

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

การศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง ผู้วิจัยได้ศึกษา ความหมายและขอบเขตของ องค์ความรู้ที่เกี่ยวข้องกับภูมิปัญญาท้องถิ่น ศึกษาองค์ความรู้ที่เกี่ยวข้องกับภูมิปัญญาท้องถิ่นภาคใต้ สำหรับด้านเทคโนโลยี ผู้วิจัยได้ศึกษา เทคโนโลยีเว็บ 2.0 เทคนิคการทำเหมืองข้อมูล การทำเหมือง ข้อมูลเว็บ เทคนิคการทำเหมืองข้อมูลการใช้เว็บไซต์ ตลอดจนศึกษางานวิจัยที่เกี่ยวข้องในประเด็นที่ จะเป็นประโยชน์ ด้านขั้นตอนและวิธีการดำเนินการวิจัย การวิเคราะห์ผล และการประเมิน ประสิทธิภาพของระบบ

2.1 ภูมิปัญญาท้องถิ่น

ภูมิปัญญาชาวบ้าน หมายถึง ความรู้ของชาวบ้าน ซึ่งเรียนรู้มาจากพ่อแม่ ปู่ย่าตายาย ญาติพี่น้อง หรือผู้มีความรู้ในหมู่บ้านตามท้องถิ่นต่างๆ ความรู้เหล่านี้สอนให้เด็กเคารพผู้ใหญ่ มีความกตัญญูรู้คุณ พ่อแม่ และผู้มีพระคุณ มีความเอื้ออาทรต่อผู้อื่น รู้จักช่วยเหลือแบ่งปันเกื้อกูลซึ่งกันและกัน

ความรู้ที่เป็นภูมิปัญญาเป็นความรู้ที่มีคุณธรรม สอนให้คนเป็นคนดี สอนให้คนเคารพธรรมชาติ รู้จักพึ่งพาอาศัยธรรมชาติโดยไม่ทำลาย เพื่อการพัฒนาที่ยั่งยืน ในด้านคุณธรรมจริยธรรมนั้น ภูมิปัญญาท้องถิ่นเป็นความรู้ที่ประกอบไปด้วยคุณธรรม ซึ่งสอดคล้องกับวิถีชีวิตของชาวบ้าน ไม่แบ่งแยก หากแต่ทุกอย่างมีความสัมพันธ์กัน การทำมาหากิน การอยู่ร่วมกันในชุมชน การปฏิบัติ ศาสนา พิธีกรรมและประเพณีความรู้เป็นคุณธรรม เมื่อผู้คนใช้ความรู้นั้น เพื่อสร้างความสัมพันธ์ที่ดี ระหว่าง คนกับคน คนกับธรรมชาติ นอกจากนั้นภูมิปัญญาชาวบ้านเป็นความรู้เรื่องการทำมาหากิน เช่น การจับปลา การจับสัตว์ การทอผ้า ทอเสื่อ การสานตะกร้า การทำเครื่องปั้นดินเผา การทำ เครื่องมือทางการเกษตร นอกจากนั้น ยังมีศิลปดนตรี การฟ้อนรำ และการละเล่นต่าง ๆ การรักษา โรคด้วยวิธีต่างภูมิปัญญาเหล่านี้เป็นความรู้ความสามารถที่บรรพบุรุษได้สร้างสรรค์ และถ่ายทอดมาให้ เรามีวิธีการหลายอย่างที่ทำให้ความรู้เหล่านี้เกิดประโยชน์แก่สังคมปัจจุบัน

สำหรับคำว่า ภูมิปัญญาท้องถิ่นไทย (Thai Local Wisdom) คือ องค์ความรู้ของกลุ่มบุคคล ท้องถิ่น และรวมถึงงานศิลปวัฒนธรรมพื้นบ้านที่มีอยู่ในประเทศไทย แบ่งออกเป็น 2 ประเภท คือ ประเภทองค์ความรู้ของกลุ่มบุคคลท้องถิ่น เช่น การผลิตอาหารและเครื่องดื่ม การผลิตผลิตภัณฑ์ จากสมุนไพร การผลิตผลิตภัณฑ์จากวัสดุเหลือใช้ และการผลิตผลิตภัณฑ์จากไม้ หิน โลหะ แก้ว เซรามิค ดินเผา เครื่องหนัง และอื่นๆ และประเภทงานศิลปวัฒนธรรมพื้นบ้าน เช่น เรื่องเล่าพื้นบ้าน กวีนิพนธ์พื้นบ้าน ปริศนา พื้นบ้าน เพลงพื้นบ้าน ดนตรีพื้นบ้าน การฟ้อนรำพื้นบ้าน ละครพื้นบ้าน จิตรกรรมพื้นบ้าน ประติมากรรมพื้นบ้าน หัตถกรรมพื้นบ้าน เครื่องแต่งกายพื้นบ้าน และสิ่งทอ พื้นบ้าน (โครงการพัฒนาศูนย์กลางความรู้แห่งชาติ, 2550)

2.2 ภูมิปัญญาท้องถิ่นภาคใต้

ภูมิปัญญาท้องถิ่นภาคใต้เป็นมรดกทางวัฒนธรรมอันเกิดจากการพัฒนา การปรับตัว ปรับวิถีชีวิตของคนในภาคใต้ที่ประกอบด้วยคนไทยและอีกหลายชาติพันธุ์ที่อยู่ร่วมกันในคาบสมุทร ประกอบด้วยเชื้อชาติมาลายู เชื้อชาติจีน และเชื้อชาติอินเดียฝ่ายใต้ สำหรับกลุ่มชนที่มีจำนวนมากที่สุดคือ คนไทย

จากสภาพภูมิศาสตร์ที่มีทะเลกว้างใหญ่ขนานอยู่ทั้งสองข้าง มีทรัพยากรธรรมชาติที่อุดมสมบูรณ์ ทั้งในทะเล และบนแผ่นดิน อันล้วนเป็นเขตมรสุมใกล้เส้นศูนย์สูตร มีผู้คนหลายชาติ หลายภาษา หลายวัฒนธรรมเดินทางมาทั้งทางบกและทางทะเลเพื่อมาตั้งหลักแหล่ง ทำมาค้าขายติดต่อกันยาวนานกว่าพันปี มีการตั้งถิ่นฐานทำมาหากินกันหลายลักษณะ ทั้งบริเวณชายทะเล ที่ราบระหว่างชายทะเลกับเทือกเขา และตามสายน้ำน้อยใหญ่จำนวนมากที่ไหลจากเทือกเขาลงสู่ทะเลทั้งสองด้าน ภูมิปัญญาของภาคใต้จึงมีความหลากหลาย ทั้งที่ได้พัฒนาจากการปรับตัวให้เข้ากับสิ่งแวดล้อม ธรรมชาติ หรือคนต่างถิ่นที่อพยพมาจากแหล่งอารยธรรมต่างๆ จนหลอมรวมกัน เกิดเป็นภูมิปัญญาประจำถิ่น ซึ่งมีอยู่ในหลายลักษณะ คือ ภูมิปัญญาในด้านความสัมพันธ์และการพึ่งพา ภูมิปัญญาในด้านหัตถกรรมพื้นบ้าน ภูมิปัญญาในด้านสมุนไพร ภูมิปัญญาในด้านทัศนคติ ภูมิปัญญาในการปลูกฝังคุณธรรม และภูมิปัญญาในการดำรงชีพ

ชุมชนภาคใต้ที่มีพื้นที่ตั้งถิ่นฐานค่อนข้างหลากหลายคือ มีทั้งที่ราบตามแนวชายฝั่ง ปากอ่าว ท่าเรือ ที่ราบเชิงเขา หลังเขา ตามแนวสายน้ำน้อยใหญ่ บนเกาะ ประกอบกับความอุดมสมบูรณ์ของผืนแผ่นดิน ทำให้คนภาคใต้ได้เรียนรู้และสั่งสมประสบการณ์ ความสามารถ อย่างมากมายในการจัดการและปรับตัวให้สามารถดำรงอยู่อย่างกลมกลืนกับสภาพแวดล้อมซึ่งมีอยู่หลายประการดังนี้

2.2.1 การขุดสระน้ำ เพื่อให้ได้น้ำจืดใสสะอาดไว้กินไว้ใช้ตลอดทั้งปี เนื่องจากอยู่ใกล้ทะเล ยามที่น้ำขึ้น น้ำเค็มจะไหลเอ่อเข้ามาในแผ่นดิน จะอาศัยอาบกินก็ไม่สะดวก ดังนั้นจึงเกิดภูมิปัญญาในการหาพื้นที่ขุดบ่อน้ำ

2.2.2 การปลูกต้นไม้บริเวณบ้าน คนภาคใต้มีความเชื่อทำนองเดียวกับภาคอื่นว่าการปลูกต้นไม้ในบริเวณบ้านต้อง เลือกปลูกเฉพาะไม้ที่เป็นมงคลและปลูกให้ถูกทิศทาง โดยมีประโยชน์แฝงไว้เพื่อให้เกิดความร่มเย็นทั้งกายใจ ส่วนที่ห้ามปลูกในบริเวณบ้าน เป็นพวกไม้ใหญ่ที่แผ่กิ่งก้านสาขามาก หรือชื่อไม่เป็นมงคล เช่น เต่าร้าง ลั่นทม ความเชื่อนี้เกิดขึ้นโดยอาศัยความรู้ประสบการณ์เกี่ยวกับการโคจรของดวงอาทิตย์ ทิศทางของลมมรสุม ความหนักเบาของฝน ซึ่งจะแตกต่างกันไปในแต่ละวัน รวมถึงความเข้าใจธรรมชาติของต้นไม้ในแต่ละพันธุ์ด้วย

2.2.3 การปลูกสร้างบ้านเรือน จะมีลักษณะแตกต่างจากภาคอื่นๆ เช่น คนภาคใต้นิยมแผ้วถางพื้นที่บริเวณบ้านให้เตียนเรียบจนเห็นเป็นพื้นทรายขาวสะอาด มีเหตุผลด้านสภาพแวดล้อมคือ การเดินเข้าออกสะดวก ปลอดภัยจากสัตว์ร้าย เช่น งู ตะขาบ แมงป่อง ที่มีอยู่อย่างชุกชุม

2.2.4 อุปายในการครองชีพ สังคมของคนภาคใต้มีเคล็ดหรืออุปายในการดำรงชีพและการทำมาหากินตามสภาพแวดล้อมหลายประการ เช่น เครื่องหมายแสดงเจตจำนง แต่เดิมคนภาคใต้ไม่รู้หนังสือหรือแม้จะรู้บ้าง แต่การสื่อความหมายโดยใช้เครื่องหมายบอกเจตจำนงก็ยังนิยมใช้กันอย่างแพร่หลาย ยอมรับเป็นกติกาแห่งสังคมที่อยู่กันแบบพึ่งพาอาศัย ที่นิยมใช้กัน คือ เครื่องหมาย “ห้าม” “ขอ” “ขัดใจ” ห้าม หรือภาษาถิ่นที่เรียกว่า ปักกำ กายยัง หรือ กายแยะ เป็นเครื่องหมายที่จะนำไปปักไว้ในที่หวงห้าม เช่น ห้ามจับปลาในหนองน้ำ

2.2.5 ภูมิปัญญาในด้านความสัมพันธ์และการพึ่งพา ซึ่งจากสภาพภูมิศาสตร์ของภาคใต้ที่มีอยู่อย่างหลากหลายเป็นตัวกำหนดสำคัญที่ทำให้ ชุมชนแตกต่างกันไปจากภาคอื่นๆ ทั้งลักษณะการตั้งถิ่นฐาน และการทำมาหากิน อย่างสำคัญประการหนึ่งคือชุมชนต่างๆ ของคนภาคใต้ ไม่อาจอยู่ได้ด้วยการพึ่งพาตนเองโดยลำพัง ชาวสวนผลไม้ สวนยาง และเหมืองแร่ในป่า ต้องการข้าว กะปิ น้ำปลา กุ้งแห้ง จากหมู่บ้านพื้นราบหรือชายฝั่ง

2.2.6 ภูมิปัญญาในด้านหัตถกรรมพื้นบ้าน ซึ่งหัตถกรรม พื้นบ้านภาคใต้เป็นภูมิปัญญา และเป็นมรดกทางวัฒนธรรมของสังคมเกษตรกรรมที่ได้ มีการสืบสานความรู้ ความสามารถและความชำนาญต่างๆ มายังคนรุ่นปัจจุบัน สิ่งใดประดิษฐ์ขึ้นใช้แล้วไม่ได้ผลหรือมีสิ่งอื่นที่ดีกว่าก็จะเสื่อมสลายไป สิ่งใดใช้ได้ดีและสะดวกต่อการผลิตสิ่งนั้นก็ยังคงอยู่ อีกทั้งยังมีการสอดแทรกคุณค่าทางด้านศิลปะลงในงานด้วย หัตถกรรมพื้นบ้านภาคใต้เป็นการนำเอาวัตถุดิบในธรรมชาติ โดยเฉพาะพืชพรรณที่มีอยู่มากมาย ทั้งบนบกและในน้ำ เช่น ไม้ไผ่ หวาย กระจูด กก มะพร้าว ลิเพา รวมถึงปาล์ม นำมาดัดแปลงเป็นเครื่องใช้ไม้สอยต่างๆ เพื่อใช้ในชีวิตประจำวัน

2.2.7 ภูมิปัญญาในด้านสมุนไพร ซึ่งจากอดีตกาลภาคใต้เป็นดินแดนที่มีทรัพยากรธรรมชาติมากมาย แม้ในปัจจุบันจะมีการบุกรุกพื้นที่ป่า แต่ก็ยังคงความอุดมสมบูรณ์และเต็มไปด้วยความหลากหลายทางชีวภาพ ความรู้และประสบการณ์ของบรรพบุรุษในการใช้ประโยชน์ทรัพยากรได้ถ่ายทอดจากคน รุ่นหนึ่งไปสู่อีกรุ่นหนึ่ง ซึ่งสมุนไพรถือเป็นภูมิปัญญาที่คนภาคใต้ได้รับการถ่ายทอดมาจนปัจจุบัน ด้วยเหตุที่คนภาคใต้ส่วนใหญ่ใช้ชีวิตคลุกคลีอยู่กับป่าดงพงไพร ความจำเป็นเมื่อเจ็บป่วยบังคับให้คนเหล่านี้ต้องเสี่ยงชีวิตเก็บเอาสิ่ง ต่างๆ ทั้งที่เป็น พืช สัตว์ หรือสินแร่ มาทดลองใช้ในลักษณะลองผิดลองถูก จนค้นพบในสรรพคุณของสิ่งที่ใช้บำบัดรักษา

2.2.8 ภูมิปัญญาในด้านทัศนคติ โดยภาพรวมแล้วทัศนคติต่อสิ่งแวดล้อมของคนภาคใต้ไม่ได้แตกต่างไปจากภูมิภาคอื่น เว้นแต่จะผูกพันแนบแน่นอยู่กับธรรมชาติที่เป็นคาบสมุทร รวมทั้งการได้แลกเปลี่ยนสังสรรค์กับคนจากวัฒนธรรมอื่นอย่างต่อเนื่องกันมา เป็นเวลานานนับพันปี ทำให้เกิดการผสมผสานของวัฒนธรรมอยู่ตลอดเวลา ซึ่งสภาพดังกล่าวนี้เป็นลักษณะเด่นที่มีอิทธิพลต่อทัศนคติที่เป็นภูมิปัญญา เฉพาะถิ่น และถือปฏิบัติเป็นบรรทัดฐานของคนภาคใต้

2.2.9 การปกครอง ในอดีตหัวเมืองปักษ์ใต้ไม่ได้ถูกปกครองอย่างใกล้ชิดจากเมือง ราชธานี แต่จะอาศัยเมืองหลักอย่างนครศรีธรรมราชและสงขลาเป็นศูนย์กลางปกครองดูแลอีก ดังนั้นความสัมพันธ์เชิงอำนาจระหว่างเมืองหลวงกับหัวเมืองปักษ์ใต้ จึงไม่ใช่การรวมศูนย์อำนาจ หัวเมืองต่างๆ มีอิสระในการปกครองและดูแลกันเอง ถึงแม้ว่าจะไม่เป็นแบบนี้ทั้งหมด แต่หลายๆ หมู่บ้านก็เกิดจากการสร้างตัวด้วยลำแข้งของตนเองท่ามกลางสภาพธรรมชาติที่ เกื้อกูลเป็นอันดี คนภาคใต้จึงมีทัศนะในการปกครองแบบพึ่งพาตนเอง รักอิสระ รักความเป็นธรรม

2.2.10 สถานภาพ สังคมของคนภาคใต้ให้ความสำคัญกับเพศชายสูงกว่าเพศหญิง ขณะเดียวกันก็มีความคาดหวังให้ผู้หญิงเป็นผู้มีคุณค่าสมเป็นกุลสตรีไทย ซึ่งมีทัศนะต่างๆ เช่น การยกย่องบุรุษ เพศในการเป็นหัวหน้าครอบครัว พร้อมกันนั้นก็คาดหวังว่าบุรุษต้องเป็นผู้มีความรู้ มีคุณธรรม และต้องเป็นคนจริง คือ เชื่อถือได้ไม่เหลวไหล

2.2.11 การปลูกฝังคุณธรรม ในความเป็นสังคมเปิดที่มีเครือข่ายระบบความสัมพันธ์กว้างขวาง และต้องปรับตัว เข้ากับความเปลี่ยนแปลง พร้อมๆ กับการที่ต้องรักษาความเชื่อแบบธรรมเนียมที่เป็น

บรรทัดฐานของสังคมไว้เพื่อให้การดำเนินชีวิตราบรื่นตามครรลองที่สืบทอดกันมา สังคมคนภาคใต้จึงมีกรรมวิธีปลูกฝังรักษาความเชื่อและบรรทัดฐานของสังคมไว้ใน หลายๆ วิธี ด้วยการอาศัย บ้าน ชุมชน วัด พิธีกรรม วรรณกรรมท้องถิ่น เป็นต้นถ่ายทอด โดยเฉพาะอย่างยิ่งวรรณกรรมคำสอน สุภาษิต ตลอดจนการเล่นต่างๆ เมื่อคนเติบโตขึ้นท่ามกลางค่านิยมและแบบอย่างความประพฤติ ในสภาพแวดล้อมทางวัฒนธรรมและกิจกรรมตามชนบประเพณีที่ได้ทำหน้าที่ถ่ายทอด หรือต่อย้ำ ก็จะทำให้เกิดการซึมซับรับเอาโดยอัตโนมัติ ซึ่งคนภาคใต้มีวิธีการปลูกฝังและรักษาบรรทัดฐานของสังคมดังนี้

ในปัจจุบันการพัฒนาของเทคโนโลยีสารสนเทศส่งผลให้เกิดการหลั่งไหลทางวัฒนธรรมจากต่างประเทศเข้าสู่สังคมไทยทำให้คนรุ่นใหม่มุ่งศึกษาเฉพาะองค์ความรู้ที่เป็นในการประกอบวิชาชีพ แต่ละเลยที่จะศึกษาองค์ความรู้ที่เป็นภูมิปัญญาท้องถิ่นภาคใต้ และองค์ความรู้เหล่านั้น ก็สูญหายไปตามกาลเวลา

งานวิจัยชิ้นนี้ ผู้วิจัยมีความตั้งใจเพื่อรักษาองค์ความรู้ให้มีการสืบทอดถึงลูกหลาน โดยการรวบรวมองค์ความรู้และจัดเก็บอย่างเป็นระบบเพื่อให้องค์ความรู้ที่ตกทอดถึงลูกหลานต่อไป ซึ่งเทคโนโลยีที่ผู้วิจัยนำเทคโนโลยี เว็บ 2.0 ในการรวบรวมบทความภูมิปัญญาท้องถิ่นภาคใต้ เพื่อเปิดโอกาสให้ผู้ใช้งานได้มีเพิ่มเติมองค์ความรู้เข้าสู่ระบบการจัดเก็บบทความ (โครงการพัฒนาศูนย์กลางความรู้แห่งชาติ, 2550)

2.3 เทคโนโลยีเว็บ

เว็บ 2.0 (Web 2.0) เป็นคำที่ใช้อธิบายถึงลักษณะของเทคโนโลยีเว็ลด์ไวด์เว็บ และการออกแบบเว็บไซต์ในปัจจุบัน ที่มีลักษณะส่งเสริมให้เกิดการแบ่งปันข้อมูล การพัฒนาในด้านแนวความคิดและการออกแบบ รวมถึงการร่วมสร้างข้อมูลในโลกของอินเทอร์เน็ต แนวคิดเหล่านี้นำไปสู่การพัฒนาและการปฏิวัติรูปแบบเทคโนโลยีที่นำไปสู่เว็บเซอร์วิสหลายอย่าง เช่น บล็อก เครือข่ายสังคมออนไลน์ วิกี คำว่า เว็บ 2.0 เริ่มเป็นที่รู้จักในวงกว้าง คำว่า เว็บ 2.0 นั้นเป็นคำกล่าวเรียกลักษณะของเว็ลด์ไวด์เว็บในปัจจุบัน ตามลักษณะของผู้ใช้งาน โปรแกรมเมอร์และผู้ให้บริการ

เว็บ 2.0 เป็นแพลตฟอร์มหนึ่ง ที่อยู่เหนือการใช้งานของซอฟต์แวร์ โดยไม่ยึดติดกับตัวซอฟต์แวร์ เหมือนระบบคอมพิวเตอร์ที่ผ่านมา โดยมีข้อมูล ที่เกิดจากผู้ใช้งานหลายคน เช่น บล็อก ซึ่งเป็นตัวผลักดันความสำเร็จของเว็บไซต์อีกต่อหนึ่ง ซึ่งเว็บไซต์ในปัจจุบันมีลักษณะการสร้างโดยผู้ใช้ที่อิสระ และแยกจากกัน ภายใต้ซอฟต์แวร์ตัวเดียวกัน เพื่อสรรค์สร้างระบบให้ก่อเกิดประโยชน์ในองค์กรรวม โดยตัวอย่างของระดับของเว็บ 2.0 ออกเป็นสี่ระดับ ดังนี้

ระดับ 3 - ระดับของการใช้งานจากผู้ทั่วไปในอินเทอร์เน็ต ซึ่งเป็นลักษณะของการสื่อสารของมนุษย์ภายใต้เว็บไซต์เดียวกัน ตัวอย่างเช่น วิกีพีเดีย สไกป์ อีเบย์ แครกส์ลิสต์

ระดับ 2 - ระดับการจัดการทั่วไปที่สามารถใช้งานได้โดยไม่จำเป็นต้องผ่านอินเทอร์เน็ต แต่เมื่อนำมาใช้งานออนไลน์ นั้น จะมีประโยชน์มากขึ้นจากการเชื่อมโยงผู้ใช้งานเข้าด้วยกัน ซึ่งไอโรลลีย ยกตัวอย่างเว็บไซต์ ฟลิคเกอร์ เว็บไซต์อัปโหลดภาพที่มีการใช้งานเชื่อมโยงระหว่างภาพ และเช่นเดียวกันระหว่างผู้ใช้งาน

ระดับ 1 - ระดับการจัดการทั่วไปที่สามารถใช้งานได้โดยไม่จำเป็นต้องผ่านอินเทอร์เน็ต แต่มีความสามารถเพิ่มขึ้นนำมาใช้งานออนไลน์ ตัวอย่าง เช่น กูเกิลดอคส์ และ ไอทูนส์

ระดับ 0 - ระดับที่สามารถใช้งานได้ทั้งออนไลน์และออฟไลน์ เช่น แมปเคเวสต์ และ กูเกิล แมปส์ ซึ่งแอปพลิเคชันหลายตัวที่ใช้ในการติดต่อสื่อสารอย่าง อีเมล เมสเซนเจอร์ ไม่ได้ถูกจำกัดอยู่ในลักษณะของเว็บ 2.0 แต่อย่างใด

ในปัจจุบันภายใต้เทคโนโลยีเว็บ 2.0 การเจริญเติบโตเว็บไซต์มีอัตราสูงขึ้นเนื่องจากเทคโนโลยีเว็บ 2.0 ผู้ใช้เป็นผู้สร้างเนื้อหาในเว็บไซต์ได้ด้วยตนเอง โดยเนื้อหาในเว็บไซต์โดยส่วนใหญ่ได้จากผู้ใช้งานเว็บไซต์ช่วยกันสร้าง ซึ่งแต่เดิมนั้นเนื้อหาในเว็บไซต์ผู้พัฒนาเว็บไซต์จะเป็นผู้สร้าง

งานวิจัยชิ้นนี้ ผู้วิจัยได้พัฒนานวัตกรรมนั้น ได้นำเอาเทคโนโลยีเว็บ 2.0 มาใช้ในการพัฒนาเครื่องมือเพื่อใช้ในการรวบรวมภูมิปัญญาท้องถิ่นจากปราชญ์ท้องถิ่นในลักษณะรูปแบบของเว็บบล็อก เพื่อให้เกิดการแลกเปลี่ยนความรู้ระหว่างผู้เป็นเจ้าขององค์ความรู้และผู้สนใจโดยมีเว็บบล็อกเป็นตัวกลาง (เนคเทค, 2549)

2.4 เว็บบล็อก

บล็อก (blog) มาจากคำว่า เว็บล็อก (weblog) เป็นรูปแบบเว็บไซต์ประเภทหนึ่ง ซึ่งถูกเขียนขึ้นในลำดับที่เรียงตามเวลาในการเขียน ซึ่งจะแสดงข้อมูลที่เขียนล่าสุดไว้แรกสุด บล็อกโดยปกติจะประกอบด้วย ข้อความ ภาพ ลิงค์ ซึ่งบางครั้งจะรวมสื่อต่างๆ ไม่ว่าจะเป็น เพลง หรือวิดีโอในหลายรูปแบบได้ จุดที่แตกต่างของบล็อกกับเว็บไซต์โดยปกติคือ บล็อกจะเปิดให้ผู้เข้ามาอ่านข้อมูลสามารถแสดงความคิดเห็นต่อท้ายข้อความที่เจ้าของบล็อกเป็นคนเขียนซึ่งทำให้ผู้เขียนสามารถได้ผลตอบกลับโดยทันที

บล็อกเป็นเว็บไซต์ที่มีเนื้อหาหลากหลายขึ้นอยู่กับเจ้าของบล็อก โดยสามารถใช้เป็นเครื่องมือสื่อสาร การประกาศข่าวสาร การแสดงความคิดเห็น การเผยแพร่ผลงาน ในหลายด้านไม่ว่า อาหาร การเมือง เทคโนโลยี หรือข่าวปัจจุบัน นอกจากนี้บริษัทเอกชนหลายแห่งได้มีการจัดทำบล็อกของทางบริษัทขึ้น เพื่อเสนอแนวความเห็นใหม่ โดยมีการเขียนบล็อกออกมาในลักษณะเดียวกับข่าวสั้น และได้รับการตอบรับจากทางลูกค้าที่แสดงความเห็นตอบกลับเข้าไป เพื่อพัฒนาผลิตภัณฑ์ บล็อกได้เริ่มมีบทบาทมากขึ้นในปัจจุบันในวงการสื่อมวลชนในหลายประเทศ เนื่องจากระบบแก้ไขที่เรียบง่าย และสามารถตีพิมพ์เรื่องราวได้โดยไม่ต้องใช้ความรู้ในการเขียนเว็บไซต์ โดยนอกเหนือจากที่ผู้เขียนข่าวส่งผลงานให้กับทางสื่อแล้ว ยังได้มาเขียนข่าวในอีกช่องทางหนึ่งในการเผยแพร่ข้อมูลหรือแนวความคิด โดยการเขียนบล็อกสามารถเผยแพร่ข้อมูล สู่ประชาชนได้รวดเร็วและเสียค่าใช้จ่ายน้อยกว่าสื่อในด้านอื่น

ผู้ใช้งานจะแก้ไขและบริหารผ่านทางเว็บเบราว์เซอร์เหมือนการใช้งานและอ่านเว็บไซต์ทั่วไป แต่จะมีรูปแบบบริหารที่แตกต่างกัน เช่นบางระบบที่มีบรรณาธิการของบล็อก ผู้เขียนหลายคนจะส่งเรื่องเข้าทางบล็อก และจะต้องรอให้บรรณาธิการอนุมัติให้บล็อกเผยแพร่ก่อน บล็อกถึงจะแสดงผลในเว็บไซต์นั้นได้ เช่น บล็อกซอฟต์แวร์ หรือ บล็อกแวร์ เป็นซอฟต์แวร์ที่ใช้ในอินเทอร์เน็ต ในลักษณะของระบบจัดการเนื้อหาเว็บ ที่ผู้พัฒนาซอฟต์แวร์และผู้เขียนหรือดูแลบล็อกจะแยกจากกันต่างหาก ส่งผลให้ผู้เขียนบล็อกสามารถใช้งานได้ทันทีโดยไม่ต้องมีพื้นฐานความรู้ในด้านเอชทีเอ็มแอล หรือการทำเว็บไซต์ ซอฟต์แวร์บางส่วนเป็นซอฟต์แวร์โอเพนซอร์ส ซึ่งเป็นซอฟต์แวร์ที่ไม่สงวนลิขสิทธิ์ ผู้พัฒนาสามารถนำมาปรับแก้เป็นของตนเอง ติดตั้งไว้ใช้เป็นบล็อกส่วนตัว หรือเผยแพร่ให้คนอื่นมาใช้งานได้ (เนคเทค, 2549)

ในงานวิจัยขั้นนี้ผู้วิจัยได้พัฒนาเว็บไซต์ ในรูปแบบเว็บบล็อก ซึ่งเป็นเทคโนโลยีเว็บ 2.0 เพื่อใช้ในการรวบรวมบทความภูมิปัญญาท้องถิ่นจากประชาชนท้องถิ่นในเขตพื้นที่จังหวัดสุราษฎร์ธานีและพื้นที่ใกล้เคียง นำบทความมาวิเคราะห์โดยใช้เทคนิคการทำเหมืองข้อมูล เพื่อหาความสัมพันธ์ของบทความที่มีการรวบรวมไว้ในเว็บบล็อก

2.5 เหมืองข้อมูล

ข้อมูลสารสนเทศมีความสำคัญ การเผยแพร่และสื่อสารข้อมูล ข่าวสารที่ตรงกับความต้องการของผู้ใช้จึงเป็นสิ่งจำเป็น การประยุกต์เทคโนโลยีสารสนเทศเพื่อช่วยในการสื่อสารข้อมูลจำนวนมากให้แก่ผู้ใช้ เช่น การให้บริการเว็บไซต์เพื่อเผยแพร่ข้อมูลข่าวสารและแลกเปลี่ยนความรู้ จึงเป็นเครื่องมือที่สำคัญในการสื่อสารข้อมูลถึงผู้ใช้จำนวนมาก ดังนั้นการศึกษาเกี่ยวกับพฤติกรรมของผู้ใช้บริการเว็บไซต์ จะช่วยให้องค์กรสามารถนำข้อมูลมาใช้ในการวางแผนพัฒนาเว็บไซต์ ให้ตรงกับความต้องการใช้งานหรือใช้ในการวางแผนกลยุทธ์ เพื่อสร้างความได้เปรียบทางการแข่งขัน

2.5.1 การทำเหมืองข้อมูล

เหมืองข้อมูล (Data Mining) คือ กระบวนการที่เกี่ยวข้องข้อมูลจำนวนมากเพื่อค้นหา รูปแบบและความสัมพันธ์ ที่ซ่อนอยู่ในชุดข้อมูลนั้น ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำไปประยุกต์ใช้ในงานหลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์และการแพทย์รวมทั้งในด้านเศรษฐกิจและสังคม

การทำเหมืองข้อมูลเปรียบเสมือนวิวัฒนาการหนึ่งในการจัดเก็บและตีความหมาย ข้อมูล จากเดิมที่มีการจัดเก็บข้อมูลอย่างง่าย มาสู่การจัดเก็บในฐานข้อมูลที่สามารถดึงข้อมูลสารสนเทศ มาใช้จนถึงการทำเหมืองข้อมูลที่สามารถค้นพบความรู้ที่ซ่อนอยู่ในข้อมูล

2.5.2 ขั้นตอนการทำเหมืองข้อมูล ประกอบด้วยขั้นตอนการทำงานย่อยที่จะเปลี่ยนข้อมูลดิบให้กลายเป็นความรู้ ประกอบด้วยขั้นตอนดังนี้

2.5.2.1 Data Cleaning เป็นขั้นตอนสำหรับการคัดข้อมูลที่ไมเกี่ยวข้องออกไป

2.5.2.2 Data Integration เป็นขั้นตอนการรวมข้อมูลที่มีหลายแหล่งให้เป็นข้อมูลชุดเดียวกัน

2.5.2.3 Data Selection เป็นขั้นตอนการดึงข้อมูลสำหรับการวิเคราะห์จากแหล่งที่บันทึกไว้

2.5.2.4 Data Transformation เป็นขั้นตอนการแปลงข้อมูลให้เหมาะสมสำหรับการใช้งาน

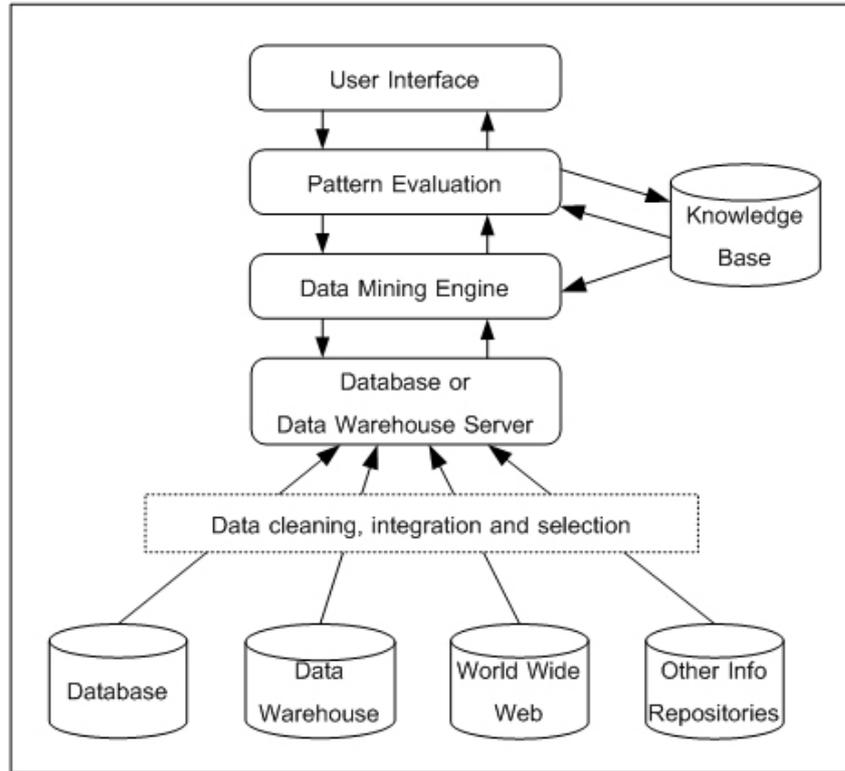
2.5.2.5 Data Mining เป็นขั้นตอนการค้นหารูปแบบที่เป็นประโยชน์จากข้อมูลที่มี

2.5.2.6 Pattern Evaluation เป็นขั้นตอนการประเมินรูปแบบที่ได้จากการทำเหมืองข้อมูล

2.5.2.7 Knowledge Representation เป็นขั้นตอนการนำเสนอความรู้ที่ค้นพบ โดยใช้เทคนิคในการนำเสนอเพื่อให้เข้าใจ

2.5.3 ส่วนประกอบของระบบการทำเหมืองข้อมูล

สถาปัตยกรรมของระบบการทำเหมืองข้อมูล ประกอบด้วยส่วนประกอบที่สำคัญดังภาพที่ 2 - 1



ภาพที่ 2 - 1 สถาปัตยกรรมระบบการทำเหมืองข้อมูล
(ที่มา: อุดุลย์ ยิ้มงาม, 2552)

2.5.3.1 Database, Data Warehouse, World Wide Web และ Other Info Repositories เป็นแหล่งข้อมูลสำหรับการทำเหมืองข้อมูล

2.5.3.2 Database หรือ Data Warehouse Server ทำหน้าที่นำเข้าข้อมูลตามคำขอของผู้ใช้

2.5.3.3 Knowledge Base ได้แก่ ความรู้เฉพาะด้านในงานที่จะเป็นประโยชน์ต่อการสืบค้น หรือประเมินความน่าสนใจของรูปแบบผลลัพธ์ที่ได้

2.5.3.4 Data Mining Engine เป็นส่วนประกอบหลักประกอบด้วยโมดูลที่รับผิดชอบงานทำเหมืองข้อมูลประเภทต่างๆ ได้แก่ การหากฎความสัมพันธ์ การจำแนกประเภท การจัดกลุ่ม

2.5.3.5 Pattern Evaluation Module ทำงานร่วมกับ Data Mining Engine โดยใช้มาตรวัดความน่าสนใจในการกลั่นกรองรูปแบบผลลัพธ์ที่ได้ เพื่อให้การค้นหามุ่งเน้นเฉพาะรูปแบบที่น่าสนใจ

2.5.3.6 Graphic User Interface ส่วนติดต่อประสานระหว่างผู้ใช้กับระบบการทำเหมืองข้อมูล ช่วยให้ผู้ใช้สามารถระบุงานทำเหมืองข้อมูลที่ต้องการทำ ดูข้อมูลหรือโครงสร้างการจัดเก็บข้อมูล ประเมินผลลัพธ์ที่ได้

2.5.4 ประเภทข้อมูลที่ใช้ทำเหมืองข้อมูล

2.5.4.1 Relational Database เป็นฐานข้อมูลที่จัดเก็บอยู่ในรูปแบบของตาราง โดยในแต่ละตารางจะประกอบไปด้วยแถวและคอลัมน์ ความสัมพันธ์ของข้อมูลทั้งหมดสามารถแสดงได้โดย Entity Relationship Model

2.5.4.2 Data Warehouses เป็นการเก็บรวบรวมข้อมูลจากหลายแหล่งมาเก็บไว้ในรูปแบบเดียวกันและรวบรวมไว้ในที่ๆ เดียวกัน

2.5.4.3 Transactional Database ประกอบด้วยข้อมูลที่แต่ละทรานแซกชันแทนด้วยเหตุการณ์ในขณะใดขณะหนึ่ง เช่น ใบเสร็จรับเงิน จะเก็บข้อมูลในรูปแบบชื้อลูกค้าและรายการสินค้าที่ลูกค้ารายชื้อ

2.5.4.4 Advanced Database เป็นฐานข้อมูลที่จัดเก็บในรูปแบบอื่นๆ เช่น ข้อมูลแบบ Object-Oriented ข้อมูลที่เป็น Text File ข้อมูลมัลติมีเดีย ข้อมูลในรูปของ Web การทำเหมืองข้อมูลเว็บ (Web Mining)
(อดุลย์ ยิ้มงาม, 2552)

2.6 การทำเหมืองข้อมูลเว็บ

การทำเหมืองข้อมูลเว็บ คือ การใช้เทคนิคการทำเหมืองข้อมูลเพื่อค้นหาและสกัดข้อมูลและสารสนเทศจากเอกสารเว็บและบริการบนเว็บโดยอัตโนมัติ เพื่อนำความรู้ที่ได้มาแก้ปัญหาที่ต้องการทั้งทางตรงและทางอ้อม นอกจากนี้ยังได้แบ่งประเภทของการทำเหมืองข้อมูลเว็บโดยพิจารณาจากข้อมูลที่นำมาวิเคราะห์ออกเป็น 3 ประเภท คือ Web Content Mining, Web Structure Mining และ Web Usage Mining

2.6.1 Web Content Mining เป็นการค้นหาข้อมูลที่มีประโยชน์จากข้อมูลที่อยู่ภายในเว็บ เช่น ข้อความ รูปภาพ เป็นต้น โดย Web Content Mining สามารถแบ่งออกเป็น 2 ประเภทตามมุมมองคือ มุมมองทางด้านการสืบค้นสารสนเทศ (Information Retrieval) และมุมมองทางด้านฐานข้อมูล (Database) สำหรับเป้าหมายของ Web Content Mining จากมุมมองของการสืบค้นสารสนเทศคือการทำเหมืองข้อมูลเว็บเพื่อปรับปรุงการหา ข้อมูลหรือกรองข้อมูลให้ผู้ใช้โดยพิจารณาจากข้อมูลที่ผู้ใช้อ้างอิงหรือร้องขอในขณะที่เป้าหมายของ Web Content Mining ในมุมมองของฐานข้อมูลส่วนใหญ่

พยายามจำลองข้อมูลบนเว็บและรวมข้อมูลนั้น เพื่อให้การสอบถามทำงานดีขึ้นมากกว่าการใช้คำสั่งหลักเป็นตัวค้นหาเพียงอย่างเดียว

2.6.2 Web Structure Mining เป็นวิธีการที่พยายามค้นหารูปแบบโครงสร้างการเชื่อมโยงที่สำคัญและซ่อนอยู่ในเว็บ ซึ่งรูปแบบนี้จะขึ้นอยู่กับรูปแบบการเชื่อมโยงเอกสารภายในเว็บ โดยนำรูปแบบที่ได้มาใช้เพื่อจัดกลุ่มเว็บเพจและใช้สร้างข้อมูลสารสนเทศที่เป็นประโยชน์ เช่น นำมาใช้ในการปรับโครงสร้างของเว็บให้สามารถให้บริการผู้ใช้ได้อย่างรวดเร็ว

2.6.3 Web Usage Mining เป็นวิธีการที่พยายามค้นหาความหมายของข้อมูลที่สร้างจากช่วงการทำงานหนึ่ง ของผู้ใช้หรือสร้างจากพฤติกรรมของผู้ใช้เรียกอีกชื่อหนึ่งว่า Web Log Mining โดยในขณะที่ Web Content Mining และ Web Structure Mining ใช้ประโยชน์จากข้อมูลจริง หรือข้อมูลพื้นฐานบนเว็บแต่ Web Usage Mining ทำการค้นหาความรู้จากข้อมูลการติดต่อสื่อสารระหว่างกันของผู้ใช้ที่ติดต่อกับเว็บ โดย Web Usage Mining ทำการรวบรวมข้อมูลจากบันทึกในการดำเนินการต่างๆ เช่น บันทึกการใช้งานของ Proxy (Proxy Server Log) ข้อมูลการลงทะเบียน (Registration Data) หรือข้อมูลอื่นอันเป็นผลจากการทำงานร่วมกันมาใช้วิเคราะห์ ดังนั้น Web Usage Mining จึงเป็นวิธีการทำงานที่เน้นใช้เทคนิคที่สามารถทำนายพฤติกรรมของผู้ใช้ใน ขณะที่ผู้ใช้ทำงานกับเว็บ กระบวนการทำงานของ Web Usage Mining สามารถแบ่งออกเป็น 2 วิธี คือ

2.6.3.1 ทำการจับคู่ข้อมูลการใช้งานของเครื่องให้บริการเว็บให้อยู่ในรูปของตารางความสัมพันธ์ ก่อนที่นำข้อมูลนี้มาปรับใช้กับเทคนิคการทำเหมืองข้อมูลการใช้เว็บ

2.6.3.2 ใช้ประโยชน์จากข้อมูลในบันทึกการใช้งานโดยตรงซึ่งจะใช้เทคนิคการเตรียมข้อมูล (Preprocessing) เพื่อเตรียมข้อมูลก่อนหาความสัมพันธ์ (Pattern Discovery) และวิเคราะห์รูปแบบ (Pattern Analysis) (Yang, Qiang, Hui Wang, & Wei Zhang, 2002 และ อุดลย์ ยี่งาม, 2552)

2.7 เทคนิคการทำเหมืองข้อมูลการใช้เว็บไซต์

เทคนิคสำหรับการทำเหมืองข้อมูล การใช้เว็บไซต์ดังนี้

2.7.1 Statistical Analysis การวิเคราะห์ทางด้านสถิติ เป็นวิธีการที่จะค้นหาความรู้เกี่ยวกับผู้เยี่ยมชมเว็บไซต์ โดยการวิเคราะห์ส่วนของไฟล์ เป็นวิธีหนึ่งที่สามารถบอกชนิดความแตกต่างเกี่ยวกับการวิเคราะห์ทางสถิติบน ตัวแปร เช่น หน้าที่ดู เวลาการดูและความยาวของเส้นทางเว็บไซต์ เครื่องมือการวิเคราะห์การจราจรเว็บส่วนมากจะรายงานข้อมูลเกี่ยวกับสถิติ เช่น ความถี่ของหน้าที่เข้าถึงบ่อยๆ เวลาเฉลี่ยของหน้าหรือความยาวเฉลี่ยของเส้นทางเว็บไซต์ รายงานนี้อาจจะรวมถึงการวิเคราะห์ข้อผิดพลาดที่จำกัด เช่น การสืบทายการที่ไม่ให้สิทธิเข้าถึงหรือการค้นหา ยูอาร์แอล (url) ที่ไม่สมบูรณ์ รูปแบบของความรู้สามารถเป็นประโยชน์อย่างมากสำหรับการปรับปรุงระบบ การเพิ่มความปลอดภัยของระบบ ทำให้สะดวกในการปรับปรุงเว็บไซต์และการสนับสนุนการตัดสินใจทางการตลาด

2.7.2 Association Rule เป็นการค้นหาความสัมพันธ์ของข้อมูล โดยค้นหาความสัมพันธ์ของข้อมูลทั้งสองชุดหรือมากกว่าสองชุดขึ้นไปไว้ด้วยกัน ความสำคัญของกฎทำกวัดโดยใช้ข้อมูลสองตัวด้วยกันคือค่าสนับสนุน (Support) ซึ่งเป็นเปอร์เซ็นต์ของการดำเนินการที่กฎสามารถนำไปใช้ หรือ

เป็นเปอร์เซ็นต์ของการดำเนินการที่กฎที่ใช้มีความถูกต้อง และข้อมูลตัวที่สองที่นำมาใช้วัดคือค่าความมั่นใจ (Confidence) ซึ่งเป็นจำนวนของกรณีที่ถูกถูกต้องโดยสัมพันธ์กับจำนวนของกรณีที่กฎสามารถนำไปใช้ได้ ในการหาความสัมพันธ์นั้นจะมีขั้นตอนวิธีการหาหลายวิธีด้วยกัน แต่ขั้นตอนวิธีที่เป็นที่รู้จักและใช้อย่างแพร่หลายคือ อัลกอริทึม Apriori

2.7.3 Clustering คือ การจัดกลุ่มข้อมูลซึ่งมีลักษณะคล้ายกัน โดยการแบ่งประเภทจะวิเคราะห์ข้อมูลตามต้นแบบ การแบ่งกลุ่มเป็นการวิเคราะห์โดยไม่พิจารณาจัดกลุ่มตามประเภท แต่จะใช้ขั้นตอนวิธีการจัดกลุ่มเพื่อค้นหากลุ่มที่สามารถยอมรับได้เพื่อจัด เข้ากลุ่ม กล่าวคือ กลุ่มของวัตถุมีการสร้างขึ้นโดยเปรียบเทียบวัตถุที่มีความเหมือนกันจัดเข้ากลุ่มเดียวกัน

2.7.4 Classification เป็นการจัดแบ่งประเภทของข้อมูล โดยหาชุดต้นแบบหรือชุดของการทำงานที่อธิบายและแบ่งประเภทข้อมูล วัตถุประสงค์เพื่อให้สามารถใช้เป็นต้นแบบทำนายประเภทของวัตถุหรือข้อมูลที่ไม่มีการระบุประเภทหรือชนิดของข้อมูล ซึ่งต้นแบบสร้างจากการวิเคราะห์ชุดของข้อมูลฝึกสอน (Training Data) โดยอาจจะเป็นกลุ่มข้อมูลที่มีการระบุประเภทหรือกลุ่มเรียบร้อยแล้ว รูปแบบของต้นแบบแสดงได้หลายแบบเช่น Classification Rules, Decision Trees หรือ Neural Networks เป็นต้น

2.7.5 Sequential Pattern เป็นเทคนิคการสืบค้นรูปแบบลำดับเหตุการณ์ เป็นการหาความสัมพันธ์ของข้อมูลระหว่างทรานแซกชันซึ่งทำให้มีเวลาเข้ามา เกี่ยวข้อง ด้วยรูปแบบลำดับที่น่าสนใจ จะแสดงว่าถ้าเกิดเหตุการณ์นี้แล้วหรือพบกลุ่มของข้อมูลชุดนี้แล้วมีแนวโน้ม ที่จะเกิดเหตุการณ์อะไรหรือพบกลุ่มของข้อมูลใดตามมาจากภายหลัง

2.7.6 Dependency Modeling เป็นแบบจำลองการพึ่งพา เป็นการค้นพบรูปแบบอื่นๆ ที่มีประโยชน์ในการทำเหมืองข้อมูลเว็บ เป้าหมายคือการพัฒนาต้นแบบที่สามารถทำได้ของตัวแทนที่สำคัญของการพึ่งพา ตัวแปรต่างๆ ในโดเมนเว็บ เช่น อาจจะใช้วิธีการสร้างแบบจำลองสถานที่แตกต่างกัน ให้ผู้เยี่ยมชมได้เห็นในขณะที่เลือกซื้อสินค้าทางออนไลน์ มีหลายเทคนิคการเรียนรู้ที่เป็นไปได้ ถูกใช้เป็นแบบของการค้นหาพฤติกรรมของผู้ใช้ เช่น ตัวแบบฮิดเดนมาร์คอฟ (Hidden Markov Model) และแนวคิดข่ายงานความเชื่อเบย์ (Bayesian Belief Network)

ในงานวิจัยขั้นนี้ผู้วิจัยได้วิเคราะห์ความสัมพันธ์ (Data Mining) ของเนื้อหาภายในบทความตลอดจนการจัดหมวดหมู่ของบทความ (Clustering) เพื่อการอธิบาย เป็นการค้นหารูปแบบจากกลุ่มข้อมูล ซึ่งจะมีความสัมพันธ์หรือลักษณะที่เชื่อมโยงกันของข้อมูล (Association Rule) โดยใช้เทคนิคเหมืองข้อมูล (Data Mining) จากบทความที่มีการจัดเก็บรวบรวมไว้ในเว็บบล็อก เพื่อนำการวิเคราะห์มาพัฒนาเป็นโมเดลในการพัฒนาระบบธนาคารสมอง คลังปัญญา เพื่อจัดเก็บภูมิปัญญาท้องถิ่น และสร้างเป็นระบบจัดเก็บและสืบค้นเพื่อพัฒนาเป็นนวัตกรรมต่อไป (Mohan, E.K. & Han, 2011)

2.8 งานวิจัยที่เกี่ยวข้อง

2.8.1 งานวิจัยในประเทศ

สถาพร สังข์ศิริ (2553) การวิจัยครั้งนี้ มีวัตถุประสงค์เพื่อพัฒนาหาประสิทธิภาพการ ค้นหาข้อมูล และหาความพึงพอใจในระบบสารสนเทศเพื่อจัดการความรู้ภูมิปัญญาท้องถิ่น กรณีศึกษา

การเลี้ยงปลาในกระชัง แม่น้ำสะแกกรัง จังหวัดอุทัยธานี โดยการพัฒนาระบบสารสนเทศดังกล่าว ช่วยส่งเสริมและสนับสนุนการนำแนวคิดการจัดการความรู้และภูมิปัญญาท้องถิ่น ที่มีอยู่เป็นจำนวนมาก ซึ่งเป็นประโยชน์ต่อสังคมและชุมชน และข้อมูลดังกล่าวยังไม่ได้จัดเก็บให้เป็นระบบเรียบร้อย และง่ายต่อการสืบค้น ระบบสารสนเทศที่พัฒนาขึ้น ได้นำออนโทโลยี (Ontology) มาออกแบบโครงสร้าง ที่ได้รับการรับรองจากผู้เชี่ยวชาญด้านประมง แล้วทำการวิเคราะห์ข้อมูล โดยใช้ภาษา OWL (Web Ontology Language) ที่เป็นโครงสร้าง และระบบ RAP (RDF API For PHP) ที่สนับสนุนการวิเคราะห์ตามกระบวนการจัดการความรู้ ให้ตรงต่อความต้องการของผู้ใช้ กลุ่มตัวอย่างที่ใช้ระบบสารสนเทศที่พัฒนาเป็น ประชาชนในพื้นที่ตำบลอุทัยใหม่ ตำบลสะแกกรัง ตำบลเกาะเทโพ ตำบลท่าซุง ตำบลหาดตะนง ตำบลท่าซุง และบุคคลทั่วไป รวม 40 คน โดยทำการเลือกแบบเจาะจง (Purposive Sampling)

ผลการวิจัยพบว่า ระบบมีค่าความแม่นยำ (Precision) อยู่ที่ 0.9503 ประสิทธิภาพในการค้นหาข้อมูลตามวงจรการเลี้ยงปลาในกระชังได้อย่างแม่นยำอยู่ถึงร้อยละ 95.03 สำหรับค่าความระลึก (Recall) อยู่ที่ 0.9752 สามารถบ่งชี้ว่า ระบบสารสนเทศค้นหาข้อมูลที่มีอยู่จริงได้อย่างถูกต้องถึงร้อยละ 97.52 ระบบมีอัตราความผิดพลาดในกรณีที่ไม่สามารถค้นหาข้อมูลที่มีอยู่ในระบบ ประมาณ ร้อยละ 2.48 สำหรับประสิทธิภาพโดยภาพรวมของระบบ ผู้วิจัยได้นำค่าความแม่นยำและความระลึภมาคำนวณหาค่าความเหวี่ยง (F-measure) พบว่าระบบมีประสิทธิภาพในการค้นหาในภาพรวมอยู่ที่ร้อยละ 94.57 ระบบมีประสิทธิภาพในการค้นหาในภาพรวมอยู่ในระดับดีมาก สำหรับผลการประเมินความพึงพอใจที่มีต่อระบบสารสนเทศที่พัฒนาขึ้น อยู่ในระดับมาก

ศิริกาญจนา พิลาบุตร (2551) ได้ทำวิจัยเรื่อง การสร้างกฎข้อบังคับของฐานข้อมูลโดยการทำให้เหมือนข้อมูล ดังนี้ ระบบฐานข้อมูลเชิงสัมพันธ์ (Relational Database System) เป็นระบบฐานข้อมูลที่มีผู้นิยมใช้กันมาก ในการจัดเก็บข้อมูลลงในฐานข้อมูลให้มีความถูกต้องตามหลักของระบบฐานข้อมูลและความต้องการของแต่ละระบบงานเป็นเรื่องยากหากจะต้องป้องกันความผิดพลาดดังกล่าวโดยการเขียนโปรแกรม การจัดเก็บข้อมูลที่ไม่ถูกต้องลงในฐานข้อมูลจะส่งผลให้การประมวลผลจากข้อมูลดังกล่าวเกิดความผิดพลาด และต้องใช้เวลาในการแก้ไขจัดการข้อมูลให้มีความถูกต้อง ดังนั้น ระบบฐานข้อมูล จึงมีตัวคำสั่งเบสทริกเกอร์ (Database Trigger) หรือกฎข้อบังคับของฐานข้อมูลทำหน้าที่ตรวจสอบข้อบังคับของข้อมูลในการประมวลผลคำสั่ง SQL (Structured Query Language) ประเภท DML (Data Manipulation Language) ซึ่งเป็นคำสั่งที่ใช้ในการเปลี่ยนแปลงข้อมูลในฐานข้อมูล ให้มีความถูกต้องตามความต้องการของระบบ ในปัจจุบันการสร้างตัวเบสทริกเกอร์ให้ตรงตามความต้องการของระบบนั้น กระทำได้ด้วยผู้ดูแลจัดการฐานข้อมูลเป็นผู้กำหนดความถูกต้องของข้อมูล โดยพิจารณาจากความต้องการของระบบงาน ซึ่งอาจจะเกิดการผิดพลาดและสิ้นเปลืองเวลาได้หากมีการเปลี่ยนแปลงความต้องการของระบบงานขึ้นใหม่ ดังนั้นในงานวิจัยนี้จึงเสนอแนวคิดในการสร้างตัวเบสทริกเกอร์ โดยใช้กฎ (Rule) ที่ได้จากการทำเหมืองข้อมูล (Data Mining) มาเป็นตัวสร้างตัวเบสทริกเกอร์ขึ้นมาเอง เพื่อการลดระยะเวลาและเพิ่มความถูกต้องสูงสุดให้กับฐานข้อมูล

ธนากร สุวรรณวัฒน์ (2550) สารนิพนธ์ฉบับนี้ทำการพัฒนาระบบบริการช่วยเหลือเพื่อตอบปัญหาการติดตั้งและการใช้งานอุปกรณ์ระบบ Call Center โดยใช้ออนโทโลยีเป็นแนวทางในการ

พัฒนาระบบเพื่อให้บริการในการเรียกค้นข้อมูลปัญหาและวิธีการแก้ไขปัญหาของการติดตั้งและการใช้งานอุปกรณ์โดยระบบงานจะทำการค้นหาข้อมูลในฐานความรู้ซึ่งมีการจัดเก็บข้อมูลความรู้ซึ่งอยู่ในรูปแบบไฟล์เชิงความหมายที่ได้ถูกอธิบายขึ้นโดยใช้ออนโทโลยี ซึ่งไฟล์มีการอธิบายข้อมูลปัญหาและวิธีการแก้ไขปัญหาของการติดตั้งและการใช้งานอุปกรณ์ระบบ Call Center ซึ่งถูกรวบรวมขึ้นจากคู่มือการติดตั้งและการใช้งานอุปกรณ์และจากประสบการณ์ของผู้เชี่ยวชาญในการให้บริการช่วยเหลือระบบบริการช่วยเหลือเพื่อตอบปัญหาให้บริการในการค้นหาข้อมูลโดยสามารถกำหนดวิธีการ ค้นหาใน 2 ลักษณะ คือ การค้นหาด้วยการระบุคีย์เวิร์ดและการค้นหาโดยพิจารณาความสัมพันธ์ของปัญหาจากการอนุมานออนโทโลยี ซึ่งการค้นหาด้วยการอนุมานนี้ทำให้สามารถทำการค้นหาข้อมูลที่เกี่ยวข้องได้อย่างยืดหยุ่น และจากการทดลองด้วยการกำหนดคำสั่งในการค้นหาพบว่าการค้นหาโดยพิจารณาความสัมพันธ์ของปัญหาจากการอนุมานออนโทโลยีมีค่าเฉลี่ยรีคอลสูงกว่าการค้นหาด้วยการระบุคีย์เวิร์ด และมีค่าเฉลี่ยความแม่นยำเท่าเดิมหรือเพิ่มขึ้น เมื่อเปรียบเทียบกับการค้นหาด้วยการระบุคีย์เวิร์ด

นฤป รังงาม (2550) ได้นำเสนอระบบการวิเคราะห์คำอัตโนมัติ เพื่อสร้างระบบที่สามารถทำนายสาเหตุของปัญหาการใช้งานอินเทอร์เน็ตของลูกค้า และนำเสนอแนวทางแก้ไข โดยประยุกต์ใช้เทคนิคการสกัดคำถาม (Data Extraction) ใช้ร่วมกับเทคนิคการทำเหมืองข้อมูล (Data Mining) เทคนิคการสกัดคำถามนั้นใช้สำหรับสกัดข้อมูลที่สัมพันธ์กันจากคำถามที่เกี่ยวข้องกับการใช้งานอินเทอร์เน็ตลูกค้า เพื่อใช้เป็นข้อมูลนำเข้าในขั้นตอนการทำเหมืองข้อมูลต่อไป ขั้นตอนของการทำเหมืองข้อมูลของงานวิจัยนี้แบ่งเป็น 2 กรณี คือ (1) กรณีที่สามารถเชื่อมต่ออินเทอร์เน็ตได้ (Online) จะใช้การจัดกลุ่มข้อมูลแบบ 2 ขั้นตอน (SOM และ K-means อัลกอริทึม) ในการจัดกลุ่มของลูกค้า และใช้ OLAP Cube ในการวิเคราะห์เพื่อหาความน่าจะเป็นของสาเหตุที่ทำให้เกิดปัญหา (2) กรณีที่ไม่สามารถเชื่อมต่ออินเทอร์เน็ตได้ (Offline) จะใช้โปรแกรมตัวแทน (Software Agent) ในการตรวจสอบปัญหาและใช้ต้นไม้ตัดสินใจ (Decision Tree) เพื่อหาคำตอบของปัญหาที่เกิดขึ้น ในงานวิจัยนี้ใช้ค่า Precision ในการวัดความแม่นยำ โดยผลที่ได้จากการใช้ OLAP Cube จะให้ค่าความแม่นยำ ร้อยละ 70.46 ส่วนการใช้ Decision Tree ได้ให้ค่าความแม่นยำร้อยละ 87.56 ผลลัพธ์ที่ได้จากการใช้ระบบนั้นนอกจากจะช่วยลดปริมาณการโทรสอบถามมายังพนักงาน Call Center แล้ว ยังเอื้อประโยชน์แก่ลูกค้าทั่วไปที่ไม่มีความรู้ในเรื่องอินเทอร์เน็ตให้สามารถตรวจสอบแก้ไข ปัญหาด้วยตนเองได้

นิภาวรรณ ลาภบุญเรือง (2550) ได้ทำวิจัยเรื่อง การทำเหมืองข้อมูลชนิด Expressed Sequence Tags (ESTs) เพื่อการพัฒนาเครื่องหมาย SSR สำหรับถั่วลิสง (*Arachis Hypogaea* L.) โดยมีรายละเอียดดังนี้ EST-derived SSRs คือ SSRs ที่มีตำแหน่งอยู่ภายในลำดับนิวคลีโอไทด์ชนิด Expressed Sequence Tags (ESTs) การพัฒนาเครื่องหมาย eSSRs นั้นต้องอาศัยเครื่องมือทางเทคโนโลยีชีวสารสนเทศร่วมกับเทคนิคปฏิบัติการทางชีวโมเลกุล เครื่องหมาย eSSRs ที่พัฒนาขึ้นใหม่สามารถใช้เป็นเครื่องหมายพันธุกรรมสำหรับการศึกษพันธุศาสตร์ และการปรับปรุงพันธุ์ของพืชปลูกได้ งานวิจัยครั้งนี้มีวัตถุประสงค์เพื่อพัฒนาเครื่องหมาย eSSRs สำหรับถั่วลิสง (*Arachis Hypogaea* L.) โดยจัดเก็บ ESTs ของถั่วลิสงที่สร้างจาก eDNA Library ของพันธุ์ C34-24 จำนวน 492 ชิ้น จากฐานข้อมูลสาธารณะ GenBank การวิเคราะห์ ESTs เริ่มต้นจากการประเมินคุณภาพคือไม่มีการ

ปนเปื้อนจากลำดับเบสของเวกเตอร์ (vector) ขึ้น ESTs มีความยาวอย่างน้อย 100 คู่เบส มีนิวคลีโอไทด์ที่ไม่ทราบชนิด (N) ต่ำกว่า 5 เปอร์เซ็นต์ และไม่มีบริเวณ Low Complexity ผลจากการวิเคราะห์พบว่า ESTs จำนวน 445 ขึ้นผ่านเกณฑ์ประเมินคุณภาพ จากนั้นนำ ESTs ที่มีคุณภาพสูงไปจัดคลัสเตอร์ โดยพิจารณาความเหมือนของ ESTs ด้วยค่า E Value ซึ่งเท่ากับ 1×10^{-200} (ยกกำลัง -200) ขึ้นไป ผลการจัดคลัสเตอร์ได้ ESTs จำนวน 356 คลัสเตอร์ รวมความยาวทั้งสิ้น 199,745 คู่เบส เมื่อนำ ESTs 356 คลัสเตอร์ไปค้นหา eSSRs โดยใช้ซอฟต์แวร์ SSRIT โดยใช้เกณฑ์ตัดลิ้น eSSRs ให้ dimer ถึง hexamer มีจำนวนซ้ำเท่ากับหรือมากกว่า 4 ซ้ำ ได้ผลว่า 94 คลัสเตอร์ จาก 356 คลัสเตอร์ คิดเป็น 26.40 เปอร์เซ็นต์ มี eSSRs 107 ตำแหน่ง คิดเป็นความถี่ 1 ต่อ 1,800 คู่เบส ชนิดของ eSSRs ที่พบมากที่สุดคือ dimer เท่ากับ 59.81 เปอร์เซ็นต์ และ motif ที่พบมากที่สุดคือ TATA เท่ากับ 17.76 เปอร์เซ็นต์ และเมื่อเปลี่ยนเกณฑ์ตัดลิ้น eSSRs ให้ dimer จำนวนเท่ากับหรือมากกว่า 6 ซ้ำ และ trimer ถึง hexamer จำนวนเท่ากับหรือมากกว่า 4 ซ้ำ พบ 41 คลัสเตอร์ จาก 356 คลัสเตอร์ ซึ่งคิดเป็น 11.52 เปอร์เซ็นต์ มี eSSRs 47 ตำแหน่ง คิดเป็นความถี่ 1 ต่อ 4,200 คู่เบส ชนิดของ eSSRs ที่พบมากที่สุดคือ trimer เท่ากับ 80.85 เปอร์เซ็นต์ และ motif ที่พบมากที่สุดคือ AAT/ATT , ATC/GAT , GAA/TTC และ TCA/TGA เท่ากับ 8.51 เปอร์เซ็นต์เท่านั้น จากนั้นออกแบบไพรเมอร์ที่จำเพาะต่อ eSSRs จำนวน 14 คู่ และไพรเมอร์ไปทดสอบกับถั่วลิสง 12 พันธุ์ คือ PI 109839 , Khon Kaen 60-3 , PI 262064 , Khon Kean 4 , NC Ac 17090 , Tainan 9 , PI 155246 , ICGV 98324 , ICG 11109 , ICG 13099 , A 13 และ IC 34 พบว่าให้จำนวนแอลลีลทั้งหมด 47 แอลลีลเฉลี่ย 3.36 แอลลีลต่อโลคัส แอลลีลมีขนาด 115 ถึง 2,241 คู่เบส และค่า Polymorphic Information Content (PIC) เท่ากับ 0 ถึง 0.82 ค่าเฉลี่ยของ PIC รวมเท่ากับ 0.39

พิจิตรา จอมศรี (2549) ได้ทำวิจัยเรื่อง การทำนายเนื้อหาของเว็บโดยใช้เทคนิคเหมืองข้อมูล กรณีศึกษา มหาวิทยาลัยศิลปากร ดังนี้ งานวิจัยนี้ได้นำเทคนิคการค้นหาคำสัมพันธ์ ซึ่งเป็นเทคนิคหนึ่งในเทคนิคเหมืองข้อมูลมาประยุกต์ใช้สร้างตัวแบบเพื่อทำนายข้อมูลการเรียกใช้เว็บในอนาคต โดยผู้วิจัยได้จัดเก็บข้อมูล 2 ส่วนคือ การเรียกใช้เว็บภายในมหาวิทยาลัยศิลปากรจากระบบพร็อกซี เซิร์ฟเวอร์ และจัดทำฐานข้อมูลหมวดหมู่เว็บ แล้วนำข้อมูลทั้ง 2 ส่วนมาสร้างความสัมพันธ์ โดยงานวิจัยนี้ นำข้อมูลวัน เวลา หมวดเว็บ และเว็บ มาค้นหาคำสัมพันธ์เพื่อสร้างตัวแบบ โดยพิจารณาตัวแบบจากความเชื่อมั่นและค่าสนับสนุน ผู้วิจัยได้ทำการทดลองศึกษาตัวแบบ โดยผลของการศึกษาตัวแบบพบว่าโมเดลที่สร้างขึ้นสามารถทำนายเนื้อหาเว็บที่จะถูกเรียกใช้ในวันถัดมาได้ ผลของการใช้เทคนิคเหมืองข้อมูลพบว่าตัวแบบที่สร้างขึ้นสามารถทำนายเนื้อหาเว็บที่จะถูกเรียกใช้ได้ โดยมีความถูกต้องร้อยละ 66.67% นอกจากนี้ผู้วิจัยยังนำเสนอระบบการทำนายเนื้อหาของเว็บโดยใช้เทคนิคเหมืองข้อมูลซึ่งสร้างอยู่บนขั้นตอนวิธีที่นำเสนอและอธิบายถึงผลการทดลองบนข้อมูลจริงด้วย ซึ่งงานวิจัยนี้สามารถทำนายเนื้อหาเว็บที่ถูกเรียกใช้ในอนาคตได้ จึงสามารถเพิ่มประสิทธิภาพการทำงานของระบบพร็อกซีเซิร์ฟเวอร์ได้ ซึ่งถ้าประสิทธิภาพการเรียกใช้เว็บเพิ่มขึ้นและสามารถลดปริมาณข้อมูลในระบบเครือข่ายได้ อย่างไรก็ตามเทคนิคนี้ยังไม่สามารถครอบคลุมการทำงานในช่วงเหตุการณ์ที่ไม่เป็นปกติ เช่น อุบัติเหตุ และเทศกาลต่างๆ เป็นต้น

สุวนีย์ กุลกรนิธธรรม (2549) วิทยานิพนธ์นี้มีวัตถุประสงค์เพื่อพัฒนาตัวแบบสำหรับจัดกลุ่มหลักสูตรตามกลุ่ม สาขาวิชา ISCED ในกลุ่มสาขาวิชาวิทยาศาสตร์ ด้วยเทคนิคเหมืองข้อมูล โดยใช้

เทคนิคจำแนกประเภทข้อมูลแบบเบย์ ข้อมูลที่ใช้ในการพัฒนาตัวแบบเป็นข้อมูลหลักสูตรในกลุ่มสาขาวิชาวิทยาศาสตร์จำนวน 7 สาขาวิชาของสถาบันอุดมศึกษาของรัฐจำนวน 8 แห่งที่ผ่านการรับรองมาตรฐานหลักสูตรจากสำนักงานคณะกรรมการการอุดมศึกษาระหว่างปี พ.ศ. 2537-2547 โดยใช้คำอธิบายรายวิชาภาษาอังกฤษในการพัฒนาตัวแบบ กลุ่มสาขาวิชา ISCED ที่ใช้เป็นเกณฑ์ในการจัดกลุ่มหลักสูตรเป็นมาตรฐานขององค์กร UNESCO ตามเกณฑ์มาตรฐานของปี พ.ศ.2540 การวิจัยได้แบ่งขั้นตอนการทำงานเป็น 2 ส่วน คือ ขั้นตอนการสร้างการเรียนรู้ข้อมูล และขั้นตอนการทดสอบข้อมูล โดยใช้ทฤษฎีการจำแนกประเภทข้อมูลแบบเบย์มาใช้ในการหาความน่าจะเป็นของแต่ละกลุ่มสาขาวิชา

ผลของการวิจัยภายใต้สัดส่วนข้อมูลระหว่าง Training Set และ Validation Set พบว่า ถ้าแบ่งสัดส่วนข้อมูล Training Set และ Validation Set เป็นร้อยละ 60:40 ตัวแบบที่พัฒนาขึ้นมีค่าความถูกต้องเฉลี่ยเป็นร้อยละ 78.57 ถ้าแบ่งสัดส่วนข้อมูลเป็นร้อยละ 70:30 ตัวแบบมีค่าความถูกต้องเฉลี่ยเป็นร้อยละ 85.71 และถ้าแบ่งสัดส่วนเป็นร้อยละ 80:20 ตัวแบบมีค่าความถูกต้องเฉลี่ยเป็นร้อยละ 90.48 สรุปผลได้ว่าเมื่อสัดส่วนของ Training Set มากขึ้นทำให้มีความถูกต้องของผลลัพธ์มากขึ้นตามลำดับ ดังนั้นในงานวิจัยนี้จึงสามารถนำตัวแบบที่พัฒนาขึ้นมาใช้ในการจัดกลุ่มหลักสูตรตามกลุ่มสาขาวิชา ISCED ในกลุ่มสาขาวิชาวิทยาศาสตร์ได้

บุญมา เฟ่งชวน (2548) เนื่องจากความต้องการของตลาดมีการเปลี่ยนแปลงไปตามสภาพการเจริญเติบโตของเศรษฐกิจ ณ ช่วงเวลาหนึ่ง ตลาดแรงงานอาจต้องการบุคคลที่มีความรู้ความสามารถด้านวิศวกรรมศาสตร์ คอมพิวเตอร์ แต่เมื่อเวลาผ่านไปความต้องการของตลาดก็อาจเปลี่ยนไป ดังนั้น สถาบันการศึกษาซึ่งมีหน้าที่ผลิตบัณฑิตจะต้องมีการกำหนดเป้าหมายการผลิตบัณฑิต เพื่อรองรับการขยายตัวของเศรษฐกิจ การวิจัยครั้งนี้จึงได้นำข้อมูลภาวะการณ์มีงานทำของบัณฑิตมาวิเคราะห์โดยใช้เทคนิคเหมืองข้อมูล คือ การค้นหากฎความสัมพันธ์ ซึ่งมีวัตถุประสงค์เพื่อศึกษาและพัฒนาตัวแบบเพื่อใช้ในการทำนายแนวโน้มเลือกอาชีพแรกหลังสำเร็จการศึกษาของนักศึกษาระดับปริญญาตรี และนำข้อมูลรายบุคคลนักศึกษาของสถาบัน อุดมศึกษาของรัฐ ปีการศึกษา 2548 มาวิเคราะห์กับตัวแบบที่สร้างได้ และนำเสนอในรูปของรายงานแบบตารางและกราฟเพื่อนำไปสนับสนุนการตัดสินใจในด้านการผลิตบัณฑิต

ผลการวิจัยพบว่าตัวแบบที่พัฒนาขึ้นมีค่าเฉลี่ยร้อยละความถูกต้องเป็น 65.39 ถ้าแบ่งสัดส่วนข้อมูล Training Set เป็น ร้อยละ 60 และค่าเฉลี่ยร้อยละความถูกต้องเป็น 74.72 ถ้าแบ่งสัดส่วนข้อมูล Training Set เป็นร้อยละ 70 และค่าเฉลี่ยร้อยละความถูกต้องเป็น 81.89 ถ้าแบ่งสัดส่วนข้อมูล Training Set เป็นร้อยละ 80 จะเห็นได้ว่าเมื่อสัดส่วนของ Training Set มากขึ้นทำให้มีความถูกต้องของผลลัพธ์มากขึ้น ซึ่งสอดคล้องกับการใช้เทคนิคการสุ่มตัวอย่างที่เป็นตัวแทนของประชากรมากขึ้น ซึ่งไม่เพียงแต่ให้สารสนเทศมากขึ้นเท่านั้น แต่ยังให้ความถูกต้องมากขึ้นด้วย ดังนั้นในงานวิจัยนี้จึงสามารถนำตัวแบบมาหาแนวโน้มการเลือกอาชีพแรกหลังสำเร็จการศึกษาของนักศึกษาระดับปริญญาตรีได้ และนำตัวแบบมาหาแนวโน้มการเลือกอาชีพแรกหลังสำเร็จการศึกษาของนักศึกษาระดับปริญญาตรีได้ และนำข้อมูลรายบุคคลนักศึกษามาประมวลผลกับข้อมูลตัวแบบเพื่อนำไปสนับสนุนการตัดสินใจด้านการผลิตบัณฑิตระดับปริญญาตรีต่อไป

กฤษณะ ไวยมัย และ ชีระวัฒน์ พงษ์ศิริปริดา (2546) ได้นำเทคนิคการจำแนกประเภทข้อมูลมาสร้างตัวจำแนกข้อมูล ในการวิจัยได้นำเทคนิคการจำแนกประเภท และเทคนิคการค้นหากฎความสัมพันธ์ มาประยุกต์ใช้ในการจัดสรรกฎหมายที่เหมาะสมกับคดีความ โดยนำเทคนิคการจำแนกประเภทข้อมูลมาสร้างตัวจำแนกข้อมูลจากกฎเกณฑ์ที่ได้จากเทคนิคการค้นหากฎความสัมพันธ์อีกทีหนึ่ง ซึ่งตัวจำแนกข้อมูลที่ได้นี้จะสามารถนำไปใช้ทำนายคดีความแต่ละคดีว่าควรใช้กฎหมายฉบับใดในการพิจารณา ผลการวิจัยที่ได้แสดงให้เห็นว่าการสร้างตัวจำแนกตามวิธีที่เสนอนี้ได้ประสิทธิภาพดีกว่าการสร้างตัวจำแนกตามเทคนิค Data Classification แบบปกติ นอกจากนี้ วิธีดังกล่าวยังช่วยแก้ปัญหาต่างๆที่เกิดขึ้นจากการใช้เทคนิคการจำแนกประเภทข้อมูลโดยทั่วไปได้อีกด้วย

ธนาวิทย์ รักษรรณานนท์ (2546) ได้ทำวิจัยเรื่อง การพัฒนาระบบสืบค้นความรู้ : การหา กฎความสัมพันธ์บนฐานข้อมูลเชิงวัตถุแบบแน่น ดังนี้ บ่อยครั้งที่ข้อมูลหนึ่งๆสามารถสร้างลำดับชั้นได้มากกว่า 1 รูปแบบและ/หรือลำดับชั้นของวัตถุถูกเปลี่ยนแปลงเพื่อให้เหมาะสมอยู่ตลอดเวลา แต่การหาความสัมพันธ์หลายระดับชั้นในปัจจุบันซึ่งเป็นแนวทางการค้นหาแบบบนลงล่างทั้งสิ้นมีขีดจำกัดที่สำคัญคือลำดับชั้นของวัตถุต้องไม่มีการเปลี่ยนแปลงแต่หากพบว่าการเปลี่ยนแปลงจะต้องหากฎความสัมพันธ์หลายระดับชั้นแบบล่างขึ้นบน ที่สามารถเปลี่ยนแปลงลำดับชั้นของวัตถุได้บ่อยครั้งเท่าที่ผู้ใช้ต้องการ โดยไม่จำเป็นต้องอ่านข้อมูลจริงแม้แต่บ่อยและอ่านเพียงข้อมูลกลุ่มวัตถุที่เกิดขึ้นบ่อยในลำดับชั้นล่างสุดเพียงครั้งเดียวเท่านั้น ตลอดจนรองรับการหาความสัมพันธ์ระดับชั้นแบบหลายมิติและการหากฎความสัมพันธ์ข้ามชั้นด้วย จากผลการทดลองแสดงให้เห็นว่าวิธีที่นำเสนอมีประสิทธิภาพและรองรับข้อมูลปริมาณมาก โดยขั้นตอนการพัฒนาของระบบสืบค้นความรู้เป็นปัจจัยสำคัญอีกประการหนึ่งที่ส่งผลโดยตรงต่อประสิทธิภาพการทำงาน โดยรวมของระบบ ระบบสืบค้นความรู้เกือบทั้งหมดในปัจจุบันเป็นระบบที่พัฒนาขึ้นแยกจากระบบจัดการฐานข้อมูลหรือเป็นระบบที่พัฒนาแบบหลวม ซึ่งประสิทธิภาพของระบบจะถูกจำกัดโดยความสามารถในการจัดการข้อมูลของโปรแกรมประยุกต์แทน ซึ่งทำให้ระบบไม่มีประสิทธิภาพสูงเท่าที่ควร งานวิจัยนี้จึงพัฒนาระบบสืบค้นความรู้บนข้อมูลเชิงวัตถุที่สามารถหากฎความสัมพันธ์สำหรับข้อมูลที่มีโครงสร้างซับซ้อนได้ โดยระบบนี้ได้รับพัฒนาระบบสืบค้นความรู้แบบแน่นบนระบบจัดการฐานข้อมูลที่โครงสร้างซับซ้อนได้ ทำให้ระบบสามารถเข้าถึงข้อมูลที่อยู่ภายในฐานข้อมูลได้โดยตรงและทำให้การสืบค้นความรู้ทำได้รวดเร็ว นอกจากนี้ระบบสามารถสืบค้นความรู้ที่มีหลายมิติและหลายลำดับชั้นได้โดยไม่ต้องการความรู้พื้นฐาน (Background Knowledge) เพิ่มเติม ผลจากการหากฎความสัมพันธ์ที่ได้จากระบบสืบค้นความรู้ที่พัฒนาขึ้นจะได้เป็นกฎความสัมพันธ์ที่ได้จากระบบสืบค้นความรู้ที่พัฒนาขึ้นจะได้เป็นกฎความสัมพันธ์ที่ประกอบด้วยความสัมพันธ์หลายมิติและความสัมพันธ์ที่ประกอบด้วยความสัมพันธ์หลายมิติและความสัมพันธ์ของวัตถุในหลายระดับชั้น

สุรชัย วิวัฒน์เจริญชัย (2546) ได้ทำวิจัยเรื่อง การขุดค้นข้อมูลของ การทำธุรกรรมอิเล็กทรอนิกส์ของธนาคารแห่งประเทศไทย วิเคราะห์พฤติกรรมกรรมการใช้งานโดยอัลกอริทึมเคมีน และโครงข่ายประสาทเทียม ดังนี้ ในยุคที่ข้อมูลข่าวสารมีความสำคัญ เทคโนโลยีสารสนเทศได้มีการพัฒนาขึ้นอย่างรวดเร็วและกลายเป็นสิ่งที่จำเป็นต่อการทำธุรกิจทุกประเภท ในปัจจุบันธนาคารบางแห่งในประเทศไทยเริ่มเปิดบริการลูกค้าในการธุรกรรมผ่านทางอินเทอร์เน็ต การบริการนี้มีผู้นิยมใช้พอสมควร อย่างไรก็ตามการศึกษาพฤติกรรมของลูกค้าที่มาใช้บริการมีปริมาณน้อย จึงควรมี

การศึกษาเกี่ยวกับพฤติกรรมของลูกค้าในการทำธุรกรรมของธนาคารผ่านทางอินเทอร์เน็ต เพื่อเป็นข้อมูลขององค์กรและจะนำมาใช้วางแผนกลยุทธ์เพื่อทำให้เกิดความได้เปรียบทางการแข่งขัน งานวิจัยนี้มีวัตถุประสงค์ในการศึกษาพฤติกรรมของลูกค้า ที่ทำธุรกรรมทางอินเทอร์เน็ต จึงนำหลักการขุดค้นข้อมูล (Data Mining) สำหรับการแบ่งกลุ่มลูกค้าเพื่อแยกประเภทลูกค้าที่มีพฤติกรรมในการทำธุรกรรมอิเล็กทรอนิกส์ที่แตกต่างกัน โดยการเปรียบเทียบการแบ่งกลุ่มโดยใช้อัลกอริทึม 2 ประเภท คือ K-meaning Algorithm และ Neural Network Algorithm และใช้ข้อมูลจากแฟ้ม System Log ของลูกค้าธนาคารพาณิชย์ที่เข้ามาทำธุรกรรมทางอินเทอร์เน็ตในช่วงเวลา 1 ปี ผลที่ได้คือสามารถแบ่งกลุ่มลูกค้าและวิเคราะห์พฤติกรรมการใช้อินเทอร์เน็ตของข้อมูลแต่ละกลุ่มได้ชัดเจน และสามารถสรุปว่าการแบ่งกลุ่มข้อมูลโดยใช้ K-means Algorithm และ Neural Networks Algorithm มีความเหมาะสมสำหรับข้อมูลที่มีต่อปริมาณที่แตกต่างกัน ท้ายที่สุดมีข้อเสนอแนะปริมาณข้อมูลที่เหมาะสมสำหรับเลือกใช้อัลกอริทึมดังกล่าว

ธีระวัฒน์ พงษ์ศิริปริดา (2544) ในการพิจารณาคดีความทั่วไปของนักกฎหมาย หรือ ทนายความ จำเป็นต้องอ้างอิงตัวบทกฎหมายของประเทศ ซึ่งกฎหมายเหล่านี้มีปริมาณมาก ยากต่อการศึกษาย่างครบถ้วนเพื่อเลือกกฎหมายฉบับที่เหมาะสมสำหรับประกอบการพิจารณาคดีความ งานวิจัยนี้จึงเกิดขึ้นมาเพื่อแก้ไขปัญหาดังกล่าว โดยการสร้างระบบที่สามารถทำนายกฎหมายฉบับที่เหมาะสมสำหรับแต่ละ คดีความให้อย่างอัตโนมัติ โดยอาศัยเทคนิคของ Data Mining ในงานวิจัยฉบับนี้เราเลือกใช้เทคนิค Data Classification และ Association Rule Discovery ซึ่งเป็นเทคนิคที่สำคัญของ Data Mining มาใช้กับงานด้านกฎหมาย โดยการสร้างระบบจัดสรรกฎหมายที่เหมาะสมใช้กับคดีความโดยอัตโนมัติ เพื่อช่วยการทำงานของนักกฎหมายให้มีประสิทธิภาพมากขึ้น และช่วยให้ประชาชนทั่วไปที่ไม่มีความรู้ทางด้านกฎหมาย ได้สามารถหากกฎหมายที่เหมาะสมกับคดีความที่กำลังสนใจอยู่

เราได้เริ่มต้นทำการทดลอง โดยการนำโปรแกรมสำเร็จรูปที่มีความสามารถทางด้าน Data Classification มาทดสอบกับคดีความจำนวน 30,000 คดี ผลการทดลองที่ได้ไม่เป็นที่น่าพอใจ เนื่องจากเปอร์เซ็นต์ความถูกต้องจากการทำนายต่ำมาก ต่อมาเราได้พัฒนาระบบขึ้นมาใหม่ โดยในขั้นตอนการเตรียมข้อมูล เราได้ปรับปรุงเทคนิคการตัดคำด้วยพจนานุกรม โดยใช้พจนานุกรมวลีที่สร้างขึ้นด้วยเทคนิค Suffix Array เปอร์เซ็นต์ความถูกต้องที่ได้จากระบบใหม่นี้ให้ผลดีกว่าระบบเดิมที่ใช้การตัดคำด้วยพจนานุกรมคำเดียว ส่วนในขั้นตอนการสร้างตัวทำนายผลกฎหมาย เราได้นำเทคนิค Association Rule Discovery มาใช้ร่วมกับเทคนิค Data Classification ซึ่งเปอร์เซ็นต์ความถูกต้องเฉลี่ยจากการทำนายที่ได้จากระบบใหม่ ก็สูงกว่าระบบเดิมที่สร้างตัวทำนายด้วยเทคนิค Data Classification แบบที่นิยมกันทั่วไปคือ Decision Tree

นอกจากเปอร์เซ็นต์ความถูกต้องเฉลี่ยที่เพิ่มมากขึ้นในระบบใหม่แล้ว ด้วยเทคนิคใหม่ที่เสนอช่วยแก้ปัญหาต่างๆ ที่ระบบเดิมไม่สามารถทำได้ คือ การทำนายผลครั้งละมากกว่า 1 ผล และการทำนายผลแบบมีลำดับชั้น เนื่องจากว่าในแต่ละคดีความจะอ้างอิงกฎหมายมากกว่า 1 ฉบับ และการอ้างอิงกฎหมายจะแบ่งเป็น อ้างอิงชื่อกฎหมาย และฉบับที่ของกฎหมายนั้นๆ ทำให้ระบบใหม่ที่เสนอนี้สามารถรองรับความต้องการจริงได้เพิ่มมากขึ้น

ชิตชนก ส่งศิริ (2544) ระบบการศึกษาในปัจจุบันได้พัฒนาอย่างต่อเนื่อง มีการคิดค้นแนวคิดและทฤษฎี รวมถึงการนำเทคโนโลยีใหม่ ๆ มาใช้เพื่อยกระดับมาตรฐานการศึกษา แต่พบว่ยังมีปัญหาบางประการที่ไม่ได้รับการแก้ไขเท่าที่ควร อาทิเช่น ปัญหานิสิตเลือกเรียนในสาขาวิชาที่ไม่ตรงกับความสามารถที่แท้จริง ซึ่งเป็นผลให้การผลิตบุคลากรในสาขานั้นๆ มีประสิทธิภาพไม่สูงเท่าที่ควร ประกอบกับได้มีเทคนิค Data Mining เกิดขึ้นมา ซึ่งในปัจจุบันได้มีการนำไปประยุกต์ใช้กับด้านต่างๆ มากขึ้นแต่ยังไม่มีการนำมาประยุกต์ใช้ในด้านการศึกษาอย่างจริงจัง ซึ่งตามสถาบันการศึกษาส่วนใหญ่มีข้อมูลที่ถูกรวบรวมและจัดเก็บไว้อย่างดีเป็นจำนวนมาก แต่ไม่ได้นำข้อมูลที่มีอยู่มาใช้ให้เกิดประโยชน์อย่างคุ้มค่า จึงสมควรอย่างยิ่งที่จะนำ ข้อมูลที่เก็บไว้มาใช้ให้ก่อเกิดประโยชน์สูงสุด งานวิจัยนี้ได้ นำ เสนอการใช้เทคนิค Classification ในการชี้แนะแนวทางการเลือกภาควิชาที่เหมาะสมที่สุดกับลักษณะและความสามารถของนิสิตแต่ละคน ผลลัพธ์ของงานวิจัยค่อนข้างน่าพอใจ โดยโมเดลการจัดภาควิชาที่ได้มีความถูกต้องมากกว่า 80 เปอร์เซ็นต์

ชนวัฒน์ ศรีสอาน (2551) ได้นำเสนอการวิเคราะห์ฟอแมลคอนเท็กซ์ (Formal Context) ขององค์ความรู้ด้านสังคมศาสตร์ เพื่อนำผลการวิเคราะห์ข้อมูล ไปพัฒนาจัดระบบการจัดหมวดหมู่บทความ ระบบพิจารณาความถี่ของคำ ระบบการสร้างความสัมพันธ์ด้วยกำหนดเมทาเดตา (Metadata) เพื่อสร้างเป็นตัวแทน (Agent) เนื่องจากองค์ความรู้มีความสัมพันธ์กัน จึงกำหนดรูปแบบตารางประกอบไปด้วย 3 ส่วน คือ ออบเจกต์ (Object) แอตทริบิว (Attribute) ความสัมพันธ์ (Relation) ทฤษฎี ฟอแมลคอนเท็กซ์ (Formal Context) ประกอบด้วย 2 ส่วนได้แก่ ขอบเขต (extent) หรือ เซต (Set) ของวัตถุในตารางที่เกิดร่วมกัน คือวัตถุในตาราง และ เจตนา (Intent) คือ ลักษณะแอตทริบิว (attribute) ร่วมกันของวัตถุ การนิยามตัวแปรต่างๆดังนี้

$O =$ ออบเจกต์ (Object) $A =$ แอตทริบิว (Attribute) $R =$ ความสัมพันธ์ (Relation)
กำหนดให้ $X \subseteq O$ และ $Y \subseteq A$

สมการ $P(X) = \{a \in A \mid \forall o(X: (o, a)(R))\}$ หมายถึง แอตทริบิวร่วมของ X (2-1)

สมการ $Q(X) = \{a \in A \mid \forall o(Y: (o, a)(R))\}$ หมายถึง แอตทริบิวร่วมของ Y (2-2)

ดังนั้น ในการออกแบบระบบจัดเก็บและสืบค้นจะได้คอนเซพแลททิซ (Concept Lattice) กำหนดการสืบค้นผ่าน เมทาเดตา (Metadata) และการค้นหาจากเนื้อความ (Text Search) เพื่อให้สามารถสืบค้นภูมิปัญญาท้องถิ่นที่มีเนื้อหาสัมพันธ์กับคำค้น ซึ่งอยู่ต่างหมวดหมู่

จากการศึกษางานวิจัยภายในประเทศ สามารถสรุปเทคนิคในการทำเหมืองข้อมูล (Data mining) ได้ดังนี้

2.8.1.1 เหมืองข้อมูล (Data mining) เพื่อการทำนาย เป็นการนำความรู้ที่เรียนรู้มาจากข้อมูลที่มีอยู่เพื่อประโยชน์ในการทำนาย ข้อมูลใหม่ที่จะเกิดขึ้นในอนาคต เช่น จากชุดข้อมูล เทคนิคเหมืองข้อมูล (Data mining) สามารถเรียนรู้จากข้อมูลเหล่านี้และค้นหาโมเดลที่สามารถใช้อธิบายลักษณะของ เป็นลำดับชั้นจากโมเดลที่ได้นี้สามารถนำไปใช้ทำนายข้อมูลใหม่ได้เป็นประเภท และลำดับ

2.8.1.2 เหมืองข้อมูล (Data Mining) เพื่อ การอธิบาย เป็นการค้นหารูปแบบที่น่าสนใจจากกลุ่มข้อมูล รูปแบบนี้มักจะเป็นความสัมพันธ์หรือลักษณะที่เชื่อมโยงกันของข้อมูล การทำแบบนี้

ต่างจากแบบแรกตรงที่ผู้ใช้ไม่ได้กำหนดล่วงหน้าว่าจะให้โปรแกรม Data Mining ค้นหาแบบหรือโมเดลของอะไร แต่ให้ค้นหาทุกรูปแบบที่น่าสนใจจากข้อมูล

2.8.1.3 เหมืองข้อมูล (Data Mining) เพื่อการค้นหาความสัมพันธ์ (Association Rule) ซึ่งเป็นเทคนิคหนึ่งในเทคนิคเหมืองข้อมูล เช่น วัน เวลา หมวดเว็บ และเว็บ มาค้นหาความสัมพันธ์เพื่อสร้างตัวแบบ ได้ศึกษาจากงานวิจัยของ พิจิตรา จอมศรี ธนาวินท์ รักธรรมานนท์ และธีระวัฒน์ พงษ์ศิริปริดา

2.8.1.4 เหมืองข้อมูล (Data Mining) เพื่อ การแบ่งกลุ่มลูกค้า คัดแยกประเภทลูกค้าที่มีพฤติกรรมในการทำธุรกรรมอิเล็กทรอนิกส์ที่แตกต่างกัน โดย K-meaning Algorithm ได้ศึกษาจากงานวิจัยของ สุรชัย วิวัฒน์เจริญชัย

2.8.1.5 การนำเทคนิคเหมืองข้อมูล (Data Mining) มาประยุกต์ใช้ในการพัฒนาระบบจัดเก็บและสืบค้นสารสนเทศ ได้ศึกษาจากงานวิจัยของ นฤป รังงาม สถาพร สังข์ศิริ วาทีณี ดวงอ่อนนาม ธนากร สุวรรณวัฒน์ และ บุญมา เฟ่งชวน โดยนฤป รังงาม ได้นำเทคนิค OLAP Cube มาประยุกต์ใช้ และสถาพร สังข์ศิริ ได้นำเทคนิคออนโทโลยี (Ontology) มาประยุกต์ใช้ในการพัฒนาระบบจัดเก็บและสืบค้นสารสนเทศ

2.8.1.6 การทดสอบประสิทธิภาพระบบ โดยการหาค่าความแม่นยำ (Precision) และการหาค่าความระลึก (Recall) ซึ่งเป็นวิธีการทดสอบประสิทธิภาพของระบบจัดเก็บและสืบค้น ได้ศึกษาจากงานวิจัยของ สถาพร สังข์ศิริ วาทีณี ดวงอ่อนนาม และธนากร สุวรรณวัฒน์

2.8.1.7 การวิเคราะห์ความสัมพันธ์ (Association Rule) ได้ศึกษากระบวนการขั้นตอนการวิเคราะห์ข้อมูลจากผลงานของชนวัฒน์ ศรีสอ้าน เพื่อเปรียบเทียบความน่าเชื่อถือกับการวิเคราะห์ฟอโมลคอนเท็กซ์ (Formal Context) ของบทความทั้งหมด

2.8.2 งานวิจัยต่างประเทศ

David M W Powers (2011) ได้ศึกษามาตรวัดการประเมินผลการเรียกคืนสารสนเทศ นั้น คือ F-Factors ซึ่งเป็นมาตรวัดการประเมินผล ความแม่นยำในการค้นคืนสารสนเทศ F- Factors ความแม่นยำและค่าความคาดเคลื่อนนั้น ไม่ควรใช้โดยไม่มีข้อมูลที่ชัดเจนของ และมีอคติ โอกาสที่จะได้ผลลัพธ์การประเมินค่าทางสถิติ สอดคล้องกับกรณีศึกษาทำได้ยาก มาตรวัดเหล่านี้ ส่งผลให้ประสิทธิภาพระบบแย่ง วัตถุประสงค์ของงานวิจัยขึ้นนี้ จึงดำเนินการ ที่ดีขึ้น ภายใต้ปรากฏการณ์การวัดมาตรการใช้ ศึกษาเกี่ยวกับ แนวคิดและ มาตรวัดหลายโมเดล ที่สะท้อนให้เห็นถึง ความน่าจะเป็นในการทำนาย ที่ได้รับ เปรียบเทียบกับ ข้อมูลความผิดปกติ เช่น วัดแบบจับคู่ สำหรับความน่าจะเป็น ที่คาดการณ์ว่าจะถูกตรวจจับ ในข้อผิดพลาดของการวัด และแสดงให้เห็นถึง การเชื่อมต่อระหว่างแนวความคิดแจ้งความผิดปกติ ที่มีความสัมพันธ์ และสำคัญ เช่นเดียวกับที่ พวกเขา ใช้งานง่าย ความสัมพันธ์กับ การเรียกคืน และ ความแม่นยำ

Mohan, E.K. & Han (2011) ได้ทำการศึกษาวิธีที่ทำให้แคชสามารถใช้ข้อมูลร่วมกันได้ ซึ่งพร้อมๆกันด้วยคงทำงานอย่างมีประสิทธิภาพที่วัดจากระดับความน่าเชื่อถือของ hit ratio งานวิจัยนี้ กำหนดให้แคชเป็นเหมือนแคชส่วนตัวโดยขึ้นอยู่กับการรูปแบบการเข้าใช้ของผู้ใช้ ซึ่งผู้วิจัยได้ ออกแบบสถาปัตยกรรมของแคช โดยมี Web cache server ทำหน้าที่เป็นตัวจัดการระบบ Web Proxy Cache และทำการจัดเก็บข้อมูลที่ได้จาก Proxy Cache แต่ละตัว โดย Proxy Cache จะแชร์

ข้อมูลและกำหนดรูปแบบการใช้งานของ Client โดยการสุ่ม งานวิจัยนี้ผู้วิจัยได้จัดกลุ่มตัวแทน (Agent) ของผู้ใช้เป็น 4 กลุ่มโดยวิธีอัลกอริทึม Classification และใช้อัลกอริทึม Apriori ของ Association Rule เพื่อบอกลำดับ ความถี่และเวลาในการใช้เว็บจาก Cache Log ซึ่งการวิจัยครั้งนี้ถือเป็นการศึกษารูปแบบและจัดเตรียม แคมป์ที่มีประสิทธิภาพให้กับผู้ใช้ในการค้นหาข้อมูลโดยซอฟต์แวร์นี้สามารถทำนายเวลาที่ผู้ใช้โปรแกรม กับการใช้ทรัพยากรคอมพิวเตอร์อย่างเหมาะสม

Yang, Wang, & Zhang (2002) ได้เสนอรูปแบบสำหรับการทำนายเหตุการณ์ที่เกิดขึ้นบนเว็บ ซึ่งเป็นการทำนายช่วงเวลาไหนที่จะมีการเรียกใช้เว็บที่เคยถูกเรียกใช้แล้วอีกครั้ง เช่น เมื่อสำรวจพบว่าผู้ใช้เรียกใช้เว็บ A และ B น้อยกว่า เว็บ C แล้ว ระบบอาจทำนายว่า เว็บ C ควรที่จะเป็นเพจต่อไปที่จะถูกเรียกใช้ภายใน 10-20 วินาที จากเวลาปัจจุบัน ซึ่งเป็นวิธีการจำแนกประเภทโดยกฎความสัมพันธ์ (Association Rule of Classification Method) ผู้วิจัยเก็บข้อมูลจริงจากเว็บเซิร์ฟเวอร์ ของ NASA เป็นเวลา 1 เดือน โดยนำ อัลกอริทึม Classification มาใช้ในขั้นตอนของการเตรียมข้อมูลและใช้ประยุกต์สร้างรูปแบบการทำนายเวลาที่เกิดขึ้นที่เรียกว่า อัลกอริทึม Moving-Window ที่พัฒนามาจากจากกฎความสัมพันธ์ (LHS \rightarrow RHS) ซึ่งการวิจัยนี้ได้ศึกษาความแตกต่างของวิธี Temporal Regain Prediction ซึ่งเป็นเทคนิคที่พัฒนามาจากกฎการค้นหาความสัมพันธ์ โดยทำเปรียบเทียบความแตกต่างของวิธี Temporal Regain Prediction 3 วิธี คือ วิธีแรก Naïve วิธีที่สอง Confidence Interval Bases วิธีที่สาม Minimal Temporal Region โดยผลการทดลองปรากฏว่า วิธี Confidence Interval Bases และ วิธี Minimal Temporal Region ได้ผลการทดลองที่ใกล้เคียงกัน ซึ่งวิธี Minimal Temporal Region จะมีความถูกต้องแม่นยำกว่าเพียงเล็กน้อย

Agrawal, Rakesh, & Ramakrishnan Srikant (1994) มีผู้วิจัยเสนอแนวคิดใหม่ในการหาความสัมพันธ์ของฐานข้อมูลซึ่งเป็นเทคนิคหนึ่งที่สำคัญของเหมืองข้อมูล (Data Mining) งานวิจัยได้อธิบายถึง วิธีการ และจุดเด่นของอัลกอริทึมใหม่ 2 อัลกอริทึม คือ อัลกอริทึม Apriori Tid และอัลกอริทึม Apriori Hybrid ที่ใช้ในการหาความสัมพันธ์ในฐานข้อมูลการขายสินค้าขนาดใหญ่ อัลกอริทึมใหม่ทั้งสองอัลกอริทึมนี้อ้างอิงมาจากอัลกอริทึม Apriori ซึ่งผู้วิจัยได้แสดงการพัฒนาอัลกอริทึม Apriori Hybrid จากการรวมอัลกอริทึม Apriori และ Apriori Tid มีการเปรียบเทียบประสิทธิภาพของอัลกอริทึมใหม่กับอัลกอริทึมแบบเดิมที่ใช้ในการหา กฎความสัมพันธ์ เช่น อัลกอริทึม AIS อัลกอริทึม SETM และอัลกอริทึม Apriori การทดสอบ ประสิทธิภาพของอัลกอริทึมต่าง ๆ นี้ได้ทดสอบกับฐานข้อมูลที่มีขนาดและรูปแบบต่าง ๆ กัน จากการทดสอบประสิทธิภาพของอัลกอริทึมใหม่พบว่ามีประสิทธิภาพดีกว่าอัลกอริทึมที่เคยมีการนำเสนอมาแล้วมากกว่า 3 เท่าสำหรับฐานข้อมูลที่มีข้อมูลเป็นจำนวนมาก

จากการศึกษางานวิจัยต่างประเทศ สามารถสรุปเทคนิคในการทำเหมืองข้อมูล (Data Mining) ได้ดังนี้

2.8.2.1 งานวิจัยต่างประเทศเป็นการเปรียบเทียบอัลกอริทึมของการหาความสัมพันธ์ และการแบ่งกลุ่มเพื่อทดสอบความถูกต้องแม่นยำของอัลกอริทึม ตลอดจนการเปรียบเทียบอัลกอริทึมของ กฎความสัมพันธ์ และหาอัลกอริทึม นำอัลกอริทึมใหม่ใช้สำหรับการทำเหมืองข้อมูล (Data Mining)

2.8.2.2 จากงานวิจัยของ Yang ที่นำข้อมูลล็อกไฟล์ (Log File) มาหาความสัมพันธ์เพื่อพยากรณ์การใช้เว็บไซต์ และได้ประยุกต์วิธีการดังกล่าว โดยนำข้อมูลสถิติการสืบค้น มาวิเคราะห์สร้างเป็นหมวดหมู่ ประกอบกับศึกษาจากเทคนิคการทำเหมืองข้อมูล (Data Mining) ที่กล่าวไว้ในผลงานของ Han คือ การวิเคราะห์หมวดหมู่ (K-means Cluster Analysis) และการวิเคราะห์ความสัมพันธ์ (Association Rule) สำหรับการวิเคราะห์หมวดหมู่ ได้จำแนกกรณีการค้นข้อมูล กำหนดจำนวนกลุ่ม ตัวแปรที่ใช้จะต้องเป็นตัวแปรเชิงปริมาณ เป็น สเกลอันตรภาค (Interval Scale) สเกลอัตราส่วน (Ratio Scale) ในส่วนของการวิเคราะห์ความสัมพันธ์ (Association Rule) ได้สร้างกฎความสัมพันธ์ระหว่างค่าค้นในแต่ละบทความ สำหรับผู้ใช้สนใจค้นหาหลายบทความแบบต่อเนื่องเกิดขึ้นร่วมกันหรือพร้อมกันในฐานะข้อมูล หลังจากนั้นนำบทความทั้งหมดวิเคราะห์ฟอแมลคอนเท็กซ์ (Formal Context) เพื่อทำการเปรียบเทียบความน่าเชื่อถือกับหมวดหมู่ที่ได้จากการวิเคราะห์หมวดหมู่ (K-means Cluster Analysis) และการวิเคราะห์ความสัมพันธ์ (Association Rule) เพื่อทำการสรุปผลและนำผลการวิเคราะห์ไปพัฒนาเป็นระบบจัดเก็บและสืบค้น ตลอดจนทดสอบประสิทธิภาพระบบ โดยการหาค่าความแม่นยำ (Precision) การหาค่าความระลึก (Recall) ซึ่งเป็นวิธีการทดสอบประสิทธิภาพของระบบจัดเก็บและสืบค้น นอกจากนี้ได้ศึกษาการประเมินค่า F-Factors จากงานวิจัยของ David M W Powers

2.9 ผลการศึกษางานวิจัย

ผลการศึกษางานวิจัยจะเป็นเทคนิคและหลักการที่จะนำไปใช้ในงานวิจัยต่อไปประกอบด้วยเทคนิคและหลักการดังนี้

2.9.1 กฎจำแนกประเภทข้อมูล (Data Classification)

เป็นการแบ่งหมวดหมู่โดยทำการกำหนดสิ่งที่เป็นลักษณะเด่นในแต่ละหมวดหมู่ซึ่งจะแบ่งข้อมูลตามความคล้ายคลึงกันจากตัวอย่างข้อมูลที่มีอยู่ (Supervised Learning) - Classification เป็นการจัดแบ่งประเภทของข้อมูล โดยหาชุดต้นแบบหรือชุดของการทำงานที่อธิบายและแบ่งประเภทข้อมูล วัตถุประสงค์เพื่อให้สามารถใช้เป็นต้นแบบทำนายประเภทของวัตถุหรือข้อมูลที่ไม่มีการระบุประเภทหรือชนิดของข้อมูล ซึ่งต้นแบบสร้างจากการวิเคราะห์ชุดของข้อมูลฝึกสอน (Training Data) โดยอาจจะเป็นกลุ่มข้อมูลที่มีการระบุประเภทหรือกลุ่มเรียบร้อยแล้ว รูปแบบของต้นแบบแสดงได้หลายแบบเช่น Classification Rules, Decision Trees หรือ Neural Networks เป็นต้น

2.9.2 กฎการแบ่งกลุ่มข้อมูล (Cluster Analysis)

การแบ่งกลุ่มเป็นเทคนิคที่ใช้จำแนกหรือแบ่งเป็นกรณี (คน สัตว์ สิ่งของ หรือ องค์กร) หรือ แบ่งตัวแปรเป็นกลุ่มย่อยตั้งแต่ 2 กลุ่มขึ้นไป โดยกรณีที่อยู่กลุ่มเดียวกันจะมีลักษณะที่เหมือนกันหรือคล้ายกัน ส่วนกรณีที่อยู่ห่างกันจะมีลักษณะที่แตกต่างกัน ดังนั้น การพิจารณาเลือกลักษณะตัวแปรที่จะนำมาใช้แบ่งกลุ่มกรณีใดๆ จึงมีความสำคัญ นอกจากนั้นกรณี ไตๆกรณีหนึ่งจะอยู่ในกลุ่มหนึ่งเพียงกลุ่มเดียว เทคนิค Cluster Analysis แบ่งเป็นหลายประเภทโดยเทคนิคที่ใช้กันมากมี 2 เทคนิค

2.9.2.1 เทคนิค Hierarchical Cluster Analysis เป็นเทคนิคที่นิยมใช้กันมากในการแบ่งกลุ่ม หรือแบ่งกลุ่มตัวแปรที่มีจำนวนไม่มาก สามารถใช้ได้กับตัวแปรเชิงคุณภาพและตัวแปรเชิงปริมาณซึ่งไม่จำเป็นต้องทราบว่าตัวแปรนั้นอยู่กลุ่มใด และมีจำนวนกลุ่มมากเท่าใด

2.9.2.2 เทคนิค K-means Cluster Analysis เป็นเทคนิคการจำแนกกรณีออกเป็นกลุ่มย่อย โดยจะใช้กับ Case ที่มีจำนวนมากและต้องกำหนดจำนวนกลุ่ม ตัวแปรที่ใช้จะต้องเป็นตัวแปรเชิงปริมาณคือ เป็นสเกลอันดับ (Interval Scale) หรือ สเกลอัตราส่วน (Ratio Scale) เท่านั้น ผู้วิเคราะห์จะต้องทำข้อมูลให้เป็นมาตรฐานก่อน

2.9.3 กฎการวิเคราะห์ความสัมพันธ์ (Association rule) เป็นการหาความสัมพันธ์ระหว่างข้อมูลด้วยกันเองซึ่งมีพื้นฐานมาจากการเกิดขึ้น ร่วมกันหรือพร้อมกันในฐานข้อมูลรูปแบบของการค้นหาความสัมพันธ์

รูปแบบทั่วไปของการค้นหาความสัมพันธ์ คือ A->B โดยที่

A: เป็นเงื่อนไข หรือ LHS (Left - Hand Side)

B: เป็นผลลัพธ์ที่เกิดขึ้น หรือ RHS (Right - Hand Side)

หรืออยู่ในรูปของ “ถ้า.....แล้ว” (If.....Then....) เช่น

A -> B; if A Then B เป็นกฎที่ 1 (2-3)

B -> A; if B Then A เป็นกฎที่ 2 (2-4)

การประเมินค่าของกฎจะใช้ค่าสนับสนุน(Support) และค่าความเชื่อมั่น (Confidence) โดยที่ ค่าสนับสนุน คือ เปอร์เซนต์ของข้อมูลที่มีเงื่อนไขและผลลัพธ์สอดคล้องตามกฎต่อจำนวนข้อมูลทั้งหมด สามารถเขียนเป็นสมการดังนี้

$$\text{ค่าสนับสนุน(A,B)} = \frac{\text{จำนวนของ Transaction (A,B)}}{\text{จำนวน Transaction ทั้งหมด}} \quad (2-5)$$

โดยที่ A หมายถึง เหตุการณ์ที่ใช้เป็นเงื่อนไขในการหาผลลัพธ์

B หมายถึง เหตุการณ์ที่เป็นผลลัพธ์

Transaction (A,B) หมายถึง เหตุการณ์ที่ประกอบด้วยเหตุการณ์ A และ B

ค่าความเชื่อมั่น คือ เปอร์เซนต์ของข้อมูลที่มีเงื่อนไขและผลลัพธ์สอดคล้องตามกฎต่อจำนวนข้อมูลทั้งหมดที่เป็นเงื่อนไข สามารถเขียนเป็นสมการดังนี้

$$\text{ค่าความเชื่อมั่น (A,B)} = \frac{\text{จำนวนของ Transaction (A,B)}}{\text{จำนวน Transaction (A)}} \quad (2-6)$$

2.9.4 การประเมินประสิทธิภาพจากการคำนวณ Recall และ Precision

ศุภชัย ตั้งวงศ์ศานต์ (2551) ได้อธิบายถึงการประเมินประสิทธิภาพจากการคำนวณ Recall และ Precision ดังนี้

2.9.4.1 การหาค่าความระลึก (Recall) ถูกกำหนดให้เป็นอัตราส่วนของเอกสารที่เกี่ยวข้องที่ถูกดึงออกมาจากจำนวนเอกสารที่เกี่ยวข้องทั้งหมด โดยมีสูตรดังนี้

$$\frac{\text{จำนวนข้อมูลที่ถูกดึงที่ค้นคืนได้}}{\text{จำนวนข้อมูลทั้งหมดที่ค้นคืนออกมาได้}} \times 100 \quad (2-7)$$

2.9.4.2 การหาค่าความแม่นยำ (Precision) เป็นอัตราส่วนของเอกสารที่ถูกดึงออกมาแล้วข้อความ จากจำนวนเอกสารที่ถูกดึงออกมาทั้งหมด ในทางปฏิบัติ, สารสนเทศที่ต้องการของผู้ใช้แต่ละคนนั้นย่อมแตกต่างกันไป ซึ่งผู้ใช้บางคนอาจต้องการค่าความระลึกที่สูง กล่าวคือ ทุกสิ่งที่ถูกดึงออกมาเป็นเรื่องที่น่าสนใจ ในขณะที่อีกคนหนึ่งอาจต้องการถ้าหากมีเส้นตัดผ่าน Document Collection เพื่อที่จะแยกแยะรายการเอกสารที่ถูกดึงออกมา ให้ออกจากรายการเอกสารที่ไม่ได้ถูกดึงออกมา โดยมีสูตรดังนี้

$$\frac{\text{จำนวนข้อมูลที่ถูกดึงที่ค้นคืนได้}}{\text{จำนวนข้อมูลที่ถูกดึงทั้งหมดในฐานข้อมูล}} \times 100 \quad (2-8)$$

ในการประเมินประสิทธิภาพของระบบจัดเก็บและสืบค้นที่พัฒนาขึ้น ผู้วิจัยจะประเมินค่าความระลึก และค่าความแม่นยำ โดยค่าที่ได้จะได้นั้นจะใช้ในการอธิบายประสิทธิภาพในการสืบค้นบทความที่มีการจัดเก็บในระบบธนาคารสมองคลังปัญญาท้องถิ่น