรูปของตัวเลข ตัวอักษร สัญลักษณ์ รูปภาพ เป็นต้น เมื่อผ่านการประมวลผลด้วยวิธีการต่าง ๆ ข้อมูลจะกลายเป็นสารสนเทศ (Information) ที่อยู่ในรูปแบบที่พร้อมนำไปใช้ประโยชน์ในการปฏิบัติงาน และการตัดสินใจ จากนั้นรูปแบบของสารสนเทศที่มีการนำประสบการณ์ วิจารณญาณ ความคิด ค่านิยม และปัญญาของมนุษย์มาวิเคราะห์ เพื่อประยุกต์ใช้ในการทำงานการตัดสินใจ และการแก้ปัญหาความสัมพันธ์ของข้อมูล สารสนเทศจะกลายเป็นความรู้ (Knowledge) สามารถแสดงความสัมพันธ์ได้ในภาพที่ 2.5 ดังนั้นสรุปได้ว่าการสั่งสมสารสนเทศปริมาณมาก จนเกิดเป็นข้อสรุปที่สามารถนำไปใช้ให้เกิดประโยชน์จนก่อให้เกิดความรู้ (Becerra-Fernandez, I. and Sabherwal, R., 2015)



ภาพที่ 2.5 ความสัมพันธ์ระหว่างข้อมูล สารสนเทศ และความรู้

หมายเหตุ.จาก Knowledge Management Systems and Processes (หน้า 21),

โดย Becerra-Fernandez, I. and Sabherwal, R., 2015, Taylor & Francis

ความรู้แบ่งเป็น 2 ประเภท คือ 1) ความรู้ที่ชัดแจ้ง (Explicit Knowledge) มักอยู่ในรูปแบบของการจัดเก็บ และสืบค้นข้อมูลจากเอกสาร หรือฐานข้อมูลที่มีอยู่ และ 2) ความรู้ที่อยู่ในตัวบุคคล (Tacit Knowledge) มักอยู่ในรูปแบบของการส่งเสริมให้เกิดการบันทึก ถ่ายทอดข้อมูลความรู้ของแต่ละบุคคลให้กับผู้อื่น เช่นในรูปแบบของการเขียนบล็อก หรือวิกิ เป็นต้น อย่างไรก็ดี การรวบรวมองค์ความรู้ที่กระจัดกระจายอยู่ในตัวบุคคล หรือเอกสาร มาพัฒนาให้เป็นระบบ เพื่อให้ทุกคนสามารถเข้าถึงความรู้ได้อย่างมีประสิทธิภาพ เป็นที่มาของการจัดการความรู้ (Knowledge Management : KM) (North, K and Kumta, G., 2015)

การจัดการความรู้เชิงความหมาย (Semantic Knowledge Management) (Acampora, G., et al., 2020) เป็นรูปแบบการจัดการความรู้ในอีกรูปแบบหนึ่ง ที่มุ่งเน้นการจัดเก็บองค์ความรู้ในรูปแบบที่คอมพิวเตอร์สามารถเข้าใจและนำไปประมวลผลได้ เรียกว่าออนโทโลยี (Ontology) ซึ่งวิธีการหนึ่งที่ใช้ในการพัฒนาคือโดยการใช้กระบวนการทางวิศวกรรมความรู้ (Knowledge Engineering) การจัดการความรู้เชิงความหมายจำเป็นต้องอาศัยแหล่งความรู้ที่มีอยู่

ทั้งที่อยู่ในรูปแบบของเอกสารอ้างอิง (Reference documents) และจากผู้เชี่ยวชาญเฉพาะสาขา (Domain experts) ดังนั้นการจัดการความรู้เชิงความหมายจึงต้องมีการผสมผสานทั้งการจัดการความรู้ที่ชัดแจ้ง (Explicit Knowledge) และการจัดการความรู้ที่อยู่ในตัวบุคคล (Tacit Knowledge) เข้าด้วยกันในรูปแบบฐานความรู้สำหรับโปรแกรมคอมพิวเตอร์

หากพิจารณาตามทิศทางการพัฒนาเทคโนโลยีเว็บแล้ว (Acampora, G., et al., 2020) จะพบว่าเทคโนโลยีเว็บในยุคเริ่มต้นมุ่งเน้นที่การจัดการความรู้ที่ชัดแจ้ง ในรูปแบบของการจัดเก็บและสืบค้นข้อมูลจากเอกสาร HTML และ ฐานข้อมูลจากเว็บไซต์ต่าง ๆ เป็นหลัก ในขณะที่เว็บยุคที่ 2 มุ่งเน้นที่การจัดการความรู้ที่อยู่ในตัวบุคคลมากยิ่งขึ้น ในรูปแบบของการเขียนบล็อก และวิกิ รวมทั้งเว็บไซต์เครือข่ายสังคม ดังเช่น Facebook และ Twitter เป็นต้น ในเว็บยุคที่ 3 จะมุ่งเน้นที่การจัดการความรู้เชิงความหมายมากยิ่งขึ้น เพื่อนำไปสู่การพัฒนาโปรแกรมตัวแทนที่มีความชาญฉลาด (Intelligent Agents) เพื่อมาช่วยในการประกอบกิจกรรมและสืบค้นข้อมูลของผู้ใช้ได้ดียิ่งขึ้น เรียกว่าเว็บเชิงความหมาย (Semantic Web) ที่มีการเชื่อมโยงข้อมูลของเว็บผู้พัฒนาและเว็บของแหล่งข้อมูลอื่นที่สัมพันธ์กัน ทำให้เกิดระบบสืบค้นที่มีประสิทธิภาพสามารถสืบค้นข้อมูลได้อย่างรวดเร็วและตรงประเด็น ภายใต้ความสัมพันธ์ของคำที่มีความหมายต่อกันและสามารถเชื่อมโยงไปยังข้อมูลที่ต้องการอย่างแท้จริง ส่งผลให้เกิดนวัตกรรมการสืบค้นข้อมูลผ่านฐานข้อมูลขนาดใหญ่ที่มีการเชื่อมโยงความสัมพันธ์ของข้อมูล รายละเอียดดังตารางที่ 2.3

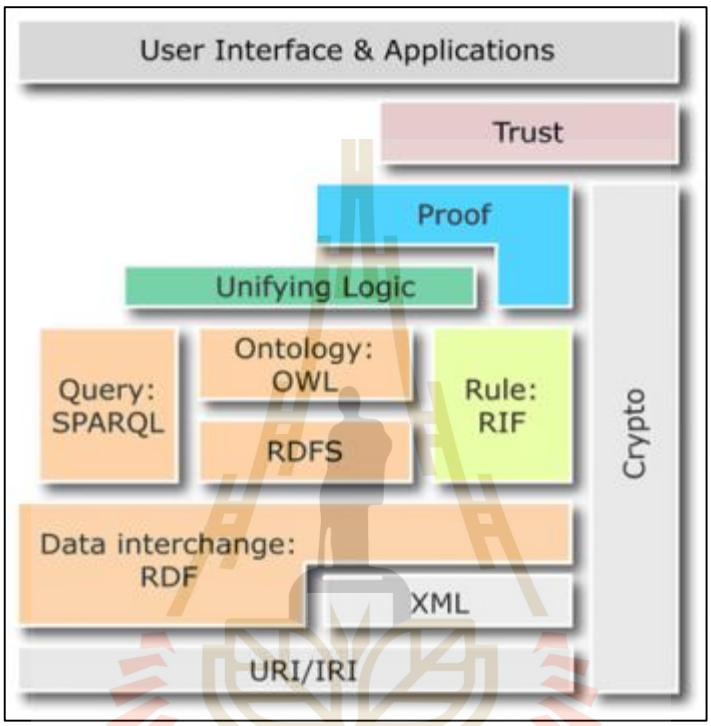
ตารางที่ 2.3 เทคโนโลยีเว็บและการจัดการความรู้

เทคโนโลยี	การจัดการความรู้	การประยุกต์ใช้งาน
เว็บ 1.0	Explicit Knowledge Management	HTML Document, Web-Database
เว็บ 2.0	Tacit Knowledge Management	WiKi, Blog, Social Network
เว็บ 3.0	Semantic Knowledge Management	Intelligent Agents, Question Answering, Recommended Systems and Group Recommendation

2.3.2 ฐานความรู้ด้วยตัวแบบออนโทโลยี

ในทางวิทยาการคอมพิวเตอร์ ออนโทโลยี (Ontology) คือ รูปแบบ ข้อกำหนดที่ชัดเจนของแนวคิดที่มีการแบ่งปันหรือใช้ร่วมกัน ซึ่งจากความหมายดังกล่าวออนโทโลยีจะต้องมี รูปแบบ (Formal) ที่เครื่องจักรสามารถประมวลผลได้ ความชัดเจน (Explicit) ของรูปแบบที่มีการกำหนดไว้ และ แบ่งปัน (Shared) ตามที่ได้มีการตกลงที่จะใช้งานร่วมกัน (Studer, Benjamins and Fensel,

1998) โดยออนโทโลยีเป็นส่วนประกอบที่สำคัญอย่างหนึ่งของเว็บเชิงความหมาย (Semantic Web) ซึ่งเป็นเครื่องมือสำหรับบูรณาการและจัดระเบียบข้อมูล เชิงความหมายทำให้มีการเชื่อมโยงเครือข่ายของข้อมูลบนเว็บไซต์ต่างๆและส่งผลให้การสืบค้นเป็นไปอย่างมีประสิทธิภาพและสามารถสร้างความสัมพันธ์ให้กับข้อมูลที่มาจากแหล่งที่ต่างกันเป็นฐานข้อมูลออนโทโลยี (Ontology) ที่เชื่อมโยงกัน ดังภาพที่ 2.6



ภาพที่ 2.6 สถาปัตยกรรมของเว็บเชิงความหมาย

ที่มา: https://lst.nectec.or.th/oam/img/semanticwebarch.png

จากภาพที่ 2.6 แสดงภาษาที่ใช้ในเว็บเชิงความหมายและสถาปัตยกรรมของเว็บเชิงความหมาย แสดงรายละเอียดจากล่างขึ้นบน (Bottom-Up Architecture) มีรายละเอียดดังนี้

- URI (Uniform Resource Identifier) เป็นการระบุกลไกในการเข้าถึงทรัพยากร (ชื่อแฟ้ม) และแหล่งเก็บทรัพยากร
- XML (Extensive Markup Language) เป็นภาษาที่ใช้สำหรับการเขียนภาษาที่ใช้นิยามความหมายของเอกสารหรือข้อมูล (Markup) โดยที่เอกสารนั้นมีการใช้ Metadata เพื่อบอกหน้าที่และประเภทของข้อมูลของส่วนต่าง ๆ ของเอกสารนั้นได้โดยชัดเจนด้วยการเพิ่ม Metadata เข้าไปในเอกสารสามารถทำให้โครงสร้างของเอกสารชัดเจนขึ้นและทำให้การประมวลผลเอกสารเป็นไปโดยง่ายและไม่จำเป็นที่จะต้องอาศัยมนุษย์เพื่อตีความเอกสาร

• RDF (Record Description Framework) เป็นภาษามาตรฐานที่ใช้อธิบายลักษณะและความสัมพันธ์ของข้อมูลคล้ายกับแผนภาพอีอาร์โดยใช้โครงสร้างและรูปแบบของ XML

• RDFS (RDF Schema) เป็นภาษาที่ใช้นิยามหรือกำหนดโครงสร้างของ RDF ในการอธิบายลักษณะของข้อมูลในรูปแบบของคลาสที่อยู่บนมาตรฐานของ RDF

• OWL (Web Ontology Language) เป็นภาษาอธิบายข้อมูลในเชิง Ontology โดยพิจารณาจากความสัมพันธ์ระหว่างสิ่งต่าง ๆ ของความหมาย

• SPARQL เป็นโปรโตคอลและภาษาสืบค้นข้อมูลของแหล่งข้อมูลเว็บเชิงความหมาย

• RIF (Rule Interchange Format) เป็นกฎของ Semantic Web Stack ในแต่ละชั้น

• Logic ใช้เพื่อพัฒนา Ontology ให้สามารถเขียนแอพพลิเคชันที่มีลักษณะเฉพาะและแสดงวิธีการอธิบายองค์ความรู้

• Proof เป็นชั้นที่มีการพิสูจน์ว่าเอกสารนั้นถูกต้องตามกฎ

• Trust เป็นชั้นที่มีความสำคัญของเว็บในการได้รับความน่าเชื่อถือเมื่อมีระบบความปลอดภัยที่ดี

จะเห็นได้ว่าเว็บเชิงความหมายจึงเป็นวิสัยทัศน์ของทิศทางการพัฒนาเทคโนโลยีเว็บในยุคปัจจุบัน โดยมีองค์กร W3C เป็นผู้กำหนดแนวทางการพัฒนามาตรฐานของเว็บเชิงความหมายเป็นเว็บของข้อมูลที่มีการเชื่อมโยงข้อมูลกัน ในลักษณะของเครือข่ายเชิงความหมายด้วยเทคโนโลยีเว็บเชิงความหมาย ที่ทำให้ผู้ใช้สร้างข้อมูลสร้างคำศัพท์ฐานความรู้ (Ontology) และข้อกำหนดในการจัดการข้อมูลโดยใช้เทคโนโลยีต่าง ๆ เช่น RDF, SPARQL, OWL เพื่อนำไปสู่การพัฒนาโปรแกรมประยุกต์ที่มีความชาญฉลาดการสืบค้นข้อมูลที่อิงตามความหมายที่ผู้ใช้กำหนด

ดังนั้น เว็บเชิงความหมายไม่ใช่สภาพแวดล้อมและไม่ใช่การประมวลผลแบบใหม่ แต่เป็นส่วนขยายของเว็บที่มีอยู่ ข้อมูลเชิงความหมายที่ให้ "ความหมาย" ของข้อมูลที่เครื่องคอมพิวเตอร์จะอ่านและประมวลผลได้จะถูกจัดวางเป็นชั้น ๆ อุปสรรคที่เกิดขึ้นคือความซับซ้อนซึ่งประกอบด้วยแง่มุมต่าง ๆ มากมายของการสร้างเว็บเชิงความหมาย เนื่องจากความหมายของข้อมูลเป็นนามธรรมมากกว่าเอกสารหรือฐานข้อมูล ดังนั้นหากนำมาสร้างเป็นแบบจำลองเชิงโครงสร้างที่มีความสัมพันธ์กันย่อมส่งผลให้มีขนาดมหาศาลกว่าจะครอบคลุมความเชื่อมโยงทั้งหมดได้ อีกทั้งความหมายของคำไม่มีลักษณะตายตัว ไม่มีลักษณะเฉพาะเจาะจง และไม่มีมาตรฐาน ตลอดจนความหมายเปลี่ยนแปลงไปได้ตลอดเวลาตามสถานการณ์และตามยุคสมัย จึงเป็นอุปสรรคสำคัญของ “ข้อมูลเชิงความหมาย”

ข้อมูลเชิงความหมายจะถูกจัดเก็บในออนโทโลยี ซึ่งเป็นฐานความรู้สำหรับโปรแกรมคอมพิวเตอร์ที่จัดเก็บองค์ความรู้เฉพาะทาง (Domain Knowledge) ที่ส่วนใหญ่จะเกิดจาก

การพัฒนาขึ้นโดยวิศวกรความรู้ (Knowledge Engineers) เป็นแนวทางในการพัฒนา จัดเก็บ และแบ่งปันองค์ความรู้จากผู้เชี่ยวชาญเฉพาะทาง (Domain Experts) โดยมีวัตถุประสงค์หลัก เพื่อให้สามารถนำความรู้เฉพาะทางไปประยุกต์ใช้กับโปรแกรมคอมพิวเตอร์ได้อย่างชาญฉลาด ใน 4 ประการ คือ1) เพิ่มความอัตโนมัติของกระบวนการ (Automation) 2) ลดภาระของมนุษย์ (Reduced workloads) 3) เพิ่มความแม่นยำและลดข้อผิดพลาดในกระบวนการทำงาน (Reduced errors) และ 4) สามารถนำไปประยุกต์ใช้งานได้ในโปรแกรมและระบบสารสนเทศต่าง ๆ ได้กว้างขวางยิ่งขึ้น (Interoperability) จากคุณลักษณะสำคัญของฐานความรู้ที่สามารถแบ่งปันและใช้ซ้ำได้ (Share and Reuse) จากมาตรฐานของภาษาที่ใช้ในการพัฒนาออนโทโลยี (Ontology language) ที่สามารถแบ่งปันและแลกเปลี่ยนข้อมูลได้บนเครือข่ายเว็บด้วยภาษา OWL (Web ontology language) ตามแนวทางของเว็บความหมาย อ้างอิงจากมาตรฐาน RDF (Resource description framework) โดยใช้รูปแบบภาษา XML (Extensible Markup Language) และมาตรฐานการอ้างอิงข้อมูล URI โดยมาตรฐานเหล่านี้ล้วนเป็นองค์ประกอบสำคัญต่อการพัฒนาเว็บเชิงความหมาย เครื่องมือสนับสนุนการพัฒนาออนโทโลยี (Ontology Editor) ในปัจจุบันที่ได้รับความนิยม เช่น โปรแกรม Protégé ซึ่งพัฒนาโดย มหาวิทยาลัยสแตนฟอร์ด โปรแกรมHozo ซึ่งพัฒนาโดยมหาวิทยาลัยโอซากา เป็นต้น โดยเครื่องมือเหล่านี้เป็นเครื่องมือสนับสนุนกระบวนการวิศวกรรมความรู้ที่ช่วยให้ผู้เชี่ยวชาญเฉพาะทางสามารถถ่ายทอดและจัดเก็บองค์ความรู้ในรูปแบบของออนโทโลยีได้สะดวก และง่ายดายยิ่งขึ้น

จากนิยามของออนโทโลยีจะเห็นถึงองค์ประกอบของออนโทโลยีที่ประกอบด้วยส่วนต่าง ๆ สรุปได้ดังนี้ (Calero, C., Ruiz, F. and Piattini, M., 2006)

1) แนวคิด (Concept) หมายถึง ขอบเขตของความรู้เรื่องใดเรื่องหนึ่งหรือความคิดทั่วไปหรือนามธรรมในโดเมนที่สนใจ และสามารถอธิบายรายละเอียดได้ เช่น รายวิชา (Subject) แบบฝึกหัด (Exercise) บทเรียนสอนเสริม (Tutorial) และการสอนเสริม (Tutoring) เป็นต้น

2) คุณลักษณะ (Property) หมายถึง คุณสมบัติต่าง ๆ ที่นำมาใช้อธิบายแนวคิด เช่น ประเภทรายวิชา (SubjectType) ระดับบทเรียน (TutorialLevel) ชื่อบทเรียน (TutorialName) เป็นต้น

3) ความสัมพันธ์ (Relationship) หมายถึง รูปแบบของความสัมพันธ์กันระหว่างแนวคิด โดยมีการระบุความสัมพันธ์ไว้เป็นแบบต่าง ๆ ได้แก่

3.1) ความสัมพันธ์แบบลำดับชั้น (Subclass of หรือ Is-a hierarchy) คือความสัมพันธ์ที่มีคุณสมบัติการถ่ายทอด คุณสมบัติของแนวคิดแม่ไปยังแนวคิดลูก เช่น Database is-a Subject ซึ่งอธิบายได้ว่าฐานข้อมูล (Database) เป็นรายวิชา (Subject)

3.2) ความสัมพันธ์แบบเป็นส่วนหนึ่ง (Part-of) คือ ความสัมพันธ์ที่หมายถึงการเป็นส่วนประกอบเช่น SumExercise part-of FunctionExercise ซึ่งอธิบายได้ว่า แบบฝึกหัดการหาค่าผลรวม (SumExercise) เป็นส่วนหนึ่งของแบบฝึกหัดฟังก์ชัน (FunctionExercise)

3.3) ความสัมพันธ์เชิงความหมาย (Syn-of) คือ ความสัมพันธ์ที่แสดงถึงแนวคิดที่มีความเหมือนเชิงความหมายต่อกัน เช่น Degree Syn-of Education ซึ่งอธิบายได้ว่า ระดับการศึกษา (Degree) มีความหมายเดียวกันกับการศึกษา (Education) สามารถใช้แทนกันได้

3.4) ความสัมพันธ์การเป็นตัวแทน (Instance-of) คือ ความสัมพันธ์ที่แสดงถึงการเป็นตัวแทน หรือสมาชิกของแนวคิด นอกจากนี้ออนโทโลยียังประกอบไปด้วยความสัมพันธ์เชิงความหมายอื่น ๆ ที่สอดคล้องกับแนวคิดซึ่งกำหนดโดยผู้เชี่ยวชาญ

4) ข้อกำหนดในการสร้างความสัมพันธ์ (Axiom) หมายถึง เงื่อนไขหรือข้อกำหนดเฉพาะ หรือตรรกะในการแปลงความสัมพันธ์ระหว่างแนวคิดกับคุณสมบัติ หรือแนวคิดกับแนวคิดเพื่อให้แปลงความหมายได้ถูกต้อง ซึ่งเป็นกลไกสำคัญที่ใช้ในการอนุมานความรู้เพื่อสร้างความรู้ใหม่จากออนโทโลยี

5) ตัวอย่างข้อมูล (Instances) หมายถึง คำศัพท์ที่มีการกำหนดความหมายไว้ในออนโทโลยีเรื่องนั้น ๆ ดังนั้นจะเห็นได้ว่าออนโทโลยีเกี่ยวข้องกับการให้นิยามสำหรับสิ่งที่สนใจที่ประกอบด้วยแนวคิด คุณลักษณะของแนวคิด และความสัมพันธ์ระหว่างแนวคิดภายในขอบเขตที่กำหนด ซึ่งเป็นมุมมองระดับแนวคิด

2.3.3 การออกแบบออนโทโลยี

กระบวนการออกแบบออนโทโลยีที่พัฒนาตามแนวคิดของ Noy and McGuinness (2001) แบ่งได้เป็นขั้นตอนย่อยทั้ง 7 ขั้นตอนดังนี้

1) ขั้นกำหนดโดเมนที่สนใจและขอบเขตของออนโทโลยี โดยกำหนดขอบเขตเรื่องของออนโทโลยีและขอบเขตการใช้งานของออนโทโลยีให้ครอบคลุมคำถามของการใช้งาน ผู้ออกแบบต้องตอบคำถามให้ได้ว่าจะใช้ออนโทโลยีที่ออกแบบมาเพื่องานอะไรหรือเพื่อตอบคำถามใด และใครจะเป็นผู้ใช้ เช่น ออนโทโลยีในโดเมนไวน์และอาหาร ผู้ออกแบบสร้างขึ้นเพื่อประยุกต์ใช้ในการแนะนำอาหารและไวน์

2) ขั้นพิจารณานำออนโทโลยีที่มีอยู่แล้วกลับมาใช้ใหม่ บางครั้งอาจพิจารณานำออนโทโลยีที่มีอยู่แล้วมาใช้ใหม่ หากออนโทโลยีนั้นสามารถตอบปัญหาการใช้งานของเราได้ การค้นหาออนโทโลยีที่เผยแพร่สามารถค้นหาผ่านเว็บของผู้ให้บริการ เช่น Ontolingua, DAML หรือเว็บไซต์ของเจ้าของออนโทโลยีนั้น

3) ขั้นระบุคำสำคัญในออนโทโลยี ขั้นตอนนี้เป็นการสร้างคำศัพท์ที่สำคัญเพื่อใช้อธิบายออนโทโลยี เช่น ออนโทโลยีไวน์มีคำสำคัญที่เกี่ยวข้อง เช่น ผู้ผลิตไวน์ (Winery) ที่ตั้ง (Location) สีของไวน์ (A wine's color)

4) ขั้นระบุคลาสและลำดับชั้นของคลาส สามารถทำได้ 3 วิธี ได้แก่ 4.1. แบบบนลงล่าง (Top-down) วิธีนี้เริ่มจากคำศัพท์ทั่วไปนำมาสร้างเป็นคลาส เช่น คำว่าไวน์หรืออาหารนำมาสร้างคลาสไวน์ (Wine) และคลาสอาหาร (Food) จากนั้นค่อยนึกถึงคำศัพท์ที่มีความเฉพาะมากขึ้น เช่น ไวน์ขาว (White wine) หรือ ไวน์แดง (Red wine) กำหนดให้เป็นคลาสย่อยของไวน์ 4.2. แบบล่างขึ้นบน (Bottom-up) วิธีนี้จะเริ่มจากคำศัพท์ที่มีความเฉพาะมากกว่าก่อน แล้วค่อยนำคำศัพท์มารวมกลุ่มเป็นคลาสย่อยของคำศัพท์ที่มีความเฉพาะน้อยกว่า และ 4.3. แบบผสมผสาน (Combination) ใช้ทั้งวิธี Top-down และ Bottom-up ผสมกัน

5) ขั้นระบุคุณสมบัติของแต่ละคลาสหรือสล็อตของคลาส (Slots) เช่น คลาสไวน์มีสล็อตสี (Color) รสสัมผัส (Body) กลิ่น (Flavor) ของไวน์ ที่อธิบายสี รสสัมผัส และกลิ่นของไวน์แต่ละชนิด

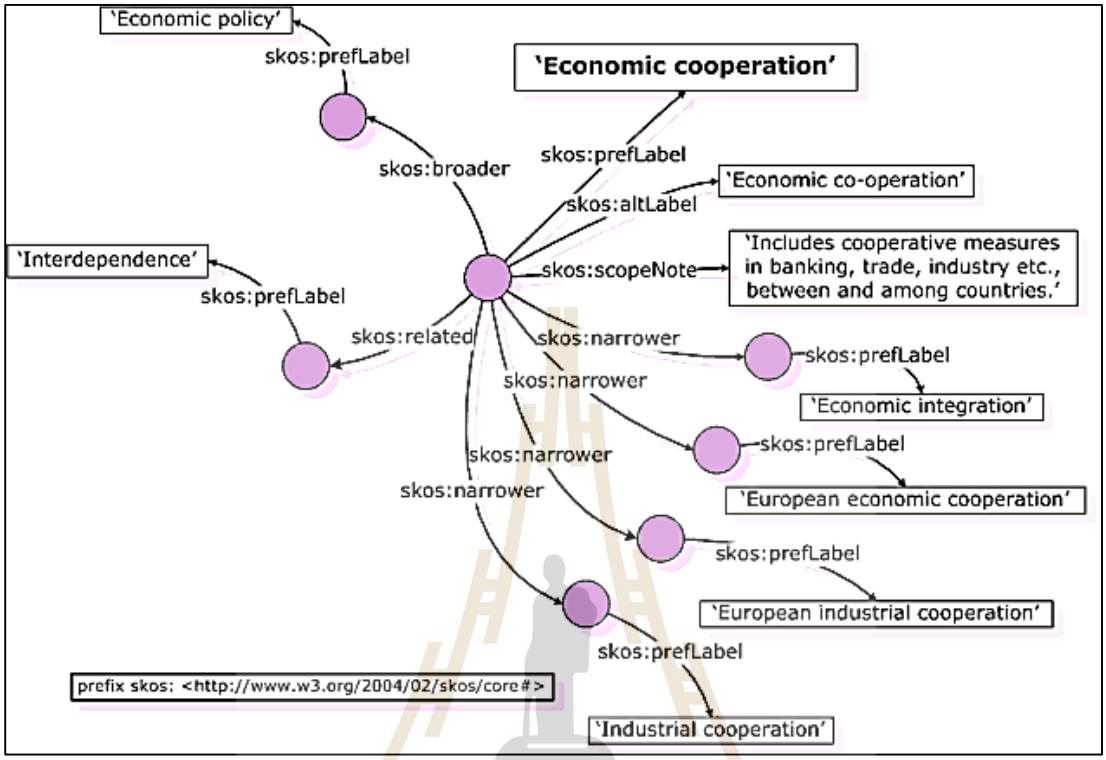
6) ขั้นระบุข้อจำกัดของสล็อต (Facets) ได้แก่ ชนิดของข้อมูล (Data type) จำนวนค่าของข้อมูลที่มีได้ (Cardinality) เช่น สล็อตชื่อไวน์ (Name) ชนิดข้อมูลเป็นอักขระ (String) มีได้แค่ 1 ค่า (Cardinality = 1) หมายความว่าไวน์หนึ่งชนิดจะมีชื่อได้เพียง 1 ชื่อต่อข้อมูล 1 ตัว

7) ขั้นสร้างตัวอย่างข้อมูล เป็นการสร้างตัวอย่างข้อมูลของแต่ละคลาสโดยมีขั้นตอนได้แก่เลือกคลาสสร้างข้อมูลในคลาสที่เลือก และเติมค่าในสล็อต เช่น Chateau-Morgon-Beaujolais เป็นตัวอย่างข้อมูลของคลาส Beaujolais และมีค่าในสล็อตที่อธิบายคุณสมบัติของคลาส เช่น มีรสสัมผัสแบบนุ่ม และมีสีแดง

2.3.4 การจัดความรู้ให้เป็นระบบ

ระบบจัดเก็บความรู้ (Knowledge Storage System) เป็นกระบวนการรวบรวมรายละเอียดความรู้ที่จัดเก็บในฐานความรู้ เพื่อให้สามารถค้นความรู้กลับคืนมาได้อย่างสะดวกและรวดเร็ว หากข้อมูลเชิงความหมายจะถูกจัดเก็บในออนโทโลยี ซึ่งฐานความรู้สำหรับโปรแกรมคอมพิวเตอร์ที่จัดเก็บองค์ความรู้เฉพาะทาง นิยมใช้เอสเคโอเอส (Simple Knowledge Organization System: SKOS) (Allemang, D. and Hendler, J., 2011) เป็นตัวแบบสำหรับสร้างคลังคำ หรือมาตรฐานข้อมูลที่ใช้ร่วมกัน พื้นฐานของสคอสประกอบไปด้วยความสัมพันธ์ดังนี้ คือ คำหลัก (skos:preLabel) คำเหมือน (skos:altLabel) คำที่กว้างกว่า (skos:broader) คำที่แคบกว่า (skos:narrower) และคำที่เกี่ยวข้อง (skos:related) เป็นแบบจำลองที่ใช้งานง่ายเหมาะสำหรับการแทนความรู้แบบใช้งานร่วมกัน พื้นฐานของสคอส ประกอบไปด้วย Concepts, Labels,Preferred Labels, Alternate Labels, Hidden

Labels (Fulvio and Laura, 2005) ตัวอย่างของการนิยามถึงความสัมพันธ์ของคำตามแบบโครงสร้างของเอสเคโอเอส แสดงได้ดัง ภาพที่ 2.7



ภาพที่ 2.7 ตัวอย่างความสัมพันธ์ของคำตามแบบโครงสร้างของเอสเคโอเอส

การจัดทำดัชนีของคำสำคัญหรือคำควบคุมที่ใช้สำหรับการค้นหา สามารถใช้วิธีการ การให้หมายเหตุเชิงความหมาย (Semantic annotation) ซึ่งเป็นการให้หมายเหตุหรือป้ายระบุ เพื่อให้ข้อมูลมีความหมาย โดยเป็นการให้หมายเหตุของเอกสารทั้งหมด หรือการให้หมายเหตุบางส่วนของเอกสารก็ได้ วิธีการให้หมายเหตุสามารถทำด้วยตนเองหรือโดยอัตโนมัติก็ได้ (Giannopoulos, Bikakis, et al, n.d.) กระบวนการในการให้หมายเหตุหรือป้ายระบุนั้น เป็นการให้ความหมายสำหรับเอกสารเพื่อให้สามารถประมวลผลทางคอมพิวเตอร์ได้ (Kiyavitskaya, Zeni, et al, 2005) การให้หมายเหตุเชิงความหมาย เป็นวิธีการเข้าถึงข้อมูลผ่านทางข้อมูลจากแหล่งภายนอกที่มีความหลากหลาย โดยความรู้ที่หลากหลายเหล่านั้นสามารถจับคู่กับคำอธิบายอย่างเป็นทางการได้ ทำให้มีความหมายเพิ่มมากขึ้น และเกิดการเชื่อมโยงซึ่งกันและกันของข้อมูลเหล่านั้น (Kiryakov, Popov, et al, 2004)

การออกแบบและพัฒนาตัวแบบค้นคืนเชิงความหมายสำหรับผลไม้เศรษฐกิจไทย จะมีการออกแบบฐานความรู้ตามกระบวนการวิศวกรรมออนโทโลยีเพื่อใช้จัดเก็บความรู้ต่าง ๆ เพื่อให้การค้นคืนได้อย่างมีประสิทธิภาพ มีการจัดทำคลังคำตามกระบวนการของการจัดความรู้ให้เป็นระบบ การให้หมายเหตุหรือป้ายคำแก่ความรู้ที่จัดเก็บ โดยมีการนำตัวแบบเอสเคโอเอสมาใช้สำหรับสร้าง

ความสัมพันธ์ของคำต่าง ๆ เพื่อช่วยในการค้นหา โดยจะมีการนำคำสำคัญมาจัดเป็นหมวดหมู่ และมีการเชื่อมโยงคำสำคัญไปยังความรู้ที่อยู่ในฐานความรู้อีกด้วย

2.4 การค้นคืนจากฐานความรู้ในออนโทโลยี

2.4.1 ภาษาสำหรับการค้นคืนจากฐานความรู้ในออนโทโลยี

ในการค้นคืนความรู้จะใช้ภาษาสปาร์เคิล (SPARQL Protocol and RDF Query Language: SPARQL) (Allemang, D. and Hendler, J., 2011) ซึ่งเป็นภาษาที่ใช้ในการสืบค้นข้อมูลในออนโทโลยี เพื่อใช้กับเทคโนโลยีเว็บเชิงความหมาย โดยการสืบค้นนั้นอยู่ในรูปแบบที่เรียกว่า อาร์ดีเอฟทริปเปิ้ล (RDF Triple) ซึ่งประกอบไปด้วยโครงสร้าง 3 ส่วน คือ ประธาน (Subject) ภาคแสดง (Predicate) และกรรม (Object) ภาษาสปาร์เคิลสามารถดึงข้อมูลโดยใช้ SELECT และ WHERE ซึ่งSELECT จะอธิบายคุณลักษณะ (Attribute) ที่มีคำนำหน้าโดย “?” ส่วน WHERE ตามด้วยเงื่อนไขสำหรับการดึงข้อมูลดังรูปแบบ

```
SELECT varlist WHERE {bgp}
by    varlist = (v1,v2,...,vn)
          bgp = Basic Graph Pattern
```

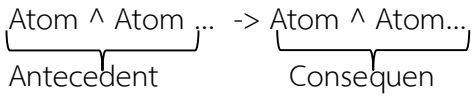
ตัวอย่างภาษาสปาร์เคิลในการสืบค้นหัวข้อ (Topic) ของกรอบเนื้อหา (CourseDomain) โดยแสดงผลเรียงลำดับตามหัวข้อดังนี้

```
SELECT ?Subject ?Course
    WHERE {?Subject :isSubjectOf ?Course
            }
    order by ?Subject
```

2.4.2 ภาษากฎสำหรับฐานความรู้ในออนโทโลยี

ในการดำเนินการโดยใช้กฎสำหรับการสร้างความรู้ขึ้นมาใหม่นี้ จะใช้ภาษาเอสดับบลิวอาร์แอล (Semantic Web Rule Language: SWRL) (Hebeler, J. et al., 2009) ซึ่งเป็นภาษากฎที่มีพื้นฐานของการรวมกันระหว่างภาษารูลเอ็มแอล (RuleML) และโอดับบิวแอลซึ่งพัฒนามาเพื่อทำงานร่วมกับเทคโนโลยีเว็บเชิงความหมายที่ใช้ในการอนุมานความรู้ใหม่จากฐานความรู้ที่มีอยู่แล้ว และสามารถดึงข้อมูลจากออนโทโลยีหรือดึงความรู้มาใช้ได้ทันที เป็นภาษาที่ใช้อนุมานความรู้ใหม่ ซึ่งได้จากฐานความรู้ในออนโทโลยี โดยการใช้กฎเพื่อหาความรู้ให้กับสมาชิกในคลาส (Individual) ภาษาเอสดับบิวอาร์แอลเป็นภาษาที่สามารถเขียนกฎให้อยู่ในรูปแบบประโยคได้ โดยเลือกใช้เครื่องมือให้เหตุผล (Rule Reasoner) ได้ไม่จำกัด รูปแบบของการเขียนกฎให้เขียนในรูปคู่ของเหตุ (Antecedent) และผล (Consequent) ส่วนเหตุ เรียกอีกอย่างหนึ่งว่า ส่วนตัว (Body) ส่วนผล เรียกอีกอย่างหนึ่งว่า

ส่วนหัว (Head) คู่ของเหตุและผลนี้มีตัวเชื่อมระหว่างกันที่สามารถเชื่อมหน่วย (Atom) ได้มากกว่าหนึ่งหน่วย (W3C, www, 2004) การเขียนกฎของเอสดับบิวอาร์แอลสามารถเขียนให้อยู่ในรูปแบบประโยคได้โดยไม่จำกัดด้วยเครื่องมือให้เหตุผล (Rule Reasoner) ที่มีรูปแบบการเขียนกฎคือ



โดยเหตุ (Antecedent) หรือส่วนตัว (Body) และ ผล (Consequent) หรือส่วนหัว (Head) (Segaran, T., Evans, C., and Taylor, J., 2009) จะประกอบด้วยตัวเชื่อมหน่วย (Atom) ได้มากกว่าหนึ่ง โดยหน่วยคือ นิพจน์ในรูปแบบ *p(arg1, arg2,..., argn)*โดย p คือ ภาคแสดงของคลาส (OWL classes) คุณสมบัติ (Properties) หรือชนิดข้อมูล (Data types) และ *arg1, arg2,..., argn* คือ อาร์กิวเมนต์ของนิพจน์ ได้แก่ ข้อมูล (OWL Individuals) ค่าข้อมูล (Data Values) หรือตัวแปร (Variables)

ตัวอย่างของการเขียนกฎการอนุมาน

Learner(?x) ^ hasTopicSession(?x, ?y) ^ Ex-Easy(?y) ^ Exp-Easy(?y) ⇒ Beginner(?y)

จากกฎ หมายถึง เมื่อ *x* เป็นสมาชิกของคลาสผู้เรียน (Learner) และ *x* มีผลการทดสอบหัวข้อ (hasTopicSession) เป็น *y* โดย *y* เป็นสมาชิกของคลาสตัวอย่างระดับง่าย (Ex-Easy) และแบบฝึกปฏิบัติระดับง่าย (Exp-Easy) แล้ว ดังนั้นถือว่าในผลการทดสอบหัวข้อ *y* ของผู้เรียน x เป็นผู้เรียนระดับเริ่มต้น(Beginner) โดย

Learner, Ex-Easy, Exp-Easy, Beginner คือ หน่วยที่เป็นคลาส (OWL Named Classes)

hasTopicSession คือ หน่วยที่เป็นออปเจ็กต์ (OWL Object Properties)

?x, ?y คือ ตัวแปรที่แสดงค่าของข้อมูล (Variable of OWL Individuals)

นอกจากนี้ภาษาเอสดับบิวอาร์แอลยังมีหน่วย (Atom) ที่สามารถนำเข้ามาใช้งานได้เพิ่มเติม (SWRL Built-Ins Atom) ซึ่งเป็นนิพจน์ที่มีอาร์กิวเมนต์ได้มากกว่าหนึ่งอาร์กิวเมนต์ในการดำเนินการ ซึ่งนำหน้าด้วย swrlb ตัวอย่างคือ

TopicSession(?x) ^ PosttestScore(?x, ?PostScore) ^

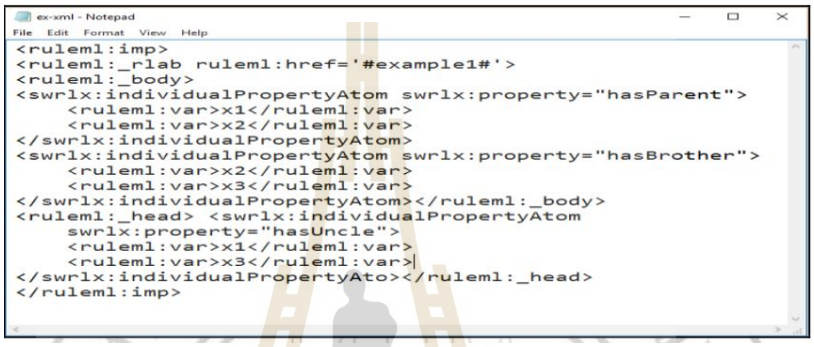
swrlb:greaterThanOrEqual(?PostScore, 1) ^

swrlb:lessThanOrEqual(?PostScore, 49) ^PosttestPeriod(?x, 1) -> Post-Weak(?x)

จากกฎ หมายถึง ผลการทดสอบหัวข้อ *x* ที่มีคะแนนทดสอบหลังเรียน (PosttestScore) มากกว่าหรือเท่ากับ 1 แต่น้อยกว่าหรือเท่ากับ 49 และมีระยะเวลาของการทำการทดสอบ (PosttestPeriod) อยู่ในเวลาที่กำหนด (1) แล้ว จะมีระดับผลการทดสอบหลังเรียนคืออ่อน (Post-Weak) ซึ่งการสร้างกฎด้วยภาษาเอสดับบิวอาร์แอลรูปแบบของภาษาที่เขียนนั้นมนุษย์จะสามารถเข้าใจได้ แต่จะมีการสั่งให้เครื่องสามารถปฏิบัติตามได้ในรูปแบบของภาษาเอ็กซ์เอ็มแอล ตัวอย่างเช่น

hasParent(?x1, ?x2) ^ hasBrother(?x2,?x3) ⇒hasUncle(?x1,?x3)

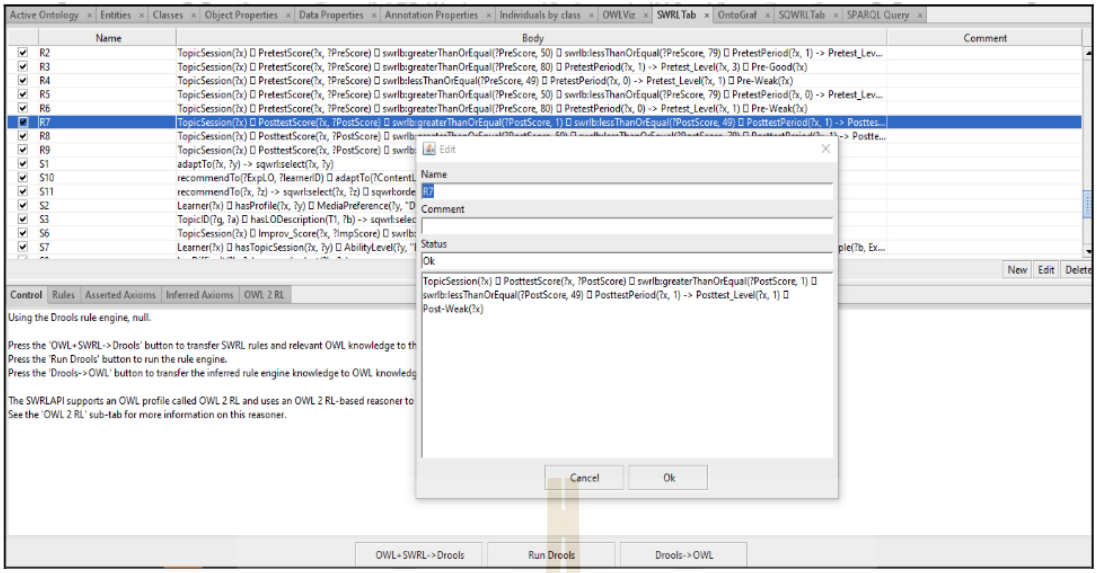
จะแสดงด้วยรูปแบบของภาษาเอ็กซ์เอ็มแอล ดังภาพที่ 2.8



ภาพที่ 2.8 รูปแบบของภาษาเอ็กซ์เอ็มแอลของการสร้างกฎ

การสร้างกฎด้วยภาษาเอสดับบิวแอลโดยใช้โปรแกรมโปรเทเจ (Protégé) มีเครื่องมือที่ช่วยในการทำงาน รวมถึงขั้นตอนของการนำกฎที่ได้ไปประมวลผลโดยผ่านเครื่องมือของการประมวลผลกฎ (Rule Engine) คือปุ่ม 'OWL+SWRL->Drools' ที่จะเปลี่ยนรูปกฎและความรู้ด้วยภาษาโอดับบิวแอลที่เกี่ยวข้องสู่เครื่องมือการประมวลผล เพื่อนำกฎไปใช้ในฐานความรู้ และปุ่ม 'Run Drools' ในการรันกฎและปุ่ม 'Drools->OWL' ในการเปลี่ยนรูปเพื่ออนุมานกฎเป็นฐานความรู้ต่อไป ดังแสดงหน้าจอตัวอย่างการทำงานของโปรแกรมโปรเทเจเวอร์ชัน 5 ดังภาพที่ 2.9

มหาวิทยาลัยเทคโนโลยีสุรนารี



ภาพที่ 2.9 หน้าจอการสร้างกฎด้วยภาษาเอสดับบิวอาร์แอลด้วยโปรแกรมโปรเทเจ

นอกจากนี้ ได้มีการพัฒนาภาษามาเพื่อสำหรับการสร้างกฎและการสืบค้นข้อมูลในฐานความรู้โดยไม่ต้องมีการบันทึกข้อมูลนั้นลงในฐานความรู้ แต่ให้เราสามารถนำความรู้นั้นมาแสดงผลได้ในทันที ซึ่งเป็นการรวมกันของภาษาการสืบค้นและภาษากฎไว้ในภาษาเดียวกัน ซึ่งในการสืบค้นภาษากฎสามารถทำได้โดยใช้ภาษาเอสคิวดับบิวอาร์แอล (Semantic Query Web Rule Language: SQWRL) มีลักษณะไวยากรณ์ของภาษาคล้ายกับภาษาสอบถามเชิงโครงสร้างรวมกับภาษากฎไว้ด้วยกัน ดังตัวอย่างด้านล่างที่แสดงภาษาเอสคิวดับบิวอาร์แอลในการสืบค้นผลการทดสอบของผู้เรียนที่มีความก้าวหน้าในการเรียน

```
TopicSession(?x) ^Improv_Score(?x, ?ImpScore) ^
swrlb:greaterThan(?ImpScore, 0) -> sqwrl:select(?x, ?ImpScore)
```

2.5 การวัดค่าความคล้ายคลึง

2.5.1 การวัดค่าความคล้ายคลึงเชิงมุมโคไซน์

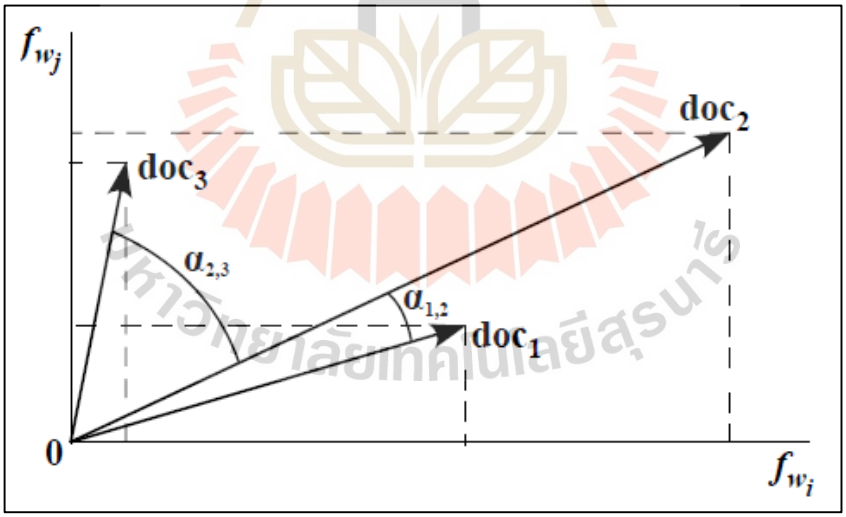
การวัดค่าความคล้ายคลึงเชิงมุมโคไซน์ (Cosine Similarity Measurement) เป็นวิธีการหาค่าความคล้ายคลึงตามทฤษฎีแบบจำลองเวกเตอร์สเปซ (Vector Space Model: VSM) ซึ่งใช้การแทนเอกสารด้วยระบบเวกเตอร์และหลักการทางสถิติในการวัด ดังสมการที่ 2.1

$$\text{Cosine Similarity} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \qquad \text{(สมการที่ 2.1)}$$

โดยที่ Similarity คือ ค่าคะแนนความคล้ายคลึงกัน

A_i	คือ เวกเตอร์ของเอกสาร A ค่าที่เป็นไปได้คือ 0 หรือ 1
B_i	คือ เวกเตอร์ของเอกสาร B ค่าที่เป็นไปได้คือ 0 หรือ 1
$\sum_{i=1}^{n} A_i B_i$	คือ ค่าผลรวมของเอกสาร A และเอกสาร B
$\sqrt{\sum_{i=1}^{n} A_i^2}$	คือ ค่ารากที่สองของผลรวมเวกเตอร์เอกสาร A ยกกำลัง 2
$\sqrt{\sum_{i=1}^{n} B_i^2}$	คือ ค่ารากที่สองของผลรวมเวกเตอร์เอกสาร B ยกกำลัง 2

การวัดค่าความคล้ายคลึงเชิงมุมโคไซน์เป็นวิธีการเปรียบเทียบความคล้ายคลึงของเอกสาร 2 เอกสาร โดยแต่ละเอกสารจะถูกแทนด้วยเวกเตอร์ขนาดเอ็น (N-Dimensional Vector) ซึ่งเก็บค่าน้ำหนักคำแต่ละคำในเอกสารนั้น (N-Dimensional Vector in Term Space) การเปรียบเทียบความคล้ายคลึงของเอกสารจะเปรียบเทียบโดยดูจากมุมโคไซน์ของมุมระหว่าง 2 เวกเตอร์ของเอกสาร หากเอกสารทั้งสองเอกสารคล้ายคลึงกันมากเวกเตอร์ของเอกสารทั้ง 2 จะทับกันเกือบสนิทมุมจึงมีค่าน้อย ค่าโคไซน์ที่ได้จะมีค่ามาก หรือค่าใกล้กับ 1 ในทางตรงกันข้าม หากเอกสารนั้นมีความแตกต่างกันมาก ค่าโคไซน์ที่ได้จะมีค่าน้อย หรือค่าเข้าใกล้กับ 0 เนื่องจากการซ้อนทับกันน้อยนั่นเอง ดังภาพที่ 2.10



ภาพที่ 2.10 การวัดค่าความคล้ายคลึงเชิงมุมโคไซน์

หมายเหตุ.จาก Text Mining with Machine Learning Principles and Techniques (หน้า 165), โดย Zˇizˇk a, Darˇena and Svoboda, 2020, Boca Raton: CRC Press.

จากภาพที่ 2.10 เทคนิคการวัดค่าความคล้ายคลึงเชิงมุมโคไซน์ (Cosine similarity) เป็นวิธีการหาคล้ายคลึงกันจากค่าความต่างของมุมของวัตถุสองวัตถุที่เกิดขึ้นบนพื้นที่เวกเตอร์ วิธีการนี้เป็นที่นิยมและมีประสิทธิภาพสูงในการวัดความคล้ายคลึงระหว่างวัตถุสองวัตถุ หรือระหว่างเอกสารทั้งสองเอกสารและถูกนำมาประยุกต์ใช้กับศาสตร์ทางด้านการค้นคืนข้อมูลอย่างแพร่หลาย เนื่องจากวิธีการนี้จะมีประสิทธิภาพในกรณีที่เอกสารสองเอกสารมีความยาวไม่เท่ากัน หรือทำให้มีความยุติธรรมต่อเอกสารที่สั้นกว่า ซึ่งค่าความคล้ายคลึงกันแบบโคไซน์นี้จะมีค่าอยู่ระหว่าง 0-1 เท่านั้น โดยแต่ละเอกสารจะถูกแทนด้วยเวกเตอร์ซึ่งเก็บค่าน้ำหนักคำแต่ละคำในเอกสารนั้น การเปรียบเทียบความคล้ายคลึงของเอกสารจะพิจารณาจากมุมโคไซน์ของมุมระหว่างสองเวกเตอร์ของเอกสาร ทั้งนี้ค่าความคล้ายคลึงสูงสุดที่วัดด้วยวิธีนี้จะมีค่าเท่ากับ 1 ความหมายคือเวกเตอร์ทั้งสองทำมุมระหว่างกัน 0 องศา นั้นคือเวกเตอร์ทั้งสองมีทิศทางเดียวกันหรือมีความคล้ายคลึงกัน 100 เปอร์เซ็นต์ และถ้าค่าความคล้ายคลึงมีค่าเท่ากับ 0 เวกเตอร์ทั้งสองจะทำมุมระหว่างกัน 90 องศา หมายถึงเวกเตอร์ทั้งสองหรือเอกสารทั้งสองไม่มีความคล้ายคลึงกันเลย

ตัวอย่างกระบวนการการวัดค่าความคล้ายคลึงเชิงมุมโคไซน์ เพื่อเปรียบเทียบความคล้ายคลึงกันของ 2 เอกสาร ได้แก่

Document 1: Deep Learning can be hard

Document 2: Deep Learning can be simple

ขั้นตอนแรก คือการแปลงข้อความให้อยู่ในรูปแบบเวกเตอร์ ประกอบด้วย 0 และ 1 โดยการเปรียบเทียบเนื้อหาระหว่างเอกสาร โดยการนับความถี่ของคำ (Term Frequency) ที่ปรากฏในแต่ละเอกสารที่นำมาเปรียบเทียบกัน (Z̆iz̆k a, Dar̆ena and Svoboda, 2020) จากนั้นนำคำหรือดัชนีทั้งหมดในเอกสารทั้ง 2 มารวมกัน ถ้าเอกสารมีคำเหมือนกันให้นำมาสร้างเป็นดัชนีเพียงคำเดียว และนำความถี่ของคำในแต่ละเอกสารเข้าเวกเตอร์ทั้ง 2 เวกเตอร์ จากนั้นทำการคำนวณค่าความคล้ายคลึงมุมโคไซน์ได้ ดังตารางที่ 2.4

ตารางที่ 2.4 ตัวอย่างข้อมูลสำหรับคำนวณค่าความคล้ายคลึงมุมโคไซน์

Word	Document1	Document2
1. Deep	1	1
2. Learning	1	1
3. can	1	1
4. be	1	1
5. hard	1	0
6. simple	0	1

จากตารางที่ 1 ผลลัพธ์แปลงข้อความให้อยู่ในรูปแบบเวกเตอร์ คือ

Document 1: [1, 1, 1, 1, 1, 0] กำหนดให้แทนด้วย A

Document 2: [1, 1, 1, 1, 0, 1] กำหนดให้แทนด้วย B

จะได้เอกสารที่อยู่ในรูปแบบเวกเตอร์ 2 เอกสาร คือ Document1 (A) และ Document (B) มีขนาดเท่ากับ 6 เวกเตอร์

ขั้นตอนที่ 2 คือการวัดค่าความคล้ายคลึงเชิงมุมโคไซน์ จากสูตร

Cosine similarity (CS) = (A . B) / (||A|| ||B||)

คำนวณหา dot product ระหว่าง A และ B:

1.1 + 1.1 + 1.1 + 1.1 + 1.0 + 0.1 = 4

คำนวณหาขนาดของเวกเตอร์ A:

$\sqrt{1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 0^2}$ = 2.2360679775

คำนวณหาขนาดของเวกเตอร์ B:

$\sqrt{1^2 + 1^2 + 1^2 + 1^2 + 0^2 + 1^2}$ = 2.2360679775

คำนวณหาค่าความคล้ายคลึงเชิงมุมโคไซน์:

(4) / (2.2360679775*2.2360679775) = 0.80

ดังนั้นค่าความคล้ายคลึงเชิงมุมโคไซน์ระหว่างเอกสารที่ 1 และเอกสารที่ 2 เท่ากับ 80%

2.5.2 การวัดค่าสัมประสิทธิ์ความคล้ายคลึงของแจ็กการ์ด

ค่าสัมประสิทธิ์ความคล้ายคลึงของแจ็กการ์ด (Jaccard similarity coefficient) (Jaccard, 1908, 1912) เป็นสถิติที่ประยุกต์แนวคิดในทฤษฎีเซต (Set Theory) เพื่อนำมาใช้เปรียบเทียบความคล้ายคลึงและความหลากหลายของกลุ่มตัวอย่าง เมื่อแรกเริ่มค่าสัมประสิทธิ์ความคล้ายคลึงของแจ็กการ์ดถูกเสนอขึ้นเพื่อเปรียบเทียบความหลากหลายในเชิงพฤกษศาสตร์ ต่อมาจึงแพร่หลายไปสู่วงการอื่น ๆ โดยเฉพาะอย่างยิ่งในงานค้นคืนสารสนเทศ แนวคิดของค่าสัมประสิทธิ์ความคล้ายคลึงของแจ็กการ์ดคือการวัดค่าความคล้ายคลึงระหว่างกลุ่มประชากร 2 กลุ่ม โดยคำนวณจากอินเตอร์เซกชัน (Intersection) ในทฤษฎีเซต กล่าวคือขนาดของประชากรที่ทั้งสองกลุ่มตัวอย่างมีร่วมกัน หารด้วยยูเนียน (Union) ในทฤษฎีเซต ซึ่งได้แก่ขนาดของประชากรทั้งหมดจากทั้งสองกลุ่มตัวอย่าง ดังแสดงในสมการที่ 2.2

$$\text{Jaccard similarity coefficient (A,B)} = \frac{|\boldsymbol{A} \cap \boldsymbol{B}|}{|\boldsymbol{A} \cup \boldsymbol{B}|} \qquad \text{(สมการที่ 2.2)}$$

ทั้งนี้ค่าสัมประสิทธิ์ความคล้ายคลึงของแจ็กการ์ดจะมีค่าระหว่าง 0 ถึง 1 หากเอกสารทั้งสองมีความคล้ายคลึงกันมาก ค่าสัมประสิทธิ์จะสูงขึ้นจนเข้าใกล้ 1 ในทางตรงกันข้าม หากค่าสัมประสิทธิ์มีค่าน้อย หรือเข้าใกล้ 0 มากเท่าใด หมายถึงเอกสารทั้งสองไม่มีความคล้ายคลึงกันเลย เนื่องจากไม่มีคำที่ตรงกันในทั้งสองเอกสาร

ตัวอย่างกระบวนการการวัดค่าสัมประสิทธิ์ความคล้ายคลึงของแจ็กการ์ด เพื่อเปรียบเทียบความคล้ายคลึงกันของ 2 เอกสาร ได้แก่

Document 1: Deep Learning can be hard

Document 2: Deep Learning can be simple

ขั้นตอนแรก แยกคำในเอกสารทั้งสองออกให้เป็นอิสระต่อกัน แล้วแทนให้อยู่ในรูปเซท

Document 1 = {'Deep ', 'Learning', 'can', 'be', 'hard'} *แทนด้วย A*

Document 2 = {'Deep ', 'Learning', 'can', 'be', 'simple'} *แทนด้วย B*

ขั้นตอนที่ 2 ทำการคำนวณหาค่าอินเตอร์เซกชัน (Intersection) และยูเนียน (Union) ของเอกสาร 1 (A) และเอกสาร 2 (B) เพื่อคำนวณหาค่าสัมประสิทธิ์ความคล้ายคลึงของแจ็กการ์ดระหว่างสองเอกสาร ด้วยการแทนค่าในสมการที่ 2.2

Jaccard similarity coefficient (A,B) =

$$\frac{\{\text{Deep},\text{Learning},\text{can},\text{be},\text{hard}\} \cap \{\text{Deep},\text{Learning},\text{can},\text{be},\text{simple}\}}{\{\text{Deep},\text{Learning},\text{can},\text{be},\text{hard}\} \cup \{\text{Deep},\text{Learning},\text{can},\text{be},\text{simple}\}}$$

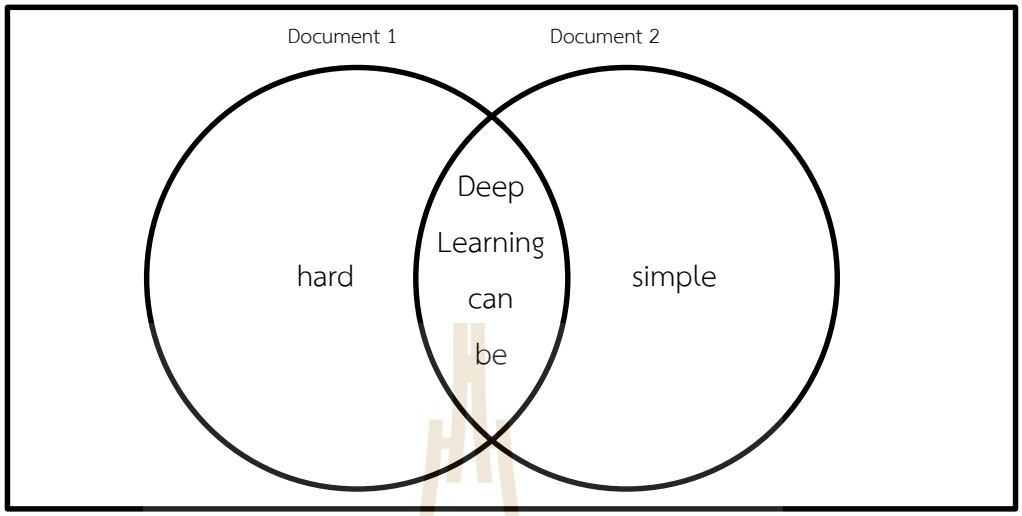
$$= \frac{\{\text{Deep},\text{Learning},\text{can},\text{be}\}}{\{\text{Deep},\text{Learning},\text{can},\text{be},\text{hard},\text{simple}\}}$$

$$= \frac{4}{6}$$

$$= 0.6667$$

ดังนั้นค่าสัมประสิทธิ์ความคล้ายคลึงของแจ็กการ์ดระหว่างเอกสารที่ 1 และเอกสารที่ 2 เท่ากับ 66.67% ดังแสดงได้ดังภาพที่ 2.11 แสดงตัวอย่างความคล้ายคลึงกันของสองเอกสาร

มหาวิทยาลัยเทคโนโลยีสุรนารี



ภาพที่ 2.11 ตัวอย่างความคล้ายคลึงของสองเอกสาร

ในการวัดความคล้ายคลึงของสิ่งของต่าง ๆ จะมีหลากหลายวิธี ซึ่งในงานวิจัยนี้จะใช้วิธีการวัดความคล้ายคลึงของแจ็กการ์ดในการหาความคล้ายคลึงระหว่างคำค้นจากผู้ใช้กับคำสำคัญที่เกี่ยวข้องกับรายการความรู้ และจะมีการใช้การวัดค่าความคล้ายคลึงเชิงมุมโคไซน์มาใช้การวัดความคล้ายคลึงของรายการความรู้กับรายการคำที่ได้จากคำค้นของผู้ใช้

2.6 การประเมินประสิทธิภาพของการค้นหา

การวัดความสามารถในการค้นคืนสารสนเทศแบบประเมินประสิทธิภาพเอกสารที่ถูกเลือก แต่มิได้เรียงลำดับความคล้ายคลึง (Manning, C.D., Raghavan, P., and Schütze, H., 2008) ซึ่งหาได้จากตาราง Confusion Matrix แสดงผลการประเมินผลลัพธ์การทำนายกับผลลัพธ์จริง ๆ ที่โมเดลจำแนกได้ ดังตารางที่ 2.5

ตารางที่ 2.5 Confusion Matrix

	Predict class	
Actual class	Class=yes	Class=no
Class=Yes	TP	FN
Class=No	FP	TN

จากตารางที่ 2.5 เมื่อมีการทำนาย 2 ประเภท ผลการทำนายทั้งหมดที่เป็นไปได้จะมี 4 ชนิด ดังนี้

- TP (True Positive) หมายถึง ค่าคลาสเป้าหมายคือ Yes แบบโมเดลทำนายว่า Yes
- FN (False Negative) หมายถึง ค่าคลาสเป้าหมายคือ Yes แบบโมเดลทำนายว่า No
- FP (False Positive) หมายถึง ค่าคลาสเป้าหมายคือ No แบบโมเดลทำนายว่า Yes
- TN (True Negative) หมายถึง ค่าคลาสเป้าหมายคือ No แบบโมเดลทำนายว่า No

จากนั้นนำมาคำนวณหา 3 ค่าที่จำเป็นต่อการวัดประสิทธิภาพโมเดล ประกอบด้วย

1) ค่าความถูกต้อง (Accuracy) คือ จำนวนข้อมูลที่ทำนายถูกโดยพิจารณารวมทุกคลาส

2) ความแม่นยำ (Precision) คือ ค่าของตัวแบบที่ทำนายให้ถูกต้องที่บอกถึงประสิทธิภาพของระบบในการค้นคืนเอกสารโดยดูจากอัตราส่วนของจำนวนเอกสารที่ถูกต้องจากเอกสารที่ถูกเลือกมาทั้งหมด โดยพิจารณาแยกทีละคลาส

3) ค่าความระลึก (Recall) คือ จำนวนการกระทำด้วยกันแบบที่ตรงกับความเป็นจริง เพื่อประสิทธิภาพของการค้นคืนเอกสารโดยดูจากอัตราส่วนจำนวนเอกสารที่ถูกต้องที่เลือกมาต่อจำนวนเอกสารที่ถูกต้องทั้งหมด โดยพิจารณาแยกทีละคลาส ดังสมการที่ 2.3, 2.4 และ 2.5

$$\text{Accuracy} = \frac{\mathbf{TP+TN}}{\mathbf{TP+FP+TN+FN}} \qquad \text{(สมการที่ 2.3)}$$

$$\text{Precision} = \frac{\mathbf{TP}}{\mathbf{TP+FP}} \qquad \text{(สมการที่ 2.4)}$$

$$\text{Recall} = \frac{\mathbf{TP}}{\mathbf{TP+FN}} \qquad \text{(สมการที่ 2.5)}$$

อย่างไรก็ดีค่าความถูกต้องและค่าความระลึกเป็นประโยชน์ต่อผู้ใช้ต่างกลุ่มกันไป โดยผู้ใช้จะให้ความสำคัญกับค่าความถูกต้องมากกว่าค่าความครบถ้วน เนื่องจากผู้ใช้ส่วนใหญ่ต้องการให้ข้อมูลที่เป็นผลลัพธ์ที่ตรงกับสิ่งที่ต้องการค้นหา แต่ในทางกลับกัน ผู้ที่พัฒนาระบบค้นหาข้อมูลจะเน้นไปที่ค่าความระลึก เนื่องจากต้องการให้ระบบค้นหาและเลือกเอกสารที่ถูกต้องให้ได้มากที่สุด ดังนั้นอีกค่าหนึ่งที่นำมาใช้ในการรวมค่าความถูกต้องและค่าความระลึกเข้าด้วยกันโดยหาค่าเฉลี่ยที่ดีที่สุด (Weighted Harmonic Mean) เรียกว่า ค่าอัตรารู้จำ (F-measure) ที่เกิดจากการเปรียบเทียบกันระหว่างค่าความแม่นยำและค่าความระลึกของแต่ละคลาสเป้าหมาย เพื่อเป็นมาตรวัดเดี่ยว (Single Metric) ที่วัดความสามารถของโมเดลแทนค่าความแม่นยำและค่าความระลึก ดังสมการที่ 2.6

มหาวิทยาลัยเทคโนโลยีสุรนารี

$$\text{F-measure} = 2 * \frac{\textbf{Precision} * \textbf{Recall}}{\textbf{Precision} + \textbf{Recall}} \qquad \text{(สมการที่ 2.6)}$$

ในการประเมินประสิทธิภาพการค้นคืนของงานวิจัยนี้ จะใช้แนวทางในการประเมินด้วยการหาค่าความแม่นยำ (Precision) ค่าความละรึก (Recall) และค่าอัตรารู้จำ (F-measure) ซึ่งเป็นค่าที่ใช้กันโดยทั่วไปในการประเมินสิทธิภาพของการค้นคืนต่าง ๆ

2.7 เอกสารและงานวิจัยที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้องเพื่อให้การค้นคืนข้อมูลจากฐานความรู้เป็นไปอย่างมีประสิทธิภาพ เพื่อการค้นคืนความรู้เป็นไปอย่างรวดเร็ว ถูกต้อง และตรงตามความต้องการของผู้ใช้จะแบ่งออกเป็น 3 ประเด็นหลัก ได้แก่ 1) การพัฒนาฐานความรู้ 2) การค้นหาเชิงความหมาย และ 3) การประเมินประสิทธิภาพการค้นคืน

2.7.1 การพัฒนาฐานความรู้

การพัฒนาฐานความรู้สำหรับการค้นคืนในปัจจุบันจะเน้นการออกแบบและพัฒนาออนโทโลยี ซึ่งพบว่ามีการนำออนโทโลยีมาใช้ในขอบเขตงานหลากหลายสาขาวิชาชีพ ดังเช่น การเกษตร การศึกษา การจัดการสำนักงาน และการแพทย์ เป็นต้น ดังจะเห็นได้จากการนำมาประยุกต์ใช้จัดเก็บความรู้ด้านการเกษตร โดยนำออนโทโลยีมามาประยุกต์ใช้ในการเป็นแหล่งเชื่อมโยงไปยังเนื้อหาครอบคลุมข้อมูลในขอบเขตด้านผลไม้ของประเทศไทย ได้แก่ ลักษณะทางพฤกษศาสตร์ การปลูกและการบำรุงรักษา โรคและแมลง การแปรรูป การตลาดและส่งออก (ศุภกฤษฎิ์ นิวัฒนากูล, 2556) การออกแบบและพัฒนาออนโทโลยีสำหรับรวบรวมความรู้สำหรับการผลิตลำไยคุณภาพสูงซึ่งได้กำหนดเป็น 7 ระยะ ได้แก่ ระยะเตรียมการ ระยะก่อนออกดอก ระยะออกดอก ระยะติดผล ระยะให้ผลช่วงแรก ระยะให้ผลช่วงกลาง และระยะเก็บเกี่ยว (Bangkhomned and Payakpate, 2020) นอกจากนี้ยังพบโครงการ FoodOn ที่นำออนโทโลยีไปประยุกต์ใช้ในการจัดเก็บความรู้สำหรับอาหารที่ประกอบด้วยทั้งพืชและสัตว์ และยังสามารถให้มีการแบ่งปันการใช้ข้อมูลร่วมกันได้อีกด้วย (Dooley, Griffiths, Gosal, et al., 2018)

2.7.2 การค้นคืนเชิงความหมาย

- **แนวคิดการค้นคืน**

การค้นคืนจากฐานความรู้ที่ออกแบบในรูปแบบของออนโทโลยีในปัจจุบันได้มีการนำเสนอการค้นคืนตามแนวคิด Concept-based ซึ่งเป็นการค้นคืนด้วยคลาสหรือคอนเซ็ปต์ รวมถึงคุณสมบัติของสิ่งต่าง ๆ ที่ได้ออกแบบไว้เพื่อช่วยการค้นคืนที่ไม่ได้ขึ้นอยู่กับคำที่ปรากฏในเอกสารดังที่ผ่านมา ดังจะเห็นได้จากงานวิจัยของ Bangkhomned and Payakpate (2020) และ

Formica, Pourabbas, and Taglino (2020) ที่มีการนำเสนอรูปแบบการค้นคืนโดยการสร้างส่วนต่อประสานผู้ใช้ให้เลือกการค้นหาตามโครงสร้างของคลาสและคุณสมบัติที่ได้ออกแบบสำหรับการจัดเก็บความรู้ในโดเมนต่าง ๆ นอกจากนี้ยังมีระบบ QSST (Quran Semantic Search Tool) ซึ่งนำเสนอโดย Ensaf and Eyad (2020) ที่นำเสนอการค้นคืนแบบ Concept-based สำหรับการค้นคืนในออนโทโลยีคัมภีร์กุรอ่าน

- **กระบวนการค้นคืน**

เพื่อให้การค้นหาได้อย่างรวดเร็วและตรงตามความต้องการมากขึ้น ได้มีการจัดทำคลังคำเพื่อให้สามารถนำคำสำคัญจากผู้ใช้มาเปรียบเทียบกับคำสำคัญที่จัดเก็บไว้ ซึ่งกระบวนการนี้เป็นการจัดความรู้ให้เป็นระบบ โดยตัวแบบที่นำมาใช้สำหรับการจัดความรู้ให้เป็นระบบ ได้แก่ เอสเคโอเอส ดังจะเห็นได้จากงานวิจัยของศุภกฤษฎ์ นิวัฒนากูล (2556) มีการนำโครงสร้างอรรถาภิธานศัพท์ (Thesaurus) ไปใช้ในการจัดทำคลังคำ โดยใช้รายการคำจากอรรถาภิธานศัพท์เกษตรไทย

ในกระบวนการค้นคืนได้มีการพัฒนาเฟรมเวิร์คสำหรับการค้นหาเพื่อนำมาใช้เป็นเครื่องมือสำหรับการจัดการและค้นหาข้อมูลในออนโทโลยี ดังเช่น การค้นหาข้อมูลในออนโทโลยีด้วยภาษา SPQRQL บนพื้นฐานของเฟรมเวิร์ค OAM (Ontology Application Management Framework) โดยจะทำการสร้างส่วนต่อประสานผู้ใช้สำหรับการค้นหาตามคลาสและคุณสมบัติที่ได้ออกแบบไว้ (Bangkhomned and Payakpate, 2020) ซึ่งสอดคล้องกับงานวิจัยของ ทิพวิมล ชมภูคำ. (2020) ที่ได้ออกแบบออนโทโลยีและการค้นหาเชิงความหมายสำหรับโรคติดเชื้อในเด็กโดยทำงานบนพื้นฐานของเฟรมเวิร์ค OAM เช่นเดียวกัน

จากงานวิจัยที่ค้นพบมีการจัดเรียงลำดับผลลัพธ์ตามความคล้ายคลึงระหว่างเอกสารกับคำสำคัญ ซึ่งมีเทคนิควิธีการคำนวณที่หลากหลาย เช่น การวัดความคล้ายคลึงแบบโคซายน์ (Cosine Similarity) ซึ่งเป็นวิธีการหาคล้ายคลึงกันจากค่าความต่างของมุมของข้อมูลที่เกิดขึ้นบนพื้นที่เวกเตอร์สเปซดังงานวิจัยของ Ensaf and Eyad (2020) และการวัดความคล้ายคลึงด้วยสัมประสิทธิ์ของแจ็คการ์ด (Jaccard) ซึ่งเป็นวิธีการเปรียบเทียบ หรือระบุความใกล้เคียงของกลุ่มข้อมูลในกระบวนการจัดกลุ่มข้อมูล โดยไม่ต้องใช้ค่าตัวแปรความถี่ของข้อมูล ดังงานวิจัยของศุภกฤษฎ์ นิวัฒนากูล (2556) นอกจากนี้ยังพบว่ามีการกำหนดวิธีการคำนวณสำหรับงานเป็นการเฉพาะ ดังเช่นวิธีการ $SemSim^{e}$ (Formica, Pourabbas, and Taglino, 2020) ที่มีการคำนวณค่าความคล้ายคลึงโดยมีการกำหนดค่าน้ำหนักของคำ ซึ่งใช้ทดสอบสำหรับการคำนวณทางด้านโดเมนการท่องเที่ยว

2.7.3 การประเมินประสิทธิภาพการค้นคืน

การประเมินประสิทธิภาพการค้นคืนข้อมูล สารสนเทศ หรือความรู้จากออนโทโลยีที่พัฒนาขึ้นมา ที่พบในงานวิจัยส่วนใหญ่จะใช้การประเมินประสิทธิภาพด้วยการคำนวณค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าอัตรารู้จำ (F-measure) ดังจะเห็นได้จากงานวิจัยของ

Bangkhomned and Payakpate (2020) ที่เป็นการค้นหาความรู้จากออนโทโลยีการผลิตลำไยคุณภาพสูง งานวิจัยของ Formica, Pourabbas, and Taglino (2020) สำหรับการค้นหาความรู้จากออนโทโลยีการท่องเที่ยว งานวิจัยของ Ensaf and Eyad (2020) สำหรับการค้นหาความรู้จากออนโทโลยีคัมภีร์กุรอ่าน งานวิจัยของ ทิพวิมล ชมภูคำ (2020) สำหรับการค้นหาความรู้จากออนโทโลยีโรคติดเชื้อในเด็ก และงานวิจัยของ ศุภกฤษฎ์ นิวัฒนากูล (2556) ที่ค้นหาความรู้จากออนโทโลยีผลไม้เศรษฐกิจของไทย และพบว่ามีบางงานวิจัยที่มีการประเมินประสิทธิภาพการค้นคืนของระบบที่พัฒนาขึ้นด้วยการสอบถามความคิดเห็นจากผู้เกี่ยวข้องกับระบบ ดังงานวิจัยของ ของทิพวิมล ชมภูคำ (2020) และศุภกฤษฎ์ นิวัฒนากูล (2556)

2.8 กรอบแนวคิดการวิจัย

กรอบแนวคิดการวิจัยผู้วิจัยนำเสนอในรูปแบบการเปรียบเทียบงานวิจัยที่เกี่ยวข้อง ซึ่งสามารถจำแนกได้ 3 ประเด็น คือ 1) การพัฒนาฐานความรู้ 2) การค้นคืนเชิงความหมาย ซึ่งประกอบด้วยแนวคิดการค้นคืน และกระบวนการค้นคืน และ 3) การประเมินประสิทธิภาพการค้นคืน แสดงดัง

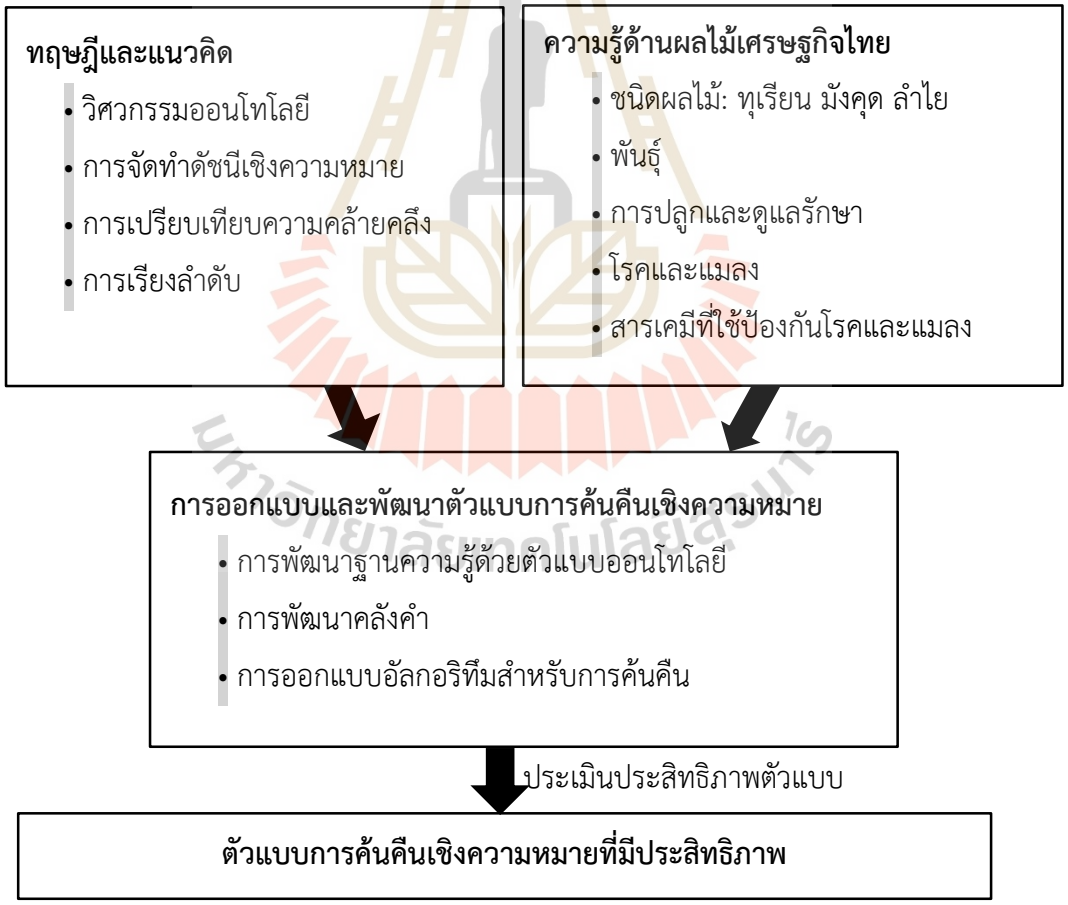
ตารางที่ 2.6 สรุปเปรียบเทียบงานวิจัยที่เกี่ยวข้องตัวแบบค้นคืนเชิงความหมาย

ผู้วิจัย	การพัฒนาฐานความรู้	การค้นคืนเชิงความหมาย		การประเมินประสิทธิภาพการค้นคืน
		แนวคิดการค้นคืน	กระบวนการค้นคืน	
Bangkhomned and Payakpate (2020)	ออนโทโลยีการผลิตลำไยคุณภาพสูง	Concept-based searching	ค้นหาด้วยภาษา SPARQL ด้วยเฟรมเวิร์ค OAM	Precision, recall and F-measure
Formica, Pourabbas, and Taglino (2020)	ออนโทโลยีการท่องเที่ยว	Concept-based searching	ระบุน้ำหนักคำ และกำหนดวิธีการ SemSim^{e} สำหรับการคำนวณความคล้ายคลึง	Precision and recall
Ensaf and Eyad (2020)	ออนโทโลยีคัมภีร์กุรอ่าน	Concept-based searching	คำนวณความคล้ายคลึงด้วย Cosine similarity	Precision, recall and F-measure
ทิพวิมล ชมภูคำ. (2020)	ออนโทโลยีโรคติดเชื้อในเด็ก	Concept-based searching	ค้นหาด้วยภาษา SPARQL ด้วยเฟรมเวิร์ค OAM	Precision, recall and F-measure
ศุภกฤษฎ์ นิวัฒนากูล (2556)	ออนโทโลยีผลไม้เศรษฐกิจไทย	Concept-based searching	จัดการคลังคำด้วย SKOS และคำนวณค่าความคล้ายคลึงของผลลัพธ์ด้วย Jaccard similarity	Precision, recall and F-measure

ตารางที่ 2.6 สรุปเปรียบเทียบงานวิจัยที่เกี่ยวข้องตัวแบบค้นคืนเชิงความหมาย

ผู้วิจัย	การพัฒนาฐานความรู้	การค้นคืนเชิงความหมาย		การประเมินประสิทธิภาพการค้นคืน
		แนวคิดการค้นคืน	กระบวนการค้นคืน	
ผู้วิจัย	ออนโทโลยีผลไม้เศรษฐกิจไทย	Concept-based and keyword-based searching	จัดการคลังคำด้วย SKOS และ SWRL การหาความคล้ายคลึงของคำค้นด้วย Jaccard similarity และคำนวณค่าความคล้ายคลึงของผลลัพธ์ด้วย Cosine coefficient	Precision, recall and F-measure

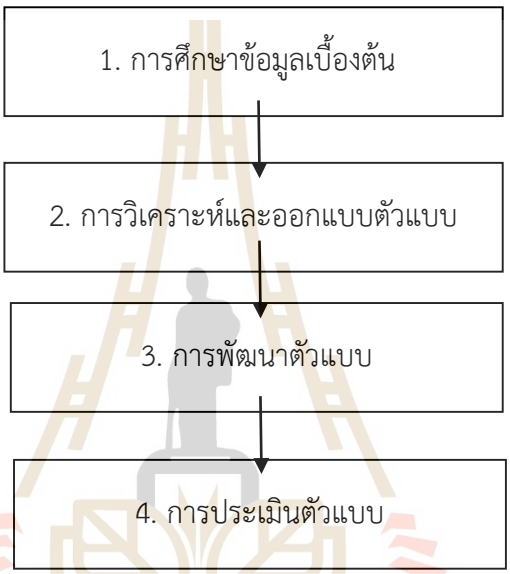
จากตารางที่ 2.6 ผู้วิจัยนำเสนอในรูปแบบการเปรียบเทียบงานวิจัยที่เกี่ยวข้องกับงานวิจัยนี้ ซึ่งมีการนำเอาขั้นตอนวิธีที่หลากหลายมาใช้สำหรับการออกแบบและพัฒนาตัวแบบค้นคืนเชิงความหมายสำหรับผลไม้เศรษฐกิจไทย ซึ่งสามารถสรุปเป็นกรอบแนวคิดแสดงดังภาพที่ 2.12 ดังนี้



ภาพที่ 2.12 กรอบแนวคิดการพัฒนาตัวแบบการค้นคืนเชิงความหมาย

บทที่ 3 วิธีการดำเนินการวิจัย

การดำเนินการวิจัยในครั้งนี้ เป็นการวิจัยและพัฒนา โดยมีการประยุกต์ใช้ทฤษฎี หลักการและเทคนิคที่เกี่ยวข้องกับการค้นหาเชิงความหมาย การจัดการฐานความรู้ด้วยออนโทโลยี และการค้นหาความรู้ในออนโทโลยีด้วยภาษา SPARQL และ SWRL มาประยุกต์ใช้ให้เหมาะสมต่อการพัฒนาแบบจำลองการค้นคืนเชิงความหมายสำหรับผลไม้เศรษฐกิจไทยควบคู่กับการพัฒนาระบบตามกระบวนการของวงจรการพัฒนาระบบ System Development Life Cycle (SDLC) โดยมีลำดับขั้นตอนในการวิจัยดังภาพที่ 3.1 ดังนี้



ภาพที่ 3.1 ขั้นตอนการดำเนินการวิจัย

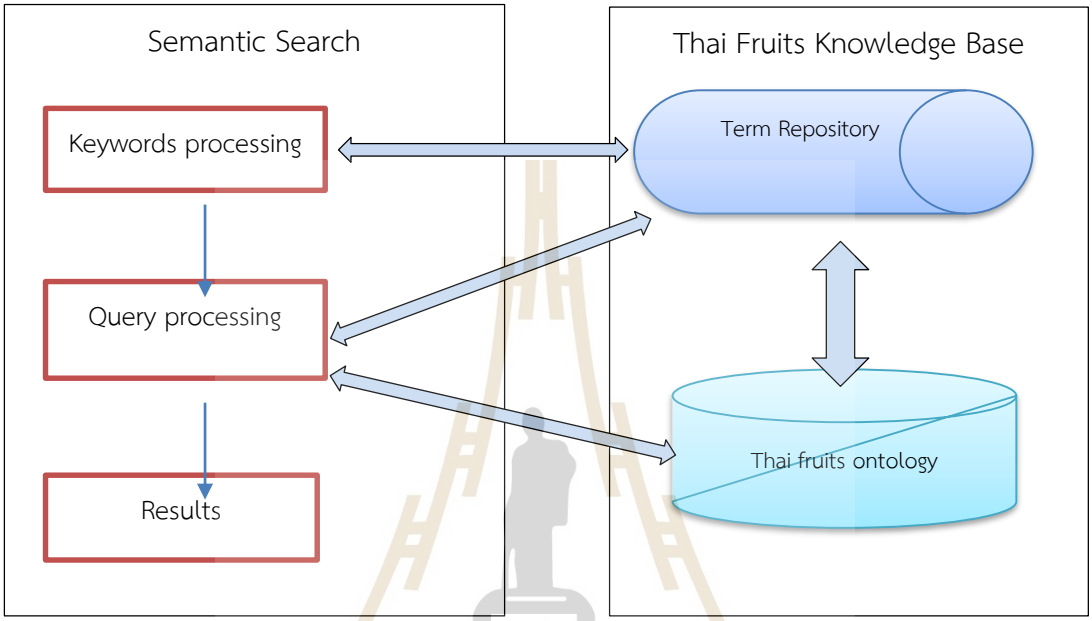
3.1 การศึกษาข้อมูลเบื้องต้น

การศึกษาข้อมูลเบื้องต้นจากเอกสารที่เกี่ยวข้องจากหนังสือ เอกสาร และข้อมูลต่าง ๆ บนอินเทอร์เน็ต ประกอบด้วย ข้อมูลการค้นคืนเชิงความหมาย ข้อมูลผลไม้เศรษฐกิจไทย และข้อมูลการวิจัยที่เกี่ยวข้อง เพื่อนำข้อมูลไปใช้ในการวิเคราะห์และออกแบบฐานความรู้และออกแบบตัวแบบ

การศึกษาความรู้ทางด้านผลไม้เศรษฐกิจไทยสำหรับงานวิจัยนี้ ได้ใช้ออนโทโลยีด้านผลไม้เศรษฐกิจที่วิจัยและพัฒนาขึ้นโดย ศุภกฤษฎิ์ นิวัฒนากูล (2556) มาประยุกต์ใช้ในการออกแบบฐานความรู้ที่อยู่ในรูปแบบภาษาโอดับเบิ้ลยูแอล ซึ่งมีโครงสร้างของฐานความรู้ผลไม้เศรษฐกิจไทยที่ประกอบไปด้วยข้อมูล พันธุ์ การปลูกและดูแลรักษา โรคและแมลง สารเคมีที่ใช้ป้องกันโรคและแมลง ซึ่งเป็นข้อมูลที่ผู้ใช้ส่วนใหญ่สนใจในการค้นหาข้อมูล (ศุภกฤษฎิ์ นิวัฒนากูล และฉัตรภัสร์ ฐิติอัคราวงศ์, 2560)

3.2 การวิเคราะห์และออกแบบตัวแบบ

เป็นการวิเคราะห์และออกแบบตัวแบบ จะเริ่มจากการการออกแบบโครงสร้างฐานความรู้ การออกแบบหน้าจอนำเข้าข้อมูลจากผู้ใช้ หน้าจอผลลัพธ์ของระบบ รวมถึงการออกแบบขั้นตอนวิธี (Algorithm) ของการทำงานในส่วนต่าง ๆ ของตัวแบบค้นหาเชิงความหมาย ซึ่งมีกรอบการทำงานของตัวแบบค้นหาเชิงความหมายในงานวิจัยนี้ มีส่วนประกอบสำคัญ ดังภาพที่ 3.2



ภาพที่ 3.2 กรอบการทำงานของตัวแบบการค้นคืนเชิงความหมาย

จากภาพที่ 3.2 การทำงานของตัวแบบค้นคืนเชิงความหมายออนไลน์สำหรับผลไม้เศรษฐกิจไทย ประกอบด้วย 2 ส่วน ดังนี้

1. ฐานความรู้ผลไม้เศรษฐกิจไทย (Thai Fruits Knowledge Base) เป็นการเตรียมข้อมูลเบื้องต้นสำหรับการจัดเก็บความรู้และการจัดทำความรู้ที่จัดเก็บไว้ให้เป็นระบบเพื่อสนับสนุนการค้นคืนเชิงความหมาย ประกอบด้วย

1.1 คลังคำ (Term repository) ใช้ตัวแบบ Simple knowledge organization system (SKOS) สำหรับสร้างคลังคำ คำศัพท์ที่จัดเก็บจะนำมาจากฐานความรู้ผลไม้เศรษฐกิจไทย โดยนำมาจำแนกความสัมพันธ์ตามลำดับชั้น จัดกลุ่มตามความสัมพันธ์ และมีการเชื่อมโยงระหว่างคำศัพท์ที่มีความสัมพันธ์กัน รวมถึงกำหนดคุณสมบัติเพื่อเชื่อมความสัมพันธ์ระว่างคลังคำกับความรู้ในออนโทโลยีผลไม้เศรษฐกิจไทย

1.2 ออนโทโลยีผลไม้เศรษฐกิจไทย (Thai fruits ontology) ได้นำออนโทโลยีผลไม้เศรษฐกิจของ ศุภกฤษฎิ์ นิวัฒนากูล (2556) มาเป็นออนโทโลยีเริ่มต้น จากนั้นทำการปรับปรุงและประยุกต์ใช้เป็นฐานความรู้ที่จัดเก็บความรู้จากเอกสารและข้อมูลจากแหล่งต่าง ๆ ของผลไม้เศรษฐกิจ

ไทย 3 ชนิด ได้แก่ ทุเรียน มังคุด และลำไย โดยมีการจัดเก็บข้อมูลในด้าน พันธุ์ การปลูกและดูแลรักษา โรคและแมลง สารเคมีที่ใช้ป้องกันโรคและแมลง ซึ่งเป็นข้อมูลที่ผู้ใช้ส่วนใหญ่สนใจในการค้นหา (ศุภกฤษฎิ์ นิวัฒนากูล และฉัตรภัสร์ ฐิติอัคราวงศ์, 2560) การออกแบบออนโทโลยีผลไม้เศรษฐกิจไทย ใช้แนวทางการออกแบบตามกระบวนการวิศวกรรมออนโทโลยี (Ontology engineering) ตามแนวคิดของ Noy and McGuinness (2001)

2. การค้นหาเชิงความหมาย (Semantic search) ประกอบด้วย 3 กระบวนการ ดังนี้

2.1 การประมวลผลคำสำคัญ (Keywords processcing) เป็นกระบวนการกำหนดกลุ่มคำค้นโดยการนำคำที่ผู้ใช้ระบุมาค้นหาคำที่อยู่ในคลังคำ จากนั้นจะทำการเลือกกลุ่มคำเพื่อมาใช้สำหรับการค้นหาแล้วส่งต่อไปยังกระบวนการการประมวลผลคำค้นต่อไป

2.2 การประมวลผลคำค้น (Query processing) เป็นกระบวนการในการสร้างคำค้นในรูปแบบภาษา SPARQL หรือการสร้างคำค้นพร้อมกับกฎการค้นคืนในฐานความรู้ด้วยภาษา SQWRL

2.3 การแสดงผลลัพธ์ (Result processing) เป็นการนำผลลัพธ์ที่ได้จากค้นหาในขั้นตอนการประมวลผลคำค้นมาแสดงให้กับผู้ใช้

3.3 การพัฒนาตัวแบบ

ในงานวิจัยนี้ได้พัฒนาตัวแบบการค้นคืนเชิงความหมายโดยการค้นหาความรู้ผลไม้ทางเศรษฐกิจ ซึ่งใช้ตัวแบบหรือภาษา ดังต่อไปนี้

1) SKOS สำหรับการจัดระบบคลังคำ
2) SPARQL เป็นภาษาสำหรับการค้นหาความรู้
3) SWRL เป็นภาษาสำหรับกฎการจัดการความรู้

การนำซอฟต์แวร์มาใช้ในการพัฒนาตัวแบบ จะเน้นการนำซอฟต์แวร์ที่เป็น Open Source หรือ Freeware เป็นหลักในการพัฒนาตัวแบบในรูปแบบของ Web application ดังต่อไปนี้

1) Protégé สำหรับเป็นเครื่องมือในการสร้างออนโทโลยี นำข้อมูลเข้า ทดสอบการค้นหาเบื้องต้น
2) Tomcat 9.0 สำหรับเป็น Web server
3) JSP (Java Server Pages) สำหรับเป็นภาษาหลักในการพัฒนาตัวแบบ โดยการเขียนโปรแกรมในรูปแบบภาษา JAVA

3.4 การประเมินตัวแบบ

การประเมินตัวแบบ เป็นการประเมินความถูกต้องและความสามารถของตัวแบบการค้นหาเชิงความหมาย โดยการดำเนินการเริ่มจากกำหนดข้อมูลที่ต้องการค้นหา หลังจากนั้นจะทำการค้นหาตามกระบวนการที่ได้ออกแบบไว้ แล้วนำข้อมูลมาประเมินเพื่อวิเคราะห์โดยใช้การคำนวณค่าความแม่นยำ (Precision Value) ค่าความระลึก (Recall Value) และอัตราการรู้จำ (F-Measure)

ค่าความแม่นยำ (Precision) คือค่าที่แสดงให้เห็นว่าตัวแบบที่พัฒนาขึ้นมีความแม่นยำอย่างไร ดังสมการที่ 3.1

$$Precision = \frac{TP}{TP+FP} \times 100\% \qquad \text{(สมการที่ 3.1)}$$

จากสมการ อธิบายได้ดังนี้

TP (True Position) แทน ความรู้ที่เลือกและเกี่ยวข้อง

FP (False Position) แทน ความรู้ที่เลือกแต่ไม่เกี่ยวข้อง

ค่าความระลึก (Recall) คือค่าที่แสดงให้เห็นว่าตัวแบบได้ทำการแสดงผลลัพธ์ออกมาแล้วมีความครอบคลุมความรู้ที่ต้องการค้นหาอย่างไร ดังสมการที่ 3.2

$$Recall = \frac{TP}{TP+FN} \times 100\% \qquad \text{(สมการที่ 3.2)}$$

จากสมการ อธิบายได้ดังนี้

TP (True Position) แทน ความรู้ที่เลือกและเกี่ยวข้อง

FN (False Negative) แทน ความรู้ที่ไม่เลือกแต่เกี่ยวข้อง

อัตราการรู้จำ (F-measure) คือค่าเฉลี่ยที่ให้ความสำคัญกับความแม่นยำและค่าความระลึก คำนวณได้จากสมการที่ 3.3

$$F\text{-}Measure = \frac{2 \times Recall \times Precision|}{Precision + Recall} \qquad \text{(สมการที่ 3.3)}$$

มหาวิทยาลัยเทคโนโลยีสุรนารี

บทที่ 4 ผลการวิจัยและอภิปรายผล

การดำเนินการวิจัยครั้งนี้ ผู้วิจัยเสนอผลการวิจัยการพัฒนาตัวแบบการค้นคืนเชิงความหมายสำหรับผลไม้เศรษฐกิจไทย ดังนี้

4.1 ผลการออกแบบฐานความรู้สำหรับผลไม้เศรษฐกิจไทย

4.1.1 คลังคำ

4.1.2 ออนโทโลยีผลไม้เศรษฐกิจไทย

4.2 ผลการพัฒนาตัวแบบการค้นคืนเชิงความหมาย

4.2.1 การประมวลผลคำสำคัญ

4.2.2 การประมวลผลคำค้น

4.2.3 การแสดงผลลัพธ์

4.2.4 การประเมินผลตัวแบบ

4.3 ผลการพัฒนาระบบการค้นคืนเชิงความหมาย

4.4 การอภิปรายผล

4.1 ผลการออกแบบฐานความรู้สำหรับผลไม้เศรษฐกิจไทย

4.1.1 คลังคำ

ในการออกแบบและพัฒนาคลังคำโดยใช้ตัวแบบเอสเคโอเอส (Simple Knowloedge Organization System: SKOS) ซึ่งเป็นตัวแบบสำหรับเป็นมาตรฐานในการทำให้เป็นระบบเป็นการดำเนินการจัดทำดัชนีเชิงความหมาย เพื่อเป็นการแลกเปลี่ยนหรือใช้ข้อมูลร่วมกันกับระบบอื่น ๆ ในอนาคต โดยมีการดำเนินการในการพัฒนาคลังคำ ดังนี้

1) การระบุคำสำคัญ

คำสำคัญที่ได้มาจากความรู้ผลไม้เศรษฐกิจไทย ซึ่งนำมาเป็นคำเดี่ยว ๆ เพื่อช่วยในการค้นหาความรู้ในฐานความรู้ และได้นำคำสำคัญดังกล่าวมาจัดการให้เป็นระบบ โดยการนำมาเป็นคำสำคัญที่อยู่ภายใต้คลาส skos:Concept และกำหนดให้เป็นคำหลักโดยใช้คุณสมบัติ skos:prefLabel ซึ่งเป็นคำที่ใช้กันโดยทั่วไป และในบางกรณีอาจมีคำอื่นที่มีความหมายเดียวกันก็จะนำมากำหนดคุณสมบัติไว้เป็นคำรองโดยใช้คุณลักษณะ skos:altLabel ซึ่งอาจไม่มี หรือมีมากกว่า 1 คำก็ได้ ตัวอย่างการระบุคำสำคัญแสดงได้ดังต่อไปนี้

ตัวอย่างการระบุคำสำคัญ

กำหนด “การควบคุมทรงพุ่ม” ให้อยู่ภายใต้คลาส skos:Concept

กำหนด “การควบคุมทรงพุ่ม” ให้เป็นคำหลักโดยใช้คุณลักษณะ skok:prefLabel

กำหนด “การตัดแต่งทรงพุ่ม” ให้เป็นคำรองโดยใช้คุณลักษณะ skos:altLabel

2) การระบุหมวดหมู่คำสำคัญ

เพื่อให้การจัดกลุ่มคำสำคัญที่จัดเก็บไว้ให้เป็นหมวดหมู่เพื่อให้สะดวกต่อการค้นหาหรือค้นคืนผลลัพธ์ได้ตรงกับความต้องการมากขึ้น ในงานวิจัยนี้จะใช้คุณลักษณะที่เอสเคโอเอสกำหนดมาให้โดยเริ่มจากกำหนดชื่อของหมวดหมู่คำสำคัญขึ้นมาก่อน จากนั้นนำไว้อยู่ภายใต้คลาส skos:Collection และสามารถกำหนดคำหลักและคำรองโดยใช้คุณลักษณะ skos:prefLabel และ skos:altLabel ได้เช่นเดียวกัน ตัวอย่างการระบุหมวดหมู่คำสำคัญแสดงได้ดังต่อไปนี้

ตัวอย่างการระบุหมวดหมู่คำสำคัญ

กำหนด “การดูแลรักษา” ให้อยู่ภายใต้คลาส skos:Collection

กำหนด “การดูแลรักษา” ให้เป็นคำหลักโดยใช้คุณลักษณะ skok:prefLabel

กำหนด “ดูแลรักษา” ให้เป็นคำรองโดยใช้คุณลักษณะ skok:altLabel

3) การเชื่อมโยงความสัมพันธ์ของคำสำคัญกับหมวดหมู่คำสำคัญ

หลังจากมีการระบุคำสำคัญไว้ที่คลาส skos:Concept และหมวดหมู่คำสำคัญไว้ที่คลาส skos:Collection เรียบร้อยแล้ว ต่อไปจะต้องมีการเชื่อมโยงหมวดหมู่คำสำคัญไปยังคำสำคัญที่เกี่ยวข้องอยู่ภายใต้หมวดหมู่คำสำคัญที่กำหนดขึ้นมานั้น ซึ่งสามารถใช้คุณลักษณะ skos:hasMember มาทำการเชื่อมโยงได้ ตัวอย่างการเชื่อมโยงความสัมพันธ์ของคำสำคัญกับหมวดหมู่คำสำคัญแสดงได้ดังต่อไปนี้

ตัวอย่างการเชื่อมโยงความสัมพันธ์ของคำสำคัญกับหมวดหมู่คำสำคัญ

กำหนดให้หมวดหมู่ “การดูแลรักษา” skos:hasMember “ควบคุมทรงพุ่ม”

skos:hasMember “ให้น้ำ”

skos:hasMember “ใส่ปุ๋ย”

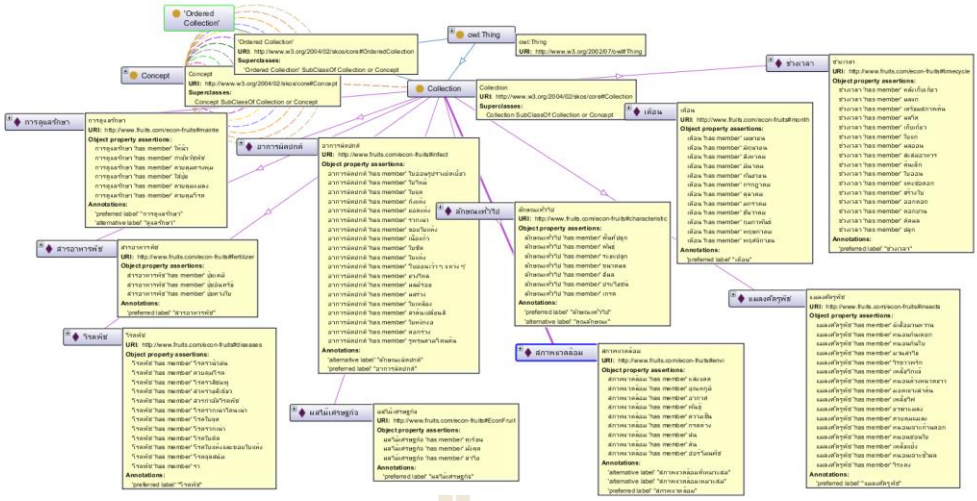
skos:hasMember “กำจัดวัชพืช”

skos:hasMember “ควบคุมโรค”

skos:hasMember “ควบคุมแมลง”

การระบุคำสำคัญ หมวดหมู่คำสำคัญ และการเชื่อมโยงความสัมพันธ์ของคำสำคัญดำเนินการโดยใช้โปรแกรมโปรเทเจ ซึ่งแสดงตัวอย่างของการจัดการต่าง ๆ เกี่ยวกับคลังคำได้ดังภาพที่ 4.1

มหาวิทยาลัยเทคโนโลยีสุรนารี



ภาพที่ 4.1 ตัวอย่างความสัมพันธ์ของคำในคลังคำตามตัวแบบเอสเคโอเอส

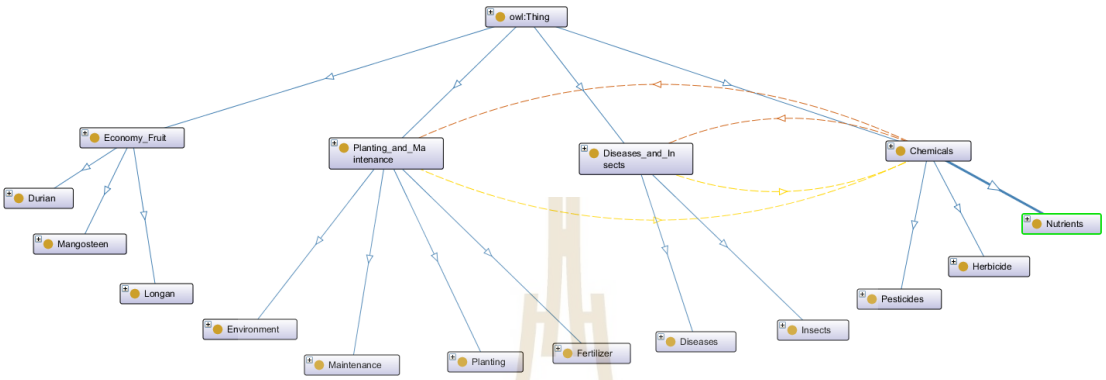
การออกแบบและพัฒนาคลังคำที่จัดทำขึ้นมาในขั้นแรกนี้ประกอบไปด้วย 10 คอลเล็คชัน 101 คอนเซ็ปต์ และสามารถเพิ่มเติมได้ในภายหลังตามความเหมาะสม ในแต่ละคอลเล็คชันมีคอนเซ็ปต์ที่เป็นสมาชิก ดังนี้

1) ผลไม้เศรษฐกิจ จำนวน 3 คอนเซ็ปต์
2) ลักษณะทั่วไป จำนวน 7 คอนเซ็ปต์
3) สภาพแวดล้อม จำนวน 9 คอนเซ็ปต์
4) ช่วงเวลา จำนวน 16 คอนเซ็ปต์
5) เดือน จำนวน 12 คอนเซ็ปต์
6) การดูแลรักษา จำนวน 6 คอนเซ็ปต์
7) สารอาหารพืช จำนวน 3 คอนเซ็ปต์
8) อาการผิดปกติ จำนวน 19 คอนเซ็ปต์
9) โรคพืช จำนวน 10 คอนเซ็ปต์
10) แมลงศัตรูพืช จำนวน 16 คอนเซ็ปต์

4.1.2 ออนโทโลยีผลไม้เศรษฐกิจไทย

ผลการออกแบบออนโทโลยีผลไม้เศรษฐกิจไทยตามแนวทางของกระบวนการวิศวกรรมออนโทโลยี ตามโครงสร้างที่ ศุภกฤษฎิ์ นิวัฒนากูล (2556) ได้ออกแบบไว้มาเป็นต้นแบบของคลาสในระดับบนต่อจากคลาส owl:Thing ซึ่งเป็นคลาสในระดับบนสุดหรือระดับราก (Root) แต่เนื่องจากข้อมูลที่ผู้ใช้ส่วนใหญ่ต้องการจะเป็นข้อมูลเกี่ยวกับการปลูก การดูแลรักษา การป้องกันกำจัดศัตรูพืช ตามที่ ศุภกฤษฎิ์ นิวัฒนากูล และฉัตรภัสร์ ฐิติอัคราวงศ์ (2560) ได้ศึกษาถึงประเภทของความรู้ผลไม้

เศรษฐกิจไทยที่ส่วนใหญ่ต้องการค้นหา จึงได้เลือกมาเฉพาะคลาสที่เกี่ยวข้องจำนวน 4 คลาส จากนั้นได้ดำเนินการออกแบบคลาสย่อยเพิ่มเติมเพื่อให้สอดคล้องกับข้อคำถามของการค้นคืนของตัวแบบการค้นหาเชิงความหมายในงานวิจัยนี้ โดยออกแบบคลาสดังกล่าวตามหลักการอนุกรมวิธาน (Taxonomy) ดังภาพที่ 4.2 ซึ่งเป็นการแสดงคลาสและความสัมพันธ์ของคลาสในระดับต่าง ๆ ที่ออกแบบไว้ในโปรแกรมโปรเทเจ



ภาพที่ 4.2 ลำดับชั้นของคลาสหลักและคลาสย่อยของออนโทโลยีผลไม้เศรษฐกิจไทย

คลาสหลักและคลาสย่อยของออนโทโลยีผลไม้เศรษฐกิจไทย มีการจัดเก็บความรู้ต่าง ๆ เพื่อช่วยการจัดกลุ่มและการจำแนกความรู้เป็นหมวดหมู่สำหรับรองรับการค้นหาข้อมูลให้มีความสะดวกมากขึ้น โดยความรู้ที่จัดเก็บในแต่ละคลาสจะมีขอบเขตของข้อมูลแสดงดังตารางที่ 4.1

ตารางที่ 4.1 ขอบเขตข้อมูลที่จัดเก็บของออนโทโลยีผลไม้เศรษฐกิจไทย

ชื่อคลาส	ขอบเขตข้อมูล
owl:Thing	เป็นคลาสระดับราก (Root) ของออนโทโลยีที่อยู่ในรูปแบบภาษา OWL
Δ Economy_Fruit	ผลไม้เศรษฐกิจ เป็นคลาสหลักของผลไม้เศรษฐกิจไทยที่นำมาเป็นฐานความรู้ในงานวิจัยนี้
• Durain	ทุเรียน เป็นคลาสย่อยที่รวบรวมข้อมูลทั้งหมดที่เกี่ยวกับทุเรียนที่อยู่ในฐานความรู้ที่สร้างขึ้น
• Mangosteen	มังคุด เป็นคลาสย่อยที่รวบรวมข้อมูลทั้งหมดที่เกี่ยวกับมังคุดที่อยู่ในฐานความรู้ที่สร้างขึ้น
• Longan	ลำไย เป็นคลาสย่อยที่รวบรวมข้อมูลทั้งหมดที่เกี่ยวกับลำไยที่อยู่ในฐานความรู้ที่สร้างขึ้น
Δ Planting_and_Maintenance	การปลูกและบำรุงรักษา เป็นคลาสหลักของสภาพแวดล้อมการปลูก การดูแลรักษา รวมถึงการให้ปุ๋ยและอาหารเสริม

ตารางที่ 4.1 ขอบเขตข้อมูลที่จัดเก็บของออนโทโลยีผลไม้เศรษฐกิจไทย

ชื่อคลาส	ขอบเขตข้อมูล
• Environment	สภาพแวดล้อม เป็นคลาสย่อยสำหรับรวบรวมความรู้เกี่ยวกับสภาพแวดล้อมสำหรับการปลูกไม้ผลเศรษฐกิจไทย เช่น สภาพดิน ปริมาณน้ำ ภูมิอากาศ ความเป็นกรดเป็นด่าง เป็นต้น
• Planting	การปลูก เป็นคลาสย่อยสำหรับรวบรวมความรู้เกี่ยวกับระยะปลูก วิธีการปลูก การดูแลเบื้องต้นสำหรับการปลูก เป็นต้น
• Maintenance	การดูแลรักษา เป็นคลาสย่อยสำหรับการรวบรวมความรู้เกี่ยวกับการดูแลรักษาทุกระยะของการผลิต เช่น ระยะก่อนออกดอก ระยะออกดอก ระยะติดผล หลังการเก็บผลผลิต เป็นต้น
• Fertilizer	การให้ปุ๋ย เป็นคลาสย่อยสำหรับการรวบรวมความรู้เกี่ยวกับการให้ปุ๋ยและธาตุอาหารในระยะต่าง ๆ
Δ Diseases_and_Insects	โรคและแมลง เป็นคลาสหลักสำหรับรวบรวมความรู้เกี่ยวกับโรคและแมลงศัตรูพืช
• Diseases	โรค เป็นคลาสย่อยสำหรับการรวบรวมความรู้เกี่ยวกับโรคพืช เช่น ชื่อโรคพืช อาการ สาเหตุ วิธีการป้องกัน การกำจัด เป็นต้น
• Insects	แมลง เป็นคลาสย่อยสำหรับการรวบรวมความรู้เกี่ยวกับแมลงศัตรูพืช เช่น ชื่อแมลง การป้องกัน การกำจัด เป็นต้น
Δ Chemicals	สารเคมี เป็นคลาสหลักสำหรับการรวบรวมความรู้เกี่ยวกับสารกำจัดแมลงศัตรูพืช สารกำจัดโรคพืช และธาตุอาหารเสริมสำหรับพืช
• Pesticides	สารกำจัดแมลงศัตรูพืช เป็นคลาสย่อยสำหรับการรวบรวมความรู้เกี่ยวกับสารกำจัดแมลงศัตรูพืชประเภทต่าง ๆ
• Herbicide	สารกำจัดโรคพืช เป็นคลาสย่อยสำหรับการรวบรวมความรู้เกี่ยวกับสารกำจัดโรคพืช
• Nutrients	ธาตุอาหารพืช เป็นคลาสย่อยสำหรับการรวบรวมความรู้เกี่ยวกับธาตุอาหารเสริมพืชเพื่อนำมาใช้ประโยชน์กับไม้ผล เช่น สร้างใบ กระตุ้นออกดอก ขยายขนาดผล เป็นต้น

ในกระบวนการออกแบบออนโทโลยีขั้นต่อไปหลังจากที่ได้โครงสร้างของคลาสที่แล้ว จะเป็นการกำหนดคุณสมบัติของออนโทโลยีซึ่งแบ่งได้เป็น 2 ประเภท ได้แก่ Datatype properties และ Object properties ซึ่งคุณสมบัติทั้งสองประเภทนี้จะใช้เชื่อมโยงความสัมพันธ์ของข้อมูลภายในออนโทโลยีที่พัฒนาขึ้น และใช้สำหรับการเชื่อมโยงความสัมพันธ์กับคำต่าง ๆ ที่ถูกจัดเก็บอยู่ในคลังคำ สำหรับรายละเอียดของคลังคำจะได้อธิบายในหัวข้อ 4.2 ต่อไป หลักการออกแบบคุณสมบัติจะใช้หลักการให้มีคุณสมบัติน้อยที่สุดเท่าที่จะเป็นไปได้แต่จะต้องให้สามารถยังคงอธิบายหรือเชื่อมโยงข้อมูลต่าง ๆ ได้อย่างชัดเจน ในการนี้ ได้ทำการออกแบบคุณสมบัติตามลักษณะของการใช้งานแบ่งได้ออกเป็น 2 ประเภท ดังนี้

1) คุณสมบัติสำหรับจัดเก็บรายละเอียดความรู้ภายในฐานความรู้

ได้ออกแบบคุณสมบัติประเภท Datatype properties มาจำนวน 1 คุณสมบัติเพื่อจัดเก็บรายละเอียดต่าง ๆ ของความรู้ในแต่ละคลาสสำหรับช่วยในการค้นหาผลลัพธ์เพื่อแสดงให้ตรงกับความต้องการของผู้ใช้ การออกแบบคุณสมบัติแสดงได้ดังตารางที่ 4.2

ตารางที่ 4.2 รายละเอียดของ Data type properties ในฐานความรู้

ชื่อคุณสมบัติ	โดเมน	ขอบเขต	รายละเอียดเพิ่มเติม
hasDescription	Planting_and_Maintenance Diseases_and_Insects Chemicals	String	ใช้สำหรับอธิบายรายละเอียดของความรู้ที่จัดเก็บในฐานความรู้

ในการเชื่อมโยงความสัมพันธ์ระหว่างข้อมูลในฐานความรู้จะมีการเชื่อมโยงระหว่างคลาส Planting_and_Maintenance และ คลาส Diseases_and_Insects กับคลาส Chemicals ดังนั้น ในการวิจัยนี้ จึงได้ออกแบบคุณสมบัติแบบ Object properties เพื่อเชื่อมโยงระหว่างคลาสดังกล่าว รายละเอียดแสดงดังตารางที่ 4.3

ตารางที่ 4.3 รายละเอียดของ Object properties ในฐานความรู้

ชื่อคุณสมบัติ	โดเมน	ขอบเขต	รายละเอียดเพิ่มเติม
useChemical	Planting_and_Maintenance Diseases_and_Insects	Chemicals	ใช้สำหรับเชื่อมโยงการใช้กับสารเคมีที่ใช้
useChemicalFor	Chemicals	Planting_and_Maintenance Diseases_and_Insects	ใช้สำหรับเชื่อมโยงสารเคมีที่ใช้กับการใช้ โดยกำหนดเป็น

ตารางที่ 4.3 รายละเอียดของ Object properties ในฐานความรู้

ชื่อคุณสมบัติ	โดเมน	ขอบเขต	รายละเอียดเพิ่มเติม
			inverse property ของ useChemical

2) คุณสมบัติเพื่อเชื่อมโยงความสัมพันธ์ข้อมูลในฐานความรู้กับรายการคำในคลังคำ

เป็นการออกแบบสำหรับการเชื่อมโยงความสัมพันธ์ระหว่างข้อมูลในฐานความรู้ที่เป็นรายละเอียดต่าง ๆ ซึ่งส่วนใหญ่เป็นคำอธิบาย หรืออาจจะเป็นคำเฉพาะ เช่น ชื่อยา สูตรปุ๋ย เป็นต้น และมักจะมีส่วนขยายตามมาด้วย เช่น อัตราการใช้ สาเหตุของการเกิดโรคพืช วิธีการดูแลรักษา เป็นต้น โดยการออกแบบจะมีคุณสมบัติหลักสำหรับการเชื่อมโยง และภายใต้คุณสมบัติหลักนี้จะมีคุณสมบัติย่อยเพื่อใช้การอธิบายรายละเอียดของการเชื่อมโยง และเพื่อช่วยการค้นหาให้มีความถูกต้อง และตรงกับความต้องการของผู้ใช้มากขึ้น รายละเอียดการออกแบบแสดงดังตารางที่ 4.4

ตารางที่ 4.4 คุณสมบัติเพื่อเชื่อมโยงความสัมพันธ์ระหว่างฐานความรู้กับคลังคำ

ชื่อคุณสมบัติ	โดเมน	ขอบเขต	รายละเอียดเพิ่มเติม
hasConcept	*3 คลาสหลัก	skos:Concept	เป็นคุณสมบัติหลักสำหรับเชื่อมโยงจากฐานความรู้ไปยังรายการคำในคลังคำ
forTreatment	*3 คลาสหลัก	skos:Concept	เป็นคุณสมบัติย่อยภายใน hasConcept สำหรับเชื่อมโยงไปยังรายการคำในคลังคำเพื่อแสดงคุณลักษณะย่อยในการใช้งานเพื่อการรักษา
forPrevent	*3 คลาสหลัก	skos:Concept	เป็นคุณสมบัติย่อยภายใน hasConcept สำหรับเชื่อมโยงไปยังรายการคำในคลังคำเพื่อแสดงคุณลักษณะย่อยในการใช้งานเพื่อการป้องกัน
forPractice	*3 คลาสหลัก	skos:Concept	เป็นคุณสมบัติย่อยภายใน hasConcept สำหรับเชื่อมโยงไปยังรายการคำในคลังคำเพื่อแสดงคุณลักษณะย่อยในการใช้งานเพื่อการปฏิบัติหรือดำเนินการใด ๆ
forExplain	*3 คลาสหลัก	skos:Concept	เป็นคุณสมบัติย่อยภายใน hasConcept สำหรับเชื่อมโยงไปยังรายการคำในคลัง

ตารางที่ 4.4 คุณสมบัติเพื่อเชื่อมโยงความสัมพันธ์ระหว่างฐานความรู้กับคลังคำ

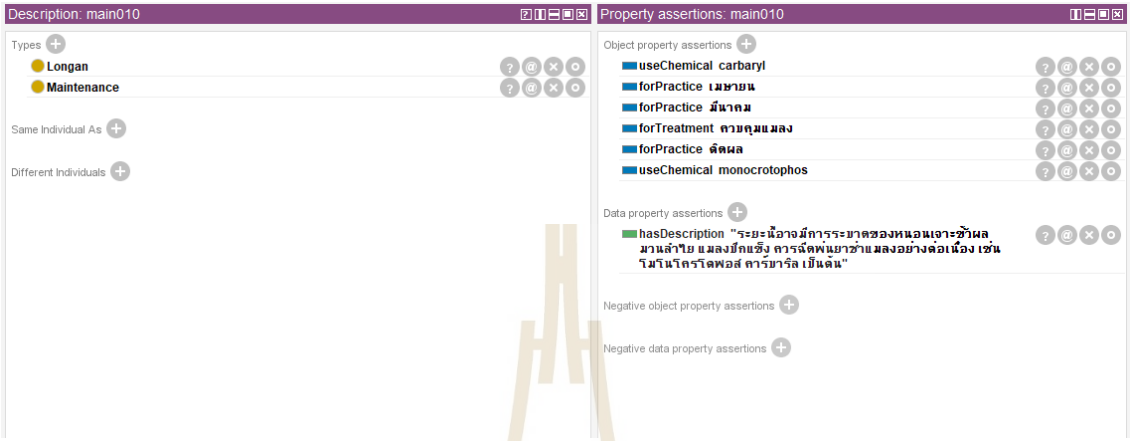
ชื่อคุณสมบัติ	โดเมน	ขอบเขต	รายละเอียดเพิ่มเติม
			คำเพื่อแสดงคุณลักษณะย่อยในการอธิบายรายละเอียดเพิ่มเติมเ
isConceptOf	skos:Concept	*3 คลาสหลัก	เป็นคุณสมบัติหลักสำหรับเชื่อมโยงจากรายการคำในคลังคำกลับมายังฐานความรู้ มีการกำหนดให้เป็น Inverse property ของ hasConcept
treatmentFor	skos:Concept	*3 คลาสหลัก	เป็นคุณสมบัติย่อยภายใน isConceptOf มีการกำหนดให้เป็น Inverse property ของ forTreatment
preventFor	skos:Concept	*3 คลาสหลัก	เป็นคุณสมบัติย่อยภายใน isConceptOf มีการกำหนดให้เป็น Inverse property ของ forPrevent
practiceFor	skos:Concept	*3 คลาสหลัก	เป็นคุณสมบัติย่อยภายใน isConceptOf มีการกำหนดให้เป็น Inverse property ของ forPractice
explainFor	skos:Concept	*3 คลาสหลัก	เป็นคุณสมบัติย่อยภายใน isConceptOf มีการกำหนดให้เป็น Inverse property ของ forExplain

หมายเหตุ *3 คลาสหลัก ได้แก่ Planting_and_Maintenance, Diseases_and_Insects และ Chemicals

หลังจากการออกแบบคลาสและคุณสมบัติตามรูปแบบออนโทโลยีแล้ว จะใช้โปรแกรมโปรเทเจในการสร้างคลาสและคุณสมบัติต่าง ๆ จากนั้นจะทำการระบุรายละเอียดของข้อมูลดังตัวอย่างแสดงดัง

ภาพที่ 4.3 เมื่อมีการสร้างข้อมูลขึ้นมาหนึ่งชุดจะมีการกำหนดให้อยู่ในคลาสที่เกี่ยวข้อง ซึ่งตามตัวอย่างจะได้แก่คลาส Longan และ Maintenance กำหนดคำอธิบายข้อมูลด้วยคุณสมบัติ hasDescription สำหรับในกรณีของข้อมูลที่เกี่ยวข้องกับการใช้สารเคมี จะมีการกำหนดความสัมพันธ์ไปยังข้อมูลในคลาสที่เกี่ยวข้อง ดังตัวอย่างที่ปรากฏจะมีการใช้คุณสมบัติ useChemical เพื่อกำหนดความสัมพันธ์ไปยังข้อมูล monocrotophos และ carbaryl ที่อยู่ภายใต้คลาส Peticides สำหรับการสร้างความสัมพันธ์ไปยังคอนเซปต์ตามตัวแบบ SKOS จะกระทำผ่านคุณสมบัติย่อยของ hasConcept

ดังเช่นการกำหนดคุณสมบัติย่อย forPractice สำหรับ “มีนาคม” และ “เมษายน” เพื่อบอกช่วงเวลาสำหรับการดำเนินการ กำหนดคุณสมบัติย่อย forPractice สำหรับ “ติดผล” เพื่อบอกระยะของการดำเนินการ และกำหนดคุณสมบัติย่อย forTreatment สำหรับ “ควบคุมแมลง” เพื่อบอกวัตถุประสงค์ในการใช้



ภาพที่ 4.3 ตัวอย่างการระบุรายละเอียดข้อมูลในโปรแกรมโปรเทเจ

ในการนำข้อมูลเข้าคลังคำและรายการความรู้ต่าง ๆ จะนำความรู้ที่เผยแพร่ทั้งเป็นเอกสารและเผยแพร่บนเครือข่ายอินเทอร์เน็ต นำมาสกัดเป็นความรู้ย่อย ๆ แล้วนำเข้าฐานความรู้ โดยความรู้ทั้งหมดที่นำเข้าเกี่ยวกับทุเรียน มังคุด และลำไย นำมากจาเอกสารหรือสื่อที่เผยแพร่บนอินเทอร์เน็ตของสถาบันวิจัยพืชสวน กรมส่งเสริมการเกษตร (2564) กรมวิชาการเกษตร กระทรวงเกษตรและสหกรณ์ (2564) คลังข้อมูลสารสนเทศระดับภูมิภาค (ภาคใต้) สำนักงานพัฒนาการวิจัยการเกษตร (องค์การมหาชน) (2564) ศูนย์วิจัยพืชสวนจันทบุรี (2558, 2564) กรมส่งเสริมการเกษตร กระทรวงเกษตรและสหกรณ์ (2564) คู่มือนักวิชาการส่งเสริมการเกษตร (2564) สำนักส่งเสริมและฝึกอบรม มหาวิทยาลัยเกษตรศาสตร์ (2564)

เมื่อดำเนินการระบุข้อมูลในออนโทโลยีทั้งหมดแล้ว จะมีการสร้างกฎด้วยภาษา SWRL เพื่อทำการสร้างข้อมูลให้กับฐานความรู้ ดังตัวอย่างกฎต่อไปนี้

useChemical(?X,?Y) -> useChemicalFor(?X,?Y)

forTreatment(?X,?Y) -> treatmentFor(?X,?Y)

forPractice(?X,?Y) -> practiceFor(?X,?Y)

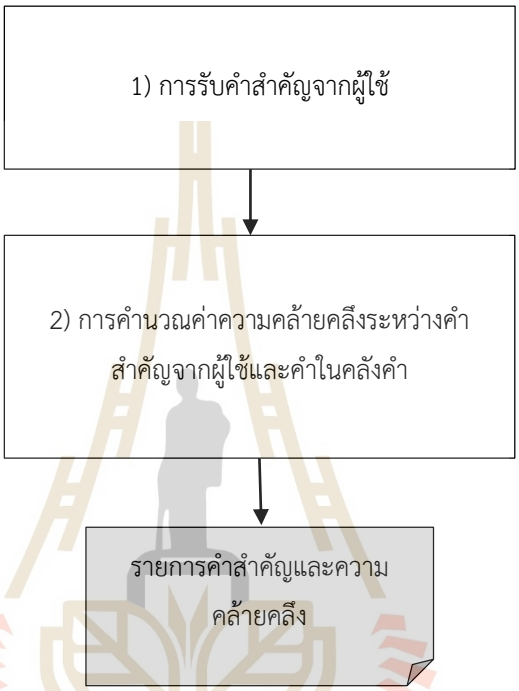
forPrevent(?X,?Y) -> preventFor(?X,?Y)

forExplain(?X,?Y) -> ExplainFor(?X,?Y)

4.2 ผลการพัฒนาตัวแบบการค้นคืนเชิงความหมาย

4.2.1 การประมวลผลคำสำคัญ

การประมวลผลคำสำคัญเป็นกระบวนการที่เริ่มจากการรับคำสำคัญจากผู้ใช้ที่ระบุเข้ามา โดยอาจเป็นคำสำคัญเดี่ยว ๆ หรืออาจมีคำสำคัญตั้งแต่สองคำขึ้นไป จากนั้นก็จะนำคำสำคัญที่ผู้ใช้ระบุมาเหล่านั้นไปดำเนินการเพื่อที่จะได้ผลลัพธ์ออกมาเป็นรายการคำที่จะใช้สำหรับการค้นหา โดยการประมวลผลคำสำคัญมีขั้นตอนแสดงดังภาพที่ 4.4



ภาพที่ 4.4 ขั้นตอนการประมวลผลคำสำคัญ

การทำงานในแต่ละขั้นตอนการประมวลผลคำสำคัญ มีรายละเอียดดังนี้

1) การรับคำสำคัญจากผู้ใช้

เมื่อผู้ใช้ประสงค์ที่จะค้นหาความรู้จากฐานความรู้ที่พัฒนาขึ้น จะต้องมีการกำหนดคำสำคัญสำหรับการค้นหา โดยอาจจะเป็นคำสำคัญเดี่ยว ๆ หรืออาจมีหลายคำสำคัญประกอบกันก็ได้ โดยผู้ใช้จะทำการระบุโดยแยกคำสำคัญออกจากกันโดยการเว้นวรรค เพื่อเป็นการแสดงตัวอย่างของการดำเนินการในกระบวนการรับคำสำคัญจากผู้ใช้ ได้กำหนดให้ผู้ใช้มีการค้นหาด้วยข้อความ “มังคุด ปลูกระยะ ดิน” จากนั้น จะทำการแบ่งคำออกเป็นคำย่อยโดยการเว้นวรรคแต่ละคำ จะได้คำค้นจากผู้ใช้เป็น “มังคุด” “ปลูกระยะ” และ “ดิน” เพื่อจะได้นำไปใช้ในกระบวนการต่อไป

2) การคำนวณค่าความคล้ายคลึงระหว่างคำสำคัญจากผู้ใช้กับคำในคลังคำ

โดยจะมีการค้นหาและเปรียบเทียบความคล้ายคลึงกับคำสำคัญที่อยู่ภายใต้คลาส skos:Concept ดังตัวอย่างภาษา SPARQL สำหรับการค้นหาคำสำคัญที่อยู่ในคลังคำ ดังต่อไปนี้

```
SELECT ?concept ?keyword
WHERE {
  { ?concept rdf:type skos:Concept .
    ?concept skos:prefLabel ?keyword . }
  UNION { ?concept rdf:type skos:Concept .
    ?concept skos:altLabel ?keyword . }
}
```

จากตัวอย่างแสดงดังตารางที่ 4.5 เป็นการแสดงกลุ่มคำจากผู้ใช้และกลุ่มคำสำคัญที่ได้จากการค้นหาในคลังคำ

ตารางที่ 4.5 กลุ่มคำจากผู้ใช้และกลุ่มคำสำคัญในคลังคำ

รายการคำค้นจากผู้ใช้	**รายการคำสำคัญในคลังคำ**
มังคุด ปลูกระยะ ดิน	ทุเรียน มังคุด ดิน ระยะปลูก ระยะผลเติบโต การกำจัดแมลง

เมื่อได้รายการคำสำคัญจากผู้ใช้และคำในคลังคำจากขั้นตอนที่ผ่านมา จะมีการนำสูตรการคำนวณค่าสัมประสิทธิ์ความคล้ายคลึงของแจ็กการ์ด (Jaccard similarity coefficient) (Jaccard, 1908, 1912) มาประยุกต์ใช้เพื่อเปรียบเทียบความคล้ายคลึงระหว่างคำค้นสองคำที่กำหนด โดยสูตรที่ใช้จะเป็นตามสมการที่ 4.1 ดังต่อไปนี้

$$\text{ค่าความคล้ายคลึงของ A และ B} = \frac{A \cap B}{A \cup B} \qquad \text{(สมการที่ 4.1)}$$

โดยที่ A หมายถึง จำนวนอักขระในคำค้น A ที่รับมาจากผู้ใช้

B หมายถึง จำนวนอักขระในคำสำคัญ B ที่อยู่ในคลังคำ

เมื่อได้ผลการคำนวณแล้ว จะทำการคัดเลือกคำค้นเพื่อไปใช้ในขั้นตอนต่อไปเฉพาะคำค้นที่มีค่าความคล้ายคลึงของคำสำคัญตั้งแต่ 0.7 ขึ้นไป ซึ่งหมายถึงว่าคำสำคัญจากผู้ใช้จะต้องมีความ

คล้ายคลึงกันกับคำในคลังคำตั้งแต่ร้อยละ 70 เป็นต้นไปจึงจะนำไปเป็นคำค้น ค่าดังกล่าวนี้เป็นค่าที่ได้ทดสอบแล้วพบว่าจะทำให้ได้คำหลักที่เหมาะสมที่สุด ซึ่งเป็นค่าที่ใกล้เคียงกับงานวิจัยของ ศุภกฤษฎิ์ นิวัฒนากูล (2556) ที่กำหนดค่าไว้ที่ 0.75 ขึ้นไป

จากตัวอย่างแสดงในตารางที่ 4.6 เป็นการคำนวณค่าความคล้ายคลึงของคำสำคัญจากผู้ใช้กับคำในคลังคำ

ตารางที่ 4.6 ตัวอย่างการคำนวนค่าความคล้ายคลึงของคำสำคัญจากผู้ใช้กับคำในคลังคำ

คำสำคัญจากผู้ใช้ (A)	**คำในคลังคำ (B)**	**จำนวนอักขระรวมทั้งหมด ($A \cup B$)**	**จำนวนอักขระที่เหมือนกัน ($A \cap B$)**	**ค่าความคล้ายคลึง**	**ผลการพิจารณา**
มังคุด	ทุเรียน	12	1	0.08	
	มังคุด	6	6	1.00	คำค้น
	ดิน	8	1	0.13	
	ระยะปลูก	13	0	0.00	
	ระยะผลเติบโต	16	0	0.00	
	การกำจัดแมลง	13	4	0.31	
ปลูกระยะ	ทุเรียน	12	2	0.17	
	มังคุด	13	0	0.00	
	ดิน	10	0	0.00	
	ระยะปลูก	7	7	1.00	คำค้น
	ระยะผลเติบโต	13	4	0.31	
	การกำจัดแมลง	15	2	0.13	
ดิน	ทุเรียน	9	1	0.11	
	มังคุด	8	1	0.13	
	ดิน	3	3	1.00	คำค้น
	ระยะปลูก	10	0	0.00	
	ระยะผลเติบโต	13	0	0.00	

มหาวิทยาลัยเทคโนโลยีสุรนารี

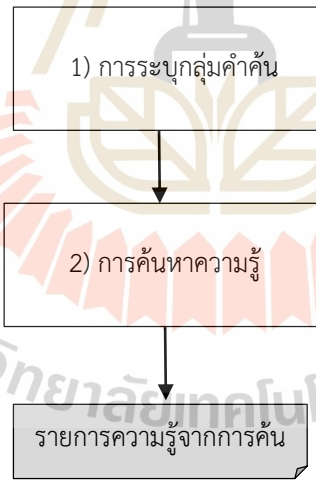
ตารางที่ 4.6 ตัวอย่างการคำนวนค่าความคล้ายคลึงของคำสำคัญจากผู้ใช้กับคำในคลังคำ

คำสำคัญจากผู้ใช้ (A)	คำในคลังคำ (B)	จำนวนอักขระรวมทั้งหมด ($A \cup B$)	จำนวนอักขระที่เหมือนกัน ($A \cap B$)	ค่าความคล้ายคลึง	ผลการพิจารณา
	การกำจัดแมลง	13	1	0.08	

หลังสิ้นสุดขั้นตอนการประมวลผลคำสำคัญ จะได้รายการคำค้นเพื่อที่จะส่งต่อไปยังขั้นตอนต่อไป จากตัวอย่างข้างต้นจะได้รายการคำค้นของคำค้นจากผู้ใช้ “มังคุด ปลูกระยะ ดิน” เป็นรายการคำค้นประกอบด้วย “มังคุด” “ระยะปลูก” และ “ดิน” เนื่องจากมีค่าความคล้ายคลึง 1.0 ทั้งหมด ซึ่งเป็นไปตามเกณฑ์ที่กำหนดให้คำค้นมีความคล้ายคลึงตั้งแต่ 0.7 ขึ้นไป

4.2.2 การประมวลผลคำค้น

การประมวลผลคำค้นจะเป็นขั้นตอนของการนำรายการคำค้นที่ได้จากขั้นตอนที่ผ่านมาเพื่อระบุคำค้นด้วยภาษา SPARQL สำหรับการค้นหาความรู้ผลไม้เศรษฐกิจไทยในฐานความรู้ที่พัฒนาขึ้น การประมวลผลคำค้นมีขั้นตอนการดำเนินการแสดงดังภาพที่ 4.5



ภาพที่ 4.5 ขั้นตอนการประมวลผลคำค้น

การดำเนินการในแต่ละขั้นตอนการประมวลผลคำค้น มีรายละเอียดดังนี้

1) การระบุกลุ่มคำค้น

การระบุกลุ่มคำค้นจะเริ่มต้นจากการนำคำค้นที่ได้จากขั้นตอนที่ผ่านมาจำแนกกลุ่มของคำค้นซึ่งอธิบายด้วย skos:Concept ว่าถูกจำแนกอยู่ในกลุ่มใดตาม skos:Collection ดังตัวอย่างคำสั่งภาษา SPARQL ที่แสดงดังต่อไปนี้

```
SELECT ?concept ?collection
WHERE {
  ?col rdf:type skos:Collection .
  ?col skos:member ?id .
  ?col skos:prefLabel ?collection .
  ?id skos:prefLabel ?concept .
  FILTER regex(?concept, "มังคุด")
}
```

ตารางที่ 4.7 ตัวอย่างการจำแนกกลุ่มของรายการคำค้น

คำค้น	**จำแนกกลุ่ม (skos:Collection)**
มังคุด	ผลไม้เศรษฐกิจ
ระยะปลูก	ลักษณะทั่วไป
ดิน	สภาพแวดล้อม

จากตารางที่ 4.7 เป็นตัวอย่างของการจำแนกกลุ่มของรายการคำค้น

2) การค้นหาความรู้

การค้นหาความรู้จะมีการดำเนินการ 2 ขั้นตอน ได้แก่ ขั้นตอนแรกเป็นการค้นหาแบบ Concept-based และขั้นตอนที่สองเป็นการค้นหาแบบ Keyword-based โดยขั้นตอนแรกการค้นหาแบบ Concept-based จะเป็นการค้นหาจากรายการคำค้นที่ได้จากขั้นตอนที่ผ่านมา ซึ่งจะมีการตรวจสอบก่อนว่ามีคำค้นที่อยู่ในกลุ่ม “ผลไม้เศรษฐกิจ” หรือไม่ หากมีคำค้นอยู่ในกลุ่มนี้ จะหมายถึงการค้นหาเฉพาะรายการความรู้ของผลไม้เศรษฐกิจตามที่ระบุเท่านั้น จึงจะต้องมีการค้นหาโดยใช้คำค้นในกลุ่ม “ผลไม้เศรษฐกิจ” ควบคู่ไปกับคำค้นอื่น ๆ แต่หากไม่มีคำค้นที่อยู่ในกลุ่ม “ผลไม้เศรษฐกิจ” ก็จะนำคำค้นต่าง ๆ ไปค้นหาความรู้ในฐานความรู้ จากตัวอย่างที่แสดงข้างต้น จะพบคำค้น “มังคุด” อยู่ในกลุ่ม “ผลไม้เศรษฐกิจ” ก็จะมีการค้นหาความรู้ที่ถูกกำกับด้วยคำค้น “ระยะปลูก” และ “ดิน” ที่เป็นรายการความรู้ของ “มังคุด” เท่านั้น การค้นหาความรู้ในฐานความรู้โดยใช้ภาษา SPARQL เพื่อเป็นการดึงความรู้ที่มีการเชื่อมโยงกับรายการคำค้น ดังตัวอย่างที่แสดงเป็นการค้นคืนความรู้ด้วยคำค้น “ระยะปลูก” และมีคำค้น “มังคุด” อยู่ในกลุ่ม “ผลไม้เศรษฐกิจ”

```
SELECT ?id ?concept ?result
WHERE {
  ?id rdf:type :Mangosteen .
  {  ?id :hasDescription ?result .
    ?id :forExplain ?con .
    ?con skos:prefLabel ?concept . }
  UNION { ?id :hasDescription ?result .
    ?id :forPractice ?con .
    ?con skos:prefLabel ?concept . }
  UNION { ?id :hasDescription ?result .
    ?id :forPrevent ?con .
    ?con skos:prefLabel ?concept . }
  UNION { ?id :hasDescription ?result .
    ?id :forTreatment ?con .
    ?con skos:prefLabel ?concept . }
  FILTER regex(?concept, "ระยะปลูก") .
}
```

หรือสามารถใช้การค้นหาดังตัวอย่างด้านล่าง ซึ่งจะทำให้ได้ผลลัพธ์ที่ไม่แตกต่างกัน แต่คำสั่งด้านบนสามารถนำไปใช้อธิบายถึงความรู้เป็นกลุ่มต่าง ๆ ดังเช่น การป้องกัน การกำจัด การรักษา เป็นต้น

```
SELECT ?id ?concept ?result
WHERE {
  ?id rdf:type :Mangosteen .
  ?id :hasDescription ?result .
   ?id :hasConcept ?con .
   ?con skos:prefLabel ?concept .
    FILTER regex(?concept, "ระยะปลูก") .
}
```

ตัวอย่างการค้นหาความรู้จากรายการคำค้น “ระยะปลูก” และ “ดิน” ที่เป็นรายการความรู้ของ “มังคุด” โดยภาษา SPARQL จะได้ผลลัพธ์แสดงดังตารางที่ 4.8

มหาวิทยาลัยเทคโนโลยีสุรนารี

ตารางที่ 4.8 ตัวอย่างผลลัพธ์ที่ได้จากการค้นหาในฐานความรู้ตามรายการคำค้นที่กำหนด

รายการความรู้	รายละเอียดความรู้	คำค้น
Km001	ดินร่วนปนทราย ระยะปลูกที่เหมาะสม 8-12 x 8-12 เมตร	มังคุด, ดิน, ระยะปลูก
Km002	การปลูกที่มีการระบายน้ำได้ดี มีความอุดมสมบูรณ์จะทำให้ต้นเติบโตเร็วมากขึ้น	มังคุด, ดิน
Km003	การปลูกระยะห่างของแต่ละต้นขึ้นอยู่กับสภาพของดินด้วย	มังคุด, ระยะปลูก

ขั้นตอนที่สอง เป็นการค้นหาแบบ Keyword-based ซึ่งเป็นการค้นหาความรู้โดยคำสำคัญจากผู้ใช้ และเช่นเดียวกัน จะต้องมีการตรวจสอบก่อนว่ามีคำค้นที่อยู่ในกลุ่ม “ผลไม้เศรษฐกิจ” หรือไม่ หากมีก็จะได้มีการกำหนดการค้นหาความรู้โดยควบคู่กับกลุ่ม “ผลไม้เศรษฐกิจ” ตามที่ผู้ใช้ระบุ การค้นหาความรู้ในฐานความรู้ตามคำสำคัญจากผู้ใช้ จะเป็นการนำคำสำคัญจากผู้ใช้ไปค้นหาคำที่อยู่ในรายละเอียดของรายการความรู้ ดังตัวอย่างเป็นการค้นคืนความรู้ที่ปรากฏคำสำคัญจากผู้ใช้ “ปลูกระยะ” และ “ดิน” โดยมีรายการความรู้อยู่ภายใต้ “มังคุด” ที่อยู่ในกลุ่ม “ผลไม้เศรษฐกิจ” จากตัวอย่างข้างต้น สามารถแสดงการค้นหาความรู้ด้วยภาษา SPARQL ดังต่อไปนี้

```
SELECT ?id ?desc
WHERE {
  ?id rdf:type :Mangosteen .
  ?id :hasDescription ?desc .
  FILTER regex(?desc, “ดิน”)
}
```

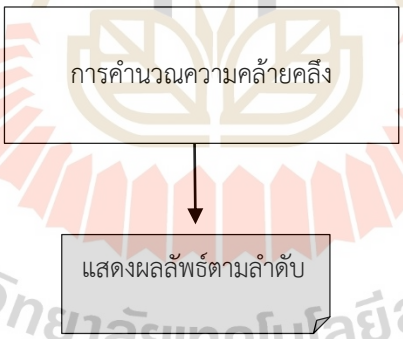
ตัวอย่างการค้นหาความรู้โดยใช้คำสำคัญจากผู้ใช้ “ปลูกระยะ” และ “ดิน” ที่เป็นรายการความรู้ของ “มังคุด” จะได้ผลลัพธ์แสดงดังตารางที่ 4.9

ตารางที่ 4.9 ตัวอย่างผลลัพธ์ที่ได้จากการค้นหาในฐานความรู้โดยใช้คำสำคัญจากผู้ใช้

รายการความรู้	**รายละเอียดความรู้**	**คำสำคัญจากผู้ใช้**
Km004	ควรมีการกำจัดแมลงศัตรูพืชในการปลูกระยะเริ่มต้นอย่างสม่ำเสมอ	มังคุด, ปลูกระยะ
Km001	ดินร่วนปนทราย ระยะปลูกที่เหมาะสม 8-12 x 8-12 เมตร	มังคุด, ดิน
Km003	การปลูกระยะห่างของแต่ละต้นขึ้นอยู่กับสภาพของดินด้วย	มังคุด, ดิน, ปลูกระยะ

4.2.3 การแสดงผลลัพธ์

การแสดงผลลัพธ์จะเป็นกระบวนการรับผลลัพธ์ที่ได้จากการค้นหาซึ่งจะมีข้อมูลเป็นรายการผลลัพธ์ของแต่ละคำค้น จากนั้นจะนำมาคำนวณหาความคล้ายคลึงของแต่ละรายการแล้วนำมาแสดงผลลัพธ์เรียงลำดับตามความคล้ายคลึงจากมากไปหาน้อย โดยจะทำการแสดงผลเฉพาะรายการความรู้ที่มีค่าความคล้ายคลึงตั้งแต่ 0.7 ขึ้นไป ขั้นตอนการดำเนินการแสดงตามภาพที่ 4.6



ภาพที่ 4.6 ขั้นตอนการประมวลผลลัพธ์

ในการคำนวณค่าความคล้ายคลึงของแต่ละรายการความรู้จะใช้การคำนวณค่าความคล้ายคลึงเชิงมุมโคไซน์ (Cosine Similarity Measurement) ดังสมการที่ 4.2

$$\text{Cosine Similarity} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \qquad \text{(สมการที่ 4.2)}$$

โดยที่ Similarity คือ ค่าคะแนนความคล้ายคลึงกัน

A_i คือ เวกเตอร์ของคำค้นที่กำหนดและจากผู้ใช้ ค่าที่เป็นไปได้คือ 0 หรือ 1

B_i คือ เวกเตอร์ของคำสำคัญที่ระบุในรายการความรู้ค่าที่เป็นไปได้คือ 0 หรือ 1

$\sum_{i=1}^{n} A_i B_i$ คือ ค่าผลรวมของคำค้นที่กำหนดและคำค้นจากรายการความรู้

$\sqrt{\sum_{i=1}^{n} A_i^2}$ คือ ค่ารากที่สองของผลรวมเวกเตอร์คำค้นที่กำหนด ยกกำลัง 2

$\sqrt{\sum_{i=1}^{n} B_i^2}$ คือ ค่ารากที่สองของผลรวมเวกเตอร์คำสำคัญจากรายการความรู้ ยกกำลัง 2

ในการเริ่มต้นคำนวณค่าความคล้ายคลึงของแต่ละรายการความรู้โดยวัดค่าความคล้ายคลึงเชิงมุมโคไซน์จะเริ่มจากการหารูปแบบเวกเตอร์ของคำค้นที่กำหนดและคำสำคัญที่ระบุแต่ละรายการความรู้โดยกำหนดค่า “0” ให้กับคำสำคัญที่ไม่มีในรายการคำค้นที่กำหนด และกำหนดค่า “1” ให้กับคำสำคัญที่มีในรายการคำค้นที่กำหนด ดังตัวอย่างการระบุรูปแบบเวกเตอร์ของคำค้นที่กำหนดและคำสำคัญจากรายการความรู้ Km001 โดยกำหนดให้ “C” เป็นการค้นแบบ Concept-based และ “K” เป็นการค้นแบบ Keyword-based ดังนี้

คำค้นที่กำหนด (A) = {C:มังคุด,C:ระยะปลูก,C:ดิน,K:ปลูกระยะ,K:ดิน}
คำสำคัญในรายการความรู้ Km001 (B) = {C:มังคุด, C:ระยะปลูก,C:ดิน,K:ดิน}
รูปแบบเวกเตอร์ของรายการความรู้ Km001 ได้แก่ {1,1,1,0,1)

การดำเนินการวัดค่าความคล้ายคลึงเชิงมุมโคไซน์ สามารถคำนวณได้จากสูตรที่ปรากฏดังสมการที่ 4.3

Cosine similarity (CS) = (A . B) / (||A|| ||B||) (สมการที่ 4.3)

โดยที่ A เป็นเวกเตอร์ของคำค้นที่กำหนด ได้แก่ {1,1,1,1,1)
B เป็นเวกเตอร์ของรายการความรู้ Km001 ได้แก่ {1,1,1,0,1)

คำนวณหา dot product ระหว่าง A และ B:

1.1 + 1.1 + 1.1 + 1.0 + 1.1 = 4

คำนวณหาขนาดของเวกเตอร์ A:

$\sqrt{1^2 + 1^2 + 1^2 + 1^2 + 1^2}$ = 2.23607

คำนวณหาขนาดของเวกเตอร์ B:

$$\sqrt{1^2 + 0^2 + 1^2 + 0^2 + 1^2} = 2.00000$$

คำนวณหาค่าความคล้ายคลึงเชิงมุมโคไซน์:

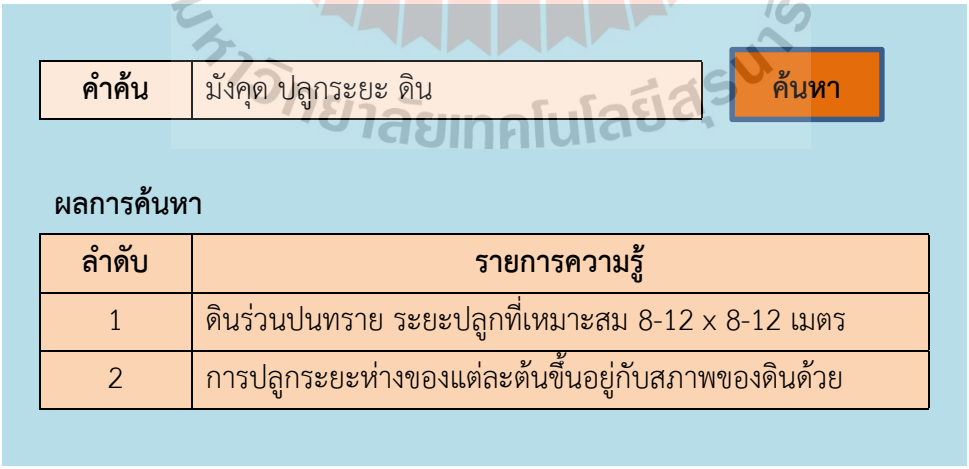
(3) / (2.23607*2.00000) = 0.89443

ดังนั้นค่าความคล้ายคลึงเชิงมุมโคไซน์ระหว่างคำค้นที่กำหนดและคำสำคัญที่ระบุในการความรู้ Km001 เท่ากับ 89.44% ดังแสดงในตารางที่ 4.10

ตารางที่ 4.10 ข้อมูลสำหรับคำนวณค่าความคล้ายคลึงเชิงมุมโคไซน์ของรายการความรู้

คำค้น	Km001	Km002	Km003	Km004
ผลไม้เศรษฐกิจ (C): มังคุด	1	1	1	1
คำค้น (C): ระยะปลูก	1	0	1	0
คำค้น (C): ดิน	1	1	0	0
คำสำคัญจากผู้ใช้ (K): ปลูกระยะ	0	0	1	1
คำสำคัญจากผู้ใช้ (K): ดิน	1	0	1	0
รูปแบบเวกเตอร์	{1,1,1,0,1}	{1,0,1,0,0}	{1,1,0,1,1}	{1,0,0,1,0}
ค่าความคล้ายคลึงเชิงมุมโคไซน์	0.89443	0.63246	0.89443	0.63246

จากตัวอย่างการคำนวณตามตารางที่ 4.10 ข้อมูลสำหรับคำนวณค่าความคล้ายคลึงเชิงมุมโคไซน์ของรายการความรู้ เมื่อได้ผลลัพธ์มาแล้วจะนำมาแสดงผลโดยเลือกเฉพาะรายการความรู้ที่มีค่าความคล้ายคลึงตั้งแต่ 0.70 ขึ้นไปมาแสดงผล โดยแสดงผลตามลำดับค่าความคล้ายคลึงเฉลี่ยจากมากไปหาน้อยในกรณีที่มีค่าความคล้ายคลึงเฉลี่ยเท่ากัน จะแสดงตามลำดับที่ได้มีการสืบค้นมา ตัวอย่างผลลัพธ์แสดงดังภาพที่ 4.7



ภาพที่ 4.7 ตัวอย่างผลลัพธ์ที่ได้จากการค้นหาด้วยคำค้นจากผู้ใช้ “มังคุด ปลูกระยะ ดิน”

4.2.4 การประเมินผลตัวแบบ

การประเมินตัวแบบที่ได้ออกแบบมาโดยการกำหนดคำค้นและทดสอบการค้นหาจากฐานความรู้ที่พัฒนามาโดยดำเนินการในโปรแกรมโปรเทเจ โดยได้ทดลองจำนวน 30 ชุดคำค้น จากนั้นนำผลลัพธ์มาวิเคราะห์และคำนวณค่าความแม่นยำ (Precision) และค่าความระลึก (Recall) เท่ากับร้อยละ 91.39, 83.44 ตามลำดับ และเมื่อพิจารณาค่าอัตราการรู้จำ (F-measure) เป็นการหาค่าเฉลี่ยที่ให้ความสำคัญกับความแม่นยำและความระลึกพบว่าผลการประเมินตัวแบบได้ค่าเท่ากับร้อยละ 87.23 ซึ่งเป็นผลลัพธ์ที่ได้เกินกว่าร้อยละ 80 ทั้งค่าความแม่นยำ ค่าความระลึก และค่าอัตราการรู้จำแสดงให้เห็นถึงตัวแบบสามารถดำเนินการในขั้นตอนการพัฒนาเป็นระบบในขั้นตอนต่อไปได้ ผลการคำนวณดังแสดงตารางที่ 4.11 สำหรับรายละเอียดการคำนวณดูได้จากภาคผนวก ก

ตารางที่ 4.11 ผลการประเมินตัวแบบ

คำค้น (Query)	การประเมินตัวแบบ		
	Precision	Recall	F-measure
30	91.39	83.44	87.23

ที่มา : การคำนวณภาคผนวก ก

4.3 ผลการพัฒนาระบบการค้นคืนเชิงความหมาย

ในการพัฒนาระบบการค้นคืนเชิงความหมายสำหรับผลไม้เศรษฐกิจไทย เป็นการพัฒนาในรูปแบบโปรแกรมประยุกต์บนเว็บ (Web application) โดยใช้โปรแกรม Apache Tomcat 9.0 สำหรับเป็นแม่ข่ายบนเว็บ พัฒนาโดยใช้ภาษา JAVA ตามรูปแบบการพัฒนาโปรแกรมประยุกต์บนเว็บด้วย JSP (Java Server Pages) โดยมีโครงสร้างไฟล์ต่าง ๆ ที่สำคัญดังนี้

- thaifo
  - **index.jsp**
  - WEB-INF
    - classes
      - fruits
        - econ_fruits
          - **econ_fruits.java**
          - econ_fruits.class
      - skos.rdf
      - EconFruits.owl
    - **myTag.tld**

มหาวิทยาลัยเทคโนโลยีสุรนารี

สำหรับรายละเอียดของไฟล์ index.jsp, myTag.tld และ econ_fruits.java สามารถดูรายละเอียดได้ที่ภาคผนวก ข ในการทำงานของระบบการค้นคืนเชิงความหมายสำหรับผลไม้เศรษฐกิจไทย มีตัวอย่างหน้าจอแสดงดังภาพที่ 4.8 และสามารถเข้าถึงได้ที่ http://thaifo.net:8080/thaifo/

ภาพที่ 4.8 ตัวอย่างหน้าจอการค้นหาเชิงความหมายสำหรับผลไม้เศรษฐกิจไทย

4.4 การประเมินผลระบบและการอภิปรายผล

ในการประเมินระบบการค้นคืนเชิงความหมายจะดำเนินการโดยการกำหนดคำค้นทั้งหมด 400 รายการ แบ่งออกเป็น 4 กลุ่ม ได้แก่

กลุ่มที่ 1 รายการคำค้นชื่อไม้ผลเศรษฐกิจและคำสำคัญ 1 คำค้น จำนวน 150 รายการ แบ่งเป็นผลไม้เศรษฐกิจ 3 ชนิด ๆ ละ 50 รายการ

กลุ่มที่ 2 รายการคำค้นชื่อไม้ผลเศรษฐกิจและคำสำคัญ 2 คำค้น จำนวน 150 รายการ แบ่งเป็นผลไม้เศรษฐกิจ 3 ชนิด ๆ ละ 50 รายการ

กลุ่มที่ 3 รายการคำค้นที่ประกอบด้วยคำสำคัญ 2 คำค้น จำนวน 50 รายการ

กลุ่มที่ 4 รายการคำค้นที่ประกอบด้วยคำสำคัญ 3 คำค้น จำนวน 50 รายการ

จากนั้นจะดำเนินการกำหนดคำค้นสอดคล้องกับแต่ละกลุ่มแล้วนำไปค้นหาในระบบการค้นคืนเชิงความหมายสำหรับผลไม้เศรษฐกิจไทยที่พัฒนาขึ้นมา แล้วให้ผู้เชี่ยวชาญที่มีประสบการณ์ในการทำสวนผลไม้ที่เกี่ยวข้องไม่น้อยกว่า 5 ปี มาทำการตรวจสอบผลการค้นคืนที่ได้มากับรายการความรู้ที่อยู่ในฐานความรู้ที่เกี่ยวข้องที่ไม่ได้นำมาแสดง แล้วนำข้อมูลมาคำนวณค่าความแม่นยำ ค่าความระลึก และค่าอัตราการรู้จำ แล้วนำมาเปรียบเทียบผลที่ได้ ผลการประเมินระบบปรากฏดังตารางที่ 4.12 สำหรับรายละเอียดของการค้นแสดงดังภาคผนวก ค

ตารางที่ 4.12 ผลการประเมินระบบการค้นคืนเชิงความหมาย

กลุ่มรายการคำค้น	จำนวนรายการ	ผลการประเมิน		
		Precision	Recall	F-measure
กลุ่ม 1 (ระบุไม้ผลและ 1 คำค้น)	150	97.22	91.83	94.45
1) ทุเรียน	*50*	*97.99*	*89.51*	*93.56*
2) มังคุด	*50*	*96.80*	*90.79*	*93.70*
3) ลำไย	*50*	*96.86*	*95.18*	*96.01*
กลุ่ม 2 (ระบุไม้ผลและ 2 คำค้น)	150	37.25	97.14	53.85
1) ทุเรียน	*50*	*48.71*	*96.75*	*64.80*
2) มังคุด	*50*	*37.61*	*94.67*	*53.83*
3) ลำไย	*50*	*25.43*	*100.00*	*40.55*
กลุ่ม 3 (2 คำค้น)	50	33.64	99.00	50.22
กลุ่ม 4 (3 คำค้น)	50	69.44	100.00	81.96
เฉลี่ย	400	63.31	95.74	76.22

ที่มา : การคำนวณภาคผนวก ค

ผลการประเมินในภาพรวมของระบบการค้นคืนเชิงความหมายสำหรับผลไม้เศรษฐกิจไทย ค่าความความแม่นยำร้อยละ 63.31 ค่าความระลึกร้อยละ 95.74 และค่าอัตราการรู้จำร้อยละ 76.22 พบว่าค่าความระลึกได้สูงเกณฑ์ที่กำหนดไว้ ซึ่งได้แก่ร้อยละ 80 ในขณะที่ค่าความแม่นยำ และค่าอัตราการรู้จำได้ต่ำกว่าเกณฑ์ที่กำหนดไว้ แสดงให้เห็นถึงการค้นหาของระบบว่าสามารถนำรายการความรู้มาแสดงเป็นผลลัพธ์ได้มากถึงร้อยละ 95.74 แต่ในรายการผลลัพธ์มีรายการที่ตรงกับความต้องการของผู้ใช้เพียงร้อยละ 63.31 เท่านั้น ซึ่งการค้นหาของระบบที่พัฒนาขึ้น จะใช้หลักการค้นหาแบบ Concept-based และ keyword-based ร่วมกัน เมื่อเปรียบเทียบกับการค้นหาแบบ Concept-based เพียงอย่างเดียวที่มีฐานความรู้ด้านเดียวกัน ได้แก่ด้านผลไม้เศรษฐกิจไทย พบว่างานวิจัยของศุภกฤษฎิ์ นิวัฒนากูล (2556) ได้นำเสนอการค้นคืนเชิงความหมายสำหรับผลไม้เศรษฐกิจไทย ได้ผลการประเมินค่าความแม่นยำ ค่าความระลึก และค่าอัตราการรู้จำ เท่ากับร้อยละ 96.84, 83.54 และ 87.55 ตามลำดับ ดังนั้น จะเห็นได้ว่า การนำหลักการค้นหาแบบ keyword-based มาร่วมในการค้นหาด้วยจะทำให้สามารถเพิ่มรายการความรู้ที่นำไปแสดงครอบคลุมได้มากขึ้น ดังจะเห็นได้จากค่าความระลึกเพิ่มขึ้นจากร้อยละ 83.54 เป็นร้อยละ 95.74 แต่ก็ส่งผลต่อความแม่นยำที่ลดลงจากร้อยละ 96.84 เป็นร้อยละ 63.31

การค้นหาแบบ Concept-based จะทำการดึงเฉพาะรายการความรู้ที่เกี่ยวข้องมาแสดง ทำให้ผลลัพธ์ที่ได้มีจำนวนน้อยแต่มีความเกียวข้องมาก ในขณะที่อาจมีรายการความรู้ในฐานความรู้ไม่ได้ถูก

นำมาแสดงเป็นผลลัพธ์ด้วย ในขณะที่การค้นหาแบบ keyword-based จะมาช่วยนำรายการความรู้ที่อยู่ในฐานความรู้มาแสดงเพิ่มมากขึ้น จะทำให้รายการความรู้ถูกดึงออกมาเป็นจำนวนมาก แต่รายการความรู้ดังกล่าวอาจไม่มีความเกี่ยวข้องกับความรู้ที่ต้องการค้นหา ซึ่งสอดคล้องกับผลการวิจัยที่พบว่ามีความรู้ถูกนำมาแสดงผลลัพธ์เป็นจำนวนมาก แต่มีรายการความรู้ที่ไม่เกี่ยวข้องถูกนำมาแสดงด้วย หากพิจารณาในประเด็นการค้นหาแบบที่ใช้ Concept-based เพียงอย่างเดียว จะพบงานวิจัยที่แสดงให้เห็นถึงผลประเมินที่ได้ค่าความแม่นยำที่สูงกว่า แต่ค่าความระลึกจะน้อยกว่า ดังงานวิจัยของ Bangkhomned and Payakpate (2020) พบว่าผลการประเมินการค้นคืนจากออนโทโลยีการผลิตลำไยคุณภาพสูง ได้ค่าความแม่นยำร้อยละ 100 ค่าความระลึกร้อยละ 93.50 และค่าอัตราการรู้จำร้อยละ 96.64 และงานวิจัยของ ทิพวิมล ชมภูคำ (2020) ได้นำเสนอการค้นคืนสารสนเทศเชิงความหมายสำหรับข้อมูลโรคติดเชื้อในเด็ก ได้ผลการประเมินค่าความแม่นยำ ค่าความระลึก และค่าอัตราการรู้จำร้อยละ 100, 92.30 และ 96.0 ตามลำดับ

การค้นหาความรู้แบบที่ระบุชื่อผลไม้และคำค้น 1 คำ จะทำให้ได้ผลลัพธ์ที่มีความเกี่ยวข้องมากขึ้น และรายการผลลัพธ์ที่ถูกนำมาแสดงจากฐานความรู้ก็มากด้วยเช่นกัน จะเห็นได้จากการค้นหาด้วยรายการคำค้นของกลุ่ม 1 ที่มีการระบุไม้ผลและ 1 คำค้น จะได้ผลประเมินเป็นค่าความแม่นยำร้อยละ 97.22 ค่าความระลึกร้อยละ 91.83 และค่าอัตราการรู้จำร้อยละ 94.45 ซึ่งทั้งหมดเกินเกณฑ์ที่กำหนดไว้ร้อยละ 80 เนื่องจากว่ามีการกำหนดชนิดผลไม้ไว้อย่างชัดเจน และจำนวนคำสำคัญที่ค้นหาจะมีเพียง 1 คำค้นเท่านั้น เมื่อพิจารณาเป็นรายชนิดผลไม้ พบว่าการค้นหาความรู้เกี่ยวกับทุเรียนจะให้ค่าความแม่นยำสูงสุดร้อยละ 97.99 หมายถึงในผลลัพธ์การค้นหาจะมีความถูกต้องมากถึงร้อยละ 97.99 ในขณะที่การค้นหาความรู้เกี่ยวกับลำไยจะให้ค่าความระลึกสูงสุดร้อยละ 95.18 หมายถึงระบบสามารถนำความรู้ที่อยู่ในฐานความรู้มาแสดงเป็นผลลัพธ์ได้มากถึงร้อยละ 95.18 และเมื่อพิจารณาถึงค่าอัตราการรู้จำแล้ว จะพบว่าการค้นหาความรู้ลำไยจะให้ผลการค้นคืนสูงสุดร้อยละ 96.01 ซึ่งผลการวิจัยนี้เป็นไปทิศทางเดียวกันของงานวิจัยที่มีการค้นหาความรู้ในฐานความรู้ที่เฉพาะเจาะจง ดังเช่น งานวิจัยของ Ensaf and Eyad (2020) ได้นำเสนอการประเมินระบบการค้นคืนคัมภีร์กุรอานที่พัฒนาขึ้นมาได้ค่าความแม่นยำ ค่าความระลึก และค่าอัตราการรู้จำ เท่ากับร้อยละ 100 ทั้งหมด โดยเป็นการค้นหาในรูปแบบ Concept-based ในฐานความรู้ที่เฉพาะเจาะจง

การระบุคำค้นในการค้นหาตั้งแต่ 2 คำค้นร่วมกันขึ้นไป จะทำให้รายการความรู้ที่ถูกนำมาแสดงมาค่าความแม่นยำที่ลดลง แต่ค่าความระลึกไม่มีความแตกต่างกัน จากการประเมินการค้นหาด้วย 2 คำค้นแบบไม่ระบุประเภทผลไม้ พบว่ามีค่าความแม่นยำร้อยละ 33.64 ค่าความระลึกร้อยละ 99.00 และค่าอัตราการรู้จำร้อยละ 50.22 และการประเมินการค้นหาด้วย 3 คำค้นแบบไม่ระบุประเภทผลไม้ พบว่ามีค่าความแม่นยำร้อยละ 69.44 ค่าความระลึกร้อยละ 100 และค่าอัตราการรู้จำร้อยละ 81.96 จะเห็นได้ว่าผลการประเมินเป็นไปในทิศทางเดียวกัน ได้แก่ค่าความแม่นยำจะต่ำกว่าเกณฑ์ ในขณะที่ค่าความระลึกจะอยู่สูงกว่าเกณฑ์ที่กำหนด ทั้งนี้ ผลการค้นหาขึ้นอยู่กับคำค้นที่นำมารวมกันด้วย เพราะ

หากคำค้นที่กว้างและอยู่ในขอบเขตที่ต่างกัน ก็จะทำให้ได้ผลลัพธ์ที่ไม่ตรงกับความต้องการตามที่ผู้ค้นกำหนดไว้ แต่ผลลัพธ์ของการค้นจะถูกนำมาแสดงเกือบจะทั้งหมด ดังจะเห็นได้จากค่าความระลึกของการค้นด้วย 2 คำค้นเท่ากับร้อยละ 99.00 และการค้นด้วย 3 คำค้นเท่ากับร้อยละ 100

การระบุประเภทผลไม้สำหรับการค้นด้วยนั้น กรณีประเภทผลไม้ที่ต่างกันไม่มีผลทำให้ผลการค้นหาแตกต่างกันมากนัก ผลการค้นหาเป็นไปในทางเดียวกันตามกลุ่มคำค้นที่ค้นหา จะเห็นได้ว่ากลุ่มที่ระบุผลไม้และ 1 คำค้น ก็จะได้ผลการประเมินไปในทิศทางเดียวกันที่ค่าความแม่นยำ ค่าความระลึก และค่าอัตราการรู้จำสูงกว่าเกณฑ์ที่กำหนดไว้ร้อยละ 80 ทั้งหมด และในขณะเดียวกันสำหรับการค้นหาในกลุ่มที่ระบุผลไม้และ 2 คำค้น จะพบว่าค่าความแม่นยำลดลงทุกชนิดผลไม้ และต่ำกว่าเกณฑ์ที่กำหนดไว้ ในขณะที่ค่าความระลึกของพืชทั้ง 3 ชนิดมีอัตราที่สูงและเข้าใกล้ 100 ซึ่งหมายถึงมีการนำรายการความรู้มาแสดงครอบคลุมเกือบจะทั้งหมด ในการที่จะเพิ่มประสิทธิภาพของการค้นคืนนั้นสามารถทำได้หลากหลายวิธี ดังเช่นวิธีของ Formica, Pourabbas, and Taglino (2020) ได้นำเสนอการเพิ่มประสิทธิภาพของผลลัพธ์ด้วยทำให้ค่าความแม่นยำ และค่าความระลึกเพิ่มขึ้นไปเป็นร้อยละ 100 ได้ โดยการเพิ่มการกำหนดน้ำหนักของคำสำคัญที่อยู่ในดัชนีของรายการคำ

อย่างไรก็ตามผลลัพธ์การค้นคืนจะมีประสิทธิภาพมากน้อยเพียงใดไม่เพียงแต่ขึ้นอยู่กับเทคนิควิธีการค้นคืน การจัดการฐานความรู้ และเลือกใช้เทคโนโลยีฐานความรู้ที่เหมาะสมเท่านั้น การเพิ่มจำนวนชนิดของผลไม้ หรือความรู้ที่มากขึ้นเป็นไปได้ว่าจะส่งผลต่อประสิทธิภาพของค่าความแม่นยำ ค่าความระลึก และค่าอัตราการรู้จำ ที่ลดลงหรือเพิ่มขึ้น แต่ประสิทธิภาพจะลดลงหรือเพิ่มมากขึ้นมากน้อยเพียงใดนั้น ขึ้นอยู่กับการจัดทำดัชนีของคำสำคัญหรือคำควบคุมในตัวแบบสร้างคลังคำเพื่อให้ครอบคลุม รองรับการค้นหาที่มีระบุการเชื่อมโยงไปยังความรู้ที่จัดเก็บที่เพิ่มมากขึ้น ดังนั้นกรณีที่ต้องการเพิ่มจำนวนความรู้ที่มากขึ้น ควรจัดทำดัชนีของคำสำคัญให้ครอบคลุมกับความรู้ เพื่อให้การค้นคืนมีประสิทธิภาพมากขึ้น การเพิ่มประเภทของผลไม้สามารถทำได้ไม่จำกัดเนื่องจากโครงสร้างออนโทโลยีได้ออกแบบรองรับไว้สำหรับการขยายในอนาคต สำหรับโปรแกรมที่พัฒนาขึ้นมาได้คำนึงถึงการเพิ่มจำนวนผลไม้ในอนาคตไว้ด้วยเช่นกัน อย่างไรก็ตาม จะต้องมีการปรับแก้โปรแกรมเพียงเล็กน้อยก็จะสามารถรองรับกับการทำงานได้ โดยส่วนการทำงานหลักของโปรแกรมไม่จำเป็นต้องมีการปรับแก้ใด ๆ ทำให้โปรแกรมที่พัฒนามีความยืดหยุ่นต่อการเพิ่มประเภทของผลไม้ได้ในอนาคต

บทที่ 5 สรุปผลการวิจัยและอภิปรายผล

จากการศึกษาวิจัยและพัฒนาตัวแบบการค้นคืนเชิงความหมายสำหรับผลไม้เศรษฐกิจไทยครั้งนี้มีผลการวิจัยและข้อเสนอแนะมีดังนี้

5.1 สรุปผลการวิจัย

งานวิจัยนี้ได้วิเคราะห์ ออกแบบ และพัฒนาตัวแบบการค้นคืนเชิงความหมายสำหรับผลไม้เศรษฐกิจไทย มีกระบวนการที่สำคัญ ดังนี้

1) กระบวนการออกแบบฐานความรู้

ฐานความรู้ที่พัฒนาขึ้นมาประกอบด้วยสองส่วนที่สำคัญ ได้แก่ คลังคำ และรายการความรู้ โดยทั้งสองส่วนจะมีการเชื่อมโยงระหว่างกันด้วยโครงสร้างของตัวแบบออนโทโลยีที่พัฒนาขึ้นมาและใช้ตัวแบบเอสเคโอเอสที่เป็นตัวแบบสำหรับการจัดการรายการคำต่าง ๆ ที่นำมาใช้ในการระบุถึงรายการความรู้ต่าง ๆ

คลังคำมีกระบวนการดำเนินการเริ่มจากการระบุคำสำคัญที่ได้มาจากรายการความรู้ผลไม้เศรษฐกิจไทย นำมาจัดโครงสร้างภายใต้คลาส skos:Concept และคุณสมบัติต่าง ๆ แต่ละคำสำคัญจะเรียกว่าคอนเซ็ปต์ จากนั้นจะนำคำสำคัญต่าง ๆ เหล่านั้นมาระบุหมวดหมู่คำสำคัญอยู่ภายใต้คลาส skos:Collection หรือที่เรียกว่าคอลเล็คชัน แล้วมีการเชื่อมโยงความสัมพันธ์ของคำสำคัญกับหมวดหมู่คำสำคัญ เพื่อที่จะได้ใช้สำหรับการเรียกแสดงข้อมูลตามความต้องการของผู้ใช้โดยจำแนกตามหมวดหมู่และคำสำคัญได้ โดยคลังคำที่พัฒนาขึ้นประกอบไปด้วย 101 คอนเซ็ปต์ซึ่งแต่ละคอนเซ็ปต์อาจมีคำสำคัญรวมอยู่ด้วยมากกว่า 1 คำสำคัญ ซึ่งเป็นคำที่มีความหมายเดียวกัน และแต่ละคอนเซ็ปต์จะเชื่อมโยงความสัมพันธ์กับคอลเล็คชันที่กำหนดไว้ 10 คอลเล็คชัน ประกอบด้วย ผลไม้เศรษฐกิจ 3 คอนเซ็ปต์ ลักษณะทั่วไป 7 คอนเซ็ปต์ สภาพแวดล้อม 9 คอนเซ็ปต์ ช่วงเวลา 16 คอนเซ็ปต์ เดือน 12 คอนเซ็ปต์ การดูแลรักษา 6 คอนเซ็ปต์ สารอาหารพืช 3 คอนเซ็ปต์ อาการผิดปกติ 19 คอนเซ็ปต์ โรคพืช 10 คอนเซ็ปต์ และแมลงศัตรูพืช 16 คอนเซ็ปต์

การออกแบบออนโทโลยีผลไม้เศรษฐกิจไทยใช้แนวทางของกระบวนการวิศวกรรมออนโทโลยี มีคลาสในระดับบนสุด หรือระดับราก (Root) ได้แก่คลาส owl:Thing และมีคลาสหลักจำนวน 4 คลาส คลาสย่อย 12 คลาส รวมทั้งหมด 16 คลาส 4 คลาสหลัก ประกอบด้วยคลาส Economy_Fruit ซึ่งเป็นคลาสของผลไม้เศรษฐกิจไทย คลาส Planting_and_Maintenance เป็นคลาสหลักของสภาพแวดล้อม การปลูก การดูแลรักกา รวมถึงการให้ปุ๋ยและอาหารเสริม คลาส Diseases_and_Insects เป็นคลาสสำหรับรวมรวมรายการความรู้เกี่ยวกับโรคและแมลงศัตรูพืช และ

คลาส Chemicals เป็นคลาสหลักของการรวบรวมรายการความรู้เกี่ยวกับสารเคมีกำจัดแมลงศัตรูพืช โรคพืช และธาตุอาหารเสริมสำหรับพืช

นอกจากนี้ยังมีการออกแบบคุณสมบัติประเภท Datatype properties จำนวน 1 คุณสมบัติ ได้แก่ hasDescription เพื่อใช้สำหรับอธิบายรายละเอียดของคววามรู้ที่จัดเก็บในฐานความรู้ และออกแบบคุณสมบัติประเภท Object properties จำนวน 12 คุณสมบัติ โดยมี 2 คุณสมบัติที่ใช้สำหรับเชื่อมโยงข้อมูลในฐานความรู้ และอีก 10 คุณสมบัติสำหรับเชื่อมโยงรายการความรู้กับคำสำคัญในคลังคำ ในจำนวนคุณสมบัติที่กำหนดไว้จะมีคุณสมบัติทีเรียกว่า Inverse properties รวมอยู่ด้วย เพื่อใช้สำหรับการเชื่อมโยงข้อมูลระหว่างความรู้และคำสำคัญให้สามารถค้นหาได้สะดวกและรวดเร็วขึ้น

หลังจากพัฒนาโครงสร้างออนโทโลยีแล้ว ได้มีการนำเข้าความรู้และมีการเชื่อมโยงความรู้กับคลังคำเพื่อจะช่วยในการการค้นหาความรู้ต่อไป การนำเข้าข้อมูลในออนโทโลยีจะมีการนำเข้าข้อมูลบางส่วน และบางส่วนจะใช้การสร้างกฎด้วยภาษา SWRL เพื่อทำการเชื่อมโยงข้อมูลให้กับฐานความรู้ทั้งระหว่างฐานความรู้และระหว่างฐานความรู้และคลังคำ ซึ่งจะทำให้ลดเวลาในการนำเข้ารายการความรู้ อีกทั้งจะมีความถูกต้องมากขึ้นเนื่องจากอาจมีความผิดพลาดจากการระบุการเชื่อมโยงที่ไม่ครบถ้วน หรือเชื่อมโยงไม่ถูกต้อง

2) กระบวนการพัฒนาตัวแบบการค้นคืนเชิงความหมาย

ในการพัฒนาตัวแบบการค้นคืนเชิงความหมายจะเริ่มจากการประมวลคำสำคัญ การประมวลคำค้น การแสดงผลลัพธ์ และการประเมินผลตัวแบบ การประมวลคำสำคัญจะเริ่มจาก การรับข้อความจากผู้ใช้ที่ระบุเข้ามาเพื่อใช้สำหรับการค้นหา นำมาระบุเป็นคำสำคัญโดยการเว้นวรรค จากนั้นจะนำไปคำนวณค่าความคล้ายคลึงระหว่างคำสำคัญจากผู้ใช้กับคำสำคัญในคลังคำด้วยสูตร การคำนวณค่าสัมประสิทธิ์ความคล้ายคลึงของแจ็กการ์ด (Jaccard similarity coefficient) (Jaccard, 1912) เมื่อได้ผลการคำนวณแล้วจะคัดเลือกคำค้นจากคำสำคัญที่มีค่าความคล้ายคลึงตั้งแต่ 0.7 ขึ้นไป เพื่อเป็นรายการคำค้นสำหรับการค้นหาต่อไป

การประมวลคำค้น ประกอบด้วย 2 กระบวนการ ได้แก่ การระบุกลุ่มคำค้น และการค้นหาความรู้ การระบุกลุ่มคำค้น เป็นการนำรายการคำค้นที่ได้จากขั้นตอนการประมวลผลคำสำคัญมาระบุกกลุ่มคำค้นก่อนว่าอยู่ในคอลเล็คชันใด จากนั้น จะทำการตรวจสอบว่ามีคำค้นที่อยู่ในกลุ่ม “ผลไม้เศรษฐกิจ” หรือไม่ ซึ่งหมายถึงมีการระบุถึงชนิดไม้ผลที่ต้องการค้นหา หากพบว่ามีคำค้นที่อยู่ในกลุ่ม “ผลไม้เศรษฐกิจ” จะกำหนดให้ค้นคืนเฉพาะรายการความรู้ตามชนิดผลไม้ที่ผู้ใช้ระบุมาเท่านั้น แต่หากไม่มีคำค้นที่ระบุถึงชนิดผลไม้ก็จะทำการค้นหาโดยใช้คำสำคัญจากรายการคำค้น การค้นหาความรู้จะเป็นการใช้ในการสร้างคำสั่งสำหรับค้นหาด้วยภาษา SPARQL การค้นหาความรู้จะมีการดำเนินการสองขั้นตอน ได้แก่ การค้นหาแบบ Concept-based เป็นการค้นหาตามรายการคำสำคัญได้มาจากกระบวนการประมวลผลคำสำคัญ และการค้นหาแบบ Keyword-bassed เป็นการค้นหาตาม

คำค้นที่ผู้ใช้ระบุ ผลลัพธ์ที่ได้จากค้นหาในขั้นตอนนี้จะเป็นรายการความรู้ที่ประกอบไปด้วยชนิดผลไม้ คำสำคัญที่ระบุในรายการความรู้ และรายละเอียดของรายการความรู้

การแสดงผลลัพธ์เป็นกระบวนการที่นำรายการผลลัพธ์ที่ได้จากการค้นหาแบบ Concept-based และ Keyword-based มารวมเป็นรายการผลลัพธ์เดียวกันโดยมีการคำนวณค่าความคล้ายคลึงของแต่ละรายการผลลัพธ์กับรายการคำค้นและคำค้นจากผู้ใช้โดยการคำนวณค่าความคล้ายคลึงเชิงมุมโคไซน์ (Cosine Similarity Measurement) โดยจะมีการหาค่ารูปแบบเวกเตอร์ของแต่ละรายการความรู้ที่ค้นคืนมาได้ จากนั้นจะนำไปหาค่า Dot product ระหว่างรายการคำค้นและคำค้นจากผู้ใช้กับการระบุรายการคำค้นและคำค้นจากผู้ใช้ในแต่ละรายการความรู้ แล้วคำนวณหาขนาดของเวกเตอร์ของทั้งสองรายการ แล้วนำไปคำนวณหาค่าความคล้ายคลึงเชิงมุมโคไซน์ตามสูตรที่กำหนด เมื่อได้ผลลัพธ์มาแล้วจะมาแสดงตามลำดับจากค่าที่ได้มากที่สุดลงไปหาค่าน้อยที่สุด โดยจะคัดเลือกเฉพาะรายการความรู้ที่มีค่าความคล้ายคลึงเชิงมุมโคไซน์ตั้งแต่ 0.7 ขึ้นไปมาแสดงผลลัพธ์ส่วนที่ได้ ต่ำกว่านี้จะตัดออกไม่นำมาเป็นผลลัพธ์ ในการพัฒนาตัวแบบนี้จะมีการประเมินตัวแบบโดยการทดลองการค้นหา 30 ชุดคำค้นที่แตกต่างกัน แล้วนำมาวิเคราะห์ค่าความแม่นยำได้ร้อยละ 91.39 ค่า ความระลึกร้อยละ 83.44 และค่าอัตราการรู้จำร้อยละ 87.23

3) การพัฒนาระบบการค้นคืนเชิงความหมายและการประเมินผลระบบ

ในการพัฒนาระบบการค้นคืนเชิงความหมายสำหรับผลไม้เศรษฐกิจไทย จะทำงานในรูปแบบโปรแกรมประยุกต์บนเว็บ (Web application) โดยใช้ภาษา JAVA ตามรูปแบบการพัฒนาโปรแกรมประยุกต์บนเว็บด้วย JSP (Java Server Pages) และมีการประเมินผลระบบโดยผู้เชี่ยวชาญเป็นผู้กำหนดคำค้นและตรวจสอบรายการผลลัพธ์ มีการทดสอบการค้นทั้งหมด 400 รายการ แบ่งออกเป็น 4 กลุ่ม ได้แก่ กลุ่มที่ 1 ค้นหาตามชื่อผลไม้เศรษฐกิจและคำสำคัญ 1 คำค้น จำนวน 150 รายการแบ่งเป็นผลไม้ชนิดละ 50 รายการ ผลการประเมินได้ค่าความแม่นยำร้อยละ 97.22 ค่าความระลึกร้อยละ 91.83 และค่าอัตราการรู้จำร้อยละ 94.45 กลุ่มที่ 2 ค้นหาตามชื่อผลไม้เศรษฐกิจและคำสำคัญ 2 คำค้น จำนวน 150 รายการแบ่งเป็นผลไม้ชนิดละ 50 รายการ ผลการประเมินได้ค่าความแม่นยำร้อยละ 37.25 ค่าความระลึกร้อยละ 97.14 และค่าอัตราการรู้จำร้อยละ 53.85 กลุ่มที่ 3 ค้นหาตามคำสำคัญ 2 คำค้นโดยไม่ระบุชนิดผลไม้ จำนวน 50 รายการ ผลการประเมินได้ค่าความแม่นยำร้อยละ 33.64 ค่าความระลึกร้อยละ 99.00 และค่าอัตราการรู้จำร้อยละ 50.22 กลุ่มที่ 4 ค้นหาตามคำสำคัญ 3 คำค้นโดยไม่ระบุชนิดผลไม้ จำนวน 50 รายการ ผลการประเมินได้ค่าความแม่นยำร้อยละ 69.44 ค่าความระลึกร้อยละ 100.00 และค่าอัตราการรู้จำร้อยละ 81.96 กลุ่มที่ 3 ผลการประเมินระบบในภาพรวมได้ค่าความแม่นยำร้อยละ 63.31 ค่าความระลึกร้อยละ 95.74 และค่าอัตราการรู้จำร้อยละ 76.22

5.2 ข้อเสนอแนะ

ข้อเสนอแนะสำหรับกระบวนการออกแบบและพัฒนาตัวแบบการค้นคืนเชิงความหมายสำหรับผลไม้เศรษฐกิจไทยมีดังต่อไปนี้

1) การพัฒนาคลังคำจะใช้คำสำคัญที่ได้มาจากเอกสารความรู้ ซึ่งเป็นเอกสารทางวิชาการ มีการใช้คำหรือภาษาที่เป็นทางการเป็นส่วนมาก ทำให้คำสำคัญต่าง ๆ ที่อยู่ในคลังคำเป็นคำที่เป็นทางการเช่นเดียวกัน แต่คำที่ผู้ใช้ทำการค้นหาซึ่งเป็นชาวสวนที่ทำสวนผลไม้ จะใช้คำที่ใช้กันทั่วไป ไม่เป็นทางการ และใช้กันหลากหลายในแต่ละพื้นที่ ทำให้การค้นหาอาจไม่ได้ผลลัพธ์ตามความต้องการของผู้ใช้ เนื่องจากภาษาที่อธิบายความรู้ กับภาษาที่ผู้ใช้งานจริงระบุเข้า ดังนั้น ควรมีการศึกษาเปรียบเทียบระหว่างคำหรือภาษาทางการกับคำหรือภาษาท้องถิ่นที่ผู้ใช้มักจะใช้เป็นส่วนใหญ่ เพื่อที่จะให้การค้นหาด้วยคำสำคัญไม่ว่าจะเป็นการค้นหาด้วยคำค้นที่เป็นทางการ หรือคำค้นที่ภาษาท้องถิ่น ก็จะทำให้ได้ผลลัพธ์ที่ตรงกับความต้องการ และครอบคลุมผลลัพธ์ที่นำมาแสดงมากขึ้น

2) ในการเพิ่มประสิทธิภาพของการค้นคืน โดยเฉพาะการค้นคืนเพื่อให้มีความแม่นยำและครอบคลุมกับรายการผลลัพธ์ที่มีอยู่ให้มากที่สุด การให้น้ำหนักคำสำคัญที่อธิบายของแต่ละรายการความรู้จะทำให้การค้นหาได้มีความแม่นยำมากขึ้น เพราะคำสำคัญแต่ละคำจะให้น้ำหนักที่แตกต่างกันในแต่ละรายการความรู้ การศึกษาถึงการให้น้ำหนักคำสำคัญจำแนกตามรายการความรู้ หรือกลุ่มรายการความรู้จึงน่าจะทำให้การค้นคืนมีประสิทธิภาพเพิ่มมากขึ้นได้

3) ออนโทโลยีมีการออกแบบมาสำหรับรองรับการแสดงตามรายการที่ต้องการ (Browsing) การจัดทำระบบเพื่อให้ผู้ใช้เลือกแสดงตามรายการที่ต้องการ และให้มีการประเมินผลลัพธ์แบบแสดงตามรายการด้วย จะทำให้เป็นทางเลือกของผู้ใช้ในการเข้าถึงความรู้ที่ต้องการได้อย่างรวดเร็ว และถูกต้อง โดยที่ผู้ใช้ไม่จำเป็นต้องมีการกำหนดคำค้นเอง เพราะการกำหนดคำค้นเองของผู้ใช้มีข้อจำกัดในแต่ละบุคคล เพราะขึ้นอยู่กับแต่ละประสบการณ์ของแต่ละบุคคลที่จะกำหนดคำค้นให้สอดคล้องหรือตรงตามคำค้นที่มีอยู่ในระบบที่จัดทำขึ้นได้

ข้อเสนอแนะสำหรับการประยุกต์ใช้ตัวแบบการค้นคืนเชิงความหมายสำหรับผลไม้เศรษฐกิจไทยควรมีการเพิ่มจำนวนชนิดผลไม้ให้มากขึ้น เพิ่มวิธีการค้นคืนให้หลากหลายมากขึ้น ดังเช่น การค้นคืนตามระยะเวลา การค้นคืนตามลำดับการทำงาน การค้นคืนตามวัตถุประสงค์ของการใช้งาน เป็นต้น

บรรณานุกรม

กรมวิชาการเกษตร กระทรวงเกษตรและสหกรณ์. (2564). ***การจัดการการผลิตมังคุด***. [ออนไลน์]. เข้าถึงได้จาก: https://www.doa.go.th/hort/wp-content/uploads/2020/01/การผลิตมังคุด.pdf

กรมส่งเสริมการเกษตร กระทรวงเกษตรและสหกรณ์. (2564), ***การผลิตมังคุดนอกฤดู***. [ออนไลน์]. เข้าถึงได้จาก:http://www.agriman.doae.go.th/km62/download/1287leaflet.pdf

คลังข้อมูลสารสนเทศระดับภูมิภาค (ภาคใต้) สำนักงานพัฒนาการวิจัยการเกษตร (องค์การมหาชน). (2564). ***มังคุด***. [ออนไลน์]. เข้าถึงได้จาก: https://www.arda.or.th/kasetinfo/south/mangosteen/controller/01-02.php

คลังข้อมูลสารสนเทศระดับภูมิภาค (ภาคใต้) สำนักงานพัฒนาการวิจัยการเกษตร (องค์การมหาชน). (2564). ***ทุเรียน***. [ออนไลน์]. เข้าถึงได้จาก: https://www.arda.or.th/kasetinfo/south/durian/controller/01-03.php

คู่มือนักวิชาการส่งเสริมการเกษตร. (2564). ***ลำไย***. [ออนไลน์]. เข้าถึงได้จาก: http://www.agriman.doae.go.th/home/t.n/t.n1/3fruit_Requirement/02_Longan.pdf

ทิพวิมล ชมภูคำ. (2020). ***ต้นแบบออนโทโลยีเพื่อการค้นคืนสารสนเทศเชิงความหมายสำหรับข้อมูลโรคติดเชื้อในเด็ก***. การประยุกต์ใช้เทคโนโลยีสารสนเทศ. 6, 2 (ธ.ค. 2020), 72-81.

ศุภกฤษฎิ์ นิวัฒนากูล และฉัตรภัสร์ ฐิติอัคราวงศ์. (2560). ***ระบบถามตอบเชิงความหมายออนไลน์สำหรับผลไม้เศรษฐกิจของไทย***. รายงานการวิจัย สาขาวิชาเทคโนโลยีสารสนเทศ สำนักวิชาเทคโนโลยีสังคม มหาวิทยาลัยเทคโนโลยีสุรนารี.

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

ศูนย์วิจัยพืชสวนจันทบุรี. (2558). ***เทคโนโลยีการผลิตทุเรียนให้มีคุณภาพ.*** กองแผนงานและวิชาการ กรมวิชาการเกษตร กระทรวงเกษตรและสหกรณ์.

ศูนย์วิจัยพืชสวนจันทบุรี. (2564). ***โครงการวิจัย วิจัยและพัฒนาเทคโนโลยีการเพิ่มประสิทธิภาพการผลิตมังคุดคุณภาพ.*** [ออนไลน์]. เข้าถึงได้จาก: https://www.doa.go.th/research/attachment.php?aid=2395

สถาบันวิจัยพืชสวน กรมส่งเสริมการเกษตร. (2564). ***การปลูกมังคุด***. [ออนไลน์]. เข้าถึงได้จาก: https://esc.doae.go.th/ebooks/download-pdf/mangosteen2.pdf

สำนักงานเศรษฐกิจการเกษตร. (2564). ***ตัวชี้วัดเศรษฐฐิจการเกษตรของไทย ปี 2563***. [ออนไลน์]. เข้าถึงได้จาก:https://www.oae.go.th/assets/portals/1/files/jounal/2564/indicator2563.pdf.

สำนักส่งเสริมและฝึกอบรม มหาวิทยาลัยเกษตรศาสตร์. (2564). ***การปลูกลำไย.*** [ออนไลน์]. เข้าถึงได้จาก:https://eto.ku.ac.th/neweto/e-book/plant/tree_fruit/puklamyai.pdf

Acampora, G., Pedrycz, W., Vasilakos, A. V. and Vitiello, A. (2020). ***Computational Intelligence for Semantic Knowledge Management***. Cham: Springer Nature Switzerland.

Allemang, D. and Hendler, J. (2011). ***Semantic Web for the Working Ontologist***. 2nd ed. Waltham: Elsevier Inc.

Bangkhomned, W., and Payakpate, J.. (2020). **Applying Ontology Knowledge Representation Technology and Semantic Searching Methods to Support the Production of High Quality Longan Fruit**. Information Science and Applications. pp.(601-612).

Becerra-Fernandez, I. and Sabherwal, R., (2015). ***Knowledge Management Systems and Processes.*** 2nd ed. New York: Taylor & Francis: Springer-Verlag Berlin Heidelberg.

Calero, C., Ruiz, F. and Piattini, M. (2006). ***Ontologies for Software Engineering and Software Technology***. Berlin: Springer-Verlag.

d'Aquin, M., Ding, L., & Motta, E. (2011). ***Semantic web search engines***. In Handbook of Semantic Web Technologies, John Domingue, Dieter Fensel, & James A. Hendler. p.659-700. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg.

Domingue, J., Fensel, D., & Hendler, J. A. (Eds). ***Handbook of Semantic Web Technologies,*** p.661-679. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg.

Dooley, D.M., Griffiths, E.J., Gosal, G.S. et al. (2018). ***FoodOn: a harmonized food ontology to increase global food traceability, quality control and data integration***. npj Sci Food 2, 23. https://doi.org/10.1038/s41538-018-0032-6

Ensaf, M. and Eyad, S., (2020). ***QSST: A Quranic Semantic Search Tool based on word embedding***. Journal of King Saud University - Computer and Information Sciences. 10.1016/j.jksuci.2020.01.004.

Esmaili, K. S., & Abolhassani, H. (2006). ***A categorization scheme for semantic web search engines***. In the Proceedings of IEEE International Conference on Computer Systems and Applications.

Finnie, T. & et al. ***Swoogle: A Search and Metadata Engine for the SemanticWeb***. In the Proceedings of the 2004 ACM CIKM International Conference on Information and Knowledge Management, Washington, DC, USA, November 8-13, 2004

Formica, A, Pourabbas, E., and Taglino, F. (2020). ***Semantic Search Enhanced with Rating Scores***. Future Internet, 12, 67, 1-10.

Giannopoulos, G., Bikakis, N., Dalamagas, T., Sellis, T. (n.d.). ***GoNTogle: A Tool for Semantic Annotation and Search***. Retrieved from http://web.imis.athena-innovation.gr/~dalamag/pub/gbds-eswc10.pdf

Harth, A., Hogan, A., Umbrich, J., & Decker, S. (2008). ***Building a semantic web search engine: Challenges and solutions***. In the Proceedings of the 3rd XTech Conference, 2008.

Hebeler, J., Fisher, M., Blace, R. and Perez-Lopez, A. (2009). ***Semantic Web Programming***. Indiana: Wiley Publishing, Inc.

Jaccard, P. (1908). ***Nouvelles Recherches Sur La Distribution Florale***. Bulletin de la Société vaudoise des Sciences Naturelles, 44, 223-270.

Jaccard, P. (1912). ***The distribution of the flora in the alpine zone***. New Phytologist, 11, 37-50.

Kiryakov, A., Popov, B., Terziev, I., Manov, D., Ognyanoff, D. (2004). ***Semantic annotation, indexing, and retrieval***. Web Semantics: Science, Services and Agents on the World Wide Web 2 (2004) 49–79. Retrieved from http://infosys3.elfak.ni.ac.rs/nastava/attach/SemantickiWebKurs/sdarticle.pdf

Kiyavitskaya, N., Zeni, N., Cordy, J. R., Mich, L., Mylopoulos, J. (2005). ***Semi-Automatic Semantic Annotations for Web Documents***. Proceedings of the 2nd Italian Semantic Web Workshop (SWAP 2005). University of Trento: Trento, Italy. Retrieved from http://ftp.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-166/27.pdf

Madhu, G., Govardhan, A., & Rajinikanth, T. V. (2011). ***Intelligent semantic web search engines: A brief survey***. International Journal of Web & Semantic Technology (IJWesT), 2(1), 34-42.

Maitaouthong, T. (2019*). **Semantic Web Search Engines***. Journal of Information Science. 37(1). pp. 94-112.

Malve, A., & Chawan, P. M. (2015**). *A comparative study of keyword and Semantic based search engine***. International Journal of Innovative Research in Science, Engineering and Technology**,** 4(11), 11156-11161.

Manning, C.D., Raghavan, P., and Schütze, H., (2008). ***Introduction to Information Retrieval***, 1st ed. Cambridge: Cambridge University Press.

McConnell, Jeffery J. (2001). ***Analysis of Algorithms : An Active Learning Approach***. Canada : Jones and Barlett, pp. 163-168.

North, K and Kumta, G. (2015). ***Knowledge Management Value Creation through Organizational Learning***. 2nd ed. Cham: Springer International Publishing AG.

Noy, N. F. and McGuinness, D. L. (2001). ***Ontology Development 101: A Guide to Creating Your First Ontology.*** Knowledge Systems Laboratory. 32. pp. 1-25.

Ontotext. (2013). ***Semantic Annotation***. Retrieved from http://www.ontotext.com/kim/semantic-annotation

Rashid, J., & Nisar, M. W. (2016). ***A study on semantic searching, semantic search engines and technologies used for semantic search engines***. I.J.Information Technology and Computer Science, 10, 82-89.

Salton, G. (1989) ***Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer***. Addison-Wesley Publishing.

Segaran, T., Evans, C., and Taylor, J. (2009). ***Programming the Semantic Web***. Sebastopol: O'Reilly Media, Inc.

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

Storey, Veda C. & Burton-Jones, Andrew & Vijayan, Sugumaran and Purao, Sandeep. (2003). ***Making the Web More Semantic: A Methodology for Context-Aware Query Processing***. Proceeding of the Twenty-Second International Conference on Conceptual Modeling (ER2003).

Studer, R., Benjamins V.R. and Fensel, D. (1998). ***Knowledge Engineering: Principles and Methods***. Data and Knowledge Engineering**.** vol.25, pp. 161-197.

Sudeepthi, G., Anuradha, G., & Babu, M. S. P. (2012). ***A survey on Semantic Web search engine***. International Journal of Computer Science Issues**,** 9(2), 241-245.

The World Wide Web Consortium (W3C). (2010). ***Search engines.*** Retrieved 18 January 2021, from ttps://www.w3.org/wiki/Search_engines

Z̆iz̆ka, J., Dar̆ena, F., and Svoboda, A., (2020). ***Text Mining with Machine Learning Principles and Techniques***. CRC Press.

มหาวิทยาลัยเทคโนโลยีสุรนารี

ภาคผนวก ก
ผลการประเมินตัวแบบการค้นคืนเชิงความหมาย

มหาวิทยาลัยเทคโนโลยีสุรนารี

คำถาม	จำนวนคำตอบ		Context classification			Precision	Recall	F-measure
	ค้นพบ	เกี่ยวข้อง	TP	FP	FN			
Q1	3	3	3	0	1	100.00	75.00	85.71
Q2	4	4	4	0	0	100.00	100.00	100.00
Q3	6	6	6	0	0	100.00	100.00	100.00
Q4	3	3	3	0	2	100.00	60.00	75.00
Q5	1	1	1	0	0	100.00	100.00	100.00
Q6	4	3	3	1	0	75.00	100.00	85.71
Q7	6	6	6	0	0	100.00	100.00	100.00
Q8	4	2	2	2	0	50.00	100.00	66.67
Q9	1	1	1	0	0	100.00	100.00	100.00
Q10	2	2	2	0	0	100.00	100.00	100.00
Q11	1	1	1	0	1	100.00	50.00	66.67
Q12	4	4	4	0	0	100.00	100.00	100.00
Q13	4	4	4	0	0	100.00	100.00	100.00
Q14	2	2	2	0	2	100.00	50.00	66.67
Q15	1	1	1	0	0	100.00	100.00	100.00
Q16	4	3	3	1	0	75.00	100.00	85.71
Q17	5	5	5	0	0	100.00	100.00	100.00
Q18	3	2	2	1	0	66.67	100.00	80.00
Q19	1	1	1	0	0	100.00	100.00	100.00
Q20	1	1	1	0	0	100.00	100.00	100.00
Q21	6	6	6	0	1	100.00	85.71	92.31
Q22	4	4	4	0	0	100.00	100.00	100.00
Q23	6	6	6	0	4	100.00	60.00	75.00
Q24	2	2	2	0	3	100.00	40.00	57.14
Q25	4	1	1	3	1	25.00	50.00	33.33
Q26	4	2	2	2	3	50.00	40.00	44.44
Q27	4	4	4	0	4	100.00	50.00	66.67
Q28	8	8	8	0	0	100.00	100.00	100.00

มหาวิทยาลัยเทคโนโลยีสุรนารี

คำถาม	จำนวนคำตอบ		Context classification			Precision	Recall	F-measure
	ค้นพบ	เกี่ยวข้อง	TP	FP	FN			
Q29	4	4	4	0	1	100.00	80.00	88.89
Q30	5	5	5	0	3	100.00	62.50	76.92
เฉลี่ย						91.39	83.44	87.23

มหาวิทยาลัยเทคโนโลยีสุรนารี

ภาคผนวก ข
รายละเอียดโปรแกรม

มหาวิทยาลัยเทคโนโลยีสุรนารี

Index.jsp

```
<%@page contentType="text/html;charset=UTF-8"%>
<% request.setCharacterEncoding("UTF-8"); %>
<%@ taglib uri="WEB-INF/mytags.tld" prefix="m" %>
<html>
<head>
  <META http-equiv="Content-Type" content="text/html;charset=UTF-8">
  <title>Semantic search model for economic fruits!!! </title>
</head>
<body>
  <center>
 <h1>!!!  Semantic search model for economic fruits  !!!</h1>
   <blockquote>
     <h2>คำชี้แจง </h2>
     <h3>ระบบการสืบค้นความรู้เกี่ยวกับผลไม้เศรษฐกิจของไทยนี้เป็นส่วนหนึ่งของผลการวิจัย <br>
    ประยุกต์ใช้ฐานความรู้ด้วยตัวแบบออนโทโลยีกับภาษาค้นหาและภาษากฎสำหรับการค้นคืนเชิงความหมาย <br>
  ความรู้ผลไม้เศรษฐกิจสามชนิด ได้แก่ ทุเรียน มังคุด และลำไย เน้นการผลิตและการดูแลรักษานำเสนอเป็นตัวอย่าง
<br> <br>
   สามารถระบุชื่อผลไม้และคำค้นประกอบการค้นหาได้ ดังเช่น ทุเรียน พันธุ์ปลูก </h3>
        <form action="index.jsp" method="POST">
          <input type="text" name="qstring" value="${param.qstring}" size="50">          <br><br>
       <input type="submit" value="ค้นหา">
      </form>
      <br>
      </center>
      <m:sparqllab qstring="${param.qstring}"></m:sparqllab>
   </blockquote>
</html>
```

myTag.tld

```
<?xml version="1.0" encoding="ISO-8859-1" ?>

<!DOCTYPE taglib
        PUBLIC "-//Sun Microsystems, Inc.//DTD JSP Tag Library 1.2//EN"
    "http://java.sun.com/j2ee/dtd/web-jsptaglibrary_1_2.dtd">

<taglib>

  <tlib-version>1.0</tlib-version>
  <jsp-version>1.2</jsp-version>
  <short-name>simple</short-name>
  <uri>http://tomcat.apache.org/example-taglib</uri>

<tag>
  <name>sparqllab</name>
  <tag-class>fruits.econ_fruits.econ_fruits</tag-class>
  <attribute>
    <name>qstring</name>
    <required>true</required>
    <rtexprvalue>true</rtexprvalue>
  </attribute>
</tag>

</taglib>
```

มหาวิทยาลัยเทคโนโลยีสุรนารี

Econ_fruits.java

```
package fruits.econ_fruits;

import java.util.ArrayList;
import java.util.Calendar;
import java.util.Collections;
import java.util.Iterator;
import java.util.List;
import java.util.StringTokenizer;
import java.util.TreeSet;
import java.util.HashSet;
import java.lang.Math;

import javax.servlet.jsp.JspException;
import javax.servlet.jsp.JspWriter;
import javax.servlet.jsp.tagext.TagSupport;

import com.hp.hpl.jena.query.Query;
import com.hp.hpl.jena.query.QueryExecution;
import com.hp.hpl.jena.query.QueryExecutionFactory;
import com.hp.hpl.jena.query.QueryFactory;
import com.hp.hpl.jena.query.QuerySolution;
import com.hp.hpl.jena.query.ResultSet;
import com.hp.hpl.jena.query.ResultSetFormatter;
import com.hp.hpl.jena.rdf.model.Literal;
import com.hp.hpl.jena.rdf.model.Model;
import com.hp.hpl.jena.util.FileManager;

public class econ_fruits extends TagSupport {

   private String qstring = "";
   public void setQstring(String qstring) {
       this.qstring = qstring;
   }
```

มหาวิทยาลัยเทคโนโลยีสุรนารี

```
public int doStartTag() throws JspException {
    String qs, qs1;
    String keyword_list, sim_s, str_tmp, str_tmp_1, str_tmp_2, str_tmp_3;
    String keyword_list_1, str_fruit="",result_list_1, str_sf="", str_inp="";
    String str_sim;
    String result_1, result_2;
    int flag_show = 5; // 0 = not show and 1.. = show
    JspWriter out = pageContext.getOut();
    try {
        if (!qstring.equals("")) {
            // To prepare the keyword(s) from user
            StringTokenizer st = new StringTokenizer(qstring," ");
            ArrayList<String> qs_s = new ArrayList<String>();
            while (st.hasMoreTokens())
            {
             qs = st.nextToken();
               qs_s.add(qs);
            }
            // To declare the variables
            char[] qs_c, key_c, inter_c, union_c, item_cos;
            char sim_c;
            int[] keyw_sim;
            float sim, union, intersect;
            float[][] sim_list;
            int num_key = 0,num_list,num_list_1, num_fruit;
            int flag_add; // For adding keyword in the list
            int flag_fruit=0; // For searching specific fruit
            QuerySolution soln_2;
            Literal var_1, var_2, var_3;
            HashSet<Character> union_qs = new HashSet<Character>(); // For keywords
            HashSet<Character> union_inp = new HashSet<Character>(); // For input data
            HashSet<Character> union_all = new HashSet<Character>(); // For all data
            HashSet<Character> inters_all = new HashSet<Character>(); // For all data
            HashSet<String> key_list = new HashSet<String>(); // For all data
            HashSet<String> search_list =  new HashSet<String>(); //  For keywords and their
collection
```

```
HashSet<String> fruit_list = new HashSet<String>(); // For keywords contain economic fruits
HashSet<String> result_con = new HashSet<String>(); // For the list of result from concepts
HashSet<String> result_des = new HashSet<String>(); // For the list of result from description
HashSet<String> result_all = new HashSet<String>(); // For the list of all result

HashSet<String> result_all_1 = new HashSet<String>(); // For the list of all result

HashSet<String> result_cosine = new HashSet<String>(); // For the list of all result
// To find concepts by using keywords from users
FileManager.get().addLocatorClassLoader(econ_fruits.class.getClassLoader());
Model model = FileManager.get().loadModel("EconFruits.owl");
qstring = "PREFIX skos: <http://www.w3.org/2004/02/skos/core#>\r\n" +
        "PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>\r\n" +
        "PREFIX ef: <http://www.fruits.com/econ-fruits#>\r\n" +
        "SELECT ?keyword \r\n" +
        "WHERE { \r\n" +
        "?x rdf:type skos:Concept . \r\n" +
        "{?x skos:prefLabel ?keyword .} \r\n" +
        "UNION {?x skos:altLabel ?keyword . } \r\n" +
        "}";

Query query = QueryFactory.create(qstring);
QueryExecution qexec = QueryExecutionFactory.create(query, model);
ResultSet results = qexec.execSelect();

// To calculate the similarity value
if (flag_show<1) out.print("<br><h3> แสดงรายละเอียดจากการค้นหา: </h3><br>");
List<String> list = results.getResultVars();
if (flag_show<1) out.print("<table>");
while (results.hasNext()) {
    union_qs.clear();
    flag_add = 0;
    if (flag_show<1) out.print("<tr>");
    QuerySolution soln = results.nextSolution();
```

```
Literal var1 = soln.getLiteral(list.get(0));
keyword_list = ""+var1;
if (flag_show<1) out.print("<td>" + keyword_list + " -> ");

qs_c = keyword_list.toCharArray();
for (int j=0; j<qs_c.length; j++ ) {
    union_qs.add(qs_c[j]);
}
if (flag_show<1) {
    out.print(" Union=" + union_qs.size());
    Iterator u_ks = union_qs.iterator();
    while (u_ks.hasNext()) {
        out.print(u_ks.next() + "|");
    }
}
if (flag_show<1) out.print("</td>");
if (flag_show<1) out.print("<td>");
sim_s = "";
for (int i=0; i < qs_s.size(); i++) {
    union_inp.clear();
    union_all.clear();
    if (flag_show<1) out.print("<td>" + qs_s.get(i) + " -> ");
    qs1 = qs_s.get(i).toString();
    key_c = qs1.toCharArray();
    for (int j=0; j<key_c.length; j++ ) {
        union_inp.add(key_c[j]);
        // if (flag_show<1) out.print(key_c[j] + "|");
    }
    if (flag_show<1) {
        out.print(" Union=" + union_inp.size());
        Iterator u_kin = union_inp.iterator();
        while (u_kin.hasNext()) {
            out.print(u_kin.next() + "|");
        }
    }
```

```
        union_all.addAll(union_inp);
        union_all.addAll(union_qs);
        if (flag_show<1) {
            out.print(" Union All =" + union_all.size());
            Iterator u_qs = union_all.iterator();
            while (u_qs.hasNext()) {
                out.print(u_qs.next() + "|");
            }
        }
        if (flag_show<1) out.print("</td>");

        // Intersection
        if (flag_show<1) out.print("<td>");
        union_inp.retainAll(union_qs);
        if (flag_show<1) {
            out.print(" Intersection all =" + union_inp.size());
            Iterator u_ins = union_inp.iterator();
            while (u_ins.hasNext()) {
                out.print(u_ins.next() + "|");
            }
        }

        union = union_all.size();
        intersect = union_inp.size();
        sim = intersect / union;
        if (sim>=0.70) {
            flag_add=1;
            str_inp = qs_s.get(i);
            sim_s = sim_s + "|" + String.valueOf(sim)  + "|" + qs_s.get(i);

        }
        if (flag_show<1) out.print("U=" + union + " I=" + intersect + " Sim=" + sim);
    }
    if (flag_add==1) key_list.add(keyword_list + "|" + "0.0");
    if (flag_show<1) out.print("</td>");
    if (flag_show<1) out.print("</td>");
    if (flag_show<1) out.print("</tr>");
```

มหาวิทยาลัยเทคโนโลยีสุรนารี

```
    num_key++;
}
if (flag_show<2) out.print("</table>");

// To classify the collection of searching keywords
if (flag_show<2) out.print("<table>");
Iterator key_search = key_list.iterator();
while (key_search.hasNext()) {
    num_list=1;
    str_tmp = key_search.next().toString();
    if (flag_show<2) out.print("<tr><td> str_tmp = " + str_tmp + "</td></tr>");
    if (flag_show<2) out.print("<tr>");
    StringTokenizer k_list = new StringTokenizer(str_tmp,"|");
    ArrayList<String> ks_list = new ArrayList<String>();
    str_tmp_1 = "";
   while (k_list.hasMoreTokens())  {
    str_sim = k_list.nextToken();
    if (num_list==0) { // to find collection
        if (flag_show<2) out.print("<td>");
        if (flag_show<2) out.print(str_tmp_1 + " -> " + str_sim);
        if (flag_show<2) out.print("</td>");
        qstring = "PREFIX skos: <http://www.w3.org/2004/02/skos/core#>\r\n" +
                  "PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>\r\n" +
                  "PREFIX ef: <http://www.fruits.com/econ-fruits#>\r\n" +
                  "SELECT ?collection \r\n" +
                  "WHERE { \r\n" +
                  "?col rdf:type skos:Collection . \r\n" +
                  "?col skos:member ?id . \r\n" +
                  "?col skos:prefLabel ?collection . \r\n" +
                  "?id skos:prefLabel ?concept . \r\n" +
                  "FILTER regex(?concept,'" + str_tmp_1 + "') . \r\n" +
                  "}";

        if (flag_show<2) out.print("qstring = " + qstring + "<-END");

            query = QueryFactory.create(qstring);
            qexec = QueryExecutionFactory.create(query, model);
```

```
                    results = qexec.execSelect();
                    list = results.getResultVars();

                    while (results.hasNext()) {
                        QuerySolution soln_1 = results.nextSolution();
                        Literal var1_1 = soln_1.getLiteral(list.get(0));
                        keyword_list_1 = ""+var1_1;
                        if ( flag_show<2)  out. print( " <td> keyword_list_1 = " +
keyword_list_1.toString() + "</td>");
                        if (keyword_list_1.equals("ผลไม้เศรษฐกิจ")) {
                            flag_fruit++; // For searching in specific fruit
                            fruit_list.add(str_tmp_1);
                            if (flag_show<2) out.print("<td> -> Fruit domain search </td>");
                        } else {
                            search_list.add(str_tmp_1 + "|" + str_sim);
                        }
                    } // End while results
                    num_list=1;
                } else if (num_list==1) { // End if num_list = 0
                    str_tmp_1 = str_sim;
                    num_list = 0;
                }

                // num_list++;
               }

               if (flag_show<2) out.print("</tr>");
            }   // end while key_search
            if (flag_show<2) out.print("</table>");

            if (flag_show<3) out.print("<table>");
            if (flag_show<3) out.print("<tr><td> Show concepts </td?</tr>");
            // for economic fruit
            key_search = fruit_list.iterator();
            while (key_search.hasNext()) {
                str_tmp_2 = key_search.next().toString();
                if (flag_show<3) out.print("<tr><td>Econ fruit = " + str_tmp_2 + "</td></tr>");
```

มหาวิทยาลัยเทคโนโลยีสุรนารี

```
}
// For keyword list
key_search = search_list.iterator();
while (key_search.hasNext()) {
    str_tmp_2 = key_search.next().toString();
    if (flag_show<3) out.print("<tr><td>Keyword = " + str_tmp_2 + "</td></tr>");
}
if (flag_show<3) out.print("</table>");
// Searching by the list of keywords
// Case the keyword contains some econ frutis
if (flag_show<3) out.print("<table>");
if (flag_show<3) out.print("<tr><td> Show searching keywords </td?</tr>");
if (flag_fruit==1) { // case keyword contain some econ fruits
    Iterator fruit_search = fruit_list.iterator();
    while (fruit_search.hasNext()) {
    num_list=0; // For creating only first column
        str_tmp = fruit_search.next().toString();

        StringTokenizer f_list = new StringTokenizer(str_tmp,"|");
        if (flag_show<3) out.print("<tr>");
       while (f_list.hasMoreTokens())  {
        str_tmp_1 = f_list.nextToken();
        str_sf = str_tmp_1.toString();
            if (str_tmp_1.matches("มังคุด")) str_fruit = "ef:Mangosteen";
            else if (str_tmp_1.matches("ทุเรียน")) str_fruit = "ef:Durian";
            else if (str_tmp_1.matches("ลำไย")) str_fruit = "ef:Longan";
            else str_fruit = "None";
            if (flag_show<3) out.print("<tr><td>Fruit = " + str_fruit + "</td></tr>");
        if (num_list==0) {
                key_search = search_list.iterator(); //To start keyword search
                while (key_search.hasNext()) {
                num_list_1 = 1; // For creating only fist column
                    str_tmp_2 = key_search.next().toString();
                    str_tmp_3="";
                    StringTokenizer c_list = new StringTokenizer(str_tmp_2,"|");
                    while (c_list.hasMoreTokens()) {
                        str_sim = c_list.nextToken();
```

```
                    if (flag_show<3) out.print("<tr><td>str_tmp_3 = " + str_tmp_3 + "
str_sim = " + str_sim + "</td></tr>");
                    if (num_list_1==0) {
                        if (flag_show<3) out.print("<tr><td>Keyword in num_list_1=0 =
" + str_tmp_3 + "</td></tr>");
                        qstring = " PREFIX skos:
<http://www.w3.org/2004/02/skos/core#>\r\n" +
                                "PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-
ns#>\r\n" +
                                "PREFIX ef: <http://www.fruits.com/econ-fruits#>\r\n" +
                                "SELECT ?concept ?result  \r\n" +
                                "WHERE { \r\n" +
                                " ?id rdf:type " + str_fruit + " . \r\n" +
                                " { ?id ef:hasDescription ?result . \r\n" +
                                " ?id ef:forExplain ?con . \r\n" +
                                " ?con skos:prefLabel ?concept . } \r\n" +
                                " UNION { ?id ef:hasDescription ?result . \r\n" +
                                " ?id ef:forPractice ?con . \r\n" +
                                " ?con skos:prefLabel ?concept . } \r\n" +

                                " UNION { ?id ef:hasDescription ?result . \r\n" +
                                " ?id ef:forPrevent ?con . \r\n" +
                                " ?con skos:prefLabel ?concept . } \r\n" +

                                " UNION { ?id ef:hasDescription ?result . \r\n" +
                                " ?id ef:forTreatment ?con . \r\n" +
                                " ?con skos:prefLabel ?concept . } \r\n" +

                                " FILTER regex(?concept, '" + str_tmp_3 + "') . \r\n" +
                                "}";

                    if ( flag_show<3)  out. print( " <tr><td>Query = " + qstring +
"</td></tr>");

                    // Add the result to the list
                        query = QueryFactory.create(qstring);
                        qexec = QueryExecutionFactory.create(query, model);
```

```
                                    results = qexec.execSelect();
                                    list = results.getResultVars();
                                    while (results.hasNext()) {
                                        soln_2 = results.nextSolution();
                                        var_1 = soln_2.getLiteral(list.get(0));
                                        result_1 = ""+var_1;
                                        var_2 = soln_2.getLiteral(list.get(1));
                                        result_2 = ""+var_2;
                                        result_con.add(str_sf + "|" + result_2 + "|" + result_1 + "|" +
str_sim.toString());
                                        if (flag_show<3) out.print("<tr><td>Result:" + str_sf.toString()
+ " -> " +
                                            result_2.toString() + " ," + result_1.toString() + ", " +
str_sim.toString() + "</td></tr>");
                                    } // end while results
                                    num_list_1=1;
                                } else if (num_list_1==1) {
                                    str_tmp_3 = str_sim;
                                    num_list_1=0;
                                } // end if num_list_1 = 0
                            } // end while c_list
                        } // end whie key_search
                        num_list=1;
                    } // end if num_list = 0
                } // end while f_list

                // Searching keyword in description
                num_key = qs_s.size();
                for (int i=0;i<num_key;i++) {
                    str_tmp_3 = qs_s.get(i);
                    if ( ! ( str_tmp_3. equals( " ทุเรียน " ) || str_tmp_3. equals( " มังคุด " ) ||
str_tmp_3.equals("ลำไย") )) {
                        if (flag_show<3) out.print("<tr><td>Search Keyword in description = " +
str_tmp_3 + "</td></tr>");
                        qstring = "PREFIX skos: <http://www.w3.org/2004/02/skos/core#>\r\n" +
                            "PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>\r\n" +
                            "PREFIX ef: <http://www.fruits.com/econ-fruits#>\r\n" +
```

```
                "SELECT ?concept ?result  \r\n" +
                "WHERE { \r\n" +
                " ?id rdf:type " + str_fruit + " . \r\n" +
                " { ?id ef:hasDescription ?result . \r\n" +
                " ?id ef:forExplain ?con . \r\n" +
                " ?con skos:prefLabel ?concept . } \r\n" +
                " UNION { ?id ef:hasDescription ?result . \r\n" +
                " ?id ef:forPractice ?con . \r\n" +
                " ?con skos:prefLabel ?concept . } \r\n" +

                " UNION { ?id ef:hasDescription ?result . \r\n" +
                " ?id ef:forPrevent ?con . \r\n" +
                " ?con skos:prefLabel ?concept . } \r\n" +

                " UNION { ?id ef:hasDescription ?result . \r\n" +
                " ?id ef:forTreatment ?con . \r\n" +
                " ?con skos:prefLabel ?concept . } \r\n" +

                " FILTER regex(?result, '" + str_tmp_3 + "') . \r\n" +
                "}";

        if (flag_show<3) out.print("<tr><td>Query = " + qstring + "</td></tr>");

        query = QueryFactory.create(qstring);
            qexec = QueryExecutionFactory.create(query, model);
            results = qexec.execSelect();
            list = results.getResultVars();
            while (results.hasNext()) {
                soln_2 = results.nextSolution();
                var_1 = soln_2.getLiteral(list.get(0));
                result_1 = ""+var_1;
                var_2 = soln_2.getLiteral(list.get(1));
                result_2 = ""+var_2;
                result_des.add(str_sf + "|" + result_2 + "|" + str_tmp_3 + "|" + result_1
);
                if (flag_show<3) out.print("<tr><td>Result:" + str_sf.toString() + " -> " +
```

```
                                result_2.toString() + " ," + result_1.toString() + "</td></tr>");

                            } // end while results
                        } // end if fruits
                    } // end for num_key
                } // end while fruit_search
            } else { // case keyword don't contain any econ fruits
                key_search = search_list.iterator();
                while (key_search.hasNext()) {
                num_list_1 = 0; // For creating only fist column
                    str_tmp_2 = key_search.next().toString();
                    str_tmp_3="";
                    StringTokenizer c_list = new StringTokenizer(str_tmp_2,"|");
                    while (c_list.hasMoreTokens()) {
                        str_sim = c_list.nextToken();
                        str_tmp_3 = String.valueOf(str_sim);
                        if (flag_show<3) out.print("<tr><td>str_tmp_3 = " + str_tmp_3 + " str_sim = "
+ str_sim + "</td></tr>");
                        if (num_list_1==0) {
                            for (int i=0;i<3;i++) {
                                if (i==0) str_fruit = "ef:Mangosteen";
                                else if (i==1) str_fruit = "ef:Durian";
                                else if (i==2) str_fruit = "ef:Longan";
                                else str_fruit = "None";
                                if (flag_show<3)  out.print("<tr><td>Keyword in num_list_1=0 =  "  +
str_tmp_3 + "</td></tr>");
                                qstring = "PREFIX skos: <http://www.w3.org/2004/02/skos/core#>\r\n"
+
                                        " PREFIX  rdf:   <http: / / www. w3. org/ 1999/ 02/ 22- rdf- syntax-
ns#>\r\n" +
                                        "PREFIX ef: <http://www.fruits.com/econ-fruits#>\r\n" +
                                        "SELECT ?concept ?result  \r\n" +
                                        "WHERE { \r\n" +
                                        " ?id rdf:type " + str_fruit + " . \r\n" +
                                        " { ?id ef:hasDescription ?result . \r\n" +
                                        " ?id ef:forExplain ?con . \r\n" +
                                        " ?con skos:prefLabel ?concept . } \r\n" +
```

```
                    " UNION { ?id ef:hasDescription ?result . \r\n" +
                    " ?id ef:forPractice ?con . \r\n" +
                    " ?con skos:prefLabel ?concept . } \r\n" +

                    " UNION { ?id ef:hasDescription ?result . \r\n" +
                    " ?id ef:forPrevent ?con . \r\n" +
                    " ?con skos:prefLabel ?concept . } \r\n" +

                    " UNION { ?id ef:hasDescription ?result . \r\n" +
                    " ?id ef:forTreatment ?con . \r\n" +
                    " ?con skos:prefLabel ?concept . } \r\n" +

                    " FILTER regex(?concept, '" + str_tmp_3 + "') . \r\n" +
                    "}";

                if (flag_show<3) out.print("<tr><td>Query = " + qstring + "</td></tr>");
                    if (i==0) str_fruit = "มังคุด";
                    else if (i==1) str_fruit = "ทุเรียน";
                    else if (i==2) str_fruit = "ลำไย";
                    else str_fruit = "None";

                // Add the result to the list
                    query = QueryFactory.create(qstring);
                    qexec = QueryExecutionFactory.create(query, model);
                    results = qexec.execSelect();
                    list = results.getResultVars();
                    while (results.hasNext()) {
                        soln_2 = results.nextSolution();
                        var_1 = soln_2.getLiteral(list.get(0));
                        result_1 = ""+var_1;
                        var_2 = soln_2.getLiteral(list.get(1));
                        result_2 = ""+var_2;
                        result_con.add(str_fruit.toString() + "|" + result_2 + "|" + result_1 +
"|" + str_sim.toString());
                        if (flag_show<3) out.print("<tr><td>Result:" + str_fruit.toString() + "
-> " +
```

```
                                    result_2.toString() + " ," + result_1.toString() + ", " +
str_sim.toString() + "</td></tr>");
                                } // end while results
                                num_list_1=1;
                            } // end for i
                        } else if (num_list_1==1) {
                            str_tmp_3 = str_sim;
                            num_list_1=0;
                        } // end if num_list_1 = 0
                    } // end while c_list
                } // end while key_search

                // Searching keyword in description
                num_key = qs_s.size();
                for (int j=0;j<num_key;j++) {
                    str_tmp_3 = qs_s.get(j);
                    for (int i=0;i<3;i++) {
                        if (i==0) str_fruit = "ef:Mangosteen";
                        else if (i==1) str_fruit = "ef:Durian";
                        else if (i==2) str_fruit = "ef:Longan";
                        else str_fruit = "None";
                        //  if (!(str_tmp_3.equals("ทุเรียน") || str_tmp_3.equals("มังคุด") ||
str_tmp_3.equals("ลำไย") )) {
                            if (flag_show<3) out.print("<tr><td>Search Keyword in description = " +
str_tmp_3 + "</td></tr>");
                            qstring = "PREFIX skos: <http://www.w3.org/2004/02/skos/core#>\r\n" +
                                "PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>\r\n" +
                                "PREFIX ef: <http://www.fruits.com/econ-fruits#>\r\n" +
                                "SELECT ?concept ?result  \r\n" +
                                "WHERE { \r\n" +
                                " ?id rdf:type " + str_fruit + " . \r\n" +
                                " { ?id ef:hasDescription ?result . \r\n" +
                                " ?id ef:forExplain ?con . \r\n" +
                                " ?con skos:prefLabel ?concept . } \r\n" +
                                " UNION { ?id ef:hasDescription ?result . \r\n" +
                                " ?id ef:forPractice ?con . \r\n" +
```

```
                " ?con skos:prefLabel ?concept . } \r\n" +

                " UNION { ?id ef:hasDescription ?result . \r\n" +
                " ?id ef:forPrevent ?con . \r\n" +
                " ?con skos:prefLabel ?concept . } \r\n" +

                " UNION { ?id ef:hasDescription ?result . \r\n" +
                " ?id ef:forTreatment ?con . \r\n" +
                " ?con skos:prefLabel ?concept . } \r\n" +

                " FILTER regex(?result, '" + str_tmp_3 + "') . \r\n" +
                "}";

            if (flag_show<3) out.print("<tr><td>Query = " + qstring + "</td></tr>");
                if (i==0) str_fruit = "มังคุด";
                else if (i==1) str_fruit = "ทุเรียน";
                else if (i==2) str_fruit = "ลำไย";
                else str_fruit = "None";

            query = QueryFactory.create(qstring);
                qexec = QueryExecutionFactory.create(query, model);
                results = qexec.execSelect();
                list = results.getResultVars();
                while (results.hasNext()) {
                    soln_2 = results.nextSolution();
                    var_1 = soln_2.getLiteral(list.get(0));
                    result_1 = ""+var_1;
                    var_2 = soln_2.getLiteral(list.get(1));
                    result_2 = ""+var_2;
                    //result_des.add(str_fruit + "|" + result_2 + "|" + result_1 + "|" + "0.4" );
                    result_des.add(str_fruit + "|" + result_2 + "|" + str_tmp_3 + "|" + result_1
);
                    if (flag_show<3) out.print("<tr><td>Result:" + str_fruit.toString() + " ->
" +
                        result_2.toString() + " ," + result_1.toString() + "</td></tr>");

                } // end while results
```

```
                    // } // end if fruits
                } // end for search by description
            } // end for num_key

        } // end search without any fruit

        // for result list by concept
        if (flag_show<4) out.print("<table>");
        if ( flag_show<4)  out. print( " <tr><td> Show the list of result by searching by concepts
</td></tr>");
        ArrayList<String> array_con = new ArrayList<String>();
        ArrayList<String> array_des = new ArrayList<String>();
        String str_result;
        int num_result;
        Iterator list_of_result = result_con.iterator();
        while (list_of_result.hasNext()) {
            str_tmp_2 = list_of_result.next().toString();
            StringTokenizer row_list = new StringTokenizer(str_tmp_2,"|");
            str_result="";
            num_result=0;
            if (flag_show<4) out.print("<tr>");
            while (row_list.hasMoreTokens()) {
                str_tmp_3 = row_list.nextToken();
                array_con.add(str_tmp_3);
                if (num_result==0) {
                    str_result= str_tmp_3;
                    num_result=1;
                } else if (num_result==1) {
                    str_result = str_result + "|" + str_tmp_3;
                    num_result=2;
                } else if (num_result==2) {
                    str_result = str_result + "|" + str_tmp_3;
                    result_all.add(str_result);
                    num_result=3;
                }
                if (flag_show<4) out.print("<td>" + str_tmp_3.toString() + "</td>");
            }
```

```
            if (flag_show<4) out.print("</tr>");
        }
        if (flag_show<4) out.print("</table>");

        // for result list by text in description
        if (flag_show<4) out.print("<table>");
        if (flag_show<4) out.print("<tr><td> Show the list of result by searching by text in description </td></tr>");
        list_of_result = result_des.iterator();
        while (list_of_result.hasNext()) {
            str_tmp_2 = list_of_result.next().toString();
            StringTokenizer row_list = new StringTokenizer(str_tmp_2,"|");
            if (flag_show<4) out.print("<tr>");
            num_result=0;
            str_result="";
            while (row_list.hasMoreTokens()) {
                str_tmp_3 = row_list.nextToken();
                array_des.add(str_tmp_3);
                if (num_result==0) {
                    str_result= str_tmp_3;
                    num_result=1;
                } else if (num_result==1) {
                    str_result = str_result + "|" + str_tmp_3;
                    num_result=2;
                } else if (num_result==2) {
                    str_result = str_result + "|" + str_tmp_3;
                    result_all.add(str_result);
                    num_result=3;
                }
                if (flag_show<4) out.print("<td>" + str_tmp_3.toString() + "</td>");
            }
            if (flag_show<4) out.print("</tr>");
        }
        if (flag_show<4) out.print("</table>");

        // for keyword list
        if (flag_show<4) out.print("<table>");
```

```
            if (flag_show<4) out.print("<tr><td> Show the list of concepts and their similarity value
</td></tr>");
            Iterator list_of_keyword = key_list.iterator();
            while (list_of_keyword.hasNext()) { // for print keyword and similarity value
                str_tmp_2 = list_of_keyword.next().toString();
                StringTokenizer kr_list = new StringTokenizer(str_tmp_2,"|");
                if (flag_show<4) out.print("<tr>");
               while (kr_list.hasMoreTokens())  {
                str_tmp_3 = kr_list.nextToken();
                if (flag_show<4) out.print("<td>" + str_tmp_3 + "</td>");
               }
                if (flag_show<4) out.print("</tr>");
            }
            if (flag_show<4) out.print("</table>");

            //show the result list
            if (flag_show<5) {
                out.print("<table>");
                out.print("<tr><td>Show the result from concept </td></tr>");
                out.print("<tr>");
                for (int i=0; i<array_con.size(); i++) {
                    if (i%4==0) out.print("</tr><tr>");
                    out.print("<td> i=" + i + " ->" + array_con.get(i) + "</td>");
                }
                out.print("</tr>");
                out.print("<tr><td>Show the result from description </td></tr>");
                out.print("<tr>");
                for (int i=0; i<array_des.size(); i++) {
                    if (i%4==0) out.print("</tr><tr>");
                    out.print("<td> i=" + i + " ->" + array_des.get(i) + "</td>");
                }
                out.print("</tr>");
                out.print("</table>");
            } // end if flag_show

            // Manage the list in result before calculation cosine
            String result_before="";
```

มหาวิทยาลัยเทคโนโลยีสุรนารี

```
Iterator result_all_list_before = result_all.iterator();
Double cosine_value=0.0;
if (flag_show<5) out.print("<table>");
if ( flag_show<5)  out. print( " <tr><td>Show all result in the list before calculation cosine</td></tr>");
while (result_all_list_before.hasNext()) { // for print keyword and similarity value
    str_tmp_2 = result_all_list_before.next().toString();
    StringTokenizer result_before_token = new StringTokenizer(str_tmp_2,"|");
    result_before="";
    num_result=0;
    if (flag_show<5) out.print("<tr>");
   while (result_before_token.hasMoreTokens())  {
    str_tmp_3 = result_before_token.nextToken();
    if (num_result%3==0) result_before = str_tmp_3;
    if (num_result%3==1) result_before = result_before + "|" + str_tmp_3;
    num_result++;
   }
if (flag_show<5) out.print("<tr><td> result before = " + result_before + "</td></tr>");

    if (flag_show<5) out.print("</tr>");
    result_cosine.add(result_before);

}
out.print("</table>");

// Calculation the cosine value
String str_fruit_qry="",str_text_qry="",str_result_cos="";
String str_keyword_con="",str_keyword_des="";
String str_concept_list="",str_cos="";
int cos_qry=0,cos_inp=0,flag_sim=0;
Iterator result_cosine_iter = result_cosine.iterator();
if (flag_show<5) out.print("<table>");
if (flag_show<5) out.print("<tr><td>Show the calculation of cosine value </td></tr>");


while (result_cosine_iter.hasNext()) { // for print keyword and similarity value
    str_tmp_2 = result_cosine_iter.next().toString();
```

```
//str_result_sim= String.valueOf(str_tmp_2);
StringTokenizer result_cosine_token = new StringTokenizer(str_tmp_2,"|");
num_result=0;
// Collect the data from result list
while (result_cosine_token.hasMoreTokens())  {
str_tmp_3 = result_cosine_token.nextToken();
if (num_result%3==0) str_fruit_qry = String.valueOf(str_tmp_3);
if (num_result%3==1) str_text_qry = String.valueOf(str_tmp_3);
num_result++;
} // end while result_cosine_token
if (flag_show<5) out.print("<tr><td>.....</td></tr>");
if (flag_show<5) out.print("<tr><td> fruit query = " + str_fruit_qry + "</td>");
if (flag_show<5) out.print("<td> text query = " + str_text_qry + "</td></tr>");

// for matching concept between query and database
num_result=0;
if (flag_fruit==1) {
    str_cos = "1";
    cos_inp = 1;
    cos_qry = 1;
} else {
    str_cos = "";
    cos_inp = 0;
    cos_qry = 0;
}

Iterator keyword_list_iter = key_list.iterator();
while (keyword_list_iter.hasNext()) {
    str_tmp_2 = keyword_list_iter.next().toString();
    StringTokenizer keyword_list_token = new StringTokenizer(str_tmp_2,"|");
    flag_sim=0;
    str_concept_list = keyword_list_token.nextToken();
    if ( flag_show<5)  out. print( " <tr><td> concept list=  "  +  str_concept_list +
"</td><td>Text= " + str_text_qry + "</td></tr>");
    // Searching in result list by concept
    for (int i=0; i<array_con.size(); i+=4) {
```

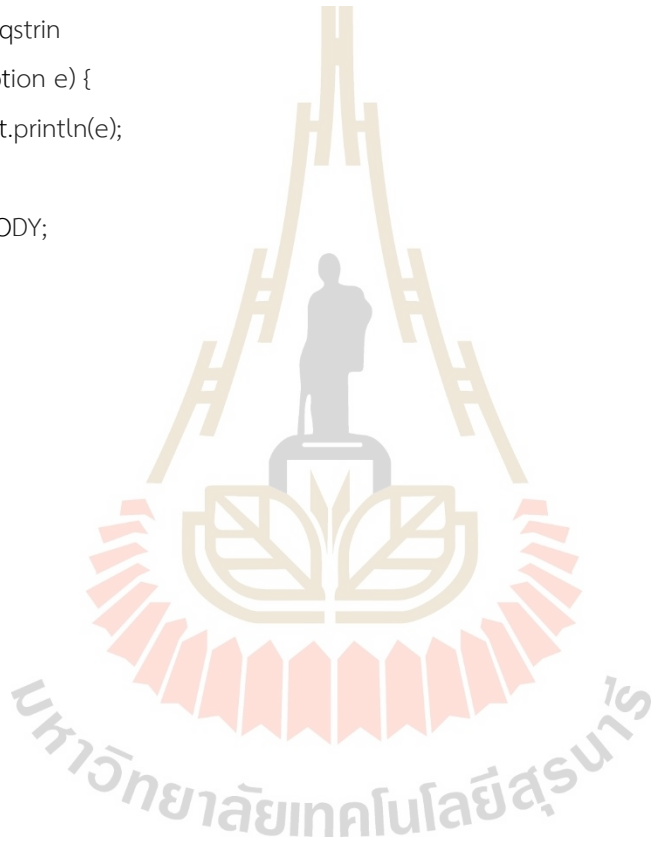
มหาวิทยาลัยเทคโนโลยีสุรนารี

```
                    if( array_con. get( i+ 1) . contentEquals( str_text_qry) &&
array_con.get(i+2).contentEquals(str_concept_list)) {
                        str_cos = str_cos + "1";
                        cos_qry++;
                        cos_inp++;
                        flag_sim=1;
                        str_keyword_con = str_concept_list;
                        if (flag_show<5) out.print("<tr><td>Match con = " + array_con.get(i+2) +
"</td>");
                        if ( flag_show<5) out. print( " <td>Match text " + array_con. get( i+ 1) +
"</td></tr>");
                    } else {
                        //if (flag_show<5) out.print("<tr><td>No Match con = " + array_con.get(i+2)
+ "</td>");
                        //if (flag_show<5) out.print("<td>No Match text " + array_con.get(i+1) +
"</td></tr>");
                    }


                } // end for array_con
                    if (flag_sim!=1) {
                    str_cos = str_cos + "0";
                    cos_inp++;
                    }
                    flag_sim=0;
                } // end keyword_list_iter
                if (flag_show<5) out.print("<tr><td> Dot product= " + str_cos + "</td>");
                if (flag_show<5) out.print("<td>cos_qry= " + String.valueOf(cos_qry) + "</td>");
                if (flag_show<5) out.print("<td>cos_inp= " + String.valueOf(cos_inp) + "</td></tr>");


                // Search in result list by text in description
                for (int j=0;j<qs_s.size();j++) {
                    //str_keyword_inp = qs_s.get(j);
                    flag_sim=0;
                    for (int i=0; i<array_des.size();i+=4) {
```

```
                    if( array_des. get( i+ 1) . contentEquals( str_text_qry) &&
array_des.get(i+2).contentEquals(qs_s.get(j))) {
                        str_cos = str_cos + "1";
                        cos_qry++;
                        cos_inp++;
                        flag_sim=1;
                        str_keyword_des = qs_s.get(j);
                        if (flag_show<5) out.print("<tr><td>Match word = " + array_des.get(i+2) +
"</td>");
                        if ( flag_show<5) out. print( " <td>Match text " + array_des. get( i+ 1) +
"</td></tr>");
                    } else {
                        //if (flag_show<5) out.print("<tr><td>No Match word = " + array_des.get(i+2)
+ "</td>");
                        //if (flag_show<5) out. print( " <td>No Match text " + array_des. get( i+ 1) +
"</td></tr>");
                    }
                    } // end array_des
                    if (flag_sim!=1) {
                    str_cos = str_cos + "0";
                    cos_inp++;
                    }
                    flag_sim=0;
                } // End for qs_s
                if (flag_show<5) out.print("<tr><td> Dot product= " + str_cos + "</td>");
                if (flag_show<5) out.print("<td>cos_qry= " + String.valueOf(cos_qry) + "</td>");
                if (flag_show<5) out.print("<td>cos_inp= " + String.valueOf(cos_inp) + "</td></tr>");

                // Calculate cosine value each item
                cosine_value = (cos_qry / (Math.sqrt(cos_qry) * Math.sqrt(cos_inp))) * 100;
                str_result_cos = String.valueOf(cosine_value) + "|" + str_fruit_qry + "|" + str_text_qry
+ "|"
                    + String.valueOf(Math.round(cosine_value+.5) + "|" + str_keyword_con + "|" +
str_keyword_des);
                if (cosine_value>=70.0) {
                    result_all_1.add(str_result_cos);
```

```
                    if (flag_show<5)  out.print("<tr><td> New value of result:  "  +  str_result_cos +  "
Cosine = " +
                        String.valueOf(cosine_value) + "</td></tr>");
                } else {
                    if (flag_show<5) out.print("<tr><td> No add to a new result: " + str_result_cos +
" Cosine = " +
                        String.valueOf(cosine_value) + "</td></tr>");
                }

            } // end while result_cosine_iter
            if (flag_show<5) out.print("</table>");

            // Show the result
            int item_list=1;
            String str_result_sim="";
            if (result_all_1.isEmpty()) {
                out.print("<br><h3> ไม่พบผลลัพธ์จากคำค้นที่กำหนด....!!! </h3><br>");
            } else {
                item_list = 1;
                TreeSet<String> result_sort = new TreeSet<String>(Collections.reverseOrder());
                result_sort.addAll(result_all_1);
                Iterator result_display = result_sort.iterator();
                out.print("<br><h3> แสดงผลลัพธ์ที่ได้จากการค้นหา: </h3><br>");
                out.print("<table border='1'>");
                out.print("<th>ลำดับ</th><th>ไม้ผลเศรษฐกิจไทย</th><th>รายการความรู้</th>");

                while (result_display.hasNext()) { // for print keyword and similarity value
                    str_tmp_2 = result_display.next().toString();
                    str_result_sim=str_tmp_2;
                    StringTokenizer result_display_token = new StringTokenizer(str_tmp_2,"|");
                    out.print("<tr>");
                    out.print("<td align='center'>" + String.valueOf(item_list) + "</td>");
                    num_result=0;
                    item_list++;
                   while (result_display_token.hasMoreTokens())  {
                    str_tmp_3 = result_display_token.nextToken();
                    if (num_result==1) out.print("<td align='center'>" + str_tmp_3 + "</td>");
```

```
                    if (num_result==2) out.print("<td>" + str_tmp_3 + "</td>");
                    //if (num_result==3) out.print("<td align='center'>" + str_tmp_3 + "%</td>");

                    num_result++;
                   } // end while result_display_token
                    out.print("</tr>");
                    }
                    out.print("</table>");
            }

            } // end if qstrin
        } catch (Exception e) {
            System.out.println(e);
        }
        return SKIP_BODY;
    }
}
```



ภาคผนวก ค
ผลการประเมินระบบค้นคืนเชิงความหมาย

Case	TP	FP	FN	P	R	F
Q1	2	0	0	100.00	100.00	100.00
Q2	1	0	0	100.00	100.00	100.00
Q3	1	0	0	100.00	100.00	100.00
Q4	6	0	4	100.00	60.00	75.00
Q5	15	0	1	100.00	93.75	96.77
Q6	4	0	0	100.00	100.00	100.00
Q7	3	1	0	75.00	100.00	85.71
Q8	1	0	1	100.00	50.00	66.67
Q9	3	0	0	100.00	100.00	100.00
Q10	6	0	0	100.00	100.00	100.00
Q11	4	1	0	80.00	100.00	88.89
Q12	2	0	0	100.00	100.00	100.00
Q13	5	0	0	100.00	100.00	100.00
Q14	1	0	0	100.00	100.00	100.00
Q15	1	0	2	100.00	33.33	50.00
Q16	1	0	3	100.00	25.00	40.00
Q17	2	0	5	100.00	28.57	44.44
Q18	3	0	0	100.00	100.00	100.00
Q19	3	0	0	100.00	100.00	100.00
Q20	7	2	0	77.78	100.00	87.50
Q21	8	0	0	100.00	100.00	100.00

Case	TP	FP	FN	P	R	F
Q22	3	0	0	100.00	100.00	100.00
Q23	2	0	0	100.00	100.00	100.00
Q24	2	0	0	100.00	100.00	100.00
Q25	5	0	0	100.00	100.00	100.00
Q26	3	0	1	100.00	75.00	85.71
Q27	2	0	0	100.00	100.00	100.00
Q28	1	0	0	100.00	100.00	100.00
Q29	2	0	0	100.00	100.00	100.00
Q30	3	0	2	100.00	60.00	75.00
Q31	1	0	1	100.00	50.00	66.67
Q32	7	0	0	100.00	100.00	100.00
Q33	4	0	0	100.00	100.00	100.00
Q34	1	0	0	100.00	100.00	100.00
Q35	3	0	0	100.00	100.00	100.00
Q36	8	0	0	100.00	100.00	100.00
Q37	2	0	0	100.00	100.00	100.00
Q38	4	0	0	100.00	100.00	100.00
Q39	1	0	1	100.00	50.00	66.67
Q40	5	0	0	100.00	100.00	100.00
Q41	1	0	1	100.00	50.00	66.67
Q42	7	0	0	100.00	100.00	100.00

Case	TP	FP	FN	P	R	F
Q43	1	0	0	100.00	100.00	100.00
Q44	2	0	0	100.00	100.00	100.00
Q45	2	0	0	100.00	100.00	100.00
Q46	1	0	0	100.00	100.00	100.00
Q47	2	0	0	100.00	100.00	100.00
Q48	4	0	0	100.00	100.00	100.00
Q49	2	1	0	66.67	100.00	80.00
Q50	1	0	0	100.00	100.00	100.00
Q51	4	3	0	57.14	100.00	72.73
Q52	2	1	0	66.67	100.00	80.00
Q53	1	2	0	33.33	100.00	50.00
Q54	1	1	0	50.00	100.00	66.67
Q55	1	1	0	50.00	100.00	66.67
Q56	1	1	0	50.00	100.00	66.67
Q57	1	1	0	50.00	100.00	66.67
Q58	1	1	0	50.00	100.00	66.67
Q59	1	1	0	50.00	100.00	66.67
Q60	1	2	0	33.33	100.00	50.00
Q61	6	2	0	75.00	100.00	85.71
Q62	1	4	0	20.00	100.00	33.33
Q63	1	4	0	20.00	100.00	33.33
Q64	2	2	0	50.00	100.00	66.67
Q65	1	3	0	25.00	100.00	40.00
Q66	1	4	0	20.00	100.00	33.33

Case	TP	FP	FN	P	R	F
Q67	2	2	0	50.00	100.00	66.67
Q68	1	3	1	25.00	50.00	33.33
Q69	1	1	0	50.00	100.00	66.67
Q70	1	2	0	33.33	100.00	50.00
Q71	3	13	0	18.75	100.00	31.58
Q72	2	0	0	100.00	100.00	100.00
Q73	1	7	0	12.50	100.00	22.22
Q74	2	4	0	33.33	100.00	50.00
Q75	3	7	0	30.00	100.00	46.15
Q76	2	8	0	20.00	100.00	33.33
Q77	3	0	5	100.00	37.50	54.55
Q78	5	2	0	71.43	100.00	83.33
Q79	2	3	0	40.00	100.00	57.14
Q80	2	0	0	100.00	100.00	100.00
Q81	1	1	0	50.00	100.00	66.67
Q82	1	3	0	25.00	100.00	40.00
Q83	1	1	0	50.00	100.00	66.67
Q84	1	1	0	50.00	100.00	66.67
Q85	1	3	0	25.00	100.00	40.00
Q86	1	3	0	25.00	100.00	40.00
Q87	1	0	0	100.00	100.00	100.00
Q88	2	0	0	100.00	100.00	100.00
Q89	2	8	0	20.00	100.00	33.33
Q90	1	0	0	100.00	100.00	100.00

มหาวิทยาลัยเทคโนโลยีสุรนารี

Case	TP	FP	FN	P	R	F
Q91	1	0	0	100.00	100.00	100.00
Q92	1	1	0	50.00	100.00	66.67
Q93	3	6	0	33.33	100.00	50.00
Q94	1	2	1	33.33	50.00	40.00
Q95	1	0	0	100.00	100.00	100.00
Q96	2	16	0	11.11	100.00	20.00
Q97	3	0	0	100.00	100.00	100.00
Q98	5	4	0	55.56	100.00	71.43
Q99	1	16	0	5.88	100.00	11.11
Q100	1	5	0	16.67	100.00	28.57
Q101	2	3	0	40.00	100.00	57.14
Q102	1	0	0	100.00	100.00	100.00
Q103	1	1	0	50.00	100.00	66.67
Q104	2	0	0	100.00	100.00	100.00
Q105	2	0	0	100.00	100.00	100.00
Q106	1	0	0	100.00	100.00	100.00
Q107	3	0	0	100.00	100.00	100.00
Q108	2	1	3	66.67	40.00	50.00
Q109	3	0	1	100.00	75.00	85.71
Q110	1	0	0	100.00	100.00	100.00
Q111	1	0	3	100.00	25.00	40.00
Q112	1	0	0	100.00	100.00	100.00
Q113	1	0	0	100.00	100.00	100.00
Q114	5	1	0	83.33	100.00	90.91

Case	TP	FP	FN	P	R	F
Q115	8	0	0	100.00	100.00	100.00
Q116	2	0	0	100.00	100.00	100.00
Q117	1	0	0	100.00	100.00	100.00
Q118	1	0	0	100.00	100.00	100.00
Q119	1	0	0	100.00	100.00	100.00
Q120	7	0	0	100.00	100.00	100.00
Q121	3	0	0	100.00	100.00	100.00
Q122	3	0	0	100.00	100.00	100.00
Q123	3	0	0	100.00	100.00	100.00
Q124	3	0	0	100.00	100.00	100.00
Q125	3	0	0	100.00	100.00	100.00
Q126	11	0	2	100.00	84.62	91.67
Q127	2	0	0	100.00	100.00	100.00
Q128	2	0	0	100.00	100.00	100.00
Q129	2	0	0	100.00	100.00	100.00
Q130	2	0	0	100.00	100.00	100.00
Q131	1	0	0	100.00	100.00	100.00
Q132	1	0	0	100.00	100.00	100.00
Q133	4	0	0	100.00	100.00	100.00
Q134	1	0	0	100.00	100.00	100.00
Q135	1	0	1	100.00	50.00	66.67
Q136	5	0	0	100.00	100.00	100.00
Q137	3	0	0	100.00	100.00	100.00
Q138	3	0	0	100.00	100.00	100.00

มหาวิทยาลัยเทคโนโลยีสุรนารี

Case	TP	FP	FN	P	R	F
Q139	2	0	3	100.00	40.00	57.14
Q140	3	0	0	100.00	100.00	100.00
Q141	3	0	0	100.00	100.00	100.00
Q142	2	0	0	100.00	100.00	100.00
Q143	2	0	0	100.00	100.00	100.00
Q144	2	0	1	100.00	66.67	80.00
Q145	1	0	0	100.00	100.00	100.00
Q146	2	0	0	100.00	100.00	100.00
Q147	1	0	2	100.00	33.33	50.00
Q148	1	0	3	100.00	25.00	40.00
Q149	1	0	0	100.00	100.00	100.00
Q150	1	0	0	100.00	100.00	100.00
Q151	1	9	0	10.00	100.00	18.18
Q152	2	8	0	20.00	100.00	33.33
Q153	1	8	0	11.11	100.00	20.00
Q154	1	3	0	25.00	100.00	40.00
Q155	1	1	0	50.00	100.00	66.67
Q156	1	3	0	25.00	100.00	40.00
Q157	1	5	2	16.67	33.33	22.22
Q158	1	4	1	20.00	50.00	28.57
Q159	2	4	1	33.33	66.67	44.44
Q160	1	3	0	25.00	100.00	40.00
Q161	1	4	0	20.00	100.00	33.33
Q162	2	6	0	25.00	100.00	40.00

Case	TP	FP	FN	P	R	F
Q163	1	3	0	25.00	100.00	40.00
Q164	1	6	0	14.29	100.00	25.00
Q165	2	2	0	50.00	100.00	66.67
Q166	1	3	0	25.00	100.00	40.00
Q167	1	2	0	33.33	100.00	50.00
Q168	1	1	0	50.00	100.00	66.67
Q169	1	4	0	20.00	100.00	33.33
Q170	1	2	0	33.33	100.00	50.00
Q171	1	5	0	16.67	100.00	28.57
Q172	2	6	0	25.00	100.00	40.00
Q173	1	1	0	50.00	100.00	66.67
Q174	1	0	2	100.00	33.33	50.00
Q175	2	2	2	50.00	50.00	50.00
Q176	1	6	0	14.29	100.00	25.00
Q177	2	3	0	40.00	100.00	57.14
Q178	1	4	0	20.00	100.00	33.33
Q179	2	6	0	25.00	100.00	40.00
Q180	4	1	0	80.00	100.00	88.89
Q181	4	2	0	66.67	100.00	80.00
Q182	2	4	0	33.33	100.00	50.00
Q183	1	1	0	50.00	100.00	66.67
Q184	3	1	0	75.00	100.00	85.71
Q185	2	3	0	40.00	100.00	57.14
Q186	1	5	0	16.67	100.00	28.57

มหาวิทยาลัยเทคโนโลยีสุรนารี

Case	TP	FP	FN	P	R	F
Q187	1	1	0	50.00	100.00	66.67
Q188	3	14	0	17.65	100.00	30.00
Q189	2	0	0	100.00	100.00	100.00
Q190	1	1	0	50.00	100.00	66.67
Q191	1	1	0	50.00	100.00	66.67
Q192	2	4	0	33.33	100.00	50.00
Q193	1	3	0	25.00	100.00	40.00
Q194	3	4	0	42.86	100.00	60.00
Q195	3	1	0	75.00	100.00	85.71
Q196	1	0	0	100.00	100.00	100.00
Q197	1	6	0	14.29	100.00	25.00
Q198	3	4	0	42.86	100.00	60.00
Q199	1	3	0	25.00	100.00	40.00
Q200	1	4	0	20.00	100.00	33.33
Q201	16	2	0	88.89	100.00	94.12
Q202	1	6	0	14.29	100.00	25.00
Q203	2	3	1	40.00	66.67	50.00
Q204	3	0	0	100.00	100.00	100.00
Q205	2	0	0	100.00	100.00	100.00
Q206	3	0	0	100.00	100.00	100.00
Q207	4	0	1	100.00	80.00	88.89
Q208	2	0	0	100.00	100.00	100.00
Q209	5	0	0	100.00	100.00	100.00
Q210	4	0	0	100.00	100.00	100.00

Case	TP	FP	FN	P	R	F
Q211	2	0	0	100.00	100.00	100.00
Q212	2	0	0	100.00	100.00	100.00
Q213	2	0	0	100.00	100.00	100.00
Q214	7	0	0	100.00	100.00	100.00
Q215	2	0	1	100.00	66.67	80.00
Q216	2	0	0	100.00	100.00	100.00
Q217	2	0	1	100.00	66.67	80.00
Q218	1	0	0	100.00	100.00	100.00
Q219	2	0	0	100.00	100.00	100.00
Q220	8	0	0	100.00	100.00	100.00
Q221	4	0	0	100.00	100.00	100.00
Q222	1	0	0	100.00	100.00	100.00
Q223	10	0	0	100.00	100.00	100.00
Q224	2	0	0	100.00	100.00	100.00
Q225	1	0	0	100.00	100.00	100.00
Q226	5	0	3	100.00	62.50	76.92
Q227	9	0	0	100.00	100.00	100.00
Q228	1	0	0	100.00	100.00	100.00
Q229	1	0	0	100.00	100.00	100.00
Q230	2	0	0	100.00	100.00	100.00
Q231	1	0	0	100.00	100.00	100.00
Q232	2	0	1	100.00	66.67	80.00
Q233	3	0	0	100.00	100.00	100.00
Q234	8	0	0	100.00	100.00	100.00

มหาวิทยาลัยเทคโนโลยีสุรนารี

Case	TP	FP	FN	P	R	F
Q235	1	0	0	100.00	100.00	100.00
Q236	5	0	0	100.00	100.00	100.00
Q237	2	0	0	100.00	100.00	100.00
Q238	1	0	0	100.00	100.00	100.00
Q239	3	0	0	100.00	100.00	100.00
Q240	3	0	0	100.00	100.00	100.00
Q241	2	0	0	100.00	100.00	100.00
Q242	1	0	0	100.00	100.00	100.00
Q243	2	0	0	100.00	100.00	100.00
Q244	1	0	1	100.00	50.00	66.67
Q245	1	0	0	100.00	100.00	100.00
Q246	2	0	0	100.00	100.00	100.00
Q247	3	0	0	100.00	100.00	100.00
Q248	3	0	0	100.00	100.00	100.00
Q249	7	0	0	100.00	100.00	100.00
Q250	1	0	0	100.00	100.00	100.00
Q251	1	5	0	16.67	100.00	28.57
Q252	1	7	0	12.50	100.00	22.22
Q253	1	6	0	14.29	100.00	25.00
Q254	4	5	0	44.44	100.00	61.54
Q255	2	2	0	50.00	100.00	66.67
Q256	2	5	0	28.57	100.00	44.44
Q257	1	2	0	33.33	100.00	50.00
Q258	1	3	0	25.00	100.00	40.00

Case	TP	FP	FN	P	R	F
Q259	3	6	0	33.33	100.00	50.00
Q260	2	3	0	40.00	100.00	57.14
Q261	1	4	0	20.00	100.00	33.33
Q262	1	5	0	16.67	100.00	28.57
Q263	1	4	0	20.00	100.00	33.33
Q264	2	4	0	33.33	100.00	50.00
Q265	1	3	0	25.00	100.00	40.00
Q266	2	4	0	33.33	100.00	50.00
Q267	2	1	0	66.67	100.00	80.00
Q268	1	0	0	100.00	100.00	100.00
Q269	1	10	0	9.09	100.00	16.67
Q270	2	9	0	18.18	100.00	30.77
Q271	1	2	0	33.33	100.00	50.00
Q272	1	7	0	12.50	100.00	22.22
Q273	2	7	0	22.22	100.00	36.36
Q274	2	3	0	40.00	100.00	57.14
Q275	2	2	0	50.00	100.00	66.67
Q276	3	7	0	30.00	100.00	46.15
Q277	3	14	0	17.65	100.00	30.00
Q278	2	7	0	22.22	100.00	36.36
Q279	1	9	0	10.00	100.00	18.18
Q280	1	9	0	10.00	100.00	18.18
Q281	1	4	0	20.00	100.00	33.33
Q282	1	3	0	25.00	100.00	40.00

มหาวิทยาลัยเทคโนโลยีสุรนารี

Case	TP	FP	FN	P	R	F
Q283	2	7	0	22.22	100.00	36.36
Q284	1	7	0	12.50	100.00	22.22
Q285	1	8	0	11.11	100.00	20.00
Q286	1	16	0	5.88	100.00	11.11
Q287	2	6	0	25.00	100.00	40.00
Q288	1	9	0	10.00	100.00	18.18
Q289	1	7	0	12.50	100.00	22.22
Q290	2	5	0	28.57	100.00	44.44
Q291	2	9	0	18.18	100.00	30.77
Q292	1	11	0	8.33	100.00	15.38
Q293	1	8	0	11.11	100.00	20.00
Q294	1	10	0	9.09	100.00	16.67
Q295	2	1	0	66.67	100.00	80.00
Q296	1	2	0	33.33	100.00	50.00
Q297	1	6	0	14.29	100.00	25.00
Q298	1	6	0	14.29	100.00	25.00
Q299	1	3	0	25.00	100.00	40.00
Q300	1	9	0	10.00	100.00	18.18
Q301	1	0	0	100.00	100.00	100.00
Q302	2	6	0	25.00	100.00	40.00
Q303	1	4	0	20.00	100.00	33.33
Q304	1	3	0	25.00	100.00	40.00
Q305	1	10	0	9.09	100.00	16.67
Q306	2	7	0	22.22	100.00	36.36

Case	TP	FP	FN	P	R	F
Q307	2	7	0	22.22	100.00	36.36
Q308	1	2	0	33.33	100.00	50.00
Q309	1	2	0	33.33	100.00	50.00
Q310	1	2	0	33.33	100.00	50.00
Q311	1	6	0	14.29	100.00	25.00
Q312	1	2	0	33.33	100.00	50.00
Q313	1	0	0	100.00	100.00	100.00
Q314	1	8	0	11.11	100.00	20.00
Q315	1	4	0	20.00	100.00	33.33
Q316	1	5	1	16.67	50.00	25.00
Q317	3	3	0	50.00	100.00	66.67
Q318	1	4	0	20.00	100.00	33.33
Q319	1	2	0	33.33	100.00	50.00
Q320	1	5	0	16.67	100.00	28.57
Q321	2	7	0	22.22	100.00	36.36
Q322	1	4	0	20.00	100.00	33.33
Q323	1	8	0	11.11	100.00	20.00
Q324	2	2	0	50.00	100.00	66.67
Q325	2	1	0	66.67	100.00	80.00
Q326	1	1	0	50.00	100.00	66.67
Q327	2	1	0	66.67	100.00	80.00
Q328	1	7	0	12.50	100.00	22.22
Q329	1	3	0	25.00	100.00	40.00
Q330	1	7	0	12.50	100.00	22.22

มหาวิทยาลัยเทคโนโลยีสุรนารี

Case	TP	FP	FN	P	R	F
Q331	2	7	0	22.22	100.00	36.36
Q332	1	3	0	25.00	100.00	40.00
Q333	1	6	0	14.29	100.00	25.00
Q334	1	3	0	25.00	100.00	40.00
Q335	2	4	0	33.33	100.00	50.00
Q336	1	9	0	10.00	100.00	18.18
Q337	1	6	0	14.29	100.00	25.00
Q338	1	7	0	12.50	100.00	22.22
Q339	1	0	0	100.00	100.00	100.00
Q340	1	0	0	100.00	100.00	100.00
Q341	2	0	0	100.00	100.00	100.00
Q342	1	1	0	50.00	100.00	66.67
Q343	1	2	0	33.33	100.00	50.00
Q344	1	4	0	20.00	100.00	33.33
Q345	1	2	0	33.33	100.00	50.00
Q346	1	2	0	33.33	100.00	50.00
Q347	1	7	0	12.50	100.00	22.22
Q348	1	3	0	25.00	100.00	40.00
Q349	2	7	0	22.22	100.00	36.36
Q350	1	4	0	20.00	100.00	33.33
Q351	2	2	0	50.00	100.00	66.67
Q352	1	3	0	25.00	100.00	40.00
Q353	1	3	0	25.00	100.00	40.00
Q354	1	5	0	16.67	100.00	28.57

Case	TP	FP	FN	P	R	F
Q355	1	5	0	16.67	100.00	28.57
Q356	1	3	0	25.00	100.00	40.00
Q357	1	4	0	20.00	100.00	33.33
Q358	1	6	0	14.29	100.00	25.00
Q359	1	0	0	100.00	100.00	100.00
Q360	1	6	0	14.29	100.00	25.00
Q361	1	5	0	16.67	100.00	28.57
Q362	2	1	0	66.67	100.00	80.00
Q363	2	2	0	50.00	100.00	66.67
Q364	2	0	0	100.00	100.00	100.00
Q365	1	0	0	100.00	100.00	100.00
Q366	1	0	0	100.00	100.00	100.00
Q367	1	0	0	100.00	100.00	100.00
Q368	3	1	0	75.00	100.00	85.71
Q369	1	0	0	100.00	100.00	100.00
Q370	2	2	0	50.00	100.00	66.67
Q371	1	4	0	20.00	100.00	33.33
Q372	1	0	0	100.00	100.00	100.00
Q373	1	2	0	33.33	100.00	50.00
Q374	2	0	0	100.00	100.00	100.00
Q375	1	3	0	25.00	100.00	40.00
Q376	1	1	0	50.00	100.00	66.67
Q377	1	4	0	20.00	100.00	33.33
Q378	1	2	0	33.33	100.00	50.00

มหาวิทยาลัยเทคโนโลยีสุรนารี

Case	TP	FP	FN	P	R	F
Q379	1	1	0	50.00	100.00	66.67
Q380	1	3	0	25.00	100.00	40.00
Q381	1	0	0	100.00	100.00	100.00
Q382	1	0	0	100.00	100.00	100.00
Q383	1	0	0	100.00	100.00	100.00
Q384	1	0	0	100.00	100.00	100.00
Q385	1	0	0	100.00	100.00	100.00
Q386	1	0	0	100.00	100.00	100.00
Q387	1	0	0	100.00	100.00	100.00
Q388	1	0	0	100.00	100.00	100.00
Q389	1	0	0	100.00	100.00	100.00

Case	TP	FP	FN	P	R	F
Q390	1	0	0	100.00	100.00	100.00
Q391	1	0	0	100.00	100.00	100.00
Q392	1	0	0	100.00	100.00	100.00
Q393	1	1	0	50.00	100.00	66.67
Q394	1	0	0	100.00	100.00	100.00
Q395	1	0	0	100.00	100.00	100.00
Q396	1	0	0	100.00	100.00	100.00
Q397	1	0	0	100.00	100.00	100.00
Q398	1	0	0	100.00	100.00	100.00
Q399	1	0	0	100.00	100.00	100.00
Q400	1	0	0	100.00	100.00	100.00

หมายเหตุ : Case หมายถึง กรณีทดสอบ

TP หมายถึง จำนวนผลลัพธ์ที่ค้นคืนได้และเกี่ยวข้อง (True positive)

FP หมายถึง จำนวนผลลัพธ์ที่ค้นคืนได้แต่ไม่เกี่ยวข้อง (False positive)

FN หมายถึง จำนวนผลลัพธ์ที่ค้นคืนไม่ได้แต่เกี่ยวข้อง (False negative)

P หมายถึง ค่าความแม่นยำ (Precision)

R หมายถึง ค่าความระลึก (Recall)

F หมายถึง ค่าอัตราการรู้จำ (F-Measure)

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 และฉัตรภัสร์ ฐิติอัคราวงศ์. (2560). ระบบถามตอบเชิงความหมายออนไลน์สำหรับผลไม้เศรษฐกิจของไทย.
- 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