

การจัดหมวดหมู่เว็บเพจที่คล้ายคลึงกันโดยการรวบรวมกราฟแสดงความสัมพันธ์ของ  
คำศัพท์กับการโปรแกรมตรรกะเชิงอุปนัย

Similarity Web Pages Classification by Combining Relation Graphs to Inductive Logic  
Programming

ภริตา บุญปลูก, สุกรี สินธุภิญโญ

บทคัดย่อ

การเติบโตอย่างรวดเร็วของหน้าเว็บทำให้เกิดปัญหาในการค้นหาสารสนเทศในหน้าเว็บเหล่านั้น ซึ่งทางแก้ปัญหานี้ของปัญหานี้คือการจัดหน้าเว็บเหล่านี้ให้เป็นกลุ่มของความสนใจ วิธีการทั่วไปที่ใช้จัดกลุ่มหน้าเว็บคือการใช้โครงสร้างของหน้าเว็บเหล่านั้นเช่น ไฮเปอร์ลิงค์ คำสำคัญ แท็กเมตา เป็นต้น ซึ่งเป็นลักษณะสำคัญที่ใช้ในการจัดกลุ่มในบทความนี้ผู้วิจัยได้นำเสนอวิธีการที่ใช้การโปรแกรมตรรกะเชิงอุปนัย (Inductive Logic Programming) หรือ ILP เพื่อจำแนกคู่ของหน้าเว็บที่เหมือนกันโดยใช้ข้อความจากหน้าเว็บประกอบกับพจนานุกรมเชิงความหมายจากเวิร์ดเน็ต (WordNet) โดยนำมาเขียนใหม่เป็นกราฟความสัมพันธ์ของคำศัพท์ที่กำหนดขึ้นโดยผู้วิจัย ผลการทดลองพบว่าวิธีการนำเสนอให้อัตราความถูกต้องสูงกว่าวิธีการอื่นที่ทดสอบในการทดลองครั้งนี้

## **Abstract**

The rapid growth of web pages makes difficulties for searching for information stored in them. One solution of this problem is to organize those web pages into several groups of interest. The common method in organizing the pages is a use of structure of the pages, for example, hyperlinks, keywords, meta tag, etc., as important features to cluster them. In this paper, we propose a novel method which uses Inductive Logic Programming (ILP) to classify tuples of similar pages using the text in the pages combining to semantic dictionary from WordNet which are rewritten in our relation graph format. The results obtained from the experiments show that the proposed method yields the better accuracy rate than other methods run in our experiments.