

## บทที่ 2

### เอกสารที่เกี่ยวข้อง

การพัฒนาระบบแนะนำหนังสือที่เกี่ยวข้องกับการสืบค้นสารสนเทศห้องสมุดโดยใช้เทคนิคอะโพรอริอัลกอริทึม ประกอบด้วยสาระสำคัญและเอกสารที่เกี่ยวข้องดังต่อไปนี้

#### 2.1 เทคนิคการทำเหมืองข้อมูล

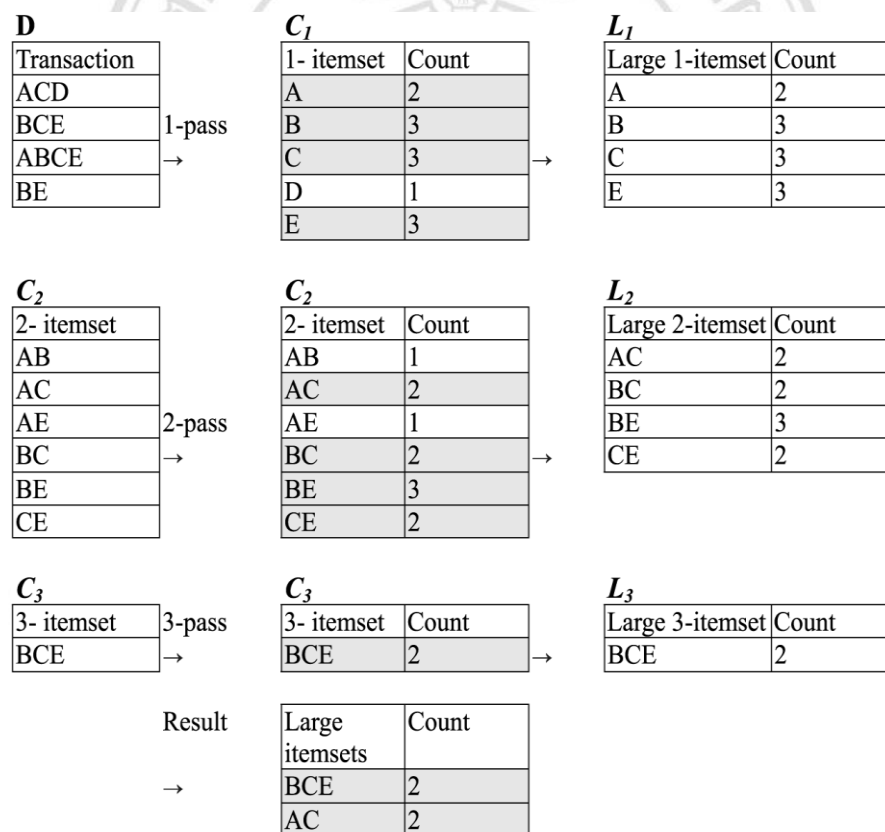
เทคนิคการทำเหมืองข้อมูล (Data Mining) ในปัจจุบันมีหลายรูปแบบโดยทั่วไปแล้วแบ่งออกเป็น 4 เทคนิคคือ 1) การทำเหมืองข้อมูลแบบการหากฎการรวมกลุ่ม 2) การทำเหมืองข้อมูลแบบจำแนกประเภท 3) การทำเหมืองข้อมูลแบบการจัดกลุ่ม 4) การนำเสนอข้อมูลในรูปแบบของรูปภาพสองมิติ และในการศึกษาครั้งนี้ ผู้ศึกษาได้เลือกการทำเหมืองข้อมูลแบบการจัดกลุ่มโดยใช้เทคนิคอะโพรอริอัลกอริทึมเป็นเครื่องมือในการจัดกลุ่มคำสำคัญของสารสนเทศห้องสมุด

2.1.1 กฎการรวมกลุ่มเป็นเทคนิคหนึ่งของการทำเหมืองข้อมูลที่สำคัญและสามารถนำไปประยุกต์ใช้ได้จริงกับงานต่างๆ หลักการทำงานของวิธีนี้คือการค้นหาความสัมพันธ์ของข้อมูลจากข้อมูลขนาดใหญ่ที่มีอยู่ไปใช้ในการวิเคราะห์ หรือทำนายปรากฏการณ์ต่างๆ หรือมาจากการวิเคราะห์พฤติกรรมการซื้อขายของลูกค้า (Market Basket Analysis) ซึ่งประเมินจากข้อมูลในตารางที่รวบรวมไว้ ผลการวิเคราะห์ที่ได้จะเป็นคำตอบของปัญหา ซึ่งการวิเคราะห์แบบนี้เป็นการใช้กฎการรวมกลุ่มเพื่อหาความสัมพันธ์ระหว่างข้อมูล (จิรยุทธ ไชยจารุวนิช, 2546) การค้นหากฎการรวมกลุ่มนิยมใช้อะโพรอริอัลกอริทึม โดยมีรายละเอียดในการทำงานของอัลกอริทึมดังต่อไปนี้

1) การทำงานของอะโพรอริอัลกอริทึม (Apriori Algorithm) อะโพรอริอัลกอริทึมถือเป็นอัลกอริทึมที่นิยมใช้ในการหากฎการรวมกลุ่มของข้อมูล หากฐานข้อมูลมีการเพิ่มข้อมูลเข้ามา หรือมีการเปลี่ยนแปลงอะโพรอริอัลกอริทึม จะต้องนำข้อมูลทั้งหมดมารวมกันก่อนจึงจะสามารถนำข้อมูลทั้งหมดไปค้นหากฎการรวมกลุ่มใหม่ทั้งหมดโดยมีลักษณะการทำงานคือสร้างแคนดิเดทไอเทมเซต (Candidate Itemsets) ขึ้นมาก่อน ซึ่งจะเริ่มจากแคนดิเดทไอเทมเซตที่มีขนาดเท่ากับ k-Itemsets โดยที่  $k = \{1, 2, 3, \dots, n\}$  ซึ่งแคนดิเดทไอเทมเซตในระดับที่ k จะได้มาจากการ join กันระหว่างไอ

ไอเทมเซตขนาดใหญ่ (Large Itemsets) ในระดับที่  $k-1$  เช่น ไอเทมเซต {ABC} เป็นผลมาจากการนำไอเทมเซต {AB} join กับ ไอเทมเซต {AC} จากนั้นจึงทำการหาไอเทมเซตขนาดใหญ่ในระดับที่  $k$  โดยจะทำการอ่านข้อมูลจากฐานข้อมูลเพื่อทำการนับค่าสนับสนุน (Support Factor) ของแคนดิเดทไอเทมเซตแต่ละตัวในระดับที่  $k$  นั้น จากนั้นจึงทำการตรวจสอบว่าแคนดิเดทไอเทมเซตใดมีค่าสนับสนุนมากกว่าหรือเท่ากับค่าสนับสนุนน้อยสุด (min\_sup) จะถือว่าแคนดิเดทไอเทมเซตนั้นมีคุณสมบัติเป็นไอเทมเซตขนาดใหญ่ในระดับที่  $k$  แล้วทำการสร้างแคนดิเดทไอเทมเซตหา ไอเทมเซตขนาดใหญ่ในระดับที่  $k+1$  ต่อไป โดยจะทำเช่นนี้ไปเรื่อยๆ จนกว่าจะไม่สามารถที่จะสร้างแคนดิเดทไอเทมเซตในระดับที่  $k+1$  ได้ (Agrawal and Srikant, 1994)

Han *et al.* (2000) ได้เสนอตัวอย่างการสร้างกฎการรวมกลุ่มโดยใช้อะโพอริอัลกอริทึมเป็นเครื่องมือดังภาพที่ 2.1



ภาพที่ 2.1 ตัวอย่างการสร้างกฎการรวมกลุ่มโดยใช้อะโพอริอัลกอริทึมเป็นเครื่องมือ

จากภาพที่ 2.1 “D” คือตารางฐานข้อมูลไอเทมเซต “C<sub>1</sub>” คือแคนดิเดทไอเทมเซตรอบแรก “C<sub>2</sub>” คือแคนดิเดทไอเทมเซตรอบที่สอง “C<sub>3</sub>” คือแคนดิเดทไอเทมเซตรอบที่สาม “L<sub>1</sub>” คือไอเทมเซตขนาดใหญ่รอบแรก “L<sub>2</sub>” คือไอเทมเซตขนาดใหญ่รอบที่สอง “L<sub>3</sub>” คือไอเทมเซตขนาดใหญ่รอบที่สาม

ตั้งค่านับสนับสนุนน้อยที่สุดเท่ากับสอง การทำงานของอะไพอริอัลกอริทึมในรอบแรกอัลกอริทึมจะตั้งแคนดิเดทไอเทมเซตเท่ากับหนึ่งก่อนแล้วทำการนับความถี่ของแต่ละไอเทมเซต จากภาพในตาราง “C1” ไอเทมเซต “D” มีความถี่เท่ากับหนึ่งซึ่งมีค่าความถี่น้อยกว่าค่านับสนับสนุนน้อยที่สุดจึงไม่ใช่ไอเทมเซตขนาดใหญ่อัลกอริทึมจะไม่นำไอเทมเซต “D” ไปพิจารณาในรอบต่อไป ในรอบที่สองอัลกอริทึมจะทำการสร้างแคนดิเดทไอเทมเซตขึ้นมาใหม่โดยเพิ่มจำนวนสมาชิกในแคนดิเดทไอเทมเซตอีกจำนวนหนึ่งสมาชิก (“C2”) แล้วทำการนับความถี่ของไอเทมเซตอีกครั้ง ไอเทมเซตที่มีความถี่น้อยกว่าค่านับสนับสนุนน้อยที่สุดจะไม่นำไปพิจารณาในรอบการทำงานต่อไป เซตที่มีความถี่มากกว่าหรือเท่ากับสองจะถูกเก็บไว้ในตาราง “L2” และในรอบถัดไป อัลกอริทึมจะสร้างแคนดิเดทไอเทมเซตที่มีจำนวนสมาชิกเท่ากับสามโดยวิธีการ “join” แล้วทำการนับความถี่ เซตที่มีความถี่มากกว่าหรือเท่ากับสองจะถูกเก็บไว้ในตาราง “L3” ซึ่งอัลกอริทึมจะทำงานลักษณะนี้แบบวนลูปไปเรื่อยๆ จนไม่สามารถสร้างแคนดิเดทไอเทมเซตได้ จากตัวอย่างสร้างแคนดิเดทไอเทมเซตที่มีสมาชิกมากที่สุดเท่ากับสาม

2) ค่าความเชื่อมั่น (Confidence) คือค่าที่แสดงความเชื่อมั่นของกฎการรวมกลุ่มที่เมื่อมีรูปแบบที่เกิดขึ้นทางซ้ายมือแล้วจะเกิดรูปแบบทางขวามือขึ้นเป็นจำนวนที่เปอร์เซ็นต์ การคำนวณค่าความเชื่อมั่นสามารถคำนวณได้จากสูตร

$$\text{confidence(LHS} \Rightarrow \text{RHS)} = \frac{\text{support(LHS, RHS)}}{\text{support(LHS)}}$$

จำนวนความถี่ของรูปแบบที่เกิดขึ้นซ้ายมือและขวามือในข้อมูลทั้งหมดหารด้วยความถี่ของรูปแบบที่เกิดขึ้นซ้ายมือในจำนวนข้อมูลทั้งหมด จากตัวอย่างเซต {A,C} มีค่าความเชื่อมั่นเท่ากับ 100% คำนวณได้จาก

รูปแบบ A และรูปแบบ C มีความถี่ทั้งหมดเท่ากับ 2 ในข้อมูล 4 ข้อมูล

รูปแบบ A มีความถี่ทั้งหมดเท่ากับ 2 ในข้อมูล 4 ข้อมูล

$$(2/4)/(2/4) = 1 \text{ หรือ } 100\%$$

หมายความว่า ร้อยละหนึ่งร้อยเมื่อเกิดรูปแบบ A แล้วจะเกิดรูปแบบ C ด้วย

3) ข้อดีและข้อเสียของอะไพอริอัลกอริทึม มีดังต่อไปนี้

— ข้อดีของอะโพอริอัลกอริทึมคือสามารถทำงานกับฐานข้อมูลขนาดเล็กได้ดี โดยกำหนดค่าสนับสนุนน้อยสุดมีค่ามากๆ และมีจำนวนชิ้นข้อมูลน้อย มีกระบวนการทำงานในการค้นหากฎการรวมกลุ่มที่มีลักษณะง่ายและไม่ซับซ้อน

— ข้อเสียของอะโพอริอัลกอริทึมคือกระบวนการทำงานของอะโพอริอัลกอริทึมนั้นต้องอ่านข้อมูลจำนวนหลายครั้ง เพื่อใช้ตรวจสอบแคนดิเดทไอเทมเซตซึ่งอาจทำให้เกิดปัญหาคอขวด และยังใช้เนื้อที่ของหน่วยความจำค่อนข้างมากในการสร้างแคนดิเดทไอเทมเซตขึ้นมาใหม่ในแต่ละรอบการทำงาน (ฟูโคละห์ คีอมอง, 2553)

## 2.2 ระบบแนะนำ

ระบบแนะนำ (Recommender Systems) คือระบบที่แนะนำข้อมูล ผลิตภัณฑ์หรือผู้คนที่ให้กับผู้ใช้ระบบ อ้างอิงจากสมมุติฐานการเรียนรู้ข้อมูล ความชอบหรือความต้องการ ณ เวลานั้นของผู้ใช้ โดยระบบแนะนำกลายเป็นหัวข้องานวิจัยที่สำคัญ ความสำเร็จในช่วงแรกๆ ของการทำระบบแนะนำเกิดจากการเจริญเติบโตของธุรกิจซื้อขายสินค้าและบริการบนเว็บไซต์ ตัวอย่างของแอปพลิเคชันที่ใช้ระบบแนะนำในการเลือกซื้อหนังสือ ซีดีเพลง หรือผลิตภัณฑ์อื่นๆ ได้แก่ เว็บไซต์ต่อเมซอนดอทคอม และเว็บไซต์ซีดีนาวดอทคอม (CDNow.com) เป็นต้น ทั้งนี้เพื่อให้ลูกค้าเกิดความภักดีต่อองค์กรและเป็นการเพิ่มโอกาสทางธุรกิจ (จาริณู ถิ่นพังงาและไพรสน ศิริพูล, 2552)

โดยทั่วไปวิธีการพัฒนาระบบแนะนำข้อมูลนั้นใช้เทคนิคทางด้านการทำเหมืองข้อมูล ซึ่งเป็นการกรองข้อมูลแบบพึ่งพาผู้ร่วมเป็นหลัก เป็นเทคนิคหนึ่งที่ประสบความสำเร็จมากและนิยมนำมาใช้ในระบบพาณิชย์อิเล็กทรอนิกส์ เพื่อแนะนำสินค้าและบริการให้กับลูกค้า วิธีการนี้จะใช้เทคนิคทางสถิติในการค้นหาเซตของลูกค้าที่เรียกว่า กลุ่มสมาชิกข้างเคียง (Neighbor) ที่มีประวัติความชอบในตัวสินค้าที่เหมือนกันกับลูกค้าเป้าหมาย เช่น มีลักษณะการซื้อสินค้าประเภทเดียวกัน ข้อมูลเหล่านี้จะถูกนำมาใช้ในระบบเพื่อนำไปวิเคราะห์และแนะนำสินค้าให้ตรงกับความต้องการของลูกค้าเป้าหมายต่อไป (วงกต ศรีอุไร ชูชาติ หฤไชยะศักดิ์และจิรารัตน์ สิทธิวรชาติ, 2551)

จากการศึกษานี้ ผู้ศึกษาได้นำเทคนิคการทำเหมืองข้อมูลมาประยุกต์ใช้ในงานห้องสมุดเพื่อพัฒนาระบบแนะนำหนังสือในห้องสมุด โดยวิธีการจัดกลุ่มค่าสำคัญของหนังสือ เพื่อแนะนำหนังสือที่มีค่าสำคัญเหมือนกับหนังสือที่ผู้ใช้บริการต้องการค้นหา โดยมีงานวิจัยที่เกี่ยวข้องดังต่อไปนี้

### 2.3 งานวิจัยที่เกี่ยวข้อง

O'Mahony and Smyth (2007) พัฒนาการใช้ระบบแนะนำหลักสูตร (Course Recommender System) เพื่อนำมาใช้ในการลงทะเบียนออนไลน์ของมหาวิทยาลัยคอลเลจดับลิน (College-Dublin's) โดยการร่างปัจจัยที่มีอิทธิพลต่อทางเลือกของนักศึกษาและเสนอแนวทางเพื่อใช้ในการพิจารณาคำตอบ วิธีการประเมินกระทำโดยใช้ข้อมูลประวัติการลงทะเบียนของนักศึกษา ซึ่งการออกแบบเบื้องต้นนี้แสดงให้เห็นถึงประสิทธิภาพของการดำเนินการที่มีแนวโน้มสัมฤทธิ์ผล

Trestian *et al.* (2009) ศึกษาการจำแนกลักษณะความสัมพันธ์ระหว่างความสนใจ แอปพลิเคชันของผู้คนและพื้นที่ให้บริการ เพื่ออำนวยความสะดวกในการค้นหาผู้คน ธุรกิจและสิ่งต่างๆ ซึ่งในรายงานนี้ได้นำกฎของเหมืองข้อมูล (Rule Mining) มาช่วยในการจัดกลุ่มสเปกตรัมเพื่อศึกษาความสัมพันธ์ระหว่างผู้ใช้ที่มีจำนวนมากกว่า 280,000 คน กับการใช้เครือข่ายโทรศัพท์มือถือในระบบ 3G ในเขตเมืองที่มีขนาดใหญ่

Hong *et al.* (2010) ศึกษาวิธีการแนะนำผู้ใช้เพื่อสนองความต้องการของตัวผู้ใช้เองโดยนำตัวแทนตอบโต้อิเล็กทรอนิกส์ (Electronic Conversational Agents :ECAs) มาใช้เพื่อโต้ตอบกับผู้ใช้ โดยใช้หลักการของอะไพออร์อัลกอริทึม จากการศึกษาครั้งนี้ได้ออกแบบระบบตัวแทนที่โต้ตอบได้ (Interactive Agent System) เพื่อสนับสนุนการแนะนำฐานความรู้และเพื่อแนะนำผลิตภัณฑ์โดยใช้ความรู้จากประวัติผู้ใช้ ซึ่งการเข้าถึงฐานความรู้มุ่งเน้นระบบที่อธิบายสาเหตุของผลการแนะนำ ด้วยเหตุนี้จึงศึกษาการนำตัวแทนตอบโต้อิเล็กทรอนิกส์มาใช้เพื่อส่งเสริมและอธิบายผลการแนะนำ เป็นผลให้ระบบที่เสนอขึ้นนี้สามารถบรรลุวัตถุประสงค์ของการโฆษณาสินค้าและได้รับการตอบกลับ (Feedback) จากผู้ใช้ เพื่อสาธิตการใช้งานผู้วิจัยได้ใช้ระบบตัวแทนดังกล่าวนี้แนะนำและโต้ตอบกับผู้ใช้บนเว็บไซต์นวนิยายอิเล็กทรอนิกส์ (e-Novel) ชื่อเรื่องแองเจิลซิตี (Angel City) เพื่อแนะนำนวนิยายอิเล็กทรอนิกส์ให้แก่ผู้อ่าน

จริยาพร จำปาและสมชาย ปราการเจริญ (2553) ได้ศึกษางานวิจัยนี้โดยการนำแนวคิดเรื่องการทำเหมืองข้อมูลมาวิเคราะห์ข้อมูลสารสนเทศของห้องสมุด ประกอบด้วยข้อมูลส่วนบุคคลของผู้ใช้บริการรหัส หมวดหมู่ของทรัพยากรที่ถูกยืม วันที่ยืมและวันที่คืนของผู้ใช้บริการแต่ละรายเพื่อเป็นการเรียนรู้พฤติกรรมการใช้ทรัพยากรห้องสมุดและนำข้อมูลเหล่านั้นมาประมวลผลเพื่อค้นหาความสัมพันธ์ของผู้ใช้บริการกับทรัพยากรที่ถูกยืม แล้วนำผลที่ได้ไปบริหารจัดการและจัดสรรงบประมาณที่มีอยู่อย่างจำกัดในการจัดซื้อทรัพยากรให้เหมาะสมและตรงความต้องการของ

ผู้ใช้บริการ โดยการใช้อะไพออร์อัลกอริทึม ข้อมูลที่นำมาใช้เป็นข้อมูลการยืมคืนทรัพยากรห้องสมุด ระหว่างปี พ.ศ. 2545-2552 จำนวน 4,500 รายการ

อดุลยรัตน์ วรรณแสง (2553) พัฒนาการประยุกต์ใช้ทฤษฎีเหมืองข้อมูลในบริการยืม-คืนของห้องสมุด กรณีศึกษาสำนักหอสมุดมหาวิทยาลัยธรรมศาสตร์ เนื่องจากสำนักหอสมุดมหาวิทยาลัยธรรมศาสตร์มีสารนิเทศจำนวนมาก ทำให้ผู้ยืมหนังสือไม่สามารถใช้ทรัพยากรที่มีได้อย่างมีประสิทธิภาพ ดังนั้นจึงจำเป็นต้องค้นหากฎการรวมกลุ่มโดยใช้เทคนิคการจัดกลุ่ม (Cluster) ซึ่งเป็นการเพิ่มศักยภาพให้ห้องสมุดสามารถให้บริการทรัพยากรที่มีอยู่ให้แก่ผู้ใช้บริการได้อย่างมีประสิทธิภาพมากขึ้น

การศึกษาครั้งนี้มุ่งเน้นการทำเหมืองข้อมูลโดยใช้เทคนิคอะไพออร์อัลกอริทึมในการสร้างกฎการรวมกลุ่มของคำสำคัญ เพื่อแนะนำหนังสือที่มีคำสำคัญเหมือนกับคำสำคัญของหนังสือที่ผู้ใช้บริการค้นหา ซึ่งแตกต่างจากงานที่ได้กล่าวมาข้างต้นที่เน้นการศึกษาพฤติกรรมของผู้ใช้บริการหรือลูกค้าเพื่อจัดกลุ่มสินค้าและบริการแล้วจึงแนะนำสินค้าและบริการที่มีความสัมพันธ์กับสินค้าและบริการที่ผู้ใช้บริการหรือลูกค้าค้นหา

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่  
Copyright © by Chiang Mai University  
All rights reserved