

ชื่อ : นางวิไลพร เลิศมหาเกียรติ
 ชื่อวิทยานิพนธ์ : การค้นคืนสารสนเทศแบบสหความสัมพันธ์ตามหมวดหมู่
 ระบบทศนิยมดิวอี้
 สาขาวิชา : เทคโนโลยีสารสนเทศ
 มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
 อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก : ผู้ช่วยศาสตราจารย์ ดร.อนิราช มิ่งขวัญ
 ปีการศึกษา : 2553

บทคัดย่อ

การค้นคืนสารสนเทศในปัจจุบัน เน้นการนำคำสำคัญในส่วนคำค้นของผู้ใช้ไปเปรียบเทียบกับและค้นหาในเอกสาร เพื่อหาเอกสารที่เกี่ยวข้อง ซึ่งผลลัพธ์ปริมาณมากที่ได้นั้นแม้จะตรงกับคำค้นของผู้ใช้ แต่ยังคงมีผลลัพธ์จำนวนมากที่มีเนื้อหาไม่ตรงตามความต้องการของผู้ใช้ในงานวิจัยนี้ จึงมุ่งนำกรอบความรู้การจัดหมวดหมู่ระบบทศนิยมดิวอี้ (DDC) มาพัฒนาเป็นเทคนิคการวิเคราะห์หมวดหมู่แบบสหความสัมพันธ์ (DDC Multiple Relations) โดยวิเคราะห์ความสัมพันธ์ของคำสำคัญในเอกสาร และคำค้นของผู้ใช้กับกรอบความรู้ระบบทศนิยมดิวอี้ คำนวณน้ำหนักและหาสัดส่วนความสัมพันธ์ของเนื้อหาในแต่ละหมวด เพื่อนำมาแสดงเป็นสัดส่วนของความสัมพันธ์สำหรับใช้เปรียบเทียบความคล้ายคลึงระหว่างคำสำคัญในส่วนของเอกสาร และในส่วนคำค้นของผู้ใช้ ซึ่งการนำเทคนิคการวิเคราะห์หมวดหมู่แบบสหความสัมพันธ์ จะทำให้สามารถนำเสนอเอกสารที่มีรูปแบบสหความสัมพันธ์ของเนื้อหาที่ตรงกัน หรือมีความคล้ายคลึงกับรูปแบบสหความสัมพันธ์ของคำค้นของผู้ใช้ และสามารถใช้คำค้นในภาษาใด ภาษาหนึ่ง เพื่อค้นหาเอกสารจากฐานข้อมูลสากลที่มีการจัดหมวดหมู่ตามระบบทศนิยมดิวอี้ได้ ตลอดจนสามารถคัดเลือกและรวบรวมเนื้อหาเฉพาะส่วนที่เกี่ยวข้องกับคำค้นเพื่อนำเสนอต่อผู้ใช้ ระบบนี้ถูกนำมาทดสอบกับหนังสือจำนวน 1334 เล่ม และคำค้น 100 คำค้นจากสื่อการรายการคำค้นของผู้ใช้ และค้นคืนสารสนเทศจากคำร้อยละความสัมพันธ์ของหมวดหมู่ของหนังสือกับคำค้น โดยมีค่าความแม่นยำ (Precision) เท่ากับ 0.74

Name : Mrs.Wilaiporn Lertmahakiat
Thesis Title : Information Retrieval Using DDC Multiple Relations
Major Field : Information Technology
King Mongkut's University of Technology North Bangkok
Thesis Advisor : Assistant Professor Dr.Anirach Mingkhwan
Academic Year : 2010

Abstract

Nowadays, information retrieval focuses on using keyword matching by comparing with the keywords in the documents database in order to get the relevant results. This method returns a large number of results though some of them did not have any contents related to user needs. As for this research, Dewey Decimal Classification (DDC) was adopted as a basic to develop classification technique DDC-Multiple Relations (DDC-MR) which classifies documents contents and query keywords by using all of the relations. Pearson's Coefficient correlation was used to calculate those of the books and queries into a percentage of multiple relations and to compare with the similarities. By using those techniques, any related details could be represented not just the keywords. This system is able to search not only the results that do not match the query keywords, but it also returns them in different language by using Call Number as keyword. This system was tested with 1334 books and 100 queries from user logs, and retrieved by indicating the percentage of the Call Number and suggesting similar related contents, which the precision is equal to 0.74.