

## บทที่ 6

### บทสรุป ปัญหาและข้อเสนอแนะ

สำหรับบทนี้จะกล่าวถึงบทสรุปของงานวิจัย ปัญหาและอุปสรรคที่เกิดขึ้นระหว่างการทำวิจัยซึ่งนำเสนอวิธีการหรือข้อเสนอแนะเพื่อแก้ปัญหาดังกล่าว นอกจากนี้ยังอธิบายถึงข้อจำกัดของแม่แบบที่นำเสนอ รวมทั้งข้อเสนอแนะสำหรับผู้สนใจพัฒนางานวิจัยนี้ต่อไป

#### 6.1 สรุปงานวิทยานิพนธ์

งานวิทยานิพนธ์นี้นำเสนอแม่แบบ MineCache ซึ่งเป็นแม่แบบของเว็บแคชซึ่งที่ปรับปรุงประสิทธิภาพการทำงานด้วยการใช้เทคนิคของการทำเหมืองข้อมูลบันทึกการเข้าใช้งานเว็บในการทำนายกฎความเชื่อมโยงของความสัมพันธ์ในการร้องขอ web object ของผู้ใช้งาน นั่นคือทำนายรูปแบบการร้องขอข้อมูลในอนาคตของผู้ใช้ เนื่องจากเทคนิคการทำเหมืองข้อมูลการเข้าใช้งานเว็บสามารถค้นหารูปแบบหรือความสัมพันธ์ที่มีประโยชน์จากข้อมูลบันทึกการเข้าใช้งานซึ่งเป็นตัวแทนพฤติกรรมของผู้ใช้ ในการค้นหารูปแบบการร้องขอจะใช้ขั้นตอนวิธี Apriori สำหรับการทำนายเพราะเป็นขั้นตอนวิธีหาความกฎเชื่อมโยงที่เหมาะสมสำหรับการดำเนินการกับข้อมูลจำนวนมากดังเช่นข้อมูลจากบันทึกการเข้าใช้งาน กฎความเชื่อมโยงที่ได้นำมาร้องขอข้อมูลเก็บไว้ล่วงหน้าก่อนมีการร้องขอจริง จากการทดสอบการทำงานแสดงให้เห็นว่าแม่แบบ MineCache มีประสิทธิภาพการทำงานที่ดี โดยแม่แบบ MineCache มีค่า hit ratio สูงกว่าการทำงานของเว็บแคชเชิงปกติ นอกจากนี้ค่า byte hit ratio ก็มีค่าสูงขึ้นเช่นเดียวกัน รายละเอียดผลสรุปของงานวิทยานิพนธ์ประกอบด้วย การสรุปการทำงานของแม่แบบ MineCache นำเสนอรายละเอียดในหัวข้อที่ 6.1.1 และสรุปผลการทำนายนำเสนอในหัวข้อที่ 6.1.2

### 6.1.1 สรุปการทำงานของแม่แบบ MineCache

ผลที่ได้จากงานวิจัยคือแม่แบบ โดยสถาปัตยกรรมของแม่แบบที่ได้ประกอบด้วยการทำงาน 3 ส่วนหลักคือ

1. หน่วยการทำงานเว็บแคชชิง เป็นส่วนการทำงานหลักที่ให้บริการกับผู้ใช้เมื่อผู้ใช้ร้องขอเว็บใดๆ ทำหน้าที่เป็นตัวแทนในการติดต่อกับเครือข่ายภายนอกให้กับผู้ใช้บริการ และบันทึก web object ที่มีการร้องขอไว้เพื่อให้ในอนาคตหากมีผู้ใช้ร้องขอ web object เดียวกันนี้ก็จะส่ง web object นั้นๆ ไปยังผู้ใช้ได้ทันที โดยในการบันทึก web object ที่มีการร้องขอมีการแยกการจัดเก็บระหว่าง web object ที่เกิดจากการร้องขอใช้งานตามปกติบันทึก web object ใน Normal cache กับ web object ที่ทำการร้องขอก่อนล่วงหน้า ซึ่งบันทึก web object ใน Prefetched cache
2. หน่วยการทำเหมืองข้อมูลบันทึกการใช้งานเว็บ ทำหน้าที่ทำนาย web object ซึ่งในงานวิทยานิพนธ์นี้คือ URL ที่ผู้ใช้จะร้องขอในอนาคตอันใกล้ด้วยการพิจารณาความสัมพันธ์ของการร้องขอจากข้อมูลบันทึกการใช้งานเว็บที่โปรแกรมเว็บแคชชิงบันทึก
3. หน่วยการดึงข้อมูลล่วงหน้า ทำหน้าที่ร้องขอข้อมูลที่ได้จากกฎความเชื่อมโยงล่วงหน้า ซึ่งร้องขอผ่านโปรแกรมเว็บแคชชิงของแม่แบบ โดยบันทึก web object ที่ตอบกลับในส่วนของ Prefetched cache แยกจากการทำงานปกติเพื่อไม่ให้เกิดความสับสนในการหากฎในครั้งต่อไป

### 6.1.2 สรุปผลของการทำงานของแม่แบบ MineCache

การทำงานแต่ละหน่วยภายในแม่แบบที่นำเสนอจะมีการทำงานประสานกัน ดังนั้นแม่แบบ MineCache ที่นำเสนอสามารถให้บริการแก่ผู้ใช้บริการ สร้างกฎความเชื่อมโยง และร้องขอ web object ที่ทำนายมาเก็บล่วงหน้าได้ ทำให้ประสิทธิภาพการทำงานของเว็บแคชชิงดีขึ้น โดยผลสรุปของการทำงานสามารถสรุปผลตามประเด็นต่อไปนี้คือ ประสิทธิภาพของแม่แบบนั้นคือค่า hit ratio ขนาดของแคชที่จะใช้ ความน่าเชื่อถือของกฎความเชื่อมโยงที่ได้จากการทำนาย ซึ่งรายละเอียดของแต่ละประเด็นนำเสนอดังต่อไปนี้

1. ประสิทธิภาพของแม่แบบ เนื่องจากวัตถุประสงค์ของแม่แบบที่นำเสนอคือเพิ่มประสิทธิภาพของเว็บแคชชิง ซึ่งประสิทธิภาพการทำงานสามารถวัดได้หลายประเด็น แต่ประเด็นที่มุ่งเน้นในงานวิจัยคือการเพิ่มค่า hit ratio ให้สูงขึ้น จากการทดสอบการทำงานดังที่เสนอในบทที่ 5 แสดงให้เห็นว่าแม่แบบ MineCache ที่เสนอทำให้ค่า hit ratio สูงขึ้น โดยอัตราส่วนของค่า hit ratio ที่เพิ่มขึ้นเฉลี่ยประมาณร้อยละ 75
2. ขนาดของแคชที่เหมาะสม เนื่องจากผลการทำงานของแม่แบบที่นำเสนอสำหรับการวัดประสิทธิภาพของระบบจะมีปัจจัยที่ส่งผลต่อประสิทธิภาพนั่นก็คือ ขนาดของแคชที่จะบันทึก ซึ่งถ้าขนาดของแคชสามารถบันทึกข้อมูลได้มากจะทำให้ค่า hit ratio เพิ่มขึ้นตามไปด้วยดังที่ได้แสดงในตารางที่ 5-1 แต่อาจมีประเด็นเรื่องของความใหม่ของข้อมูลและพื้นที่ที่จะใช้บันทึกซึ่งหมายถึงตัวอุปกรณ์ทางกายภาพ แต่ถ้าขนาดของแคชมีขนาดเล็กข้อมูลที่บันทึกก็จะน้อยส่งผลให้ค่า hit ratio ต่ำ แต่มีความเป็นไปได้ที่ข้อมูลจะมีความทันสมัยเพราะไม่มีการเก็บข้อมูลไว้นานๆ เพราะฉะนั้นการเลือกขนาดของแคชที่ใช้ขึ้นอยู่กับความต้องการหรือนโยบายที่กำหนดว่าต้องการเน้นความสำคัญที่ประเด็นใด ถ้าต้องการค่า hit ratio สูงโดยมีความพร้อมทางด้านอุปกรณ์และไม่คำนึงถึงความทันสมัยของข้อมูลสามารถทำการกำหนดขนาดของแคชให้มากได้ แต่ถ้ามีข้อจำกัดเรื่องค่าใช้จ่ายสำหรับอุปกรณ์และต้องการให้ผู้ใช้บริการมีความพอใจในระดับหนึ่งก็สามารถกำหนดขนาดของแคชให้น้อยได้ ดังนั้นการกำหนดขนาดจึงขึ้นอยู่กับความต้องการและนโยบาย
3. ความน่าเชื่อถือของกฎความเชื่อมโยงที่ได้จากการทำนาย เนื่องจากการทำเหมืองข้อมูลบันทึกการใช้งานเว็บต้องเสียทรัพยากรบางส่วนทั้งพื้นที่สำหรับจัดเก็บกฎความเชื่อมโยงที่ได้ ทั้งเวลาในการประมวลผลและดำเนินการ หรือแม้กระทั่งพื้นที่ของแคชที่บันทึก web object ที่ร้องขอล่วงหน้าจากการทำนาย เพราะฉะนั้นหากกฎความเชื่อมโยงที่หาได้ไม่มีการร้องขอจริงตามที่คาดการณ์จะทำให้สูญเสียทรัพยากรโดยเปล่าประโยชน์ แต่จากการทดสอบการทำงานพบว่ากฎความเชื่อมโยงที่ทำนายได้เมื่อทำการร้องขอ web object ดังกล่าวมาก่อนล่วงหน้าจะได้รับการร้องขอจริงโดยมีเปอร์เซ็นต์ของการได้รับการร้องขอสูงถึงร้อยละ 96 ดังแสดงรายละเอียดตารางที่ 5-4 ในบทที่ 5

อย่างไรก็ตามถึงแม้ในการทดสอบจะใช้ข้อมูลจริงจากบันทึกการใช้งานของเว็บ แคลซิงของมหาวิทยาลัยในการทดสอบ แต่การทดสอบจำกัดเฉพาะในห้องทดลองเท่านั้นเนื่องจากไม่สามารถนำแม่แบบไปทดสอบกับระบบเครือข่ายจริงของมหาวิทยาลัย เพราะฉะนั้นผลการทำงานที่ได้จึงเป็นทำงานในสภาพแวดล้อมจำลอง แต่สามารถขยายการทำงานเพื่อใช้งานจริงในอนาคตได้เนื่องจากสภาพแวดล้อมของระบบจำลองมีความใกล้เคียงกับระบบจริง

## 6.2 อุปสรรคและปัญหาที่พบในงานวิทยานิพนธ์

อุปสรรคและปัญหาที่พบในการทำวิจัยนี้ มีดังนี้

1. เครื่องมือที่เลือกใช้สำหรับการทดลองในตอนเริ่มการทำงานไม่เหมาะกับงาน ทำให้ต้องเสียเวลาในการศึกษาเครื่องมือใหม่ กล่าวคือเลือกใช้โปรแกรม squid ซึ่งเป็น โปรแกรมประยุกต์ที่มีความสามารถทำงานเป็นเว็บแคชซิง และ โปรแกรม squid เป็น โปรแกรมประยุกต์ Open source ที่ได้รับการพัฒนาความสามารถในการทำงานเพิ่มเติมหลายการทำงานเช่น ทำงานเป็น Web server เป็นต้น ด้วยสาเหตุข้างต้นทำให้โปรแกรม squid เป็นโปรแกรมขนาดใหญ่ที่มีส่วนการทำงานหลายส่วน ซึ่งแต่ละส่วนการทำงานก็จะมีการทำงานที่เกี่ยวข้องกัน เพราะฉะนั้นทำให้ไม่สามารถจัดการเปลี่ยนแปลงโปรแกรมตามแนวคิดที่เสนอในงานวิจัย โดยเฉพาะอย่างยิ่งในงานวิจัยมีวัตถุประสงค์เพื่อเสนอแม่แบบเว็บแคชซิงที่มีประสิทธิภาพ ดังนั้นโปรแกรมที่ใช้จึงควรทำงานเป็นเว็บแคชซิงอย่างเดียวเพื่อให้เห็นภาพของการทำงานได้อย่างชัดเจน
2. การทดลองการทำงานต้องใช้ข้อมูลทดสอบที่มีขนาดใหญ่ และเนื่องจากใช้ข้อมูลจริงในการทดลองหากภูความเชื่อมโยง ทำให้ต้องใช้เวลาในการเก็บข้อมูลจริงมาทดสอบ และต้องใช้เวลาในการทดสอบเพื่อหาค่าในการทดลองที่เหมาะสม
3. เนื่องจากข้อมูลที่ใช้ในการหาภูความเชื่อมโยงประกอบด้วย เวลาการร้องขอ หมายเลข IP ที่ทำการร้องขอ และ URL ที่มีการร้องขอ ซึ่งข้อมูล URL เป็นข้อมูลที่เป็นสายอักขระที่มีความยาวของข้อมูลค่อนข้างยาว และใช้เนื้อที่ในการจัดเก็บมาก เมื่อทำการการหาภูความเชื่อมโยงจากข้อมูลดังกล่าวจะทำให้เกิดปัญหาโปรแกรมหยุดทำงานเอง เนื่องจากขนาดของผลลัพธ์ที่ได้เกินกว่าขนาดที่ระบบปฏิบัติการอนุญาต

4. ต้องใช้เวลาในการทดสอบเนื่องจากการทดสอบที่ต้องติดต่อกับเว็บ ซึ่งการทำงานจะขึ้นอยู่กับสภาพของเครือข่าย
5. ลักษณะการทำงานของภาษาโปรแกรมที่ใช้ส่งผลต่อการทำงานรวมทั้งระบบเนื่องจากโปรแกรมเว็บแคชซิงที่ใช้ในงานวิจัยพัฒนาด้วยภาษาจาวา ซึ่งมีข้อดีในเรื่องของการที่สามารถใช้งานบนระบบปฏิบัติการใดก็ได้ แต่ก็ยังเป็นประเด็นที่ทำให้เกิดปัญหาเช่นกัน เนื่องจากโปรแกรมที่พัฒนาด้วยภาษาจาวาจะทำการแปลรหัสโปรแกรมด้วยเครื่องมือเฉพาะที่เรียกว่า Java Virtual Machine (JVM) จัดการการทำงานทั้งหมดของโปรแกรมในภาษาจาวา ซึ่งจะใช้หน่วยความจำในการทำงานมากกว่าตัวแปลภาษาแบบอื่นๆ และการจัดการเรื่องการการกินหน่วยความจำที่ไม่ได้ใช้งานแล้วผู้พัฒนาโปรแกรมจาวาไม่สามารถจัดการได้เองต้องให้ JVM จัดการให้ เพราะฉะนั้นจึงมีปัญหาในเรื่องของหน่วยความจำไม่พอเมื่อโปรแกรมทำงานไปในระยะเวลาานาน

### 6.3 การแก้ปัญหาและข้อเสนอแนะสำหรับปัญหาที่พบในงานวิทยานิพนธ์

จากปัญหาที่เกิดขึ้นระหว่างการทำวิจัยดังกล่าวข้างต้น หัวข้อนี้นำเสนอวิธีการแก้ปัญหาที่พบหรือ นำเสนอแนวทางหรือข้อเสนอแนะสำหรับบางปัญหาที่ยังไม่สามารถแก้ปัญหาได้สมบูรณ์ดังต่อไปนี้

1. ปัญหาการเลือกใช้งาน โปรแกรมเว็บแคชซิงตอนเริ่มต้นที่เลือกใช้โปรแกรม squid และพบปัญหาเรื่องของการความซับซ้อนของโปรแกรมที่มีการพัฒนาเพิ่มเติม จึงได้ทำการศึกษาโปรแกรมเว็บแคชซิงอื่นๆ ที่มีความซับซ้อนไม่มาก เพื่อให้สามารถจัดการเปลี่ยนแปลง และปรับปรุงการทำงานตามแนวคิดที่นำการทำเหมืองข้อมูลมาประยุกต์ใช้ได้อย่างสะดวก ดังนั้นจึงเลือกโปรแกรมชื่อ Jproxy ที่พัฒนาด้วยภาษาจาวา เนื่องจากเป็น โปรแกรมที่ทำงานเฉพาะส่วนของการให้บริการเป็นเว็บแคชซิงและมีความสามารถของ proxy ด้วยซึ่งเหมาะสมกับความต้องการและสามารถปรับปรุงโปรแกรมได้ง่าย
2. การแก้ปัญหาในเรื่องของขนาดของข้อมูลที่ได้จากการหาความเชื่อมโยง เนื่องจากข้อมูลที่ใช้เป็นตัวอักษรที่มีขนาดของข้อมูลใหญ่ แก้ปัญหาโดยการแปลงข้อมูล URL ให้อยู่ในรูปของตัวเลขจำนวนเต็ม หรือเข้ารหัสข้อมูลให้

เป็นตัวเลขนั่นเอง นอกจากจะใช้พื้นที่การเก็บข้อมูลลดลงแล้วยังทำให้สามารถทำงานกับข้อมูลจำนวนมากได้ ช่วยลดระยะเวลาในการทำงานให้น้อยลง

3. ปัญหาจากเรื่องการจัดการหน่วยความจำในระหว่างการทำงานของตัวแปลภาษาจาวา เนื่องจากโปรแกรมที่พัฒนาด้วยจาวามีข้อดีคือทำงานบนระบบปฏิบัติการใดก็ได้ ซึ่งผู้วิจัยต้องการนำเสนอแนวคิดของการทำงานเท่านั้นจึงใช้โปรแกรมขนาดเล็กที่ไม่ซับซ้อน หากต้องการแก้ปัญหาในประเด็นข้างต้นมีข้อเสนอแนะให้ปรับเปลี่ยนโปรแกรมที่พัฒนาด้วยภาษาจาวาเป็นภาษาซี เพื่อลดการใช้ทรัพยากรในการแปลรหัสโปรแกรมถ้าต้องการใช้งานจริง หรือเพิ่มศักยภาพการทำงานของแม่แบบให้มากขึ้น
4. ปัญหาเรื่องของสภาพเครือข่ายเป็นปัจจัยเหนือการควบคุม จึงยังไม่สามารถสรุปวิธีการแก้ปัญหาที่ชัดเจนได้ แต่มีข้อเสนอแนะคืออาจทำการทดสอบการทำงานในช่วงระยะเวลาตอนกลางคืนหรือช่วงหยุด เนื่องจากเป็นช่วงเวลาที่มีการใช้งานเครือข่ายน้อยอาจช่วยลดระยะเวลาของการทดสอบให้สามารถทำงานได้เร็วขึ้น

#### 6.4 ข้อจำกัดของแม่แบบในงานวิทยานิพนธ์

เนื่องจากงานวิทยานิพนธ์นี้นำเสนอแม่แบบของระบบการทำงานเพื่อเสนอแนวคิดในการพัฒนาประสิทธิภาพของเว็บแคชชิง ด้วยการนำเทคนิคการทำเหมืองข้อมูลบันทึกการเข้าใช้งานเว็บซึ่งเป็นเทคนิคที่พิจารณาความสัมพันธ์จากข้อมูลเว็บจำนวนมากเท่านั้น จึงนำเครื่องมือที่มีอยู่แล้วมาพัฒนาปรับปรุงตามแนวคิดที่นำเสนอเพื่อให้สามารถเห็นภาพว่าสามารถนำมาใช้งานได้จริง เพราะฉะนั้น โปรแกรมเว็บแคชชิงที่ใช้ทดสอบจึงเป็น โปรแกรมขนาดเล็กที่สามารถรองรับการทำงานได้ในระดับหนึ่ง นอกจากนี้ถึงแม้จะใช้ข้อมูลจริงในการทดลองการทำงานบางส่วนแต่ไม่สามารถนำระบบที่นำเสนอไปทดลองใช้งานจริงได้ เพราะโปรแกรมไม่สามารถรองรับการทำงานจริงที่เกิดขึ้นได้ สาเหตุตามที่กล่าวมาแล้วข้างต้น

#### 6.5 ข้อเสนอแนะ

จากการทำวิทยานิพนธ์และศึกษางานวิจัยที่เกี่ยวข้องผู้วิจัยขอเสนอแนะประเด็นในการนำงานวิจัยไปพัฒนาต่อในอนาคตดังนี้

1. การขยายความสามารถของขั้นตอนการเตรียมข้อมูลเพื่อให้สามารถวิเคราะห์ข้อมูล web object ที่ไม่สามารถบันทึกได้อันได้แก่ web object ประเภทที่มีการประมวลผลที่เครื่องให้บริการเช่น php, asp, cgi เป็นต้นและเพิ่มความสามารถในการแบ่งกลุ่มข้อมูลและแบ่งช่วงการทำงานให้เหมาะสม ซึ่งจะสามารถทำให้สร้างกฎความเชื่อมโยงที่น่าสนใจมากยิ่งขึ้น
2. งานวิจัยนี้สร้างระบบ MineCache เพื่อเป็นแม่แบบในการทำงานเท่านั้น เพราะฉะนั้นบางส่วนการทำงานจึงมีข้อจำกัดในการเอาไปทดสอบสำหรับการพัฒนา หรือนำไปใช้ในระบบจริง ตัวอย่างเช่น โปรแกรมเว็บแคชชิง Jproxy เป็นต้น ดังนั้นหากต้องการนำไปใช้งานกับระบบจริงควรมีการปรับปรุงประสิทธิภาพของโปรแกรมให้ดีขึ้น โดยคำนึงถึงเรื่องการจัดการหน่วยความจำโดยเฉพาะการคืนหน่วยความจำที่ไม่ใช้แล้ว หรือพัฒนาโปรแกรมด้วยภาษาโปรแกรมอื่นๆเช่น ภาษาซี เป็นต้น

## 6.6 สรุป

งานวิทยานิพนธ์นี้เสนอแนวความคิดการเพิ่มประสิทธิภาพการทำงานของเว็บแคชชิงด้วยการทำเหมืองข้อมูลบันทึกการเข้าใช้งานเว็บ โดยการเลือกใช้เทคนิคการทำเหมืองข้อมูลการเข้าใช้งานเว็บเนื่องจากต้องการทำนายรูปแบบการร้องขอข้อมูลโดยพิจารณาจากพฤติกรรมการใช้งานของผู้ใช้ที่บันทึกรูปแบบการร้องขอในบันทึกการเข้าใช้งาน การพัฒนาแม่แบบการทำงานเว็บแคชชิงในงานวิทยานิพนธ์มีขั้นตอนทำงานคือการทำนายความสัมพันธ์ของการร้องขอข้อมูลของผู้ใช้โดยใช้ขั้นตอนวิธี Apriori ซึ่งเป็นขั้นตอนวิธีการทำนายหากฎความเชื่อมโยงที่เหมาะสมกับข้อมูลจำนวนมากดังเช่นข้อมูลในบันทึกการเข้าใช้งานเว็บเพื่อทำนายว่า web object ใดที่จะมีการร้องขอในอนาคตอันใกล้ และร้องขอ web object ดังกล่าวที่ได้จากการทำนายมาเก็บล่วงหน้าก่อนมีการร้องขอจริง ซึ่งแม่แบบ MineCache ที่นำเสนอแสดงให้เห็นว่าเว็บแคชชิงมีประสิทธิภาพการทำงานดีขึ้น โดยอัตราการพบข้อมูลที่ต้องการ (hit ratio) มีค่าสูงขึ้นเฉลี่ยประมาณร้อยละ 75 และอัตราการพบข้อมูลจำนวนไบต์ (byte hit ratio) มีค่าสูงขึ้นเฉลี่ยประมาณร้อยละ 20