

56363202 : สาขาวิชาสารสนเทศศาสตร์ทางสุขภาพ

คำสำคัญ : ชื่อเครื่องยา / ยาแผนโบราณ / การเปรียบเทียบสายอักขระ / อัลกอริธึมเลเวนชเตย์ / อัลกอริธึมซิมิลาร์เท็กซ์ / เอนแซมเบลเทคนิค

ดวงกมล ภักดีสัตยพงศ์ : การใช้วิธีการเปรียบเทียบสายอักขระในการแก้ไขและสร้างมาตรฐานการระบุชื่อเครื่องยาในสูตรตำรับยาแผนโบราณ. อาจารย์ที่ปรึกษาวิทยานิพนธ์ : ภก.รศ. ดร.วีรยุทธ์ เลิศนที. 157 หน้า.

ชื่อเครื่องยาที่มีการขึ้นทะเบียนเป็นยาแผนโบราณตามที่ปรากฏในฐานข้อมูลสำนักงานคณะกรรมการอาหารและยามีนับจำนวน และความหลากหลายของข้อมูลมาก งานวิจัยนี้มีวัตถุประสงค์เพื่อ 1) วิเคราะห์รูปแบบชื่อเครื่องยาที่ปรากฏในฐานข้อมูลทะเบียนตำรับยา 2) นำวิธีการเปรียบเทียบสายอักขระไปใช้แก้ไขข้อผิดพลาด ของการระบุชื่อเครื่องยา รวมทั้งแนะนำแนวทางสร้างมาตรฐานการระบุชื่อเครื่องยาในสูตรตำรับยาแผนโบราณ โดยชื่อเครื่องยาที่นำมาศึกษาเป็นชื่อเครื่องยาที่ได้รับการขึ้นทะเบียนเป็นยาแผนโบราณตั้งแต่พฤษภาคม พ.ศ.2526 – กันยายน พ.ศ.2556 และปรากฏในฐานข้อมูลสำนักงานคณะกรรมการอาหารและยา จากผลการศึกษาพบชื่อเครื่องยาตามแบบอักขระ 3 แบบ คืออักขระไทย อักขระอังกฤษ และอักขระผสมระหว่างไทยและอังกฤษ หากจำแนกตามภาษาและสำเนียงที่ใช้ปรากฏว่ามีภาษาที่ใช้ทั้งสิ้น 5 ภาษาและสำเนียง คือภาษาไทย ภาษาอังกฤษ ภาษาละติน ภาษาจีนสำเนียงแต้จิ๋ว และภาษาจีนสำเนียงจิ้นกลาง เมื่อจำแนกตามลักษณะของชื่อเครื่องยานั้นอาจพบชื่อเครื่องยาในลักษณะที่มีชื่อแหล่งให้เครื่องยาซึ่งอาจระบุส่วนที่ใช้หรือไม่ได้ระบุส่วนที่ใช้ ชื่อวิทยาศาสตร์ที่มีหรือไม่มีส่วนของผู้ตั้งชื่อวิทยาศาสตร์ ชื่อเครื่องยาที่มีการระบุลักษณะเฉพาะ และ/หรือกรรมวิธีการผลิต นอกจากนั้นแล้วยังพบปัญหาจากชื่อเครื่องยาที่มีคำขยายซึ่งไม่ทำให้ชื่อเครื่องยามีความหมายผิดไปและชื่อเครื่องยาที่มีการสลับตำแหน่งของส่วนที่ใช้ได้ร่วมอยู่ด้วย ผลการศึกษาแสดงให้เห็นถึงความหลากหลายในเชิงรูปแบบที่มีเป็นจำนวนมากของชื่อเครื่องยาและการสะกดผิดของชื่อเครื่องยาในฐานข้อมูล ทำให้เกิดปัญหาในการค้นชื่อเครื่องยา จึงได้นำวิธีการเปรียบเทียบสายอักขระมาเพื่อช่วยในการหาชื่อเครื่องยาที่ถูกต้อง วิธีการเปรียบเทียบสายอักขระ 2 วิธีที่ได้รับความนิยมคือ อัลกอริธึมซิมิลาร์เท็กซ์ และอัลกอริธึมเลเวนชเตย์ได้ถูกนำมาใช้เพื่อแก้ไขปัญหา การประเมินผลการจับคู่ทำโดยการวัดค่าเอฟวัน (F1) ค่าความระลึก (recall) และ ค่าความเที่ยงตรง (precision) โดยผลจากการใช้อัลกอริธึมเดี่ยว ยังไม่ดีพอ ยังมีปัญหาบางอย่างที่ไม่ได้รับการแก้ไข เทคนิคเอนแซมเบลจึงถูกนำมาใช้เพื่อเพิ่มประสิทธิภาพของอัลกอริธึมเดี่ยว ผลจากการศึกษาพบว่าผลความต่างค่าเฉลี่ยของเอฟวัน (F1) ความระลึก (recall) และค่าความเที่ยงตรง (precision) ระหว่างการเปรียบเทียบสายอักขระด้วยอัลกอริธึมที่ได้รับการปรับปรุงด้วยเทคนิคเอนแซมเบล และอัลกอริธึมเดี่ยวที่ดีที่สุด คือร้อยละ 2.42, 3.05 และ 1.57 ตามลำดับ ซึ่งเป็นความแตกต่างอย่างมีนัยสำคัญที่ระดับความเชื่อมั่นเท่ากับ 95% โดยชื่อเครื่องยาที่มีปัญหาในส่วนที่มีคำขยายที่ไม่ได้ให้ความหมายและตำแหน่งของส่วนที่ใช้ จะได้รับการแก้ไขด้วยอัลกอริธึมที่ได้รับการปรับปรุงด้วยเทคนิคเอนแซมเบล เมื่อสามารถระบุชื่อเครื่องยาที่ถูกต้อง ชื่อเครื่องยาดังกล่าวจะถูกเปลี่ยนให้อยู่ในรูปแบบมาตรฐานได้ และทำให้รูปแบบชื่อเครื่องยาที่หลากหลายเป็นรูปแบบเดียวกัน กล่าวโดยสรุปได้ว่าวิธีการเปรียบเทียบสายอักขระสามารถปรับแก้ชื่อเครื่องยาให้ถูกต้องและเป็นแนวทางในการสร้างมาตรฐานชื่อเครื่องยาต่อไป

สาขาวิชาสารสนเทศศาสตร์ทางสุขภาพ

ลายมือชื่อนักศึกษา.....

ลายมือชื่ออาจารย์ที่ปรึกษาวิทยานิพนธ์

บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร

ปีการศึกษา 2557

56363202 : MAJOR : HEALTH INFORMATICS

KEY WORD : CRUDE DRUG NAME / TRADITIONAL MEDICINE / STRING MATCHING
TECHNIQUE / LEVENSHTINE ALGORITHM / SIMILAR TEXT
ALGORITHM / ENSEMBLE TECHNIQUE

DUANGKAMOL PAKDEESATTAYAPONG : USING STRING MATCHING
TECHNIQUES FOR CORRECTING AND STANDARDIZING NAMES OF CRUDE DRUGS IN
TRADITIONAL MEDICINE FORMULAE. THESIS ADVISOR : ASSOC.PROF.VEERAYUTH
LERTNATTEE, Ph.D. 157 pp.

The large number of crude drug names is registered as traditional medicines in the database of Thai Food and Drug Administration. These crude drug names have various patterns. The aims of this research are 1) to analyze patterns of crude drug names as well as 2) to correct and to provide the way for standardizing crude drug names by string matching techniques. Crude drug names appeared in Thai FDA database during from May 1983 to September 2013, were applied in the research. The result of study found that three types of alphabets were used for representing crude drug names, i.e., Thai alphabet, English alphabet and the combination of Thai and English alphabets. Five languages including dialects were found to represent crude drug names, i.e., Thai, English, Latin, Chinese dialects of Chaozhou and Mandarin. Several patterns of crude drug names were shown, e.g., names of sources with or without their parts used and scientific names with or without author names. Some characteristics and/or manufacturing processes might be included. Moreover, problems from modifiers that have no significant meaning and the position of a part used in crude drug names were presented. The study showed the diversity in formats of crude drug names and misspellings of crude drug names found in the database. Therefore, Problems of finding a crude drug may be occurred. In order to match registered crude drug names with standard names in a database, two well-known string matching algorithms, i.e., similar text and Levenshtein were investigated. The F-measure (F1), recall and precision are measures used for evaluating the matching results. The results from single algorithms indicated that crude drug names in the test set were moderately matched with those of the standard set. To increase performance of these single algorithms, the ensemble algorithm for matching crude drug names was proposed. Differences of means from F1, recall and precision between the ensemble algorithm and the best from single algorithm were 2.42%, 3.05% and 1.57%, respectively that were significantly different to match crude drug names at 95% confidence level. Problems of crude drug names with worthless modifiers and the position of a part used can be improved by the ensemble algorithm. When correct names were matched, crude drug names could be set to a standard pattern. Various patterns were standardized to the single format. In conclusion, string matching techniques are useful for correcting and providing the way to standardize crude drug names.

Program of Health Informatics
Student's signature
Thesis Advisor's signature

Graduate School, Silpakorn University
Academic Year 2014