

บทคัดย่อ

รหัสโครงการ : MRG5080125

ชื่อโครงการ : การพัฒนาระบบค้นคืนข้อมูลบนอินเทอร์เน็ตของข้อมูลสมุนไพรไทย

ชื่อนักวิจัย : วีรยุทธ์ เลิศนที มหาวิทยาลัยศิลปากร

ธนารักษ์ ชีระมันคง สถาบันเทคโนโลยีนานาชาติสิรินธร มหาวิทยาลัยธรรมศาสตร์

E-mail Address : verayuth@email.pharm.su.ac.th, verayuths@hotmail.com

ระยะเวลาโครงการ : 1 ธันวาคม 2549 – 30 พฤศจิกายน 2552

สารสนเทศด้านสมุนไพรเป็นสารสนเทศที่เกี่ยวข้องกับพืชที่ใช้เป็นยา อินเทอร์เน็ตเป็นแหล่งค้นคว้าข้อมูลด้านสมุนไพรที่ดีอย่างหนึ่ง ตัวสืบค้นแบบทั่วไปเป็นเครื่องมือที่ได้รับความนิยมในการหาข้อมูลสมุนไพรบนอินเทอร์เน็ต อย่างไรก็ตาม เป็นการยากที่จะสืบค้นข้อมูลสมุนไพรให้ได้ข้อมูลที่มีความเกี่ยวข้องและน่าเชื่อถือ ปัญหาอย่างหนึ่งก็คือคำศัพท์ทางวิชาการที่ใช้ในการสืบค้นข้อมูลหน้าเว็บที่เกี่ยวข้องกับด้านแพทย์พื้นบ้านไทยนั้นนิยมแสดงด้วยภาษาไทย ซึ่งเป็นการยากที่จะหาข้อมูลที่มีประโยชน์และไม่ลำเอียง แม้จะเป็นผู้ที่ได้รับการฝึกสอนมาอย่างดี เพื่อแก้ปัญหานี้ งานวิจัยนี้จึงได้พัฒนาตัวสืบค้นสมุนไพรไทย (THSearch) โดยใช้กลวิธีสามประการ อันได้แก่ ๑) มีการใช้แหล่งข้อมูลที่มีประโยชน์รวมทั้งลิงค์ของจากแหล่งข้อมูล ที่ผ่านการประเมินผลด้วยเกณฑ์การประเมินที่จัดทำขึ้น ๒) การเพิ่มรายชื่อของสมุนไพรเข้าไปในพจนานุกรมเพื่อเพิ่มประสิทธิภาพในการตัดคำภาษาไทย และ ๓) การกำหนดชื่อท้องถิ่นสำหรับสมุนไพรหนึ่ง เพื่อให้เกิดการขยายส่วนคำถาม โดยชื่อท้องถิ่นนี้สามารถป้อนเข้าในระบบได้โดยตรง หรือรวบรวมมาจาก คุยเอิร์บ เว็บไซต์ที่รวบรวมชื่อท้องถิ่นของพืชสมุนไพร โดยอาศัยหลักการด้านปัญญารวมหมู่ ซึ่งจะทำให้ได้คำศัพท์ที่น่าเชื่อถือและเปลี่ยนแปลงได้ตามกาลเวลา สำหรับตัวสืบค้นข้อมูลสมุนไพรไทยนี้พัฒนาบนระบบปฏิบัติการลินุกซ์ทะเล และมีการใช้ mnoGoSearch เป็นต้นแบบของตัวสืบค้น ผลการทดลองจากวิจัยนี้คือได้ต้นแบบตัวสืบค้นที่สามารถสืบค้นข้อมูลสมุนไพรไทย ที่ค้นคืนข้อมูลที่มีคุณค่าและน่าเชื่อถือโดยสามารถใช้ชื่อพืชสมุนไพรทั้งแบบทั่วไปและชื่อท้องถิ่น

คำหลัก : สารสนเทศทางสมุนไพร ตัวสืบค้น อินเทอร์เน็ต การจัดประเภทเอกสาร

Abstract

Project Code: MRG5080125

Project Title : A Development of the Thai-enable Herbal Search Engine

Investigator : Verayuth Lertnattee, Silpakorn University

Thanaruk Theeramunkong, Sirindhorn International Institute of Technology,
Thammasat Univeristy

E-mail Address : verayuth@email.pharm.su.ac.th, verayuths@hotmail.com,
thanaruk@siit.tu.ac.th

Project Period : 1 December 2006 – 30 November 2009

Herbal information is a special type of information dealing with medicinal herbs. The Internet is an excellent source of herbal information. General-purpose search engines are frequently used as a tool for finding herbal information on the Internet. However, it is hard to find relevance and reliable herbal information. A set of technical terms for searching herbal information is a major problem. Most Thai Web documents on traditional medicines were written in Thai. It is hard to find the useful and unbiased information even well-trained users. To solve these problems, Thai-enable search engine for herbal information (THSearch) is constructed. We use three strategies to improve our searching: (1) a list of useful websites and their links which are evaluated by a set of guidelines, (2) a Thai word segmentation is used and technical terms both in Thai about herbal information are added into the dictionary for improving the indexing process in Thai and (3) a set of synonyms for local names is built for helping users from a lot of keywords of the same term. These local names can be manually created or collected from KUIHerb, a collective intelligence system for herbal information. KUIHerb provides herbal vocabulary which is dynamically and confidentially applied for searching improvement on the Thai herbal search engine. Our system is implemented using Linux TLE (Thai Language Extension) as an operating system and mnoGoSearch as search engine software. The results show that Thai herbal search engine is very useful and easy to use. It returns useful and reliable information by using a query with a common name or local names of an herb.

Keywords : herbal information, search engine, Internet, text classification