

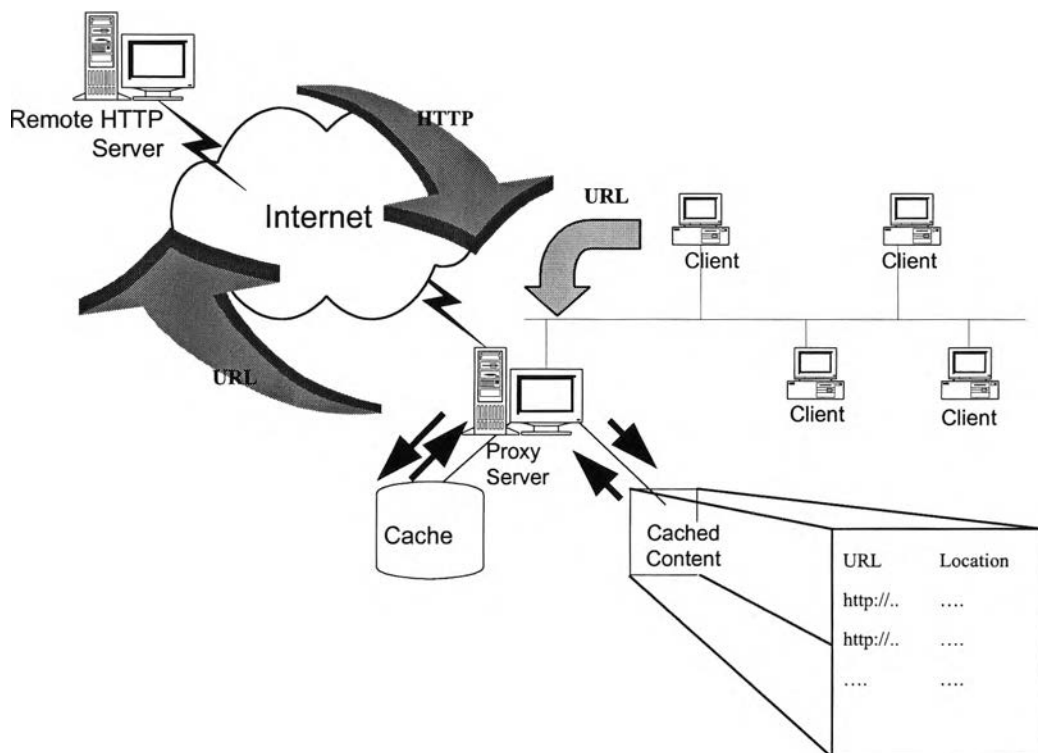
# บทที่ 1

## บทนำ



### 1.1 ความเป็นมา

ปัจจุบันมีความนิยมใช้งานอินเทอร์เน็ตเป็นเพิ่มขึ้นอย่างมากทำให้องค์กรต่างๆ ประสบปัญหาแบนด์วิดท์ไม่เพียงพอ การเพิ่มแบนด์วิดท์ต้องเสียค่าใช้จ่ายสูงและไม่สามารถแก้ปัญหาได้อย่างแท้จริง ดังนั้นการแก้ปัญหาดังกล่าวจึงเน้นไปที่การเพิ่มประสิทธิภาพในการใช้งานแบนด์วิดท์ที่มีอยู่ให้ได้ ประสิทธิภาพสูงสุด การติดตั้งเว็บพร็อกซีหรือเว็บแคช (Web Proxy/Web Cache) เป็นวิธีการหนึ่งที่จะเพิ่มประสิทธิภาพการใช้งานแบนด์วิดท์ให้มีประสิทธิภาพสูงขึ้น เว็บพร็อกซีจะทำหน้าที่ไปนำข้อมูลจากเซิร์ฟเวอร์ที่เครื่องคอมพิวเตอร์ลูกข่ายที่ใช้บริการเว็บเรียกขอ แทนที่เครื่องคอมพิวเตอร์ลูกข่ายแต่ละเครื่องจะเรียกขอข้อมูลไปยังเครื่องคอมพิวเตอร์เซิร์ฟเวอร์ผู้ให้บริการข้อมูลนั่นเอง และจะเก็บสำเนาข้อมูลเหล่านั้นไว้ระยะหนึ่ง (ซึ่งเรียกว่า “การแคชข้อมูล”) เมื่อมีเครื่องลูกข่ายเครื่องอื่นๆ เรียกขอข้อมูลเดียวกัน เว็บพร็อกซีก็จะส่งข้อมูลที่เก็บไว้ไปให้เครื่องลูกข่ายนั้นได้ทันที โดยไม่จำเป็นต้องขอข้อมูลจากเครื่องคอมพิวเตอร์ที่ให้บริการข้อมูลนั้นใหม่ [1] ช่วยลดปริมาณการใช้แบนด์วิดท์และลดเวลาตอบสนองที่ปรากฏต่อเครื่องลูกข่ายได้ การทำงานของเว็บพร็อกซีแสดงในรูปที่ 1.1



รูปที่ 1.1 แสดงการทำงานของเว็บพร็อกซี

การทำงานเครื่องคอมพิวเตอร์ลูกข่าย โดยโปรแกรมเว็บเบราว์เซอร์ (Web Browser) เช่น Internet Explorer หรือ Netscape Communicator จะเรียกขอข้อมูลในรูปแบบ http หรือ ftp โดยระบุ ยูอาร์แอลหรือยูอาร์ไอ (URL or URI) ไปยังเครื่องคอมพิวเตอร์ที่ทำหน้าที่เป็นพร็อกซีเซิร์ฟเวอร์หรือเว็บพร็อกซี เว็บพร็อกซีจะตรวจสอบว่าภายในแคชของตนมียูอาร์แอลดังที่เครื่องลูกข่ายต้องการหรือไม่ ถ้าหากว่ามีก็จะส่งข้อมูลไปยังเครื่องลูกข่ายทันที แต่ถ้าหากว่าไม่มีข้อมูล กรณีที่เป็นเว็บพร็อกซีเดี่ยวๆ ก็จะส่งการเรียกขอไปบนอินเทอร์เน็ต เพื่อขอข้อมูลไปยังเซิร์ฟเวอร์ที่ให้บริการตามยูอาร์แอลนั้น กรณีที่เป็นกลุ่มของเว็บพร็อกซีก็จะมีการสอบถามภายในกลุ่มเว็บพร็อกซีด้วยกันก่อนว่ามีข้อมูลหรือไม่ ถ้าไม่มีก็จะส่งการเรียกขอไปยังอินเทอร์เน็ตเช่นเดียวกันกับกรณีแรก เมื่อได้รับข้อมูลกลับมาจากเซิร์ฟเวอร์ที่ให้บริการข้อมูลตามยูอาร์แอลที่ขอไปแล้ว เว็บพร็อกซีก็จะส่งข้อมูลเหล่านั้นต่อไปให้กับเครื่องลูกข่ายที่ขอข้อมูลมา รวมทั้งเก็บข้อมูลที่รับมาใหม่นี้ไว้ในแคช จากนั้นก็จะรวบรวมยูอาร์แอลจากการเรียกขอจากเครื่องลูกข่ายต่อไป

## 1.2 ความสำคัญของยูอาร์แอล

เมื่อเครื่องคอมพิวเตอร์ลูกข่ายที่ใช้บริการเว็บต้องการข้อมูลจะส่งการเรียกขอข้อมูลไปยังเครื่องคอมพิวเตอร์เซิร์ฟเวอร์ที่ให้บริการข้อมูลนั้นว่าต้องการข้อมูลอะไร สิ่งที่ใช้ในการระบุถึงข้อมูลก็คือ ยูอาร์แอล (URL:Universal Resource Locator) และ ยูอาร์ไอ (URI:Universal Resource Identifier) ยูอาร์แอล (URL) เป็นสิ่งที่บอกตำแหน่งของข้อมูลว่าอยู่ที่เซิร์ฟเวอร์ใดบนอินเทอร์เน็ต ส่วน ยูอาร์ไอ (URI) จะระบุถึงตำแหน่งข้อมูลที่ต้องการบนเซิร์ฟเวอร์นั้น [2] ยูอาร์แอลมีบทบาทที่สำคัญสำหรับเว็บพร็อกซี เนื่องจากเว็บพร็อกซีที่มีการแคชข้อมูลจะต้องเก็บยูอาร์แอลของข้อมูลที่แคชเอาไว้ เมื่อเครื่องคอมพิวเตอร์ลูกข่ายขอข้อมูล โดยการระบุยูอาร์แอลหรือยูอาร์ไอของข้อมูลที่ต้องการไปยังเว็บพร็อกซี เว็บพร็อกซีจะตรวจสอบว่ามีข้อมูลตามยูอาร์แอลดังกล่าวอยู่ในแคชหรือไม่ ถ้าหากมีก็จะส่งข้อมูลไปยังเครื่องคอมพิวเตอร์ลูกข่ายที่เรียกขอข้อมูลมาทันที แต่ถ้าหากไม่มีเว็บพร็อกซีก็จะขอไปยังเครื่องคอมพิวเตอร์เซิร์ฟเวอร์ที่ให้บริการข้อมูลตามยูอาร์แอลนั้น และนำข้อมูลที่ส่งไปยังเครื่องคอมพิวเตอร์ลูกข่ายที่เรียกขอข้อมูลมาต่อไป

ในองค์กรขนาดใหญ่อาจมีการติดตั้งเว็บพร็อกซีมากกว่า 1 ตัว เว็บพร็อกซีเหล่านี้จะทำงานร่วมกัน กล่าวคือเมื่อเครื่องคอมพิวเตอร์ลูกข่ายของเว็บพร็อกซีตัวใดตัวหนึ่งเรียกขอข้อมูลและเว็บพร็อกซีไม่มีข้อมูล เว็บพร็อกซีนั้นจะสอบถามเว็บพร็อกซีอื่นๆ ที่ทำงานร่วมกันก่อนว่ามีข้อมูลหรือไม่ ก่อนที่จะไปเรียกขอข้อมูลนั้นจากเซิร์ฟเวอร์ที่ให้บริการข้อมูล ถ้าหากข้อมูลที่เครื่องคอมพิวเตอร์ลูกข่ายเรียกขอนั้นอยู่ในกลุ่มของเว็บพร็อกซีที่ทำงานร่วมกันก็จะได้รับข้อมูลที่เร็วขึ้น การสอบถามข้อมูล

---

\* ในเอกสารนี้จะใช้ ยูอาร์แอล(URL) ในความหมายที่รวมทั้งยูอาร์ไอ(URI) และ ยูอาร์แอล(URL)

ระหว่างเว็บพริกซ์ก็จะสอบถามโดยใช้ยูอาร์แอล [3] ด้วยเช่นกัน การทำงานร่วมกันของเว็บพริกซ์อีกลักษณะหนึ่งคือ แต่ละเว็บพริกซ์จะกระจายยูอาร์แอลของข้อมูลของตัวเองแคชไว้ให้กับเว็บพริกซ์อื่นๆ ทราบ เว็บพริกซ์แต่ละตัวก็จะสามารถคาดการณ์ได้ว่า ข้อมูลที่ต้องการควรจะอยู่ที่เว็บพริกซ์ตัวใด

นอกจากยูอาร์แอลจะใช้ในการเรียกขอและสอบถามถึงข้อมูลที่คอมพิวเตอร์ลูกข่ายต้องการแล้ว ในการวิเคราะห์การทำงานเพื่อปรับแต่งและศึกษาคุณลักษณะของเว็บพริกซ์ รวมไปถึงการคำนวณค่าใช้จ่าย จะต้องอาศัยข้อมูลที่เว็บพริกซ์เก็บในระหว่างการทำงานหรือ ล็อกไฟล์ (log files) เช่น การวิเคราะห์อัตราการพบข้อมูลที่เครื่องลูกข่ายเรียกขอในแคช (hit ratio) เป็นต้น ข้อมูลในล็อกไฟล์ส่วนหนึ่งจะเก็บยูอาร์แอลของข้อมูลที่เครื่องคอมพิวเตอร์ลูกข่ายเรียกขอในแต่ละครั้งเอาไว้ การวิเคราะห์ก็จะใช้ยูอาร์แอลนี้ในการคำนวณเพื่อหาอัตราการพบข้อมูลที่เครื่องคอมพิวเตอร์ลูกข่ายขอในแคช หรือใช้ยูอาร์แอลในการคำนวณค่าใช้จ่าย นั่นเอง

จากการทำงานของเว็บพริกซ์จะเห็นได้ว่ายูอาร์แอลมีความสัมพันธ์กับประสิทธิภาพโดยรวมในการทำงานของพริกซ์ เนื่องจากในการทำงานของเว็บพริกซ์ต้องมีการประมวลผลกับยูอาร์แอลจำนวนมากบ่อยๆ ดังนั้นประสิทธิภาพในการประมวลผลยูอาร์แอลจึงสะท้อนประสิทธิภาพการทำงานของเว็บพริกซ์ได้

### 1.3 ปัญหาในการวิเคราะห์ยูอาร์แอล

ยูอาร์แอลเป็นสตริงตัวอักษรที่มีความยาวไม่แน่นอน และไม่ได้มีการกำหนดความยาวสูงสุดของยูอาร์แอลไว้ [2] นอกจากนี้ยูอาร์แอลยังมีลักษณะเฉพาะตัว เช่น มักขึ้นต้นด้วย "http://www..." หรือ "ftp://ftp..." ยูอาร์แอลของเว็บไซต์ที่มีความนิยมสูงก็จะมีโอกาสพบมาก ลักษณะเฉพาะเหล่านี้มีผลต่อการประมวลผลยูอาร์แอลซึ่งส่งผลกระทบต่อประสิทธิภาพการทำงานของเว็บพริกซ์ดังนี้

- เว็บพริกซ์จะต้องทำการเปรียบเทียบยูอาร์แอลเพื่อตรวจสอบว่ามีข้อมูลในแคชหรือไม่ [4] ซึ่งถ้าหากยูอาร์แอลมีความยาวมากๆ ย่อมทำให้เสียเวลาในการประมวลผลและเสียทรัพยากรระบบเป็นจำนวนมาก
- โดยปกติแล้วเว็บพริกซ์จะเก็บข้อมูลที่แคชเอาไว้ในฮาร์ดดิสก์ที่มีขนาดใหญ่หลายๆ ดังนั้นจะต้องใช้เนื้อที่ฮาร์ดดิสก์จำนวนมากในการเก็บยูอาร์แอลของสิ่งที่อยู่ในแคชด้วย
- การวิเคราะห์ไฟล์ที่พริกซ์เขียนในระหว่างการทำงานหรือล็อกไฟล์ (log file) เช่น การหาอัตราการพบข้อมูลในแคชจะเสียเวลานาน ไม่มีประสิทธิภาพ และสิ้นเปลืองทรัพยากรของระบบ
- ในการสอบถามข้อมูลระหว่างเว็บพริกซ์ โดยส่งยูอาร์แอลระหว่างกัน จะเกิดข้อมูลไหลในเครือข่ายเป็นจำนวนมากทำให้ประสิทธิภาพของเครือข่ายลดลง

- การตรวจสอบยูอาร์แอลของเว็บพริ็อกที่ถูกจำกัดด้วยเงื่อนไขของเวลา ถ้าหากการค้นหาเปรียบเทียบยูอาร์แอลและการดึงข้อมูลจากแคชช้ากว่าการไปนำข้อมูลจากเซิร์ฟเวอร์โดยตรงแล้ว การประมวลผลนั้นก็ไร้ประโยชน์

จากปัญหาที่เกิดขึ้นกับการเก็บ เปรียบเทียบ และประมวลผลยูอาร์แอลตามที่กล่าวมา จึงมีความพยายามในการเพิ่มประสิทธิภาพในการประมวลผลยูอาร์แอลในเว็บพริ็อกซี เช่น สควิด(Squid) [4] จึงมีการย่อขนาดหรือเข้ารหัสยูอาร์แอล [5] เพื่อให้ได้ผลลัพธ์เป็นสตริงที่สั้นลงโดยใช้ MD5[6] ซึ่งเป็นกรรมวิธีในการเข้ารหัสที่เรียกว่าการแฮช (Hashing) แบบหนึ่ง เพื่อเพิ่มประสิทธิภาพในการประมวลผลยูอาร์แอล

Summary Cache [18] ซึ่งเป็นงานวิจัยที่เสนอโปรโตคอลที่ลดโอเวอร์เฮด (Overhead) เพื่อใช้ในการสื่อสารระหว่างพริ็อกซีที่ทำงานร่วมกันแทนโปรโตคอล ICP (Internet Cache Protocol) ที่ใช้ทำงานกันอยู่

นอกจากนั้นยังมีการใช้ฟังก์ชันแฮชเพื่อเข้ารหัสยูอาร์แอลในการเลือกที่จะใช้เว็บพริ็อกซีตัวใดในแคชอาเรย์ (Cache Array Routing Protocol – CARP) [20] ซึ่งคุณสมบัติของฟังก์ชันแฮชที่เลือกใช้ นอกจากจะส่งต่อประสิทธิภาพโดยรวมของแคชอาเรย์แล้ว ยังส่งผลต่อการกระจายภาระ (load balancing) ให้กับเว็บพริ็อกซีแต่ละตัวในแคชอาเรย์นั้นด้วย

อย่างไรก็ดี ในการประมวลผลยูอาร์แอลของเว็บพริ็อกซีนั้นมีความต้องการที่แตกต่างกันขึ้นอยู่กับแอปพลิเคชัน ตัวอย่างเช่น การสอบถามหรือกระจายข้อมูลระหว่างเว็บพริ็อกซีที่ทำงานร่วมกันนั้นสามารถระบุยูอาร์แอลของข้อมูลที่มีในแคชคร่าวๆ โดยไม่จำเป็นต้องถูกต้อง 100% ก็เพียงพอ แต่ปริมาณข้อมูลที่ส่งระหว่างกันนั้นจะต้องน้อยและสามารถระบุได้อย่างรวดเร็วว่าในแคชมีข้อมูลที่ต้องการหรือไม่ ในกรณีเช่นนี้ต้องเลือกวิธีการเข้ารหัสยูอาร์แอลที่ทำงานได้เร็วและรหัสที่ได้มีขนาดสั้น โดยอาจมีการชนกันของรหัสบ้าง [7]

อีกตัวอย่างหนึ่ง คือ การวิเคราะห์ล็อกไฟล์ของเว็บพริ็อกซี ซึ่งต้องการวิธีการเข้ารหัสที่ให้รหัสที่สั้นเพื่อประหยัดทรัพยากรระบบในการประมวลผลและจะต้องไม่เกิดการชนกันของรหัสเพื่อความถูกต้องในการวิเคราะห์ผล แต่ความเร็วในการเข้ารหัสยูอาร์แอลไม่ใช่สิ่งสำคัญ ดังนั้นควรเลือกวิธีในการเข้ารหัสที่มีรหัสสั้นและต้องไม่เกิดการชนหรือซ้ำกันของรหัสเลย เป็นต้น

## 1.4 แนวคิดในการวิจัย

จากการพิจารณาและตัวอย่างแอปพลิเคชันเว็บพริคซ์ที่กล่าวมา จะพบว่าสิ่งที่ต้องคำนึงในการเลือกวิธีการเข้ารหัสที่มีผลต่อประสิทธิภาพของเว็บพริคซ์มีดังต่อไปนี้

- **ความเร็วในการเข้ารหัส**

การตรวจสอบข้อมูลว่าอยู่ในแคชหรือไม่ ถ้าหากเวลาที่ใช้ในการค้นหายูอาร์แอลของข้อมูลที่ต้องการว่าอยู่ในแคชหรือไม่นั้น มากกว่าเวลาที่ใช้ในการไปนำข้อมูลมาจากเซิร์ฟเวอร์ที่ให้บริการข้อมูลนั้นแล้ว การค้นหายูอาร์แอลนั้นก็ไร้ประโยชน์ เป็นที่สังเกตว่าในการค้นหาไม่ต้องการความถูกต้อง 100% เนื่องจากเมื่อมีการตรวจสอบยูอาร์แอลว่ามีข้อมูลอยู่ในแคชแล้วก็ตาม แต่เมื่อมีการเข้าถึงข้อมูลเพื่อดึงมาใช้งานจริง ปรากฏว่าข้อมูลนั้นอาจถูกแทนที่ไปแล้วก็ได้ ดังนั้นความเร็วในการค้นหาข้อมูลจึงเป็นสิ่งสำคัญที่สุด

- **ขนาดของรหัสที่ได้**

จากแนวโน้มราคาฮาร์ดดิสก์ที่ลดลงเรื่อยๆ ส่งผลให้เว็บพริคซ์มีการใช้ฮาร์ดดิสก์ที่มีขนาดใหญ่มากขึ้นในการแคชข้อมูล ดัชนีที่ใช้ในการระบุตำแหน่งข้อมูลในแคชจึงมีขนาดใหญ่ขึ้นตามไปด้วย การจัดการกับดัชนีเหล่านั้นต้องใช้ทรัพยากรมาก ดังนั้นดัชนีที่มีขนาดเล็กเท่าใดก็ทำให้จัดการแคชได้มีประสิทธิภาพมากขึ้นเท่านั้น ในกรณีเช่นนี้ความยาวของรหัสที่ได้จากยูอาร์แอลจึงเป็นสิ่งสำคัญ

- **การชนกันของรหัส**

การชนกันของข้อมูลหมายถึงการซ้ำกันของรหัสที่ได้จากการเข้ารหัส ในการวิเคราะห์อัตราการพบข้อมูลในแคชจะมีการเปรียบเทียบยูอาร์แอลที่ต้องการความถูกต้อง 100% เพื่อให้สามารถคำนวณอัตราดังกล่าวได้ถูกต้องตามความเป็นจริง ในกรณีเช่นนี้การเข้ารหัสยูอาร์แอลที่เกิดการชนกันของรหัสจึงไม่สามารถใช้ได้

จากปัจจัยต่างๆ ดังที่กล่าวมานั้น การพิจารณาว่าวิธีการเข้ารหัสที่เหมาะสมในการเข้ารหัสยูอาร์แอล ควรจะมีคุณสมบัติดังนี้ คือ

1. ทำงานได้รวดเร็ว
2. ขนาดของรหัสที่ได้ต้องสั้น (อย่างน้อยต้องสั้นกว่ายูอาร์แอลก่อนเข้ารหัส)
3. เกิดการชนกันของข้อมูลที่ได้จากฟังก์ชันแฮชน้อยหรือไม่เกิดเลย
4. เป็นวิธีการที่เป็นที่รู้จักกันดี สามารถทำความเข้าใจและเขียนโปรแกรมได้ง่าย

ดังที่กล่าวไปแล้วว่า สควิด (Squid) ซึ่งเป็นเว็บพร็อกซีที่ใช้งานโดยทั่วไป ใช้วิธีการ MD5 ซึ่งเป็นที่รู้จักกันดีในการเข้ารหัสเพื่อย่อขนาดยูอาร์แอล แต่เป็นที่น่าสังเกตว่า MD5 ถูกออกแบบมาเพื่อสร้างลายเซ็นอิเล็กทรอนิกส์ (Electronic Fingerprint) ซึ่งมีความปลอดภัยสูง แต่อาจจะทำงานได้ช้า และโดยที่ไม่มีงานวิจัยยืนยันว่า MD5 นั้นเหมาะสมที่จะนำมาใช้เข้ารหัสยูอาร์แอลในทุกกรณี ส่วนฟังก์ชันแฮชแบบง่าย ๆ เช่น การหารด้วย 256 (Division Method) สามารถทำงานได้เร็วแต่เกิดการชนของข้อมูลสูง แต่อย่างไรก็ดีก็อาจมีความเหมาะสมกับงานบางงานดังที่กล่าวมาแล้ว

อนึ่งมีงานวิจัยหลายงานได้ทำการศึกษาถึงความเหมาะสมในการเข้ารหัสข้อมูลเพื่อใช้งานต่างๆ แต่สำหรับการเข้ารหัสยูอาร์แอลนั้น ยังไม่มีงานวิจัยที่ศึกษาถึงความเหมาะสมในการเข้ารหัสยูอาร์แอล ทั้งยังเห็นได้ว่าการเข้ารหัสแต่ละแบบมีคุณลักษณะที่เป็นข้อดีและข้อด้อยที่แตกต่างกันออกไป ผู้วิจัยจึงมีแนวความคิดในการศึกษาถึงข้อดีและข้อเสียของวิธีการหรืออัลกอริทึมในการเข้ารหัสต่างๆ เพื่อเข้ารหัสยูอาร์แอล โดยนำวิธีการเข้ารหัสต่างๆ มาทดสอบเข้ารหัสยูอาร์แอล ซึ่งนำมาจากล็อกไฟล์ในรูปแบบของสควิด (Squid format) และนำผลที่ได้มาวิเคราะห์ในแง่ต่างๆ ดังต่อไปนี้ คือ

- เวลาที่ใช้ในการเข้ารหัส
- ขนาดของรหัสที่ได้
- จำนวนครั้งที่เกิดการชนของรหัส

จากนั้นจะเปรียบเทียบและวิเคราะห์ผล เพื่อนำผลการวิเคราะห์ที่ได้ใช้เป็นแนวทางในการเลือกวิธีการเข้ารหัสหรือฟังก์ชันแฮชที่มีความเหมาะสมสำหรับการเข้ารหัสยูอาร์แอลต่อไป

## 1.5 งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้องกับการนำฟังก์ชันแฮชมาใช้กับแอปพลิเคชันบนเว็บโดยเฉพาะอย่างยิ่งกับเว็บพร็อกซี ยกมาพอสังเขป ดังนี้

1.5.1 Signature Indexing การศึกษาวิจัยเกี่ยวกับการใช้ Signature file หรือ ไฟล์ที่เก็บข้อมูลที่เป็น message digest เพื่อใช้ในฐานะข้อมูลที่เป็น ตัวอักษร [15][16][17] และฐานข้อมูลชนิดอื่นๆ รวมทั้งมีการวิจัยถึงความเหมาะสมในการใช้เพื่อบีบอัดข้อมูลที่มีรูปแบบต่างๆ เช่น ข้อมูลภาพ และข้อมูลเสียง คำว่า Signature ก็คือการสร้างลายเซ็นอิเล็กทรอนิกส์ ซึ่งในการสร้างจะใช้ฟังก์ชันแฮช (กลุ่ม MD หรือ Message Digest)

1.5.2 Suitability of Signature Indexing Over the World Wide Web [7] ความเหมาะสมในการใช้การทำดัชนีแบบ Signature บนเครือข่ายเว็ลด์ไวด์เว็บ เป็นการศึกษาเพื่อเปรียบเทียบคุณลักษณะของการทำดัชนีแบบ Signature ที่มีความจำเป็นใช้ในการเข้าถึงดัชนีบนเว็บและความเหมาะสมของการใช้งานบนเว็บ โดยการนำ Signature ของข้อมูลก็

คือการใช้ฟังก์ชันแฮชในการเข้ารหัสแบบหนึ่งนั่นเอง นอกจากนี้ยังมีการเสนอให้ใช้ Signature ในการค้นหา เพิ่มข้อมูลใหม่ และการแทนที่ข้อมูล ในแคชด้วย [18][19]

1.5.3 Summary Cache งานวิจัยที่ศึกษาการใช้งานพรีอิกซีแคชร่วมกันคือ Summary Cache [18] ในงานวิจัยนี้มีการนำเสนอโปรโตคอลใหม่ที่มีการใช้การย่อข้อมูลของข้อมูลในแคชของแต่ละพรีอิกซี โดยใช้วิธีการแทนเซตข้อมูลวิธีการหนึ่งชื่อ Bloom Filter และมีการใช้งานฟังก์ชันแฮช MD5 (MD 5 Hash Functions) เพื่อใช้เป็นดัชนีของยูอาร์แอลและใช้ข้อมูลที่ย่อนี้ในการสื่อสารระหว่างเว็บพรีอิกซี เพื่อลดโอเวอร์เฮดของการทำงานให้ดีกว่า ICP [3]

1.5.4 Super Proxy [20] เป็นงานวิจัยที่เกี่ยวข้องกับการใช้ฟังก์ชันแฮชกับยูอาร์แอล และนำค่าแฮชมาใช้ในการตัดสินใจในการเลือกหรือขอข้อมูลตามยูอาร์แอลนั้น งานวิจัยได้ศึกษาถึงการนำค่าแฮชของยูอาร์แอลในการทำงานร่วมกันระหว่างเว็บพรีอิกซี โดยไม่มีการเสนอโปรโตคอลใหม่แต่อย่างใด

## 1.6 ขอบเขตการวิจัย

ในการทำวิจัย ผู้วิจัยได้เลือกศึกษาวิธีการหรืออัลกอริทึมในการเข้ารหัสดังต่อไปนี้ คือ

### 1. กลุ่มวิธี Cyclic Redundancy Check

- 1.1 CRC-16
- 1.2 CRC-CCITT
- 1.3 CRC-32

### 2. การเข้ารหัส (Coding)

- 2.1 Huffman Coding

### 3. กลุ่มวิธีการแฮช (Hashing)

#### กลุ่ม MD

- 3.1 MD2
- 3.2 MD4
- 3.3 MD5
- 3.4 SHA-1

#### กลุ่มฟังก์ชันแฮชอย่างง่าย

- 3.5 Division method
- 3.6 Midsquare method
- 3.7 Folding method

### 3.8 Digit Analysis method

ในการวิจัยจะวิเคราะห์หาคุณลักษณะของวิธีการเข้ารหัสแต่ละตัว ที่มีผลต่อการเข้ารหัสยูอาร์แอลในด้านต่างๆ ดังนี้

- ความเร็วในการทำงาน
- ขนาดของผลลัพธ์ที่ได้จากการแฮช หรือการเข้ารหัส
- จำนวนครั้งที่เกิดการชนกันของข้อมูล

โดยจะทำการทดลองเข้ารหัสยูอาร์แอลซึ่งเก็บมาจากล็อกไฟล์ของเว็บพริคซีต่างๆ ดังนี้

- ข้อมูลการใช้เว็บของสำนักเทคโนโลยีสารสนเทศ จุฬาลงกรณ์มหาวิทยาลัย

### 1.7 วัตถุประสงค์

เพื่อศึกษาเปรียบเทียบวิธีการหรืออัลกอริทึมในการเข้ารหัสยูอาร์แอลเพื่อเสนอแนะข้อดีและข้อเสียของแต่ละวิธี

### 1.8 ประโยชน์ที่จะได้รับ

งานวิจัยจะนำเสนอข้อดี ข้อเสียของการเข้ารหัสแบบต่างๆ เพื่อใช้เป็นข้อเสนอแนะและพิจารณาในการเลือกวิธีการเข้ารหัสในการเข้ารหัสยูอาร์แอลที่มีความเหมาะสมต่อแอปพลิเคชันเว็บพริคซี

### 1.9 โครงสร้างวิทยานิพนธ์

วิทยานิพนธ์ฉบับนี้จัดแบ่งออกเป็น 6 บท ในแต่ละบทมีเนื้อหาดังต่อไปนี้

**บทที่ 1** จะเป็นกล่าวถึงความเป็นมาของปัญหา และความน่าสนใจของปัญหาที่ก่อให้เกิดงานวิจัยนี้ขึ้น จากนั้นจะกล่าวถึงงานวิจัยที่มีเนื้อหาเกี่ยวข้องกับวิทยานิพนธ์ฉบับนี้ ถัดจากส่วนของงานวิจัยที่เกี่ยวข้อง จะเป็นขอบเขตของงานวิจัยว่ามีวิธีการหรืออัลกอริทึมใดที่จะนำมาทดสอบ ใช้ข้อมูลในการทดสอบจากที่ใด และผลการวิจัยครอบคลุมถึงเรื่องใดบ้าง ทำการวิเคราะห์ผลการทดลองโดยใช้สิ่งใดเป็นเกณฑ์วัด

ถัดจากหัวข้อในส่วนของขอบเขตการวิจัยแล้ว จะเป็นวัตถุประสงค์ของงานวิจัยนี้ และสุดท้ายจะเป็นประโยชน์ที่คาดว่าจะได้รับจากผลการวิจัย



- บทที่ 2** จะเป็นการกล่าวถึงแนวความคิดที่ใช้และทฤษฎีที่เกี่ยวข้อง ซึ่งจะเป็นการอธิบายถึงอัลกอริทึมที่นำมาทดสอบโดยสังเขป มีการอ้างอิงถึงทฤษฎีหรือสูตรคณิตศาสตร์ที่ใช้ในอัลกอริทึมเพื่อช่วยในการทำความเข้าใจ
- บทที่ 3** จะเป็นการกล่าวถึงแนวทางการทำวิจัย ซึ่งจะเป็นการอธิบายถึงขั้นตอนในการทำวิจัย วิธีการพิจารณาเลือกข้อมูล การเลือกภาษาโปรแกรม จำนวนครั้งที่ทำการทดสอบซ้ำ และผลการทดลองที่จะได้รับ และความน่าเชื่อถือของการทดลอง
- บทที่ 4** จะเป็นการนำเสนอผลการทดลอง และข้อมูลต่างๆที่ได้จากการทดลอง
- บทที่ 5** กล่าวถึงการวิเคราะห์ผลการทดลอง โดยนำผลการทดลองจากบทที่ 4 มาเปรียบเทียบและใช้เหตุผล เพื่อให้เห็นภาพของผลการทดลองในมุมมองต่างๆได้ชัดเจนยิ่งขึ้น
- บทที่ 6** การวิเคราะห์ผลการทดลองในบทที่ 5 จะถูกนำมาสรุปผลการวิจัยที่ได้ และเสนอแนะถึงวิธีการในการเลือกอัลกอริทึมในการเข้ารหัสซึ่งเหมาะสมกับการใช้งานที่มีความต้องการที่แตกต่างกันไป นอกจากนั้นยังกล่าวถึงปัญหาที่เกิดขึ้นในระหว่างการทำวิจัย และข้อเสนอแนะ หัวข้อสุดท้ายจะเป็นการเสนอแนวทางในการทำวิจัยต่อไปในอนาคต

นอกจากในส่วนของเนื้อหาวิทยานิพนธ์ทั้ง 6 บทแล้ว ยังมีภาคผนวก ซึ่งเป็นการอธิบายผลลัพธ์ที่ได้จากโปรแกรม ซอร์สโค้ดโปรแกรมที่ผู้วิจัยพัฒนาขึ้นในการวิจัย รวมทั้งรายละเอียดในการคอมไพล์โปรแกรม