

บทคัดย่อ

ในงานวิจัยนี้ คณะผู้วิจัยได้พัฒนาระบบค้นหาสำหรับการบริการข้อมูลเชิงสุขภาพในอำเภอหัวหิน ระบบค้นหาได้อาศัยเทคโนโลยีการประมวลแบบขนานและกลุ่มเมฆ เชื่อมต่อกับข้อมูลท่องเที่ยวเชิงสุขภาพที่จัดเก็บในรูปแบบออนโทโลยี สนับสนุนการเทคโนโลยีเชิงความหมาย คณะผู้วิจัยได้นำระบบขึ้นไปวางบนกลุ่มเมฆในลักษณะงานประยุกต์โดยการสนับสนุนของ กสท. (ตะวันตก) ภายในแพลตฟอร์มของ IRIS

การประมวลแบบขนานที่ใช้ในงานวิจัยนี้มี 2 ด้านได้แก่ ด้านเสาะแสวงหา ดึงข้อมูล และเก็บข้อมูล โดยใช้เทคโนโลยี Java Concurrent (Java 1.7) ร่วมกับการพัฒนาโปรแกรมสำหรับการจัดทำกลไกการดึงรวบรวมข้อมูลหน้าเว็บ นอกจากนี้ได้ใช้เฟรมเวิร์ก MapReduce ในการพัฒนาส่วนของการรวมข้อมูล อีกด้านได้แก่การออกแบบการแทนค่าข้อมูลสำหรับการประมวลควิรีภายในสถาปัตยกรรม GPU ซึ่งเป็นแบบใช้หน่วยความจำร่วมกันแบบหลายแกน ในรายงานได้นำเสนอกลไกทั้งหมดตั้งแต่การรับข้อมูลจากผู้ใช้งานแปลงไปเป็นควิรีสำหรับการค้นหา การนำเข้าข้อมูลใน GPU สำหรับค้นหา และการนำผลลัพธ์การค้นหาจากการประมวลผลการค้นหาแบบขนานมาแสดง

การวัดผลด้านประสิทธิภาพมีในการค้นหาข้อมูลเชิงความหมายพบว่าใช้เวลาประมาณ 0.06 วินาทีในการค้นหาแต่ละควิรี และ Speedup ที่ได้จากการใช้เครื่องที่มีจำนวน 4 แกนในการดึงข้อมูลแบบขนานประมาณ 3.6 เท่า และการใช้ MapReduce ในการรวมไฟล์ข้อมูลขนาด 4 GB ประมาณ 6-7 วินาที สำหรับการค้นหาใน GPU ด้วยวิธีการแทนค่าข้อมูลที่น่าเสนอนั้น ทำให้สามารถนำเข้าข้อมูลได้มากกว่า และสามารถนำเข้าข้อมูลได้ถึง 18 ล้านเทอม ใช้เวลาค้นหา 34 มิลลิวินาที สำหรับด้านความพึงพอใจในการค้นหาเว็บเชิงความหมายจากผู้ใช้ 200 คน ได้คะแนนเฉลี่ยด้านความถูกต้องของข้อมูล และความพอใจในหน้าจอการใช้งาน 3.96/4.00

Abstract

In this project, Researchers develop the search engine for health information services in Hua Hin district. The search engine employs the parallel processing and cloud computing. The search engine connects with the health tourism data which is stored in a prototype ontology format. It supports the semantic web technology. Researchers particularly develop the search engine as a cloud application on IRIS sponsored by CAT Telecom Co. Ltd. (Thailand).

The parallel processing used in this research is two folds. First, it is for the purpose of data extraction and collection using Java Concurrent Package (1.7) together with the development of the parallel programs to implement the web crawler engine. Also, the MapReduce framework is used to combine the variety of data extraction. Secondly, the data representation for parallel query processing is designed. The parallel query processing is based on the shared-memory architecture which is on the GPU platform. The research reports the whole process from the user queries until the search results are derived and given back to the user using the representation and the parallel query.

Researchers measured the performance of the semantic query and find that it takes about 0.06 seconds to query the semantic web. Researchers measured the performance by using the parallel processing using the speedup, the speedup obtained by the parallel web crawler is about 3.6 for using 4-core computer and using MapReduce to merge the data about 4 GB takes only 6-7 seconds. The prototype of the searching in GPU can perform the search with larger data than using the search directly with the traditional RDF format. It can search 18 million terms using 34 milliseconds for searching. The user satisfaction of the semantic web development is on average 3.96/4.0 from 200 users.