

# บทที่ 1

## บทนำ

### 1.1 ความสำคัญและที่มาของงานวิจัย

อินเทอร์เน็ต (Internet) เกิดขึ้นจากแนวคิดของหน่วยงานรัฐบาลกลางสหรัฐชื่อ Advanced Research Project Agency (ARPA) ซึ่งมีวัตถุประสงค์หลักเพื่อสร้างเครือข่ายที่สามารถเชื่อมโยงเครื่องคอมพิวเตอร์สำหรับงานวิจัยด้านการทหารในห้องปฏิบัติการของมหาวิทยาลัยต่างๆ ให้สามารถติดต่อกันได้ ดังนั้นเครือข่ายอินเทอร์เน็ตในตอนต้นผู้ใช้งานส่วนใหญ่จึงเป็นกลุ่มของนักวิจัยในมหาวิทยาลัยและสถาบันวิจัยของสหรัฐอเมริกา แต่ต่อมาเครือข่ายนี้ถูกแยกออกมาใช้งานสำหรับบุคคลทั่วไปไม่จำเป็นต้องเป็นอาจารย์หรือนักวิจัยก็สามารถใช้งานได้ จากสาเหตุข้างต้นทำให้อินเทอร์เน็ตกลายเป็นเครือข่ายสาธารณะขนาดใหญ่ซึ่งใครก็สามารถใช้งานได้ถ้ามีเครื่องคอมพิวเตอร์และอุปกรณ์สำหรับเชื่อมต่อ บริการบนอินเทอร์เน็ตมีหลายประเภทแต่ที่รู้จักกันอย่างดีคือ เวิลด์ ไวด์ เว็บ (World Wide Web) หรือเรียกสั้นๆว่าเว็บ (WWW) โดยเว็บคือแหล่งรวมของเอกสารที่มีการเชื่อมโยงกันทำให้ผู้ใช้เว็บสามารถอ่านเอกสารอื่นที่เกี่ยวข้องได้อย่างสะดวก นอกจากความสามารถในการเชื่อมโยงเอกสารต่างๆแล้วเว็บยังมีโปรแกรมสำหรับอำนวยความสะดวกในการค้นหาข้อมูลและแสดงเอกสารเรียกว่าเว็บเบราว์เซอร์ (web browser) ซึ่งโปรแกรมนี้สามารถแสดงข้อมูลได้หลากหลายประเภทไม่จำกัดเฉพาะตัวอักษรเท่านั้น อาจเป็นรูปภาพหรือเสียงก็ได้ จากความสามารถข้างต้นทำให้เว็บกลายเป็นบริการบนอินเทอร์เน็ตที่ได้รับการตอบรับเป็นอย่างดีจากผู้ใช้

การขยายขอบเขตการใช้อินเทอร์เน็ตออกไปนอกเหนือจากแวดวงของกลุ่มนักวิชาการและนักวิจัยส่งผลให้การใช้งานเว็บสูงขึ้นตามจำนวนผู้ใช้บริการที่เพิ่มขึ้นซึ่งนำไปสู่ปัญหาบางประการเช่น ปัญหาความคับคั่งของข้อมูลบนเส้นทางสื่อสาร หรือการค้นหาข้อมูลข่าวสารที่ต้องการใช้เวลานานเป็นต้น จากปัญหาดังกล่าวนักวิจัยที่เกี่ยวข้องได้นำเสนอวิธีการแก้ปัญหาต่างๆเพื่อเพิ่มประสิทธิภาพการทำงานของเว็บ เทคโนโลยีหนึ่งที่น่าสนใจคือเทคโนโลยีเว็บแคชชิง (web caching) ซึ่งเป็นเทคนิคที่ทำให้เครื่องคอมพิวเตอร์หนึ่งเครื่อง หรือมากกว่าหนึ่งเครื่องทำหน้าที่เป็นแหล่งเก็บข้อมูลจากการร้องขอผ่านเว็บ โดยพื้นที่สำหรับจัดเก็บข้อมูลนี้เรียกว่าแคช (cache) นอกจากการบันทึกข้อมูลที่ร้องขอแล้วกลุ่มวิจัยเกี่ยวกับเว็บแคชชิงนำความสามารถของการให้บริการแทนหรือที่เรียกว่า proxy มาเพิ่มความสามารถของเว็บแคชชิงให้ทำงานได้

มากกว่าการเก็บข้อมูลกล่าวคือเว็บแคชซึ่งสามารถทำหน้าที่ใช้ร้องขอข้อมูลแทนผู้ใช้ได้ เพราะฉะนั้นเว็บแคชจึงเป็นเทคนิคที่ใช้ในการร้องขอข้อมูลจากผู้ใช้บริการ (server) แทนผู้ใช้บริการ (client) และบันทึกข้อมูลนี้เพื่อใช้งานในอนาคตเช่น ผู้ใช้บริการเว็บรายอื่นต้องการข้อมูลนี้ด้วย เป็นต้น

เว็บแคชซึ่งช่วยเพิ่มประสิทธิภาพการทำงานของเว็บโดยลดระยะเวลาการรอข้อมูลตอบกลับเนื่องจากเครื่องให้บริการเว็บแคชซึ่งจะติดตั้งอยู่ใกล้กับเครื่องผู้ใช้บริการมากกว่าเครื่องให้บริการจริง ดังนั้นระยะทางในการเดินทางจากเครื่องให้บริการเว็บแคชถึงเครื่องผู้ใช้บริการสั้นกว่าระยะทางจากเครื่องให้บริการจริงมาถึงเครื่องผู้ใช้บริการ นอกจากนี้เว็บแคชยังช่วยลดปริมาณงานที่ผู้ใช้บริการต้องดำเนินการเนื่องจากถ้าข้อมูลที่ผู้ใช้บริการต้องการมีอยู่ในแคช เครื่องให้บริการจริงไม่จำเป็นต้องส่งข้อมูลมาให้ แต่เครื่องให้บริการเว็บแคชซึ่งส่งข้อมูลนั้นๆ ให้กับผู้ใช้ ดังนั้นถ้าสามารถปรับปรุงประสิทธิภาพของเว็บแคชซึ่งให้ดีขึ้นจะส่งผลให้ประสิทธิภาพของเว็บดีขึ้นด้วยเช่นกัน ในการวัดประสิทธิภาพการทำงานของเว็บแคชซึ่งสามารถพิจารณาได้หลายประเด็นแต่ประเด็นที่นิยมนำมาใช้วัดคืออัตราการพบข้อมูลที่ต้องการในแคช (hit ratio) และอัตราข้อมูลจำนวนไบต์ที่พบในแคช (byte hit ratio) ซึ่งถ้า hit ratio มีค่าสูงหมายความว่าข้อมูลที่ผู้ใช้ต้องการมีในแคช และสามารถส่งไปให้เครื่องผู้รับบริการได้ทันทีทำให้ผู้ใช้ไม่ต้องรอนาน

แคชของเว็บแคชซึ่งจะกำหนดพื้นที่บางส่วนของฮาร์ดดิสก์ (hard disk) สำหรับจัดเก็บข้อมูลที่ร้องขอด้วยเหตุนี้แคชจึงมีขนาดจำกัดตามที่กำหนด ดังนั้นเมื่อเว็บแคชซึ่งทำงานไประยะเวลาหนึ่งข้อมูลจะบันทึกในพื้นที่ที่กำหนดจนเต็มทำให้ข้อมูลใหม่ๆ ไม่สามารถบันทึกในแคชได้ จากสาเหตุข้างต้นจึงมีการนำขั้นตอนการแทนที่ (replacement algorithm) มาใช้เพื่อจัดสรรพื้นที่สำหรับบันทึกข้อมูลของแคชที่มีขนาดจำกัดให้มีประสิทธิภาพ ซึ่งขั้นตอนการแทนที่จะมีเกณฑ์การพิจารณาสำหรับเลือกลบข้อมูลที่บันทึกในแคชที่อาจไม่มีผู้ชมใดต้องการแล้ว และบันทึกข้อมูลใหม่ที่ผู้ใช้ต้องการในอนาคตบันทึกในแคชตัวอย่างเช่น ถ้าข้อมูลใดได้รับการร้องขอบ่อยก็จะทำการเก็บข้อมูลนั้นไว้ แต่ถ้าข้อมูลใดได้รับการร้องขอน้อยจะลบข้อมูลนี้ออกจากแคชหากต้องการพื้นที่สำหรับบันทึกข้อมูลใหม่เป็นต้น ขั้นตอนวิธีการแทนที่เป็นปัจจัยที่ส่งผลต่อประสิทธิภาพการทำงานของเว็บแคชซึ่งจะช่วยเพิ่มค่า hit ratio ในแคชให้สูงขึ้นได้ แต่จากการศึกษาในงานวิจัย [Abrams, 1995] พบว่าไม่ว่าจะใช้ขั้นตอนวิธีการแทนที่วิธีใดก็ตามค่า hit ratio ที่เป็นไปได้ที่มากที่สุดจะมีค่าประมาณ 30 ถึง 50 เปอร์เซ็นต์เท่านั้น ดังนั้นหากสามารถทำนายการร้องขอที่จะเกิดขึ้นในอนาคตและนำข้อมูลดังกล่าวบันทึกในแคชก่อนการใช้งานจริงส่งผลให้ค่า hit ratio เพิ่มขึ้น โดยการทำนายคำร้องขอจากประวัติของพฤติกรรมการใช้งานจะช่วยให้การทำนายใกล้เคียงกับการใช้งานจริง ซึ่งการทำนายการร้องขอล่วงหน้าจากพฤติกรรมของผู้ใช้สามารถทำได้

โดยการวิเคราะห์ข้อมูลบันทึกการเข้าใช้งานเว็บ หรือที่เรียกว่า web access log เนื่องจากข้อมูลข้างต้นเป็นรูปแบบการใช้งานของผู้ใช้ซึ่งเป็นตัวแทนพฤติกรรมการใช้งานจริงการค้นหาความสัมพันธ์และทำนาย

การขยายตัวของการใช้งานเว็บที่เพิ่มขึ้นส่งผลให้ผู้ใช้ประสบปัญหาในการใช้บริการเว็บ เว็บแคชจึงจึงเป็นทางเลือกหนึ่งในการเพิ่มประสิทธิภาพเว็บโดยหากปรับปรุงการทำงานของเว็บแคชจึงให้ดีขึ้นก็จะส่งผลให้เว็บให้บริการได้อย่างเต็มที่มากยิ่งขึ้น แต่เว็บแคชยังมีข้อด้อยในเรื่องเนื้อที่จัดเก็บมีขนาดจำกัด การแทนที่ข้อมูลในแคชจึงเป็นวิธีที่นำมาใช้สำหรับแก้ปัญหาแต่ก็ยังไม่สามารถเพิ่มค่า hit ratio ของเว็บแคชจึงให้สูงขึ้นมากกว่าเดิมเนื่องจากการทำงานที่ใช้ค่าข้อมูลทางสถิติในการพิจารณาซึ่งไม่ได้เป็นตัวแทนพฤติกรรมของผู้ใช้บริการ ดังนั้นถ้าสามารถทำนายข้อมูลที่จะมีการร้องขอสอดคล้องตามความต้องการของผู้ใช้บริการงานจริงจะช่วยเพิ่มประสิทธิภาพการทำงานของเว็บแคชจึงให้ดีขึ้น ส่งผลให้การทำงานของเว็บดีขึ้นเช่นกัน

งานวิทยานิพนธ์นี้นำเสนอแม่แบบการทำงานเพื่อเพิ่มประสิทธิภาพเว็บแคชจึงให้มีค่า hit ratio สูงขึ้น โดยการดึงข้อมูลล่วงหน้าจากการทำนายซึ่งใช้ข้อมูล web access log เพื่อค้นหาความสัมพันธ์และทำนายข้อมูลที่จะได้รับการร้องขอในอนาคตแทนการใช้ค่าทางสถิติในการพิจารณา สำหรับการทำนายจะใช้เทคนิคการทำเหมืองข้อมูลเว็บซึ่งอธิบายรายละเอียดในบทที่ 2 เมื่อทำนายข้อมูลที่จะมีการใช้ในอนาคดแล้วทำการร้องขอข้อมูลดังกล่าวบันทึกในแคชของเว็บแคชจึงล่วงหน้า หากสามารถทำนายการร้องขอที่จะเกิดขึ้นล่วงหน้าได้ถูกต้องหรือแม่นยำในระดับหนึ่งจะส่งผลให้ค่า hit ratio สูงขึ้น และความเร็วในการได้รับข้อมูลตอบกลับจากผู้ให้บริการจะรวดเร็วขึ้นตามไปด้วย

สำหรับเอกสารส่วนที่เหลือจะกล่าวถึงการตรวจเอกสารในหัวข้อที่ 1.2 และหัวข้อที่ 1.3 นำเสนอวัตถุประสงค์ของการทำวิจัย สำหรับหัวข้อที่ 1.4 นำเสนอขอบเขตของการดำเนินการวิจัย และขั้นตอนการดำเนินการในหัวข้อที่ 1.5 สำหรับในหัวข้อที่ 1.6 นำเสนอระยะเวลาการดำเนินงาน หัวข้อที่ 1.7 เครื่องมือและอุปกรณ์ที่ใช้ในงานวิจัย หัวข้อที่ 1.8 นำเสนอประโยชน์ที่คาดว่าจะได้รับ และในหัวข้อสุดท้ายหัวข้อที่ 1.9 นำเสนอบทสรุป

## 1.2 การตรวจเอกสาร

เว็บแคชจึงเป็นเทคนิคหนึ่งสำหรับปรับปรุงประสิทธิภาพการทำงานของเว็บ หลักการทำงานของเว็บแคชจึงคือการแทนที่ข้อมูลเพื่อให้จัดการข้อมูลที่บันทึกในแคชซึ่งมีขนาดจำกัด งานวิจัยส่วนใหญ่ของเว็บแคชจึงจะมุ่งเน้นการพัฒนาวิธีการแทนที่ข้อมูลในแคชโดยใช้

ค่าทางสถิติในการพิจารณาตัดสินใจ web object ออกจากแคชเพื่อบันทึก web object ใหม่เช่น ขั้นตอนวิธี Greedy-Dual-Size-Frequency (GDSF) [Ludmila, 1998] ซึ่งมีวิธีการพิจารณาค่าของเพจ  $p$  หนึ่งจากการคำนวณด้วยสมการ  $K(p) = L + F(p) * C(p) / S(p)$  โดย  $L$  คือปัจจัยช่วงชีวิต (aging factor) ของ web object ที่บันทึกในแคช  $F(p)$  คือความถี่ของการเข้าถึงเพจ  $p$  ที่เกิดขึ้นก่อนหน้า และ  $C(p)$  คือค่าใช้จ่ายที่ต้องเสียไปเมื่อดึงข้อมูลเพจ  $p$  และ  $S(p)$  คือขนาดของเพจ  $p$  เป็นต้น แต่การทำงานของเว็บแคชซึ่งก็ยังมีข้อจำกัดบางประการ จึงมีการนำเทคนิคการทำนายข้อมูลมาใช้เพื่อแก้ปัญหาและเพิ่มประสิทธิภาพการทำงานในแง่ของอัตราการพบข้อมูลที่ต้องการให้มีค่าสูงขึ้น

การทำเหมืองข้อมูลบันทึกการใช้งานเว็บเป็นทางเลือกหนึ่งในการปรับปรุงประสิทธิภาพการทำงานของเว็บแคชซึ่ง เนื่องจากการวิเคราะห์และทำนายความสัมพันธ์จากข้อมูลรูปแบบการใช้งานจริงซึ่งเก็บอยู่ในบันทึกการเข้าใช้งาน การทำเหมืองข้อมูลเว็บเป็นการนำเทคนิคการทำเหมืองข้อมูลมาใช้กับข้อมูลเว็บ โดยงานวิจัยของ Bonchi และคณะ [Bonchi, 2001] นำเสนอการประยุกต์ใช้เทคโนโลยีคลังข้อมูล (data warehouse) และการทำเหมืองข้อมูลกับเว็บแคชซึ่งเพื่อสร้างเว็บแคชอัจฉริยะ (intelligent web caching) ซึ่งเสนอวิธีการสร้างขั้นตอนวิธีสำหรับการทำนายรูปแบบของการร้องขอ แนวคิดพื้นฐานคือการขยายความสามารถของขั้นตอนวิธี Least Recently Used (LRU) โดยสร้างรูปแบบการเข้าถึงเว็บด้วยการใช้เทคนิคการทำเหมืองข้อมูลสองวิธีคือ association rule และ decision tree และมีการเตรียมข้อมูลโดยใช้โปรแกรม Data Transformation Services (DTS) ซึ่งเป็นเครื่องมือสำหรับจัดการกับข้อมูลในฐานข้อมูล ผลลัพธ์จากการทดลองสำหรับขั้นตอนใหม่ที่น่าเสนอในงานวิจัยนี้แสดงให้เห็นว่าค่า hit ratio เพิ่มขึ้นเมื่อเทียบกับการทำงานด้วยขั้นตอนวิธี LRU ปกติ การดำเนินการวิจัยนี้ต้องใช้โปรแกรมสำหรับจัดการข้อมูลจำนวนมากอาจทำให้เพิ่มปริมาณงานที่ต้องทำนอกเหนือจากการหาความสัมพันธ์จากข้อมูลจำนวนมากแล้วยังต้องใช้ทรัพยากรบางส่วนสำหรับโปรแกรม DTS เพื่อเตรียมข้อมูล และงานวิจัยของ Qing Yang และคณะ [Yang, 2001] นำเสนอวิธีการทำเหมืองข้อมูลเว็บสำหรับบันทึก web object และใช้ขั้นตอนวิธีที่น่าเสนอเพื่อเพิ่มประสิทธิภาพของระบบเว็บแคชซึ่งให้ดีขึ้น งานวิจัยนี้พัฒนาขั้นตอนวิธีการทำนายจาก n-gram โดย n-gram คือกลุ่มของคำร้องขอข้อมูลที่มีขนาด  $n$  สำหรับทำนายการร้องขอเว็บในอนาคตนำมาใช้เพิ่มความสามารถการทำงานของเว็บแคชซึ่งที่ใช้ขั้นตอนวิธี GDSF สำหรับแทนที่ข้อมูลในแคช การเตรียมข้อมูลของงานวิจัยนี้จะทำการแยก web object ต่างๆ ในเว็บเช่น รูปภาพ หรือข้อมูลเสียง เป็นต้น ออกจากเอกสาร HTML เนื่องจากพิจารณาว่า web object ข้างต้นเป็นข้อมูลที่ไม่ได้เกิดจากการร้องขอโดยตรงแต่จะมาพร้อมกับการร้องขอเว็บเพจจึงไม่นำ web object ดังกล่าวมาพิจารณาเพื่อวิเคราะห์ความสัมพันธ์ หลังจากคัดเหลือเฉพาะลำดับของการร้องขอเว็บเพจที่เป็นเอกสาร HTML แล้วกำหนดความยาวเพื่อสร้างเป็นกลุ่มของคำร้อง

ขนาด  $n$  หรือเรียกว่า  $n$ -gram ซึ่งพิจารณาความสัมพันธ์ในรูปแบบการทำเหมืองข้อมูลตามลำดับ แต่การแยกข้อมูลประเภทรูปภาพ หรือเสียงออกจากข้อมูลร้องขอเนื่องจากถือว่าเป็นข้อมูลที่ได้มาพร้อมกับการร้องขอเว็บในงานวิจัยนี้อาจข้ามประเด็นการร้องขอที่ผู้ใช้ตั้งใจร้องขอข้อมูลที่เป็นรูปภาพโดยตรงทำให้การวิเคราะห์คลาดเคลื่อนจากรูปแบบการใช้งานในความเป็นจริง

งานวิจัยที่จะนำเสนอในรายงานวิทยานิพนธ์นี้จะใช้เทคนิคการทำเหมืองข้อมูลเว็บ โดยใช้ข้อมูลบันทึกการใช้งานเว็บแคชซึ่งมีการบันทึกในเพิ่มข้อมูลธรรมดาไม่ต้องใช้โปรแกรมการจัดการเฉพาะแต่จะใช้โปรแกรม script ที่มีขนาดโปรแกรมเล็กกว่าช่วยในการเตรียมข้อมูล นอกจากนี้ยังพิจารณาความสัมพันธ์ของข้อมูลที่ร้องขอทั้งหมดไม่แยกพิจารณาเพื่อให้การวิเคราะห์ครอบคลุมรูปแบบการร้องขอ

### 1.3 วัตถุประสงค์ของงานวิจัย

1. เพื่อศึกษาหาวิธีการในการเพิ่มประสิทธิภาพการทำงานของเว็บแคช
2. เพื่อสร้างแม่แบบการทำงานเว็บแคชซึ่งที่เพิ่มประสิทธิภาพการทำงาน โดยใช้การทำนายความสัมพันธ์ของการร้องขอข้อมูลจากผู้ใช้

### 1.4 ขอบเขตการดำเนินการ

1. พัฒนาแม่แบบการทำงานเว็บแคชซึ่งที่สามารถทำงานขั้นพื้นฐานและทำหน้าที่เป็นตัวกลางที่ใช้ติดต่อกับเครือข่ายอื่นได้
2. ทำนายการใช้งานเว็บในอนาคตของผู้ใช้จากพฤติกรรมของผู้ใช้เพื่อให้ได้ถูกความเชื่อมโยงที่มีความถูกต้องแม่นยำที่สุดเท่าที่จะเป็นไปได้
3. บันทึกข้อมูลที่ร้องขอตามข้อมูลทำนายในแคชล่วงหน้าและพิจารณาเลือกคัดข้อมูลออกจากแคชเพื่อให้การทำงานของเว็บแคชซึ่งมีประสิทธิภาพ โดยให้มีค่า hit ratio และ byte hit ratio สูงขึ้น

### 1.5 ขั้นตอนการดำเนินงานวิจัย

1. ศึกษาและค้นคว้าเอกสารที่เกี่ยวข้องกับงานวิทยานิพนธ์
2. รวบรวมแนวคิด และเลือกเทคโนโลยีสำหรับการพัฒนาระบบและทดลองการทำงาน
3. ออกแบบสถาปัตยกรรมแม่แบบเว็บแคชซึ่งที่ทำงานร่วมกับการทำเหมืองข้อมูลเว็บจากแนวคิดที่ทำการศึกษา
4. พัฒนาแม่แบบเพื่อนำเสนอแนวคิดโดยปรับปรุงโปรแกรมที่เป็นเครื่องมือให้ทำงานรองรับกันได้
5. ทดสอบการทำงานของแม่แบบ
6. สรุปผลการทดสอบ และจัดทำเอกสารประกอบการวิจัย

### 1.6 ระยะเวลาการดำเนินงาน

เมษายน พ.ศ. 2548 – มีนาคม พ.ศ. 2549

แผนการดำเนินงาน

ขั้นตอน	เม.ย. 48	พ.ค. 48	มิ.ย. 48	ก.ค. 48	ส.ค. 48	ก.ย. 48	ต.ค. 48	พ.ย. 48	ธ.ค. 48	ม.ค. 49	ก.พ. 49	มี.ค. 49
1.	←	→										
2.		←	→									
3.			←	→								
4.					←	→						
5.							←	→				
6.									←	→		

## 1.7 เครื่องมือและอุปกรณ์ที่ใช้ในงานวิจัย

### ฮาร์ดแวร์

เครื่อง ไมโครคอมพิวเตอร์ 2 เครื่อง

- หน่วยประมวลผลกลาง Pentium 4 และ Pentium II
- ระบบปฏิบัติการ Linux Red Hat 9.0
- ระบบปฏิบัติการ Microsoft Windows 2000 Professional
- หน่วยความจำอย่างน้อย 128 MB
- ฮาร์ดดิสก์อย่างน้อย 60 GB

### ซอฟต์แวร์

- ตัวแปลภาษา Java Development Kit (JDK) เวอร์ชัน 1.4.2 หรือสูงกว่า บนระบบปฏิบัติการ Unix
- ตัวแปลภาษา Perl บนระบบปฏิบัติการ Unix
- โปรแกรมเว็บแคชชิง
- โปรแกรมการทำเหมืองข้อมูลเว็บ
- โปรแกรมร้องขอข้อมูลเว็บ

## 1.8 ประโยชน์ที่คาดว่าจะได้รับ

ได้แม่แบบเว็บแคชชิงที่ใช้ความรู้ในการทำนายรูปแบบการใช้งานของผู้ใช้จากบันทึกการเข้าใช้งานเว็บเพื่อประสิทธิภาพการทำงานของเว็บแคชชิง

## 1.9 บทสรุป

เว็บแคชชิงเป็นเทคโนโลยีหนึ่งซึ่งช่วยเพิ่มประสิทธิภาพการทำงานของเว็บแต่เนื่องจากเว็บแคชชิงมีข้อจำกัดในเรื่องพื้นที่การจัดเก็บข้อมูล อันเป็นปัจจัยประการหนึ่งที่ส่งผลต่อประสิทธิภาพการทำงาน จึงมีความพยายามที่จะปรับปรุงประสิทธิภาพโดยการใช้ขั้นตอนวิธีการแทนที่เพื่อจัดการข้อมูลในแคชให้สามารถใช้พื้นที่จำกัดได้อย่างดี อย่างไรก็ตามขั้นตอนการแทนที่ยังไม่สามารถเพิ่มประสิทธิภาพของเว็บแคชชิงได้มากนัก เทคนิคการดึงข้อมูลล่วงหน้าเป็นเทคนิคหนึ่งที่น่าสนใจเพื่อขยายขีดความสามารถของเว็บแคชชิง การดึงข้อมูลล่วงหน้ามีหลักการทำงานคือ

ทำนายข้อมูลที่จะมีการร้องขอในอนาคตแล้วทำการร้องขอข้อมูลนั้นๆล่วงหน้า ในงานวิทยานิพนธ์นี้จึงได้เสนอให้มีการนำเทคนิคการทำเหมืองข้อมูลเว็บมาช่วยในการทำนายคำร้องขอของผู้ใช้จากข้อมูลบันทึกการใช้งานเว็บที่เป็นเสมือนตัวแทนของผู้ใช้ เพื่อให้คำร้องขอที่ทำนายใกล้เคียงกับความเป็นจริงมากที่สุด โดยรายละเอียดของทฤษฎี และแนวคิดที่ใช้ในงานวิจัยจะนำเสนอในบทที่ 2 สำหรับในบทที่ 3 นำเสนอการออกแบบสถาปัตยกรรมของแม่แบบการทำงานซึ่งแสดงภาพรวมของสถาปัตยกรรมและรายละเอียดของการทำงานในแต่ละส่วนการทำงาน ในบทที่ 4 นำเสนอการพัฒนาแม่แบบตามที่ได้ออกแบบ สำหรับบทที่ 5 นำเสนอผลการทดสอบการทำงาน และบทที่ 6 นำเสนอผลสรุปผลการวิจัยพร้อมทั้งข้อเสนอแนะและงานวิจัยในอนาคต