การพัฒนาเหมืองความรู้เชิงความหมายสมุนไพรไทยเพื่อการบำบัดรักษา

The Development of Semantic Thai Herb Knowledge Mining for Treatment

อนุพงษ์ สุขประเสริฐ[1] ดวงจันทร์ สีหาราช[2*] เจษฎาพร ปาคำวัง[2] สุภาพร วิสุงเร[3] และ จิตรนันท์ ศรีเจริญ[2]

Anupong Sukparseart[1], Duangchan Siharad[2*], Jetsadaporn Pakamwang[2], Supaporn Wisungre[3] and Jitranan Sricharoen[2]

Received: 20 June 2022, Revised: 28 January 2023, Accepted: 18 August 2023

บทคัดย่อ

การใช้สมุนไพรไทยเพื่อการบำบัดรักษานับเป็นการแพทย์ทางเลือก ที่ช่วยดูแลสุขภาพเบื้องต้นให้แก่ประชาชน ให้บรรเทาจากอาการโรคต่าง ๆ ได้ ประเด็นสำคัญนี้ นำไปสู่การวิจัย โดยมีวัตถุประสงค์ 1) เพื่อสร้างออนโทโลยีและกฎในการสืบค้นความรู้เชิงความหมายสมุนไพรไทย 2) เพื่อพัฒนาเหมืองความรู้เชิงความหมายสมุนไพรไทย และ 3) เพื่อประเมินประสิทธิภาพในการสกัดความรู้เชิงความหมาย และจำแนกข้อมูลสมุนไพรไทย ทำการรวบรวมองค์ความรู้สมุนไพรไทยจาก 100 เว็บไซต์ เพื่อนำมาสร้างออนโทโลยี และทำการสกัดความรู้เชิงความหมายด้วยภาษาเอสดับบลิวอาร์แอล ร่วมกับการประมวลผลภาษาธรรมชาติ ทำการวัดประสิทธิภาพในการสกัดความรู้เชิงความหมาย และประสิทธิภาพการทำนายเพื่อจำแนกชื่อสมุนไพร ด้วยอัลกอริทึมการเรียนรู้ของเครื่อง ได้แก่ นิวรอลเนตเวิร์ต เอสวีเอ็ม เคเนียเรสเนสเบอร์ และแผนผังต้นไม้เพื่อการตัดสินใจ ผลการวิจัยพบว่า ออนโทโลยีสมุนไพรไทย แบ่งเป็น 2 ชั้นความรู้ และมีคลาสความรู้จำนวนทั้งหมด 8 คลาส สามารถสร้างกฎ เพื่อดึงโครงสร้างความรู้ที่ช่วยระบุรูปแบบข้อมูลของคลาสคำตอบสมุนไพร 3 ชนิด ได้แก่ ไพล ย่านาง และฟ้าทะลายโจร กระบวนการเรียนรู้ของเครื่องบนเหมืองความรู้เชิงความหมายสมุนไพรไทย ประกอบด้วย 4 ส่วน ได้แก่ 1) การประมวลผลเชิงความหมายด้วยออนโทโลยี 2) การสกัดเหมืองความรู้ 3) การสร้างชุดข้อมูลสมุนไพร และ 4) การทำนายผลลัพธ์ด้วยการเรียนรู้ของเครื่อง ผลการประเมินประสิทธิภาพในการค้นคืนความรู้เชิงความหมายสมุนไพร

---

[1] สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏเพชรบูรณ์ อำเภอเมือง จังหวัดเพชรบูรณ์ 67000

[1] Major of Information Technology and Communication, Faculty of Science and Technology, Phetchabun Rajabhat University, Mueang, Phetchabun 67000, Thailand.

[2] สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏเพชรบูรณ์ อำเภอเมือง จังหวัดเพชรบูรณ์ 67000

[2] Major of Computer Science, Faculty of Science and Technology, Phetchabun Rajabhat University, Mueang, Phetchabun 67000, Thailand.

[3] สาขาวิชาแพทย์แผนไทย คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏเพชรบูรณ์ อำเภอเมือง จังหวัดเพชรบูรณ์ 67000

[3] Major of Thai Traditional Medicine, Faculty of Science and Technology, Phetchabun Rajabhat University, Mueang, Phetchabun 67000, Thailand.

[*] ผู้นิพนธ์ประสานงาน ไปรษณีย์อิเล็กทรอนิกส์ (Corresponding author, e-mail): kob_siharad@pcru.ac.th

ไทยโดยใช้ออนโทโลยี และกฎ ได้ค่าเอฟเมสเซอร์ เท่ากับ 94.8% และประสิทธิภาพในการทำนาย เพื่อจำแนกคลาสคำตอบสมุนไพร ด้วยอัลกอริทึมนิวรอลเนตเวิร์ต ได้ค่าความถูกต้องสูงที่สุด เท่ากับ 90.0% ส่วนอัลกอริทึมเคเนียเรสเนสเบอร์ ได้ค่าความถูกต้องต่ำที่สุด เท่ากับ 82.2%

คำสำคัญ: เหมืองความรู้, ออนโทโลยี, การเรียนรู้ของเครื่อง, สมุนไพรไทย

ABSTRACT

Using Thai herbs for treating is an alternative medicine that helps take care of people's preliminary health and relieves them from diseases and symptoms. This is an important topic that leads to operating this research aimed at: 1) creating the ontology and the rules for searching the semantic Thai herb knowledge, 2) developing semantic Thai herb knowledge mining, and 3) assessing the efficiency of extracting Thai herb semantic knowledge and classifying the Thai herbs' class. The Thai herb knowledge was collected from 100 websites to build an ontology and extract the semantic knowledge using SWRL with natural language processing. The performance is measured in extracting the semantic knowledge and predicting the classification of herb classes using the neural network (NN), the support vector machine (SVM), the K-nearest neighbor (KNN), and the decision tree (DT). The results show that Thai herb ontology was divided into 2 knowledge classes and comprised eight knowledge nodes. SWRL rules were created to extract the knowledge structures identifying the herb data pattern in 3 types: Zingiber montanum, Tiliacora triandra, and Andrographis paniculate. The ML processing on semantic Thai herb knowledge mining consisted of 4 parts: 1) semantic processing on ontology, 2) extracting the knowledge mining, 3) creating the Thai herb dataset, and 4) predicting the results using ML. The results of assessing the semantic Thai herb knowledge retrieval performance using the ontology and the SWRL rules yielded an F-measure value of 94.8%. The NN had the highest accuracy in predicting Thai herb classes at 90.0% and the KNN had the lowest accuracy at 82.2%.

Key words: knowledge mining, ontology, machine learning, Thai herb

บทนำ

องค์ความรู้สมุนไพรไทยมีความสำคัญเป็นอย่างมากในทางการรักษาโรคและการดูแลสุขภาพ องค์ความรู้เหล่านี้จะทำให้ผู้ที่ประสงค์จะใช้สมุนไพรมีความเข้าใจและสามารถนำสมุนไพรไปใช้ได้อย่างถูกต้องและปลอดภัย ซึ่งในปัจจุบันผู้คนได้หันมาให้ความสำคัญและมีความสนใจเกี่ยวกับการใช้สมุนไพรมากขึ้น โดยกระทรวงสาธารณสุขมีนโยบายส่งเสริมการใช้ประโยชน์จากสมุนไพรในงานสาธารณสุขมูลฐาน (Ministry of Public Health, 2008) ตามพระราชบัญญัติคุ้มครองและส่งเสริมภูมิปัญญาแพทย์แผนไทย ให้มีการคุ้มครองภูมิปัญญาด้านการใช้สมุนไพรไทยที่ระบุไว้ในพระราชบัญญัติ ปี พ.ศ. 2542 จึงทำให้มหาวิทยาลัยและสถาบัน

ทางด้านสาธารณสุขต่าง ๆ ได้มีการวิจัยพืชสมุนไพรอย่างต่อเนื่อง ทำให้เนื้อหาและข้อมูลด้านสมุนไพรไทยมีจำนวนมากอยู่ในอินเทอร์เน็ต ซึ่งมาจากแหล่งข้อมูลที่หลากหลาย ทำให้ข้อมูลมีการกระจัดกระจาย และมีความขัดแย้งกัน โดยที่บางแหล่งข้อมูลให้ข้อมูลข้อห้ามการใช้สมุนไพรในปริมาณที่มากเกินความจำเป็น แต่ในขณะเดียวกันบางแหล่งข้อมูลไม่ได้แนะนำวิธีการใช้และข้อห้ามของสมุนไพรต่าง ๆ จึงทำให้เกิดการสื่อสารที่ผิดพลาดไปยังผู้ใช้ ทั้งนี้อาจเนื่องมาจากข้อมูลด้านสมุนไพรไทยที่ถูกเผยแพร่ออกไปอาจยังไม่ได้รับการตรวจสอบอย่างเป็นระบบจากผู้เชี่ยวชาญหรือวิธีการคัดกรององค์ความรู้ด้วยกระบวนการทางคอมพิวเตอร์ ส่งผลให้การสืบค้นข้อมูลผ่านระบบเสิร์ชเอ็นจินโดยทั่วไปอาจจะทำให้ได้ข้อมูลไม่ตรงกับความต้องการของผู้สืบค้น และองค์ความรู้ด้านสมุนไพรไทยส่วนใหญ่ที่พบบนเว็บไซต์ มักจะขาดการเชื่อมโยงและการจัดหมวดหมู่เพื่อให้ผู้ใช้เกิดความเข้าใจที่ชัดเจนอย่างคลอบคลุมในการนำองค์ความรู้ไปใช้ประโยชน์ได้เบ็ดเสร็จในแหล่งความรู้เดียว

การแก้ไขปัญหาเหล่านี้ จะสามารถประยุกต์ใช้ออนโทโลยี ในการระบุแนวคิดหลักและความสัมพันธ์ระหว่างแนวคิดในโดเมนความรู้ของสมุนไพรไทย (Tungkwampian *et al*., 2015) ร่วมกับการนำเทคนิคการทำเหมืองข้อมูลมาใช้ค้นหาความรู้ใหม่จากข้อมูลที่มีปริมาณมาก โดยมีการสกัดข้อมูลต่าง ๆ เพื่อให้สามารถนำข้อมูลเหล่านี้ไปใช้ประโยชน์ได้ ดังงานวิจัยของ Prompukdee *et al.* (2017) ได้นำเทคนิคเหมืองข้อมูลมาประยุกต์ใช้ในการแพทย์แผนไทย โดยสกัดชุดข้อมูลการรักษาอาการของผู้ป่วยที่แตกต่างกัน และใช้การค้นหากฎการเชื่อมโยงเพื่อนำไปใช้สร้างแบบจำลอง เพื่อวินิจฉัยผู้ป่วยจากอาการที่บันทึกไว้ จากการวิเคราะห์ถึงข้อจำกัดในการสกัดองค์ความรู้จากเหมืองข้อมูลแบบเดิม มาสร้างชุดข้อมูลเพื่อส่งต่อให้อัลกอริทึมการเรียนรู้ของเครื่อง (ML : Machine learning) ใช้ในการทำนายเชิงจำแนกและจัดกลุ่มข้อมูลความรู้ พบว่า การสกัดข้อมูลยังขาดคุณภาพที่ดี และคุณลักษณะของข้อมูลบางตัวไม่ได้มีความเกี่ยวข้อง/ไม่เกิดประโยชน์ที่แท้จริงในการบ่งบอกถึงคลาสคำตอบได้อย่างมีประสิทธิภาพ ส่งผลให้โมเดล ML ที่นำไปใช้ประโยชน์ขาดคุณภาพและมีอัตราความแม่นยำในการทำนายที่ไม่เป็นที่น่าพอใจตามไปด้วย ในปัจจุบันมีการนำ ML มาใช้สร้างวิธีการที่ทำให้คอมพิวเตอร์เรียนรู้ได้จากข้อมูลตัวอย่างหรือสภาพแวดล้อม เพื่อสร้างตัวแบบและนำไปใช้ทำนายข้อมูลใหม่ได้ และ ML สามารถใช้ในการทำเหมืองข้อมูลในด้านธุรกิจ ใช้ในการวิเคราะห์ข้อมูลและค้นหารูปแบบการดำเนินธุรกิจต่าง ๆ ได้ (Xiuyi and Yuxia, 2018) และการทำเหมืองข้อมูลเชิงความหมาย (Semantic Mining) เป็นวิธีการสกัดและวิเคราะห์ความรู้จากข้อมูลตัวอักษร เพื่อให้คอมพิวเตอร์นำไปใช้ได้โดยตรง โดยจะสกัดคุณสมบัติ และความสัมพันธ์ทางความหมายระหว่างข้อความ โดยใช้เทคนิคการประมวลผลภาษาธรรมชาติออนโทโลยี ร่วมกับการสร้างฐานกฎเชิงความหมาย และทำการประยุกต์ใช้การเรียนรู้ของเครื่องมาช่วยจำแนกหรือจัดกลุ่มคำตอบเป้าหมายที่ผู้ใช้ต้องการทราบจากคุณลักษณะ (Feature) ของข้อความต่าง ๆ ที่สกัดออกมาได้ (Tian *et al*., 2023)

การพัฒนาเหมืองความรู้เชิงความหมาย (Semantic Mining) จะช่วยให้เข้าถึงความรู้จากแหล่งต่าง ๆ บนเครือข่ายอินเทอร์เน็ตอย่างเป็นระบบ ส่งผลให้การทำเหมืองข้อมูลเพื่อการสืบค้นข้อมูลมาทำการวิเคราะห์และทำนายผลลัพธ์ได้ถูกต้องและตรงกับความต้องการของผู้ใช้มากยิ่งขึ้น ทั้งนี้การประมวลผลข้อมูลเชิงความหมายจะช่วยให้การสกัดข้อมูลจากแหล่งข้อมูลหลายแหล่ง เพื่อนำมาสร้างชุดข้อมูลที่มีคุณลักษณะที่ครบถ้วน พร้อมส่งต่อเข้าสู่การวิเคราะห์

ด้วยวิธีการทางสถิติ และการทำนายข้อมูลด้วยเทคนิค ML ให้มีความแม่นยำมากยิ่งขึ้น เพื่อนำไปสู่การพัฒนาซอฟต์แวร์ที่สามารถใช้ประโยชน์ในชุมชนท้องถิ่นได้อย่างแท้จริง

วิธีดำเนินการวิจัย

การดำเนินงานวิจัยนี้ ได้กำหนดวิธีดำเนินการวิจัยที่สอดคล้องกับวัตถุประสงค์ของการวิจัย ประกอบด้วย 3 ขั้นตอน ดังนี้

1. การสร้างออนโทโลยีและกฎในการสืบค้นความรู้เชิงความหมายสมุนไพรไทย

ทำการศึกษารวบรวมองค์ความรู้สมุนไพรไทยจากแหล่งความรู้ออนไลน์บนเว็บไซต์จำนวน 100 แหล่ง อาทิ 1) wikipedia.com 2) ananhosp.go.th 3) mnre.go.th 4) rspg.or.th 5) nittm.com 6) hsri.or.th 7) rama.mahidol.ac.th 8) thaihealth.or.th 9) hfocus.org และ 10) saranukromthai.or.th โดยการคัดเลือกเว็บไซต์เหล่านี้ จากผู้เชี่ยวชาญด้านสมุนไพรไทย จำนวน 3 ท่าน ที่ผ่านตามเกณฑ์การนำเสนอข้อมูลสารสนเทศที่มีความถูกต้องและเชื่อถือได้ เพื่อนำมาสร้างออนโทโลยี โดยใช้โปรแกรม Protégé สำหรับรวบรวมและนิยามข้อมูลเชิงความหมายเกี่ยวกับสมุนไพรไทย โดยมีขั้นตอนการสร้างออนโทโลยีจำแนกได้ 4 ขั้นตอน ได้แก่ 1) กำหนดโดเมน และขอบเขตของออนโทโลยีจากองค์ความรู้ที่สืบค้นได้จากเว็บไซต์ต่าง ๆ 2) แจกแจงคำสำคัญเพื่อกำหนดเป็นคลาส และแอทริบิวต์ในออนโทโลยี 3) กำหนดคุณสมบัติของคลาส และความสัมพันธ์ระหว่างคลาส และ 4) สร้างรายการข้อมูล (Instant) เพื่อเก็บรวบรวมในคลาสต่าง ๆ ทั้งนี้วิธีการสร้างออนโทโลยีดังกล่าวมีความสอดคล้องกับงานวิจัยของ Noy and McGuinness (2001) และออนโทโลยีได้รับการตรวจสอบความถูกต้องครบถ้วน และความสอดคล้องกับองค์ความรู้ต้นฉบับจากเว็บไซต์ โดยผู้เชี่ยวชาญ 3 ท่านในข้างต้น จากนั้นทำการสร้างกฎเพื่อสืบค้นความรู้ด้วยภาษา SWRL (Semantic web rule language) โดยใช้เครื่องมือส่วนขยาย (Extension) ที่เปิดให้ติดตั้งเพิ่มเติม และใช้งานบนโปรแกรม Protégé ทั้งนี้ SWRL เป็นภาษาที่มีความสามารถเข้าถึงความรู้ และแสดงกฎสำหรับการอนุมานความรู้ทำให้เกิดข้อมูลใหม่ (Kliangkhlao *et al.*, 2015) ซึ่งสามารถนำมาใช้ในการสืบค้นข้อมูลเชิงความหมายได้อย่างมีประสิทธิภาพ

2. การพัฒนากระบวนการเรียนรู้ของเครื่องบนเหมืองความรู้เชิงความหมายสมุนไพรไทย

นำผลลัพธ์ของคำสำคัญที่เกี่ยวข้องกันเชิงความหมาย จากการสืบค้นความรู้ในออนโทโลยีด้วยกฎ SWRL ในขั้นตอนที่ 1 ไปใช้สกัดข้อมูลเพื่อสร้างเหมืองความรู้จากแหล่งความรู้บนเว็บไซต์ 100 แหล่ง โดยใช้ Web Scraping บนภาษา Python ในการดึงข้อมูล (Hillen, 2019; Yaiprasert and Yusakul, 2022) ตามโครงสร้างเชิงความหมายที่สร้างด้วยออนโทโลยี ซึ่งในขั้นตอนนี้จะใช้การประมวลผลภาษาธรรมชาติ (NLP : Natural Language Processing) เพื่อตัดข้อความจากประโยคความรู้บนเว็บไซต์ แล้วใช้อัลกอริทึม Name Matching Algorithm ด้วยเทคนิค Levenshtein Distance ซึ่งเป็นวิธีการวัดหาค่าความต่างกันของ 2 ข้อความ โดยวัดจากจำนวนอักขระที่จะต้องทำการตัดออก แทรก และแทนที่อักขระในข้อความ จนได้อักขระเหมือนข้อความต้นแบบ (Po, 2020) ในการคัดเลือกเฉพาะคำสำคัญที่มีอัตราความคล้าย (Similarity rate) กับชื่อคลาสความรู้จากออนโทโลยี ในอัตราค่า ratio 0.7 (70%) ขึ้นไป เพื่อนำมาใช้สร้างคุณลักษณะ (Feature) และค่าในแต่ละคุณลักษณะ (Value) บนชุดข้อมูล (Dataset) สมุนไพรไทย ซึ่งประกอบด้วย Feature ต่าง ๆ อาทิ ส่วนประกอบที่ใช้ประโยชน์ได้ อาการ/โรค วิธีการนำมาใช้ประโยชน์ ห้ามรับประทานเกินกี่วัน และปริมาณที่

แนะนำให้รับประทาน เพื่อระบุไปยังคลาสคำตอบ คือ ชื่อสมุนไพร ที่นำไปใช้ในกระบวนการทำนายเชิงการจำแนกด้วยอัลกอริทึม ML ใน 4 เทคนิค ดังนี้

2.1 Neural Network (NN) เป็น Algorithm หลักของการเรียนรู้แบบ Deep learning ซึ่งมีกระบวนการประมวลผลดังสมการ 1 และ 2 (Jordan, 2017)

$$a = g(z) = g(\sum_{i=1}^{n} x_i w_i + b),\ a = g(z) = \frac{1}{1+e^{-z}},\ a = g(z)\begin{cases}0 & if \quad z \le 0\\ 1 & if \quad z > 0\end{cases} \quad (1)$$

$$\frac{\partial E_{total}}{\partial_{w5}}, \frac{\partial E_{total}}{\partial_{w5}} = \frac{\partial E_{total}}{\partial out_{o1}} * \frac{\partial out_{o1}}{\partial net_{o1}} * \frac{\partial net_{o1}}{\partial w_5} \quad (2)$$

จากสมการ (1) ให้ a เป็น perceptron ที่คำนวณผลรวมของค่า input (x_i) คูณด้วยค่าน้ำหนัก (w_i) และรวมด้วยค่าไบแอส (b) ในชั้นโครงข่าย (Layer) ต่าง ๆ ตามโครงสร้างของ Neural network จากนั้นค่า Output ในชั้นสุดท้าย จะถูกนำมาคำนวณค่าความน่าจะเป็นตั้งแต่ค่า 0 ถึง 1 ด้วยฟังก์ชัน Sigmoid คือ $z = 1/(1+e^{-z})$ เพื่อตัดสินคลาสคำตอบ โดยจะตรวจสอบว่า ถ้า z <= 0 ทำนายคลาส 0, z>0 ทำนายคลาส 1 ในกระบวนการ Feed forward หากได้ผลลัพธ์ที่ผิดพลาดจากคลาสคำตอบจริง จะทำการ Back propagation ดังสมการ (2) เพื่อกระจายค่าความผิดพลาดย้อนกลับด้วยกฎ Chain rule เพื่อทำการปรับ w_i และ b ในชั้นต่าง ๆ และวนรอบทำซ้ำในกระบวนการ Feed forward จนครบจำนวนรอบการประมวลผลที่กำหนดไว้

2.2 Support Vector Machine (SVM) เป็นเทคนิคในการจดจำรูปแบบเพื่อแก้ปัญหาการจัดกลุ่ม (classification problem) (Wang *et al.*, 2009) โดยใช้การหาสัมประสิทธิ์ของสมการเพื่อสร้างเส้นแบ่งที่สุดในการแยกกลุ่มข้อมูลที่ถูกป้อนเข้าสู่อัลกอริทึม (optimal separating hyper plane) เมื่อเราพิจารณาข้อมูลที่ประกอบด้วยข้อมูล 2 กลุ่มดังสมการที่ 3

$$D = \{(x_i, y_i); i = 1,2, \dots, n\} \quad (3)$$

เมื่อ $x_i = (x_{i1}, x_{i2}, \dots, x_{im}) \in R^m$

$y_i \in \{1, -1\}$ โดย 1 คือ ข้อมูลกลุ่ม 1 และ -1 คือ ข้อมูลกลุ่ม 2

ซึ่งเป็นการกำหนดกลุ่มเป้าหมายให้ SVM โดยที่ SVM นั้นมุ่งเป้าเพื่อหาฟังก์ชันการตัดสินใจที่สามารถแบ่งแยกค่าที่ไม่ทราบได้ ดังสมการที่ 4

$$f(x) = \text{sign}\{\sum_{k=1}^{n_v} w_k \varphi_k(x)\varphi_k(x_k) + b\} \quad (4)$$

$$\varphi(x) = [\varphi_1(x_1), \varphi_2(x_2), \dots, \varphi_n(x_{n_v})]^T \quad (5)$$

กลุ่มข้อมูล x จากสมการที่ 5 ไม่สามารถแบ่งแยกได้ด้วยสมการเส้นตรงแต่จะถูกแปลงให้อยู่ในรูปแบบที่สามารถใช้สมการเส้นตรงแบ่งแยกได้ โดยใช้เคอร์เนลฟังก์ชัน (kernel function) ดังสมการที่ 6

$$K(x, x_k) = \varphi(x)\varphi(x_k) \quad (6)$$

เมื่อ $\varphi(x)$ แทน ฟังก์ชันสำหรับแปลงข้อมูลที่ไม่เป็นเชิงเส้นให้เป็นข้อมูลที่อยู่ในรูปเชิงเส้นสามารถแบ่งแยกได้

w_k แทน ค่าน้ำหนักที่เชื่อมโยงจาก feature space ไปสู่ output space

B แทน ค่าโน้มเอียง (bias)

x_k แทน ซัพพอร์ตเวกเตอร์

โดย $k = 1,2, \dots, n_v$

n_v แทน จำนวนซัพพอร์ตเวกเตอร์

วิธีการที่ใช้ในการหาเส้นแบ่งที่ดีที่สุด คือ การเพิ่มเส้นขอบ (margin) ให้กับเส้นแบ่งทั้งสองข้าง และ สร้างเส้นขอบที่สัมผัสกับค่าข้อมูลใน feature space ที่ใกล้ที่สุดดังนั้นเส้นแบ่งที่มีเส้นขอบกว้างที่สุดจึงเป็นเส้นแบ่งที่ดีที่สุดและเรียกตำแหน่งการสัมผัสข้อมูลที่ใกล้ที่สุดจากการเพิ่มขอบนี้ว่า “ซัพพอร์ต เวกเตอร์” (support vector) เนื่องจากในบางกรณีการแบ่งแยกกลุ่มไม่สามารถทำได้ถูกต้องโดยสมบูรณ์

ดังนั้น จึงต้องมีการกำหนดตัวแปรสำหรับยอมรับค่าความผิดพลาดโดยการเพิ่มตัวแปร ξ (slack variable) ดังสมการที่ 7 และ 8 ดังนี้

$$W^T\mathrm{x} + \mathrm{b} \geq \mathrm{y} - \xi_i \text{ เมื่อกำหนดให้ y = 1} \quad (7)$$

$$W^T\mathrm{x} + \mathrm{b} \leq \mathrm{y} + \xi_i \text{ เมื่อกำหนดให้ y = -1} \quad (8)$$

จากการกำหนดค่า $\xi_i > 0$ ทำให้โครงสร้างของซัพพอร์ตเวกเตอร์แมทชีนบรรลุวัตถุประสงค์ใน 2 ส่วนคือการเพิ่มระยะแบ่งแยกให้มากที่สุดและลดข้อผิดพลาดในการทำนายให้ต่ำที่สุด ดังสมการที่ 9

$$\mathrm{M}inimize \frac{1}{2}\|w\|^2 + c\sum_{i=1}^{N}\xi_i \quad (9)$$

โดยที่ $:y_i(W^T\varphi(x) + b) + \xi_i - 1 \geq 0$
$\xi_i \geq 0, i = 1,2,\ldots,N$

2.3 K-Nearest Neighbor (KNN) เป็นอัลกอริทึมการเรียนรู้ที่ไม่ซับซ้อน และมีประสิทธิภาพสำหรับการจำแนกข้อมูลและการรับรู้วัตถุซึ่ง KNN จะคำนวณระยะทางโดยปริภูมิแบบยุคลิด (Sun *et al.*, 2009) ดังสมการที่ 10

$$D(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_n - q_n)^2} \quad (10)$$

2.4 Decision Tree (DT) สามารถนำเสนอชุดข้อมูล และกระบวนการประมวลผลเพื่อสร้างเหมืองความรู้ในแผนภาพผังงาน ดังนี้

ตารางที่ 1 ชุดข้อมูลสร้างโมเดล Decision Tree

Feature 1	Feature 2	Feature 3	Class
T	T	T	Y
F	F	F	N
T	T	F	Y
F	F	T	N

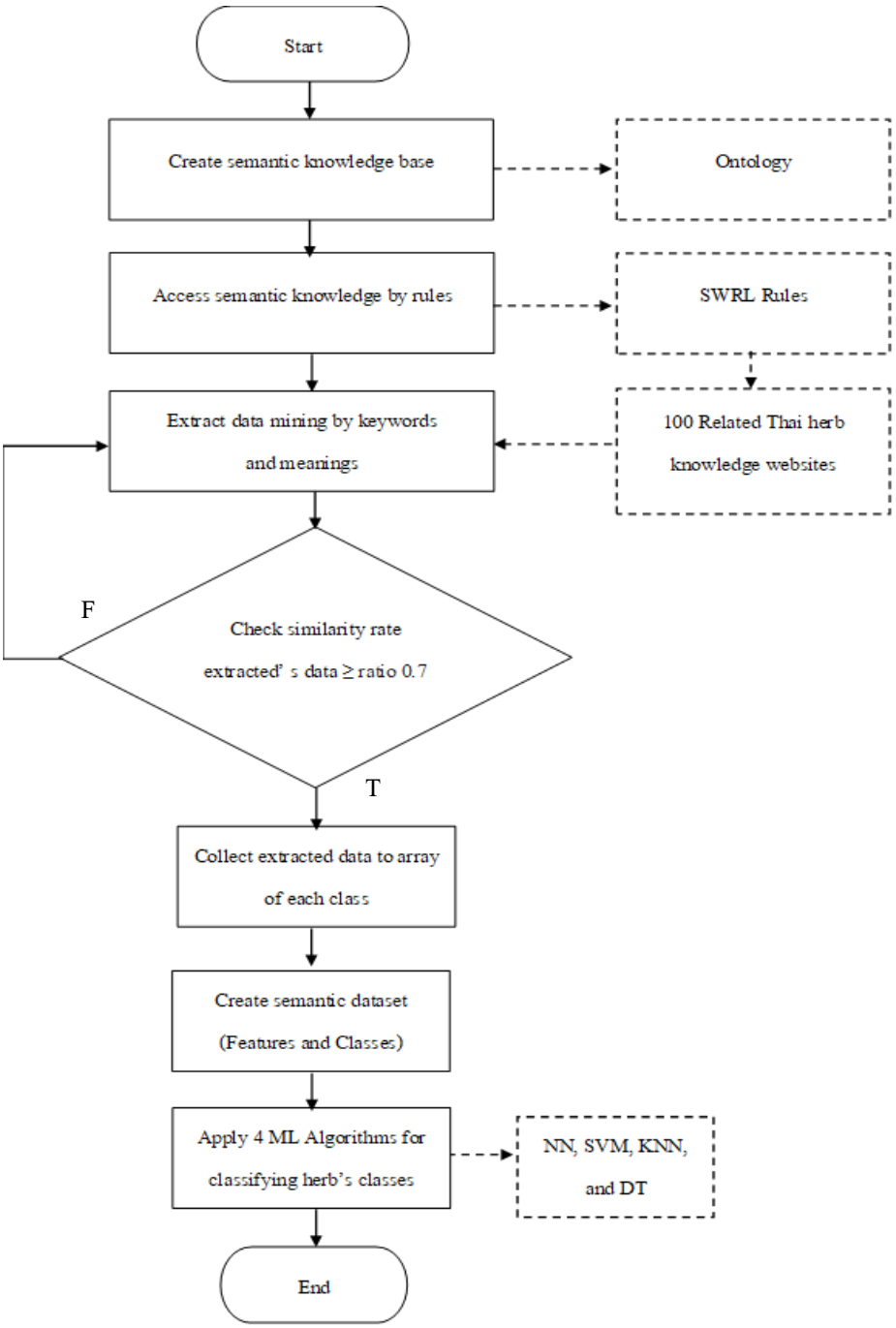
จากตารางที่ 1 สามารถสร้างกฎของเงื่อนไขการตัดสินใจด้วยโมเดล Decision Tree ดังนี้

R1 = IF (Feature 1 = T) AND (Feature 2 = T) THEN CLASS = Y

R2 = IF (Feature 1 = T) AND (Feature 2 = F) THEN CLASS = N

R3 = IF (Feature 1 = F) AND (Feature 2 = F) THEN CLASS = N

จากขั้นตอนการดำเนินงานข้างต้น สามารถนำเสนอขั้นตอนวิธีการพัฒนากระบวนการเรียนรู้ของเครื่องบนเหมืองความรู้เชิงความหมายสมุนไพรไทย ในรูปแบบผังงาน (Flowchart) ดังนี้



ภาพที่ 1 Flowchart ของกรอบแนวคิดกระบวนการเรียนรู้ของเครื่องบนเหมืองความรู้เชิงความหมายสมุนไพรไทย

จากภาพที่ 1 สามารถอธิบายกรอบแนวคิดของกระบวนการเรียนรู้ของเครื่องบนเหมืองความรู้เชิงความหมายสมุนไพรไทย ได้ 4 ขั้นตอน ดังนี้

1. สร้างใช้ออนโทโลยีโดยใช้โปรแกรม Protégé เพื่อรวบรวมองค์ความรู้สมุนไพรไทยเพื่อการบำบัดรักษา

2. สร้างฐานกฎเชิงความหมาย SWRL สำหรับคลาสไพล ย่านาง และฟ้าทะลายโจร โดยใช้ Library บนโปรแกรม Protégé ในการอนุมานความรู้ในออนโทโลยี แล้วนำไปสกัดความรู้สมุนไพร จาก 100 เว็บไซต์ ทั้งนี้ในขั้นตอนการสกัดความรู้จะใช้ Name Matching Algorithm ด้วยเทคนิค Levenshtein Distance เป็นวิธีการวัดหาค่าความต่างกันของ 2 ข้อความ โดยวัดจากจำนวนอักขระที่จะต้องทำการตัดออก แทรก และแทนที่อักขระในข้อความ จนได้อักขระเหมือนข้อความต้นแบบ (Po, 2020) จากนั้นนำทั้งสอง 2 ข้อความมาเปรียบเทียบกันด้วยอัตราความคล้าย $\geq$ ratio 0.7 เพื่อคัดเลือกคำสำคัญหรือข้อความบนเว็บไซต์ ที่มีความคล้ายกับชื่อคลาสหรือข้อมูลที่เก็บไว้ในคลาสต่าง ๆ ของออนโทโลยีตามที่อนุมานได้

3. สร้างชุดข้อมูลเชิงความหมาย (semantic dataset) ซึ่งประกอบด้วยคุณลักษณะ (feature) และค่าความถี่ในการค้นพบคำสำคัญ (value) ของแต่ละคุณลักษณะ

4. ส่งชุดข้อมูลให้กับอัลกอริทึมการเรียนรู้ของเครื่อง (ML) ได้แก่ Neural network, SVM, KNN และ Decision tree เพื่อใช้ในการจำแนกคลาสคำตอบ โดยที่อัลกอริทึม ML ทั้ง 4 ชนิด จะถูกตั้งค่าพารามิเตอร์เพื่อให้กระบวนการเรียนรู้และทำนายเกิดประสิทธิภาพที่ดี ดังนี้

4.1 Neural network ตั้งค่าพารามิเตอร์ hidden_layer_sizes = (3,3), activation='relu', optimizer = 'adam' และวนรอบฝึกสอนโมเดล จำนวน 500 รอบ

4.2 SVM ตั้งค่าพารามิเตอร์ kernel='linear' และการกำหนดระยะขอบในการจำแนกข้อมูล (C=1.0)

4.3 KNN ตั้งค่าพารามิเตอร์ ที่ k=3

4.4 Decision tree ตั้งค่าพารามิเตอร์ criterion="entropy" และ max_depth=3

3. การประเมินประสิทธิภาพในการสกัดความรู้เชิงความหมาย และจำแนกข้อมูลสมุนไพรไทย

แบ่งขั้นตอนการดำเนินงานเป็น 2 ส่วน ดังนี้

3.1 การประเมินประสิทธิภาพในการสกัดความรู้เชิงความหมาย

ใช้คำสำคัญเชิงความหมายที่สืบค้นได้จากออนโทโลยี นำไปใช้สกัด/ค้นคืนข้อมูลสมุนไพรไทย 3 ชนิด ได้แก่ ไพล ย่านาง และฟ้าทะลายโจร บนเว็บไซต์ 100 แหล่ง ซึ่งในการเลือกสมุนไพรไทย 3 ชนิด มาใช้ในการทดสอบประสิทธิภาพในครั้งนี้ เนื่องจากเป็นสมุนไพรที่เป็นที่รู้จักของประชาชนอย่างแพร่หลาย และสามารถนำมาใช้ในการบำบัดรักษาโรคที่พบได้ทั่ว ๆ ไป ซึ่งประชาชนสามารถนำองค์ความรู้ที่สกัดได้จากเหมืองความรู้เชิงความหมายไปใช้ได้ในชีวิตประจำวัน จากนั้นนำผลลัพธ์ที่ได้มาคำนวณหาค่าประสิทธิภาพ โดยแทนสูตรของรายการผลลัพธ์ ดังนี้ A คือ จำนวนข้อมูลสมุนไพรที่สืบค้นได้และถูกต้องตามความสนใจ B คือ จำนวนข้อมูลสมุนไพรที่ถูกต้องตามความสนใจแต่ไม่ถูกสืบค้น และ C คือ จำนวนข้อมูลสมุนไพรที่สืบค้นได้ แต่ไม่ถูกต้องตามความสนใจ (Miao and Zhu, 2021) จากนั้นนำค่าต่าง ๆ เหล่านี้ มาคำนวณหาค่าต่าง ๆ เพื่อสรุปผลคุณภาพในการสกัดความรู้ใน 3 ส่วน ดังนี้

3.1.1 ค่าความแม่นยำ (Precision) คือ การหาค่าความแม่นยำของโมเดล โดยทำการพิจารณาแยกทีละคลาส ดังสมการที่ 11

$$\text{Precision} = A/(A+C) \times 100 \quad (11)$$

3.1.2 ค่าความระลึก (Recall) คือ การวัดค่าความถูกต้องของโมเดล โดยทำการพิจารณาแยกทีละคลาส ดังสมการที่ 12

$$\text{Recall} = A/(A+B) \times 100 \quad (12)$$

3.1.3 ค่าอัตราความถูกต้องในภาพรวม (F-measure) คือ ความสามารถในการทำนายของโมเดลจากการเฉลี่ยค่า Precision และ Recall ดังสมการที่ 13

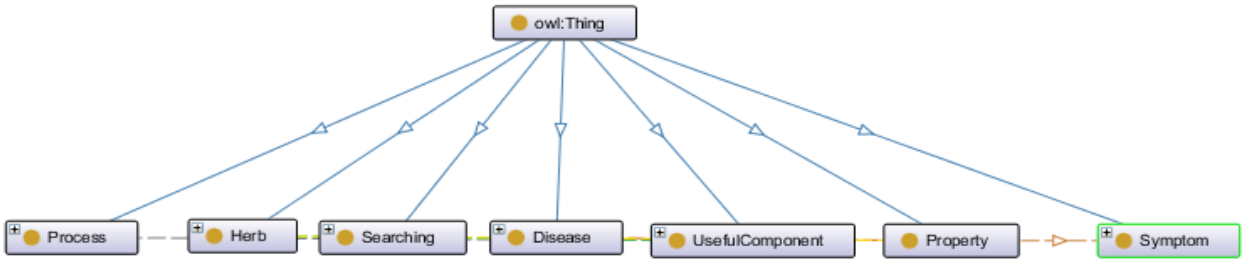
$$F\text{-measure} = 2 \times (Precision \times Recall) / (Precision + Recall) \quad (13)$$

3.2 การประเมินประสิทธิภาพในการทำนายเพื่อจำแนกชื่อสมุนไพร

จากข้อมูลคุณลักษณะเชิงความหมายดังที่นำเสนอในข้อ 2 เพื่อทำนายคลาสคำตอบชื่อสมุนไพรด้วยการเรียนรู้ของเครื่อง โดยทำการทดสอบใน 4 อัลกอริทึม ได้แก่ Neural Network, SVM, KNN และ Decision Tree โดยกระบวนการฝึกสอนและทดสอบความถูกต้องในการทำนาย โดยใช้วิธีการแบ่งข้อมูล (Split Test) ในอัตรา 70:30 (Pongsanguan *et al.*, 2018; Ibrahim, 2022) จากตัวอย่างข้อมูลในเหมืองความรู้ที่สืบค้นบนเว็บไซต์ 100 แหล่ง จำนวนทั้งหมด 300 ตัวอย่าง จาก 3 คลาสคำตอบ ได้แก่ ไพล (Zingiber montanum) ย่านาง (Tiliacora triandra) และฟ้าทะลายโจร (Andrographis paniculata) โดยได้ข้อมูลที่ใช้ฝึกสอน จำนวน 210 ตัวอย่าง และข้อมูลทดสอบจำนวน 90 ตัวอย่าง จากนั้นทำการวัดค่าความถูกต้องของแบบจำลองโดยใช้ค่าความถูกต้องของข้อมูล (Accuracy)

ผลการวิจัยและวิจารณ์ผล

1. การสร้างออนโทโลยีและกฎในการสืบค้นความรู้เชิงความหมายสมุนไพรไทย



ภาพที่ 2 ฐานความรู้ออนโทโลยีในการสืบค้นความรู้เชิงความหมายสมุนไพร

ภาพที่ 2 ฐานความรู้ออนโทโลยีในการสืบค้นความรู้เชิงความหมายสมุนไพร แบ่งเป็น 2 ชั้นความรู้ และมีคลาสความรู้จำนวนทั้งหมด 8 คลาส ประกอบด้วย 1) คลาส Process อธิบายเกี่ยวกับวิธีการปรุงสมุนไพรเพื่อใช้ในการบำบัดรักษาโรค 2) คลาส Herb อธิบายเกี่ยวกับรายละเอียดต่าง ๆ ของสมุนไพร เช่น ชื่อสมุนไพรท้องถิ่น ชื่อสมุนไพรทางวิทยาศาสตร์ พื้นที่ ๆ พบ คำอธิบายรูปลักษณ์ของสมุนไพร และลิงค์เชื่อมโยงไปยังภาพถ่ายสมุนไพร เป็นต้น 3) คลาส Searching รวบรวมสถิติในการค้นหาข้อมูลสมุนไพรชนิดต่าง ๆ จากผู้ใช้งานผ่าน Google Search engine เพื่อใช้ระบุความสำคัญในการสกัดข้อมูลผ่านกฎที่สอดคล้องกับความต้องการรับรู้ข้อมูลของผู้ใช้งาน 4) คลาส Disease และ Symptom อธิบายเกี่ยวกับโรคและอาการที่สามารถใช้สมุนไพรบำบัดรักษาได้ เช่น โรคอีสุกอีใส โรคความดันโลหิตสูง และอาการเลือดกำเดาไหล เป็นต้น 5) คลาส UseFulComponent อธิบายเกี่ยวกับส่วนประกอบของ

สมุนไพรที่ใช้ประโยชน์ในการรักษาโรคและอาการต่าง ๆ ได้ (Chamnongsri, 2019) เช่น ย่านางมีสรรพคุณในการรักษาโรคอีสุกอีใส เป็นต้น และในส่วนของส่วนประกอบที่ใช้ประโยชน์ได้ จะมีการปรุง เช่น ย่านางมีส่วนประกอบของใบ และรากสามารถนำไปใช้ปรุงยาสมุนไพรได้ และ 6) คลาส Property อธิบายเกี่ยวกับสรรพคุณและผลข้างเคียงของสมุนไพรชนิดต่าง ๆ ที่ผู้ใช้ควรต้องศึกษาก่อนการใช้

โครงสร้างออนโทโลยีที่พัฒนาขึ้น มีความสอดคล้องกับโครงสร้างออนโทโลยีในงานวิจัยของ Chamnongsri (2019) ที่ให้ความหมายเกี่ยวกับชื่อสมุนไพรท้องถิ่น ชื่อทางวิทยาศาสตร์ สรรพคุณ และส่วนประกอบที่ใช้ประโยชน์ในการรักษาโรคและอาการต่าง ๆ ได้ แต่จะมีความแตกต่างตรงที่งานวิจัยในครั้งนี้ มีการสร้างคลาสความรู้เพิ่มเติมในการรวบรวมสถิติการค้นหาข้อมูลสมุนไพรจากผู้ใช้ เพื่อให้ระบบสามารถสกัดความรู้จากเหมืองบนเว็บ (Web Mining) ที่ตอบสนองต่อความต้องการของผู้ใช้บนพื้นฐานของค่าคะแนนความนิยมในการสืบค้น และงานวิจัยของ Tungkwampian *et al*. (2015) ได้สรุปว่า การประยุกต์ใช้ออนโทโลยี จะช่วยระบุความสัมพันธ์ในเชิงความหมายระหว่างแนวคิดในโดเมนความรู้ของสมุนไพรไทย ได้อย่างมีประสิทธิภาพ ดังนั้นออนโทโลยี จึงนับเป็นแนวคิดที่มีความเหมาะสมอย่างยิ่งในการนำมาใช้พัฒนาเครื่องมือสกัด หรือขุดเจาะความรู้ที่เกี่ยวข้องกันในเชิงความหมายที่อยู่กระจายในหลายแหล่งข้อมูล

จากภาพที่ 2 สามารถสร้างกฎ SWRL เพื่อเข้าถึงโครงสร้างความรู้เชิงความหมายสมุนไพรไทยบนออนโทโลยี ได้ดังนี้

```
Herb(?h) ^ UsefulComponent(?c) ^ Process(?pc) ^ Interdict(?ir ) ^ HasComp(?h, ?c)
^ IsProceeded(?c, ?pc) ^ Interdict(?h, ?ir) ^
Disease_name(?h, ?ds) ^ Symptom_name(?h, ?st) ^
RecommendedQuantity(?h, ?rq) -> sqwrl:select(?h, ?c, ?pc, ?ir, ?ds, ?st, ?rq) ^ sqwrl:orderBy(?c)
```

ภาพที่ 3 กฎ SWRL เพื่อเข้าถึงโครงสร้างความรู้เชิงความหมายสมุนไพรไทย

จากภาพที่ 3 กฎ SWRL ใช้ดึงโครงสร้างฐานความรู้ออนโทโลยีเพื่อระบุว่าสมุนไพรแต่ละชนิดมีสรรพคุณและวิธีการนำส่วนประกอบต่าง ๆ ไปปรุงใช้ประโยชน์ จากนั้นนำผลลัพธ์ที่ได้ส่งออกเป็นเป็นไฟล์ CSV จากนั้นส่งไฟล์เข้าสู่การประมวลผลเพื่อสกัดความรู้ในเหมืองความรู้เชิงความหมาย สร้างชุดข้อมูล และใช้ ML ในการทำนายผลลัพธ์ ตามขั้นตอนวิธีที่สร้างไว้ในขั้นตอนที่ 2 โดยสามารถอธิบายโครงสร้างกฎ SWRL ได้ดังนี้

1. Herb(?h) ^ UsefulComponent(?c) ^ Process(?pc) ^ Interdict(?ir) คือ การประกาศตัวแปร ?h เพื่อเข้าถึงข้อมูลสมุนไพร ?c เพื่อเข้าถึงข้อมูลส่วนประกอบของสมุนไพรที่นำมาใช้ประโยชน์ ?pc เพื่อเข้าถึงข้อมูลวิธีการปรุงสมุนไพร และ ?ir เพื่อเข้าถึงข้อมูลผลข้างเคียงของสมุนไพร

2. HasComp(?h, ?c) คือ การนิยามความหมายว่า สมุนไพรแต่ละชนิดมีส่วนประกอบที่นำมาใช้ประโยชน์ได้แตกต่างกันออกไป

3. IsProceeded(?c, ?pc) คือ การนิยามความหมายว่า ส่วนประกอบที่นำมาใช้ประโยชน์ของสมุนไพร มีวิธีการปรุงเพื่อนำมาใช้บำบัดรักษาโรคและอาการต่างๆ

4. Interdict (?h, ?ir) คือ การนิยามความหมายว่า สมุนไพรแต่ละชนิดมีผลข้างเคียงที่ผู้ใช้ควรระมัดระวังในการใช้

5. Disease_name(?h,?ds)^Symptom_name (?h, ?st) คือ การนิยามความหมายว่า สมุนไพรแต่ละชนิด สามารถใช้บำบัดรักษาโรคและอาการต่างๆ ได้แตกต่างกันออกไป

6. RecommendedQuantity(?h, ?rq) คือ การนิยามความหมายว่า สมุนไพรแต่ละชนิดถูกจะถูกแนะนำปริมาณการใช้หรือรับประทานในแต่ละวันและไม่ควรเกินกี่วัน เพื่อไม่ให้ส่งผลเสียต่อร่างกาย

7. sqwrl:select(?h, ?c, ?pc, ?ir, ?ds, ?st, ?rq) ^ sqwrl:orderBy(?c) คือ การแสดงผลลัพธ์ที่ได้จากการอนุมานด้วยกฎตามองค์ประกอบข้อ 1-6 ด้านบน โดยจะแสดงข้อมูลในฟิลด์ต่างๆ ได้แก่ ชื่อสมุนไพร ส่วนประกอบ วิธีปรุง ผลข้างเคียง โรคและอาการ และคำแนะนำเกี่ยวกับปริมาณการใช้หรือรับประทานในแต่ละวัน โดยจะแสดงผลลัพธ์เรียงตามกลุ่ม (orderBy) ของส่วนประกอบของสมุนไพร

2. การพัฒนากระบวนการเรียนรู้ของเครื่องบนเหมืองความรู้เชิงความหมายสมุนไพรไทย

สามารถนำเสนอขั้นตอนวิธีการประมวลผลด้วยซูโดโค้ด (Pseudo code) ดังนี้

```
Ontology = herb.owl
Rule = SWRL command
WordsfromWeb =  words on 100 herb knowledge  websites
Keywords  = SemanticProcess(Ontology, Rule)
Features, Value=  KnowledgeMiningProcess(Keywords, WordsfromWeb )
HerbDataset = CreateDataset(Features, Value)
HerbdatasetTrain, HerbdatasetTest  = TrainTest_Split(HerbDataset, 70, 30)
ML(HerbdatasetTrain, HerbdatasetTest)
SemanticProcess(Ontology, Rule):
        Results = Rule -> Ontology
        For  Keyword  in Results :
           i += 1
            Keywords[i]  =  Keyword
        Return Keywords
KnowledgeMiningProcess(Keywords, WordsfromWeb_TH) :
        WordsfromWeb = Eng(WordsfromWeb_TH)
        Sim = LVDistanct(Keywords, WordsfromWeb)
        if(Sim>=0.7) then
        Features  = WordsfromWeb
        Value = count(Features)
        Return  Features, Value
CreateDataset(Features, Value) :
        HerbDataset <- Add(Features, Value)
        Return HerbDataset
TrainTest_Split(HerbDataset, TrainingProportion, TestingProportion) :
        Train, Test  = train_test_split (HerbDataset, TrainingProportion,TestingProportion)
        Return Train, Test
ML(Train, Test) :
        NN(Train, Test)
        SVM(Train, Test)
        KNN(Train, Test)
        DT(Train, Test)
```

ภาพที่ 4 Pseudo code กระบวนการเรียนรู้ของเครื่องบนเหมืองความรู้เชิงความหมายสมุนไพรไทย

จากภาพที่ 4 Pseudo code แบ่งฟังก์ชันการประมวลผลออกเป็น 4 ส่วน ดังนี้

1. Semantic process ทำหน้าที่ในการใช้กฎ SWRL ดึงข้อมูลจากออนโทโลยี เพื่อให้ได้คำสำคัญจากคลาสความรู้เชิงความหมาย ซึ่งงานวิจัยของ Polenghi *et al.* (2022) ได้ใช้กฎ SWRL ดำเนินการอัตโนมัติกับออนโทโลยีในการให้เหตุผลขององค์ความรู้เชิงความหมาย งานวิจัยของ Aldana Martín *et al.* (2022) ได้ใช้ SWRL ในการระบุความสัมพันธ์ระหว่างแต่ละคลาสความรู้ เพื่อให้เกิดการเชื่อมโยงสาเหตุไปยังผลที่ตามมาขององค์ความรู้ที่สนใจได้อย่างมีประสิทธิภาพ และงานวิจัยของ Ruger *et al.* (2022) ได้ใช้ภาษา SWRL สำหรับการสร้างกฎการอนุมานแบบมีเงื่อนไข เพื่อดึงความรู้เชิงความหมายในออนโทโลยี แสดงเป็นคำแนะนำด้านการให้บริการข้อมูลงานวิจัยตามโดเมนความรู้ที่ผู้ใช้สนใจได้อย่างถูกต้องและคลอบคลุม

2. Knowledge mining process ทำหน้าที่ในการรับค่าคำสำคัญ เพื่อนำมาเปรียบเทียบค่าความคล้ายกับข้อความที่อยู่บนเว็บไซต์สมุนไพรไทย จำนวน 100 แหล่งโดยข้อมูล/ข้อความต่าง ๆ จะผ่านขั้นตอนการแปลข้อมูลจากภาษาไทยเป็นภาษาอังกฤษก่อนโดยใช้ไลบรารี่ Python googletrans แล้วทำการตัดข้อความในประโยคโดยใช้ไลบรารี่ Python wordsegment เพื่อนำข้อมูลมาเปรียบเทียบความคล้ายกัน ตามแนวคิด Levenshtein distance จากนั้นเลือกข้อความที่มีความคล้ายมากกว่าค่า ratio 0.7 นำมาสร้างคุณลักษณะ และกำหนดค่าของคุณลักษณะ (value) โดยการนับคุณลักษณะ (คำสำคัญ) ที่พบในองค์ความรู้บนเว็บไซต์ต่าง ๆ ซึ่งสอดคล้องกับงานวิจัยของ Ghosh *et al.* (2022) ได้ใช้อัลกอริทึม Levenshtein distance ในการสกัดระหว่าง 2 ข้อความที่มีอัตราความคล้ายกันในระดับ ratio ที่กำหนดไว้

3. Create Dataset โดย feature และ value จากการสืบค้นเหมืองข้อมูล (ส่วนที่ 2) ถูกส่งมาเพื่อสร้างชุดข้อมูลสมุนไพร ยกตัวอย่างเช่น การสร้างชุดข้อมูลจากเหมืองความรู้เกี่ยวกับคลาสไพล (Zingiber montanum) มีความหมายเกี่ยวกับการมีสรรพคุณรักษาอาการเลือดดำเดาไหล โดยใช้เหง้า (Rhizome/Rhizomes) มาทำการต้ม (Boil) คลาสใบย่านาง (Tiliacora triandra) ใช้ใบนำมาต้มเพื่อรักษาอาการโรคอีสุกอีใส (Chickenpox) และคลาสฟ้าทะลายโจร (Andrographis paniculata) ใช้ใบนำมาต้มเพื่อรักษาอาการเป็นไข้ตัวร้อน (Flu) โดยไม่ควรรับประทานเกิน 7 วัน (EatDays) และใช้ไม่เกิน 2 เม็ดต่อวัน (EatTablets)

4. Machine learning (ML) ทำการแบ่งข้อมูลเพื่อใช้ในกระบวนการฝึกสอนและทดสอบความถูกต้องในการทำนาย ในอัตรา 70:30 (Pongsanguan *et al.*, 2018; Ibrahim, 2022) จากนั้นส่งให้อัลกอริทึม Neural network, SVM, KNN และ Decision tree เพื่อประมวลผลและทดสอบประสิทธิภาพ ทั้งนี้ในการประยุกต์ใช้เทคนิคเหมืองข้อมูลเชิงความหมายร่วมกับการทำนายผลลัพธ์ด้วยการเรียนรู้ของเครื่องในงานวิจัยนี้ สามารถสร้างระบบการแนะนำสมุนไพรไทยเพื่อบำบัดรักษาอาการโรคบนข้อมูลแบบไร้โครงสร้างได้อย่างมีประสิทธิภาพ โดยมีความสอดคล้องกับงานวิจัยของ Prompukdee *et al.* (2017) ที่ได้นำเทคนิคเหมืองข้อมูลมาประยุกต์ใช้ในการแพทย์แผนไทย โดยสกัดชุดข้อมูลการรักษาอาการของผู้ป่วยที่แตกต่างกัน และใช้การค้นหากฎการเชื่อมโยง เพื่อนำไปใช้สร้างแบบจำลองด้วยอัลกอริทึมการเรียนรู้ของเครื่อง (ML : Machine learning) เพื่อใช้ในการทำนายเชิงจำแนกและจัดกลุ่มข้อมูลความรู้ ทั้งนี้สามารถนำเสนอโครงสร้างชุดข้อมูลที่สร้างได้จากการทำสกัดเหมืองความรู้เชิงความหมายสมุนไพรไทย ดังตารางที่ 2

ตารางที่ 2 ชุดข้อมูลที่สกัดได้จากเหมืองความรู้สมุนไพรไทย

Rhizome	Rhizomes	Boil	Noseblee d	leaf	Chickenpox	EatDays	EatTablets	flu	class
4	1	1	1	0	0	0	0	0	Zingiber montanum
0	0	3	2	0	1	0	0	0	Tiliacora triandra
0	0	1	0	1	0	7	2	1	Andrographis paniculata

IsProceeded IsProceeded Herb->HasProp-> Property-> TreatD-> Disease, ->TeatS->Symptom

3. ผลการประเมินประสิทธิภาพของกระบวนการเรียนรู้ของเครื่องบนเหมืองความรู้เชิงความหมายสมุนไพรไทย

แบ่งเป็น 2 ส่วน ดังนี้

3.1 ผลการประเมินประสิทธิภาพในการสกัดความรู้เชิงความหมาย สรุปได้ดังนี้

ตารางที่ 3 ประสิทธิภาพในการสกัด/ค้นคืนความรู้เชิงความหมายสมุนไพรไทยแหล่งความรู้บนเว็บไซต์ 100 แหล่ง

รายการทดสอบ	**ผลลัพธ์ในการสกัด/ค้นคืนความรู้ด้วยค่าเป้าหมายที่กำหนด**		
	จำนวนข้อมูลสมุนไพรที่สืบค้นได้และถูกต้องตามความสนใจ (A)	**จำนวนข้อมูลสมุนไพรที่ถูกต้องตามความสนใจแต่ไม่ถูกสืบค้น (B)**	**จำนวนข้อมูลสมุนไพรที่สืบค้นได้ แต่ไม่ถูกต้องตามความสนใจ (C)**
1. ไพล (Zingiber montanum) ค่าเป้าหมาย ได้แก่ เหง้า (rhizome), ต้ม (boil), ฝน (sting), กำเดาไหล (nose bleed)	300 (rhizome, boil, nose bleed)	95 (sting)	0
2. ย่านาง (Tiliacora triandra) ค่าเป้าหมาย ได้แก่ ใบ (leaf), ต้ม (boil), อีสุกอีใส (chickenpox)	300 (leaf, boil, chickenpox)	0	0

ตารางที่ 3 (ต่อ)

รายการทดสอบ	ผลลัพธ์ในการสกัด/ค้นคืนความรู้ด้วยค่าเป้าหมายที่กำหนด		
	จำนวนข้อมูลสมุนไพรที่สืบค้นได้และถูกต้องตามความสนใจ (A)	**จำนวนข้อมูลสมุนไพรที่ถูกต้องตามความสนใจแต่ไม่ถูกสืบค้น (B)**	**จำนวนข้อมูลสมุนไพรที่สืบค้นได้ แต่ไม่ถูกต้องตามความสนใจ (C)**
3. ฟ้าทะลายโจร (Andrographis paniculata) ค่าเป้าหมาย ได้แก่ ใบ (leaf), ต้ม (boil), ไม่ควรทานเกิน 7 วัน (should not eat more 7 days), ไข้หวัด (flu), 2 เม็ด (2 tablets)	300 (leaf, boil, flu)	70 (should not eat more 7 days, 2 tablets)	0
4. ส่วนประกอบของสมุนไพรที่ช่วยบำบัดโรคความดันโลหิตสูงได้ ค่าเป้าหมาย ได้แก่ ใบ (leaf), ย่านาง(Tiliacora triandra)	300 (leaf, Tiliacora triandra)	0	0
5. สมุนไพรที่นิยมสืบค้นและช่วยบำบัดโรคได้มากกว่า 1 โรค ค่าเป้าหมาย ได้แก่ ย่านาง (Tiliacora triandra), ใบ (leaf), จำนวนครั้งที่สืบค้นย่านาง(Searching), โรคความดันโลหิตสูง (High blood pressure), โรคอีสุกอีใส (Chickenpox)	300 (Tiliacora triandra, Leaf, Searching, High blood pressure, Chickenpox)	0	0
Mean	**300**	**33**	**0**
Precision	100%		
Recall	90.1%		
F-measure	94.8%		

จากตารางที่ 3 ประสิทธิภาพในการสกัด/ค้นคืนความรู้เชิงความหมายเกี่ยวกับสมุนไพรไทย โดยใช้ออนโทโลยี และกฎ SWRL ทำให้ได้ข้อมูล/คำสำคัญที่สามารถนำมาใช้สร้าง Feature และ Value ในชุดข้อมูลเชิงความหมายของสมุนไพรทั้ง 3 ชนิด เพื่อส่งไปประมวลผลด้วยอัลกอรึทึม ML ได้อย่างมีประสิทธิภาพ โดยมีค่า F-measure เท่ากับ 94.8%

3.2 ผลการประเมินประสิทธิภาพในการทำนาย เพื่อจำแนกคลาสคำตอบชื่อสมุนไพรด้วย ML สรุปได้ดังนี้

ตารางที่ 4 ค่าร้อยละความถูกต้องของการทำนายเชิงจำแนกชื่อสมุนไพร จาก 4 อัลกอริทึม

รายการทดสอบ	**ค่าร้อยละความถูกต้องในการทำนาย**			
	Neural network	**SVM**	**KNN**	**Decision tree**
1. ไพล (Zingiber montanum)	91.1%	88.8%	83.3%	85.5%
2. ย่านาง (Tiliacora triandra)	90.0%	91.1%	83.3%	85.5%
3. ฟ้าทะลายโจร (Andrographis paniculata)	88.9%	87.8%	80.0%	82.2%
Average percentage	**90.0%**	**89.2%**	**82.2%**	**84.4%**
	(1)	**(2)**	**(4)**	**(3)**

จากตารางที่ 4 ประสิทธิภาพในการทำนายเพื่อจำแนกคลาสคำตอบสมุนไพร 3 ชนิด ด้วย Neural network ได้ค่าความถูกต้อง (Accuracy) สูงที่สุดเท่ากับ 90.0% รองลงมาคือ SVM ค่าความถูกต้องเท่ากับ 89.2% Decision tree ค่าความถูกต้อง เท่ากับ 84.4% และ KNN ค่าความถูกต้องเท่ากับ 82.2% ตามลำดับ

สรุป

ออนโทโลยีสมุนไพรไทย แบ่งเป็น 2 ชั้นความรู้ และมีคลาสความรู้จำนวนทั้งหมด 8 คลาส และในงานวิจัยนี้ได้สร้างกฎ SWRL เพื่อดึงโครงสร้างความรู้ที่ช่วยระบุรูปแบบ (Pattern) ข้อมูลที่คลาสคำตอบชนิดสมุนไพร 3 ชนิด ได้แก่ ไพล ย่านาง และฟ้าทะลายโจร ได้อย่างเป็นระบบ และสอดคล้องกับขอบเขตความสนใจที่แท้จริง จากนั้นโครงสร้างเชิงความหมาย จะถูกส่งเข้าสู่กระบวนการประมวลผลเพื่อสกัดความรู้ในเหมืองความรู้เชิงความหมาย สร้างชุดข้อมูล และใช้ ML ในการทำนายผลลัพธ์ ซึ่งวิธีการบูรณาการออนโทโลยี และ ML เพื่อสร้างเหมืองความรู้เชิงความหมายสมุนไพรไทย ทั้งนี้ผลจากการบูรณาการเทคนิคที่โดดเด่นเข้าด้วยกัน ส่งผลให้ผลการประเมินประสิทธิภาพในการสกัด/ค้นคืนความรู้เชิงความหมายเกี่ยวกับสมุนไพรไทย ได้ค่า F-measure ในระดับที่น่าพอใจ โดยมีค่าเท่ากับ 87.9% ซึ่งผลจากความสามารถและความแม่นยำในระดับสูงของการสกัดและจำแนกผลลัพธ์บนเหมืองข้อมูลเชิงความหมายที่พัฒนาขึ้น จะช่วยให้นักพัฒนาสามารถใช้อัลกอริทึมที่คิดค้นขึ้นไปใช้สร้างเหมืองข้อมูลทางสุขภาพในด้านต่าง ๆ ได้อย่างมีประสิทธิภาพ และสามารถสร้างชุดข้อมูล (Dataset) เชิงความหมายที่มีคุณภาพขึ้นมาใช้กับงานต่าง ๆ ได้เอง

กิตติกรรมประกาศ

คณะผู้วิจัยขอขอบคุณมหาวิทยาลัยราชภัฏเพชรบูรณ์ที่สนับสนุนทุนวิจัยในครั้งนี้ รวมถึงผู้เชี่ยวชาญที่ให้ข้อมูลที่เป็นประโยชน์ในการวิจัยและพัฒนาเทคโนโลยีที่ทันสมัยและมีประสิทธิภาพ

เอกสารอ้างอิง

Aldana Martín, J.F., García Nieto, J., Roldán García, M.M. and Aldana Montes, J.F. 2022. Semantic modelling of Earth Observation remote sensing. **Expert Systems with Applications** 187(C): 115838.

Chamnongsri, N. 2019. Ontology Development for Thai Herbal Medicine Recorded in Ancient Documents. **Journal of Information Science** 37(4): 18-43. (in Thai)

Ghosh, S., Ghosh, A., Ghosh, B. and Roy, A. 2022. Plagiarism Detection in the Bengali Language: A Text Similarity-Based Approach**,** pp. 1-4. *In* **3RD International Conference on Engineering and Advancement in Technology (ICEAT 2022).** Lisbon, Portugal.

Hillen, J. 2019. Web scraping for food price research. **British Food Journal** 121(12): 3350-3361.

Ibrahim, A.F. 2022. Application of various machine learning techniques in predicting coal wettability for CO2 sequestration purpose. **International Journal of Coal Geology** 252(14): 103951.

Jordan, J. 2017. **Neural networks: training with backpropagation.** Available Source: https://www. jeremyjor dan. me/neural-networks-training, June 9, 2022.

Kliangkhlao, M., Kochakornjarupong, D. and Boonchoom, V. 2015. An Ontology Development for Personal Health Adaptation Related to Climate Change. **Thaksin University Journal** 18(1): 23-31. (in Thai)

Miao, J. and Zhu, W. 2021. Precision–recall curve (PRC) classification trees. **Evolutionary Intelligence** 15(3): 1545-1569.

Ministry of Public Health. 2008. **Notification of Ministry of Public Health 2008 Title: Management plan to protect herbs in Phu Pha Kood Conservation Area Mukdahan Province According to the Protection and Promotion of Thai Traditional Medicine Knowledge Act the Protection and Promotion of Thai Traditional Medicine Wisdom Act 1999**. Government Gazette vol.125, Special Part 42. (dated February 28, 2008). (in Thai)

Noy, N. and McGuinness, D. 2001. **Ontology development 101: A guide to creating your first ontology.** Stanford University, Stanford.

Po, D.K. 2020. Similarity Based Information Retrieval Using Levenshtein Distance Algorithm. **International Journal of Advances in Scientific Research and Engineering** 6(4): 6-17.

Polenghi, A., Roda, I., Macchi, M. and Pozzetti, A. 2022. An ontological modelling of multi-attribute criticality analysis to guide Prognostics

and Health Management program development. **Autonomous Intelligent Systems** 2(1): 1-16.

Pongsanguan, W., Thinsungnoen, T. and Thinsungnoen, M. 2018. Development of Model for Diabetes Mellitus Using Decision Tree Technique. **Journal of Science and Technology** 1(1): 1-8. (in Thai)

Prompukdee, C., Rungrattanaubol, J. and Na-udom, A. 2017. Applying Data Mining Techniques on Thai Traditional Medicine Treatment. **NU. International Journal of Science** 14(1): 24-36. (in Thai)

Ruger, Y., Kabmala, M. and Chansanam, W. 2022. Ontology Rules for Library Research Services Recommendation System. **International Journal of Information and Education Technology** 12(9): 831-839.

Sun, B., Du, J. and Gao, T. 2009. Study on the Improvement of K-Nearest-Neighbor Algorithm, pp. 390-393. *In* **International Conference on Artificial Intelligence and Computational Intelligence 2009.** Shanghai, China.

Tian, D., Li, M., Zhang, X., Han, S. and Shen, Y. 2023. Intelligent Question Answering Method for Construction Safety Hazard Knowledge Based on Deep Semantic Mining. **Automation in Construction** 145: 1-13.

Tungkwampian, W., Theerarungchaisri, A. and Buranarach, M. 2015. Development of Thai herbal medicine knowledge base using ontology technique. **The Thai Journal of Pharmaceutical Sciences** 39(3): 102-109.

Wang, S., Mathew, A., Chen, Y., Xi, L., Ma, L. and Lee, J. 2009. Empirical analysis of support vector machine ensemble classifiers. **Journal of Expert Systems with Applications** 36(3): 6466-6476.

Xiuyi, T. and Yuxia, G. 2018. Research on Application of Machine Learning in Data Mining. **Journal of Materials Science and Engineering** 6(392): 1-4.

Yaiprasert, C. and Yusakul, G. 2022. Artificial intelligence for target symptoms of Thai herbal medicine by web scraping. **International Journal of Data and Network Science** 6(3): 1013-1022. (in Thai)