



การใช้เทคนิคเหมืองข้อมูลเพื่อพัฒนาระบบการจัดเก็บและสืบค้น ภูมิปัญญาท้องถิ่นภาคใต้ในพื้นที่จังหวัดสุราษฎร์ธานี

ณรงค์ศักดิ์ จายางกูร^{1*} อโศก ศรีสวัสดิ์¹ และ จิตติมา ศิลประชาวงศ์¹

บทคัดย่อ

งานวิจัยชิ้นนี้ ได้รวบรวมบทความภูมิปัญญาท้องถิ่นภาคใต้ในจังหวัดสุราษฎร์ธานี เพื่อทำการเผยแพร่ผ่านเว็บไซต์ บันทึกสถิติการสืบค้นจากการค้นตรงผ่านเว็บไซต์ และการสืบค้นผ่านเว็บกูเกิ้ล มีการรวบรวมข้อมูลสถิติการสืบค้นเพื่อทำการวิเคราะห์หมวดหมู่ ด้วยเทคนิคเคมีนคลัสเตอร์ริง จากผลการวิเคราะห์ชี้ให้เห็นว่า ควรจะจัดหมวดหมู่ 15 กลุ่ม เพราะระดับค่า Sig เท่ากับ .000 ค่าผลรวมกำลังสองของค่าความคลาดเคลื่อนเท่ากับ 0.81 ใกล้เคียงค่า 0 ข้อมูลในแต่ละกลุ่มกระจุกตัวดี ถือว่ายอมรับได้ แต่แต่ละกลุ่มมีความสอดคล้องกันของเนื้อหากับความถี่ของคำค้นในแต่ละบทความ มีการนำบทความทั้งหมดวิเคราะห์หมวดหมู่ และวิเคราะห์ความสัมพันธ์ด้วยการวิเคราะห์ฟอมอลคอนเท็กซ์จากออบเจกต์ 15 ออบเจกต์ กับแอตทริบิว 245 แอตทริบิว ผลการวิเคราะห์ได้กฎความสัมพันธ์ 319 กฎ ที่มีความ

สัมพันธ์กับคำค้นของบทความและมีการซ้อนทับของคำค้นในเนื้อหาบทความใช้คำค้นร่วมกัน 45 บทความ แทบไม่ใช้ร่วมกัน 87 บทความ และใช้คำค้นร่วมกันมากที่สุด 6 บทความ ผลการวิเคราะห์หมวดหมู่ด้วยการวิเคราะห์ฟอมอลคอนเท็กซ์ 15 กลุ่ม ใช้เป็นหมวดหมู่ได้ โดยนำหมวดหมู่ และกฎความสัมพันธ์ ไปวิเคราะห์ออกแบบและพัฒนาระบบการจัดเก็บและสืบค้นประกอบด้วยระบบการจัดหมวดหมู่บทความ ระบบจัดทำบทความ ระบบการสร้างความสัมพันธ์ด้วยกำหนดเมทาเดตา และระบบสืบค้น ผลการทดสอบหาประสิทธิภาพของระบบได้ค่าความแม่นยำ (Precision) 98% ระดับดีมาก ค่าความระลึก (Recall) 81% ระดับดี

คำสำคัญ: ระบบการจัดเก็บและสืบค้น เหมืองข้อมูล วิเคราะห์หมวดหมู่ วิเคราะห์ความสัมพันธ์ ฟอมอลคอนเท็กซ์

¹ อาจารย์ สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏสุราษฎร์ธานี

* ผู้นิพนธ์ประสานงาน โทรศัพท์ 08-9199-0395 อีเมล: narongchatcc@hotmail.com



An Applied Data Mining Technique to Develop an Information Storage and Retrieval for Intelligence of Locality for the Southern Part of Thailand in Suratthani Province

Narongsak Chayangkoon^{1*} Asok Srisawat¹ and Jittima Silprachawong¹

Abstract

This study was conducted for the research and collection of the articles about local wisdom in Suratthani Province, Southern Thailand which were published on Google website. The searched articles were collected, analyzed, and categorized by K-Means Cluster Technique. The analysis results indicated that the data can be categorized into 15 groups due to the statistical significant value of 0.000 and the sum square error value of 0.81 which was near zero. The data of cluster were concentrated which was reliable. The clusters were congruent with the article contents and the frequency of keywords. All articles were analyzed based on the association rule and categorized by the formal context analysis resulting in 15 objects with 245 attributes. The analysis results generated into 319 association rules relating

to the keywords of articles. There were keyword overlaps in 45 articles with the edge count of 87 articles. The highest keyword overlaps were in 6 articles. The results of formal context analysis of 15 groups could be used for category analysis. The categories and the association rules could be used for design, analysis, and development of information storage and retrieval system. This consists of the category system, the article creating system, the metadata related to article system, and the retrieval system. The efficacy of the system and the precision ratio of the prototype was 98% or at the highest level, while the recall ratio of the prototype was 81% or at high level.

Keywords: Information Storage and Retrieval, Data Mining, Cluster Analysis, Association Rules, Formal Context

¹ Lecturer, Information School, Faculty of Science and Technology, Suratthani Rajabhat University.

* Corresponding Author, Tel. 08-9199-0395, E-mail: narongsak@sru.ac.th

1. บทนำ

จากแผนพัฒนาเศรษฐกิจและสังคมแห่งชาติ ฉบับที่ 11 มุ่งให้เกิดระบบฐานข้อมูลเพื่อการจัดเก็บ ภูมิปัญญาท้องถิ่น สามารถนำภูมิปัญญาท้องถิ่นมาใช้ให้เกิดประโยชน์ เพื่อพัฒนาและต่อยอดภูมิปัญญาท้องถิ่น ให้เป็นนวัตกรรมให้มีการเพิ่มมูลค่าทางเศรษฐกิจ [1] ซึ่งทิศทางการพัฒนาดังกล่าว สอดคล้องกับนโยบาย มหาวิทยาลัยราชภัฏสุราษฎร์ธานี ที่ได้กำหนดภารกิจ สำคัญ คือเป็นสถาบันการศึกษาเพื่อพัฒนาท้องถิ่นให้เกิด สิ่งสมดังอยู่ของวัฒนธรรมท้องถิ่น โดยเฉพาะภูมิปัญญา ท้องถิ่นภาคใต้ ที่เกิดจากคนไทยหลายชาติพันธุ์ที่อาศัย อยู่ร่วมกัน ได้แก่ คนมาเลย์ คนจีน และอินเดียฝ่ายใต้ แต่ภูมิปัญญาเหล่านี้ ยังคงอยู่กับปราชญ์ในพื้นที่ขาด การรวบรวมอย่างเป็นระบบทางคณะผู้วิจัยจึงได้พัฒนาระบบเพื่อจัดเก็บและสืบค้นภูมิปัญญาท้องถิ่นภาคใต้ โดยใช้เทคนิคการทำเหมืองข้อมูล (Data Mining) มาช่วย ในการจัดการวิเคราะห์หมวดหมู่และความสัมพันธ์ของ บทความภูมิปัญญาท้องถิ่นภาคใต้ เพื่อพัฒนาระบบจัดเก็บ และสืบค้นให้ง่ายต่อการสืบค้น

จากการลงพื้นที่เพื่อสำรวจภูมิปัญญาท้องถิ่นจาก ปราชญ์ท้องถิ่นในพื้นที่จังหวัดสุราษฎร์ธานี คณะผู้วิจัย ได้รวบรวมภูมิปัญญาท้องถิ่นสร้างเป็นบทความและ เผยแพร่สู่สังคม โดยใช้เว็บไซต์เป็นตัวกลางเพื่อให้เกิด การแลกเปลี่ยนความรู้และเผยแพร่ความรู้แก่ผู้สนใจ พร้อมจัดเก็บสถิติการเข้าใช้จากการค้นหาข้อมูลคำค้น ของผู้ใช้งาน จากการค้นหาตรงในเว็บไซต์และการสืบค้น ผ่านกูเกิ้ล (Google) ใช้เวลาในการจัดเก็บข้อมูล 1 ปี 7 เดือน และเริ่มศึกษาวิธีการทำเหมืองข้อมูลบนเว็บ (Web Mining) จากงานวิจัยของ Yang [2] ที่นำข้อมูลล็อกไฟล์ (Log File) มาหาความสัมพันธ์เพื่อพยากรณ์การใช้เว็บไซต์ และได้ประยุกต์วิธีการดังกล่าว โดยนำข้อมูลสถิติการสืบค้น มาวิเคราะห์สร้างเป็นหมวดหมู่สำหรับภูมิปัญญาท้องถิ่น ภาคใต้ ประกอบกับศึกษาจากเทคนิคการทำเหมืองข้อมูล ที่กล่าวไว้ในผลงานของ Han [3] คือการวิเคราะห์หมวดหมู่ (K-means Cluster Analysis) และการวิเคราะห์ความสัมพันธ์

(Association Rule) สำหรับการวิเคราะห์หมวดหมู่ ได้จำแนก กรณีการค้นหาข้อมูล โดยใช้กับกรณี 1 กลุ่มไอพีแอดเดรส (IP Address) คือ 1 กลุ่มผู้ใช้งาน และกำหนดจำนวนกลุ่ม ตัวแปรที่ใช้จะต้องเป็นตัวแปรเชิงปริมาณ เป็นสเกล อันตรภาค (Interval Scale) สเกลอัตราส่วน (Ratio Scale) ในส่วนของการวิเคราะห์ความสัมพันธ์ ได้สร้างกฎ ความสัมพันธ์ระหว่างคำค้นในแต่ละบทความ มีพื้นฐาน การศึกษาจากการค้นหาของกรณี 1 กลุ่มไอพีแอดเดรส สำหรับผู้ใช้สนใจค้นหาหลายบทความแบบต่อเนื่อง เกิดขึ้นร่วมกันหรือพร้อมกันในฐานข้อมูล หลังจากนั้น นำบทความทั้งหมดวิเคราะห์ ฟอแมลคอนเท็กซ์ (Formal Context) เพื่อทำการเปรียบเทียบความน่าเชื่อถือกับ หมวดหมู่ที่ได้จากการวิเคราะห์หมวดหมู่และการวิเคราะห์ ความสัมพันธ์ เพื่อทำการสรุปผลและนำผลการวิเคราะห์ ไปพัฒนาเป็นระบบจัดเก็บและสืบค้นภูมิปัญญาท้องถิ่น ภาคใต้ ตลอดจนทดสอบประสิทธิภาพระบบ โดยการหาค่า ความแม่นยำ (Precision) การหาค่าความระลึก (Recall) ซึ่งเป็นวิธีการทดสอบประสิทธิภาพของระบบจัดเก็บ และสืบค้น ได้ศึกษาจากงานวิจัยของ Sungsisri [4] Duangonnam [5] Suwannawat [6] Powers [7]

การวิเคราะห์หมวดหมู่และการวิเคราะห์ความสัมพันธ์ ได้ศึกษากระบวนการขั้นตอนการวิเคราะห์ข้อมูลจากงาน วิจัยของ Songsiri [8] เพื่อเปรียบเทียบความน่าเชื่อถือ กับการวิเคราะห์ฟอแมลคอนเท็กซ์ (Formal Context) ของบทความทั้งหมด ตามวิธีการที่กล่าวในผลงานของ Srisa-an [9] เพื่อนำผลการวิเคราะห์ข้อมูลไปพัฒนาระบบ การจัดเก็บและสืบค้น ระบบการจัดหมวดหมู่บทความ ระบบจัดทำบทความ ระบบพิจารณาความถี่ของคำ ระบบการสร้างความสัมพันธ์ด้วยกำหนดเมทาเดตา (Metadata) เพื่อสร้างเป็นตัวแทน (Agent) ของบทความ เนื่องจากภูมิปัญญาท้องถิ่นภาคใต้มีความสัมพันธ์กัน เช่น ระหว่างอาหาร และยารักษาโรค หรือนิทานพื้นบ้าน หนึ่งตระกูล เพลงกล่อมเด็ก ดังนั้น ในการออกแบบระบบ จัดเก็บและสืบค้น ได้ใช้คอนเซพแลททิซ (Concept Lattice) กำหนดการสืบค้นผ่านเมทาเดตา และเนื้อหา

บทความ (Text Search) เพื่อให้สามารถสืบค้นภูมิปัญญาท้องถิ่นที่มีเนื้อหาสัมพันธ์กับคำค้น ซึ่งอยู่ต่างหมวดหมู่ โดยตั้งชื่อระบบนี้ว่า ธนาคารสมองคลังปัญญาภูมิปัญญาท้องถิ่น

2. วิธีการดำเนินงานวิจัย

2.1 สืบค้นสืบค้นปราชญ์ท้องถิ่น และลงพื้นที่เก็บข้อมูลองค์ความรู้ เพื่อรวบรวมข้อมูลภูมิปัญญาท้องถิ่นภาคใต้ในพื้นที่จังหวัดสุราษฎร์ธานี

2.2 พัฒนาเว็บไซต์ให้สามารถจัดเก็บบทความจัดเก็บสถิติการสืบค้น การระบุคำค้นของผู้ใช้ และสถิติการเข้าถึงบทความ จากการค้นโดยตรงในเว็บไซต์ และการสืบค้นผ่านกูเกิ้ล

2.3 นำบทความภูมิปัญญาท้องถิ่นภาคใต้บรรจุลงเว็บไซต์ เพื่อเผยแพร่ให้กับเครือข่ายห้องสมุดมหาวิทยาลัยราชภัฏทั่วประเทศ เพื่อให้เป็นที่รู้จักและได้เข้าสืบค้นข้อมูลในงานประชุม Workshop on Uninet Network and Computer Application: Wunca ครั้งที่ 24 โครงการอบรมเมทาเดตา สำหรับการจัดการองค์ความรู้ วันที่ 28-29 กรกฎาคม 2554 ใช้เวลาจัดเก็บข้อมูล 1 ปี 7 เดือน

2.4 เตรียมข้อมูลเพื่อวิเคราะห์ ส่งออกข้อมูลสถิติการค้นหาข้อมูล และการเข้าใช้งาน จัดรูปแบบ สถิติการค้นหาข้อมูลของผู้ใช้ เพื่อให้อยู่ในรูปแบบของตัวเลขและตรรกศาสตร์เพื่อวิเคราะห์ทางสถิติด้วยโปรแกรม Weka

2.4.1 วิเคราะห์หมวดหมู่ภูมิปัญญาท้องถิ่นภาคใต้จากสถิติการสืบค้น ด้วยอัลกอริทึมเคมีน คลัสเตอร์ริง (K-means Clustering Algorithm) โดยประยุกต์กระบวนการวิเคราะห์มาจากงานวิจัยของ Songsiri [8] ดังนี้

1) Pre-processing ขั้นตอนการเตรียมข้อมูลให้เหมาะสมและให้อยู่ในรูปแบบที่สามารถนำไปใช้งานได้ โดยกรณี 1 กลุ่ม ไอพีแอดเดรส คือ 1 กลุ่มผู้ใช้งานในแต่ละบทความจะมีตัวแทนของบทความ คิดจากคำค้นที่มีความถี่สูงเป็นตัวแทนของเนื้อหาบทความการใช้งานของ 1 กลุ่ม ไอพีแอดเดรส ความถี่ของ 1 กลุ่ม ไอพีแอดเดรส แทนการใช้ของ 1 กลุ่มผู้ใช้ ความถี่ของการอ่านบทความ

คือการใช้งานตัวแทนของบทความ

2) ขั้นตอน Data Mining เป็นการพิจารณาระหว่าง K-means Clustering Algorithm กับ Hierarchical Cluster Algorithm เพื่อคัดเลือกอัลกอริทึมที่นำมาใช้ในการวิเคราะห์ โดยเลือกใช้ K-means Clustering Algorithm เพราะรายการข้อมูล (Data Items) ที่ปริมาณมาก มีจำนวน 7,481 รายการ จึงเหมาะสมกว่า Hierarchical Cluster Algorithm ประกอบกับรายการข้อมูล ในแต่ละรายการ ไม่ได้บอกว่าย่อยกลุ่มใด เป็นผลจากการจัดกลุ่มของ K-means Clustering Algorithm จากการหาค่าเฉลี่ยแบบ K-means Cluster การกำหนดจำนวนกลุ่ม K กลุ่ม และการกำหนดจุดศูนย์กลางเริ่มต้นจำนวน K จุด

3) Post-processing กำหนดจำนวนกลุ่ม แปลงข้อมูลจากไฟล์เอสคิวแอล (sql) ให้อยู่ในรูปแบบไฟล์ซีเอสวี (csv) เพื่อวิเคราะห์โดยโปรแกรม Weka โดยทดลองกำหนดกลุ่ม K เพื่อการวิเคราะห์ คือกรณีการแบ่งกลุ่มเป็น 10 หมวดหมู่ ค่า Mean Square Error ต่ำสุดเท่ากับ .123 และสูงสุดเท่ากับ .181 ค่า Sig เท่ากับ .000 สำหรับกรณีการแบ่งกลุ่ม เป็น 15 หมวดหมู่ ค่า Mean Square Error ต่ำสุดเท่ากับ .033 และสูงสุดเท่ากับ .143 ค่า Sig เท่ากับ .000 สำหรับกรณีการแบ่งกลุ่ม เป็น 20 หมวดหมู่ ค่า Mean Square Error ต่ำสุดเท่ากับ .000 และสูงสุดเท่ากับ .143 ค่า Sig เท่ากับ .020 ในส่วนของประเมินและวัดประสิทธิภาพจำนวนหมวดหมู่หลัก วัดจากค่าผลรวมของกำลังสองของความคลาดเคลื่อน (Sum of Square Error: SSE) ที่ใกล้เคียงค่า 0 ที่สุด และค่าระยะห่าง (Distance) ที่ต่างจากจุดศูนย์กลาง (Center) ของแต่ละคลัสเตอร์ เพื่อวัดการกระจุกตัวของกลุ่มตลอดจนแปรค่าทางสถิติเป็นจำนวนหมวดหมู่หลัก

2.4.2 การวิเคราะห์ความสัมพันธ์ จากการรวบรวมข้อมูลผลการค้นหาข้อมูลแบบต่อเนื่อง จำนวน 57 รายการ ที่สัมพันธ์กัน ไม่เพียงพอที่จะวิเคราะห์ความสัมพันธ์ มีปริมาณที่น้อยเกินไปสำหรับการทำเหมืองข้อมูล ส่งผลให้ขาดความน่าเชื่อถือของผลลัพธ์ จึงเลือกวิเคราะห์ความสัมพันธ์ โดยใช้การวิเคราะห์ฟอมอลคอนเท็กซ์

2.4.3 การวิเคราะห์ห้วงหมวดหมู่ นำผลข้อมูลที่ได้จากการวิเคราะห์มาเปรียบเทียบกับผลการวิเคราะห์บทความโดยใช้กระบวนการวิเคราะห์ฟอแมลคอนเท็กซ์

2.5 กระบวนการวิเคราะห์ฟอแมลคอนเท็กซ์ [9] มีขั้นตอนดังนี้

2.5.1 นำบทความที่รวบรวมจากการลงพื้นที่จัดเก็บในรูปแบบแฟ้มไฟล์นามสกุล txt เพื่อทำแยกคำ ซึ่งเป็นวิธีการหนึ่งที่ใช้ในงานวิจัยของ Pongsiripreeda [10] แต่แตกต่างกันที่งานวิจัยชิ้นนี้เลือกใช้อัลกอริทึม Swath [11] ในการแยกคำต่างๆ ในบทความ และบันทึกไฟล์สำหรับการวิเคราะห์ด้วยโปรแกรมนับคำ โดยไฟล์ที่ได้จากการตัดคำด้วยอัลกอริทึม Swath [11] จะแยกคำต่างๆ ในบทความมีการตัดคำแยกด้วยตัวคั่น “|” เพื่อให้สามารถนำไฟล์ที่ได้ไปนับคำด้วยโปรแกรม ซึ่งผลลัพธ์การตัดคำที่ได้จะแสดงในรูปที่ 1

2.5.2 การเขียนโปรแกรมเพื่อนับคำในบทความสำหรับการวิเคราะห์คำคั่นในบทความ และนับความถี่ของคำคั่น เพื่อจัดบทความเข้าสู่หมวดหมู่ที่ได้วิเคราะห์ไว้ดังแสดงไว้ในรูปที่ 2

2.5.3 การวิเคราะห์ฟอแมลคอนเท็กซ์ ได้กำหนดรูปแบบตารางประกอบด้วย 3 ส่วน คือออบเจกต์ (Object) แอตทริบิวต์ (Attribute) ความสัมพันธ์ (Relation) ซึ่งฟอแมลคอนเท็กซ์ ประกอบด้วย 2 ส่วน ได้แก่ ขอบเขต (Extent) หรือเซต (Set) ของวัตถุในตารางที่เกิดร่วมกันคือวัตถุในตารางและเจตนา (Intent) คือลักษณะแอตทริบิวต์ร่วมกันของวัตถุการนิยามตัวแปรต่างๆ ดังนี้

1) O = ออบเจกต์ A = แอตทริบิวต์ R = ความสัมพันธ์ กำหนดให้ $X \subseteq O$ และ $Y \subseteq A$ ดังสมการที่ 1

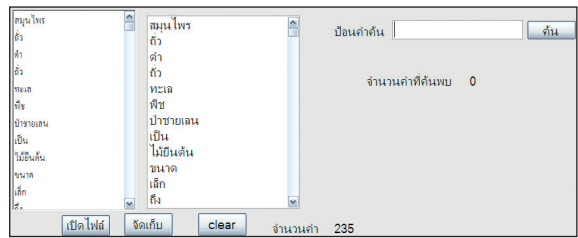
$$P(X) = \{a \in A \mid \forall o(X:(o, a)(R) \text{ Attribute } X\}$$

$$Q(X) = \{a \in A \mid \forall o(Y:(o, a)(R) \text{ Attribute } Y\} \quad (1)$$

อธิบายได้ว่า Context $K:=(O,A,R)$ ออบเจกต์ประกอบด้วย $O:=\{01 \text{ สมุนไพร} 02 \text{ ยารักษาโรค} 03 \text{ ผักและอาหารพื้นบ้าน} 04 \text{ การเลี้ยงเด็ก} 05 \text{ เพลงกล่อมเด็ก}$

สมุนไพร
ถั่ว|ดำ| ถั่ว|ทะเล| |พืช|ป่าชายเลน| |เป็น|ไม้ยืนต้น|

รูปที่ 1 การแยกคำด้วยอัลกอริทึม Swath



รูปที่ 2 โปรแกรมเพื่อนับคำในบทความ

	A0001	A0002	A0003	A0004	A0005	A0006	A0007
01 สมุนไพร	X	X	X	X	X	X	X
02 ยารักษาโรค	X	X	X	X	X	X	X
03 ผักและอาหารพื้นบ้าน	X	X					
04 การเลี้ยงเด็ก							X
05 เพลงกล่อมเด็ก							
06 นันทนาการ							
07 สุภาพบุรุษ							
08 ศาสนา							
09 ประเพณีวัฒนธรรม							
10 ศาสนา							
11 หัตถกรรม							
12 การดำรงชีวิต		X					
13 ดนตรี							
14 การละเล่น							
15 การแสดง							

รูปที่ 3 การวิเคราะห์ฟอแมลคอนเท็กซ์ (Formal Context) ด้วยโปรแกรม Conexp

06 นันทนาการ 07 สุภาพบุรุษ 08 ศาสนา 09 ประเพณีวัฒนธรรม 10 ศาสนา 11 หัตถกรรม 12 เครื่องมือเครื่องใช้การดำรงชีวิต 13 ดนตรี 14 การละเล่น 15 การแสดง เพื่อจัดรูปแบบข้อมูลเพื่อทำการวิเคราะห์ฟอแมลคอนเท็กซ์ระหว่างออบเจกต์ จำนวน 15 ออบเจกต์ กับแอตทริบิวต์ จำนวน 245 แอตทริบิวต์ บันทึกไฟล์ในรูปแบบซีเอสวี

2) ทำการวิเคราะห์ด้วยโปรแกรม Conexp ระหว่างออบเจกต์จำนวน 15 ออบเจกต์ กับแอตทริบิวต์จำนวน 245 แอตทริบิวต์ ดังรูปที่ 3

3) จากคอนเซ็ปต์ภายในบริบท (Context) จะได้คุณสมบัติ ดังสมการที่ 2

$$X \subseteq G: X \rightarrow X: = \{m \in M \mid g|m \text{ สำหรับ } g \in X\}$$
$$Y \subseteq M: Y \rightarrow Y: = \{g \in G \mid g|m \text{ สำหรับ } m \in X\} \quad (2)$$

4) แปลงฟอมอลคอนเท็กซ์ให้เป็นคอนเซพแลททิส

5) ขั้นตอนการสร้าง เริ่มสร้างโหนดแรก แอดทริบิวที่มีออบเจกต์ ร่วมกันมากที่สุดก่อน จึงสร้างโหนดที่มีออบเจกต์ร่วมมากที่สุด แต่ไม่มีแอดทริบิวร่วมกันของ Concept ใดๆ และสร้างโหนดที่มีออบเจกต์ ร่วมกัน แต่ไม่มีแอดทริบิวร่วมกันของ Concept สุดท้ายเอาแอดทริบิวตามลักษณะของโหนดแต่ละคู่มา Union กัน เพื่อให้ได้โหนดใหม่

2.6 สรุปผลหมวดหมู่ และนำหมวดหมู่ที่ได้จากการวิเคราะห์ ไปตรวจสอบโดยผู้ทรงคุณวุฒิด้านภูมิปัญญาท้องถิ่นภาคใต้ และสรุปความสัมพันธ์จากคอนเซพแลททิส สำหรับพัฒนาระบบจัดเก็บและสืบค้นภูมิปัญญาท้องถิ่นภาคใต้

2.7 จากคอนเซพแลททิส ออกแบบระบบฐานข้อมูลของระบบ ระบบผังการไหล (DFD) ระบบการจัดเก็บและสืบค้น ประกอบด้วย ระบบการจัดหมวดหมู่บทความ (Cluster Analysis) ระบบจัดทำบทความ ระบบพิจารณาความถี่ของคำ (Frequency Transaction) ระบบการสร้างความสัมพันธ์ และระบบการสืบค้น ออกแบบเชิงแนวคิด (Conceptual Design) เชิงตรรกะ (Logical Design) ออกแบบเชิงกายภาพ (Physical Design)

2.8 พัฒนาระบบจัดเก็บและสืบค้น ระบบการจัดหมวดหมู่บทความ โดยพิจารณาตามความถี่ของคำที่ปรากฏเพื่อสร้าง หมวดหมู่เพื่อทำการแบ่งกลุ่มบทความ พัฒนาระบบความสัมพันธ์ของบทความ ด้วยกำหนดเมทาเดตาให้กับบทความที่มีความเกี่ยวข้องกันของเนื้อหาและคำค้นบางส่วน พัฒนาระบบการสืบค้นเพื่อค้นหาคำค้นที่มีความเกี่ยวข้องกันจากคำค้นเดียวกัน พัฒนาระบบต้นแบบ (Prototype)

2.9 ทดสอบหาประสิทธิภาพของระบบ เป็นการประเมินค่า Recall และ Precision เป็นมาตรวัดการเรียกคืนและความถูกต้องแม่นยำวิธีการประเมินประสิทธิภาพ

ระบบจัดเก็บและสืบค้น ซึ่งเป็นวิธีการที่ใช้ในงานวิจัยของ Sungsisri [4] Duangonnam [5] Suwannawat [6] และ Powers [7]

2.9.1 การหาค่าความแม่นยำคำนวณจากสูตรที่ 3 [5]

$$\frac{\text{จำนวนข้อมูลที่ต้องการที่ค้นคืนได้}}{\text{จำนวนข้อมูลทั้งหมดที่ค้นคืนออกมาได้}} \quad (3)$$

2.9.2 การหาค่าความระลึก คำนวณจากสูตรที่ 4 [5]

$$\frac{\text{จำนวนข้อมูลที่ต้องการที่ค้นคืนได้}}{\text{จำนวนข้อมูลที่ต้องการทั้งหมดในฐานข้อมูล}} \quad (4)$$

3. ผลการวิจัย

3.1 จากการวิเคราะห์หมวดหมู่ ได้ผลการแบ่งกลุ่มข้อมูล (Cluster Analysis) การแบ่งหมวดหมู่จากคำค้นควรแบ่งเป็น 20 กลุ่ม ค่าผลรวมของกำลังสองของความคลาดเคลื่อน (Sum Square Error: SSE) ใกล้เคียงค่า 0 ที่สุด แต่เมื่อเทียบเคียงกับค่าความถี่ของคำค้นในบทความ แต่ปรากฏว่าผลการวิเคราะห์หมวดหมู่ไม่สอดคล้องกับบทความมีความขัดแย้งกัน เมื่อนำผลการแบ่งกลุ่ม เป็น 15 กลุ่ม เทียบเคียงกับค่าความถี่ของคำค้นในบทความแล้วข้อมูลมีความสอดคล้องกันจึงสรุปได้ว่า ควรจัดแบ่งหมวดหมู่ภูมิปัญญาท้องถิ่นภาคใต้เป็นจำนวน 15 หมวดหมู่ และดูค่าผลรวมของกำลังสองของความคลาดเคลื่อนใกล้เคียงค่า 0 ที่สุด เท่ากับ 0.81 ค่า Sig เท่ากับ .000 แสดงว่าข้อมูลในแต่ละกลุ่มกระจุกตัวดีถือว่ายอมรับได้ และเมื่อพิจารณาจากความถี่ของคำที่ปรากฏในบทความ จึงกำหนดเป็นหมวดหมู่ ได้แก่ 01 สมุนไพร 02 ยารักษาโรค 03 ผักและอาหารพื้นบ้าน 04 การเลี้ยงเด็ก 05 เพลงกล่อมเด็ก 06 นิทานพื้นบ้าน 07 สุภาษิต คำพังเพย 08 ศาสนา 09 ประเพณีวัฒนธรรม 10 คาถา 11 หัตถกรรม 12 เครื่องมือเครื่องใช้การดำรงชีวิต 13 ดนตรี 14 การละเล่น 15 การแสดง ทั้ง 15 หมวดหมู่ใช้เป็นหมวดหมู่เริ่มต้น ในการวิเคราะห์ค่า

ฟอมอลคอนเท็กซ์ของบทความทั้งหมด

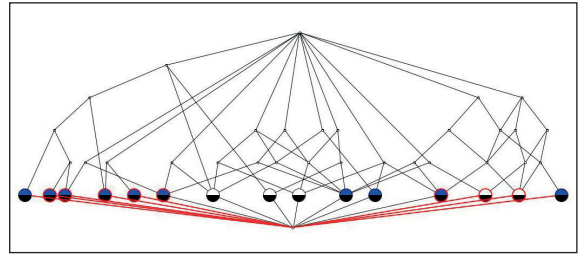
3.2 ผลการวิเคราะห์ฟอมอลคอนเท็กซ์ จากการวิเคราะห์ด้วยโปรแกรม Conexp ระหว่างออบเจกต์ 15 ออบเจกต์ กับแอตทริบิว 245 แอตทริบิว ได้ผลการวิเคราะห์ฟอมอลคอนเท็กซ์ คือการวิเคราะห์ความสัมพันธ์จากออบเจกต์ 15 ออบเจกต์กับแอตทริบิว 245 แอตทริบิว ได้ผลการวิเคราะห์ดังนี้

3.2.1 การแปลงคอนเซพแลททิส จะถูกนำไปประยุกต์ใช้กับการวิเคราะห์ความสัมพันธ์ และการแบ่งกลุ่มข้อมูล

จากรูปที่ 4 อธิบายได้ว่าคอนเซพแลททิสที่ได้ทั้งหมด 45 เคาท (Count) มีความสัมพันธ์ตามคอนเซพแลททิส จำนวน 45 (Node) บทความขอบคอนเซพแลททิส (Edge Count) จำนวน 87 (Node) คอนเซพแลททิสที่มากที่สุด 6 โหนด (Node) โหนดหลัก 15 โหนด

3.2.2 จากคอนเซพแลททิส มีการเชื่อมโยงความสัมพันธ์ของคำค้นบทความ มีการซ้อนทับของคำค้นในบทความ บทความใช้คำค้นร่วมกัน 45 บทความ ใช้คำค้นร่วมกันบางส่วนแทบไม่ใช้ร่วมกัน 87 บทความ ใช้คำค้นร่วมกันมากที่สุด 6 บทความ ปรากฏหมวดหมู่ใน 01 สมุนไพร บทความที่ 1-29, 02 ยารักษาโรค บทความที่ 1-11, 03 ผักและอาหารพื้นบ้าน 1-29, 04 การเลี้ยงเด็ก บทความที่ 3, 06 นิทานพื้นบ้าน บทความที่ 1-12, 07 สุภาษิต บทความที่ 1-12, 08 ศาสนา บทความที่ 12-15, 09 ประเพณีวัฒนธรรม บทความที่ 26-30, 11 หัตถกรรม บทความที่ 15-31, 12 เครื่องมือเครื่องใช้ การดำรงชีวิต บทความที่ 1-8, 14 การละเล่น หากพิจารณา จากบทความที่ 3 หมวดหมู่ 04 เรื่องแอกเวียน จากผลการตัดคำด้วยอัลกอริทึม Swath ก็ทราบได้ว่าคำนั้นได้ 2 คำ ปรากฏเด่นชัด “ชาวบ้าน” และ “ภูมิปัญญา” ในส่วนของ 87 บทความใช้คำค้นร่วมกันบางส่วนแทบไม่ใช้ร่วมกัน

3.2.3 โหนดหลัก มีจำนวน 15 โหนด เท่ากับผลการแบ่งกลุ่มข้อมูล ด้วยเทคนิค K-means Cluster จำนวน 15 หมวดหมู่ที่มีการวิเคราะห์ และเมื่อตรวจสอบ



รูปที่ 4 คอนเซพแลททิส (Concept Lattice)

```
9 < 1 > A0002 C0023 C0029 = [100%] => < 1 > A0013 A0024 A0027 A0028 A0029 C0001  
C0002 C0003 C0004 C0005 C0006 C0007 C0008 C0009 C0010 C0011 C0012 C0013 C0014  
C0015 C0016 C0017 C0018 C0019 C0020 C0021 C0022 C0024 C0025 C0026 C0027 C0028  
D0001 D0002 L0003;
```

รูปที่ 5 กฎความสัมพันธ์ของบทความ

โดยผู้ทรงคุณวุฒิด้านภูมิปัญญาท้องถิ่นภาคใต้ เห็นว่าเนื้อหาบทความมีความสอดคล้องและตรงกับหมวดหมู่

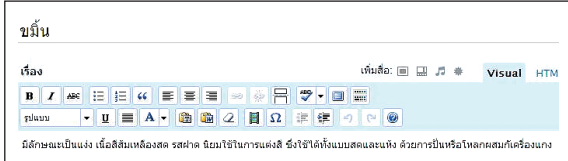
3.2.4 การเชื่อมโยงความสัมพันธ์ของบทความที่มีเนื้อหาซ้อนทับกันบางส่วน ตามกฎความสัมพันธ์วิเคราะห์ และได้กฎความสัมพันธ์ของบทความเพื่อทำการเชื่อมต่อในการสืบค้นด้วยเมทาเดตา ที่เกิดจากการตัดคำด้วยอัลกอริทึม Swath ทั้งหมด 319 ความสัมพันธ์ทั้ง 319 ความสัมพันธ์จะนำไปสร้างความสัมพันธ์ของบทความดังอธิบายด้วยตัวอย่างกฎความสัมพันธ์ข้อที่ 9 ที่ได้แสดงไว้ในรูปที่ 5

จากรูปที่ 5 แสดงความสัมพันธ์ของบทความของบทความเรื่องขมิ้น (A0002) หมวดหมู่ 01 สมุนไพร มีความสัมพันธ์กับบทความเรื่องเครื่องแกง (C0023) และบทความเรื่องแกงเหลือง (C0029) หมวดหมู่ 03 ผักและอาหารพื้นบ้าน

3.3 ผลการพัฒนาพัฒนาระบบการจัดเก็บและสืบค้นโดยระบบต้นแบบ ได้พัฒนาผลลัพธ์ที่ได้วิเคราะห์การแบ่งกลุ่มข้อมูลของภูมิปัญญาท้องถิ่นภาคใต้ ในข้อ 3.1 ตลอดจนผลการวิเคราะห์ฟอมอลคอนเท็กซ์ และคอนเซพแลททิส ในข้อ 3.2.1, 3.2.2, 3.2.3 และข้อ 3.2.4 ซึ่งสามารถสรุปและแยกเป็นกระบวนการทำงานของระบบ (Processing) ที่สำคัญได้ดังนี้



รูปที่ 6 ระบบจัดหมวดหมู่บทความ



รูปที่ 7 ระบบจัดทำบทความ

3.3.1 ระบบการจัดหมวดหมู่บทความ

จากรูปที่ 6 อธิบายได้ว่ากระบวนการทำงานของระบบ สำหรับกำหนดหมวดหมู่ให้กับบทความภูมิปัญญาท้องถิ่นภาคใต้ จะถูกเรียกใช้โดยผู้เขียนบทความ ซึ่งพิจารณาหมวดหมู่บทความจะพิจารณาจากความถี่ของคำที่กำหนดเป็นเมทาเดตา เมื่อมีการบันทึกรายการบทความเข้าสู่หมวดหมู่บทความ

3.3.2 ระบบจัดทำบทความ

จากรูปที่ 7 อธิบายได้ว่าระบบจัดทำบทความสามารถการเพิ่ม ลบ แก้ไขได้ บทความถูกเรียกใช้ โดยผู้สร้างบทความควบคู่กับการพิจารณาหมวดหมู่บทความ จะพิจารณาจากความถี่ของคำที่กำหนดเป็นเมทาเดตา

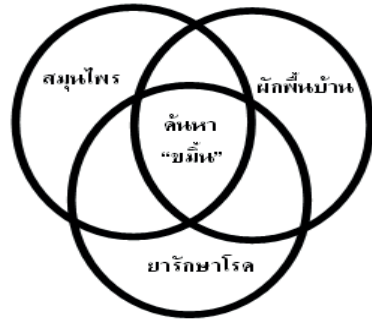
3.3.3 ระบบการสร้างความสัมพันธ์ด้วยกำหนดเมทาเดตา

จากรูปที่ 8 อธิบายได้ว่าการกำหนดเมทาเดตาที่ช่วยสร้างความสัมพันธ์ระหว่างบทความ ในแต่ละหมวดหมู่ โดยเมทาเดตาเกิดจากการวิเคราะห์ความถี่ของคำในบทความ เพื่อจัดเข้าหมวดหมู่ของบทความ รวมทั้งสร้างความสัมพันธ์ของบทความจากคอนเซพแลททิสซึ่งได้จากการวิเคราะห์ฟอมอลคอนเท็กซ์ เพื่อสืบค้นบทความที่มีความสัมพันธ์กันในเชิงคำค้น (Keywords) และความสัมพันธ์เชิงเนื้อหา (Content)

3.3.4 ระบบสืบค้นบทความสามารถทำการสืบค้นเนื้อหาบทความในคำสำคัญ (Keyword) ที่ปรากฏ



รูปที่ 8 การกำหนดเมทาเดตา ให้กับบทความ



รูปที่ 9 การค้นหาข้อมูลที่จัดเก็บต่างหมวดหมู่

ในเนื้อหาบทความ โดยไม่ต้องระบุฟิลด์ (Field) เพื่อทำการสืบค้น เนื่องจากระบบมีการสืบค้นจาก 2 ส่วน คือ สืบค้นผ่านเมทาเดตา และเนื้อหาบทความ (Text Search) เพราะภูมิปัญญาท้องถิ่นภาคใต้จะมีความสัมพันธ์และเชื่อมโยงกัน เช่น ขมิ้น สมุนไพร ยารักษาโรค ผักพื้นบ้าน ดังตัวอย่างที่แสดงไว้ในรูปที่ 9

3.4 ผลการทดสอบหาประสิทธิภาพของระบบจัดเก็บและสืบค้นภูมิปัญญาท้องถิ่นภาคใต้

3.4.1 ค่าความแม่นยำ จากข้อมูลเอกสารทั้งหมดที่ค้นได้ จำนวน 203 รายการ พบว่าจำนวนรายการเอกสารที่ค้นได้ข้อมูลที่ตรงกับความต้องการ จำนวน 186 รายการ และตรงต่อความต้องการบางส่วน 13 รายการ รวมทั้งหมด 199 รายการ ดังนั้น ค่าความแม่นยำเฉลี่ยเท่ากับ 0.98 หรือร้อยละ 98 ซึ่งได้ผลลัพธ์ออกมาอยู่ในระดับดีมาก

3.4.2 ค่าความระลึก จากจำนวนข้อมูลผลลัพธ์ที่ถูกต้อง และข้อมูลผลลัพธ์ที่ถูกต้องแต่ไม่ได้สืบค้นจำนวน 245 รายการ พบว่าจำนวนรายการเอกสารเกี่ยวข้องที่สามารถค้นได้จำนวน 199 รายการ ดังนั้นการหาค่าความระลึกเฉลี่ยเท่ากับ 0.81 หรือร้อยละ 81 ซึ่งได้ผลลัพธ์ออกมาอยู่ในระดับดี

4. อภิปรายผลและสรุป

4.1 ผลการวิเคราะห์หมวดหมู่ ได้ประยุกต์ขั้นตอนการวิเคราะห์จากงานวิจัยของ Songsiri [8] Kulkarineetham [12] และการวิเคราะห์ฟอมอลคอนเท็กซ์ ตามวิธีการที่กล่าวไว้ในผลงานของ Srisa-an [9] ซึ่งได้ผลการวิเคราะห์หมวดหมู่ เป็น 15 หมวดหมู่ ใกล้เคียงกับค่าความถี่ของค่าคั่นที่สุด เมื่อตรวจสอบโดยผู้ทรงคุณวุฒิด้านภูมิปัญญาท้องถิ่นภาคใต้ เห็นว่าบทความมีความสอดคล้องและตรงกับหมวดหมู่ บทความมีความสอดคล้องกับหมวดหมู่ จึงสรุปได้ว่าควรมีจัดแบ่งหมวดหมู่ภูมิปัญญาท้องถิ่นภาคใต้ 15 หมวดหมู่ และเมื่อนำบทความเข้าสู่หมวดหมู่ ควรจะใช้หลักการพิจารณาอื่นๆ เพิ่มเติม เช่น ความถี่ของค่าที่ปรากฏ เพื่อให้แน่ชัดว่าเนื้อหานั้นจะโน้มเอียงไปหมวดหมู่ใด

4.2 ผลการวิเคราะห์ความสัมพันธ์ จากข้อมูลการค้นหาแบบต่อเนื่อง จำนวน 57 รายการ ไม่สามารถวิเคราะห์ความสัมพันธ์วิเคราะห์ได้ เนื่องจากปริมาณที่น้อยเกินไปสำหรับการทำเหมืองข้อมูล ส่งผลให้ขาดความน่าเชื่อถือของผลลัพธ์ เมื่อเปรียบเทียบกับงานวิจัยของ Songsiri [8] Ruk-ngam [13] ดังนั้น ในการประยุกต์ความสัมพันธ์เพื่อไปพัฒนาระบบต้นแบบ จึงเลือกใช้ผลจากการวิเคราะห์ความสัมพันธ์ที่ได้จากการวิเคราะห์ผลการวิเคราะห์ ฟอมอลคอนเท็กซ์จากคอนเซพแลททิส ซึ่งเป็นกระบวนการที่กล่าวไว้ในผลงานของ Srisa-an [9]

4.3 ผลการวิเคราะห์ความสัมพันธ์ ด้วยการวิเคราะห์ด้วยฟอมอลคอนเท็กซ์ จะให้ความสัมพันธ์ที่ละเอียดมาก เนื่องจากวิเคราะห์จากบทความทั้งหมดของการจัดเก็บและสืบค้นบทความที่จัดเก็บไว้ แต่การวิเคราะห์ด้วยฟอมอลคอนเท็กซ์นั้นจะเหมาะกับการวิเคราะห์บทความที่ระบบนี้ไม่มีกรณีอื่นไหว หากผู้วิจัยท่านอื่นจะเริ่มต้นการวิเคราะห์ความสัมพันธ์ ควรใช้ฟอมอลคอนเท็กซ์ เพราะให้รายละเอียดมากพอที่จะเชื่อมโยงบทความได้ทั้งหมด

4.4 ผลการแปลงคอนเซพแลททิส จะถูกนำไปประยุกต์ใช้กับการวิเคราะห์ความสัมพันธ์และการแบ่งกลุ่ม

ข้อมูล ในการทำเหมืองข้อมูล เพื่อเชื่อมโยงการค้นหาบทความให้สามารถแสดงบทความที่มีเนื้อหาสัมพันธ์กัน แตกต่างหมวดหมู่กันได้ ตัวอย่าง การค้นคำว่า “ขมิ้น” จะปรากฏในหมวดหมู่ 01 สมุนไพร 02 ยารักษาโรค และ 03 ผักพื้นบ้าน จากผลการวิเคราะห์หมวดหมู่ ได้ 15 หมวดหมู่ มีความสอดคล้องกับการสร้างคอนเซพแลททิส ซึ่งมีโหนดหลัก มีจำนวน 15 โหนด ซึ่งวิธีการนี้ได้ถูกนำไปประยุกต์ใช้ในการวิเคราะห์ในหลากหลายสาขาวิชา เช่น จิตวิทยา สังคมวิทยา มนุษยวิทยา ตามวิธีการที่กล่าวไว้ในผลงานของ Srisa-an [9]

4.5 การพัฒนาระบบจัดเก็บและสืบค้นภูมิปัญญาท้องถิ่นภาคใต้ เป็นระบบต้นแบบ ได้พัฒนาจากผลการวิเคราะห์กฎความสัมพันธ์กฎการแบ่งกลุ่มข้อมูล และคอนเซพแลททิส ซึ่งใช้เมทาเดตาในการเชื่อมโยงบทความ จึงง่ายต่อการสืบค้น ไม่ว่าจะพิมพ์เป็นคำค้น หรือพิมพ์วลี ก็จะสามารถแสดงบทความที่เกี่ยวข้องกับคำนั้น แม้จะอยู่ต่างหมวดหมู่ จึงเป็นการนำเทคนิคเหมืองข้อมูลมาประยุกต์ใช้ในการจัดการภูมิปัญญาท้องถิ่นสอดคล้องงานวิจัยของ Sungsisri [4] Duangonnam [5] Suwannawat [6] Phengsuan [14] ที่นำเทคนิคเหมืองข้อมูลมาประยุกต์ในการพัฒนาระบบจัดเก็บและสืบค้น

4.6 ผลการทดสอบหาประสิทธิภาพระบบที่พัฒนาขึ้น ค่าความแม่นยำ เฉลี่ยเท่ากับ 0.98 หรือร้อยละ 98 อยู่ในระดับดีมาก โดยกระบวนการวิเคราะห์ประยุกต์จากงานวิจัยของ Sungsisri [4] และได้ค่าความแม่นยำใกล้เคียงกับ Sungsisri [4] ซึ่งนำเอาออนโทโลยี (Ontology) ออกแบบโครงสร้างระบบสืบค้น ตลอดจนสอดคล้องกับงานวิจัยของ Ruk-ngam [13] ได้ใช้เทคนิค OLAP Cube มาประยุกต์ใช้ในการจัดทำระบบสืบค้น ซึ่งจะแตกต่างจากงานวิจัยชิ้นนี้ที่ทำการสืบค้นผ่านเมทาเดตา และเนื้อหาบทความ ในส่วนของค่าความระลึก ซึ่งประยุกต์มาจากวิธีการในงานวิจัยของ Sungsisri [4] ได้ค่าความระลึกเฉลี่ยเท่ากับ 0.81 หรือร้อยละ 81 อยู่ในระดับดี แต่น้อยกว่าผลการคำนวณของ Sungsisri [4] ที่ใช้นำเอาออนโทโลยีมาออกแบบโครงสร้างระบบสืบค้น แต่งานวิจัยชิ้นนี้

ยังขาดการประเมินค่า F-Factor เมื่อเปรียบเทียบกับงานวิจัยของ Powers [7] นอกจากนี้ผลการทดสอบประสิทธิภาพค่าความระลึก น้อยกว่าค่าความแม่นยำ เนื่องจากการทดสอบประสิทธิภาพระบบจากคำค้น คำว่า “รักษาโรค” แต่ในเมทาเดตาของบทความจัดเก็บคำว่า “แก้โรค” ซึ่งเป็นการเขียนบทความโดยใช้ศัพท์ภาษาใต้ ซึ่งเป็นข้อมูลที่ติดมาจากการลงพื้นที่และถูกกำหนดให้เป็นเมทาเดตาโดยโปรแกรมนับคำ จึงทำให้ไม่ปรากฏบทความ “รักษาโรค” หรือ คำว่า “ขนม” ซึ่งจะอยู่ในหมวดผักและอาหารพื้นบ้าน ซึ่งจะมีบทความเกี่ยวกับอาหารคาว อาหารหวาน แกง และผัก อยู่ในหมวดหมู่นั้นด้วย ทำให้ข้อมูลผลลัพธ์ที่ถูกต้องและข้อมูลผลลัพธ์ที่ถูกต้องแต่ไม่ได้สืบค้นมีค่าสูง ค่าความระลึกมีน้อยกว่าค่าความแม่นยำ (Precision) จึงทำให้ค่าประสิทธิภาพต่างกัน ดังนั้นหากต้องการปรับให้ค่าสองค่าใกล้เคียงกันจะต้องเพิ่มเมทาเดตา คำว่า “ยา” และ “รักษาโรค” ในบทความการรักษาโรคชนิดต่างๆ และเพิ่มเมทาเดตา คำว่า “ขนม” ในบทความเกี่ยวกับอาหารคาวบ้างอย่าง ด้วยการป้อนข้อมูลเข้าสู่ระบบ เพราะค่าเหล่านี้ไม่ปรากฏในบทความ

5. กิตติกรรมประกาศ

ขอบพระคุณ มหาวิทยาลัยราชภัฏสุราษฎร์ธานี และสำนักงานคณะกรรมการวิจัยแห่งชาติ

เอกสารอ้างอิง

- [1] Office of the National Economic and Social Development Broad, *Summary of the Derection of The Eleventh National Development Plan*, Bangkok, 2010 (in Thai).
- [2] Q. Yang, H. Wang, and W. Zhang, “Web-log Mining for Quantitative Temporal-Event Prediction,” *IEEE Computational Intelligence Bulletin*, vol. 1, no.1, pp.10-18, Dec. 2002.
- [3] J. Han and M. Kamber, *Data mining concepts and techniques*, 3rd ed. San Francisco: Morgan Kaufman Publishers, 2011.
- [4] S. Sungsisri, “The Development an Information System for Knowledge Management of Intelligence of Locality. Case Study of Floating-Basket Fishery in the Sakaekrang River,” M.S. Thesis, Department of Computer Education, Faculty of Technical Education King Mongkut’s University of Technology North Bangkok, 2010 (in Thai).
- [5] W. Duangonnam, “Searching Fluke Egg using Image-based Retrieval Technique,” M.S. Thesis, Department of Information Technology, Faculty of Information Technology King Mongkut’s Institute of Technology North Bangkok, 2005 (in Thai).
- [6] T. Suwannawat, “An Ontology-Based Help Desk Service of Installing and Using Call Center Devices,” M.S. Thesis, Department of Computer and Information Science, The Faculty of Applied Science King Mongkut’s Institute of Technology North Bangkok, 2007 (in Thai).
- [7] D.M.W. Powers, “Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation,” *Journal of Machine Learning Technologies*, vol. 2, pp.37-63, Feb. 2011.
- [8] C. Songsiri, T. Rakthanmanon, and K. Waiyamai, “Applying a Data Mining Technique to Help Students in Selecting Their Majors,” The 39th Kasetsart University Annual Conference, Kasetsart University, 2001 (in Thai).
- [9] C. Srisa-an, *Database Data Warehouse and Data Mining*. Pathum Thani: Rangsit University, 2008 (in Thai).
- [10] T. Pongsiripreeda, “Automatic Legislation



- Selection for Lawsuits,” M.S. Thesis, Department of Computer Engineering, The Faculty of Engineering Kasetsart University, 2001 (in Thai).
- [11] P. Charoenpornasawat, “Feature-Based Thai Word Segmentation,” M.S. Thesis, Department of Computer Engineering, Faculty of Engineering Chulalongkorn University, 1999 (in Thai).
- [12] S. Kulkarineetham, “Isced Curriculum Categorization Using Data mining: A Case Study of Science Category,” M.S. Thesis, Department of Computing, The Faculty of Science Silpakorn University, 2006 (in Thai).
- [13] N. Ruk-ngam, “Question Analysis and Answering System for the Internet Usage by Using Data Mining Techniques,” M.S. Thesis, Department of Computer Science, The Faculty of Science Kasetsart University, 2007 (in Thai).
- [14] B. Phengsuan, “Applied Data Mining Techniques to Develop a Decision Support System for Graduate Production of Bachelor’s Degrees,” M.S. Thesis, Department of Computing, The Faculty of Science Silpakorn University, 2005 (in Thai).