แบบจำลองการสืบค้นเชิงความหมายสำหรับความรู้ด้านพืชไร่

นายเจษฎา สิงห์ทองชัย

มหาวิทยาลัยเทคโนโลยีสุรนารี

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาการสารสนเทศดุษฎีบัณฑิต

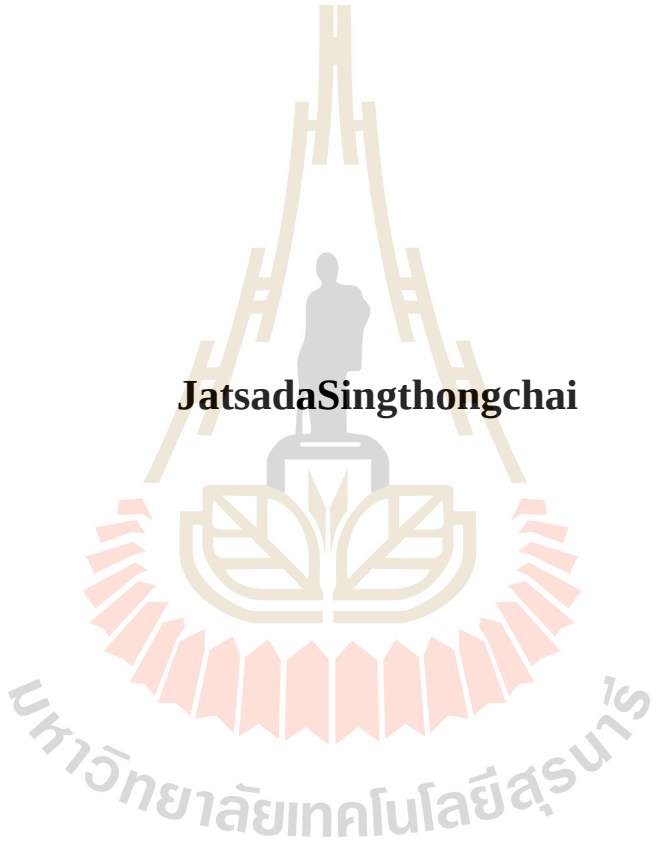
สาขาวิชาเทคโนโลยีสารสนเทศ

มหาวิทยาลัยเทคโนโลยีสุรนารี

ปีการศึกษา2559

A SEMANTIC SEARCH MODEL FOR AGRONOMY KNOWLEDGE

JatsadaSingthongchai



A Thesis Submitted in Partial Fulfillment of the Requirements for the

Degree of Doctor of Information Science in Information Technology

Suranaree University of Technology

Academic Year 2016

แบบจำลองการสืบค้นเชิงความหมายสำหรับความรู้ด้านพืชไร่

มหาวิทยาลัยเทคโนโลยีสุรนารี อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาดุษฎีบัณฑิต

คณะกรรมการสอบวิทยานิพนธ์

(รองศาสตราจารย์ ดร.วีรพงษ์ พลนิกรกิจ)
ประธานกรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.ศุภกฤษฎิ์ นิวัฒนากูล)
กรรมการ (อาจารย์ที่ปรึกษาวิทยานิพนธ์)

(รองศาสตราจารย์ ดร.นิตยา เกิดประสพ)
กรรมการ

(อาจารย์ ดร.ชูชาติ หฤไชยะศักดิ์)
กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.จิติมนต์ อั่งสกุล)
กรรมการ

(ศาสตราจารย์ ดร.ชูกิจ ลิมปีจำนงค์)
รองอธิการบดีฝ่ายวิชาการและนวัตกรรม

(รองศาสตราจารย์ ดร.วีรพงษ์ พลนิกรกิจ)
คณบดีสำนักวิชาเทคโนโลยีสังคม

มหาวิทยาลัยเทคโนโลยีสุรนารี

เจษฎา สิงห์ทองชัย : แบบจำลองการสืบค้นเชิงความหมายสำหรับความรู้พืชไร่ (A SEMANTIC SEARCH MODEL FOR AGRONOMY KNOWLEDGE) อาจารย์ที่ปรึกษา : ผู้ช่วยศาสตราจารย์ ดร.ศุภกฤษฎิ์ นิวัฒนากูล, 159 หน้า.

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อออกแบบและพัฒนาแบบจำลองการสืบค้นเชิงความหมายสำหรับความรู้ด้านพืชไร่ ประกอบด้วย 2 ส่วน คือการจัดความรู้ให้เป็นระบบ และการสืบค้นเชิงความหมาย 3 ขั้นตอน ได้แก่ การกำหนดลำดับความสำคัญคำค้น การให้น้ำหนักคำสำคัญ และการวัดความคล้ายคลึงเชิงความหมาย โดยรวบรวมข้อมูลความรู้ด้านพืชไร่ของประเทศไทย 3 ชนิด คือ มันสำปะหลัง อ้อย และข้าวโพด และเลือกรายการความรู้ที่มีคำอธิบายครบถ้วนจากแหล่งข้อมูลที่มีความน่าเชื่อถือ และมีการเผยแพร่ตั้งแต่ปี พ.ศ. 2550 เป็นต้น นำมาจัดเก็บในฐานความรู้ที่ประยุกต์ใช้โครงสร้างของคลาสในออนโทโลยีไม้ผลเศรษฐกิจ ซึ่งประเมินประสิทธิภาพการสืบค้นด้วยค่าความแม่นยำ ค่าความระลึก ค่าความถ่วงดุล ค่าความแม่นยำเฉลี่ย และค่าความผิดพลาด ผลการวิจัยพบว่า

1. การออกแบบจำลองการสืบค้นเชิงความหมายสำหรับความรู้ด้านพืช ประกอบด้วยการทำงาน 2 ส่วน คือ การจัดความรู้ให้เป็นระบบ และการสืบค้นเชิงความหมาย

2. การประเมินประสิทธิภาพการให้น้ำหนักคำสำคัญด้วยการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ มีค่าความแม่นยำเฉลี่ย เท่ากับ 0.906 และมีค่าความผิดพลาด เท่ากับ 0.097 ซึ่งให้ผลดีกว่าทั้งในด้านของความแม่นยำเฉลี่ย และมีค่าความผิดพลาดที่น้อยกว่าวิธีการให้น้ำหนักด้วยไอดีเอฟ ซึ่งเป็นไปตามสมมติฐานที่ตั้งไว้

3. การประเมินประสิทธิภาพการวัดความคล้ายคลึงเชิงความหมาย

3.1 ผลการเปรียบเทียบประสิทธิภาพการวัดความคล้ายคลึงจากการสืบค้น ใช้ค่าความถ่วงดุลที่สะท้อนจากค่าความแม่นยำ และค่าความระลึกของรูปแบบการทดลองคำนวณค่าระยะทางสั้นที่สุดระหว่างคำค้นกับคำในความรู้หรือเอกสาร มาคำนวณกับน้ำหนักคำสำคัญด้วยการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ และค่าลำดับความสำคัญคำค้นจากผู้ใช้ และค่าลำดับความสำคัญคำค้นจากตำแหน่งของคำค้นที่ปรากฏในความรู้ โดยให้ค่าความถ่วงดุลสูงที่สุด มีค่าความถ่วงดุลเท่ากับ 0.8596

3.2 ผลการเปรียบเทียบประสิทธิภาพการวัดความคล้ายคลึงจากการสืบค้น ใช้ค่าความแม่นยำเฉลี่ยของรูปแบบการทดลองค่าระยะทางสั้นที่สุดระหว่างคำค้นกับคำในความรู้หรือเอกสาร มาคำนวณกับสัดส่วนของจำนวนเอกสารทั้งหมด จำนวนของเอกสารที่ปรากฏคำที่สนใจ ค่าลำดับความสำคัญคำค้นจากผู้ใช้ และค่าลำดับความสำคัญคำค้นจากตำแหน่งของคำค้นที่ปรากฏในความรู้ โดยให้ค่าความแม่นยำเฉลี่ยสูงที่สุด มีค่าความแม่นยำเฉลี่ย เท่ากับ 0.8661

สาขาวิชาเทคโนโลยีสารสนเทศ ลายมือชื่อนักศึกษา_____

ปีการศึกษา 2559 ลายมือชื่ออาจารย์ที่ปรึกษา_____

JATSADA SINGTHONGCHAI : A SEMANTIC SEARCH MODEL FOR AGRONOMY KNOWLEDGE. THESIS ADVISOR : ASST. PROF. SUPHAKIT NIWATTANAKUL, Ph.D., 159 PP.

SEMANTIC SEARCH/ TERMS WEIGHTING/ CORRELATION COEFFICIENT/ SIMILARITY

This research aims to design and develop semantic search model for agronomy knowledge : SeeAgrow. It consists of two parts of a knowledge management and a 3 steps semantic search which is keywords degree identification, terms' weighting, and similarity evaluation. Information about tapioca, sugarcane, and corn has been gathered from reliable sources. The information has been confirmed to be completed and publicized since 2007 in a database with ontology application. Precision, Recall, F-measure, Mean Reciprocal Rank (MRR) and Sum of Squared Errors of Prediction (SSE) of the search model is evaluated.

1. Semantic Search Model for Agronomy Knowledge (SeeAgrow) comprises two parts of knowledge management and semantic search.

2. Weighting Efficiency Evaluation with Correlation Coefficient (Corr) has MRR at 0.906 and SSE at 0.097. This obviously indicates better MRR performance with less SSE than Inverse Document Frequency (IDF) method according to the stated hypothesis.

3. The Efficiency of Semantic Similarity Evaluation.

3.1 The result of comparison in evaluating semantic similarity from F-measure that reflexes Precision and Recall with the application of shortest path between Keywords and corpus merging with the calculation of term weighting by Corr,

Sequence of Words (SW) and Position of Words (PW) provides the highest F-measure value at 0.8596.

3.2 The result of comparison in evaluating semantic similarity from MRR with the application of shortest path between Keywords and corpus merging with the calculation of Inverse Document Frequency (IDF) Sequence of Words (SW) and Position of Words (PW) provides the highest MRR value at 0.8661.

มหาวิทยาลัยเทคโนโลยีสุรนารี

School of Information Technology

Academic Year 2016

Student's Signature _____

Advisor's Signature S. Nimit

กิตติกรรมประกาศ

วิทยานิพนธ์นี้สำเร็จลุล่วงด้วยดี ผู้วิจัยขอกราบขอบพระคุณ บุคคล และกลุ่มบุคคลต่าง ๆ ที่ได้กรุณาให้คำปรึกษา แนะนำ ช่วยเหลืออย่างดียิ่ง ทั้งในด้านวิชาการ และด้านการดำเนินงานวิจัย ดังต่อไปนี้

ผู้ช่วยศาสตราจารย์ ดร.ศุภกฤษฎิ์ นิวัฒนากูล อาจารย์ที่ปรึกษาวิทยานิพนธ์ และอาจารย์ ดร.นิศาชล จำนงศรี ที่ให้คำปรึกษาในการทำงานวิจัย การจัดรูปแบบ และช่วยตรวจทานความถูกต้องของวิทยานิพนธ์ ขอขอบคุณรองศาสตราจารย์ ดร.วีรพงษ์พลนิกรกิจ รองศาสตราจารย์ ดร.นิตยา เกิดประสพอาจารย์ ดร.ชูชาติ หฤไชยะศักดิ์ และผู้ช่วยศาสตราจารย์ ดร.จิติมนต์ อั่งสกุล ที่ให้คำชี้แนะต่าง ๆ อันเกี่ยวข้องกับวิทยานิพนธ์นี้ ตลอดจนผู้เชี่ยวชาญที่ประเมินตรวจสอบความถูกต้องของแบบสอบถามคำค้นด้านพืชไร่

มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน ที่ช่วยสนับสนุนทุนการศึกษาทุนวิจัยและค่าใช้จ่ายนอกจากนี้ขอขอบคุณครูอาจารย์ทั้งในอดีตและปัจจุบันที่ให้ความรู้แก่ผู้วิจัยจนประสบความสำเร็จในชีวิต และพี่ ๆ เพื่อน ๆ น้อง ๆ ที่ให้คำปรึกษา เป็นกำลังใจในการทำวิทยานิพนธ์ตลอดมา

ท้ายที่สุดคุณประโยชน์จากการวิจัยครั้งนี้ ผู้วิจัยขอกราบขอบพระคุณ และเป็นเครื่องแสดงความกตัญญุตาแด่ บิดา มารดา ตลอดจนบุรพาจารย์ และผู้มีพระคุณทุกท่านที่ได้ให้การศึกษาอบรมสั่งสอนให้แสงสว่าง นำทางสู่ปัญญา และคุณธรรมอันเป็นเครื่องชี้นำสู่ความสำเร็จของชีวิต เพื่อเป็นสิริมงคลแก่ผู้วิจัยตลอดไป

เจษฎา สิงห์ทองชัย

มหาวิทยาลัยเทคโนโลยีสุรนารี

สารบัญ

หน้า

บทคัดย่อ (ภาษาไทย) ..... ก

บทคัดย่อ (ภาษาอังกฤษ) ..... ข

กิตติกรรมประกาศ..... ง

สารบัญ..... จ

สารบัญตาราง..... ช

สารบัญรูป..... ซ

บทที่

1 บทนำ..... 1

1.1 ความสำคัญและที่มาของปัญหาการวิจัย..... 1

1.2 วัตถุประสงค์ของการวิจัย..... 5

1.3 สมมติฐานการวิจัย..... 5

1.4 ข้อตกลงเบื้องต้น..... 5

1.5 ขอบเขตของการวิจัย..... 5

1.6 ประโยชน์ที่คาดว่าจะได้รับ..... 6

1.7 คำอธิบายศัพท์..... 6

2 ปริทัศน์วรรณกรรมและงานวิจัยที่เกี่ยวข้อง..... 7

2.1 การจัดระบบความรู้ด้านพืชไร่..... 8

2.2 แนวคิดการค้นคืนสารสนเทศ..... 20

2.3 แบบจำลองการสืบค้น..... 30

2.4 การวิเคราะห์คำสำคัญ..... 43

2.5 การวัดความคล้ายคลึง..... 48

2.6 กรอบแนวคิดการวิจัย..... 56

3 วิธีดำเนินการวิจัย..... 63

3.1 การศึกษาปัญหา..... 64

3.2 การออกแบบและพัฒนาแบบจำลองการสืบค้นเชิงความหมาย..... 64

มหาวิทยาลัยเทคโนโลยีสุรนารี

สารบัญ (ต่อ)

หน้า

3.2.1 การออกแบบและพัฒนากระบวนการจัดความรู้ให้เป็นระบบ..... 64

3.2.2 การออกแบบและพัฒนากระบวนการสืบค้นเชิงความหมาย..... 69

3.3 การประเมินประสิทธิภาพการสืบค้นเชิงความหมาย..... 75

4 ผลการวิจัย..... 82

4.1 ผลการออกแบบจำลองการสืบค้นเชิงความหมาย..... 82

4.1.1 การจัดความรู้ให้เป็นระบบ..... 84

4.1.2 การสืบค้นเชิงความหมาย..... 94

4.2 ผลการประเมินประสิทธิภาพการสืบค้นเชิงความหมาย..... 104

5 สรุป อภิปรายผล และข้อเสนอแนะ..... 113

5.1 สรุปผลการวิจัย..... 113

5.2 อภิปรายผลการวิจัย..... 114

5.3 ข้อจำกัดของการวิจัย..... 119

5.4 การประยุกต์ผลการวิจัย..... 119

5.5 ข้อเสนอแนะในการวิจัยครั้งต่อไป..... 120

รายการอ้างอิง..... 121

ภาคผนวก

ภาคผนวก ก แบบสำรวจเพื่อการวิจัยแบบสอบถามความต้องการใช้คำค้นสำหรับสืบค้นความรู้ด้านพืชไร่..... 133

ภาคผนวก ข ผลการหาประสิทธิภาพเครื่องมือและผลการทดลอง..... 139

ประวัติผู้เขียน..... 159

สารบัญตาราง

ตารางที่ **หน้า**

2.1 ตัวอย่างแหล่งจัดเก็บและการเข้าถึงความรู้ด้านพืชไร่..... 8
2.2 ชุดหน่วยข้อมูลย่อยดับลินคอร์ 15 อิลีเมนต์ (Element) 15
2.3 การคำนวณความน่าจะเป็นของความสัมพันธ์ระหว่างเอกสารกับคำค้น…..... 30
2.4 ตัวอย่างวิธีการวัดความคล้ายคลึงแบบต่าง ๆ..... 48
2.5 สรุปเปรียบเทียบงานวิจัยที่เกี่ยวข้อง..... 56
3.1 รูปแบบการทดลอง..... 73
4.1 ตัวอย่างผลการคำนวณความถี่ของคำสำคัญที่ปรากฏในแต่ละความรู้..... 88
4.2 ตัวอย่างค่าความสัมพันธ์ระหว่างคำในโครงสร้างด้วยการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์..... 89
4.3 ตัวอย่างค่าน้ำหนักคำด้วยการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ต่อความรู้หรือเอกสาร..... 91
4.4 ตัวอย่างผลลัพธ์การคำนวณระยะทางระหว่างคำในโครงสร้าง..... 92
4.5 ตัวอย่างการวิเคราะห์ตำแหน่งของคำค้นที่ปรากฏในความรู้..... 95
4.6 ตัวอย่างผลลัพธ์การคำนวณระยะทางระหว่างคำในโครงสร้าง..... 96
4.7 ตัวอย่างการคำนวณระยะทางระหว่างคำค้นกับคำในความรู้หรือเอกสาร ()..... 97
4.8 การวัดความคล้ายคลึงของคำค้น k1 เทียบกับคำในความรู้หรือเอกสาร..... 98
4.9 การวัดความคล้ายคลึงของคำค้น k2 เทียบกับคำในความรู้หรือเอกสาร..... 98
4.10 ข้อมูลทั่วไปของผู้ตอบแบบสอบถามคำค้นสำหรับสืบค้นความรู้ด้านพืชไร่..... 104
4.11 ตัวอย่างการวัดประสิทธิภาพด้วยค่าความแม่นยำเฉลี่ย (MRR) จากการสืบค้น..... 106
4.12 ตัวอย่างประสิทธิภาพการให้น้ำหนักคำค้น..... 107

สารบัญรูป

รูปที่ **หน้า**

2.1 ตัวอย่างอรรถาภิธานของ “Agriculture”..... 12
2.2 ตัวอย่างอรรถาภิธานศัพท์เกษตรไทยของคำ “มันสำปะหลัง” 12
2.3 ตัวอย่างข่ายงานเชิงความหมาย..... 13
2.4 การนิยามศัพท์ในเอสเคโอเอส (SKOS) 14
2.5 ลักษณะ CYC Ontology และ WordNet Ontology ในระดับ High Level..... 17
2.6 ตัวอย่างข้อมูลออนโทโลยีเฉพาะเกี่ยวกับพืชไร่..... 18
2.7 แนวคิดการค้นคืนสารสนเทศ..... 21
2.8 ตัวอย่างกระบวนการค้นคืนสารสนเทศนำเสนอโดยเฟรค..... 23
2.9 ตัวอย่างกระบวนการค้นคืนสารสนเทศนำเสนอโดยเบย์ซา เยตส์ และริบริโร เนโท… 24
2.10 ตัวอย่างแสดงการดึงข้อมูลโดยใช้ภาษาสปาร์เก้ิลแบบเงื่อนไขเดียว..... 27
2.11 ตัวอย่างแสดงการดึงข้อมูลโดยใช้ภาษาสปาร์เก้ิลแบบเปรียบเทียบหลายข้อมูล..... 27
2.12 การสืบค้นเชิงความหมายจากการวัดความคล้ายคลึงระหว่างเอกสาร..... 37
2.13 ขั้นตอนการสืบค้นเชิงความหมาย..... 39
2.14 ขั้นตอนส่วนแสดงผลลัพธ์การสืบค้น..... 40
2.15 วิธีการสืบค้นเชิงความหมายของ ลีย์, มิน และโฮ..... 40
2.16 ตัวอย่างคำค้นที่ระบุจากผู้ใช้ และตำแหน่งที่ปรากฏในเอกสาร..... 44
2.17 ตัวอย่างของกราฟมีทิศทาง..... 50
2.18 ตัวอย่างของกราฟไม่มีทิศทาง..... 51
2.19 ตัวอย่างของกราฟที่มีน้ำหนัก..... 51
2.20 การหาเส้นทางจากโหนด A ไป B..... 53
2.21 การหาเส้นเชื่อมจากโหนดต่างๆ ถึงโหนด B..... 53
2.22 การหาเส้นเชื่อมจากโหนดต่าง ๆ ถึงโหนด F..... 54
2.23 เส้นทาง AFC ถูกตัดทิ้ง..... 54
2.24 แสดงเส้นทาง ADB ถูกตัดทิ้ง..... 55
2.25 เส้นทาง ABG ถูกตัดทิ้ง..... 55

สารบัญรูป (ต่อ)

รูปที่ **หน้า**

2.26 ผลลัพธ์เส้นทางที่สั้นที่สุด..... 55
2.27 กรอบแนวคิดการวิจัย..... 62
3.1 ระเบียบวิธีการวิจัย..... 63
3.2 กรอบแนวคิดขั้นตอนการจัดความรู้ให้เป็นระบบ..... 64
3.3 คำสำคัญที่เลือกจากความรู้ด้านพืชไร่..... 65
3.4 ตัวอย่างโครงสร้างความสัมพันธ์ของคำในความรู้ด้านพืชไร่..... 66
3.5 ตัวอย่างการนิยามศัพท์ด้านพืชไร่ในเอสเคโอเอส..... 67
3.6 ตัวอย่างโครงข่ายความสัมพันธ์ของคำในความรู้ด้านพืชไร่..... 68
3.7 ตัวอย่างการกำหนดความสัมพันธ์ของคำในคลังคำ..... 69
3.8 ตัวอย่างลำดับคำค้นจากผู้ใช้และตำแหน่งคำค้นที่ปรากฏในความรู้..... 70
3.9 กรอบแนวคิดขั้นตอนการกำหนดลำดับความสำคัญคำค้น..... 70
3.10 กรอบแนวคิดขั้นตอนการให้น้ำหนักคำสำคัญ..... 71
3.11 กรอบแนวคิดขั้นตอนการวัดความคล้ายคลึง..... 72
3.12 กระบวนการประเมินประสิทธิภาพการสืบค้น..... 75
4.1 สถาปัตยกรรมของแบบจำลองการสืบค้นเชิงความหมาย (SeeAgrow)..... 83
4.2 โครงสร้างของคลาสในการจัดเก็บความรู้ด้านพืชไร่..... 84
4.3 ตัวอย่างผลการคำนวณความถี่คำที่ปรากฏในความรู้โดยใช้ภาษาพีเอชพี..... 85
4.4 ขั้นตอนการวิเคราะห์ให้น้ำหนักคำ..... 86
4.5 ตัวอย่างโครงสร้างความสัมพันธ์ของคำที่เกี่ยวข้องกับความรู้ด้านพืชไร่..... 87
4.6 ตัวอย่างการกำหนดความสัมพันธ์ของคำในคลังคำ..... 91
4.7 ลำดับคำสำคัญของตำแหน่งที่ปรากฏคำค้น..... 95
4.8 ตัวอย่างคำค้นที่ผู้ใช้ระบุเข้ามากับคำค้นที่ปรากฏในความรู้หรือเอกสาร..... 96
4.9 ผลการเปรียบเทียบค่าความแม่นยำเฉลี่ย (MRR)..... 108
4.10 ผลการเปรียบเทียบค่าความผิดพลาด (SSE)..... 108
4.11 ผลเปรียบเทียบค่าความแม่นยำเฉลี่ย และค่าความผิดพลาด..... 109
4.12 เปรียบเทียบประสิทธิภาพค่าความถ่วงดุลรูปแบบการทดลอง A และ B..... 111
4.13 เปรียบเทียบประสิทธิภาพค่าความแม่นยำเฉลี่ยรูปแบบการทดลอง A และ B..... 112

บทที่ 1
บทนำ

1.1 ความสำคัญและที่มาของปัญหาการวิจัย

ปัจจุบันความก้าวหน้าทางด้านเทคโนโลยีสารสนเทศได้มีบทบาทสำคัญในการ สร้างโอกาส ความเสมอภาค และเท่าเทียมกันในการเข้าถึงข้อมูล สารสนเทศ หรือความรู้ต่าง ๆ ที่กระจายอยู่ทั่วไปบนเครือข่ายอินเทอร์เน็ต (Internet) ซึ่งมีแนวโน้มเพิ่มมากขึ้นอย่างรวดเร็ว ซึ่งไกรศักดิ์ เกสร (2554: 1) ได้อธิบายว่าข้อมูลที่กระจายอยู่ทั่วไปบนอินเทอร์เน็ตมีทั้งข้อมูลที่เป็นตัวอักษร (Textual Information) และข้อมูลมัลติมีเดีย (Multimedia Information) เช่น รูปภาพ (Image) วิดีโอ (Video) และเสียง (Audio) ซึ่งผู้ใช้ส่วนใหญ่จะทำการค้นหาข้อมูลโดยใช้โปรแกรมค้นหา (Search engine) เช่น กูเกิ้ล (Google) ยาฮู (Yahoo) และบิง (Bing) เป็นต้น

การค้นหาข้อมูลจะใช้คำสำคัญจากการระบุคำค้นของผู้ใช้เพื่อให้พบข้อมูล สารสนเทศ หรือความรู้ที่เกี่ยวข้องกับคำสำคัญให้มากที่สุด และสามารถเข้าถึงความรู้ได้หลากหลาย โดยเฉพาะความรู้ด้านการเกษตรที่มีความสำคัญต่อการผลิตและสร้างมูลค่าเพิ่มให้กับสินค้าเกษตรของประเทศ จากรายงานของสำนักงานเศรษฐกิจการเกษตร (2558: 15-20) พบว่าประเทศไทยมีการส่งออกสินค้าเกษตรประเภท ข้าว มันสำปะหลัง และ อ้อย เป็นลำดับต้น ๆ ของประเทศ ดังนั้นการเข้าถึงความรู้ด้วยโปรแกรมค้นหาจึงเป็นทางเลือกลำดับต้น ๆ แต่ โปรแกรมค้นหาในปัจจุบันไม่ได้ถูกออกแบบมาเพื่อรองรับการสืบค้นความรู้ด้านการเกษตรดังกล่าว ซึ่งจากการศึกษาของฮาห์นและมนิ (Hahn and Mani, 2000) พบว่าผลลัพธ์ที่ได้จากการสืบค้นด้วยโปรแกรมค้นหายังไม่ตรงกับความต้องการ ซึ่งการใช้คำสำคัญเพียงอย่างเดียวไม่พอต่อการเข้าถึงความรู้ และได้ผลลัพธ์ตามต้องการ

ปัญหาของโปรแกรมค้นหา คือผลลัพธ์ที่ได้จากการค้นหามีปริมาณมาก หรือข้อมูลที่ได้เกินความจำเป็นทำให้การตัดสินใจของผู้ใช้ยากต่อการเลือกคำตอบ และเสียเวลาในการเลือกข้อมูลหรือเอกสารที่เกี่ยวข้อง สอดคล้องกับ วิไลพร เลิศมหาเกียรติ , ภูริวัตร คัมภีรภาพัฒน์ และอนิราช มิ่งขวัญ (2551) และศุภกฤษฎิ์ นิวัฒนากูล (2556) ที่ว่าโปรแกรมค้นหาไม่สามารถประมวลผลความหมายจากการระบุคำสำคัญของผู้ใช้ และให้ผลลัพธ์ไม่ตรงประเด็น เนื่องจากโปรแกรมค้นหาไม่มีการกำหนดตัวแทนความรู้ตามที่ผู้ใช้ระบุคำสำคัญที่ครบถ้วนสมบูรณ์ เช่น ผู้ใช้ระบุคำสำคัญชื่อบุคคล และอ้อยที่เป็นพืชไร่ เป็นต้น ทำให้ผลลัพธ์ที่ได้มีจำนวนมากเพราะไม่มีคำค้น หรือตัว

แทนที่ใช้ในการเชื่อมโยงถึงความรู้ที่ผู้ใช้ต้องการ จากปัญหาดังกล่าวจึงเป็นสิ่งที่นักวิจัยให้ความสนใจแก้ปัญหา โดยการใช้แบบจำลองการสืบค้น

แบบจำลองการสืบค้นที่ใช้ในโปรแกรมค้นหามีหลากหลายแนวคิด ซึ่งแบบจำลองของซัลตอน (Salton, 1989) ก็เป็นแนวคิดหนึ่งที่ได้รับความนิยมและถูกนำไปใช้ในการสืบค้นข้อมูล ได้แก่แบบจำลองปริภูมิเวกเตอร์ (Vector Space Model) ที่ให้น้ำหนักคำ (Term) ด้วยความถี่ของคำที่ปรากฏในเอกสาร (Term Frequency: TF) และคำนวณสัดส่วนของจำนวนเอกสารทั้งหมดกับจำนวนเอกสารที่ปรากฏคำที่สนใจ (Inverse Document Frequency: IDF) หรือเรียกว่า TF-IDF ข้อดีคือ วัดความคล้ายคลึงระหว่างคำค้นกับเอกสารได้ แม้ว่าเอกสารนั้นไม่มีคำค้นปรากฏอยู่ทุกคำ แต่ข้อจำกัดคือไม่สามารถวิเคราะห์ความหมายของคำได้เนื่องจากการกำหนดดัชนีไม่มีความสัมพันธ์กัน ซึ่งคิโมโต และอิเวดิร่า (Kimoto and Iwadera , 1990) ได้เสนอแนวทางแก้ปัญหาด้วยการเพิ่มความสัมพันธ์ภายในเอกสาร ต่อมาอาเมียร์ และโมแรด (Amir and Mourad, 2013), ลีย์, มิน และโฮ (Lee, Min and Oh, 2014) ได้เสนอแบบจำลองการสืบค้นเพื่อแก้ปัญหาข้างต้น ซึ่งมีส่วนประกอบ ได้แก่ ระบบจัดเก็บความรู้ ออนโทโลยี การวิเคราะห์คำสำคัญ การให้น้ำหนักคำ และการวัดความคล้ายคลึง

ระบบจัดเก็บความรู้ (Knowledge Storage System) เป็นกระบวนการรวบรวมรายละเอียดความรู้ที่จัดเก็บในฐานความรู้ เพื่อให้สามารถค้นความรู้กลับคืนมาได้อย่างสะดวกและรวดเร็ว ซึ่งฮอดจ์ (Hodge, 2000) ได้อธิบายว่า ปัจจุบันมีการจัดความรู้ให้เป็นระบบโดยใช้ตัวแบบการแทนความรู้ที่เรียกว่า เอสเคโอเอส (Simple Knowledge Organization System: SKOS) สามารถกำหนดความสัมพันธ์ระหว่างคำ หรือข้อความต่าง ๆ ในรูปแบบลำดับชั้น (Hierarchical) ซึ่งความสัมพันธ์สามารถแสดงระหว่างคำหรือข้อความ ที่มีข้อมูลเกี่ยวกับคำเหมือน (Synonym) คำตรงข้าม (Antonym) คำที่ออกเสียงหรือสะกดคล้ายกัน (Homonym) คำทั่วไป (Generic Term) และคำที่เกี่ยวข้อง (Associative Term) สอดคล้องกับฟุลวิโอ และลอร่า (Fulvio and Laura, 2005) อธิบายว่าเอสเคโอเอสเป็นเครื่องมือสำหรับสร้างคลังคำหรือมาตรฐานข้อมูลที่ใช้ร่วมกัน เหมาะสำหรับแก้ไขปัญหาความแตกต่างของข้อมูลที่หลากหลาย นอกจากนี้การจัดระบบความรู้สมัยใหม่ได้นำออนโทโลยี (Ontology) เข้ามาช่วยในการจัดการความรู้เพื่ออธิบายความหมายในขอบเขตเรื่องที่สนใจ

ออนโทโลยีเป็นแนวคิดการจัดฐานความรู้ให้อยู่ในรูปแบบโครงสร้างลำดับชั้น ซึ่งโจนส์, เบนซ์-เคพอน และวิชเซอร์ (Jones, Bench-Capon and Visser, 1998), นอย และแม็คกินนิสส์ (Noy and McGuinness, 2001), แกสสิวิค, ดูจ์ริค และเดเวดซิค (Gasevic, Djuric and Devedzic, 2009) ได้อธิบายว่าออนโทโลยีเป็นโครงสร้างความรู้เชิงความหมาย นำมาใช้อธิบายแนวคิดโดเมนหนึ่ง ๆ เพื่อให้เข้าใจความหมายสิ่งนั้น ๆ ในทิศทางเดียวกัน และที่สำคัญออนโทโลยีสามารถจัดเก็บคำค้นที่มีความหมายเดียวกันไว้เป็นกลุ่ม เพื่อใช้ขยายคำค้นจากผู้ใช้ ซึ่งสามารถแก้ปัญหาความกำกวมของ

คำสำคัญ เช่น หนึ่งคำมีหลายความหมาย หรือคำหลาย ๆ คำมีความหมายเดียวกัน สอดคล้องกับ จุฑามาศ เทียนสอาด และอรวรรณ อิ่มสมบัติ (2552), เฟนเซลล์ (Fensel, 2004), อาเมียร์และโมแรด (Amir and Mourad, 2013), ศุภกฤษฎิ์ นิวัฒนากูล (2556), ลีย์, มิน และโฮ (Lee, Min and Oh, 2014) ที่กล่าวว่าออนโทโลยีมีความจำเป็นต่อการประมวลผลเชิงความหมายโดยเฉพาะการสืบค้นและสกัดข้อมูล ดังนั้นออนโทโลยีจึงมีความสำคัญต่อกระบวนการสืบค้นข้อมูล

การกำหนดลำดับความสำคัญคำค้นเป็นขั้นตอนหนึ่งในแบบจำลองการสืบค้นที่ใช้ในโปรแกรมค้นหา โดยใช้การพิจารณาจาก 2 กรณี คือ พิจารณาจากลำดับคำค้นของผู้ใช้ที่ระบุเข้ามา (Sequence of Words) และจากตำแหน่งของคำที่ปรากฏในเอกสาร (Position of Words) ซึ่ง ลีย์, มิน และโฮ (Lee, Min and Oh, 2014), วิไลพร เลิศมหาเกียรติ, ภูริวัตร คัมภีรภาพพัฒน์ และอนิราช มิ่งขวัญ (2551) ได้กล่าวถึงปัญหาในขั้นตอนนี้เกิดจาก (1) ผู้ใช้สะกดคำไม่ถูกต้องตามหลักไวยกรณ์ (2) ผู้ใช้เลือกคำสำคัญไม่ตรงกับคำในดัชนี และ (3) การสืบค้นไม่สามารถวิเคราะห์ถึงความหมายของคำสำคัญได้ ซึ่งการแก้ปัญหาดังกล่าว เดียร์เวสเตอร์ (Deerwester,1990) ได้เสนอเทคนิควิธีการแยกคำที่ผู้ใช้สนใจจากเอกสารมาสร้างเป็นดัชนี (Index) ต่อมา อาเมียร์และโมแรด (Amir and Mourad, 2013) ได้สร้างดัชนีด้วยแนวคิดกราฟ นอกจากนี้ศุภกฤษฎิ์ นิวัฒนากูล (2556) ได้เสนอวิธีแก้ปัญหาดังกล่าวด้วยการวัดความคล้ายคลึงกันของแจ็คการ์ด (Jaccard Similarity Coefficient) สำหรับเปรียบเทียบความคล้ายคลึงคุณสมบัติของเซตข้อมูล และการให้น้ำหนักคำ

การให้น้ำหนักคำ (Term Weighting) เป็นขั้นตอนหนึ่งที่สำคัญของแบบจำลองการสืบค้น โดยเรสนิค (Resnik, 1999) และณัฐพร หอมเมือง (2550) ได้กล่าวว่า น้ำหนักของคำแต่ละคำ จะคำนวณโดยเปรียบเทียบกับคำทั้งหมดในเอกสาร ซึ่งมีการคำนวณให้น้ำหนักคำใน (1) เอกสาร และ (2) คำค้นจากผู้ใช้ที่ระบุเข้ามา ซึ่งค่าน้ำหนักเหล่านั้นจะถูกนำไปแทนในสมการเพื่อคำนวณความคล้ายคลึงระหว่างคำค้นกับเอกสาร เรียกว่าการวัดความคล้ายคลึงระหว่างเอกสาร ซึ่งแนวคิดดังกล่าวจะสอดคล้องกับวิธีการให้น้ำหนักคำของ ซัลตอน (Salton, 1989) แต่การให้น้ำหนักคำดังกล่าว จะไม่ให้ความสำคัญกับความหมายของคำ กลุ่มคำ และถือว่าคำแต่ละคำเป็นอิสระต่อกัน ซึ่ง ดีแจเมล (Djamel, 2009), ปฏิคม ทองจริง (2552), ลีย์, มิน และโฮ (Lee, Min and Oh, 2014) ได้เสนอแนวทางแก้ปัญหาโดยการนำออนโทโลยีเข้ามาช่วยสำหรับการกำหนดความหมายของคำเพื่อให้การค้นหาตรงกับความต้องการของผู้ใช้มากขึ้น

การวัดความคล้ายคลึงระหว่างคำกับเอกสารในปัจจุบัน สามารถคำนวณได้หลายวิธี แต่วิธีที่นิยมใช้ คือการวัดความคล้ายคลึงเชิงมุม ซึ่งใช้เทคนิคการแทนเอกสารด้วยระบบเวกเตอร์ และหลักการทางสถิติ ข้อดีคือสามารถวัดความคล้ายคลึงระหว่างคำค้นกับเอกสารได้ แม้ว่าเอกสารนั้นไม่มีคำค้นปรากฏอยู่ทุกคำ แต่มีข้อจำกัดคือไม่สามารถวิเคราะห์ความหมายของคำได้เนื่องจากการกำหนดดัชนีไม่มีความสัมพันธ์ กัน จากปัญหาดังกล่าวมีนักวิจัยจำนวนหนึ่งได้นำเสนอแนวคิด

สำหรับแก้ปัญหาข้างต้นโดยการนำออนโทโลยีมาช่วยในการอธิบายความหมายของคำ ดังงานวิจัยของดีแจเมล (Djamel, 2009), ซันดราเซร์คารัน , โจเซฟซันและเบนจามินส์ (Chandrasekaran, Josephson and Benjamins, 1999) แมดฮุรา และคณะ (Madhura, et. al., 2012) อาเมียร์และโมแรด (Amir and Mourad, 2013) และงานวิจัยของ ลีย์, มิน และโฮ (Lee, Min and Oh, 2014) เป็นต้น นอกจากนี้ พิลาวัณย์ พลับรู้การและกฤษณะ ไวยมัย (2544) ศุภกฤษฎิ์ นิวัฒนากูล (2556) ได้เสนอเทคนิควิธีการคำนวณค่าคำนวณความคล้ายคลึงกันระหว่างเอกสารเชิงความหมาย ด้วยการหาค่าความเกี่ยวข้องระหว่างคำสองคำ

จากที่กล่าวมาข้างต้นสรุปได้ว่าโปรแกรมค้นหาในปัจจุบันมักพบปัญหา (1) ผู้ใช้สะกดคำไม่ถูกต้องตามหลักไวยกรณ์ (2) ผู้ใช้เลือกคำสำคัญไม่ตรงกับคำในดัชนี และ (3) กระบวนการสืบค้นไม่สามารถวิเคราะห์ความหมายของคำสำคัญ และแนวทางการแก้ปัญหานี้ ลีย์, มิน และโฮ (Lee, Min and Oh, 2014) ได้อธิบายว่าควร ออกแบบขั้นตอนวิธีการจัดการความรู้ การวิเคราะห์คำสำคัญ และ การวัดความคล้ายคลึงระหว่างคำกับเอกสาร ให้เหมาะสม ซึ่งสอดคล้องกับแนวคิดของเฟรค และคณะ (Frakes, et. al., 1992), เบย์ซา เยตส์ และริบริโร เนโท (Baeza-Yates and Ribeiro-Neto, 1999), นคร โคตรโสภา (2549) และศุภกฤษฎิ์ นิวัฒนากูล (2556) ได้ศึกษาออกแบบเทคนิควิธีการสำหรับแก้ปัญหาข้างต้น เพื่อให้ได้ผลลัพธ์ที่เหมาะสมและตรงกับความต้องการของผู้ใช้ แต่เทคนิควิธีการที่นำเสนอยังมีข้อจำกัดในส่วนของการประมวลความหมายของคำ ในขอบเขตโดเมนเรื่องที่สนใจ อาจมีสาเหตุจากการเพิ่มขึ้นของ ปริมาณข้อมูล สารสนเทศ และความรู้บนเครือข่ายอินเทอร์เน็ต จึงทำให้เกิดความท้าทายของนักวิจัยในการคิดค้นขั้นตอนวิธีสำหรับการสืบค้น เพื่อให้ได้ผลลัพธ์ที่เหมาะสมและตรงกับความต้องการของผู้ใช้

ดังนั้นงานวิจัยนี้จึงพัฒนาแบบจำลองการสืบค้นเชิงความหมายสำหรับความรู้ด้านพืชไร่ เพื่อหาแนวทาง และออกแบบขั้นตอนวิธีการสืบค้นเชิงความหมาย ประกอบด้วยสองส่วน ส่วนแรกคือการจัดความรู้ให้เป็นระบบ และส่วนที่สองคือการสืบค้นเชิงความหมาย และสิ่งที่ได้รับจากการวิจัยนี้ คือได้ขั้นตอนวิธีสำหรับการออกแบบและพัฒนาระบบสืบค้นข้อมูลสำหรับความรู้ด้านพืชไร่อันมีผลลัพธ์ที่ตรงกับความต้องการของผู้ใช้

1.2 วัตถุประสงค์ของการวิจัย

เพื่อออกแบบและพัฒนาแบบจำลองการสืบค้นเชิงความหมายสำหรับความรู้ด้านพืชไร่ 2 ส่วน คือ การจัดความรู้ให้เป็นระบบ และการสืบค้นเชิงความหมาย 3 ขั้นตอน ได้แก่ การกำหนดลำดับความสำคัญคำค้น การให้น้ำหนักคำสำคัญ และการวัดความคล้ายคลึง

1.3 สมมติฐานการวิจัย

แบบจำลองการสืบค้นเชิงความหมายสำหรับความรู้ด้านพืชไร่ มีค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) ค่าความถ่วงดุล (F- Measure) ค่าความแม่นยำเฉลี่ย (Mean Reciprocal Rank: MRR) มากกว่าร้อยละ 80 ขึ้นไป และค่าความผิดพลาด (Sum of Squared Errors of Prediction: SSE) ไม่เกิน 0.1

1.4 ข้อตกลงเบื้องต้น

1.4.1 ความรู้พืชไร่ที่นำมาเป็นข้อมูลทดสอบได้จากการสืบค้นด้วยโปรแกรมค้นหาบนเครือข่ายอินเทอร์เน็ตจากแหล่งข้อมูลที่มีความน่าเชื่อถือ และได้เลือกรายการความรู้ที่มีคำอธิบายครบถ้วนตามความต้องการที่มีการเผยแพร่ตั้งแต่ปี พ.ศ. 2550 เป็นต้นมา เช่น ฐานข้อมูลงานวิจัยการเกษตรไทย ฐานข้อมูลในโครงการเครือข่ายห้องสมุดในประเทศไทย สำนักงานคณะกรรมการการอุดมศึกษา และฐานข้อมูลของสำนักงานคณะกรรมการวิจัยแห่งชาติ เป็นต้น

1.4.2 ฐานความรู้พืชไร่ ประยุกต์ใช้ออนโทโลยีไม้ผลเศรษฐกิจที่พัฒนาขึ้นโดย ศุภกฤษฎิ์ นิวัฒนากูล (2556) โดยใช้โครงสร้างที่ออกแบบคลาสไว้เท่านั้น ซึ่งไม่ใช่ความรู้ของไม้ผลเศรษฐกิจ

1.4.3 คำสำคัญที่จะนำไปใช้สำหรับการสืบค้นความรู้ต้องเป็นคำที่สะกดถูกต้อง และการระบุคำค้นต้องแยกคำสำคัญโดยการเว้นวรรค

1.5 ขอบเขตของการวิจัย

1.5.1 จัดเก็บความรู้ด้านพืชไร่ของประเทศไทย 3 ชนิด คือ มันสำปะหลัง อ้อย และข้าวโพด ซึ่งเป็นบทความวิจัย หรือรายงานการวิจัย หรือวิทยานิพนธ์ โดยจัดเก็บชื่อเรื่อง คำสำคัญ คำอธิบายความรู้ (บทคัดย่อ) และแหล่งอ้างอิงต้นฉบับ จำนวนไม่น้อยกว่า 240 รายการ

1.5.2 ออกแบบและพัฒนาแบบจำลองการสืบค้นเชิงความหมายสำหรับความรู้ด้านพืชไร่ ประกอบด้วยการจัดความรู้ให้เป็นระบบ และการสืบค้นเชิงความหมาย 3 ขั้นตอน ได้แก่ (1) การกำหนดลำดับความสำคัญคำค้น (2) การให้น้ำหนักคำสำคัญ และ (3) การวัดความคล้ายคลึง ซึ่งจัดลำดับผลลัพธ์ตามผลการเปรียบเทียบความคล้ายคลึงระหว่างเอกสารกับคำค้น

1.6 ประโยชน์ที่คาดว่าจะได้รับ

ได้แบบจำลองการสืบค้นเชิงความหมายสำหรับความรู้ด้านพืชไร่ที่สามารถแก้ปัญหาความกำกวมของคำสำคัญที่ใช้สืบค้นความรู้ เพื่อให้ได้ผลลัพธ์ที่ตรงกับความต้องการของผู้ใช้ ซึ่งประกอบด้วยการจัดความรู้ให้เป็นระบบ และการสืบค้นเชิงความหมาย 3 ขั้นตอน คือ (1) การกำหนดลำดับความสำคัญคำค้น (2) การให้น้ำหนักคำสำคัญ และ (3) การวัดความคล้ายคลึง โดยแบบจำลองในงานวิจัยนี้สามารถนำไป เป็นแนวทางสำหรับการพัฒนาระบบสืบค้นความรู้ด้านที่เกี่ยวข้อง และอื่น ๆ ต่อไป

1.7 คำอธิบายศัพท์

1.7. 1 แบบจำลองการสืบค้นเชิงความหมาย (Semantic Search Model) หมายถึงขั้นตอนวิธีการจัดความรู้ให้เป็นระบบ และการสืบค้นเชิงความหมาย ประกอบด้วยขั้นตอน (1) การกำหนดลำดับความสำคัญคำค้น (2) การให้น้ำหนักคำสำคัญ และ (3) การวัดความคล้ายคลึง

1.7.2 ความรู้ด้านพืชไร่ (Agronomy Knowledge) หมายถึงเอกสารที่เป็น บทความวิจัย หรือรายงานการวิจัย หรือวิทยานิพนธ์ ที่เกี่ยวข้องกับมันสำปะหลัง อ้อย และข้าวโพด ที่จัดเก็บในรูปแบบออนโทโลยี

1.7.3 การอธิบายความรู้ (Explain Knowledge) หมายถึงการอธิบายเนื้อหา (Content) ของความรู้ด้านพืชไร่ที่เกี่ยวกับชื่อเรื่อง คำสำคัญ คำอธิบายความรู้ (บทคัดย่อ) และแหล่งอ้างอิงต้นฉบับ โดยใช้เมทาดาทา (Metadata) ดับลินคอร์ (Dublin Core)

1.7.4 คลังคำ (Word Repository) หมายถึง การรวบรวมคำด้านพืชไร่จากอรรถาภิธานศัพท์เกษตรไทยมาจัดความสัมพันธ์ในรูปแบบลำดับชั้น โดยใช้ตัวแบบเอสเคโอเอส

มหาวิทยาลัยเทคโนโลยีสุรนารี

บทที่ 2
ปริทัศน์วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

การออกแบบและพัฒนาแบบจำลองการสืบค้นเชิงความหมายสำหรับความรู้ด้านพืชไร่ได้มีการศึกษา ค้นคว้าแนวคิดและงานวิจัยที่เกี่ยวข้องจากตำรา เอกสารและบทความตามหัวข้อ ดังนี้

การจัดระบบความรู้ด้านพืชไร่

2.1 การจัดระบบความรู้ด้านพืชไร่

2.1.1 ความรู้ด้านพืชไร่

2.1.2 การจัดระบบความรู้

2.1.3 การจัดระบบความรู้ด้วยออนโทโลยี

2.1.4 งานวิจัยที่เกี่ยวข้องกับการจัดระบบความรู้โดยใช้ออนโทโลยี

2.2 แนวคิดการค้นคืนสารสนเทศ

2. 2.1 ความหมายการค้นคืนสารสนเทศ

2.2.2 รูปแบบการค้นคืนสารสนเทศ

2.2.3 ภาษาสืบค้นความรู้

2.2.4 งานวิจัยที่เกี่ยวข้องกับแนวคิดการค้นคืนสารสนเทศ

2.3 แบบจำลองการสืบค้น

2.3.1 วิวัฒนาการของแบบจำลองการสืบค้น

2.3.2 แนวคิดการสืบค้นเชิงความหมาย

2.3.3 รูปแบบการสืบค้นเชิงความหมาย

2.3.4 งานวิจัยที่เกี่ยวข้องกับแนวคิดการสืบค้นเชิงความหมาย

2.4 การวิเคราะห์คำสำคัญ

2.4.1 การกำหนดความสำคัญของคำค้น

2.4.2 การให้น้ำหนักคำ

2.5 การวัดความคล้ายคลึง

2.6 กรอบแนวคิดการวิจัย

มหาวิทยาลัยเทคโนโลยีสุรนารี

2.1 การจัดระบบความรู้ด้านพืชไร่

2.1.1 ความรู้ด้านพืชไร่

การเกษตรเป็นอาชีพหลักของประเทศไทย ประชาชนส่วนใหญ่ประกอบอาชีพเกษตรกรรม ทำให้การเกษตรมีความสำคัญในการพัฒนาเศรษฐกิจและการพัฒนาประเทศ เนื่องจากรายได้ส่วนใหญ่ของประเทศมาจากสินค้าเกษตรเป็นสำคัญ แต่เศรษฐกิจของประเทศจะเจริญเติบโตมากน้อยเพียงใดขึ้นอยู่กับความรู้ของเกษตรกรว่ามีเพียงใด ไม่ว่าจะเป็น ประสบการณ์ กระบวนการบูรณาการ การแลกเปลี่ยน ถ่ายทอดความรู้ และการเลือกใช้สารสนเทศในการจัดการความรู้ที่เหมาะสม เพื่อให้เกิดองค์ความรู้จนกระทั่งเกิดเป็นความรู้ใหม่ เพื่อพัฒนาผลผลิต และผลิตภัณฑ์ทั้งทางด้านคุณภาพและราคา ให้ก้าวทันต่อกลุ่มประเทศคู่แข่ง โดยเฉพาะในกลุ่มประเทศอาเซียนด้วยกันเอง

จากรายงาน ของสำนักงานเศรษฐกิจการเกษตร กระทรวงเกษตรและสหกรณ์ ปี 2558 ประเทศไทยมีการขยายตัวเศรษฐกิจการเกษตร และในอนาคตมีแนวโน้มเพิ่มขึ้นเรื่อย ๆ โดยเฉพาะผลผลิตในกลุ่มของพืชไร่ เช่น ข้าว มันสำปะหลัง อ้อย และข้าวโพด เป็นต้น เนื่องจากสภาพดินฟ้าอากาศที่เอื้ออำนวยและเป็นข้อได้เปรียบต่อการแข่งขันด้านการเกษตรในกลุ่มประเทศอาเซียน แม้การขยายตัวเศรษฐกิจด้านการเกษตรของกลุ่มพืชไร่จะเพิ่มขึ้น แต่การจัดการความรู้สำหรับพืชไร่ นับเป็นเรื่องค่อนข้างใหม่ จึงพบรายงานไว้ไม่มากนัก (สถาบันส่งเสริมการจัดการความรู้เพื่อสังคม , 2550) ดังนั้นการจัดการความรู้สำหรับด้านพืชไร่จึงมีความจำเป็น เร่งด่วนเพื่อส่งเสริมและสนับสนุนให้มีการแลกเปลี่ยนความรู้ด้านพืชไร่ให้สามารถรองรับการขยายตัวก้าวสู่ประชาคมอาเซียน ซึ่งการจัดเก็บความรู้ด้านพืชไร่ของประเทศไทยมีการจัดเก็บอยู่ในรูปแบบเอกสารและข้อมูลอิเล็กทรอนิกส์ โดยสามารถเข้าถึงแหล่งจัดเก็บความรู้ได้ดังตัวอย่างแสดงดังตารางที่ 2.1

ตารางที่ 2.1 ตัวอย่างแหล่งจัดเก็บและการเข้าถึงความรู้ด้านพืชไร่

แหล่งจัดเก็บความรู้	การเข้าถึงความรู้
กรมส่งเสริมการเกษตร	http://km.doae.go.th/
ศูนย์จัดการความรู้ด้านการเกษตรเพื่อการพัฒนา	http://www.agkmstou.com/2008/index.php
ห้องสมุด สสวท.	http://www.ipst.ac.th/library/index.asp
ห้องสมุดความรู้ด้านการเกษตร	http://www.doae.go.th/library/index.html

มหาวิทยาลัยเทคโนโลยีสุรนารี

ตารางที่ 2.1 ตัวอย่างแหล่งจัดเก็บและการเข้าถึงความรู้ด้านพืชไร่ (ต่อ)

แหล่งจัดเก็บความรู้	การเข้าถึงความรู้
สถาบันวิจัยมูลนิธิสถาบันวิจัยเพื่อการพัฒนาประเทศไทย	http://www.info.tdri.or.th
ศูนย์ความรู้ด้านการเกษตร มหาวิทยาลัยเกษตรศาสตร์	http://www.egov.go.th/Pages/Directory/e-Service/Public/ListServicesByCatTargetDetails.aspx?sc=01199
ฐานการเรียนรู้ด้านการเกษตร มหาวิทยาลัยแม่โจ้	http://www.mju.ac.th/mju2010/Learning_Project.php
ศูนย์สารสนเทศทางการเกษตรแห่งชาติ	http://thaiagris.lib.ku.ac.th/index.php/menu-database
สำนักวิจัยและส่งเสริมวิชาการการเกษตร	http://www2.it.mju.ac.th/dbresearch/raen/
สำนักงานกองทุนสนับสนุนการวิจัย (สกว.)	http://www.trf.or.th/research
ฐานข้อมูลวิจัยไทย	http://www.thairesearch.in.th
ศูนย์ความรู้งานวิจัยการเกษตร	http://kmag.ku.ac.th/
เครือข่ายห้องสมุดในประเทศไทย (สกอ.)	http://dcms.thailis.or.th/dcms/basic.php
สำนักงานเศรษฐกิจการเกษตร	http://www.oae.go.th/main.php?filename=index

กรมวิชาการเกษตร (2544), ปรีชา สุริยพันธุ์ (2544), อุดม เนียบวัน (2544), อัปสร เปลี่ยนสินไชย (2544), ดนัย ศุภาพร (2537), โชติ สิทธิบุศย์ (2526), สุนี ศรีสิงห์ และ เสน่ห์ นิลมณี (2537) และ อรุณี วงษ์กอบรัษฎ์ (2526) ได้อธิบายเกี่ยวกับโครงสร้างความรู้ด้านพืชไร่ประกอบด้วย 1) ลักษณะทางพฤกษศาสตร์ 2) การปลูกและการบำรุงรักษา 3) โรคและแมลง 4) การแปรรูป และ 5) การตลาดและส่งออก

การออกแบบและพัฒนาออนโทโลยีด้านไม้ผลเศรษฐกิจ (ThaiFO Ontology) ของศุภกฤษฎิ์ นิวัฒนากูล (2556) ประกอบด้วยคลาสแทนข้อมูลไม้ผลเศรษฐกิจ คลาสแทนข้อมูลการปลูกและบำรุงรักษา คลาสแทนข้อมูลโรคและแมลง คลาสแทนข้อมูลสารเคมี คลาสแทนข้อมูลการแปรรูป คลาสแทนข้อมูลเครื่องจักรหรือเครื่องมือ คลาสแทนข้อมูลการตลาดและการส่งออก และคลาสแทนข้อมูลเอกสารความรู้ไม้ผลเศรษฐกิจ ซึ่งการออกแบบโครงสร้างออนโทโลยี

ดังกล่าวมีความใกล้เคียงกับโครงสร้างความรู้ด้านพืชไร่ จึงเป็นเหตุผลสำคัญที่ทำให้งานวิจัยนี้มีการประยุกต์ใช้ออนโทโลยีดังกล่าว

2.1.2 การจัดระบบความรู้

มีนักวิชาการหลาย ๆ ท่าน ที่ได้อธิบายเกี่ยวกับการจัดระบบความรู้ ไว้ดังนี้

สำนักคณะกรรมการพัฒนาระบบราชการ และสถาบันเพิ่มผลผลิตแห่งชาติ (2548) ได้อธิบายการจัดระบบความรู้ไว้ว่า เป็นการวางโครงสร้างความรู้เพื่อเตรียมพร้อมสำหรับการเก็บความรู้อย่างเป็นระบบในอนาคต สอดคล้องกับโกศล ดีศิธรรม (2546) ที่ได้อธิบายการจัดเก็บความรู้ (Knowledge Storage) ว่าเป็นการวางโครงสร้างของการเก็บรวบรวมความรู้ เพื่อเตรียมพร้อมสำหรับการเก็บความรู้อย่างเป็นระบบเป็นหมวดหมู่เก็บลงในฐานข้อมูล ซึ่งศุภกฤษฎิ์ นิวัฒนากูล และนิศาชล จำนงศรี (2556) ได้อธิบายการจัดระบบความรู้ ไว้ว่าเป็นกระบวนการในการรวบรวมรายละเอียดของความรู้ที่จัดเก็บในฐานความรู้ เพื่อให้สามารถค้นความรู้กลับคืนมาได้อย่างสะดวกและรวดเร็ว และตรงกับความต้องการของผู้ใช้มากที่สุด และการจัดระบบความรู้ที่นิยมใช้ คือ เคโอเอส (Knowledge Organization System: KOS) ตามที่ ฮอดจ์ (Hodge, 2000) นำเสนอไว้ มี 3 ประเภท ดังนี้

1. รายการข้อความ (Term Lists)

รายการข้อความ หมายถึง การรวบรวมคำสำคัญในขอบเขตทั่วไปหรือเฉพาะเจาะจงไว้ในรูปแบบแฟ้มข้อมูล หรืออาจจะเป็นฐานข้อมูล เพื่อที่จะค้นหาและนำไปใช้ได้โดยสะดวก ซึ่งรายการข้อความ อาจประกอบด้วย

1.1 รายการหลักฐาน (Authority Files) เป็นการแสดงรายการของคำหรือข้อความที่ใช้ในการควบคุมชื่อที่แตกต่างกันสำหรับโดเมนใดโดเมนหนึ่ง เช่น รายชื่อสัตว์น้ำ รายชื่ออุปกรณ์ประกอบอาหาร เป็นต้น

1.2 อภิธานศัพท์ (Glossaries) เป็นรายการของคำหรือข้อความพร้อมกับความหมายของคำหรือข้อความเหล่านั้น

1.3 พจนานุกรม (Dictionaries) เป็นรายการคำตามลำดับตัวอักษรพร้อมกับความหมาย และอาจมีรายละเอียดเพิ่มเติม เช่น คำเหมือน คำตรงข้าม ประเภทของคำ เป็นต้น

1.4 อักขรานุกรมภูมิศาสตร์ (Gazetteers) เป็นรายการของชื่อสถานที่ ซึ่งอาจเป็นชื่อเมือง ภูเขา ถนน ธนาคาร สถานีบริการน้ำมันเชื้อเพลิง เป็นการรวบรวมรายชื่อสถานที่ไว้เพื่อให้สะดวกต่อการค้นหาในภายหลัง

2. การจัดกลุ่มและจำแนก (Classifications and Categories)

การจัดกลุ่มและจำแนก เป็นการจัดมาตรฐานของการจำแนกสิ่งต่าง ๆ ที่อยู่ในโลกนี้ เพื่อให้ผู้ที่เกี่ยวข้องนำไปใช้ในการควบคุมและจัดหมวดหมู่ของวัตถุต่าง ๆ อันจะทำให้มีความเป็นมาตรฐานเดียวกัน จะทำให้ค้นหาได้สะดวกยิ่งขึ้น การจัดกลุ่มและจำแนก อาจประกอบด้วย

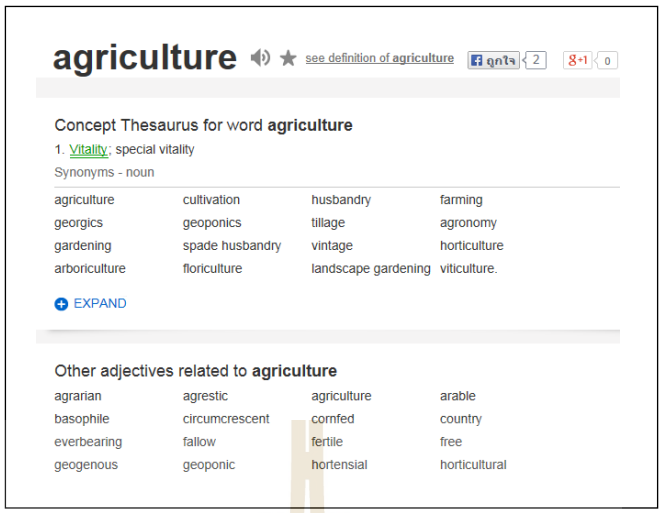
2.1 หัวเรื่อง (Subject Headings) เป็นกลุ่มของข้อความควบคุมเพื่อเป็นตัวแทนหัวเรื่องของรายการหรือวัตถุต่าง ๆ ที่ได้รวบรวมไว้ ดังเช่น ระบบหัวเรื่องของห้องสมุดรัฐสภาอเมริกัน หรือที่รู้จักในชื่อ LCSH (Library of Congress Subject Headings) เพื่อใช้ในการควบคุมหัวเรื่องของรายการหนังสือในห้องสมุด

2.2 รูปแบบการจัดกลุ่ม อนุกรมวิธาน และรูปแบบการจำแนก (Classification Schemes, Taxonomies and Categorization Schemes) เป็นการจัดกลุ่มและจำแนกตามหัวเรื่องที่ได้กำหนดไว้ตามลำดับชั้น การจัดกลุ่มของวัตถุตามคุณลักษณะเฉพาะ ดังเช่น อนุกรมวิธานสัตว์บกจะเป็นระบบการจำแนกประเภทของสัตว์บกตามลำดับชั้นที่ได้กำหนดไว้

3 . รายการความสัมพันธ์ (Relation Lists)

รายการความสัมพันธ์ เป็นระบบการจัดความรู้ให้เป็นระบบที่ซับซ้อนมากกว่าสองกลุ่มที่ได้กล่าวมาแล้วข้างต้น เนื่องจากว่ารายการข้อความเป็นเพียงคำหรือข้อความเท่านั้น ไม่มีการสร้างความสัมพันธ์ระหว่างคำหรือข้อความเหล่านั้น ในขณะที่การจัดกลุ่มและการจำแนกจะมีการสร้างความสัมพันธ์ระหว่างคำหรือข้อความเหล่านั้นแบบเป็นลำดับชั้น แต่รายการความสัมพันธ์จะสามารถสร้างความสัมพันธ์ระหว่างคำหรือข้อความมากกว่านั้น ทำให้การจัดการรายการความสัมพันธ์จะมีความยุ่งยากและซับซ้อนมากกว่า ซึ่งรายการความสัมพันธ์ที่ใช้กันโดยทั่วไปในปัจจุบัน ประกอบด้วย

3.1 อรรถาภิธาน (Thesaurus) เป็นการกำหนดความสัมพันธ์ระหว่างคำต่าง ๆ ที่ได้รวบรวมไว้ ซึ่งความสัมพันธ์ที่สามารถแสดงระหว่างคำหรือข้อความเหล่านั้น อาจมีข้อมูลเกี่ยวกับคำเหมือน (Synomym) คำตรงข้าม (Antonym) คำพ้องรูปพ้องเสียง (Homonym) คำที่ใช้ทั่วไป (Generic Term) คำที่เป็นส่วนหนึ่งของ (Part/Whole Term) คำที่เกี่ยวข้อง (Asssociative Term) และข้อมูลอื่น ๆ จากตัวอย่างอรรถาภิธานของคำว่า “Agriculture” จะทำให้ทราบถึงข้อมูลต่าง ๆ ที่เกี่ยวข้องกับคำนี้ ดังเช่น จะได้ทราบว่าคำเหมือน คำตรงข้าม มีคำใดบ้าง ตัวอย่างดังรูปที่ 2.1



ที่มา: http://thesaurus.com/browse

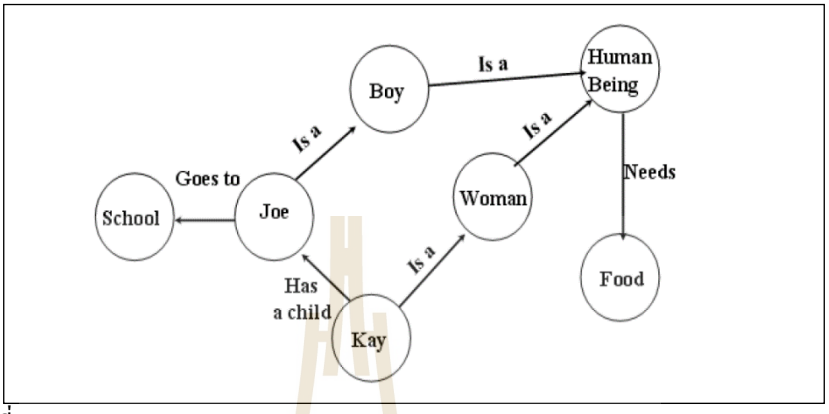
รูปที่ 2.1 ตัวอย่างอรรถาภิธานของ “Agriculture”

องค์การอาหารและเกษตรแห่งสหประชาชาติ (Food and Agriculture Organization – FAO) ได้มีการกำหนดคำศัพท์เพื่อช่วยในการค้นหาความรู้ทางเกษตรในรูปแบบอรรถาภิธานศัพท์สากล AGORVOC Thesaurus ซึ่งรองรับ 22 ภาษา รวมทั้งภาษาไทย (FAO, www, 2013) ซึ่งต่อมา อารีย์ ธัญกิจจานุกิจ และคณะ (2556) ได้ทำการวิจัยและพัฒนาอรรถาภิธานศัพท์เกษตรไทย เพื่อเป็นคำศัพท์มาตรฐานสำหรับกำหนดดรรชนีสารสนเทศเกษตรของไทย โดยจะช่วยเพิ่มประสิทธิภาพในการค้นคืนสารสนเทศทางการเกษตรในการขยายขอบเขตคำค้น ค้นคำศัพท์ที่เป็นคำพ้องเสียงและคำพ้องความหมายซึ่งปัจจุบันมีคำศัพท์ที่ให้บริการมากกว่า 28,000 คำ แสดงดังรูปที่ 2.2

ที่มา: http://pikul.lib.ku.ac.th/www/agrovoc_th_adv.php

รูปที่ 2.2 ตัวอย่างอรรถาภิธานศัพท์เกษตรไทยของคำ “มันสำปะหลัง”

3.2 ข่ายงานความหมาย (Semantic Network) เป็นการจัดเก็บข้อมูลในลักษณะกราฟซึ่งประกอบด้วยโหนด (Node) ซึ่งเป็นตัวแทนแนวคิดหรือสิ่งใดสิ่งหนึ่ง และเส้นเชื่อม (Edge) ซึ่งใช้แทนความสัมพันธ์เชิงความหมายระหว่างโหนดหรือแนวคิดนั้น ๆ ตัวอย่างดังรูปที่ 2.3



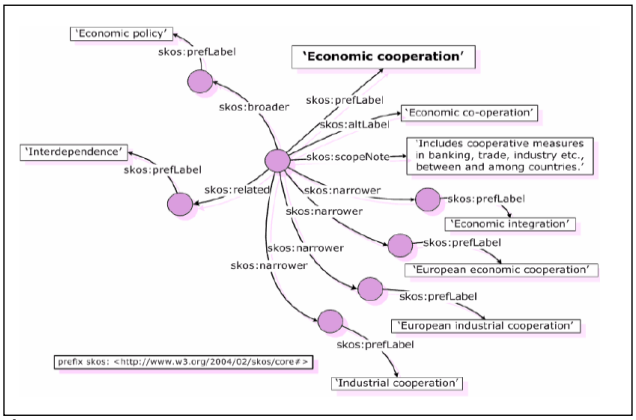
ที่มา: http://www.no-poor.com/JavaandAi/bookAI/

รูปที่ 2.3 ตัวอย่างข่ายงานเชิงความหมาย

3.3 ออนโทโลยี (Ontology) ในทางวิทยาการคอมพิวเตอร์ สตูเดอร์ เบนจามินส์ และเฟนเซล (Studer, Benjamins and Fensel, 1998) ได้ให้ความหมายไว้ว่า ออนโทโลยี คือ รูปแบบข้อกำหนดที่ชัดเจนของแนวคิดที่มีการแบ่งปันหรือใช้ร่วมกัน ซึ่งจากความหมายดังกล่าวออนโทโลยีจะต้องมี รูปแบบ (Formal) ที่เครื่องจักรสามารถประมวลผลได้ ความชัดเจน (Explicit) ของรูปแบบที่มีการกำหนดไว้ และแบ่งปัน (Shared) ตามที่ได้มีการตกลงที่จะใช้งานร่วมกัน

ปัจจุบันการจัดความรู้ให้เป็นระบบ อยู่ในความสนใจของนักวิจัยสำหรับการพัฒนาระบบความรู้ของแต่ละองค์กรเพื่อแก้ไขปัญหาการใช้ศัพท์ควบคุมและการกำหนดนโยบายให้ใช้ระบบหมวดที่เป็นมาตรฐานสากล และการตกลงใช้ระบบมาตรฐานเดียวกันสำหรับหน่วยงาน โดยปัจจุบันมีการนำเทคโนโลยีการจัดเก็บความรู้เข้ามาช่วยจัดการความรู้ให้เป็นระบบ เช่น ใช้ตัวแบบการแทนความรู้ หรือเรียกว่า เอสเคโอเอส (Simple Knowledge Organization Systemm: SKOS) และเมทาดาทา (Metadata) สำหรับอธิบายให้ทราบรายละเอียดของข้อมูล ดังนี้

1. ตัวแบบการแทนความรู้ เอสเคโอเอส (SKOS) ฟุลวิโอ และลอร่า (Fulvio and Laura, 2005) ได้กล่าวไว้ว่า เอสเคโอเอส (SKOS) คือตัวแบบหรือเครื่องมือสำหรับสร้างคลังคำหรือมาตรฐานข้อมูลที่ใช้ร่วมกันในการแก้ไขปัญหาความแตกต่างของข้อมูลที่หลากหลาย โดยเน้นไปที่คำสำคัญ คำพ้อง คำกว้างกว่า (Broader Term-BT) คำแคบกว่า (Narrower Term-NT) และคำเกี่ยวข้อง (Related Term-RT) เป็นแบบจำลองที่ใช้งานง่ายเหมาะสำหรับการแทนความรู้แบบใช้งานร่วมกัน พื้นฐานของ เอสเคโอเอส (SKOS) ประกอบไปด้วย Concepts, Labels, Preferred Labels, Alternate Labels, Hidden Labels แสดงตัวอย่างดังรูปที่ 2.4



ที่มา: ไมล์ (Miles, 2005)

รูปที่ 2.4 การนิยามศัพท์ในเอสเคโอเอส (SKOS)

2. เมทาดาทา คือ การอธิบายให้ทราบรายละเอียดของข้อมูล ซึ่งมีคุณสมบัติ 2 ประการ คือ 1) เป็นข้อมูลที่มีโครงสร้าง และ 2) ต้องอธิบายถึงทรัพยากรสารสนเทศนั้น ๆ เพื่อช่วยในการค้นหาสารสนเทศให้มีความสะดวกง่ายต่อการเชื่อมโยงไปยังสารสนเทศต่าง ๆ รองรับการใช้งานได้หลากหลายด้วยความแตกต่างทางฮาร์ดแวร์ และซอฟต์แวร์

2.1 องค์ประกอบของสารสนเทศ ซึ่งบุญเลิศ อรุณพิบูลย์, สุภาพร ชัยธัมมะปกรณ์ และฐิติมา ธรรมบำรุง (2550) ได้อธิบายถึงความจำเป็นที่ต้องมีเมทาดาทา เนื่องจากสารสนเทศที่สร้างขึ้นมีองค์ประกอบ 3 ลักษณะ ดังนี้

2.1.1 เนื้อหา (Content) ของงาน เกี่ยวกับชื่อเรื่อง หัวเรื่อง ต้นฉบับ (แหล่งที่มา) ภาษา เรื่องที่เกี่ยวข้อง และขอบเขต

2.1 .2 บริบท (Context) ของสารสนเทศ เกี่ยวกับทรัพย์สินทางปัญญาของงาน เช่น ผู้เขียน ผู้สร้างสรรค์ผลงาน สำนักพิมพ์ ผู้มีส่วนร่วมในผลงาน และสิทธิในงานนั้น ๆ

2.1.3 โครงสร้าง (Structure) ของข้อมูลเกี่ยวกับ วัน เดือน ปีที่สร้างผลงาน ประเภทของเนื้อหา รูปแบบการนำเสนอผลงาน และตัวบ่งชี้หรือตัวระบุถึงทรัพยากร หน้าที่ และประเภทของเมทาดาทา

2.2 การทำเมทาดาทา จะต้องเข้าใจถึงประเภทของเมทาดาทาซึ่งจำแนกตามหน้าที่ได้ 3 ประการ ดังนี้

2.2.1 เมทาดาทาเชิงการบริหาร (Administrative Metadata) เป็นการกำหนดเมทาดาทาเพื่อใช้ในการจัดการทรัพยากรสารสนเทศ ทั้งในระยะสั้น และระยะยาว เช่น เป็นเมทาดาทาของข้อมูลเกี่ยวกับการจัดหาสารสนเทศนั้น ๆ การสร้าง การปรับปรุงสารสนเทศ เป็นเมทาดาทาเกี่ยวกับการสงวนรักษาข้อมูล เช่น ข้อมูลเกี่ยวกับการย้ายสภาพ เป็นต้น

2.2.2 เมทาดาทาเชิงพรรณนา (Descriptive Metadata) เป็นเมทาดาทาที่อธิบายเกี่ยวกับทรัพยากรสารสนเทศนั้น ๆ เช่น การลงรายการ ชื่อเรื่อง ชื่อผู้แต่ง หัวเรื่อง การจัดหมวดหมู่

2.2.3 เมทาดาทาเชิงโครงสร้าง (Structure Metadata) เป็นเมทาดาทาที่จะแสดงหรือนำสารสนเทศไปถึงผู้ใช้ เช่น เมทาดาทาของข้อมูลเกี่ยวกับอุปกรณ์หรือเครื่องที่ใช้ในการอ่านภาพ ข้อมูลการแปลงเป็นดิจิทัล เป็นต้น

เมทาดาทาที่รู้จักกันดี ได้แก่ ดับลินคอร์ ซึ่งปัจจุบันดับลินคอร์ได้รับการประกาศเป็นมาตรฐานสากล ISO15836-2003 และมีชุดหน่วยข้อมูลย่อยดับลินคอร์ แบ่งได้เป็น 3 กลุ่ม แสดงดังตารางที่ 2.2 ดังนี้

ตารางที่ 2.2 ชุดหน่วยข้อมูลย่อยดับลินคอร์ 15 อิลีเมนต์ (Element)

เนื้อหาของทรัพยากรสารสนเทศ	**รูปแบบที่ปรากฏให้ใช้งาน**	**ทรัพย์สินทางปัญญา**
Title (ชื่อเรื่อง)	Format (รูปแบบ)	Creator (เจ้าของผลงาน)
Subject (หัวเรื่อง)	Language (ภาษา)	Contributor (ผู้ร่วมงาน)
Source (ต้นฉบับ)	Identifier (รหัส/ตัวบ่งชี้ทรัพยากรสารสนเทศ)	Publisher (สำนักพิมพ์)
Description (ลักษณะ)	Date (วันที่)	Rights (สิทธิ)
Relation (เรื่องที่เกี่ยวข้อง)		
Coverage (ขอบเขต)		
Type (ประเภท)		

จากความสำคัญการจัดระบบความรู้ที่กล่าวมาข้างต้น สรุปได้ว่า การจัดระบบความรู้เป็นกระบวนการรวบรวมความรู้จัดเก็บในฐานความรู้ เพื่อสะดวกต่อการสืบค้น และเทคโนโลยีที่นำมาจัดเก็บความรู้ในปัจจุบัน คือ เอสเคโอเอส (SKOS) เป็นเครื่องมือสำหรับการสร้างคลังคำ และงานวิจัยนี้ได้นำเทคโนโลยีดังกล่าวมาจัดระบบความรู้พืชไร่

2.1.3 การจัดระบบความรู้ด้วยออนโทโลยี

ออนโทโลยี คือ การอธิบายความหมายของสิ่งต่าง ๆ และสามารถจัดหมวดหมู่เอกสารของข้อมูลได้ในขอบเขตความสนใจในเรื่องใดเรื่องหนึ่ง ซึ่งจุดมุ่งหมายของออนโทโลยีเปรียบเหมือนการสร้างพจนานุกรมร่วมของคำเพื่อการแลกเปลี่ยนสารสนเทศ ดังงานวิจัยของ

มหาวิทยาลัยเทคโนโลยีสุรนารี

แพตตัน และออสคาร์ (Paton and Oscar, 1999) ได้ทำการจัดเก็บฐานความรู้อยู่ในรูปแบบของออนโทโลยี ซึ่งช่วง แรก ๆ ใช้รูปแบบของภาษาเอ็กซ์เอ็มแอล (eXtensible Markup Language: XML) โดยชาลอแทม และคณะ (Chalortham, et. al., 2008) ได้อธิบายว่าการใช้รูปแบบภาษาโอดับเบิ้ลยูแอล (Web Ontology Language: OWL)ที่สนับสนุนออนโทโลยีสามารถนำมาเชื่อมโยงข้อมูล และการใช้ข้อมูลร่วมกันได้ดีกว่าระบบฐานข้อมูล สอดคล้องกับ บรอตัน (Broughton, 2006) ที่กล่าวว่าออนโทโลยีเป็นการทำโครงสร้างและการนำเสนอความรู้ในระบบการจัดการความรู้ ระบบการจัดความรู้เป็นเสมือนเครื่องมือที่มีคุณลักษณะเป็นดัชนีที่ช่วยในการจัดการและเข้าถึงเนื้อหา (Subject Access) โดยเป็นภาษาควบคุม (Controlled Vocabularies) ที่ประกอบด้วยคำศัพท์ (Vocabularies /Terms) และไวยากรณ์ (Syntax) หรือ กฎ (Rules) เพื่อประโยชน์ในการกำหนดรูปแบบคำศัพท์ที่เหมาะสม ทำให้สามารถเข้าถึงสาระสำคัญของความรู้ที่มีอยู่และความรู้อื่น ๆ ที่เกี่ยวข้องได้

นอกจากนี้ ฮันเตอร์ (Hunter, 2001), แคทเฟียร์ (Kashyap, 1999) และ เฟนเซลล์ (Fensel, 2004) อธิบายไว้ว่าออนโทโลยีมีบทบาทสำคัญในการบรรยายเชิงความหมายเพื่อความเข้าใจร่วมกันในขอบเขตความรู้เรื่องใดเรื่องหนึ่งให้สอดคล้องตรงกัน โดยใช้แนวคิดเดียว (Concept) เพื่อลดหรือการตัดทอนแนวคิด หรือคำศัพท์ที่สับสน และสนับสนุนการแลกเปลี่ยน การค้นคืนข้อมูลสารสนเทศ และโจนส์ , เบนช์-เคพอน และวิชเซอร์ (Jones, Bench-Capon and Visser, 1998), นอย และแม็คกินนิสส์ (Noy and McGuinness, 2001) ได้อธิบายว่าออนโทโลยีมีความสำคัญต่อการพัฒนาระบบฐานความรู้ (Knowledge Based Systems) เนื่องจากการสร้างออนโทโลยีทำให้สามารถแบ่งปันความรู้และใช้ความรู้ร่วมกัน และสามารถนำความรู้มาใช้ใหม่ได้

ออนโทโลยีมีการกำหนดโครงสร้างความรู้ที่ชัดเจนของสิ่งที่เกี่ยวข้องกับแนวคิด (Concepts) เพื่อนำมาใช้บรรยายหรือแสดงความคิดในขอบเขตความรู้เรื่องใดเรื่องหนึ่ง (Domain) สามารถแสดงให้เห็นถึงคุณสมบัติที่เกี่ยวข้องกับสิ่งนั้น ๆ ซึ่งรูปแบบโดยทั่วไปของออนโทโลยีจะอยู่ในรูปโครงสร้างข้อมูลแบบมีลำดับชั้น (Hierarchical Data Structure) และมีการจัดกลุ่มแนวคิดแบบการจัดหมวดหมู่ (Taxonomy) ในลักษณะความสัมพันธ์แบบพ่อกับลูก (Parent Child) ดังนั้นโครงสร้างออนโลยีจึงประกอบด้วย การกำหนดนิยามความหมายหรือการอธิบายเรื่องราวโดยใช้แนวคิด (Concept) คุณสมบัติ (Properties) ความสัมพันธ์ (Relationships) แอกเซียม (Axioms) และอินสแตนซ์ (Instances)

1. แนวคิด หมายถึงแนวคิดในขอบเขตความรู้เรื่องใดเรื่องหนึ่ง ซึ่งสามารถเป็นอะไรก็ได้ที่เกี่ยวกับสิ่งที่ถูกกล่าวถึงและสามารถอธิบายรายละเอียดได้

2. คุณสมบัติ หมายถึงคุณสมบัติต่าง ๆ ที่เกี่ยวข้องสัมพันธ์กับแนวคิด (Concept) สอดคล้องกับการกล่าวของเจคอป (Jacob, 2003) ที่ว่านำมาใช้ในการอธิบายแนวคิด เช่น ราคาของรถมือสอง (Price of a Used Car) และระยะทางการขับ (Mileage)

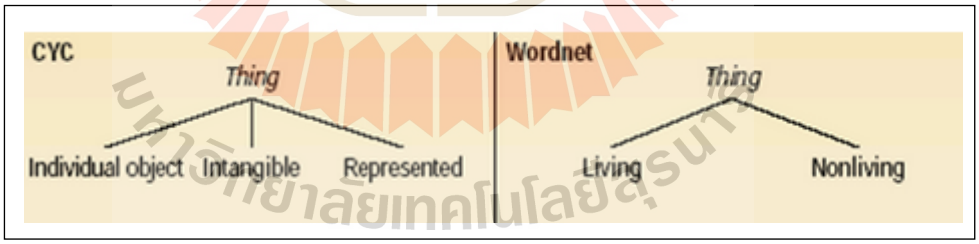
3. ความสัมพันธ์ หมายถึงการนำเสนอประเภทของการปฏิสัมพันธ์ (Interaction) ระหว่างแนวคิดกับแนวคิดในขอบเขตความรู้ที่สนใจ และมีการกำหนดแบบแผนความสัมพันธ์ มีลักษณะต่าง ๆ

4. แอกเซียม หมายถึงตรรกะของการแปลงความสัมพันธ์ระหว่างแนวคิด (Concepts) กับคุณสมบัติ (Properties) เพื่อการแปลความหมายที่ถูกต้อง

5. อินสแตนซ์ (Instances) หมายถึงคำศัพท์หรือตัวข้อมูลที่มีการกำหนดนิยามความหมายร่วมกันทั้งหมดในออนโทโลยี

ไฮย์, ไกร์เบอร์ และ วีลลินกา (Heijst; Schreiber and Wielinga, 1997) การ์รีโน (Guarino, 1997), การ์รีโน (Guarino, 1998) และเฟนเซล (Fensel, 2004) ได้กล่าวถึงประเภทออนโทโลยีสามารถจำแนกได้ตามเนื้อหา ตั้งแต่ระดับทั่วไปหรือระดับสูงสุด ไปสู่ระดับที่มีความเฉพาะเจาะจงทั้งในด้านเนื้อหาและการใช้งาน โดยออนโทโลยีมีบทบาทในฐานะของคำศัพท์ที่ใช้ร่วมกันในกลุ่มชุมชนสารสนเทศเดียวกัน หรือในกลุ่มความรู้เฉพาะด้าน และสำนักงานพัฒนาการวิจัยการเกษตร (2558) ได้แบ่งออนโทโลยีออกเป็น 2 ประเภท ดังนี้

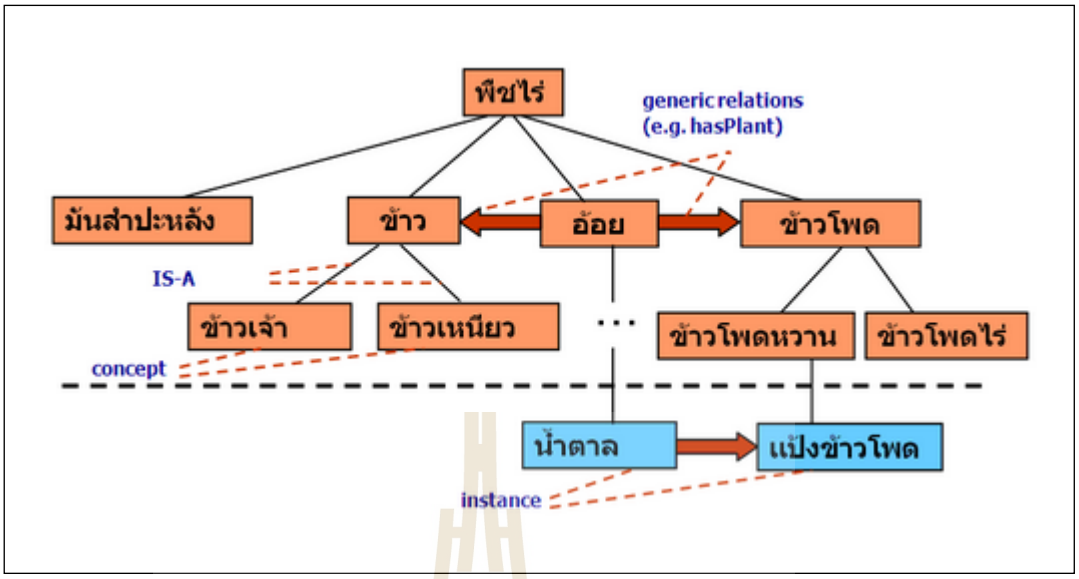
1. ออนโทโลยีทั่วไป คือออนโทโลยีที่มีจุดรวมของโครงสร้างของความรู้ในเชิงกว้าง มีความเป็นนามธรรมค่อนข้างสูงเป็นความรู้เชิงมหภาพสามารถถูกเรียกใช้งาน (Reuse) คุณลักษณะต่าง ๆ ของออนโทโลยีไปใช้งานได้โดยไม่ต้องสร้างใหม่ เช่น WordNet, CYC และ SENSUS เป็นต้น แสดงดังรูปที่ 2.5



ที่มา: สำนักงานพัฒนาการวิจัยการเกษตร (2558)

รูปที่ 2.5 ลักษณะ CYC Ontology และ WordNet Ontology ในระดับ High Level

2. ออนโทโลยีเฉพาะด้าน คือออนโทโลยีที่ใช้ในการแทนความรู้เฉพาะด้านเฉพาะงานซึ่งเหมาะสำหรับการนำไปใช้งานเฉพาะด้าน และจะขึ้นอยู่กับความสัมพันธ์ที่มีของแต่ละระบบงาน อาทิ ออนโทโลยีข้าว และ ออนโทโลยีมันสำปะหลัง เป็นต้น



ที่มา: สำนักงานพัฒนาการวิจัยการเกษตร (2558)

รูปที่ 2.6 ตัวอย่างข้อมูลออนโทโลยีเฉพาะเกี่ยวกับพืชไร่

ออนโทโลยี จัดเก็บอยู่ในรูปแบบภาษาโอดับเบิ้ลยูแอล ซึ่ง ไคยน์ และคณะ (Klyne, et. al., 2004) ได้อธิบายไว้ว่าภาษาโอดับเบิ้ลยูแอลได้รับการพัฒนาขึ้นเพื่อปิดจุดด้อยของภาษาอาร์ดีเอฟ (Resource Description Framework: RDF) โดยการพัฒนาการออกแบบภาษาอาร์ดีเอฟได้ถูกเพิ่มเติมด้วย ภาษาโอดับเบิ้ลยูแอล ซึ่งเป็นเครื่องมือสำหรับการอธิบายความสัมพันธ์ระหว่างคลาส การกำหนดคุณลักษณะ จำนวนสมาชิกและการตั้งค่าที่เกี่ยวข้องกับคุณสมบัติ ฯลฯ ในทางปฏิบัติการพัฒนาเว็บเชิงความหมาย (Semantic Web) ไม่สามารถเขียนภาษาอาร์ดีเอฟหรือ โอดับเบิ้ลยูแอล ได้ด้วยมือ ซึ่งต้องอาศัยซอฟแวร์สำหรับการเขียนโค้ดอัตโนมัติ

จากความสำคัญของ การจัดระบบความรู้ด้วยออนโทโลยี สรุปได้ว่า ออนโทโลยีมีการกำหนดโครงสร้างความรู้ในขอบเขตเรื่องใดเรื่องหนึ่ง ซึ่งสามารถอธิบายความรู้เชิงความหมายเพื่อให้มีความเข้าใจร่วมกันโดยใช้แนวคิด คุณสมบัติ ความสัมพันธ์ แอกเซียม และอินสแตนซ์ และการจัดเก็บปัจจุบันอยู่ในรูปแบบภาษาโอดับเบิ้ลยูแอล ดังนั้นงานวิจัยนี้จึงได้ทำการเลือกออนโทโลยีสำหรับกำหนดความรู้ด้านพืชไร่ เพื่ออธิบายความรู้และรองรับการสืบค้นเชิงความหมาย

2.1.4 งานวิจัยที่เกี่ยวข้องกับการจัดระบบความรู้โดยใช้ออนโทโลยี

จากการศึกษางานวิจัยที่เกี่ยวข้องกับการจัดระบบความรู้โดยใช้ออนโทโลยี มีนักวิจัยหลาย ๆ ท่านนำเอาออนโทโลยีมาจัดระบบความรู้ เพื่อสร้างความเข้าใจร่วมกันในขอบเขตเรื่องที่สนใจให้ตรงกัน ซึ่งเป็นการแก้ปัญหา ความกำกวมของคำสำคัญรองรับการสืบค้นเชิง

ความหมาย ดังงานวิจัยของ โรสริน อัคนิจ และคณะ (2554) ได้นำออนโทโลยีมาจัดระบบความรู้การแปรรูปข้าว ประกอบด้วยการสร้างความสัมพันธ์ 9 กลุ่ม ประกอบด้วย (1) มาตรฐานข้าวไทย (2) คุณภาพในแต่ละด้านของข้าว (3) กระบวนการผลิตผลิตภัฑณ์ข้าว (4) สมบัติของข้าว (5) ประเภทของผลิตภัณฑ์ (6) องค์ประกอบข้าว (7) ผลิตภัณฑ์จากข้าว (8) เครื่องมือในการตรวจวัด และ (9) เครื่องจักรในการแปรรูป ซึ่งแนวคิดในออนโทโลยีการแปรรูปข้าวมีทั้งหมด 907 มโนทัศน์ และคำแทนมโนทัศน์มี 1863 คำ ซึ่งออนโทโลยีที่พัฒนาขึ้นมีคุณสมบัติในการขยายคำค้นโดยสามารถดึงความรู้ที่เกี่ยวข้องจากการระบุคำสำคัญของผู้ใช้ ผลการวิจัยพบว่าการจัดระบบความรู้ด้วยออนโทโลยีสามารถเพิ่มประสิทธิภาพการสืบค้นตรงกับความต้องการของผู้ใช้มากขึ้น

นอกจากนี้ ปฏิคม ทองจริง (2552) ได้วิจัยการจัดระบบความรู้ด้านสมุนไพรไทยโดยใช้ออนโทโลยี จัดเก็บในรูปแบบของภาษาโอดับเบิ้ลยูแอล ซึ่งได้นำความสัมพันธ์ของคำและความหมายของคำที่เกี่ยวข้องกับสมุนไพรไทยมาใช้ในการค้นคืนข้อมูล สอดคล้องกับงานวิจัยของไมล์ และคณะ (Miles, et. al., 2007) ได้เสนอเทคโนโลยีที่จะนำมาจัดเก็บความรู้ คือเอสเคโอเอส (SKOS) ที่เป็นเครื่องมือสำหรับการสร้างคลังคำ โดยการใช้ศัพท์ควบคุมสร้างข้อมูลด้วยโครงสร้างดับลินคอร์ด้วยภาษาเอ็กซ์เอ็มแอล และอาร์ดีเอฟ โดยประยุกต์ใช้รายการคำจากอรรถาภิธานศัพท์สำหรับการจัดกลุ่ม (Classification Schemes) คำต่าง ๆ ในการจัดเก็บข้อมูล เพื่อให้ระบบการสืบค้นสามารถสืบค้นให้ผลลัพธ์ที่มีความถูกต้องตามความต้องการและรวดเร็ว และงานวิจัยของ ศุภกฤษฎิ์ นิวัฒนากูล (2556) ได้จัดเก็บความรู้ด้วยเทคโนโลยีและวิธีการดังกล่าวซึ่งออกแบบออนโทโลยีความรู้ด้านไม้ผลเศรษฐกิจของประเทศไทย โดยใช้เอสเคโอเอสสำหรับการจัดเก็บความรู้ และนำรายการคำจากอรรถาภิธานศัพท์เกษตรไทย ส่งผลให้ประสิทธิภาพการสืบค้นเพิ่มขึ้นมีความถูกต้องแม่นยำตรงกับความต้องการของผู้ใช้

จากการศึกษางานวิจัยที่เกี่ยวข้องกับการจัดระบบความรู้ด้วยออนโทโลยี สรุปได้ว่า ปัจจุบันมีการนำออนโทโลยีเข้ามาช่วยในการจัดระบบความรู้ เพื่อรองรับการสืบค้นเชิงความหมาย เพราะออนโทโลยีมีการกำหนดโครงสร้างความรู้ในขอบเขตที่สนใจชัดเจน และที่สำคัญมีการใช้คำศัพท์ร่วมกันในกลุ่มความรู้เดียวกัน หรือกลุ่มความรู้เฉพาะด้านได้

2.2 แนวคิดการค้นคืนสารสนเทศ

2.2.1 ความหมายการค้นคืนสารสนเทศ

การค้นคืนสารสนเทศ (Information Retrieval) เป็นคำที่มีขอบเขตกว้างและมีความหมายที่หลากหลาย ซึ่งจันทิมา เขียวแก้ว (2557) ได้อธิบายเกี่ยวกับการค้นคืนสารสนเทศเป็นกระบวนการดึงหรือค้นหาข้อมูลย้อนหลังที่มีการจัดเก็บไว้ตามหัวข้อตามความสนใจผู้ใช้ ซึ่งสามารถค้นจากเครื่องคอมพิวเตอร์โดยตรงหรือเครื่องที่เชื่อมโยงกับเครือข่ายอินเทอร์เน็ต โดยมีการทำงาน 2 ส่วนคือ การจัดเก็บสารสนเทศ และการค้นคืนสารสนเทศ สอดคล้องกับ ศุภชัย ตั้งวงศ์ศานต์ (2551) ที่ได้อธิบายว่าการค้นคืนสารสนเทศ หมายถึง การกระทำใด ๆ ที่คัดเลือกสารสนเทศจากแหล่งเก็บข้อมูลที่ต้องการ โดยเป็นความรู้เกี่ยวกับเรื่องใดเรื่องหนึ่งที่ผู้ใช้สนใจ ซึ่งกระบวนการค้นคืนสารสนเทศที่ดีควรมีการจัดเก็บสารสนเทศที่ดี และใช้เทคนิควิธีการเข้าถึงข้อมูลที่เหมาะสม โดยอุไร ทองหัวไผ่ (2551) ได้อธิบายเพิ่มเติมว่าผลลัพธ์ที่ได้จากการค้นคืนอาจเป็นข้อมูล หรือรายการเอกสารที่บรรจุความรู้ตามความต้องการของผู้ใช้ โดยผลลัพธ์ที่ได้จากการค้นคืนจะมีความถูกต้อง รวดเร็ว และตรงกับความต้องการของผู้ใช้ได้นั้นก็ขึ้นอยู่กับกระบวนการจัดเก็บสารสนเทศ การเข้าถึงข้อมูล และการใช้เทคนิควิธีการค้นคืนสารสนเทศต่าง ๆ เพื่อนำเสนอรูปแบบผลลัพธ์ให้กับผู้ใช้

ต่อมา สิทธิโชค ปัญญาฤกษ์ชัย และศิพาณี นุชิตประสิทธิ์ชัย (2552) ได้สนับสนุนแนวคิดดังกล่าวโดยอธิบายว่าการค้นคืนสารสนเทศเป็นกระบวนการที่เกี่ยวข้องกับการเก็บสารสนเทศ การเข้าถึงความรู้ และการจัดรูปแบบการนำเสนอให้ตรงกับความต้องการของผู้ใช้ และปริศนา มัชฌิมา (2552) ได้อธิบายว่าการค้นคืนสารสนเทศต้องใช้เทคโนโลยีสารสนเทศที่เหมาะสมมาเป็นเครื่องมือสำหรับการจัดเก็บ โดยใช้เทคนิควิธีอย่างเป็นขั้นตอนสำหรับเข้าถึง เพื่อนำเสนอรูปแบบผลลัพธ์ให้มีความสะดวก รวดเร็ว ถูกต้องและตรงกับความต้องการของผู้ใช้มากที่สุด และปัจจุบันผู้ใช้นิยมเข้าถึงสารสนเทศต่าง ๆ จากการใช้โปรแกรมค้นหา บนเครือข่ายอินเทอร์เน็ต เช่น กูเกิ้ล (Google) ยาฮู (Yahoo) และ บิง (Bing) ซึ่งโปรแกรมดังกล่าวมีการพัฒนาอย่างต่อเนื่องเพื่อให้ได้ความรู้ที่ตรงกับความต้องการของผู้ใช้

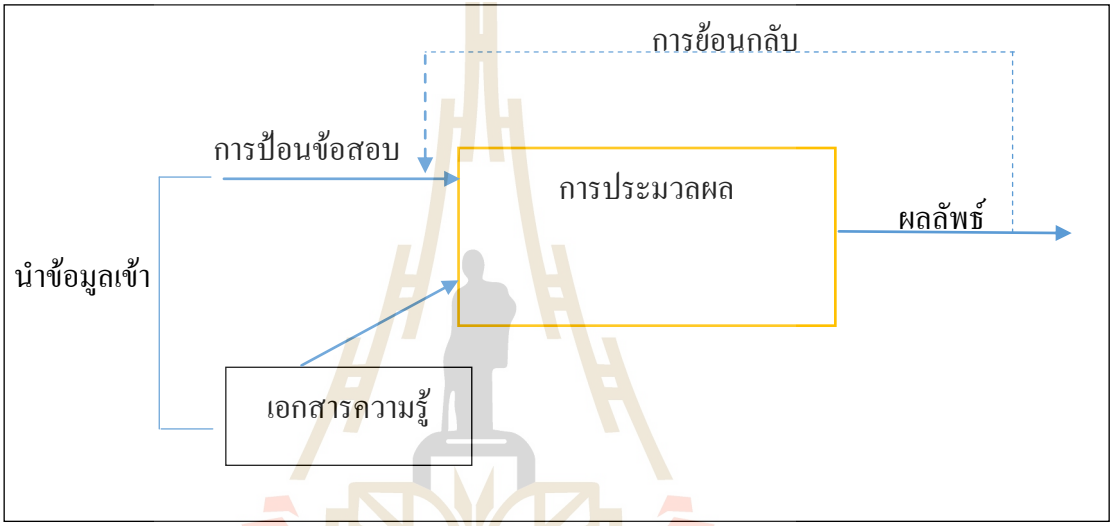
จากความหมายที่กล่าวมาข้างต้น สรุปได้ว่า การค้นคืนสารสนเทศ หมายถึง กระบวนการค้นหาความรู้ที่เกี่ยวข้องกับคำค้นใน 3 ขั้นตอน ได้แก่ การกำหนดลำดับความสำคัญคำค้น การให้น้ำหนักคำสำคัญ และการวัดความคล้ายคลึง โดยสามารถนำเสนอรูปแบบของผลลัพธ์ได้ถูกต้อง รวดเร็ว และตรงกับความต้องการของผู้ใช้

2.2.2 รูปแบบการค้นคืนสารสนเทศ

การค้นคืนสารสนเทศมีกระบวนต่าง ๆ ที่หลากหลาย ซึ่งนักวิชาการหลาย ๆ ท่าน ได้สรุปไว้ดังนี้

1. กระบวนการค้นคืนสารสนเทศ

อุไร ทองหัวไผ่ (2551) ได้อธิบายแนวคิดการค้นคืนสารสนเทศ ประกอบด้วย 3 ส่วน ได้แก่ การนำเข้าข้อมูล การประมวลผล และผลลัพธ์ แสดงดังรูปที่ 2.7



ที่มา: อุไร ทองหัวไผ่ (2551)

รูปที่ 2.7 แนวคิดการค้นคืนสารสนเทศ

1.1 ส่วนนำเข้าข้อมูล (Input) คือส่วนของการป้อนข้อสอบถาม (Query) จากผู้ใช้ ซึ่งเป็นคำสำคัญ หรือภาษาธรรมชาติ หรืออาจเป็นการนำเอกสารความรู้มาใช้ในการประมวลผลโดยใช้เมทาดาทา (Metadata) สำหรับอธิบายรายละเอียดความรู้นั้น ๆ ซึ่งความรู้ที่นำเข้ามาเป็นข้อมูลที่เกี่ยวข้องกับเอกสาร เช่น

1.1.1 การนำเข้าสารสนเทศที่เป็นความหมายของเอกสารที่อยู่ภายนอก เช่น ผู้แต่ง (Author) ชื่อเรื่อง (Title) แหล่งที่มา (Source: หนังสือ วารสาร หนังสือพิมพ์ และบทความ) วันที่ (Date) ISBN สำนักพิมพ์ (Publisher) และความยาว (Length)

1.1.2 การนำเข้าเนื้อความที่มีความหมาย เช่น บทคัดย่อ (Abstract) คำสำคัญ (Keywords) รหัสของหัวเรื่อง (Subject Codes) ซึ่งอาจเป็น Library of Congress หรือ Dewey Decimal หรือ UMLS (Unified Medical Language System)

1.1.3 การนำเข้าคำสำคัญของหัวเรื่อง ซึ่งอาจเป็นลำดับชั้นของคำค้นมาตรฐาน (Hierarchical Taxonomies of Standardized Semantic Terms)

1.1.4 การนำเข้าสารสนเทศของเว็บ (Web Information) เช่นแท็กเมทาในเอชทีเอ็มแอล ดังนี้

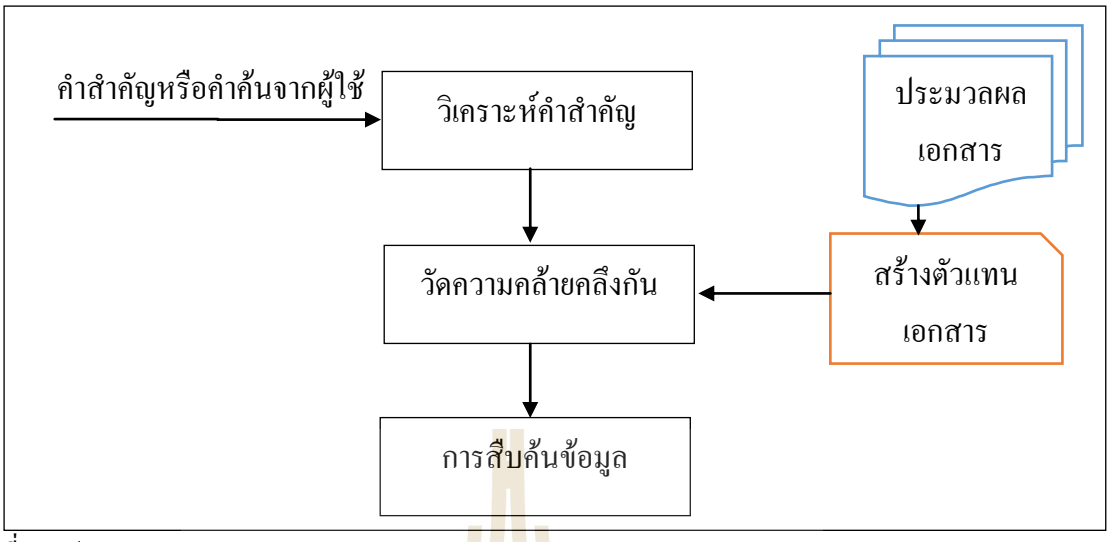
<META NAME="keywords" CONTENT=" Agronomy, Cassava, Sugarcane, Maize">

การค้นคืนสารสนเทศจะนำสารสนเทศที่กล่าวมาข้างต้นผ่านกระบวนการประมวลผลแบบเชื่อมโดยตรงกับระบบคอมพิวเตอร์ โดยผู้ใช้สามารถโต้ตอบหรือปฏิสัมพันธ์กับระบบโดยตรง

1.2 ส่วนการประมวลผล (Processor) คือส่วนของการจัดโครงสร้างของสารสนเทศในรูปแบบที่เหมาะสม ประกอบด้วย การสร้างตัวแทนเอกสาร การแบ่งกลุ่มของเอกสาร การจัดเก็บสารสนเทศ และการดึงข้อมูลตามที่ผู้ใช้ต้องการ ซึ่งกระบวนการจะมีการนำข้อสอบถามไปเปรียบเทียบกับตัวแทนเอกสารที่มีอยู่ เพื่อดึงเอกสารหรือความรู้ที่ใกล้เคียงแสดงผลลัพธ์ให้แก่ผู้ใช้

1.3 ส่วนผลลัพธ์ (Output) คือเอกสารหรือความรู้ที่ได้จากระบบอาจเป็นข้อความสั้น ๆ เช่น ชื่อเอกสาร หมายเลขเอกสาร ชื่อผู้แต่ง คำสำคัญ และรายละเอียดโดยย่อ เป็นต้น โดยผู้ใช้สามารถพิจารณาจากข้อมูลต่าง ๆ ที่ได้จากระบบ ในกรณีที่ได้ผลลัพธ์จำนวนมากเกินไปหรือไม่ใกล้เคียงกับสิ่งที่ต้องการ ผู้ใช้สามารถปรับปรุงข้อสอบถามใหม่เพื่อให้ข้อสอบถามนั้นสืบค้นสารสนเทศได้ตรงกับความต้องการมากที่สุด ซึ่งเป็นการย้อนกลับ (Feedback) ดังนั้นผลลัพธ์ที่ได้จึงขึ้นอยู่กับข้อสอบถาม (Query) ของผู้ใช้

ส่วนประกอบต่าง ๆ ของแนวคิดการค้นคืนสารสนเทศที่กล่าวมาข้างต้นต้องมีกระบวนการต่าง ๆ อันจะส่งผลให้การค้นคืนสารสนเทศได้ผลลัพธ์ตรงกับความต้องการของผู้ใช้ ซึ่งเฟรค (Frake, 1992) ได้เสนอกระบวนการค้นคืนสารสนเทศประกอบด้วย 4 ส่วน ได้แก่ การสร้างตัวแทนเอกสาร หรือดัชนี (Index) การวิเคราะห์คำสำคัญ การวัดความคล้ายคลึง และการสืบค้นข้อมูล ดังรูปที่ 2.8



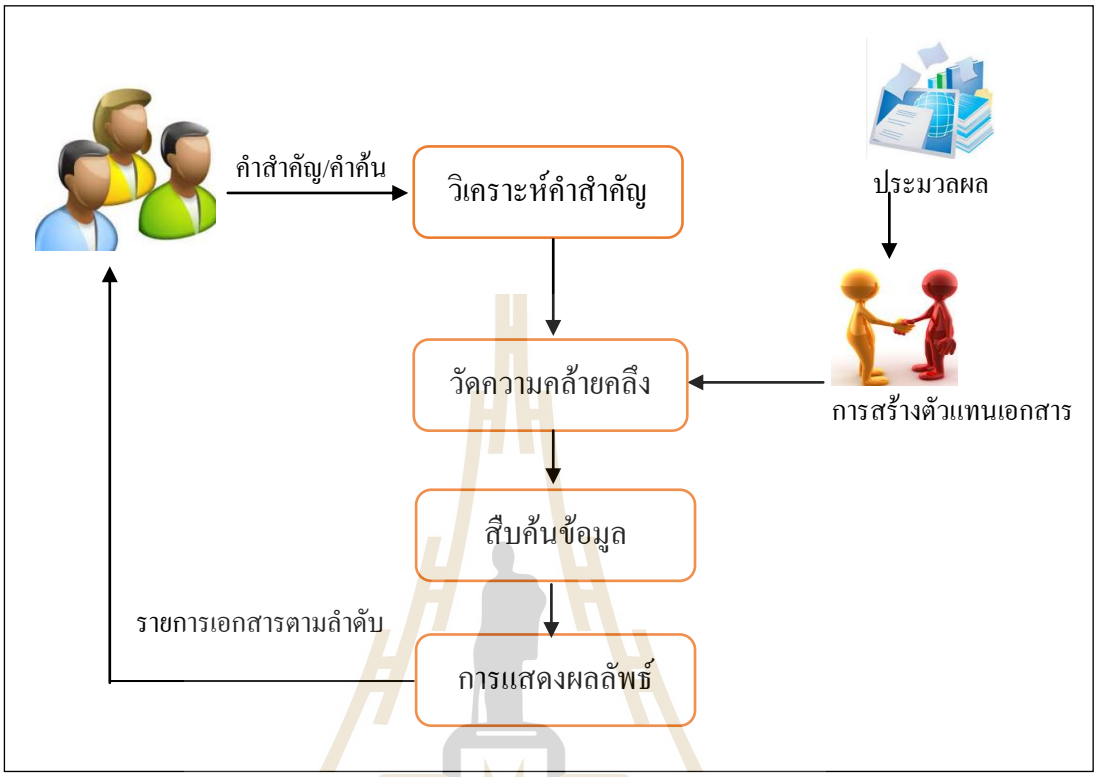
ที่มา: เฟรค (Frake, 1992)

รูปที่ 2.8 ตัวอย่างกระบวนการค้นคืนสารสนเทศนำเสนอโดยเฟรค

ต่อมา เบย์ซา เยตส์ และริบริโร เนโท (Baeza-Yates and Ribeiro-Neto, 1999) ได้สนับสนุนแนวคิดกระบวนการค้นคืนสารสนเทศดังกล่าวข้างต้น โดยอธิบายไว้ว่าเมื่อผู้ใช้งานระบุคำสำคัญหรือคำค้นเข้ามา คำสำคัญเหล่านั้นจะผ่านกระบวนการวิเคราะห์คำสำคัญ เช่น การให้ความสำคัญกับลำดับของคำ การจัดกลุ่มคำ และการตรวจหลักไวยกรณ์ของคำให้ถูกต้อง เป็นต้น และกระบวนการวัดความคล้ายคลึง เช่น การวัดความคล้ายคลึงระหว่างคำ กลุ่มคำ และคำกับเอกสาร เป็นต้น ซึ่งกระบวนการนี้มีการเปรียบเทียบความคล้ายคลึง (Similarity Computation) ของตัวแทนคำสำคัญกับตัวแทนเอกสาร จากนั้นทำการสืบค้นข้อมูล และแสดงผลลัพธ์รายการเอกสารโดยการจัดอันดับตามความคล้ายคลึงของเอกสาร

หลักการสืบค้นข้อมูล สารสนเทศจะมีกระบวนการประมวลผลเอกสาร (Document/Text Operations) จากการจัดหมวดหมู่ แยกประเภท การจัดเรียง เช่นการใช้กระบวนการตัดคำ (Word Segmentation) โดยเป็นการแบ่งตัวอักษรจากข้อความเพื่อหาขอบเขตของแต่ละหน่วยคำ แล้วนำไปเทียบกับพจนานุกรม และจากนั้นก็จะทำการสร้างตัวแทนเอกสารหรือดัชนี (Index or Document Representation) คือการสร้างตัวแทนเอกสารในรูปแบบต่าง ๆ อย่างมีมาตรฐานและจัดเก็บในรูปของฐานข้อมูล เช่น การใช้ระบบดัชนีคำเพื่อช่วยให้การสืบค้นที่รวดเร็วขึ้นเนื่องจากลดกระบวนการสืบค้นคำจากข้อมูลต้นฉบับโดยตรง แต่มีการสืบค้นจากตารางจัดเก็บดัชนีแทนเพื่อความรวดเร็ว ซึ่งการสร้างตัวแทนสำคัญ หรือคำค้น (Query Representation) ได้แก่ 1) การขยายคำค้น (Query Expansion) คือการทำให้การค้นมีประสิทธิภาพมากขึ้นด้วยการใช้ตัวดำเนินการ (Operation) บางตัวเพิ่มเข้าไป เช่น หากต้องการข้อมูลใด ๆ สามารถเพิ่มการค้นด้วย +train และ/หรือ +car ร่วมด้วย 2) การเปลี่ยนคำค้น (Query Transformation) คือการป้อนความเกี่ยว

พันธ์ย้อนกลับ (Relevance Feedback) ซึ่งเป็นการเปลี่ยนคำค้นเก่าให้เป็นคำค้นใหม่สำหรับการสืบค้นในแต่ละรอบ เพื่อให้ได้ผลลัพธ์ที่แตกต่างไปจากเดิม แสดงดังรูปที่ 2.9



ที่มา: เบย์ซา เยตส์ และริบริโร เนโท (Baeza-Yates and Ribeiro-Neto, 1999)

รูปที่ 2.9 ตัวอย่างกระบวนการค้นคืนสารสนเทศนำเสนอโดยเบย์ซา เยตส์ และริบริโร เนโท

จากแนวคิดการค้นคืนสารสนเทศที่กล่าวมามีข้อดีคือสามารถลดขนาดของการจัดเก็บเอกสารหรือความรู้สอดคล้องกับงานวิจัยของ ศุภกฤษฎิ์ นิวัฒนากูล (2556) ที่สามารถทำให้การค้นคืนสารสนเทศ มีความถูกต้อง รวดเร็ว และการที่จะได้ผลลัพธ์ตรงกับความต้องการผู้ใช้นั้นควรคำนึงถึงเทคนิควิธีการค้นคืนสารสนเทศ

สรุป แนวคิดการค้นคืนสารสนเทศ ประกอบด้วย 3 ส่วน ได้แก่ การนำเข้าข้อมูล การประมวลผล และการแสดงผลลัพธ์ ซึ่งในแต่ละส่วนจะมีกระบวนการต่าง ๆ อันจะส่งผลให้การค้นคืนได้ผลลัพธ์ตรงกับความต้องการ ได้แก่ การวิเคราะห์คำสำคัญ การวัดความคล้ายคลึง การสืบค้นข้อมูล และการแสดงผลลัพธ์ ซึ่งงานวิจัยนี้ได้นำแนวคิดข้างต้นมาประยุกต์ใช้ในงานวิจัยดังนี้

1. ส่วนการนำเข้าข้อมูล เป็นการป้อนข้อสอบถามจากผู้ใช้ที่เป็นคำสำคัญในลักษณะคำเดี่ยวที่เป็นทั้งภาษาไทย และภาษาอังกฤษ โดยสกัดคำจากชื่อเรื่อง (Title) คำอธิบายความรู้โดยย่อ (Descriptioin/Abstract) คำสำคัญ (Keywords/Subject) และแหล่งอ้างอิงต้นฉบับ

(Source) ของความรู้ด้านพืชไร่ และใช้ดับลินคอร์เมทาดาทา (Dublin Core Metadata) สำหรับอธิบายรายละเอียดความรู้

2. ส่วนการประมวลผล (Processor) เป็นกระบวนการวิเคราะห์คำสำคัญ ได้แก่ 1) การวิเคราะห์คำสำคัญ ประกอบด้วยการกำหนดคำสำคัญ และการให้น้ำหนักคำ 2) การวัดความคล้ายคลึงเชิงความหมายระหว่างคำ

3. ส่วนการแสดงผลลัพธ์ (Output) เป็นกระบวนการจัดลำดับผลลัพธ์ตามผลการเปรียบเทียบความคล้ายคลึงกันของเอกสาร

2.2.3 ภาษาสืบค้นความรู้

1 . ภาษาสอบถามอาร์ดีคิวแอล (RDQL)

รูดอสลอว์ (Radossslaw, 2004) และซีบอร์น (Seaborne, 2004) ได้อธิบายไว้ว่า ภาษาสอบถามอาร์ดีคิวแอลเป็นภาษาในการสืบค้นข้อมูลบนโครงสร้างข้อมูลอาร์ดีเอฟ โดยรูปแบบของภาษาอาร์ดีคิวแอล มีลักษณะที่คล้ายคลึงกับภาษาในการสอบถามข้อมูลแบบโครงสร้างเอสคิวแอล (SQL) ของฐานข้อมูลเชิงสัมพันธ์ ต่างกันตรงที่อ้างอิงข้อมูลทุกครั้งจะต้องอ้างถึงยูอาร์ไอ (URI) ด้วย

อาร์ดีคิวแอล มีรูปแบบของการกำหนดตัวแปรในภาษาจะใช้เครื่องหมายคำถาม (?) นำหน้าชื่อตัวแปรเสมอ และต้องมีเครื่องหมายคอมม่า (,) คั่นระหว่างตัวแปรแต่ละตัว ภาษาอาร์ดีคิวแอล มีรูปแบบประกอบไปด้วย ดังนี้

- SELECT clause เป็นการระบุรายการของตัวแปรที่ต้องการจากเงื่อนไข WHERE ซึ่งตัวแปรในภาษาอาร์ดีคิวแอล จะขึ้นต้นด้วยเครื่องหมายคำถาม (?) และถ้ามีหลายตัวจะคั่นด้วยเครื่องหมาย คอมม่า (,) เช่น SELECT ?name, ?email, ?age, ?tel_number หรือถ้าต้องการตัวแปรทั้งหมด จะใช้เครื่องหมายดอกจัน * เช่น SELECT *
- FROM clause ใช้ในการระบุหรือไม่ต้องการระบุเส้นทาง (Path) ตำแหน่งที่อยู่ของเว็บไซต์ (Uniform Resource Locator: URL)
- WHERE clause เป็นการระบุรายการในรูปแบบ Triple ที่ต้องการผลลัพธ์ คือ (Subject,Predicate,Object) เช่น WHERE (?Resource,<http://www.w3.org/2001/vcard-rdf/3.0/EMAIL>,"jatsada.si@ksu.ac.th") หมายถึงต้องการค้นหา?Resource ที่มีอีเมลเป็นjatsada.si@ksu.ac.th
- USING ช่วยในการอธิบายยูอาร์ไอ ที่ใช้ให้อยู่ในรูปแบบชื่อสั้น ๆ เพื่อให้ง่ายต่อการอ่านคำสั่ง และการพิมพ์ เช่น WHERE (?Resource, vCard:EMAIL, radol@gmx.de") USING vCard FOR <http://www.w3.org/2001/vcard-rdf/3.0/> หมายถึงใช้ vCard เป็นชื่อย่อของ URI http://www.w3.org/2001/vcard-rdf/3.0/

รูปแบบการเขียนภาษาอาร์ดีคิวแอล มีโครงสร้าง ดังนี้

SELECT variables listing FROM rdf documents WHERE patterns AND filter expressions USING prefix declaration

การสืบค้นของภาษาอาร์ดีคิวแอลใช้หลักในการสืบค้นโดยอาศัยการเชื่อมโยงของความรู้จำนวนมากที่เชื่อมต่อกันอยู่ในกราฟจนกว่าจะสามารถหาการเชื่อมโยงไปจนถึงความรู้ที่เป็นผลลัพธ์ที่ต้องการ

ตัวอย่างการเขียนภาษาอาร์ดีคิวแอล

SELECT ?A, ?B

WHERE (Subject1,Predicate1,Object1 or Literal1), (Subject2,Predicate2,Object2 or Literal2),(…)

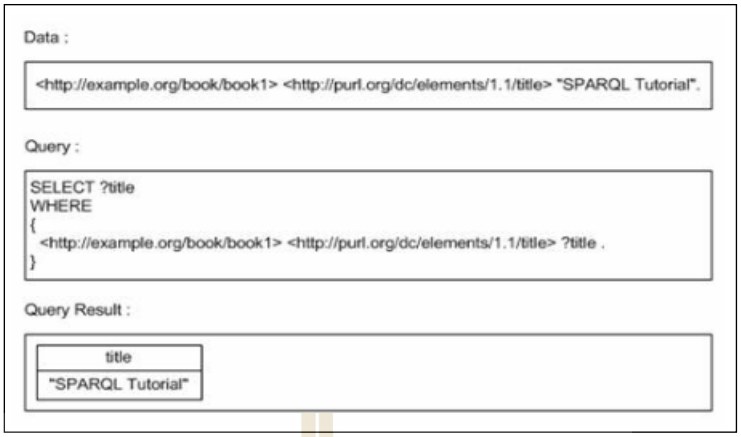
USING ชื่อ Prefix FOR <URI ของ NameSpace>

จากตัวอย่างในการสืบค้นของภาษาอาร์ดีคิวแอลข้างต้นจะพบว่าตัวแปร ?A และ ?B เป็นตัวแปรที่ใช้แทนผลลัพธ์ที่ได้จากการสืบค้นข้อมูลจากเอกสาร โดยมีเงื่อนไขตามที่กำหนด และในส่วน (Subject1, Predicate1, Object1 or Literal1) แสดงถึงลำดับในการอ้างอิงถึงในแต่ละส่วนของความรู้ตามลำดับ

2 . ภาษาสอบถามสปาร์เก้ิล (SPARQL)

สปาร์เก้ิล เป็นภาษาสำหรับการสืบค้น ความรู้จากออนโทโลยี สามารถทำการดึงข้อมูลมาแสดง หรือเรียกว่าภาษาสอบถามความรู้ หรือ “Query Language” ซึ่งอยู่บนพื้นฐานของข้อมูลที่เป็นไปในรูปแบบของกราฟ ซึ่งมีลักษณะในรูปแบบของอาร์ดีเอฟ และโอดับเบิ้ลยูแอล สอดคล้องกับการศึกษาของเวงและคณะ (Wang, et. al., 2009) ได้ทำการศึกษาปัญหาเกี่ยวกับการเขียนคำสั่งสำหรับสืบค้นความรู้ และได้มีกระบวนการสืบค้นจากแหล่งข้อมูลที่หลากหลาย มีกระบวนการทำงาน 2 ส่วน คือ 1) เป็นส่วนการหาความสัมพันธ์ภายในตารางทั้งหมดแล้วนำมาแยกส่วนเป็น Minimal Connectable Unit (MCU) ซึ่งหมายถึงกลุ่มย่อยความรู้ตามข้อกำหนดของแหล่งจัดเก็บที่ได้อธิบายไว้ 2) ทำการเชื่อมโยง (JION) MCUs เพื่อสืบค้นเชิงความหมายตามรูปแบบของกราฟ โดยที่ Q เป็นคำสั่งภาษาสอบถามสปาร์เก้ิล ซึ่ง D_p เป็นการอธิบายแหล่งจัดเก็บตาราง $D_p = (Q_p, \mu)$ ซึ่ง P คือตาราง , μ คือความสัมพันธ์

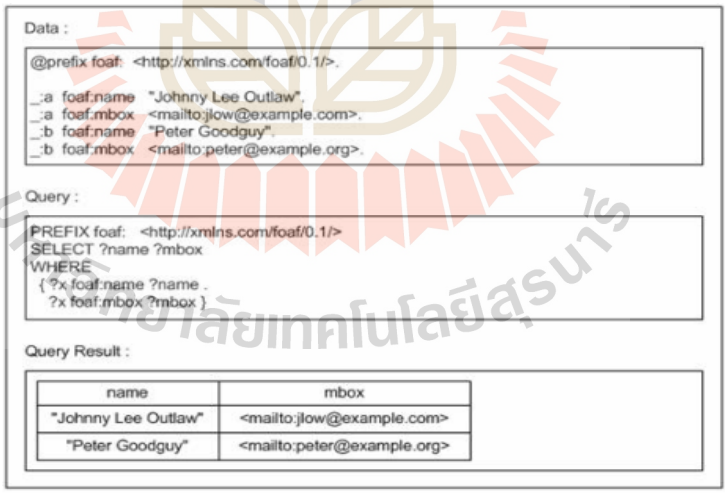
สมมุติกรณีที่เราต้องการค้นหาหัวข้อเรื่องของหนังสือจากข้อมูลที่ให้ในลักษณะของกราฟที่เป็นอาร์ดีเอฟ โดยภาษาที่จะดึงข้อมูลแบ่งเป็นสองส่วนคือ SELECT และ WHERE ซึ่ง SELECT จะไปอธิบายตัวแปรที่จะไปปรากฏบนผลลัพธ์ และ WHERE จะเป็นเงื่อนไข แสดงดังรูปที่ 2.10



ที่มา: http://www.w3.org/TR/rdf-sparql-query/

รูปที่ 2.10 ตัวอย่างแสดงการดึงข้อมูลโดยใช้ภาษาสปาร์เก้ิลแบบเงื่อนไขเดียว

กรณีที่ต้องการค้นหาชื่อและอีเมล ซึ่งเป็นการค้นหาแบบเปรียบเทียบหลายข้อมูล โดยข้อมูลที่มีนั้นมีสองส่วนคือ ชื่อ และอีเมล ในการดึงข้อมูลนั้นก็กำหนดส่วนที่ต้องการจะดึงข้อมูล โดยใช้ PREFIX แล้ว SELECT ข้อมูลที่ต้องการคือ ทั้งชื่อ และอีเมล โดยระบุเงื่อนไข WHERE ที่ต้องการชื่อ และอีเมลของกลุ่มคนนั้น แสดงดังรูปที่ 2.11



ที่มา: http://www.w3.org/TR/rdf-sparql-query/

รูปที่ 2.11 ตัวอย่างแสดงการดึงข้อมูลโดยใช้ภาษาสปาร์เก้ิลแบบเปรียบเทียบหลายข้อมูล

จากความสำคัญของภาษาสืบค้นความรู้ที่กล่าวมาข้างต้น สรุปได้ว่า ภาษาที่ใช้สำหรับการสืบค้นความรู้มีการใช้ภาษาสอบถามอาร์ดีคิวแอลสำหรับการค้นคืนสารสนเทศบนโครงสร้างข้อมูลอาร์ดีเอฟ ซึ่งใช้หลักการสืบค้นโดยอาศัยการเชื่อมโยงของความรู้ที่เชื่อมต่อกันอยู่

ในกราฟ ซึ่งภาษาสอบถามสปาร์เก้ิลสามารถค้นคืนสารสนเทศบนโครงสร้างข้อมูลอาร์ดีเอฟ และโอดับเบิ้ลยูแอล ซึ่งสามารถหาความสัมพันธ์ของความรู้และทำการเชื่อมโยงความรู้ตามรูปแบบของกราฟได้ จึงเป็นเหตุผลที่งานวิจัยนี้ได้ทำการเลือกใช้ภาษาสอบถามสปาร์เก้ิลสำหรับสอบถามความรู้ด้านพืชไร่

2. 2.4 งานวิจัยที่เกี่ยวข้องกับแนวคิดการค้นคืนสารสนเทศ

จากการศึกษางานวิจัยกับการนำเสนอแนวคิด เทคนิควิธีต่าง ๆ สำหรับการสืบค้นสารสนเทศมีหลากหลายเทคนิค เพื่อเพิ่มประสิทธิภาพในการค้นคืนสารสนเทศ ดังงานวิจัยของบังอร กลับบ้านเกาะ (2543) ที่ได้นำเสนอการใช้คำสำคัญเป็นหลักในการค้นหา ซึ่งคำสำคัญจะเป็นตัวชี้ไปยังเอกสารใด ๆ โดยประยุกต์ใช้ขั้นตอนวิธีจีเนติก (Genetic Algorithms) ภายใต้แบบจำลองปริภูมิเวกเตอร์ซึ่งแต่ละเอกสารแทนด้วยเวกเตอร์ของคำสำคัญ และข้อสอบถาม (Query) แทนด้วยเวกเตอร์ของคำที่ใช้ในข้อสอบถาม การใช้เวกเตอร์ในการสืบค้นทำได้โดยการจับคู่ระหว่างเอกสารกับข้อสอบถามแล้วทำการคำนวณหาความคล้ายคลึง ซึ่งกรณีที่ปรากฏคำ ณ ตำแหน่งใดในเอกสารจะให้ค่าเป็น 1 หากไม่ปรากฏให้ค่าเป็น 0 ซึ่งกระบวนการของจีเนติกจะนำข้อสอบถามแทนด้วยโครโมโซม เช่นเมื่อผู้ใช้ระบุข้อสอบถามเข้ามา(Q1) เพื่อค้นคืนเอกสาร(D1) และ (D2) ประกอบด้วยคำดังนี้

Q1: การวัดความคล้ายคลึงเชิงความหมาย

D1: การวัดความคล้ายคลึงเชิงความหมายระหว่างคำ

D1: ความคล้ายคลึงของเอกสารด้านพืชไร่

สามารถจัดเรียงเป็นชุดของคำสำคัญ คือ การ วัด ความ คล้ายคลึง เชิง ความหมาย ระหว่าง คำ ของ เอกสาร ด้าน พืช ไร่

นำเสนอในรูปแบบโครโมโซม ซึ่งความยาวของโคโมโซมจะขึ้นอยู่กับจำนวนคำสำคัญของชุดเอกสารทั้งหมดที่ตรงกับข้อสอบถาม ดังนี้

D1: 1111111100000 *D2:* 0011000011111

ดังนั้น โครโมโซม D1 และ D2 เป็นโครโมโซมชุดแรกเรียกว่าประชากรต้นกำเนิด และมีความยาวเท่ากับ 16 บิท จากนั้นนำโครโมโซมชุดแรกไปหาค่าความเหมาะสมโดยใช้ฟิตเนสฟังก์ชันของ ซัลตอน (Salton, 1989) ก่อนจะนำเข้าสู่กระบวนการจีเนติก ดังนี้ (1) คัดเลือกโคโมโซมที่มีค่าเหมาะสมที่สุดไว้ (Selection) (2) ใช้ค่าโครโมโซมใหม่สำหรับการคัดเลือกครั้งถัดไป

(Crossover) (3) นำโครโมโซมเก่ามาสุ่มแก้ไขบางส่วนของโครโมโซม (Mutation) โดยดำเนินการจนกระทั่งได้โครโมโซมที่เหมาะสมเพื่อนำไปค้นคืนสารสนเทศต่อไป สอดคล้องกับงานวิจัยของ ดิก ลุน ลีย์_ (Dik Lun Lee, 1997) ที่ได้อธิบายไว้ว่า เทคนิควิธีที่ใช้ในการค้นคืนสารสนเทศส่วนใหญ่จะอาศัยการใช้ดัชนีคำในการสืบค้นเพียงอย่างเดียว และเป็นเรื่องที่ไม่ดีนักที่ว่าการใช้เพียงดัชนีคำเพียงอย่างเดียวไม่สามารถที่จะสืบค้นสารสนเทศ โดยจับใจความในเนื้อหาของเอกสารได้ ส่งผลให้ประสิทธิภาพการสืบค้นต่ำ การแก้ไขคือการนำหลักการของ แบบจำลองปริภูมิเวกเตอร์ มาใช้เพื่อหาจุดสมดุลที่ดีระหว่างประสิทธิภาพในการประมวลผลกับประสิทธิภาพในการค้นคืนเอกสาร โดยแสดงรูปแบบในระดับของเอกสารที่สัมพันธ์กัน

ต่อมาเรมีย์ สเตต้า และคณะ (Raymie Stata, et. al., 2000) ได้สนับสนุนแนวคิดดังกล่าว โดยกำหนดองค์ประกอบของเวกเตอร์ (Vector Element) และระยะห่างเอกสาร (Document Space) คำนวณน้ำหนักคำโดยใช้ความถี่ที่ปรากฏในเอกสารบ่งบอกถึงความเกี่ยวข้องของคำกับเอกสาร และยังมีการสนับสนุนแนวคิดดังกล่าวจากการวิจัยของ วรวีร์ เกษร (2553) ที่ให้ค่าน้ำหนักคำโดยใช้ความถี่ที่ปรากฏในเอกสารมาใช้กับหลักการของแบบจำลองปริภูมิเวกเตอร์และใช้การวัดความคล้ายคลึงของเอกสารด้วยการประมวลผลภาษาธรรมชาติ ซึ่งการวัดความคล้ายคลึงใช้หลักการวัดค่าความคล้ายคลึงเชิงมุมด้วยโคซายน์ ผลการวิจัยที่กล่าวมาพบว่าระบบสืบค้นสารสนเทศมีค่าความแม่นยำและค่าความถูกต้องของข้อมูลที่ระดับค่าแตกต่างกัน ระบบสืบค้นสามารถทำงานได้ถูกต้อง สอดคล้องกับการวิจัยของ มานพ จงเจริญใจ (2541) ได้นำเสนอเทคนิควิธีการค้นคืนสารสนเทศประกอบด้วย 3 ขั้นตอน คือ (1) การสร้างดัชนี (2) การจัดลำดับการสืบค้น และ (3) การสืบค้นย้อนกลับ โดยใช้ความถี่ของคำที่ปรากฏในเอกสารกำหนดการสร้างดัชนี และการจัดลำดับการสืบค้นเพื่อให้ได้ค่าน้ำหนักคำรวมของแต่ละเอกสารแล้วนำผลจากการคำนวณน้ำหนักคำที่ได้มาทำการจัดลำดับผลลัพธ์ตามค่าน้ำหนักคำ และส่วนของการสืบค้นย้อนกลับ ผู้ใช้สามารถปรับปรุงข้อสอบถามใหม่เพื่อให้ข้อสอบถามนั้นสืบค้นสารสนเทศได้ตรงกับความต้องการ

จากการศึกษางานวิจัยที่เกี่ยวข้องสรุปได้ว่า แนวคิดการค้นคืนสารสนเทศจะใช้กระบวนการของการวิเคราะห์คำสำคัญ อันได้มาซึ่งคำค้น หรือตัวแทนที่จะบ่งบอกถึงเอกสารใด ๆ โดยมีการประยุกต์ใช้ขั้นตอนวิธีต่าง ๆ ภายใต้ของแนวคิดของแบบจำลองการค้นคืนสารสนเทศที่เหมาะสม เช่น การประยุกต์ใช้ขั้นตอนวิธีจีนีติก การให้น้ำหนักคำโดยใช้ความถี่ของคำที่ปรากฏในเอกสาร กับจำนวนของเอกสารทั้งหมดที่มีคำ ๆ นั้นปรากฏอยู่ และ ค่าสัดส่วนของจำนวนเอกสารทั้งหมดกับจำนวนของเอกสารที่ปรากฏคำที่สนใจ ภายใต้แบบจำลองปริภูมิเวกเตอร์ ดังนั้นการประยุกต์ใช้ความถี่ของคำที่ปรากฏในเอกสารจึงเป็นแนวทางใช้งานวิจัยครั้งนี้

2.3 แบบจำลองการสืบค้น

2.3.1 วิวัฒนาการของแบบจำลองการสืบค้น

ธนุสศักดิ์ ธัญญสิริ (2543) ได้อธิบายถึงวิวัฒนาการของแบบจำลองการสืบค้น แบ่งออกเป็น 2 ประเภทใหญ่ ๆ ได้แก่ การสืบค้นแบบดั้งเดิมที่อาศัยวิธีทางสถิติ และการค้นคืนสารสนเทศอัจฉริยะ ดังนี้

1. การสืบค้นเอกสารแบบดั้งเดิม (Traditional Retrieval) เป็นการสืบค้นที่ใช้วิธีทางสถิติโดยใช้ดัชนีในระดับคำ และวลี ซึ่งประกอบด้วย

1.1 การจับคู่แบบเทียบตรง (Exact Match) เป็นการจับคู่ที่ให้ผลลัพธ์เป็น 0 ในกรณีที่ไม่พบเอกสาร และ 1 ในกรณีที่พบเอกสาร วิธีการนี้เป็นการประยุกต์แบบจำลองตรรกะ (Boolean Model) ซึ่ง ซัลตอน (Salton, 1989) และซาร์เดส (Zadeh, 1993) ได้อธิบายไว้ว่าเป็นแบบจำลองการสืบค้นในยุคแรก ๆ ที่พัฒนาขึ้น ผลลัพธ์ที่ได้จากการสืบค้นมีเพียง 2 คำตอบ ได้แก่ จับคู่ได้ (Match) และจับคู่ไม่ได้ (No Match) เท่านั้น

แบบจำลองดังกล่าวข้างต้นมีข้อดีที่สามารถเพิ่มเติมดัชนีได้โดยไม่ต้องทำตัวแทนของเอกสาร แต่มีข้อเสีย ได้แก่การค้นหาช้า และผลลัพธ์ที่ได้มีมากหรือน้อยเกินไป ดังนั้นแบบจำลองนี้ไม่นิยมใช้ในปัจจุบัน

1.2 การจับคู่แบบไม่เทียบตรง (Inexact Match) วิธีการนี้เป็นการใช้แบบจำลองความน่าจะเป็น (Probabilistic Model) แบบจำลองปริภูมิเวกเตอร์ (Vector Space Model) และแบบจำลองลาเทนซีแมนติก (Latent Semantic Model) ดังนี้

1.2.1 แบบจำลองความน่าจะเป็น (Probabilistic Model) เป็นแบบจำลองที่ มารอน และคูนส์ (Maron and Kuhns, 1960) , ครอฟ และแฮร์เปอร์ (Croft and Harper, 1979) ได้อธิบายไว้ว่าแบบจำลองความน่าจะเป็นมีการพิจารณาความน่าจะเป็นของคำค้น X ในชุดของเอกสารที่สัมพันธ์กับคำค้น X และเอกสารที่ไม่สัมพันธ์กับคำค้น X โดยมีการนำค่าความน่าจะเป็นที่ได้มาคำนวณค่าถ่วงน้ำหนักเพื่อใช้สำหรับคำนวณค่าความคล้ายคลึงของแต่ละเอกสารและคำค้น ซึ่งความน่าจะเป็นของการเกิดความสัมพันธ์ของคำค้น แสดงดังตารางที่ 2.3

ตารางที่ 2.3 การคำนวณความน่าจะเป็นของความสัมพันธ์ระหว่างเอกสารกับคำค้น

การปรากฏของคำค้น X ในเอกสาร	จำนวนเอกสารที่สัมพันธ์กับคำค้น X	จำนวนเอกสารที่ไม่สัมพันธ์กับคำค้น X
ในกรณีที่ปรากฏ	*r*	*n-r*
ในกรณีที่ไม่ปรากฏ	*R-r*	(*N-R*)-(*n-r*)
ผลรวมของเอกสารที่ปรากฏ	*R*	*N-R*

เมื่อ

N คือ จำนวนเอกสารทั้งหมด

R คือ จำนวนของเอกสารทั้งหมดที่สัมพันธ์กับคำค้น X

n คือ จำนวนเอกสารทั้งหมดที่สืบค้นได้

r คือ จำนวนเอกสารที่สืบค้นได้ และสัมพันธ์กับคำค้น X

โดยวิธีการนี้จะนำมาคำนวณหาค่าถ่วงน้ำหนักของ โจนส์(Jones, 1990) ดังสมการที่ 2.1

$$w_{ij} = tf_{ij} \cdot tr_j \quad (2.1)$$

เมื่อ

w_{ij} คือ ค่าถ่วงน้ำหนักของคำค้น j ในเอกสาร i

tf_{ij} คือ ความถี่ของคำค้น j ในเอกสาร i

tr_j คือ ค่าความน่าจะเป็นของการเกิดคำค้น j

จากตารางที่ 2.1 โจนส์(Jones, 1990) อธิบายไว้ว่าค่า tr_j ควรเป็นดังสมการที่ 2.2

$$tr_j = \log \frac{\left(\frac{r}{R-r}\right)}{\left(\frac{n-r}{(N-n-R+r)}\right)} \quad (2.2)$$

ข้อดีของแบบจำลองนี้ คือสามารถวัดค่าความคล้ายคลึงระหว่างคำค้นกับเอกสารได้ แม้ว่าเอกสารนั้นไม่มีคำค้นปรากฏอยู่ทุกคำ แต่มีข้อเสียไม่สามารถวิเคราะห์ความหมายของคำรวมถึงคำที่มีความหมายใกล้เคียงกันจากการระบุคำสำคัญของผู้ใช้ได้ เนื่องจากการกำหนดศัพท์ดัชนีเป็นอิสระต่อกันไม่มีความสัมพันธ์กัน ดังนั้นผู้ใช้ต้องใช้คำค้นที่เหมาะสมสำหรับการค้นคืนสารสนเทศ

1.2.2 แบบจำลองปริภูมิเวกเตอร์(Vector Space Model) เป็นแบบจำลองที่รากาวาน และวอง (Raghavan and Wong, 1986) ได้อธิบายหลักการทำงานไว้ว่าให้แทนเอกสาร (Document) และคำ (Term) ที่ผู้ใช้สอบถาม (User Query) เข้ามาด้วย เวกเตอร์ จากนั้นทำการเปรียบเทียบหาเอกสารที่มี เวกเตอร์ คล้ายคลึงกับคำค้นของผู้ใช้ให้มากที่สุด ซึ่งซัลตอน (Salton , 1989) และลีย์ Lee (1995) ได้อธิบายว่าการแทนคำที่ปรากฏในเอกสารมีการแทนคำให้อยู่ในรูปแบบของเวกเตอร์ที่มีการให้ค่าน้ำหนักของคำ (Term) โดยใช้ความถี่ของคำที่สนใจที่ปรากฏในเอกสาร (Term Frequency: TF) และค่าสัดส่วนของจำนวนเอกสารทั้งหมดกับจำนวนของเอกสารที่ปรากฏ

คำที่สนใจ (Inverse Document Frequency: IDF) หรือเรียกว่า TF-IDF ซึ่งสามารถแสดงตัวอย่างการแทนเอกสารและคำที่ผู้ใช้สอบถามด้วยหลักการแบบจำลองปริภูมิเวกเตอร์ดังนี้

ตัวอย่างเช่น ในชุดเอกสารประกอบด้วยเอกสาร *D1, D2* ซึ่งในแต่ละเอกสารประกอบด้วยคำ ดังนี้

D1 : การวัดความคล้ายคลึงเชิงความหมายระหว่างคำ

ผลการตัดคำ D1 : การ|วัด|ความ|คล้ายคลึง|เชิง|ความหมาย|ระหว่าง|คำ

D2 : ความคล้ายคลึงของเอกสารด้านพืชไร่

ผลการตัดคำ D2 : ความ|คล้ายคลึง|ของ|เอกสาร|ด้าน|พืช|ไร่

จากนั้น นำคำที่ปรากฏในแต่ละเอกสาร D1, D2 มาทำการคำนวณความถี่ของคำดังนี้

ความถี่ของคำที่ปรากฏเอกสาร *D1*

การ	วัด	ความ	คล้ายคลึง	เชิง	ความหมาย	ระหว่าง	คำ
1	1	1	1	1	1	1	1

ความถี่ของคำที่ปรากฏในเอกสาร *D2*

ความ	คล้ายคลึง	ของ	เอกสาร	ด้าน	พืช	ไร่
1	1	1	1	1	1	1

โดยมีคำศัพท์ในระบบคือ – การ – วัด – ความ – คล้าย – คลึง – เชิง – ความหมาย – ระหว่าง – คำ – ของ – เอกสาร - ด้าน – พืช - ไร่

จากนั้นก็จะนำเอกสารทั้ง D1 และ D2 มาทำเป็นเวกเตอร์

ค่าของเอกสาร D1 เมื่อนำมาทำเป็นเวกเตอร์

การ	วัด	ความ	คล้าย คลึง	เชิง	ความ หมาย	ระหว่าง	คำ	ของ	เอกสาร	ด้าน	พืช	ไร่
1	1	1	1	1	1	1	1	0	0	0	0	0

เวกเตอร์ของเอกสาร D1 : (1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0)

ค่าของเอกสาร D2 เมื่อนำมาทำเป็นเวกเตอร์

การ	วัด	ความ	คล้าย คลึง	เชิง	ความ หมาย	ระหว่าง	คำ	ของ	เอกสาร	ด้าน	พืช	ไร่
0	0	1	1	0	0	0	0	1	1	1	1	1

เวกเตอร์ของเอกสาร D2: (0, 0, 1, 1, 0, 0, 0, 0, 1, 1, 1, 1, 1)

เมื่อผู้ใช้ระบุข้อสอบถามเข้ามาในระบบเพื่อค้นคืนสารสนเทศระบบจะนำข้อสอบถามนั้นมาทำเป็นเวกเตอร์ เช่นเดียวกับเอกสาร D1 และ D2 แล้วจึงนำเวกเตอร์ที่ได้มาทำการเปรียบเทียบกันว่าเอกสารใดมีความคล้ายคลึงกับข้อสอบถาม

- สมมุติ ผู้ใช้ระบุข้อสอบถาม Q1 เข้ามาเป็น "การวัดความคล้ายคลึงเชิงความหมาย"

ผลการตัดคำ: การ |วัด|ความ|คล้ายคลึง|เชิง|ความหมาย

ค่าของข้อสอบถาม Q1 เมื่อนำมาทำเป็นเวกเตอร์

การ	วัด	ความ	คล้าย คลึง	เชิง	ความ หมาย	ระหว่าง	คำ	ของ	เอกสาร	ด้าน	พืช	ไร่
1	1	1	1	1	1	0	0	0	0	0	0	0

เวกเตอร์ของข้อสอบถาม Q1: (1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0)

จากนั้นนำเวกเตอร์ของข้อสอบถามจากผู้ใช้ Q1 เปรียบเทียบกับเวกเตอร์ของเอกสาร D1 และ D2 โดยการคำนวณผลคูณภายในของทั้ง 2 เวกเตอร์ ซึ่งทำให้ทราบว่าข้อสอบถามที่ระบุเข้ามา Q1 จะตรงกับเอกสารใดมากที่สุด ดังนี้

เวกเตอร์ Q1 (1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0)

เวกเตอร์ D1 (1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0)

Q1 * D1 = (1*1) + (1*1) + (1*1) + (1*1) + (1*1) + (1*1) + (0*1) + (0*1) + (0*0) + (0*0) +
(0* 0) + (0*0) + (0*0)

ดังนั้น Q1*D1 = 6

เวกเตอร์ Q1 (1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0)

เวกเตอร์ D2 (0, 0, 1, 1, 0, 0, 0, 0, 1, 1, 1, 1, 1)

Q1 * D2 = (1*0) + (1*0) + (1*1) + (1*1) + (1*0) + (1*0) + (0*0) + (0*0) + (0*1) + (0*1) +

(0* 1) + (0*1) + (0*1)

ดังนั้น Q1*D2 = 2

สรุปได้ว่าข้อสอบถาม Q1 จะตรงกับเอกสาร D1 มากกว่าเอกสาร D2 เนื่องจากผลคูณภายในมีค่าสูงกว่า

ข้อดีของแบบจำลองนี้ คือเป็นการจับคู่แบบวัดความคล้ายคลึง เช่นเดียวกับแบบจำลองทฤษฎีความน่าจะเป็นทำให้สามารถสืบค้นเอกสารได้ถึงแม้ว่าเอกสารนั้นจะไม่มีคำค้นปรากฏอยู่ครบทุกคำ แต่มีข้อเสียในการใช้ความถี่ของคำที่ปรากฏในเอกสารเพียงอย่างเดียวไม่มีการคำนึงถึงความหมายของคำ โดยจะถือว่าแต่ละคำเป็นอิสระต่อกัน ซึ่งความเป็นจริงแล้วคำแต่ละคำอาจมีความหมายเดียวกัน หรือใกล้เคียงกัน

1.2.3 แบบจำลองลาเทนซีแมนติก (Latent Semantic Model: LSI) เป็นแบบจำลองที่มีการพิจารณาถึงความสัมพันธ์ของคำนั้น ๆ กับเอกสารใด ๆ ในเชิงความหมายด้วยการคำนวณทางสถิติ โดยมีการกำหนดคำภายในเอกสารว่าคำใด ๆ มีความหมายเหมือนกัน เช่น กรณีที่ใช้คำสำคัญ “อ้อย” แบบจำลองนี้สามารถประมวลผลได้ว่า อ้อยที่เป็นพืชไร่ หรือ พืชให้น้ำตาล ชื่อคน หรือชื่อสถานที่ได้อย่างถูกต้องถึงแม้ว่าเอกสารนั้น ๆ จะไม่ปรากฏคำที่ต้องการสืบค้นก็ตาม โดยแบบจำลองนี้ได้เพิ่มขีดความสามารถให้กับการสร้างคำค้นเมตริกของเอกสารที่นอกเหนือจากพิจารณาความถี่ของคำที่ปรากฏอยู่ในเอกสารเพียงอย่างเดียว

ข้อดีของแบบจำลองนี้ คือสามารถสืบค้นเชิงความหมายโดยพิจารณาจากความสัมพันธ์ของคำกับเอกสารได้ และเหมาะสำหรับการค้นหาประเภทเอกสารที่เหมือนกันแต่อยู่ในหลายภาษา (Multi-Language) แต่มีข้อเสียในการค้นคืนกับภาษาไทย ที่ยังไม่สามารถแก้ไขปัญหาความกำกวมของคำศัพท์ภาษาไทยได้ เนื่องจากภาษาไทยเป็นภาษาที่ไม่มีช่องว่างระหว่างคำหรือการเว้นวรรคระหว่างคำเหมือนภาษาอังกฤษ ซึ่งสิทธิโชค ปัญญาฤกษ์ชัย และศิพาณี นุชิตประสิทธิ์ชัย (2009) และไกรศักดิ์ เกษร (2013) ได้กล่าวไว้ว่างานต่าง ๆ สำหรับการประมวลผลภาษาไทยนั้นต้องมีกระบวนการตัดคำเป็นพื้นฐาน (Word Segmentation)

2. การค้นคืนสารสนเทศอัจฉริยะ (Intelligent Information Retrieval) เป็นการสืบค้นที่อาศัยเทคนิคการประมวลผลภาษาธรรมชาติประยุกต์ใช้ร่วมกับฐานความรู้ ซึ่งการประมวลผลภาษาธรรมชาติ (Natural Language Processing) เป็นการวิเคราะห์และประมวลผลภาษาของมนุษย์ให้คอมพิวเตอร์เข้าใจ โดยผลการวิจัยของ เฮิร์ท และเทอร์เทล (Hersh and Turtle, 1989)

พบว่าการสืบค้นโดยใช้การประมวลผลภาษาธรรมชาติให้ประสิทธิภาพสำหรับการสืบค้นได้ดีกว่าการสืบค้นแบบทั่วไป และพื้นฐานการประมวลผลทางภาษามีกระบวนการเริ่มจากการตัดคำจากข้อความ (Tokenization) และการแก้ไขคำค้นที่สะกดผิด (Word Approximation) ปัจจุบันการประมวลผลภาษาธรรมชาติได้ถูกนำมาใช้ในงานด้านต่าง ๆ เช่น การค้นคืนข้อมูล (Information Retrieval) การแปลภาษาด้วยเครื่อง (Machine Translation) และงานวิจัยของ อรสา ขันเอียดและคณะ (2552) ที่ได้ประยุกต์ใช้กับการประมวลผลประโยคด้วยคอมพิวเตอร์ที่พิจารณาเลือกสูตรไวยกรณ์ และเทคนิคการแจงประโยค เป็นต้น

ข้อดีของการสืบค้นแบบนี้ คือผู้ใช้งานไม่จำเป็นต้องมีความรู้ในศาสตร์ความรู้ที่ต้องการสืบค้นข้อมูลมากนัก แต่มีข้อเสียในการพิจารณาถึงความหมายคำ เพราะการให้ผลลัพธ์ตรงกับความต้องการของผู้ใช้นั้นควรมีการพิจารณาถึง ความหมายของคำที่โยงไปถึงความสัมพันธ์ของกลุ่มคำที่เกี่ยวข้องกับคำที่ผู้ใช้ต้องการค้นคืน

สรุปได้ว่า แบบจำลองการสืบค้นแต่ละวิธีมีข้อดี และข้อเสียแตกต่างกันออกไป ซึ่งการประยุกต์ใช้ควรคำนึงบริบทในความรู้นั้น ๆ เช่นภาษา ประเภทความรู้ และแหล่งข้อมูล เป็นต้น และที่สำคัญให้มีการ พิจารณาเกี่ยวกับความหมายของคำรวมถึงคำที่มีความหมายใกล้เคียงกันจากการระบุคำสำคัญของผู้ใช้ อันจะทำให้ผลลัพธ์ตรงกับความต้องการ ซึ่งเป็นแนวคิดที่จะถูกนำมาใช้สำหรับการวิจัยในครั้งนี้

2.3.2 แนวคิดการสืบค้นเชิงความหมาย

การพัฒนาระบบสืบค้นเชิงความหมาย (Semantic Search) เป็นกระบวนการ ค้นหาคำหรือกลุ่มคำที่เกี่ยวข้อง กับความรู้หรือเอกสาร ซึ่งยังใช้แนวคิดจากการค้นคืนสารสนเทศ (Information Retrieval) แต่การสืบค้นเชิงความหมายได้ปรับเปลี่ยนเทคนิควิธีสำหรับการวิเคราะห์ความหมายของคำรวมถึงคำที่มีความหมายใกล้เคียงกันจากการระบุคำสำคัญของผู้ใช้ เพื่อให้ผลลัพธ์ที่ได้ตรงกับความต้องการของผู้ใช้ เช่น การปรับเปลี่ยนวิธีการทำดัชนีที่สามารถสนับสนุนการสืบค้นเชิงความหมายได้แก่ วิธีการของแบบจำลองลาเทนซีแมนติก และวิธีการออนโทโลยี (ไกรศักดิ์ เกสร, 2554) ดังนี้

1) การทำดัชนีโดยใช้แบบจำลองลาเทนซีแมนติกเป็นวิธีการที่มีหลักการทำงานอยู่บนพื้นฐานของการคำนวณทางสถิติ โดยพิจารณาจากการปรากฏร่วมของคำต่าง ๆ (Co-Occurrence) เป็นเทคนิคการทำดัชนีแบบใหม่นำเสนอโดย เดียร์เวสเตอร์ (Deerwester, 1990) โดยใช้หลักการความสัมพันธ์ของคำสำคัญและเอกสารมาใช้ในการทำดัชนี นอกจากจะพิจารณาความถี่ของคำที่ปรากฏในเอกสารหนึ่งแล้ว ยังจะพิจารณาถึงความสัมพันธ์ของคำนั้น ๆ กับเอกสารอื่น ๆ ในคอลเล็คชั่นเรียกว่า เชิงความสัมพันธ์ (Interrelationship) วิธีการนี้จะทำการพิจารณาเอกสารใด ๆ

ที่มีจำนวนคำสำคัญปรากฏร่วมกันสูงถือว่าเอกสารเหล่านั้นมีความสัมพันธ์กันเชิงความหมาย (Semantically Related) และเอกสารใด ๆ ที่มีคำสำคัญปรากฏร่วมกันอยู่จำนวนน้อยจะถือว่ามีความสัมพันธ์กันเชิงความหมายต่ำ เนื่องจากว่าเอกสารใด ๆ อาจจะมีความสัมพันธ์กันเชิงความหมายถึงแม้ว่าเอกสารนั้น ๆ จะไม่มีคำสำคัญที่ซ้ำกันเลย ซึ่งวิธีนี้จะสามารถหาความสัมพันธ์ของเอกสารเหล่านั้นได้จากการคำนวณทางสถิติ ดังนั้นวิธีการนี้จึงหาคำตอบของคำค้นที่มีประสิทธิภาพดีกว่าวิธีการหาเอกสารแบบเปรียบเทียบสายอักขระถึงแม้ว่าเอกสารนั้น ๆ จะไม่ปรากฏคำสำคัญในคำค้นเลยก็ตาม ดังตัวอย่างเช่นต้องการทราบว่าเอกสารที่เกี่ยวกับยี่ห้อรถยนต์รุ่นบีเอ็มดับเบิ้ลยู (BMW) โตโยต้า(Toyota) และ ฮอนด้า (Honda) เป็นเอกสารที่ควรจะเป็นคำตอบของคิวรีรถยนต์ แม้ว่าเอกสารเหล่านั้นจะไม่มีคำว่า “รถยนต์” ปรากฏอยู่

2) การทำดัชนีโดยใช้ออนโทโลยีเป็นแนวคิดการจัดฐานความรู้ให้อยู่ในรูปแบบโครงสร้างลำดับชั้น โดยประกอบด้วยแนวคิดต่าง ๆ และความสัมพันธ์ของคอนเซ็ปต์หรือแนวคิด ซึ่งซันดาร์เซการัน , โจส์เซฟชัน และเบนจามินส์ (Chandrasekaran, Josephson and Benjamins, 1999) ได้อธิบายว่าความสัมพันธ์ของคอนเซ็ปต์ต่าง ๆ ในออนโทโลยีเป็นตัวกำหนดความหมายที่แท้จริงของคำสำคัญต่างๆ โดยจะช่วยแก้ปัญหาความกำกวมของคำสำคัญ เช่น หนึ่งคำมีหลายความหมาย หรือคำหลาย ๆ คำมีความหมายเดียวกัน และ แกสสิวิค , ดูจ์ริค และเดเวดซิค (Gasevic, Djuric and Devedzic, 2009) ได้อธิบายเพิ่มว่าออนโทโลยีสามารถใช้เป็นตัวกำหนดลำดับชั้นของแนวคิดต่าง ๆ ภายในโดเมนที่สนใจมีการเก็บข้อมูลได้หลายรูปแบบ และเป็นที่รู้จักคืออาร์ดีเอฟและโอดับเบิ้ลยูแอล ซึ่งใช้ภาษาที่ใช้ในการสอบถามความรู้ คือ ภาษาสอบถามสปาร์เก็ล สอดคล้องกับ อรวรรณ อิ่มสมบัติ และคณะ (2552) ได้อธิบายว่า ออนโทโลยีมีความจำเป็นต่อการประมวลผลเชิงความหมายโดยเฉพาะการสืบค้นและการสกัดข้อมูล ซึ่งสามารถช่วยให้การทำงานของระบบประมวลผลเชิงความหมายทำงานได้อย่างมีประสิทธิภาพ

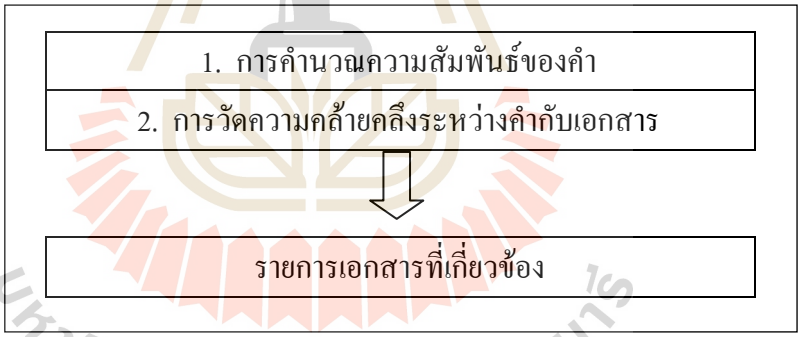
ต่อมาศุภกฤษฎิ์ นิวัฒนากูล (2556) ได้อธิบายแนวคิดการสืบค้นเชิงความหมายไว้ว่าการค้นหาความรู้หรือเอกสารที่เกี่ยวข้องจะใช้คำศัพท์จากผู้ใช้เพียงอย่างเดียวอาจจะทำให้การสืบค้นได้เอกสารความรู้ไม่ครบถ้วน เพราะถ้าหากเอกสารที่เกี่ยวข้องไม่ได้มีคำสำคัญตามที่ผู้ใช้กำหนด จะทำให้ไม่สามารถค้นคืนเอกสารความรู้นั้นได้ ดังนั้น วิธีการสืบค้นควรมีการวิเคราะห์ถึงความหมายของคำ และความหมายที่ใกล้เคียงกับคำค้น ให้สามารถเข้าถึงเอกสารที่เกี่ยวข้องได้ตรงกับความต้องการของผู้ใช้ สอดคล้องกับการอธิบายของพิลาวัณย์ พลับรู้การ และ กฤษณะ ไวยมัย (2544) ที่ว่า คำแต่ละคำเป็นอิสระต่อกัน ซึ่งในความเป็นจริงแล้วคำแต่ละคำอาจมีความใกล้เคียงทางความหมายต่อกัน

สรุป แนวคิดการสืบค้นเชิงความหมายเป็นการนำเสนอเทคนิควิธีการที่ให้ระบบค้นคืนสารสนเทศมีความสามารถวิเคราะห์คำจากความหมาย รวมไปถึงคำที่ใกล้เคียงอันส่งผลให้

ได้ผลลัพธ์ตรงกับความต้องการ ซึ่งวิธีการ สร้างดัชนีที่สนับสนุนการสืบค้นนี้ ได้แก่ Latent Semantic Indexing (LSI) และออนโทโลยี (Ontology) สำหรับงานวิจัยนี้สร้างดัชนีด้วยออนโทโลยี เพราะการจัดฐานความรู้ด้านพืชไร่มีรูปแบบโครงสร้างลำดับชั้น และใช้ภาษาที่รองรับสำหรับการสืบค้น คือ ภาษาสอบถามสปาร์เกิ้ล

2.3.3 รูปแบบการสืบค้นเชิงความหมาย

ปัจจุบันการสืบค้นเชิงความหมาย เป็นแนวคิดที่คิดค้นขึ้นมาเพื่อแก้ปัญหาให้ระบบสืบค้นมีความสามารถวิเคราะห์ความหมายของคำ อันจะส่งผลต่อประสิทธิภาพในการสืบค้น ซึ่งมีนักวิจัยจำนวนมากที่ให้ความสำคัญกับการแก้ปัญหาดังกล่าวดังการวิจัยของ คิโมโต และ อิวาดีรา (Kimoto and Iwadera, 1990) ได้เสนอเทคนิควิธีการแยกคำที่ผู้ใช้สนใจจากเอกสารมาสร้างเป็นดัชนี (Index) และสร้างความสัมพันธ์ภายในเอกสาร (Co- Occurrence) พิจารณาจากคำสำคัญ กว้างกว่า (Broader) คำแคบกว่า (Narrower) คำเกี่ยวข้อง (Related Term-RT) ซึ่งพิลาวัณย์ พลับรู้การ และ กฤษณะ ไวยมัย (2544) ได้เสนอเทคนิคการวัดความคล้ายคลึงระหว่างเอกสาร ด้วยการหาค่าความเกี่ยวข้องระหว่างคำสองคำจากระยะทางน้อยที่สุดของคำ ซึ่งมี 2 ขั้นตอน แสดงดังรูปที่ 2.12



ที่มา : พิลาวัณย์ พลับรู้การ และ กฤษณะ ไวยมัย (2544)

รูปที่ 2.12 การสืบค้นเชิงความหมายจากการวัดความคล้ายคลึงระหว่างเอกสาร

จากรูปที่ 2.12 แสดงขั้นตอนของการวัดความคล้ายคลึงระหว่างเอกสารโดยใช้แนวทางเชิงความหมาย ดังนี้

1. การคำนวณความสัมพันธ์ของคำ ซึ่งเป็นการหาค่าความเกี่ยวข้องระหว่างคำสองคำจากระยะห่างน้อยที่สุด ค่าความเกี่ยวข้องระหว่างคำเป็นสัดส่วนผกผันกับระยะทางที่สั้นที่สุดที่เชื่อมถึงกันในโครงข่ายอภิธานคำศัพท์ หากค่าระยะทางที่สั้นที่สุดระหว่างสองคำมีค่าน้อยจะถือว่าคำสองคำนั้นมีความเกี่ยวพันกันทางความหมายค่อนข้างมาก ในทางตรงกันข้าม หาก

ระยะทางที่สั้นที่สุดระหว่างคำสองคำมีค่ามาก หมายถึงคำทั้งสองมีความเกี่ยวพันกันทางความหมายน้อย การเปลี่ยนค่าระยะห่างที่น้อยที่สุดระหว่างคำเป็นค่าความเกี่ยวข้องจะอยู่ในรูปสมการคณิตศาสตร์ที่แปรผกผันกัน งานวิจัยที่ศึกษาการวัดความคล้ายคลึงทางความหมายระหว่างคำในโครงข่ายโดยพิจารณาระยะทาง (Edge Based) ดังสมการที่ 2.3

$$y(x) = 0.9-\log(x) \quad (2.3)$$

โดยที่ x เป็นระยะทางที่สั้นที่สุดระหว่างสองคำและ y(x) เป็นการคำนวณความคล้ายคลึงระหว่างคำสองคำ

2. การวัดความคล้ายคลึงระหว่างคำกับเอกสาร เป็นการนำค่าจากการคำนวณความคล้ายคลึงระหว่างคำสองคำมาทำการคำนวณความคล้ายคลึงระหว่างคำกับเอกสาร โดยใช้วิธีการวัดความคล้ายคลึงเชิงมุม (Cosine Similarity) ซึ่งมีการกำหนดให้เอกสารใดๆ ประกอบด้วยเซตของคำนำมาเปรียบเทียบกัน เช่น เซตคำ $T = \{t_1, t_2,\ldots, t_n\}$ เปรียบเทียบกับเอกสาร d_i ซึ่งการคำนวณความคล้ายคลึงของระหว่างเอกสารกับคำค้น ดังสมการที่ 2.4

$$S = \sum_{k=1}^{p} s_k / \sqrt{(n_D * n_i)} \quad (2.4)$$

เมื่อ

S คือ ค่าความคล้ายคลึงระหว่างเอกสาร D และ d_i

p คือ จำนวนกลุ่มย่อยที่มีความคล้ายคลึงร่วมกันระหว่าง D และ d_i

s_k คือ ค่าความคล้ายคลึงในกลุ่มย่อยที่ k

n_D คือ จำนวนกลุ่มที่แตกต่างกันของคำในเอกสาร D

n_i คือ จำนวนกลุ่มที่แตกต่างกันของคำในเอกสาร d_i

ต่อมาสุภกฤษฎิ์ นิวัฒนากูล (2556) ได้นำเสนอแบบจำลองสำหรับการสืบค้นเชิงความหมายโดยการประยุกต์ใช้ฐานความรู้ที่เป็นออนโทโลยีสำหรับอธิบายความหมาย ความสัมพันธ์ของคำ และโครงสร้างของข้อมูลผ่านภาษาโอดับเบิ้ลยูแอล (Web Ontology Language: OWL) และใช้เทคนิควิธีการวัดความคล้ายคลึงระหว่างคำด้วยสัมประสิทธิ์ความคล้ายคลึงของแจ็คการ์ด สำหรับเปรียบเทียบความคล้ายคลึงระหว่างชุดของข้อมูล และนำเสนอขั้นตอนวิธีค้นหาเชิงความหมายประกอบด้วย2 ส่วน ดังนี้

1. การจัดการฐานความรู้ ที่ใช้รูปแบบของออนโทโลยีเพื่อใช้จัดเก็บ อธิบายความรู้ และการจัดทำคลังคำหรือดัชนีคำสำคัญเพื่อใช้สำหรับช่วยในการสืบค้นความรู้ ดังนี้

1.1 ฐานความรู้ เป็นการสร้างฐานความรู้ที่สามารถอธิบายรูปแบบโครงสร้างความสัมพันธ์ระหว่างข้อมูลความสัมพันธ์ของคำค้น (Term) จากอรรถาภิธานศัพท์เกษตรไทย

1.2 คลังคำ เป็นแหล่งรวบรวมคำต่าง ๆ โดยคำศัพท์แต่ละคำมีการจัดความสัมพันธ์ในลักษณะลำดับชั้น (Hierarchical Relation) ที่มีความหมายเหมือนกัน แต่เขียนต่างกัน มีการประยุกต์ใช้ตัวแบบเอสเคโอเอส (Simple Knowledge Organization System: SKOS) สำหรับสร้างคลังคำ ซึ่งมีความสัมพันธ์ของคำดังการอธิบายของ ฟุลวิโอ และลอร่า (Fulvio and Laura, 2005) คือ คำสำคัญ (skos: preLabel) คำเหมือน (skos: altLabel) คำกว้างกว่า (skos: broader) คำแคบกว่า (skos: narrower) คำเกี่ยวข้อง (skos: related)

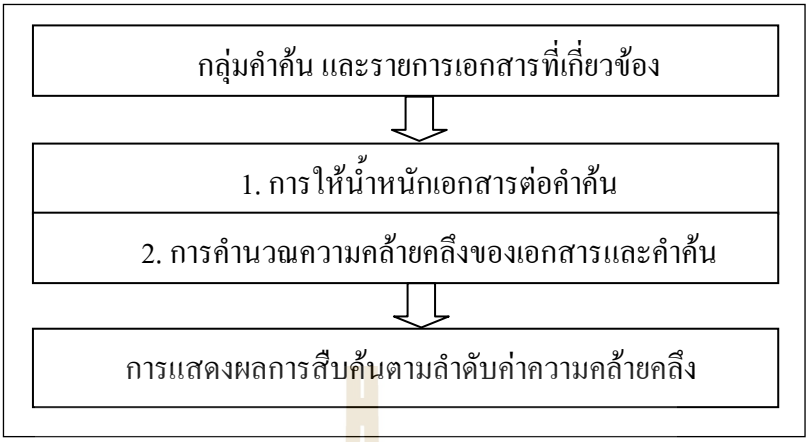
2. การสืบค้นเชิงความหมาย ประกอบด้วย 3 กระบวนการ ดังนี้

2.1 การวิเคราะห์คำสำคัญ เป็นส่วนของการวิเคราะห์คำสำคัญที่มีการวัดความคล้ายคลึงของคำค้น แก้ไขคำค้นที่สะกดผิดจากผู้ใช้

2.2 การสืบค้นเชิงความหมาย เป็นส่วนของการใช้คลังคำช่วยในการกำหนดความสำคัญของคำสำคัญสำหรับการสืบค้น ซึ่งมี 3 ขั้นตอนแสดงดังรูปที่ 2.13 ดังนี้

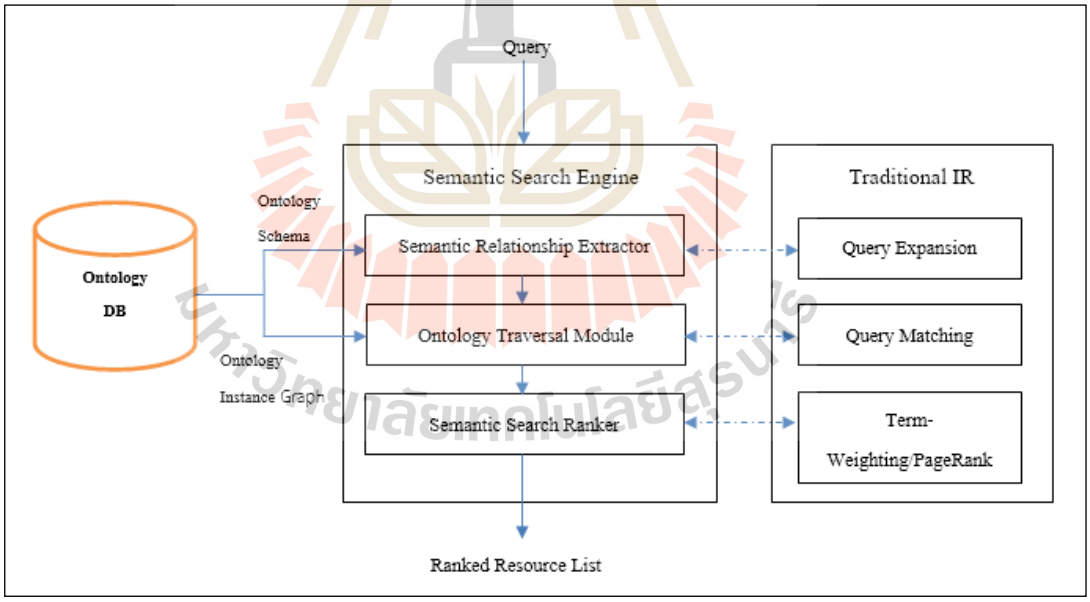
รูปที่ 2.13 ขั้นตอนการสืบค้นเชิงความหมาย

2.3 การแสดงผลลัพธ์ เป็นส่วนของการนำผลลัพธ์ที่ได้จากการสืบค้นความรู้มาจัดเรียงลำดับผลลัพธ์ ซึ่งมี 2 ขั้นตอน แสดงดังรูปที่ 2.14 ดังนี้



รูปที่ 2.14 ขั้นตอนส่วนแสดงผลลัพธ์การสืบค้น

นอกจากนี้ ลีย์, มิน และโฮ (Lee, Min and Oh, 2014) ได้นำเสนอแบบจำลองการสืบค้นเชิงความหมายที่มีการประยุกต์ใช้กับออนโทโลยี แสดงดังรูปที่ 2.15 ดังนี้



รูปที่ 2.15 วิธีการสืบค้นเชิงความหมายของ ลีย์, มิน และโฮ (Lee, Min and Oh, 2014)

จากรูปที่ 2.15 แสดงให้เห็นว่าการค้นคืนแบบดั้งเดิม (Traditional Information Retrieval) เป็นการค้นคืนในลักษณะของการจับคู่แบบเทียบตรงกับข้อสอบถาม (Query Matching) โดยขยายขอบเขตของคำ (Query Expansion) ซึ่งมีการให้ค่าน้ำหนักของคำ (Term) โดยใช้ความถี่ของคำที่สนใจที่ปรากฏในเอกสารเท่านั้น แต่สำหรับวิธีการสืบค้นแบบใหม่ที่นำเสนอ แบบจำลองการสืบค้นเชิงความหมายจะประกอบด้วย4 ส่วน ได้แก่ (1) ฐานความรู้ที่จัดเก็บในรูปแบบออนโทโลยี (Ontology Database) ที่เป็นรูปแบบโครงสร้างความสัมพันธ์ระหว่างข้อมูลและความสัมพันธ์ของคำในลักษณะ คำสำคัญ คำเหมือน คำกว้างกว่า คำแคบกว่า และคำเกี่ยวข้อง (2) การขยายขอบเขตของคำสำคัญ (Semantic Relationship Extractor) ซึ่งอาจมีกระบวนการวิเคราะห์ลำดับคำสำคัญ จัดกลุ่มคำที่เกี่ยวข้อง และเป็นส่วนของการแก้ไขคำค้นที่สะกดผิดจากผู้ใช้ระบุคำสำคัญ (Query) เข้ามา (3) การสืบค้นเอกสารจากฐานความรู้ที่ตรงกับคำสำคัญ (Ontology Traversal Module) ซึ่งเป็นส่วนของการวัดความคล้ายคลึงระหว่างคำกับเอกสารในเชิงความหมาย และ(4) การจัดลำดับผลลัพธ์ในเชิงความหมาย (Semantic Search Ranker) ซึ่งการจัดลำดับมีการเรียงผลลัพธ์ตามลำดับค่าความคล้ายคลึงกันระหว่างเอกสารกับคำสำคัญ

จากการศึกษาแนวคิดทฤษฎีที่เกี่ยวข้องกับการสืบค้นเชิงความหมายสรุปได้ว่า รูปแบบการสืบค้นเชิงความหมายเป็นแนวคิดที่อยู่บนพื้นฐานการสืบค้นความรู้ที่เก็บอยู่ในรูปแบบออนโทโลยี ซึ่งมีองค์ประกอบที่สำคัญคือ การจัดความรู้ให้เป็นระบบ และขั้นตอนวิธีสำหรับการสืบค้นเชิงความหมาย ดังนั้น จึงเป็นแนวคิดสำหรับงานวิจัยนี้ที่จะนำเสนอ 1 กระบวนการจัดความรู้ให้เป็นระบบ และเสนอขั้นตอนวิธีสำหรับการสืบค้นเชิงความหมาย ได้แก่ (1) การกำหนดลำดับความสำคัญคำค้น (2) การให้น้ำหนักคำสำคัญ และ (3) การวัดความคล้ายคลึง โดยจัดลำดับผลลัพธ์ตามผลการเปรียบเทียบความคล้ายคลึงของเอกสาร เพราะทั้ง 3 ขั้นตอนเป็นกระบวนการที่สำคัญสำหรับการสืบค้นเชิงความหมาย อันจะส่งให้ได้ผลลัพธ์ตรงกับความต้องการของผู้ใช้

2.3.4 งานวิจัยที่เกี่ยวข้องกับแนวคิดการสืบค้นเชิงความหมาย

จากการศึกษางานวิจัยกับการนำเสนอแนวคิด เทคนิควิธีต่าง ๆ สำหรับการสืบค้นเชิงความหมายมีหลากหลายเทคนิค เพื่อ แก้ปัญหาของคำสำคัญที่มีความหมายเหมือนกัน หรือคำใกล้เคียงกัน ดังงานวิจัยของ นคร โคตรโสภา (2549) ได้เสนอแนวคิดการสืบค้นเอกสารเชิงความหมายบนข้อมูลภาษาเอ็กซ์เอ็มแอล (Extensible Markup Language: XML) ประกอบด้วย 3 ขั้นตอน คือ (1) การทำดัชนีคำสำคัญ ซึ่งขั้นตอนนี้ได้ทำการแบ่งเอกสารเอ็กซ์เอ็มแอลเป็นกลุ่มเพื่อสร้างความสัมพันธ์ของเอกสารตามคำสำคัญโดยกำหนดระดับความสำคัญและให้น้ำหนักคำจากความหมายของคำสำคัญ เพื่อหาระดับความสัมพันธ์ระหว่างกลุ่มเอกสารกับคำสำคัญ (2) การค้นคืนสารสนเทศ และ (3) การจัดลำดับผลลัพธ์ ซึ่งขั้นตอนนี้ใช้หลักการวัดความคล้ายคลึงเชิง

ความหมายของข้อมูลคำสำคัญในเอกสารกับความหมายของข้อสอบถามที่ใช้สืบค้น เพื่อจัดเรียงลำดับเอกสารตามค่าความคล้ายคลึง ผลการวิจัยด้วยวิธีการที่นำเสนอทำให้สามารถสืบค้นสารสนเทศได้ตามความหมายของคำสำคัญ

ต่อมา มนัสนันท์ ปัญญามี (2552) ได้นำเสนอแนวคิดการจัดกลุ่มข้อมูลของเว็บเพจตามความหมาย โดยหาความสัมพันธ์ในเชิงความหมายของคำสำคัญที่ผู้ใช้ระบุเข้ามา และใช้ฐานข้อมูลจากเวิร์ดเน็ต (WordNet) เพื่อจัดกลุ่มคำของเว็บเพจที่มีความหมายเหมือนกัน เรื่องเดียวกัน เกี่ยวข้องหรือเชื่อมโยงกันเชิงความหมายไว้ด้วยกัน ซึ่งรวมไปถึงเอกสารใด ๆ ที่ไม่มีคำสำคัญหรือคำค้นที่ผู้ใช้ระบุเข้ามาปรากฏอยู่ในเอกสารก็ตาม ผลการวิจัยพบว่าการสืบค้นสามารถค้นคืนความรู้หรือเอกสารได้มีประสิทธิภาพเพิ่มสูงขึ้น สอดคล้องกับการวิจัยของพิลาวัณย์ พลับรู้การ และ กฤษณะ ไวยมัย (2544) ที่มีการนำเสนอวิธีการหาความสัมพันธ์ของคำในโครงข่าย เพื่อวัดความคล้ายคลึงระหว่างเอกสาร โดยกำหนดตัวแทนเอกสารด้วยวิธีการแบบ เวกเตอร์ และหลักการทางสถิติในการวัด ซึ่งวิธีการดังกล่าวสามารถให้ผลลัพธ์สืบค้นเชิงความหมายจากคำสำคัญได้ตรงกับความต้องการของผู้ใช้

ต่อมาณัฐพร หอมเมือง (2550) ได้แนะนำหลักการให้ค่าน้ำหนักคำมาคำนวณร่วมกับการนับความถี่ของคำ ซึ่งกล่าวไว้ว่าคำที่มีความสำคัญหรือใช้เป็นตัวแทนของเอกสารที่ดีควรปรากฏอยู่เป็นจำนวนมากในเนื้อหาของเอกสารใด ๆ และปรากฏอยู่น้อยมากในชุดเอกสารที่เหลือทั้งหมด แต่ถ้าคำนั้นปรากฏเป็นจำนวนมากในทุก ๆ เอกสารแสดงว่าคำดังกล่าวไม่สามารถใช้เป็นตัวแทนของเอกสารใด ๆ ได้ ซึ่งเรียกคำเหล่านี้ว่าคำหยุด (Stop Word) ดังนั้นการให้ค่าน้ำหนักคำ ๆ ใด ๆ ในเอกสารฉบับหนึ่งควรพิจารณาจากความถี่ของคำที่ปรากฏในเอกสารนั้นและจำนวนของเอกสารทั้งหมดที่มีคำ ๆ นั้นปรากฏอยู่ สอดคล้องกับ เรสนิค (Resnik, 1999) ที่ได้กล่าวว่าการให้ค่าน้ำหนักคำแต่ละคำ (ทั้งคำที่มีในข้อสอบถามและคำในเอกสาร) จะคำนวณโดยเปรียบเทียบกับคำทั้งหมดในเอกสาร ค่าน้ำหนักคำในเอกสารจะได้รับการดำเนินการในขั้นตอนการเตรียมฐานความรู้ ส่วนค่าน้ำหนักคำในข้อสอบถามจะดำเนินการในแต่ละครั้งที่มีการสืบค้นด้วยวิธีการแบบเวกเตอร์ เมื่อคำนวณค่าน้ำหนักคำของข้อสอบถามได้เรียบร้อยแล้ว ค่าน้ำหนักเหล่านั้นจะถูกนำไปแทนในสมการเพื่อคำนวณระดับความคล้ายคลึงระหว่างข้อสอบถามและเอกสาร เรียกว่า การวัดความคล้ายคลึงระหว่างเอกสารกับคำสำคัญ

นอกจากนี้ อรวรรณ อุไรเรืองพันธุ์ และสมจิตร อาจอินทร์ (2009) ได้มีการสรุปเอกสารเชิงความหมายโดยใช้ออนโทโลยี ซึ่งใช้ฐานข้อมูลคำศัพท์จากเวิร์ดเน็ต (WordNet) โดยใช้หลักการ Lexical Chain สำหรับจัดหมวดหมู่ของกลุ่มคำเหมือน ซึ่งมี 2 ขั้นตอน ได้แก่ การหาคำสำคัญ (Candidate Word) และการค้นหากลุ่มคำสำคัญ (Lexical Chain) โดยทำการจัดหมวดหมู่กลุ่มคำที่มีความหมายใกล้เคียงกันไว้ด้วยกัน และพิจารณาความสัมพันธ์ระหว่างกลุ่มคำที่มี

ความหมายเหมือนกันรวมไปถึงใกล้เคียงกันแล้วทำการวัดความคล้ายคลึงระหว่างคำกับเอกสาร ผลการวิจัยพบว่าเทคนิควิธีที่นำเสนอทำให้ผลลัพธ์จากการสืบค้นมีประสิทธิภาพสูงขึ้น แต่มีข้อจำกัดคือใช้กับเอกสารที่เป็นอังกฤษเท่านั้น ต่อมา ศุภกฤษฎิ์ นิวัฒนากูล (2556) ได้นำเสนอเทคนิควิธีการสืบค้นเชิงความหมายที่สามารถใช้ได้กับการเข้าถึงความรู้หรือเอกสารภาษาไทยสำหรับความรู้ทางการเกษตร โดยใช้ออนโทโลยีอธิบายความหมาย ความสัมพันธ์และโครงสร้างของข้อมูลผ่านภาษาโอดับเบิ้ลยูแอล (Web Ontology Language: OWL) ประกอบด้วย 3 ส่วน คือ (1) การวิเคราะห์คำสำคัญ (2) การสืบค้นเชิงความหมาย มี 3 ขั้นตอน คือ กำหนดความสำคัญของคำค้น คำนวณน้ำหนักของคำค้น และ หาความสัมพันธ์ของคำค้นกับคลังคำ ซึ่งจะทำให้ได้กลุ่มคำค้นและรายการเอกสารที่เกี่ยวข้อง และ (3) การแสดงผลลัพธ์ มี 2 ขั้นตอน คือ การคำนวณน้ำหนักเอกสารต่อคำค้น และ การคำนวณความคล้ายคลึงของเอกสารต่อคำค้น ผลการวิจัยพบว่ากระบวนการที่นำเสนอสามารถทำให้เพิ่มประสิทธิภาพสำหรับการสืบค้นเชิงความหมาย ได้ผลลัพธ์ตรงกับความต้องการของผู้ใช้

จากการศึกษางานวิจัยที่เกี่ยวข้องกับการสืบค้นเชิงความหมาย สรุปได้ว่าส่วนใหญ่มีการจัดความรู้ หรือฐานความรู้ที่สามารถอธิบายความสัมพันธ์ของแนวคิดต่าง ๆ ด้วยออนโทโลยี เพราะสามารถกำหนดความหมายที่แท้จริงของคำสำคัญต่าง ๆ โดยอธิบายความหมาย ความสัมพันธ์และโครงสร้างของข้อมูลผ่านภาษาโอดับเบิ้ลยูแอล และงานวิจัยนี้นำเสนอเทคนิควิธีการวิเคราะห์คำสำคัญ การให้น้ำหนักคสำคัญ และการวัดความคล้ายคลึง ซึ่งเป็นขั้นตอนวิธีสำหรับการสืบค้นเชิงความหมาย

2.4 การวิเคราะห์คำสำคัญ

2.4.1 การกำหนดความสำคัญของคำค้น

การกำหนดคำค้นเป็นขั้นตอนหนึ่งที่สำคัญที่จะทำให้ผู้ใช้สามารถเลือกเอกสารที่ตนต้องการออกจากเอกสารที่ไม่ต้องการได้อย่างครบถ้วน ซึ่ง เฮิร์สต์ (Hearst, 1999) ได้อธิบายการกำหนดคำสืบค้นสามารถแบ่งตามขั้นตอนดังนี้

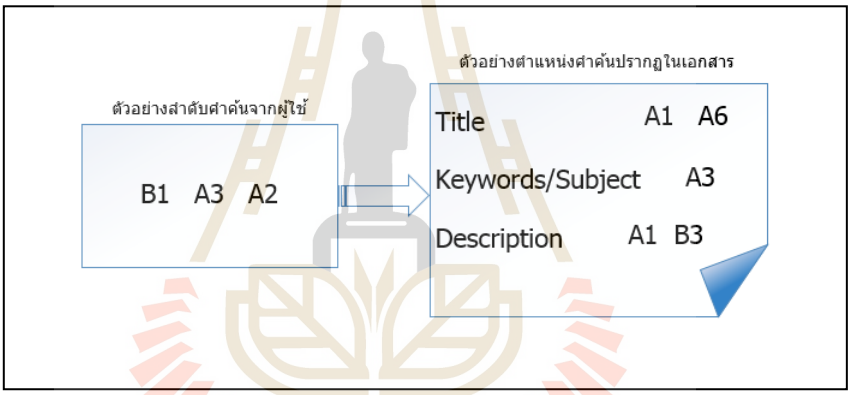
1. กำหนดคำสำคัญโดยเลือกจากชื่อเรื่องของของเอกสารความรู้
2. การหาคำพ้องต่าง ๆ ของแต่ละคำสำคัญ
3. กำหนดคำเชื่อมระหว่างคำสำคัญต่าง ๆ อย่างเหมาะสม ซึ่งอาจจะนำมาจากอภิธานศัพท์ที่มีการสร้างความสัมพันธ์ของคำไว้แล้ว

นอกจากนี้ สุธรรม อุมาแสงทองกุล (2541) ได้อธิบายไว้ว่าการกำหนดคำค้นสามารถกำหนดโดยวิเคราะห์จากชื่อเรื่อง คำสำคัญ และเนื้อหาของเอกสารความรู้ และการจัดลำดับความสำคัญของคำสำคัญ สมจิน เปียโคกสูง (2553) ได้กล่าวไว้ว่า การจัดลำดับความสำคัญของคำ

สำคัญอาศัยกระบวนการกำหนดน้ำหนักของดรรชนี (Term Weight) และแบจำลองการค้นคืน (Retrieval Model) สามารถกำหนดได้ดังนี้

1. ลำดับคำค้นจากที่ผู้ใช้ป้อนเข้ามา (Sequence of Words) ซึ่งในกระบวนการสืบค้นนั้น ผู้ใช้สามารถใช้หลายคำค้นเพื่อระบุถึงเอกสารที่ต้องการได้อย่างละเอียดชัดเจนมากขึ้น ดังนั้นการเรียงลำดับคำค้นจึงส่งผลต่อการสื่อความด้วยว่าคำใดมีความสำคัญก่อนหลัง

2. ตำแหน่งของคำที่ปรากฏในเอกสาร (Position of Words) ซึ่งในแต่ละเอกสารมีการปรากฏของคำอยู่ในส่วนต่าง ๆ เช่น ชื่อเรื่อง (Title) คำสำคัญ (Keywords/Subject) และเนื้อหา (Description/Abstract) เป็นต้น ซึ่งการให้ความสำคัญอาจมีการให้ความสำคัญจากชื่อเรื่อง คำสำคัญ และเนื้อหา ตามลำดับ และการกำหนดความสำคัญของคำให้กับส่วนต่าง ๆ ของเอกสารที่คำนั้นปรากฏอยู่ หมายถึงการให้ความสำคัญต่อคำต่าง ๆ และความเกี่ยวข้องของคำนั้นกับเอกสารด้วย ดังรูปที่ 2.16



รูปที่ 2.16 ตัวอย่างคำค้นที่ระบุจากผู้ใช้ และตำแหน่งที่ปรากฏในเอกสาร

จากรูปที่ 2.16 พบว่าผู้ใช้ระบุคำค้นเข้ามาสืบค้นเอกสาร จำนวน 3 คำ คือ B1, A3, และ A2 ซึ่งคำค้นปรากฏอยู่ในตำแหน่งของเอกสาร คือตำแหน่งชื่อเรื่อง คำสำคัญ และเนื้อหา

จากความสำคัญของการกำหนดคำค้นที่กล่าวมาข้างต้น สรุปว่า การกำหนดคำค้นเริ่มจากการวิเคราะห์เนื้อหาเอกสาร แล้วดำเนินการกำหนดน้ำหนักคำโดยใช้ความถี่ของคำที่ปรากฏในเอกสาร ซึ่งมีขั้นตอนดังนี้ (1) แยกคำสำคัญของเรื่องที่ต้องการค้น (2) หาคำพ้องต่าง ๆ ของคำสำคัญ และ (3) กำหนดคำเชื่อมระหว่างคำสำคัญ ซึ่งงานวิจัยนี้ได้กำหนดลำดับความสำคัญของคำค้นจาก (1) การกำหนดลำดับความสำคัญของคำค้นจากผู้ใช้โดยแยกคำเป็นคำเดี่ยว และ (2) การกำหนดความสำคัญของคำค้นโดยพิจารณาจากตำแหน่งของคำที่ปรากฏในเอกสาร

2.4.2 การให้น้ำหนักคำ

การให้น้ำหนักคำเป็นขั้นตอนสำคัญที่จะส่งผลต่อการค้นคืนสารสนเทศให้มีประสิทธิภาพ ซึ่งมีนักวิชาการหลาย ๆ ท่านได้อธิบายเกี่ยวกับการให้น้ำหนักคำโดย เรสนิค (Resnik, 1999) ได้กล่าวว่า น้ำหนักของคำแต่ละคำ(ทั้งคำที่มีในข้อสอบถามและคำในเอกสาร) จะคำนวณโดยเปรียบเทียบกับคำทั้งหมดในเอกสาร ค่าน้ำหนักคำในเอกสารจะได้รับการดำเนินการในขั้นตอนการเตรียมฐานความรู้ ส่วนค่าน้ำหนักคำในข้อสอบถามจะดำเนินการในแต่ละครั้งที่มีการสืบค้น เมื่อคำนวณค่าน้ำหนักคำของข้อสอบถามได้เรียบร้อยแล้วค่าน้ำหนักเหล่านั้นจะถูกนำไปแทนในสมการเพื่อคำนวณระดับความคล้ายคลึงระหว่างข้อสอบถามและเอกสาร เรียกว่า วิธีวัดความคล้ายคลึงระหว่างเอกสารคือ เวกเตอร์ ซึ่งธีรศักดิ์ สังข์ศรี (2548) ได้กล่าวว่า การให้ค่าน้ำหนักคำ ๆ หนึ่งในเอกสารฉบับหนึ่งจะพิจารณาจากการนับความถี่ของคำที่ปรากฏในเอกสารนั้น ๆ และจำนวนของเอกสารทั้งหมดที่มีคำ ๆ นั้นปรากฏอยู่ แต่นั่นไม่ได้หมายถึงว่าคำที่ปรากฏอยู่มากในเอกสารใด ๆ จะมีความสำคัญเสมอ เพราะคำเหล่านั้นอาจเป็นคำหยุด (Stop Word) เช่น คำว่า “และ” “หรือ” “ซึ่ง” เป็นต้น และวิธีการให้น้ำหนักคำที่นิยมใช้กันมาก คือ การกำหนดค่าน้ำหนักคำในเอกสารตามแนวคิดของ ซัลตอน (Salton, 1989) ที่ได้อธิบายว่าการแทนคำที่ปรากฏในเอกสารจะแทนคำให้อยู่ในรูปแบบของเวกเตอร์ ซึ่งมีการให้ค่าน้ำหนักของคำ (Term) โดยใช้ความถี่ของคำที่สนใจที่ปรากฏในเอกสาร (Term Frequency: TF) และการหาค่าส่วนกลับความถี่ของเอกสารทั้งหมดกับจำนวนของเอกสารที่ปรากฏคำที่สนใจ (Inverse Document Frequency: IDF) หรือเรียกว่า TF-IDF แสดงดังสมการที่ 2.5

$$\mathrm{IDF} = \log\left(\frac{\mathrm{N}}{\mathrm{df}}\right) \tag{2.5}$$

เมื่อ

IDF คือ ค่าส่วนกลับความถี่ของเอกสาร

N คือ จำนวนเอกสารในชุดเอกสารทั้งหมด

df คือ ค่าความถี่ของเอกสาร (Document Frequency) ของคำแต่ละคำ

ถ้าคำใดปรากฏในเอกสารทุกเอกสารจะมีค่าเท่ากับ N ซึ่งทำให้ค่าส่วนกลับความถี่ของเอกสารมีค่าเท่ากับศูนย์ ซึ่งกระบวนการของTF-IDF สามารถคำนวณหาค่าน้ำหนักคำได้ดังสมการที่ 2.6

$$w = \mathrm{tf} * \log\left(\frac{\mathrm{N}}{\mathrm{df}}\right) \tag{2.6}$$

มหาวิทยาลัยเทคโนโลยีสุรนารี

เมื่อ

tf คือ ความถี่ของคำที่ปรากฏในแต่ละเอกสาร

N คือ จำนวนเอกสารในชุดเอกสารทั้งหมด

df คือ ค่าความถี่ของเอกสาร (Document Frequency) ของคำแต่ละคำ

ณัฐพร หอมเมือง (2550) ได้นำหลักการให้ค่าน้ำหนักคำมาคำนวณร่วมกับการนับความถี่ของคำ โดยกล่าวถึงหลักการนี้ว่า คำที่มีความสำคัญหรือใช้เป็นตัวแทนของเอกสารที่ดีควรจะปรากฏอยู่เป็นจำนวนมากในเนื้อหาของเอกสารเฉพาะฉบับนั้น และปรากฏอยู่น้อยมากในชุดเอกสารที่เหลือทั้งหมด แต่ถ้าคำนั้นปรากฏเป็นจำนวนมากในทุก ๆ เอกสารแสดงว่าคำดังกล่าวไม่สามารถใช้เป็นตัวแทนของเอกสารใด ๆ ได้ และงานวิจัยของ คมคิด ชัชราภรณ์, ธรา อั่งสกุล และจิติมนต์ อั่งสกุล (2012) ที่ได้ประยุกต์เทคนิคการเลือกคุณลักษณะ โดยใช้วิธีการให้น้ำหนักของคำในเอกสารแบบ TF-IDF (Term Frequency-Inverse Document Frequency) ซึ่งเป็นวิธีที่ใช้ประเมินความสำคัญของคำหรือคุณลักษณะนั้น ๆ ในเอกสารที่เป็นชุดฝึก (Training Set) ทั้งหมด และงานวิจัยของ เทียน ,ฉ่าว, เวง และ ไค (Tian, Zhao, Wang, and Dai, 2015) ใช้วิธีการให้น้ำหนักคำด้วย TF-IDF ในการวิเคราะห์ข้อความบนหน้าเว็บ ซึ่งอ้างอิงถึง จิง ฮวง ซิ (Jing, Huang and Shi, 2002) ได้อธิบายว่าคุณลักษณะที่มีค่าน้ำหนัก TF-IDF มาก หมายความว่า คุณลักษณะนั้นมีประสิทธิภาพในการจำแนกเอกสารมากเช่นกัน

การให้น้ำหนักด้วยวิธี TF-IDF เป็นวิธีการที่นิยมนำมาใช้ในการสร้างแบบจำลองเอกสาร เนื่องจากมีประสิทธิภาพให้ค่าความแม่นยำในการจำแนกหมวดหมู่สูง และวิธีการคำนวณไม่ซับซ้อนใช้หน่วยประมวลผลไม่มากนัก แต่วิธีการให้น้ำหนักดังกล่าวยังมีจุดด้อย ซึ่ง จิระวิชิตชัย (2556) ได้อธิบายไว้ว่าการให้น้ำหนักแบบนี้ไม่ได้คำนึงถึงความหมายของคำ และถือว่าคำแต่ละคำเป็นอิสระต่อกัน ซึ่งความเป็นจริงแล้วคำแต่ละคำอาจมีความหมายเดียวกัน ดังนั้นการให้น้ำหนักคำควรมีการพิจารณาจากความถี่ของคำที่ปรากฏในเอกสาร และควรมีการพิจารณาจากความสัมพันธ์ของคำ ดังงานวิจัยของคิโมโต และอิเวดิร่า (Kimoto and Iwadera , 1990) ที่ได้นำเสนอวิธีสำหรับการค้นคืนโดยใช้เทคนิคการแยกคำจากเอกสารตัวอย่างที่ผู้ใช้สนใจมาทำเป็นดัชนี (Index) และสร้างความสัมพันธ์ภายในเอกสาร (Co-Occurrence) และมีการเพิ่มความสัมพันธ์ภายในเอกสารจากการคำนวณน้ำหนักของโหนด และเส้นเชื่อมระหว่างโหนดที่เป็นไปได้จากคำสำคัญที่ผู้ใช้ระบุเข้ามา ซึ่งการเลือกโหนด และเส้นทางเชื่อมโยงระหว่างโหนดที่เหมาะสมควรมีการคำนวณน้ำหนักโหนดและน้ำหนักของเส้นทางระหว่างโหนด เพื่อหาความเกี่ยวข้องของคำค้น

จากความสำคัญของการให้น้ำหนักคำที่กล่าวมาข้างต้น สรุปว่า เทคนิควิธีการให้น้ำหนักคำที่นิยมสำหรับการค้นคืนสารสนเทศในปัจจุบัน คือ วิธีการให้น้ำหนักคำแบบ TF-IDF

แต่การให้น้ำหนักแบบนี้จะไม่สามารถให้ความสำคัญกับความหมายของคำ หรือกลุ่มคำ จะถือว่าแต่ละคำเป็นอิสระต่อกัน ซึ่งนักวิจัยได้พยายามศึกษาขั้นตอนวิธีการให้น้ำหนักคำโดยพิจารณาจากความสัมพันธ์ของคำมาสร้างความสัมพันธ์ภายในเอกสารและวิธีการหาความสัมพันธ์ของคำจากโครงสร้างคำเพื่อเพิ่มประสิทธิภาพการสืบค้นให้ตรงกับความต้องการของผู้ใช้มากขึ้น

ดังนั้น งานวิจัยนี้จึงมีแนวคิดที่จะนำเสนอเทคนิควิธีการให้น้ำหนักคำที่สามารถสื่อถึงความหมายของคำอย่างแท้จริง อันจะเพิ่มประสิทธิภาพการสืบค้นให้ตรงกับความต้องการของผู้ใช้มากยิ่งขึ้น โดยใช้ทฤษฎี การหาความสัมพันธ์ระหว่างตัวแปรสองตัว (Bivariate Correlation) ซึ่งดีคอฟฟ์ (Diekhoff, 1992) เรียกว่าการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) ซึ่งเป็นวิธีการศึกษาความสัมพันธ์ระหว่างตัวแปรตั้งแต่ 2 ตัวขึ้นไป (หรือข้อมูล 2 ชุดขึ้นไป) ว่ามีความสัมพันธ์กันหรือไม่ และสัมพันธ์กันอย่างไร ความสัมพันธ์อาจเป็นไปในทิศทางเดียวกัน หรือทิศทางตรงข้ามกัน ตัวอย่างการศึกษาความสัมพันธ์ เช่น การหาความสัมพันธ์ระหว่างอายุและความดันโลหิต ความสัมพันธ์ระหว่างส่วนสูงกับน้ำหนัก เป็นต้น สำหรับการพิจารณาความสัมพันธ์ระหว่างตัวแปรว่ามีมากน้อยเพียงใดนั้น จะใช้ค่าสัมประสิทธ์สหสัมพันธ์เป็นค่าที่วัดความสัมพันธ์ มีค่าอยู่ระหว่าง - 1 ถึง 1 กรณีมีค่าเข้าใกล้ - 1 หรือ 1 แสดงถึงการมีความสัมพันธ์กันในระดับสูง แต่หากมีค่าเข้าใกล้ 0 แสดงถึงการมีความสัมพันธ์กันในระดับน้อย หรือไม่มีเลย ดังสมการที่ 2.7

$$Correl(X,Y) = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \sum (y-\bar{y})^2}} \qquad (2.7)$$

การพิจารณาค่าสัมประสิทธ์สหสัมพันธ์ โดยทั่วไปใช้เกณฑ์ของฮินเคอล์ (Hinkle, 1998: 118) ดังนี้

ค่า r	ระดับของความสัมพันธ์
.91 - 1.00	มีความสัมพันธ์กันมากที่สุด
.71 - .90	มีความสัมพันธ์กันในระดับมาก
.51 - .70	มีความสัมพันธ์กันในระดับปานกลาง
.31 - .50	มีความสัมพันธ์กันในระดับต่ำ
.00 - .30	มีความสัมพันธ์กันในระดับต่ำที่สุด

มหาวิทยาลัยเทคโนโลยีสุรนารี

2.5 การวัดความคล้ายคลึง

ระบบสืบค้นข้อมูลหรือสารสนเทศเกี่ยวกับเอกสารความรู้ใด ๆ มีแนวคิดที่ว่า คำหรือคำค้นที่คล้ายคลึงกันจะมีความสัมพันธ์ในการค้นหาจากข้อสอบถาม โดยปกติแล้วการสืบค้นความรู้ที่ต้องการจะมีการจับคู่กับคำของข้อสอบถาม สอดคล้องกับการอธิบายของ แจสวินเดอร์ พลาวินเดอร์ และโยเกรท์ (Jaswinder, Parvinder, and Yogesh, 2014) ที่กล่าวถึงวิธีการวัดความคล้ายคลึง ดังตารางที่ 2.4

ตารางที่ 2.4 ตัวอย่างวิธีการวัดความคล้ายคลึงแบบต่าง ๆ

วิธีวัดความคล้ายคลึง	การวัดค่าแบบ ไบนารี่เวกเตอร์	การวัดค่าแบบ ให้น้ำหนักเวกเตอร์
Dice's Coeffiecient	$2 \mid X \cap Y \mid / \mid X \mid + \mid Y \mid$	$\frac{2\sum_{i=1}^{t} X_i Y_i}{\sum_{i=1}^{t} X_i^2 + \sum_{i=1}^{t} Y_i^2}$
Jaccard's Coeffiecient	$\mid X \cap Y \mid / \mid X \cup Y \mid$	$\frac{\sum_{i=1}^{t} X_i Y_i}{\sqrt{\sum_{i=1}^{t} X_i^2 + \sum_{i=1}^{t} Y_i^2}}$
Cosine Coeffiecient	$\mid X \cap Y \mid / \mid X \mid 1/2 * \mid Y \mid ½$	$\frac{\sum_{i=1}^{t} X_i Y_i}{\sum_{i=1}^{t} X_i^2 + \sum_{i=1}^{t} Y_i^2 - \sum_{i=1}^{t} X_i}$

การวัดความคล้ายคลึงของเอกสารจะอาศัยค่าน้ำหนักคำสำคัญที่มีอยู่ในแต่ละเอกสารและค่าน้ำหนักคำของข้อสอบถาม โดยค่าน้ำหนักคำสำคัญในเอกสารคำนวณได้จากขั้นตอนการเตรียมข้อมูล และค่าน้ำหนักคำของข้อสอบถามคำนวณในขั้นตอนการเปรียบเทียบเอกสารในแต่ละครั้ง เมื่อคำนวณได้ค่าน้ำหนักคำในเอกสารและค่าน้ำหนักคำของข้อสอบถามจะนำค่าดังกล่าวไปคำนวณค่าความคล้ายคลึง ซึ่งสามารถคำนวณได้หลายวิธี เช่น การคำนวณค่าความคล้ายคลึงของเอกสารโดยใช้วิธีแบบเวกเตอร์ (Vector Based Similarity Measurement) ดังสมการที่ 2.8

$$sim(Q, D_j) = \frac{\sum_{\forall j}(w_{Q.J} \cdot w_{i,j})}{\sqrt{\sum_{\forall j} w_{Q.j^2}} \sqrt{\sum_{\forall j} w_{i.j^2}}} \qquad (2.8)$$

เมื่อ

$sim(Q, D_j)$ แทน ค่าความคล้ายคลึงระหว่างข้อสอบถามกับเอกสาร j

$\sum_{\forall j}(w_{Q.J} \cdot w_{i,j})$ แทน ค่าผลรวมระหว่างค่าน้ำหนักคำของข้อสอบถาม Q กับค่าน้ำหนักคำสำคัญ i ในเอกสาร j

$\sqrt{\sum_{\forall j} w_{Q.j^2}}$ แทน ผลรวมของค่าน้ำหนักทั้งหมดของข้อสอบถาม Q

$\sqrt{\sum_{\forall j} w_{i.j^2}}$ แทน ผลรวมของค่าน้ำหนักทั้งหมดของคำสำคัญ i ใน เอกสาร j

ปัจจุบันการวัดความคล้ายคลึงที่ได้รับความนิยม คือวิธีการวัดความคล้ายคลึงเชิงมุม (Cosine Coeffiecient) วิธีการวัดความคล้ายคลึงแบบนี้เป็นวิธีการเปรียบเทียบความคล้ายคลึงของเอกสาร 2 เอกสาร โดยแต่ละเอกสารจะถูกแทนด้วยเวกเตอร์ขนาดเอ็น (N-Dimensional Vector) เก็บค่าน้ำหนักคำแต่ละคำในเอกสารนั้น ซึ่งการเปรียบเทียบความคล้ายคลึงของเอกสารเปรียบเทียบโดยดูจากการทำมุมระหว่าง 2 เวกเตอร์ของเอกสาร หากเอกสารทั้งสองเอกสารคล้ายคลึงกันมาก เวกเตอร์ของเอกสารทั้ง 2 จะทับกันเกือบสนิทมุมซึ่งจะมีค่าน้อย ค่าความคล้ายคลึงจะมีค่าสูง กรณีที่มีค่าเท่ากับ 1 นั่นหมายถึง เวกเตอร์ทั้งสองทำมุมระหว่างกัน 0 องศา หรือกล่าวได้ว่าเวกเตอร์ทั้งสองมีทิศทางไปในแนวเดียวกัน ซึ่งคำนวณได้จากสมการที่ 2.9

$$Similarity = \frac{d * d'}{|d||d'|} \tag{2.9}$$

เมื่อ

$d * d'$ แทน เวกเตอร์ของโปรดักซ์ (Product)

สอดคลองกับงานวิจัยของ พิลาวัณย์ พลับรู้การ และ กฤษณะ ไวยมัย (2544) ได้นำเสนอการนำค่าความคล้ายคลึงระหว่างคำมาใช้คำนวณค่าความคล้ายคลึงระหว่างเอกสาร ซึ่งการหาค่าความเกี่ยวข้องระหว่างคำสองคำจากระยะห่างน้อยที่สุด เป็นการเปลี่ยนค่าระยะห่างที่น้อยที่สุดระหว่างคำเป็นค่าความเกี่ยวข้องจะอยู่ในรูปสมการคณิตศาสตร์ที่แปรผกผันกัน ซึ่งการวัดความคล้ายคลึงทางความหมายระหว่างคำในโครงข่ายโดยพิจารณาระยะทาง (Edge Based) น้อยที่สุดระหว่างคำในเอกสารเปรียบเทียบกับทุกคำในเอกสาร และจัดกลุ่มโดยเลือกกำหนดกลุ่มให้กับเอกสาร และหาค่าความคล้ายคลึงเฉลี่ยภายในกลุ่ม และนอกจากนั้นยังมีงานวิจัยของ สุภกฤษฎิ์ นิวัฒนากูล (2556) ที่ได้นำเสนอเทคนิควิธีการวัดความคล้ายคลึงเชิงความหมายระหว่างเอกสารกับคำค้น การกำหนดค่าความสำคัญของคำค้น คำนวณหาค่าน้ำหนักของคำค้น การหาเส้นทางสั้นที่สุดตามโครงสร้าง และคำนวณหาค่าความคล้ายคลึงเชิงความหมายของเอกสารเพื่อใช้ในการลำดับความสำคัญของเอกสารให้ตรงตามความต้องการของผู้ใช้ได้มากที่สุด ดังสมการที่ 2.10

$$S_k = \sum(Q_i * D_k) \quad (2.10)$$

เมื่อ

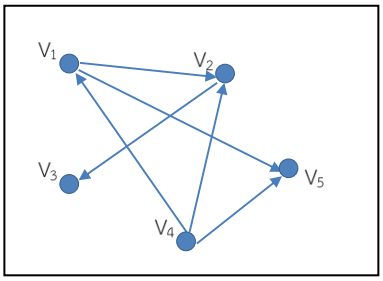
S_k แทน คะแนนความคล้ายคลึงเชิงความหมายของเอกสารที่ k

Q_i แทน ค่าน้ำหนักของคำค้นที่ i

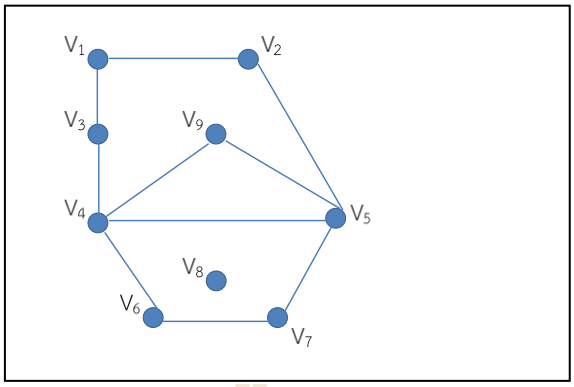
D_k แทน ค่าน้ำหนักของเอกสารต่อคำค้นที่ k

นอกจากนี้งานวิจัยของ รอย และ คริสทินา (Roi and Christina, 2012) ได้ประยุกต์นำเสนอเทคนิคการหาความสัมพันธ์ระหว่างคำ จากแนวคิดทฤษฎี กราฟ ซึ่งเป็นทฤษฎีทางคณิตศาสตร์ มีองค์ประกอบสองส่วน คือ จุดยอด (Vertex) และ เส้นเชื่อม (Edge) กล่าวคือ กราฟ คือเซ็ตของโหนด และเส้นเชื่อม ซึ่งโครงสร้างข้อมูลกราฟ (Graphs) ประกอบด้วยเซต V ซึ่งเราเรียกว่า จุดของกราฟ (Vertices of G) และเซต E ซึ่งแยกออกมาจากเซต V เราเรียกว่าด้านของกราฟ (Edges of G) หรือกล่าวอีกลักษณะหนึ่งว่า กราฟ คือ เซตของจุด (Points) และเซตของเส้น (Lines) ซึ่งเส้นจะเป็นตัวเชื่อมโยงจากจุดหนึ่งไปยังอีกจุดหนึ่ง คริสเทนเซน และ อัลเบิทร์ (Christensen and Albert, 2007) ได้อธิบายนิยามกราฟ คือ Graph G = (V, E) ประกอบด้วยเซ็ตของจุดยอด V และเซ็ตของเส้นเชื่อม E แต่ละเส้นแบ่งออกเป็น 2 ประเภท ได้แก่ ได้แก่ กราฟไม่มีทิศทาง (Undirected Graphs) และกราฟมีทิศทาง (Directed Graphs)

กราฟมีทิศทาง คือกราฟที่เป็นเส้นทางเชื่อมจากโหนดหนึ่งไปยังโหนดอื่น ๆ ได้แบบมีทิศทาง ซึ่งใช้เส้นเชื่อมมีลูกศรกำกับทิศทาง ดังรูปที่ 2.17 ซึ่ง กราฟไม่มีทิศทาง คือกราฟที่เป็นเส้นเชื่อมจากโหนดหนึ่งไปยังโหนดอื่น ๆ แบบไม่มีทิศทาง ซึ่งใช้เส้นเชื่อมไม่มีลูกศรกำกับทิศทาง และมองความสัมพันธ์ของ 2 โหนดแบบไปและกลับ แสดงดังรูปที่ 2.18

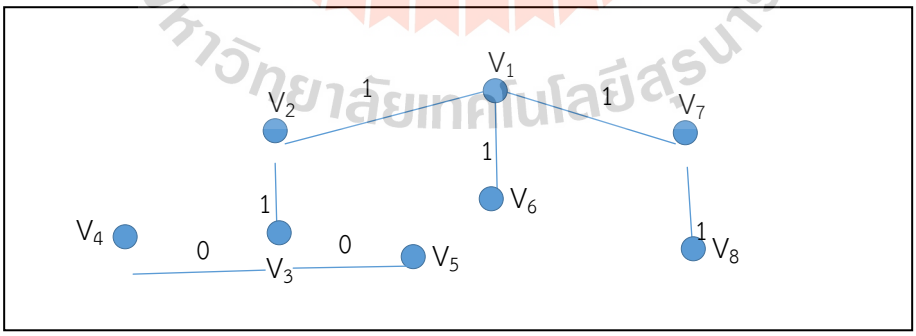


รูปที่ 2.17 ตัวอย่างของกราฟมีทิศทาง



รูปที่ 2.18 ตัวอย่างของกราฟไม่มีทิศทาง

กราฟมีทิศทางใน รูปที่ 2.17 กรณีเริ่มจาก V_1 : V_1-> V_2, V_1->V_5, V_1-> V_2->V_3 แต่ไม่สามารถไปถึง V_4 ได้ หากเริ่มจาก V_5 ก็ไม่สามารถค้นหาเส้นทางหรือเส้นเชื่อมไปยังโหนด อื่น ๆ ได้ แต่กราฟไม่มีทิศทางใน รูปที่ 2.18 ออเดอร์ และแรนยิ (Erdos and Renyi , 1961) ได้อธิบายว่า โหนดที่เชื่อมกันสามารถเข้าถึงกันได้หมด หากเริ่มจาก V_1 สามารถเข้าถึงได้ทุกโหนดยกเว้น V_8 เพราะ V_8 ไม่เชื่อมต่อกับใครเลย นอกจากนี้กราฟอาจจะมีน้ำหนัก หรือมีตัวเลขที่ถูกกำหนดในแต่ละด้าน ซึ่งเรียกว่า กราฟที่มีน้ำหนัก (Weighted Edges) โดยน้ำหนักอาจหมายถึงระยะทาง ทั้งนี้ขึ้นอยู่กับความสัมพันธ์ของกราฟ ซึ่งความสัมพันธ์ของกราฟดังกล่าวจะหมายถึงโครงสร้างความสัมพันธ์ระหว่างคำในคลังคำ และเป็นกราฟที่มีน้ำหนักโดยใช้ระยะทางระหว่างคำ แสดงดังรูปที่ 2.19



รูปที่ 2.19 ตัวอย่างของกราฟที่มีน้ำหนัก

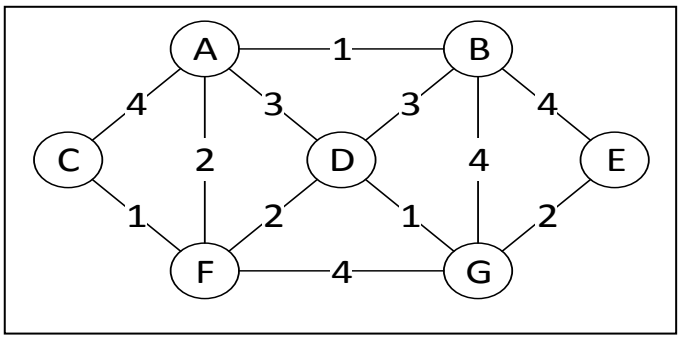
จากรูปที่ 2.19 ระยะทางของการเดินทางที่สั้นที่สุดจาก V_1 ไปยังจุดยอดอื่น ๆ คือ $V_1 \rightarrow V_2 = 1$, $V_1 \rightarrow V_3 = 2$, $V_1 \rightarrow V_4 = 2$, $V_1 \rightarrow V_5 = 2$, $V_1 \rightarrow V_6 = 1$, $V_1 \rightarrow V_7 = 1$, $V_1 \rightarrow V_8 = 2$ ซึ่งการหาเส้นทางที่สั้นที่สุดมีขั้นตอนวิธี (Algorithms) หลายแบบที่แตกต่างกันสำหรับการนำไปหาเส้นทางที่สั้นที่สุดระหว่าง 2 จุดยอด ในกราฟที่มีน้ำหนัก ซึ่งขั้นตอนวิธีที่นิยมใช้ในปัจจุบัน คือ ไดก์สตรา (Dijkstra Algorithm)

ไดก์สตรา (Dijkstra, 1959) ได้อธิบายเกี่ยวกับเทคนิควิธีไดก์สตราไว้ว่าเป็นการหาระยะทางที่สั้นที่สุดจากจุดยอดจุดเริ่มต้นไปยังจุดที่สอง และหาไปเรื่อย ๆ จนกว่าจะเจอเส้นทางที่สั้นที่สุด กล่าวคือจะค้นหาระยะทางจากเส้นเชื่อมระหว่างสองโหนดที่มีผลรวมน้ำหนักน้อยที่สุด ซึ่งขั้นตอนวิธีนี้จะอาศัยชุดของการวนซ้ำ โดยชุดของจุดที่แตกต่างกันจะถูกสร้างโดยการเพิ่มจุดเข้าไปในการวนซ้ำแต่ละรอบ ดังนี้

1. กำหนด PriorityQueue = {จุดยอด}
2. เมื่อ PriorityQueue ว่าง ให้ทำ
   - 1.1 ดึงสมาชิกตัวแรกในโครงสร้างข้อมูลออกมา (ให้ชื่อว่า X)
   - 1.2 หาก X เคยถูกเลือกเป็นเส้นทางแล้ว กลับไปข้อ 1.1
   - 1.3 เลือกโหนด X ให้เป็นเส้นทางจริง
   - 1.4 สำหรับโหนดใด ๆ ที่เชื่อมต่อกับ X ให้ทำดังนี้
     - 1.4.1 คำนวณระยะทางรวมของโหนด X มายังโหนดนั้น ๆ
     - 1.4.2 นำทุกเส้นทางไปไว้ใน แถวเรียงต่อกัน

ขั้นตอนวิธีไดก์สตรามีการมีการจัดลำดับการเข้าและออกข้อมูลอย่างเป็นลำดับโดยใช้เส้นทางรวมจากจุดเริ่มต้นมายังโหนดนั้น ๆ เพื่อระบุความสำคัญของสมาชิกที่อยู่ใน ลำดับ และแม็คคอนเนลล์ (McConnell, 2001) ได้อธิบายเกี่ยวกับการหาระยะทางที่สั้นที่สุด (Shortest Path) คือการหาเส้นทางสั้นที่สุดระหว่างโหนด (Node) จากเส้นเชื่อม (Edge) ค้นหาจากทุกเส้นทางที่เป็นไปซึ่งเป็นกราฟที่มีน้ำหนักและค่าน้ำหนักทั้งหมดเป็นค่าบวก มีข้อดีคือง่ายต่อการปรับค่าระยะทางที่สั้นที่สุด ซึ่งสามารถแสดงตัวอย่างการคำนวณหาเส้นทางสั้นที่สุด ดังนี้

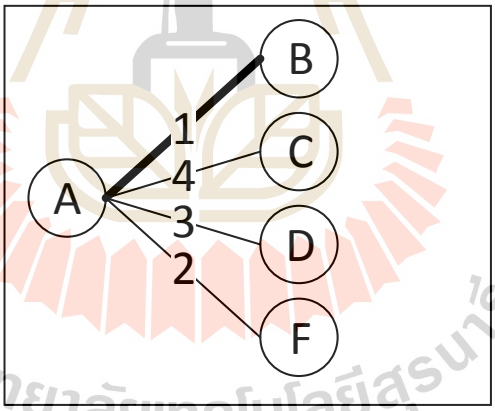
ซาน และนูน (Zhan and Noon, 1998) ได้อธิบายการคำนวณหาเส้นทางที่สั้นที่สุดจากโหนด A ไปยังโหนด G ตามโครงสร้างต่อไปนี้



รูปที่ 2.20 การหาเส้นทางจากโหนด A ไป B

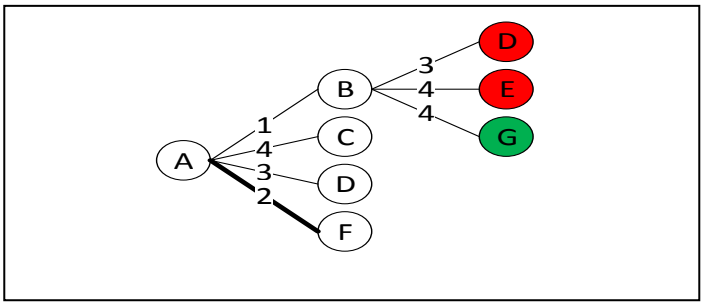
ศุภกฤษฎิ์ นิวัฒนากูล (2556) อธิบายการคำนวณหาเส้นทางที่สั้นที่สุดจากโหนด A ไป B จากโครงสร้างในรูปที่ 2.20 ดังนี้

1. กำหนดโหนด A เป็นโหนดเริ่มต้น จากนั้นหาเส้นเชื่อมจากโหนดต่าง ๆ ถึงโหนด A เลือกเส้นเชื่อมที่มีค่าน้ำหนักต่ำสุดในที่นี้คือโหนด B มีค่าน้ำหนักเป็น 1



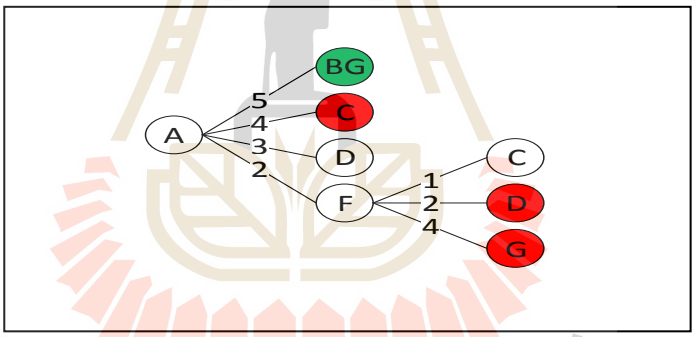
รูปที่ 2.21 การหาเส้นเชื่อมจากโหนดต่างๆ ถึงโหนด B

2. หาเส้นเชื่อมจากโหนดต่าง ๆ ถึงโหนด B หากเส้นเชื่อมมีโหนดร่วมกับเส้นทางอื่นให้ดูค่าน้ำหนักรวมจากจุดเริ่มต้นแล้วเลือกเส้นทางที่มีค่าน้ำหนักต่ำสุดไว้แล้วตัดเส้นทางอื่นทิ้ง ดังเส้นทาง ABD กับ เส้นทาง AD ซึ่งเส้นทาง ABD ถูกตัดทิ้งเนื่องจากมีค่าน้ำหนักสูงกว่า กรณีพบว่ามีเส้นทางจากโหนดต้นทางถึงปลายทางแล้ว เส้นทางอื่น ๆ ที่มีค่าน้ำหนักรวมต่ำสุดสูงกว่าเส้นทางนั้นจะถูกตัดทิ้งด้วย ดังเส้นทาง ABE จากนั้นหาเส้นเชื่อมจากโหนดต่าง ๆ ถึงโหนด A เลือกเส้นเชื่อมที่มีค่าน้ำหนักต่ำสุดในที่นี้คือโหนด F มีค่าน้ำหนักเป็น 2



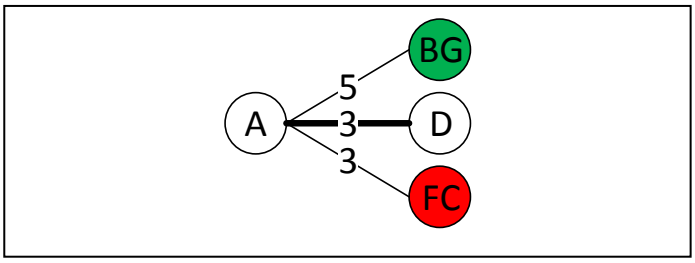
รูปที่ 2.22 การหาเส้นเชื่อมจากโหนดต่าง ๆ ถึงโหนด F

3. หาเส้นเชื่อมจากโหนดต่างๆ ถึงโหนด F เส้นทาง AC ถูกตัดทิ้งเนื่องจากมีค่าน้ำหนักรวมสูงกว่าเส้นทาง AFC เส้นทาง AFD ถูกตัดทิ้งเนื่องจากมีค่าน้ำหนักรวมสูงกว่าเส้นทาง AD และเส้นทาง AFG ถูกตัดทิ้งเนื่องจากมีค่าน้ำหนักรวมสูงกว่าเส้นทาง ABG



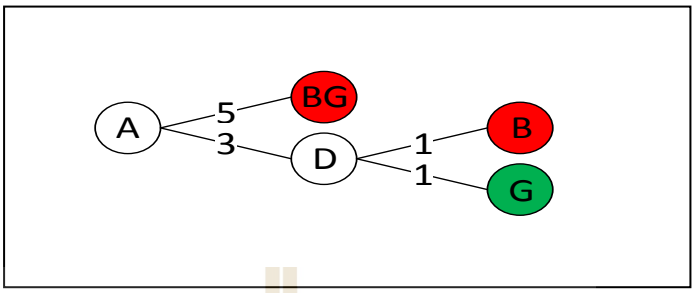
รูปที่ 2.23 เส้นทาง AFC ถูกตัดทิ้ง

4. เส้นทาง AFC ถูกตัดทิ้งเนื่องจากไม่มีโหนดเชื่อมต่อ จากนั้นหาเส้นเชื่อมจากโหนดต่าง ๆ ถึงโหนด A เลือกเส้นเชื่อมที่มีค่าน้ำหนักต่ำสุดในที่นี้คือโหนด D มีค่าน้ำหนักเป็น 3



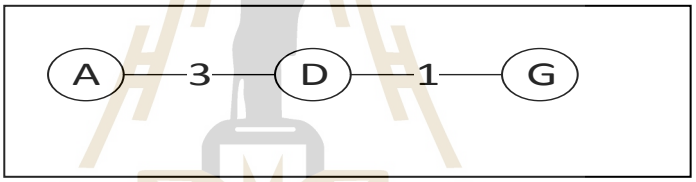
รูปที่ 2.24 เส้นทาง ADB ถูกตัดทิ้ง

5. เส้นทาง ADB ถูกตัดทิ้งเนื่องจากมีค่าน้ำหนักรวมสูงกว่าเส้นทาง AB และ เส้นทาง ABG ถูกตัดทิ้งเนื่องจากมีค่าน้ำหนักรวมสูงกว่าเส้นทาง ADG



รูปที่ 2.25 เส้นทาง ABG ถูกตัดทิ้ง

6. ผลลัพธ์สุดท้ายคือเส้นทางสั้นที่สุด คือ ADG มีค่าน้ำหนักรวมเท่ากับ 4



รูปที่ 2.26 ผลลัพธ์เส้นทางที่สั้นที่สุด

จากความสำคัญของการวัดความคล้ายคลึงที่กล่าวมาข้างต้น สรุปว่าการวัดความคล้ายคลึงเป็นขั้นตอนหนึ่งสำหรับการค้นคืนสารสนเทศ ซึ่งมีการวัดความคล้ายคลึงของคำสำคัญและนำผลการคำนวณไปใช้กับการคำนวณความคล้ายคลึงระหว่างเอกสารกับคำค้น โดยปัจจุบันเทคนิควิธีการวัดความคล้ายคลึงอยู่ในความสนใจของนักวิจัย คือ เทคนิควิธีการค้นหาเส้นทาง จากคำหนึ่งไปยังคำที่ต้องการที่มีความเกี่ยวข้องกันเชิงความหมายในคลังคำ โดยมีการคำนวณจากเส้นเชื่อม (Edge) ระหว่างโหนด (Node) ที่มีผลรวมน้ำหนักน้อยที่สุด กล่าวคือ ถ้าคำสองคำมีค่าระยะทางห่างกันน้อยหรือสั้นที่สุดจะถือว่าคำสองคำนั้นมีความเกี่ยวพันกันทางความหมายค่อนข้างมาก แต่ถ้าคำสองคำมีค่าระยะทางห่างกันมากจะถือว่าคำทั้งสองมีความเกี่ยวพันกันทางความหมายน้อย ดังนั้นเอกสารที่คล้ายคลึงกันไม่จำกัดอยู่เพียงแค่คำเดียวกันต้องปรากฏในแต่ละเอกสารเหมือนกัน ซึ่งคำแต่ละคำอาจมีความหมายเหมือนกัน หรือใกล้เคียงกันเชิงความหมาย

สรุป การวัดความคล้ายคลึงของเอกสารใช้ค่าน้ำหนักจากคำสำคัญที่ปรากฏในเอกสารและน้ำหนักคำของข้อสอบถามจากผู้ใช้ มีหลากหลายวิธี เช่น วิธีเวกเตอร์ การวัดความคล้ายคลึงเชิงมุม

และจากการคำนวณหาเส้นทางที่สั้นที่สุดระหว่างคำในโครงสร้าง ซึ่งงานวิจัยนี้ได้นำเสนอแนวคิดการวัดความคล้ายคลึงจากระยะทางที่สั้นที่สุดของความสัมพันธ์กันระหว่างคำสองคำในโครงสร้าง และมีการพิจารณาร่วมกับการกำหนดลำดับความสำคัญคำค้น และการให้น้ำหนักคำสำคัญ

2.6 กรอบแนวคิดการวิจัย

งานวิจัยนี้ได้เปรียบเทียบงานวิจัยที่เกี่ยวข้อง โดยสามารถจำแนกได้ 2 ประเด็น คือ การจัดความรู้ให้เป็นระบบ และการสืบค้นเชิงความหมาย แสดงดังตารางที่ 2.5

ตารางที่ 2.5 สรุปเปรียบเทียบงานวิจัยที่เกี่ยวข้อง

งานวิจัย	การจัดความรู้ให้เป็นระบบ	การสืบค้นเชิงความหมาย			การประเมินประสิทธิภาพ
		การกำหนดลำดับความสำคัญของคำค้น	การให้น้ำหนักคำสำคัญ	การวัดความคล้ายคลึง	
พิลาวัณย์ พลับรู้การและกฤษณะ ไวยมัย (2544)	อภิธานศัพท์โดเมนคอมพิวเตอร์	กำหนดความสำคัญของคำในโครงสร้างจากคำกว้างกว่า, คำแคบกว่า และคำเกี่ยวข้อง	ค่าถ่วงน้ำหนักความสัมพันธ์ระหว่างกลุ่มคำในโครงสร้าง	คำนวณค่าความคล้ายคลึงจากจากโครงข่ายความสัมพันธ์ของคำ	ค่าความแม่นยำและค่าระลึก
วุฒิชัย ปียะพันธวงศ์ (2545)	ฐานข้อมูลเชิงสัมพันธ์	ให้คะแนนตำแหน่งของคำสำคัญที่ปรากฏในเอกสาร	TF		-ขนาดฐานข้อมูลเทียบกับขนาดเอกสาร -ระยะเวลาการทำดัชนี
ณัฐพร หอมเมือง (2550)	ฐานข้อมูลเชิงสัมพันธ์	-	TF-IDF	แบบจำลองปริภูมิเวกเตอร์	ค่าความแม่นยำ
สิทธิโชค ปัญญาฤกษ์ชัย และศิพาณี นุชิตประสิทธิ์ชัย (2552)	ฐานข้อมูลเชิงสัมพันธ์	-	TF-IDF	N-Gram	ค่าความแม่นยำและค่าระลึก

มหาวิทยาลัยเทคโนโลยีสุรนารี

ตารางที่ 2.5 สรุปเปรียบเทียบงานวิจัยที่เกี่ยวข้อง (ต่อ)

งานวิจัย	การจัดความรู้ให้เป็นระบบ	การสืบค้นเชิงความหมาย			การประเมินประสิทธิภาพ
		การกำหนดลำดับความสำคัญของคำค้น	การให้น้ำหนักคำสำคัญ	การวัดความคล้ายคลึง	
ธนิตา วงศ์กาฬสินธุ์ และ งามนิจ อาจอินทร์ (2552)	ออนโทโลยี			คำนวณค่าความคล้ายคลึงระหว่างเอ็นทิตี้ของ Source กับ Target	ค่าความแม่นยำ, ค่าระลึก และค่าความถ่วงดุล
อรวรรณ อุไรเรืองพันธุ์ และ สมจิตร อาจอินทร์ (2552)	ออนโทโลยี	สกัดคำหลักและจัดหมวดหมู่ของคำ	ค่าถ่วงน้ำหนักความสัมพันธ์ระหว่างกลุ่มคำในโครงสร้าง	-	ค่าความแม่นยำ, ค่าระลึก และค่าความถ่วงดุล
จุฑาวรรณ สิทธิโชคสถาพร (2555)	ออนโทโลยี	-	-	-	เปรียบเทียบการสืบค้นจากฐานความรู้ออนโทโลยีกับฐานข้อมูลเชิงสัมพันธ์
จิรัฏฐา ภูบุญอบ และ วรวิทย์ สังฆทิพย์ (2555)	ออนโทโลยี	-	-	วัดความคล้ายคลึงระหว่างคลาสของออนโทโลยีแบบเมตริก	ค่าความแม่นยำ, ค่าระลึก และค่าความถ่วงดุล
บารมี โอสธีรกุล และ ธวัชชัย งามสันติวงศ์ (2555)	ฐานข้อมูลเชิงสัมพันธ์	-	TF-IDF	วัดความคล้ายคลึงเชิงมุม	ค่าความแม่นยำ และค่าระลึก
พยุง มีสัจ วาทินี นุ้ยเพียร และ ผุสดี บุญรอด (2556)	ฐานข้อมูลเชิงสัมพันธ์ ร่วมกับออนโทโลยี	ตัดคำสำคัญของบทความในส่วนของชื่อเรื่อง ผู้แต่ง และคำสำคัญมาสร้างเป็นตัวแทนเอกสาร และคัดเลือกคุณลักษณะ	เทคนิคการจำแนก	หากฎความสัมพันธ์เชิงความหมาย	ค่าความแม่นยำ, ค่าระลึก และค่าความถ่วงดุล

ตารางที่ 2.5 สรุปเปรียบเทียบงานวิจัยที่เกี่ยวข้อง (ต่อ)

งานวิจัย	การจัดความรู้ให้เป็นระบบ	การสืบค้นเชิงความหมาย			การประเมินประสิทธิภาพ
		การกำหนดลำดับความสำคัญของคำค้น	การให้น้ำหนักคำสำคัญ	การวัดความคล้ายคลึง	
ศุภกฤษฎิ์ นิวัฒนากูล (2556)	ออนโทโลยี	กำหนดความสำคัญของคำในโครงสร้างจากคำกว้างกว่า, คำแคบกว่า และคำเกี่ยวข้อง	ค่าถ่วงน้ำหนักความสัมพันธ์ระหว่างกลุ่มคำในโครงสร้าง	แจ็คการ์ด	ค่าความแม่นยำ, ค่าระลึก และค่าความถ่วงดุล
ชารินี พรหมภักดี มารุต บูรณรัช และจรัสศรี รุ่งรัตนาอุบล (2558)	ออนโทโลยี	-	-	-	ค่าความแม่นยำ
Cristiano, Daniel and Marcus (2004)	ออนโทโลยี	ความสัมพันธ์ของคุณสมบัติและคำศัพท์	TF-IDF	-	ค่าความแม่นยำ
Enrico, Sebastiano and Elisabetta (2008)	ออนโทโลยี	สกัดคำหลักและการจัดหมวดหมู่ของคำ	ค่าถ่วงน้ำหนักความสัมพันธ์ระหว่างกลุ่มคำในโครงสร้าง	วัดความคล้ายคลึงเชิงมุม	ค่าความแม่นยำและค่าระลึก
Jihyun, Jun and Chin (2009)	ออนโทโลยีจากข้อมูล 2 แหล่ง คือ DBLP (http://dblp.uni-trier.de/xml/) and IMDB (http://www.im-db.com/interface)	-	ค่าถ่วงน้ำหนักความสัมพันธ์ระหว่างกลุ่มคำในโครงสร้าง	-	จัดลำดับความถูกต้องด้วย Semantic Search Rank (SSR)
A.K. Sharma, Neelam and Bharti (2010)	ฐานข้อมูลเชิงสัมพันธ์	ความสำคัญของคำค้นจากผู้ใช้ และตำแหน่งคำที่ปรากกฏในเอกสาร	-	-	ค่าความแม่นยำ

ตารางที่ 2.5 สรุปเปรียบเทียบงานวิจัยที่เกี่ยวข้อง (ต่อ)

งานวิจัย	การจัดความรู้ให้เป็นระบบ	การสืบค้นเชิงความหมาย			การประเมินประสิทธิภาพ
		การกำหนดลำดับความสำคัญของคำค้น	การให้น้ำหนักคำสำคัญ	การวัดความคล้ายคลึง	
G.R.J. Srinivas, Niket and Vasudeva (2010)	สกัดข้อมูลจากเว็บไซต์		TF-IDF	วัดความคล้ายคลึงเชิงมุม	ค่าความแม่นยำ, ค่าระลึก และค่าความถ่วงดุล
Madhura, Charitha and Dilina (2012)	สกัดข้อมูลจากเว็บไซต์	สกัดดัชนีตัวแทนเอกสารจากเมทาดาทา		ความใกล้เคียงของ แท็กกับคำสำคัญ	ค่าความแม่นยำและค่าระลึก
Amir and Mourad (2013)	ออนโทโลยี		คำนวณจากเซ็ตของโหนด และเส้นเชื่อม	คำนวณความคล้ายคลึงจากความสัมพันธ์ของคำในโครงสร้าง	ค่าความแม่นยำ, ค่าระลึก และค่าความถ่วงดุล
Lee, Min and Oh (2014)	ออนโทโลยี	ขยายขอบเขตคำสำคัญ และสร้างกลุ่มคำที่มีความเกี่ยวข้อง	ค่าถ่วงน้ำหนักความสัมพันธ์ระหว่างกลุ่มคำในโครงสร้าง	ความคล้ายคลึงระหว่างชุดของข้อมูลเชิงความหมาย	ค่าความแม่นยำเฉลี่ย
Dharminder and Ashwani (2015)	สกัดข้อมูลจากเว็บไซต์	-	TF	-	ค่าความแม่นยำ
งานวิจัยนี้	ออนโทโลยี	ลำดับคำค้นจากผู้ใช้ และตำแหน่งของคำค้นที่ปรากฏในเอกสาร	การวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ และ IDF	ระยะทางระหว่างคำในโครงสร้าง	ค่าความแม่นยำเฉลี่ย, ค่าความผิดพลาด ค่าความแม่นยำ ค่าระลึก และค่าความถ่วงดุล

จากตารางที่ 2. 5 งานวิจัยนี้นำเสนอขั้นตอนวิธีที่หลากหลายมาใช้สำหรับการออกแบบและพัฒนาแบบจำลองการสืบค้นความรู้ เพื่อให้การสืบค้นมีประสิทธิภาพและตอบสนองต่อความต้องการของผู้ใช้ รายละเอียดดังนี้

1. การจัดความรู้ให้เป็นระบบ

1.1 ออนโทโลยี

ปัจจุบันมีการนำออนโทโลยีมาใช้ในขอบเขตงานที่หลากหลาย เช่น การเกษตร การศึกษา และการแพทย์ เป็นต้น งานวิจัยนี้ใช้ออนโทโลยีด้านไม้ผลเศรษฐกิจที่วิจัยและพัฒนาขึ้น

โดย ศุภกฤษฎิ์ นิวัฒนากูล (2556) มาประยุกต์ใช้ในการออกแบบฐานความรู้ที่อยู่ในรูปแบบภาษาโอดับเบิ้ลยูแอล โดยนำเอกสารความรู้เกี่ยวข้องกับพืชไร่ที่มีคำอธิบายครบถ้วนเผยแพร่บนเครือข่ายอินเทอร์เน็ตตั้งแต่ปี พ.ศ. 2550 จำนวน 240 เอกสารมาวิเคราะห์เพื่อหาคำสำคัญที่อยู่ในขอบเขตพืชไร่ จากนั้นจึงนำคำสำคัญที่วิเคราะห์ได้มาจัดเก็บในออนโทโลยี ซึ่งลักษณะข้อมูลประกอบด้วย ชื่อเรื่อง คำอธิบายความรู้โดยย่อ คำสำคัญ และแหล่งอ้างอิงต้นฉบับ โดยใช้เมทาดาทาสำหรับอธิบายรายละเอียดของข้อมูล

1.2 คลังคำ

คลังคำ หรือดัชนีคำสำคัญ คือคำที่กำหนดขึ้นมาเพื่อใช้แทนการค้นหาเนื้อหาของเอกสารหรือความรู้ ซึ่งคำหนึ่งคำมีหลายความหมาย คำหลาย ๆ คำมีความหมายเดียวกัน และหนึ่งคำหรือกลุ่มคำเรียกได้หลายชื่อ เป็นต้น การใช้ดัชนีเป็นตัวแทนสำหรับการค้นหามีจุดประสงค์เพื่อหลีกเลี่ยงการค้นหาแบบเรียงลำดับ และจะช่วยลดเวลาในการค้นหาข้อมูลที่ต้องการ โดยการจัดทำดัชนีคำสำคัญที่พบในงานวิจัยมีหลากหลายเทคนิค เช่น (1) การทำดัชนีแบบอินเวิร์ด ซึ่งประกอบด้วยพจนานุกรม (Dictionary) และการประกาศ (Posting) เก็บไว้ในหน่วยความจำหลักและมีตัวชี้ (Pointer) ชี้ไปยังการประกาศต่าง ๆ ที่เก็บไว้ในหน่วยความจำสำรอง คำสำคัญจะถูกเรียงลำดับตามตัวอักษรพร้อมกับความถี่ของคำคำนั้นที่ปรากฏในเอกสารต่าง ๆ คำเดียวกันจะนับความถี่รวมกัน ส่วนนี้เรียกว่า“ดิกชันนารี” และมีตัวชี้ไปยังการเก็บรายการเอกสารต่าง ๆ ที่คำนั้น ๆ ปรากฏ (2) การทำดัชนีแบบแอลเอสไอ และ (3) การทำดัชนีโดยใช้ออนโทโลยี เป็นต้นสำหรับงานวิจัยนี้ได้จัดทำดัชนีโดยใช้ ออนโทโลยี และใช้ตัวแบบการแทนความรู้สำหรับสร้างคลังคำ เพื่อแก้ปัญหาความแตกต่างของข้อมูลที่หลากหลาย และความสัมพันธ์ระหว่างคำ ซึ่งนำคำศัพท์มาจากอรรถาภิธานศัพท์เกษตรไทย ที่รวบรวมคำต่าง ๆ เกี่ยวกับด้านการเกษตร และมีการเพิ่มเติมคำที่เกี่ยวข้องจากเอกสารความรู้ด้านพืชไร่ของประเทศไทย

2. การสืบค้นเชิงความหมาย

การสืบค้นเชิงความหมายสำหรับความรู้ด้านพืชไร่สำหรับการวิจัยครั้งนี้ ประกอบด้วย 3 ส่วน ได้แก่ (1) การกำหนดลำดับความสำคัญของคำค้น (2) การให้น้ำหนักคำสำคัญ และ (3) การวัดความคล้ายคลึง ดังนี้

2.1 การกำหนดลำดับความสำคัญของคำค้น

การกำหนดลำดับความสำคัญของคำค้นจะมีขยายขอบเขตคำสำคัญ และสร้างกลุ่มคำตามแนวคิดของเฮิร์สต์ (Hearst, 1999) ซึ่งงานวิจัยนี้ได้นำแนวคิดการกำหนดลำดับความสำคัญของคำค้น 2 กรณี ได้แก่ พิจารณาลำดับคำค้นจากผู้ใช้ และ ตำแหน่งของคำค้นที่ปรากฏในเอกสาร

มหาวิทยาลัยเทคโนโลยีสุรนารี

2.2 การให้น้ำหนักคำสำคัญ

การให้น้ำหนักคำสำคัญในงานวิจัยนี้ใช้แนวคิดของดีคอฟฟ์ (Diekhoff, 1992) วิเคราะห์สัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) โดยใช้สถิติหาความสัมพันธ์ระหว่างตัวแปรสองตัว คือคำสองคำที่ปรากฏในเอกสารเดียวกันว่ามีความสัมพันธ์กันหรือไม่ และสัมพันธ์กันอย่างไร โดยทดลองเปรียบเทียบกับวิธีการที่นักวิจัยส่วนใหญ่นิยมใช้ ได้แก่ IDF

2.3 การวัดความคล้ายคลึง

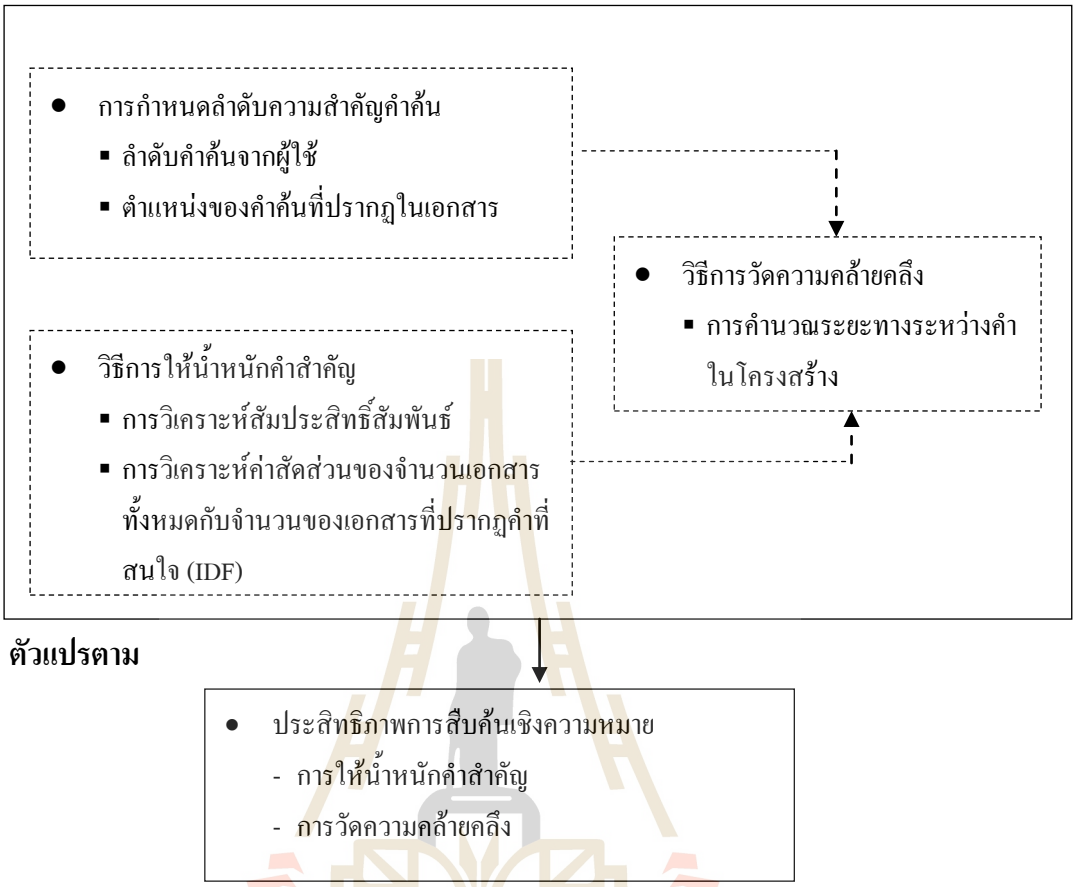
การวัดความคล้ายคลึงได้ประยุกต์จากแนวคิดของพิลาวัณย์ พลับรู้การ และกฤษณะ ไวยมัย (2544) และศุภกฤษฎิ์ นิวัฒนากูล (2556) โดยการวัดความคล้ายจากระยะทางที่สั้นที่สุดของความสัมพันธ์กันระหว่างคำสองคำในโครงสร้าง ซึ่งใช้ขั้นตอนวิธีไดก์สตรา สำหรับค้นหาเส้นทางจากโหนดหนึ่งไปยังโหนดที่ต้องการในโครงสร้าง

3. การประเมินประสิทธิภาพ

การประเมินประสิทธิภาพของการสืบค้นมีอยู่หลายวิธี ซึ่งแฟรงค์เคส , วิลเลี่ยม , เบร์ซา- เยต์ส และริคาร์โด (Frankes, William, Baeza-Yates and Ricardo, 1992) ได้กล่าวว่าวิธีที่นิยมใช้สำหรับการสืบค้น ได้แก่ การใช้ค่าความแม่นยำเฉลี่ย ค่าความผิดพลาด ค่าความแม่นยำ ค่าความระลึก และ ค่าความถ่วงดุล ซึ่งเป็นเหตุผลที่งานวิจัยนี้ได้นำวิธีการประเมินดังกล่าวมาใช้สำหรับการประเมินประสิทธิภาพของการสืบค้น แบ่งเป็น 2 ประเด็น คือ (1) การประเมินประสิทธิภาพการให้น้ำหนักคำสำคัญประเมินโดยใช้ค่าความแม่นยำเฉลี่ย และค่าความผิดพลาด เพราะต้องการวัดประสิทธิภาพการจัดอันดับของผลลัพธ์ที่ค้นพบอย่างถูกต้อง และค่าความผิดพลาดที่เกิดขึ้นจากการให้น้ำหนักคำสำคัญซึ่งค่าน้ำหนักบางคำอาจมีค่าต่ำส่งผลต่อการเรียงลำดับข้อมูล และ (2) การวัดความคล้ายคลึงประเมินโดยใช้ค่าความแม่นยำ ค่าความระลึก และค่าความถ่วงดุล เพราะต้องการ พิจารณาให้ความสำคัญกับความแม่นยำ ความถูกต้องครบถ้วน และค่าเฉลี่ยรวมของข้อมูลที่สืบค้นได้

จากการศึกษาเปรียบเทียบงานวิจัยที่เกี่ยวข้อง สรุปเป็นกรอบแนวคิดสำหรับการวิจัยดังรูปที่ 2.27

ตัวแปรอิสระ

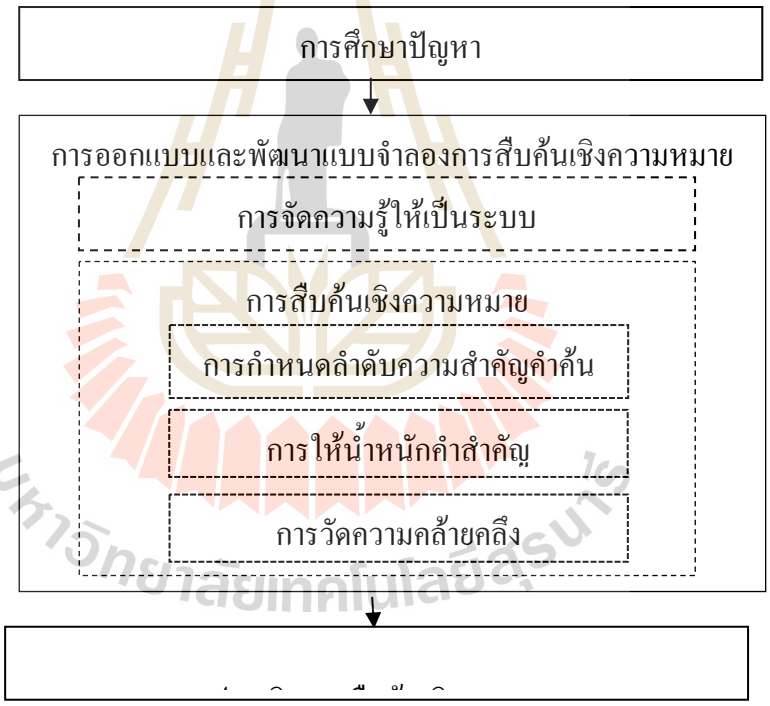


รูปที่ 2.27 กรอบแนวคิดการวิจัย

มหาวิทยาลัยเทคโนโลยีสุรนารี

บทที่ 3
วิธีดำเนินการวิจัย

งานวิจัยนี้เป็นการวิจัยเชิงพัฒนาแบบจำลองการสืบค้นเชิงความหมายสำหรับความรู้ด้านพืชไร่ ซึ่งประยุกต์ขั้นตอนจากวงจรการพัฒนาระบบ (System Development Life Cycle: SDLC) เพื่อให้การดำเนินการวิจัยมีระเบียบแบบแผนที่เหมาะสม งานวิจัยนี้ได้แบ่งการวิจัยออกเป็น 3 ระยะ ตามกรอบการวิจัย ได้แก่ (1) การศึกษาปัญหา (2) การออกแบบและพัฒนาแบบจำลองการสืบค้นเชิงความหมาย (3) การประเมินผลการสืบค้นเชิงความหมาย ดังรูปที่ 3.1



รูปที่ 3.1 ระเบียบวิธีการวิจัย

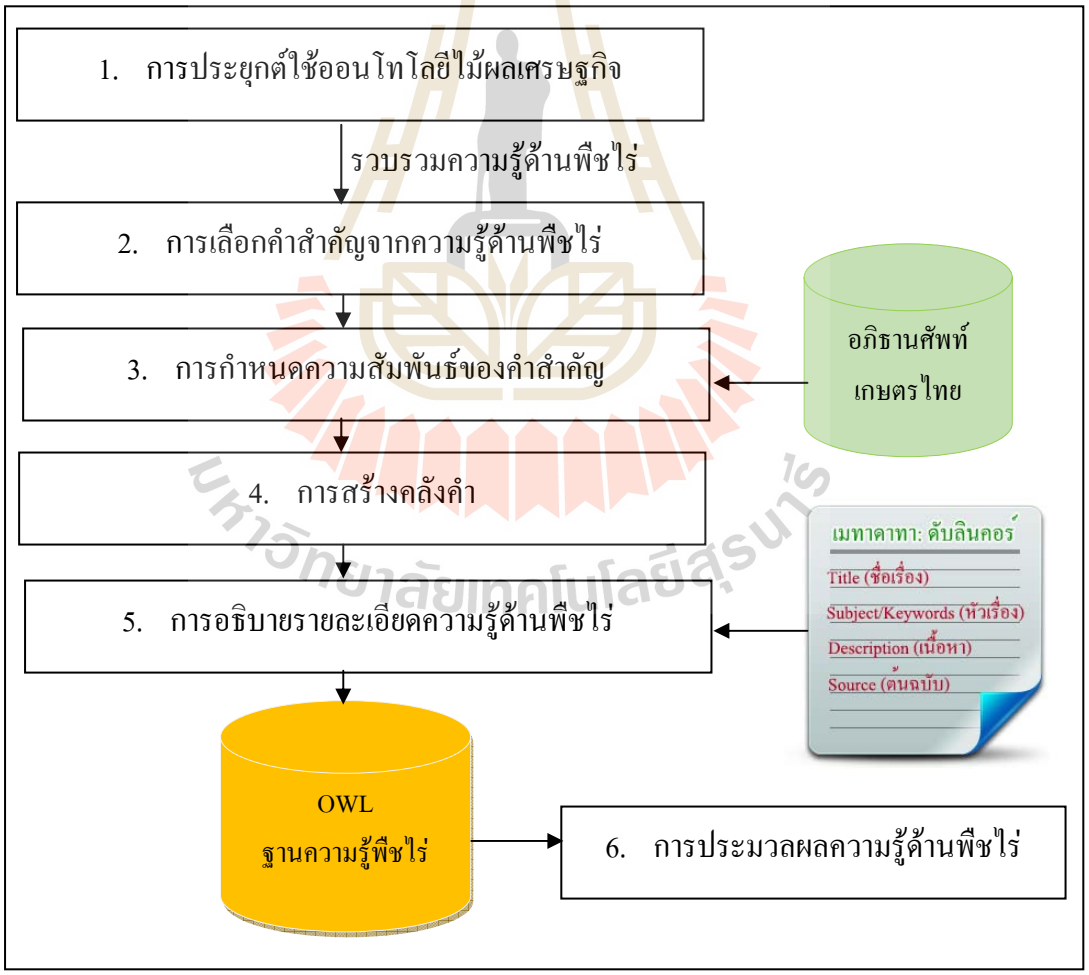
3.1 การศึกษาปัญหา

งานวิจัยนี้ได้ศึกษาปัญหาของวิธีการสืบค้นความรู้ และเก็บรวบรวมปัญหาจากแนวคิดและงานวิจัยที่เกี่ยวข้องนำมาวิเคราะห์ สรุปประเด็นปัญหาที่ค้นพบ ซึ่งมี2 ประเด็นหลัก ได้แก่ การจัดความรู้ให้เป็นระบบ และการสืบค้นความรู้

3.2 การออกแบบและพัฒนาแบบจำลองการสืบค้นเชิงความหมาย

การออกแบบและพัฒนาแบบจำลองการสืบค้นเชิงความหมายสำหรับความรู้ด้านพืชไร่มี 2 กระบวนจากปัญหาที่ค้นพบ ได้แก่การจัดความรู้ให้เป็นระบบ และการสืบค้นเชิงความหมาย ดังนี้

3.2.1 การออกแบบและพัฒนากระบวนการจัดความรู้ให้เป็นระบบมีขั้นตอนการออกแบบและพัฒนาดังรูปที่ 3.2



รูปที่ 3.2กรอบแนวคิดขั้นตอนการจัดความรู้ให้เป็นระบบ

1. เครื่องมือการออกแบบและพัฒนาแบบจำลองการจัดความรู้ให้เป็นระบบ

1.1 โปรเทเจ้3.48 ใช้สำหรับออกแบบ และพัฒนาฐานความรู้โดยออนโทโลยี ซึ่งเก็บอยู่ในรูปแบบภาษาโอดับเบิ้ลยูแอล

1.2 เอสเคโอเอส ใช้สำหรับสร้างคลังคำ

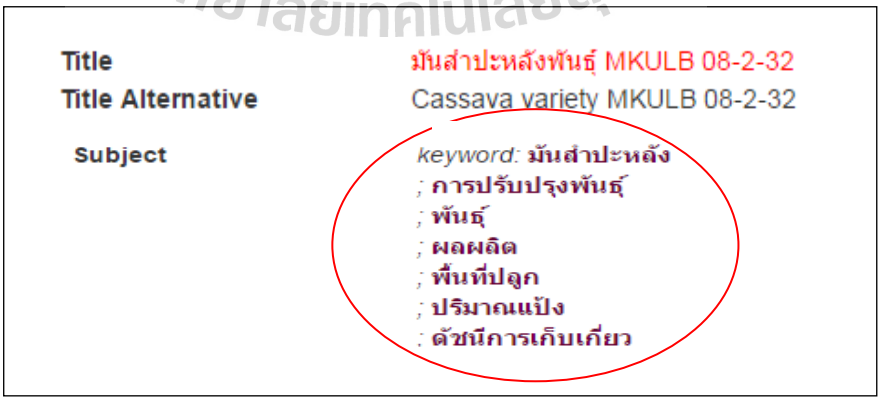
1.3 ดับลินคอร์ใช้เป็นเมทาดาเมทาดาทา(Metadata) สำหรับ การอธิบายให้ทราบรายละเอียดของข้อมูลถึงความรู้ด้านพืชไร่

2. กรอบแนวคิดขั้นตอนการจัดความรู้ให้เป็นระบบ

กรอบแนวคิดการจัดความรู้ให้เป็นระบบในรูปที่ 3.2 มี 6 ขั้นตอน ได้แก่ (1) การประยุกต์ใช้ออนโทโลยีไม้ผลเศรษฐกิจ(2) การเลือกคำสำคัญจากความรู้ด้านพืชไร่(3) การกำหนดความสัมพันธ์ของคำสำคัญ(4) การสร้างคลังคำ(5) การอธิบายรายละเอียดความรู้ด้านพืชไร่และ 6) การประมวลผลความรู้ด้านพืชไร่

ขั้นตอนที่ 1 การประยุกต์ใช้ออนโทโลยีไม้ผลเศรษฐกิจ (ThaiFO Ontology) เนื่องจากออนโทโลยีได้ออกแบบโครงสร้างคลาส จำนวน 8 คลาส (ศุภกฤษฎิ์ นิวัฒนากูล, 2556) ที่สามารถประยุกต์ใช้งานร่วมกับการจัดการความรู้พืชไร่ได้โดยเฉพาะคลาสแทนข้อมูลเอกสารความรู้ (Documents)

ขั้นตอนที่ 2การเลือกคำสำคัญจากความรู้ด้านพืชไร่ ได้เลือกคำสำคัญจากการกำหนดไว้ในความรู้ด้านพืชไร่ ดังรูปที่ 3.3

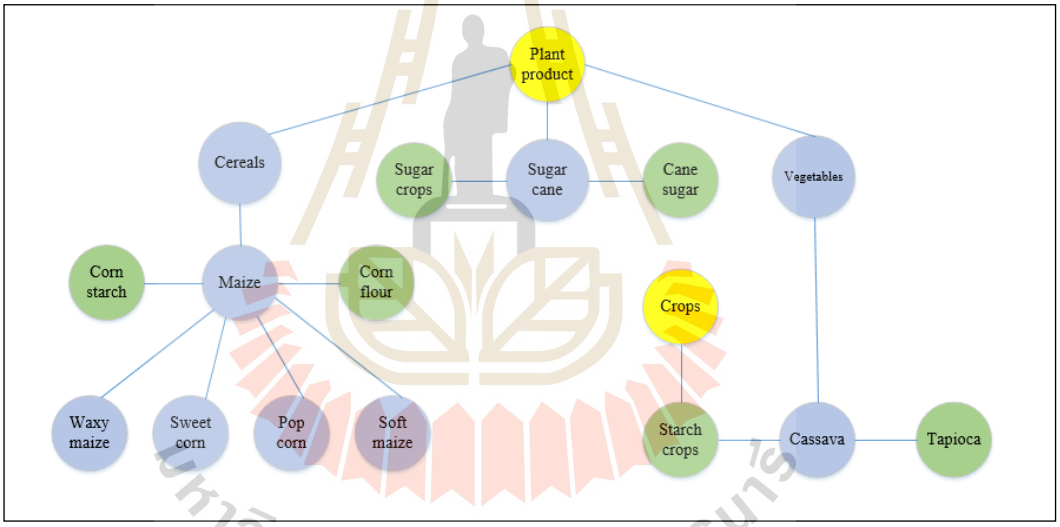


รูปที่ 3.3คำสำคัญที่เลือกจากความรู้ด้านพืชไร่

มหาวิทยาลัยเทคโนโลยีสุรนารี

จากรูปที่ 3.3เป็นตัวอย่างคำสำคัญที่เลือก ซึ่งงานวิจัยนี้ได้เลือกจากความรู้ด้านพืชไร่ จำนวน 240 รายการได้คำสำคัญ จำนวน735 คำ และคัดเลือกคำสำคัญเพื่อนำไปสร้างและจัดเก็บในคลังคำ โดยใช้เกณฑ์ คือ (1) คำสำคัญนั้นต้องปรากฏความถี่ในคำสำคัญของความรู้ตั้งแต่ 3คำขึ้นไป และ (2) คำสำคัญนั้นต้องปรากฏการกำหนดความสัมพันธ์ของคำในอภิธานศัพท์เกษตรไทย ซึ่งขั้นตอนนี้ได้คำสำคัญผ่านเกณฑ์ที่กำหนดไว้ข้างต้น จำนวน 181 คำ

ขั้นตอนที่ 3การกำหนดความสัมพันธ์ของคำสำคัญ ซึ่งคำสำคัญที่ได้ในขั้นตอนที่ 2นำไปสืบค้นในอภิธานศัพท์เกษตรไทยเพื่อดูความสัมพันธ์ของคำหลัก คำกว้างกว่า คำแคบกว่า คำที่เกี่ยวข้อง และคำเหมือน เพื่อนำไปสร้างในคลังคำตัวอย่างความสัมพันธ์ของคำสำคัญดังรูปที่ 3.4



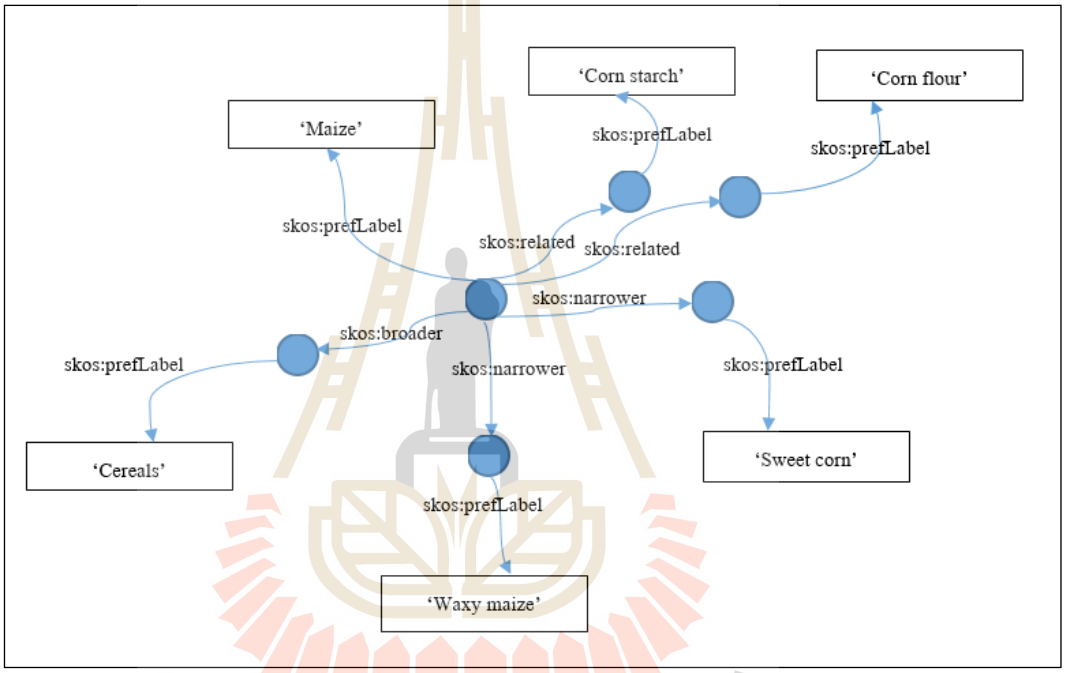
ที่มา : อภิธานศัพท์เกษตรไทย (http://pikul.lib.ku.ac.th/agrovoc/agrovoc_th_adv.php)

รูปที่3.4ตัวอย่างโครงสร้างความสัมพันธ์ของคำในความรู้ด้านพืชไร่

จากรูปที่ 3.4ตัวอย่างโครงสร้างความสัมพันธ์ของคำ “ผลิตผลจากพืช (Plant Product)”เป็นคำกว้างกว่า (Skos:Broader) มีธัญพืช (Cereals) อ้อย (Sugarcane) และผัก (Vegetable) เป็นคำแคบกว่า (Skos:Narrower) โดยธัญพืช (Cereals) มีข้าวโพด (Maize) เป็นคำแคบกว่า โดยมีคำที่เกี่ยวข้อง(Skos:Related) กับข้าวโพด คือ สตาร์ข้าวโพด (Corn starch) และ แป้งข้าวโพด (Corn Flour) ซึ่งข้าวโพด (Maize) มีข้าวโพดข้าวเหนียว (Waxy Maize) ข้าวโพดหวาน (Sweet Corn)ข้าวโพดคั่ว (Pop Corn)และ ข้าวโพดแป้งอ่อน(Soft Maize) เป็นคำแคบกว่า ส่วนอ้อย

(Sugarcane) มีพืชน้ำตาล (Sugar Crops) และ น้ำตาลอ้อย (Cane Sugar) เป็นคำที่เกี่ยวข้อง ส่วนผัก (Vegetable) มีมันสำปะหลัง (Cassava) เป็นคำแคบกว่า โดยมีคำที่เกี่ยวข้อง คือแป้งมันสำปะหลัง (Tapioca) และพืชให้แป้ง (Starch Crops) เป็นคำที่เกี่ยวข้อง โดยมีพืชปลูก (Crops) เป็นคำกว้างกว่า

ขั้นตอนที่ 4การสร้างคลังคำ สร้างจากความสัมพันธ์ของคำสำคัญที่ได้ในขั้นตอนที่ 3นำมานิยามความสัมพันธ์ของคำศัพท์ที่เกี่ยวข้องกับความรู้ด้านพืชไร่ และสร้างคลังคำโดยใช้ตัวแบบการแทนความรู้เอสเคโอเอส ดังรูปที่ 3.5



รูปที่ 3.5 ตัวอย่างการนิยามศัพท์ด้านพืชไร่ในเอสเคโอเอส

ขั้นตอนที่ 5การอธิบายรายละเอียดความรู้ด้านพืชไร่ อธิบายความรู้ด้วยเมทาดาทา โดยใช้ชุดหน่วยข้อมูลย่อยดับลินคอร์ 4 อิลีเมนต์ ได้แก่ ชื่อเรื่อง คำสำคัญ คำอธิบายความรู้ที่อธิบายความรู้เฉพาะบทคัดย่อ และแหล่งอ้างอิงต้นฉบับ สอดคล้องกับบุญเลิศอรุณพิบูลย์ , สุภาพรชัยธัมมะปกรณ์และฐิติมาธรรมบำรุง (2550) ได้อธิบายไว้ว่าอธิบายเกี่ยวกับทรัพยากรสารสนเทศใด ๆ ควรใช้ชุดหน่วยข้อมูลย่อยตั้งแต่ 4 อิลีเมนต์ขึ้นไป

ขั้นตอนที่ 6การประมวลผลความรู้ด้านพืชไร่ เป็นขั้นตอนการคำนวณเก็บไว้ในรูปแบบของไฟล์ข้อความ ประกอบด้วย(1) การคำนวณความถี่ของคำที่ปรากฏในความรู้พืชไร่ (2) การคำนวณน้ำหนักคำด้วยการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ตามแนวคิดของดีคอฟฟ์

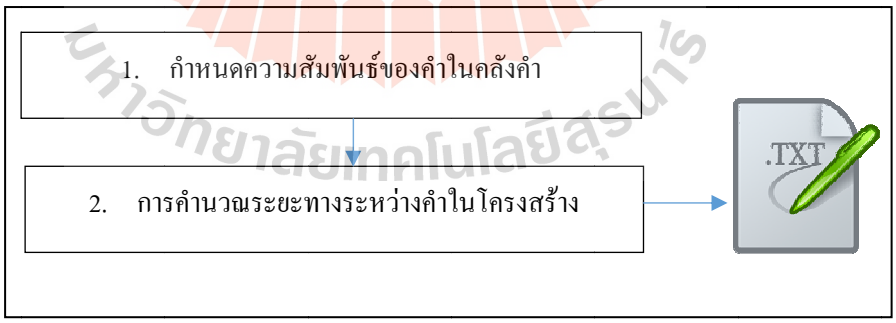
(Diekhoff, 1992) (3) การคำนวณน้ำหนักคำด้วยการวิเคราะห์ IDFและ (4) การคำนวณระยะทางสั้นที่สุดระหว่างคำในโครงสร้าง

- การคำนวณความถี่ของคำที่ปรากฏในความรู้ คือ การนับความถี่ของคำสำคัญที่ปรากฏในความรู้ด้านพืชไร่จำนวน 240 รายการ
- การคำนวณน้ำหนักคำในโครงสร้างงานวิจัยนี้คำนวณน้ำหนักคำ 2 วิธี ดังนี้

1) วิเคราะห์สัมประสิทธิ์สหสัมพันธ์จากแนวคิดของดีคอฟฟ์ (Diekhoff, 1992) คือ การวิเคราะห์หาความสัมพันธ์ระหว่างคำสองคำในโครงสร้างว่ามีความสัมพันธ์กันมากน้อยเพียงใด ดังสมการที่ 2.7ซึ่งการพิจารณาค่าสัมประสิทธ์สหสัมพันธ์ใช้เกณฑ์ของฮินเคอล์ (Hinkle, 1998 : 118)

(2) การวิเคราะห์ IDF คือวิเคราะห์ค่าสัดส่วนของจำนวนเอกสารทั้งหมดกับจำนวนของเอกสารที่ปรากฏคำที่สนใจซึ่งไม่ใช้ TF-IDF เพราะ TF แสดงถึงความสำคัญของคำสำคัญใด ๆ ที่มีความสำคัญต่อความรู้หรือเอกสารหนึ่ง ๆ ส่วน IDF จะแสดงถึงความสำคัญของคำต่อการสืบค้นเอกสารหนึ่ง ๆ เมื่อเทียบกับทุกเอกสาร และงานวิจัยนี้ต้องการคำนวณค่าน้ำหนักของคำเพื่อแสดงให้เห็นว่าคำนั้นสำคัญต่อการสืบค้นความรู้มากน้อยเท่าใด โดยไม่ได้เฉพาะเจาะจงต่อความรู้หรือเอกสารหนึ่ง ๆ จึงไม่มีความจำเป็นที่ต้องใช้ค่า TF โดยใช้สมการที่ 2.5

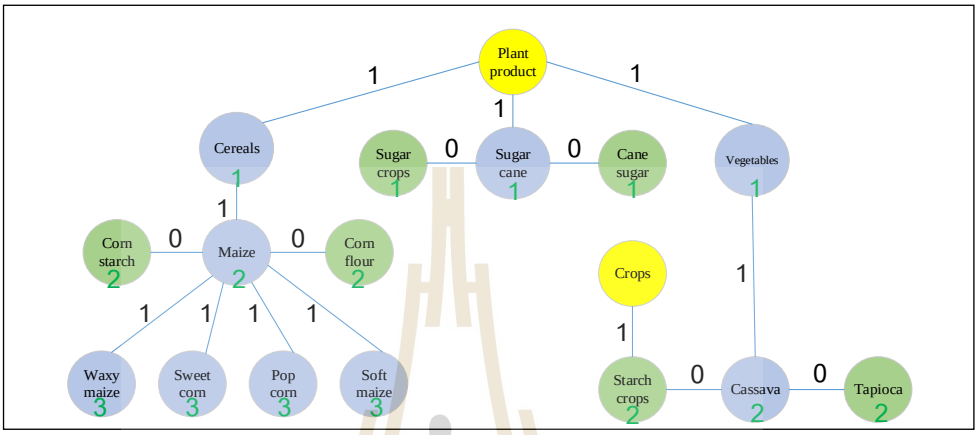
- การคำนวณระยะทางสั้นที่สุดระหว่างคำในโครงสร้าง คือ การหาระยะทางจากเส้นเชื่อมระหว่างโหนดที่มีผลรวมระยะทางน้อยที่สุดระหว่างคำในโครงสร้าง ซึ่งมีขั้นตอนการคำนวณดังรูปที่ 3.6



รูปที่ 3.6กรอบแนวคิดการคำนวณระทางสั้นที่สุดระหว่างคำในโครงสร้าง

จากรูปที่ 3.6 การคำนวณระยะทางสั้นที่สุดระหว่างคำในโครงสร้างมี 2 ขั้นตอน คือ กำหนดความสัมพันธ์ของคำในคลังคำ และคำนวณระยะทางระหว่างคำในโครงสร้างโดยเก็บในรูปแบบไฟล์ข้อความ ดังนี้

1. กำหนดความสัมพันธ์ของคำในคลังคำ โดยกำหนดคำหลัก (skos: preLabel) คำกว้างกว่า (Skos: Broader) คำแคบกว่า (Skos: Narrower) มีค่าเท่ากับ 1(พิลาวัณย์พลับรู้การ และกฤษณะ ไวยมัย, 2544) และคำเกี่ยวข้อง (Related Term: RT)คำเหมือน (Skos:altLabel) มีค่าเท่ากับ 0 แสดงตัวอย่างดังรูป3.7



รูปที่ 3.7ตัวอย่างการกำหนดความสัมพันธ์ของคำในคลังคำ

2. การคำนวณระยะทางระหว่างคำในโครงสร้าง คำณวณจากขั้นตอนวิธีของไดก์สตรา (Dijkstra, E., 1959) ดังนี้

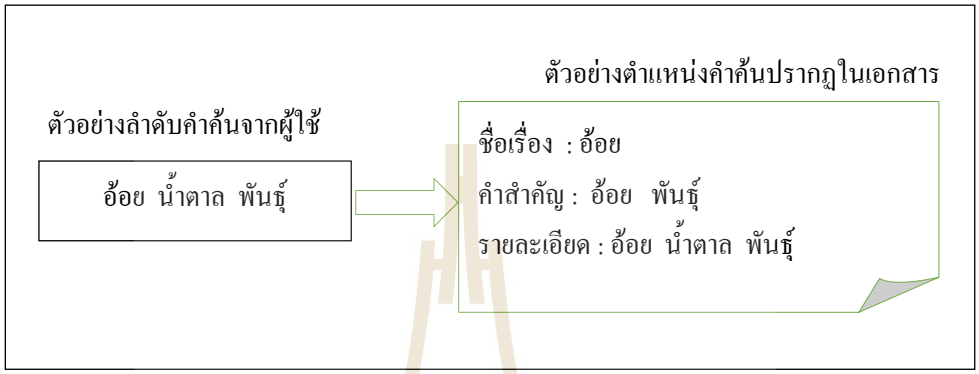
มหาวิทยาลัยเทคโนโลยีสุรนารี

3.2.2 การออกแบบและพัฒนากระบวนการสืบค้นเชิงความหมาย

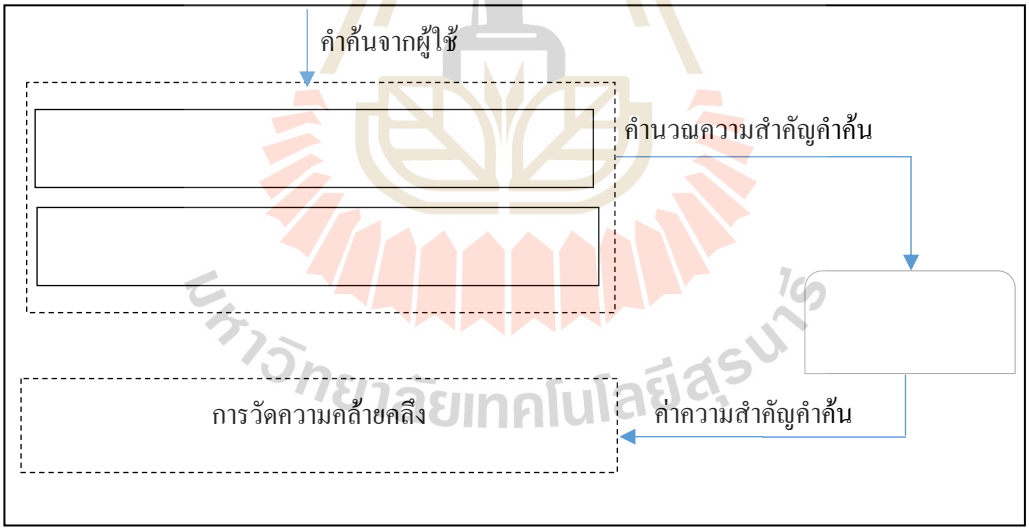
กรอบแนวคิดการออกแบบและพัฒนาแบบจำลองการสืบค้นเชิงความหมาย งานวิจัยนี้มี 3 ขั้นตอน ได้แก่ การกำหนดลำดับความสำคัญคำค้นการให้น้ำหนักคำสำคัญ และการวัดความคล้ายคลึง ดังนี้

1. การกำหนดลำดับความสำคัญของคำค้น

การกำหนดลำดับความสำคัญคำค้นงานวิจัยนี้ให้ความสำคัญกับคำค้น 2 ลักษณะ ได้แก่ลำดับคำค้นจากผู้ใช้ (Sequence of Words: SW) และตำแหน่งของคำค้นที่ปรากฏในความรู้ (Position of Words: PW)ตัวอย่างดังรูปที่ 3.8 และมีขั้นตอนดังรูปที่ 3.9



รูปที่ 3.8ตัวอย่างลำดับคำค้นจากผู้ใช้และตำแหน่งคำค้นที่ปรากฏในความรู้



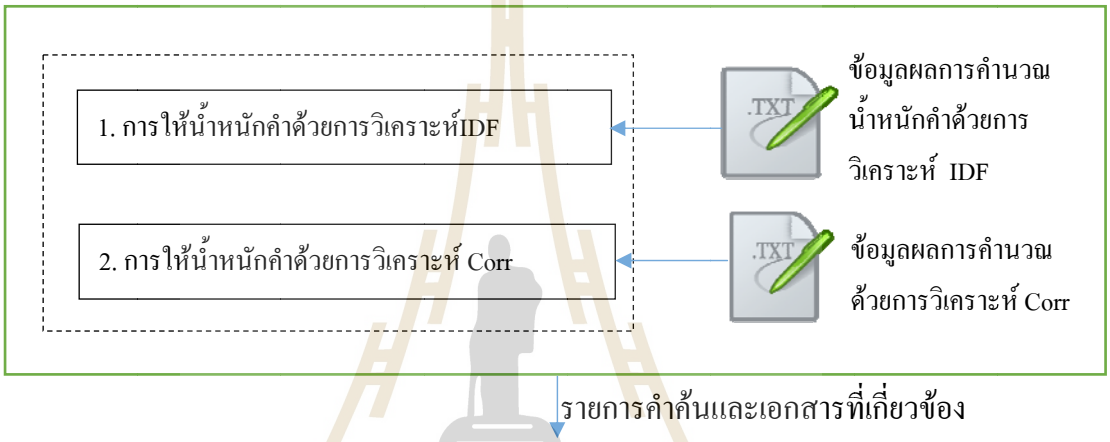
รูปที่ 3.9กรอบแนวคิดขั้นตอนการกำหนดลำดับความสำคัญคำค้น

1.1 ลำดับคำค้นจากผู้ใช้คือการให้ความสำคัญกับลำดับคำค้นที่12 และ 3 ตามลำดับคำค้นจากผู้ใช้ที่ระบุเข้ามาเพื่อการสืบค้นเช่น เมื่อผู้ใช้ระบุคำค้นเข้ามา3 คำ ดังนั้นค่าความสำคัญจากลำดับคำค้น ที่ 12และ 3 ตามลำดับ

1.2 ตำแหน่งของคำค้นที่ปรากฏในความรู้คือการให้ความสำคัญกับการตำแหน่งของคำค้นที่ปรากฏในส่วนต่าง ๆ ของความรู้ที่ผู้ใช้ระบุเข้ามา ได้แก่ ชื่อเรื่อง (Title)

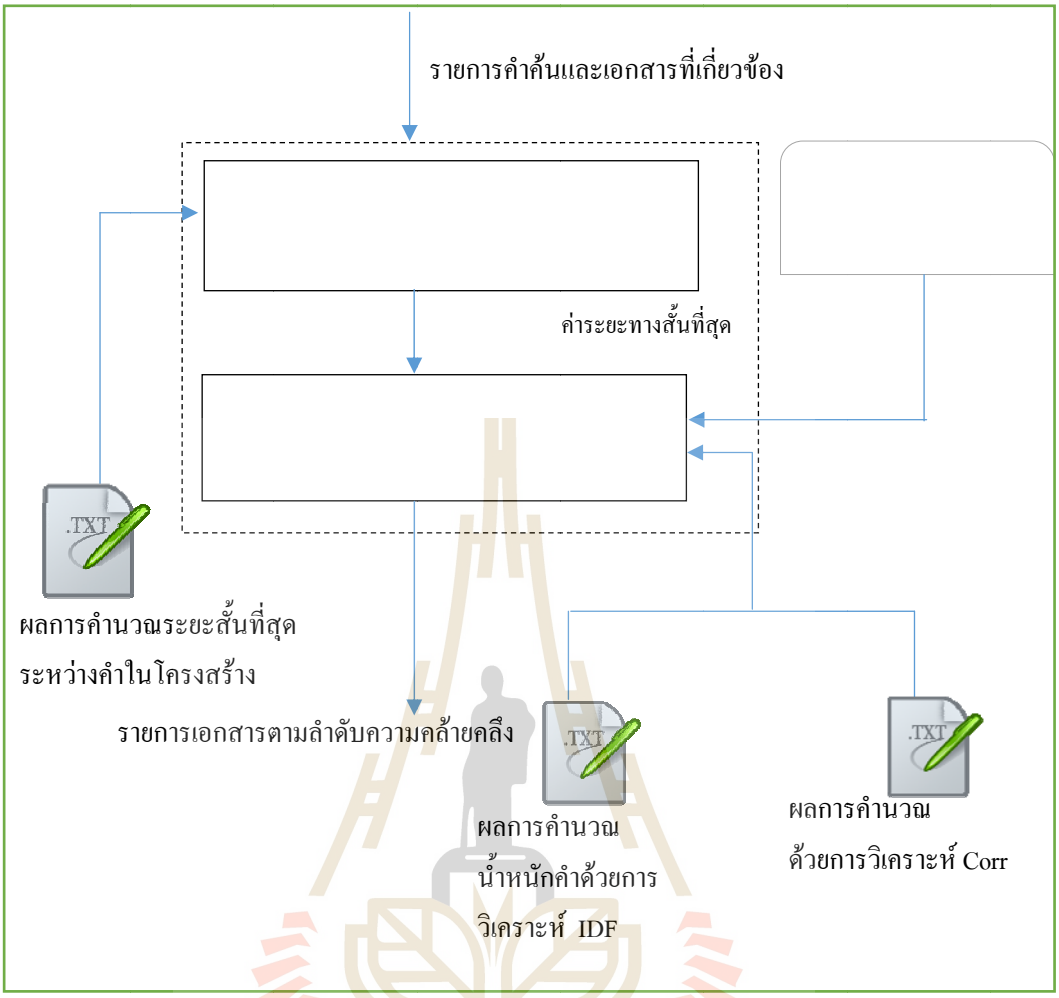
มีความสำคัญเท่ากับ 1 คำสำคัญ (Keywords/Subject) เท่ากับ 2 และรายละเอียด (Description/Abstract)เท่ากับ 3ตามลำดับ

2. การให้น้ำหนักคำสำคัญคือการนำผลการคำนวณน้ำหนักคำในขั้นตอนที่ 6การประมวลผลความรู้ด้านพืชไร่ ซึ่งเป็นขั้นตอนหนึ่งของการจัดความรู้ให้เป็นระบบที่เก็บไว้ในรูปแบบไฟล์ข้อความมาใช้ โดยคำนวณ(1) น้ำหนักคำด้วยการวิเคราะห์ค่าสัดส่วนของจำนวนเอกสารทั้งหมดกับจำนวนของเอกสารที่ปรากฏคำที่สนใจ (IDF) และ (2) น้ำหนักคำด้วยการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ (Corr) ดังรูปที่3.10



รูปที่ 3.10กรอบแนวคิดขั้นตอนการให้น้ำหนักคำสำคัญ

3. การวัดความคล้ายคลึงคือการวัดความคล้ายคลึงระหว่างเอกสารกับคำค้นประกอบด้วย 2 ขั้นตอน ได้แก่ ได้แก่ (1) การคำนวณความคล้ายคลึงระหว่างคำค้นกับคำในความรู้และ (2) การคำนวณความคล้ายคลึงระหว่างความรู้กับคำค้น ซึ่งกระบวนการวัดความคล้ายคลึงนี้จะใช้ข้อมูลจากผลการคำนวณ(1) ระยะทางสั้นที่สุดระหว่างคำในโครงสร้าง(2) การให้น้ำหนักคำด้วยการวิเคราะห์ IDF และ (3) การให้น้ำหนักคำด้วยการวิเคราะห์ Corr ที่มีการคำนวณเก็บไว้ในรูปแบบของไฟล์ข้อความ จากกระบวนการจัดความรู้ให้เป็นระบบ ขั้นตอนที่ 6 การประมวลผลความรู้ด้านพืชไร่ ดังรูปที่ 3.11



รูปที่ 3.11กรอบแนวคิดขั้นตอนการวัดความคล้ายคลึง

จากรูปที่ 3.11 การวัดความคล้ายคลึง มี 2ขั้นตอน ได้แก่ (1) การคำนวณระยะทางสั้นที่สุดระหว่างคำค้นกับคำในความรู้หรือเอกสารทั้งหมด และ(2) การคำนวณความคล้ายคลึงระหว่างความรู้กับคำค้น ดังนี้

3.1 นำผลการคำนวณระยะทางสั้นที่สุดระหว่างคำในโครงสร้างที่เก็บในรูปแบบไฟล์ข้อความมาคำนวณค่าระยะทางระหว่างคำค้นกับคำในความรู้ทั้งหมด โดยใช้สมการที่ 3.5ซึ่งประยุกต์มาจากแนวคิดของศุภกฤษฎิ์ นิวัฒนากูล (2556)

$$Dis_i = \exp\left(-\min_dis(d_i, q_j)\right) \quad (3.1)$$

เมื่อ

Dis_i แทน ค่าระยะทางสั้นที่สุดระหว่างคำค้นกับคำในความรู้หรือเอกสาร

$\min_dis(d_i, q_j)$ แทน ค่าระยะทางน้อยที่สุดในกลุ่มระหว่างคำค้นกับคำในเอกสาร d_i กับคำค้น q_j

3.2 จากการคำนวณ3.1จะได้ค่าระยะทางสั้นที่สุดระหว่างคำค้นกับคำในความรู้ทั้งหมด โดยค่าดังกล่าวนำมาคำนวณความคล้ายคลึงระหว่างความรู้กับคำค้น ซึ่งงานวิจัยนี้ได้คำนวณตามรูปแบบการทดลอง ดังตารางที่ 3.1

ตารางที่ 3.1 รูปแบบการทดลอง

รูปแบบการทดลอง	**การกำหนดลำดับความสำคัญ**		**วิธีการให้น้ำหนักคำ**		**การวัดความคล้ายคลึง**
	ลำดับคำค้นจากผู้ใช้ (SW)	**ตำแหน่งคำที่ปรากฏในเอกสาร (PW)**	**Corr**	**IDF**	**ระยะทางระหว่างคำในโครงสร้าง**
A การให้น้ำหนักคำสำคัญด้วยการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์					
A1			✓		✓
A2	✓		✓		✓
A3		✓	✓		✓
A4	✓	✓	✓		✓
B การให้น้ำหนักคำสำคัญด้วยการวิเคราะห์สัดส่วนของจำนวนเอกสารทั้งหมดกับจำนวนของเอกสารที่ปรากฏคำที่สนใจ					
B1				✓	✓
B2	✓			✓	✓
B3		✓		✓	✓
B4	✓	✓		✓	✓

**Corr หมายถึง การวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient)*

IDF หมายถึง การวิเคราะห์สัดส่วนของจำนวนความรู้ทั้งหมดกับจำนวนของความรู้ที่ปรากฏคำที่สนใจ (Inverse document frequency)

SW หมายถึง ลำดับคำค้นจากผู้ใช้ (Sequence of words)

PW หมายถึงตำแหน่งของคำค้นที่ปรากฏในเอกสาร (Position of words)

รูปแบบการทดลอง A คือนำผลการคำนวณน้ำหนักคำสำคัญด้วยการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ (Corr)มาใช้สำหรับการวัดความคล้ายคลึง4ลักษณะดังนี้

A1 คือการนำค่าระยะทางสั้นที่สุดระหว่างคำค้นกับคำในความรู้หรือเอกสาร (Dis_i)มาคำนวณกับค่าน้ำหนักคำสำคัญด้วยวิเคราะห์สัมประสิทธิ์สหสัมพันธ์

A2 คือ การนำค่าระยะทางสั้นที่สุดระหว่างคำค้นกับคำในความรู้หรือเอกสาร (Dis_i)มาคำนวณกับค่าน้ำหนักคำสำคัญด้วยวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ และค่าลำดับความสำคัญคำค้นจากผู้ใช้ (SW)

A3 คือ การนำค่าระยะทางสั้นที่สุดระหว่างคำค้นกับคำในความรู้หรือเอกสาร (Dis_i)มาคำนวณกับค่าน้ำหนักคำสำคัญด้วยวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ และค่าลำดับความสำคัญคำค้นจากตำแหน่งของคำค้นที่ปรากฏในความรู้ (PW)

A4 คือการนำค่าระยะทางสั้นที่สุดระหว่างคำค้นกับคำในความรู้หรือเอกสาร (Dis_i)มาคำนวณกับค่าน้ำหนักคำสำคัญด้วยวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ และค่าลำดับความสำคัญคำค้นจากผู้ใช้ และค่าลำดับความสำคัญคำค้นจากตำแหน่งของคำค้นที่ปรากฏในความรู้

รูปแบบการทดลอง Bคือนำผลการคำนวณน้ำหนักคำสำคัญด้วยการวิเคราะห์สัดส่วนของจำนวนเอกสารทั้งหมดกับจำนวนของเอกสารที่ปรากฏคำที่สนใจ (IDF) มี 4 ลักษณะดังนี้

B1 คือการนำค่าระยะทางสั้นที่สุดระหว่างคำค้นกับคำในความรู้หรือเอกสาร (Dis_i)มาคำนวณกับสัดส่วนของจำนวนเอกสารทั้งหมดกับจำนวนของเอกสารที่ปรากฏคำที่สนใจ

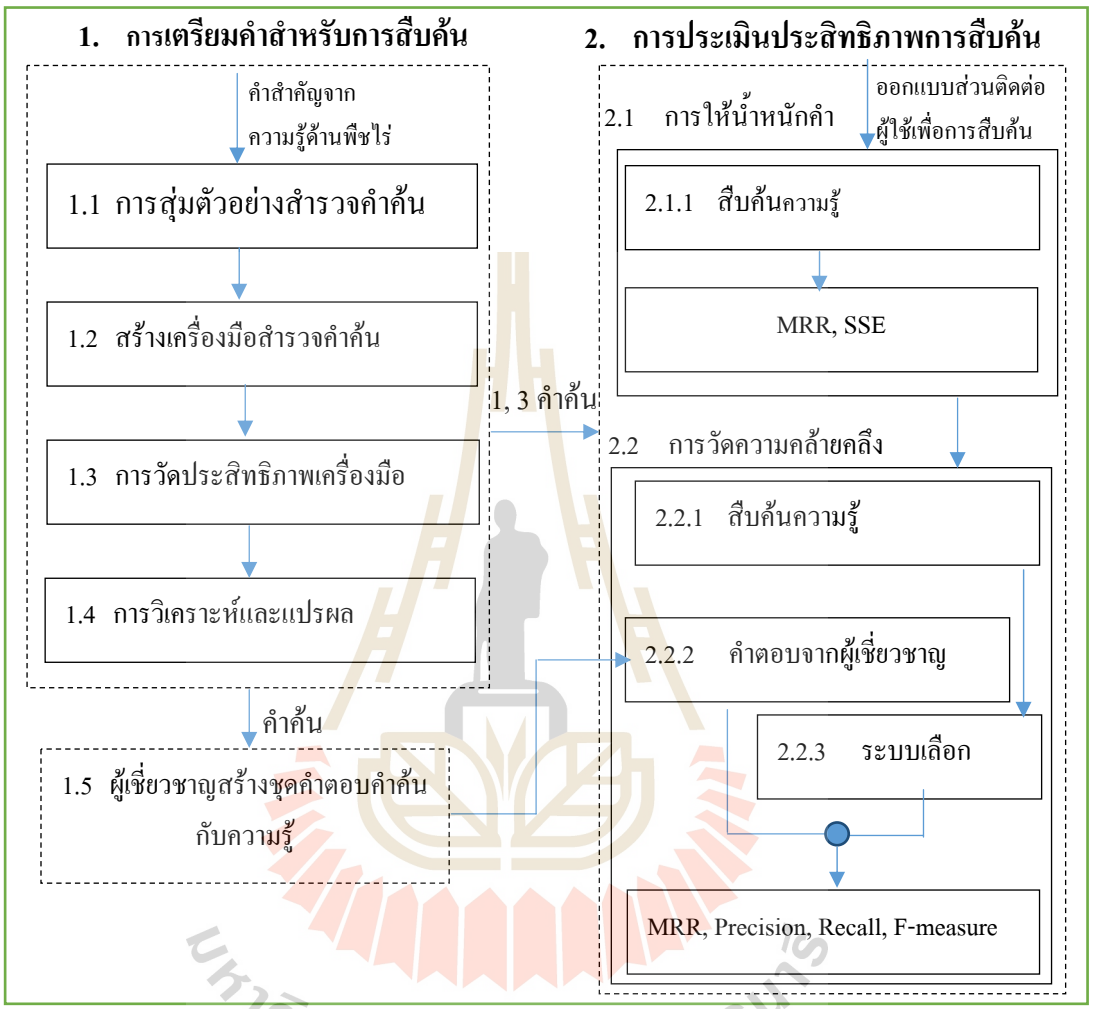
B2 คือการนำค่าระยะทางสั้นที่สุดระหว่างคำค้นกับคำในความรู้หรือเอกสาร (Dis_i)มาคำนวณกับสัดส่วนของจำนวนเอกสารทั้งหมดกับจำนวนของเอกสารที่ปรากฏคำที่สนใจ และค่าลำดับความสำคัญคำค้นจากผู้ใช้ (SW)

B3 คือการนำค่าระยะทางสั้นที่สุดระหว่างคำค้นกับคำในความรู้หรือเอกสาร (Dis_i)มาคำนวณกับสัดส่วนของจำนวนเอกสารทั้งหมดกับจำนวนของเอกสารที่ปรากฏคำที่สนใจ และค่าลำดับความสำคัญคำค้นจากตำแหน่งของคำค้นที่ปรากฏในความรู้ (PW)

B4 คือการนำค่าระยะทางสั้นที่สุดระหว่างคำค้นกับคำในความรู้หรือเอกสาร (Dis_i)มาคำนวณกับสัดส่วนของจำนวนเอกสารทั้งหมดกับจำนวนของเอกสารที่ปรากฏคำที่สนใจ และค่าลำดับความสำคัญคำค้นจากผู้ใช้ และค่าลำดับความสำคัญคำค้นจากตำแหน่งของคำค้นที่ปรากฏในความรู้

3.3 การประเมินประสิทธิภาพการสืบค้นเชิงความหมาย

การประเมินการสืบค้นเชิงความหมายสำหรับความรู้ด้านพืชไร่ ดังรูปที่ 3.12



รูปที่ 3.12กระบวนการประเมินประสิทธิภาพการสืบค้น

3.3.1 การเตรียมคำสำหรับการสืบค้น

การเตรียมคำค้นเป็นกระบวนการวิเคราะห์เลือกคำที่ต้องการเพื่อใช้สำหรับประเมินประสิทธิภาพการสืบค้น งานวิจัยนี้เป็นการวิจัยเชิงสำรวจ รวบรวมข้อมูลโดยการใช้แบบสอบถาม

1. กลุ่มตัวอย่างสำรวจคำค้นความรู้ด้านพืชไร่

กลุ่มตัวอย่างที่ใช้ในการสำรวจคำค้นงานวิจัยครั้งนี้ ได้แก่ประชาชนเขตอำเภอเมืองกาฬสินธุ์ จำนวน 384คน โดยการสุ่มตัวอย่างแบบบังเอิญดังสูตรที่ 3.2

$$n = \frac{P(1-P)(Z)^2}{e^2} \quad (3.2)$$

เมื่อ n แทน ขนาดของกลุ่มตัวอย่าง
P แทน ค่าเปอร์เซ็นที่ต้องการจะสุ่มจากประชากรทั้งหมด
e แทน ค่าเปอร์เซ็นต์ความคาดเคลื่อนของการสุ่มตัวอย่าง
Z แทน ระดับความเชื่อมั่น

$$n = \frac{(.50)(1-.50)(1.96)^2}{.05^2}$$

$$n = 384$$

2. เครื่องมือในการสำรวจคำค้นความรู้ด้านพืชไร่

เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูลจากการสำรวจคำค้นที่ต้องการใช้เพื่อการสืบค้นในงานวิจัยนี้ ได้แก่ แบบสอบถาม (Questionnaire) ประกอบด้วย

ตอนที่1 ข้อมูลพื้นฐานของผู้ตอบแบบสอบถามเพื่อใช้ในการอ้างอิงแหล่งข้อมูล ได้แก่ เพศ อายุ และอาชีพ เป็นคำถามแบบเลือกตอบ (Check List) ซึ่งเลือกได้คำตอบเดียว

ตอนที่ 2 ข้อมูลคำค้นที่ผู้ตอบแบบสอบถามต้องการใช้สำหรับสืบค้นความรู้ด้านพืชไร่ที่เกี่ยวข้องกับมันสำปะหลัง อ้อย และข้าวโพด ซึ่งเป็นคำที่ได้มาจากขั้นตอนที่ 2 การเลือกคำสำคัญจากความรู้ด้านพืชไร่ในการออกแบบและพัฒนาแบบจำลองการจัดความรู้ให้เป็นระบบจากความรู้ด้านพืชไร่ จำนวน 240 รายการ ซึ่งได้คำสำคัญ จำนวน 735 คำสำคัญ และคัดเลือกนำไปใช้ในแบบสอบถามจำนวน 181 คำเป็นคำถามแบบเลือกตอบ (Check List)ซึ่งสามารถเลือกได้หลายคำตอบ

ตอนที่ 3 ข้อเสนอแนะสำหรับคำค้นอื่น ๆ ที่เกี่ยวข้องกับความรู้พืชไร่ ซึ่งเป็นคำถามแบบปลายเปิด

3. การวัดประสิทธิภาพเครื่องมือในการสำรวจคำค้นความรู้ด้านพืชไร่

การหาประสิทธิภาพของแบบสอบถาม งานวิจัยนี้ได้ทดสอบคุณภาพของเครื่องมือก่อนนำไปใช้จริง คือการทดสอบความเที่ยงตรงของเนื้อหา (Validity) โดยนำแบบสอบถามไปให้ผู้เชี่ยวชาญพิจารณาในด้านความเที่ยงตรง ความเหมาะสมของเนื้อหา ความเหมาะสมในการ

ใช้ภาษา และความชัดเจนของคำถาม เพื่อให้แบบสอบถามสามารถสื่อความหมายได้ตรงตามความประสงค์ของงานวิจัย ซึ่งผู้เชี่ยวชาญสำหรับพิจารณาความเที่ยงตรงมีจำนวน3คน ประกอบด้วย

- รองศาสตราจารย์ ดร.สายตา บุญโฉม อาจารย์ประจำคณะบริหารธุรกิจและเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน (ผู้เชี่ยวชาญด้านงานวิจัยและสถิติ)
- ผู้ช่วยศาสตราจารย์ ดร.เนติรัฐ วีระนาคินทร์ อาจารย์ประจำคณะวิทยาการเทคโนโลยีสารสนเทศ มหาวิทยาลัยมหาสารคาม (ผู้เชี่ยวชาญด้านงานวิจัยและสถิติ)
- ผู้ช่วยศาตราจารย์ ดร.สถิรวงค์ แจ่มจรรยา อาจารย์ประจำคณะเทคโนโลยีอุตสาหกรรมเกษตร มหาวิทยาลัยกาฬสินธุ์ (ผู้เชี่ยวชาญด้านความรู้พืชไร่)

โดยการให้คะแนนรายข้อของผู้เชี่ยวชาญมี 3 ค่า คือ

ค่า +1 แน่ใจว่าข้อคำถามนั้นสามารถใช้วัดค่าตัวแปรที่ศึกษาได้

ค่า 0 ไม่แน่ใจว่าข้อคำถามนั้นสามารถใช้วัดค่าตัวแปรที่ศึกษาได้

ค่า -1 แน่ใจว่าข้อคำถามนั้นไม่สามารถใช้วัดค่าตัวแปรที่ศึกษาได้

การคำนวณค่าดัชนีความสอดคล้อง (Index of Consistency: IOC) โดยเลือกเฉพาะข้อคำถามที่มีค่าดัชนีความสอดคล้องตั้งแต่ 0.66 ขึ้นไป เพื่อให้ได้ข้อคำถามที่มีคุณภาพ ซึ่งการหาค่าความเที่ยงตรงเชิงเนื้อหาของแบบทดสอบแต่ละข้อ โดยใช้สูตร (สมนึก ภัททิยธนี ,2546: 167) ดังนี้

$$IOC = \frac{\sum_{i=1}^{n} R_i}{N} \qquad (3.15)$$

เมื่อ

IOC แทน ดัชนีความสอดคล้องระหว่างจุดประสงค์กับเนื้อหา หรือระหว่างเนื้อหากับจุดประสงค์

R_i แทน คะแนนความคิดเห็นของผู้เชี่ยวชาญ

N แทน จำนวนผู้เชี่ยวชาญทั้งหมด

มหาวิทยาลัยเทคโนโลยีสุรนารี

ผลการทดสอบความเที่ยงตรงเชิงเนื้อหาของแบบสอบถามในงานวิจัยนี้มีค่าเฉลี่ย IOC เท่ากับ 0.91

4. การวิเคราะห์และแปรผลข้อมูลการสำรวจคำค้นความรู้ด้านพืชไร่

การวิเคราะห์ข้อมูลจากแบบสอบถามการสำรวจคำค้นความรู้พืชไร่ วิเคราะห์โดยใช้โปรแกรมสำเร็จรูปทางคอมพิวเตอร์ และนำเสนอในรูปแบบตารางประกอบการแปลความหมายเ พื่ออธิบายข้อมูลในด้านต่าง ๆ

ตอนที่ 1ข้อมูลพื้นฐานของผู้ตอบแบบสอบถามเพื่อใช้ในการอ้างอิงแหล่งข้อมูล ได้แก่ เพศ อายุ และอาชีพ

ตอนที่ 2ข้อมูลคำค้นที่ผู้ตอบแบบสอบถามต้องการใช้สำหรับสืบค้นความรู้ด้านพืชไร่ที่เกี่ยวข้องกับมันสำปะหลัง อ้อย และข้าวโพด

ตอนที่ 3 ข้อเสนอแนะสำหรับคำค้นอื่น ๆ ที่เกี่ยวข้องกับความรู้พืชไร่ ซึ่งเป็นคำถามแบบปลายเปิด

การวิเคราะห์ข้อมูลโดยการแจกแจงความถี่ (Frequencies) การหาค่าสัดส่วนหรือร้อยละ (Percentage)

5. การสร้างชุดคำตอบ

การสร้างชุดคำตอบ (เฉลย) งานวิจัยนี้ใช้คำค้นจากการวิเคราะห์และแปรผลข้อมูลที่ได้ค่าร้อยละ 25ขึ้นไปมาใช้เป็นคำค้น โดยผู้เชี่ยวชาญเลือกและจับคู่1-3 คำค้นเพื่อนำไปใช้สำหรับสืบค้นดังนี้

5.1 ผู้เชี่ยวชาญเลือกคำตอบความรู้ หรือเอกสารใด ๆ ที่ตรงหรือเกี่ยวข้องกับคำค้น โดยเรียงลำดับความคล้ายคลึงจากมากไปหาน้อย ซึ่งการเลือกคำตอบหรือความรู้ที่ถูกต้องกับคำค้น ผู้เชี่ยวชาญต้องเลือก 2ใน 3คน จึงถือว่าเป็นคำตอบที่ถูกต้อง และกรณีที่ผู้เชี่ยวชาญมีคำตอบไม่ตรงกันหรือไม่ชัดเจนจะมีการจัดลำดับคำตอบให้เลือกเพื่อให้ได้คำตอบกับคำค้นนั้น ๆ

5.2 ระบบเลือกคำตอบความรู้ หรือเอกสารใด ๆ ที่ตรงหรือเกี่ยวข้องกับคำค้น โดยเรียงลำดับความคล้ายคลึงจากมากไปหาน้อย ซึ่งเกณฑ์การเลือกความรู้ หรือเอกสารใด จากงานวิจัยเกี่ยวข้องจะใช้เกณฑ์ตั้งแต่ 0.5 –1 (ศุภกฤษฎิ์ นิวัฒนากูล, 2556)แต่ความเป็นจริงแล้วความรู้หรือเอกสารที่ไม่ถูกเลือกโดยมีค่าน้อยกว่า 0.5 ก็ยังถือว่าความรู้หรือเอกสารนั้นมีความเกี่ยวข้องกับคำค้น ดังนั้นงานวิจัยนี้จึงนำเสนอแนวคิดการเลือกความรู้หรือเอกสารแบบไดนามิกเกณฑ์ (Dynamic Gains)โดยใช้ค่าคะแนนต่ำสุดของความรู้หรือเอกสารค้นคืนได้มาเป็นเกณฑ์

มหาวิทยาลัยเทคโนโลยีสุรนารี

3.3.2 การประเมินประสิทธิภาพการสืบค้น

1. เครื่องมือในการออกแบบสำหรับการสืบค้น

- ภาษาพีเอชพี (PHP) ใช้สำหรับการพัฒนาส่วนติดต่อผู้ใช้งานและเพื่อแสดงผลลัพธ์ของการสืบค้นเชิงความหมายสำหรับความรู้ด้านพืชไร่ผ่านเว็บบราวเซอร์
- แร็พ (RDF API for PHP: RAP) ใช้สำหรับเชื่อมต่อระหว่างโครงสร้างออนโทโลยีกับการสืบค้นความรู้
- แซมป์5.6.30 (XAMMP) ใช้สำหรับสร้างเว็บเซิร์ฟเวอร์ (Web Server) สำเร็จรูปบนระบบปฏิบัติการไมโครซอฟท์ วินโดวส์
- สปาร์เก้ิล ใช้สำหรับสอบถามสืบค้นความรู้ที่เอื้อต่อการประมวลผลหรือข้อคำถามอย่างเป็นอัตโนมัติ

2. การประเมินประสิทธิภาพของแบบจำลองการสืบค้น

2.1 การให้น้ำหนักคำสำคัญ ประเมินประสิทธิภาพการสืบค้นด้วยค่าความแม่นยำเฉลี่ย (Mean Reciprocal Rank:MRR)และค่าความผิดพลาด (Sum of Squared Errors of Prediction: SSE) ดังนี้

- ค่าความแม่นยำเฉลี่ย คือการวัดประสิทธิภาพของการจัดอันดับจากแนวคิดของ คลาสเวลล์ และ ฮาวคิง(Craswell and Hawking, 2002) ดังสมการที่ 3.3

$$MRR = \sum_{i=1}^{N} \frac{1}{N \times r_i} \quad (3.3)$$

เมื่อ

N คือ จำนวนคำค้นทั้งหมด

r_i คือ ตำแหน่งของผลลัพธ์ที่ค้นพบอย่างถูกต้องของจำนวนคำค้นที่i

- ค่าความผิดพลาด คือการคำนวณเพื่อรวมค่าความผิดพลาดของการให้น้ำหนักคำสำคัญ ซึ่งความคาดหวังของงานวิจัยนี้ต้องการให้ค่าน้ำหนักมีค่าเข้าใกล้ 1 มากที่สุดโดยค่าที่ได้ยิ่งมีค่าน้อยยิ่งดีซึ่งค่าน้ำหนักที่ได้จะส่งผลต่อการเรียงลำดับข้อมูล ดังสมการที่ 3.4

$$RSS = \sum_{i=1}^{n} (y_i - f(x_i))^2 \quad (3.4)$$

มหาวิทยาลัยเทคโนโลยีสุรนารี

2.2 การวัดความคล้ายคลึงประเมินประสิทธิภาพการสืบค้นด้วยค่าความแม่นยำเฉลี่ย (Mean Reciprocal Rank:MRR)ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และ ค่าความถ่วงดุล (F-measure)ซึ่งเหตุผลที่มีการเลือกด้วยวิธีวัดการประเมินดังกล่าว เนื่องจากเป็นวิธีการที่นิยมใช้ สอดคล้องกับแฟรงค์เคส, วิลเลี่ยม, เบร์ซา เยต์ส และริคาร์โด (Frankes, William, Baeza-Yates and Ricardo, 1992) และ ไมโอ, ดวน, เซียง และ ไจโอ (Miao, Duan, Zhang, and Jiao, 2009) อธิบายไว้ดังนี้

- ความแม่นยำ คือ การพิจารณาความถูกต้องของข้อมูลที่สืบค้นได้จากอัตราส่วนระหว่างจำนวนเอกสารที่ค้นคืนถูกต้องกับจำนวนเอกสารที่ค้นคืนมาได้ทั้งหมด ดังสมการที่ 3.5

$$Precision = \frac{TP}{TP+FP} \times 100\% \qquad (3.5)$$

เมื่อ

TP (True Prositive) แทน เอกสารที่เลือก และถูกเลือกโดยผู้เชี่ยวชาญ

FP (False Prositive) แทน เอกสารที่เลือก แต่ไม่ถูกเลือกโดยผู้เชี่ยวชาญ

- ค่าความระลึก คือ การพิจารณาความครบถ้วนของข้อมูลเมื่อเทียบกับข้อมูลที่ควรได้ทั้งหมด จากอัตราส่วนระหว่างจำนวนเอกสารที่ค้นคืนถูกต้องกับจำนวนเอกสารที่ถูกต้องทั้งหมดของระบบ (เน้นคำตอบที่ถูกต้องและแสดงผลลัพธ์มากที่สุด)ดังสมการที่ 3.6

$$Recall = \frac{TP}{TP+FN} \times 100\% \qquad (3.6)$$

เมื่อ

TP (True Prositive) แทน เอกสารที่เลือก และถูกเลือกโดยผู้เชี่ยวชาญ

FN (False Negative) แทน เอกสารที่ไม่เลือก แต่ถูกเลือกโดยผู้เชี่ยวชาญ

- ค่าความถ่วงดุล คือ ค่าเฉลี่ยที่ให้ความสำคัญกับความแม่นยำและความครบถ้วนเท่า ๆ กัน โดยเป็นการเปลี่ยนค่าความถูกต้องและค่าความแม่นยำมารวมเป็นหนึ่งเดียวดังสมการที่ 3.7

$$F = (2 \times Precision \times Recall)/(Precision+Recall) \quad (3.7)$$

บทที่ 4
ผลการวิจัย

การวิจัยแบบจำลองการสืบค้นเชิงความหมายสำหรับความรู้ด้านพืชไร่ในครั้งนี้ มีผลการวิเคราะห์ข้อมูลตามลำดับดังนี้

4.1 ผลการออกแบบจำลองการสืบค้นเชิงความหมาย

4.1.1 การจัดความรู้ให้เป็นระบบ

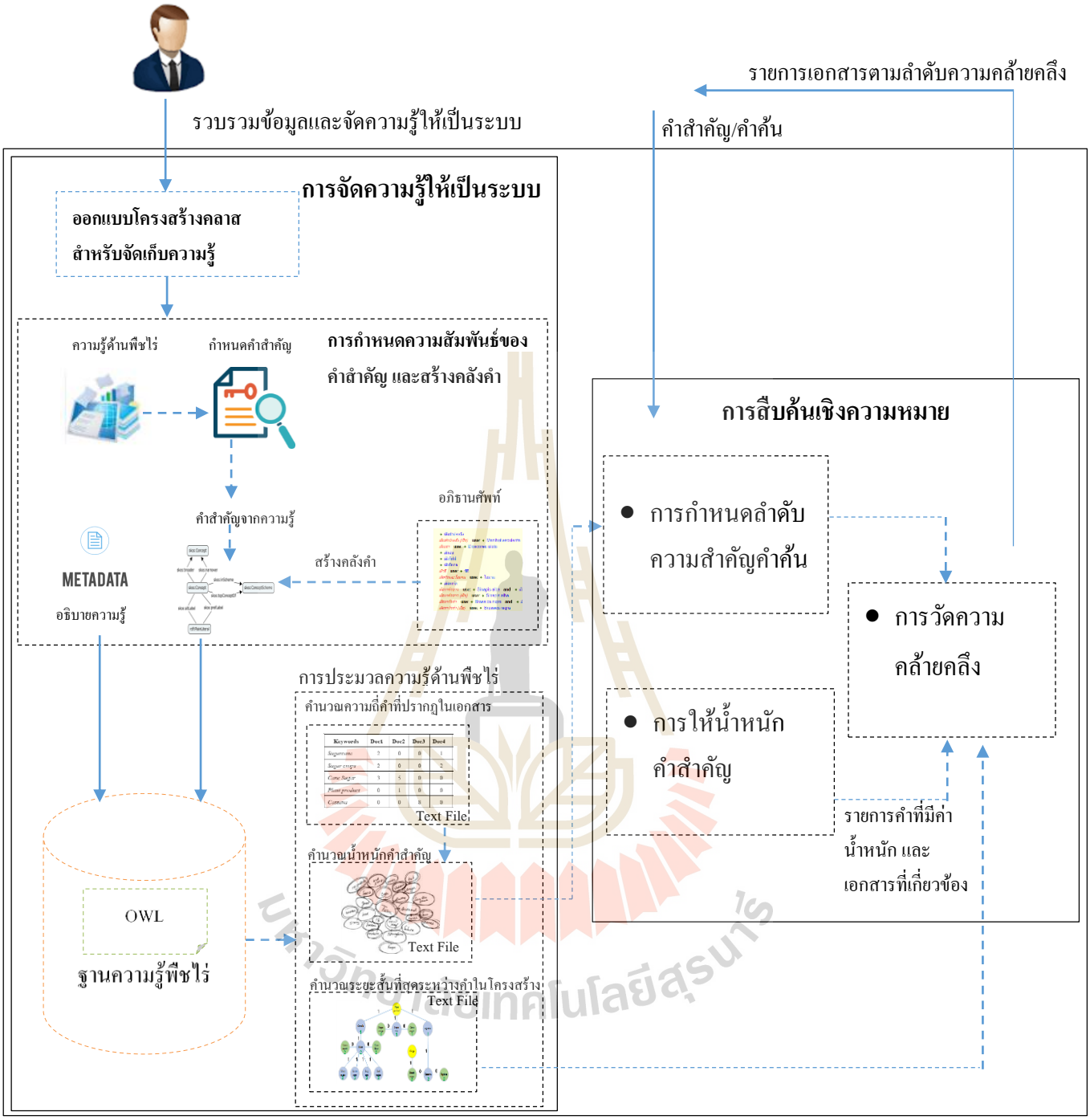
4.1.2 การสืบค้นเชิงความหมาย

4.2 ผลการประเมินประสิทธิภาพการสืบค้นเชิงความหมาย

4.1 ผลการออกแบบจำลองการสืบค้นเชิงความหมาย

จากการศึกษาปัญหา วิเคราะห์และออกแบบจำลองการสืบค้นเชิงความหมาย สำหรับความรู้ด้านพืชไร่ (Semantic Search Model for Agronomy Knowledge: SeeAgrow) ทำให้ได้สถาปัตยกรรมของแบบจำลองการสืบค้นเชิงความหมาย ดังรูปที่ 4.1

มหาวิทยาลัยเทคโนโลยีสุรนารี



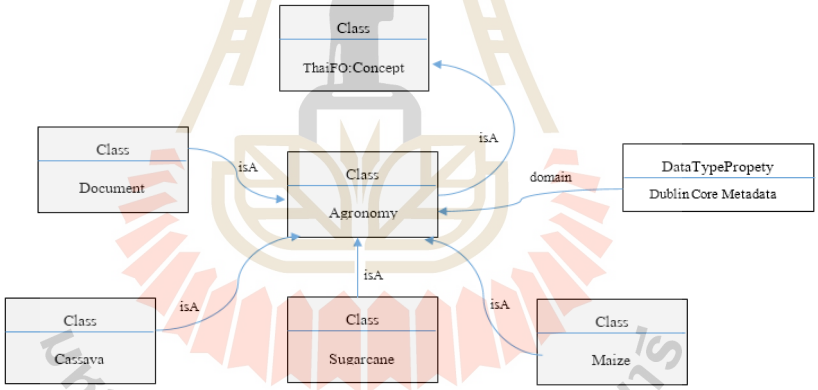
รูปที่ 4.1 สถาปัตยกรรมของแบบจำลองการสืบค้นเชิงความหมาย (SeeAgrow)

จากรูปที่ 4.1 สถาปัตยกรรมแบบจำลองการสืบค้นเชิงความหมายสำหรับความรู้ด้านพืชไร่ (SeeAgrow) ประกอบด้วยการทำงาน 2 กระบวนการ คือ การจัดความรู้ให้เป็นระบบ และการสืบค้นเชิงความหมาย

4.1.1 การจัดความรู้ให้เป็นระบบ

1) การออกแบบโครงสร้างคลาสจัดเก็บความรู้ด้านพืชไร่

ผลการออกแบบโครงสร้างคลาสจัดเก็บความรู้ในกระบวนการจัดความรู้ให้เป็นระบบ ซึ่งประยุกต์ใช้ออนโทโลยีไม้ผลเศรษฐกิจ (ThaiFO Ontology) งานวิจัยนี้ได้สร้างคลาสจัดเก็บกลุ่มความรู้ด้านพืชไร่เพิ่มเติม จำนวน 4 คลาส โดยประยุกต์แนวคิดการออกแบบข้อมูลออนโทโลยีเฉพาะเกี่ยวกับพืชไร่ของสำนักงานพัฒนาการวิจัยการเกษตร (องค์การมหาชน, 2558) ได้แก่ คลาสแทนความรู้ด้านพืชไร่ (Agronomy) ประกอบด้วย คลาสมันสำปะหลัง (Cassava) คลาสอ้อย (Sugarcane) คลาสข้าวโพด (Maize) โครงสร้างดังรูปที่ 4.2



รูปที่ 4.2 โครงสร้างของคลาสในการจัดเก็บความรู้ด้านพืชไร่

จากรูปที่ 4.2 การจัดเก็บความรู้ด้านพืชไร่จัดเก็บความรู้ในคลาสความรู้หรือเอกสาร (Document) ซึ่งเป็นโครงสร้างคลาสในออนโลยีไม้ผลเศรษฐกิจ และได้เชื่อมความสัมพันธ์ของความรู้พืชไร่ 3 คลาส คือ คลาสมันสำปะหลัง (Cassava) คลาสอ้อย (Sugarcane) คลาสข้าวโพด (Maize)

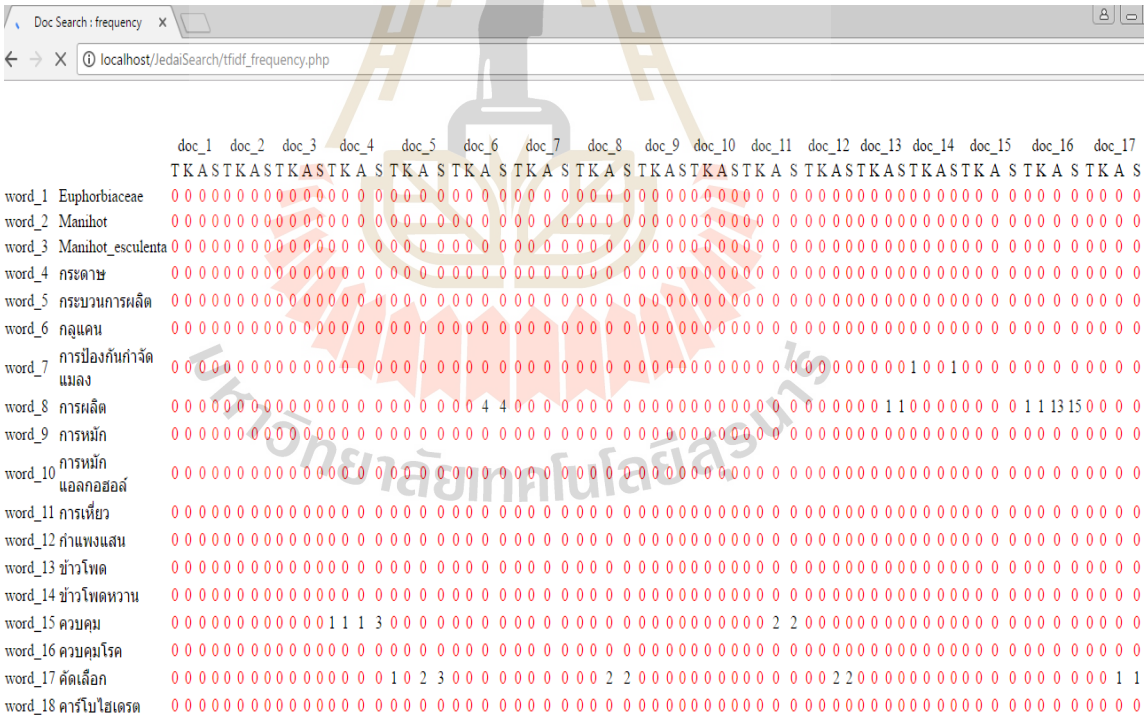
2) การกำหนดความสัมพันธ์ของคำสำคัญ และสร้างคลังคำ

ผลการกำหนดความสัมพันธ์ของคำสำคัญ และสร้างคลังคำจากขั้นตอนการเลือกคำสำคัญจากความรู้ด้านพืชไร่ 240 รายการ ได้จำนวน 181 คำ และเมื่อนำคำดังกล่าวไปกำหนดความสัมพันธ์ของคำหลัก คำกว้างกว่า คำแคบกว่า คำที่เกี่ยวข้อง และคำเหมือน จากอภิธานศัพท์เกษตรไทยทำให้ได้ จำนวน 1,525 คำ จัดเก็บในคลังคำสำหรับงานวิจัยนี้

3) การประมวลความรู้ด้านพืชไร่

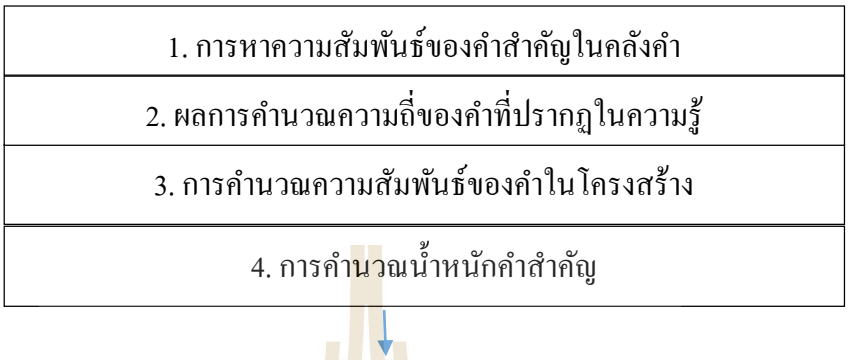
ผลการประมวลความรู้ด้านพืชไร่ (1) การคำนวณความถี่ของคำที่ปรากฏในความรู้พืชไร่ (2) การคำนวณน้ำหนักคำ และ (3) การคำนวณระยะทางสั้นที่สุดระหว่างคำในโครงสร้าง ปรากฏตัวอย่างวิธีการคำนวณ และผลการคำนวณที่เก็บเป็นไฟล์ข้อความ ดังนี้

3.1) การคำนวณความถี่ของคำที่ปรากฏในความรู้ด้านพืชไร่ ปรากฏตัวอย่างดังรูปที่ 4.3



รูปที่ 4.3 ตัวอย่างผลการคำนวณความถี่คำที่ปรากฏในความรู้โดยใช้ภาษาพีเอชพี

3.2 การคำนวณน้ำหนักคำสำคัญ จากการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ของดีคอฟฟ์ (Diekhoff, 1992) มีขั้นตอน และตัวอย่างวิธีการคำนวณดังนี้



รูปที่ 4.4 ขั้นตอนการวิเคราะห์ให้น้ำหนักคำ

3.2.1 การหาความสัมพันธ์ของคำสำคัญในคลังคำ คือการคำนวนวณความสัมพันธ์ของคำในคลังคำทั้งหมด โดยไม่คำนึงถึงอันดับคำนวณจากสมการที่ 4.1

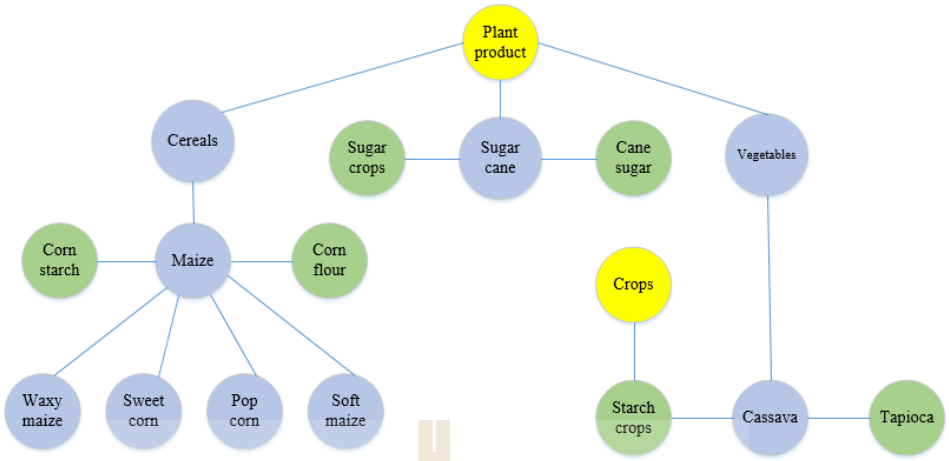
$$\frac{n!}{(n-r)!r!} \tag{4.1}$$

เมื่อ

n แทน จำนวนสมาชิก

r แทน จำนวนกลุ่มหรือหมู่

ตัวอย่างที่ 4.1 การหาความสัมพันธ์ของคำสำคัญในคลังคำ จากตัวอย่างโครงสร้างความสัมพันธ์ของคำในคลังคำ ดังรูปที่ 4.5



รูปที่ 4.5 ตัวอย่างโครงสร้างความสัมพันธ์ของคำที่เกี่ยวข้องกับความรู้ด้านพืชไร่
ที่มา: อภิธานศัพท์เกษตรไทย (http://pikul.lib.ku.ac.th/agrovoc/agrovoc_th_adv.php)

จากรูปที่ 4.5 โครงสร้างความสัมพันธ์ของคำ จำนวน 17 คำ สร้างเซตของคำ ดังนี้

S={ *Plant product, Cereals, Maize, Corn starch, Corn flour, Waxy maize, Sweet corn, Pop corn, Soft maize, Sugar cane, Sugar crops, Cane sugar, Vegetables, Cassava, Tapioca, Starch crops, Crops* }

จากเซตของคำสามารถคำนวณหาความสัมพันธ์ระหว่างคำ 2 คำ ว่ามีความสัมพันธ์กันหรือไม่ โดยแทนค่าในสมการที่ 4.1 ดังนี้

$$\frac{17!}{(17-2)!2!}$$

$$= \frac{17x16x15x14x13x12x11x10x9x8x7x6x5x4x3x2x1}{(15x14x13x12x11x10x9x8x7x6x5x4x3x2x1)2x1}$$

$$= \frac{355687428096000}{1307674368000x2}$$

$$= \frac{355687428096000}{2615348736000} \quad = \quad 136 \text{ สัมพันธ์}$$

3.2.2 ผลการคำนวณความถี่ของคำที่ปรากฏในความรู้ คือการนำไฟล์ข้อความที่เก็บผลการคำนวณความถี่ของคำที่ปรากฏในความรู้มาใช้ ปรากฏตัวอย่างดังตารางที่ 4.1

ตารางที่ 4.1 ตัวอย่างผลการคำนวณความถี่ของคำสำคัญที่ปรากฏในแต่ละความรู้

คำ	Doc1	Doc2	Doc3	Doc4	Doc5	Doc6	Doc7	Doc8	Doc9
Sugarcane	2	0	0	1	0	0	1	0	0
Sugar crops	2	0	0	2	1	0	0	1	0
Cane Sugar	3	5	0	0	0	0	0	0	2
Plant product	0	1	0	0	1	0	0	0	1
Cassave	0	0	8	0	0	2	0	0	0

3.2.3 การคำนวณความสัมพันธ์ของคำในโครงสร้าง จากการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ของดีคอฟฟ์ (Diekhoff, 1992) มีค่าสัมประสิทธิ์สหสัมพันธ์อยู่ในช่วง 0 ถึง -1 งานวิจัยนี้ปรับค่าสัมประสิทธิ์สัมพันธ์ระหว่างคำสองคำให้อยู่ในช่วงเกณฑ์ 0 ถึง 1 ตามเกณฑ์ของฮินเคอล์ (Hinkle, 1998 : 118) ดังสมการที่ 4.2

$$corr(x,y) = \left| \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}} \right| \quad (4.2)$$

เมื่อ

x = ความถี่ของคำ x ในแต่ละเอกสาร

$\bar{x}$ = ค่าเฉลี่ยของความถี่คำ $\bar{x}$ ในทุกเอกสาร

y = ความถี่ของคำ y ในแต่ละเอกสาร

$\bar{y}$ = ค่าเฉลี่ยของความถี่คำ $\bar{y}$ ในทุกเอกสาร

ตัวอย่างที่ 4.2 การวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ระหว่างคำ *“Sugarcane”* และ *“Sugar crops” (ใช้ตัวอย่างความสัมพันธ์ของคำในโครงสร้างรูปที่ 4.5 และ ผลการคำนวณความถี่ตารางที่ 4.1)* ดังนี้

มหาวิทยาลัยเทคโนโลยีสุรนารี

$$\sum(Sugarcane - \overline{Sugarcane})(Sugar\ crops - \overline{Sugar\ crops})$$
$$= (2-0.44)(2-0.67) + (0-0.44)(0-0.67) + (0-0.44)(0-0.67) + (1-0.44)(2-0.67) + (0-0.44)(1-0.67) + (0-0.44)(0-0.67) + (1-0.44)(1-0.67) + (0-0.44)(1-0.67) + (0-0.44)(0-0.67) = 3.33$$

$$\sum(Sugarcane - \overline{Sugarcane})^2$$
$$= (2-0.44)^2 + (0-0.44)^2 + (0-0.44)^2 + (1-0.44)^2 + (0-0.44)^2 + (0-0.44)^2 + (1-0.44)^2 + (0-0.44)^2 + (0-0.44)^2 = 4.22$$

$$\sum(Sugar\ crops - \overline{Sugar\ crops})^2 = (2-0.67)^2 + (0-0.67)^2 + (0-0.67)^2 + (2-0.67)^2 + (1-0.67)^2 + (0-0.67)^2 + (1-0.67)^2 + (1-0.67)^2 + (0-0.67)^2 = 6.00$$

$$\frac{\sum(Sugarcane - \overline{Sugarcane})(Sugar\ crops - \overline{Sugar\ crops})}{\sqrt{\sum(Sugarcane - \overline{Sugarcane})^2 \sum(Sugar\ crops - \overline{Sugar\ crops})^2}} = \frac{3.33}{\sqrt{4.22 \times 6.00}}$$

$$corr(Sugarcane, Sugar\ crops) = |0.66| = 0.66$$

จากตัวอย่างวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ระหว่างคำ *"Sugarcane"* และ *"Sugar crops"* พบว่าคำ *Sugarcane* และ *Sugar crops* มีค่าสัมประสิทธ์สหสัมพันธ์กันระดับปานกลาง เท่ากับ 0.66 ตามเกณฑ์ของฮินเคอล์ (Hinkle, 1998 : 118) และคำใดก็แล้วแต่ถ้ามีความเหมือนกันค่าสัมประสิทธ์สหสัมพันธ์จะเท่ากับ 1 เสมอ เช่น *Sugarcane* กับ *Sugarcane* หรือ *Sugar crops* กับ *Sugar crops* เป็นต้น ซึ่งผลการวิเคราะห์น้ำหนักคำด้วยการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ ปรากฏดังตารางที่ 4.2

ตารางที่ 4.2 ตัวอย่างค่าความสัมพันธ์ระหว่างคำในโครงสร้างด้วยการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์

คำ	k1	k2	k3	k4	k5	k6	k 7	k8	k9	k10
k1	1	0.5	0.24	0.24	0	0	0.36	0.08	0.06	0.46
k2	-	1	0.12	0.12	0.5	0.24	0.18	0.18	0.58	0.22
k3	-	-	1	1	0.24	0.24	0.18	0.18	0.24	0.8

ตารางที่ 4.2 ตัวอย่างค่าความสัมพันธ์ระหว่างคำในโครงสร้างด้วยการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ (ต่อ)

คำ	k1	k2	k3	k4	k5	k6	k 7	k8	k9	k10
k4	-	-	-	1	0.24	0.24	0.18	0.18	0.24	0.8
k5	-	-	-	-	1	0	0	0.36	0.62	0.12
k6	-	-	-	-	-	1	0	0.74	0.48	0.12
k7	-	-	-	-	-	-	1	0.26	0.18	0.32
k8	-	-	-	-	-	-	-	1	0.36	0.12
k9	-	-	-	-	-	-	-	-	1	0.08
k10	-	-	-	-	-	-	-	-	-	1

** k1, k2 , k3, k4, k5, k6, k7, k8, k9, k10 คือ Plant product, Cereals, Maize, Corn starch, Corn flour, Waxy maize, Sweet corn, Pop corn, Soft maize, Sugar cane (อ้างอิงตัวอย่างโครงสร้างความสัมพันธ์ของคำรูปที่ 4.5)*

3.2.3 การคำนวณน้ำหนักคำสำคัญ เป็นการคำนวณน้ำหนักคำสำคัญของแต่ละคำที่ผู้ใช้ป้อนเข้ามาไปคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคำค้นจากผู้ใช้กับคำที่ปรากฏในเอกสาร ซึ่งขั้นตอนนี้จะนำผลการคำนวณน้ำหนักคำในโครงสร้างมาใช้ในการคำนวณหาค่าน้ำหนักแต่ละครั้ง ดังสมการที่ 4.3 และปรากฏตัวอย่างผลการคำนวณดังตารางที่ 4.3

$$corr(x)_j = \frac{\sum corr(x,y_j)}{T_j} \tag{4.3}$$

เมื่อ

$corr(x)_j$ = ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง คำค้น x กับเอกสาร j

$corr(x,y_j)$ = ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง คำค้น x กับคำ y ในเอกสาร

x = คำค้น

y_j = แต่ละคำในเอกสาร j

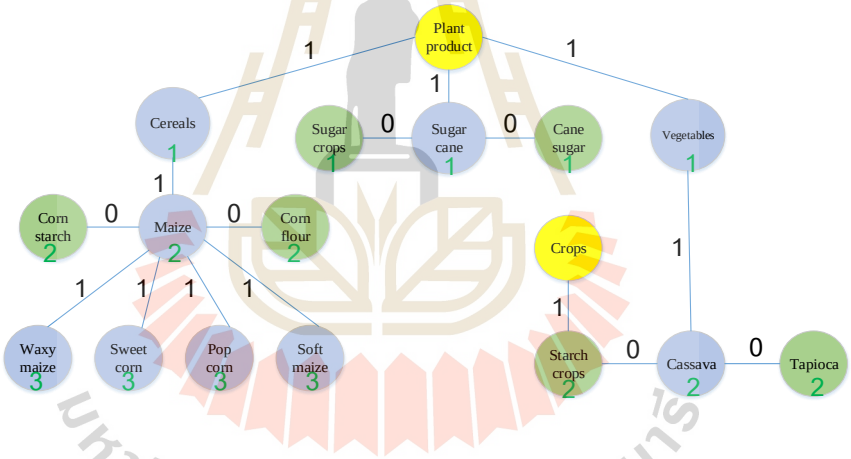
T_j = จำนวนคำในเอกสาร j

ตารางที่ 4.3 ตัวอย่างค่าน้ำหนักคำด้วยการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ต่อความรู้หรือเอกสาร

คำ	Doc1 ปรากฏคำ k1, k3	Doc2 ปรากฏคำ k2,k4
k1	0.62	0.37
k2	0.31	0.31
k3	0.62	0.62
k4	0.62	0.62

3.3 การคำนวณระยะทางสั้นที่สุดระหว่างคำในโครงสร้าง

การคำณวณระยะทางสั้นที่สุดระหว่างโหนดของคำในโครงสร้างคำนวณตามขั้นตอนวิธีของ ไดก์สตรา (Dijkstra, E., 1959) ปรากฏตัวอย่างดังรูปที่ 4.6 ผลการคำนวณดังตารางที่ 4.4



รูปที่ 4.6 ตัวอย่างการกำหนดความสัมพันธ์ของคำในคลังคำ

ตัวอย่างที่ 4.3 การคำนวณระยะทางระหว่างคำในโครงสร้างรูปที่ 4.6 ดังนี้

<u>รอบที่ 1</u>

PriorityQueue = {startVertex}

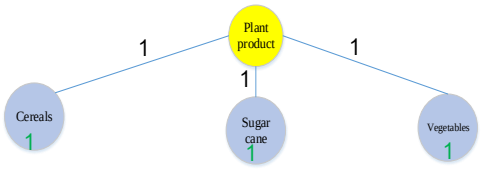
PQ={(Plant product,Plant product,0)}

Step 1: ดึงสมาชิกตัวแรกออกมา PQ={}

Step 2: เลือก Plant product ให้เป็นเส้นทางจริง PQ={}

Step 3: สำหรับทุกโหนดที่ต่อกับ Plant product คำนวณระยะทางแล้วนำกลับไปเก็บใน PQ

PQ={(Plant product,Cereals,1),(Plant product,Sugar cane,1),(Plant product,Vegetables,1)}



<u>รอบที่ 2</u>

Step 4: ดึง (Plant product, Sugar cane,1) สมาชิกตัวแรกออกมา PQ={(Plant product,Cereals,1),(Plant product,Vegetables,1)}

Step 5: เลือก Sugar cane ให้เป็นเส้นทางจริง PQ={(Plant product, Cereals,1),(Plant product,Vegetables,1)}

Step 6: สำหรับทุกโหนดที่ต่อกับ Sugar cane คำนวณระยะทางแล้วนำกลับไปเก็บใน PQ PQ={(Sugar cane,Sugar crops,1),(Sugar cane,Cane sugar,1) ,(Plant product, Cereals,1),(Plant product,Vegetables,1)}

<u>รอบที่... N</u>

Step... N: ไม่มีสมาชิกใน PQ สิ้นสุดการคำนวณ

ตารางที่ 4.4 ตัวอย่างผลลัพธ์การคำนวณระยะทางระหว่างคำในโครงสร้าง

คำ	k1	k2	k3	k4	k5	k6	k 7	k8	k9	k10
k1	0	1	2	2	2	3	3	3	3	1
k2	1	0	1	1	1	2	2	2	2	2
k3	2	1	0	0	0	1	1	1	1	3
k4	2	1	0	0	0	1	1	1	1	3
k5	2	1	0	0	0	1	1	1	1	3
k6	3	2	1	1	1	0	2	2	2	4
k7	3	2	1	1	1	2	0	2	2	4
k8	3	2	1	1	1	2	2	0	2	4
k9	3	2	1	1	1	2	2	2	0	4
k10	1	2	3	3	3	4	4	4	4	0

** k1, k2 , k3, k4, k5, k6, k7, k8, k9, k10 คือ Plant product, Cereals, Maize, Corn starch, Corn flour, Waxy maize, Sweet corn, Pop corn, Soft maize, Sugar cane (อ้างอิงตัวอย่างการกำหนดความสัมพันธ์ของคำรูปที่ 4.6)*

มหาวิทยาลัยเทคโนโลยีสุรนารี

จากตารางที่ 4.4 พบว่า ถ้าคำสองคำมีค่าระยะทางห่างกันน้อยหรือสั้นที่สุดจะถือว่าคำสองคำนั้นมีความเกี่ยวพันกันทางความหมายค่อนข้างมาก แต่ถ้าคำสองคำมีค่าระยะทางห่างกันมากจะถือว่าคำทั้งสองมีความเกี่ยวพันกันทางความหมายน้อย

ผลกระบวนการจัดการความรู้ให้เป็นระบบจากการรวบรวมเอกสารความรู้ด้านพืชไร่ ซึ่งประยุกต์ใช้ออนโทโลยีไม้ผลเศรษฐกิจ โดยกำหนดคำสำคัญจากความรู้นำมาสร้างคลังคำที่ใช้ตัวแบบการแทนความรู้เอสเคโอเอส และนิยามศัพท์ที่เกี่ยวข้องกับคำสำคัญจากอภิธานศัพท์เกษตรไทย ใช้การอธิบายรายละเอียดความรู้ด้วยเมทาดาทาดับลินคอร์ ซึ่งมีการคำนวณ (1) ความถี่ของคำที่ปรากฏในเอกสาร (2) น้ำหนักคำด้วยการวิเคราะห์ค่าสัดส่วนของจำนวนเอกสารทั้งหมดกับจำนวนของเอกสารที่ปรากฏคำสนใจ (3) น้ำหนักคำด้วยการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ และ (4) ระยะทางสั้นที่สุดระหว่างคำในโครงสร้าง ได้ข้อสังเกตจากผลการวิเคราะห์สามารถสรุปได้ดังนี้

1. การให้น้ำหนักคำสำคัญด้วยการวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์ พบว่า คำใด ๆ ที่มีความสัมพันธ์กัน หรือเกี่ยวข้องกันในลักษณะคำกว้างกว่า คำแคบกว่า คำที่เกี่ยวข้อง และคำเหมือนจะมีค่าสัมประสิทธิ์สหสัมพันธ์เข้าใกล้ 1 และกรณีคำสองคำนั้นเป็นคำเหมือนกันจะมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 1 ซึ่งแสดงว่าคำสองคำมีความสัมพันธ์กันในระดับมากที่สุด

2. การคำนวณระยะทางสั้นที่สุดระหว่างโหนดของคำในโครงสร้างโดยการหาระยะทางจากเส้นเชื่อมระหว่างโหนดของความสัมพันธ์คำในโครงสร้าง พบว่า คำใด ๆ ก็แล้วแต่ที่มีค่าระยะทางที่น้อยจะมีความเกี่ยวพันหรือสัมพันธ์กันทางความหมายมาก เช่นระยะทางระหว่างคำ k1 ผลิตผลจากพืช *(Plant product)* กับคำ k1 ผลิตผลจากพืช *(Plant product)* ซึ่งเป็นคำเหมือนกันมีค่าระยะทางเท่ากับ 0 และระยะทางระหว่างคำ k1 ผลิตผลจากพืช *(Plant product)* กับคำ k2 ธัญพืช *(Cereals)* ซึ่ง k1 เป็นคำที่กว้างกว่าของคำ k2 มีค่าระยะทางเท่ากับ 1

4.1.2 สืบค้นเชิงความหมาย

กระบวนการออกแบบและพัฒนาแบบจำลองการสืบค้นเชิงความหมาย เป็นผลที่ได้จาก (1) การกำหนดลำดับความสำคัญคำค้น (2) การวัดความคล้ายคลึงจากการนำผลการคำนวณให้น้ำหนักคำสำคัญ และนำผลคำนวณระยะทางระหว่างคำในโครงสร้างมาใช้ ซึ่งได้ตัวอย่าง และผลการคำนวณ ดังนี้

มหาวิทยาลัยเทคโนโลยีสุรนารี

1) การกำหนดลำดับความสำคัญของคำค้น

กรณีผู้ใช้ป้อนข้อสอบถาม หรือคำค้นเข้ามา ซึ่งจะดำเนินการคำนวณวิเคราะห์กำหนดลำดับความสำคัญของคำค้นโดยเก็บไว้ในหน่วยความจำ 2 กรณี คือวิเคราะห์ลำดับคำค้นจากผู้ใช้ และวิเคราะห์ตำแหน่งของคำค้นที่ปรากฏในเอกสาร ตัวอย่างการคำนวณดังนี้

1.1) ลำดับคำค้นจากผู้ใช้ คำนวณได้จากสมการที่ 4.4

$$Q_i = \frac{n-i+1}{\sum i} \tag{4.4}$$

เมื่อ

Q_i แทน ค่าน้ำหนักคำค้นที่ i

n แทน จำนวนของคำค้น

i แทน ลำดับของคำค้น

$\sum i$ แทน ผลรวมค่าลำดับของคำค้นทั้งหมด

ตัวอย่างที่ 4.4 การคำนวณลำดับคำค้นจากผู้ใช้ กรณีผู้ใช้ระบุคำค้นเข้ามา 3 คำค้น ได้แก่ อ้อย น้ำตาล และพันธุ์ ตามลำดับ ดังนั้นลำดับความสำคัญของคำค้น อ้อย มีความสำคัญเท่ากับ 1 น้ำตาล เท่ากับ 2 และพันธุ์ เท่ากับ 3 ตามลำดับ ซึ่งผลรวมค่าลำดับคำค้นทั้งหมด เท่ากับ 6 แทนค่าในสมการที่ 4.4 ดังนี้

$$Q_{\text{อ้อย}} = \frac{3-1+1}{6} \quad \text{ดังนั้น} \quad Q_{\text{อ้อย}} = 0.50$$

$$Q_{\text{น้ำตาล}} = \frac{3-2+1}{6} \quad \text{ดังนั้น} \quad Q_{\text{น้ำตาล}} = 0.33$$

$$Q_{\text{พันธุ์}} = \frac{3-3+1}{6} \quad \text{ดังนั้น} \quad Q_{\text{พันธุ์}} = 0.17$$

ดังนั้นลำดับความสำคัญของคำค้น อ้อย น้ำตาล และพันธุ์ เท่ากับ 0.50, 0.33 และ 0.17 ตามลำดับ

มหาวิทยาลัยเทคโนโลยีสุรนารี

1.2) ตำแหน่งของคำค้นที่ปรากฏในความรู้ คำนวณได้จากสมการที่ 4.5

$$p_i = \frac{w_i}{\sum w_i} \tag{4.5}$$

เมื่อ

p_i แทน ค่าน้ำหนักของตำแหน่งที่ i

w_i แทน น้ำหนักของตำแหน่งที่ i ที่กำหนดไว้เบื้องต้น

$\sum w_i$ แทน ผลรวมค่าน้ำหนักของตำแหน่งที่ i ที่กำหนดไว้เบื้องต้น

ตัวอย่างที่ 4.5 การคำนวณตำแหน่งของคำค้นที่ปรากฏในความรู้ กรณีผู้ใช้ระบุคำค้นเข้ามา 3 คำค้น ได้แก่ อ้อย น้ำตาล และพันธุ์ ตามลำดับ เช่น ลำดับความสำคัญของคำค้นอ้อยที่ปรากฏในชื่อเรื่อง เท่ากับ 1 อ้อย ปรากฏในคำสำคัญ เท่ากับ 2 และอ้อยที่ปรากฏในรายละเอียด เท่ากับ 3 เป็นต้น ซึ่งค่าความสำคัญของตำแหน่งที่ปรากฏคำค้น ดังรูปที่ 4.7

ชื่อเรื่อง : อ้อย(1)
คำสำคัญ : อ้อย(2) พันธุ์(2)
รายละเอียด : อ้อย(3) น้ำตาล(3) พันธุ์(3)

รูปที่ 4.7 ลำดับคำสำคัญของตำแหน่งที่ปรากฏคำค้น

จากรูปที่ 4.7 พบว่า คำค้น อ้อย ปรากฏในตำแหน่งชื่อเรื่อง คำสำคัญ และรายละเอียด ซึ่งผลรวมลำดับความสำคัญของตำแหน่งที่ปรากฏคำค้น อ้อย มีค่า (1+2+3) เท่ากับ 6 “พันธุ์” (2+3) เท่ากับ 5 และ น้ำตาล เท่ากับ 3 สามารถแทนค่าในสมการที่ 4.4 ปรากฏผลการคำนวณดังตารางที่ 4.5

มหาวิทยาลัยเทคโนโลยีสุรนารี

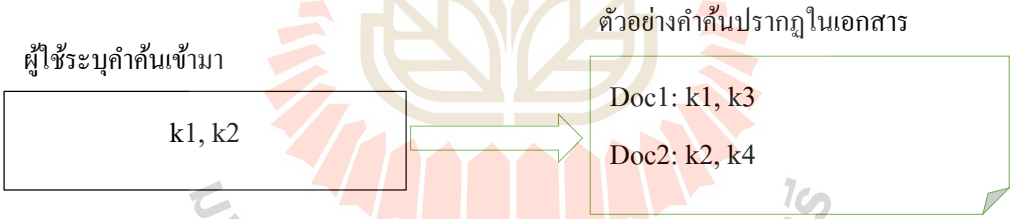
ตารางที่ 4.5 ตัวอย่างการวิเคราะห์ตำแหน่งของคำค้นที่ปรากฏในความรู้

3 คำค้น	ตำแหน่งคำค้นที่ปรากฏ/การกำหนดลำดับความสำคัญ		
	ชื่อเรื่อง	คำสำคัญ	รายละเอียด
อ้อย	$\frac{1}{6} = 0.17$	$\frac{2}{6} = 0.33$	$\frac{3}{6} = 0.5$
น้ำตาล			$\frac{1}{3} = 1$
พันธุ์		$\frac{2}{5} = 0.4$	$\frac{3}{5} = 0.6$

2) การวัดความคล้ายคลึง กระบวนการนี้ได้จากการนำผลการคำนวณ (1) กำหนดลำดับความสำคัญของคำค้น (2) การให้น้ำหนักคำสำคัญ และ (3) การคำนวณระยะทางระหว่างคำในโครงสร้างมาทดลองตามรูปแบบการทดลองในตารางที่ 3.1 ตัวอย่างการคำนวณ ดังนี้

ตัวอย่างที่ 4.6 การคำนวณวัดความคล้ายคลึง

เมื่อผู้ใช้ระบุคำค้นเข้า จำนวน 2 คำ ได้แก่ k1 *(Plant product)* กับ k2 *(Cereals, Maize)* เพื่อการค้นคืนความรู้ ซึ่งคำค้นปรากฏในความรู้หรือเอกสาร ดังรูปที่ 4.7



รูปที่ 4.8 ตัวอย่างคำค้นที่ผู้ใช้ระบุเข้ามากับคำค้นที่ปรากฏในความรู้หรือเอกสาร

1) นำผลการคำนวณระยะทางระหว่างคำในโครงสร้างมาคำนวณระยะทางระหว่างคำค้นกับคำในความรู้หรือเอกสาร ดังตารางที่ 4.6 *(อ้างอิงจากตารางผลลัพธ์การคำนวณระยะทางตารางที่ 4.4)*

ตารางที่ 4.6 ตัวอย่างผลลัพธ์การคำนวณระยะทางระหว่างคำในโครงสร้าง

ระยะทางระหว่างคำในโครงสร้าง	k1	k2	k3	k4
k1	0	1	2	2
k2	1	0	1	1
k3	2	1	0	0
k4	2	1	0	0

* k1, k2 , k3, k4 คือ *Plant product, Cereals, Maize, Corn starch, Corn flour*

จากตารางที่ 4.6 ค่าระยะทางระหว่างคำค้น k1 *(Plant product)* กับ k2 *(Cereals, Maize)* เท่ากับ 1 ซึ่งค่าดังกล่าวนำไปคำนวณระยะทางระหว่างคำค้นกับคำในความรู้ทั้งหมดได้ในสมการที่ 4.6 ประยุกต์มาจากแนวคิดของศุภกฤษฎิ์ นิวัฒนากูล (2556) ดังนี้

$$Dis_i = \exp\left(-\text{min_dis}(d_i, q_j)\right) \tag{4.6}$$

$$Dis_i = \exp(-1) \quad = 0.367879$$

ตารางที่ 4.7 ตัวอย่างการคำนวณระยะทางระหว่างคำค้นกับคำในความรู้หรือเอกสาร (Dis_i)

คำค้น q_j	**คำในความรู้หรือเอกสาร** d_i			
	k1	k2	k3	k4
k1	1	0.367879	0.135335	0.135335
k2	0.367879	1	0.367879	0.367879
k3	0.135335	0.367879	1	1
k4	0.135335	0.367879	1	1

* k1, k2 , k3, k4 คือ *Plant product, Cereals, Maize, Corn starch, Corn flour*

2) นำค่าการคำนวณระยะทางระหว่างคำค้นกับคำในความรู้หรือเอกสาร (Dis_i) ในตารางที่ 4.6 ไปคำณวณความคล้ายคลึงกับทุกความรู้หรือเอกสารตามรูปแบบการทดลองในตารางที่ 3.1 ซึ่งการคำนวณรูปแบบการทดลอง A1 คำนวณได้จากสมการที่ 4.7

มหาวิทยาลัยเทคโนโลยีสุรนารี

$$\mathrm{Sim}_{\mathrm{k-corr}} = \left[\frac{\sum\left[\sqrt{\frac{\sum[\mathrm{Dis}_i{}^2]}{\mathrm{T}}} \times Corr_j\right]}{\mathrm{N}}\right] \quad (4.7)$$

เมื่อ

$\mathrm{Sim}_{\mathrm{k-corr}}$ แทน ค่าความคล้ายคลึงเชิงคความหมายจากการให้น้ำหนักคำสำคัญด้วยการวิเคราะห์ Corr

Dis_i แทน ค่าระยะทางสั้นที่สุดระหว่างคำค้นกับคำในความรู้

T แทน จำนวนคำในความรู้

N แทน จำนวนคำสำคัญในแต่ละความรู้

$Corr_j$ แทน ค่าน้ำหนักคำสำคัญด้วยการวิเคราะห์ Corr

ตารางที่ 4.8 การวัดความคล้ายคลึงของคำค้น k1 เทียบกับคำในความรู้หรือเอกสาร

คำค้นจากผู้ใช้ k1	เทียบกับคำในความรู้หรือเอกสาร			เทียบกับคำในความรู้หรือเอกสาร		
	Doc1	Dis_i	$\mathrm{Dis}_i{}^2$	Doc2	Dis_i	$\mathrm{Dis}_i{}^2$
	k1	1	1	k2	0.367879	0.135335
	k3	0.135335	0.018316	k4	0.135335	0.018316
$\sum[\mathrm{Dis}_i{}^2]$			1.018316			0.153651
$\frac{\sum[\mathrm{Dis}_i{}^2]}{\mathrm{T}}$	1.018316/2= 0.509158 T คือจำนวนคำใน Doc1 (k1,k3)			0.153651/2 = 0.076825 T คือจำนวนคำใน Doc2 (k1,k4)		
$\sqrt{\frac{\sum[\mathrm{Dis}_i{}^2]}{\mathrm{T}}}$		0.713553081			0.277173231	
$\sqrt{\frac{\sum[\mathrm{Dis}_i{}^2]}{\mathrm{T}}} \times Corr_j$		**(k1Doc1)** 0.713553081×0.62= 0.44240291			**(k1Doc2)** 0.277173231×0.37= 0.10255410	

* $Corr_j$ *อ้างอิงผลการคำนวณน้ำหนักคำสำคัญด้วยการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ (Corr) ตารางที่ 4.3*

ตารางที่ 4.9 การวัดความคล้ายคลึงของคำค้น k2 เทียบกับคำในความรู้หรือเอกสาร

คำค้นจากผู้ใช้ k2	เทียบกับคำในความรู้หรือเอกสาร			เทียบกับคำในความรู้หรือเอกสาร		
	Doc1	Dis_i	Dis_i^2	Doc2	Dis_i	Dis_i^2
	k1	0.367879	0.135335	k2	1	1
	k3	0.367879	0.135335	k4	0.367879	0.135335
$\sum[Dis_i^2]$			0.270671			1.135335
$\frac{\sum[Dis_i^2]}{T}$	0.270671/2= 0.135335 T คือจำนวนคำใน Doc1 (K1,K3)			1.135335/2 = 0.567668 T คือจำนวนคำใน Doc2 (K1,K4)		
$\sqrt{\frac{\sum[Dis_i^2]}{T}}$		0.713553081			0.277173231	
$\sqrt{\frac{\sum[Dis_i^2]}{T}} \times Corr_j$		**(k2Doc1)** 0.713553081×0.31= 0.44240291			**(k2Doc2)** 0.277173231×0.31= 0.17184740	

* $Corr_j$ *อ้างอิงผลการคำนวณน้ำหนักคำด้วยการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ (Corr) ตารางที่ 4.3*

ฉะนั้น $\sum\left[\sqrt{\frac{\sum[Dis_i^2]}{T}} \times Corr_j\right]$ ของDoc1 =(k1Doc1) + (k2Doc1) = 0.88480582

Doc2 =(k1Doc2) + (k2Doc2) = 0.27440150

$\left[\frac{\sum\left[\sqrt{\frac{\sum[Dis_i^2]}{T}} \times Corr_j\right]}{N}\right]$ ของ Doc1 =(k1Doc1) + (k2Doc1) /จำนวนคำค้น

= (k1Doc1) + (k2Doc1) /2 = 0.44240291

Doc2 =(k1Doc2) + (k2Doc2) /จำนวนคำค้น

=(k1Doc2) + (k2Doc2) /2 = 0.13720075

มหาวิทยาลัยเทคโนโลยีสุรนารี

ดังนั้น กรณีผู้ใช้ป้อนคำค้น k1, k2 เข้ามาค้นคืนความรู้หรือเอกสาร ความคล้ายคลึงเชิงความหมายกับความรู้ Doc1 เท่ากับ 0.44240291 และ Doc2 เท่ากับ 0.13720075

จากตัวอย่างการคำนวณรูปแบบการทดลอง A1 งานวิจัยนี้ได้ทดลองวัดความคล้ายคลึงเชิงความหมายตามรูปแบบการทดลองในตารางที่ 3.1 ดังนี้

A2 คำนวณได้จากสมการที่ 4.8

$$\mathrm{Sim}_{\mathrm{k-corr-sw}} = \left[\frac{\sum\left[\sqrt{\frac{\sum[\mathrm{Dis}_i{}^2]}{\mathrm{T}}} \times Corr_j \times SW_j\right]}{\mathrm{N}}\right] \tag{4.8}$$

เมื่อ

$\mathrm{Sim}_{\mathrm{k-corr-}sw}$ แทน ค่าความคล้ายคลึงเชิงคความหมายจากการให้น้ำหนักคำสำคัญด้วยการวิเคราะห์ Corr กับ SW

Dis_i แทน ค่าระยะทางสั้นที่สุดระหว่างคำค้นกับคำในความรู้

T แทน จำนวนคำในเอกสาร

N แทน จำนวนคำสำคัญในแต่ละความรู้

$Corr_j$ แทน ค่าน้ำหนักคำสำคัญด้วยการวิเคราะห์ Corr

SW_j แทน ค่าความสำคัญของคำค้นจากผู้ใช้

A3 คำนวณได้จากสมการที่ 4.9

$$\mathrm{Sim}_{\mathrm{k-corr-pw}} = \sum\left[\frac{\sum\left[\sqrt{\frac{\sum[\mathrm{Dis}_i{}^2]}{\mathrm{T}}} \times Corr_j\right]}{\mathrm{N}} \times PW_j\right] \tag{4.9}$$

เมื่อ

$Sim_{\text{k}-\text{corr}-pw}$ แทน ค่าความคล้ายคลึงเชิงคความหมายจากการให้น้ำหนักคำสำคัญด้วยการวิเคราะห์ Corr กับ PW

Dis_i แทน ค่าระยะทางสั้นที่สุดระหว่างคำค้นกับคำในความรู้

T แทน จำนวนคำในความรู้

N แทน จำนวนคำสำคัญในแต่ละความรู้

$Corr_j$ แทน ค่าน้ำหนักคำสำคัญด้วยการวิเคราะห์ Corr

PW_j แทน ค่าความสำคัญของคำค้นที่ปรากฏในความรู้

A4 คำนวณได้จากสมการที่ 4.10

$$Sim_{\text{k}-\text{corr}-\text{sw}-\text{pw}} = \sum \left[\frac{\sum \left[\sqrt{\frac{\sum [\text{Dis}_i{}^2]}{\text{T}}} \times Corr_j \times SW_j \right]}{\text{N}} \times PW_k \right] \quad (4.10)$$

เมื่อ

$Sim_{\text{k}-\text{corr}-\text{sw}-pw}$ แทน ค่าความคล้ายคลึงเชิงคความหมายจากการให้น้ำหนักคำสำคัญด้วยการวิเคราะห์ Corr กับ SW และ PW

Dis_i แทน ค่าระยะทางสั้นที่สุดระหว่างคำค้นกับคำในความรู้

T แทน จำนวนคำในความรู้

N แทน จำนวนคำสำคัญในแต่ละเอกสาร

$Corr_j$ แทน ค่าน้ำหนักคำสำคัญด้วยการวิเคราะห์ Corr

SW_j แทน ค่าน้ำหนักลำดับคำค้นจากผู้ใช้

PW_k แทน ค่าความสำคัญของคำค้นที่ปรากฏในความรู้

มหาวิทยาลัยเทคโนโลยีสุรนารี

<u>รูปแบบการทดลอง B</u> คือนำผลการคำนวณน้ำหนักคำสำคัญด้วยการวิเคราะห์สัดส่วนของจำนวนเอกสารทั้งหมดกับจำนวนของเอกสารที่ปรากฏคำที่สนใจ (IDF) มี 4 รูปแบบ มีสมการ ดังนี้

B1 คำนวณได้จากสมการที่ 4.11

$$Sim_{k-idf} = \left[\frac{\sum \left[\sqrt{\frac{\sum [\mathrm{Dis}_i{}^2]}{T}} \times IDF_j \right]}{N} \right] \quad (4.11)$$

เมื่อ

Sim_{k-idf} แทน ค่าความคล้ายคลึงเชิงคความหมายจากการให้น้ำหนักคำสำคัญด้วยการวิเคราะห์ IDF

Dis_i แทน ค่าระยะทางสั้นที่สุดระหว่างคำค้นกับคำในความรู้

T แทน จำนวนคำในความรู้

N แทน จำนวนคำสำคัญในแต่ละความรู้

IDF_j แทน ค่าน้ำหนักคำสำคัญด้วยการวิเคราะห์ IDF

B2 คำนวณได้จากสมการที่ 4.12

$$Sim_{k-idf-sw} = \left[\frac{\sum \left[\sqrt{\frac{\sum [\mathrm{Dis}_i{}^2]}{T}} \times IDF_j \times SW_j \right]}{N} \right] \quad (4.12)$$

เมื่อ

$Sim_{k-idf-sw}$ แทน ค่าความคล้ายคลึงเชิงคความหมายจากการให้น้ำหนักคำสำคัญด้วยการวิเคราะห์ IDF กับ SW

Dis_i	แทน	ค่าระยะทางสั้นที่สุดระหว่างคำค้นกับคำในความรู้
T	แทน	จำนวนคำในความรู้
N	แทน	จำนวนคำสำคัญในแต่ละความรู้
IDF_j	แทน	ค่าน้ำหนักคำสำคัญด้วยการวิเคราะห์ IDF
SW_j	แทน	ค่าความสำคัญของคำค้นจากผู้ใช้

B3 คำนวณได้จากสมการที่ 4.13

$$Sim_{k-idf-pw} = \sum \left[\frac{\sum \left[\sqrt{\frac{\sum [\mathrm{Dis}_i{}^2]}{T}} \times IDF_j \right]}{N} \times PW_k \right] \quad (4.13)$$

เมื่อ

$Sim_{k-idf-PW}$	แทน	ค่าความคล้ายคลึงเชิงคความหมายจากการให้น้ำหนักคำสำคัญด้วยการวิเคราะห์ IDF กับ PW
Dis_i	แทน	ค่าระยะทางสั้นที่สุดระหว่างคำค้นกับคำในความรู้
T	แทน	จำนวนคำในความรู้
N	แทน	จำนวนคำสำคัญในแต่ละความรู้
IDF_j	แทน	ค่าน้ำหนักคำสำคัญด้วยการวิเคราะห์ IDF
PW_j	แทน	ค่าความสำคัญของคำค้นที่ปรากฏในความรู้

B4 คำนวณได้จากสมการที่ 4.14

$$Sim_{k-idf-sw-pw} = \sum \left[\frac{\sum \left[\sqrt{\frac{\sum [\mathrm{Dis}_i{}^2]}{T}} \times IDF_j \times SW_j \right]}{N} \times PW_k \right] \quad (4.14)$$

มหาวิทยาลัยเทคโนโลยีสุรนารี

เมื่อ

$Sim_{k-idf-SW-PW}$	แทน	ค่าความคล้ายคลึงเชิงคความหมายจากการให้น้ำหนักคำสำคัญด้วยการวิเคราะห์ IDF กับ SW และ PW
Dis_i	แทน	ค่าระยะทางสั้นที่สุดระหว่างคำค้นกับคำในความรู้
T	แทน	จำนวนคำในความรู้
N	แทน	จำนวนคำสำคัญในแต่ละความรู้
IDF_j	แทน	ค่าน้ำหนักคำสำคัญด้วยการวิเคราะห์ IDF
SW_j	แทน	ค่าความสำคัญของคำค้นจากผู้ใช้
PW_j	แทน	ค่าความสำคัญของคำค้นที่ปรากฏในความรู้

4.2 การประเมินประสิทธิภาพการสืบค้นเชิงความหมาย

การประเมินประสิทธิภาพการสืบค้นเชิงความหมายเริ่มจากผู้ใช้ใส่คำสำคัญในการค้นหาเข้ามาผ่านส่วนติดต่อผู้ใช้งานเปรียบเทียบกับความรู้หรือเอกสารที่เป็นคำตอบที่มีการสร้างชุดเฉลยไว้แล้วจากผู้เชี่ยวชาญ แสดงผลคำตอบนั้นผ่านเว็บเบราว์เซอร์ พัฒนาโดยใช้ภาษาพีเอช และ RDF API for PHP สำหรับเชื่อมต่อระหว่างโครงสร้างออนโทโลยีกับการสืบค้นความรู้ สอบถามความรู้จากการประมวลข้อคำถามหรือคำค้นโดยใช้ภาษาสปาร์เก็ล เพื่ออำนวยความสะดวกในการทดสอบระบบ โดยการทดสอบระบบจะนำเสนอเป็น 3 ส่วน ได้แก่ (1) การสำรวจคำค้นความรู้ด้านพืชไร่โดยผู้ใช้ (2) การประเมินประสิทธิภาพการให้น้ำหนักคำสำคัญ และ (3) การประเมินประสิทธิภาพการวัดความคล้ายคลึงเชิงความหมาย โดยมีรายละเอียดดังนี้

4.2.1 ผลประเมินการสำรวจคำค้นความรู้ด้านพืชไร่โดยผู้ใช้

1) ข้อมูลพื้นฐานของผู้ตอบแบบสอบถาม โดยผู้ตอบแบบสอบถาม ได้แก่ ประชาชนเขตอำเภอเมืองกาฬสินธุ์ จำนวน 384 คน ดังตารางที่ 4.10

ตารางที่ 4.10 ข้อมูลทั่วไปของผู้ตอบแบบสอบถามคำค้นสำหรับสืบค้นความรู้ด้านพืชไร่

ข้อมูลทั่วไปของผู้ตอบแบบสอบถาม	จำนวน	ร้อยละ
1. เพศ		
ชาย	171	44.50
หญิง	213	55.50
รวม	384	100.00
2. อายุ		
ต่ำกว่า 20 ปี	79	20.60
21-40 ปี	145	37.80
41-60 ปี	119	31.00
ตั้งแต่ 60 ปีขึ้นไป	41	10.70
รวม	384	100.00
3. อาชีพ		
เกษตรกร	132	34.40
นักเรียนหรือนักศึกษา	108	28.10
ธุรกิจส่วนตัวหรือค้าขาย	85	22.10
ข้าราชการหรือพนักงานรัฐวิสาหกิจ	39	10.20
ลูกจ้างหรือพนักงานบริษัท	20	5.20
รวม	384	100.00

จากตารางที่ 4.10 พบว่า ผู้ตอบแบบสอบถามคำค้นสำหรับความรู้ด้านพืชไร่ ส่วนใหญ่เป็นเพศหญิง ร้อยละ (55.50) อายุระหว่าง 21-40 ปี ร้อยละ (37.80) รองลงมา 41-60 ปี ร้อยละ (31.00) อาชีพเกษตรกร ร้อยละ (34.40) รองลงมานักเรียนหรือนักศึกษา ร้อยละ (28.10)

2) ความคิดเห็นเกี่ยวกับคำค้นสำหรับสืบค้นความรู้ด้านพืชไร่

แบบสอบถามคำค้นสำหรับความรู้ด้านพืชไร่ มีจำนวน 181 คำ ให้ผู้ตอบแบบสอบถามเลือกคำค้นที่จะนำมาใช้สืบค้นความรู้ พบว่ามีผู้ตอบแบบสอบถามคำค้นมีความคิดเห็นเกี่ยวกับคำค้นสำหรับความรู้ด้านพืชไร่ที่ได้ร้อยละ 25 ขึ้นไป มีจำนวน 100 คำ *(รายละเอียดภาคผนวก ข)* โดยผู้เชี่ยวชาญเลือก 1-3 คำค้น แนวคิดมาจาก เจษฎา สิงห์ทองชัยและคณะ (Jatsada Singthongchai, et. al., 2012) ได้คำค้นจำนวน 100 ชุด เพื่อใช้ในการประเมินประสิทธิภาพการ

สืบค้น 2 ส่วน ได้แก่ การประเมินประสิทธิภาพการให้น้ำหนักคำสำคัญ และประเมินประสิทธิภาพการวัดความคล้ายคลึง ดังนี้

4.2.2 การประเมินประสิทธิภาพการให้น้ำหนักคำสำคัญ

การให้น้ำหนักคำสำคัญเริ่มจากผู้ใช้ใส่คำสำคัญในการค้นเข้ามาผ่านส่วนติดต่อผู้ใช้งาน แสดงผลลัพธ์ผ่านเว็บเบราว์เซอร์ โดยเปรียบเทียบวิธีการให้น้ำหนักคำด้วย การวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ (Corr) กับการให้น้ำหนักด้วยการวิเคราะห์สัดส่วนของจำนวนความรู้ทั้งหมดกับจำนวนของความรู้ที่ปรากฏคำที่สนใจ (IDF) ความคาดหวังในกระบวนการให้น้ำหนักคำต้องการให้มีค่าเข้าใกล้ 1 และค่าน้ำหนักที่ได้จะส่งผลต่อการเรียงลำดับข้อมูล การประเมินประสิทธิภาพการให้น้ำหนักคำ ประกอบด้วย (1) ค่าความแม่นยำเฉลี่ย (MRR) ซึ่งคำค้นจะถูกนำไปค้นคืนกับความรู้หรือเอกสารที่เป็นคำตอบจากการสร้างชุดเฉลยไว้แล้วของผู้เชี่ยวชาญเทียบอันดับกับคำตอบที่ระบบเลือก (2) ค่าความผิดพลาด (SSE) โดยมีรายละเอียดดังนี้

ตัวอย่างที่ 4.7 การวัดประสิทธิภาพด้วยค่าความแม่นยำเฉลี่ย (MRR)

ค่าความแม่นยำเฉลี่ยเป็นการวัดประสิทธิภาพของการจัดอันดับของคำตอบ หรือผลลัพธ์จากการสืบค้นความรู้ ซึ่งผู้เชี่ยวชาญมีการทำเฉลยจัดลำดับคำตอบความรู้หรือเอกสารไว้แล้ว โดยเปรียบเทียบกับการจัดอันดับคำตอบของระบบ ซึ่งมีตัวอย่างการคำนวณ ดังนี้

ตารางที่ 4.11 ตัวอย่างการวัดประสิทธิภาพด้วยค่าความแม่นยำเฉลี่ย (MRR) จากการสืบค้น

คำค้นจากผู้ใช้	อันดับคำตอบ		คำนวณการจัดอันดับ
	ผู้เชี่ยวชาญเลือก	ระบบเลือก	
(1) กรณีเลือกคำตอบการจัดอันดับตรงกัน			
อ้อย	Doc1	Doc1	1/(4×1) =0.25
	Doc2	Doc2	1/(4×1) =0.25
	Doc3	Doc3	1/(4×1) =0.25
	Doc4	Doc4	1/(4×1) =0.25
ค่าความแม่นยำเฉลี่ย (MRR)			1

ตารางที่ 4.11 ตัวอย่างการวัดประสิทธิภาพด้วยค่าความแม่นยำเฉลี่ย (MRR) จากการสืบค้น (ต่อ)

คำค้นจากผู้ใช้	อันดับคำตอบ		คำนวณการจัดอันดับ
	ผู้เชี่ยวชาญเลือก	ระบบเลือก	
(2) กรณีเลือกคำตอบการจัดอันดับไม่ตรงกัน			
อ้อย	Doc1	Doc1	1/(4×1) =0.25
	Doc3	Doc2	1/(4×2) =0.125 *(Doc3 อยู่ต่ำกว่า 1 อันดับ)*
	Doc2	Doc3	1/(4×1) =0.25
	Doc4	Doc4	1/(4×1) =0.25
ค่าความแม่นยำเฉลี่ย (MRR)			0.875
(3) กรณีเลือกคำตอบการจัดอันดับไม่ตรงกัน			
อ้อย	Doc3	*Doc1*	1/(4×3) =0.083 *(Doc3 อยู่ต่ำกว่า 2 อันดับ)*
	Doc1	Doc2	1/(4×1) =0.25 *(Doc1 อยู่สูงกว่า 1 อันดับ)*
	Doc4	Doc3	1/(4×2) =0.125 *(Doc4 อยู่ต่ำกว่า 1 อันดับ)*
	Doc2	Doc4	1/(4×1) =0.25 *(Doc2 อยู่สูงกว่า 2 อันดับ)*
ค่าความแม่นยำเฉลี่ย (MRR)			0.71
* คำตอบที่ระบบเลือกกรณีอยู่สูงกว่าคำตอบที่ผู้เชี่ยวชาญเลือกเป็นการจัดลำดับที่ดีกว่า			

ตัวอย่างที่ 4.8 การคำนวณค่าความผิดพลาด (SSE)

เมื่อผู้ใช้ระบุคำ “อ้อย” “การปลูก” เป็นคำค้น สมมุติน้ำหนักคำสำคัญ เท่ากับ 0.25 และ 0.5 ตามลำดับ ซึ่งมีคำสำคัญ จำนวน 2 คำ การคำณวณค่าความผิดพลาดแทนค่าในสมการที่ 3.4 ดังนี้

ค่าความผิดพลาด = $((1-\text{ค่าน้ำหนักคำสำคัญที่ 1})^{\text{จำนวนคำค้น}}(1-\text{ค่าน้ำหนักคำสำคัญที่ 1})^{\text{จำนวนคำค้น}}$

ค่าความผิดพลาด = $((1-0.25)^2+(1-0.5)^2)/2$

ซึ่งค่าน้ำหนักคำสำคัญที่ใกล้ 1 แล้วนำไปลบ (–) กับ 1 จะได้ค่าความผิดพลาดที่น้อย

ตารางที่ 4.12 ตัวอย่างประสิทธิภาพการให้น้ำหนักคำค้น

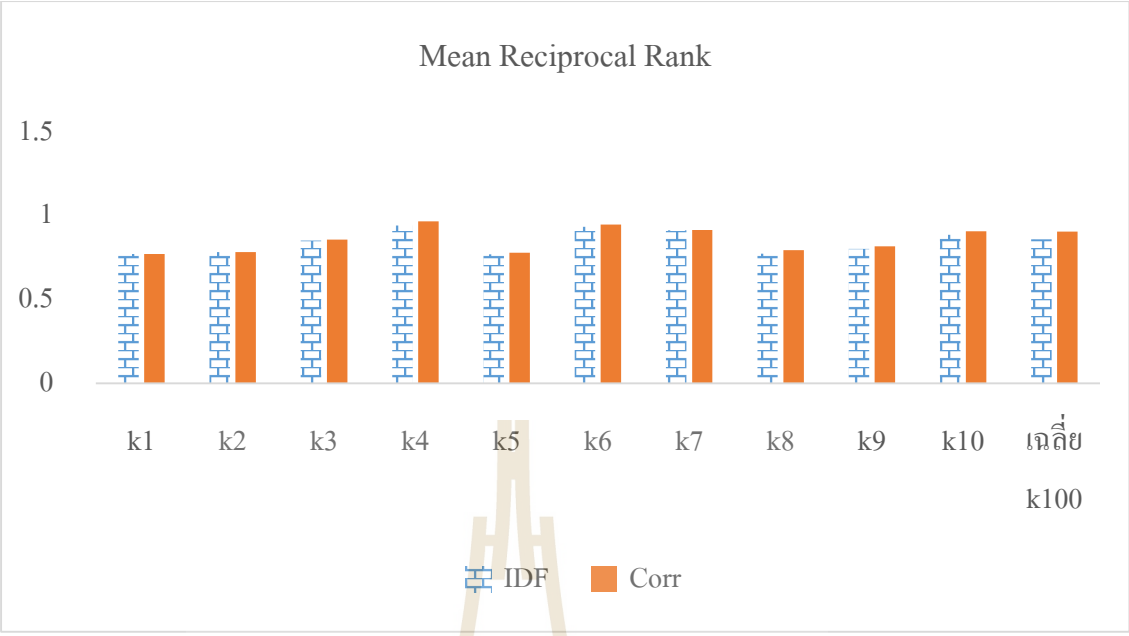
1-3 คำค้น	ความแม่นยำเฉลี่ย		ความผิดพลาด	
	IDF	Corr	IDF	Corr
พืชไร่	0.773	0.773	0.080	0.087
มันสำปะหลัง พันธุ์ ดินที่เหมาะสม	0.784	0.784	0.100	0.093
อ้อย การปลูก	0.853	0.859	0.098	0.093
น้ำตาล การแปรรูป	0.942	0.968	0.102	0.102
ข้าวโพด เทคโนโลยีหลังการเก็บเกี่ยว	0.771	0.780	0.092	0.086
ข้าวโพด พันธุ์	0.934	0.948	0.102	0.098
ข้าวโพด การบำรุงรักษา วิธีเพาะปลูก	0.913	0.916	0.102	0.099
ปุ๋ย	0.775	0.796	0.090	0.087
อ้อย โรค สารป้องกันกำจัดแมลง	0.803	0.819	0.090	0.088
มันสำปะหลัง สารฆ่าแมลง	0.888	0.909	0.095	0.096
เฉลี่ย 100 คำค้น	**0.860**	**0.906**	**0.104**	**0.097**

ที่มา : จากการศึกษา (รายละเอียดภาคผนวก ข)

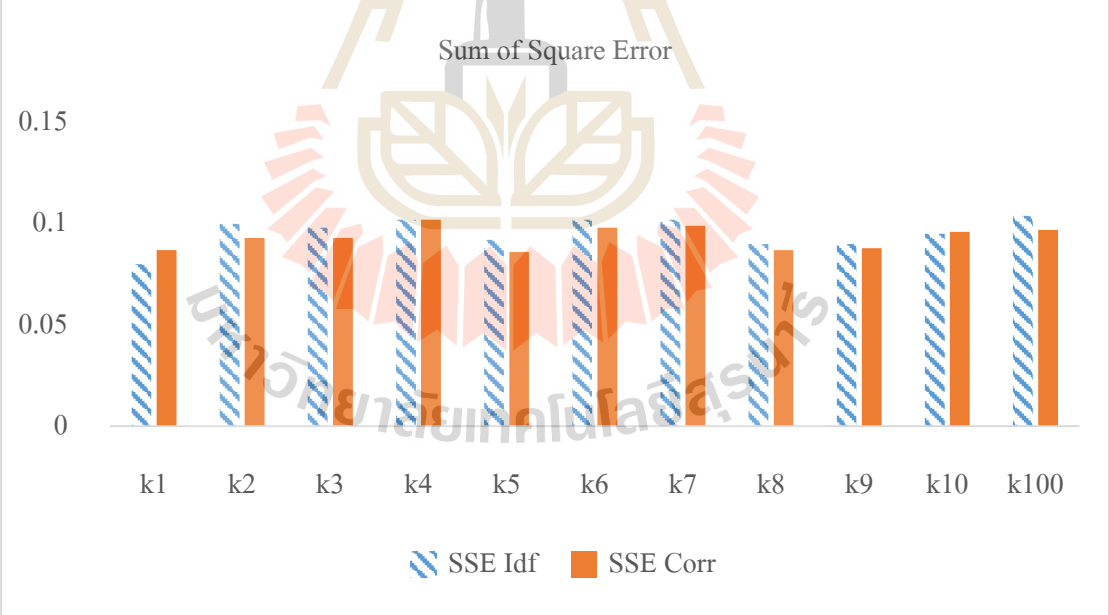
**Corr หมายถึง การให้น้ำหนักด้วยการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient)*

IDF หมายถึง การให้น้ำหนักด้วยการวิเคราะห์สัดส่วนของจำนวนความรู้ทั้งหมดกับจำนวนของความรู้ที่ปรากฏคำที่สนใจ (Inverse document frequency)

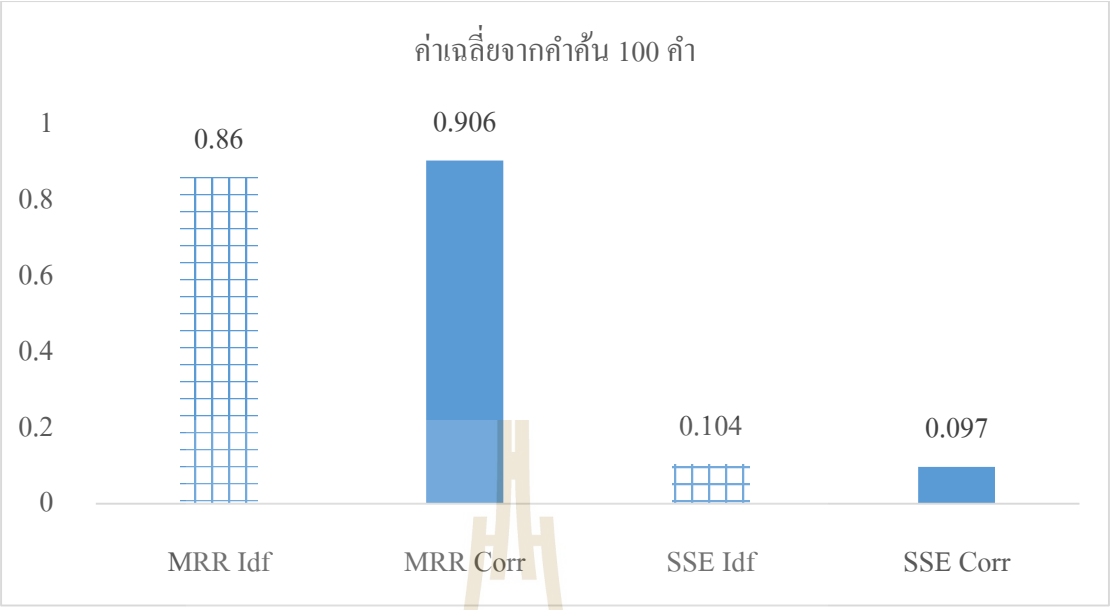
จากตารางที่ 4.12 พบว่าการให้น้ำหนักคำด้วยการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ (Corr) มีค่าความแม่นยำเฉลี่ย (MRR) เท่ากับ 0.906 และมีค่าความผิดพลาด (SSE) เท่ากับ 0.097 ซึ่งให้ผลดีกว่าทั้งในด้านของความแม่นยำเฉลี่ย และมีค่าความผิดพลาดที่น้อยกว่าวิธีการให้น้ำหนักด้วย IDF แสดงดังรูปที่ 4.9 – 4.11



รูปที่ 4.8 ผลการเปรียบเทียบค่าความแม่นยำเฉลี่ย (MRR)



รูปที่ 4.9 ผลการเปรียบเทียบค่าความผิดพลาด (SSE)



รูปที่ 4.10 ผลเปรียบเทียบค่าความแม่นยำเฉลี่ย และค่าความผิดพลาด

ข้อค้นพบสำหรับการในกระบวนการนี้พบว่าขั้นตอนการให้น้ำหนักคำด้วยการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ (Corr) มีประสิทธิภาพของการจัดอันดับแม่นยำเฉลี่ยดีกว่าการให้น้ำหนักคำด้วยการวิเคราะห์ค่าสัดส่วนของจำนวนเอกสารทั้งหมดกับจำนวนของเอกสารที่ปรากฏคำที่สนใจ (IDF) และจากการคำนวณค่าความผิดพลาด พบว่าขั้นตอนวิธีการให้น้ำหนักคำด้วยการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์มีความผิดพลาดต่ำกว่าวิธีการให้น้ำหนักด้วย IDF

4.2.3 การประเมินประสิทธิภาพการวัดความคล้ายคลึงเชิงความหมาย

งานวิจัยนี้ได้ทดลองประเมินประสิทธิภาพการวัดความคล้ายคลึงเชิงความหมายด้วยค่าความแม่นยำเฉลี่ย (MRR) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และ ค่าความถ่วงดุล (F-measure) ตามรูปแบบการทดลองตารางที่ 3.1 ใน 2 รูปแบบ คือ

(1) รูปแบบการทดลอง A ที่นำผลการคำนวณน้ำหนักคำสำคัญด้วยการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ (Corr) มาใช้สำหรับการวัดความคล้ายคลึงเชิงความหมาย แบ่ง 4 ลักษณะ ได้แก่

(A1) การนำค่าระยะทางสั้นที่สุดระหว่างคำค้นกับคำในความรู้หรือเอกสาร โดยวัดระยะทางระหว่างโหนดของคำในโครงสร้าง มาคำนวณกับค่าน้ำหนักคำสำคัญด้วยวิเคราะห์สัมประสิทธิ์สหสัมพันธ์

(A2) การนำค่าระยะทางสั้นที่สุดระหว่างคำค้นกับคำในความรู้หรือเอกสาร โดยวัดระยะทางระหว่างโหนดของคำในโครงสร้าง มาคำนวณกับค่าน้ำหนักคำสำคัญด้วยวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ และค่าลำดับความสำคัญคำค้นจากผู้ใช้ (SW)

(A3) การนำค่าระยะทางสั้นที่สุดระหว่างคำค้นกับคำในความรู้หรือเอกสาร โดยวัดระยะทางระหว่างโหนดของคำในโครงสร้าง มาคำนวณกับค่าน้ำหนักคำสำคัญด้วยวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ และค่าลำดับความสำคัญคำค้นจากตำแหน่งของคำค้นที่ปรากฏในความรู้ (PW)

(A4) การนำค่าระยะทางสั้นที่สุดระหว่างคำค้นกับคำในความรู้หรือเอกสาร โดยวัดระยะทางระหว่างโหนดของคำในโครงสร้าง มาคำนวณกับค่าน้ำหนักคำสำคัญด้วยวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ และค่าลำดับความสำคัญคำค้นจากผู้ใช้ และค่าลำดับความสำคัญคำค้นจากตำแหน่งของคำค้นที่ปรากฏในความรู้

(2) รูปแบบการทดลอง B ที่นำผลการคำนวณน้ำหนักคำสำคัญด้วยการวิเคราะห์สัดส่วนของจำนวนเอกสารทั้งหมดกับจำนวนของเอกสารที่ปรากฏคำที่สนใจ (IDF) ได้แก่

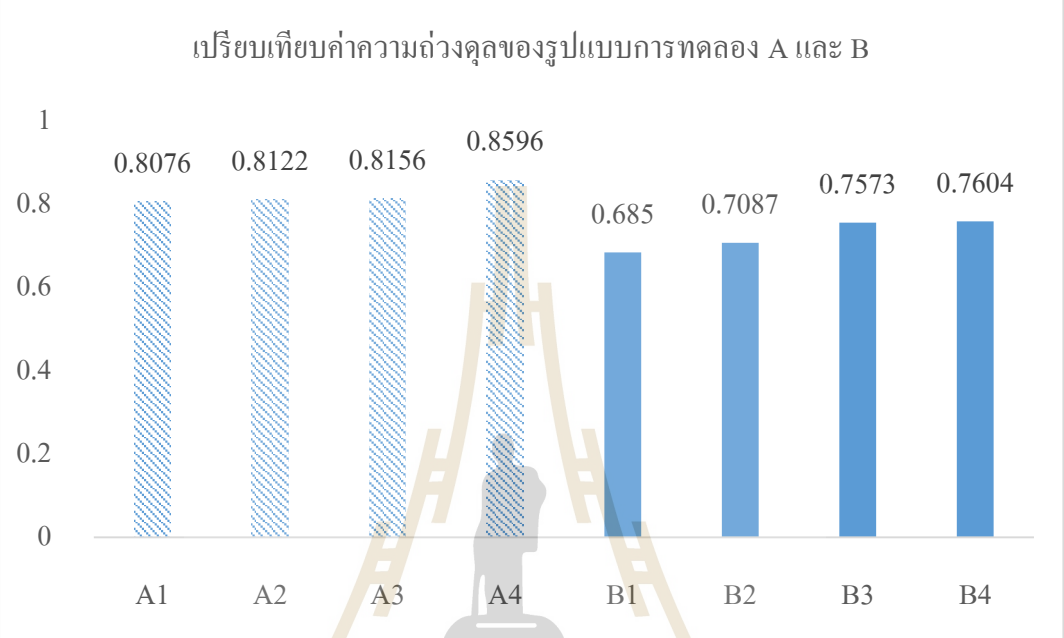
(B1) การนำค่าระยะทางสั้นที่สุดระหว่างคำค้นกับคำในความรู้หรือเอกสาร โดยวัดระยะทางระหว่างโหนดของคำในโครงสร้าง มาคำนวณกับสัดส่วนของจำนวนเอกสารทั้งหมดกับจำนวนของเอกสารที่ปรากฏคำที่สนใจ

(B2) การนำค่าระยะทางสั้นที่สุดระหว่างคำค้นกับคำในความรู้หรือเอกสาร โดยวัดระยะทางระหว่างโหนดของคำในโครงสร้าง มาคำนวณกับสัดส่วนของจำนวนเอกสารทั้งหมดกับจำนวนของเอกสารที่ปรากฏคำที่สนใจ และค่าลำดับความสำคัญคำค้นจากผู้ใช้ (SW)

(B3) การนำค่าระยะทางสั้นที่สุดระหว่างคำค้นกับคำในความรู้หรือเอกสาร โดยวัดระยะทางระหว่างโหนดของคำในโครงสร้าง มาคำนวณกับสัดส่วนของจำนวนเอกสารทั้งหมดกับจำนวนของเอกสารที่ปรากฏคำที่สนใจ และค่าลำดับความสำคัญคำค้นจากตำแหน่งของคำค้นที่ปรากฏในความรู้ (PW)

(B4) การนำค่าระยะทางสั้นที่สุดระหว่างคำค้นกับคำในความรู้หรือเอกสาร โดยวัดระยะทางระหว่างโหนดของคำในโครงสร้าง มาคำนวณกับสัดส่วนของจำนวนเอกสารทั้งหมดกับจำนวนของเอกสารที่ปรากฏคำที่สนใจ และค่าลำดับความสำคัญคำค้นจากผู้ใช้ และค่าลำดับความสำคัญคำค้นจากตำแหน่งของคำค้นที่ปรากฏในความรู้

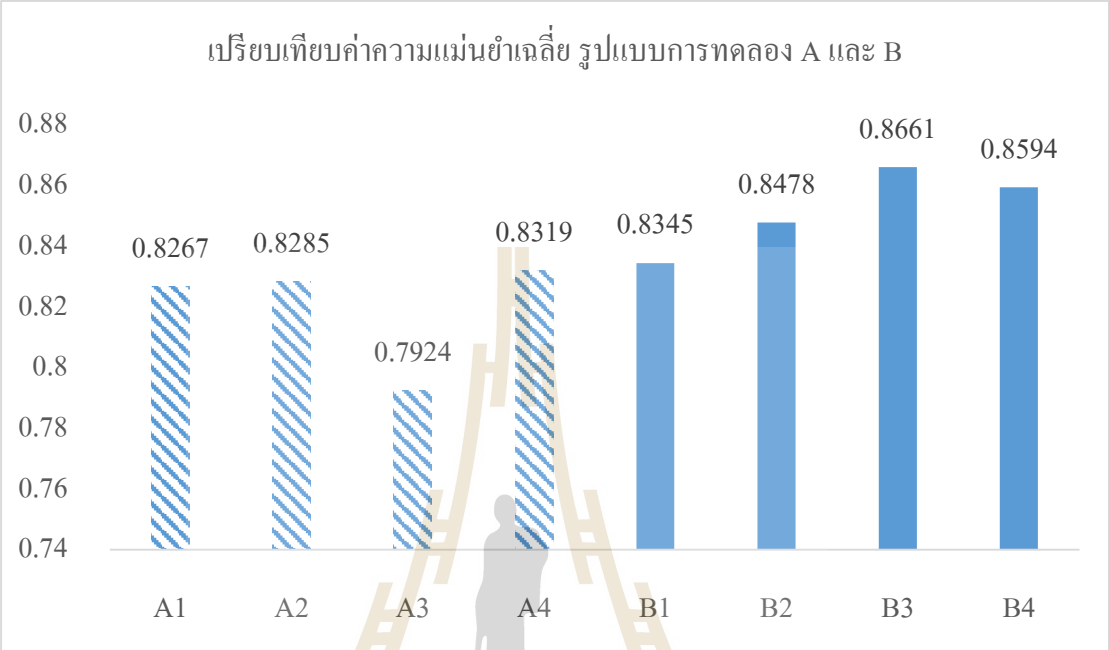
ผลการประเมินประสิทธิภาพการวัดความคล้ายคลึงเชิงความหมาย จากการทดลองด้วย 1-3 คำค้น แนวคิดมาจาก เจษฎา สิงห์ทองชัยและคณะ (Jatsada Singthongchai, et. al., 2012) ปรากฏดังนี้



รูปที่ 4.11 เปรียบเทียบประสิทธิภาพค่าความถ่วงดุลรูปแบบการทดลอง A และ B

จากรูปที่ 4.11 ผลการเปรียบเทียบประสิทธิภาพการวัดความคล้ายคลึงจากการสืบค้นโดยใช้ค่าความถ่วงดุล (F-measure) ที่สะท้อนจากค่าความแม่นยำ (Precision) และ ค่าความระลึก (Recall) ของรูปแบบการทดลอง 8 ลักษณะ คือ A1, A2, A3, A4, B1, B2, B3, B4 มีผลต่อประสิทธิภาพการสืบค้นที่แตกต่างกัน พบว่ารูปแบบการทดลอง A4 ให้ผลการสืบค้นมีค่าความถ่วงดุลสูงกว่าการทดลองทุกลักษณะ มีค่าความถ่วงดุลเท่ากับ 0.8596 เพราะการกำหนดเกณฑ์เลือกความรู้หรือเอกสารแบบไดนามิกเกณฑ์ที่เลือกจากค่าต่ำสุดของความรู้หรือเอกสารที่ค้นพบ (เลือกทุกความรู้ที่พบว่าเกี่ยวข้อง) ทำให้ได้ผลลัพธ์ที่ได้มีปริมาณมาก มีความครบถ้วนของความรู้เมื่อเทียบกับความรู้ที่ควรได้ทั้งหมด จากอัตราส่วนระหว่างจำนวนความรู้ที่ค้นคืนถูกต้องกับจำนวนความรู้ที่ถูกต้องทั้งหมดของระบบจึงส่งผลให้ค่าความระลึก (Recall) มีค่าเป็น 1 เสมอจึงส่งผลต่อความถ่วงดุล และที่สำคัญการทดลองลักษณะของ A4 ได้นำค่าการคำนวณระยะทางระหว่างโหนดของคำในโครงสร้างลักษณะคำกว้างกว่า คำแคบกว่า คำที่เกี่ยวข้อง และคำเหมือน ที่คำนึงถึงความหมายของคำมาคำนวณร่วมกับการให้ความสำคัญกับผู้ใช้จากลำดับคำค้น และความสำคัญของ

คำศัพท์ที่ปรากฏในตำแหน่งต่าง ๆ ของความรู้อาจเป็นสาเหตุที่สำคัญที่ทำให้ความครบถ้วนของเนื้อหาความรู้ที่เกี่ยวข้องกับคำค้น และผลลัพธ์ที่เกี่ยวข้องค่อนข้างสูง



รูปที่ 4.12 เปรียบเทียบประสิทธิภาพค่าความแม่นยำเฉลี่ยรูปแบบการทดลอง A และ B

จากรูปที่ 4.12 ผลการเปรียบเทียบประสิทธิภาพการวัดความคล้ายคลึงจากการสืบค้นโดยใช้ค่าความแม่นยำเฉลี่ย (MRR) ของรูปแบบการทดลอง 8 ลักษณะ คือ A1, A2, A3, A4, B1, B2, B3, B4 มีผลต่อประสิทธิภาพการสืบค้นที่แตกต่างกัน พบว่ารูปแบบการทดลอง B3 ให้ผลการสืบค้นมีค่าความแม่นยำเฉลี่ย สูงกว่าการทดลองทุกลักษณะ มีค่าความแม่นยำเฉลี่ย เท่ากับ 0.8661 และข้อสังเกตุคือการทดลองลักษณะ B1, B2, B3 มีค่าความแม่นยำเฉลี่ยสูงกว่าการทดลอง A1, A2, A3, A4 ทุกกรณี เพราะรูปแบบการทดลอง B1, B2, B3 และ B4 มีการคำนวณน้ำหนักคำสำคัญมีการใช้ความถี่ของคำที่ปรากฏในความรู้หรือเอกสารเพียงอย่างเดียวไม่มีการคำนึงถึงความหมายของคำ โดยถือว่าแต่ละคำเป็นอิสระต่อกัน ซึ่งความเป็นจริงแล้วคำแต่ละคำอาจมีความสัมพันธ์กันเชิงความหมาย หรือใกล้เคียงกัน จึงเป็นสาเหตุสำคัญไม่สามารถเชื่อมโยงไปยังความรู้หรือเอกสารที่เกี่ยวได้ครบถ้วน จึงทำให้ผลลัพธ์ที่ได้มีปริมาณน้อยส่งผลต่อค่าความแม่นยำเฉลี่ยที่สูง

บทที่ 5
สรุป อภิปราย และข้อเสนอแนะ

ในส่วนของบทที่ 5นี้เป็นการสรุป และอภิปรายผลการวิจัยออกแบบและพัฒนาแบบจำลองการสืบค้นเชิงความหมายสำหรับความรู้ด้านพืชไร่ พร้อมทั้งได้ข้อเสนอแนะ และแนวทางการประยุกต์งานวิจัยครั้งต่อไป ซึ่งมีรายละเดียดดังต่อไปนี้

5.1 สรุปผลการวิจัย

5.2 อภิปรายผลการวิจัย

5.3 ข้อจำกัดของการวิจัย

5.4 การประยุกต์ผลการวิจัย

5.5 ข้อเสนอแนะ

5.1 สรุปผลการวิจัย

งานวิจัยนี้ได้ศึกษาและวิเคราะห์ขั้นตอนวิธีการสืบค้นความรู้ด้านพืชไร่ โดยใช้ออนโทโลยีอธิบายความหมาย ความสัมพันธ์และโครงสร้างของข้อมูลผ่านภาษาโอดับเบิ้ลยูแอล (Web Ontology Language: OWL) ซึ่งประยุกต์การจัดความรู้ให้เป็นระบบโดยใช้ออนโทโลยีด้านไม้ผลเศรษฐกิจที่วิจัยและพัฒนาขึ้นโดยศุภกฤษฎิ์ นิวัฒนากูล (2556)มีรายละเอียดของการสรุปผลการวิจัยดังนี้

5.1.1 การออกแบบจำลองการสืบค้นเชิงความหมายสำหรับความรู้ด้านพืชไร่ (Semantic Search Model for Agronomy Knowledge: SeeAgrow) ประกอบด้วยการทำงาน 2ส่วน คือ การจัดความรู้ให้เป็นระบบ และการสืบค้นเชิงความหมาย

5.1.1.1 การจัดความรู้ให้เป็นระบบ ประกอบด้วย 3 ขั้นตอนหลัก คือ 1) การออกแบบโครงสร้างคลาสจัดเก็บความรู้ด้านพืชไร่ 2) การกำหนดความสัมพันธ์ของคำสำคัญ และสร้างคลังคำ และ 3) การประมวลความรู้ด้านพืชไร่ ได้แก่ 3.1) การคำนวณความถี่ของคำที่ปรากฏในความรู้พืชไร่ 3.2) การคำนวณน้ำหนักคำ และ 3.3) การคำนวณระยะทางสั้นที่สุดระหว่างคำในโครงสร้าง

5.1.1.2 การสืบค้นเชิงความหมาย ประกอบด้วยกระบวนการ 3 ส่วน คือ 1) การกำหนดลำดับความสำคัญของคำค้นมี 2 กระบวนการ คือวิเคราะห์จากลำดับคำค้นจากผู้ใช้และตำแหน่งของคำค้นที่ปรากฏในเอกสาร2) การให้น้ำหนักคำสำคัญ และ 3) การวัดความคล้ายคลึงเชิงความหมาย

5.1.2 การประเมินประสิทธิภาพ

5.1.2.1 การประเมินประสิทธิภาพการให้น้ำหนักคำสำคัญ ด้วยการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ (Corr)มีค่าความแม่นยำเฉลี่ย (MRR) เท่ากับ 0.906 และมีค่าความผิดพลาด (SSE) เท่ากับ 0.097 ซึ่งให้ผลดีกว่าทั้งในด้านของความแม่นยำเฉลี่ย และมีค่าความผิดพลาดที่น้อยกว่าวิธีการให้น้ำหนักด้วย IDF

5.1.2.2 การประเมินประสิทธิภาพการวัดความคล้ายคลึงเชิงความหมาย

1) ผลการเปรียบเทียบประสิทธิภาพการวัดความคล้ายคลึงจากการสืบค้นโดยใช้ค่าความถ่วงดุล (F-measure)ที่สะท้อนจากค่าความแม่นยำ (Precision) และ ค่าความระลึก (Recall) ของรูปแบบการทดลอง 8 ลักษณะ คือ A1, A2, A3, A4, B1, B2, B3, B4 มีผลต่อประสิทธิภาพการสืบค้นที่แตกต่างกัน พบว่ารูปแบบการทดลอง A4 ให้ผลการสืบค้นมีค่าความถ่วงดุลสูงกว่าการทดลองทุกลักษณะ มีค่าความถ่วงดุลเท่ากับ 0.8596และข้อสังเกตุคือการทดลองลักษณะ A1, A2, A3, A4ที่ให้น้ำหนักคำด้วยการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ (Corr) มีค่าความถ่วงดุลสูงกว่าการทดลอง A1, A2, A3, A4 ทุกกรณี

2) ผลการเปรียบเทียบประสิทธิภาพการวัดความคล้ายคลึงจากการสืบค้นโดยใช้ค่าความแม่นยำเฉลี่ย (MRR)ของรูปแบบการทดลอง 8 ลักษณะ คือ A1, A2, A3, A4, B1, B2, B3, B4 มีผลต่อประสิทธิภาพการสืบค้นที่แตกต่างกัน พบว่ารูปแบบการทดลอง B3 ให้ผลการสืบค้นมีค่าความแม่นยำเฉลี่ย สูงกว่าการทดลองทุกลักษณะ มีค่าความแม่นยำเฉลี่ย เท่ากับ 0.8661 และข้อสังเกตุคือการทดลองลักษณะ B1, B2, B3, B4 ที่ให้น้ำหนักคำด้วยการวิเคราะห์ IDF มีค่าความแม่นยำเฉลี่ยสูงกว่าการทดลอง A1, A2, A3, A4 ทุกกรณี

5.2 อภิปรายผลการวิจัย

การออกแบบและพัฒนาแบบจำลองการสืบค้นเชิงความหมายสำหรับความรู้ด้านพืชไร่ ประกอบด้วย 2ส่วน คือการจัดความรู้ให้เป็นระบบ และการสืบค้นเชิงความหมาย โดยจัดลำดับผลลัพธ์ตามผลการเปรียบเทียบความคล้ายคลึงระหว่างเอกสารความรู้ด้านพืชไร่ ซึ่งผลที่ได้จากการวิจัยในครั้งนี้มีข้ออภิปรายดังนี้

5.2.1 ผลที่ได้จากการวิจัย เป็นไปตามวัตถุประสงค์ คือ

1. การออกแบบจำลองการสืบค้นเชิงความหมายสำหรับความรู้ด้านพืชไร่ (Semantic Search Model for Agronomy Knowledge: SeeAgrow) งานวิจัยนี้ได้วิเคราะห์กระบวนการทำงาน 2 ส่วน คือการจัดความรู้ให้เป็นระบบ และการสืบค้นเชิงความหมายซึ่งการออกแบบการจัดความรู้ให้เป็นระบบประกอบด้วย 3 ขั้นตอนหลัก คือ 1)การออกแบบโครงสร้างคลาสจัดเก็บความรู้ด้านพืชไร่ 2)การกำหนดความสัมพันธ์ของคำสำคัญ และสร้างคลังคำ และ3) การประมวลความรู้ด้านพืชไร่ ได้แก่ 3.1) การคำนวณความถี่ของคำที่ปรากฏในความรู้พืชไร่ 3.2) การคำนวณน้ำหนักคำ และ 3.3) การคำนวณระยะทางสั้นที่สุดระหว่างคำในโครงสร้าง ซึ่งเก็บไว้ในรูปแบบของไฟล์ข้อความ ที่เป็นเช่นนี้มีสาเหตุมาจากผู้วิจัยได้ทำการศึกษาปัญหาจากแนวคิดการสืบค้นความรู้ เอกสารและงานวิจัยที่เกี่ยวข้องแล้วพบปัญหาใน 2 ประเด็นหลัก ได้แก่ การจัดความรู้ให้เป็นระบบ และการสืบค้นความรู้ ซึ่งการจัดความรู้ให้เป็นระบบผู้วิจัยได้ประยุกต์ใช้ออนโทโลยีไม้ผลเศรษฐกิจ (ThaiFO Ontology) และการสร้างคลังคำที่ใช้ตัวแบบการแทนความรู้เอสเคโอเพราะมีโครงสร้างใกล้เคียงกับการจัดระบบความรู้ด้านพืชไร่ โดยแนวคิดการจัดความรู้ให้เป็นระบบดังกล่าวผู้วิจัยได้ศึกษาแนวคิดจากนักวิจัยหลาย ๆ ท่าน เช่น นอย และแม็คกินนิสส์ (Noy and McGuinness, 2001), แกสสิวิค, ดูจ์ริค และเดเวดซิค (Gasevic, Djuric and Devedzic, 2009) และจุฑามาศ เทียนสอาด และอรวรรณ อิ่มสมบัติ(2552) ซึ่งแนวคิดการคำนวณน้ำหนักคำด้วยการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ ได้ประยุกต์จากดีคอฟฟ์ (Diekhoff, 1992) และการคำนวณระยะทางสั้นที่สุดระหว่างคำในโครงสร้างใช้แนวคิดของ ไดก์สตรา (Dijkstra, E., 1959) เพราะขั้นตอนวิธีดังกล่าวมีการจัดลำดับการเข้าและออกข้อมูลอย่างเป็นลำดับโดยใช้เส้นทางรวมจากจุดเริ่มต้นมายังโหนดนั้น ๆ นอกจากนี้พิลาวัณย์ พลับรู้การ และ กฤษณะ ไวยมัย (2544) และศุภกฤษฎิ์ นิวัฒนากูล (2556) ได้สนับสนุนแนวคิดการคำนวณหาเส้นทางสั้นที่สุดตามโครงสร้างดังกล่าว เพื่อใช้ในการคำนวณหาค่าความคล้ายคลึงเชิงความหมายของเอกสาร อันส่งผลต่อลำดับความสำคัญของเอกสารที่ตรงตามความต้องการของผู้ใช้ นอกจากนี้การจัดความรู้ให้เป็นระบบโดยใช้ออนโทโลยีสามารถจัดเก็บคำค้นที่มีความหมายเดียวกันไว้เป็นกลุ่ม เพื่อใช้ขยายคำค้นจากผู้ใช้ส่งผลให้การออกแบบดังกล่าวสามารถแก้ปัญหาความกำกวมของคำสำคัญ สอดคล้องกับการวิจัยของอาเมียร์และโมแรด (Amir and Mourad, 2013), ศุภกฤษฎิ์ นิวัฒนากูล (2556), ลีย์, มิน และโฮ (Lee, Min and Oh, 2014) ที่กล่าวว่าออนโทโลยีมีความจำเป็นต่อการประมวลผลเชิงความหมายโดยเฉพาะการสืบค้นความรู้

2. การสืบค้นเชิงความหมาย ผู้วิจัยได้ออกแบบใน 3 ขั้นตอน ได้แก่การกำหนดลำดับความสำคัญคำค้น การให้น้ำหนักคำสำคัญ และการวัดความคล้ายคลึง เหตุผลที่เป็นเช่นนี้มีสาเหตุมาจาก

2.1 การกำหนดลำดับความสำคัญคำค้นเป็นกระบวนการแรกในส่วนนำเข้าข้อมูล คือการป้อนข้อสอบถามจากผู้ใช้ ซึ่งเป็นคำสำคัญ หรือภาษาธรรมชาติ ซึ่งข้อสอบถามดังกล่าวควรมีการวิเคราะห์ถึงลำดับคำสำคัญ และตำแหน่งคำสำคัญที่ปรากฏในเอกสาร สอดคล้องกับแนวคิดการสืบค้นสารสนเทศของอุไร ทองหัวไผ่ (2551) และเบย์ซา เยตส์ และริบริโร เนโท (Baeza-Yates and Ribeiro-Neto, 1999)ได้อธิบายไว้ว่าเมื่อผู้ใช้งานระบุคำสำคัญเข้ามา คำสำคัญเหล่านั้นจะผ่านกระบวนการวิเคราะห์คำสำคัญ เช่น การให้ความสำคัญกับลำดับของคำการจัดกลุ่มคำ และการตรวจหลักไวยกรณ์ของคำให้ถูกต้อง เป็นต้นนอกจากนี้ผู้วิจัยได้กำหนดตัวแทนเอกสารจัดเก็บในรูปแบบออนโทโลยีเพื่อช่วยให้การสืบค้นมีความรวดเร็วเนื่องจากลดการสืบค้นจากคำสำคัญจากข้อมูลต้นฉบับโดยตรง สอดคล้องกับเฟรค(Frake, 1992)ที่เสนอให้การสร้างตัวแทนเอกสาร หรือดัชนี เป็นกระบวนการหนึ่งที่สำคัญในการสืบค้นสารสนเทศ

2.2 การให้น้ำหนักคำสำคัญได้ทำการวิเคราะห์หาความสัมพันธ์ระหว่างคำสองคำว่ามีความสัมพันธ์กันมากน้อยเพียงใด โดยใช้การวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ที่สามารถสื่อถึงความหมายของคำอย่างแท้จริง อันจะเพิ่มประสิทธิภาพการสืบค้นให้ตรงกับความต้องการของผู้ใช้มากยิ่งขึ้น สอดคล้องกับแนวคิดของดีคอฟฟ์ (Diekhoff, 1992) ได้อธิบายว่าการศึกษาความสัมพันธ์ระหว่างตัวแปรว่ามีความสัมพันธ์กันหรือไม่ และสัมพันธ์กันอย่างไร ความสัมพันธ์อาจเป็นไปในทิศทางเดียวกัน หรือทิศทางตรงข้ามกันจะมีการพิจารณาความสัมพันธ์ระหว่างตัวแปร 2 ตัวขึ้นไป และจะใช้ค่าสัมประสิทธ์สหสัมพันธ์เป็นค่าที่วัดความสัมพันธ์

2.3 การวัดความคล้ายคลึงได้ทำการวิเคราะห์หาค่าความเกี่ยวข้องระหว่างคำสองคำจากระยะห่างน้อยที่สุดของเส้นเชื่อม (Edge) ระหว่างโหนด (Node) ที่มีผลรวมระยะทางน้อยที่สุดระหว่างคำในโครงสร้างที่จัดเก็บในคลังคำ ด้วยขั้นตอนวิธีไดก์สตรา ซึ่งเป็นแนวคิดที่สอดคล้องกับทฤษฎีกราฟมีองค์ประกอบสองส่วน คือ จุดยอด และ เส้นเชื่อมสอดคล้องกับคริสเทนเซน และ อัลเบิทร์ (Christensen and Albert, 2007) พิลาวัณย์ พลับรู้การ และ กฤษณะ ไวยมัย (2544) อาเมียร์ และโมแรด (Amir and Mourad, 2013)และศุภกฤษฎิ์ นิวัฒนากูล (2556)ได้นำเสนอการคำนวณความคล้ายคลึงจากความสัมพันธ์ของคำในโครงสร้าง

5.2.2 ผลการประเมินประสิทธิภาพการสืบค้น 2 กระบวนการ พบว่า

1. การให้น้ำหนักคำด้วยการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ (Corr)มีค่าความแม่นยำเฉลี่ย (MRR) เท่ากับ 0.906 และมีค่าความผิดพลาด (SSE) เท่ากับ 0.097 ซึ่งให้ผลดีกว่าทั้งในด้านของความแม่นยำเฉลี่ย และมีค่าความผิดพลาดที่น้อยกว่าวิธีการให้น้ำหนักด้วย IDF เป็นไปตามสมมติฐานที่ตั้งไว้ ที่เป็นเช่นนี้มีสาเหตุมาจากกระบวนการแตกต่างกันของการให้น้ำหนักคำสำคัญซึ่งการให้น้ำหนักคำด้วยการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ (Corr) ได้คำนึงถึง

ความหมาย และความเกี่ยวข้องของคำสำคัญลักษณะคำกว้างกว่า คำแคบกว่า คำที่เกี่ยวข้อง และคำเหมือน ที่มีความสัมพันธ์กันในโครงสร้างคลังคำ และงานวิจัยนี้ยังได้นำความถี่ของคำที่ปรากฏในเอกสารมาพิจารณาร่วมกับการให้น้ำหนักคำสำคัญตามแนวคิดของ ซอลตั่น และบัคลี่ย์(Salton, G. and C. Buckley, 1988) อธิบายว่าวิธีการให้น้ำหนักสำหรับคำที่มีความสำคัญหรือใช้เป็นตัวแทนของเอกสารที่ควรจะปรากฏอยู่ และการให้น้ำหนักคำเป็นขั้นตอนสำคัญที่จะส่งผลต่อการค้นคืนเอกสารที่มีประสิทธิภาพสอดคล้องกับการอธิบายของ เรสนิค (Resnik, 1999) ที่ว่าการให้ค่าน้ำหนักของคำแต่ละคำ (ทั้งคำที่มีในข้อสอบถามและคำในความรู้หรือเอกสาร) จะคำนวณโดยเปรียบเทียบกับคำทั้งหมดในเอกสารค่าน้ำหนักคำในเอกสารจะได้รับการดำเนินการในขั้นตอนการเตรียมฐานความรู้ส่วนค่าน้ำหนักคำในข้อสอบถามจะดำเนินการในแต่ละครั้งที่มีการสืบค้นเมื่อคำนวณค่าน้ำหนักคำของข้อสอบถามได้เรียบร้อยแล้วค่าน้ำหนักเหล่านั้นจะถูกนำไปแทนในสมการเพื่อคำนวณระดับความคล้ายคลึงระหว่างข้อสอบถามและเอกสาร

2. ประสิทธิภาพการวัดความคล้ายคลึงเชิงความหมาย

2.1 ผลการเปรียบเทียบประสิทธิภาพการวัดความคล้ายคลึงจากการสืบค้นโดยใช้ค่าความถ่วงดุล (F-measure)ของรูปแบบการทดลอง 8 ลักษณะ คือ A1, A2, A3, A4, B1, B2, B3, B4 มีผลต่อประสิทธิภาพการสืบค้นที่แตกต่างกัน พบว่ารูปแบบการทดลอง A4 ให้ผลการสืบค้นมีค่าความถ่วงดุลสูงกว่าการทดลองทุกลักษณะ มีค่าความถ่วงดุลเท่ากับ 0.8596ที่เป็นเช่นนี้เพราะการทดลองลักษณะของ A4 ได้นำค่าการคำนวณระยะทางระหว่างโหนดของคำในโครงสร้างลักษณะคำกว้างกว่า คำแคบกว่า คำที่เกี่ยวข้อง และคำเหมือน ที่คำนึงถึงความหมายของคำมาคำนวณร่วมกับการให้ความสำคัญกับผู้ใช้จากลำดับคำค้น และความสำคัญของคำศัพท์ที่ปรากฏในตำแหน่งต่าง ๆ ของความรู้อาจเป็นสาเหตุที่สำคัญที่ทำให้ความครบถ้วนของเนื้อหาความรู้ที่เกี่ยวข้องกับคำค้น และผลลัพธ์ที่เกี่ยวข้องค่อนข้างสูง

2.2 ผลการเปรียบเทียบประสิทธิภาพการวัดความคล้ายคลึงจากการสืบค้นโดยใช้ค่าความแม่นยำเฉลี่ย (MRR)ของรูปแบบการทดลอง 8 ลักษณะ คือ A1, A2, A3, A4, B1, B2, B3, B4 มีผลต่อประสิทธิภาพการสืบค้นที่แตกต่างกัน พบว่ารูปแบบการทดลอง B3 ให้ผลการสืบค้นมีค่าความแม่นยำเฉลี่ย สูงกว่าการทดลองทุกลักษณะ มีค่าความแม่นยำเฉลี่ย เท่ากับ 0.8661 ที่เป็นเช่นนี้เพราะเพราะรูปแบบการทดลอง B1, B2, B3 และ B4 มีการคำนวณน้ำหนักคำสำคัญมีการใช้ความถี่ของคำที่ปรากฏในความรู้หรือเอกสารเพียงอย่างเดียวไม่มีการคำนึงถึงความหมายของคำ โดยถือว่าแต่ละคำเป็นอิสระต่อกัน ซึ่งความเป็นจริงแล้วคำแต่ละคำอาจมีความสัมพันธ์กันเชิงความหมาย หรือใกล้เคียงกัน จึงเป็นสาเหตุสำคัญไม่สามารถเชื่อมโยงไปยังความรู้หรือเอกสารที่เกี่ยวได้ครบถ้วน จึงทำให้ผลลัพธ์ที่ได้มีปริมาณน้อยส่งผลต่อค่าความแม่นยำเฉลี่ยที่สูง

จากการสังเกตของผู้วิจัย พบว่าการทดลองลักษณะ A1, A2, A3, A4ที่ให้น้ำหนักคำด้วยการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ (Corr) มีค่าความถ่วงดุลสูงกว่าการทดลอง A1, A2, A3, A4 ทุกกรณี เนื่องจากการทดลองลักษณะดังกล่าวทำให้ผลลัพธ์ที่พบมีปริมาณมากจากการกำหนดเกณฑ์เลือกความรู้หรือเอกสารแบบไดนามิกเกณฑ์ที่เลือกจากค่าต่ำสุดของความรู้หรือเอกสารที่ค้นพบทำให้ได้ผลลัพธ์ที่ได้มีปริมาณมาก มีความครบถ้วนของความรู้เมื่อเทียบกับความรู้ที่ควรได้ทั้งหมด จากอัตราส่วนระหว่างจำนวนความรู้ที่ค้นคืนถูกต้องกับจำนวนความรู้ที่ถูกต้องทั้งหมดของระบบจึงส่งผลให้ค่าความระลึก (Recall) มีค่าเป็น 1 เสมอ แต่ในทางกลับกันการทดลองลักษณะ B1, B2, B3, B4 ที่ให้น้ำหนักคำด้วยการวิเคราะห์ IDF มีค่าความแม่นยำเฉลี่ยสูงกว่าการทดลอง A1, A2, A3, A4 ทุกกรณี ซึ่งรูปแบบการทดลอง B1, B2, B3 และ B4 ไม่ได้ให้ความสำคัญกับการค้นหาคำ กลุ่มคำ โดยไม่มีการคำนึงถึงความหมายของคำและถือว่าคำแต่ละคำเป็นอิสระต่อกันผลลัพธ์ที่ได้จากการสืบค้นจึงน้อยส่งผลต่อความแม่นยำเฉลี่ยที่สูงกว่ารูปแบบการทดลอง A1, A2, A3, A4 ที่คำนึงถึงความสัมพันธ์ของคำสำคัญในขั้นตอนการให้น้ำหนักคำ และการคำนวณระยะทางสั้นที่สุดระหว่างโหนดของคำในโครงสร้างโดยการหาระยะทางจากเส้นเชื่อมระหว่างโหนดความสัมพันธ์ของคำในโครงสร้างในขั้นตอนการวัดความคล้ายคลึงเชิงความหมาย และพบว่า คำใด ๆ ก็แล้วแต่ที่มีค่าระยะทางที่น้อยจะมีความเกี่ยวพันหรือสัมพันธ์กันทางความหมายมาก แต่ถ้าหากคำสองคำมีค่าระยะทางห่างกันมากจะถือว่าคำทั้งสองมีความเกี่ยวพันกันทางความหมายน้อยวิธีการนี้จะแก้ปัญหาความสัมพันธ์ระหว่างคำเป็นคำหนึ่งคำมีหลายความหมาย คำหลาย ๆ คำมีความหมายเดียวกัน และหนึ่งคำหรือกลุ่มคำเรียกได้หลายชื่อ สอดคล้องกับงานวิจัยของ พิลาวัณย์ พลับรู้การ และกฤษณะ ไวยมัย (2544) ได้เสนอเทคนิคการวัดความคล้ายคลึงระหว่างเอกสาร ด้วยการหาค่าความเกี่ยวข้องระหว่างคำสองคำจากระยะทางน้อยที่สุดของคำ รอย และ คริสทินา (Roi and Christina, 2011)ที่ได้นำเสนอเทคนิคการหาความสัมพันธ์ระหว่างคำโดยใช้แนวคิดทฤษฎีกราฟและสอดคล้องกับงานวิจัยของ ศุภกฤษฎิ์ นิวัฒนากูล (2556) ที่ได้นำเสนอเทคนิควิธีการวัดความคล้ายคลึงเชิงความหมายระหว่างเอกสารกับคำค้น การกำหนดค่าความสำคัญของคำค้นคำนวณหาค่าน้ำหนักของคำค้น การหาเส้นทางสั้นที่สุดตามโครงสร้าง และคำนวณหาค่าความคล้ายคลึงเชิงความหมายของเอกสาร เพื่อให้ความรู้ที่ได้ตรงตามความต้องการของผู้ใช้ แต่อย่างไรก็ตามแบบจำลองการสืบค้นเชิงความหมายที่ผู้วิจัยนำเสนอ สามารถนำไปประยุกต์ใช้สำหรับการค้นคืนเชิงความหมายในโดเมนอื่นๆ ได้ แต่ควรจัดความรู้ให้เป็นระบบในส่วนของโครงสร้างของความรู้ และการเพิ่มคำสำคัญในคลังคำ

ดังนั้นงานวิจัยนี้ได้ข้อสรุปดังนี้

1. การนำฐานความรู้ด้านพืชไร่ที่ประยุกต์จากออนโทโลยีไม้ผลเศรษฐกิจในงานวิจัยนี้ไปใช้ควรพิจารณาถึงโครงการการออกแบบจัดเก็บความรู้ในโดเมนที่ใกล้เคียงกัน

2. วิธีการวิเคราะห์สัมประสิทธิ์สัมพันธ์ (Corr) จากแนวคิดของดีคอฟฟ์ (Diekhoff, 1992) สามารถประยุกต์ใช้กับการให้น้ำหนักคำสำคัญสำหรับการสืบค้นความรู้เป็นที่ยอมรับได้

3. การวัดความคล้ายคลึงจากรูปแบบการทดลอ A4 ที่ได้คำนวณค่าระยะทางสั้นที่สุดระหว่างคำค้นกับคำในความรู้หรือเอกสารมาคำนวณกับค่าน้ำหนักคำสำคัญด้วยวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ และค่าลำดับความสำคัญคำค้นจากผู้ใช้ และค่าลำดับความสำคัญคำค้นจากตำแหน่งของคำค้นที่ปรากฏในความรู้เป็นที่ยอมรับได้

5.3 ข้อจำกัดของการวิจัย

5.3.1 จำนวนความรู้ที่ใช้ทดสอบมีจำนวน 240 รายการ ซึ่งควรเพิ่มความรู้ เพื่อยืนยันความถูกต้องแม่นยำในการสืบค้น

5.3.2 การสร้างคลังคำจะมีการนำคำศัพท์มาจากอรรถาภิธานศัพท์เกษตรไทย ซึ่งเป็นอรรถาภิธาน (Thesaurus) ด้านการเกษตร โดยนำคำศัพท์มาจัดเก็บโดยใช้ตัวแบบการแทนความรู้หรือเรียกว่า เอสเคโอเอส (Simple Knowledge Organization System: SKOS)ซึ่งจะมีความซับซ้อนยุ่งยากในการนำคำศัพท์เข้ามาจัดเก็บ และคลังคำยังมีคำศัพท์น้อย อาจจะส่งผลต่อการสืบค้นความรู้โดยไม่ครอบคลุมเนื้อหาในความรู้

5.4 การประยุกต์ผลการวิจัย

แบบจำลองการสืบค้นเชิงความหมายสำหรับความรู้ด้านพืชไร่สามารถนำแนวคิดทฤษฎีที่ผู้วิจัยนำเสนอไปประยุกต์ใช้ในการสืบค้นเชิงความหมายให้มีประสิทธิภาพ ได้ข้อมูลถูกต้องตรงกับความต้องการของผู้ใช้ และสามารถปรับใช้กับความรู้ในโดเมนอื่นๆ หรือโดเมนที่สนใจได้ แต่ต้องมีการประยุกต์หรือปรับปรุงออนโทโลยี ดังนี้

5.4.1 ควรจะมีการเพิ่มคลาสที่เกี่ยวข้องในโดเมนที่สนใจ และเพิ่มออฟเจคพร็อพเพอร์ตี้ที่สามารถเชื่อมโยงระหว่างคลาสในโดเมนที่สนใจกับคลาสความรู้

5.4.2 การใช้เมทาดาทาที่เป็นดับลินคอร์สำหรับอธิบายให้ทราบรายละเอียดของข้อมูลอาจจะใช้ชุดหน่วยข้อมูลย่อยใน 15 อิลีเมนต์หรือเลือกใช้ตามความเหมาะสมกับความรู้ในโดเมนที่สนใจ และรูปแบบของเอกสารควรเป็นข้อมูลที่มีโครงสร้าง

5.4.3 เพิ่มคำศัพท์ที่เกี่ยวข้องในโดเมนที่สนใจหรือเกี่ยวข้องสำหรับการจัดเก็บให้ครอบคลุมกับเนื้อหา ซึ่งควรพิจารณาจากรายการของคำหรือข้อความพร้อมกับความหมายของคำจากอรรถาภิธานที่เป็นการกำหนดความสัมพันธ์ระหว่างคำต่าง ๆ ที่ได้รวบรวมไว้แล้วอย่างเป็นมาตรฐานในโดเมนนั้นๆ

5.5 ข้อเสนอแนะในการวิจัยครั้งต่อไป

งานวิจัยในอนาคตควรจะมีการออกแบบและพัฒนาเทคนิควิธีการปรับเปลี่ยนการให้น้ำหนักคำที่ให้ความสำคัญกับความหมายของคำ และการวัดความคล้ายคลึงโดยการหา และปรับเปลี่ยนค่าความสัมพันธ์ระหว่างคำในคลังคำรวมไปถึงการวัดความคล้ายคลึงระหว่างคำระหว่างเอกสารกับคำค้น ซึ่งอาจจะส่งผลให้การสืบค้นความรู้มีความถูกต้อง แม่นยำตรงกับความต้องการของผู้ใช้มากยิ่งขึ้น และที่สำคัญควรเพิ่มคำศัพท์ในคลังคำให้ครอบคลุมเนื้อหา มีการเพิ่มจำนวนความรู้สำหรับใช้ทดสอบกับการสืบค้นและมีการเปรียบเทียบวิธีการให้น้ำหนักคำค้น และการวัดความคล้ายคลึงกับวิธีการอื่น ๆ ที่เหมาะสมนอกเหนือจากวิธีการที่นำเสนอในงานวิจัยนี้

มหาวิทยาลัยเทคโนโลยีสุรนารี

รายการอ้างอิง

กระทรวงเกษตรและสหกรณ์สำนักงานเศรษฐกิจเกษตร. (2557).**ภาวะและแนวโน้มเศรษฐกิจการเกษตร.**[ออนไลน์]. ได้จาก: http://www.oae.go.th/download/bapp/OutlookQ3OutlookQ3_2556.pdf

โกศล ดีศีลธรรม. (2546). **การจัดการความรู้แห่งโลกธุรกิจยุคใหม่.**กรุงเทพฯ: ซีเอ็ดยูเคชั่น.

ไกรศักดิ์ เกสร. (2554). การค้นหาข้อมูลเชิงความหมาย: แนวคิดใหม่ของโปรแกรมการค้นหา (Search Engine) และแนวทางการพัฒนาในอนาคต. **วารสารวไลยอลงกรณ์ปริทัศน์.** 1(2): 2-6.

คธาทิพย์ บำรุงกิจ. (2549). **การสืบค้นเชิงความหมายบนฐานของ OWL**. การศึกษาอิสระ. วิทยาศาสตรมหาบัณฑิต. ขอนแก่น: มหาวิทยาลัยขอนแก่น.

คมคิด ชัชราภรณ์, ธรา อั่งสกุล และจิติมนต์ อั่งสกุล. (2012). แบบจำลองการจัดหมวดหมู่สถานที่ท่องเที่ยวโดยใช้เทคนิคการเรียนรู้ของเครื่อง. **Suranaree Journal of Science and Technology**. 12(6).

ฆนาศัย กรึงไกรและชุลีรัตน์ จรัสกุลชัย. (2550).**การจัดกลุ่มเอกสารข้อความภาษาไทยด้วยขั้นตอนวิธี Spherical K-means แบบขนานบนพิรุณลีนุกซ์คลัสเตอร์.** วิทยานิพนธ์. วิทยาศาสตรมหาบัณฑิต. กรุงเทพฯ: มหาวิทยาลัยเกษตรศาสตร์.

จิรัฏฐา ภูบุญอบ และ วรวิทย์ สังฆทิพย์. (2555). การออกแบบสถาปัตยกรรมออนโทโลยีเพื่อบูรณาการข้อมูลสำหรับการบริหารงานกิจการนิสิต มหาวิทยาลัยมหาสารคาม. **วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยมหาสารคาม ในการประชุมวิชาการมหาวิทยาลัยมหาสารคาม ครั้งที่ 9.**มหาสารคาม: มหาวิทยาลัยมหาสารคาม.(หน้า 271-280).

จิรัสย์มาลาวงษ์และอานนท์รุ่งสว่าง. (2552).**การเพิ่มประสิทธิภาพระบบสืบค้นเนมเพจด้วยการวิเคราะห์ชื่อที่ปรากฏเด่นชัด**. วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์มหาวิทยาลัยเกษตรศาสตร์.

จุฑามาศ เทียนสอาด และอรวรรณ อิ่มสมบัติ. (2552). ระบบผู้เชี่ยวชาญเพื่อวินิจฉัยและให้คำแนะนำผู้ป่วยไตวายเรื้อรังโดยใช้ฐานความรู้ออนโทโลยี.**Journal of InformationScience and Techology.** 3(2): 22-30.

จุฑาวรรณ สิทธิโชคสถาพร. (2555). **ต้นแบบออนโทโลยีเพื่อการค้นคืนสารสนเทศเชิงความหมายสำหรับงานสารบรรณอิเล็กทรอนิกส์ กรณีศึกษา งานบริหารและธุรการคณะแพทย์ศาสตร์ มหาวิทยาลัยสงขลานครินทร์**. วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต. สงขลา: มหาวิทยาลัยสงขลานครินทร์.

ชารินี พรหมภักดี, มารุต บูรณรัช และ จรัสศรี รุ่งรัตนาอุบล. (2558). ระบบสืบค้นข้อมูลการรักษาด้านการแพทย์แผนไทยด้วยฐานความรู้ออนโทโลยี. **วารสารสังคมศาสตร์**.4(2): 61-70.

ชุลีรัตน์ จรัสกุลชัย, เจษฎา กันทะเสนา และสถาพร คิ้วสุวรรณสุข. (2544). **รายงานการวิจัยเรื่องการจัดกลุ่มข้อความสำหรับเอกสารภาษาไทย.**รายงานการวิจัยสารสนเทศอัจฉริยะและฐานข้อมูล คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์.

ชัชวาล วงศ์ประเสริฐ. (2548). **การจัดการความรู้ในองค์กรธุรกิจ.** กรุงเทพฯ: เอ็กซเปอร์เน็ท.

โชติสิทธิบุศย์. (2526). **การบำรุงรักษาดินและการใช้ปุ๋ย**. เอกสารเลขที่ 7 งานทะเบียนและประมวลสถิติกองแผนงานและวิชาการกรมวิชาการเกษตรกระทรวงเกษตรและสหกรณ์.

ณัฐพร หอมเมือง. (2550). **การพัฒนาระบบถามตอบภายใต้แบบจำลองเวคเตอร์สเปซแบบผสมผสาน.** สารนิพนธ์วิทยาศาสตรมหาบัณฑิต ภาควิชาวิทยาการคอมพิวเตอร์และสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

ดนัยศุภาพร. (2537). **พฤกษศาสตร์และพันธุ์ศาสตร์ของมันสำปะหลัง**. เอกสารวิชาการ ศูนย์วิจัยพืชไร่ระยองสถาบันวิจัยพืชไร่กรมวิชาการเกษตรกระทรวงเกษตรและสหกรณ์.

ธนิตา วงศ์กาฬสินธุ์ และ งามนิจ อาจอินทร์. (2009). การเชื่อมโยงออนโทโลยีบนโดเมน E-learning โดยใช้ WordNetบนพื้นฐานของการวัดค่าความคล้ายคลึงเชิงความหมาย. **การประชุมวิชาการระดับชาติ NCCIT 2009ครั้งที่ 5**. (หน้า 218-224).กรุงเทพฯ: มหาวิทยาลัยพระจอมเกล้าพระนครเหนือ.

ธนุสศักดิ์ ธัญญสิริ. (2543). **การออกแบบ และพัฒนาระบบต้นแบบสำหรับการสืบค้นข้อมูลสารสนเทศภาษาไทยด้วยดัชนีหลายระดับ.**วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต. กรุงเทพฯ: มหาวิทยาลัยเกษตรศาสตร์.

นคร โคตรโสภา.(2549). **วิธีการสืบค้นเอกสารเชิงความหมายบนข้อมูลXML**. การศึกษาอิสระวิทยาศาสตรมหาบัณฑิต. ขอนแก่น: มหาวิทยาลัยขอนแก่น.

บังอร กลับบ้านเกาะ. (2543).การค้นคืนสารสนเทศออนไลน์โดยใช้จีเนติกอัลกอริทึม.**NECTEC Technical Journal.** 2(7): 2-6.

บารมี โอสธีรกุล และ ธวัชชัย งามสันติวงศ์. (2555). ระบบเทียบโอนรายวิชาโดยใช้แบบจำลองเวกเตอร์สเปซ.**การประชุมเสนอผลงานวิจัยระดับบัณฑิตศึกษามหาวิทยาลัยสุโขทัยธรรมาธิราช ครั้งที่ 2.**(หน้า 1-8).นนทบุรี: มหาวิทยาลัยสุโขทัยธรรมาธิราช.

บุญชม ศรีสะอาด. 2545). **การวิจัยเบื้องต้น.** พิมพ์ครั้งที่ 7. กรุงเทพฯ: สุวีริยาสาส์น.

บุญเลิศอรุณพิบูลย์, สุภาพรชัยธัมมะปกรณ์และ ฐิติมาธรรมบำรุง. (2550).**เริ่มต้นกับเมทาดาทา.** ศูนย์บริการความรู้วิทยาศาสตร์และเทคโนโลยี สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ.

ปฏิคม ทองจริง.(2552).**ออนโทโลยีสำหรับการรวมข้อมูลเชิงความหมายของความรู้ด้านสมุนไพรไทย.**วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยธรรมศาสตร์.

ปรีชาพราหมณีย์. (2544). **การใช้ปุ๋ยและการจัดการดินในไร่เอกสารวิชาการการปลูกดูแลรักษาพันธุ์อ้อย.** สถาบันวิจัยพืชไร่กรมวิชาการเกษตรกระทรวงเกษตรและสหกรณ์.

พยุง มีสัจ วาทินี นุ้ยเพรียร และผุสดี บุญรอด. (2553). ระบบการค้นคืนเชิงความหมายจากข้อมูลบรรณานุกรมโดเมน.**Information Systems JOURNAL OFINFORMATION SCIENCE AND TECHNOLOGY.** 4(1): 11-20.

พรธิดา วิเชียรปัญญา. (2552).**การจัดการความรู้: พื้นฐานและการประยุกต์ใช้.** พิมพ์ครั้งที่ 2. กรุงเทพฯ: ธรรมกมลการพิมพ์.

พันธุ์ปิติ เปี่ยมสง่า, จิติมนต์ เขียนดวงจันทร์ และคณะ. (2545). **การศึกษาและพัฒนาระบบข้อมูลทางการเกษตรโลกWAICENT.** ภาควิชาวิศวกรรมคอมพิวเตอร์มหาวิทยาลัยเกษตรศาสตร์.

พิลาวัณย์ พลับรู้การ และกฤษณะ ไวยมัย. (2544). **รายงานการวิจัยเรื่องการวัดความคล้ายคลึงระหว่างเอกสารโดยใช้แนวทางด้านความหมาย.** รายงานวิจัยระบบคอมพิวเตอร์และเครื่อข่ายสื่อสาร คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์.

เพ็ญพรรณ อัศวนพเกียรติ, อรินทิพย์ ธรรมชัยพิเนต และกฤษณะ ไวยมัย. (2546). “ออนโทโลยีชีวภาพ: ระบบสำหรับสืบค้นและวิเคราะห์ข้อมูลทางด้านชีววิทยา”**Proceedings of 41thKasetsart University Annual Conference: Engineering and Architecture**, 277-285.

มนัสนันท์ ปัญญามี.(2552). **การจัดกลุ่มผลการสืบค้นเอกสารบนเว็บเพื่อการสืบค้นเชิงความหมาย.** วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น.

มานพ จงเจริญใจ. (2541). **การพัฒนาระบบค้นคืนสารสนเทศด้วยวิธีค้นคืนแบบจัดลำดับแลแบบค้นคืนย้อนกลับโดยใช้แถวลำดับแพ็ต.** วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

ราชบัญฑิตยสถาน. (2542). **พจนานุกรม ฉบับราชบัณฑิตยสถาน 2542.** กรุงเทพฯ: นานมีบุ๊คส์.

โรสริน อัคนิจ, ธนิต พุทธพงศ์ศิริพร, น้ำฝน ลำดับวงศ์และอารีย์ ธัญกิจจานุกิจ. (2554). การวิจัยการพัฒนาออนโทโลยีเพื่อการจัดการความรู้ด้านการแปรรูปข้าว. **วารสารเกษตรมหาวิทยาลัยเกษตรศาสตร์**. 27(3): 267-274.

วรวีร์เกสร. (2553). **การวัดความคล้ายคลึงของเอกสารภาษาไทยโดยใช้การประมวลผลภาษาธรรมชาติ.** การค้นคว้าอิสระวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยเชียงใหม่.

วิมลพร ใบสนธิ์ (บรรณาธิการ). (2550).**สังคมสุขสันต์มหัศจรรย์ KM.** กรุงเทพฯ: สถาบันส่งเสริมการจัดการความรู้เพื่อสังคม (สคส.).

วิจารณ์ พานิช. (2547). **การจัดการความรู้.** กรุงเทพฯ: สถาบันการจัดการความรู้เพื่อสังคม.

วิจารณ์ พานิช. (2548). **การจัดการความรู้ ฉบับนักปฏิบัติ.** กรุงเทพฯ: สุขภาพใจ.

วิไลพร เลิศมหาเกียรติ, ภูริวัตร คัมภีรภาพพัฒน์ และอนิราช มิ่งขวัญ. (2551).“รูปแบบการแสดงผลการค้นคืนของเครื่องมือการสืบค้นสารนิเทศบนอินเตอร์เน็ต” **วารสารวิชาการพระจอมเกล้าพระนครเหนือ**, 18(1): 89-98.

วุฒิชัย ปิยะพันธวงศ์. (2545). **การสืบค้นข้อมูลอิงแบบจำลองออปเจคสำหรับการทำดัชนีเอกสารเอ็กซ์เมแอล.** วิทยานิพนธ์. วิศวกรรมศาสตรมหาบัณฑิต. กรุงเทพฯ: มหาวิทยาลัยเกษตรศาสตร์.

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

ศุภกฤษฎิ์ นิวัฒนากูล และนิศาชล จำนงศรี. (2556). **การพัฒนารูปแบบการจัดการความรู้ในชุมชนเศรษฐกิจฐานรากในเขตอีสานใต้ ประเทศไทย**.รายงานการวิจัยฉบับสมบูรณ์สาขาวิชาเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีสุรนารี.

ศุภชัย ตั้งวงศ์ศานต์. (2553). **ระบบการจัดเก็บและการสืบค้นสารสนเทศด้วยคอมพิวเตอร์.** กรุงเทพฯ: พิทักษ์การพิมพ์.

ศุภวรรณ รัตนมุณี. (2547). การกำหนดคำค้นสารสนเทศ. **วารสารวิทยบริการมหาวิทยาลัยสงขลานครินทร์**. 15(2/3):39-47.

สถาบันส่งเสริมการจัดการความรู้เพื่อสังคม (สคส.). (2550). **ตามรอยความสำเร็จ KM ประเทศไทย.** รายงานประจำปี 2550.กรุงเทพฯ: กองทุนสนับสนุนการวิจัย (สกว.).

สมจิน เปียโคกสูง. (2553). การพัฒนาระบบนำทางความรู้เพื่อการเข้าถึงเนื้อหาในสื่อสิ่งพิมพ์. **Journal of Information Science, Suranaree University of Technology**.28(3): 9-20.

สิทธิโชค ปัญญาฤกษ์ชัย และศิพาณี นุชิตประสิทธิ์ชัย. (2009). ระบบการค้นคืนสารสนเทศโดยใช้เทคนิค N-Gram. **การประชุมวิชาการระดับชาติ NCCIT 2009 ครั้งที่ 5**.(หน้า307-312). กรุงเทพฯ: มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

สุธรรม อุมาแสงทองกุล. (2541). **การศึกษาผลการสืบค้นสารสนเทศที่ใช้เทคนิคดรรชนีแบบกำหนดค่าน้ำหนักที่จัดทำโดยมนุษย์.**กรุงเทพฯ: มหาวิทยาลัยธรรมศาสตร์.

สุนีศรีสิงห์และเสน่ห์นิลมณี. (2537). **โรคมันสำปะหลังและการป้องกันกำจัด**. เอกสารวิชาการศูนย์พืชไร่กรมวิชาการเกษตรกระทรวงเกษตรและสหกรณ์.

สินีนาถ นาคโต และไพลิน เงินประเสริฐ. (2549). **การค้นหาเอกสารงานวิจัยเชิงความหมาย.** ปริญญานิพนธ์วิทยาศาสตรบัณฑิต. กรุงเทพฯ: สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

สิริรัตน์ ประกฤติกรชัย. (2550). **การสร้างต้นแบบออนโทโลยีของพืชสมุนไพรไทย.** สารนิพนธ์วิทยาศาสตรมหาบัณฑิต. กรุงเทพฯ: มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

สำนักงานคณะกรรมการพัฒนาระบบราชการ และสถาบันเพิ่มผลผลิตแห่งชาติ.(2548). **คู่มือการจัดทำแผนการจัดการความรู้**. กรุงเทพฯ: สำนักพิมพ์สำนักงานคณะกรรมการพัฒนาระบบราชการ และสถาบันเพิ่มผลผลิตแห่งชาติ.

อรรวรรณ อุไรเรืองพันธุ์ และสมจิตร อาจอินทร์. (2009). การสรุปเอกสารเชิงความหมายโดยใช้ออนโทโลยี.In **Proceedings of the Computing and Information Technology**. (pp. 294-299). Bangkok: Thailand.

อุดมเนียบวัน. (2544). **เอกสารวิชาการพันธุ์อ้อยการปลูกดูแลรักษาอ้อย**. สถาบันวิจัยพืชไร่กรมวิชาการเกษตรกระทรวงเกษตรและสหกรณ์.

อรุณีวงษ์กอบรัษฎ์. (2526).**แมลงศัตรูพืชและการป้องกันกำจัด.**เอกสารเล่มที่ 7 งานทะเบียนและประมวลสถิติกองแผนงานและวิชาการกรมวิชาการเกษตรกระทรวงเกษตรและสหกรณ์.

อัปสรเปลี่ยนสินไชย. (2544).**โรคอ้อยที่สำคัญและการจัดการเพื่อควบคุม.** "เอกสารวิชาการการป้องกันกำศัตรูอ้อย" สถาบันวิจัยพืชไร่กรมวิชาการเกษตรกระทรวงเกษตรและสหกรณ์.

อารีย์ ธัญกิจจานุกิจ และคณะ.(2556). **การพัฒนาอรรถาภิธานและฐานข้อมูลองค์ความรู้เกษตรไทย.** ศูนย์สนเทศทางการเกษตรแห่งชาติ สำนักหอสมุด มหาวิทยาลัยเกษตรศาสตร์. [ออนไลน์]. ได้จาก:http://www. slideshare.net/boonlert/thai-arguriculture.

Amir and Mourad. (2013). A generalized frame work for Ontology-based information retrieval. **IEEE: Application to a Public-TransportationSystem**.(pp. 165-169). Sousse, Tunisia.

Baeza-Yates, R., and Ribeiro-Neto, B. (1999). **Modern Information Retrieval**. Addison-Wesley Longman.

Bennett, A. and Bennett, D. (2004). Learning Knowledge Management and Knowledge Workers in Organizational Survival in the New World. **The Intelligent Complex Adaptive System**. Boston: Elsvier.

Blake, C. and Pratt, W. (2001). Better Rules, fewer features: A Semantic Approach to Selecting Features from Text. In **Proceedings of the IEEE Data Ming Conference** (pp. 59-66). San Jose, California.

Bo, M., Yating, Y., Xi, Z., and Junlin, Z. (2010). An Ontology-Based Semantic Retrieval Model for Uyghur Search Engine. In **Proceeding of the 2nd Web Society Symposium on.** (pp. 191-196). Beijing, China.

Broughton, V. (2006). The need for a faceted classification as the basis of all methods of information retrieval. In **Proceedings of the New Information Perspective**, 58(1/2): 49-72.

Chalortham, N., Leesawat, P., Buranarach, M., and Supnithi, T. (2008). Ontology Development or Pharmaceutical Tablet Production Expert System. In **Proceeding of ECTI-CON** (pp. 205-208) Krabi, Bangkok.

Chandrasekaran, B., Josephson, J. R. and Benjamins, V. R. (1999). What are Ontologies, and Why do we need them?. **IEEE Intelligent Systems**, 14(1):20-26.

Chongchong, Z., Jing, W., Wei, H., Xiao, Y. and Xiaofeng, W. (2010). An Ontology-Based Semantic Search Model Study. In **Proceeding of the 3rdInternational Symposium on Knowledge Acquisition and Modeling.**(pp. 182-185). Wuhan, China, China.

Choudhary, S. G., Kalmegh, S. R., Deshmukh, S. N. (2013). Semantic Search Algorithms based on Page Rank and Ontology: **A Review.** In **ProceedingOf the 3rd International Conference on Intelligent Computational Systems.** (pp. 17-20). Hong Kong: China.

Craswell,N. and Hawking, D.(2002). Overview of the TREC-2002 Web Track.**Technical report In Text Retrieval Conference**. Gaithersburg,Maryland.

Cristiano, R., Daniel, S. and Marcus, P. (2004). A Hybrid Approach for Searching in theSemanticWeb. **ACM 1-58113-844-X/04/0005.**(pp. 374-383). New York, USA.

Christensen, C., & Albert, R. (2007). Using Graph Concepts to Understand the Organizationof Complex Systems. **International Journal of Bifurcation and Chaos**, 17(7): 2201–2214.

Croft, W. B. and Harper, D. J. (1979). Using Probabilistic Models of DocumentRetrieval Without Relevance Information, (pp.285-295). **Documentation.**35(4).

Davenport, T.H. and Prusak, I. (1998). **Working Knowledge : How OrganizationManage What they Know.** Cambridge. MA Harvard Business School Press.

Deerwester, S.(1990). Indexing by Latent Semantic Analysis. **Journal of the AmericanSociety for Information Science.** 41(3): 91-407.

Dik, L. (1997). **Document Ranking and the Vector-Space Model.**Hong Kong University ofScience and Technology HUEI CHUANG, Information Dimensions, Hong Kong.

Dijkstra, E. W. (1959). "A Note on Two Problems in Connexion with Graphs". **NumerischeMathematik, (**1): 269–271.

Djamel, N. (2009). A Model Based Agents and Ontology for Semantic InformationSearch. In **Proceeding of the 3rd International Conference on Research Challenges in Information Science.**(pp. 421-426). Fez, Morocco.

Dharminder, S. and Ashwani, S. (2015). Semantic Search's Impacts on Searching & Rankingof WebPages. **International Journal of Advanced Research In Computer and Communication Engineering.** 4(7): 105-107.

Enrico Francesconi, E., Sebastiano, F. and Elisabetta, M. (2008). A Framework forSemantic Mapping between Thesauri. In **Proceedings of the ICEGOV2008Conference**. (pp. 251-256). Cairo, Egypt.

Erdos, P., &Renyi, A. (1961). On the Evolution of Random Graphs. **Bulletin Institution of International Statistics**. (38): 343–347.

FAO. (2013). AGROVOC. **Food and Agricultural Organization of the United Nations**. [Online]. Available: http://aims.fao.org/website/AGROVOC-Thesaurus/sub

Fensel, D. (2004). Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce**. IEEE Intelligent Systems.** 16(1): 8-14.

Foskett, A.C. (1996). **The Subject Approach to Information.** London: Library Association Pub.

Frakes, William, B., Baeza-Yates, and Ricardo. (1992). **Information Retrieval Data Structures & Algorithms.** USA: Prentice Hall PTR.

Fulvio, C., Laura, F. (2005). **SKOS -Simple Knowledge Organization System.** [Online]. Available: http://elite.polito.it.

Gasevic, D., Djuric, D. and Devedzic, V. (2009).**Model Driven Engineering and Ontology Development.** 2nd ed. London, United Kingdom: Springer.

G. R. J., Srinivas, Niket, T. and Vasudeva, V. (2010). A Weighted Tag Similarity MeasureBased on a Collaborative Weight Model. In **Proceedings of the SMUC'10 Conference**. (pp. 79-86). Toronto, Ontario, Canada.

Guarino, N. (1997). Semantic matching: formal ontological distinctions for information organization, extraction, and integration. In M. T. Pazienza (ed.).**Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology**. [n.p.]: Springer Verlag.

Guarino, N. (1997). Understanding, building, and using ontologies: a commentary to using explicit ontologies. In van Heijst, Schreiber, and Wielinga(Eds.). **International Journal of Human and Computer Studies**. (46): 293-310.

Guarino. G. (1998). "**Formal Ontology in Information System**." Guarino N. 1 st International.

Guha, R., McCool, R. and Miller, E. (2003). **Semantic Search**. WWW2003, May 20-24 2003, Hungary. [Online]. Available: http://www2003.org/cdrom/papers/refereed/p779/ess.html.

Guha, S., Rastogi.R, and Shim.K. (1999). ROCK: A Robust Clustering Algorithm for Categorical Attributes, In **Proceedings of International Conference on Data Engineering (ICDE).** (pp. 512-521). Sydney, Australia.

Haruo, K. , Toshiaki, I. (1990). Construction of a dynamic Thesaurus andIts Use for Associated Information Retrieval. In **Proceedings of the 13th annual international ACM SIGIR conference on Research and development in information retrieval.** (pp. 227-240). New York, USA.

Hearst, Marti.(1999). **Modern Information.** [Online]. Available: http://www.sims.berkeley.edu/~hearst/irbook/10/node6.html.

Heijst, G. V., Schreiber, A.T. and Wielinga, B.J. (1997). Using explicit ontologies in KBS development.**International Journal of Human Computer Studies**. 46(2/3), 183-292.

Hodge, G. (2000). **Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files.**[Online]. Available: http://www.clir.orgpubs/reports/pub91 /contents.html.

Hunter, J. (2001). MetaNet–A metadata term theasuarus to enable semantic Interoperabilitybetween metadata domains. **Journal of Digital Information**. 1(8): 42, [Online]. Available: http://jodi.tamu.edu/ Articles/v01/i08/Hunter/.

Idri, A. and Abran, A. (2001). A Fuzzy Logic Based Set of Measures for Software Project Similarity: Validation and Possible Improvements, In **Proceedings of METRICS.** (pp. 85-96). London, England.

Jatsada, S. andet. al. (2012). Development for Prae-wa Silk Knowledge Base using Ontology, In **Proceedings of IEEE International Conference on Computer Science and Automation Engineering.** (25-27). Zhangjiajie, China.

Jihyun, L., Jun-ki, M. and Chin-Wan, C. (2009). An Effective Semantic Search Technique using Ontology. In **Proceedings of the www 2009 MADRIDI Poster Sessions**. (pp. 1057-1058). Madrid, Spain.

Jones, D.M., Bench-Capon, T.J.M. andVisser, P.R.S. (1998). Methodologies for Ontology Development. In **Proceeding of the IFIP 17th International Conferenceon IT & KNOWS.** (pp. 62-75). Budapest, Hungary.

Kashyap, V. (1999). Design and creation of ontologies for environmental information retrieval. In**Proceeding of the 12th Workshop on Knowledge Acquisition, Modeling and Management**. (pp. 1-18). Banff, Alberta, Canada.

Kesorn, K., et. Al. (2011). Visual Content Representation Using Semantically SimilarVisual Words. **Expert Systems with Applications**. 38(9): 11472-11481.

Kimoto H, Iwadera T. (1990) Construction of a Dynamic Thesaurus and Its Use for Associated Information Retrieval. In: **Proceedings of Research and Development in Information Retrieval**.(pp.227-240). New York, USA.

Kimoto H, Iwadera T. (1991). A Dynamic Thesaurus and Its Application to AssociatedInformation Retrieval. In: **Proceedings of International Joint Conferenceon Neural Networks** *(*pp. I-19). Seattle, USA.

Klyne, G., and Carroll Jeremy, J. (2004). **Resource Description Framwork (RDF): Concepts and Abstract Syntax**, W3C Candidate Recommendation. [Online]. Available: http://www.w3.org/TR/rdf-concepts.

Kunmei, W., Zhengding, L., Ruixuan, L., Xiaolin, S. and Zhigang, W. (2006). A Semantic Search Conceptual Model and Application in Security AccessControl. In **Proceedingof the 1st Asian Semantic Web Conference.** (pp. 366-375). Beijing, China.

Lancaster, F. W. (1968). **Information Retrieval Systems: Characteristics, TestingAnd Evaluation.** New York: John Wiley.

Lancaster, F. W. (2003). **Indexing and Abstracting in theory and Practice.** (3rded), London: Facet Pub.

Lee, J. H. (1995). Combining Multiple Evidence from Different Properties of Weighting Schemes, In **Proceedings of the 18thAnnual International ACM SIGIR Conference on Research and Developmentin Information Retrieval.** (pp.180-188). New York, USA.

Lee, J., Min, J., Oh, A., Chung, C. (2014). Effective Ranking and Search Techniquesfor Web Resources Considering Semantic Relationships. **JournalInformation Processing and Management**. 50(2014):132–155.

Li, Y., Bandar, Z. A., McLean, D., (2003). An Approach for Measuring Semantic Similarity between Words Using Multiple Information Sources. **Journal IEEE Transactions on Knowledge and Data Engineering.** 15(4):871-882.

Madhura, J., Charitha, D., Dilina, F., et. al. (2012).iSeS : Intelligent Semantic SearchFramework. In **Proceedings of the EATIS'12 Conference.** (pp. 215-222).Valencia, Spain.

Manning, C. D., Raghavan, P. and Schutze, H. (2008).**Introduction to Information Retrieval**. London, United Kingdom: Cambridge University Press.

Maron, M. E. and Kuhns, J. L. (1960). On Relevance, Probabilistic Indexing andInformation Retrieval. **Association for Computing Machinery**. 7(3): 216-244.

McConnell, Jeffrey J. (2001). **Analysis of Algorithms: An Active LearningApproach.** (pp. 163-168). Canada, Jones and Bartlett.

Miao, D., Duan, Q., Zhang, H., and Jiao, N. (2009). Rough set based hybrid algorithm for text classification. **Expert Systems with Applications**. 36(5): 9168-9174.

Miles, A. &Brickley, D. (2005). SKOS Core Guide. **W3C Recommendation**.[Online]. Available: http://www.w3.org/TR/2005/WD-swbp-skos-core-guide-20051102/.

Muhammad, S., Nadeem, Y., Hikmat, U. K., Imran Saeed. M, and Malik Sikandar, H. K. (2009).Relational WordNet Model for Semantic Search in Holy Quran.In **Proceeding of the International Conference on EmergingTechnologies.**(pp. 29-34). Islamabad, Pakistan.

Nielsen, J., and Mack, R. L. (Eds.). (1994). **Usability Inspection Methods,** John Wiley & Sons, NewYork, 173-202.

Niwattanakul, S. (2008).**Access to Knowledge Based-on an Ontology Model**, Ph.D. Thesis, University of La Rochelle.

Niwattanakul, S., Singthongchai, J., Naenudorn, E. and Wanapu, S. (2013). Using of Jaccard Coefficient for Keywords Similarity. In **Proceeding International MutiConf. Engineers and Computer Scientists** (pp. 380-384). HongKong, China.

Noy, N. F., McGuinness, D. L. (2001).**Ontology Development: A Guide to Creating Your First Ontology**.[Online]. Available: http://www.stanford.edu/people/dlm/papers/ontology/ontology101/ontology101-noy-mcguinness.html.

Nyamsuren, E.,Ho-Jin, C. (2009).Building a Semantic Model of a Textual Document for Efficient Search and Retrieval. In **Proceeding of the 11th International Conference on Advanced Communication Technology.** (pp. 238-302)**.** Phoenix Park: France.

Okkyung, C., SeokHyun, Y., Myeongeun, O., and Sangyong, H. (2003). SemanticWeb Search Model for Information Retrieval of the Semantic Data. **Journal Lecture Notes in Computer Science.** 27(13):588-593.

Paton, N. W., and Oscar, D. (1999). Active Database System. **ACM ComputingSurveys.** 3(1): 63-103.

Platzer, C. and Dustdar, S. (2005)**.** A Vector Space Search Engine for Web Services.Vienna University of Technology, **Distributed System Group.** Austria.

Radossslaw, O. (2004). **RDQL Tutorial:This tutorial is part of the RAP-RDF API for PHP documentation.** [Online]. Available: http:/sites.wiwiss.fu-berlin.De/suhl/bizer/rdfapi/tutorial/rdql_tutorial.htm.

Raymie, S., Krishna, B., and Farzin, M. (2000). The Term Vector Database: Fast Access to Indexing Terms for Web Pages. In **Proceeding of the 9th International WWW Conference on Computer Networks.**North-Holland Publishing Co.

Resnik, P. (1999). Semantic Similarity in a Taxonomy: An Information-Based Measure and its Application to Problems of Ambiguity on Natural Language, **Journal Artificial Intelligent Research. (**11): 95-130.

Richard, J. (2000). **Measuring Search Effectiveness**. [Online]. Available: http://Newadonis.Creighton.edu/hsl/searching/Recall-Precision.html.

Roi Blanco, Christina Lioma. (2012). Graph-base term weighting for information retrieval. **Information Retrieval.** (15):54-92.

Salton, G. (1989). **Automatic Text Processing: the Transformation, Analysis, and Retrieval of Information by Computer**. Addison-Wesley Publishing.Samir Tartir, I. BudakArpinar, Michael Moore, Amit P. Sheth, Boanerges Aleman.

Meza. (2005). **OntoQA: Metric-Based Ontology Quality nalysis**.Department of Computer Science University of Georgia Athens, GA 30602USA.

Sharma, A.K., Duhan, N. and Sharma, B. (2010). A Semantic Search System using Query Dfinitions. In **Proceedings of the IITM' 10 Conference.** (pp. 279-283).Allahabad, UP, India.

Singthongchai, J., Niwattanakul, S. and Chamnongsri, N. (2016). Semantic Term Weighting Method by Using Correlation Coefficient. **INFORMATION: An International Interdiscriplinary Journal**. 19(1): 91-106.

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

Tirawongkusol, T., and Waiyamai, K. (2009). **Ontology-based Interface Thai Herbs Search System (THSS) Ontology-based User Interface Module**. Thesis of master degree,Kasetsart University.

Tru , H. C. and Vuong , M. N. (2010). Semantic Search by Latent Ontological Features. **Journal New Generation Computing.** 30(2012): 53-71.

Tran, T., Cimiano, P., Rudolph, S., Studer, R. (2007). Ontology-Based Interpretationof Keywords for Semantic Search. **Journal Lecture Notes in Computer Science.** 48(25):523-536.

Wang, J., Miao, Z., Zhang, Y., and Zhou, B. (2009). **Querying Heterogeneous Relational Database using SPARQL**. In **Proceeding of the International Conferenceon Computer and Information Science**. (pp. 475480). Shanghai, China.

W3c. (2006). **W3C Semantic Web Activity.** Available: http://www.w3org/2001/sw.

XiangFeng, W., Quan, Z. (2007). Approach of Text Search Based on Semantic Parsing Model. In **Proceeding of the 4th International Conference on Fuzzy Systems and Knowledge Discovery.** (355-359). Washington, DC, USA.

Yang, Y., Xiong, J., and Wang, S. (2011). A Semantic Search Engine Based on SKOS Model Ontology in Agriculture. **Journal IFIP International Federation for Information Processing.** 344(1): 110-118.

Zadeh, L. A. (1993). Fuzzy sets. In Dubois D., H. Prade, and R. R. Yager, eds. **Readings in Fuzzy Sets for Intelligent Systems**. Morgan, Kaufmann.

มหาวิทยาลัยเทคโนโลยีสุรนารี

ภาคผนวก ก

แบบสำรวจเพื่อการวิจัย

แบบสอบถามความต้องการใช้คำค้นสำหรับสืบค้นความรู้ด้านพืชไร่

มหาวิทยาลัยเทคโนโลยีสุรนารี

แบบสอบถามความต้องการใช้คำค้นสำหรับสืบค้นความรู้ด้านพืชไร่

แบบสอบถามชุดนี้จัดทำขึ้นเพื่อใช้ประกอบการทำวิทยานิพนธ์เรื่อง แบบจำลองการสืบค้นเชิงความหมายสำหรับความรู้ด้านพืชไร่ซึ่งเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาการสารสนเทศดุษฎีบัณฑิตสาขาวิชาเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีสุรนารี

ผู้วิจัยใคร่ขอความอนุเคราะห์จากท่าน โปรดตอบแบบสอบถามให้ตรงกับความคิดเห็นของท่าน เพื่อจะช่วยให้งานวิจัยดำเนินไปด้วยความถูกต้อง และข้อมูลการตอบแบบสอบถามครั้งนี้จะไม่ถูกนำเสนอหรือเปิดเผยเป็นรายบุคคล โดยจะนำไปใช้เพื่อสรุปผลการวิจัยในภาพรวมเกี่ยวกับการคัดเลือกข้อมูลที่ได้มาซึ่งคำค้นไปใช้สำหรับสืบค้นความรู้ด้านพืชไร่เท่านั้น แบบสอบถามชุดนี้แบ่งออกเป็น3 ตอน ดังนี้

ตอนที่ 1 ข้อมูลของผู้ตอบแบบสอบถาม

ตอนที่ 2 คำค้นที่ต้องการใช้สำหรับสืบค้นความรู้ด้านพืชไร่ (มันสำปะหลัง, อ้อย, ข้าวโพด)

ตอนที่ 3 คำค้นเสนอแนะสำหรับความรู้ด้านพืชไร่

ตอนที่ 1 ข้อมูลส่วนบุคคล

คำชี้แจง โปรดใส่เครื่องหมาย ✓ ลงใน ( ) ที่ตรงกับท่าน

1. เพศ

( )ชาย ( ) หญิง

2. อายุ

( ) ต่ำกว่า 20 ปี ( ) 21-40 ปี

( ) 41-60 ปี ( ) ตั้งแต่ 60 ปีขึ้นไป

3. อาชีพ

( ) เกษตรกร ( ) นักเรียนหรือนักศึกษา

( ) ธุรกิจส่วนตัวหรือค้าขาย ( ) ข้าราชการหรือพนักงานรัฐวิสาหกิจ

( ) ลูกจ้างหรือพนักงานบริษัท

มหาวิทยาลัยเทคโนโลยีสุรนารี

ตอนที่ 2 คำค้นที่ต้องการใช้สำหรับสืบค้นความรู้ด้านพืชไร่ (มันสำปะหลัง, อ้อย, ข้าวโพด)

คำชี้แจง โปรดใส่เครื่องหมาย ✓ ใน ( ) ซึ่งเป็นคำค้นที่ตรงกับความคิดเห็นของท่านที่จะนำไปใช้สำหรับสืบค้นความรู้ด้านพืชไร่เกี่ยวข้องกับมันสำปะหลัง อ้อย และข้าวโพด (เลือกได้มากกว่า 1 คำค้น)

( ) พืช	( ) พืชไร่	( ) พืชเพาะปลูก
( ) พืชเศรษฐกิจ	( ) พืชสำหรับปลูก	( ) พืชที่มีประโยชน์
( ) พืชให้สปอร์	( ) พืชไร้ดอก	( ) พืชปลูก
( ) ผัก	( ) ผลิตผลจากพืช	
( ) มันสำปะหลัง	( )*Tapioca*	( ) พืชให้แป้ง
( )*Cassava*	( ) ผักรับประทานราก	( ) แป้ง
( ) ผักรับประทานราก	( ) แป้งธัญพืช	
( ) อ้อย	( ) น้ำตาล	( ) พืชน้ำตาล
( ) อ้อยโรงงาน	( )Sugarcane	( ) สารให้ความหวาน
( ) น้ำตาลอ้อย	( ) น้ำตาลทรายแดง	( ) น้ำตาลทรายขาว
( ) น้ำตาลทรายบริสุทธิ์	( ) สารที่ใช้แทนน้ำตาล	( ) เพอริลลาร์ทีน
( ) สเตวิโอไซด์	( ) ทัวมาทิน	( ) น้ำอ้อย
( ) ข้าวโพด	( ) ข้าวโพดข้าวเหนียว	( ) ข้าวโพดหวาน
( ) ข้าวโพดเลี้ยงสัตว์	( ) ข้าวโพดฝักอ่อน	( )Maize
( ) ข้าวโพดฝักสด	( ) แป้งข้าวโพด	( ) ธัญพืช
( ) พันธุ์	( ) พันธุ์ปลูก	( ) พันธุ์นำเข้า
( ) พันธุ์ดั้งเดิม	( ) พันธุ์ลูกผสม	( ) พืชพันธุ์พื้นเมือง
( ) พันธุ์พืชเฉพาะถิ่น	( ) สายพันธุ์บริสุทธิ์	( ) พันธุ์ไร้เมล็ด
( ) พันธุ์พืชท้องถิ่น	( ) พันธุ์ที่ให้ผลผลิตสูง	( ) พันธุ์ที่เพาะปลูก
( ) การปรับปรุงพันธุ์พืช	( ) การผสมพันธุ์	( ) การปลูกเชื้อ
( ) กรรมวิธีปรับปรุงพันธุ์	( ) วิธีการปลูกเชื้อ	
( ) การดูแลรักษา	( ) การบำรุงรักษา	( ) การได้รับสารอาหาร
( ) วิธีเพาะปลูก	( ) ปุ๋ย	( ) การใส่ปุ๋ย

มหาวิทยาลัยเทคโนโลยีสุรนารี

( ) การใส่ปุ๋ยในฤดูเดียว	( ) การหว่านปุ๋ย	( ) การหว่านปุ๋ยในร่อง
( ) การผสมปุ๋ย	( ) การใส่ปุ๋ยปรับปรุงดิน	( ) การเพิ่มผลผลิต
( ) ดินที่เหมาะสม	( ) ฤดูปลูก	
( ) เทคโนโลยีที่เหมาะสม	( ) เทคโนโลยีการผลิตที่ใช้ต้นทุนต่ำ	( ) เทคโนโลยีชีวภาพด้านพืช
( ) เทคโนโลยีการผลิตที่เหมาะสม		
( ) การเก็บรักษา	( ) การเก็บรักษาพืชผล	( ) การเก็บ
( ) การอนุรักษ์ (การเก็บรักษา)	( ) คุณภาพการเก็บ	( ) ระบบหลังการเก็บเกี่ยว
( ) สรีรวิทยาหลังการเก็บเกี่ยว	( ) สถานที่เก็บรักษา	( ) การสูญเสียจากการปฏิบัติต่อผลิตผล
( ) การสูญเสียเนื่องจากการเก็บรักษา	( ) ความเสียหายหลังการเก็บเกี่ยว	( ) การเสื่อมสภาพทางชีวภาพ
( ) การสูญเสียในกระบวนการผลิต	( ) กิจกรรมที่เกิดขึ้นหลังการเก็บเกี่ยว	( ) การควบคุมหลังการเก็บเกี่ยว
( ) การเน่าเสีย (การเสื่อมสภาพ)		
( ) เครื่องมือและอุปกรณ์	( ) เครื่องเก็บเกี่ยว	( ) เครื่องเก็บเกี่ยวแบบวางราย
( ) เครื่องมือหลังการเก็บเกี่ยว	( ) เครื่องกะเทาะเปลือก	( ) เครื่องฉีด
( ) เครื่องมือทำความสะอาด	( ) เครื่องตัดหญ้า	( ) เครื่องล้าง
( ) เครื่องตัดและปรับสภาพพืช	( ) เครื่องฉีดสารเข้าต้นไม้	( ) เครื่องทำความสะอาดราก
( ) เครื่องฉีดสารลงดิน	( ) เครื่องอบแห้ง	( ) เครื่องคัดแยก
( ) เครื่องแยกชนิด		
( ) โรค	( ) โรคที่เกิดจากปรสิต	( ) โรคจากปรสิต
( ) โรคราสนิม	( ) โรคราน้ำค้าง	( ) โรคที่เกิดจากเชื้อรา

มหาวิทยาลัยเทคโนโลยีสุรนารี

( ) โรคราแป้ง	( ) โรคใบจุด	( ) โรคพืช
( ) โรคหนอนเน่าของผึ้ง	( ) เชื้อรา	( ) เชื้อราแป้ง
( ) โรคแอนแทรกโนส	( ) โรคเน่าเปื่อย	( ) โรคใบขาว
( ) โรคหนอนเน่าอเมริกันของผึ้ง	( ) โรคติดเชื้อพยาธิไส้เดือน	( ) โรคติดเชื้อพยาธิไส้เดือนตัวผึ้ง
( ) โรคราเขม่าดำ	( ) โรคจุดด่าง	( ) โรครากหรือหัวเน่า
( ) โรคของแมลง	( ) ศัตรูในโรงเก็บ	( ) ศัตรูหลังการเก็บเกี่ยว
( ) แมลงศัตรู	( ) ไร	( ) ไรที่เป็นอันตราย
( ) ไรศัตรู	( ) ไส้เดือนฝอย	( ) ไรที่ทำให้เกิดอันตราย
( )แมลงหวี่ขาว	( ) เพลี้ยแป้งมัน	( ) ไรที่ทำให้เกิดความเสียหาย
( ) ตัวเบียน	( ) แมลงช้างปีกใส	( ) ศัตรูพืชหรือสัตว์
( ) ตัวห้ำ	( ) ด้วงเต่า	( ) ศัตรูพืช
( ) สารป้องกันกำจัดแมลง	( ) สารฆ่าแมลง	( ) สารควบคุมการเจริญเติบโตของพืช
( ) สารป้องกันกำจัดศัตรูพืช	( ) สารป้องกันกำจัดเพลี้ย	( ) สารป้องกันกำจัดแมลงจากพืช
( ) อาร์โบริไซด์	( ) ยาฆ่าหญ้า	( ) สารป้องกันกำจัดวัชพืช
( ) การควบคุมศัตรูพืช (หลังเก็บเกี่ยว)	( ) การฉีด	( ) การให้ทางใบ
( ) การฉีดสารลงดิน	( ) วิธีการใช้	
( ) ผลิตภัณฑ์	( ) การแปรรูป (ของผลิตภัณฑ์)	( ) ผลิตภัณฑ์แปรรูปจากพืช
( ) เอทานอล	( ) กระบวนการผลิต	( ) ผลิตภัณฑ์จากสตาร์ช
( ) ตลาดผลิตผลทางการเกษตร	( ) ตลาดในประเทศ	( ) ตลาดโลก
( ) ตลาดสินค้าเกษตร	( ) ราคาขายส่ง	( ) ราคาขายปลีก

มหาวิทยาลัยเทคโนโลยีสุรนารี

( ) ราคาผู้บริโภค	( ) ราคาผู้ผลิต	( ) ราคาสูงสุด
( ) การส่งออก	( ) การค้าระหว่างประเทศ	( ) ราคา

ตอนที่ 3คำค้นเสนอแนะสำหรับความรู้ด้านพืชไร่

1. มันสำปะหลัง

.....

.....

.....

2. อ้อย

.....

.....

.....

3. ข้าวโพด

.....

.....

.....

ขอขอบคุณในความกรุณาการตอบแบบสอบถาม

เจษฎา สิงห์ทองชัย

สาขาวิชาเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีสุรนารี

มหาวิทยาลัยเทคโนโลยีสุรนารี

มหาวิทยาลัยเทคโนโลยีสุรนารี

ภาคผนวก ข

ผลการหาประสิทธิภาพเครื่องมือและผลการทดลอง

มหาวิทยาลัยเทคโนโลยีสุรนารี

ตารางที่ ข-1 ผลการประเมินความสอดคล้องระหว่างข้อคำถามคำค้นความรู้ด้านพืชไร่ของผู้เชี่ยวชาญ

คำค้น	ผู้เชี่ยวชาญ			ค่า IOC
	คนที่ 1	คนที่ 2	คนที่ 3	
พืช	+1	+1	+1	1.00
พืชไร่	+1	+1	+1	1.00
พืชเพาะปลูก	+1	+1	0	0.66
พืชเศรษฐกิจ	+1	+1	+1	1.00
พืชสำหรับปลูก	+1	+1	+1	1.00
พืชที่มีประโยชน์	+1	0	+1	0.66
พืชให้สปอร์	+1	+1	+1	1.00
พืชไร้ดอก	+1	+1	0	0.66
พืชปลูก	+1	+1	+1	1.00
ผัก	+1	+1	+1	1.00
ผลิตผลจากพืช	+1	+1	+1	0.66
มันสำปะหลัง	+1	+1	0	0.66
Tapioca	+1	+1	+1	1.00
พืชให้แป้ง	+1	0	+1	0.66
Cassava	+1	+1	+1	1.00
ผักรับประทานราก	+1	+1	0	0.66
แป้ง	+1	+1	+1	1.00
ผักรับประทานราก	+1	+1	+1	1.00
แป้งธัญพืช	0	+1	+1	0.66
อ้อย	+1	+1	+1	1.00
น้ำตาล	+1	+1	+1	1.00
พืชน้ำตาล	+1	0	+1	0.66

มหาวิทยาลัยเทคโนโลยีสุรนารี

ตารางที่ ข-1 ผลการประเมินความสอดคล้องระหว่างข้อคำถามคำค้นความรู้ด้านพืชไร่ของผู้เชี่ยวชาญ(ต่อ)

คำค้น	ผู้เชี่ยวชาญ			ค่า IOC
	คนที่ 1	คนที่ 2	คนที่ 3	
อ้อยโรงงาน	0	+1	+1	0.66
Sugarcane	0	+1	+1	0.66
สารให้ความหวาน	+1	+1	+1	1.00
น้ำตาลอ้อย	+1	+1	+1	1.00
น้ำตาลทรายแดง	+1	+1	+1	1.00
น้ำตาลทรายขาว	+1	+1	+1	1.00
น้ำตาลทรายบริสุทธิ์	+1	+1	+1	1.00
สารที่ใช้แทนน้ำตาล	+1	+1	+1	1.00
เพอริลลาร์ทีน	+1	0	+1	0.66
สเตวิโอไซด์	+1	+1	0	0.66
ทัวมาทิน	+1	+1	+1	1.00
น้ำอ้อย	+1	+1	+1	1.00
ข้าวโพด	+1	+1	0	0.66
ข้าวโพดข้าวเหนียว	+1	+1	0	0.66
ข้าวโพดหวาน	+1	+1	+1	1.00
ข้าวโพดเลี้ยงสัตว์	+1	+1	+1	1.00
ข้าวโพดฝักอ่อน	+1	+1	+1	1.00
Maize	+1	+1	+1	1.00
ข้าวโพดฝักสด	+1	+1	+1	1.00
แป้งข้าวโพด	+1	+1	+1	1.00
ธัญพืช	+1	+1	0	0.66
พันธุ์	+1	+1	+1	1.00
พันธุ์ปลูก	+1	+1	+1	1.00
พันธุ์นำเข้า	+1	+1	0	0.66
พันธุ์ดั้งเดิม	+1	+1	0	0.66

มหาวิทยาลัยเทคโนโลยีสุรนารี

ตารางที่ ข-1 ผลการประเมินความสอดคล้องระหว่างข้อคำถามคำค้นความรู้ด้านพืชไร่ของผู้เชี่ยวชาญ(ต่อ)

คำค้น	ผู้เชี่ยวชาญ			ค่า IOC
	คนที่ 1	คนที่ 2	คนที่ 3	
พันธุ์ลูกผสม	0	+1	+1	0.66
พืชพันธุ์พื้นเมือง	+1	+1	+1	1.00
พันธุ์พืชเฉพาะถิ่น	+1	+1	+1	1.00
สายพันธุ์บริสุทธิ์	+1	+1	+1	1.00
พันธุ์ไร้เมล็ด	+1	+1	+1	1.00
พันธุ์พืชท้องถิ่น	+1	+1	+1	1.00
พันธุ์ที่ให้ผลผลิตสูง	+1	+1	+1	1.00
พันธุ์ที่เพาะปลูก	0	+1	+1	0.66
การปรับปรุงพันธุ์พืช	0	+1	+1	0.66
การผสมพันธุ์	+1	+1	+1	1.00
การปลูกเชื้อ	+1	+1	+1	1.00
กรรมวิธีปรับปรุงพันธุ์	+1	+1	+1	1.00
วิธีการปลูกเชื้อ	+1	+1	+1	1.00
การดูแลรักษา	+1	+1	+1	1.00
การบำรุงรักษา	+1	+1	+1	1.00
การได้รับสารอาหาร	+1	+1	+1	1.00
วิธีเพาะปลูก	+1	+1	+1	1.00
การบำรุงรักษา	+1	+1	+1	1.00
การได้รับสารอาหาร	+1	+1	+1	1.00
วิธีเพาะปลูก	+1	+1	+1	1.00
ปุ๋ย	+1	+1	+1	1.00
การใส่ปุ๋ย	+1	+1	+1	1.00
การใส่ปุ๋ยในฤดูเดียว	+1	+1	+1	1.00
การหว่านปุ๋ย	+1	+1	+1	1.00
การหว่านปุ๋ยในร่อง	+1	+1	+1	1.00

มหาวิทยาลัยเทคโนโลยีสุรนารี

ตารางที่ ข-1 ผลการประเมินความสอดคล้องระหว่างข้อคำถามคำค้นความรู้ด้านพืชไร่ของผู้เชี่ยวชาญ(ต่อ)

คำค้น	ผู้เชี่ยวชาญ			ค่า IOC
	คนที่ 1	คนที่ 2	คนที่ 3	
การผสมปุ๋ย	+1	+1	+1	1.00
การใส่ปุ๋ยปรับปรุงดิน	+1	+1	+1	1.00
การเพิ่มผลผลิต	+1	+1	0	0.66
ดินที่เหมาะสม	+1	+1	+1	1.00
ฤดูปลูก	+1	+1	+1	1.00
เทคโนโลยีที่เหมาะสม	+1	0	+1	0.66
เทคโนโลยีการผลิตที่ใช้ต้นทุนต่ำ	0	+1	+1	0.66
เทคโนโลยีชีวภาพด้านพืช	+1	+1	0	0.66
เทคโนโลยีการผลิตที่เหมาะสม	+1	+1	+1	1.00
การเก็บรักษา	+1	+1	+1	1.00
การเก็บรักษาพืชผล	+1	0	+1	0.66
การเก็บ	+1	0	+1	0.66
การอนุรักษ์ (การเก็บรักษา)	+1	+1	+1	1.00
คุณภาพการเก็บ	+1	+1	+1	1.00
ระบบหลังการเก็บเกี่ยว	+1	0	+1	0.66
สรีรวิทยาหลังการเก็บเกี่ยว	+1	0	+1	0.66
สถานที่เก็บรักษา	+1	+1	+1	1.00
การสูญเสียจากการปฏิบัติต่อผลิตผล	+1	+1	+1	1.00
การสูญเสียเนื่องจากการเก็บรักษา	+1	+1	+1	1.00
ความเสียหายหลังการเก็บเกี่ยว	+1	+1	0	0.66
การเสื่อมสภาพทางชีวภาพ	+1	+1	0	0.66
การสูญเสียในกระบวนการผลิต	+1	+1	0	0.66
กิจกรรมที่เกิดขึ้นหลังการเก็บเกี่ยว	+1	+1	0	0.66

มหาวิทยาลัยเทคโนโลยีสุรนารี

ตารางที่ ข-1 ผลการประเมินความสอดคล้องระหว่างข้อคำถามคำค้นความรู้ด้านพืชไร่ของผู้เชี่ยวชาญ(ต่อ)

คำค้น	ผู้เชี่ยวชาญ			ค่า IOC
	คนที่ 1	คนที่ 2	คนที่ 3	
การควบคุมหลังการเก็บเกี่ยว	0	+1	+1	0.66
การเน่าเสีย (การเสื่อมสภาพ)	0	+1	+1	0.66
เครื่องมือและอุปกรณ์	+1	+1	+1	1.00
เครื่องเก็บเกี่ยว	+1	+1	+1	1.00
เครื่องเก็บเกี่ยวแบบวางราย	+1	+1	+1	1.00
เครื่องมือหลังการเก็บเกี่ยว	+1	+1	0	0.66
เครื่องกะเทาะเปลือก	+1	+1	+1	1.00
เครื่องฉีด	+1	+1	+1	1.00
เครื่องมือทำความสะอาด	+1	+1	0	0.66
เครื่องตัดหญ้า	+1	+1	+1	1.00
เครื่องล้าง	+1	+1	+1	1.00
เครื่องตัดและปรับสภาพพืช	+1	+1	0	0.66
เครื่องฉีดสารเข้าต้นไม้	+1	+1	+1	1.00
เครื่องทำความสะอาดราก	+1	+1	+1	1.00
เครื่องฉีดสารลงดิน	+1	+1	+1	1.00
เครื่องอบแห้ง	+1	+1	+1	1.00
เครื่องคัดแยก	+1	+1	+1	1.00
เครื่องแยกชนิด	+1	+1	+1	1.00
โรค	+1	+1	+1	1.00
โรคที่เกิดจากปรสิต	0	+1	+1	0.66
โรคจากปรสิต	+1	+1	+1	1.00
โรคราสนิม	+1	+1	+1	1.00
โรคราน้ำค้าง	+1	+1	+1	1.00
โรคที่เกิดจากเชื้อรา	+1	+1	+1	1.00

มหาวิทยาลัยเทคโนโลยีสุรนารี

ตารางที่ ข-1 ผลการประเมินความสอดคล้องระหว่างข้อคำถามคำค้นความรู้ด้านพืชไร่ของผู้เชี่ยวชาญ(ต่อ)

คำค้น	ผู้เชี่ยวชาญ			ค่า IOC
	คนที่ 1	คนที่ 2	คนที่ 3	
โรคราแป้ง	+1	+1	+1	1.00
โรคใบจุด	+1	+1	+1	1.00
โรคพืช	+1	+1	+1	1.00
โรคหนอนเน่าของผึ้ง	+1	0	+1	0.66
เชื้อรา	+1	+1	+1	1.00
เชื้อราแป้ง	+1	+1	+1	1.00
โรคแอนแทรกโนส	+1	+1	+1	1.00
โรคเน่าเปื่อย	+1	+1	+1	1.00
โรคใบขาว	+1	+1	+1	1.00
โรคหนอนเน่าอเมริกันของผึ้ง	+1	+1	0	0.66
โรคติดเชื้อพยาธิไส้เดือน	+1	+1	0	0.66
โรคติดเชื้อพยาธิไส้เดือนตัวผึ้ง	+1	0	+1	0.66
โรคราเขม่าดำ	+1	+1	+1	1.00
โรคจุดด่าง	+1	+1	+1	1.00
โรครากหรือหัวเน่า	+1	+1	+1	1.00
โรคของแมลง	+1	+1	+1	1.00
ศัตรูในโรงเก็บ	+1	+1	+1	1.00
ศัตรูหลังการเก็บเกี่ยว	+1	+1	+1	1.00
แมลงศัตรู	+1	+1	+1	1.00
ไร	+1	+1	+1	1.00
ไรที่เป็นอันตราย	0	+1	+1	0.66
ไรศัตรู	+1	+1	+1	1.00
ไส้เดือนฝอย	+1	+1	+1	1.00
ไรที่ทำให้เกิดอันตราย	+1	+1	+1	1.00
แมลงหวี่ขาว	+1	+1	+1	1.00

มหาวิทยาลัยเทคโนโลยีสุรนารี

ตารางที่ ข-1 ผลการประเมินความสอดคล้องระหว่างข้อคำถามคำค้นความรู้ด้านพืชไร่ของผู้เชี่ยวชาญ(ต่อ)

คำค้น	ผู้เชี่ยวชาญ			ค่า IOC
	คนที่ 1	คนที่ 2	คนที่ 3	
เพลี้ยแป้งมัน	+1	+1	+1	1.00
ไรที่ทำให้เกิดความเสียหาย	0	+1	+1	0.66
ตัวเบียน	+1	+1	+1	1.00
แมลงช้างปีกใส	+1	+1	+1	1.00
ศัตรูพืชหรือสัตว์	+1	+1	+1	1.00
ตัวห้ำ	+1	+1	+1	1.00
ด้วงเต่า	+1	+1	+1	1.00
ศัตรูพืช	+1	+1	+1	1.00
สารป้องกันกำจัดแมลง	+1	+1	+1	1.00
สารฆ่าแมลง	+1	+1	+1	1.00
สารควบคุมการเจริญเติบโตของพืช	+1	+1	0	0.66
สารป้องกันกำจัดศัตรูพืช	+1	+1	+1	1.00
สารป้องกันกำจัดเพลี้ย	+1	+1	+1	1.00
สารป้องกันกำจัดแมลงจากพืช	+1	+1	+1	1.00
อาร์โบริไซด์	+1	+1	+1	1.00
ยาฆ่าหญ้า	+1	+1	+1	1.00
สารป้องกันกำจัดวัชพืช	+1	+1	+1	1.00
การควบคุมศัตรูพืช (หลังเก็บเกี่ยว)	0	+1	+1	0.66
การฉีด	+1	+1	+1	1.00
การให้ทางใบ	+1	+1	+1	1.00
การฉีดสารลงดิน	+1	+1	+1	1.00
วิธีการใช้	+1	+1	+1	1.00
ผลิตภัณฑ์	+1	+1	+1	1.00

มหาวิทยาลัยเทคโนโลยีสุรนารี

ตารางที่ ข-1 ผลการประเมินความสอดคล้องระหว่างข้อคำถามคำค้นความรู้ด้านพืชไร่ของผู้เชี่ยวชาญ(ต่อ)

คำค้น	ผู้เชี่ยวชาญ			ค่า IOC
	คนที่ 1	คนที่ 2	คนที่ 3	
การแปรรูป (ของผลิตภัณฑ์)	+1	0	+1	0.66
ผลิตภัณฑ์แปรรูปจากพืช	+1	+1	+1	1.00
เอทานอล	+1	+1	+1	1.00
กระบวนการผลิต	+1	+1	+1	1.00
ผลิตภัณฑ์จากสตาร์ช	+1	+1	+1	1.00
ตลาดผลิตผลทางการเกษตร	+1	+1	+1	1.00
ตลาดในประเทศ	+1	+1	+1	1.00
ตลาดโลก	+1	+1	+1	1.00
ตลาดสินค้าเกษตร	+1	+1	+1	1.00
ราคาขายส่ง	+1	+1	+1	1.00
ราคาขายปลีก	+1	+1	+1	1.00
ราคาผู้บริโภค	+1	+1	+1	1.00
ราคาผู้ผลิต	+1	+1	+1	1.00
ราคาสูงสุด	+1	+1	+1	1.00
การส่งออก	+1	+1	+1	1.00
การค้าระหว่างประเทศ	+1	+1	+1	1.00
ราคา	+1	+1	+1	1.00
เฉลี่ย				0.91

มหาวิทยาลัยเทคโนโลยีสุรนารี

ตารางที่ ข-3ผลการวิเคราะห์คำค้นจากผู้ใช้ที่มีร้อยละ 25 ขึ้นไป

คำค้น	จำนวน	ร้อยละ	คำค้น	จำนวน	ร้อยละ
พืชไร่	384	100.0	พืชเศรษฐกิจ	200	52.1
พืชเพาะปลูก	282	73.4	พืชปลูก	186	48.4
ผลิตผลจากพืช	160	41.7	มันสำปะหลัง	380	99.0
Cassava	152	39.6	พืชให้แป้ง	201	52.3
แป้งธัญพืช	180	46.9	อ้อย	376	97.9
น้ำตาล	222	57.8	พืชน้ำตาล	234	60.9
อ้อยโรงงาน	305	79.4	Sugarcane	167	43.5
สารให้ความหวาน	102	26.6	น้ำตาลทรายบริสุทธิ์	102	26.6
น้ำตาลอ้อย	285	74.2	น้ำอ้อย	298	77.6
ข้าวโพด	367	95.6	Maize	167	43.5
ข้าวโพดข้าวเหนียว	321	83.6	ข้าวโพดหวาน	210	54.7
ข้าวโพดเลี้ยงสัตว์	136	35.4	ข้าวโพดฝักอ่อน	160	41.7
ข้าวโพดฝักสด	160	41.7	แป้งข้าวโพด	180	46.9
พันธุ์	300	78.1	พันธุ์ปลูก	231	60.2
พันธุ์นำเข้า	181	47.1	พันธุ์ดั้งเดิม	188	49.0
พันธุ์ลูกผสม	216	56.3	พืชพันธุ์พื้นเมือง	144	37.5
พันธุ์พืชท้องถิ่น	162	42.2	พันธุ์ที่ให้ผลผลิตสูง	205	53.4
การปรับปรุงพันธุ์พืช	286	74.5	การดูแลรักษา	301	78.4
การบำรุงรักษา	316	82.3	วิธีเพาะปลูก	318	82.8
ปุ๋ย	302	78.6	การใส่ปุ๋ย	275	71.6
การหว่านปุ๋ย	256	66.7	การหว่านปุ๋ยในร่อง	243	63.3
การผสมปุ๋ย	299	77.9	การใส่ปุ๋ยปรับปรุงดิน	240	62.5
การเพิ่มผลผลิต	289	75.3	ดินที่เหมาะสม	109	28.4
เทคโนโลยีการผลิตที่ใช้ต้นทุนต่ำ	300	78.1	เทคโนโลยีการผลิตที่เหมาะสม	266	69.3
การเก็บรักษา	245	63.8	การเก็บรักษาพืชผล	346	90.1

มหาวิทยาลัยเทคโนโลยีสุรนารี

ตารางที่ ข-3ผลการวิเคราะห์คำค้นจากผู้ใช้ที่มีร้อยละ 25 ขึ้นไป (ต่อ)

คำค้น	จำนวน	ร้อยละ	คำค้น	จำนวน	ร้อยละ
คุณภาพการเก็บ	177	46.1	ระบบหลังการเก็บเกี่ยว	98	25.5
สถานที่เก็บรักษา	128	33.3	การสูญเสียเนื่องจากการเก็บรักษา	160	41.7
ความเสียหายหลังการเก็บเกี่ยว	180	46.9	การเสื่อมสภาพทางชีวภาพ	118	30.7
การควบคุมหลังการเก็บเกี่ยว	286	74.5	การเน่าเสีย (การเสื่อมสภาพ)	276	71.9
เครื่องมือและอุปกรณ์	284	74.0	เทคโนโลยีหลังการเก็บเกี่ยว	155	40.4
เครื่องเก็บเกี่ยว	233	60.7	เครื่องกะเทาะเปลือก	147	38.3
เครื่องมือหลังการเก็บเกี่ยว	200	52.1	เครื่องฉีด	190	49.5
เครื่องอบแห้ง	275	71.6	เครื่องคัดแยก	168	43.8
โรค	244	63.5	โรคจากปรสิต	119	31.0
โรคราสนิม	244	63.5	โรคราน้ำค้าง	244	63.5
โรคที่เกิดจากเชื้อรา	182	47.4	โรคราแป้ง	175	45.6
โรคใบจุด	172	44.8	เชื้อราแป้ง	169	44.0
เชื้อรา	155	40.4	โรคใบขาว	133	34.6
โรคแอนแทรกโนส	235	61.2	โรคเน่าเปื่อย	116	30.2
โรคติดเชื้อพยาธิไส้เดือน	100	26.0	โรคราเขม่าดำ	119	31.0
โรคของแมลง	279	72.7	ไส้เดือนฝอย	105	27.3
เพลี้ยแป้งมัน	110	28.6	ไรที่ทำให้เกิดความเสียหาย	122	31.8
แมลงหวี่ขาว	115	29.9	สารป้องกันกำจัดแมลง	230	59.9
สารฆ่าแมลง	140	36.5	สารป้องกันกำจัดศัตรูพืช	177	46.1
สารป้องกันกำจัดเพลี้ย	118	30.7	สารป้องกันกำจัดวัชพืช	130	33.9
การฉีด	109	28.4	การแปรรูป (ของผลิตภัณฑ์)	160	41.7

มหาวิทยาลัยเทคโนโลยีสุรนารี

ตารางที่ ข-3ผลการวิเคราะห์คำค้นจากผู้ใช้ที่มีร้อยละ 25 ขึ้นไป (ต่อ)

คำค้น	จำนวน	ร้อยละ	คำค้น	จำนวน	ร้อยละ
ผลิตภัณฑ์แปรรูปจากพืช	105	27.3	ตลาดสินค้าเกษตร	142	37.0
ตลาดในประเทศ	165	43.0	การค้าระหว่างประเทศ	167	43.5
ตลาดสินค้าเกษตร	235	61.2	ราคาขายปลีก	354	92.2
การส่งออก	205	53.4	ตลาดโลก	299	77.9

- ผลการวิเคราะห์ประสิทธิภาพการสืบค้นจากวิธีการให้น้ำหนักคำค้น

ตารางที่ ข-4ผลประสิทธิภาพการให้น้ำหนักคำสำคัญ 1-3 คำค้น 100 ข้อสอบถาม

คำค้น	ความแม่นยำเฉลี่ย		ความผิดพลาด	
	IDF	***Corr***	**IDF**	***Corr***
พืชไร่พืชเศรษฐกิจ	0.773	0.773	0.080	0.087
พืชเศรษฐกิจ	0.946	0.949	0.102	0.102
มันสำปะหลัง พันธุ์ ดินที่เหมาะสม	0.784	0.784	0.100	0.093
*Cassava*พันธุ์ที่ให้ผลผลิตสูง	0.926	0.947	0.100	0.100
อ้อย การปลูก	0.853	0.859	0.098	0.093
Sugarcane	0.981	0.989	0.104	0.103
น้ำตาลSugarcane	0.942	0.968	0.102	0.102
ข้าวโพดเทคโนโลยีหลังการเก็บเกี่ยว	0.771	0.780	0.092	0.086
Maizeพันธุ์ที่ให้ผลผลิตสูง	0.945	0.957	0.103	0.101
ข้าวโพด	0.946	0.959	0.103	0.100
ข้าวโพด Maizeพันธุ์	0.934	0.948	0.102	0.098
ข้าวโพด การบำรุงรักษาวิธีเพาะปลูก	0.913	0.916	0.102	0.099
ปุ๋ย มันสำปะหลัง การหว่านปุ๋ย	0.775	0.796	0.090	0.087
การเก็บรักษาพันธุ์ ข้าวโพด	0.729	0.751	0.091	0.087
เครื่องเก็บเกี่ยวมันสำปะหลัง	0.934	0.952	0.101	0.101
อ้อย โรคสารป้องกันกำจัดแมลง	0.803	0.819	0.090	0.088
โรคของแมลงมันสำปะหลัง สารฆ่าแมลง	0.637	0.673	0.083	0.075
มันสำปะหลัง สารฆ่าแมลง	0.888	0.909	0.095	0.096

ตารางที่ ข-4ผลประสิทธิภาพการให้น้ำหนักคำสำคัญ 1-3 คำค้น 100 ข้อสอบถาม (ต่อ)

คำค้น	ความแม่นยำเฉลี่ย		ความผิดพลาด	
	IDF	*Corr*	IDF	*Corr*
การแปรรูป อ้อย	0.673	0.688	0.084	0.081
ตลาดสินค้าเกษตร พืชไร่	0.944	0.969	0.105	0.103
โรคของแมลง อ้อย สารฆ่าแมลง	0.824	0.899	0.118	0.090
โรคของแมลง ข้าวโพด สารฆ่าแมลง	0.916	0.870	0.101	0.051
การแปรรูป มันสำปะหลัง	0.825	0.891	0.129	0.097
การแปรรูป ข้าวโพด	0.889	0.897	0.101	0.136
พืชเพาะปลูกพันธุ์นำเข้าพืชให้แป้ง	0.858	0.886	0.123	0.102
พืชน้ำตาลพันธุ์ปลูกพันธุ์ที่ให้ผลผลิตสูง	0.812	0.902	0.150	0.073
อ้อยโรงงานพันธุ์ที่ให้ผลผลิตสูง	0.869	0.879	0.103	0.095
สารให้ความหวาน	0.933	0.927	0.087	0.055
น้ำตาลทรายบริสุทธิ์	0.895	0.872	0.064	0.136
น้ำตาลอ้อยน้ำอ้อย	0.918	0.951	0.121	0.143
ข้าวโพดข้าวเหนียวพันธุ์นำเข้า	0.816	0.932	0.147	0.137
ข้าวโพดข้าวเหนียวพันธุ์ปลูก	0.910	0.947	0.140	0.126
พันธุ์ ข้าวโพดเลี้ยงสัตว์	0.856	0.936	0.150	0.067
การปลูก ข้าวโพดฝักอ่อน พันธุ์	0.833	0.932	0.078	0.115
ข้าวโพดฝักสด	0.851	0.945	0.145	0.124
แป้งข้าวโพดแป้งธัญพืช	0.884	0.938	0.082	0.106
อ้อย พันธุ์ที่ให้ผลผลิตสูง	0.865	0.931	0.074	0.109
มันสำปะหลัง พันธุ์นำเข้า	0.920	0.904	0.086	0.083
ข้าวโพด พันธุ์ดั้งเดิม	0.879	0.899	0.095	0.053
พันธุ์ลูกผสมพันธุ์นำเข้าข้าวโพดฝักอ่อน	0.854	0.954	0.053	0.112
ข้าวโพด พืชพันธุ์พื้นเมืองพันธุ์พืชท้องถิ่น	0.868	0.948	0.131	0.136
การดูแลรักษา วิธีเพาะปลูก ข้าวโพด	0.896	0.891	0.089	0.123
อ้อย พันธุ์ที่ให้ผลผลิตสูง	0.806	0.931	0.149	0.094
การปรับปรุงพันธุ์พืชข้าวโพด	0.845	0.898	0.074	0.075

มหาวิทยาลัยเทคโนโลยีสุรนารี

ตารางที่ ข-4ผลประสิทธิภาพการให้น้ำหนักคำสำคัญ 1-3 คำค้น 100 ข้อสอบถาม (ต่อ)

คำค้น	ความแม่นยำเฉลี่ย		ความผิดพลาด	
	IDF	*Corr*	IDF	*Corr*
การดูแลรักษา อ้อย	0.843	0.926	0.063	0.145
วิธีเพาะปลูก ข้าวโพด	0.902	0.883	0.133	0.116
ปุ๋ย มันสำปะหลัง	0.870	0.926	0.073	0.125
การใส่ปุ๋ย อ้อย	0.925	0.941	0.149	0.055
การหว่านปุ๋ย ข้าวโพด	0.908	0.960	0.069	0.119
การหว่านปุ๋ยในร่อง อ้อย	0.921	0.876	0.145	0.096
การผสมปุ๋ยการดูแลรักษา	0.816	0.934	0.066	0.050
ดินที่เหมาะสมการใส่ปุ๋ยปรับปรุงดินการเพิ่มผลผลิต	0.911	0.893	0.067	0.069
ข้าวโพด การเพิ่มผลผลิต	0.838	0.899	0.068	0.068
ข้าวโพดฝักอ่อนเชื้อราสารป้องกันกำจัดวัชพืช	0.829	0.939	0.129	0.062
เทคโนโลยีการผลิตที่ใช้ต้นทุนต่ำ	0.921	0.923	0.144	0.053
เทคโนโลยีการผลิตที่เหมาะสม	0.941	0.920	0.103	0.087
การเก็บรักษาพืชผลการเก็บรักษา ข้าวโพดเลี้ยงสัตว์	0.933	0.955	0.066	0.075
มันสำปะหลัง	0.917	0.933	0.120	0.091
คุณภาพการเก็บ ข้าวโพดฝักสด	0.823	0.891	0.142	0.094
ระบบหลังการเก็บเกี่ยว	0.812	0.878	0.107	0.120
สถานที่เก็บรักษา พืชไร่	0.836	0.962	0.110	0.089
การสูญเสียเนื่องจากการเก็บรักษา	0.909	0.890	0.106	0.072
ความเสียหายหลังการเก็บเกี่ยว	0.823	0.916	0.141	0.068
การเสื่อมสภาพทางชีวภาพ	0.909	0.883	0.152	0.064
การควบคุมหลังการเก็บเกี่ยว	0.913	0.927	0.115	0.089
การเน่าเสีย (การเสื่อมสภาพ)	0.847	0.886	0.077	0.130
เครื่องมือและอุปกรณ์ระบบหลังการเก็บเกี่ยว มันสำปะหลัง	0.918	0.882	0.148	0.074
เทคโนโลยีหลังการเก็บเกี่ยว	0.935	0.884	0.125	0.092

ตารางที่ ข-4ผลประสิทธิภาพการให้น้ำหนักคำสำคัญ 1-3 คำค้น 100 ข้อสอบถาม (ต่อ)

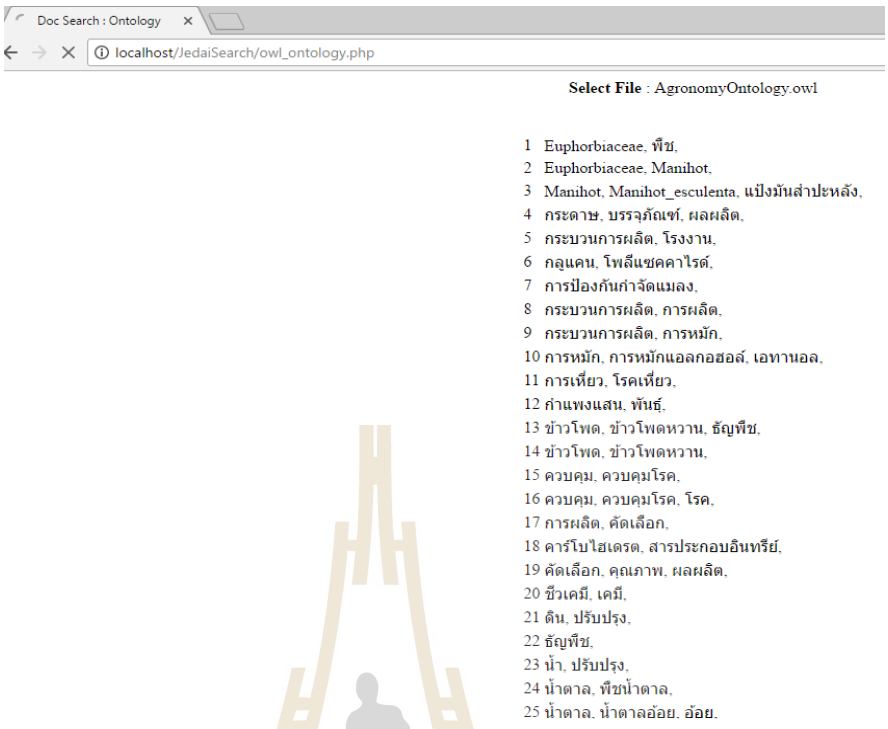
คำค้น	ความแม่นยำเฉลี่ย		ความผิดพลาด	
	IDF	*Corr*	IDF	*Corr*
เครื่องกะเทาะเปลือก	0.897	0.933	0.065	0.074
เครื่องมือหลังการเก็บเกี่ยว	0.911	0.945	0.075	0.113
เครื่องฉีดสารป้องกันกำจัดศัตรูพืช	0.937	0.879	0.111	0.144
เครื่องอบแห้ง	0.826	0.960	0.064	0.119
เครื่องคัดแยก	0.811	0.935	0.123	0.057
โรคจากปรสิต พืชไร่	0.832	0.927	0.100	0.107
โรคราสนิม อ้อย	0.882	0.921	0.100	0.127
โรคราน้ำค้าง ข้าวโพด	0.910	0.938	0.142	0.064
โรคที่เกิดจากเชื้อรา มันสำปะหลัง	0.852	0.921	0.098	0.071
โรคราแป้ง พืชไร่	0.815	0.903	0.055	0.075
โรคใบจุด พืชไร่	0.861	0.952	0.061	0.050
เชื้อราแป้ง สารป้องกันกำจัด	0.940	0.938	0.100	0.133
เชื้อรา มันสำปะหลัง	0.866	0.921	0.135	0.077
โรคใบขาว	0.927	0.902	0.086	0.107
โรคแอนแทรกโนส	0.830	0.905	0.099	0.114
โรคเน่าเปื่อย	0.847	0.889	0.115	0.068
โรคติดเชื้อพยาธิไส้เดือน	0.942	0.925	0.054	0.126
โรคราเขม่าดำ	0.857	0.945	0.147	0.089
ไส้เดือนฝอย การฉีด สารป้องกันกำจัดศัตรูพืช	0.857	0.953	0.081	0.107
เพลี้ยแป้งมัน การฉีด สารป้องกันกำจัดแมลง	0.848	0.889	0.093	0.062
ไรที่ทำให้เกิดความเสียหาย	0.858	0.961	0.082	0.090
แมลงหวี่ขาวสารป้องกันกำจัดแมลง	0.909	0.927	0.082	0.121
สารป้องกันกำจัดแมลงเพลี้ยแป้งมัน	0.889	0.947	0.053	0.101
สารป้องกันกำจัดศัตรูพืช	0.836	0.951	0.130	0.143

ตารางที่ ข-4ผลประสิทธิภาพการให้น้ำหนักคำสำคัญ 1-3 คำค้น 100 ข้อสอบถาม (ต่อ)

คำค้น	ความแม่นยำเฉลี่ย		ความผิดพลาด	
	IDF	*Corr*	IDF	*Corr*
การฉีดสารป้องกันกำจัดเพลี้ย	0.814	0.957	0.081	0.084
การฉีดสารป้องกันกำจัดวัชพืช	0.888	0.943	0.105	0.062
การฉีด	0.887	0.899	0.090	0.130
ผลิตภัณฑ์แปรรูปจากพืช	0.902	0.957	0.136	0.126
ตลาดสินค้าเกษตร	0.834	0.876	0.148	0.133
ตลาดในประเทศตลาดสินค้าเกษตร	0.899	0.949	0.090	0.122
การค้าระหว่างประเทศการส่งออก	0.861	0.937	0.150	0.128
ราคาขายปลีก	0.936	0.902	0.114	0.109
การส่งออกราคาขายปลีก	0.848	0.889	0.093	0.062
ตลาดโลกการส่งออกการค้าระหว่างประเทศ	0.858	0.961	0.082	0.090
ค่าเฉลี่ย 100 คำค้น	**0.860**	**0.906**	**0.104**	**0.097**

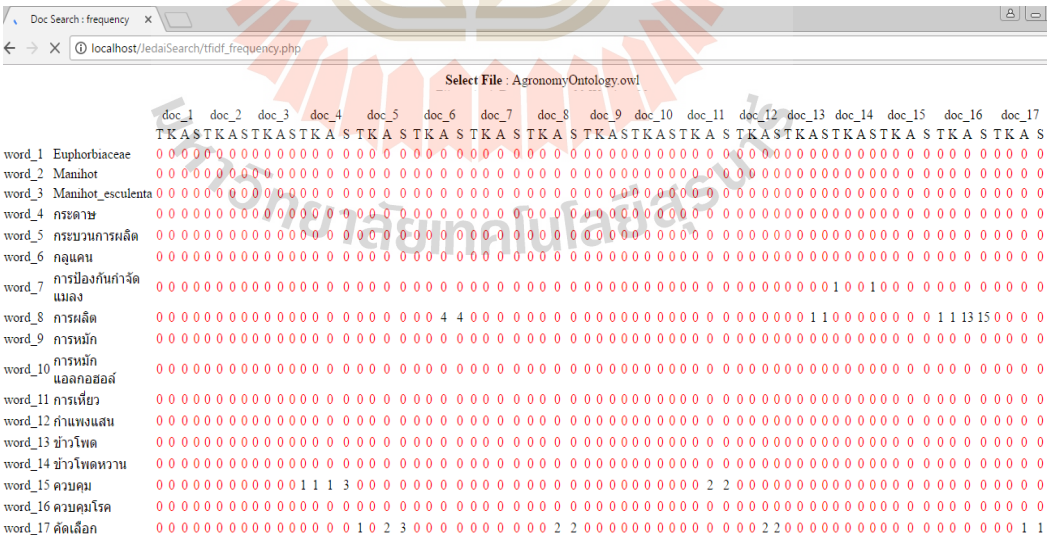
มหาวิทยาลัยเทคโนโลยีสุรนารี

● ตัวอย่างการจัดความรู้ให้เป็นระบบ : การอ่านไฟล์ OWL และคำนวณความถี่คำที่ปรากฏในเอกสาร

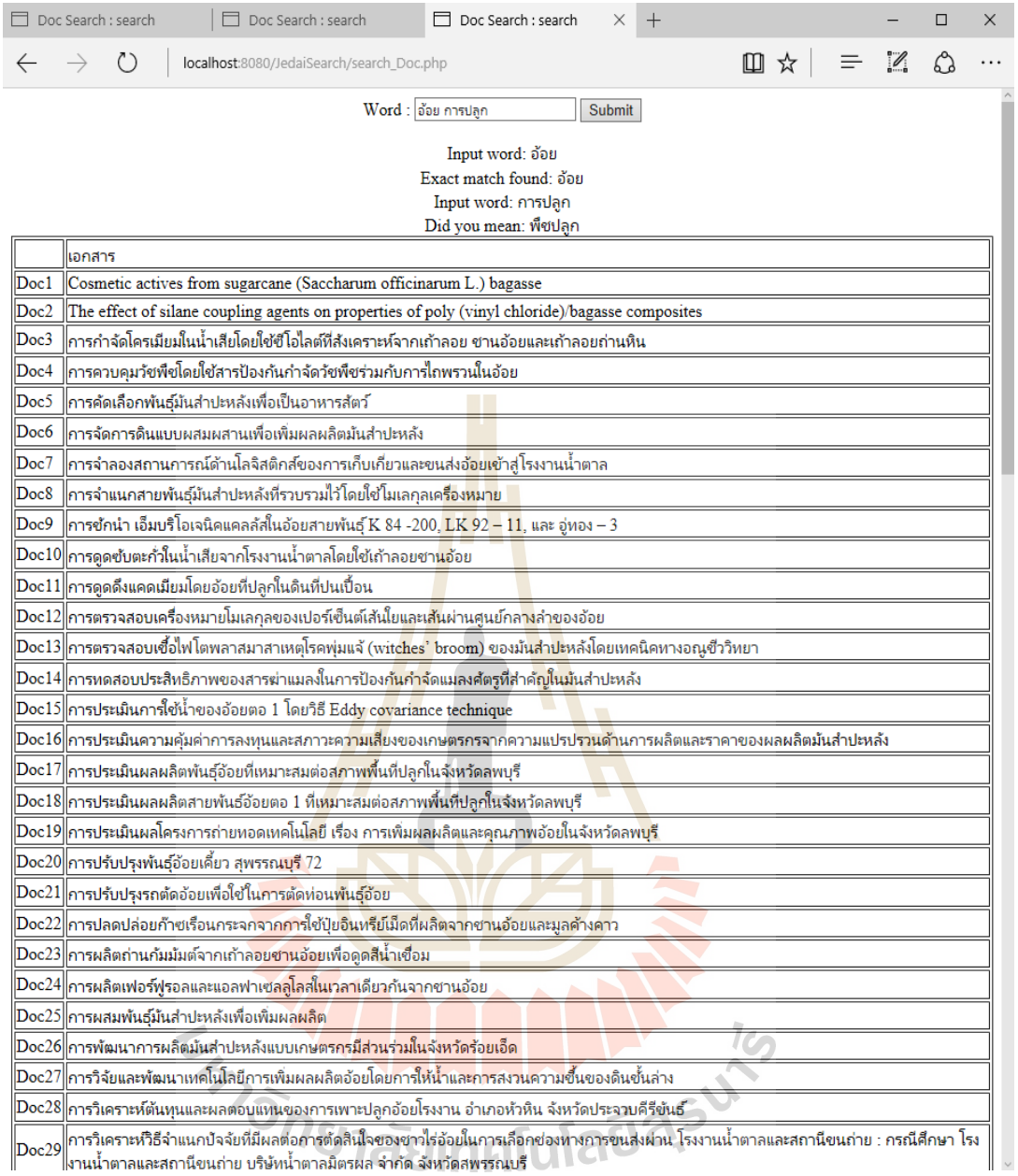


รูปที่ ข-1ตัวอย่างการอ่านไฟล์ OWL แสดงความสัมพันธ์ของคำในโครงสร้าง

- ตัวอย่างลำดับผลลัพธ์การสืบค้นความรู้จากการวัดความคล้ายคลึงด้วยคำค้น ***อ้อย การปลูก***

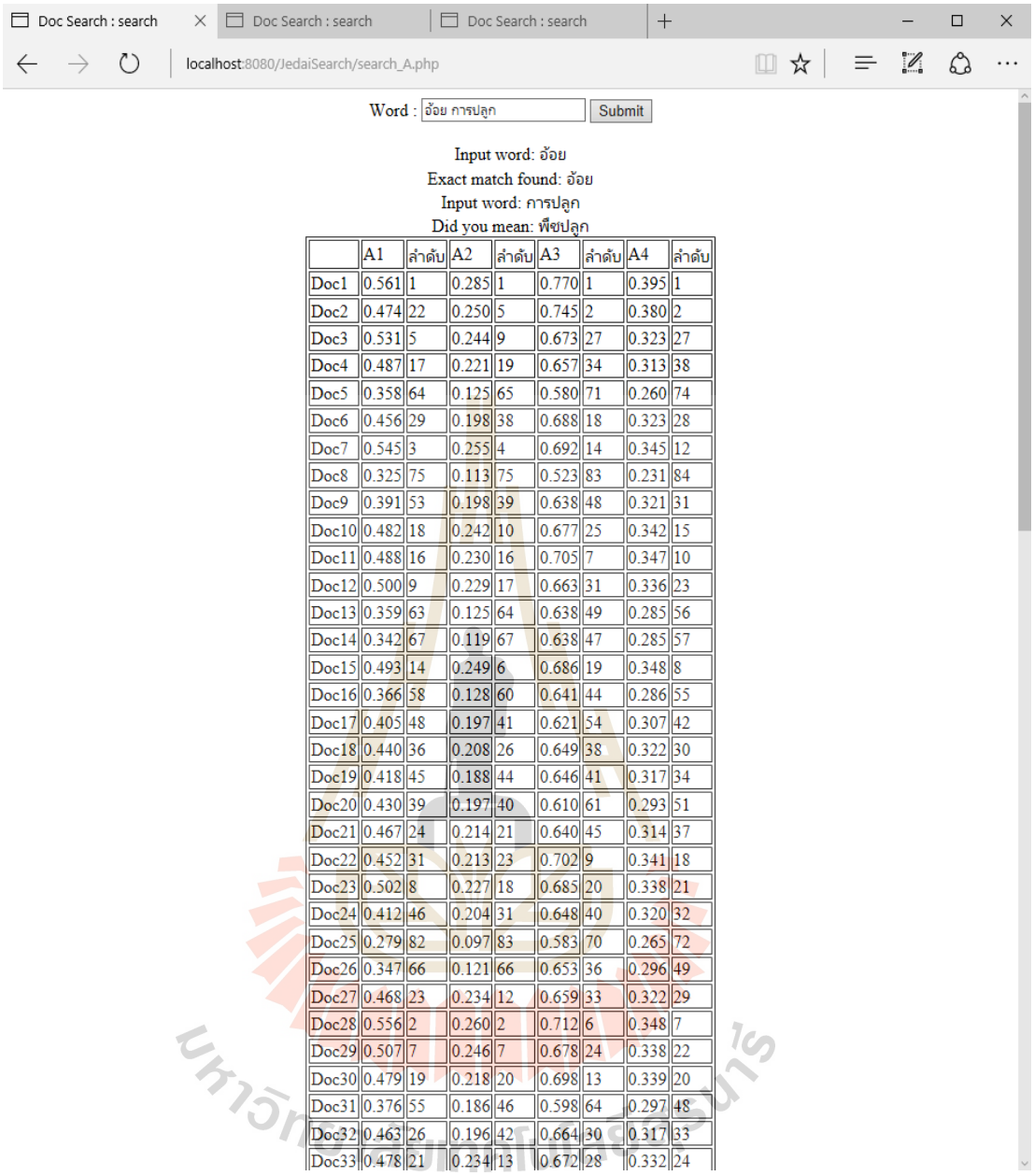


รูปที่ ข-2 ตัวอย่างคำนวณความถี่คำที่ปรากฏในเอกสาร



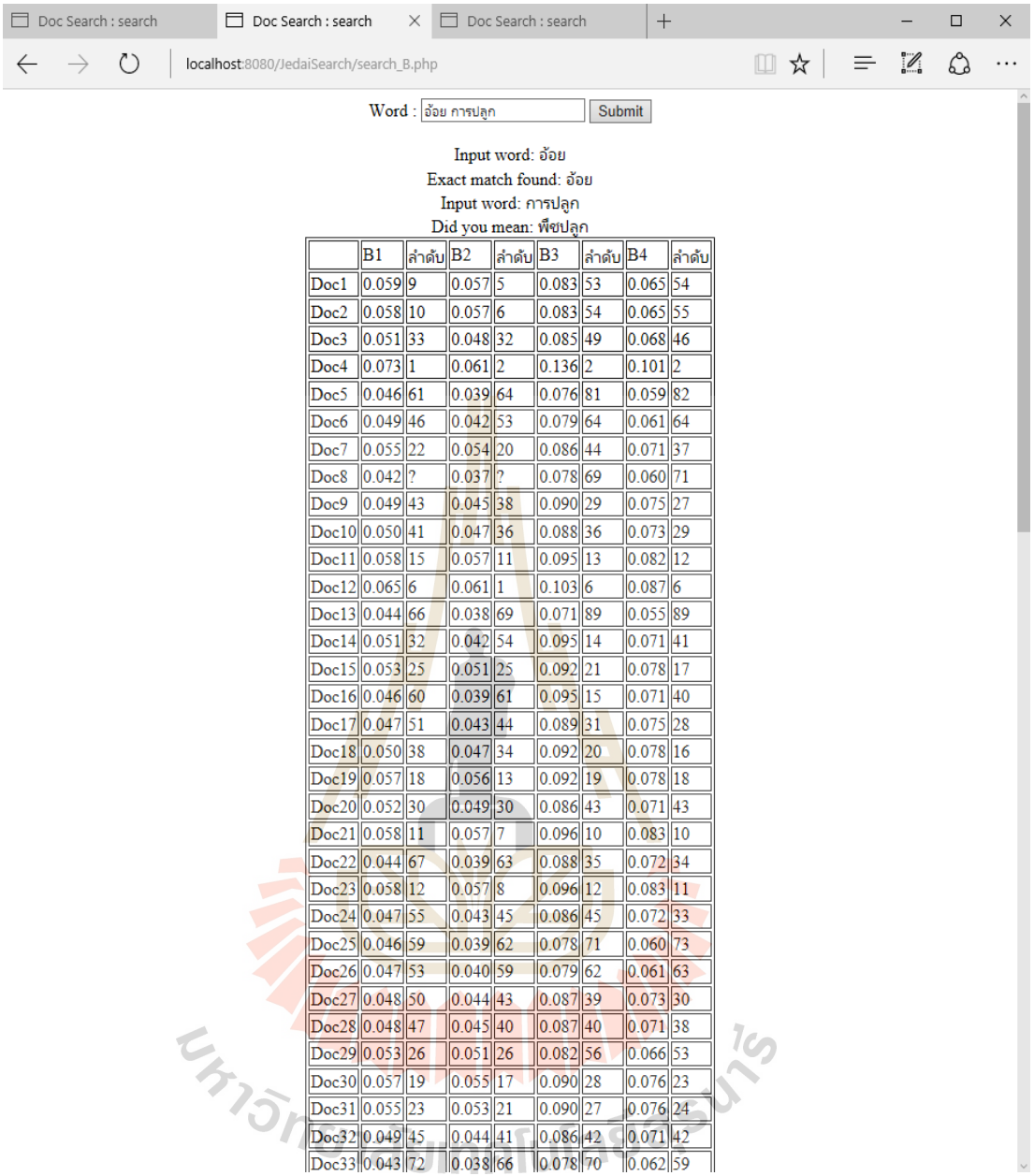
	เอกสาร
Doc1	Cosmetic actives from sugarcane (Saccharum officinarum L.) bagasse
Doc2	The effect of silane coupling agents on properties of poly (vinyl chloride)/bagasse composites
Doc3	การกำจัดโครเมียมในน้ำเสียโดยใช้ซีโอไลต์ที่สังเคราะห์จากเถ้าลอย ชานอ้อยและเถ้าลอยถ่านหิน
Doc4	การควบคุมวัชพืชโดยใช้สารป้องกันกำจัดวัชพืชร่วมกับการไถพรวนในอ้อย
Doc5	การคัดเลือกพันธุ์มันสำปะหลังเพื่อเป็นอาหารสัตว์
Doc6	การจัดการดินแบบผสมผสานเพื่อเพิ่มผลผลิตมันสำปะหลัง
Doc7	การจำลองสถานการณ์ด้านโลจิสติกส์ของการเก็บเกี่ยวและขนส่งอ้อยเข้าสู่โรงงานน้ำตาล
Doc8	การจำแนกสายพันธุ์มันสำปะหลังที่รวบรวมไว้โดยใช้โมเลกุลเครื่องหมาย
Doc9	การชักนำ เอ็มบริโอเจนิคแคลลัสในอ้อยสายพันธุ์ K 84 -200, LK 92 – 11, และ อู่ทอง – 3
Doc10	การดูดซับตะกั่วในน้ำเสียจากโรงงานน้ำตาลโดยใช้เถ้าลอยชานอ้อย
Doc11	การดูดดึงแคดเมียมโดยอ้อยที่ปลูกในดินที่ปนเปื้อน
Doc12	การตรวจสอบเครื่องหมายโมเลกุลของเปอร์เซ็นต์เส้นใยและเส้นผ่านศูนย์กลางลำของอ้อย
Doc13	การตรวจสอบเชื้อไฟโตพลาสมาสาเหตุโรคพุ่มแจ้ (witches' broom) ของมันสำปะหลังโดยเทคนิคทางอณูชีววิทยา
Doc14	การทดสอบประสิทธิภาพของสารฆ่าแมลงในการป้องกันกำจัดแมลงศัตรูที่สำคัญในมันสำปะหลัง
Doc15	การประเมินการใช้น้ำของอ้อยตอ 1 โดยวิธี Eddy covariance technique
Doc16	การประเมินความคุ้มค่าการลงทุนและสภาวะความเสี่ยงของเกษตรกรจากความแปรปรวนด้านการผลิตและราคาของผลผลิตมันสำปะหลัง
Doc17	การประเมินผลผลิตพันธุ์อ้อยที่เหมาะสมต่อสภาพพื้นที่ปลูกในจังหวัดลพบุรี
Doc18	การประเมินผลผลิตสายพันธ์อ้อยตอ 1 ที่เหมาะสมต่อสภาพพื้นที่ปลูกในจังหวัดลพบุรี
Doc19	การประเมินผลโครงการถ่ายทอดเทคโนโลยี เรื่อง การเพิ่มผลผลิตและคุณภาพอ้อยในจังหวัดลพบุรี
Doc20	การปรับปรุงพันธุ์อ้อยเคี้ยว สุพรรณบุรี 72
Doc21	การปรับปรุงรถตัดอ้อยเพื่อใช้ในการตัดท่อนพันธุ์อ้อย
Doc22	การปลดปล่อยก๊าซเรือนกระจกจากการใช้ปุ๋ยอินทรีย์เม็ดที่ผลิตจากชานอ้อยและมูลค้างคาว
Doc23	การผลิตถ่านกัมมัมต์จากเถ้าลอยชานอ้อยเพื่อดูดสีน้ำเชื่อม
Doc24	การผลิตเฟอร์ฟูรอลและแอลฟาเซลลูโลสในเวลาเดียวกันจากชานอ้อย
Doc25	การผสมพันธุ์มันสำปะหลังเพื่อเพิ่มผลผลิต
Doc26	การพัฒนาการผลิตมันสำปะหลังแบบเกษตรกรมีส่วนร่วมในจังหวัดร้อยเอ็ด
Doc27	การวิจัยและพัฒนาเทคโนโลยีการเพิ่มผลผลิตอ้อยโดยการให้น้ำและการสงวนความชื้นของดินชั้นล่าง
Doc28	การวิเคราะห์ต้นทุนและผลตอบแทนของการเพาะปลูกอ้อยโรงงาน อำเภอห้วหิน จังหวัดประจวบคีรีขันธ์
Doc29	การวิเคราะห์วิธีจำแนกปัจจัยที่มีผลต่อการตัดสินใจของชาวไร่อ้อยในการเลือกช่องทางการขนส่งผ่าน โรงงานน้ำตาลและสถานียนถ่าย : กรณีศึกษา โรงงานน้ำตาลและสถานียนถ่าย บริษัทน้ำตาลมิตรผล จำกัด จังหวัดสพรรณบุรี

รูปที่ ข-3ตัวอย่างรายชื่อเอกสารความรู้ด้านพืชไร่คำค้น อ้อย การปลูก



	A1	ลำดับ	A2	ลำดับ	A3	ลำดับ	A4	ลำดับ
Doc1	0.561	1	0.285	1	0.770	1	0.395	1
Doc2	0.474	22	0.250	5	0.745	2	0.380	2
Doc3	0.531	5	0.244	9	0.673	27	0.323	27
Doc4	0.487	17	0.221	19	0.657	34	0.313	38
Doc5	0.358	64	0.125	65	0.580	71	0.260	74
Doc6	0.456	29	0.198	38	0.688	18	0.323	28
Doc7	0.545	3	0.255	4	0.692	14	0.345	12
Doc8	0.325	75	0.113	75	0.523	83	0.231	84
Doc9	0.391	53	0.198	39	0.638	48	0.321	31
Doc10	0.482	18	0.242	10	0.677	25	0.342	15
Doc11	0.488	16	0.230	16	0.705	7	0.347	10
Doc12	0.500	9	0.229	17	0.663	31	0.336	23
Doc13	0.359	63	0.125	64	0.638	49	0.285	56
Doc14	0.342	67	0.119	67	0.638	47	0.285	57
Doc15	0.493	14	0.249	6	0.686	19	0.348	8
Doc16	0.366	58	0.128	60	0.641	44	0.286	55
Doc17	0.405	48	0.197	41	0.621	54	0.307	42
Doc18	0.440	36	0.208	26	0.649	38	0.322	30
Doc19	0.418	45	0.188	44	0.646	41	0.317	34
Doc20	0.430	39	0.197	40	0.610	61	0.293	51
Doc21	0.467	24	0.214	21	0.640	45	0.314	37
Doc22	0.452	31	0.213	23	0.702	9	0.341	18
Doc23	0.502	8	0.227	18	0.685	20	0.338	21
Doc24	0.412	46	0.204	31	0.648	40	0.320	32
Doc25	0.279	82	0.097	83	0.583	70	0.265	72
Doc26	0.347	66	0.121	66	0.653	36	0.296	49
Doc27	0.468	23	0.234	12	0.659	33	0.322	29
Doc28	0.556	2	0.260	2	0.712	6	0.348	7
Doc29	0.507	7	0.246	7	0.678	24	0.338	22
Doc30	0.479	19	0.218	20	0.698	13	0.339	20
Doc31	0.376	55	0.186	46	0.598	64	0.297	48
Doc32	0.463	26	0.196	42	0.664	30	0.317	33
Doc33	0.478	21	0.234	13	0.672	28	0.332	24

รูปที่ ข-4ตัวอย่างลำดับผลลัพธ์ความคล้ายคลึงเชิงความหมายรูปแบบการทดลอง A1, A2, A3, A4



	B1	ลำดับ	B2	ลำดับ	B3	ลำดับ	B4	ลำดับ
Doc1	0.059	9	0.057	5	0.083	53	0.065	54
Doc2	0.058	10	0.057	6	0.083	54	0.065	55
Doc3	0.051	33	0.048	32	0.085	49	0.068	46
Doc4	0.073	1	0.061	2	0.136	2	0.101	2
Doc5	0.046	61	0.039	64	0.076	81	0.059	82
Doc6	0.049	46	0.042	53	0.079	64	0.061	64
Doc7	0.055	22	0.054	20	0.086	44	0.071	37
Doc8	0.042	?	0.037	?	0.078	69	0.060	71
Doc9	0.049	43	0.045	38	0.090	29	0.075	27
Doc10	0.050	41	0.047	36	0.088	36	0.073	29
Doc11	0.058	15	0.057	11	0.095	13	0.082	12
Doc12	0.065	6	0.061	1	0.103	6	0.087	6
Doc13	0.044	66	0.038	69	0.071	89	0.055	89
Doc14	0.051	32	0.042	54	0.095	14	0.071	41
Doc15	0.053	25	0.051	25	0.092	21	0.078	17
Doc16	0.046	60	0.039	61	0.095	15	0.071	40
Doc17	0.047	51	0.043	44	0.089	31	0.075	28
Doc18	0.050	38	0.047	34	0.092	20	0.078	16
Doc19	0.057	18	0.056	13	0.092	19	0.078	18
Doc20	0.052	30	0.049	30	0.086	43	0.071	43
Doc21	0.058	11	0.057	7	0.096	10	0.083	10
Doc22	0.044	67	0.039	63	0.088	35	0.072	34
Doc23	0.058	12	0.057	8	0.096	12	0.083	11
Doc24	0.047	55	0.043	45	0.086	45	0.072	33
Doc25	0.046	59	0.039	62	0.078	71	0.060	73
Doc26	0.047	53	0.040	59	0.079	62	0.061	63
Doc27	0.048	50	0.044	43	0.087	39	0.073	30
Doc28	0.048	47	0.045	40	0.087	40	0.071	38
Doc29	0.053	26	0.051	26	0.082	56	0.066	53
Doc30	0.057	19	0.055	17	0.090	28	0.076	23
Doc31	0.055	23	0.053	21	0.090	27	0.076	24
Doc32	0.049	45	0.044	41	0.086	42	0.071	42
Doc33	0.043	72	0.038	66	0.078	70	0.062	59

รูปที่ ข-5ตัวอย่างลำดับผลลัพธ์ความคล้ายคลึงเชิงความหมายรูปแบบการทดลอง B1, B2, B3, B4

ประวัติผู้เขียน

นายเจษฎา สิงห์ทองชัย เกิดเมื่อวันที่ 28ธันวาคม 2519 ที่อำเภอสังคม จังหวัดหนองคาย สำเร็จการศึกษาในระดับมัธยมตอนต้นที่โรงเรียนสังคมวิทยา และระดับประกาศนียบัตรวิชาชีพ (ปวช.) ที่วิทยาลัยเกษตรกรรมอุดรธานี โดยปีการศึกษา 2538ได้เข้าศึกษาต่อระดับประกาศนียบัตรวิชาชีพชั้นสูง (ปวส.) สาขาวิชาคอมพิวเตอร์ธุรกิจที่สถาบันเทคโนโลยีราชมงคลน่าน และในปีการศึกษา 2540ได้เข้าศึกษาต่อระดับปริญญาตรีในสาขาวิชาธุรกิจศึกษาคอมพิวเตอร์ ที่สถาบันเทคโนโลยีราชมงคล (คลองหก) ปทุมธานี และสำเร็จการศึกษาในปีการศึกษา 2542

เริ่มปฏิบัติงานในตำแหน่งอาจารย์ สาขาวิชาคอมพิวเตอร์ธุรกิจ วิทยาลัยเกษตรและเทคโนโลยีพิจิตร จังหวัดพิจิตร และตำแหน่งอาจารย์สาขาวิชาคอมพิวเตอร์ธุรกิจวิทยาลัยการอาชีพเชียงคำ จังหวัดพะเยาตามลำดับ

ปี พ.ศ. 2557 สำเร็จการศึกษาระดับปริญญาโทสาขาวิชาเทคโนโลยีอินเทอร์เน็ตและสารสนเทศ มหาวิทยาลัยนเรศวร ปฏิบัติงานตำแหน่งอาจารย์ประจำสาขาวิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสังคม มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตกาฬสินธุ์ (มหาวิทยาลัยกาฬสินธุ์)จนถึงปัจจุบัน

ปี พ.ศ.2554 ได้ศึกษาต่อระดับปริญญาเอกสาขาวิชาเทคโนโลยีสารสนเทศ สำนักวิชาเทคโนโลยีสังคม มหาวิทยาลัยเทคโนโลยีสุรนารี ทั้งนี้ผู้วิจัยได้มีความสนใจด้านการจัดการความรู้ให้เป็นระบบ การออกแบบและพัฒนาการสืบค้นเชิงความหมาย ตลอดจนเป็นที่ปรึกษาด้านการออกแบบและพัฒนาระบบสารสนเทศให้หน่วยงานทั้งภาครัฐ และเอกชน

มหาวิทยาลัยเทคโนโลยีสุรนารี