



จลนิพนธ์

เรื่อง

ศึกษาปัญหาการสืบค้นข้อมูลพีชสมุนไพรรไทย

โดยใช้ตัวสืบค้นข้อมูลทางอินเทอร์เน็ต

โดย

นศภ. กชกร	แข่งคุ้ม	รหัส	0845001
นศภ. ชมกฤษ	สุคนธวารี	รหัส	0845032
นศภ. ศิริพร	น้อยแก้ว	รหัส	0845151
นศภ. ชนิศรา	ลือวิพันธ์	รหัส	0845188

จลนิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาบัณฑิต

ปีการศึกษา 2549

คณะศึกษาศาสตร์ มหาวิทยาลัยศิลปากร

ชื่อเรื่อง	:	ศึกษาปัญหาการสืบค้นข้อมูลทางอินเทอร์เน็ต	
ผู้จัดทำ	:	นางสาว กชกร	แข่งคุ้ม
		นางสาว ชมภุช	สุคนธวาริ
		นางสาว ศิริพร	น้อยแก้ว
		นางสาว ชนิสรา	ลือวิพันธ์
อาจารย์ที่ปรึกษา	:	ผศ.ดร. วิรุทธิ์	เลิศนที
		ผศ. สนิทพ	โฉมยา

ปีการศึกษา 2549

บทคัดย่อ

ปัจจุบันมีการเจริญเติบโตของข้อมูลข่าวสารทางอินเทอร์เน็ตอย่างรวดเร็วและในที่สุดต้องมีการสืบค้นและจัดระบบข้อมูลที่มีความเกี่ยวข้องนั้น การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อศึกษาปัญหาการสืบค้นข้อมูลของพีชสมุนไพรไทยโดยการใช้ตัวสืบค้นข้อมูลบนอินเทอร์เน็ตและเพื่อทราบแนวทางการสืบค้นข้อมูลของพีชสมุนไพรไทยอย่างมีประสิทธิภาพ โดยข้อมูลที่ใช้ศึกษา คือ ชื่อพีชสมุนไพรไทยจำนวน 41 ชนิด โดยเลือกจากหนังสือสมุนไพรไม้พื้นบ้านเล่มที่ 1-5 ซึ่งชื่อพีชสมุนไพรที่เลือกต้องเป็นชื่อที่รู้จักกันทั่วไปและเมื่อนำมาทดสอบโดยใช้ตัวสืบค้นข้อมูลบนอินเทอร์เน็ตแล้วมีปัญหาเกิดขึ้น งานวิจัยครั้งนี้จะใช้ตัวสืบค้นข้อมูลทางอินเทอร์เน็ตเป็นตัวสืบค้นข้อมูลสมุนไพรไทยจำนวน 3 ชนิด คือ www.google.co.th, www.yahoo.com, www.vivisimo.com โดยจะเก็บข้อมูลเฉพาะ 10 url แรกในการค้นหาที่ใช้ตัวสืบค้นข้อมูลบนอินเทอร์เน็ตใช้เวลาในการศึกษางานวิจัยตั้งแต่วันที่ 29 พ.ค.-31 ส.ค. 2549

จากการใช้ชื่อพีชสมุนไพรเป็นคำสำคัญซึ่งจะแบ่งคำสำคัญที่ใช้ในการสืบค้นเป็น 3 กลุ่ม คือ ชื่อพีชสมุนไพรที่นิยมใช้, ส่วนที่ใช้ของพีชสมุนไพรไทย, คำพ้องในการสืบค้นข้อมูลบนตัวสืบค้นข้อมูลทางอินเทอร์เน็ตทั้ง 3 ชนิด ดังที่ได้กล่าวมาแล้ว โดยใช้คำสำคัญทั้งหมด 123 คำ ได้ เว็บไซต์ทั้งหมดจำนวน 3,604 เว็บไซต์ แล้วนำเว็บไซต์ ที่ได้มาประเมินความเกี่ยวข้องของข้อมูลว่าเกี่ยวข้องกับพีชสมุนไพรไทยที่ใช้เป็นคำสำคัญในการสืบค้นอยู่ในระดับใดโดยมีค่าคะแนนความเกี่ยวข้องตั้งแต่ 0-5 หลังจากที่ได้ประเมินความเกี่ยวข้องของเว็บไซต์ แล้วนำข้อมูลจาก Access เข้าสู่ฐานข้อมูลใน Postgresql และต่อมาใช้ query ของ SQL

ในการสอบถามเพื่อให้ได้ข้อมูลตามที่ผู้วิจัยต้องการเมื่อได้ข้อมูลที่ต้องการแล้วนำข้อมูลที่ต้องการ
แล้วนำมาทดสอบสมมติฐาน 3 อย่างดังนี้

- (1) ตัวสืบค้นข้อมูลทางอินเทอร์เน็ต 3 ชนิดมีความแตกต่างกันหรือไม่
- (2) ทดสอบว่าคำสำคัญที่ใช้ในการสืบค้นที่แบ่งเป็น 3 กลุ่มนั้นมีความแตกต่างกันหรือไม่
- (3) จากการแบ่งประเภท Domain เป็น 9 กลุ่ม แล้วทดสอบทั้ง 9 กลุ่มว่าให้ผลแตกต่างกันหรือไม่
ซึ่งในการทดสอบความแตกต่างนี้ใช้สถิติ คือ Kruskal-Wallis Test ในการทดสอบความแตกต่าง
ผลที่ได้ปรากฏว่าชนิดของตัวสืบค้นข้อมูลบนอินเทอร์เน็ต ให้ผลในการค้นหาข้อมูลแตกต่างกัน
โดยตัวสืบค้นข้อมูลบนอินเทอร์เน็ตที่ให้ค่าเฉลี่ยของระดับคะแนนสูงสุด คือ www.google.co.th,
ส่วนประเภทของคำสำคัญที่ใช้ในการสืบค้นข้อมูลบนอินเทอร์เน็ตที่ให้ค่าเฉลี่ยของระดับคะแนนสูง
สุด คือ คำพ้อง และประเภทของ Domain ของ URL ที่ได้มาจากการสืบค้นข้อมูลบนอินเทอร์เน็ต
ให้ผลในการค้นหาข้อมูลแตกต่างกัน โดยประเภทของ Domain ของ URL ที่ให้ค่าเฉลี่ยของระดับ
คะแนนสูงสุดคือ .ac แต่ในความเป็นจริงแล้วควรใช้คำสำคัญในการสืบค้นหลายๆชนิดเพื่อ ให้ได้
ข้อมูลที่ต้องการได้อย่างครบถ้วน

Senior Project : The Study of Problems for Searching Internet Herbal
Information with the Thai Herb Name

By : Kodchakorn Khengkoom
Chumponuch Sukontavaree
Siriporn Noikeao
Chanisara Luewipan

Adviser : Assist.Prof. Veerayuth Lertnathee, Ph.D
Assist.Prof. Sinthop Shomya, Ph.D

Year : 2006

Abstract

With the fast growth of Internet information, there has been extreme needed to find and organize relevant Internet information. For this purpose, Internet search engines are constructed. General-purpose search engines such as Google or Yahoo are frequently used as a tool for finding information on Internet including herbal information. However, several problems may be occurred when we use these search engines. In this study, we address the problems of the Thai herb name as a keyword for search engines. To achieve our goal, a set of thirty Thai herbs with problems are randomly selected from the book of Samunpai Mai Peun Baan vol. 1- 5. Each herb has three search terms i.e. common name, a part of use and a popular synonym. We select the top ten documents which retrieve from the three search engines i.e. Google, Yahoo and Vivisimo. The level of relevance in Web documents is assigned by the level of score (0-5). After we finish the first data set, a set of ten herbs are randomly selected and used as a test set. The aims of the study are four folds. Firstly, we evaluate the relevant of content of Web documents and classify the problems from using Thai herb names. Secondly, we would like to compare performance of the three search engines for retrieving relevant documents.

Thirdly, the difference among three search terms is analyzed. Finally, the relevance of documents from several domains are compared. All comparisons are performed by Kruskal-Wallis test with the confidence level of 0.05.

From the results, three main types of problems are 1) Web content from owners who create Web documents 2) Thai word segmentation and 3) language encoding. There are significant differences among search engines, search terms and domains of the Web. For search engines, Google and Yahoo perform better than Vivisimo. The average score from popular synonym terms is the best. For domains, the documents from an academic domain (.ac.th) show to be the most relevant.