

ผลการทดลองและการประเมินผล

การสกัดความรู้เกี่ยวกับสรรพคุณทางยาของพืชสมุนไพร

คลังข้อมูลที่ใช้ทดสอบแบบจำลองที่ได้เสนอเกี่ยวกับการสกัดความรู้เกี่ยวกับสรรพคุณทางยาของพืชสมุนไพรด้วยการใช้เทคนิคที่แตกต่างกันสองวิธี คือวิธีการประยุกต์ใช้ความรู้ทางภาษาศาสตร์ร่วมกับการเรียนรู้ของเครื่องด้วยการใช้เทคนิค NB (Naïve Bates Classifier) และวิธีการประยุกต์ใช้ความรู้ต่างๆทางภาษาศาสตร์จนถึงระดับบทความที่ใช้ทฤษฎีเซนเทอร์ริง (Centering Theory, CT) กับคลังข้อมูล(Corpus)เอกสารทางวิชาการเกี่ยวกับพืชสมุนไพร ที่ได้จากการดาวน์โหลดจากเว็บไซต์สองแหล่งที่แตกต่างกันคือ จากสำนักงานเภสัชกรอำเภอพระพุทธรบาท (<http://phraphutthabat.saraburi.doe.go.th/herbs.htm>) และจากโครงการอนุรักษ์พันธุพืชอันเนื่องมาจากพระราชดำริสมเด็จพระเทพรัตนราชสุดาฯสยามบรมราชกุมารี (<http://www.rspg.or.th/index.htm>) โดยมีลักษณะของพฤติกรรมคลังข้อมูลของโครงการอนุรักษ์พันธุพืชเป็น semi-structure ที่มีความเป็นระเบียบมากกว่าของสำนักงานเภสัชกรอำเภอพระพุทธรบาท ดังนั้นงานวิจัยนี้ได้พัฒนาวิธีการสกัดความรู้พร้อมวิธีการหาขอบเขต EDU ที่เกี่ยวกับสรรพคุณทางยาของพืชสมุนไพรด้วยเทคนิคที่แตกต่างกันสองวิธีเพื่อศึกษาพฤติกรรมข้อมูลทางภาษามีผลต่อเทคนิคทั้งสองหรือไม่ ตารางที่ 4 แสดงค่า Precision (สมการที่6), Recall (สมการที่7), และ ค่าความถูกต้องของการหาขอบเขตของสรรพคุณทางยาของพืชสมุนไพรโดยเทคนิค NB และ CT โดยวัดความถูกต้องบนพื้นฐานของ Max Win Voting จากผู้เชี่ยวชาญ 3 ท่าน

$$Precision = \frac{\text{\# of samples correctly extracted as R}}{\text{\# of all samples output as being R}}$$
 (6)

$$Recall = \frac{\text{\# of samples correctly extracted as R}}{\text{\# of all samples holding the target relation R}}$$
 (7)

R คือ ความสัมพันธ์ที่เกี่ยวกับการระบุ EDU เกี่ยวกับสรรพคุณทางยาของพืชสมุนไพร

ตารางที่4 แสดงค่าPrecision, Recall, และ ค่าความถูกต้องของการหาขอบเขตของสรรพคุณทางยาของพืชสมุนไพรโดยเทคนิค NB และ CT

คลังข้อมูล	Lexico Syntactic Pattern for Herbal Medicinal Property Identification		Correctness of Boundary Determination	
	precision	recall	NB	CT
สำนักงานเภสัชกรอำเภอพระพุทธรบาท (500 EDUs)	82	65	90	79
โครงการอนุรักษ์พันธุพืช (500 EDUs)	92	84	93	93

จากตารางที่ 5 แสดงให้เห็นว่า ณ ที่ 95%Confident การหาขอบเขตของสรรพคุณทางยาของพืชสมุนไพรโดยเทคนิค NB ให้ผลดีกว่าการใช้ CT สำหรับคลังข้อมูลที่มีความเป็น Semi-structure น้อย ส่วนคลังข้อมูลที่มีความเป็น Semi-structure มากจะไม่มี ความแตกต่างอย่าง เป็นนัยสำคัญในระหว่างเทคนิคทั้งสองนี้

ตารางที่5 แสดงค่า t-test ของค่าความถูกต้องของการหาขอบเขตของสรรพคุณทางยาของพืชสมุนไพร ระหว่างเทคนิคที่แตกต่างกันคือ NB กับ CT

คลังข้อมูล	Correctness of Boundary Determination		t-test
	NB	CT	
สำนักงานเกษตรอำเภอพระพุทธรบาท	90	79	2.15
โครงการอนุรักษ์พันธุพืชฯ	93	93	0.27

จากตารางที่6 แสดงให้เห็นว่า ณ ที่ 95%Confident การหาขอบเขตของสรรพคุณทางยาของพืชสมุนไพรสำหรับ คลังข้อมูลที่มีความแตกต่างกันในด้านความเป็น Semi-Structure พบว่าการใช้เทคนิค NB ให้ผลไม่มีความแตกต่าง กันอย่างมีนัยสำคัญ แต่จะให้ผลแตกต่างอย่างมีนัยสำคัญสำหรับการใช้เทคนิค CT นั้นหมายความว่าหากใช้เทคนิค CT จะต้องคำนึงถึงการเลือกใช้คลังข้อมูลที่เหมาะสม

ตารางที่6 แสดงค่า t-test ของค่าความถูกต้องของการหาขอบเขตของ สรรพคุณทางยาของพืชสมุนไพรระหว่างคลังข้อมูลที่แตกต่างกัน

คลังข้อมูล	Correctness of Boundary Determination	
	NB	CT
สำนักงานเกษตรอำเภอพระพุทธรบาท	90	79
โครงการอนุรักษ์พันธุพืชฯ	93	93
t-test	0.76	2.85

จากตารางที่7 แสดงให้เห็นว่า ณ ที่ 95%Confident การระบุ EDU ที่เกี่ยวกับสรรพคุณทางยาของพืชสมุนไพรด้วย วิธีการใช้แพทเทิร์นทางภาษาศาสตร์คือ Lexico Syntactic Pattern สำหรับคลังข้อมูลที่มีความแตกต่างกันในด้าน ความเป็น Semi-Structure พบว่าค่า Precision และ ค่าRecall ที่ได้มีความแตกต่างกันอย่างมีนัยสำคัญในระหว่าง คลังข้อมูลทั้งสองที่แตกต่างกันนั้น นั้นหมายความว่าพฤติกรรมของคลังข้อมูลมีผลต่อการ ใช้ Lexico Syntactic Pattern สำหรับการระบุ EDU ที่ต้องการ

ตารางที่7 แสดงค่า t-test ของค่า Precision และ Recall ระหว่างคลังข้อมูล(Corpus)ที่แตกต่างกัน

คลังข้อมูล	Lexico Syntactic Pattern for Herbal Medicinal Property Identification	
	precision	precision
สำนักงานเกษตรอำเภอพระพุทธรบาท	82	82
โครงการอนุรักษ์พันธุพืชฯ	92	92
t-test	2.1	3.08

การตอบคำถามชนิด “ถามอะไร” ประเภทลิสต์ และประเภทเอนติตี้เกี่ยวกับสรรพคุณทางยาของพืชสมุนไพร

พืชสมุนไพรที่สกัดได้มีประมาณ 22 ชนิด คือกระชาย กระเทียม กระเพรา ขี้เหล็ก ชุมเห็ดเทศ โหระพา ขิง ข่า ตะไคร้ มะขาม พริก มะกูด มะนาว ทับทิม มังคุด คำฝอย ว่านหางจระเข้ ฟ้าทะลายโจร พลู กระจวาน กระดังงา ขมิ้น เป็นต้น คำถามที่ใช้ทดสอบมีทั้งหมด 50 คำถาม ระบบสามารถตอบได้ถูกต้อง 36 คำถาม (72%) ที่ไม่ถูกต้องส่วนใหญ่เนื่องมาจากคำถามที่ซับซ้อน