

52363201 : สาขาวิชาสารสนเทศศาสตร์ทางสุขภาพ

คำสำคัญ : การจัดแบ่งประเภทเอกสารโดยอัตโนมัติ / การจัดแบ่งประเภทเอกสารชนิดหนึ่งหมวดหมู่

/ การจัดแบ่งประเภทเอกสารชนิดหลายหมวดหมู่ / เอกสารข้อมูลยา

ชนิษฐา ลือวิพันธ์ : การจัดแบ่งประเภทเอกสารข้อมูลยาโดยอัตโนมัติตามกลุ่มเภสัชวิทยาและการรักษา. อาจารย์ที่ปรึกษาวิทยานิพนธ์ : ภก.รศ.ดร. วีรยุทธ์ เลิศนที. 91 หน้า.

ปัจจุบันเอกสารข้อมูลยามีจำนวนเพิ่มขึ้นมาก การจัดแบ่งประเภทเอกสารโดยอัตโนมัติจึงมีบทบาทสำคัญในการจัดแบ่งประเภทเอกสารให้อยู่ในกลุ่มที่เหมาะสม เอกสารข้อมูลยาสามารถจัดแบ่งประเภทตามกลุ่มเภสัชวิทยาและการรักษาให้อยู่ตั้งแต่หนึ่งกลุ่มจนถึงหลายกลุ่ม ในการศึกษาครั้งนี้ นำเอกสารข้อมูลยามาจากเว็บไซต์ RxList และ DailyMed โดยใช้ตัวจัดแบ่งประเภท 3 วิธีคือ วิธีเบย์อย่างง่าย วิธีเพื่อนบ้านใกล้ที่สุด และวิธีแบบเซนทรอยด์ในการจัดแบ่งประเภทเอกสารชนิดหนึ่งกลุ่มและชนิดหลายกลุ่ม การจัดแบ่งประเภทแบบไบนารีและการจัดแบ่งประเภทแบบครึ่งละหลายกลุ่มถูกนำมาใช้ด้วย การจัดแบ่งประเภทเอกสารแบบ 2 ขั้นตอนถูกพัฒนาขึ้นสำหรับการจัดแบ่งประเภทเอกสารชนิดหลายกลุ่ม ในขั้นที่ 1 เอกสารข้อมูลยาจะถูกจัดแบ่งประเภทตามจำนวนกลุ่มที่เป็นไปได้ ผลจากขั้นที่ 1 จะถูกใช้ในการจัดแบ่งประเภทเอกสารข้อมูลยาตามกลุ่มเภสัชวิทยาและการรักษาของขั้นที่ 2 การจัดแบ่งประเภทแบบสมบูรณ์ของทั้ง 2 ขั้นตอนจะถูกใช้ในการประเมิน จากผลการทดลองวิธีเพื่อนบ้านใกล้ที่สุดเป็นตัวจัดแบ่งประเภทที่ดีที่สุดในการจัดแบ่งประเภทเอกสารข้อมูลยาตามจำนวนกลุ่ม ในขณะที่วิธีแบบเซนทรอยด์เป็นตัวจัดแบ่งประเภทที่ดีที่สุดในการจัดแบ่งประเภทเอกสารข้อมูลยาตามกลุ่มเภสัชวิทยาและการรักษา ซึ่งการจัดแบ่งประเภทแบบ 2 ขั้นตอนสามารถประยุกต์ใช้ในการจัดแบ่งประเภทเอกสารชนิดหลายกลุ่มได้ โดยจะมีประสิทธิภาพดีกว่าการจัดแบ่งประเภทเอกสารแบบดั้งเดิมสำหรับเอกสารชนิดหนึ่งกลุ่มและชนิดหลายกลุ่ม ดังเช่นการจัดแบ่งประเภทแบบไบนารี

ภาควิชาสารสนเทศศาสตร์ทางสุขภาพ

บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร

ลายมือชื่อนักศึกษา.....

ปีการศึกษา 2555

ลายมือชื่ออาจารย์ที่ปรึกษาวิทยานิพนธ์.....

52363201: MAJOR: HEALTH INFORMATICS

KEY WORD: TEXT CATEGORIZATION, SINGLE-LABEL CLASSIFICATION, MULTI-LABEL CLASSIFICATION, DRUG MONOGRAPH CLASSIFICATION

CHANISARA LUEVIPHAN: AUTOMATIC TEXT CATEGORIZATION FOR DRUG MONOGRAPH BASED ON PHARMACOLOGIC-THERAPEUTIC CLASSES. THESIS ADVISOR: ASSOC. PROF. VERAYUTH LERTNATTEE, Ph.D. 91 pp.

With an increasing number of documents for drug information, automatic categorization is an important task for organizing these documents into suitable classes. Drug monographs can be classified into one or multiple classes based on Pharmacologic-Therapeutic classes. In this research, a set of drug monographs was drawn from the RxList and the DailyMed websites. Three classifiers, naïve Bayes, k-nearest neighbor, and centroid-based were used in single-label and multi-label classification. Binary classification and multiclass classification were used. Two-step text classification was developed for multi-label classification. The first step, drug monographs were classified into the number of possible classes. The results from the first step were applied on the second step based on their pharmacologic-therapeutic classes. The complete classification of the two-step was used for evaluation. From the experimental results, k-nearest neighbor was the best classifier to categorize drug monographs into possible classes while centroid-based classifier was the best classifier to categorize drug monographs based on pharmacologic-therapeutic classes. In conclusion, two-step text classification could be applied for multi-label classification. It was better performance than the classical method for single-label and multi-label documents, i.e., binary classification.

Department of Health-Related Informatics Graduate School, Silpakorn University

Student's signature..... Academic Year 2012

Thesis Advisor's signature.....

กิตติกรรมประกาศ

การจัดแบ่งประเภทเอกสารข้อมูลยาโดยอัตโนมัติตามกลุ่มเภสัชวิทยาและการรักษา ได้รับการสนับสนุนจาก 2 ทุน คือทุนวิจัยในโครงการวิจัยเรื่อง “การพัฒนาตัวสืบค้นข้อมูลยาผ่านเครือข่ายอินเทอร์เน็ตสำหรับบุคลากรสาธารณสุขและบุคคลทั่วไป” SURDI 52/01/23 และ SURDI 53/01/12 จากสถาบันวิจัยและพัฒนา มหาวิทยาลัยศิลปากร และทุนอุดหนุนการทำวิจัย จากงบประมาณแผ่นดิน (หมวดเงินอุดหนุนทั่วไป) และเงินงบประมาณโครงการพิเศษของบัณฑิตวิทยาลัย (สมทบ) ประจำปีงบประมาณ 2554 จึงทำให้เกิดโครงการวิทยานิพนธ์นี้ ข้าพเจ้าขอขอบพระคุณผู้เกี่ยวข้องในหน่วยงานดังกล่าวทุกท่านที่ให้การสนับสนุน ประสานงานในทุกๆ ด้านจนกระทั่งการศึกษาในครั้งนี้สำเร็จลงด้วยดี

ข้าพเจ้าขอกราบขอบพระคุณ ภก.รศ.ดร. วีรยุทธ์ เลิศนที อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ได้กรุณาให้โอกาส ให้คำแนะนำ ให้กำลังใจ ช่วยเหลือตลอดจนแก้ไขข้อบกพร่องต่างๆ ในการทำวิทยานิพนธ์จนสำเร็จลงได้ ขอขอบพระคุณ ภก.รศ.ดร. สรวง รุ่งประกายพรรณ ประธานกรรมการตรวจสอบวิทยานิพนธ์ และดร. อธิธิพันธ์ เมธเศรษฐ์ กรรมการตรวจสอบวิทยานิพนธ์ที่กรุณาให้คำแนะนำตรวจทานแก้ไขวิทยานิพนธ์อย่างสมบูรณ์

สุดท้ายนี้ต้องขอบพระคุณคุณแม่ น้องชาย คุณป้า พี่ๆ น้องๆ และเพื่อนๆ ทุกคนที่เป็นแรงผลักดันและเป็นกำลังใจให้สามารถทำวิทยานิพนธ์นี้สำเร็จลุล่วงไปด้วยดี