Bibliomining: การทำเหมืองข้อมูลในห้องสมุด

ไพโรจน์ เอี่ยมชัยมงคล

บทความนี้กล่าวถึง ความหมายของคำว่าการทำเหมืองข้อมูลในห้องสมุด หรือที่เรียกว่า Bibliomining ความหมายของคำว่าเหมืองข้อมูล ประโยชน์ของการทำเหมืองข้อมูล การประยุกต์การทำเหมืองข้อมูลกับงานด้านต่างๆ ของห้องสมุด ขั้นตอนการทำเหมืองข้อมูลในห้องสมุด ซึ่งรวมถึงวิธีการสร้างขลังข้อมูลและการทำความสะอาดข้อมูล และแนะนำซอฟท์แวร์บางตัวสำหรับการทำเหมืองข้อมูล

Bibliomining: Data Mining in Libraries

Pairoj Eamchaimongkol

This article describes about the meaning of data mining in library or "Bibliomining", meaning of data mining, advantages of data mining, applications of data mining in library, steps of taking data mining in library that included method of creating data warehouse and data cleansing. And advising of some data mining softwares.

Bibliomining: การทำเหมืองข้อมูลในห้องสมุด

ไพโรจน์ เอี่ยมชัยมงคล *

บทนำ

ปัจจุบันห้องสมุดในสถาบันการศึกษาส่วนใหญ่ของไทยนั้นได้นำระบบคอมพิวเตอร์เข้ามาช่วยงานด้านต่างๆ ในห้องสมุด เช่นงานด้านการยืมและคืนหนังสือ งานด้านการค้นหาหนังสือซึ่งมาแทนที่ตู้บัตรรายการ การค้นหาข้อมูลของวารสารวิชาการต่างๆ ผ่านระบบเครือข่ายหรือผ่านระบบอินเตอร์เน็ต จะเห็นได้ว่าห้องสมุดในปัจจุบันเป็นแหล่งรวบรวมข้อมูลด้านต่างๆ ไว้เป็นจำนวนมาก ซึ่งห้องสมุดส่วนใหญ่มักให้ความสำคัญกับแหล่งข้อมูลด้านวิชาการว่ามีเพียงพอกับความต้องการของนักศึกษาและบุคลากรในสถาบันการศึกษาของตนเองหรือไม่ แต่ไม่ได้ให้ความสำคัญกับข้อมูลต่างๆ ที่เกี่ยวข้องกับการให้บริการที่ห้องสมุดเป็นผู้เก็บไว้เอง เช่นข้อมูลจำนวนครั้งของการยืมของหนังสือแต่ละเล่ม หรือ หนังสือแต่ละเล่มนั้นมีใครเป็นผู้ยืมไปบ้าง ซึ่งจริงๆ แล้วเราสามารถนำข้อมูลเหล่านี้มาเป็นส่วนในการช่วยตัดสินใจการทำงานด้านต่างๆ ในห้องสมุดได้เป็นอย่างดี เช่นใช้ทำนายความต้องการหนังสือล่วงหน้า หรือแม้กระทั่งใช้เป็นข้อมูลช่วยในการตัดสินใจในการจัดสรรงบประมาณในการซื้อหนังสือให้สอดคล้องกับความต้องการของผู้ใช้ได้

แต่การนำข้อมูลที่เกี่ยวข้องกับการให้บริการมาใช้ได้นั้นจำเป็นจะต้องมีการเก็บข้อมูลการให้บริการเหล่านั้นไว้ในฐานข้อมูลของระบบคอมพิวเตอร์ในห้องสมุดเสียก่อน จากนั้นจึงนำข้อมูลการให้บริการเหล่านั้นผ่านการประมวลผลบางอย่างก่อน แล้วจึงค่อยนำผลที่ได้จากการประมวลผลนั้นไปใช้งานต่อ ซึ่งการประมวลผลนี้เรียกว่า Bibliomining

Bibliomining คืออะไร

Bibliomining หมายถึงการทำเหมืองข้อมูล (Data mining) กับข้อมูลต่างๆ ที่เก็บอยู่ในฐานข้อมูลห้องสมุด (Nicholson,S. & Stanton, J. 2003) ซึ่งผลที่ได้จากการทำ Bibilomining สามารถนำไปเป็นข้อมูลช่วยในการตัดสินใจหรือปรับปรุงบริการของห้องสมุด

การทำเหมืองข้อมูลคืออะไร

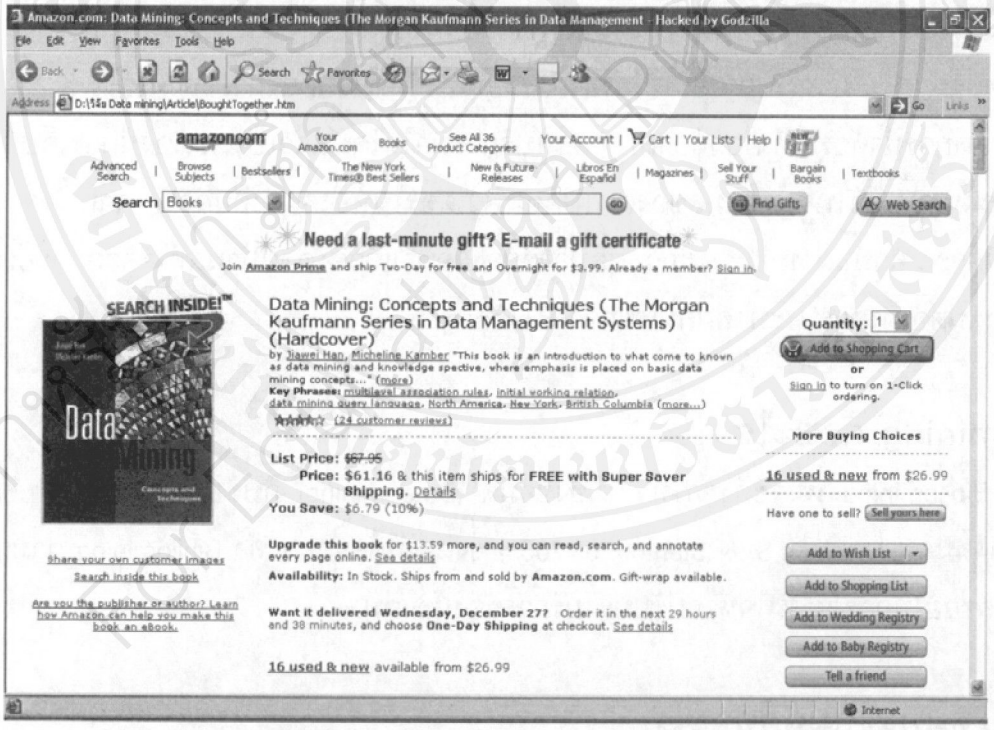
การทำเหมืองข้อมูล หรือ Data mining นั้นมีผู้ให้คำจำกัดความอยู่หลายแบบ จึงจะยกคำจำกัดความของการทำเหมืองข้อมูลให้เป็นตัวอย่างสองแบบดังนี้

* อาจารย์ประจำภาควิชาบรรณารักษศาสตร์ คณะศิลปศาสตร์ มหาวิทยาลัยธรรมศาสตร์

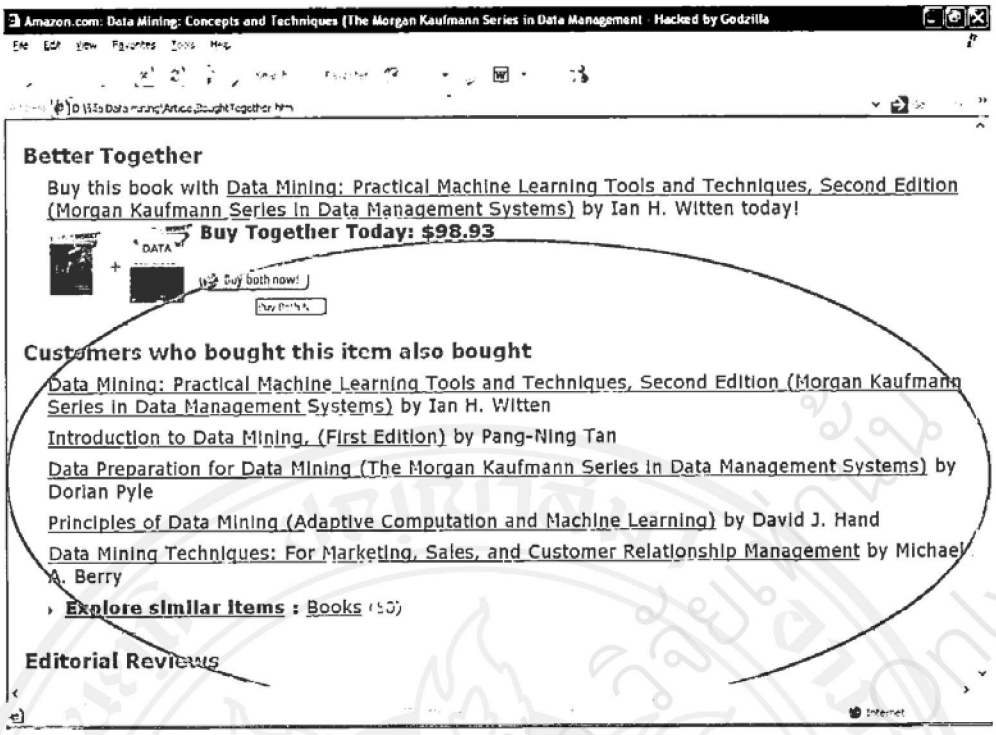
คำจำกัดความแรกนั้นให้คำนิยามของการทำเหมืองข้อมูลว่า เกี่ยวกับการวิเคราะห์ข้อมูลและค้นหารูปแบบที่ซ่อนอยู่โดยใช้วิธีอัตโนมัติ หรือ กึ่งอัตโนมัติ (Tang, Z.H. และ MacLennan, J. 2005)

สำหรับคำจำกัดความที่สองนั้นให้คำนิยามของการทำเหมืองข้อมูลว่า เป็นกระบวนการสกัดหรือ "ทำเหมือง" เพื่อเอาความรู้ออกจากข้อมูลจำนวนมาก (Han, J. และ Kamber, M. 2001)

ซึ่งปัจจุบันมีการใช้เทคนิคการทำเหมืองข้อมูลกันอย่างแพร่หลายโดยเฉพาะอย่างยิ่งในแวดวงธุรกิจ ตัวอย่างเช่นในเว็บไซต์การขายหนังสือของ Amazon นั้นจะเก็บข้อมูลการสั่งซื้อหนังสือของลูกค้าไว้ แล้วนำข้อมูลดังกล่าวมาผ่านกระบวนการทำเหมืองข้อมูล ซึ่งหลังจากทำเหมืองข้อมูลแล้วจะพบว่าเมื่อลูกค้าสั่งซื้อหนังสือ A แล้วมักจะสั่งซื้อหนังสือ B ด้วย (Lavoie,B., Demsey L. และ Connaway, L.S. 2006) เมื่อได้ข้อมูลดังกล่าวแล้วก็นำไปปรับปรุงเว็บไซต์การสั่งซื้อหนังสือ โดยเมื่อใดก็ตามที่มีลูกค้าเปิดดูเว็บไซต์เพื่อที่จะสั่งซื้อหนังสือ A แล้ว ก็จะมีส่วนหนึ่งของหน้าเว็บไซต์ทำการแนะนำหนังสือ B ให้ลูกค้าทราบด้วย (รูปที่ 1 และ 2) ทำให้ลูกค้ามีทางเลือกในการสั่งซื้อหนังสือมากขึ้น อีกทั้งยังสามารถเพิ่มยอดขายได้ด้วย



รูปที่ 1 หน้าเว็บไซต์ Amazon.com โดยแสดงรายละเอียดหนังสือที่เกี่ยวกับ Data mining (จาก www.amazon.com เมื่อวันที่ 25/12/2549)



รูปที่ 2 ส่วนต่อมาของหน้าเว็บไซต์ในรูปที่ 1 ซึ่งข้อความที่อยู่ในวงนั้นเป็นข้อมูลที่แสดงว่าเมื่อลูกค้าซื้อหนังสือเล่มดังกล่าวแล้วมักซื้อหนังสือที่แสดงอยู่ในวงด้วย
(จาก www.amazon.com เมื่อวันที่ 25/12/2549)

หรือตัวอย่างการใช้ประโยชน์จากการทำเหมืองข้อมูลอีกตัวอย่างหนึ่งคือ การวิเคราะห์รูปแบบการซื้อสินค้าในซุปเปอร์มาร์เก็ตแห่งหนึ่ง โดยนำข้อมูลการซื้อสินค้าของลูกค้าผ่านการทำเหมืองข้อมูลแล้วพบว่า จะมีลูกค้าอยู่จำนวนหนึ่งซึ่งมักซื้อผ้าอ้อมพร้อมกับเบียร์เสมอ (Larose, Daniel T. 2005) เมื่อพิจารณาข้อมูลลูกค้าที่ซื้อผ้าอ้อมพร้อมเบียร์พบว่าเป็นชายหนุ่มที่เพิ่งจะมีลูก ดังนั้นเมื่อมาซื้อผ้าอ้อมให้ลูกก็มักจะซื้อเบียร์ไปทานด้วย จากผลการทำเหมืองข้อมูลดังกล่าวทำให้ร้านซุปเปอร์มาร์เก็ตแห่งนี้ปรับเปลี่ยนรูปแบบการวางสินค้า โดยจัดวางผ้าอ้อมและเบียร์ให้อยู่ใกล้กันเพื่อความสะดวกของลูกค้าและเป็นการเพิ่มยอดขาย จะเห็นได้ว่าการทำเหมืองข้อมูลนั้นเป็นประโยชน์ในวงการธุรกิจอย่างมาก และไม่เพียงแต่ในวงการธุรกิจเท่านั้นที่ใช้เทคนิคการทำเหมืองข้อมูล แม้แต่วงการกีฬาเองก็ใช้เทคนิคการทำเหมืองข้อมูลเพื่อปรับปรุงประสิทธิภาพของทีมกีฬาของตนเอง ซึ่งในการแข่งขันบาสเก็ตบอล NBA ของสหรัฐอเมริกานั้นได้นำเทคนิคการทำเหมืองข้อมูลมาใช้ร่วมกับการบันทึกภาพระหว่างการแข่งขัน เพื่อทำการวิเคราะห์การแข่งขันในแต่ละครั้ง ยกตัวอย่างเช่นการแข่งขันบาสเก็ตบอล NBA ระหว่างทีม New York Knicks กับทีม Cleveland Cavaliers ที่มีขึ้นในวันที่ 6 มกราคม ปีค.ศ. 1995 นั้นได้มีการใช้ซอฟท์แวร์สำหรับทำเหมืองข้อมูลที่ชื่อว่า Advanced Scout ของ IBM มาวิเคราะห์การเคลื่อนที่ของนักกีฬาระหว่างการแข่งขัน ซึ่งซอฟท์แวร์ Advanced Scout วิเคราะห์ว่าเมื่อผู้เล่นของทีม Cleveland Cavaliers ที่ชื่อว่า Mark Price ลงเล่นตำแหน่งการ์ด (Guard) แล้ว ผู้เล่นของทีม Cleveland Cavaliers อีกคนหนึ่งที่ชื่อว่า John Williams จะมีโอกาสซู้ตลูกบาสเพื่อทำคะแนนถึง 4 ครั้ง และจะเป็นแบบนี้ทุกครั้งที่นักกีฬาสองคนนี้

ลงไปเล่นอยู่ในสนามพร้อมกัน ซึ่งเมื่อโค้ชของทีมนำผลการวิเคราะห์รูปแบบการเล่นดังกล่าวมาพิจารณาร่วมกับภาพวิดีโอเฉพาะช่วงที่ Mark Price และ John Williams ลงเล่นพร้อมกันพบว่าเมื่อ Mark Price ลงเล่นจะสามารถดึงนักกีฬาแนวรับของทีม New York Knicks ซึ่งเป็นทีมฝ่ายตรงข้ามให้มาคอยประกบติดการเคลื่อนไหวของ Mark Price และจะทำให้ไม่มีนักกีฬาของทีมฝ่ายตรงข้ามคอยติดตาม John Williams อันเป็นเหตุให้ John Williams มีโอกาสชู้ตลูกบาสทำคะแนนมากขึ้น ซึ่งโค้ชของทีมสามารถนำรูปแบบที่วิเคราะห์ได้นี้ไปปรับปรุงประสิทธิภาพการเล่นของทีมให้ดียิ่งขึ้น และมีทีมบาสเก็ตบอล NBA ของสหรัฐอเมริกาอยู่ 16 ทีม จาก 29 ทีมที่ใช้ซอฟท์แวร์ Advanced Scout ในการทำเหมืองข้อมูลการเล่นบาสเก็ตบอล (Larose, Daniel T. 2005)

สามารถนำเหมืองข้อมูลมาใช้กับงานด้านใดได้บ้าง

สามารถนำเหมืองข้อมูลมาประยุกต์ใช้กับงานได้หลากหลายด้าน ซึ่งแต่ละด้านก็มีความแตกต่างกันและสามารถนำไปใช้งานให้เกิดประโยชน์ที่แตกต่างกันไป และเราสามารถแบ่งงานที่สามารถนำเหมืองข้อมูลมาใช้ได้ 6 ด้าน (Larose, Daniel T. 2005) ดังนี้

1. การอธิบายหรือพรรณนา (Description)

เป็นการนำผลที่ได้จากการทำเหมืองข้อมูลไปใช้เพื่ออธิบาย หรือพรรณนาถึงรูปแบบหรือแนวโน้มที่ซ่อนอยู่ในข้อมูล ซึ่งส่วนใหญ่จะเป็นการอธิบายหรือพรรณนาด้วยค่าต่างๆ ทางสถิติเช่น ผลรวม ค่าเฉลี่ย ค่าเบี่ยงเบนมาตรฐาน

2. การประมาณ (Estimation)

เป็นการนำผลที่ได้จากการทำเหมืองข้อมูลซึ่งในกรณีนี้มักจะอยู่ในรูปของแบบจำลองชนิดต่างๆ ไปใช้ประมาณค่าต่างๆ ตัวอย่างของการใช้งานด้านนี้ได้แก่ การประมาณเกรดเฉลี่ยของนักศึกษาระดับบัณฑิตศึกษา โดยอาศัยค่าเกรดเฉลี่ยของนักศึกษาที่ได้รับเมื่อศึกษาอยู่ในระดับปริญญาตรีมาประมาณ ซึ่งมีวิธีการคือนำเอาเกรดเฉลี่ยในระดับบัณฑิตศึกษาและระดับปริญญาตรีของนักศึกษาจำนวน 1000 คน มาเขียนกราฟและนำมาหาสมการของเส้น Regression ระหว่างเกรดเฉลี่ยในระดับบัณฑิตศึกษากับเกรดเฉลี่ยในระดับปริญญาตรีได้ดังนี้ $y = 1.24 + 0.67x$ โดย y หมายถึงเกรดเฉลี่ยในระดับบัณฑิตศึกษา และ x หมายถึงเกรดเฉลี่ยในระดับปริญญาตรี ดังนั้นถ้านักศึกษามีเกรดเฉลี่ยในระดับปริญญาตรีเป็น 3.0 แล้ว สามารถประมาณเกรดเฉลี่ยในระดับบัณฑิตศึกษาได้เป็น 3.25 โดยการแทนค่าลงในสมการดังกล่าว

3. การทำนาย (Prediction)

เป็นการนำหลักการทำเหมืองข้อมูลมาทำนายผลที่จะเกิดขึ้น ความแตกต่างระหว่างการประมาณ และ การทำนาย คือ การทำนายนั้นผลที่ได้จะทำนายสิ่งที่เกิดขึ้นในอนาคต ตัวอย่าง

เช่นการทำนายราคาหุ้นในอีก 3 เดือนข้างหน้า หรือการทำนายอัตราการตายที่เกิดจากอุบัติเหตุทางการจราจรในปีหน้า ซึ่งวิธีใดก็ตามที่นำมาใช้ในการทำประมาณก็สามารถนำมาใช้ในการทำนายได้โดยต้องใช้อย่างเหมาะสม ซึ่งวิธีเหล่านี้ได้แก่การวิเคราะห์การถดถอยเชิงเส้น (Linear Regression) หรือ การหาสหสัมพันธ์ (Correlation)

4. การแบ่งประเภท (Classification)

เป็นการนำหลักการทำเหมืองข้อมูลมาแบ่งประเภทของข้อมูล โดยการแบ่งประเภทนั้นจะต้องมีตัวแปรที่ใช้เป็นเป้าหมายในการแบ่งประเภท เช่น รายได้ แล้วจึงแบ่งตัวแปรนั้นออกเป็นประเภทต่างๆ เช่น รายได้สูง รายได้ปานกลาง หรือ รายได้ต่ำ ตัวอย่างงานที่ใช้หลักการแบ่งประเภท เช่น การจัดให้นักเรียนเข้าใหม่ได้เรียนในสายวิชาที่ถนัด หรือ ธนาคารใช้หลักการแบ่งประเภทเพื่อประเมินว่าบ้านหรือที่ดินที่นำมาจำนองใดมีความเสี่ยงต่อการหลุดจำนอง

5. การจัดกลุ่ม (Clustering)

เป็นการนำหลักการทำเหมืองข้อมูลมาจัดกลุ่มข้อมูล โดยเป็นการจัดให้ข้อมูลที่มีความคล้ายคลึงกันอยู่ในกลุ่มเดียวกัน ความแตกต่างระหว่างการจัดกลุ่มกับการแบ่งประเภทอยู่ที่ว่าการจัดกลุ่มนั้นจะไม่มีตัวแปรที่ใช้เป็นเป้าหมายในการจัดกลุ่ม ตัวอย่างงานที่ใช้หลักการแบ่งกลุ่ม เช่น การแบ่งกลุ่มลูกค้าของบริษัท การแบ่งกลุ่มผู้ยืมหนังสือของห้องสมุด การแบ่งกลุ่มยีนของมนุษย์

6. การสัมพันธ์กัน (Association)

เป็นการนำหลักการทำเหมืองข้อมูลมาหาว่าข้อมูลใดที่ไปด้วยกัน หรือ สัมพันธ์กัน ซึ่งในแวดวงธุรกิจมีการใช้งานด้านนี้กันมากแต่รู้จักกันในชื่อว่า Affinity Analysis หรือ Market Basket Analysis ตัวอย่างงานที่ใช้หลักการสัมพันธ์ได้แก่กรณีที่วิเคราะห์ได้ว่ามีลูกค้ากลุ่มหนึ่งที่มักจะซื้อเบียร์กับผ้าอ้อมด้วยกัน หรือมีสินค้าใดบ้างที่ลูกค้าไม่เคยซื้อพร้อมกันเลย

การประยุกต์การทำเหมืองข้อมูลกับงานด้านห้องสมุด

การทำเหมืองข้อมูลในห้องสมุดนั้นจะช่วยให้สามารถวิเคราะห์รูปแบบของกิจกรรมต่างๆ ที่เกิดขึ้นกับห้องสมุดนั้น และรูปแบบที่วิเคราะห์ได้นั้นสามารถนำมาใช้ให้เป็นประโยชน์กับห้องสมุดได้ ซึ่งตัวอย่างของรูปแบบงานในห้องสมุดที่สามารถนำหลักการทำเหมืองข้อมูลเข้ามาประยุกต์ใช้ได้มีดังนี้

1. <u>การทำนายความต้องการของผู้ใช้บริการในอนาคต</u>

นำหลักการทำเหมืองข้อมูลด้านการอธิบายหรือพรรณนาทางสถิติมาใช้ เพื่อวิเคราะห์ว่ามีหนังสือเล่มใดหรือหมวดหมู่ใดที่มีสถิติการยืมสูง แล้วนำข้อมูลนี้มาตัดสินใจว่าควรจะสั่งซื้อหนังสือ

เล่มใดหรือหมวดหมู่ใดเพิ่ม หรือหนังสือเล่มใดควรมีหลายฉบับเพื่อให้เพียงพอกับความต้องการ ในทำนองเดียวกันหลักการนี้ก็สามารถใช้วิเคราะห์ว่าหนังสือประเภทใดหรือหมวดหมู่ใดที่มีการยืมน้อยก็จะจัดหาหนังสือเหล่านั้นเข้ามาให้บริการน้อย

2. การวางนโยบายป้องกันการคืนหนังสือล่าช้า

นำหลักการทำเหมืองข้อมูลมาวิเคราะห์รูปแบบการคืนหนังสือล่าช้าว่าผู้ใช้บริการกลุ่มใดที่มักจะคืนหนังสือล่าช้า หรือหนังสือเล่มใดที่เมื่อผู้ใช้บริการยืมไปแล้วมักจะคืนล่าช้า แล้วนำผลการวิเคราะห์ที่ได้มาวางนโยบายการยืมและคืนหนังสือใหม่เพื่อลดการคืนหนังสือล่าช้าให้น้อยลง

3. การวิเคราะห์ว่าหนังสือประเภทใดที่ไม่พอเพียงต่อความต้องการ

การวิเคราะห์ว่าหนังสือประเภทใดที่ไม่พอเพียงต่อความต้องการนั้น จะต้องใช้ข้อมูลจากสองแหล่งคือจากบรรณารักษ์ที่ให้บริการด้านการค้นคว้า หรือ จากฐานข้อมูลของระบบ OPAC สำหรับข้อมูลที่ได้จากบรรณารักษ์ที่ให้บริการด้านการค้นคว้านั้น บรรณารักษ์จะต้องบันทึกรายละเอียดเกี่ยวกับหนังสือที่ผู้ใช้ต้องการแต่หาไม่พบในห้องสมุดนั้นไว้ หรือนำผลการค้นหนังสือแต่ไม่พบและได้บันทึกไว้ในระบบ OPAC มาวิเคราะห์โดยใช้หลักทางสถิติเพื่อหาว่าหนังสือประเภทใดที่ผู้ใช้ต้องการมากที่สุดแต่ไม่มีในห้องสมุด

4. วิเคราะห์หนังสือที่มักถูกยืมไปด้วยกัน

การให้บริการคำปรึกษาส่วนใหญ่ของห้องสมุดในปัจจุบันนั้น มักจะให้บรรณารักษ์เป็นผู้แนะนำหนังสือแก่ผู้ใช้ว่ามีหนังสือเล่มใดบ้างที่มีเนื้อหาคล้ายกันเพื่อให้ผู้ใช้ไม่เสียเวลาในการค้นหาหนังสือและได้หนังสือมากเพียงพอต่อการค้นคว้า ซึ่งงานในลักษณะนี้สามารถนำหลักการทำเหมืองข้อมูลมาใช้อันได้แก่งานการสัมพันธ์กัน หรือ Market Basket Analysis ที่ได้กล่าวไว้ข้างต้น โดยการนำประวัติการยืมหนังสือมาวิเคราะห์หาว่าหนังสือเล่มใดบ้างที่มักถูกยืมด้วยกัน แล้วนำข้อมูลที่วิเคราะห์ได้นี้ใส่ไว้ในระบบ OPAC เพื่อให้เมื่อใดที่ผู้ใช้บริการสืบค้นหนังสือผ่านระบบ OPAC แล้ว ให้ระบบ OPAC แสดงรายชื่อหนังสือที่เกี่ยวข้องกันหรือมักถูกยืมด้วยกันให้ผู้ใช้ทราบด้วย ซึ่งจะเป็นการเพิ่มทางเลือกให้แก่ผู้ใช้มากขึ้น

5. การปรับปรุงเว็บไซต์ของห้องสมุดให้ตอบสนองความต้องการของผู้ใช้

การปรับปรุงเว็บไซต์นั้น จะต้องนำข้อมูลการเข้าใช้เว็บไซต์ของผู้ใช้บริการมาวิเคราะห์พฤติกรรรมและรูปแบบการใช้งาน แล้วนำผลการวิเคราะห์ที่ได้มาปรับปรุงเว็บไซต์เช่น พิจารณาว่าควรวางข้อความประกาศข่าวสารไว้ในที่ใดของหน้าเว็บไซต์จึงจะเหมาะสม ส่วนใดของเว็บไซต์ที่ต้องมีการเพิ่มคำแนะนำการใช้งาน หรือ จากการวิเคราะห์พฤติกรรมการใช้งานเว็บไซต์แล้วพบว่าส่วนใดของเว็บไซต์ที่ผู้ใช้เข้ามาใช้บริการมากที่สุด ก็จะต้องปรับปรุงเว็บไซต์ส่วนนั้นให้ผู้ใช้สามารถเข้ามาใช้บริการได้เร็วที่สุด หรือ เข้ามาใช้บริการในส่วนนั้นโดยกดปุ่มบนเมาส์ให้น้อยครั้งที่สุด ทั้งนี้การปรับปรุงเว็บไซต์ทำเพื่อให้ผู้ใช้บริการได้รับความสะดวกสบาย มีความรู้สึกอยากใช้งานเว็บไซต์ของห้องสมุด และให้ผู้ใช้บริการได้รับประโยชน์สูงสุดจากการใช้เว็บไซต์

6. การจัดกลุ่มผู้ใช้บริการ

นำหลักการทำเหมืองข้อมูลด้านการจัดกลุ่มมาจัดกลุ่มผู้ใช้บริการออกเป็นกลุ่มต่างๆ และวิเคราะห์ว่าผู้ใช้แต่ละกลุ่มนั้นใช้บริการทรัพยากรประเภทใดของห้องสมุดบ้าง แล้วนำข้อมูลที่ได้ไปจัดสรรประเภทและจำนวนของทรัพยากรให้สอดคล้องกับจำนวนของผู้ใช้แต่ละกลุ่ม เป็นการปรับปรุงการให้บริการห้องสมุดเพื่อตอบสนองความต้องการของผู้ให้ครอบคลุมทุกกลุ่ม

จากตัวอย่างรูปแบบงานข้างต้นจะเห็นว่าสามารถนำหลักการทำเหมืองข้อมูลมาประยุกต์กับงานบริการของห้องสมุดได้หลายๆ ด้าน ทั้งนี้ขึ้นกับห้องสมุดแต่ละแห่งว่าจะนำหลักการทำเหมืองข้อมูลไปประยุกต์เพื่อปรับปรุงงานด้านใดบ้างซึ่งอาจแตกต่างไปจากตัวอย่างข้างต้น แต่ทั้งนี้การนำหลักการทำเหมืองข้อมูลมาใช้ก็เพื่อปรับปรุงประสิทธิภาพด้านต่างๆ ของห้องสมุดเอง

ขั้นตอนการทำเหมืองข้อมูลในห้องสมุด

Nicholson (2003) ได้กล่าวว่ากระบวนการ Bibilomining หรือ การทำเหมืองข้อมูลในห้องสมุดนั้นประกอบด้วย 6 ขั้นตอนดังนี้

1. ระบุขอบเขต
2. ระบุแหล่งของข้อมูล
3. สร้างคลังข้อมูล
4. เลือกเครื่องมือที่จะนำมาวิเคราะห์
5. ทำเหมืองข้อมูลเพื่อค้นหารูปแบบ และสร้างรายงาน
6. วิเคราะห์ผลและนำผลที่ได้ไปประยุกต์ใช้

โดยแต่ละขั้นตอนมีรายละเอียดปลีกย่อยแตกต่างกันไป ซึ่งสามารถอธิบายรายละเอียดของแต่ละขั้นตอนได้ดังนี้

1. ระบุขอบเขต

ขั้นตอนแรกในการทำเหมืองข้อมูลในห้องสมุดคือการกำหนดขอบเขตของสิ่งที่จะนำมาทำเหมืองข้อมูล ซึ่งสิ่งที่จะนำมาทำเหมืองข้อมูลนั้นมาได้จากสองแหล่งคือ 1. จากปัญหาที่เกิดขึ้นจริงกับห้องสมุดซึ่งเรียกว่า Directed data mining และ 2.จากการที่ผู้บริหารของห้องสมุดต้องการรู้ข้อมูลต่างๆ เพิ่มเติมซึ่งเรียกว่า Undirected data mining

แบบ Directed data mining นั้นมักเป็นปัญหาที่เกิดกับการปฏิบัติงานของห้องสมุด และต้องการแก้ปัญหาที่เกิดขึ้นเหล่านั้นโดยใช้หลักการทำเหมืองข้อมูลเข้ามาช่วย ตัวอย่างของปัญหาที่เกิดขึ้นเช่น มักมีการคืนหนังสือล่าช้ากว่ากำหนดประมาณ 1 อาทิตย์เสมอ จึงนำหลักการทำเหมืองข้อมูลมาทำนายล่วงหน้าว่าผู้ใช้บริการกลุ่มใด หรือผู้ใช้บริการที่ยืมหนังสือประเภทใดที่มีโอกาสคืนหนังสือล่าช้ากว่ากำหนด แล้วนำผลที่ได้มาช่วยแก้ปัญหาการยืมคืนเพื่อที่จะให้เกิดกรณีการคืนหนังสือล่าช้ากว่ากำหนดน้อยลง

แบบ Undirected data mining นั้นเป็นข้อมูลที่ผู้บริหารของห้องสมุดอยากทราบเพื่อนำข้อมูลนั้นไปใช้ในการวางนโยบายหรือช่วยในการตัดสินใจด้านต่างๆ ซึ่งจะใช้หลักการทำเหมืองข้อมูลเพื่อค้นหาข้อมูลที่ผู้บริหารห้องสมุดต้องการดังกล่าว ตัวอย่างของข้อมูลที่ผู้บริหารของห้องสมุดอยากทราบเช่น มีผู้ใช้บริการประเภทใดบ้างที่ใช้วารสารอิเลกทรอนิกส์

2. ระบุแหล่งที่มาของข้อมูล

หลังจากทราบแล้วว่าจะนำข้อมูลใดมาทำเหมืองข้อมูลแล้ว ขั้นต่อไปได้แก่การระบุที่มาของแหล่งข้อมูลที่จะนำมาทำเหมืองข้อมูล ซึ่งข้อมูลที่จะนำมาทำเหมืองข้อมูลได้นั้นจะต้องเป็นเป็นข้อมูลที่เกี่ยวกับการทำธุรกรรม (Transaction) ต่าง ๆ กับห้องสมุดเช่นข้อมูลการยืมและคืนหนังสือ เป็นข้อมูลที่ครบถ้วนและต้องไม่เป็นข้อมูลที่ผ่านการสรุปมาแล้ว ดังนั้นข้อมูลที่จะนำมาทำเหมืองข้อมูลได้นั้นค่อนข้างหายากเนื่องจากไม่ได้มีการบันทึกไว้ ทำให้อาจต้องมีการนำข้อมูลจากแหล่งอื่นมาช่วยทำให้ข้อมูลที่จะนำมาทำเหมืองข้อมูลสมบูรณ์ยิ่งขึ้น

แหล่งข้อมูลที่จะนำมาทำเหมืองข้อมูลแบ่งเป็นสองชนิดคือ 1. แหล่งข้อมูลภายใน และ 2. แหล่งข้อมูลภายนอก สำหรับแหล่งข้อมูลภายในนั้นได้แก่ข้อมูลที่ได้เก็บไว้แล้วในระบบห้องสมุด สำหรับแหล่งข้อมูลภายนอกนั้นได้แก่ข้อมูลที่ไม่ได้เก็บไว้ในระบบห้องสมุด ตัวอย่างของข้อมูลที่ต้องนำมาจากแหล่งข้อมูลภายนอกได้แก่ ข้อมูลเกี่ยวกับอาชีพของบิดาและมารดาของนักศึกษาในสถาบันอุดมศึกษา ซึ่งข้อมูลเหล่านี้มักเก็บไว้ในระบบทะเบียนของมหาวิทยาลัยมากกว่าจะมาจัดเก็บไว้ในระบบของห้องสมุด

3. สร้างคลังข้อมูล

เมื่อได้ข้อมูลจากแหล่งต่างๆ ครบถ้วนแล้ว ขั้นตอนต่อไปคือการนำข้อมูลจากแหล่งต่างๆ มารวมกันเป็นคลังข้อมูลเดียวกัน ซึ่งคลังข้อมูลหรือ Data Warehouse หมายถึง ฐานข้อมูลที่แยกออกมาจากระบบการทำงานปัจจุบัน และบรรจุเฉพาะข้อมูลที่จะนำมาวิเคราะห์ ซึ่งข้อมูลที่บรรจุไว้ในคลังข้อมูลนั้นจะต้องเป็นข้อมูลที่ได้ทำความสะอาดแล้ว (Cleaned Data) และข้อมูลเหล่านั้นจะต้องไม่บ่งชี้ว่าใครเป็นเจ้าของข้อมูล (Nicholson,S. 2003)

ซึ่งการทำความสะอาดข้อมูล หรือ Data Cleansing ก็คือการขจัดความไม่สอดคล้องกัน (Inconsistency) ของข้อมูลให้หมดไป (กิตติพงษ์ กลมกล่อม.2548:198) ซึ่งในทางปฏิบัติแล้วข้อมูลที่ถูกบันทึกไว้ไม่เพียงแต่จะมีความผิดพลาดในเรื่องของความไม่สอดคล้องกันเท่านั้น ยังอาจเกิดความผิดพลาดในเรื่องอื่นๆ อีกเช่น ข้อมูลบางส่วนขาดหายไป หรือข้อมูลบางค่าที่ไม่แน่ใจว่าเป็นข้อมูลที่มีค่าเป็นปกติหรือไม่ และยังมีความผิดพลาดในแบบอื่นๆ ที่สามารถเกิดได้อีก ซึ่งตัวอย่างข้อมูลที่มีความผิดพลาดแบบต่างๆ ได้แสดงไว้ดังตารางที่ 1

จากข้อมูลในตารางที่ 1 พบว่า ข้อมูลเพศของพนักงานรหัส 1003 หายไป ส่วนข้อมูลอายุของพนักงานรหัส 1001 ที่เป็น 0 ปี นั้นน่าจะเป็นข้อมูลที่ไม่สอดคล้องกับความเป็นจริง เนื่องจากผู้ที่เป็นพนักงานในบริษัทต่างๆ ควรจะมีอายุมากกว่า 0 ปี สำหรับข้อมูลเบอร์โทรศัพท์แล้วพบว่าจะมีเบอร์โทรศัพท์มือถือของพนักงานรหัส 1002 ที่มีจำนวน 10 หลัก ซึ่งจริงๆ แล้วเบอร์โทรศัพท์มือ

ถือในปัจจุบันถูกปรับเป็น 10 หลักหมดแล้ว ดังนั้นควรจะปรับเบอร์โทรศัพท์มือถือของพนักงานที่เหลือเป็น 10 หลักด้วย และสำหรับตัวเลขรายได้ของพนักงานรหัส 1003 ที่มีค่าเป็น 9999 นั้นทำให้ไม่แน่ใจว่าเป็นเงินเดือนของพนักงานคนนั้นจริงๆ หรือไม่ หรือเป็นเพียงแต่ตัวเลขสมมติที่ใส่ไว้เพื่อเป็นสัญลักษณ์แจ้งความหมายบางอย่างไว้

ตารางที่ 1 ตัวอย่างข้อมูลที่มีความผิดพลาด

รหัสพนักงาน	เพศ	อายุ	เบอร์โทรศัพท์มือถือ	รายได้
1001	ชาย	0	012345678	9500
1002	หญิง	25	0814567890	10200
1003		36	093336666	9999
1004	ชาย	42	052227777	13000

ซึ่งข้อมูลที่ผิดปกติที่อยู่ในตารางที่ 1 นั้นเราจำเป็นต้องทำความสะอาดข้อมูลเพื่อให้ได้ข้อมูลที่ถูกต้องครบถ้วนเสียก่อนจึงจะสามารถนำข้อมูลไปวิเคราะห์ได้ สำหรับการแก้ไขข้อมูลให้ถูกต้องนั้นบางครั้งก็สามารถแก้ไขโดยใช้การพิจารณาเองได้ เช่นเบอร์โทรศัพท์มือถือที่ไม่ครบ 10 หลักนั้นสามารถเติมให้ครบ 10 หลักได้เอง แต่สำหรับข้อมูลบางค่าแล้วไม่สามารถพิจารณาเองได้ อาจต้องสอบถามจากพนักงานที่เป็นเจ้าของข้อมูลเอง หรืออาจต้องใช้หลักทางสถิติเพื่อหาค่าที่เหมาะสมมาใช้งานแทน ซึ่งขั้นตอนการทำความสะอาดข้อมูลโดยละเอียดรวมถึงหลักการทางสถิติที่นำมาใช้ในการทำความสะอาดข้อมูลนั้น สามารถอ่านรายละเอียดเพิ่มเติมได้จากหนังสือที่เขียนโดย Daniel T. Larose(2005) และสำหรับขั้นตอนการสร้างคลังข้อมูลโดยละเอียดนั้นสามารถศึกษาเพิ่มเติมได้จากหนังสือเขียนโดย กิตติพงษ์ กลมกล่อม(2548)

สิ่งสำคัญสิ่งหนึ่งในการทำเหมืองข้อมูลคือ การรักษาข้อมูลส่วนตัวของผู้ใช้บริการให้เป็นความลับ ดังนั้นข้อมูลที่บรรจุอยู่ในคลังข้อมูลนั้นต้องไม่มีส่วนที่บ่งชี้ว่าข้อมูลเหล่านั้นเป็นของผู้ใช้บริการคนใดบ้าง ยกตัวอย่างเช่นตารางบันทึกข้อมูลเดิมของการยืมหนังสือ (ตารางที่ 2) นั้นจะประกอบไปข้อมูลเลขเรียกหนังสือ ชื่อหนังสือ และรหัสของผู้ยืม และตารางข้อมูลเดิมของทะเบียนผู้ใช้บริการห้องสมุดเดิม (ตารางที่ 3) นั้นจะประกอบไปด้วย รหัสผู้ใช้บริการ ชื่อ-สกุลผู้ใช้บริการ ประเภทของผู้ใช้บริการ และ คณะที่ผู้ใช้บริการสังกัด

ดังนั้นเมื่อนำข้อมูลจากตารางที่ 2 และ 3 มารวมกันเป็นคลังข้อมูลใหม่จะต้องไม่มีข้อมูลที่บ่งชี้ว่าใครเป็นเจ้าของข้อมูลนี้โดยการลบในส่วนข้อมูลรหัสผู้ยืมและชื่อ-สกุล ผู้ใช้บริการออก ดังแสดงในตารางที่ 4

ตารางที่ 2 ข้อมูลการยืมหนังสือ

ตารางบันทึกข้อมูลการยืมหนังสือ		
เลขเรียกหนังสือ	**ชื่อหนังสือ**	**รหัสผู้ยืม**
BQ4570.ส6 ม63	หัวใจชาวพุทธ	49217858
TK5105.5.ร83	เรียนการใช้งานระบบเครือข่ายด้วยตนเอง	47125879
HF5145.13.อ37	การบริหารการตลาด	45231487

ตารางที่ 3 ทะเบียนผู้ใช้บริการห้องสมุด

ตารางบันทึกข้อมูลผู้ใช้บริการ			
รหัสผู้ใช้บริการ	**ชื่อ-สกุล ผู้ใช้บริการ**	**ประเภท**	**คณะที่สังกัด**
49217858	สมชาย รักงาน	ปริญญาตรี	ศิลปศาสตร์
47125879	สมศักดิ์ เรียนดี	ปริญญาโท	วิทยาศาสตร์
45231487	สมศรี เรือนงาม	ปริญญาตรี	บริหารธุรกิจ

ตารางที่ 4 คลังข้อมูล

คลังข้อมูล			
เลขเรียกหนังสือ	**ชื่อหนังสือ**	**ประเภท**	**คณะที่สังกัด**
BQ4570.ส6 ม63	หัวใจชาวพุทธ	ปริญญาตรี	ศิลปศาสตร์
TK5105.5.ร83	เรียนการใช้งานระบบเครือข่ายด้วยตนเอง	ปริญญาโท	วิทยาศาสตร์
HF5145.13.อ37	การบริหารการตลาด	ปริญญาตรี	บริหารธุรกิจ

จากหลักการการสร้างคลังข้อมูลที่กล่าวผ่านมาทั้งหมด จะสามารถพิจารณาได้ว่าขั้นตอนการสร้างคลังข้อมูลนั้นเป็นขั้นตอนที่ใช้เวลานานที่สุด และอาจเป็นขั้นตอนที่มีการลงทุนมากที่สุดกว่าขั้นตอนใดๆ ในกระบวนการทำเหมืองข้อมูลในห้องสมุด เนื่องจากจะต้องพิจารณาข้อมูลทีละรายการว่าถูกต้องหรือไม่ สอดคล้องหรือไม่และต้องทำข้อมูลให้สมบูรณ์ ซึ่งในการทำเหมืองข้อมูลนั้นต้องใช้ข้อมูลเป็นจำนวนมากในบางครั้งอาจเป็นหมื่นหรือแสนรายการทีเดียว ดังนั้นแนะนำว่าเมื่อเริ่มต้นกระบวนการทำเหมืองข้อมูลในห้องสมุดนั้น ควรเริ่มจากการระบุขอบเขตของหัวข้อในการทำเหมืองข้อมูลให้แคบ แล้วทำเหมืองข้อมูลในหัวข้อนั้นให้สำเร็จ จากนั้นค่อยเริ่มทำเหมืองข้อมูลในหัวข้อที่มีขอบเขตที่กว้างขึ้นต่อไป

4. เลือกเครื่องมือที่จะนำมาวิเคราะห์

เมื่อสร้างคลังข้อมูลได้แล้ว ขั้นต่อไปก็จะทำการวิเคราะห์ข้อมูลที่อยู่ในคลังข้อมูล ซึ่งการวิเคราะห์ข้อมูลนี้ทำได้หลายแบบตั้งแต่แบบง่ายจนไปถึงแบบที่ยาก สำหรับการวิเคราะห์แบบง่ายได้แก่การวิเคราะห์เกี่ยวกับผลรวมของตัวเลขต่างๆ ค่าเฉลี่ย หรือสัดส่วน เช่นผลรวมของจำนวนผู้ใช้บริการยืมหนังสือในรอบปีที่ผ่านมา หรือ สัดส่วนของผู้ใช้ห้องสมุดระหว่างนักศึกษาปริญญาตรีกับนักศึกษาในระดับบัณฑิตศึกษา ซึ่งการวิเคราะห์ข้อมูลในลักษณะนี้ทำได้ง่ายและใช้เวลาไม่นาน แต่ไม่สามารถวิเคราะห์รูปแบบบางอย่างที่เป็นประโยชน์และซ่อนอยู่ภายใต้ข้อมูลเหล่านี้ได้ สำหรับการวิเคราะห์ข้อมูลแบบที่ยากขึ้น และเป็นจุดประสงค์หลักของบทความชิ้นนี้คือการทำเหมืองข้อมูล เมื่อได้ข้อมูลที่สมบูรณ์แล้วจากขั้นตอนที่ 3 ก็จะนำข้อมูลนั้นมาทำเหมืองข้อมูล ซึ่งการทำเหมืองข้อมูลนั้นจะใช้ซอฟท์แวร์สำหรับทำเหมืองข้อมูลโดยเฉพาะ โดยซอฟท์แวร์เหล่านี้มีหลายชนิดหลายราคา และมีคุณสมบัติที่แตกต่างกันออกไป ซึ่งในการเลือกใช้ซอฟท์แวร์ตัวใดนั้นจะต้องพิจารณาคุณสมบัติว่าสามารถทำงานตอบสนองกับประเภทงานที่ต้องการได้หรือไม่ และอาจต้องพิจารณาเรื่องของราคาด้วย สำหรับโปรแกรม SPSS ที่นิยมใช้วิเคราะห์งานด้านสถิตินั้นก็สามารถทำงานด้านเหมืองข้อมูลได้ โดยจะต้องซื้อส่วนของโปรแกรมที่ทำเหมืองข้อมูลมาลงเพิ่มเติมในโปรแกรม SPSS สำหรับผู้ใช้ที่ไม่สามารถซื้อโปรแกรมทำเหมืองข้อมูลที่มีราคาแพงได้ ก็มีโปรแกรมทำเหมืองข้อมูลที่สามารถนำมาใช้ฟรีๆ ได้ โดยสามารถดาวน์โหลดผ่านทางอินเตอร์เน็ต ยกตัวอย่าง เช่น โปรแกรม Weka data mining suite สำหรับตารางที่ 5 นั้นจะแสดงชื่อของซอฟท์แวร์สำหรับทำเหมืองข้อมูลบางตัวและคุณสมบัติบางอย่างของซอฟท์แวร์ตัวนั้นไว้

ตารางที่ 5 รายชื่อและคุณสมบัติบางประการของซอฟท์แวร์ทำเหมืองข้อมูล

ชื่อซอฟท์แวร์	ลิขสิทธิ์	Website ของซอฟท์แวร์
Clementine	ต้องซื้อ	http://www.spss.com/clementine/
Viscovery SOMine	ต้องซื้อ	http://www.somine.info/
Oracle Data Mining	ต้องซื้อ	http://www.oracle.com/technology/products/bi/odm/index.html
MS SQL Server	ต้องซื้อ	http://www.microsoft.com/sql/technologies/dm/default.mspx
Weka	ฟรี	http://www.cs.waikato.ac.nz/ml/weka
YALE	ฟรี	http://yale.sf.net/ หรือ http://rapid-i.com/

(ข้อมูล ณ วันที่ 30 ธันวาคม 2549)

5. ทำเหมืองข้อมูลเพื่อค้นหารูปแบบ และสร้างรายงาน

หลังจากเลือกซอฟท์แวร์ได้แล้วก็นำข้อมูลไปทำเหมืองข้อมูล เมื่อซอฟท์แวร์ได้วิเคราะห์แล้วก็จะแสดงผลออกมาในรูปของรายงานทางสถิติ หรือ แบบจำลองต่างๆ

6. วิเคราะห์ผลและนำผลที่ได้ไปประยุกต์ใช้

เมื่อได้ผลมาแล้วไม่ว่าผลจะอยู่ในรูปแบบของรายงานทางสถิติ หรือแบบจำลองที่ใช้เพื่อการทำนาย หรือรูปแบบของพฤติกรรมต่างๆ ที่สามารถวิเคราะห์ออกมาได้ จะต้องนำผลเหล่านี้มาตรวจสอบความถูกต้องเสียก่อน ซึ่งการทดสอบผลนั้นจะต้องใช้ข้อมูลที่เกี่ยวข้องในระบบห้องสมุดมาทดสอบแต่จะต้องเป็นข้อมูลคนละชุดกับข้อมูลที่นำมาใช้ทำเหมืองข้อมูล เพื่อเป็นการทดสอบว่าผลที่ได้นั้นสามารถใช้กับข้อมูลตัวอื่นๆ ได้จริงหรือไม่ และสิ่งที่สำคัญที่สุดสำหรับการทดสอบความถูกต้องของผลนั้นจะต้องให้บรรณารักษ์ หรือ เจ้าหน้าที่ในห้องสมุดที่เกี่ยวข้องกับงานตรงส่วนนั้นเป็นผู้ทดสอบผล ถ้าบรรณารักษ์หรือเจ้าหน้าที่ที่เกี่ยวข้องยอมรับผลนั้นแล้วก็แสดงว่าผลนั้นถูกต้องและนำมาใช้งานได้ แต่ถ้าบรรณารักษ์หรือเจ้าหน้าที่ไม่ยอมรับผลนั้นก็จะต้องวิเคราะห์หาสาเหตุ ซึ่งสาเหตุหลักสองสาเหตุทำให้ผลที่ได้ไม่ถูกต้องคือมีความผิดพลาดของข้อมูลที่นำมาใช้ และการเลือกใช้เครื่องมือหรือซอฟท์แวร์ในการทำเหมืองข้อมูลผิดหรือไม่เหมาะสม

เมื่อทดสอบผลและเป็นผลที่เป็นที่ยอมรับก็สามารถนำผลนั้นไปใช้งานได้ แต่การนำผลไปใช้งานนั้นจะต้องคอยติดตามเสมอว่ามีตัวแปรใดที่มีอิทธิพลต่อผลบ้าง ทั้งนี้เนื่องจากเมื่อเวลาผ่านไปและอาจมีการเปลี่ยนแปลงระบบงานในห้องสมุดทำให้มีผลต่อตัวแปรเหล่านั้น เมื่อตัวแปรเหล่านั้นเปลี่ยนก็จำเป็นจะต้องเริ่มกระบวนการทำเหมืองข้อมูลขึ้นมาใหม่ เพื่อให้ได้ผลที่สอดคล้องกับรูปแบบการทำงานแบบใหม่

สรุป

การทำเหมืองข้อมูลในห้องสมุดเป็นการนำประยุกต์ความรู้ด้านฐานข้อมูล คลังข้อมูล และเหมืองข้อมูล ซึ่งเป็นวิทยาการด้านคอมพิวเตอร์มาประยุกต์ใช้กับข้อมูลต่างๆ ที่เก็บไว้ในระบบฐานข้อมูลของห้องสมุด การทำเหมืองข้อมูลในห้องสมุดจะทำให้ทราบถึงพฤติกรรม กลุ่ม หรือประเภทของผู้ใช้บริการ ซึ่งสามารถนำผลลัพธ์ที่ได้ไปปรับปรุงบริการด้านต่างๆ ของห้องสมุด ไม่ว่าจะเป็นบริการเกี่ยวกับการยืมและคืนหนังสือ หรือบริการพื้นฐานต่างๆ ที่ห้องสมุดมีก็ตาม นอกจากนี้ผลที่ได้จากการทำเหมืองข้อมูลยังสามารถนำมาประยุกต์ใช้เพื่อวางนโยบายต่างๆ ในระดับบริหารเพื่อให้ผู้บริหารห้องสมุดสามารถบริหารงานด้านต่างๆ ในห้องสมุดได้อย่างมีประสิทธิภาพ สำหรับขั้นตอนการทำเหมืองข้อมูลในห้องสมุดนั้นมี 6 ขั้นตอน ซึ่งแต่ละขั้นตอนมีรายละเอียดแตกต่างกัน อีกทั้งแต่ละขั้นตอนยังใช้เวลาการทำงานที่แตกต่างกันซึ่งขั้นตอนที่ใช้เวลามาที่สุดคือการสร้างคลังข้อมูล ซึ่งผลที่ได้จากการทำเหมืองข้อมูลจะดีหรือไม่ก็ขึ้นกับว่ามีข้อมูลที่ถูกต้อง ครบถ้วนหรือไม่ ดังนั้นหากห้องสมุดใดมีนโยบายที่จะทำเหมืองข้อมูลแล้วจะต้องมีการเก็บข้อมูลด้านต่างๆ ของห้องสมุดไว้ในฐานข้อมูลอย่างเป็นระบบ ครบถ้วน และต้องเก็บสะสมข้อมูลเหล่านั้นให้ได้ช่วงเวลาหนึ่งเพื่อให้ได้ข้อมูลที่มากพอนำไปทำเหมืองข้อมูลได้

เอกสารอ้างอิง

กิตติพงษ์ กลมกล่อม (2548). การออกแบบและพัฒนาคลังข้อมูล (Data Warehouse) พิมพ์ครั้งที่ 2. เคทีพี คอมพ์ แอนด์ คอนซัลท์. กรุงเทพฯ.

Han, J. and Kamer, M. (2001). Data Mining: Concept and Techniques. Academic Press, San Diego, CA.

Larose, Daniel T. (2005). Discovering Knowledge in Data: An Introduction to Data Mining. John Wiley & Sons Inc., Hoboken, New Jersey.

Lavoie, B. , Demsey, L. and Connaway, Lynn S. (2006). Making Data Work Harder. Library Journal, January 2006, P.40

Nicholson, S.(2003). The Bibliomining Process: Data Warehousing and Data Mining for Library Decision-Making. Information Technology and Libraries 22 (4).

Nicholson, S. and Stanton, J. (2003). Gaining strategic advantage through bibliomining: Data miming for management decisions in corporate, special, digital, and traditional libraries. In Nemati, H. & Barko, C.(Eds.). Organizational data mining: Leveraging enterprise data resources for optional performance. Hershey, PA: Idea Group Publishing. P.247-262.

Tang, Z.H. and Maclennan, J.(2005). Data Mining with SQL Server 2005. Wiley Publishing. Indianapolis. IN.

สำนักบรรณสารการพัฒนา
LIBRARY AND INFORMATION CENTER