

เทคนิคการลดขนาดข้อมูลเพื่องานจัดกลุ่มข้อมูลขนาดใหญ่

กิตติศักดิ์ เกิดประสพ นิตยา เกิดประสพ และ นายธรรมศักดิ์ เขียรนิเวศน์
หน่วยปฏิบัติการวิจัยวิศวกรรมข้อมูลและการค้นหาความรู้,
สาขาวิชาวิศวกรรมคอมพิวเตอร์ สำนักวิชาวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีสุรนารี
111 ถนนมหาวิทยาลัย ต.สุรนารี อ.เมือง จ.นครราชสีมา 30000
โทรศัพท์ 0 4422 4349 โทรสาร 0 4422 4220 E-mail: kerdpras@ccs.sut.ac.th

บทคัดย่อ

การจัดกลุ่มข้อมูลโดยอัตโนมัติใช้เทคนิคการวัดความคล้ายคลึงกันของข้อมูลแต่ละคู่ เพื่อรวมกลุ่มข้อมูลที่คล้ายกันเข้าด้วยกัน อัลกอริทึมในการจัดกลุ่มข้อมูลโดยอัตโนมัติมีหลากหลาย อัลกอริทึม เวลาที่ใช้ของแต่ละอัลกอริทึมจะแปรผันไปตั้งแต่ $O(knl)$ ไปจนถึง $O(n^2 \log n)$ เมื่อ n คือจำนวนข้อมูล, k คือจำนวนกลุ่มของข้อมูล และ l คือ จำนวนรอบของการทำงานซ้ำ เนื้อที่หน่วยความจำที่ใช้จะแปรผันไปตามแต่ละอัลกอริทึม โดยทั่วไปจะเป็นตั้งแต่ $O(k)$ ไปจนถึง $O(n^2)$ ดังนั้นยิ่งจำนวนข้อมูลมีมากการจัดกลุ่มข้อมูลโดยอัตโนมัติจะยิ่งเสียเวลาและเสียเนื้อที่หน่วยความจำมาก เมื่อปริมาณข้อมูลมากถึงระดับเนื้อที่หน่วยความจำไม่เพียงพอต่อการใช้งานวิธีการแก้ปัญหาที่ได้ผลคือลดขนาดของข้อมูลก่อนที่จะจัดกลุ่มข้อมูล งานวิจัยนี้จึงมีวัตถุประสงค์ที่จะศึกษาเปรียบเทียบเทคนิคต่างๆ ที่ใช้ในการลดขนาดข้อมูลได้อย่างมีประสิทธิภาพและเหมาะสมกับงานจัดกลุ่มข้อมูล

ตีพิมพ์ใน: Proceedings of the Research Network Development of Higher Education Alliance in Nakhon Ratchasima, Suranaree University of Technology, Nakhon Ratchasima, Thailand, June 24, 2005. pp.66-67.