


การวิเคราะห์ความสัมพันธ์ระหว่างลักษณะของยูอาร์แอลกับการใช้งานบริการเว็บแคช



นางสาวนพรัตน์ พรหมหาไชย

สถาบันวิทยบริการ

จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

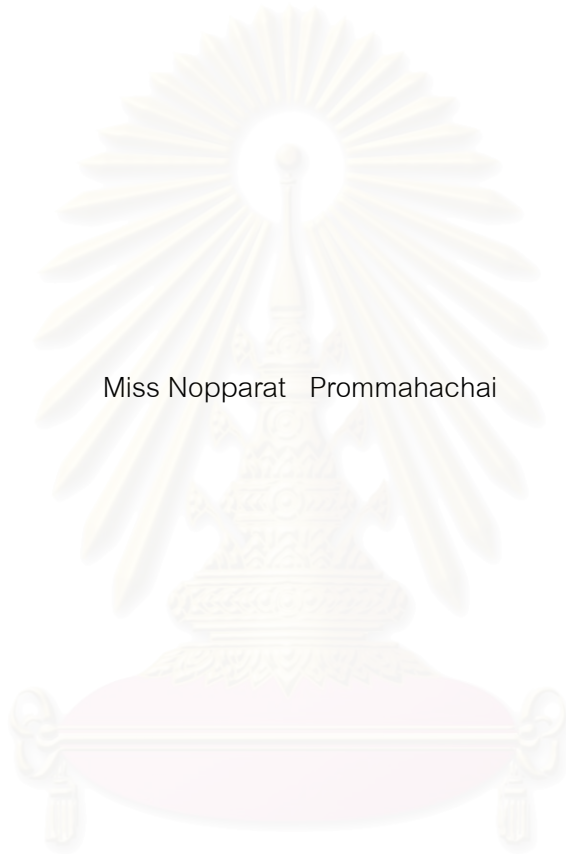
ปีการศึกษา 2543

ISBN 974-347-303-3

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

AN ANALYSIS OF RELATIONSHIP BETWEEN URL CHARACTERISTICS
AND WEB CACHE SERVICE USAGE

Miss Nopparat Prommahachai



A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Computer Science

Department of Computer Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2000

ISBN 974-347-303-3

41714398-21 : MAJOR COMPUTER SCIENCE

KEY WORD: WEB CACHE / URL CHARACTERISTIC / PATH DEPTH / LENGTH / FILE EXTENSION

NOPPARAT PROMMAHACHAI : AN ANALYSIS OF RELATIONSHIP BETWEEN URL CHARACTERISTICS AND WEB CACHE SERVICE USAGE. THESIS ADVISOR : YUNYONG TENG-AMNUAY, 48 pp. ISBN 974-347-303-3.

World Wide Web usage can be determined by analyzing important information, URL. Result from URL analysis can be used for web cache design in a more efficient way. Furthermore, URL information can also be used to develop special tool for web usage analysis. In this study, length, path depth, and file extension of URL are analyzed while proportion of number of data found in web cache to number of total service requests is used as measurement indicator.

One-month study of relation between URL information and Chulalongkorn University's web cache usage dated from 1st August to 31st August 2000 has shown that, upon approximately 64.7 million requests, which 1.44 percent of those are error information, the average URL length is 55.94 while the standard deviation is 36.29. The average value and standard deviation of requests found in web cache are 63.30 and 46.82 respectively. This means that there is more opportunity for short-length URL to be found in web cache than the long-length one. The average path depth of total service requests is 4.72 with standard deviation of 1.67 while the average path depth of requests found in web cache is 4.67 with standard deviation of 1.39. These values are almost equal to the average path depth and standard deviation of unfound requests, which are 4.98 and 12.07 respectively. From these results, it can be inferred that path depth property analysis might not give clear information. Most file extensions found in web cache are of graphic files (gif and jpg), html files (htm, html and default file) and multimedia file (mpg and mp3). Most requested file extensions while opportunity of finding those in web cache is low are cgi, pl asp, shtml, and no extension files.

Department Computer Engineering Student's signature.....

Field of study Computer Science.....Advisor's signature.....

Academic year 2000

Co-advisor's signature.....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยความช่วยเหลืออย่างดียิ่งของอาจารย์ ดร.ยรรยง เต็งอำนวยการ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ซึ่งท่านได้เอาใจใส่ให้คำแนะนำและข้อคิดเห็นต่าง ๆ ในการวิจัยด้วยดีตลอด

ทำยนี้ ผู้วิจัยใคร่ขอกราบขอบพระคุณ บิดา-มารดา ซึ่งสนับสนุนในด้านการเงิน และให้กำลังใจแก่ผู้วิจัยเสมอมาจนสำเร็จการศึกษา



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ง
บทคัดย่อภาษาอังกฤษ	จ
กิตติกรรมประกาศ	ฉ
สารบัญ	ช
สารบัญตาราง	ณ
สารบัญภาพ	ญ
บทที่	
1 บทนำ	1
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์ของการวิจัย	2
1.3 ขอบเขตของการวิจัย	2
1.4 ประโยชน์ที่คาดว่าจะได้รับ	3
1.5 วิธีดำเนินการวิจัย	3
2 เอกสารและงานวิจัยที่เกี่ยวข้อง	4
2.1 แนวคิดและทฤษฎี	4
2.1.1 องค์ประกอบของยูอาร์แอล	5
2.1.2 คุณลักษณะของยูอาร์แอล	5
2.1.3 การวิเคราะห์ทางสถิติ	6
2.2 เอกสารและงานวิจัยที่เกี่ยวข้อง	6
3 วิธีดำเนินการวิจัย	8
3.1 ข้อมูลที่นำมาใช้ในการวิจัย	8
3.1.1 ที่มาของข้อมูล	8
3.1.2 การเก็บรวบรวมข้อมูล	8
3.1.3 รูปแบบของข้อมูล	9
3.2 เครื่องมือที่ใช้ในการวิจัย	12
3.2.1 โปรแกรมวิเคราะห์ความสัมพันธ์ระหว่างความยาวของยูอาร์แอล และการใช้บริการเว็บแคช	14

สารบัญ (ต่อ)

ฎ

บทที่	หน้า
3.2.2	17
3.2.3	19
4	22
4.1	22
4.1.1	22
4.1.2	25
4.1.3	28
4.2	32
4.2.1	32
4.2.2	34
5	35
5.1	35
5.2	36
5.2.1	36
5.2.2	37
5.3	38
รายการอ้างอิง	39
ภาคผนวก	41
ภาคผนวก ก	42
ภาคผนวก ข	45
ภาคผนวก ค	46
ประวัติผู้วิจัย	48

สารบัญตาราง

ตาราง	หน้า
ตารางที่ 3.1 รายละเอียดของคำขอที่ถูกตัดออกจากการวิเคราะห์	9
ตารางที่ 3.2 ข้อมูลเกี่ยวกับคำขอที่ให้บริการ	11
ตารางที่ 3.3 ร้อยละของคำขอแบบต่าง ๆ เทียบกับคำขอทั้งหมดแจกแจงตามรหัสแคช	11
ตารางที่ 3.4 ความหมายของรหัสแคช	12
ตารางที่ 3.5 หน้าที่ขององค์ประกอบต่าง ๆ ของโปรแกรม	13
ตารางที่ 4.1 ค่าทางสถิติของความยาวยูอาร์แอล	25
ตารางที่ 4.2 ค่าทางสถิติของความลึกของเส้นทาง	28
ตารางที่ 4.3 ร้อยละของคำขอแจกแจงตามนามสกุลเพิ่มข้อมูล	29
ตารางที่ 4.4 ร้อยละของปริมาณข้อมูลให้บริการแจกแจงตามนามสกุลเพิ่มข้อมูล	31
ตารางที่ 4.5 ร้อยละของคำขอเมื่อลดคำขอที่เก็บลงเว็บแคชโดยพิจารณา จากช่วงความของยูอาร์แอลที่ค่าต่าง ๆ	33
ตารางที่ 4.6 ร้อยละของข้อมูลที่จัดเก็บเมื่อจำกัดคำขอที่เก็บลงเว็บแคช โดยพิจารณาจากช่วงความของยูอาร์แอลที่ค่าต่าง ๆ	33
ตารางที่ 5.1 จำนวนคำขอที่มีกลุ่มตัวอักษร "..." อยู่ในยูอาร์แอล	37

สารบัญภาพ

ภาพประกอบ	หน้า
รูปที่ 3.1 ลักษณะการเชื่อมต่อของอุปกรณ์เว็บแคม ในจุฬาลงกรณ์มหาวิทยาลัย	8
รูปที่ 3.2 ข้อมูลรายละเอียดการให้บริการเว็บแคมจากแฟ้มลงบันทึกการใช้งาน	9
รูปที่ 3.3 การทำงานของโปรแกรมวิเคราะห์ความยาวของยูอาร์แอล	14
รูปที่ 3.4 การทำงานของโปรแกรมส่วนของการตรวจสอบความถูกต้องของข้อมูล	15
รูปที่ 3.5 การทำงานของโปรแกรมส่วนของการบันทึกข้อมูล	16
รูปที่ 3.6 การทำงานของโปรแกรมวิเคราะห์ความลึกเส้นทางของยูอาร์แอล	17
รูปที่ 3.7 การทำงานของโปรแกรมส่วนของการนับความลึกเส้นทางของยูอาร์แอล	18
รูปที่ 3.8 การทำงานของโปรแกรมวิเคราะห์นามสกุลแฟ้มข้อมูลของยูอาร์แอล	19
รูปที่ 3.9 การทำงานของโปรแกรมส่วนของการตรวจสอบนามสกุลแฟ้มข้อมูล ของยูอาร์แอล	20
รูปที่ 4.1 ความสัมพันธ์ระหว่างร้อยละจำนวนคำขอกับความยาวของยูอาร์แอล	22
รูปที่ 4.2 ความสัมพันธ์ระหว่างร้อยละปริมาณข้อมูลให้บริการกับความยาวของยูอาร์แอล	23
รูปที่ 4.3 ความสัมพันธ์ระหว่างร้อยละของข้อมูลให้บริการที่พบในแคช กับความยาวของยูอาร์แอล	24
รูปที่ 4.4 ความสัมพันธ์ระหว่างร้อยละจำนวนคำขอกับความลึกเส้นทางของยูอาร์แอล	25
รูปที่ 4.5 ความสัมพันธ์ระหว่างร้อยละปริมาณข้อมูลให้บริการกับความลึกเส้นทาง ของยูอาร์แอล	26
รูปที่ 4.6 ความสัมพันธ์ระหว่างร้อยละของข้อมูลให้บริการที่พบในแคช กับความลึกเส้นทางของยูอาร์แอล	27
รูปที่ 4.7 ร้อยละของจำนวนคำขอแยกตามประเภทนามสกุลแฟ้มข้อมูล	30
รูปที่ 4.8 ร้อยละของปริมาณข้อมูลให้บริการแยกตามประเภทนามสกุลแฟ้มข้อมูล	32

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันอินเทอร์เน็ตได้เข้ามา มีบทบาทในชีวิตประจำวันเรามากขึ้น ด้วยความสะดวกรวดเร็ว ง่ายในการติดต่อ และความสามารถในกระจายข่าวสารไปทั่วโลก หน่วยงานหลายหน่วยงานจึงเริ่มมีการนำอินเทอร์เน็ตเข้ามาใช้ภายในองค์กร แต่เนื่องจากค่าใช้จ่ายในการเชื่อมต่อเข้ากับอินเทอร์เน็ตของประเทศไทย มีราคาแพงมากการจ้างงานจึงต้องทำให้คุ้มค่าและพยายามลดปริมาณข้อมูลที่วิ่งเข้าออกให้มากที่สุด ในการใช้งานในอินเทอร์เน็ตนั้น บริการเว็บเป็นบริการที่ได้รับความนิยมมากที่สุด จากสถิติการวัดปริมาณข้อมูลที่วิ่งผ่านแกนหลักของ MCI ระยะเวลา 1 สัปดาห์ในเดือนเมษายนปี 2541 มีปริมาณการใช้ข้อมูลที่เกี่ยวข้องกับเว็บถึง 75 เปอร์เซ็นต์ [1] การนำเว็บแคชเข้ามาใช้ จึงเป็นวิธีหนึ่งในการช่วยลดปริมาณข้อมูลที่วิ่งเข้าออกระหว่างภายในองค์กรกับโลกภายนอก โดยเมื่อมีผู้ใช้บริการส่งคำขอข้อมูลเว็บผ่านบริการเว็บแคช ตัวเว็บแคชจะเป็นผู้ที่เรียกข้อมูลเองและจะเก็บข้อมูลนั้นเอาไว้ เมื่อมีผู้อื่นร้องขอข้อมูลเดียวกันนี้อีก เว็บแคชก็จะส่งข้อมูลไปให้แทนถ้าเว็บแคชมีข้อมูลและตรวจสอบแล้วว่าข้อมูลนั้นไม่ล้าสมัย ถ้ามีผู้ร้องขอข้อมูลเดียวกันมาก ๆ และข้อมูลนั้นมีอยู่ในเว็บแคชก็จะทำให้ช่วยลดปริมาณข้อมูลที่วิ่งเข้ามาในองค์กรได้ แต่การเก็บข้อมูลในเว็บแคชก็มีข้อจำกัดเนื่องจากข้อมูลในโลกมีมากมาย มีเครื่องให้บริการข้อมูลเว็บเป็นปริมาณมากขึ้นทุกวัน เว็บแคชจึงไม่สามารถเก็บข้อมูลที่มีอยู่ทั้งหมดในโลกเพื่อให้บริการได้ นอกจากนี้ข้อมูลบางอย่างยังเป็นข้อมูลที่มีการเปลี่ยนแปลงในทุกครั้งที่ผู้ใช้มีการเรียกข้อมูลหรือข้อมูลที่เก็บอยู่ในเว็บแคชเก่าเกินไปต้องมีการดึงข้อมูลมาใหม่และเครื่องให้บริการข้อมูลเว็บก็สามารถกำหนดค่าตัวแปรในเว็บว่าให้เก็บข้อมูลเว็บของตัวเองไว้ได้นานเท่าใดจึงหมดอายุ ดังนั้นเว็บแคชจึงควรมีการเลือกข้อมูลที่มีโอกาสถูกเรียกใช้ได้อีกสูงเก็บไว้ ดังนั้นในการศึกษาพฤติกรรมการใช้งานเพื่อใช้ในการออกแบบเว็บแคชจะช่วยให้เว็บแคชมีการทำงานสอดคล้องกับพฤติกรรมของผู้ใช้มากขึ้น

ในการใช้บริการเว็บแคชนั้น สิ่งที่ผู้ใช้งานจะใช้ในการอ้างอิงถึงเว็บที่ต้องการคือ ยูอาร์แอล (URL: Universal Resource Locator) ซึ่งเป็นชื่อที่อ้างอิงถึงที่อยู่ของข้อมูลที่ใช้บริการที่ต้องการ และตัวเว็บแคชเองก็มีการเก็บข้อมูลโดยอ้างอิงกับยูอาร์แอล เช่นกัน ดังนั้นในการศึกษาพฤติกรรมการใช้งานเว็บแคชโดยพิจารณาจากลักษณะของยูอาร์แอลจึงน่าจะเป็นประโยชน์ในการ

เป็นข้อมูลช่วยออกแบบเว็บแคชให้มีประสิทธิภาพ ซึ่งตัววัดที่ง่ายที่สุดคือ จำนวนครั้งที่ร้องขอแล้วพบข้อมูลในเว็บแคช และปริมาณข้อมูลที่ร้องขอแล้วพบข้อมูลในเว็บแคช [2]

นอกจากนี้ในตัวเว็บแคชจะมีการเก็บแฟ้มลงบันทึกการใช้บริการเว็บแคชเพื่อให้สามารถนำมาวิเคราะห์ได้ภายหลัง เช่นนำมาวิเคราะห์หาประสิทธิภาพของเว็บแคชว่าสามารถช่วยลดปริมาณข้อมูลที่ต้องดึงจากอินเทอร์เน็ตมากน้อยเพียงใด สามารถวิเคราะห์พฤติกรรมของผู้ร้องขอบริการเว็บ หรือแม้แต่นำมาใช้คิดคำนวณค่าใช้จ่ายในการใช้บริการเว็บ ดังนั้นในการที่จะนำข้อมูลแฟ้มลงบันทึกการใช้บริการเว็บมาวิเคราะห์ภายในองค์กรใหญ่ ๆ ที่มีการใช้งาน อินเทอร์เน็ตมาก ๆ ทำให้ข้อมูลแฟ้มลงบันทึกการใช้งานเว็บแคชมีขนาดใหญ่ตามไปด้วย ซึ่งในการวิเคราะห์ข้อมูลของเว็บแคชปัญหาส่วนหนึ่งที่จะเกิดขึ้นคือ หน่วยความจำและเวลาที่ต้องใช้ในกระบวนการคำนวณ ซึ่งถ้ามีความเข้าใจในพฤติกรรมของการใช้งานเว็บแคชแล้วก็จะสามารถนำมาช่วยออกแบบระบบวิเคราะห์ข้อมูลได้ง่ายขึ้น หรืออาจช่วยให้การวิเคราะห์ทำได้เร็วขึ้นและใช้หน่วยความจำน้อยลง

1.2 วัตถุประสงค์ของการวิจัย

เพื่อศึกษาความสัมพันธ์ระหว่างลักษณะของ ยูอาร์แอล กับการเก็บข้อมูลปริมาณการใช้งานเว็บแคชในการใช้อินเทอร์เน็ต

1.3 ขอบเขตของการวิจัย

- 1.3.1 ข้อมูลที่จะนำมาทำการวิจัยจะเป็นข้อมูลของการใช้บริการเว็บแคชของจุฬาลงกรณ์มหาวิทยาลัย
- 1.3.2 กระบวนการที่จะทำการวิเคราะห์คือการทดสอบความสัมพันธ์ระหว่าง
 - 1) ลักษณะของ ยูอาร์แอล กับจำนวนครั้ง / ปริมาณข้อมูลที่เว็บแคชให้บริการ
 - 2) ลักษณะของ ยูอาร์แอล กับจำนวนครั้ง / ปริมาณข้อมูลที่เว็บแคชให้บริการโดยที่ข้อมูลที่ต้องการพบในแคช
- 1.3.3 ลักษณะของยูอาร์แอล หมายถึง
 - 1) ความยาวยูอาร์แอล
 - 2) ความลึกของเส้นทาง
 - 3) นามสกุลแฟ้มข้อมูล

1.4 ประโยชน์ที่คาดว่าจะได้รับ

- 1.4.1 เป็นข้อมูลช่วยในการออกแบบเว็บไซต์ให้มีประสิทธิภาพมากขึ้น
- 1.4.2 สามารถนำไปเป็นข้อมูลช่วยในการพัฒนาโปรแกรมที่ใช้คำนวณสถิติการให้บริการเว็บไซต์

1.5 วิธีดำเนินการวิจัย

- 1.5.1 ศึกษาวิธีการที่ใช้ในการเก็บข้อมูลสถิติในการให้บริการเว็บไซต์
- 1.5.2 เก็บข้อมูลการให้บริการเว็บไซต์ที่จะทำการทดสอบจากเว็บไซต์ของจุฬาลงกรณ์มหาวิทยาลัย
- 1.5.3 พัฒนาโปรแกรมในการวิเคราะห์ความสัมพันธ์ระหว่างลักษณะของยูอาร์แอล กับลักษณะของการให้บริการเว็บไซต์
- 1.5.4 ทำการวิเคราะห์และสรุปผล
- 1.5.5 ทำรายงาน



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

2.1 แนวคิดและทฤษฎี

โดยทั่วไปแล้วองค์กรต่าง ๆ ที่มีการใช้งานอินเทอร์เน็ตจะต้องเสียค่าใช้จ่ายในการเชื่อมต่อซึ่งโดยทั่วไป ค่าใช้จ่ายขึ้นอยู่กับแบนด์วิดท์ที่เชื่อมต่อกับอินเทอร์เน็ตหรือในบางกรณีก็มีการคิดค่าใช้จ่ายตามปริมาณการรับส่งข้อมูลผ่านทางเชื่อมต่อนี้ ซึ่งถ้าแบนด์วิดท์มีไม่เพียงพอต่อการใช้งานก็จะทำให้การให้บริการต่าง ๆ ในอินเทอร์เน็ตมีความล่าช้า ดังนั้นถ้าสามารถลดปริมาณข้อมูลที่วิ่งเข้าออกผ่านทางเชื่อมต่อก็จะช่วยประหยัดค่าใช้จ่ายลงได้ และในการใช้งานอินเทอร์เน็ตบริการเว็บเป็นบริการที่ได้รับความนิยมมากที่สุด ดังนั้นองค์กรต่าง ๆ จึงมักจะติดตั้งเว็บแคชเพื่อวัตถุประสงค์หลัก ๆ สองประการคือ

- 1) ลดเวลาในการรับส่งข้อมูลในการใช้บริการเว็บ
- 2) ลดปริมาณข้อมูลที่รับส่งระหว่างองค์กรกับภายนอก

แต่ถ้าข้อมูลที่ต้องการไม่มีอยู่ในเว็บแคชอาจทำให้เป็นการเพิ่มเวลาในการใช้บริการเว็บ นอกจากนี้ในบางกรณีเว็บแคชก็ไม่ได้ลดปริมาณข้อมูลที่รับส่งแต่กลับเพิ่มได้เช่นกัน เช่นเว็บแคชบางประเภทสามารถกำหนดให้ตั้งเวลาไปดึงข้อมูลจากเครื่องให้บริการข้อมูลเว็บในเวลาที่มีการใช้งานน้อยเพื่อลดการรับส่งข้อมูลในช่วงเวลาที่มีการใช้งานคับคั่ง ให้กระจายการใช้งานไปยังช่วงที่ไม่ค่อยมีการใช้งานแทน แต่ข้อมูลที่ไปดึงมานั้นไม่ได้ถูกใช้งานเลยก็เป็นได้ อย่างไรก็ตามก็ดีเมื่อมองภาพโดยรวมแล้วเว็บแคชสามารถช่วยให้การให้บริการเว็บมีคุณภาพดีขึ้นได้จากคุณสมบัติหลัก ๆ ดังที่กล่าวมาข้างต้น

ในการวิเคราะห์การให้บริการเว็บแคชนั้นใน Evaluation Content Management Techniques for Web Proxy Caches [2] ได้กล่าวถึงสิ่งที่เป็นคุณสมบัติสำคัญซึ่งสามารถบ่งชี้ถึงประสิทธิภาพของเว็บแคชคือ

- 1) อัตราส่วนของจำนวนครั้งที่เว็บแคชให้บริการโดยข้อมูลที่ต้องการพบในแคชเทียบกับจำนวนครั้งที่ให้บริการทั้งหมด
- 2) อัตราส่วนของปริมาณข้อมูลที่เว็บแคชให้บริการโดยที่ข้อมูลที่ต้องการพบในแคช เทียบกับปริมาณข้อมูลให้บริการทั้งหมด

2.1.1 องค์ประกอบของยูอาร์แอล URL (Uniform Resource Locator)

จาก RFC1738 [3] RFC1808 [4] RFC2368 [5] และข้อมูลที่น่ามาวิเคราะห์แล้ว ยูอาร์แอลมีองค์ประกอบคร่าว ๆ ดังนี้

url = [scheme:] ["/" net_loc] ["/" [path] [";" params] ["?" query] ["#" fragment]

โดยที่

scheme	โปรโตคอลที่ใช้ซึ่งจากข้อมูลที่น่ามาวิเคราะห์จะมีเพียงโปรโตคอล http
net_loc	ที่อยู่ของเครื่องที่ให้บริการข้อมูล
path	ส่วนประกอบที่เป็นเส้นทางบอกที่อยู่ของข้อมูลที่อยู่ในเครื่องที่ให้บริการ ข้อมูลซึ่งมีรูปแบบดังนี้ path = fsegment * ("/" segment)
fsegment	ข้อมูลย่อยของเส้นทางที่มีตัวอักขระอย่างน้อย 1 ตัว
segment	ข้อมูลย่อยของเส้นทาง
params	พารามิเตอร์
query	ข้อสอบถาม
fragment	ชิ้นส่วนย่อย

2.1.2 คุณลักษณะของยูอาร์แอล

คุณลักษณะของยูอาร์แอลที่งานวิจัยนี้กล่าวถึงจะมีอยู่ 3 คุณลักษณะซึ่งได้แก่

- 1) **ความยาวของยูอาร์แอล** หรือจำนวนตัวอักขระทั้งหมดที่ประกอบเป็นยูอาร์แอลนั้น ๆ
- 2) **ความลึกของเส้นทาง** ซึ่งจากองค์ประกอบของยูอาร์แอลความลึกของเส้นทางจะพิจารณาถึงส่วนที่เป็น path โดยมีตัวอักขระ "/" เป็นตัวคั่นระหว่างความลึกแต่ละชั้นของเส้นทาง ดังนั้นเมื่อนำข้อมูลยูอาร์แอลทั้งหมดมาพิจารณา จะทำการตัดส่วนที่เหลือออกเหลือจากส่วนของ path ออกก่อน
- 3) **นามสกุลของแฟ้มข้อมูล** ซึ่งจากองค์ประกอบของยูอาร์แอลนามสกุลของแฟ้มข้อมูลจะพิจารณาถึงส่วนที่เป็น path เช่นกันโดยพิจารณาข้อมูลที่มีความลึกชั้นสุดท้ายของ path ซึ่งนามสกุลแฟ้มข้อมูลจะเป็นข้อมูลสุดท้ายหลังตัวอักขระ "." ดังนั้นเมื่อนำข้อมูลยูอาร์แอลทั้งหมดมาพิจารณา จะทำการตัดส่วนที่เหลือออกเหลือจากส่วนของ path ออกก่อน

2.1.3 การวิเคราะห์ทางสถิติ

ในการวิจัยได้นำการวิเคราะห์ทางสถิติมาช่วยเสริมให้การวิเคราะห์มีความชัดเจนมากขึ้นโดยคุณสมบัติทางสถิติที่นำมาใช้มีอยู่ 2 ประเภทคือ การวัดแนวโน้มเข้าสู่ศูนย์กลางและการวัดการกระจายของข้อมูล [6]

โดยในการวัดแนวโน้มเข้าสู่ศูนย์กลางนั้นเราพิจารณาอยู่ 3 ค่าคือ

ค่าเฉลี่ยเลขคณิต คือค่าเฉลี่ยจากการคำนวณผลรวมของสมาชิกทุกตัวหารด้วยจำนวนสมาชิก

ค่ามัธยฐาน คือค่าที่อยู่กึ่งกลางของข้อมูลเมื่อนำข้อมูลทั้งหมดมาเรียงจากน้อยไปหามาก

ค่าฐานนิยม คือค่าที่มีจำนวนสมาชิกมากที่สุด

ส่วนการวัดการกระจายก็พิจารณาอยู่ 3 ค่าเช่นกันคือ

พิสัย คือผลต่างของค่าสูงสุดกับค่าต่ำสุด

ส่วนเบี่ยงเบนควอไทล์ คือผลต่างของค่าที่อยู่ลำดับที่ $\frac{3}{4}$ กับค่าที่อยู่ลำดับที่ $\frac{1}{4}$ เมื่อเรียงจากน้อยไปหามาก

ส่วนเบี่ยงเบนมาตรฐาน คือค่าที่คำนวณจากสูตรการหาค่าส่วนเบี่ยงเบนมาตรฐานทั่วไป

2.2 เอกสารและงานวิจัยที่เกี่ยวข้อง

เนื่องจากยูอาร์แอลเป็นคุณสมบัติที่สำคัญอย่างหนึ่งของการใช้บริการเว็บ ดังนั้นจึงมีเอกสารงานวิจัยที่เกี่ยวข้องกับยูอาร์แอลในรูปแบบต่าง ๆ กันดังตัวอย่างเช่นใน User's Behavior and URL Analysis at EC Sites [7] ซึ่งเป็นการศึกษาพฤติกรรมการใช้งานของผู้ใช้บริการเว็บโดยวิเคราะห์จากยูอาร์แอลเพื่อทำการจัดกลุ่มและวัดปริมาณของผู้ใช้บริการเว็บว่าเข้ามาเพื่อดู สนใจ สั่งซื้อ หรือตกลงสั่งซื้อแน่นอน โดยวิเคราะห์จากคำสำคัญในส่วนของเส้นทางที่อยู่ในยูอาร์แอล แต่เนื่องจากการวิเคราะห์ในฝั่งเครื่องที่ให้บริการเพียงที่เดียวข้อมูลที่ได้จากการวิเคราะห์จึงเป็นข้อมูลที่เจาะจงเฉพาะแห่งใดแห่งหนึ่งเท่านั้น และคำสำคัญที่อยู่ในยูอาร์แอลอาจไม่ชัดเจนเพียงพอหรือมีรูปแบบแตกต่างกันไปในแห่งอื่น ๆ

ส่วนใน Proxy Caching Based on Object Location Considering Semantic Usage [8] จะเป็นการเสนอเว็บแคชโดยใช้วิธีการใหม่ ซึ่งใช้เส้นทางของยูอาร์แอลเป็นสิ่งพิจารณาในการจัดเก็บข้อมูลและแทนที่ข้อมูลในเว็บแคช โดยตั้งสมมติฐานที่ว่า เส้นทางของยูอาร์แอลใดที่มีการใช้งานบ่อย ๆ ก็จะทำให้มีโอกาสถูกเก็บในแคชนานขึ้น ส่วนเส้นทางที่ไม่ค่อยได้ใช้ก็จะค่อย ๆ

ถูกแทนที่ด้วยข้อมูลที่ใหม่กว่าเมื่อเนื้อที่ที่ใช้เก็บข้อมูลในเว็บแคชเต็ม แต่การพัฒนาเว็บแคชที่ใช้วิธีการนี้อาจทำให้เว็บแคชมีการทำงานที่ซับซ้อนจนเกิดความล่าช้าในการให้บริการได้

ก่อนหน้านงานวิจัยของ Philippe Rochat และ Stuart Thompson นั้น ได้มีงานวิจัยที่เกี่ยวข้องกับความลึกของเส้นทางคือ Characterizing Browsing Strategies in the World-Wide Web [9] ซึ่งได้ใช้ความลึกของเส้นทางเป็นตัวแปรในการแยกแยะประเภทของผู้ใช้บริการเว็บ โดยผู้วิจัยทำการวาดกราฟความสัมพันธ์ระหว่างความถี่ของการใช้งานเทียบกับความลึกของเส้นทาง แล้วแบ่งกลุ่มผู้ใช้งานโดยพิจารณาที่ค่าความชันของกราฟ แต่เนื่องจากเป็นงานวิจัยที่เก่ามาก ผู้ใช้บริการเว็บอาจมีการเปลี่ยนแปลงพฤติกรรมไปแล้ว

นอกจากนี้ยังมีงานวิจัยอื่น ๆ ที่มีการกล่าวถึงคุณลักษณะอื่นของยูอาร์แอลอีกเช่น นามสกุลเพิ่มข้อมูล Characteristics of WWW Client-Based Traces [1] และ Web Server Workload Characterization: The Search for Invariants [10] ซึ่งในงานวิจัยส่วนใหญ่จะมีการวิเคราะห์ลักษณะการให้บริการเว็บในแง่ต่าง ๆ กันและพิจารณาจากจุดต่าง ๆ กัน แต่ส่วนใหญ่คุณสมบัตินึ่งที่งานวิจัยส่วนใหญ่นำมากล่าวถึงคือนามสกุลเพิ่มข้อมูล

นอกเหนือจากงานวิจัยต่าง ๆ ยังมีเอกสารที่เกี่ยวข้องในงานวิจัยนี้คือเอกสารของ RFC1738 [3] ซึ่งได้อธิบายองค์ประกอบของยูอาร์แอลเอาไว้ และในงานวิจัยนี้ก็ได้นำข้อมูลดังกล่าวเป็นข้อมูลหลักในการวิจัย นอกจากนี้ยังมีเอกสารปรับปรุงเพิ่มเติมเอกสาร RFC1738 [3] คือ RFC1808 [4] เพิ่มเติมเกี่ยวกับรีเลทีฟยูอาร์แอล และ RFC2368 [5] เพิ่มเติมข้อมูลเกี่ยวกับโปรเซสยี่อเล็กทรอนิกส์

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

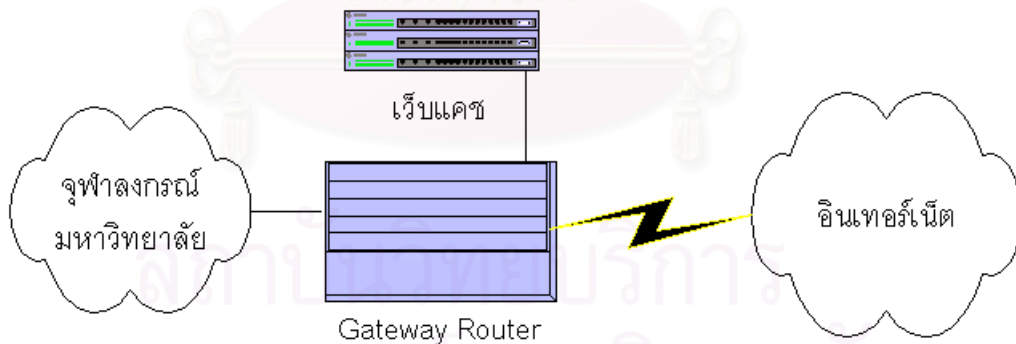
บทที่ 3 วิธีดำเนินการวิจัย

3.1 ข้อมูลที่นำมาใช้ในการวิจัย

สำหรับข้อมูลที่นำมาใช้ในการวิจัยนั้นผู้ทำการวิจัยได้แบ่งหัวข้อออกเป็น 3 หัวข้อย่อย ๆ คือ ที่มาของข้อมูลว่าข้อมูลที่นำมาวิจัยนั้นได้มาจากที่ใด การเก็บรวบรวมข้อมูลซึ่งเป็นรายละเอียดคร่าว ๆ ของข้อมูลที่นำมาใช้ และรูปแบบของข้อมูลซึ่งจะมีคำอธิบายลักษณะและความหมายของรูปแบบข้อมูลที่นำมาใช้

3.1.1 ที่มาของข้อมูล

ข้อมูลที่นำมาใช้ในการวิจัยเป็นข้อมูลจากแฟ้มลงบันทึกการให้บริการเว็บแคชของจุฬาลงกรณ์มหาวิทยาลัย ซึ่งทางมหาวิทยาลัยได้มีการนำเว็บแคชมาใช้งานให้บริการส่วนกลางในจุดที่เชื่อมต่อออกภายนอกโดยได้ใช้อุปกรณ์ที่ทำหน้าที่เว็บแคชโดยเฉพาะ ซึ่งคำขอใช้บริการเว็บทั้งหมดที่ถูกเรียกออกผ่านจุดเชื่อมต่อจะถูกเปลี่ยนทิศทางมายังตัวเว็บแคชแทน และในตัวเว็บแคชเองก็จะมีการเก็บข้อมูลแฟ้มลงบันทึกการเรียกใช้บริการเว็บเอาไว้ด้วย



รูปที่ 3.1 ลักษณะการเชื่อมต่อของอุปกรณ์เว็บแคช ในจุฬาลงกรณ์มหาวิทยาลัย

3.1.2 การเก็บรวบรวมข้อมูล

ในงานวิจัยนี้ได้ใช้ข้อมูลของการใช้งานเว็บแคชของจุฬาลงกรณ์มหาวิทยาลัยระหว่างวันที่ 1 ถึงวันที่ 31 สิงหาคม พ.ศ. 2542 โดยที่ปริมาณข้อมูลที่นำมาวิเคราะห์ทั้งหมดประมาณ 8 กิกะไบต์ มีจำนวนข้อมูลทั้งหมดประมาณ 64.7 ล้านบรรทัด และเนื่องจากเป็นการ

วิเคราะห์การใช้บริการเว็บแคชเท่านั้น ในการนำข้อมูลมาวิเคราะห์จะทำการตัดข้อมูลที่ไม่เกี่ยวข้องหรือมีความผิดพลาดออก ข้อมูลที่ถูกตัดออกมีจำนวนทั้งหมด 935,386 บรรทัด คิดเป็นร้อยละ 1.44 ของคำขอทั้งหมด ซึ่งมีรายละเอียดดังต่อไปนี้

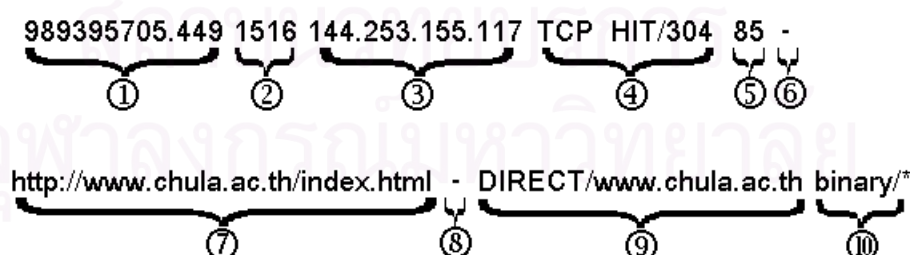
ตารางที่ 3.1 รายละเอียดของคำขอที่ถูกตัดออกจากการวิเคราะห์

ประเภทข้อมูลที่ถูกตัดออก	จำนวนคำขอ
ข้อมูลที่ไม่มีข้อมูลของยูอาร์แอล	929,238
ข้อมูลที่มีจำนวนข้อมูลไม่เท่ากับ 10 เซตข้อมูล	5,656
ข้อมูลที่มีข้อมูลของยูอาร์แอลไม่ได้ขึ้นต้นด้วย "http://"	231
ข้อมูลที่มีข้อมูลของปริมาณข้อมูลให้บริการน้อยกว่า 0	261
รวม	935,386

หมายเหตุ ข้อมูลของปริมาณข้อมูลให้บริการมีโอกาสน้อยกว่าศูนย์ได้ ในกรณีที่ปริมาณข้อมูลที่ให้บริการมีจำนวนมากกว่าค่าความสามารถในการเก็บข้อมูลของตัวแปรปริมาณข้อมูลให้บริการ หรืออาจเกิดขึ้นจากความผิดพลาดในการบันทึกข้อมูล

3.1.3 รูปแบบของข้อมูล

รูปแบบของแฟ้มลงบันทึกการใช้งานอุปกรณ์เว็บแคชนั้นทางผู้ขายผลิตภัณฑ์เว็บแคชได้อธิบายเอาไว้ในเว็บไซต์ โดยแฟ้มลงบันทึกจะเป็นแฟ้มประเภทข้อความที่มีข้อมูลแสดงเป็นบรรทัด โดยจะบันทึกข้อมูลรายละเอียดทั้งหมดของการให้บริการหนึ่งบรรทัดต่อหนึ่งคำขอ ซึ่งในหนึ่งคำขอจะมีข้อมูลรายละเอียดแบ่งเป็น 10 เซตข้อมูล แต่ละเซตข้อมูลแยกจากกันด้วยช่องว่าง



รูปที่ 3.2 ข้อมูลรายละเอียดการให้บริการเว็บแคชจากแฟ้มลงบันทึกการใช้งาน

ความหมายของข้อมูลรายละเอียดทั้ง 10 เขตข้อมูลมีดังนี้

- 1) เวลาที่ให้บริการ ซึ่งมีหน่วยเป็นวินาทีนับจากวันที่ 1 มกราคม 1970 เวลา 00:00:00 นาฬิกา
- 2) ช่วงระยะเวลาที่ให้บริการ หน่วยเป็นมิลลิวินาที
- 3) เลขที่อยู่อินเทอร์เน็ตของผู้ใช้บริการ
- 4) รหัสของแคชและสถานะของข้อมูลที่ร้องขอว่ามีอยู่ในแคชหรือไม่ เช่น
 - 4.1) TCP_HIT หมายถึงข้อมูลที่ร้องขอมืออยู่ในแคช
 - 4.2) TCP_MISS หมายถึงข้อมูลที่ร้องขอไม่มีอยู่ในแคช
 - 4.3) 200 หมายถึงรหัสสถานะ เสร็จสิ้นสมบูรณ์
 - 4.4) 304 หมายถึงรหัสสถานะห้ามแก้ไข
 - 4.5) 404 หมายถึงรหัสสถานะไม่พบข้อมูล
- 5) ปริมาณข้อมูลที่ส่งให้ผู้ให้บริการ มีหน่วยเป็นไบต์
- 6) กระบวนการที่ร้องขอไป เช่น GET หรือ POST โดย GET หมายถึงการดึงข้อมูลของเครื่องให้บริการข้อมูลเว็บ ส่วน POST หมายถึง ในคำขอจะมีการส่งข้อมูลไปยังเครื่องให้บริการข้อมูลเว็บด้วย
- 7) ยูอาร์แอลที่ร้องขอ กรณีที่ไม่มีการใส่ยูอาร์แอลที่ร้องขอเอาไว้ ข้อมูลนี้จะไม่ปรากฏ
- 8) รายละเอียดของช่องบริการที่ซีพีทีไอที่อยู่กรณีที่เปิดการทำงานให้บันทึกข้อมูลส่วนนี้
- 9) รหัสบอกลำดับชั้นของเว็บแคช
 - 9.1) DIRECT หมายถึงยูอาร์แอลนี้เรียกตรงจากเครื่องให้บริการ
 - 9.2) NONE ไม่มีลำดับชั้น
- 10) ประเภทของข้อมูลเป็นข้อความหรือเป็นข้อมูลฐานสอง

ในการวิเคราะห์ข้อมูลนั้น จะมีเขตข้อมูลที่ต้องพิจารณาอยู่ 4 ส่วนหลักคือ

ข้อมูลที่ 4 ซึ่งบอกข้อมูลรหัสแคช

ข้อมูลที่ 5 ซึ่งบอกปริมาณข้อมูลที่ตัวเว็บแคชส่งให้ผู้ร้องขอ มีหน่วยเป็นไบต์

ข้อมูลที่ 7 บอกถึงยูอาร์แอลที่ผู้ใช้บริการร้องขอข้อมูล

ข้อมูลที่ 9 รหัสบอกลำดับชั้นของเว็บแคช ซึ่งจะแบ่งออกเป็น 3 กลุ่มคือ

- 1) ข้อมูลคำขอถูกส่งต่อไปยังเครื่องให้บริการข้อมูล ซึ่งพิจารณาจากข้อมูลลำดับชั้นที่มีข้อมูลเป็น DIRECT/[Server] โดยที่ [Server] หมายถึงชื่อเครื่องที่ให้บริการข้อมูล
- 2) ข้อมูลที่มีอยู่ในแคช ซึ่งพิจารณาจากข้อมูลลำดับชั้นที่มีข้อมูลเป็น NONE/-
- 3) ข้อมูลที่มีข้อมูลลำดับชั้นเป็น DIRECT/0.0.0.0

จากข้อมูลที่น่ามาวิเคราะห์เมื่อนำมาแจกแจงประเภทตามที่ได้กล่าวมาข้างต้น
จะได้ข้อมูลดังตารางที่ 3.2 ต่อไปนี้

ตารางที่ 3.2 ข้อมูลเกี่ยวกับคำขอที่ให้บริการ

ประเภทข้อมูลที่ให้บริการ	จำนวนข้อมูล(ล้านครั้ง)	ปริมาณข้อมูล(กิกะไบต์)
ข้อมูลให้บริการทั้งหมด	63.8	431.3
ข้อมูลให้บริการที่ไม่พบในแคช	33.7	276.3
ข้อมูลให้บริการที่พบในแคช	29.9	153.8
ข้อมูลให้บริการแสดงข้อผิดพลาด	1.1	1.2

เมื่อทำการแจกแจงอย่างละเอียดโดยแยกตามรหัสแคชโดยแสดงเป็นร้อยละของ
ข้อมูลทั้งหมดจะได้ข้อมูลดังตารางที่ 3.3 ต่อไปนี้

ตารางที่ 3.3 ร้อยละของคำขอแบบต่าง ๆ เทียบกับคำขอทั้งหมดแจกแจงตามรหัสแคช

รหัสแคช	ร้อยละของจำนวนคำขอ เทียบกับจำนวนคำขอทั้งหมด				ร้อยละของปริมาณข้อมูล ให้บริการเทียบกับปริมาณข้อมูล ให้บริการทั้งหมด			
	รวม	พบใน แคช	ผิด พลาด	ไม่พบ ในแคช	รวม	พบใน แคช	ผิด พลาด	ไม่พบ ในแคช
TCP_CLIENT_REFRESH	1.34	-	-	1.34	1.68	-	-	1.68
TCP_HIT	29.38	29.38	-	-	34.61	34.61	-	-
TCP_IMS_HIT	17.54	17.54	-	-	1.05	1.05	-	-
TCP_MISS	44.44	-	0.19	44.25	52.99	-	0.26	52.73
TCP_REFRESH_HIT	0.68	0.68	0.00	-	0.50	0.50	0.00	-
TCP_REFRESH_MISS	3.03	-	0.00	3.02	4.90	-	0.00	4.90
TCP_UNKNOWN	3.60	-	0.03	3.57	4.27	-	0.02	4.25
Total	100	47.60	0.22	52.18	100	36.16	0.28	63.56

หมายเหตุ

รวม = จำนวน/ปริมาณทั้งหมดในรหัสแคชนั้น

พบในแคช = จำนวน/ปริมาณข้อมูลให้บริการที่พบในแคช

ผิดพลาด = จำนวน/ปริมาณข้อมูลที่มีข้อผิดพลาด

ไม่พบในแคช = จำนวน/ปริมาณข้อมูลให้บริการที่ไม่พบในแคช

โดยที่ความหมายของรหัสแคชต่าง ๆ มีดังตารางที่ 3.4 ต่อไปนี้

ตารางที่ 3.4 ความหมายของรหัสแคช

รหัสแคช	ความหมาย
TCP_CLIENT_REFRESH	ผู้ใช้บริการส่งข้อความ "no-cache" มายังเว็บแคชพร้อมกับคำขอ ซึ่งเว็บแคชจะทำการส่งต่อคำขอนั้นต่อไปยังเครื่องให้บริการข้อมูลโดยตรง
TCP_HIT	ข้อมูลที่ต้องการมีอยู่แล้วในแคช
TCP_IMS_HIT	ผู้ใช้บริการส่งคำขอตรวจสอบความใหม่ของข้อมูลมายังแคชพร้อมกับคำขอ ซึ่งพบข้อมูลในแคชและยังเป็นข้อมูลใหม่
TCP_MISS	ข้อมูลที่ต้องการไม่มีอยู่ในแคช
TCP_REFRESH_HIT	ข้อมูลที่ต้องการมีอยู่แล้วในแคชแต่แคชพบว่าเป็นข้อมูลเก่า จึงส่งคำขอตรวจสอบความใหม่ของข้อมูลไปยังเครื่องให้บริการข้อมูลและได้รับคำตอบว่ายังไม่มีการแก้ไข
TCP_REFRESH_MISS	ข้อมูลที่ต้องการมีอยู่แล้วในแคชแต่แคชพบว่าเป็นข้อมูลเก่า จึงส่งคำขอตรวจสอบความใหม่ของข้อมูลไปยังเครื่องให้บริการข้อมูลและได้รับข้อมูลใหม่กลับมา
TCP_UNKNOWN	ข้อมูลไม่ทราบสถานะของรหัสแคช

3.2 เครื่องมือที่ใช้ในการวิจัย

จากข้อมูลที่จะนำมาวิเคราะห์ที่เป็นข้อมูลข้อความ ผู้ทำวิจัยเห็นว่าการนำโปรแกรม Perl ซึ่งเป็นโปรแกรมที่มีฟังก์ชันในการจัดการข้อมูลเกี่ยวกับข้อความมากมาย และง่ายต่อการศึกษา ทำให้ขั้นตอนการศึกษาการใช้งานไม่นานจนเกินไป ผู้ทำวิจัยจึงเลือกโปรแกรม Perl ในการสร้างเครื่องมือที่ใช้ในการวิเคราะห์ข้อมูล ส่วนในการแสดงผลรูปภาพ หรือการคำนวณบางอย่าง ผู้ทำวิจัยเลือกใช้โปรแกรมสำเร็จรูป MS Excel ในการใช้งาน

การวิเคราะห์ข้อมูลโดยใช้โปรแกรม Perl จะแบ่งออกเป็นออกเป็นโปรแกรมหลักๆ อยู่ 3 โปรแกรมซึ่งทำหน้าที่วิเคราะห์ความสัมพันธ์ระหว่าง

- 1) ความยาวของยูอาร์แอลและการใช้บริการเว็บแคช
- 2) ความลึกเส้นทางของยูอาร์แอลและการใช้บริการเว็บแคช
- 3) นามสกุลเพิ่มข้อมูลของยูอาร์แอลและการใช้บริการเว็บแคช

นอกจากนี้ยังมีโปรแกรมช่วยด้านอื่น ๆ เช่น

- 1) โปรแกรมหาความสัมพันธ์ระหว่างข้อมูลค่าขอเทียบกับรหัสแคช
- 2) โปรแกรมคำนวณหาค่าเฉลี่ยเลขคณิตและค่าเบี่ยงเบนมาตรฐาน

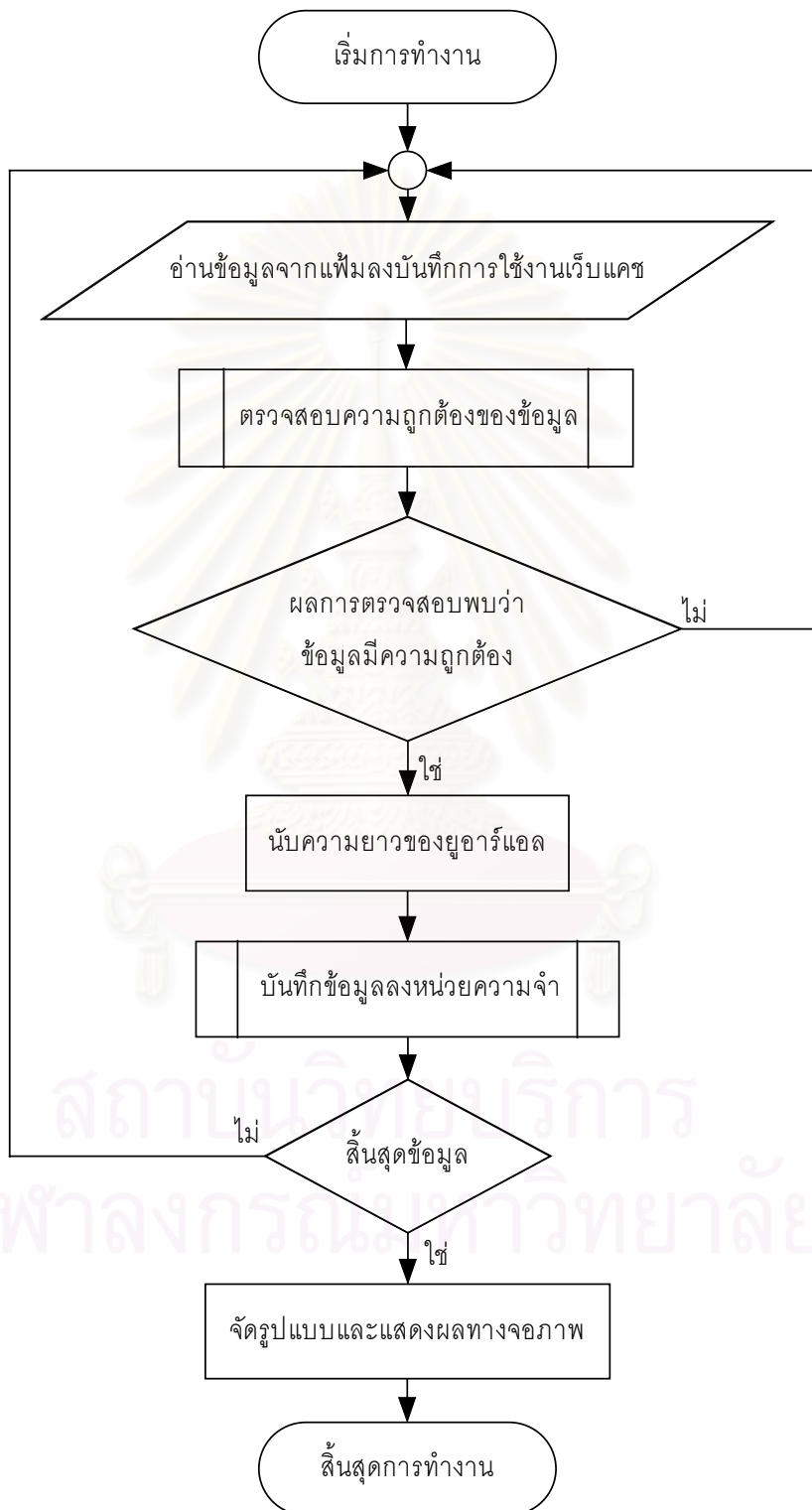
ในตัวโปรแกรมหลักเป็น 3 โปรแกรมที่มีองค์ประกอบบางส่วนเหมือนกันจึงสามารถใช้ด้วยกันได้ซึ่งองค์ประกอบเหล่านั้นเป็นดังตารางที่ 3.5 ต่อไปนี้ ตารางที่ 3.5 หน้าที่ขององค์ประกอบต่าง ๆ ของโปรแกรม

องค์ประกอบของโปรแกรม	หน้าที่
ตรวจสอบความถูกต้องของข้อมูล	ตรวจสอบว่าข้อมูลการใช้งานบรรทัดนั้น ๆ มีข้อผิดพลาดเบื้องต้นหรือไม่
บันทึกข้อมูลลงหน่วยความจำ	นำข้อมูลที่ประมวลแล้วบันทึกเข้าหน่วยความจำตามหมวดหมู่ที่เหมาะสม

แผนผังการทำงานของโปรแกรมหลัก 3 โปรแกรมที่ใช้ในการวิเคราะห์ข้อมูลมีรายละเอียดดังต่อไปนี้ ในส่วนขององค์ประกอบที่ใช้งานร่วมกันจะมีรายละเอียดดังรูปที่ 3.4 และ 3.5

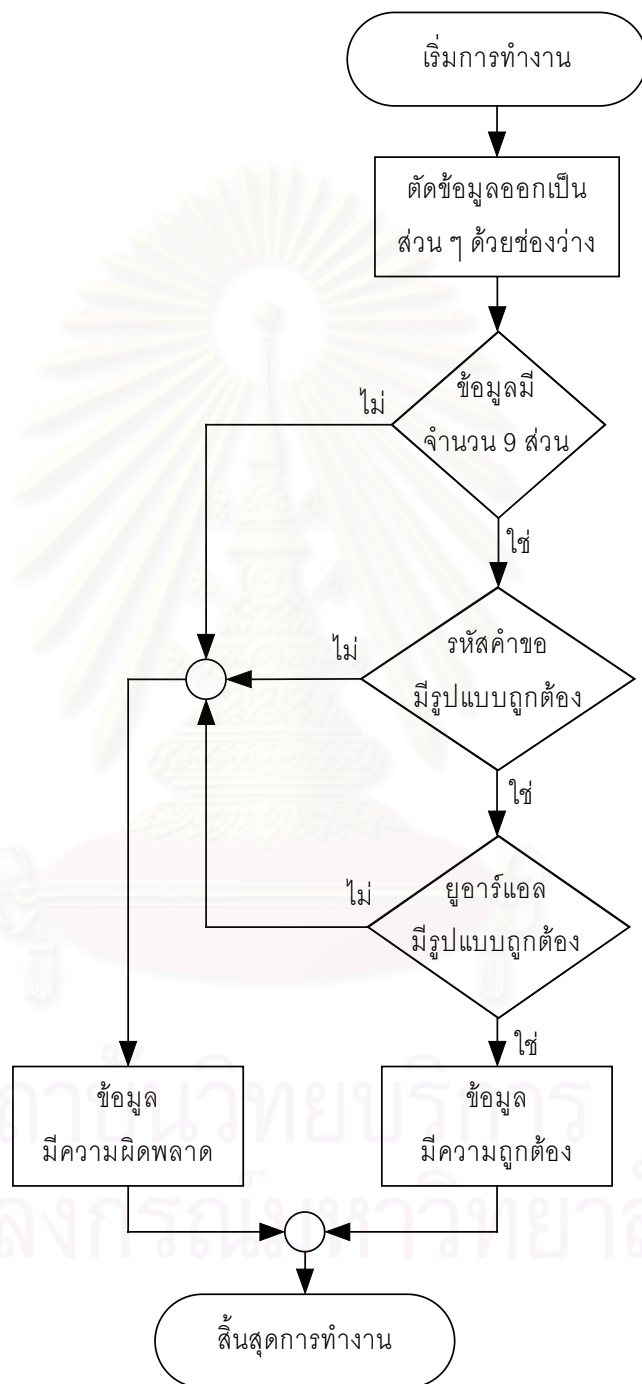
สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

3.2.1 โปรแกรมวิเคราะห์ความสัมพันธ์ระหว่างความยาวของยูอาร์แอลและการให้บริการ เว็บแคช



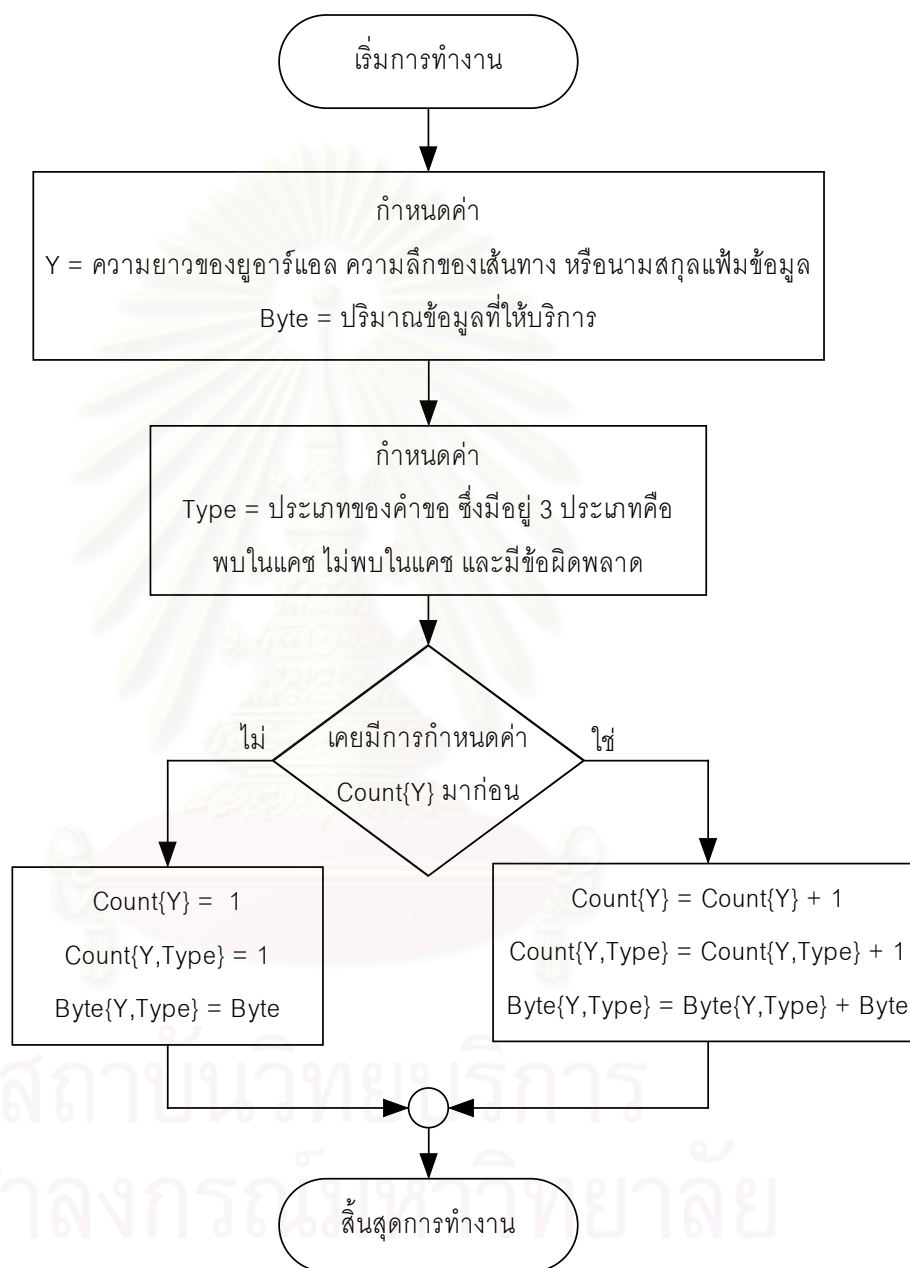
รูปที่ 3.3 การทำงานของโปรแกรมวิเคราะห์ความยาวของยูอาร์แอล

จากแผนภูมิข้างต้นในส่วนของการตรวจสอบความถูกต้องของข้อมูลจะมีรายละเอียดดังรูปที่ 3.4 ต่อไปนี้



รูปที่ 3.4 การทำงานของโปรแกรมส่วนของการตรวจสอบความถูกต้องของข้อมูล

ส่วนของการบันทึกข้อมูลลงหน่วยความจำจะมีการคำนวณก่อนบันทึกค่าตัวแปรต่าง ๆ ดังรูปที่ 3.5 ต่อไปนี้



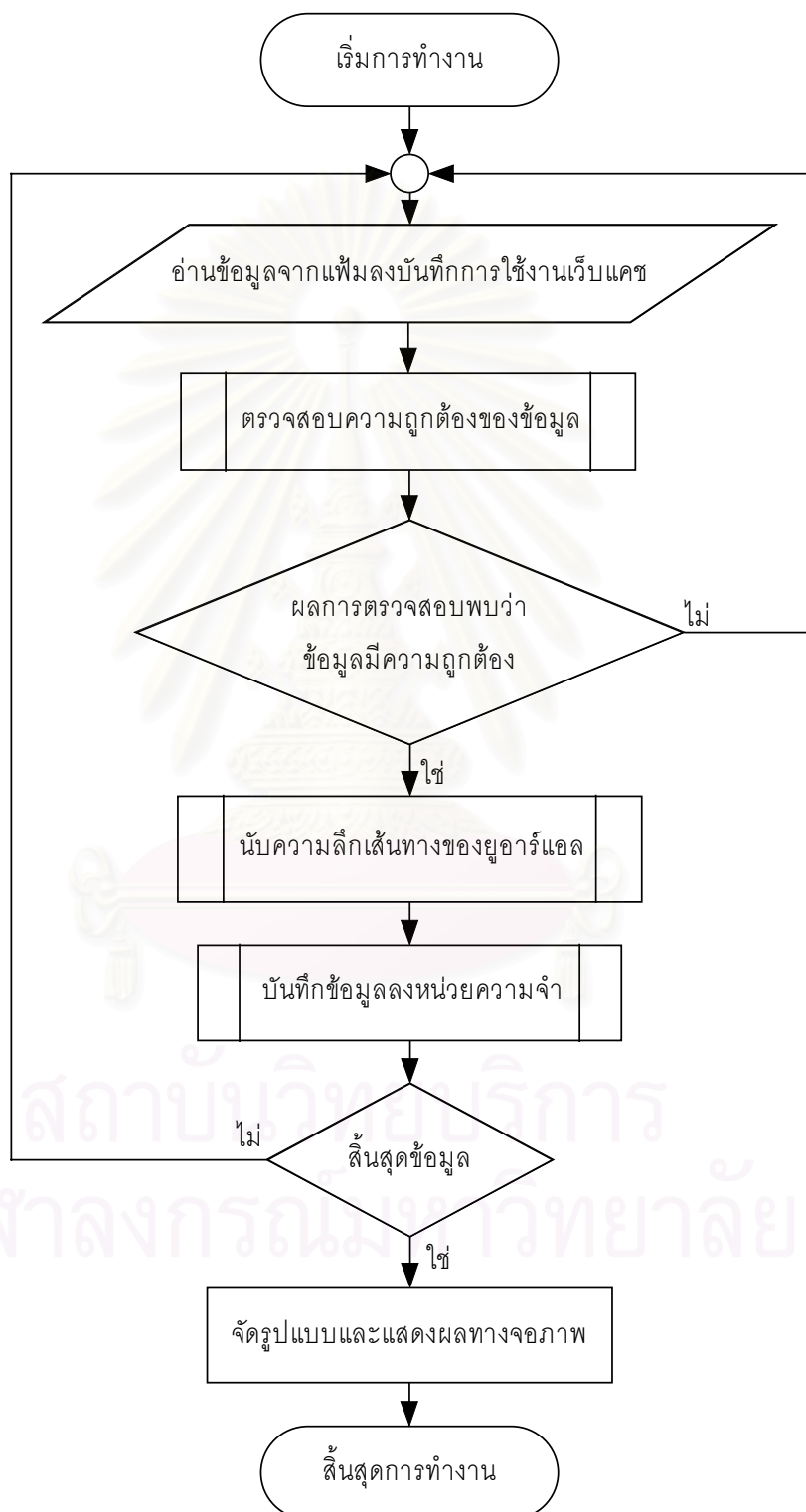
Count{Y} หมายถึง จำนวนของค่าขอของยูอาร์แอลที่ค่า Y

Count{Y,Type} หมายถึง จำนวนของค่าขอของค่า Y ที่ประเภท Type

Byte{Y,Type} หมายถึง ปริมาณข้อมูลให้บริการของค่า Y ที่ประเภท Type

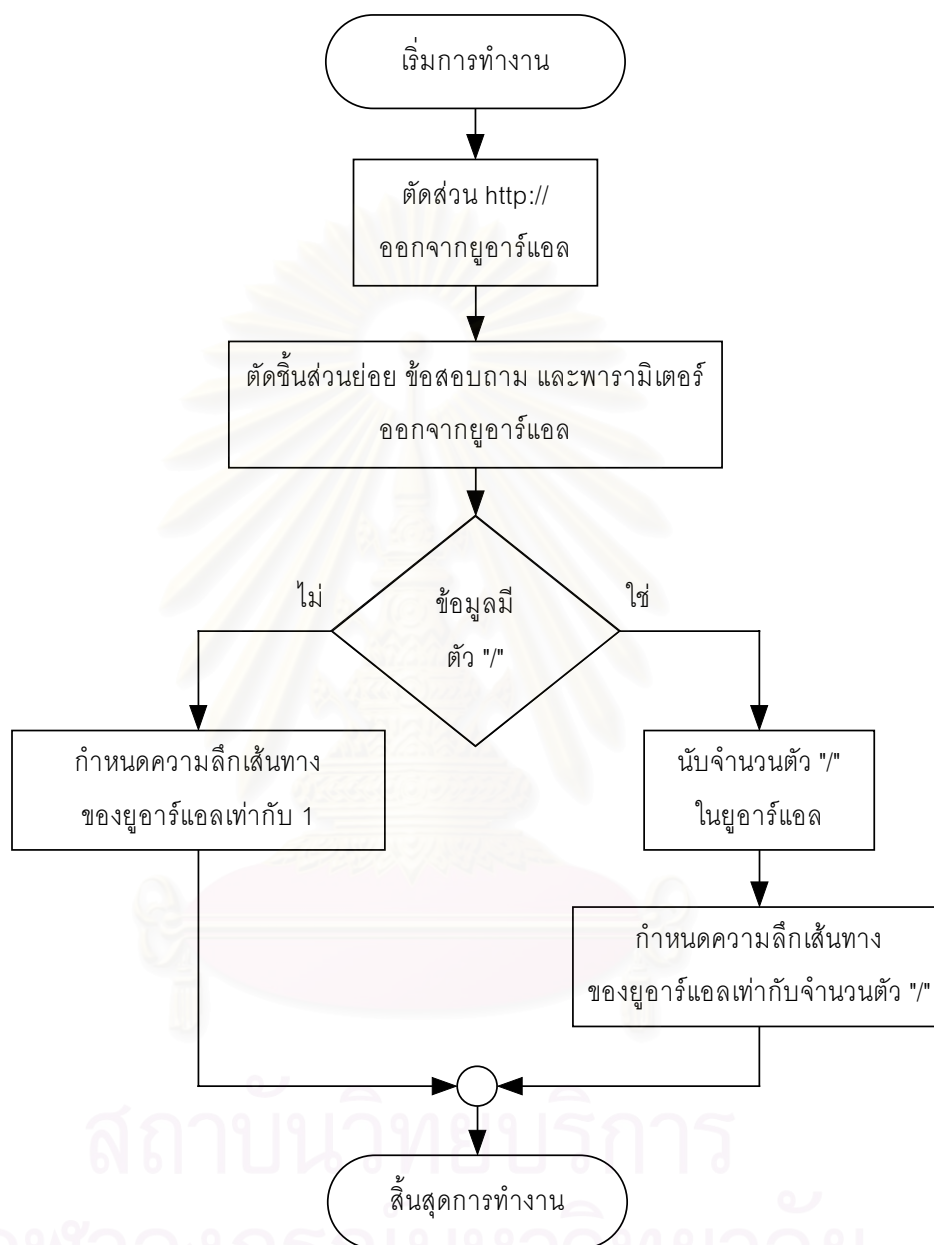
รูปที่ 3.5 การทำงานของโปรแกรมส่วนของการบันทึกข้อมูล

3.2.2 โปรแกรมวิเคราะห์ความสัมพันธ์ระหว่างความลึกเส้นทางของยูอาร์แอลและการใช้ บริการเว็บแคช



รูปที่ 3.6 การทำงานของโปรแกรมวิเคราะห์ความลึกเส้นทางของยูอาร์แอล

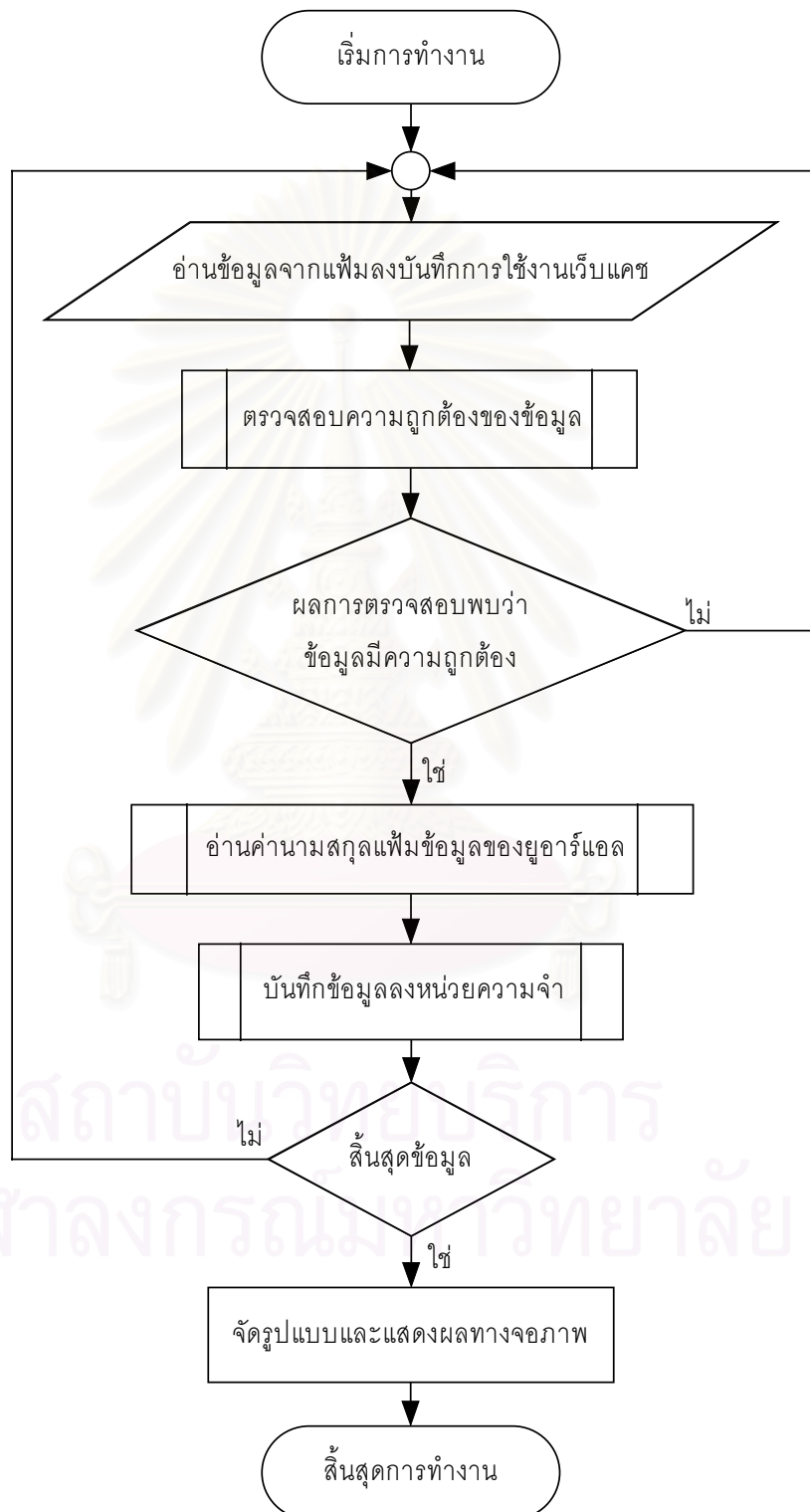
ในการทำงานส่วนของกระบวนการนับความลึกเส้นทางของยูอาร์แอลจะมีรายละเอียดดังรูปที่ 3.7 ต่อไปนี้



รูปที่ 3.7 การทำงานของโปรแกรมส่วนของกระบวนการนับความลึกเส้นทางของยูอาร์แอล

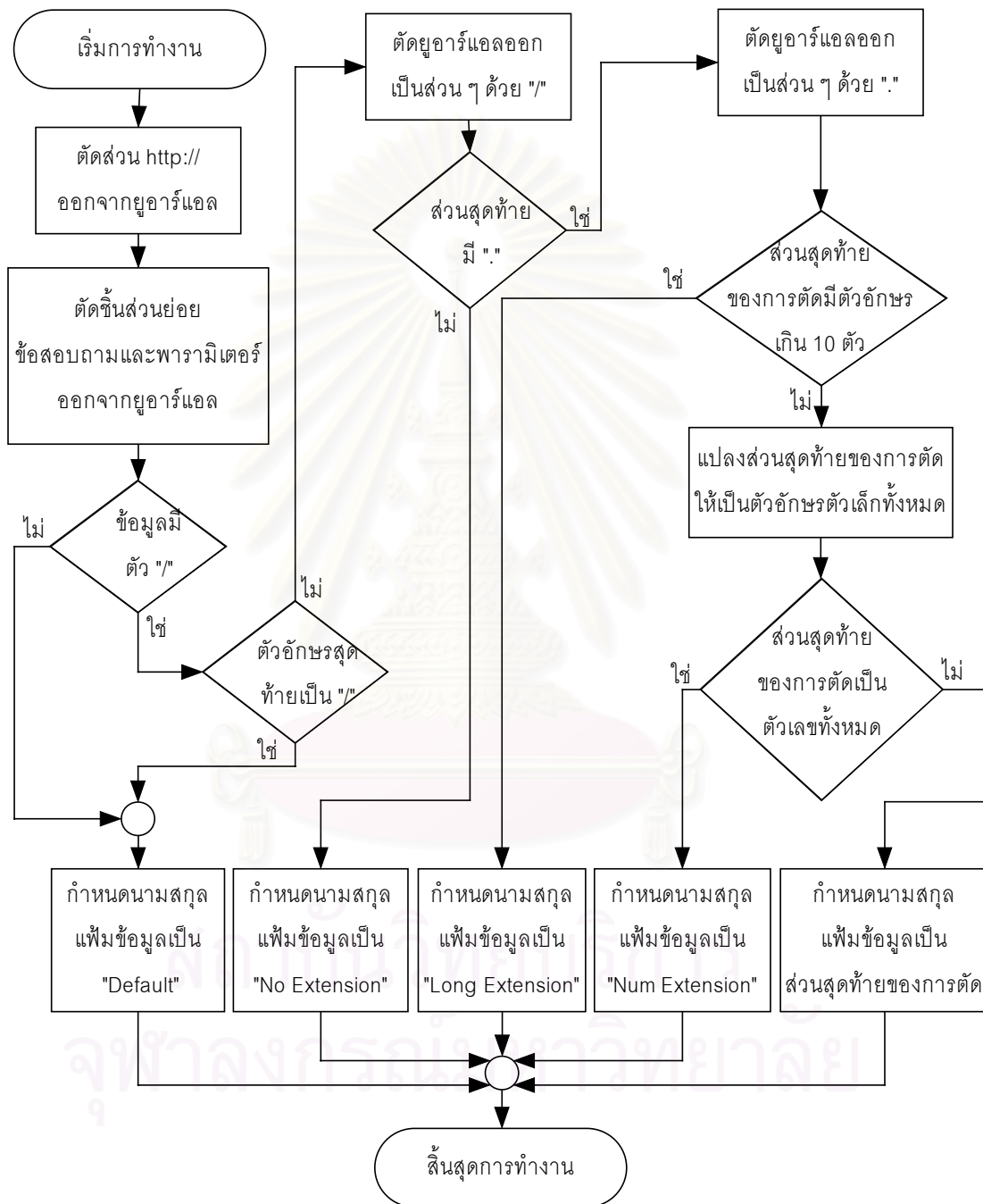
ในส่วนการทำงานของโปรแกรมตรวจสอบความถูกต้องของข้อมูล การบันทึกข้อมูลลงหน่วยความจำ จะเหมือนกันกับในโปรแกรมวิเคราะห์ความสัมพันธ์ระหว่างความยาวของยูอาร์แอลกับการใช้บริการเว็บแคช ดังรูปที่ 3.4 และ 3.5

3.3.3 โปรแกรมวิเคราะห์ความสัมพันธ์ระหว่างนามสกุลเพิ่มข้อมูลและการใช้บริการเว็บ แคช



รูปที่ 3.8 การทำงานของโปรแกรมวิเคราะห์นามสกุลเพิ่มข้อมูลของยูอาร์แอล

ในการทำงานส่วนการหาคำชื่อเพิ่มข้อมูลของยูอาร์แอลจะมีรายละเอียดดังรูปที่ 3.9 ต่อไปนี้



รูปที่ 3.9 การทำงานของโปรแกรมส่วนของการหาคำชื่อเพิ่มข้อมูลของยูอาร์แอล

ส่วนการทำงานของโปรแกรมตรวจสอบความถูกต้องของข้อมูล การบันทึกข้อมูลลงหน่วยความจำ จะเหมือนกันกับในโปรแกรมวิเคราะห์ความสัมพันธ์ระหว่างความยาวของยูอาร์แอลกับการใช้บริการเว็บแคชดังรูปที่ 3.4 และ 3.5

เมื่อนำโปรแกรมมาทำการประมวลผลข้อมูลที่ต้องการวิเคราะห์ใช้เวลาในการประมวลผลประมาณ 2 ชั่วโมงครั้งต่อ 1 โปรแกรมหลักจะได้ผลการวิเคราะห์ซึ่งจะกล่าวในบทที่ 4



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

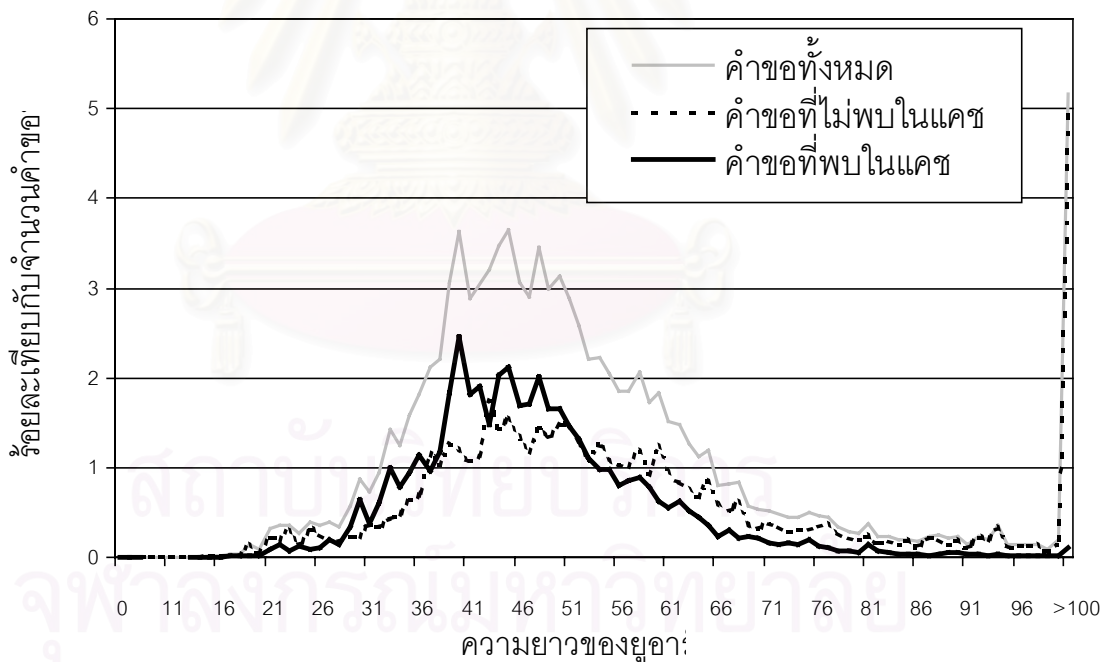
บทที่ 4 ผลการวิเคราะห์ข้อมูล

4.1 ผลการวิเคราะห์

จากการวิเคราะห์ข้อมูลโดยใช้เครื่องมือที่สร้างขึ้นมาผู้ทำการวิจัยได้แบ่งออกเป็น 2 ส่วนย่อยในการวิเคราะห์แต่ละส่วนคือ ผลจากการวิเคราะห์โดยรวม และผลจากการวิเคราะห์เมื่อพิจารณาโดยค่าทางสถิติ

4.1.1 ความสัมพันธ์ระหว่างความยาวของยูอาร์แอล กับการเก็บข้อมูลในเว็บแคช

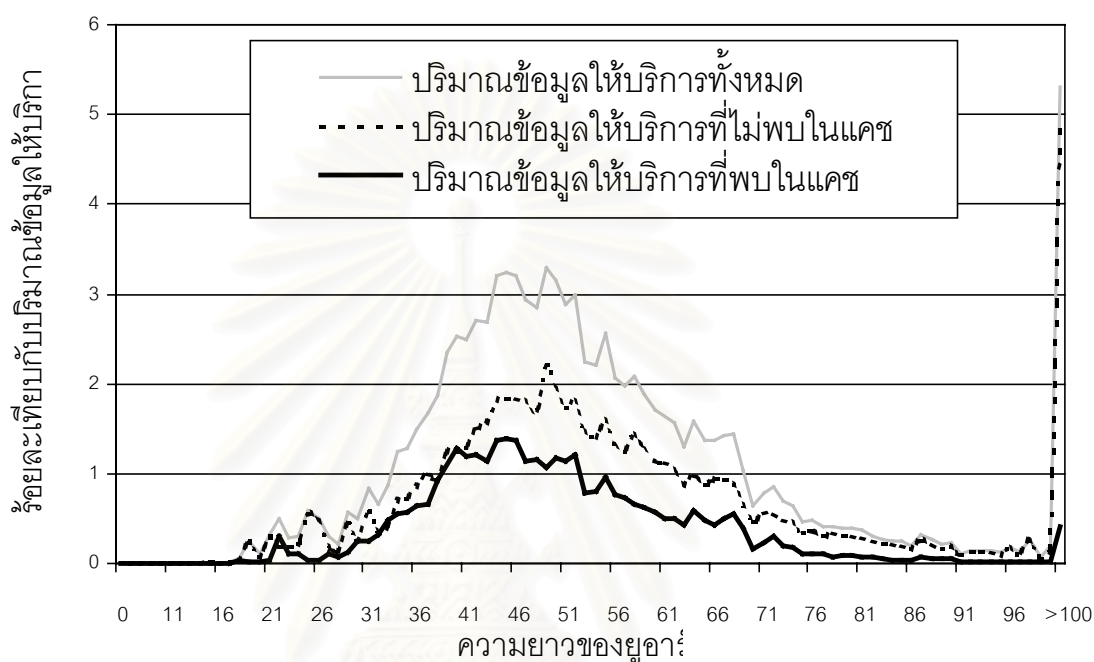
เมื่อนำข้อมูลไปประมวลผลและวาดกราฟความสัมพันธ์ระหว่างความยาวของยูอาร์แอลกับการเก็บข้อมูลในเว็บแคชได้ผลออกมาดังนี้



รูปที่ 4.1 ความสัมพันธ์ระหว่างร้อยละของจำนวนคำขอกับความยาวยูอาร์แอล

จากกราฟจะเห็นว่าที่ความยาวของยูอาร์แอลน้อยกว่าค่าเฉลี่ยกราฟคำขอตามรูปที่ 4.1 ที่พบในแคชมีแนวโน้มมากกว่ากราฟคำขอที่ไม่พบในแคช ส่วนที่ความยาวมากกว่าค่าเฉลี่ย

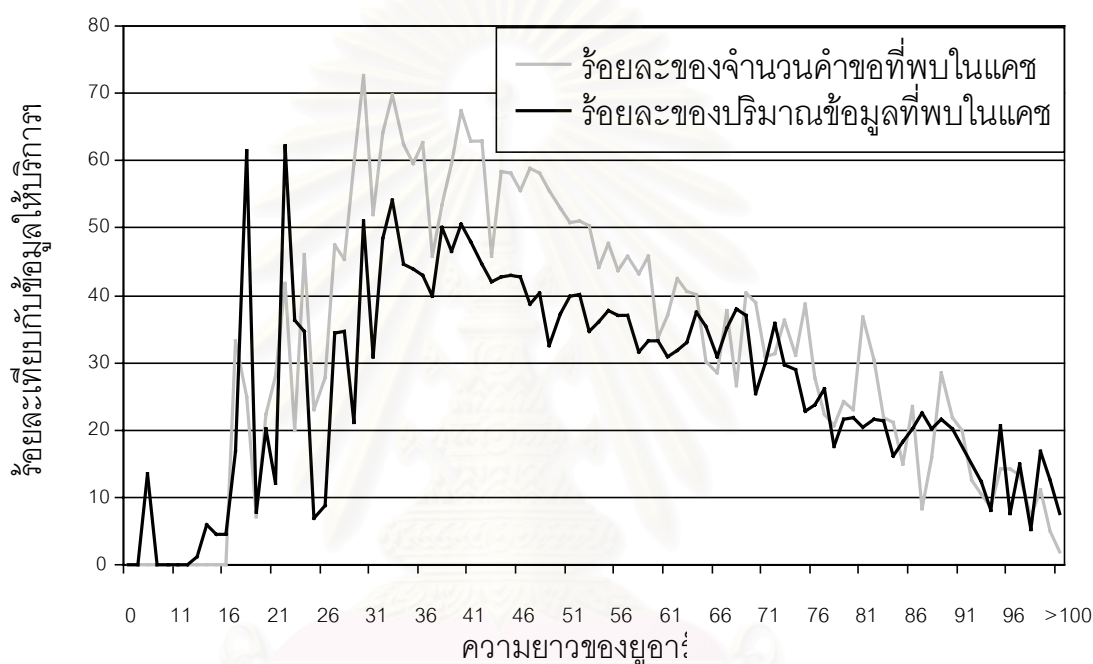
กราฟค่าขอที่ไม่พบข้อมูลในแคชสูงกว่ากราฟของค่าขอที่พบข้อมูลในแคช และที่ความยาวของยูอาร์แอลมากกว่า 100 ค่าขอเกือบทั้งหมดเป็นค่าขอที่ไม่พบข้อมูลในแคช



รูปที่ 4.2 ความสัมพันธ์ระหว่างร้อยละของปริมาณข้อมูลให้บริการกับความยาวยูอาร์แอล

เมื่อพิจารณาที่ปริมาณข้อมูลให้บริการ จากกราฟรูปที่ 4.2 ช่างต้นพบว่าที่ความยาวยูอาร์แอลน้อยกว่า 41 ปริมาณข้อมูลให้บริการที่พบและไม่พบในแคชมีค่าใกล้เคียงกัน แต่ที่ความยาวยูอาร์แอลมากกว่า 41 กราฟของปริมาณข้อมูลให้บริการที่ไม่พบในแคชมีค่าสูงกว่าปริมาณข้อมูลให้บริการที่พบในแคช และเช่นเดียวกันกับกราฟแสดงความสัมพันธ์ระหว่างร้อยละของจำนวนคำขอกับความยาวของยูอาร์แอล ที่ความยาวยูอาร์แอลมากกว่า 100 ปริมาณข้อมูลให้บริการเกือบทั้งหมดเป็นข้อมูลที่ไม่พบในแคช

จากกราฟรูปที่ 4.1 และ 4.2 เมื่อพิจารณาที่ความสัมพันธ์ระหว่างร้อยละของการพบข้อมูลในแคชเทียบกับข้อมูลที่ให้บริการทั้งหมดจะได้กราฟดังรูปที่ 4.3 ซึ่งแสดงให้เห็นว่าที่ความยาวของยูอาร์แอลต่ำ ๆ อัตราการพบข้อมูลในแคชจะมีแนวโน้มเพิ่มขึ้นถึงช่วงหนึ่ง เมื่อเลยช่วงนั้นไปแล้วจะอัตราของการพบข้อมูลในแคชจะมีแนวโน้มที่ลดลงไปเรื่อย ๆ



รูปที่ 4.3 ความสัมพันธ์ระหว่างร้อยละของข้อมูลให้บริการที่พบในแคชกับความยาวยูอาร์แอล

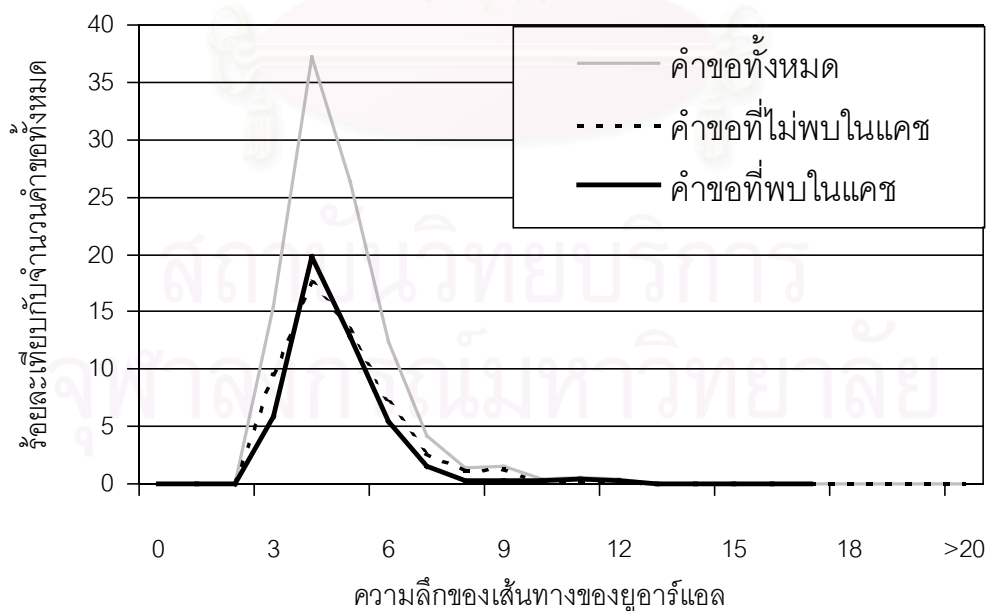
และเมื่อนำข้อมูลมาวิเคราะห์ทางสถิติ (ตารางที่ 4.1) พบว่าค่าแนวโน้มเข้าสู่ศูนย์กลางของข้อมูลที่พบในแคชมีค่าต่ำกว่าค่าแนวโน้มเข้าสู่ศูนย์กลางของข้อมูลรวม ในขณะที่ข้อมูลที่ไม่พบในแคชมีค่าสูงกว่า ส่วนค่าการกระจายของข้อมูลที่พบในแคชก็มีค่าน้อยกว่าของค่าของข้อมูลรวมและข้อมูลที่ไม่พบในแคช นั่นคือค่าของข้อมูลที่พบในแคชจะมีการกระจายอยู่ในช่วงแคบ ๆ ของความยาวค่าต่ำ ๆ เมื่อเทียบกับค่าความยาวโดยเฉลี่ยและการกระจายของข้อมูลที่ให้บริการทั้งหมดและข้อมูลที่ไม่พบในแคช

ตารางที่ 4.1 ค่าทางสถิติของความยาวยูอาร์แอล

ค่าทางสถิติ	ประเภท	ค่าขอทั้งหมด	พบข้อมูลในแคช	ไม่พบข้อมูลในแคช	มีข้อผิดพลาด
การวัดแนวโน้มเข้าสู่ส่วนกลาง	ค่าเฉลี่ยเลขคณิต	55.94	47.64	63.30	57.42
	ค่ามัธยฐาน	49	46	53	45
	ค่าฐานนิยม	45	40	43	21
การวัดการกระจาย	พิสัย	3,065	1,170	3,064	3,056
	ส่วนเบี่ยงเบนควอไทล์	9.50	6.50	11.50	14.00
	ส่วนเบี่ยงเบนมาตรฐาน	36.29	12.19	46.82	111.45
ค่าคุณสมบัติอื่น ๆ	ค่าต่ำสุด	7	7	8	16
	ค่าสูงสุด	3,072	1,177	3,072	3,072

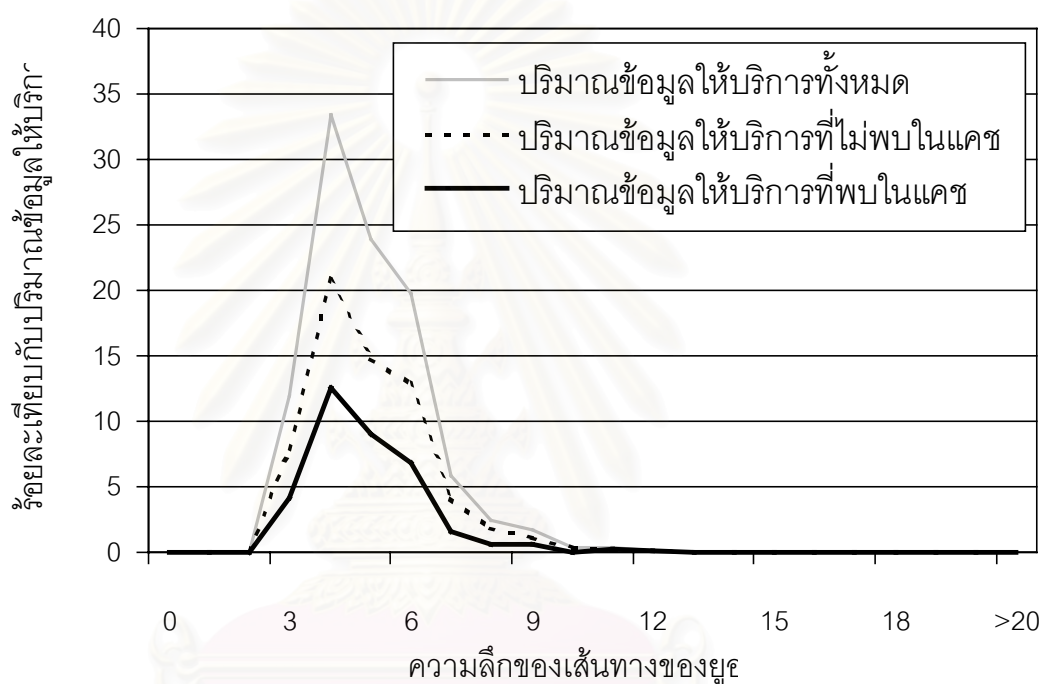
4.1.2 ความสัมพันธ์ระหว่างความลึกกับการเก็บข้อมูลในเว็บแคช

เมื่อพิจารณาความลึกเส้นทางของยูอาร์แอลจะได้ข้อมูลดังต่อไปนี้ (รูปที่ 4.4)



รูปที่ 4.4 ความสัมพันธ์ระหว่างร้อยละของจำนวนคำขอกับความลึกเส้นทาง

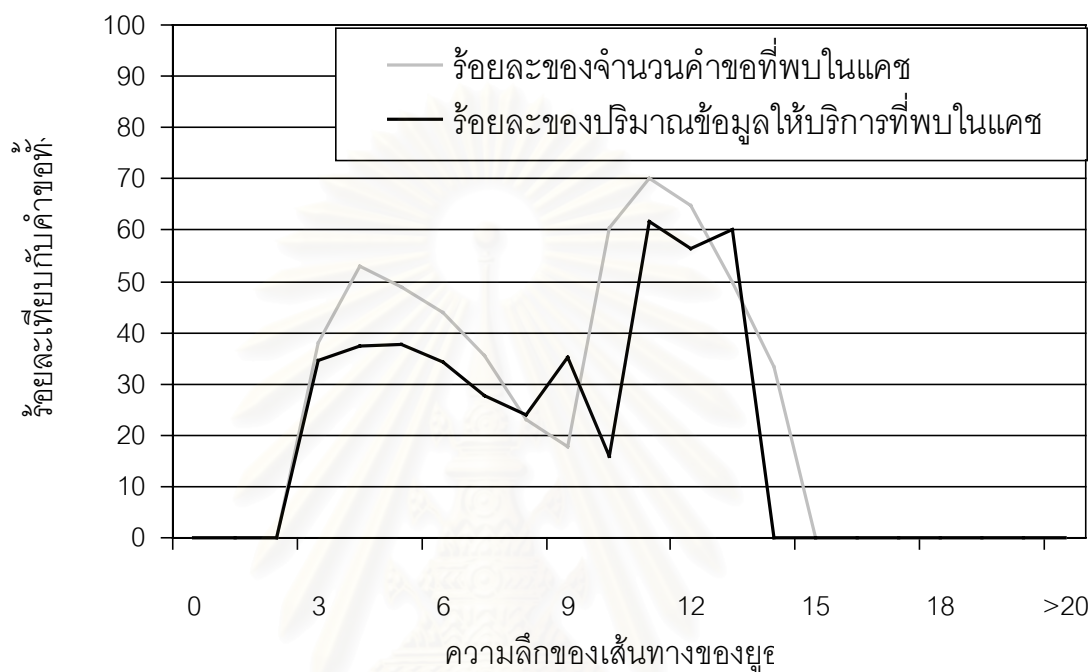
จากกราฟข้างต้นจะเห็นว่าความถี่เส้นทางของค่าขอที่พบและไม่พบในแคชมีรูปแบบของกราฟใกล้เคียงกันมาก ความแตกต่างที่พบคือกราฟของค่าขอที่พบในแคชจะมีความชันมากกว่าค่าขอที่ไม่พบข้อมูลในแคช และที่ความถี่ของเส้นทางมากกว่า 8 ขึ้นไปข้อมูลที่พบในแคชมีน้อยมาก และเมื่อพิจารณาที่ปริมาณข้อมูลให้บริการจะได้กราฟดังรูปที่ 4.5



รูปที่ 4.5 ความสัมพันธ์ระหว่างร้อยละของปริมาณข้อมูลให้บริการกับความถี่เส้นทางของยูเอชเอช

จะเห็นว่าที่ความถี่ของเส้นทางมากกว่า 20 ปริมาณข้อมูลที่ใช้บริการทั้งพบและไม่พบในแคชจะมีปริมาณน้อยมากซึ่งให้ผลคล้ายคลึงกันกับในการพิจารณาที่จำนวนค่าขอ

เมื่อนำข้อมูลจากกราฟรูปที่ 4.4 และ 4.5 มาพิจารณาที่ความสัมพันธ์ระหว่างร้อยละของการพบข้อมูลในแคชเทียบกับข้อมูลที่ให้บริการทั้งหมดจะได้กราฟดังรูปที่ 4.6



รูปที่ 4.6 ความสัมพันธ์ระหว่างร้อยละของข้อมูลให้บริการที่พบในแคชกับความลึกเส้นทางของยูอาร์แอล

เมื่อนำมาวิเคราะห์ทางสถิติ (ตารางที่ 4.2) ก็ให้ผลเดียวกับรูปภาพ คือความแตกต่างระหว่างข้อมูลให้บริการที่พบและไม่พบในแคช มีน้อยมากถึงแม้ข้อมูลที่ได้จะเป็นไปในแนวทางเดียวกันกับความยาวของยูอาร์แอล นั่นคือข้อมูลให้บริการที่พบในแคชมีค่าแนวโน้มเข้าสู่ศูนย์กลางและการกระจายน้อยกว่าข้อมูลทั้งหมดและข้อมูลที่ไม่พบในแคช แต่ค่าความแตกต่างนั้นมีค่าน้อยและบางค่าก็มีค่าเท่ากัน

ดังนั้นจากข้อมูลที่น่ามาวิเคราะห์นี้ ความลึกเส้นทางของยูอาร์แอลยังไม่มีวามเด่นชัดนักในการแยกแยะยูอาร์แอลที่มีโอกาสพบและไม่พบข้อมูลในเว็บแคช

ตารางที่ 4.2 ค่าทางสถิติของความลึกของเส้นทาง

ค่าทางสถิติ	ประเภท	ค่าขอทั้งหมด	พบข้อมูลในแคช	ไม่พบข้อมูลในแคช	มีข้อผิดพลาด
การวัดแนวโน้มเข้าสู่ส่วนกลาง	ค่าเฉลี่ยเลขคณิต	4.72	4.67	4.98	4.77
	ค่ามัธยฐาน	4	4	4	4
	ค่าฐานนิยม	4	4	4	4
การวัดการกระจาย	พิสัย	448	45	446	318
	ส่วนเบี่ยงเบนควอไทล์	0.50	0.50	1.00	0.50
	ส่วนเบี่ยงเบนมาตรฐาน	1.67	1.39	12.07	1.72
ค่าคุณสมบัติอื่น ๆ	ค่าต่ำสุด	1	1	3	1
	ค่าสูงสุด	449	46	449	319

4.1.3 ความสัมพันธ์ระหว่างนามสกุลเพิ่มข้อมูลกับการเก็บข้อมูลในเว็บแคช

เมื่อพิจารณาเฉพาะจำนวนค่าขอพบว่าสิบอันดับแรกของนามสกุลเพิ่มข้อมูลที่ปรากฏในจำนวนค่าขอทั้งหมดครอบคลุมถึงร้อยละ 93.91 และครอบคลุมค่าขอที่พบในแคชถึงร้อยละ 98.14 ดังตารางที่ 4.3 และเมื่อจัดกลุ่มของนามสกุลเพิ่มข้อมูลเป็นประเภทต่าง ๆ โดยแบ่งจากค่าร้อยละของค่าขอที่พบในแคชเทียบกับค่าขอทั้งหมดของนามสกุลเพิ่มข้อมูลประเภทนั้น ๆ เป็น สูง ปานกลาง และต่ำ โดยสูงคือพิจารณาว่าพบมากกว่าร้อยละ 50 ส่วนปานกลางคือค่ามากกว่าร้อยละ 10 ส่วนต่ำคือน้อยกว่าร้อยละ 10 จากข้อมูลของนามสกุลเพิ่มข้อมูล 10 อันดับแรกดังนี้

ประเภทของนามสกุลเพิ่มข้อมูล

ข้อมูลที่มีร้อยละของค่าขอที่พบในแคชสูง

ข้อมูลที่มีร้อยละของค่าขอที่พบในแคชปานกลาง

ข้อมูลที่มีร้อยละของค่าขอที่พบในแคชต่ำ

นามสกุลเพิ่มข้อมูล

gif และ jpg

htm html และ default file

cgi pl asp.shtml others และข้อมูลเพิ่มที่ไม่มีนามสกุล

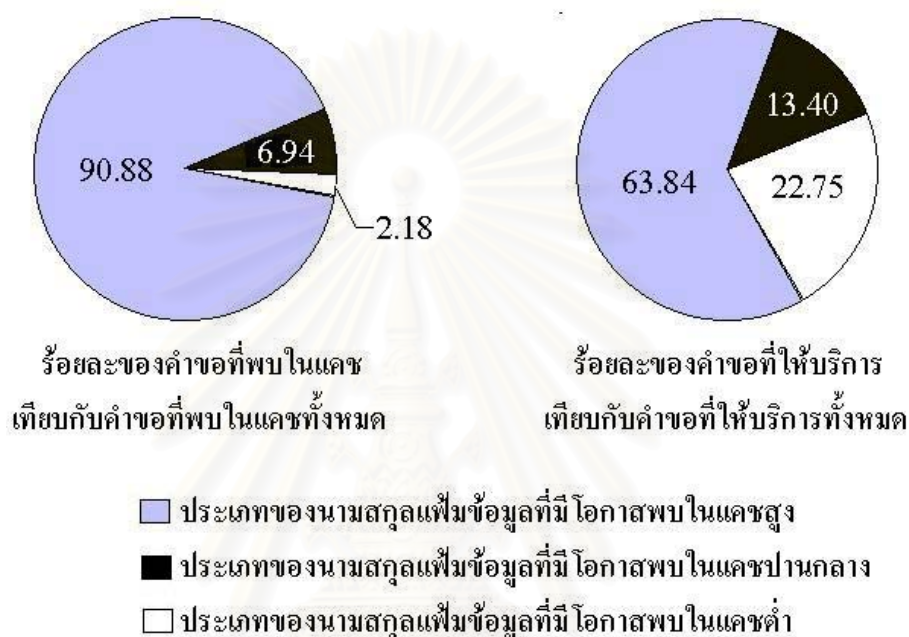
จะเห็นว่าแต่ละประเภทจะมีประเภทของแฟ้มข้อมูลที่เหมือนหรือคล้ายกัน เช่น ในประเภทแรกก็เป็นแฟ้มข้อมูลรูปภาพเหมือนกัน ส่วนประเภทที่สองส่วนใหญ่ก็เป็นแฟ้มข้อมูล ภาษาเซสที่เอ็มแอลเหมือนกัน

ตารางที่ 4.3 ร้อยละของคำขอแจกแจงตามนามสกุลแฟ้มข้อมูล

นามสกุล แฟ้มข้อมูล	ร้อยละของคำขอ เทียบกับทั้งหมด	ร้อยละของคำขอพบในแคช เทียบกับของนามสกุลนั้น	ร้อยละของคำขอพบ ในแคชเทียบกับทั้งหมด
Gif	49.24	71.72	75.28
Jpg	14.60	50.15	15.61
ไม่มีนามสกุล	9.49	1.10	0.22
Default File	6.27	19.65	2.63
Html	5.11	28.14	3.07
Pl	2.45	0.01	0.00
Cgi	2.31	1.21	0.06
Htm	2.02	29.02	1.25
Asp	1.25	0.004	0.00
Shtml	1.16	1.13	0.03
อื่น ๆ	6.09	0.01	1.86

เมื่อนำมาวาดเป็นกราฟตามประเภทที่ได้แจกแจงเอาไว้ในตอนต้น จะได้ดังรูปที่

4.7



รูปที่ 4.7 ร้อยละของจำนวนคำขอแยกตามประเภทนามสกุลเพิ่มข้อมูล

จากข้อมูลข้างต้นแสดงให้เห็นว่าคำขอที่มีเพิ่มข้อมูลนามสกุลเป็น gif กับ jpg ซึ่งมีอยู่ถึงร้อยละ 63.84 ของคำขอให้บริการทั้งหมดนั้นครอบคลุมจำนวนข้อมูลคำขอที่พบในแชทถึง 90.88 ของจำนวนคำขอที่พบในแชททั้งหมด และยังเป็นประเภทของนามสกุลเพิ่มข้อมูลที่มีโอกาสพบในแชทสูงด้วย

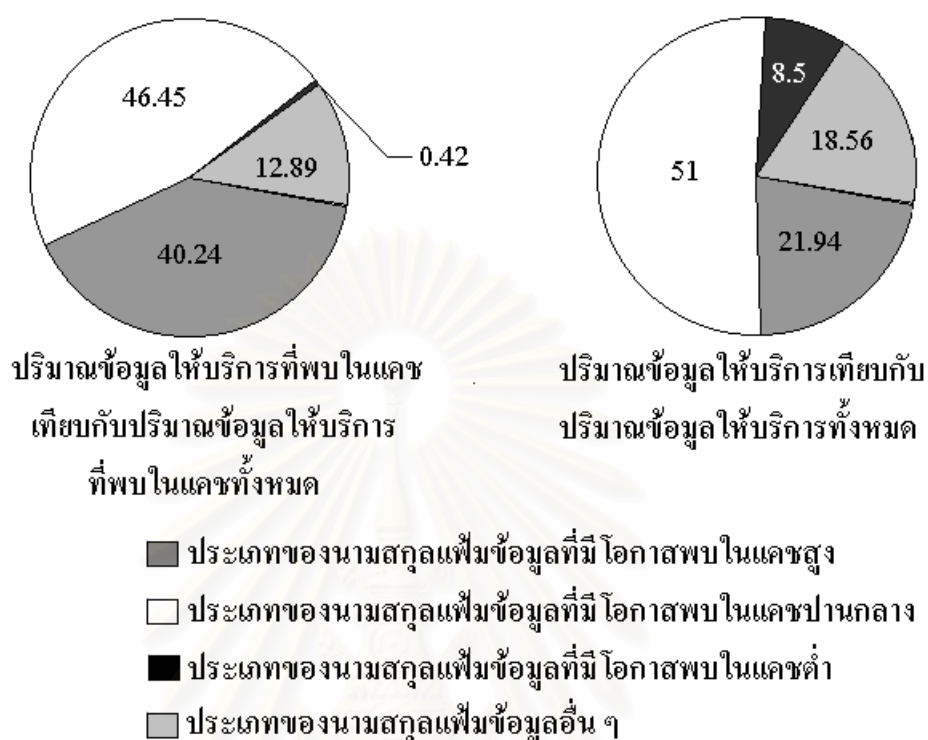
เมื่อพิจารณาที่ปริมาณข้อมูลให้บริการที่สิบอันดับแรกของปริมาณข้อมูลให้บริการทั้งหมดซึ่งครอบคลุมปริมาณข้อมูลร้อยละ 81.44 และครอบคลุมปริมาณข้อมูลที่พบในแชทร้อยละ 87.11 ดังตารางที่ 4.4 และเมื่อจัดกลุ่มของนามสกุลเพิ่มข้อมูลเป็นประเภทต่าง ๆ โดยแบ่งจากค่าร้อยละของปริมาณข้อมูลที่พบในแชทเทียบกับปริมาณข้อมูลให้บริการทั้งหมดของนามสกุลเพิ่มข้อมูลประเภทนั้น ๆ จากข้อมูลของนามสกุลเพิ่มข้อมูล 10 อันดับแรกจะแบ่งได้ 3 ประเภทเหมือนกับที่กล่าวไว้ใน การแบ่งประเภทของจำนวนคำขอ แต่พบว่านามสกุลเพิ่มข้อมูลประเภทอื่น ๆ มีปริมาณมาก ผู้ทำวิจัยจึงแยกออกมาต่างหาก

ประเภทของนามสกุลเพิ่มข้อมูล	นามสกุลเพิ่มข้อมูล
ข้อมูลที่มีร้อยละของคำขอที่พบในแคชสูง	gif
ข้อมูลที่มีร้อยละของคำขอที่พบในแคชปานกลาง	jpg, zip, exe, html, mpg, mp3 และ default file
ข้อมูลที่มีร้อยละของคำขอที่พบในแคชต่ำ	shtml และข้อมูลเพิ่มที่ไม่มีนามสกุล
ข้อมูลอื่น ๆ	นามสกุลเพิ่มข้อมูลอื่น ๆ

จะเห็นว่าการกระจายของข้อมูลจะแตกต่างจากเมื่อพิจารณาจากจำนวนคำขอ โดยในข้อมูลประเภทที่สอง ประกอบไปด้วยประเภทนามสกุลเพิ่มข้อมูลแตกต่างกันแต่มีโอกาสเป็นข้อมูลที่พบในแคชได้ใกล้เคียงกันเมื่อเทียบกับข้อมูลให้บริการที่เป็นเพิ่มข้อมูลประเภทนั้น ๆ ตารางที่ 4.4 ร้อยละของปริมาณข้อมูลให้บริการแจกแจงตามนามสกุลเพิ่มข้อมูล

นามสกุลเพิ่มข้อมูล	ร้อยละปริมาณข้อมูลเทียบกับทั้งหมด	ร้อยละปริมาณข้อมูลพบในแคชเทียบกับของนามสกุลนั้น	ร้อยละปริมาณข้อมูลพบในแคชเทียบกับทั้งหมด
jpg	22.49	35.90	22.65
gif	21.94	65.40	40.24
Default	6.05	21.87	3.71
zip	6.03	30.84	5.21
NoExt	5.77	1.35	0.22
exe	5.41	35.53	5.39
html	4.90	18.51	2.54
mpg	4.21	43.68	5.16
shtml	2.73	2.66	0.20
mp3	1.91	33.44	1.79
Others	18.56	24.76	12.89

เมื่อนำมาวาดเป็นกราฟตามประเภทที่แบ่งเอาไว้จะได้ดังรูปที่ 4.8



รูปที่ 4.8 ร้อยละของปริมาณข้อมูลให้บริการแยกตามประเภทนามสกุลเพิ่มข้อมูล

4.2 ข้อสมมติฐานจากงานวิจัย

เมื่อพิจารณาข้อมูลที่ได้จากการวิจัยและนำมาตั้งสมมติฐานเกี่ยวกับการเก็บข้อมูลลงบนเว็บแคช พบรูปแบบที่น่าสนใจเกี่ยวกับความยาว และนามสกุลเพิ่มข้อมูลของยูอาร์แอล ดังต่อไปนี้

4.2.1 การเก็บข้อมูลลงบนเว็บแคชโดยพิจารณาความยาวของยูอาร์แอล

พบว่าที่ยูอาร์แอลยาวกว่าค่าเฉลี่ยข้อมูลส่วนใหญ่ที่ให้บริการเป็นข้อมูลที่ไม่พบในแคช ดังนั้น ถ้าเราตั้งสมมติฐานว่า ถ้าปรับปรุงการออกแบบเว็บแคชเพิ่มเติม โดยออกแบบเว็บแคชให้มีขั้นตอนวิธีในการแทนที่ข้อมูลในแคชแบบเดิมแต่เลือกการเก็บข้อมูลโดยพิจารณาจากความยาวของยูอาร์แอลแทนที่จะเก็บข้อมูลทั้งหมดลงบนตัวเว็บแคชและข้อมูลที่เลือกที่จะไม่เก็บลงบนตัวเว็บแคชควรเป็นข้อมูลที่มีโอกาสถูกเรียกข้าน้อยมาก ก็จะช่วยลดปริมาณข้อมูลที่เว็บแคชจะต้องเก็บลงจากเดิม และลดการทำงานของเว็บแคชลง โดยถ้ามีค่าขอใดที่เราเลือกที่จะไม่เก็บลงบนตัวเว็บแคช ก็จะมีการส่งค่าขอนั้นไปยังเครื่องให้บริการข้อมูลเว็บ เมื่อได้รับข้อมูลที่ต้องการก็ส่ง

ให้กับผู้ใช้บริการโดยไม่ต้องเก็บข้อมูลนั้นลงบนแคช หรืออาจให้เว็บแคชปรับให้คำขอที่ส่งไปแจ้งให้เครื่องให้บริการข้อมูลเว็บส่งกลับไปยังผู้ใช้บริการโดยตรงเลยก็ได้

จากข้อสันนิษฐานข้างต้น ถ้าในการพิจารณาความยาวของยูอาร์แอลเริ่มตั้งแต่ที่ความยาวต่ำสุดเป็นเกณฑ์และคำนวณเป็นแบบข้อมูลสะสมและพิจารณาความครอบคลุมข้อมูลคำขอที่พบในเว็บแคชที่ร้อยละ 70, 80 และ 90 ขึ้นไปดังตารางที่ 4.5 และ 4.6 พบว่าถ้าพิจารณาที่ปริมาณข้อมูลให้บริการครอบคลุมมากกว่าร้อยละ 80 ของข้อมูลให้บริการทั้งหมดซึ่งต้องออกแบบให้มีการเก็บข้อมูลเฉพาะที่ความยาวยูอาร์แอลไม่เกิน 61 จะสามารถลดปริมาณข้อมูลที่ต้องจัดเก็บลงเว็บแคชได้ร้อยละ 26.63 และลดจำนวนคำขอที่ต้องจัดการในการให้บริการร้อยละ 22.09 นอกจากนี้ ในการใช้งานจริงเมื่อปริมาณข้อมูลที่ต้องจัดเก็บลงเว็บแคชลดลงก็จะทำให้ตัวเว็บแคชมีเนื้อที่ว่างสำหรับเก็บข้อมูลของคำขออื่น ๆ หรือข้อมูลคำขอบางคำขอซึ่งอาจถูกเรียกได้อีกภายหลังก็ไม่ต้องถูกแทนที่ ทำให้ในภาพรวมของเว็บแคชหลังจากการปรับปรุงกระบวนการเก็บข้อมูลอาจมีโอกาสที่พบข้อมูลในแคชเพิ่มขึ้นจากเดิมก็เป็นได้

ตารางที่ 4.5 ร้อยละของคำขอเมื่อลดคำขอที่เก็บลงเว็บแคชโดยพิจารณาจากช่วงความยาวของยูอาร์แอลที่ค่าต่าง ๆ

ความยาวยูอาร์แอลมากที่สุด	ร้อยละของคำขอที่มีข้อมูลมีอยู่ในเว็บแคช	ร้อยละของคำขอทั้งหมด	ร้อยละของคำขอที่ลดลง
52	72.76	60.58	39.42
56	81.03	68.92	31.08
61	88.92	77.91	22.09
62	90.27	79.39	20.61

ตารางที่ 4.6 ร้อยละของข้อมูลที่จัดเก็บเมื่อจำกัดคำขอที่เก็บลงเว็บแคชโดยพิจารณาจากช่วงความยาวของยูอาร์แอลที่ค่าต่าง ๆ

ความยาวยูอาร์แอลมากที่สุด	ร้อยละของปริมาณข้อมูลให้บริการที่มีอยู่ในเว็บแคช	ร้อยละปริมาณข้อมูลให้บริการทั้งหมด	ร้อยละของข้อมูลบันทึกที่ลดลง
55	70.58	62.01	37.99
61	81.36	73.37	26.63
68	91.13	83.41	16.59

4.2.2 การเก็บข้อมูลลงบนเว็บแคชโดยพิจารณานามสกุลเพิ่มข้อมูลของยูอาร์แอล

จากข้อมูลผลการวิจัยซึ่งแบ่งประเภทของปริมาณข้อมูลให้บริการออกเป็น 4 ประเภท ถ้าเราตั้งสมมติฐานว่า ถ้าปรับปรุงการออกแบบเว็บแคชเพิ่มเติม โดยออกแบบเว็บแคชให้มีขั้นตอนวิธีในการแทนที่ข้อมูลในแคชแบบเดิมแต่เลือกการเก็บข้อมูลโดยพิจารณาจากนามสกุลเพิ่มข้อมูลที่มีโอกาสพบในแคชสูงและปานกลาง ประกอบไปด้วยเพิ่มนามสกุล gif และ jpg (เป็นรูปภาพ), zip (เป็นข้อมูลบีบอัด), exe (เป็นข้อมูลทำงาน), html (เป็นข้อมูลภาษาเอชทีเอ็มแอล), mpg และ mp3 (เป็นข้อมูลสื่อประสม) และ Default File (ส่วนใหญ่เป็นข้อมูลภาษาเอชทีเอ็มแอล) จากตารางที่ 4.4 และรูปที่ 4.6 ข้อมูลที่เลือกเก็บจะครอบคลุมปริมาณค่าขอถึงร้อยละ 86.69 ของปริมาณข้อมูลให้บริการที่พบในแคชและลดปริมาณข้อมูลที่ต้องจัดเก็บในแคชลงถึงร้อยละ 27.06



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 5

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

5.1 สรุปผลการวิจัย

จากคุณสมบัติทั้ง 3 ข้อของยูอาร์แอล ซึ่งได้แก่ ความยาว ความลึกของเส้นทาง และนามสกุลเพิ่มข้อมูล เมื่อพิจารณาจากข้อมูลที่ได้จากการวิเคราะห์การใช้งานบริการเว็บของ จุฬาลงกรณ์มหาวิทยาลัยในช่วงเดือนสิงหาคมปี 2543 พบว่า มีรูปแบบเป็นไปดังต่อไปนี้

- 1) คำขอที่ยูอาร์แอลมีความยาวน้อย ๆ มีโอกาสที่จะพบข้อมูลในแคชสูงกว่าคำขอที่ยูอาร์แอลมีความยาวมาก ๆ โดยความยาวพิจารณาเทียบกับความยาวเฉลี่ยของคำขอทั้งหมด
- 2) ความลึกเส้นทางของยูอาร์แอลของคำขอที่ให้บริการซึ่งคำขอที่พบข้อมูลในแคช และคำขอที่ไม่พบข้อมูลในแคช มีค่าเฉลี่ยใกล้เคียงกันมาก ดังนั้น ความลึกเส้นทางของยูอาร์แอล อาจเป็นคุณสมบัติที่ไม่ชัดเจนพอที่จะมีผลต่อการทำนายประเภทของยูอาร์แอลว่ามีโอกาสอยู่ในกลุ่มประเภทใดมากกว่าได้
- 3) นามสกุลเพิ่มข้อมูลเป็นคุณสมบัติที่มีรูปแบบค่อนข้างชัดเจนซึ่งสามารถแยกแยะได้ว่า กลุ่มของนามสกุลเพิ่มข้อมูลบางประเภทมีโอกาสพบข้อมูลในแคชได้มากกว่าประเภทอื่น ซึ่งหมายถึงทั้งจำนวนคำขอและปริมาณข้อมูลให้บริการ ได้แก่ เพิ่มข้อมูลประเภทรูปภาพ (gif และ jpg) และเพิ่มข้อมูลที่เป็นข้อมูลภาษาเอชทีเอ็มแอล (htm, html และ default file) กลุ่มของนามสกุลเพิ่มข้อมูลบางประเภทมีปริมาณข้อมูลให้บริการที่พบในแคชมากเป็นอันดับต้น ๆ แต่เมื่อพิจารณาที่จำนวนคำขอที่ 10 อันดับแรกไม่พบข้อมูลประเภทนั้น เนื่องจากข้อมูลส่วนใหญ่มีขนาดใหญ่เมื่อเทียบกับข้อมูลประเภทอื่น ข้อมูลประเภทนั้นได้แก่ข้อมูลที่เป็นสื่อประสม (mpg และ mp3)

5.2 อภิปรายผลการวิจัย

ผู้ทำวิจัยได้พบปัญหาในการทำงานวิจัยอยู่ 2 ประการหลัก ๆ คือ

5.2.1 การออกแบบเครื่องมือในการวัดความลึกเส้นทางของยูอาร์แอล

เนื่องจากในข้อกำหนดของรีเลย์ที่พยูอาร์แอลนั้นสามารถทำให้ผู้พัฒนาเว็บไซต์เขียนอ้างอิงยูอาร์แอลเป็นแบบรีเลย์ที่พบกับตัวฐานได้ซึ่งในการเขียนอ้างอิงถึงเส้นทางอื่นที่ไม่ใช่เส้นทางเดียวกับตัวฐานถ้าผู้พัฒนาเว็บไซต์เขียนผิดก็ทำให้ตัวแจกจ่ายค่าผิดและส่งค่ายูอาร์แอลนั้นออกไปยังเครื่องให้บริการข้อมูลเว็บ ซึ่งในที่นี้ต้องผ่านตัวเว็บแคชทำให้ตัวเว็บแคชมีการบันทึกค่าของยูอาร์แอลที่ผิดนี้ไปด้วย แต่จากการทดลองส่งค่าขอที่มีความผิดพลาดนี้ไปยังเครื่องให้บริการข้อมูลเว็บพบว่าสามารถรองรับค่าขอบริการเว็บที่ผิดนี้ได้ โดยละเอียดในส่วนที่ผิดนั้นไป แต่กลุ่มตัวอักขระที่เอาไว้อ้างอิงถึงเส้นทางอื่นที่ไม่ใช่ฐานนั้นมีตัวอักขระ “/” อยู่ด้วยซึ่งในงานวิจัยนี้ใช้เป็นตัวในการแยกความลึกของเส้นทาง ดังนั้น ข้อมูลจำนวนความลึกเส้นทางของยูอาร์แอลอาจมีข้อผิดพลาดอยู่บ้าง ซึ่งผู้ทำวิจัยได้เขียนโปรแกรมเพื่อนับจำนวนของค่าขอที่มีข้อมูลประเภทนี้อยู่ในยูอาร์แอลเอาไว้ด้วย

ตัวอย่างของยูอาร์แอลที่มีความผิดพลาดในส่วนของการอ้างอิงของเส้นทาง

ยูอาร์แอลฐาน	http://www.example.com/first/second/third/index.html
การเขียนอ้างอิง	../../../../test/index.html
ยูอาร์แอลที่เรียก	http://www.example.com/../../../../test/index.html
ยูอาร์แอลที่ให้บริการ	http://www.example.com/test/index.html

จากตารางที่ 5.1 เมื่อนับจำนวนของกลุ่มอักขระ “../” ที่มีอยู่ในยูอาร์แอลของคำขอทั้งหมดแล้วพบว่าเมื่อรวมทั้งหมด 13,046 คำขอ คิดเป็นร้อยละ 0.02 ของคำขอทั้งหมดซึ่งถือว่าน้อยมากผู้ทำวิจัยจึงไม่ได้ตัดคำขอกลุ่มนี้ไปรวมกับคำขอที่มีข้อผิดพลาด

ตารางที่ 5.1 จำนวนคำขอที่มีกลุ่มตัวอักษร “../” อยู่ในยูอาร์แอล

จำนวน “../”	จำนวนของคำขอให้บริการ
ไม่มี	63,786,211
1	10,284
2	2,236
3	415
4	62
5	46
6	0
7	1
8	1
9	1

นอกจากนี้ผู้ทำวิจัยยังสังเกตพบคำขอบางคำขอที่มียูอาร์แอลอื่นอยู่ในส่วนของเส้นทางด้วย เช่น

<http://info.netscape.com/fwd/ntserch/http://www.vjf.com>

<http://c1.thecounter.com/id=81474&size=1024&colors=16&referer=http%3A//www.elsa.com/&java=true>

จากตัวอย่างทั้งสองจะเห็นว่ามีการใส่พารามิเตอร์ลงไปในตัวยูอาร์แอลโดยไม่ได้ใช้ตัวอักษรที่เอาไว้แยกส่วนพารามิเตอร์ตามข้อกำหนด ซึ่งจะทำให้เมื่อทำการนับเส้นทางของยูอาร์แอลมีข้อผิดพลาดได้เช่นกัน แต่ข้อผิดพลาดประเภทนี้จะแยกแยะได้ยากเนื่องจากผู้พัฒนาเว็บไซต์ไม่ได้ใช้รูปแบบเช่นนี้ทั้งหมด จากการสังเกตพบว่าผู้พัฒนาเว็บไซต์มีการอ้างอิงโดยใช้คำหลักเช่น referrer หรือ url โดยอาจมีส่วนของ http:// หรือ http%3A// หรือไม่มีก็ได้ ดังนั้นผู้ทำการวิจัยจึงไม่สามารถแยกแยะจำนวนของคำขอประเภทนี้ออกมาได้ทั้งหมดซึ่งทำให้ผลการวิจัยมีความคลาดเคลื่อนไปบ้าง

5.2.2 การออกแบบเครื่องมือหาคำถามสกุลเพิ่มข้อมูลของยูอาร์แอล

ในการออกแบบเครื่องมือหาคำถามสกุลเพิ่มข้อมูลนั้นจะมีวิธีในการตัดส่วนอื่น ๆ ที่ไม่ใช่เส้นทางของยูอาร์แอลเหมือนกับการนับความลึกเส้นทางของยูอาร์แอล ดังนั้นจากตัวอย่างใน

ปัญหาของการนับความลึกเส้นทางของยูอาร์แอลซึ่งมีข้อมูลพารามิเตอร์บางอย่างปนอยู่ในส่วนของเส้นทาง ทำให้เมื่อมีการหาส่วนที่เป็นนามสกุลของแฟ้มข้อมูลมีความผิดพลาดได้เช่นกันถ้าพารามิเตอร์นั้น ๆ มีตัวอักษรซึ่งเป็นตัวบ่งชี้ส่วนของนามสกุลไฟล์ หรือตัว "." อยู่ด้วย

นอกจากนี้ในช่วงแรกของการสร้างเครื่องมือที่ใช้ในการวิจัย ผู้ทำวิจัยพบปัญหาวานามสกุลแฟ้มข้อมูลมีทั้งตัวอักษรแบบเล็กและตัวอักษรแบบใหญ่ ซึ่งในระบบยูนิคซ์จะถือว่าแตกต่างกัน แต่ในระบบปฏิบัติการวินโดวส์จะมองว่าไม่แตกต่างกัน และผู้ทำวิจัยทำการวิเคราะห์บนระบบยูนิคซ์ ดังนั้น ในโปรแกรมที่ทำการวิเคราะห์จะต้องมีกระบวนการแปลงตัวอักษรแบบใหญ่ให้เป็นแบบเล็กทั้งหมด เนื่องจากในการวิเคราะห์จะเป็นการพิจารณาประเภทของนามสกุลแฟ้มข้อมูลเท่านั้น เพราะฉะนั้นนามสกุลแฟ้มข้อมูลเดียวกันที่เป็นตัวอักษรแบบเล็กและแบบใหญ่จะถือเป็นประเภทเดียวกัน

5.3 ข้อเสนอแนะ

ในงานวิจัยนี้เป็นงานวิจัยทางด้านวิเคราะห์เท่านั้น จึงมีเพียงการตั้งสมมติฐาน ดังนั้นงานส่วนที่อาจทำต่อไปได้คือนำสมมติฐานที่ตั้งไว้มาทำการทดลองจริง ๆ ซึ่งได้แก่

- 1) ทำการปรับปรุงโครงสร้างของเว็บแคชให้มีการเก็บข้อมูลโดยพิจารณาความยาวของยูอาร์แอลก่อนทำการจัดเก็บอาจเลือกค่าช่วงความยาวต่าง ๆ กันที่จะจัดเก็บข้อมูลแล้วทดลองเปรียบเทียบกับเว็บแคชที่ไม่ได้ทำการปรับปรุงโครงสร้างว่าช่วยเพิ่มประสิทธิภาพของเว็บแคชให้ดีขึ้นได้หรือไม่
- 2) ทำการปรับปรุงโครงสร้างของเว็บแคชโดยให้มีการจัดเก็บข้อมูลเฉพาะข้อมูลที่มีนามสกุลแฟ้มข้อมูลบางประเภท เช่นทำการจัดเก็บเฉพาะข้อมูลที่เป็นรูปภาพ (gif และ jpg) ภาษาเอชทีเอ็มแอล (htm, html และ Default file) และข้อมูลสื่อประสม (mpg และ mp3) ซึ่งเมื่อเราทราบประเภทของข้อมูลก็จะทำเลือกวิธีการจัดเก็บที่เหมาะสมกับข้อมูลประเภทนั้น ๆ และทำการทดลองเปรียบเทียบกับเว็บแคชที่ยังไม่ได้ปรับปรุงโครงสร้างว่าช่วยเพิ่มประสิทธิภาพได้หรือไม่
- 3) นำข้อมูลที่ได้จากงานวิจัยไปทำการออกแบบเครื่องมือในการคิดคำนวณค่าใช้จ่ายในการใช้งานบริการเว็บแคชดังงานวิจัยของ K.G. Anagnostakis และ F.C.Harmantzis [11] โดยข้อมูลจากงานวิจัยสามารถช่วยในการทำนายว่าข้อมูลที่จะนำมาคิดค่าใช้จ่ายนั้นเป็นข้อมูลที่มีโอกาสพบในแคชหรือไม่

รายการอ้างอิง

1. Carlos R. Cunha, Azer Bestavros and Mark E. Crovella. Characteristics of WWW Client-Based Traces. Technical Report TR-95-010, Boston University Computer Science Department, June 1995.
2. Martin Arlitt, Ludmila Cherkasova, John Dilley, Rich Friedrich and Tai Jin. Evaluation Content Management Techniques for Web Proxy Caches. HP Labs Technical Reports, 1999.
3. T. Berners-Lee, L. Masinter and M. McCahill. Uniform Resource Locators (URL). RFC1738, December 1994.
4. R. Fielding. Relative Uniform Resource Locators. RFC1808, June 1995.
5. P. Hoffman, L. Masinter and J. Zawinski. The mailto URL scheme. RFC2368, July 1998.
6. บุญชม ศรีสะอาด. วิธีการทางสถิติสำหรับการวิจัย. เล่มที่ 1. พิมพ์ครั้งที่ 2. มหาสารคาม : ภาควิชาพื้นฐานของการศึกษาคณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ มหาสารคาม, 2541.
7. Yumiko Seki, Norifumi Nishikawa and Hiroshi Tsuji. User's Behavior and URL Analysis at EC Sites. IEEE, 1999.
8. Philippe Rochat and Stuart Thompson. Proxy Caching Based on Object Location Considering Semantic Usage. The 4th International Web Caching Workshop by NLANR and CAIDA, San Diego, California, March 31 - April 2, 1999.
9. Lara D. and James E. Pitkow. Characterizing Browsing Strategies in the World-Wide Web. Computer Networks and ISDN Systems, 1995.
10. Martin F. Arlitt and Carey L. Williamson. Web Server Workload Characterization: The Search for Invariants. Proceedings of the 1996 ACM SIGMETRICS Conference Philadelphia, PA, May 1996.
11. K.G. Anagnostakis and F.C.Harmantzis. Simple Usage-Based Charging of Web Cache Services. NLANR Web Cache Workshop 97, Boulder, Colorado, USA, 9-10 June 1997.

12. ข้อมูลทางบรรณานุกรมของหอสมุดแห่งชาติ ราชบัณฑิตยสถาน. ศัพท์เทคโนโลยีสารสนเทศ. พิมพ์ครั้งที่ 1. กรุงเทพมหานคร : ราชบัณฑิตยสถาน, 2542.
13. Fred Douglis, Anja Feldmann, Balachander Krishnamurthy and Jeffrey Mogul. Rate of Change and other Metrics: a Live Study of the World Wide Web. USENIX Symposium on Internetworking Technologies and Systems, December 1997.
14. Kc Claffy. Internet Measurement and Data Analysis: Topology, Workload, Performance and Routing Statistics, NAE '99 Workshop of cooperative Association for Internet Data Analysis (CAIDA), 1999.
15. Mark E. Crovella and Azer Bestavros. Self-Similarity in World Wide Web Traffic Evidence and Possible Causes. Proceedings of the 1996 ACM SIGMETRICS SIGMETRICS Conference Philadelphia, PA, May 1996.
16. ยุทธพงษ์ กัยวรรณ . พื้นฐานการวิจัย. พิมพ์ครั้งที่ 1. กรุงเทพมหานคร : สุวีริยาสาส์น, 2543.



ภาคผนวก

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ก

ตัวอย่างของข้อมูลที่ถูกตัดออกก่อนทำการวิเคราะห์

ข้อมูลที่ไม่มีข้อมูลของยูอาร์แอล

934113129.416 116 161.200.97.205 TCP_UNKOWN/0 0 - - DIRECT/std.cpc.ku.ac.th text/*

ข้อมูลที่มีจำนวนข้อมูลไม่เท่ากับ 10 เขตข้อมูล

934172270.599 6316 161.200.152.165 TCP_MISS/302 399 -

http://www.iconbazaar.com/animated/carban0 .gif - DIRECT/www.iconbazaar.com text/*

ข้อมูลที่มีข้อมูลของยูอาร์แอลไม่ได้ขึ้นต้นด้วย "http://"

935848933.849 3166 161.200.129.178 TCP_UNKNOWN/0 327 GET

HTT202.77.223.12/1.0 - DIRECT/202.77.223.12 text/*

ข้อมูลที่มีข้อมูลของปริมาณข้อมูลที่ร้องขอน้อยกว่า 0

933490701.783 74083 161.200.83.171 TCP_REFRESH_HIT/304 -1 -

http://www.kmitl.ac.th/science/Home.html - DIRECT/www.kmitl.ac.th text/*

ข้อมูลที่มีความยาวของยูอาร์แอลขนาดน้อยที่สุด 5 อันดับ

ความยาว	รายละเอียดข้อมูลคำขอ
7	935408806.733 0 161.200.129.143 TCP_HIT/0 0 GET http:// - NONE/- text/*
8	935455732.949 18666 161.200.129.15 TCP_REFRESH_MISS/0 36814 GET http:/// - DIRECT/ text/*
9	934188867.516 48033 161.200.154.42 TCP_MISS/0 0 - http://3/ - DIRECT/3 text/*
10	934195052.216 49000 161.200.122.138 TCP_MISS/0 0 - http://12/ - DIRECT/12 text/*
11	934278059.683 74733 161.200.122.138 TCP_MISS/0 0 - http://120/ - DIRECT/12 text/*

ตัวอย่างยูอาร์แอลที่มีความลึกเส้นทางของยูอาร์แอลมาก ๆ

ตัวอย่างต่อไปนี้เป็นตัวอย่างยูอาร์แอลที่มีความลึกเส้นทางของยูอาร์แอลเท่ากับ 283 และมีความยาวของยูอาร์แอลเป็น 2,470

at/6c7a1da07/fr/freehosting8-at/401b73671/fr/freehosting8-at/3bf87822d/fr/freehosting8-at/037672d95/fr/freehosting8-at/749ae103e/fr/freehosting8-at/037672d65/fr/freehosting8-at/8924ada3b/fr/freehosting8-at/749ae100e/fr/freehosting8-at/037672d25/fr/freehosting8-at/037672d15/fr/freehosting7-at/749ae105e/se/sexycute-web1000/banner.jpg



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ข

คำจำกัดความที่ใช้ในการวิจัย

คำศัพท์	คำจำกัดความ
เว็บแคช	อุปกรณ์ซึ่งทำหน้าที่บริหารและให้บริการข้อมูลเว็บต่าง ๆ ที่ผู้ใช้บริการเว็บเรียกใช้ผ่านตัวมันแทนเครื่องให้บริการข้อมูลเว็บจริง ๆ
เครื่องให้บริการข้อมูลเว็บ	ในงานวิจัยนี้หมายถึงอุปกรณ์ที่ทำหน้าที่ให้บริการข้อมูลเว็บที่มีอยู่ในตัวมันกับผู้ที่ส่งคำขอข้อมูลนั้น ๆ
ผู้ใช้บริการ	ในงานวิจัยนี้หมายถึงผู้ใช้บริการเว็บโดยส่งคำขอใช้งานผ่านโปรแกรมซึ่งทำหน้าที่เป็นตัวแจกข้อมูลที่ได้มาแสดงผลให้กับผู้ใช้งานซึ่งโดยปกติในคำขอที่ส่งไปยังเครื่องให้บริการข้อมูลเว็บนั้นจะมีรายละเอียดหนึ่งส่งไปด้วยนั่นคือ ยูอาร์แอล
ยูอาร์แอล	กลุ่มของตัวอักขระซึ่งอ้างถึงทรัพยากรที่มีอยู่และสามารถเข้าถึงได้ผ่านอินเทอร์เน็ต
ความยาวยูอาร์แอล	จำนวนตัวอักขระทั้งหมดที่ประกอบเป็นยูอาร์แอลนั้น ๆ
ความลึกเส้นทาง	จำนวนของสารบทย่อยที่ยูอาร์แอลอ้างถึง ซึ่งเส้นทางในที่นี้จะมี ความหมายคล้ายคลึงกับเส้นทางในระบบแฟ้มข้อมูลที่ใช้กันทั่วไปในระบบยูนิคเพียงแต่ในระบบยูนิคจะนับจุดเริ่มต้นที่รากของแฟ้มข้อมูล แต่ในยูอาร์แอลจะนับจุดเริ่มต้นที่ชื่อเครื่องที่ให้บริการ
นามสกุลแฟ้มข้อมูล	ข้อมูลที่บอกประเภทของแฟ้มข้อมูล ซึ่งตัวโปรแกรมส่วนใหญ่ใช้อ้างอิงเพื่อที่จะสามารถทราบได้ว่าควรจัดการกับแฟ้มข้อมูลนี้อย่างไร
จำนวนข้อมูลร้องขอ/ จำนวนข้อมูลให้บริการ	จำนวนคำขอใช้บริการเว็บ/จำนวนคำขอทั้งหมดที่เว็บแคชให้บริการ มีหน่วยเป็นครั้งซึ่งในงานวิจัยนี้จะนับจากจำนวนบรรทัดของข้อมูลแฟ้มลงบันทึกการใช้บริการเว็บ
ปริมาณข้อมูลร้องขอ/ ปริมาณข้อมูลให้บริการ	ปริมาณข้อมูลที่เว็บแคชให้บริการ/ปริมาณข้อมูลทั้งหมดที่เว็บแคชให้บริการ มีหน่วยเป็นไบต์ซึ่งในงานวิจัยนี้จะคำนวณจากผลรวมของข้อมูลในแฟ้มลงบันทึกการใช้บริการเว็บ

ภาคผนวก ค

คำศัพท์ภาษาอังกฤษและคำแปลที่ใช้ในงานวิทยานิพนธ์ [12]

คำศัพท์	คำภาษาไทย
Algorithm	ขั้นตอนวิธี
Backbone	แกนหลัก
Bandwidth	แบนด์วิดท์
Base	ฐาน
Binary	ฐานสอง
Cache	แคช
Default value	ค่าโดยปริยาย
Directory	สารบบ
Electronic mail	ไปรษณีย์อิเล็กทรอนิกส์
File	แฟ้ม
Format	รูปแบบ
Function	ฟังก์ชัน
Gigabyte	กิกะไบต์
HTML(HyperText Markup Language)	(ภาษา)เฮกซ์เอ็มแอล
Internet	อินเทอร์เน็ต
Internet address	เลขที่อยู่อินเทอร์เน็ต
Keyword	คำหลัก
Log file	แฟ้มลงบันทึก
Multimedia	สื่อประสม
Parameter	พารามิเตอร์
Parser	ตัวแจง
Path	เส้นทาง
Program	โปรแกรม
Query	ข้อสอบถาม
Relative	รีเลทีฟ

คำศัพท์(ต่อ)	คำภาษาไทย(ต่อ)
Root	ราก
Software	ซอฟต์แวร์
String	อักขระ
Text	ข้อความ
Unix	ยูนิกซ์
Update	ปรับ(ให้เป็นปัจจุบัน)
URL	ยูอาร์แอล
Web server	เครื่องให้บริการข้อมูลเว็บ
Web site	เว็บไซต์

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวนพรัตน์ พรหมมหาไชย เกิดวันที่ 22 กรกฎาคม พ.ศ.2517 ที่อำเภอเมือง จังหวัดขอนแก่น สำเร็จการศึกษาปริญญาตรีวิศวกรรมศาสตรบัณฑิต ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ มหาวิทยาลัยขอนแก่น ในปีการศึกษา 2537 และเข้าศึกษาต่อในหลักสูตร วิทยาศาสตร์มหาบัณฑิต ที่จุฬาลงกรณ์มหาวิทยาลัย เมื่อ พ.ศ. 2541 ปัจจุบันเป็นพนักงานบริษัท จัสมินอินเทอร์เน็ต อำเภอปากเกร็ด จังหวัดนนทบุรี



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย