



รายงานวิจัยฉบับสมบูรณ์

การสืบค้นข้อมูลด้วยการวัดความคล้ายตามแนวคิด
A Concept Similarity Based Information Retrieval

คณะนักวิจัย

ผู้ช่วยศาสตราจารย์ ดร. จิราภรณ์ เมืองประทับ

ที่ปรึกษางานวิจัย

รองศาสตราจารย์ ดร. วีระ บุญจริง

โครงการวิจัยนี้ได้รับทุนอุดหนุนการวิจัย
จากเงินรายได้มหาวิทยาลัยสงขลานครินทร์ ประเภทพัฒนานักวิจัย
ประจำปีงบประมาณ 2556

บทคัดย่อ

งานวิจัยนี้จะนำเสนอวิธีการวัดความคล้ายแบบใหม่ของคอนเซปต์ในคอนเซปต์แลตทิตที่สร้างขึ้นจากการวิเคราะห์แนวความคิดแบบฟอร์มมัล โดย วิธีที่นำเสนอ จะใช้การวัดของแต่ละคอนเซปต์หรือแต่ละโหนดซึ่งประกอบด้วยอินเทนและเอกเทน นอกจากนี้ยังใช้การนับความถี่ที่ปรากฏของเอกสารและคีย์เวิร์ดในแต่ละคอนเซปต์ เพื่อนำมาใช้ปรับปรุงวิธีการวัด และทำการพิสูจน์ทางคณิตศาสตร์ของวิธีที่นำเสนอเพื่อรับประกันว่ามีคุณสมบัติการเป็นมาตรวัดความคล้ายทั่วไปได้ นอกจากนี้ เพื่อแสดงให้เห็นถึงการใช้งานจึงประยุกต์ใช้กับการค้นคืนเอกสาร โดยจัดเก็บเอกสารโดยใช้การวิเคราะห์แนวความคิดแบบฟอร์มมัล และในส่วนนี้จะนำเสนออัลกอริทึมใหม่สำหรับการสร้างการจัดเก็บข้อมูลในรูปแบบของ คอนเซปต์แลตทิต และสืบค้นโดยใช้วิธีที่นำเสนอและเปรียบเทียบกับวิธีวัดความคล้ายของงานที่ผ่านมา และใช้ข้อมูลมาตรฐานซึ่งใช้ในการทดสอบความถูกต้องในการสืบค้นคือ Medline และ Cranfield ผลการทดลองพบว่าวิธีที่นำเสนอช่วยปรับปรุงประสิทธิภาพของความถูกต้องในการค้นคืนดีกว่าวิธี Jaccard index ซึ่งเป็นการวัดความคล้ายของคอนเซปต์ในคอนเซปต์แลตทิตในรูปแบบเดิม

ABSTRACT

This paper proposes a new concept similarity measure for a concept lattice in formal concept analysis (FCA). The proposed similarity employs concept approximation from both attribute extent and object intent based on appearance frequencies in formal concepts. Our approach can be applied to retrieve the relevance information because the obtained structure using FCA in concept form that are definable by conjunctions of properties. This measure is mathematically proved according to similarity measure satisfies a formal similarity metric. The performance of the proposed similarity measure is demonstrated in applications to an information retrieval system (IRS). Also, this paper provides the algorithm to build information structure for IRS. Two test collections, the Medline and the Cranfield collection, are used in the performance evaluation. The results show that the new concept similarity, using equally objects and attributes, improves the precision vs. recall curve, when compared to a traditional similarity measure.

Keywords: Formal concept analysis, concept similarity, information retrieval, similarity measure, concept approximation