# ระบบการค้นคืนเชิงความหมายจากข้อมูลบรรณานุกรมโดเมน Information System

# Semantic Search for Information System Domain Bibliographic Data

พยุง มีสัจ[1] วาทินี นุ้ยเพียร[2] และ ผุสดี บุญรอด[3]

ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

**ABSTRACT –** Searching through digital information by using keywords usually returns incorrect results and does not reflect user requirements. The researchers have developed semantic search or ontology that increases efficiency of information retrieval. However, one of the problems of semantic search is that development depends on the expert knowledge. To solve the problem above, this research proposes sub class creation via semi-automatic techniques and finding relation of class by association rules. The proposed method is composed of 4 steps: 1) feature selection using Chi-Square and Support Vector Machine Radial basis function (SVMR) in order to reduce phrases and send statistics value into a subclass to aid the semi-automatic ontology, 2) semantic terms mapping using association rule methods, 3) development of Semantic Search for Information System Domain Bibliographic Data, and 4) evaluation of the developed information retrieval system. The result of the proposed method produced precision at 97.07%.

**KEY WORDS --** Semantic Information Retrieval, Ontology, Text Mining, Feature Selection, Association Rule, Bibliographic Data

**บทคัดย่อ --** ปัญหาจากการค้นคืนโดยใช้คำสำคัญทำให้ผู้ใช้ได้ข้อมูลไม่ตรงตามความต้องการ ปัจจุบันมีนักวิจัยได้พัฒนาระบบการค้นคืนเชิงความหมายหรือออนโทโลยีเพื่อช่วยเพิ่มประสิทธิภาพในการค้นคืน แต่ปัญหาหนึ่งของการค้นคืนเชิงความหมาย คือ การสร้างระบบการค้นคืนงานส่วนใหญ่ขึ้นอยู่กับผู้เชี่ยวชาญ ดังนั้นงานวิจัยนี้จึงได้นำเสนอการสร้างคลาสย่อยแบบกึ่งอัตโนมัติและการหาความสัมพันธ์ของคลาสด้วยกฎความสัมพันธ์ โดยมีขั้นตอนดังนี้ 1) ทำการคัดเลือกคุณลักษณะด้วยการใช้เทคนิคไคสแควร์และจำแนกข้อความแบบซัพพอร์ตเวกเตอร์แมชชีนเคอร์เนลฟังก์ชันแบบเรเดียลเบสิสฟังก์ชัน และนำค่าสถิติของวลีไปสร้างคลาสย่อยในออนโทโลยีแบบกึ่งอัตโมมัติ 2) จัดเทียบคำศัพท์เชิงความหมายเข้าคลาสโดยใช้กฎความสัมพันธ์ 3) พัฒนาระบบค้นคืนข้อมูลเชิงความหมายจากข้อมูลเชิงบรรณานุกรมโดเมน Information System และ 4) ประเมินระบบการค้นคืนข้อมูลเชิงความหมายจากข้อมูลเชิงบรรณานุกรมให้ค่าความแม่นยำถึง 97.07%

**คำสำคัญ --** การค้นคืนข้อมูลเชิงความหมาย ออนโทโลยี การทำเหมืองข้อความ การคัดเลือกคุณลักษณะ กฎความสัมพันธ์ ข้อมูลบรรณานุกรม

## 1. บทนำ

ปัจจุบันมีการจัดเก็บข้อมูลเกี่ยวกับความรู้ในด้านต่าง ๆ บนอินเทอร์เน็ตซึ่งเชื่อมโยงข้อมูลมากมายเป็นฐานข้อมูลขนาดใหญ่ ในการค้นหาข้อมูลที่ต้องการผ่านเซิร์ชเอ็นจิน (Search Engine) สามารถกระทำได้โดยการสืบค้นจากการใช้คำสำคัญซึ่งคอมพิวเตอร์จะส่งคำค้นไปเปรียบเทียบกับคำที่มีอยู่ใน

เอกสารที่ถูกจัดเก็บในฐานข้อมูลต่าง ๆ บนอินเทอร์เน็ต ผลการสืบค้นเป็นการคืนเอกสารที่มีคำสำคัญตามที่ต้องการ โดยสร้างเป็นไฮเปอร์ลิงค์ไปยังเอกสารต่าง ๆ ซึ่งเซิร์ชเอ็นจินทั่วไปไม่สามารถสรุปความ ประมวลความหมาย หรือแสดงความสัมพันธ์ของคำนั้น ๆ ได้อย่างตรงประเด็น [1], [2] ทำให้ไม่สามารถค้นหาคำที่มีความหมายใกล้เคียงกัน และทำให้ผลการค้นคืนไม่ตรงตามวัตถุประสงค์ของผู้ใช้งาน การสร้างการค้นคืนที่ดีควรพิจารณาเกี่ยวกับความหมายของคำที่ใกล้เคียงกัน และลำดับของคำ

มีนักวิจัยจำนวนมากให้ความสำคัญในการใช้ความหมายของคำที่ใกล้เคียงกันในการสืบค้นเพื่อทำให้ผู้ค้นหาสามารถกำหนดรายละเอียดได้มากขึ้นในการค้นหาเอกสาร อีกทั้งใช้หลักการค้นคืนเชิงความหมาย ทำให้ได้ผลลัพธ์ แม่นยำ และสอดคล้องต่อความต้องการของผู้ใช้งานมากกว่าการสืบค้นแบบคำสำคัญสิ่งที่นักวิจัยนำมาใช้คือการใช้ออนโทโลยีเข้ามาช่วย ซึ่ง เพ็ญพรรณ [1] ได้นำมาใช้ในการสืบค้นและวิเคราะห์ข้อมูลทางด้านชีววิทยาผลลัพธ์จากงานวิจัยคือระบบสามารถสืบค้นข้อมูลที่ซับซ้อนและใช้งานง่ายเหมาะสำหรับผู้ใช้งานส่วน สินีนาถ [3] ใช้กับการค้นหาเอกสารงานวิจัยเชิงความหมายซึ่งสามารถสืบค้นเอกสารได้จากการกำหนดหัวข้อที่จะใช้ในการสืบค้น และ สิทธา [4] สร้างระบบช่วยเลือกซื้อโทรศัพท์มือถือโดยใช้เว็บเชิงความหมายขึ้นมาช่วยผู้ใช้ เนื่องจากการใช้คำสำคัญในการสืบค้นมีเอกสารที่ไม่เกี่ยวข้องแสดงผลบนหน้าเว็บทำให้ผู้ใช้งานต้องใช้เวลาในการอ่านเพื่อศึกษาข้อมูลนาน จากผลการวิจัยการนำออนโทโลยีเข้ามาใช้งานจึงช่วยให้จำนวนผลลัพธ์น้อยลงทำให้การค้นคืนตรงประเด็นมากขึ้น

จากข้อดีของออนโทโลยีที่สามารถประยุกต์ใช้ในงานด้านอื่น ๆ ร่วมทั้งการจัดการองค์ความรู้ ที่สามารถสร้างความสัมพันธ์ของคำหรือทฤษฎีที่เกี่ยวข้อง จึงนำมาประยุกต์ใช้กับระบบการค้นคืนเชิงความหมายจากงานวิจัยเพื่อให้สามารถสรุป ทฤษฎีที่เกี่ยวข้อง คำสำคัญที่ต้องศึกษาเพิ่มเติม ขั้นตอนการดำเนินการวิจัย ดังนั้นผู้วิจัยจึงพัฒนาระบบการค้นคืนเชิงความหมายจากข้อมูลบรรณานุกรมโดเมน Information System เพื่อช่วยให้นักวิจัยให้สามารถพัฒนาค้นคืนข้อมูลที่มีความสัมพันธ์กับสิ่งที่ผู้วิจัยสนใจได้ง่ายขึ้น

## 2. ทฤษฎีและวรรณกรรมที่เกี่ยวข้อง

การสร้างระบบค้นคืนเชิงความหมาย แบ่งเป็น 3 ส่วนคือ 1) การทำเหมืองข้อความ 2) การหากฎความสัมพันธ์ และ 3) ส่วนประกอบของออนโทโลยี ซึ่งภาษาที่ใช้ในการอธิบายข้อมูลเชิงความหมาย เช่น ภาษา XML และ XML Schema [5] RDF เป็นภาษาที่ใช้กำหนดโครงสร้างของข้อมูลซึ่งสามารถอธิบายความสัมพันธ์ขององค์ประกอบต่าง ๆ ที่บรรยายในเอกสารในรูปแบบของประโยค (Sentence) ซึ่งมีโครงสร้างประกอบด้วยซับเจกต์ (Subject) รีเลชั่น (Relation) และ อ็อบเจกต์ (Object) [6] และ RDF Schema [7] ภาษา OWL (Web Ontology Language) [8], [9]

### 2.1 การทำเหมืองข้อความ

การทำเหมืองข้อความเริ่มจากการตัดคำเนื่องจากการค้นคืนส่วนใหญ่ใช้หลักการค้นคืนจากคำสำคัญ ซึ่งใช้วิธีการเปรียบเทียบคำค้นกับเอกสาร ปัญหาหนึ่งที่พบ คือ ข้อมูลมีมิติมาก [10]-[12] จึงต้องใช้ขบวนการในการคัดเลือกคุณลักษณะ (Feature Selection) และทำการจำแนกข้อมูล

2.1.1 ตัดวลีใช้หลักการตัดปลาย เช่น s หรือ es เนื่องจากถ้าตัดตามหลักการของภาษาธรรมชาติอาจทำให้ความหมายของคำเปลี่ยน เช่น Data Mining ตัดแบบ NLP จะได้ผลลัพธ์ Data min

2.1.2 การคัดเลือกคุณลักษณะ (Feature Selection) คือ การนำเอกสารจากระบบต่าง ๆ เช่น เอกสารข่าวจากเว็บไซต์ บทคัดย่อมาแปลงเพื่อให้เอกสารอยู่ในรูปแบบเดียวกัน มีองค์ประกอบของเอกสารที่เหมือนกัน เพื่อใช้แทนคุณลักษณะของเอกสารโดยใช้คำเดี่ยว พยางค์ วลี กลุ่มของคำ ประโยค โดยส่วนใหญ่ใช้การคำนวณหาค่าน้ำหนักของคุณลักษณะ อาจใช้เทคนิคการวิเคราะห์ทางภาษาเข้ามาช่วย (Language Analysis) เพื่อตัดคำที่ไม่จำเป็นออก หรือตัดคำตามสถิติของคลังประโยค เพื่อลดมิติของขนาดเอกสารเทคนิคที่เลือกใช้คือ

ก) อินฟอร์เมชันเกน (IG) คือ การประเมินค่าเพื่อใช้ในการแบ่งข้อมูลด้วยการคำนวณค่า Gain สำหรับแต่ละมิติข้อมูล ถ้ามิติข้อมูลใดมีค่า Gain สูงสุด จะถูกเลือกให้เป็นกลุ่มย่อยที่มีอำนาจจำแนก ดังสมการที่ (1) แสดงการคำนวณค่า Entropy และคำนวณค่า Gain ดังสมการที่ (2) [13]

$$Entropy\,(p) = -\sum_{i=0}^{c-1} p(j\,|\,t)\log_2 p\,(j\,|\,t) \qquad (1)$$

โดยที่ $\sum i$ คือ ผลรวมของความน่าจะเป็นของค่า $j$ ที่เกิดในคลาส $t$

$(j \mid t)$ คือ ค่าความถี่ที่มีความสัมพันธ์ของกลุ่ม $j$ กับโหนด $t$

$$Gain = Entropy(p) - \left( \sum_{i=1}^{k} \frac{n_i}{n} Entropy(i) \right) \quad (2)$$

โดยที่ *Entropy (p)* คือ ค่า Entropy ของตัว Root

$\sum_{i=1}^{k} \frac{n_i}{n} Entropy\,(i)$ คือ ค่า *Entropy* ในแต่ละโหนดย่อย

ข) เกนเรโช (GR) เป็นการประเมินความน่าเชื่อถือของมิติข้อมูลโดยการวัด Gain Ratio ในแต่ละคลาสการคำนวณ GR โดยใช้ค่า *SplitINFO* ในสมการที่ (3) และการคำนวณค่าการวัด Gain Ratio [14] ดังสมการที่ (4)

$$SplitINFO = -\sum_{i=1}^{k} \frac{n_i}{n} \log_2 \frac{n_i}{n} \quad (3)$$

. $$GainRatio = \frac{\Delta INFO}{SplitINFO} \quad (4)$$

ค) ไคสแควร์ ($\chi^2$) คือ การประเมินค่าของแอททริบิวต์โดยคำนวณค่า Chi-Square ทางสถิติ [15] ดังสมการที่ (5)

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \quad (5)$$

2.1.3 การจำแนกข้อความการเรียนรู้แบบมีผลเฉลย (Supervised Learning) มีขั้นตอนในการจำแนก 2 ขั้นตอน คือ การเรียนรู้เพื่อสร้างเอกสารต้นแบบและการแยกหมวดหมู่ของเอกสารที่สนใจ โดยการตรวจหาความคล้ายกับกลุ่มเอกสารต้นแบบ

ก) เบย์ (Bayes: BN) คือ วิธีการเรียนรู้ที่ใช้หลักการของความน่าจะเป็น ซึ่งมีพื้นฐานมาจากทฤษฎีของเบย์ (Bayes's theorem) [14] เช่นกำหนดให้การเกิดของเหตุการณ์ต่าง ๆ ที่ใช้ในการจำแนกกลุ่มนั้นเป็นอิสระต่อกัน แนวคิดทฤษฎีของเบย์สามารถทำนายเหตุการณ์ที่พิจารณาได้จากการเกิดของเหตุการณ์ต่าง ๆ ได้ดังสมการที่ (6)

$$\Pr(class = c \mid X) = \Pr(X \mid class = c) \Pr(class = c) \mid \Pr(X) \quad (6)$$

ข้อสังเกต $\Pr(X)$ ไม่เปลี่ยนสำหรับค่าคลาส $c$ ที่เปลี่ยนไป

$\Pr(class = c) \approx$ ความถี่สัมพัทธ์ของตัวอย่างในคลาส $c$

$\Pr(class = c \mid X)$ สูงสุดก็ต่อเมื่อ

$\Pr(X \mid class = c) = \Pr(class = c)$ สูงสุด

ข) นาอีฟเบย์ (NB) คือ การใช้วิธีการของเบย์พร้อมสมมติฐานของการเป็นอิสระต่อกันของตัวแปรอิสระทุกตัว [16] เน้นการตั้งสมมติฐานเกี่ยวกับการเกิดขึ้นของคำในเอกสารด้วยโมเดลการจำแนกกลุ่มที่ใช้หลักความน่าจะเป็นซึ่งอยู่บนพื้นฐานของ Bayes's Theorem ดังสมการที่ (7)

$$\Pr(x_1,..x_k \mid class = c) = \Pr(x_1 \mid class = c)....\Pr(x_k \mid class = c) \quad (7)$$

ถ้าลักษณะประจำ $i$ เป็น $categorical: \Pr(x_i \mid class = c)$ ประมาณด้วยความถี่สัมพัทธ์ของตัวอย่างที่มีค่า $x_i$ ในคลาส $c$

ถ้าลักษณะประจำ $i$ เป็น $continuous: \Pr(x_i \mid class = c)$ ประมาณด้วยฟังก์ชันความหนาแน่น *Gaussian*

ค) เคเนียเรสต์เนเบอร์ (KNN) คือ การตัดสินใจของคลาสสำหรับแทนเงื่อนไขหรือกรณีใหม่ โดยการตรวจสอบจำนวนบางจำนวน หรือเงื่อนไขที่เหมือนกันหรือใกล้เคียงกันมากที่สุด ใช้เวลาในการคำนวณสูงเพราะการคำนวณเป็นการเพิ่มขึ้นแบบแฟคทอเรียลตามจุดทั้งหมดขณะที่ Decision Tree หรือโครงข่ายประสาทเทียม ประมวลผลเพื่อสร้างเงื่อนไขได้เร็วกว่าเพราะเคเนียเรสต์เนเบอร์ มีการคำนวณทุกครั้งที่มีกรณีใหม่

ดังนั้นเพื่อความรวดเร็วข้อมูลทั้งหมดที่ใช้บ่อยต้องถูกเก็บไว้ในหน่วยความจำ ชื่อว่า Memory-Based Reasoning [18]

ง) ซัพพอร์ตเวกเตอร์แมชชีน (SVM) นำเสนอโดย [19] ใช้เพื่อหาระนาบการตัดสินใจในการแบ่งข้อมูลออกเป็นสองส่วน โดยใช้สมการเส้นตรงเพื่อแบ่งเขตข้อมูล 2 กลุ่มออกจากกัน มุ่งหาผลลัพธ์ที่ดีที่สุดของการเรียนรู้ (Discriminative Training) บนการเรียนรู้จากสถิติของข้อมูล ซึ่งทำงานโดยการหาค่าระยะขอบที่มากที่สุด (Maximum Margin) ของระนาบตัดสินใจ (Decision Hyper Plane) ในการแบ่งแยกกลุ่มข้อมูลที่ใช้ฝึกฝนออกจากกัน เคอร์เนลที่พบได้บ่อย คือ โพลิโนเมียล (Polynomial) เป็นการคำนวณหาเส้นแบ่งโดยใช้สมการ เชิงเส้นที่มี Degree มากกว่าสองและเรเดียลเบสิสฟังก์ชัน (Radial basis Function) โดยมีค่า C เป็นค่าตัวแปรที่ปรับความสมดุลระหว่างการให้ความสำคัญของระยะแยกแยะสูงสุด หรือให้ความสำคัญกับค่าความผิดพลาดที่ต้องการให้ต่ำที่สุด โดยปกติค่า C จะกำหนดให้มีค่ามาก ส่วนค่า Gamma มีค่าน้อย ซึ่งสมการทั้งหมดปรากฏในหนังสือ [20] ส่วนค่าเคอร์เนลดังสมการที่ (8) และ (9)

ได้แก่ Polynomial kernel: (SVMP)

$$(\gamma \times \mu' \times v + coef0)\text{^}\deg ree \qquad (8)$$

Radial basis function kernel : (SVMR)

$$\exp(-\gamma \times |\mu - v|\text{^}2) \qquad (9)$$

จากทฤษฎีการทำเหมืองข้อความ (Text Mining) เป็นเทคนิคเพื่อค้นหารูปแบบ (Pattern) จากข้อความจำนวนมหาศาลโดยอัตโนมัติ ซึ่งสามารถคัดเลือกคุณลักษณะที่ดีที่นำมาใช้เป็นตัวแทนเอกสารโดยใช้วิธีการหาค่าสถิติ หลังจากที่ได้คำหรือวลีที่พิจารณาจากค่าสถิติที่เกิดขึ้นตาม การเรียงลำดับจากมากไปน้อย ทำให้นักวิจัยสามารถนำวลีเหล่านี้มาสร้างคลาสย่อยเพื่อใช้ในระบบการค้นคืนเชิงความหมาย และหาค่าความสัมพันธ์ที่เกิดขึ้นของข้อมูลด้วยการหาค่ากฎความสัมพันธ์

**2.2 การค้นหากฎความสัมพันธ์ (Association Rules)**

การค้นหากฎความสัมพันธ์ เป็นเทคนิคที่ใช้หากฎความสัมพันธ์ทั้งหมดของข้อมูลที่มีคุณสมบัติสอดคล้องตามค่า Minimum Support และ Minimum Confidence ซึ่งเป็นการหาความสัมพันธ์ของข้อมูลสองชุดหรือมากกว่าสองชุดขึ้นไปภายใต้กลุ่มข้อมูลขนาดใหญ่ เพื่อนำข้อความที่มีความสัมพันธ์กันมาใช้ในการจับคู่เชิงความหมายของระบบการค้นคืนเชิงความหมาย

รูปแบบการหาค่ากฎความสัมพันธ์คือ $A \rightarrow B$ (If A..Then..B)

โดยที่ A: เป็นเงื่อนไข และ B: เป็นผลลัพธ์ที่เกิดขึ้น

การประเมินค่าของกฎจะใช้ค่าสนับสนุน (Support) และค่าความเชื่อมั่น (Confidence) โดยที่

ค่าสนับสนุน คือ ร้อยละของข้อมูลที่มีเงื่อนไขและผลลัพธ์สอดคล้องกฎต่อจำนวนข้อมูลทั้งหมด สามารถเขียนเป็นสมการดังนี้

$$Support\ (A, B) = \frac{|A \cap B|}{|D|} \qquad (10)$$

โดยที่ $A$ หมายถึง เหตุการณ์ที่ใช้เป็นเงื่อนไขในการหาผลลัพธ์

$B$ หมายถึงเหตุการณ์ที่เป็นผลลัพธ์

$D$ หมายถึงจำนวน Transaction ทั้งหมด

$|A \cap B|$ หมายถึงเหตุการณ์ที่ประกอบด้วยเหตุการณ์ A และ B

ค่าความเชื่อมั่น คือ ร้อยละของข้อมูลที่มีเงื่อนไขและผลลัพธ์สอดคล้องตามกฎต่อจำนวนข้อมูลทั้งหมดที่เป็นเงื่อนไข สามารถเขียนเป็นสมการดังนี้

$$Confidence\ (A,B) = \frac{|A \cap B|}{|A|} \qquad (11)$$

โดยที่ หมายถึงเหตุการณ์ที่ประกอบด้วยเหตุการณ์ A อย่างเดียวในการเลือก

### 2.3 ส่วนประกอบหลักของออนโทโลยี

ออนโทโลยีเป็นวิธีการแทนความรู้ที่ใช้โหนดเพื่อแสดงตัวแทน(Concept) พร๊อพเพอร์ตี้ (Property) และเงื่อนไข (Restriction) สำหรับการอธิบายข้อมูลเชิงความหมาย ซึ่งความสัมพันธ์ที่นิยมใช้กันมาก เช่นความสัมพันธ์แบบจัดเป็น (Is-a) ความสัมพันธ์แบบจัดเป็นส่วนประกอบ (คุณสมบัติ) (Part-of) และความสัมพันธ์จัดเป็นคุณลักษณะ( คุณสมบัติ) (Attribute-of)

#### 2.3.1 การกำหนดคลาส (Class) คือการอธิบายคลาส (Class)

คอนเซ็ปต์ (Concept) เป็นสิ่งที่แสดงถึงสิ่งที่เราสนใจซึ่งออนโทโลยีมีคลาสต้นกำเนิดคือ owl: class โดยกำหนดให้ owl: class เป็นคลาสใหญ่ที่สามารถครอบคลุม ทุกคลาสข้อมูลผู้ใช้งานกลุ่มใดสร้างคลาสขึ้นมา จะเสมือนเป็นสมาชิกภายใต้คลาส owl: class นี้

```
</owl:Class>
    <owl:Class rdf:ID="publisher">
      <rdfs:label>publisher</rdfs:label>
      <rdfs:subClassOf rdf:resource="#Any" />
```

**รูปที่ 1.** ตัวอย่างการกำหนดคลาสและซับคลาส

2.3.2 การกำหนดพร๊อพเพอร์ตี้ (Class) คือ คุณสมบัติหรือความสัมพันธ์ (Relation) หรือสล็อต (Slot) เป็นการกำหนดความสัมพันธ์หรือคุณลักษณะของคลาสเพื่อเชื่อมโยงระหว่างคลาส ด้วยการระบุค่าพร๊อพเพอร์ตี้ที่สามารถกำหนดเป็นค่าคงที่ [21]-[22] ซึ่งมีการกำหนด 2 แบบคือ owl: DatatypeProperty และ owl: ObjectProperty

ก) การกำหนดพร๊อพเพอร์ตี้ด้วย owl: DatatypeProperty เพื่อกำหนดการอธิบายคุณสมบัติของคลาสที่เป็นค่าคงที่ เช่น การอธิบายปี ชื่อ นามสกุล [22]

ข) การกำหนดพร๊อพเพอร์ตี้ด้วย owl: ObjectProperty เพื่อกำหนดการอธิบายข้อมูล ซึ่ง ต้องการอธิบายคุณสมบัติของคลาสที่เป็นรีซอร์ส (Resource) หรือกำหนดการเชื่อมโยงความสัมพันธ์ระหว่าง 2 คลาส [22]

### 2.4 ส่วนสอบถามข้อมูล (SPARQL)

การสร้างออนโทโลยีมีหลักที่สำคัญคือการศึกษาข้อมูลที่นำมาสร้างเพื่อพิจารณาความซับซ้อนของข้อมูล ซึ่งถ้าข้อมูลที่สร้างมีความซับซ้อนมากต้องพิจาณาทั้งการเลือกใช้ภาษา และกฎที่เกิดขึ้น เพื่อนำมาใช้งานต่อ ทำการตรวจสอบการค้นคืนด้วยการใช้การสอบถามด้วย (SPARQL)

SPARQL เป็นภาษาสำหรับดึงข้อมูลมาแสดง หรือเรียกว่าภาษาสอบถาม “Query Language” อยู่บนพื้นฐานของข้อมูลที่เป็นไปในรูปแบบของกราฟ มีลักษณะในรูปแบบของ RDF OWL กราฟที่มีลักษณะที่ง่ายที่สุด ซึ่งมีการเข้าถึงข้อมูลโดยอาศัยโครงสร้าง Triple (Subject, Predicate, Object) แนวความคิดของหลักการสืบค้นด้วยภาษา SPARQL แบ่งเป็นสองส่วนหลักๆ คือ ส่วน SELECT และ WHERE ซึ่ง SELECT เป็นการอธิบายตัวแปรที่ใช้ Prefix “ ? ” เป็นตัวแปรที่เก็บค่าผลลัพธ์ และ WHERE เป็นเงื่อนไขการสืบค้นข้อมูล [23]

นิยาม: การสืบค้นออนโทโลยีด้วยภาษา SPARQL

Sparql $\longrightarrow$ SELECT varlist WHERE (bgp)

โดยที่ varlist = (v1,v2,…,vn) varlist $\subseteq$ var(bgp)

bgp คือ Basic Graph Pattern

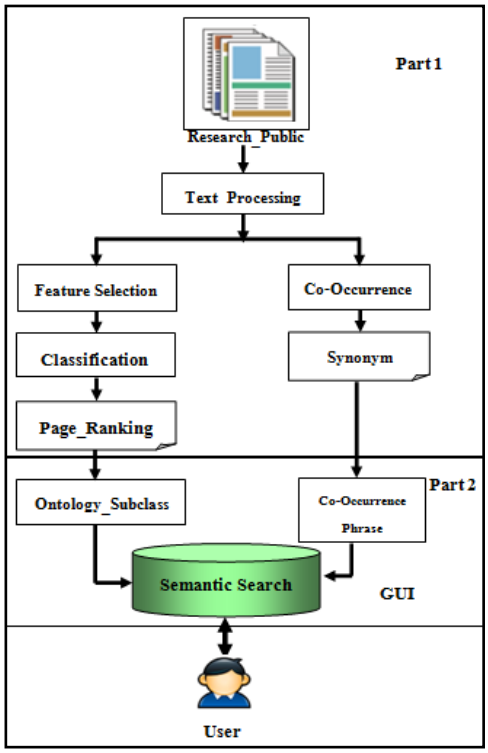
## 3. วิธีการดำเนินการวิจัย

การพัฒนาระบบการค้นคืนเชิงความหมายเป็นการประยุกต์ใช้งานเพื่อให้ผู้ใช้สามารถค้นคืนข้อมูลงานวิจัยโดเมน Information System ได้ตรงความต้องการของผู้ใช้มากที่สุดในการพัฒนาในครั้งนี้ใช้ฐานข้อมูล Mysql ร่วมกับโครงสร้างออนโทโลยีที่ได้ทำการออกแบบโดยใช้มาตรฐานเทคโนโลยีภาษาของ XML เพื่อใช้เก็บข้อมูลรูปแบบเอกสารที่มีความแตกต่างกันให้สามารถใช้งานร่วมกันได้ นำ RDF มาใช้อธิบายรูปแบบ

ความสัมพันธ์ของข้อมูล และ OWL ใช้อธิบายโครงสร้างข้อมูล โดยนิยามขอบเขตของคำที่เกี่ยวกับโครงสร้างของข้อมูล โดยมีขั้นตอนดังนี้

### 3.1 ศึกษาปัญหา รวบรวมข้อมูล

ศึกษาปัญหา รวบรวมข้อมูลงานวิจัยที่เกี่ยวข้องกับการจัดการองค์ความรู้ซึ่งใช้หลักการสร้างออนโทโลยีเป็นฐานความรู้ ทำการวิเคราะห์และออกแบบพร้อมทั้งประเมินระบบโดยแสดงเป็นขั้นตอนการทำงานตามรูปที่ 2




**รูปที่ 2.** กรอบแนวคิดระบบการค้นคืนเชิงความหมายจากข้อมูลบรรณานุกรม

ทำการเก็บข้อมูลบทความวิจัยโดยใช้โปรแกรม Zotero ตามโดเมนที่เลือกใช้คือ Information System จาก ACM [24] ตัดเฉพาะคอลัมน์ที่ต้องการ เช่น ชื่อผู้แต่ง ชื่อเรื่อง คำสำคัญ ฯลฯ และผ่านกระบวนการเตรียมข้อมูล (Text-Processing) เพื่อนำมาใช้ในฐานความรู้และสร้างออนโทโลยี นำคำสำคัญที่ผู้เขียนเลือกใช้หรือวลีไปทำเหมืองข้อความเพื่อสร้างคลาสย่อยและหาความสัมพันธ์ของความหมายตามความต้องการพื้นฐานของออนโทโลยี คือ ประกอบด้วยคลาส ความสัมพันธ์ และคุณสมบัติ

จากรูปที่ 2 ส่วนที่ 1 แบ่งการทำงานเป็น 2 ส่วนคือ 1) การทำเหมืองข้อความและการหากฎความสัมพันธ์ และ 2) การสร้างออนโทโลยี

3.1.1 การทำเหมืองข้อความ คือการตัดคำใช้คำสำคัญของบทความวิจัยมาสร้างเป็นตัวแทนเอกสารเพื่อแทนข้อมูลทั้งหมดของบทความวิจัยด้วยหลักการตัดวลี (Phrase) คือใช้หลักการตัดปลาย คัดเลือกคุณลักษณะโดยใช้ ไคสแควร์ อินฟอร์เมชันเกน และเกนเรโช ทำการจำแนกข้อความด้วยวิธีนาอีฟเบย์ เบย์เซียนเน็ตเวิร์ค เคเนียเรสเนเบอร์ และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน

**ตารางที่ 1.** ผลการวัดประสิทธิภาพโดยรวม (F-Measure) จากการคัดเลือกคุณลักษณะและจำแนกข้อมูล

| Feature | SVMP | SVMR | BN | NB | KNN |
|---|---|---|---|---|---|
| Info-Gain | 73.50 | 74.00 | 57.00 | 65.40 | 72.10 |
| Gain Ratio | 73.50 | 74.00 | 57.00 | 65.40 | 72.10 |
| Chi-square | 73.50 | 74.00 | 57.00 | 65.40 | 72.10 |

ตารางที่ 2. ตัวอย่างวลีที่มีความหมายเหมือนคลาสหลัก

| คลาส | ความหมาย | Support | Confidence |
|---|---|---|---|
| Classification | Feature Selection | 0.30 | 11.54 |
| | Dimensionality Reduction | 0.20 | 7.69 |
| Clustering | Complex Network | 0.30 | 7.50 |
| | Approximation Algorithm | 0.20 | 5.00 |
| | Frequent Mining | 0.20 | 5.00 |
| Collaborative Filtering | Recommender System | 0.79 | 47.06 |
| | Graphical Model | 0.30 | 17.65 |
| Data Stream | Frequent Itemset | 0.30 | 17.65 |
| | Concept Drift | 0.20 | 11.76 |
| Web Service | Optimization | 0.20 | 25.00 |
| | Scientific Workflow | 0.20 | 25.00 |
| | Service Composition | 0.20 | 25.00 |
| | Distributed Computing | 0.20 | 25.00 |

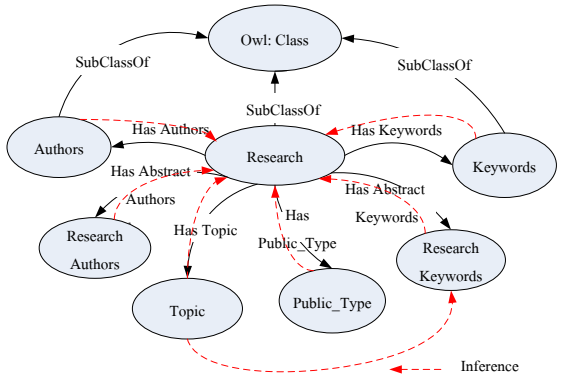
จากตารางที่ 1 สรุปว่าการคัดเลือกคุณลักษณะที่ดีคือการคัดเลือกคุณลักษณะแบบไคสแควร์เนื่องจากการคัดเลือกได้รวดเร็วกว่าวิธีการอื่น [25] และจำแนกประเภทแบบ SVMR ให้ผลการประเมินประสิทธิภาพโดยรวม 74.50% จึงนำวลีที่ได้มาสร้างคลาส ซึ่งจากค่าสถิติที่ทำการคัดเลือกได้จำนวน 174

วลีและแบ่งออกเป็น 2 กลุ่มคือ Data Mining และ Distributed Systems

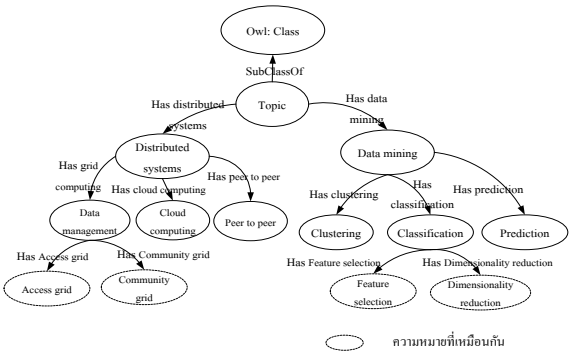
3.1.2 การหากฎความสัมพันธ์เพื่อค้นหาความหมายของวลีที่คล้ายคลึงกันโดยใช้คำสำคัญทั้งหมดมาทำการคำนวณหาค่าสนับสนุนและค่าความเชื่อมั่น โดยกำหนดค่าสนับสนุนให้มีค่ามากกว่า 0.20 ขึ้นไป โดยค่าความเชื่อมั่นเรียงจากมากไปน้อย

### 3.2 สร้างออนโทโลยี

ออนโทโลยี คือการอธิบาย เทอมของโดเมนที่สนใจในรูปแบบของคลาส (Class) ความสัมพันธ์ระหว่างคลาส (Relation) และคุณสมบัติของคลาส (Properties) แล้วนำเสนอออกมาในรูปของโหนด และความสัมพันธ์แบบลำดับชั้น [26] ผลที่ได้รับคือมีการแสดงความสัมพันธ์ตามรูปที่ 3




**รูปที่ 3.** ออนโทโลยีบทความวิจัย




**รูปที่ 4.** สร้างความสัมพันธ์ของคำที่มีความหมายเหมือนกันจากกฎความสัมพันธ์

ใช้โปรแกรม Protege [27] เป็นเครื่องมือสำหรับสร้างออนโทโลยีเป็นโปรแกรมโอเพ่นซอส สามารถสร้างไฟล์ได้หลากหลายรูปแบบเช่น XML, RDF และ OWL เป็นต้น ออนโทโลยีที่ใช้เป็นแบบ Taxonomic โดยมีความสัมพันธ์เป็นแบบ Is-a คือมีความสัมพันธ์จากพ่อไปสู่ลูก และความสัมพันธ์แบบ Part-of คือเป็นความสัมพันธ์ที่เป็นส่วนประกอบอย่างเช่น Research ประกอบด้วย ผู้แต่ง หัวเรื่อง คำสำคัญ [3] ซึ่งจากรูปที่ 3 เป็นการแสดงตัวอย่างข้อมูลเค้าร่างที่อธิบายรายละเอียดของคลาสบทความวิจัย โดยมีการประกาศคลาส และพร็อพเพอร์ตี้ เมื่อสร้างออนโทโลยีเรียบร้อยต้องมีการตรวจสอบความสัมพันธ์ของโครงสร้างเพื่อให้ได้ผลลัพธ์ที่ได้ออกมาตามโครงสร้างแบบลำดับชั้นคือจากพ่อไปสู่ลูกตามรูปที่ 3 และสร้างเป็นไฟล์ OWL เพื่อนำมาใช้งานร่วมกับการค้นคืนข้อมูล แสดงดังรูปที่ 5

```
<owl:Class rdf:ID="prediction">
 <rdfs:subClassOf>
  <owl:Restriction>
   <owl:onProperty>
    <owl:ObjectProperty rdf:ID="has_prediction"/>
   </owl:onProperty>
   <owl:someValuesFrom>
    <owl:Class>
     <owl:intersectionOf rdf:parseType="Collection">
      <owl:Class rdf:about="#prediction"/>
      <owl:Class rdf:about="#Distributed_Systems"/>
     </owl:intersectionOf>
    </owl:Class>
   </owl:someValuesFrom>
  </owl:Restriction>
 </rdfs:subClassOf>
```

**รูปที่ 5.** โครงสร้างไฟล์ OWL

### 3.3 ประเมินประสิทธิภาพ

การประเมินประสิทธิภาพใช้วิธีวัดค่าความถูกต้อง ค่าความแม่นยำ และการวัดประสิทธิภาพโดยรวม ดังสมการที่ (12) (13) และ (14) ตามลำดับ

ความแม่นยำ (Precision) เป็นอัตราส่วนของการค้นพบข้อมูลที่ถูกต้องจากจำนวนข้อมูลทั้งหมดที่ทำการค้นคืนมาได้

$$P = \frac{|Ra|}{|A|} \quad (12)$$

โดยที่ $|Ra|$ คือจำนวนข้อมูลที่ถูกต้องที่ค้นคืนออกมาได้

$|A|$ คือจำนวนข้อมูลทั้งหมดที่ค้นคืนออกมาได้

ค่าความระลึก (Recall) เป็นอัตราส่วนของการค้นพบข้อมูลที่ถูกต้องจากจำนวนข้อมูลที่ถูกต้องทั้งหมด

$$R = \frac{|Ra|}{|R|} \quad (13)$$

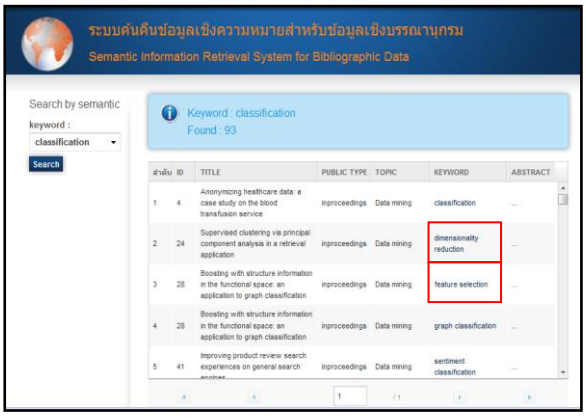
โดยที่ $|R|$ คือจำนวนข้อมูลที่ถูกต้องทั้งหมดในฐานข้อมูล

วัดประสิทธิภาพโดยรวม (F-measure) คือการวัดประสิทธิภาพโดยรวมของทั้งสองค่าระหว่างค่าความแม่นยำและค่าความระลึกซึ่งนำค่าทั้งสองมาคำนวณร่วมกัน

$$F-measure = \frac{2\times(R\times P)}{R+P} \quad (14)$$

## 4. ผลและข้อเสนอแนะ

ระบบการค้นคืนเชิงความหมายจากข้อมูลบรรณานุกรมทำการประยุกต์ใช้กับกฎความสัมพันธ์ของวลีเพื่อให้ได้ความหมายที่คล้ายคลึงกันของคำหลัก ทำการออกแบบและพัฒนาส่วนติดต่อกับผู้เชี่ยวชาญเพื่อให้ผู้เชี่ยวชาญประเมินผลลัพธ์การค้นคืนข้อมูลดังรูปที่ 6




**รูปที่ 6.** ระบบการค้นคืนเชิงความหมายจากข้อมูลบรรณานุกรม

ซึ่งจะเห็นผลลัพธ์ของคำที่มีความหมายเพิ่มขึ้น เช่น Feature Selection และ Dimensionality Reduction และทำการวัดประสิทธิภาพการค้นคืนดังตารางที่ 3

**ตารางที่ 3.** ผลลัพธ์จากการค้นคืนเชิงความหมายจากระบบการค้นคืนเชิงความหมายจากข้อมูลบรรณานุกรม

| No | คำที่ต้องการค้นคืน | \|A\| | \|Ra\| | P (%) |
|---|---|---|---|---|
| 1 | บทความวิจัยมีคำที่มีความหมายคล้ายคลึงกับ Classification | 92 | 92 | 100.00 |
| 2 | บทความวิจัยมีคำที่มีความหมายคล้ายคลึงกับ Clustering | 102 | 101 | 99.02 |
| 3 | บทความวิจัยมีคำที่มีความหมายคล้ายคลึงกับ Collaborative Filtering | 35 | 35 | 100.00 |
| 4 | บทความวิจัยมีคำที่มีความหมายคล้ายคลึงกับ Data Stream | 42 | 39 | 92.86 |
| 5 | บทความวิจัยมีคำที่มีความหมายคล้ายคลึงกับ Prediction | 38 | 37 | 97.37 |
| 6 | บทความวิจัยมีคำที่มีความหมายคล้ายคลึงกับ Cloud Computing | 16 | 15 | 93.75 |
| 7 | บทความวิจัยมีคำที่มีความหมายคล้ายคลึงกับ Data Grid | 23 | 22 | 95.65 |
| 8 | บทความวิจัยมีคำที่มีความหมายคล้ายคลึงกับ Data management | 8 | 8 | 100.00 |
| 9 | บทความวิจัยมีคำที่มีความหมายคล้ายคลึงกับ Fault Tolerance | 12 | 12 | 100.00 |
| 10 | บทความวิจัยมีคำที่มีความหมายคล้ายคลึงกับ Migration | 25 | 23 | 92.00 |
| 11 | บทความวิจัยมีคำที่มีความหมายคล้ายคลึงกับ Peer to Peer | 6 | 6 | 100.00 |
| 12 | บทความวิจัยมีคำที่มีความหมายคล้ายคลึงกับ Protocol | 13 | 13 | 100.00 |
| 13 | บทความวิจัยมีคำที่มีความหมายคล้ายคลึงกับ Scheduling | 38 | 38 | 100.00 |
| 14 | บทความวิจัยมีคำที่มีความหมายคล้ายคลึงกับ Storage | 25 | 25 | 100.00 |
| 15 | บทความวิจัยมีคำที่มีความหมายคล้ายคลึงกับ Web Service | 42 | 36 | 85.71 |
| **เฉลี่ย** | | | | **97.09** |

*|A| คือ จำนวนข้อมูลทั้งหมดที่ค้นคืนออกมาได้, |Ra| คือ จำนวนข้อมูลที่ถูกต้องที่ค้นคืนออกมาได้, P (%) = Precision

จากตารางที่ 3 ทำการประเมินโดยใช้ผู้เชี่ยวชาญจำนวน 3 ท่าน โดยทำการทดสอบการค้นคืนจากระบบที่พัฒนาขึ้นด้วยจำนวนคำค้นคืน 15 คำค้นคืน และให้ผู้เชี่ยวชาญพิจารณาผลลัพธ์ของคำที่ได้ร่วมกับชื่อบทความวิจัย โดยให้คะแนนเป็นถูก หรือ ผิด หลังจากนั้นนำค่าคะแนนของทั้ง 3 ท่านมาทำการหาค่าเฉลี่ย ซึ่งถ้ามีการให้คะแนนในผลลัพธ์ของการค้นคืนตรงมากกว่า 2 ท่านขึ้นไปถือว่าถูกต้อง เนื่องจากให้ค่ามากกว่า 0.5 คะแนน สรุปคือ ผลประเมินระบบการค้นคืนเชิงความหมายจากข้อมูลบรรณานุกรมโดยวัดค่าความแม่นยำ 97.09 %

### 4.1 ข้อเสนอแนะ

จากการจากการค้นหากฎความสัมพันธ์เพื่อให้ได้คำศัพท์ที่มีความหมายเหมือนกันโดยมีคำส่วนใหญ่เมื่อมีการจัดเทียบคู่คำศัพท์อยู่ในหัวข้อเดียวกันจะทำให้ประสิทธิภาพการค้นคืนได้ค่าความแม่นยำดีขึ้น เนื่องจากวลีที่มีความหมายเหมือนกันไม่มีจำนวนของวลีที่เพิ่มขึ้นในกลุ่มตรงข้าม เช่นคำว่า Classification มี Feature Selection และ Dimensionality Reduction จะไม่พบวลีที่เพิ่มขึ้นในกลุ่ม Distributed Systems จึงทำให้ผลของการค้นคืนคำเหล่านี้ไม่ผิดกลุ่มและถือว่าเป็นข้อมูลที่ดีส่วนคำที่ไม่สามารถแยกกลุ่มได้ถูกต้องเช่น Web Service เนื่องจากวลีเหล่านี้ไม่สามารถแยกกลุ่มได้อย่างชัดเจนจึงทำให้วลีที่มีความหมายเหมือนวลีหลักเพิ่มความหมายผิดกลุ่ม เช่น Web Service ใน Distributed Systems มีวลีที่มีความหมายเหมือนกันคือ *Optimization*, Scientific Workflow, Service Composition, Distributed Computing ซึ่งถ้าพิจารณาคำว่า *Optimization* เป็นคำที่ไม่สื่อความหมายที่ชัดเจนในการแยกกลุ่มจึงทำให้อยู่ได้ทั้งสองกลุ่มคือ Data Mining และ Distributed Systems และทำให้ประสิทธิภาพการค้นคืนลดลง ซึ่งการปรับปรุงหรือการทำให้ประสิทธิภาพการค้นคืนให้ดีขึ้นคือ 1) การใช้ตรรกะอันดับที่หนึ่งช่วยแก้ไขปัญหาเรื่องคำกำกวมซึ่งจะทำให้ผลลัพธ์ของการค้นคืนเชิงความหมายดีขึ้น 2) การค้นหาคำที่มีความหมายคล้ายคลึงกัน อาจทำได้ดีกว่านี้ด้วยการใช้เทคนิคของการจัดกลุ่ม แบบอื่น ๆ เช่น วิธีการจัดกลุ่มแบบ Clustering ในการทำเหมืองข้อความ หรือ การใช้กฎความสัมพันธ์ด้วยอัลกอริธึมอื่น เพื่อลดปัญหาเรื่องคำกำกวม โดยสามารถกำหนดเกณฑ์ในการเลือกจากค่าสถิติที่เกิดขึ้นจากอัลกอริธึมเหล่านั้น

## เอกสารอ้างอิง


[1] เพ็ญพรรณ อัศวนพเกียรติ, อรินทิพย์ ธรรมชัยพิเนต และกฤษณะ ไวยมัย, *"ออนโทโลยีชีวภาพ : ระบบสำหรับสืบคืนและวิเคราะห์ข้อมูลทางด้านชีววิทยา"*, Available from: https://pindex.ku.ac.th/file_research/ku1.pdf

[2] วิไลพร เลิศมหาเกียรติ, ภูริวัตร คัมภีรภาพพัฒน์ และอนิราช มิ่งขวัญ, "รูปแบบการแสดงผลการค้นคืนของเครื่องมือการสืบค้นสารนิเทศบนอินเตอร์เน็ต", *วารสารวิชาการพระจอมเกล้าพระนครเหนือ*, ปีที่ 18 ฉบับที่ 1, มกราคม–เมษายน 2551, หน้า 89-98

[3] สินีนาถ นาคโตและไพลิน เงินประเสริฐ, *"การค้นหาเอกสารงานวิจัยเชิงความหมาย"*, ปริญญานิพนธ์วิทยาศาสตร์บัณฑิต ภาควิชาวิทยาการคอมพิวเตอร์และสารสนเทศ คณะวิทยาศาสตร์ประยุกต์ สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ, 2549.

[4] สิทธา อาภาศิริกุลและสาริน สุภาพึ่ง, *"ระบบช่วยเลือกซื้อโทรศัพท์มือถือโดยใช้เว็บเชิงความหมาย"*, ปริญญานิพนธ์วิทยาศาสตร์บัณฑิต ภาควิชาวิทยาการคอมพิวเตอร์และสารสนเทศ คณะวิทยาศาสตร์ประยุกต์ สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ, 2549.

[5] W3C, *"Resource Description Framework"*, Available from: http://www.w3.org/XML/Schema

[6] W3C, *"Resource Description Framework"*, Available from: http://www.w3.org/RDF/

[7] W3C, *"RDF Vocabulary Description Language 1.0: RDF Schema"*, Available from: http://www.w3.org/TR/rdf-schema/

[8] W3C, *"OWL Web Ontology Language,* Available from: http://www.w3.org/TR/owl-features/

[9] X.H.Wang, et al. "Ontology Based Context Modeling and Reasoning using OWL", *Proceedings of the Second IEEE Annual Conference on Pervasive Computing and Communications Workshops (PERCOMW'04)*, 2004. pp. 18–22

[10] C. Haruechaiyasak, et al. *"Implementing News Article Category Browsing Based on Text Categorization Technique",* The 2008 IEEE/WIC/ACM International Conference on Web Intelligence (WI-08) workshop on Intelligent Web Interaction (IWI 2008), 2008. pp. 143-146

[11] V. Nuipian, P. Meesad, and P. Boonrawd, "Improve Abstract Data with Feature Selection for Classification Techniques**",** *Advanced Materials Research,* Vol. 403-408, November 2011. pp. 3699-3703

[12] K. Thongklin, S. Vanichayobon and W. Wett, "Word Sense Disambiguation and Attribute Selection Using Gain Ratio and RBF Neural Network", *IEEE International Conference Innovation and Vision for the Future in Computing & Communication Technologies (RIVF' 08)***,** Vaidnam, 2008.

[13] P.N. Tan, M. Steinbach, and K. Vipin, *Introduction to Data Mining.* United State of America : Addison Wesley, 2005.

[14] J. R. Quinlan, "Induction of Decision Trees", *Machine Learning,* Vol. 1, 1986. pp. 81-106

[15] P. Saengsiri, et al. "Comparison of Hybrid Feature Selection Models on Gene Expression Data", *IEEE International Conference on ICT and Knowledge Engineering,* 2010. pp.13-18

[16] G. L. Bretthorst, *"Bayesian Spectrum Analysis and Parameter Estimation in Lecture Notes in Statistics",* New York, Springer-Verlag*,* 1988.

[17] D. Lewis, "Naïve (Bayes) at forty: The independence assumption in information retrieval", *Lecture Notes in Computer Science*, 1998. pp. 4-15

[18] B. Links, "*A Detailed Introduction to K-Nearest Neighbor (KNN) Algorithm",* Available from: http://saravananthirumuruganathan. wordpress.com/2010/05/17/a-detailed-introduction-to-k-nearest-neighbor-knn-algorithm/

[19] V. Vapnik, *"The Nature of Statistical Learning Theory",* New York, Springer, 1995.

[20] C.J. Burges, "A tutorial on support vector machines for pattern recognition*", Data Mining and Knowledge Discovery 2,* 1998. pp. 121–167

[21] สิริรัตน์ ประกฤติกรชัย, *"การสร้างต้นแบบออนโทโลยีของพืชสมุนไพรไทย",* สารนิพนธ์วิทยาศาสตร์มหาบัณฑิต ภาควิชาวิทยาการคอมพิวเตอร์และสารสนเทศ คณะวิทยาศาสตร์ประยุกต์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ, 2550.

[22] วรวิกานต์ ปัณณะรัส, *"การประยุกต์เว็บเชิงความหมายในการสืบค้นความเชี่ยวชาญของนักวิจัย",* วิทยานิพนธ์วิทยาศาสตร์มหาบัณฑิต ภาควิชาวิทยาการคอมพิวเตอร์และสารสนเทศ คณะวิทยาศาสตร์ประยุกต์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ, 2552.

[23] W3C, *"SPARQL"***,** Available from: http://www.w3.org/TR/rdf-sparql-query/

[24] Association for Computing Machinery : ACM, Available from: http://portal.acm.org/portal.cfm

[25] P. Meesad, V. Nuipian and P. Boonrawd, "A Chi-Square-Test for Word Importance Differentiation in Text Classification", *Proceedings of 2011 International Conference on Information and Electronics Engineering (ICIEE 2011)***,** 2011. pp. 110-114

[26] G. Cui, et al. "Automatic Acquisition of Attributes for Ontology Construction", *Springer,* 2009.

[27] Protégé*,* Available from: http://protege.stanford.edu/