

จุฬาลงกรณ์มหาวิทยาลัย

ทุนวิจัย
กองทุนรัชดาภิเษกสมโภช

รายงานผลการวิจัย

การกำหนดรูปแบบบรรณานุกรมและการวางแนวทางการจัดเก็บแบบถาวรสำหรับข้อมูลต่อเนื่องที่ตัดมาเป็น
กลุ่มให้สืบค้นผ่านห้องสมุดอัตโนมัติ : กรณีศึกษาข้อมูลสถิติการใช้งานอินเทอร์เน็ต
ของจุฬาลงกรณ์มหาวิทยาลัย

โดย

อ. ดร. ยรรยง เต็งอำนวย

หัวหน้าโครงการวิจัย

สุภาพร ชัยธัมมะปกรณ์

ผู้วิจัย

นิทัศน์ สุทธิปราโมชานนท์

ผู้ช่วยวิจัย

พฤษภาคม 2545

ISBN 974-13-1875-8

กิตติกรรมประกาศ

การศึกษาวิจัยครั้งนี้ คณะผู้วิจัยขอขอบคุณ จุฬาลงกรณ์มหาวิทยาลัยที่ให้ทุนอุดหนุน การวิจัยซึ่งสนับสนุนงานวิจัยให้สำเร็จลุล่วงไปด้วยดี อ.ดร.ณัฐวุฒิ หนูไพโรจน์ ภาควิชาวิศวกรรมคอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย คุณชยา ลิ้มจิตติ สำนักเทคโนโลยีสารสนเทศ จุฬาลงกรณ์มหาวิทยาลัย ผู้ให้ความ ช่วยเหลือในการประสานงานกับแหล่งข้อมูล คุณภาณุพันธ์ สุวรรณมาตร และคุณภรศิษฐ์ อรัญสมาน ผู้เก็บ ข้อมูลสถิติการใช้อินเทอร์เน็ตและข้อมูลเว็บแคชในระยะแรก รวมทั้งฝ่ายระบบสารสนเทศ สถาบันวิทยบริการ ห้องปฏิบัติการวิศวกรรมระบบสารสนเทศ (Information System Engineering Laboratory-ISEL) ภาควิชา วิศวกรรมคอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย ที่อำนวยความสะดวกในด้านการเก็บข้อมูล สถานที่ อุปกรณ์ บุคลากร และผู้ที่ให้ความช่วยเหลืออีกหลายท่านทั้งในด้านข้อมูล และการอำนวยความสะดวกในด้านต่าง ๆ

ยรรยง เต็งอำนวยและคณะ

พฤษภาคม 2545



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ชื่อโครงการวิจัย	การกำหนดรูปแบบบรรณานุกรมและการวางแผนทางการจัดเก็บแบบถาวรสำหรับข้อมูลต่อเนื่องที่ตัดมาเป็นกลุ่มให้สืบค้นผ่านห้องสมุดอัตโนมัติ : กรณีศึกษาข้อมูลสถิติการใช้งานอินเทอร์เน็ตของจุฬาลงกรณ์มหาวิทยาลัย
ชื่อผู้วิจัย	อ.ดร. ยรรยง เต็งอำนาจ สุภาพร ชัยธัมมะปกรณ์ นิทัศน์ สุทธิปราโมชานนท์
เดือนและปีที่ทำวิจัยเสร็จ	พฤษภาคม 2545

บทคัดย่อ

ข้อมูลสถิติการใช้งานอินเทอร์เน็ตและเว็บแคชในจุฬาลงกรณ์มหาวิทยาลัยระหว่างเครือข่ายอินเทอร์เน็ตกับยูนิเน็ต ถือว่าเป็นสาระพินถิ่นซึ่งเป็นประโยชน์ต่อการวิเคราะห์ เพื่อสนับสนุนการขยายความต้องการการวางแผน และการวิจัยด้านการใช้งานอินเทอร์เน็ต แต่เนื่องจากมิได้มีการจัดเก็บอย่างเป็นระบบ ทำให้การจัดกระจายและสูญหายในที่สุด การศึกษาวิจัยนี้ จึงทำการวิเคราะห์และกำหนดรูปแบบบรรณานุกรมและแนวทางการจัดเก็บ เพื่อให้เป็นบันทึกถาวร สามารถสืบค้นผ่านระบบห้องสมุดอัตโนมัติ ในการวิจัยนี้ได้วิเคราะห์และคัดเลือกซีดี-อาร์เป็นสื่อพื้นฐานสำหรับการเก็บข้อมูล ข้อมูลสถิติต่อเนื่องจากเราเตอร์และเว็บเซิร์ฟเวอร์ถูกตัดเป็นส่วนที่มีขนาดสะดวกต่อการปฏิบัติงานแล้วผ่านการบีบอัดข้อมูลก่อนการบันทึก ได้ผลเป็นคอลเลกชันของซีดี-อาร์จำนวน 265 แผ่น และเซตข้อมูลบรรณานุกรมจำนวน 25 เซตข้อมูลหลักตามระบบมาร์ก เมตาตาตาของข้อมูลถูกบันทึกเป็นแฟ้มร่วมไปกับข้อมูลเพื่อเอื้อต่อการกำหนดระเบียบบรรณานุกรม การวิจัยอาศัยฝากคอลเลกชันในฐานะข้อมูลของระบบโปรแกรมอินโนแพคที่ใช้เป็นระบบห้องสมุดอัตโนมัติ และได้สร้างโปรแกรมเชื่อมต่อผ่านเว็บ เพื่อให้สามารถสืบค้นใช้งานข้อมูลสถิติได้ทั้งผ่านระบบอักขระและระบบเว็บ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Project Title	Design and Specification of Bibliographical Record Format and Storage Procedure for Preservation of Chunked Continuous Data Stream for Retrieval through Library Automation: A Case Study on Internet Traffic Statistics of Chulalongkorn University
Name of the Investigators	Yunyong Teng-amnuay, PH.D. Supaporn Chaithammapakorn Nitass Sutaveepramochanon
Year	2002

Abstract

Internet traffic data and Web cache data between Chulalongkorn University Network (CUNet) and University Network (UniNet) are useful local contents that support research and planning on the usage and management of the Internet gateway. The data was not well managed, causing damage such as misplaced and data lost. This research aims to analyze, design, and specify the procedure and the bibliographical tags for the storing of the data in archival form. The collection can be accessed through the library automation of Chulalongkorn University. CD-Rs are identified as appropriate media for the archive. The data stream from routers and web caches are cut into manageable-sized chunks and compressed. The result is three collections of 265 CD-Rs and 25 fields of bibliographical record based on the MARC system. Metadata files are attached to the chunks to assist the building of the bibliographical records. Existing INNOPAC database are used as the library automation. A web interface is added to extend the search and presentation capabilities of the INNOPAC and helps in the retrieval of bibliographical records.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

กิตติกรรมประกาศ	ii
บทคัดย่อภาษาไทย	iii
บทคัดย่อภาษาอังกฤษ	iv
บทที่ 1 บทนำ	1
ที่มาและความสำคัญของปัญหา	1
วัตถุประสงค์	4
ขอบเขตของการวิจัย	4
สถานที่ทำการวิจัยหรือเก็บข้อมูล	4
ผลที่จะได้รับการวิจัย	4
บทที่ 2 วรรณกรรมและงานวิจัยที่เกี่ยวข้อง	5
ปัญหาเกี่ยวกับบันทึกถาวร	5
สื่อที่ใช้บันทึก	6
เมตาดาตา	8
ลักษณะของข้อมูล	12
การใช้ประโยชน์ของข้อมูล	19
บทที่ 3 วิธีดำเนินการวิจัย	21
การวิเคราะห์และเลือกสื่อที่ใช้บันทึก	21
ระบบการบันทึกข้อมูลที่มีอยู่	28
การออกแบบการเก็บข้อมูลเพื่อให้จัดเก็บแบบถาวร	29
ระบบการบันทึกข้อมูลในแนวทางใหม่	30
การรวบรวมศึกษาและวิเคราะห์การลงเมตาดาตา	34
ทดสอบการบันทึกข้อมูล	47
การศึกษา ออกแบบและสร้าง Web interface	49
บทที่ 4 ผลการวิจัย	55
สื่อบันทึก	55
การตัดข้อมูลต่อเนื่องเป็น chunk	55
การบันทึกข้อมูลลงแผ่นบันทึกข้อมูล (ซีดี-อาร์)	56
เมตาดาตาหรือรูปแบบระเบียบบรรณานุกรม	57
ข้อมูลตัวอย่าง	58
ข้อมูลเชิงปริมาณ	61

สารบัญ (ต่อ)

บทที่ 5	สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ	66
	สรุปผลการวิจัย	66
	อภิปรายผลการวิจัย	66
	ข้อเสนอแนะที่ได้จากการวิจัย	68
	ข้อเสนอแนะเพื่องานวิจัยในอนาคต	69
	รายการอ้างอิง	71
	ประวัติผู้วิจัย	75
ภาคผนวก ก	คู่มือการปฏิบัติงานเมตาดาตาหรือชุดข้อมูลการลงรายการระเบียบบรรณานุกรมข้อมูล สถิติการใช้อินเตอร์เน็ตในจุฬาลงกรณ์มหาวิทยาลัย	1-7
ภาคผนวก ข	คู่มือการปฏิบัติงานการลงรายการระเบียบบรรณานุกรมข้อมูลสถิติการใช้อินเตอร์เน็ตใน จุฬาลงกรณ์มหาวิทยาลัย	1-7
ภาคผนวก ค	เมตาดาตาข้อมูลดิจิทัลของโครงการของห้องสมุดทั้งห้า	1-5
ภาคผนวก ง	คู่มือการปฏิบัติงานการบันทึกข้อมูลลงซีดี-อาร์	1-20
ภาคผนวก จ	Script สำหรับสนทนากับอินโนแพค	1-3



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญตาราง

ตารางที่ 1	เทคโนโลยีหลักของสื่อบันทึก	22
ตารางที่ 2	ผลิตภัณฑ์ของสื่อแต่ละประเภท	22
ตารางที่ 3	คุณสมบัติที่สำคัญของสื่อ	24
ตารางที่ 4	การเปรียบเทียบคุณลักษณะของสื่อ	25
ตารางที่ 5	พัฒนาการของระบบปฏิบัติการบนไมโครคอมพิวเตอร์	27
ตารางที่ 6	ขั้นตอนการจัดเก็บข้อมูลแทรฟฟิกตามที่เป็นอยู่เดิม	29
ตารางที่ 7	ขั้นตอนการจัดเก็บข้อมูลเว็บตามที่เป็นอยู่เดิม	29
ตารางที่ 8	ขั้นตอนการเก็บข้อมูลแทรฟฟิกแบบใหม่	34
ตารางที่ 9	ขั้นตอนการเก็บข้อมูลเว็บแบบใหม่	34
ตารางที่ 10	เปรียบเทียบความสอดคล้องของเมตาดาตาระหว่างโครงการต่าง ๆ	35
ตารางที่ 11	รายการเขตข้อมูลที่กำหนดในระเบียบบรรณานุกรม	38
ตารางที่ 12	การลงรายการแบบ bibliographical record of collection ของข้อมูลแทรฟฟิก ที่เชื่อมโยงกับบริษัทเทเลโกลป	42
ตารางที่ 13	เขตข้อมูลที่เพิ่มหรือเปลี่ยนแปลงการลงรายการแบบ bibliographical record of collection ของข้อมูลแทรฟฟิกเมื่อเลิกเชื่อมโยงกับบริษัทเทเลโกลป	44
ตารางที่ 14	การลงรายการแบบ bibliographical record of collection ของข้อมูลเว็บแคช	44
ตารางที่ 15	การลงรายการแบบ bibliographical record of conten analysis ของข้อมูล แทรฟฟิก	46
ตารางที่ 16	การสืบค้นที่เหมาะสมในระดับคอลเลกชัน	48
ตารางที่ 17	แสดงพัฒนาการของแผ่นสื่อบันทึกซีดี-อาร์	55

สารบัญภาพประกอบ

รูปที่ 1	ชื่อยูเน็ตต์เชื่อมต่อกับอินเทอร์เน็ตผ่านบริษัทเทเลโกลบ	13
รูปที่ 2	ชื่อยูเน็ตต์เชื่อมต่อกับอินเทอร์เน็ตผ่านยูนิเน็ต	14
รูปที่ 3	ตัวอย่างข้อมูลที่ถูกรวบรวมโดยโปรแกรม Netflowcollection	16
รูปที่ 4	ตัวอย่างข้อมูลของล็อกไฟล์ของ Proxy Cache Server	18
รูปที่ 5	ขั้นตอนการผลิตซีดี-อาร์ของบันทึกถาวร	31
รูปที่ 6	โครงสร้างของข้อมูลในแผ่นซีดี-อาร์	33
รูปที่ 7	เมตาดาตาของข้อมูลในซีดี-อาร์ที่บันทึกข้อมูลแทรฟฟิก หมายเลข CD TR 2000-001	37
รูปที่ 8	การลงรายการแบบหนึ่งแผ่นต่อหนึ่งระเบียบ	40
รูปที่ 9	การลงรายการแบบมีระเบียบหลักของคอลเลกชัน	41
รูปที่ 10	แสดงผลการสืบค้นจากชื่อเรื่อง	48
รูปที่ 11	การสนทนากับอินโนแพค	50
รูปที่ 12	ตัวอย่างบทในการสนทนากับอินโนแพค	50
รูปที่ 13	โครงสร้างฐานข้อมูลของการสนทนากับอินโนแพค	51
รูปที่ 14	รายการข้อมูลที่ต้องการจากอินโนแพค	51
รูปที่ 15	สรุปคุณภาพข้อมูลเป็นรายเดือน	52
รูปที่ 16	สรุปคุณภาพข้อมูลเป็นรายสัปดาห์	53
รูปที่ 17	สรุปข้อมูลในแต่ละ chunk	53
รูปที่ 18	แผ่นซีดี-อาร์ของข้อมูลแทรฟฟิกในปี 2000	62
รูปที่ 19	แผ่นซีดี-อาร์ของข้อมูลแทรฟฟิกในปี 2001	62
รูปที่ 20	แผ่นซีดี-อาร์ของข้อมูลแทรฟฟิกในปี 2002	63
รูปที่ 21	แผ่นซีดี-อาร์ของข้อมูลเว็บแคชในปี 2001	64
รูปที่ 22	แผ่นซีดี-อาร์ของข้อมูลเว็บแคชในปี 2002	64

บทที่ 1

บทนำ

ที่มาและความสำคัญของปัญหา

นับแต่ปลายศตวรรษที่ 20 เป็นต้นมา ระบบเครือข่ายสื่อสารข้อมูลสากลของโลกโดยเฉพาะอินเทอร์เน็ต ก่อให้เกิดปรากฏการณ์ใหม่ในวงการการผลิตเนื้อหาสาระในรูปของสิ่งพิมพ์ต่าง ๆ อย่างกว้างขวาง โดยเฉพาะ การส่งเสริมศักยภาพของผู้สร้างสรรค์ผลงานในการผลิตและจัดเก็บผลงานของตนในรูปแบบดิจิทัลและในการเผยแพร่ผลงานของตนในแบบอิเล็กทรอนิกส์ผ่านระบบเครือข่าย ซึ่งเป็นการข้ามขั้นตอนการเผยแพร่ผลงานโดยอาศัยองค์กรต่าง ๆ ในแนวทางปกติ คือ สำนักพิมพ์ บริษัทและร้านจัดจำหน่าย และห้องสมุด การละเว้นหรือหลีกเลี่ยงองค์กรตัวกลางในลักษณะนี้เป็นการเปิดศักราชใหม่ของการสร้างสรรค์ผลงาน ก่อให้เกิดความสะดวกรวมทั้งผู้ใช้ผลงานได้เนื้อหาสาระที่ทันสมัย รวดเร็ว เนื่องจากผลงานเหล่านั้นอยู่ภายใต้การควบคุมดูแลของผู้สร้างสรรค์สาระ สามารถแก้ไข ปรับเปลี่ยน เสริมแต่งได้ตลอดเวลา ไม่ต้องรอวัฏจักรที่ยืดยาวข้ามปีของระบบการพิมพ์และการจัดจำหน่ายหรือเผยแพร่ตามปกติ

แต่ประโยชน์ที่ได้รับจากความคล่องตัวของเครือข่ายกลับก่อให้เกิดปัญหาที่สำคัญยิ่งของยุคสมัย กล่าวคือ ผลงานและองค์ความรู้ของมวลมนุษยชาติซึ่งควรได้รับการจำแนก แยกแยะ จัดเก็บและทำนุบำรุง สำหรับการใช้ประโยชน์ทั้งในปัจจุบันและในอนาคต รวมทั้งการธำรงไว้ให้เป็นทรัพย์สินสมบัติร่วมกันในระยะยาว กลับตกเป็นภาระของผู้สร้างสรรค์ผลงาน ซึ่งไม่มีทั้งเป้าประสงค์ ความรู้ความสามารถหรือทรัพยากรที่เหมาะสมต่อการทำนุบำรุงทรัพยากรด้านสติปัญญาของมนุษย์ เกิดเป็นวิกฤตต่อองค์ความรู้ที่จะสูญหายไป [1] เมื่อตัวผู้ผลิตผลงานต้องมีภาระเพิ่มในการทำนุบำรุง และเนื่องจากเทคโนโลยีของระบบจัดเก็บข้อมูลในเชิงอิเล็กทรอนิกส์เปลี่ยนแปลงอย่างรวดเร็ว ข้อมูลที่เก็บไว้เองเกิดความเสียหาย ยากต่อการบำรุงรักษา [2] หรือเมื่อผู้สร้างสรรค์ผลงานหยุดให้บริการ เช่น บริษัทเลิกกิจการ หรือเครื่องคอมพิวเตอร์ของห้องปฏิบัติการทางวิชาการเกิดความเสียหายไม่มีงบประมาณมาทดแทน เป็นต้น

ท่ามกลางวิกฤตนี้ หน่วยงานที่เป็นสำนักพิมพ์ใหญ่ด้านวิชาการของโลกได้ดำเนินการแยกแยะและบำรุงรักษาเนื้อหาความรู้ในเชิงอิเล็กทรอนิกส์ในรูปแบบของฐานข้อมูลบนเซิร์ฟเวอร์ของบริษัท เช่น Institute of Electrical and Electronics Engineers (IEEE) Association for Computing Machinery (ACM) และบริษัท Springer-Verlag แต่กิจกรรมของหน่วยงานเหล่านี้ครอบคลุมเฉพาะเนื้อหาสาระที่ผ่านการคัดสรร มีความนิยม มีการเผยแพร่เป็นวงกว้าง และมีการเรียกเก็บค่าสมาชิกเพื่อเป็นค่าใช้จ่ายในอัตราที่สูง ซึ่งไม่ได้ครอบคลุมเนื้อหาโดยรวมหรือที่เป็นเฉพาะกิจที่มีการใช้งานต่ำ และไม่เอื้ออำนาจกิจกรรมของห้องสมุดซึ่งมีส่วนช่วยในการลดค่าใช้จ่ายแก่ผู้ใช้บริการและเป็นองค์กรที่ทำหน้าที่ทำนุบำรุงองค์ความรู้ในระยะยาวด้วย

จากสภาพการณ์เช่นนี้ สาระความรู้ และผลงานจำนวนมากตามองค์กรต่าง ๆ ที่ผลิตขึ้นโดยเฉพาะในลักษณะของสาระพื้นถิ่น (local content) โดยอาศัยเทคโนโลยีสมัยใหม่อย่างอินเทอร์เน็ต มีการเก็บไว้ในรูปของอิเล็กทรอนิกส์ และมีการเผยแพร่ผ่านเครือข่ายหรือเว็บไซต์ สาระเหล่านี้จะขาดการจำแนก แยกแยะ วิเคราะห์ จัดระเบียบ รวมไปถึงการบันทึกสิ่งที่มีคุณค่าทางถาวรและการเก็บรักษาสื่อเหล่านั้นให้อยู่ในสภาพแวดล้อม

ที่มีความเสถียร ตลอดจนการบำรุงรักษาทดแทนเทคโนโลยีของการบันทึกและการใช้สภาวะอิเล็กทรอนิกส์ที่เปลี่ยนแปลงไป

ในบรรดาสาระพินถิ่นเหล่านี้ โดยเฉพาะที่เกิดขึ้นในมหาวิทยาลัยและสถาบันอุดมศึกษา จะมีสาระอยู่ประเภทหนึ่ง ได้แก่ข้อมูลที่เป็นลักษณะของ data stream ซึ่งกำเนิดจากอุปกรณ์เครื่องวัดหรือจากบันทึกการทำงานของเครื่องมือต่าง ๆ ทั้งทางวิทยาศาสตร์และทางวิศวกรรมศาสตร์ ที่สามารถนำมาวิเคราะห์แจกแจงทางสถิติ ออกมาเป็นผลสรุปหรือรายงานที่เป็นประโยชน์ได้ในหลายลักษณะ แต่ข้อมูลเหล่านี้โดยปกติจะถูกเพิกเฉยหรือเก็บไว้เป็นช่วงเวลาสั้น ๆ จำนวนไม่มาก โดยหน่วยงานที่เป็นผู้ผลิตข้อมูลหรือจุดที่ข้อมูลถือกำเนิดขึ้น เนื่องจากข้อจำกัดในด้านทรัพยากร เช่น ที่ National Laboratory for Applied Network Research (NLANR) ที่ให้บริการเว็บแคชระดับโลกในสหรัฐอเมริกา [3] ทำให้ตกเป็นภาระของผู้นำข้อมูลไปใช้ ที่ต้องจัดเก็บข้อมูล เหล่านี้เอง เกิดความซ้ำซ้อนในการจัดเก็บ ข้อมูลที่จัดเก็บไม่ครบถ้วน เหล่านี้ขึ้นกับภารกิจหรือเป้าหมายของ ผู้ใช้ข้อมูล การจัดเก็บเหล่านี้ไม่ถาวรเนื่องจากกระบวนการจัดเก็บจะยุติลงเมื่อผู้ใช้เลิกใช้ข้อมูลนั้น อีกทั้ง ยังมีปัญหาอื่นตามมา เช่น หน่วยงานวิจัยนั้น ๆ ไม่ได้ให้บริการข้อมูลที่จัดเก็บแก่สาธารณะ หรือมีปัญหา เกี่ยวกับข้อมูลที่เป็นความลับ ซึ่งทำให้ยากต่อการบริหารและจัดการข้อมูลเหล่านั้น

ดังนั้น ในการวิจัยครั้งนี้ จึงจะใช้ศักยภาพของห้องสมุดและเทคโนโลยีของสื่อสมัยใหม่ ในการวางแผนทางเพื่อการเก็บรักษาสาระพินถิ่นที่เป็น data stream จำนวนมากอย่างเป็นระบบ โดยข้อมูลลักษณะนี้ จะปรากฏให้เห็นโดยทั่วไปในการวิจัยและทดลองทางวิทยาศาสตร์ต่าง ๆ ในลักษณะของการเฝ้าติดตามสภาพการณ์เป็นระยะยาว ในลักษณะของ data monitoring หรือ data capture เช่น การตรวจติดตามจากสัญญาณดาวเทียม ค่าระดับน้ำเพื่องานด้านอุทกศาสตร์ ปริมาณการจราจรที่ผ่านในถนนสายสำคัญของกรุงเทพฯ หรือ แม้แต่การเก็บค่าความเป็นกรดของสารละลายในห้องทดลอง เหล่านี้เป็นต้น โดยกรณีศึกษาของข้อมูลลักษณะ

ดังกล่าวที่ใช้ในงานวิจัยนี้ คือ ข้อมูลของปริมาณการเข้าใช้บริการอินเทอร์เน็ตของประชาคมจุฬาลงกรณ์มหาวิทยาลัย ซึ่งเป็นข้อมูลที่เข้ามาอย่างต่อเนื่องและมีปริมาณมาก ข้อมูลเหล่านี้หากผ่านการวิเคราะห์เป็นเวลานานพอสมควรจะเป็นประโยชน์ในการบริหารจัดการและการวิจัย หรือการวิเคราะห์พฤติกรรมการใช้งานอินเทอร์เน็ตในลักษณะต่าง ๆ ข้อมูลในลักษณะ data stream นี้ส่วนใหญ่มีจำนวนมากและเข้ามาอย่างต่อเนื่อง ในขณะที่เริ่มการทำวิจัยในครั้งนี้มีข้อมูลนี้อยู่เต็มแล้วขนาด 31.72 GB (1 GB เท่ากับ 1 พันล้านตัวอักษร) แต่เนื่องจากลักษณะการเก็บข้อมูลที่ผ่านมานั้นไม่มีระเบียบ ปัญหาที่เกิดขึ้นตามมาในภายหลัง คือ การกระจาย และการสูญหายของข้อมูลเหล่านี้ การจัดเก็บข้อมูลให้อยู่ในสภาพดีและสามารถใช้นั้นต้องคำนึงถึงค่าใช้จ่าย อายุของสื่อ ความล้าสมัยของสื่อ และความล้าสมัยของโปรแกรมที่ใช้เก็บและใช้อ่าน [4] ทั้งนี้จากกรณีศึกษาของ เจพีแอล (Jet Propulsion Laboratory- JPL) ขององค์การนาซ่า (National Aeronautic and Space Agency - NASA) พบว่ามีความเสียหายถึงร้อยละ 20 ต่อข้อมูลสำรวจดาวอังคารที่เก็บไว้ตั้งแต่ปี ค.ศ. 1976 และมหาวิทยาลัยเยลไม่สามารถอ่านข้อมูลหนังสือ 2,000 เล่มที่ถ่ายจากไมโครฟอร์มมิง optical disc [5]

การวิจัยในครั้งนี้ได้ตระหนักถึงการเปลี่ยนแปลงอันมีผลต่อความเสียหายของข้อมูลสถิติการใช้งานอินเทอร์เน็ตของจุฬาลงกรณ์มหาวิทยาลัย จึงได้พยายามศึกษาหาสื่อบันทึกที่เหมาะสมกับข้อมูลพร้อมทั้งการวางแผนในระยะยาวเพื่อการใช้งานและรองรับการเปลี่ยนแปลงของเทคโนโลยี

ในประเด็นของการนำข้อมูลการเข้าใช้อินเทอร์เน็ตในจุฬาลงกรณ์มหาวิทยาลัยไปใช้นั้น พบว่า มีการทำวิจัยโดยใช้ข้อมูลเหล่านี้อยู่จำนวนหนึ่ง ซึ่งผู้วิจัยต้องทำการเก็บข้อมูลเอง เนื่องจากโดยปกติผู้ดูแลเครือข่ายจะทิ้งข้อมูลเหล่านี้ไปเพราะไม่มีระบบจัดเก็บเพียงพอ เช่นเดียวกับกรณีของ NLANR เนื่องจากมีปริมาณข้อมูลมากแต่มีทรัพยากรจำกัด จึงไม่ทำการจัดเก็บข้อมูลเป็นภาระของนักวิจัยแต่ละรายที่ต้องขอถ่ายโอนข้อมูลมาเก็บไว้เอง [3] ดังนั้นข้อมูลในลักษณะแบบนี้จึงไม่มีการให้บริการอยู่ในจุฬาลงกรณ์มหาวิทยาลัย การดำเนินการเป็นการเฉพาะกิจสำหรับงานวิจัยนั้น ๆ ซึ่งจะเป็นภาระต่อเจ้าของงานวิจัยหรือหน่วยงานที่สังกัดอยู่ทั้งในระยะสั้น (ขณะที่ทำการวิจัย) และในระยะยาว (หลังสิ้นสุดงานวิจัยแล้ว) กลุ่มผู้ใช้ข้อมูลลักษณะนี้นอกจากจะเป็นกลุ่มนักวิจัย ยังมีกลุ่มเจ้าหน้าที่หรือหน่วยงานซึ่งบริหารระบบเครือข่าย และให้บริการอินเทอร์เน็ตของมหาวิทยาลัย ซึ่งสามารถใช้ข้อมูลที่เกิดขึ้นแบบถาวรนี้ ในการวิเคราะห์การใช้งาน การขยายความต้องการสามารถระบุประเภทของบริการที่ได้รับความนิยมเหล่านี้เป็นต้น ช่วยในการระบุปัญหา สนับสนุนการวางแผน การจัดทำงบประมาณ และการขยายการบริการอย่างเป็นระบบอีกด้วย ดังนั้น การเก็บข้อมูลและให้บริการตามแนวทางการวิจัยนี้ จึงน่าจะเป็นการลดภาระของนักวิจัยและผู้บริหารที่จะต้องอาศัยข้อมูลเหล่านี้

ปัญหาอีกประการหนึ่งคือ ไม่มีเมตาดาตาหรือลักษณะสมบัติบ่งชี้ เนื่องจากข้อมูลการใช้อินเทอร์เน็ตเป็นข้อมูลที่มีลักษณะแตกต่างจากข้อมูลในสิ่งพิมพ์ ข้อมูลรูปภาพ เสียง และรูปแบบอื่น กล่าวคือ ข้อมูลที่เก็บในฮาร์ดดิสก์นั้น มีลักษณะเป็นข้อมูลต่อเนื่องเข้ามาตลอดเวลา เมื่อข้อมูลที่อยู่บนฮาร์ดดิสก์เต็ม จะถูกตัดตอนเป็นกลุ่มแล้วถ่ายลงซีดี-อาร์ ยังไม่มีการจัดทำระเบียบบรรณานุกรมหรือเมตาดาตาใด ๆ เพื่อให้เข้าถึงข้อมูลที่เกิดขึ้นได้ ทำให้ขาดความหลากหลายในการเข้าถึงกลุ่มข้อมูลในแง่มุมต่าง ๆ เกิดการสูญหายหรือตกหล่นของข้อมูลส่วนต่าง ๆ หากไม่มีระบบที่ชัดเจนรองรับ จะทำให้ไม่อาจใช้ประโยชน์จากข้อมูลได้อย่างมีประสิทธิภาพ หรือนำออกให้บริการในวงกว้าง เนื่องจากข้อมูลไม่อยู่ในรูปแบบที่เหมาะสมต่อการค้นหาและเรียกใช้งาน จึงมีความจำเป็นในการพยายามให้เกิดระบบการจัดเก็บข้อมูลให้อยู่ในสภาพที่สมบูรณ์และสามารถใช้งานได้ในระยะยาว พร้อมมีเมตาดาตาในลักษณะที่เป็นระเบียบบรรณานุกรมให้เกิดความสะดวกในการนำข้อมูลมาใช้ [2,6] เพื่อให้สามารถใช้อินเทอร์เน็ตที่เก็บได้ จึงต้องมีหลักเกณฑ์การสร้างระเบียบบรรณานุกรมที่เหมาะสมหาข้อมูลให้ได้

กระบวนการที่สมบูรณ์ตามขั้นตอนนี้ สามารถสร้างความมั่นใจให้เกิดขึ้นกับข้อมูลข่าวสารและเนื้อหาพื้นถิ่นที่เป็นอิเล็กทรอนิกส์ ทำให้ข้อมูลไม่มีการสูญหายหรือเสียหาย สามารถใช้ประโยชน์ได้เป็นเวลานานตามลักษณะการวิจัยและการใช้งานเชิงบริหารของข้อมูลประเภทนี้ ตลอดจนสามารถเข้าถึงข้อมูลได้อย่างเจาะจงในลักษณะต่าง ๆ เช่น กำหนดช่วงเวลาที่เกิดข้อมูลเพื่อการวิเคราะห์ ตัวอย่างเช่น ต้องการทราบพฤติกรรมปริมาณการใช้งานในช่วงวันหยุดนักขัตฤกษ์ หรือในช่วงปิดภาคการศึกษา เป็นต้น

วัตถุประสงค์

ด้วยเหตุผลดังกล่าว จึงกำหนดวัตถุประสงค์ของการวิจัย เพื่อศึกษาและกำหนดรูปแบบบรรณานุกรมสำหรับข้อมูลต่อเนื่องที่ตัดมาเป็นกลุ่ม รวมถึงวางแผนทางการจัดเก็บ เพื่อให้เป็นแบบถาวรและสามารถสืบค้นผ่านห้องสมุดอัตโนมัติ โดยใช้กรณีศึกษาจากข้อมูลสถิติการใช้งานอินเทอร์เน็ตของจุฬาลงกรณ์มหาวิทยาลัย จากเครือข่ายภายใน (CUNet) ไปยังเครือข่ายระหว่างมหาวิทยาลัยของทบวงมหาวิทยาลัย หรือ เรียกว่า ยูนิเน็ต (UniNet) และจากเว็บแคชหลักของจุฬาลงกรณ์มหาวิทยาลัย

ขอบเขตของการวิจัย

1. ใช้สื่อที่มีความถาวรในการจัดเก็บข้อมูลการใช้งานอินเทอร์เน็ตของจุฬาลงกรณ์มหาวิทยาลัย คือ
 - ก. จากเกตเวย์เราเตอร์ (gateway router) ไปยังเครือข่ายระหว่างมหาวิทยาลัยของทบวงมหาวิทยาลัย
 - ข. จากเว็บแคชหลักของมหาวิทยาลัย
2. กำหนดแผ่นแบบ (template) โครงสร้างข้อมูล ลักษณะบรรณานุกรมตามหลักเกณฑ์การลงรายการของมาร์ก (MARC) และการให้บริการข้อมูลออนไลน์
3. กำหนดขั้นตอนกระบวนการจัดเก็บข้อมูลในลักษณะการเก็บถาวรดิจิทัล
4. ใช้อุปกรณ์และซอฟต์แวร์ที่เป็นผลิตภัณฑ์ที่มีใช้งานอยู่ หรือสามารถหาได้ในท้องตลาดโดยทั่วไป

สถานที่ทำการวิจัยหรือเก็บข้อมูล

สถานที่ทำการวิจัย คือ ห้องปฏิบัติการระบบสารสนเทศ ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ และ สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

สถานที่เก็บข้อมูล คือ สถาบันวิทยบริการ สำนักเทคโนโลยีสารสนเทศ จุฬาลงกรณ์มหาวิทยาลัยและเครือข่ายยูนิเน็ต

ผลที่จะได้รับการวิจัย

1. ได้กระบวนการจัดเก็บและค้นคืนข้อมูลต่อเนื่อง หรือข้อมูลที่มีคุณสมบัติในลักษณะเดียวกันสามารถนำไปใช้กับข้อมูลสถิติหรือข้อมูลการวิเคราะห์ในลักษณะตรวจจับติดตาม (monitoring) ในทางวิทยาศาสตร์ทั้งหลายได้
2. ได้ฐานข้อมูลตัวอย่างการใช้งานอินเทอร์เน็ตของประชาคมจุฬาลงกรณ์มหาวิทยาลัย และสามารถสืบค้นข้อมูลผ่านห้องสมุดเครือข่ายจุฬาลงกรณ์มหาวิทยาลัย โดยสามารถดำเนินงานต่อให้เกิดการจัดเก็บถาวรในลักษณะที่เป็นสาระพื้นถิ่นของจุฬาลงกรณ์มหาวิทยาลัยและของประเทศ
3. เป็นรากฐานในการเก็บสารสนเทศพื้นถิ่นเชิงอิเล็กทรอนิกส์ซึ่งเป็นแบบถาวร
4. เป็นแนวทางในการนำไปสู่การจัดทำห้องสมุดอิเล็กทรอนิกส์สมบูรณ์แบบที่ปราศจากกระดาษ

บทที่ 2

วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

การกำหนดรูปแบบบรรณานุกรมและการวางแผนทางการจัดเก็บแบบถาวรสำหรับข้อมูลต่อเนื่องที่ตัดมาเป็นกลุ่มนั้น สามารถแบ่งเป็นประเด็นที่เกี่ยวข้องกับงานวิจัย ดังต่อไปนี้

1. ปัญหาเกี่ยวกับบันทึกถาวร
2. สื่อที่ใช้บันทึก
3. เมตาเดตา
4. ลักษณะของข้อมูล
5. การใช้ประโยชน์ข้อมูล

ปัญหาเกี่ยวกับบันทึกถาวร

การเก็บรักษา (preservation) ของสิ่งพิมพ์หรือเนื้อหาสาระในลักษณะที่เป็นการถาวรหรืออยู่ในรูปของบันทึกถาวร (archive) ขึ้นอยู่กับสื่อที่ใช้ในการแพร่กระจายเนื้อหาของสิ่งพิมพ์ นับแต่อดีตกาลที่ผ่านมาจนถึงปัจจุบันจะใช้กระดาษที่ผ่านกรรมวิธีขจัดปริมาณของกรด เพื่อให้กระดาษไม่ย่อยสลายตัวเอง ทำให้มีอายุทนนาน เหมาะสมต่อการเก็บในลักษณะที่ถาวร ประกอบกับไม่มีปัญหาในเรื่องสิทธิ์การใช้งาน ทำให้องค์กรอิสระอย่างห้องสมุด และร้านจำหน่ายหนังสือ สามารถเป็นเจ้าของตัวเล่มของสื่อที่เป็นกระดาษได้ และทำให้สามารถจัดเก็บสาระต่าง ๆ ในระยะยาวไว้ในห้องสมุดต่าง ๆ ได้ แต่เทคโนโลยีของเครือข่ายที่พัฒนาแพร่หลายไปมากในปลายศตวรรษที่ 20 ทำให้ระบบอิเล็กทรอนิกส์เข้ามามีบทบาทอย่างสูง ในการเป็นสื่อกลางในการผลิต การเก็บ และการแพร่กระจายของเนื้อหาสาระที่พัฒนาอยู่ในรูปของดิจิทัล ซึ่งมีผลให้ผู้ประพันธ์หรือเจ้าของสาระ สามารถจัดการแจกจ่ายเนื้อหาของตนไปยังผู้อ่าน ผู้ใช้ได้โดยตรงผ่านเครือข่ายสากลอย่างอินเทอร์เน็ต เป็นการข้ามขั้นตอนของผู้พิมพ์ ผู้จำหน่าย ร้านขายและห้องสมุดไปโดยสิ้นเชิง การเก็บรักษาเอกสารจึงตกเป็นหน้าที่ของแหล่งกำเนิดข้อมูล ซึ่งไม่ได้ทำกิจการด้านการเก็บรักษา ถาวร อาจเกิดปัญหาเรื่องการเลิกกิจการ หรือมีแต่นโยบายการให้บริการข้อมูลในระยะสั้น โดยมีได้มีศักยภาพในการเก็บรักษาข้อมูลจำนวนมากในระยะยาว ดังนั้น สถาบันอิสระอย่างเช่น ห้องสมุด จึงต้องคำนึงถึงการเก็บรักษาสิ่งตีพิมพ์หรือข้อมูลเหล่านี้ในสภาพที่เป็นประโยชน์ในระยะยาว แต่การเก็บเอกสารต่าง ๆ ในระบบดิจิทัล ทำให้แนวความคิดของการเก็บและการบำรุงรักษาเปลี่ยนไป เนื่องจากปัญหาหลัก 2 ประการ [7] [8] ที่เกิดขึ้นกับเอกสารดิจิทัล คือ

1. อายุของสื่อที่บันทึก
2. ความล้าสมัยของซอฟต์แวร์และฮาร์ดแวร์

ตัวอย่างของการเกิดข้อมูลสูญหายจากปัญหาที่เกิดขึ้นได้แก่ นักวิทยาศาสตร์ของ Jet Propulsion Laboratory ขององค์การนาซ่า (NASA) พบว่า ข้อมูลสำรวจดาวอังคารที่เก็บไว้ในเทปแม่เหล็ก ตั้งแต่ ปี ค.ศ. 1976 มีความเสียหายถึงร้อยละ 20 และมหาวิทยาลัยเยล ไม่สามารถอ่านข้อมูลหนังสือ 2,000 เล่มที่ถ่ายจากไมโครฟอร์มมิง optical disc [5] รวมทั้งข้อมูลการสำรวจประชากรของสหรัฐอเมริกา ตั้งแต่ช่วงปี 1960 ซึ่งเก็บ

ไว้ในเทปแม่เหล็กที่สามารถเก็บข้อมูลในระยะยาวได้ แต่ไม่สามารถอ่านข้อมูลจากเทปแม่เหล็กที่เก็บไว้ได้ เนื่องจากเทปแม่เหล็กที่เก็บข้อมูลนั้นสามารถอ่านได้เฉพาะกับเครื่อง UNIVAC type II-A เท่านั้น ซึ่งเครื่องรุ่นนี้ไม่มีใช้ในท้องตลาดแล้วจึงต้องมีการทำสำเนาข้อมูลลงเทปแม่เหล็กที่เป็นมาตรฐานที่สามารถหาเครื่องใด ๆ อ่านได้ [2]

การเก็บรักษาเอกสารดิจิทัลจึงต้องเปลี่ยนไปตามสภาพแวดล้อม เพื่อมิให้เกิดความเสียหายต่อข้อมูล อันมีคุณค่าจากปัญหาที่เกิดจากสื่อที่บันทึก และจากโปรแกรมและอุปกรณ์ที่ใช้อ่านโดยมีกรรมวิธีต่าง ๆ จำแนกได้ ดังนี้ [2] [6] [9]

1. การปรับสภาพ (refreshing) คือ การถ่ายโอนเอกสารดิจิทัลไปยังสื่อบันทึกใหม่ เช่น เปลี่ยนจากดิสเก็ตต์ 5 ¼ นิ้ว ซึ่งมีอายุการเก็บข้อมูลเพียงไม่กี่ปีไปเป็นซีดี-รอม หรือจากซีดี-รอม เป็นดีวีดี-รอม
2. การย้ายข้อมูล (migration) คือ การย้ายข้อมูลไปรูปแบบ (format) ใหม่ของซอฟต์แวร์/ฮาร์ดแวร์อีกแบบหนึ่ง หรือจากคอมพิวเตอร์รุ่นหนึ่งที่ล้าสมัยไปยังอีกรุ่นหนึ่งที่ทันสมัยกว่า เพื่อยังคงให้ข้อมูลที่เก็บนั้นยังสามารถเข้าถึง แสดงผล และใช้ได้ แม้ว่าเทคโนโลยีได้เปลี่ยนไป
3. การสงวนรักษาเทคโนโลยี (technology preservation) คือ การเก็บรักษาฮาร์ดแวร์และสภาพแวดล้อมในการปฏิบัติงานเดิมไว้ เพื่อนำมาใช้กับเอกสารดิจิทัลที่บันทึกกับฮาร์ดแวร์ในรุ่นนั้น ๆ ซึ่งเป็นวิธีที่ค่อนข้างเป็นไปได้ยาก เนื่องจากฮาร์ดแวร์และซอฟต์แวร์ของสื่อดิจิทัลมีการเปลี่ยนแปลงอย่างรวดเร็ว ซึ่งวิธีการแก้ไขปัญหาย่างหนึ่ง คือ การเก็บเฉพาะข้อกำหนดฮาร์ดแวร์และซอฟต์แวร์ แล้วจึงทำการเลียนแบบ (emulation) ในการจำลองเทคโนโลยีเก่าโดยอาศัยอุปกรณ์ของสมัยใหม่ [10]

ตัวอย่างการสงวนรักษาเอกสารดิจิทัลของหอสมุดแห่งชาติสาขาแห่งประเทศไทย ที่เมืองโม ไอ รานา (Mo I Rana) [8] ซึ่งมีหน้าที่ในการเก็บรักษาเอกสารทุกประเภท 1 ฉบับ ตามพระราชบัญญัติ ดำเนินงานสงวนการเก็บรักษาเอกสารโดยชุดอุโมงค์ในภูเขาซึ่งเพียงอาคารแรกในอุโมงค์นี้จะมีชั้นเก็บยาวถึง 42,000 เมตร มีการควบคุมอุณหภูมิที่ 8 องศาเซลเซียส และความชื้นที่ 35 เปอร์เซ็นต์ รวมทั้งมีระบบการระบายอากาศอย่างดี เพื่อให้สามารถเก็บและคงสภาพของเอกสารให้คงทนถาวรนั่นเอง

เพราะฉะนั้น ในการดำเนินการจัดเก็บข้อมูลการใช้อินเทอร์เน็ตของประชาคมจุฬาลงกรณ์มหาวิทยาลัย ซึ่งถือว่าเป็นเอกสารดิจิทัลประเภทหนึ่ง จึงต้องคำนึงถึงสื่อบันทึกที่เหมาะสมกับข้อมูล พร้อมทั้งการวางแผนในระยะยาวเพื่อการใช้งานและรองรับการเปลี่ยนแปลงของเทคโนโลยี

สื่อที่ใช้นบันทึก

พัฒนาการทางเทคโนโลยีของสื่อที่เป็นไปอย่างรวดเร็วก่อให้เกิดความหลากหลายในเทคโนโลยีการบันทึกพื้นฐานโดยแยกออกเป็น เทปแม่เหล็ก แผ่นแม่เหล็ก แผ่น optical และเทคโนโลยีสมัยใหม่ของการเก็บข้อมูล โดยใช้ solid-state memory ซึ่งสื่อตามเทคโนโลยีพื้นฐานเหล่านี้มีคุณสมบัติเฉพาะของผลิตภัณฑ์อย่างหลากหลายในแต่ละประเภทของเทคโนโลยี ทั้งในด้านความจุ รูปทรงทางกายภาพ และราคา เป็นต้น องค์ประกอบต่าง ๆ เหล่านี้มีผลกระทบต่อการตัดสินใจคัดเลือกสื่อบันทึกที่เหมาะสมกับการเก็บในระยะยาว

1. **เทคโนโลยีพื้นฐาน** ระบบการเก็บข้อมูลโดยอาศัยสภาวะการเปลี่ยนชั่วคราวของอนุภาคสารแม่เหล็กนั้น มีใช้กันมาเป็นเวลานานนับตั้งแต่มีการประดิษฐ์คอมพิวเตอร์ขึ้นมา โดยแยกออกเป็น

การเคลือบสารแม่เหล็กบนเส้นเทป และการเคลือบลงบนแผ่นกลมทั้งแบบแผ่นแข็ง (hard disk) และแผ่นอ่อน (flexible disk) ผลิตภัณฑ์ที่ใช้เส้นเทปที่เรียกว่า เทปแม่เหล็ก (magnetic tape) นิยมใช้กันในการสำรองข้อมูล (backup) เนื่องจากมีราคาต่ำ และความเร็วไม่สูงพอจะใช้เก็บ on-line data ได้ ผลิตภัณฑ์แบบเทปแม่เหล็กมีพัฒนาการตลอดเวลา มีรูปแบบ ขนาด และความจุแตกต่างกันไป โดยทั่วไปจะได้รับการออกแบบมาทั้งสื่อและอุปกรณ์ เพื่อการใช้งานกับระบบคอมพิวเตอร์เป็นหลัก

ในส่วนของแผ่นแม่เหล็กนั้น ถ้าเป็นแบบแข็ง จะเรียกว่า จานแม่เหล็กแบบแข็ง หรือจานแม่เหล็กเพียงสั้น ๆ หรือ ฮาร์ดดิสก์ (hard disk) ส่วนถ้าเป็นแบบแผ่นอ่อนนั้น คือ จานแบบอ่อน (flexible disk) ซึ่งนิยมเรียกว่า floppy disk หรือ ดิสเก็ตต์ (diskette) ในส่วนของดิสเก็ตต์นั้น ในปัจจุบันมีมาตรฐานขนาดบรรจุ 1.44 MB เพียงขนาดเดียว แม้ว่าจะมีความพยายามผลักดันขนาดบรรจุที่สูงกว่าออกมา แต่ไม่ได้รับความนิยม ส่วนจานแม่เหล็กนั้นในปัจจุบันผลิตเป็นแบบ winchester disk ทั้งหมด คือ รวมตัวแผ่นจาน อุปกรณ์หัวอ่านมอเตอร์ขับเคลื่อนสำหรับหมุนจาน และวงจรถวลีเก็ททรอนิกส์ในการควบคุมต่าง ๆ เข้าไว้ในภาชนะชิ้นเดียวกัน มีการผนึกกันอากาศและฝุ่นละออง ทำให้สามารถเพิ่มความหนาแน่นของการเก็บข้อมูลขึ้นไปอย่างมาก และในปัจจุบันยังคงเป็นอุปกรณ์หน่วยความจำที่ได้รับความนิยมในระบบคอมพิวเตอร์โดยทั่วไป เพราะมีราคาต่ำ ความจุสูง และถ่ายโอนข้อมูลได้รวดเร็ว

ในส่วนของระบบแสง หรือ ระบบ optical ซึ่งเป็นเทคโนโลยีสมัยใหม่มีอายุเพียง 20 ปี เป็นเทคโนโลยีที่โดยพื้นฐานจะมีประสิทธิภาพสูงกว่าระบบแม่เหล็ก ทั้งในด้านความถาวร ความหนาแน่น รวมทั้งความเร็ว แต่เนื่องจากเป็นเทคโนโลยีที่ใหม่ยังต้องผ่านพัฒนาการอีกมากในปัจจุบัน จึงยังด้อยกว่าเทคโนโลยีของระบบแม่เหล็กในหลาย ๆ ด้าน ระบบเก็บข้อมูลแบบแสงนี้ ได้รับการพัฒนาควบคู่ไปกับการใช้งานแบบสื่อบันทึกอย่างภาพยนตร์และเพลง โดยในระยะแรกออกผลิตภัณฑ์เป็นจานกลมขนาด 12 นิ้ว เรียกว่า เลเซอร์ดิสก์ (laser disc) สำหรับจำหน่ายภาพยนตร์ แต่ต่อมามีผลิตภัณฑ์ที่เป็นจานขนาดย่อม สำหรับใส่เพลงจึงเรียกว่า จานขนาดย่อม (compact disc) หรือ ซีดี (CD) มีเส้นผ่าศูนย์กลาง 4 3/4 นิ้ว สะดวกต่อการจับถือและพกพามากขึ้น และเมื่อต่อมาได้มีการพัฒนาเทคโนโลยีการบีบอัดข้อมูล ทำให้สามารถบรรจุภาพยนตร์ลงในซีดีเป็น video CD หรือ VCD จึงทำให้แผ่นแบบซีดีได้รับความนิยมเกิดเทคโนโลยีของการบันทึกครั้งเดียว (CD-R - recordable) และการบันทึกหลายครั้ง (CD-RW - rewritable) ตามมาทำให้สะดวกต่อการนำมาใช้ในการเก็บข้อมูลของคอมพิวเตอร์ได้

เทคโนโลยีพื้นฐานอีกชนิดหนึ่งที่เริ่มออกสู่ท้องตลาดและมีแนวโน้มว่าจะได้รับความนิยม คือ การเอาหน่วยความจำแบบ non-volatile ที่เก็บข้อมูลได้โดยไม่ต้องใช้ไฟเลี้ยงตลอดเวลา เช่น flash memory มาใช้ในการเก็บและพกพาข้อมูล โดยอาศัยการผลิต semi-conductor chip ที่มีประสิทธิภาพและมีความหนาแน่นสูง ทำให้สามารถบรรจุข้อมูลได้ถึง 256 MB หรือ ประมาณ 1 ใน 4 GB ได้ภายใต้โครงสร้างทางกายภาพที่เล็กกะทัดรัด เช่น ขนาดหัวแม่มือ แม้ว่าเทคโนโลยีนี้จะน่าสนใจเพราะไม่มีส่วนเคลื่อนไหวในการใช้งาน (no moving part) แตกต่างจากระบบแม่เหล็กหรือ ระบบแสง แต่ความจุที่ต่ำและราคาที่สูงมาก ทำให้เทคโนโลยีนี้ยังไม่เป็นที่น่าสนใจสำหรับการเก็บข้อมูลจำนวนมากในลักษณะที่เป็นบันทึกถาวร

2. คุณสมบัติของสื่อ ในการเก็บบันทึกถาวรนั้น อายุของสื่อ เป็นปัจจัยที่สำคัญที่สุด สื่อประเภทแม่เหล็กที่มีการเสื่อมสภาพแม่เหล็กจากอิทธิพลของสนามแม่เหล็กโลก ทำให้ไม่เหมาะสมต่อการเก็บ

ในระยะยาว คุณสมบัติที่สำคัญอื่น ๆ ได้แก่ ความจุและมาตรฐานของสื่อ ในด้านความจุนั้น หากเก็บข้อมูลที่มีปริมาณที่มาก อย่างเช่นข้อมูลที่ใช้เป็นกรณีศึกษาในงานวิจัยนี้ จะต้องใช้สื่อที่มีความจุสูงในราคาที่ย่อมเยา ในขณะเดียวกัน มาตรฐานของสื่อบันทึก จะเป็นองค์ประกอบที่สำคัญ ทำให้การจัดหาเป็นไปได้โดยสะดวก อีกทั้งมาตรฐานที่ชัดเจนจะมีแนวโน้มทำให้สื่อมีการใช้งานอย่างแพร่หลาย มีราคาต่ำ

เมตาดาตา

เมตาดาตา คือ สารสนเทศที่ถูกกำหนดขึ้นเพื่ออธิบายคุณลักษณะทรัพยากรสารสนเทศ ในด้านเนื้อหา ลักษณะทางกายภาพ สถานที่ ประเภท และรูปแบบของสารสนเทศ รวมทั้งประวัติการย้ายข้อมูล วันหมดอายุ ความปลอดภัยของระบบ ความถูกต้อง รูปแบบของแฟ้มข้อมูล และความสัมพันธ์กับการจัดทำในครั้งก่อนหรือรุ่น (version) เป็นต้น ลักษณะการลงเมตาดาตาที่ค่อนข้างจะเป็นที่คุ้นเคย คือ การลงรายการทางบรรณานุกรมของทรัพยากรสารสนเทศของห้องสมุดนั่นเอง

การกำหนดเมตาดาตาเป็นองค์ประกอบสำคัญ ที่ช่วยผู้จัดการระบบหรือโปรแกรม ในการบริหารหรือจัดการสารสนเทศ และช่วยให้ผู้ใช้เข้าถึงและได้รับสารสนเทศที่ต้องการ เพราะฉะนั้น การนำเมตาดาตาเพื่อพรรณนาเอกสารดิจิทัลที่จัดเก็บแบบถาวรนั้น ควรเป็นเมตาดาตาที่เป็นมาตรฐานและง่ายต่อการใช้ของผู้จัดทำและผู้ดูแลเอกสารดิจิทัล รวมทั้งมีเมตาดาตาเพียงพอที่จะช่วยให้ผู้ใช้สารสนเทศสามารถหาสารสนเทศที่ต้องการได้

The Digital Library Forum (DLF) System Architecture Committee [11] ได้แบ่งเมตาดาตาออกเป็น 3 ประเภท คือ

1. ข้อมูลเชิงพรรณนา (description metadata) คือ เมตาดาตาที่อธิบายสารสนเทศ
2. ข้อมูลเชิงโครงสร้าง (structure metadata) คือ เมตาดาตาที่แสดงและนำสารสนเทศไปสู่ผู้ใช้
3. ข้อมูลเชิงการบริหาร (administrative metadata) คือ เมตาดาตาที่แสดงถึงการจัดการข้อมูล เช่น วันที่สร้างข้อมูล รูปแบบของแฟ้มข้อมูล และสิทธิในการใช้ เป็นต้น

ในการกำหนดเมตาดาตานั้น อาจทำได้โดยกำหนดเมตาดาตาไว้ในสารสนเทศนั้น ๆ เอง เช่น เป็นแฟ้มข้อมูล ตัวอย่างการเก็บเมตาดาตาในลักษณะนี้ได้แก่ การเก็บเมตาดาตาไว้ในเว็บเพจนั้น ๆ เมื่อมีการสำเนา ย้าย แก้ไข หรือ ลบข้อมูล เมตาดาตาจะถูกดำเนินการตามขั้นตอนนั้น ๆ ด้วยเช่นกัน หรืออาจกำหนดเมตาดาตาแยกออกมาจากสารสนเทศ หรืออาจแยกออกมาแต่ทำโยงไปถึงสารสนเทศนั้น

จากการศึกษาการกำหนดเมตาดาตาสำหรับการเก็บรักษาข้อมูลหรือเอกสารดิจิทัลในต่างประเทศนั้น พบว่า มีโครงการต่าง ๆ ที่กำหนดระเบียบปฏิบัติ แนวทาง และรากฐานในการจัดสร้างและบำรุงรักษาด้านที่ถาวรขึ้นในโลก ได้แก่

1. การศึกษาเมตาดาตาสำหรับงานปฏิบัติการ
2. โครงการของ OAIS
3. โครงการของวงการห้องสมุด

การศึกษาเมตาตาตาสำหรับงานปฏิบัติการ

McKemmish และ Perer [12] ได้ให้คำนิยามของ recordkeeping ที่เป็งานปฏิบัติการว่า เป็นมาตรฐานของข้อมูลที่เกี่ยวข้องกับความเป็นเอกลักษณ์ ความเป็นจริง โครงสร้างของเนื้อหา บริบท และความ ต้องการในการจัดการที่จำเป็นของระเบียบของข้อมูลที่ใช้ในงานปฏิบัติการโดยทั่วไป อย่างน้อยที่สุดข้อมูล บางประการข้างต้นสามารถใช้เป็นตัวจับได้ทันทีว่าระเบียบนั้นถูกสร้างขึ้นเมื่อใด ซึ่งมีประโยชน์ต่อนักวิจัยในการพิสูจน์ความถูกต้อง ความเป็นจริงของการบันทึก มีโครงการศึกษาการกำหนดเมตาตาตาสำหรับ recordkeeping หลายโครงการ อาทิ เช่น

1. The Functional Requirement for Evidence in Recordkeeping Project เป็นโครงการศึกษาวิจัยของ School of Information Science แห่งมหาวิทยาลัยพิตส์เบิร์ก ในระหว่างปี ค.ศ. 1994-1997 [13] โดยมีวัตถุประสงค์ในการศึกษาคือ พัฒนาข้อกำหนดหน้าที่ของระเบียบและกำหนด metadata specification for evidence ซึ่งเป็นที่รู้จักกันในชื่อ Reference Model for Business Acceptable Communications (BAC) โดยจัดแบ่งเป็น layer ต่าง ๆ ซึ่งเก็บรายละเอียดของข้อมูลของคำศัพท์ และสภาพการใช้ โครงสร้างข้อมูล แหล่งกำเนิดข้อมูล เนื้อหา และการใช้ระเบียบตั้งแต่สร้างข้อมูล โดยเมตาตาตาเหล่านี้จะเชื่อมโยงโดยตรงไปยังแต่ละระเบียบ และจะสามารถอธิบายเนื้อหาและบริบทของระเบียบ รวมทั้งสามารถถอดรหัสโครงสร้างของตัวเองเพื่อการใช้ในอนาคต

2. The Preservation of the Integrity of Electronic Records โดย School of Library Archival and Information Studies แห่งมหาวิทยาลัยบริติชโคลัมเบีย ได้ศึกษาโครงการนี้ระหว่างปี ค.ศ. 1994-1997 [14] โดยพัฒนาจากแนวความคิดในเรื่อง “ความน่าเชื่อถือ” และ “ความเป็นจริง” และผลจากการศึกษานี้ ได้สร้างแม่แบบ (template) 8 ชุด เพื่อช่วยพิสูจน์องค์ประกอบ (element) ที่จำเป็นของระเบียบ

3. The InterPARES Project (International Research on Permanent Authentic Records in Electronic Systems) เป็นโครงการศึกษาของ School of Library Archival and Information Studies แห่งมหาวิทยาลัยบริติชโคลัมเบีย [15] ศึกษาโดยใช้ ความน่าเชื่อถือ และความเป็นจริงของระเบียบอิเล็กทรอนิกส์มาวิเคราะห์องค์ประกอบรูปแบบของเอกสารโดยมีแม่แบบเป็นตัวบ่งชี้ลักษณะเฉพาะของระเบียบนั้นว่าถูกเปลี่ยนไปเมื่อใดและด้วยวิธีการใด และมีการสร้างขึ้นใหม่ด้วยรูปแบบใด ซึ่งจะทำให้ทราบถึงอดีต ปัจจุบัน วิธีการ รูปแบบ และการนำไปใช้

4. Australian Recording Metadata Standards เป็นโครงการศึกษาของ School of Information Management and Systems แห่งมหาวิทยาลัยโมนาช [16] ได้พัฒนา Australian Recording Metadata Schema (RKMS) โดยกำหนดและจัดทำเป็นมาตรฐานสำหรับเมตาตาตาของ recordkeeping เพื่อใช้จัดการกับระเบียบที่อยู่ในสภาพแวดล้อมที่เป็นดิจิทัลโดยนำเอามาตรฐานการลงเมตาตาตาที่เป็นมาตรฐานทั่ว ๆ ไป เช่น Dublin Core และ Australian Government Locator Service (AGLS) scheme มาเป็นองค์ประกอบพื้นฐานของระบบ

โครงการของ OAIS

ในบรรดาโครงการเพื่อกำหนดมาตรฐานสำหรับการจัดสร้าง บำรุงรักษาระบบบันทึกถาวร โครงการ OAIS หรือที่เรียกว่า Reference Model for an Open Archival Information System นับเป็นหนึ่งในโครงการที่ได้รับความสนใจมาก แม้แต่ ISO ซึ่งเป็นหน่วยงานมาตรฐานระดับโลก ก็ได้เริ่มดำเนินการเพื่อยกร่างให้ OAIS เป็นมาตรฐานระดับโลก ทั้งที่โครงการนี้ถือกำเนิดนอกวงการห้องสมุดและหน่วยงานทางด้านบันทึกถาวรทั้งหลายของโลก โดยโครงการนี้เกิดจากความจำเป็นในการเก็บรักษาข้อมูลที่ได้จากการสำรวจอวกาศ ซึ่งมีปริมาณมหาศาล และเนื่องจากเป็นโครงการที่ถือกำเนิดจากกลุ่มงานวิศวกรรม เป็นคณะกรรมการที่จัดตั้งโดยฝ่ายอำนวยการของสมาพันธ์หน่วยงานด้านอวกาศของโลกจำนวน 10 หน่วยงาน ซึ่งสมาพันธ์นี้มีหน่วยงานที่เป็นผู้สังเกตการณ์มีจำนวนสมาชิกถึง 23 หน่วยงาน มาตรฐานหรือแนวทางที่ออกมาจึงมีระเบียบและแผนแบบที่ชัดเจนเป็นขั้นตอน มีคู่มือที่เรียกว่า Red Book [17] ที่มีรูปแบบตามแนวทางของการกำหนดมาตรฐานสำคัญของโลก ทำให้สะดวกต่อการทำความเข้าใจและถือปฏิบัติ ดังจะเห็นได้จากการถูกนำไปปรับใช้ในการกำหนดเมตาเดตาสำหรับข้อมูลดิจิทัลของโครงการศึกษาหลายโครงการ เช่น The Cedars Project [18] และ The NEDLIB Project [19]

โครงการของวงการห้องสมุด

ในวงการห้องสมุดมีความตระหนักถึงการบำรุงรักษาข้อมูลดิจิทัลเพื่อเป็นการถาวรด้วยเช่นเดียวกัน ได้มีการศึกษาวิจัยเกี่ยวกับการลงเมตาเดตาสำหรับการบำรุงรักษาโดยหอสมุดแห่งชาติและหน่วยงานการศึกษาอีกหลายโครงการ ได้แก่

1. PANDORA เป็นโครงการศึกษาวิจัยการลงเมตาเดตาสำหรับการบันทึกถาวรของหอสมุดแห่งชาติออสเตรเลีย (The National Library of Australia) โดยเริ่มศึกษาตั้งแต่ ปี ค.ศ. 1996 [20] ระบบ PANDORA (Preserving and Accessing Networked Documentary Resources of Australia) อาศัยการจัดเก็บและการจัดการด้วยระบบห้องสมุด รวมทั้งมีการจัดระบบ PURLs (Persistent Uniform Resource Identifiers) มาใช้ในการบ่งชี้แต่ละรายการ ในการศึกษาครั้งนี้ได้พัฒนาข้อกำหนดภายใต้ชื่อ “Preservation Metadata for Digital Collections” โดยพิมพ์ฉบับร่างออกมาในเดือนตุลาคม ค.ศ. 1999 ซึ่งได้กำหนดข้อมูลที่เป็นระบบการจัดเก็บด้วยดิจิทัลสามารถนำไปใช้เพื่อจัดการและบำรุงรักษาข้อมูลดิจิทัล ประกอบด้วย 25 องค์ประกอบ แต่ไม่ได้มีการกล่าวถึงกลยุทธ์ทางเทคโนโลยีที่จำเป็นในการประยุกต์ใช้กับการบำรุงรักษา เช่น การย้ายข้อมูล และการจำลอง (simulation) ของเทคโนโลยีการจัดเก็บ

2. The Cedars Project ดำเนินการโดย The Consortium of University Research Libraries (CURL) [18] โดยมีวัตถุประสงค์ในการศึกษาเพื่อกล่าวถึงกลยุทธ์ รวมทั้งวิธีการและข้อปฏิบัติที่เกี่ยวข้องกับการบำรุงรักษาข้อมูลดิจิทัล ซึ่งมี 3 ประการ ได้แก่ กลยุทธ์และเทคนิคในการบำรุงรักษาข้อมูลดิจิทัล การพัฒนาข้อมูลและสิทธิในการจัดการข้อมูล และประการสุดท้ายคือ การกำหนดเมตาเดตา ทั้งนี้ได้จัดทำข้อกำหนดในการลงเมตาเดตาฉบับร่างซึ่งมีโครงสร้างจากโครงการของ OAIS

3. The NEDLIB Project (Networked European Deposit Library) เป็นความร่วมมือระหว่างหอสมุดแห่งชาติ สำนักพิมพ์ หน่วยงานทางเทคโนโลยีสารสนเทศ และหอจดหมายเหตุแห่งชาติของกลุ่ม

ประเทศในยุโรป โดยเริ่มศึกษาตั้งแต่ ปี ค.ศ. 1998 – 2000 [19] โครงการนี้ได้พัฒนาระบบการจัดเก็บสิ่งพิมพ์อิเล็กทรอนิกส์ หรือ Deposit System for Electronic Publications (DSEP) ซึ่งมีพื้นฐานมาจาก OAIS เช่นเดียวกับโครงการของ Cedars และมีการกำหนดองค์ประกอบในการลงเมตาดาตาที่จำเป็นในการจัดการการบำรุงรักษาเอาไว้ 18 องค์ประกอบหลัก และ 38 องค์ประกอบย่อย ซึ่งรวมถึงปัญหาเรื่องความล้าสมัยของเทคโนโลยีแต่จะไม่มีกรลงเมตาดาตาในเรื่องการพรรณนาข้อมูล การบริหารงาน หรือทางด้านกฎหมาย

4. The OCLC/RLG Working Group on Preservation Metadata เป็นโครงการศึกษาที่เกิดจากความร่วมมือระหว่าง OCLC Online Computer Library Center และ The Research Libraries Group ของกลุ่มห้องสมุดในรัฐโอไฮโอแห่งสหรัฐอเมริกา [21] โดยมีแนวความคิดที่ว่า ส่วนประกอบที่สำคัญของการสร้างและจัดการทรัพยากรที่เป็นดิจิทัลคือ การทำให้เกิดความมั่นใจในสารสนเทศที่จำเป็นต่อการใช้งานต่อไปว่าได้รับการสงวนรักษาในสภาพที่สามารถเข้าถึงเพื่อใช้ประโยชน์ได้ คณะทำงานชุดนี้จึงได้ร่วมกันศึกษาเมื่อ พฤษภาคม ปี ค.ศ. 1997 โดยถูกขอให้ศึกษาการพรรณนาข้อมูลที่เกี่ยวข้องกับแฟ้มดิจิทัลที่ได้รับการสงวนรักษา แต่ในการศึกษาครั้งนี้จำกัดเฉพาะแฟ้มข้อมูลภาพดิจิทัล (digital image files) เท่านั้น โดยได้กำหนดองค์ประกอบของเมตาดาตาไว้ 16 รายการ

5. Library of Congress Digital Repository Development เป็นโครงการศึกษาของหอสมุดรัฐสภาอเมริกัน (Library of Congress) [22] ในช่วงเดือนธันวาคมปี ค.ศ. 1998 ถึงเดือนเมษายน ปี ค.ศ. 1999 โดยมีจุดประสงค์ในการอธิบายถึงแกนหลักขององค์ประกอบของเมตาดาตาที่ต้องใช้ในการพัฒนาการทดสอบ และการนำไปใช้สำหรับการจัดเก็บ (repository) ที่หลากหลาย โดยเมตาดาตาที่กำหนดขึ้นนี้ประกอบด้วย เมตาดาตาเชิงพรรณนา เชิงบริหาร และเชิงโครงสร้าง รวม 76 องค์ประกอบ

โครงการศึกษาวิจัยเพื่อกำหนดเมตาดาตาสำหรับการเก็บรักษาข้อมูลหรือเอกสารดิจิทัลต่าง ๆ ดังกล่าวข้างต้น ต่างก็มีรูปแบบในการเก็บเมตาดาตามลักษณะของข้อมูลที่จัดเก็บ แต่การกำหนดเมตาดาตาข้อมูลในลักษณะนี้ในประเทศไทยนั้น ยังไม่มีการกำหนดหรือมีการศึกษาอย่างเด่นชัด การศึกษาของงานวิจัยนี้เป็นเพียงจุดเริ่มต้นในการกำหนดรูปแบบบรรณานุกรมซึ่งถือว่าเป็นเมตาดาตาประเภทหนึ่ง โดยนำข้อมูลสถิติการใช้อินเตอร์เน็ตของประชาคมจุฬาลงกรณ์มหาวิทยาลัยซึ่งเป็นข้อมูลดิจิทัลมาเป็นกรณีตัวอย่างในการลงเมตาดาตา โดยศึกษาการกำหนดเมตาดาตาผ่านระบบห้องสมุดอัตโนมัติอินโนแพค (INNOPAC) ที่ใช้กันอยู่ในระบบเครือข่ายห้องสมุดจุฬาลงกรณ์มหาวิทยาลัย (Chulalinet) [23] ทั้งนี้ การลงเมตาดาตาหรือรายการระเบียบบรรณานุกรมในระบบอินโนแพคนั้นเป็นการลงผ่านระบบรายการบรรณานุกรมแบบมาร์ก (MARC)

การลงรายการบรรณานุกรมแบบ MARC

การลงรายการทางบรรณานุกรมเป็นงานหรือหน้าที่ส่วนหนึ่งของการควบคุมทางบรรณานุกรม (bibliographic control) ของบรรณารักษ์ ระบบห้องสมุดทั่วโลกมีมาตรฐานในการลงรายการทางบรรณานุกรม ซึ่งเป็นตัวกำหนดเมตาดาตาหรือคุณสมบัติพิเศษซึ่งตัวข้อมูลสาระและสิ่งตีพิมพ์ต่าง ๆ โดยมีกำเนิดและพัฒนาการมาจาก International Standards Bibliographic Description (ISBD) และต่อมาได้รับการพัฒนาโดยหอสมุดรัฐสภาอเมริกัน เป็นระบบมาร์ก (MARC: Machine Readable Cataloguing) เป็น

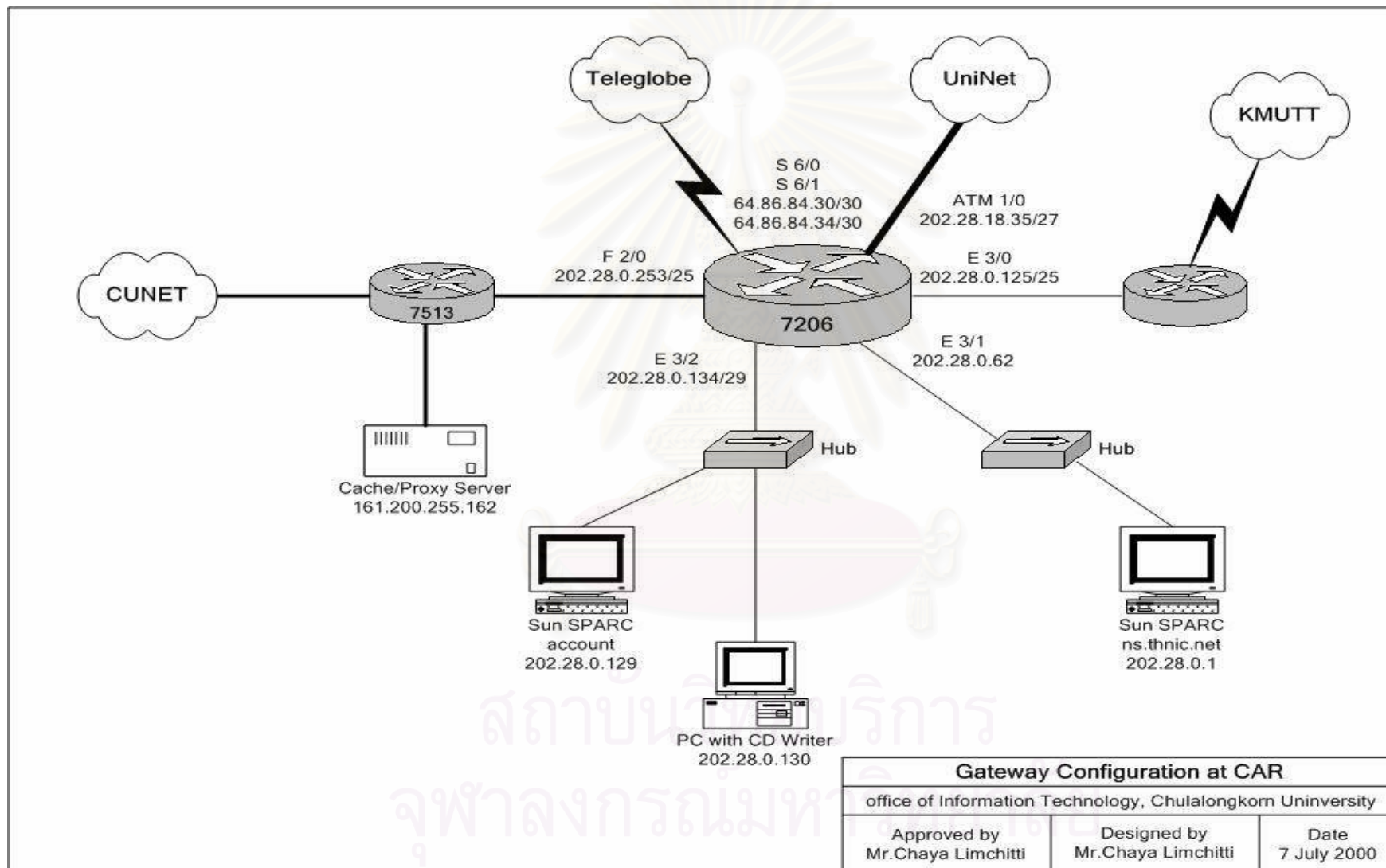
ยูเอสมาร์ก (USMARC) และเป็นมาร์ก 21 (MARC 21) ตามลำดับโดยใช้สำหรับการลงรายการทางบรรณานุกรมของทรัพยากรสารสนเทศที่อ่านได้โดยคอมพิวเตอร์ เพื่อเป็นมาตรฐานสำหรับแสดง จัดเก็บและสื่อสารข้อมูลทางบรรณานุกรมผ่านระบบอิเล็กทรอนิกส์ โดยการกำหนดเมตาดาตาผ่านตัวเลข รหัส และเขตข้อมูลต่าง ๆ ที่เป็นรายละเอียดทางบรรณานุกรม ประกอบด้วยกลุ่มของเขตข้อมูลตั้งแต่ 0xx, 1xx, 2xx, 3xx, 4xx, 5xx, 6xx, 7xx, 8xx และ 9xx

ลักษณะของข้อมูล

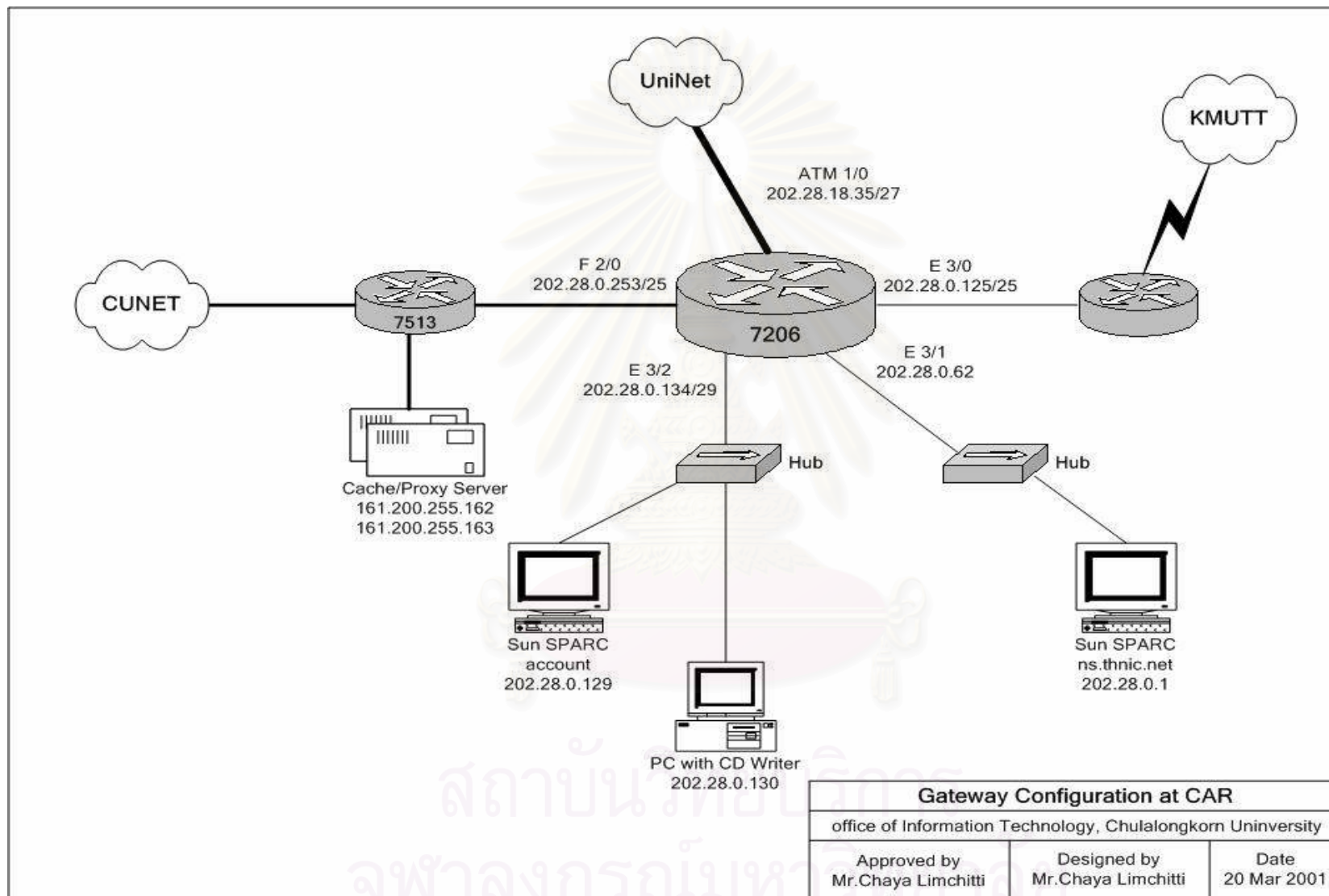
โครงสร้างของแหล่งกำเนิดข้อมูลยูนิเน็ตและซียูเน็ต

ข้อมูลสถิติการใช้อินเทอร์เน็ตที่เป็นกรณีศึกษาของงานวิจัยนี้ เป็นข้อมูลที่เกิดจากเครือข่ายจุฬาลงกรณ์มหาวิทยาลัยหรือเรียกว่า ซียูเน็ต (CUNet) ซึ่งเป็นเครือข่ายหลักของมหาวิทยาลัย เพื่อรองรับความต้องการการเชื่อมโยงและแลกเปลี่ยนข้อมูลข่าวสารระหว่างคณะ และหน่วยงานต่าง ๆ ภายในมหาวิทยาลัย นอกจากนี้ยังทำการเชื่อมต่อเครือข่ายซียูเน็ต เข้ากับเครือข่ายระหว่างมหาวิทยาลัยของทบวงมหาวิทยาลัย หรือเรียกว่า ยูนิเน็ต (UniNet) ซึ่งเป็นเครือข่ายสารสนเทศความเร็วสูงเชื่อมโยงมหาวิทยาลัย/สถาบัน และวิทยาเขตทุกแห่งทั่วประเทศ นอกจากเครือข่ายยูนิเน็ต จะทำการเชื่อมโยงสถาบันอุดมศึกษาต่าง ๆ ภายในประเทศแล้วยังมีการเชื่อมโยงไปยังเครือข่ายสากลอินเทอร์เน็ตของโลกอีกด้วย ทำให้เครือข่ายซียูเน็ต สามารถใช้งานอินเทอร์เน็ตได้โดยผ่านทางเครือข่ายยูนิเน็ต ดังมีแผนภูมิการเชื่อมโยงตามรูปที่ 1 และรูปที่ 2

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 1 ซียูเน็ตเชื่อมต่อกับอินเทอร์เน็ตผ่านบริษัทเทเลโกลบ



รูปที่ 2 ซียูเน็ตเชื่อมต่อกับอินเทอร์เน็ตผ่านยูนิเน็ต

จากรูปที่ 1 แสดงให้เห็นว่าเครือข่ายซียูเน็ตทำการเชื่อมต่อกับระบบอินเทอร์เน็ต โดยใช้ Cisco router รุ่น 7206 ซึ่งตั้งอยู่ที่สถาบันวิทยบริการ เป็น gateway ในการเชื่อมต่อกับบริษัทเทเลโกลบ (Teleglobe) ที่ความเร็ว 4 Mbps ต่อมาได้เปลี่ยนไปใช้การเชื่อมต่อกับยูนิเน็ต ในการติดต่อกับระบบอินเทอร์เน็ตแทน โดยมีความเร็ว 155 Mbps ในรูปที่ 2

ข้อมูลทั้งหมดที่รับส่งระหว่างเครือข่ายซียูเน็ต และระบบอินเทอร์เน็ตจะผ่าน Cisco router รุ่น 7513 ที่สำนักเทคโนโลยีสารสนเทศก่อน ส่วนข้อมูลประเภทอื่น ๆ จะถูกส่งผ่านเข้าสู่เครือข่ายได้โดยตรง

Cisco router ที่ใช้ IOS ตั้งแต่ version 11.1 (10.5) ขึ้นไป มีความสามารถเก็บรวบรวมข้อมูล header ของ packet ที่วิ่งผ่านเราเตอร์ได้ โดยจะทำงานร่วมกับโปรแกรม Netflowcollector ซึ่งติดตั้งอยู่บนเครื่องเซิร์ฟเวอร์เพื่อรวบรวมและสรุปข้อมูลที่ได้จาก Cisco router เพื่อใช้ทำสถิติและวิเคราะห์การใช้งานอินเทอร์เน็ตในเครือข่าย โดยงานวิจัยนี้ได้ใช้โปรแกรม Netflowcollector version 2.0 ทำงานอยู่บนเครื่อง Sun SPARC ซึ่งต่ออยู่กับ Cisco router รุ่น 7206 ทำการรวบรวมและสรุปข้อมูลที่รับจากเราเตอร์ทุกครั้งชั่วโมง และส่งต่อไปยังเครื่องคอมพิวเตอร์ที่มี CD Writer เพื่อบันทึกข้อมูลที่เตรียมเรียบร้อยแล้วลงสื่อบันทึกต่อไป สำหรับเพิ่มบันทึกการใช้งาน (log file) ของข้อมูลเว็บแคชที่อยู่บน proxy cache server จะถูกส่งไปยังเครื่องเซิร์ฟเวอร์ในเครือข่ายซียูเน็ต เพื่อทำการรวบรวมและบันทึกข้อมูลที่เรียบร้อยแล้วลงสื่อบันทึกอีกเช่นกัน

โครงสร้างของข้อมูล

ข้อมูลแทรฟฟิกและข้อมูลเว็บแคช มีลักษณะของโครงสร้างข้อมูลซึ่งสามารถสรุปได้ ดังรูปที่ 3 และมีลักษณะของแพ้มบันทึกการใช้งาน ดังรูปที่ 4

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

SOURCE 202.28.0.134|FORMAT A|AGGREGATION CallRecord|PERIOD 30|STARTTIME

1017593879|ENDTIME 1017595679|FLOWS 341669|MISSED 0|RECORDS 254624

161.200.131.156|136.186.1.23|3191|80|6|0|8|320|1|1017594830|1017594831|380

216.136.131.243|161.200.255.162|80|37461|6|0|19|24601|1|1017595068|1017595069|1268

161.200.255.162|216.136.131.243|37461|80|6|0|12|1849|1|1017595068|1017595069|1268

161.200.255.162|64.4.14.24|37920|80|6|0|5|667|1|1017595088|1017595088|468

64.4.14.24|161.200.255.162|80|37920|6|0|4|377|1|1017595088|1017595088|468

161.200.101.6|208.184.118.84|2529|80|6|0|6|360|3|1017593864|1017593957|21000

207.246.138.113|161.200.131.145|80|3545|6|0|8|1224|1|1017594619|1017594633|13292

161.200.131.145|207.246.138.113|3545|80|6|0|10|1140|1|1017594619|1017594637|17520

161.200.255.162|64.4.36.24|32289|80|6|0|5|678|1|1017594835|1017594836|468

64.4.36.24|161.200.255.162|80|32289|6|0|4|376|1|1017594835|1017594836|468

161.200.131.156|136.186.1.23|3192|80|6|0|17|680|1|1017594841|1017594843|2084

161.200.255.162|64.4.14.24|37921|80|6|0|5|644|1|1017595088|1017595088|468

}

ส่วนหัว

}

ส่วนตัว

รูปที่ 3 ตัวอย่างข้อมูลที่ถูกรวบรวมโดยโปรแกรม Netflowcollector

จุฬาลงกรณ์มหาวิทยาลัย

ตัวอย่างข้อมูลที่ได้จากการสรุปข้อมูลแบบ CallRecord โดยโปรแกรม Netflowcollector ดังรูปที่ 3 ซึ่งจะประกอบด้วย

ข้อมูลส่วนหัว จะบอกถึงรายละเอียดของการสรุปผลซึ่งประกอบด้วย

1. IP ของเราเตอร์ที่ส่งข้อมูลการจับเก็บ
2. รูปแบบการจับเก็บ A หมายถึงการจับเก็บแบบอักษระ
3. รูปแบบการสรุปรายงาน
4. ระยะเวลาที่จับเก็บ
5. เวลาที่เริ่มต้นการจับเก็บ
6. เวลาที่สิ้นสุดการจับเก็บ
7. จำนวนของโฟลว์ที่ได้รับ
8. จำนวนของโฟลว์ที่ไม่สามารถรับได้
9. จำนวนของระเบียบน

ข้อมูลส่วนตัว ประกอบด้วยระเบียบนของข้อมูลที่ได้รับการสรุปแล้วซึ่งมีเขตข้อมูลดังนี้

1. Key fields

1.1. Srcaddr	Source IP address
1.2. Dstaddr	Destination IP address
1.3. Srcport	TCP/UDP source port number
1.4. Dstport	TCP/UDP destination port number
1.5. Protocol	IP protocol type (example TCP=6, UDP=17)
1.6. ToS	IP type of service

2. Value fields

2.1. Packet count	Packet count as part of this record
2.2. Byte count	Total number of Layer 3 bytes counted as part of this record
2.3. Flow count	Total number of flows aggregated into this record
2.4. First Time Stamp	The time in UTC seconds, of the first packet summarized Into this record
2.5. Last Time Stamp	The time in UTC seconds, of the last packet summarized Into this record
2.6. Total Active Time	The sum of individual active time for all the flows summarized into the current record

```
1017594343.983 16 161.200.129.33 TCP_IMS_HIT/304 118 GET http://www.mweb.co.th/icon/a12expert.gif - NONE/- binary/*
1017594343.983 33 161.200.131.12 TCP_HIT/0 10769 GET http://www.eotoday.com/stars/scoop/pasttopresent/images/sound.swf - NONE/- binary/*
1017594344.0 0 161.200.129.30 TCP_IMS_HIT/304 90 GET http://mymail.maildozy.com/images/sort_none.gif - NONE/- binary/*
1017594344.17 0 161.200.129.33 TCP_IMS_HIT/304 118 GET http://www.mweb.co.th/icon/a13win.gif - NONE/- binary/*
1017594344.33 0 161.200.129.63 TCP_IMS_HIT/304 119 GET http://www.eotoday.com/images/menu10.gif - NONE/- binary/*
1017594344.83 516 161.200.128.9 TCP_MISS/0 4635 GET http://www.freemoviegalleries.com/asian/j2u/28/004t.jpg - DIRECT/www.freemoviegalleries.com binary/*
1017594344.83 0 161.200.129.110 TCP_HIT/0 5401 GET http://www.zonevcd.com/imageitem/jx0112/p.JPG - NONE/- binary/*
1017594344.100 16 161.200.129.33 TCP_IMS_HIT/304 118 GET http://www.mweb.co.th/icon/a14auction.gif - NONE/- binary/*
1017594344.100 50 161.200.129.30 TCP_MISS/0 693 GET http://ads.maildozy.com/adjs.php?what=zone:3&target=_blank&withText=0 - DIRECT/ads.maildozy.com
binary/*
1017594344.117 16 161.200.129.30 TCP_IMS_HIT/304 90 GET http://www.mrpalm.com/image/arrowgray.gif - NONE/- binary/*
1017594344.133 33 161.200.129.41 TCP_IMS_HIT/304 119 GET http://www.lfcmail.co.uk/images/software_multi_banner468.gif - NONE/- binary/*
1017594344.133 33 161.200.131.167 TCP_HIT/0 826 GET http://technology.mweb.co.th/image/go.gif - NONE/- binary/*
1017594344.150 466 161.200.129.87 TCP_MISS/0 443 GET http://egr.yahoo.com/serv?s=1705082353&t=1017594330&f=e - DIRECT/egr.yahoo.com binary/*
```

รูปที่ 4 ตัวอย่างข้อมูลของแฟ้มบันทึกการใช้งานของ Proxy Cache Server

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตัวอย่างข้อมูลของแฟ้มบันทึกการใช้งานของ Proxy Cache Server ดังรูปที่ 4 ประกอบด้วยเขตข้อมูลดังนี้

1. เวลาที่เกิด Request
2. ระยะเวลาที่ใช้ในการทำ Transaction
3. IP ของเครื่องที่ส่ง Request
4. ประเภท และสถานะของการ Request
5. จำนวนข้อมูล
6. วิธีการ Request
7. URL ที่ถูก Request
8. การพิสูจน์เครื่องที่ส่ง Request
9. ข้อมูลเพิ่มเติม
10. รูปแบบของเนื้อหา

การใช้ประโยชน์ของข้อมูล

ข้อมูลสถิติการใช้งานอินเทอร์เน็ตของประชาคมจุฬาลงกรณ์มหาวิทยาลัยนั้น มีประโยชน์ในการบริหารจัดการ และการวิจัยหรือการวิเคราะห์พฤติกรรมการใช้งานอินเทอร์เน็ตในลักษณะต่าง ๆ โดยกลุ่มผู้ใช้ข้อมูลแบ่งออกได้กว้าง ๆ เป็น สองประเภท คือ

1. กลุ่มเจ้าหน้าที่หรือหน่วยงาน ซึ่งบริหารระบบเครือข่ายและให้บริการอินเทอร์เน็ตของมหาวิทยาลัย สามารถใช้ข้อมูลที่เก็บถาวรเหล่านี้ในการวิเคราะห์การใช้งาน การขยายความต้องการ สามารถระบุประเภทของบริการที่ได้รับความนิยมเหล่านี้ เป็นต้น ช่วยให้ระบุปัญหาและสนับสนุนการวางแผนการจัดทำงบประมาณและการขยายการบริการอย่างเป็นระบบอีกด้วย ซึ่งสอดคล้องกับการวิเคราะห์สถิติการใช้งานของ UK National JANET Cache [24] ในการนำแฟ้มบันทึกการใช้งานมาวิเคราะห์เพื่อลดเวลาในการตอบกลับ (response time) และเพื่อลดค่าใช้จ่ายในการเชื่อมโยงการสื่อสาร นอกจากนี้ ยังมีการวิเคราะห์สถิติการใช้งานในเชิงวิจัยเพื่อสนับสนุนการบริหารของภาณุกู้พันธ์ สุวรรณมาตร [25] อีกด้วย

2. กลุ่มนักวิจัย ซึ่งใช้ข้อมูลเพื่อการวิเคราะห์การใช้งานเชิงสถิติอินเทอร์เน็ต การวิเคราะห์วิจัยเพื่อปรับปรุงอุปกรณ์ที่เกี่ยวข้อง เช่น การวิจัยประสิทธิภาพของเว็บแคชของจุฬาลงกรณ์มหาวิทยาลัย ได้แก่

- 2.1 งานวิจัยของนพรัตน์ พรหมมาไชย [26] ในการวิเคราะห์ความสัมพันธ์ระหว่างลักษณะของยูอาร์แอลกับการใช้งานบริการเว็บแคช เพื่อเป็นข้อมูลช่วยในการออกแบบเว็บแคชให้มีประสิทธิภาพมากขึ้น และสามารถนำไปเป็นข้อมูลช่วยในการพัฒนาโปรแกรมที่ใช้คำนวณสถิติการใช้งานบริการเว็บแคช

- 2.2 งานวิจัยของ พรทิว วัฒนวิฑูกร [27] การวิเคราะห์เปรียบเทียบประสิทธิภาพของขั้นตอนวิธีการแทนที่ในพรีอ็อกซีแคช เพื่อนำเสนอขั้นตอนวิธีการแทนที่ของพรีอ็อกซีแคชที่เหมาะสมกับสภาพการใช้เว็บของจุฬาลงกรณ์มหาวิทยาลัยซึ่งสามารถนำไปใช้เพื่อปรับปรุงประสิทธิภาพของเว็บแคชได้

- 2.3 งานวิจัยของ อัมพูนี พัฒนกำจร [28] การศึกษาวิธีการคงความต้องกันของเอกสารในเว็บแคช โดยใช้วิธีการส่งข้อความการเรียกขอแบบไอเอ็มเอสล่วงหน้า เพื่อนำเสนอแนวทางในการ

ลดปริมาณเอกสารที่มีค่า TTL ที่หมดอายุแล้วในเว็บแคช รวมทั้งลดปริมาณการใช้งานช่องสัญญาณที่มีอยู่ อย่างจำกัด ในช่วงเวลาที่มีการใช้เว็บมาก และยังคงความถูกต้องของเอกสารที่อยู่ในแคช ให้เหมาะสมกับ สภาพการใช้เว็บของจุฬาลงกรณ์มหาวิทยาลัย ซึ่งสามารถนำไปประยุกต์ใช้งานจริงได้

2.4 งานวิจัยของ รชช โรจนตรีคุณ [29] การวิเคราะห์ผลกระทบของจำนวนแคชและ วิธีการประสานงานที่มีประสิทธิภาพของแคชแบบกระจาย เพื่อนำเสนอถึงความเหมาะสมของจำนวนแคชในระบบ ประโยชน์ ตลอดจนข้อดีข้อเสียของการเพิ่มจำนวนแคชเข้าไปในระบบ รวมทั้งรูปแบบของการติดต่อ สื่อสารระหว่างแคชที่เลือกใช้กับสภาพการใช้เว็บจริงของจุฬาลงกรณ์มหาวิทยาลัย ซึ่งสามารถนำไปประยุกต์ ใช้จริงในระบบแคชต่อไป

ปัญหาของทั้ง 2 กลุ่ม คือ กลุ่มแรก ผู้ดูแลระบบเครือข่ายจะเก็บข้อมูลไว้ตามฮาร์ดดิสก์ของเครื่อง เซิร์ฟเวอร์ จึงมีปัญหาฮาร์ดดิสก์เต็ม และเนื่องจากเป็นอุปกรณ์มีราคาแพง การจัดซื้อจึงใช้เวลานาน จึงใช้วิธี ถ้ายโอนไปใส่ในฮาร์ดดิสก์ของเซิร์ฟเวอร์อื่น ๆ หรือของเครื่องพีซี รวมทั้งใส่ลงในเทปแม่เหล็กตามแต่จะหาได้ ซึ่งเป็นการเบียดบังฮาร์ดดิสก์ของระบบเครื่องในหน่วยงาน

ส่วนกลุ่มสอง นักวิจัยหรือนิสิตที่ทำวิจัย ต้องเก็บข้อมูลเองในลักษณะที่เป็น private lab ไม่มีการ เผยแพร่ หรือมีการจัดเก็บที่เป็นระบบ เมื่อจบการวิจัยก็ไม่มีการสานต่อในการเก็บข้อมูล

ข้อมูลที่เก็บสะสมไว้ของภาณุพันธ์ สุวรรณมาตรในการทำการวิเคราะห์และวิจัย ได้เป็นแรงจูงใจอัน สำคัญในการทำการวิจัยในครั้งนี้ เนื่องจากเป็นข้อมูลที่ควรจะให้ประโยชน์ได้หากเก็บไว้มากและนานพอ แต่ สภาพการเก็บตามความสะดวกด้วยทรัพยากรอันจำกัดและความไม่มีระบบการเก็บที่ชัดเจน ส่งผลให้คณะผู้วิจัย ของการวิจัยนี้เห็นความจำเป็นในการวางระบบบันทึกถาวรให้กับข้อมูลเหล่านี้

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 3 วิธีดำเนินการวิจัย

การวิจัยนี้เป็นการศึกษาการกำหนดรูปแบบบรรณานุกรมและการวางแผนทางการจัดเก็บแบบถาวรสำหรับข้อมูลต่อเนื่องที่ตัดมาเป็นกลุ่มให้สืบค้นผ่านห้องสมุดอัตโนมัติโดยอาศัยกรณีศึกษาเป็นข้อมูลสถิติการใช้งานอินเทอร์เน็ตของจุฬาลงกรณ์มหาวิทยาลัย การวิจัยมีขั้นตอนการดำเนินการโดยรวม ดังนี้

1. วิเคราะห์และเลือกสื่อบันทึกที่เหมาะสมในการบันทึกข้อมูลในระยะยาว
2. ศึกษาและวิเคราะห์สภาพการบันทึกข้อมูลที่มีอยู่ของปัจจุบัน
3. ออกแบบการบันทึกข้อมูลเพื่อการจัดเก็บแบบถาวร
4. รวบรวม ศึกษาและวิเคราะห์การลงเมตาดาตาในรูปแบบมาร์ก
5. กำหนดเมตาดาตาที่ใช้ประกอบข้อมูลสถิติที่บันทึก
6. ทดสอบการบันทึกข้อมูล
7. ศึกษาและสร้าง Web interface เพื่อเสริมการใช้งาน

การวิเคราะห์และเลือกสื่อที่ใช้บันทึก

ในการพิจารณาเลือกสื่อบันทึกที่เหมาะสม เพื่อนำมาบันทึกข้อมูลเพื่อการใช้งานในระยะยาวนั้น งานวิจัยนี้ได้วิเคราะห์พิจารณาโดยอาศัยรากฐานทางเทคโนโลยีของสื่อบันทึกโดยทั่วไป ซึ่งจากความหลากหลายของสื่อบันทึก จึงจะได้ทำการหาข้อมูลคุณสมบัติเปรียบเทียบระหว่างสื่อแบบต่างๆ แล้วจึงคัดเลือกสื่อที่มีความเหมาะสมที่สุดตามข้อมูลการวิเคราะห์ที่ได้ นอกจากนั้นประเด็นเรื่องอายุของสื่อจะได้นำมาพิจารณาเป็นพิเศษ เพราะเกี่ยวข้องกับกระบวนการในการเก็บรักษาในระยะยาว

เทคโนโลยีของสื่อบันทึก

ตารางที่ 1 แจกแจงตามประเภทของเทคโนโลยีที่ใช้ในการทำสื่อ ซึ่งแบ่งเป็นระบบแม่เหล็ก ระบบแสง และระบบอิเล็กทรอนิกส์ ส่วนตารางที่ 2 นั้น แจกแจงประเภทของผลิตภัณฑ์ที่ใช้เทคโนโลยีพื้นฐานแต่ละประเภท

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 1 เทคโนโลยีหลักของสื่อบันทึก

เทคโนโลยี	คุณสมบัติโดยรวม		
Magnetic	เทป	เข้าถึงข้อมูลแบบลำดับ ทำงานได้ช้า ราคาถูก	
	แผ่น	อ่อน	เข้าถึงข้อมูลแบบสุ่ม ช้า ความจุต่ำ เก็บข้อมูลได้ไม่นาน
		แข็ง	เข้าถึงข้อมูลแบบสุ่ม สามารถหมุนแผ่นได้เร็วกว่าแบบอ่อน ทำให้เข้าถึงข้อมูลได้เร็วขึ้น ความจุสูง พัฒนาการก้าวหน้าเร็วมาก
Optical	ใช้สื่อที่มีผลทางแสงหลายรูปแบบ ทำให้มีความยืดหยุ่นในการเก็บข้อมูล เทคโนโลยีปัจจุบันช้าและจุน้อยกว่าแบบแม่เหล็กแผ่นแข็ง แต่ในทางทฤษฎีจะจุได้มากและทำงานได้รวดเร็วในอนาคต		
Solid state	ทำงานโดยใช้สัญญาณไฟฟ้า ทำให้เข้าถึงข้อมูลได้เร็ว ไม่มีส่วนเคลื่อนไหว มีผลให้อายุใช้งานนาน ประหยัดพลังงาน ความจุต่ำ ต้นทุนการผลิตสูง		

ตารางที่ 2 ผลิตรหัสของสื่อแต่ละประเภท

เทคโนโลยี	ประเภท		ประเภทผลิตรหัส
Magnetic	เส้น	Cartridge	DLT tape
		Cassette	DDS tape
	แผ่น	แข็ง	Hard disk
			lomega Zip disk
			lomega Jazz disk
		อ่อน	Diskette
Optical	บันทึกตรงจากโรงงาน	CD-ROM	
		DVD-ROM	
	บันทึกได้ครั้งเดียว	CD-R	
		DVD-R	
	บันทึกได้หลายครั้ง	CD-RW	
		DVD-RW	
Solid state	Compact flash		
	SD memory card		
	Multimedia card		
	Memory stick		
	Smart media		

ในบรรดาเทคโนโลยีพื้นฐานตามตารางที่ 1 นั้น ระบบแม่เหล็กมีข้อเสียพื้นฐานอยู่ที่การเสื่อมสภาพของสถานะแม่เหล็กภายใต้เส้นแรงแสงแม่เหล็กโลก ซึ่งมีกระจายอยู่ตลอดเวลาในธรรมชาติ ทำให้ข้อมูลเกิดการเสื่อมตามกาลเวลา นอกจากนี้ การเก็บข้อมูลโดยใช้เทปแม่เหล็กยังมีข้อเสียพื้นฐานของระบบเทปแม่เหล็กโดยทั่วไป กล่าวคือ ต้องบริหาร (exercise) เทปแม่เหล็กเป็นระยะ โดยการรอกเทป เพื่อไม่ให้เนื้อเทปแม่เหล็กที่ทับซ้อนกันอยู่เกิดการหลอมติดกัน แต่ระบบแม่เหล็กนั้น ถือได้ว่าเป็นพื้นฐานของระบบเก็บข้อมูลของวงจรรวมพิวเตอร์ เพราะมีช่วงคุณสมบัติที่กว้าง นับแต่มีราคาถูก สามารถทำให้มีความจุสูง หรือในกรณีของจานแม่เหล็ก สามารถทำให้มีความเร็วในการบันทึกและอ่านได้รวดเร็ว

ความโดดเด่นของระบบแสงมีเพียง สื่อซีดี-รอม ซึ่งมีต้นกำเนิดในวงการเพลงและภาพยนตร์ โดยมีคุณสมบัติที่สำคัญ คือ รูปแบบที่กะทัดรัด มีระบบการบันทึกที่เป็นมาตรฐานและความจุที่สูง ทำให้ได้รับความนิยมอย่างรวดเร็ว เปรียบเทียบกับสื่อ optical disc ที่ออกมาก่อนหน้านี้สำหรับเก็บภาพยนตร์ ซึ่งเป็นแผ่นใหญ่แบบเดียวกับจานเสียงแบบลองเพลย์ (long-play) ที่เรียกว่า เลเซอร์ดิสก์ (laser disc) มีขนาดเทอะทะไม่สะดวกต่อการใช้งานและมีราคาแพง ความนิยมของซีดี-รอม ได้ก่อให้เกิดผล 2 ประการ คือ การผลิตจำนวนมหาศาลทั้งในเรื่องของแผ่นของสื่อ และอุปกรณ์อ่านสื่อ ทำให้เกิดการแพร่หลาย เทคโนโลยีมีราคาถูกลงกลายเป็นอุปกรณ์มาตรฐานบนเครื่องคอมพิวเตอร์ในยุคปัจจุบัน ประการที่ 2 คือ การพัฒนาศักยภาพของสื่อเกิดเป็นซีดี-อาร์ (CD-R) ซึ่งสามารถบันทึกได้ 1 ครั้ง และซีดี-อาร์ดับเบิล (CD-RW) ซึ่งสามารถบันทึกได้หลายครั้งตลอดจนความเร็วของเครื่องอ่านที่เร็วขึ้นถึง 50 เท่าของเครื่องอ่านในสมัยแรก ๆ

ปัจจุบัน คุณลักษณะทางกายภาพซึ่งสะดวกและเหมาะสมต่อการใช้งาน ได้ก่อให้เกิดปรากฏการณ์ที่น่าสนใจ คือการกำหนดรูปแบบทางกายภาพตรึงไว้เป็นมาตรฐานไม่มีการเปลี่ยนแปลง เป็นแผ่นกลมแบนขนาดเส้นผ่าศูนย์กลาง $4 \frac{3}{4}$ นิ้ว หรือ 120 มิลลิเมตร โดยใช้วิธีการสวมทับเทคโนโลยีใหม่ลงบนมาตรฐานทางกายภาพนี้ ผลที่เกิดขึ้น คือ ดีวีดี-รอม (DVD-ROM) ซึ่งมีความจุ 4.7 GB หรือ ประมาณ 9 เท่าของซีดี-รอม แต่มีขนาดของแผ่นเท่ากันทุกประการ อุปกรณ์ที่ใช้อ่าน ก็มีลักษณะที่ตัดเทียมกัน เครื่องอ่านดีวีดี-รอม สามารถอ่านซีดี-รอม จึงทำให้เกิดความต่อเนื่องของการใช้งานที่เรียกว่า downward compatibility ส่งผลให้เทคโนโลยีทั้งเก่าและใหม่ใช้โครงสร้างสนับสนุนต่าง ๆ เช่นกล่องใส่แผ่น ด้วยกันได้ รวมทั้งเป็นการยืดอายุของเทคโนโลยีเก่า คือ ซีดี-รอม ซึ่งสภาพการณ์เช่นนี้ ทำให้สภาพแวดล้อมของการใช้เทคโนโลยีนี้ ไม่เกิดการผันแปรอย่างรวดเร็วต่างจากเทคโนโลยีแบบอื่น เช่น เทปแม่เหล็กแบบ 8 track สำหรับเพลง ที่ล่าสมัยแม้แต่รูปแบบทางกายภาพของมัน พัฒนาการภายใต้มาตรฐานทางกายภาพนี้ก้าวหน้าไปอีกขั้นหนึ่ง โดยจะได้มีผลิตภัณฑ์ของแผ่นดีวีดีแบบที่ใช้เลเซอร์สีฟ้า (blue laser) ออกมาในปี ค.ศ. 2003 [30] แทนเลเซอร์สีแดง (ruby laser) ที่ใช้กันอยู่ในปัจจุบัน ซึ่งจะทำให้เก็บข้อมูลได้ถึง 27 GB ต่อแผ่นหรือประมาณ 6 เท่าของแผ่นดีวีดีปกติ ทำให้เทคโนโลยีของการเก็บข้อมูลภายใต้ข้อกำหนดทางกายภาพแบบนี้เป็นที่น่าสนใจมากที่สุดในปัจจุบัน

ในส่วนของเทคโนโลยีแบบอิเล็กทรอนิกส์ (solid-state memory) นั้น แม้ว่าในปัจจุบันจะมีความจุต่ำมาก ระดับ 1-200 MB และมีราคาสูงมาก แต่จากคุณสมบัติที่เป็น solid state คือไม่มีส่วนเคลื่อนไหวในการใช้งานและมีขนาดที่กะทัดรัดมากต่างจากระบบแม่เหล็กและระบบแสง ทำให้เริ่มได้รับความนิยมอย่างรวดเร็วในระบบที่มีข้อจำกัดในด้านขนาดทางกายภาพ เช่น กล้องถ่ายรูปดิจิทัล หรือเครื่องเสียงแบบพกพา อย่างไรก็ตามความหลากหลายในด้านมาตรฐานทางกายภาพ ความจุที่จำกัด ราคาที่สูง และพัฒนาการที่รวดเร็ว ทำให้เทคโนโลยีและสื่อประเภทนี้ไม่เหมาะสมต่อการนำมาใช้ในการบันทึกสาระจำนวนมากเป็นการถาวร

คุณสมบัติของสื่อบันทึก

ตารางที่ 3 แจกแจงคุณสมบัติต่างๆ ของสื่อบันทึกที่ควรพิจารณาในการคัดเลือก คุณสมบัติเหล่านี้ จำต้องครอบคลุมแง่มุมต่างๆ ในการนำสื่อเหล่านั้นมาใช้ จะใช้เพียงคุณสมบัติด้านความจุ ความเร็ว หรือราคา ตามหลักการในการเลือกสื่อบันทึกสำหรับเก็บข้อมูลในเครื่องคอมพิวเตอร์เพื่อการใช้งานในระยะสั้นแต่เพียงเท่านั้นหาได้ไม่ คุณสมบัติบางประการ เช่น ความสะดวกในการใช้งาน ก็เป็นคุณสมบัติที่สำคัญประการหนึ่งในการบ่งชี้ความเหมาะสมในการเก็บข้อมูลจำนวนมากเป็นระยะเวลาสั้น เพราะสื่อเหล่านั้นจำเป็นต้องอยู่นอกสภาวะแวดล้อมของการใช้งานคอมพิวเตอร์ ที่มักมีระบบไฟและระบบปรับอากาศที่ดี ควบคุมความชื้นและอุณหภูมิได้ตลอด การที่สื่อออกพ้นสภาพแวดล้อมเหล่านั้น และเข้ามาอยู่ในสภาวะแวดล้อมของการเก็บสาระในชั้นภายในห้องสมุด จำต้องผ่านการจับต้องจับถือบ่อยครั้ง รวมถึงการกระทบกระแทก การขนย้ายเป็นกอง หรือเป็นปริมาณมาก สื่อบางประเภทอย่างฮาร์ดดิสก์จะไม่สามารถทนต่อสภาวะการรบกวนเช่นนี้ได้

ตารางที่ 3 คุณสมบัติที่สำคัญของสื่อ

คุณสมบัติ	สิ่งที่พึงประสงค์
อายุ	เก็บรักษาได้เป็นเวลานาน
อายุของอุปกรณ์อ่านสื่อ	ทนทาน เทคโนโลยีไม่เปลี่ยนแปลงเร็วเกินไป
ความจุ	เก็บข้อมูลได้ปริมาณมาก
ขนาด	ไม่เทอะทะหรือเปราะบาง
ราคา	ราคาไม่แพง ทั้งตัวสื่อและอุปกรณ์ประกอบ
การเคลื่อนย้าย	จับถือเคลื่อนย้ายได้สะดวกต่อการปฏิบัติงาน
รูปแบบการบันทึก	แบบบันทึกได้ครั้งเดียวเหมาะกับการเก็บบันทึกแบบถาวร
มาตรฐาน	เป็นที่ยอมรับโดยทั่วไป มีหลากหลายยี่ห้อ ทั้งสื่อและอุปกรณ์
การเก็บรักษา	เก็บรักษาได้ง่าย ในสภาพแวดล้อมปกติ
ความเร็ว	ไม่ต้องการความเร็วมากนัก
รูปแบบการเข้าถึง	เข้าถึงข้อมูลแบบตรงเร็วกว่าเข้าถึงข้อมูลแบบลำดับ

ตารางที่ 4 แสดงถึงข้อมูลเปรียบเทียบสื่อประเภทต่าง ๆ ข้อมูลเหล่านี้บางส่วนได้จากหน้าโฆษณาขายอุปกรณ์และสื่อที่พบเห็นทั่วไปในอินเทอร์เน็ต

จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4 การเปรียบเทียบคุณลักษณะของสื่อ

เกณฑ์	Diskette	Zip disk	Jazz disk	Hard disk	Tape* (DLT)	CD-R	DVD-R
1. อายุสื่อ	N/A	2,000 cycles	10 years	5 years	30 years	70 years	100 years
2. อายุอุปกรณ์อ่านสื่อ	5 years	10,000 cycles	250,000 hours	5 years	50,000 hours	100,000 hours	100,000 hours
3. ความจุ	1.44 MB	250 MB	2 GB	Varied	40/80 GB	650 MB	4.7 GB
4. ขนาด	3 ½"	3 ½"	3 ½"	3 ½"	½"	4.72" (120 mm)	4.72" (120 mm)
5. ราคาสื่อ	7 ฿/MB	2.60 ฿/MB	2.20 ฿/MB	0.185 ฿/MB	0.10 ฿/MB	0.04 ฿/MB	N/A
6. ราคาอุปกรณ์อ่านสื่อ	440 ฿	4,200 ฿	12,000 ฿	3,700 ฿	136,000 ฿	1,400 ฿	3,200 ฿
7. ความสะดวกในการเคลื่อนย้าย และเก็บรักษา	สะดวก	สะดวก	สะดวก	ไม่สะดวก	สะดวก	สะดวก	สะดวก
8. รูปแบบการบันทึก	เขียนซ้ำได้	เขียนซ้ำได้	เขียนซ้ำได้	เขียนซ้ำได้	เขียนซ้ำได้	เขียนครั้งเดียว	เขียนครั้งเดียว
9. มาตรฐานของสื่อ และอุปกรณ์	มาตรฐาน	ไม่มาตรฐาน	ไม่มาตรฐาน	ไม่มาตรฐาน	หลายมาตรฐาน	มาตรฐาน	ยังไม่มาตรฐาน
10. ความเร็ว	500 Kb/s (2HD)	2.4 MB/s (SCSI)	8.7 MB/s (SCSI)	100 MB/s (ATA)	6/12 MB/s (SCSI)	33 MB/s (ATA)	33 MB/s (ATA)
11. รูปแบบการเข้าถึง	เข้าถึงข้อมูล แบบตรง	เข้าถึงข้อมูล แบบตรง	เข้าถึงข้อมูล แบบตรง	เข้าถึงข้อมูล แบบตรง	เข้าถึงข้อมูล แบบลำดับ	เข้าถึงข้อมูล แบบตรง	เข้าถึงข้อมูล แบบตรง

*Hi-End Representative

จากตารางที่ 4 คณะผู้วิจัยได้เปรียบเทียบคุณสมบัติของสื่อประเภทแม่เหล็กและแสงชนิดต่าง ๆ กันไว้ 7 ประเภทด้วยกัน [31][32][33][34][35][36] ในบรรดาสื่อทั้ง 7 ประเภทเหล่านี้ มีเพียงสองประเภทเท่านั้นที่มีมาตรฐานโลกชัดเจน คือ ดิสเก็ตต์และซีดี-อาร์ ในส่วนของดิสเก็ตต์นั้น แม้จะเป็นอุปกรณ์มาตรฐานประจำคอมพิวเตอร์เกือบทุกรุ่นในโลก และมีอยู่ในเครื่องคอมพิวเตอร์แบบพีซีทุกเครื่อง แต่ความจุที่ต่ำมากเพียง 1.44 MB ทำให้เหมาะกับการถ่ายโอนข้อมูลจำนวนไม่มากนัก ขนาดและอายุการใช้งานที่จำกัดทำให้ไม่เหมาะสมต่อการนำมาใช้เป็นที่สำหรับการบันทึกถาวร

ในส่วนของสื่อประเภทอื่น แม้จะมีขนาดความจุที่สูงมากและมีพัฒนาการที่ต่อเนื่องและรวดเร็วอย่างฮาร์ดดิสก์ แต่ความที่เป็นอุปกรณ์ที่ซับซ้อนโดยประกอบด้วยสื่อและอุปกรณ์อ่านบันทึกสื่อไว้เป็นชุดเดียวกัน และไม่มีความมาตรฐานในทางกายภาพของอุปกรณ์ ทำให้เกิดความไม่สะดวกต่อการเก็บรักษาในระยะยาว แม้ว่าในปัจจุบันจะได้มีการใช้ฮาร์ดดิสก์เป็นระบบสำรองข้อมูลกันมากขึ้นก็ตามก็ใช้กันเฉพาะการเก็บข้อมูลระยะสั้นเท่านั้น

นอกเหนือจากความจุ และความเป็นมาตรฐานของซีดี-อาร์แล้ว ราคาที่ถูกมากก็เป็นปัจจัยที่สำคัญประการหนึ่ง เพราะข้อมูลที่ถูกบันทึกให้ถาวรนั้นจะมีปริมาณสูง คุณสมบัติด้านราคา มีความผูกพันกับความเป็นมาตรฐานของเทคโนโลยี ทำให้เกิดความแพร่หลายอย่างสูง และมีรูปแบบให้เลือกใช้ถึง 3 แบบ คือ

1. แบบซีดี-รอม (CD-ROM) คือ บันทึกไม่ได้ อ่านได้อย่างเดียว (ROM คือ read only memory)
2. แบบบันทึกได้ครั้งเดียว อ่านได้หลายครั้ง (WORM - Write-Once-Read-Many) หรือเรียกว่า ซีดี-อาร์ (CD-R คือ Compact Disc Recordable)
3. แบบบันทึกซ้ำได้ (erasible) หรือ ซีดี-อาร์ดัดบลิว (CD-RW คือ Compact Disc Rewritable)

สื่อทั้งสามรูปแบบนี้ มีข้อดีที่สามารถใช้งาน (อ่าน) ได้โดยอุปกรณ์แบบเดียวกัน [37] ลักษณะของอุปกรณ์ประกอบที่สามารถใช้แทนกันได้ในรูปแบบที่แตกต่างกัน 3 รูปแบบนี้จะหาไม่ได้ในสื่อประเภทอื่น ไม่ว่าจะ เป็นฮาร์ดดิสก์หรือเทปแม่เหล็ก ทำให้สื่อประเภทนี้ครอบคลุมขอบเขตการใช้งานที่กว้างกว่าสื่อบันทึกประเภทอื่นอย่างมาก โดยซีดี-รอมจะใช้เป็น distribution media สำหรับเผยแพร่หรือจำหน่ายสินค้า ไม่ว่าจะ เป็นเพลง ภาพยนตร์ หรือโปรแกรมและฐานข้อมูล ซึ่งต้องการความคงทน ถาวร เปลี่ยนแปลงไม่ได้ของข้อมูล อีกทั้งยังมีราคาต้นทุนในการผลิตต่ำเมื่อผลิตจำนวนมาก ในขณะที่ซีดี-อาร์ จะใช้กับหน่วยงานที่มีข้อมูลที่ต้องการบันทึกถาวรจำนวนมาก เช่น เอกสาร จะสังเกตเห็นว่ารูปแบบทั้ง 3 ที่สามารถอ่านโดยใช้อุปกรณ์เดียวกันได้ ในลักษณะของ multi format นี้ได้กลายเป็นแนวปฏิบัติที่เป็นที่ยอมรับโดยทั่วไปในเทคโนโลยีของสื่อรุ่นถัดไป เช่น ประเภทดีวีดี (DVD คือ Digital Versatile Disc)

เหตุผลอีกประการหนึ่ง คือ เทปแม่เหล็กและฮาร์ดดิสก์ มีปัญหาที่ขนาดของสื่อที่มีพัฒนาการอย่างต่อเนื่อง มีการแข่งขันสูง ทำให้ไม่มีมาตรฐาน แต่ซีดี-รอมมีมาตรฐาน (มาตรฐานในที่นี้ หมายถึง มีข้อกำหนดทางเทคนิคที่ทำให้สื่อของยี่ห้อหนึ่งสามารถใช้อุปกรณ์อ่านบันทึกของอีกยี่ห้อหนึ่งได้)

คุณสมบัติอีกประการหนึ่ง ซึ่งทำให้สื่อแบบซีดีมีความเหมาะสม คือ รูปทรงทางกายภาพ ซึ่งมักถูกมองข้าม แต่รูปทรงทางกายภาพมีส่วนสำคัญอย่างสูงในการเก็บข้อมูลจำนวนมากไว้เป็นระยะเวลาอันยาวนาน ในสภาพของบันทึกถาวรเพราะต้องสะดวกต่อการหยิบฉวย จับต้อง สามารถวางทับซ้อนเป็นตั้ง หรือ วางเรียงเป็น

กลุ่มใหญ่ ทนทานต่อการเคลื่อนไหว การกระทบกระแทก หรือการหล่นตกในสภาพการปฏิบัติงานจริงของห้องสมุดโดยทั่วไป มีความคงทนต่อสภาวะแวดล้อม เช่น ความชื้น อุณหภูมิ และฝุ่นละออง ได้ตามสมควร ซึ่งจะพบว่าสื่อบันทึกแบบแสง โดยเฉพาะซีดี-รอมนั้น มีคุณสมบัติทางกายภาพที่เหมาะสมตามที่กล่าวมาทั้งหมด เนื่องจาก มีขนาดกะทัดรัดเบา เป็นแผ่นแบน สามารถหยิบฉวยได้ด้วยมือเดียว วางตั้งซ้อนหรือจัดเข้าชั้นได้เป็นจำนวนมากโดยไม่เปลืองเนื้อที่ ไม่ต้องการการหยิบฉวยอย่างระมัดระวัง และตัวสื่อไม่มีกลไกภายในต่างจากสื่อประเภทเทปที่มีแกนหมุนเหล่านี้ เป็นต้น

การพิจารณาอายุของสื่อ

ในด้านอายุ อายุของซีดี ประมาณได้ยาก ขึ้นกับวิธีการจัดและการเก็บรักษา [38,39] โดยปกติ จะอยู่ระหว่าง 70-200 ปี ในสภาพการใช้งานในสำนักงานหรือในบ้าน และได้ถึง 217 ปี ในสภาพแวดล้อมที่ได้รับการควบคุมอย่างดี [39] จากตารางข้างต้นนั้น มักจะไม่มีการกล่าวถึงอายุการเก็บ เพราะสื่อบันทึกส่วนใหญ่ใช้ในการสำรองข้อมูล หรือ การถ่ายโอนข้อมูล เช่น ในกรณีของดิสเก็ตต์ ซึ่งเป็นการเก็บข้อมูลในระยะสั้น ในขณะที่คุณสมบัติของซีดี มีความถาวร โดยเฉพาะในซีดี-รอมซึ่งบันทึกไม่ได้ จะมีอายุการใช้งานนานมาก เช่น ใช้เป็นสื่อในการซื้อขายเพลง และภาพยนตร์

องค์ประกอบอีกประการหนึ่ง ซึ่งมีผลต่อการกำหนดอายุของสื่อเป็นองค์ประกอบที่ไม่เกี่ยวกับคุณลักษณะทางกายภาพ เพราะข้อมูลดิจิทัลที่บันทึกลงไปบนสื่อแบบอิเล็กทรอนิกส์จำเป็นต้องมีการเข้ารหัสอักษร จัดรูปแบบ หรือมีกระบวนการดัดแปลงข้อมูลโดยใช้โปรแกรมต่าง ๆ เช่น โปรแกรมบีบอัดข้อมูลซึ่งทำงานอยู่บนระบบคอมพิวเตอร์ เช่น ไมโครคอมพิวเตอร์ที่ใช้ระบบปฏิบัติการวินโดวส์ องค์ประกอบของการจัดข้อมูลเหล่านี้มีพัฒนาการที่ค่อนข้างรวดเร็วในวงการ โปรแกรมหรือระบบเครื่องคอมพิวเตอร์ จะได้รับการพัฒนาเป็นรุ่นใหม่ออกมาทุก 6-12 เดือน และภายใน 2-3 ปี จะมีการเปลี่ยนแปลงประสิทธิภาพและลักษณะการทำงานไปโดยสิ้นเชิง เช่น พัฒนาการของระบบปฏิบัติการบนไมโครคอมพิวเตอร์ [40][41] ดังตารางที่ 5

ตารางที่ 5 พัฒนาการของระบบปฏิบัติการบนไมโครคอมพิวเตอร์

ระบบปฏิบัติการ	ปีที่เริ่มผลิต	ช่วงระยะเวลาการเปลี่ยนแปลง (ปี)
DOS	1980	5
Windows	1985	2
Windows 2.x	1987	3
Windows 3.x	1990	3
Windows 95	1995	5
Windows 98	1998	3
Windows 2000	2000	2
Windows XP	2001	1

จากตารางที่ 5 จะเห็นได้ว่า แม้แต่โปรแกรมที่เป็นระบบปฏิบัติการ ซึ่งเป็นพื้นฐานทางซอฟต์แวร์ของเครื่องคอมพิวเตอร์ ยังมีการเปลี่ยนแปลงที่ค่อนข้างรวดเร็ว โปรแกรมต่าง ๆ ที่เป็นโปรแกรมอรรถประโยชน์ ซึ่งทำงานอยู่บนระบบปฏิบัติการเหล่านี้ จะล้าสมัยไปตามระบบปฏิบัติการด้วย และแม้แต่ในช่วงของระบบปฏิบัติการตัวใดตัวหนึ่ง ผลิตภัณฑ์ที่เป็นโปรแกรมอรรถประโยชน์ยังมีแนวโน้มที่จะมีรุ่นใหม่ มีการเปลี่ยนแปลงออกมาอยู่เสมอ เพื่อให้เกิดความได้เปรียบทางการค้าต่อคู่แข่ง

ด้วยองค์ประกอบเชิงโปรแกรมที่ได้กล่าวมาแล้วนี้ จะเห็นได้ว่า ตัวกำหนดอายุของสื่อ จะเป็นเรื่องเกี่ยวกับโครงสร้างข้อมูลหรือสาระที่บรรจุในสื่อมากกว่าคุณสมบัติทางกายภาพ ในการวิจัยนี้ได้กำหนดอายุของสื่อหรือแผ่นซีดี ตามองค์ประกอบประการหลังนี้ โดยกำหนดว่าการเปลี่ยนแปลงของระบบปฏิบัติการและโปรแกรมอรรถประโยชน์ ควรจะเกิดขึ้นอย่างหักเห (discontinuous) ในช่วงประมาณ 5 ปี ดังนั้น เจ้าของ คอลเลกชันควรทำการตรวจสอบว่าสาระที่บรรจุอยู่ในสื่อ นั้น ยังสามารถใช้งานได้ เมื่อครบอายุ 5 ปี ทั้งนี้การตรวจสอบมิใช่กระทำในเชิงกายภาพว่าสามารถอ่าน bit และ byte จากแผ่นซีดีได้เท่านั้น แต่ยังหมายรวมถึงการนำ bit และ byte ที่อ่านได้ไปผ่านโปรแกรมอรรถประโยชน์ที่เกี่ยวข้องทั้งหลาย จนได้ผลลัพธ์ที่อยู่ในรูปที่นำไปใช้ประโยชน์ได้

กำหนดอายุ 5 ปีนี้เป็นเพียงการคาดการณ์ขั้นพื้นฐาน ยังต้องผ่านการวิเคราะห์และตรวจสอบในระยะยาว เมื่อคอลเลกชันมีการเริ่มให้บริการอย่างเป็นทางการ คณะนักวิจัยสังเกตเห็นว่า หากยึดอายุการเก็บเกิน 5 ปี จากตารางที่ 5 จะเห็นว่ามัลแวร์เสี่ยงสูง ที่ระบบคอมพิวเตอร์ในอนาคตจะไม่สามารถรองรับระบบปฏิบัติการหรือโปรแกรมอรรถประโยชน์ที่ใช้อยู่ได้ แต่ถ้าเก็บสื่อไว้ต่ำกว่า 5 ปี จะมีผลทำให้ระบบบำรุงรักษา คอลเลกชัน มีความซับซ้อนและเสียค่าใช้จ่ายสูงขึ้นโดยใช่เหตุ ทั้งนี้ ความเหมาะสมของช่วงอายุการเก็บก่อนที่จะมีการย้าย (migrate) ข้อมูล จะเป็นประเด็นสำคัญที่จะต้องวิเคราะห์ต่อไปในอนาคต

ระบบการบันทึกข้อมูลที่มีอยู่

จากการศึกษาสภาพการเก็บข้อมูลสถิติการใช้อินเทอร์เน็ตของจุฬาลงกรณ์มหาวิทยาลัย พบว่า มีการเก็บข้อมูลเป็น 2 ลักษณะ คือ

1. Traffic data คือ ข้อมูลการเข้า-ออกของการใช้อินเทอร์เน็ต
2. Web cache คือ ข้อมูลการเข้าใช้เว็บเพจตามหมายเลขไอพี

ข้อมูลทั้ง 2 ลักษณะดังกล่าว เกิดจากการเก็บข้อมูลของนิสิตมหาบัณฑิตของจุฬาลงกรณ์มหาวิทยาลัยเพื่อนำมาใช้ในงานวิจัยโดยข้อมูลแทรฟฟิกนั้น นำมาใช้ในงานวิเคราะห์การใช้งานเครือข่ายอินเทอร์เน็ตของจุฬาลงกรณ์มหาวิทยาลัย เพื่อเป็นต้นแบบข้อมูลสถิติประกอบการบริหารเครือข่ายของมหาวิทยาลัย เนื่องจากการเชื่อมต่อกับระบบอินเทอร์เน็ตเสียค่าใช้จ่ายสูงมาก ส่วนข้อมูลเว็บแคชนั้นใช้งานวิจัยของห้องปฏิบัติการวิศวกรรมระบบสารสนเทศ (Information System Engineering Laboratory – ISEL) [25,26,27,28] ซึ่งมุ่งเน้นการวิเคราะห์ข้อมูลการใช้งานเพื่อการปรับปรุงประสิทธิภาพของระบบแคชสำหรับการใช้เว็บในอินเทอร์เน็ต โดยมีลักษณะการเก็บข้อมูลตามตารางที่ 6 และ 7 ซึ่งเป็นสรุปขั้นตอนการจับข้อมูลทั้งแทรฟฟิกและเว็บแคช

ตารางที่ 6 ขั้นตอนการจับเก็บข้อมูลแพคเกจตามที่เป็นอยู่เดิม

ขั้นตอน	การดำเนินงาน
1	Export ข้อมูล packet ที่ผ่านเข้า / ออก router 7206
2	Aggregation ในรูปของแพคเกจข้อมูลโดยโปรแกรม NetFlow flowcollector version 2 บนเครื่อง Sun Sparc Ultra-1 ที่ใช้ระบบปฏิบัติการ Solaris 2.5.1 ทุก 30 นาที
3	นำแพคเกจมาห่อและบีบอัดลดขนาดกับเมตาเดตาที่บอกรายละเอียดเกี่ยวกับแพคเกจนั้น และแหล่งที่มาของข้อมูล
4	นำแพคเกจที่ได้มาบันทึกลงเทปแม่เหล็ก

ตารางที่ 7 ขั้นตอนการจับเก็บข้อมูลเว็บตามที่เป็นอยู่เดิม

ขั้นตอน	การดำเนินงาน
1	โอนย้ายข้อมูล log ของเว็บแคมมาที่เครื่อง PC ที่ใช้ระบบปฏิบัติการ Linux วันละ 1 ครั้ง โดยแบ่งเป็น 4 แพคเกจ ละ 6 ชั่วโมง โดยอยู่ในรูปของแพคเกจข้อมูล
2	นำแพคเกจทั้ง 4 มาต่อกันเป็น 1 แพคเกจต่อวัน
3	ทำการบีบอัดลดขนาด
4	นำแพคเกจที่ได้มาบันทึกลงแผ่นซีดี-อาร์ พร้อมกับเมตาเดตาเกี่ยวกับข้อมูลที่อยู่ภายในแผ่นซีดี-อาร์

การออกแบบการเก็บข้อมูลเพื่อให้จัดเก็บแบบถาวร

เนื่องจากข้อมูลนั้น กำเนิดจากข้อมูลต่อเนื่องตามการใช้งานอยู่ตลอดเวลา ทำให้ไม่สะดวกต่อการใช้งานตามสภาพดั้งเดิม ต้องแยกส่วนเป็นชิ้น ๆ ให้สามารถจัดการในลักษณะต่าง ๆ ได้โดยสะดวกต่อการปฏิบัติงาน เช่น บีบได้หรือบีบอัดได้ การวิจัยนี้จึงได้นำเสนอวิธีการจัดข้อมูลในรูปของ chunk

ในพจนานุกรม Webster's Third New International Dictionary Unabridged and Seven Languages Dictionary [42] ได้อธิบายคำว่า chunk ซึ่งใกล้เคียงกับลักษณะการใช้งานในงานวิจัยนี้ได้ว่า

1. A short thick piece often crudely or roughly form
2. A sizable amount; a note worthy quantity

จากคำอธิบายทั้ง 2 ลักษณะนี้ จะมองสภาพ chunk ว่า เป็นข้อมูลจำนวนหนึ่งที่เก็บในช่วงเวลาสั้น แต่มีปริมาณมากพอสมควรและอยู่ในสภาพที่เป็นข้อมูลดิบ เพราะฉะนั้นตามลักษณะของ chunk นี้ คือ ข้อมูลดิบ ซึ่งเก็บจากอุปกรณ์ซึ่งสร้างข้อมูลประเภทนี้ขึ้นมาในลักษณะที่เป็นอนุกรมตามเวลา (series by time) การตัดตอนข้อมูลอนุกรมนี้อาศัยหลักการที่เน้นความสะดวกในขั้นตอนการปฏิบัติเพื่อการจัดเก็บโดยมีข้อกำหนดดังนี้

1. เป็นวันโดยประมาณ
2. มีขนาดเหมาะสม

ข้อกำหนดที่ 1: คาบการตัดตอนข้อมูลออกเป็น chunk ข้อกำหนดนี้จำเป็นต่อการปฏิบัติงานของเจ้าหน้าที่และการนำข้อมูลไปวิเคราะห์ โดยกำหนดให้ขอบเขตของ chunk ครอบคลุม 1 วัน (24 ชั่วโมง) จากเที่ยงคืนถึงเที่ยงคืนโดยประมาณ ทั้งนี้ไม่อาจจะกำหนดช่วงเวลาให้ชัดเจนลงไปได้ เนื่องจากสภาพความต่อเนื่องของข้อมูล และข้อจำกัดของอุปกรณ์เราเตอร์ในการให้ข้อมูลออกมา การกำหนดคาบ 1 วันนี้ ทำให้สะดวกต่อการปฏิบัติงานและการนำข้อมูลไปใช้ เพราะสอดคล้องกับวันทำการปกติ อย่างไรก็ตาม จะต้องคำนึงถึงปริมาณข้อมูลที่เกิดขึ้นในแต่ละคาบเพื่อให้สัมพันธ์กับข้อกำหนดที่ 2 ด้วย หาก chunk มีขนาดใหญ่เกินไป คาบจะต้องมีขนาดเล็กลง แต่การกำหนดคาบจะยังคงให้สอดคล้องใกล้เคียงกับเวลาของวัน เช่น ลดลงจาก 24 ชั่วโมง เป็น 12 เป็น 6 และเป็น 3 ชั่วโมงตามลำดับ ตามการขยายตัวของปริมาณข้อมูลใน chunk

ข้อกำหนดที่ 2: ปริมาณข้อมูลใน chunk ปริมาณข้อมูลของ chunk ถูกกำหนดโดยอาศัยความสะดวกต่อการจัดการข้อมูลในขั้นตอนการปฏิบัติการของเจ้าหน้าที่เป็นหลัก โดยให้สัมพันธ์กับคาบเวลาการเก็บ chunk และความจุของสื่อที่ใช้ ซึ่งในการนี้ไม่มีสูตรคำนวณแน่นอน เป็นการประเมินประนีประนอมในเชิงปฏิบัติการโดยกำหนดเพียงแต่ว่า chunk ควรมีขนาดเล็กพอที่จะบรรจุได้ในเนื้อที่ไม่เกิน 60% ของแผ่นสื่อบันทึก ซึ่งในงานวิจัยนี้กำหนดให้ chunk ครอบคลุมคาบ 24 ชั่วโมงซึ่งมีขนาดประมาณ 200 MB หลังจากผ่านการบีบอัดข้อมูลแล้ว ทำให้บรรจุได้ 2 chunk ในหนึ่งแผ่นของซีดี-อาร์ ขนาด 650 MB ทั้งนี้การที่เหลือเนื้อที่ไว้ถึง 40% ของแผ่น ก็เพราะข้อมูลการใช้งานในแต่ละคาบไม่แน่นอน สามารถเพิ่มลดได้ค่อนข้างมาก อีกประการหนึ่งในการจำกัดขนาดของ chunk ไว้ที่ 200 MB เพราะทำให้สะดวกต่อการจัดการกับ chunk เช่นการถ่ายโอนข้อมูลระหว่างเครื่อง การทำการบีบอัดข้อมูล เหล่านี้เป็นต้น

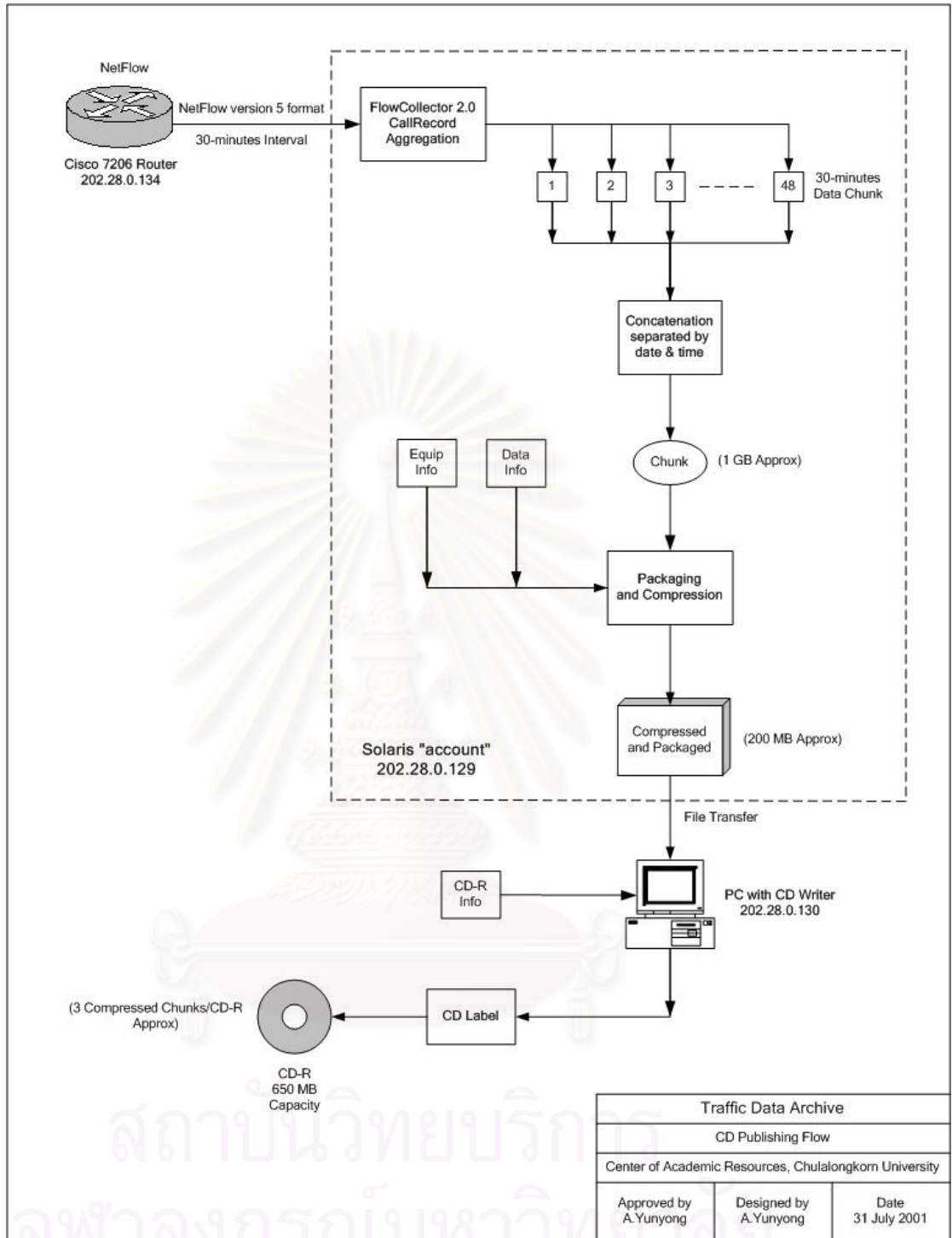
ตัวอย่างของการพิจารณาข้อมูลที่เป็นอนุกรมนี้ พบได้กับงานที่เป็นงานสำรวจ หรือสังเกตการณ์ปรากฏการณ์ธรรมชาติซึ่งมีข้อมูลเกิดขึ้นหรือตรวจจับได้ เพื่อให้บริหารงานได้จึงต้องตัดข้อมูลออกมาเป็นส่วน ๆ [43]

ระบบการบันทึกข้อมูลในแนวทางใหม่

เมื่อมีการตัดข้อมูลออกเป็นส่วนแล้ว คณะผู้วิจัยได้ศึกษาและวิเคราะห์ลักษณะการบันทึกข้อมูล ทำให้มีขั้นตอนในการบันทึกลงสื่อที่เป็นระบบมากขึ้น ดังแสดงในรูปที่ 5 และรูปที่ 6

จากรูปที่ 5 แสดงขั้นตอนการผลิต (หรือตีพิมพ์) ซีดี-อาร์ที่จะเป็นบันทึกถาวร โดยเริ่มจากข้อมูลดิบจากอุปกรณ์เครือข่ายเราเตอร์ หรืออุปกรณ์เว็บแคชโดยโปรแกรมพิเศษชื่อ FlowCollector ของบริษัทที่บรรจุอยู่ในคอมพิวเตอร์ชื่อ account ซึ่งเป็นเซิร์ฟเวอร์ทำงานภายใต้ระบบ Solaris จะทำการโอนย้ายข้อมูลอัตโนมัติทุก 30 นาที ทั้งนี้เพราะอุปกรณ์เครือข่ายเหล่านั้นมีหน่วยความจำจำกัด ไม่สามารถเก็บข้อมูลสถิติได้มากนัก

ข้อมูลที่เก็บมาทุก 30 นาทีนี้เป็น chunk ขนาดเล็ก จำนวน 48 ขึ้นต่อวัน จะถูกนำมาต่อกัน (concatenation) เป็น chunk ของวัน (24 ชั่วโมงโดยประมาณ) เมื่อผนวกด้วยเมตาเดตาของ chunk สองแฟ้มคือ ข้อมูลอุปกรณ์ (equipment information) และข้อมูลของข้อมูล (data information) ที่เป็นเมตาเดตา จะผ่านการห่อ (wrap) และบีบอัดลดขนาด (compress) โดยใช้โปรแกรม Zip ซึ่งเป็น freeware จากนั้น chunk ของแต่ละวันที่ได้นี้จะถูกส่งไปยังเครื่องพีซีเพื่อทำการบรรจุลงซีดี-อาร์ต่อไป



รูปที่ 5 ขั้นตอนการผลิตซีดี-อาร์ของบันทึกถาวร

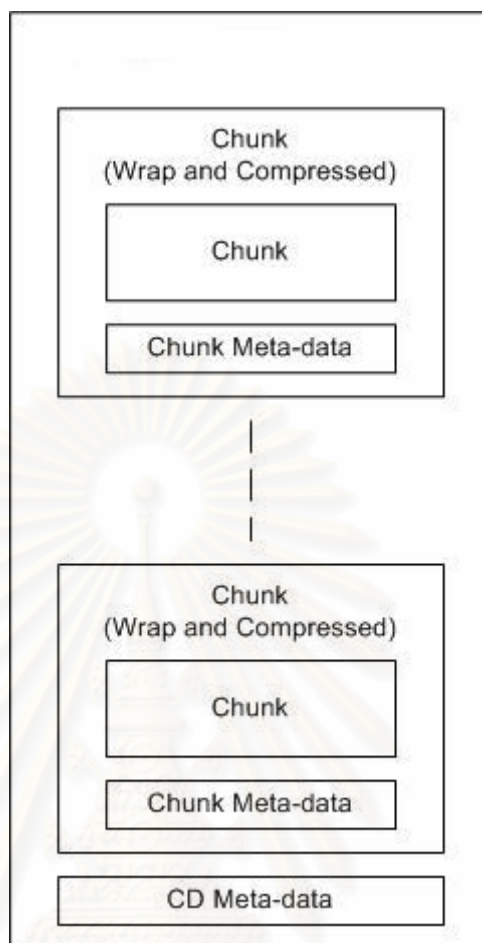
เครื่องพีซีที่มีอุปกรณ์บันทึกซีดี-อาร์ จะทำการสำเนาย้ายแฟ้ม (file transfer) ที่เป็น chunk ของวัน จากเครื่อง account จากนั้นจะรวบรวม chunk ของวันจำนวนหนึ่งให้มีปริมาณข้อมูลประมาณ 60% ของความจุของซีดี-อาร์ (ประมาณ 400 MB) จากนั้นจึงเพิ่มแฟ้มที่เป็นเมตาดาตาของซีดี-อาร์ แล้วจึงบันทึกลงซีดี-อาร์

ซีดี-อาร์ที่ผลิตได้จากกระบวนการนี้ จะถูกส่งต่อไปยังกระบวนการสร้างระเบียบบรรณานุกรมจากเมตาดาตาเหล่านี้ต่อไป

รูปที่ 6 แสดงโครงสร้างของข้อมูลที่เก็บไว้ในแผ่นซีดี-อาร์แต่ละแผ่น ที่ได้จากขั้นตอนการผลิตในรูปแบบที่ 5 โดยในแต่ละแผ่นจะประกอบด้วย chunk แต่ละวัน จำนวนหนึ่งบวกกับข้อมูลเมตาดาตาของแผ่นซีดี-อาร์นั้น ในแต่ละ chunk ของวันจะประกอบด้วยแฟ้มสองแฟ้ม ห่อรวมกันและบีบอัดไว้โดยแฟ้มแรกเป็นตัวข้อมูลสถิติของวัน และแฟ้มที่สองเป็นเมตาดาตาของแฟ้มแรก



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 6 โครงสร้างของข้อมูลในแผ่นซีดี-อาร์

ขั้นตอนการบันทึกซีดี-อาร์ ตามที่แสดงไว้ในรูปที่ 5 สามารถถอดเป็นกระบวนการ เป็นขั้นตอนสำหรับการบันทึกข้อมูลแทรฟฟิก ดังแสดงในตารางที่ 8 และสำหรับข้อมูลเว็บแคช ดังตารางที่ 9

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 8 ขั้นตอนการเก็บข้อมูลแพ็กเก็ตแบบใหม่

ขั้นตอน	การดำเนินการ
1	Export ข้อมูล packet ที่ผ่านเข้า / ออก router
2	ทำ aggregation ในรูปของแพ้มข้อมูลโดยโปรแกรม Netflow flowcollector version 2 บนเครื่อง Sun Sparc Ultra-1 ที่ใช้ระบบปฏิบัติการ Solaris 2.5.1 ทุก 30 นาที
3	นำแพ้มมาต่อกันเป็น chunk เพื่อความสะดวกในการจัดการ โดยจำนวนแพ้มที่ใช้ขึ้นกับขนาดของแพ้มและความสะดวกในการใช้งาน
4	นำ chunk มาห่อและบีบอัดลดขนาดกับเมตาเดตาที่บอกรายละเอียดเกี่ยวกับ chunk และแหล่งที่มาของข้อมูล
5	นำแพ้มที่ได้มาบันทึกลงแผ่นซีดี-อาร์พร้อมทั้งเมตาเดตาที่บอกรายละเอียดเกี่ยวกับข้อมูลที่อยู่ในแผ่นซีดี-อาร์

ตารางที่ 9 ขั้นตอนการเก็บข้อมูลเว็บแบบใหม่

ขั้นตอน	การดำเนินการ
1	โอนย้ายข้อมูล log ของเว็บแคมมาที่เครื่อง PC ที่ใช้ระบบปฏิบัติการ Linux วันละ 1 ครั้ง โดยแบ่งเป็น 4 แพ้ม ๆ ละ 6 ชั่วโมง โดยอยู่ในรูปของแพ้มข้อความ
2	นำแพ้มมาต่อกันเป็น 1 chunk ต่อวัน
3	นำมาห่อและบีบอัดลดขนาดกับเมตาเดตาที่บอกรายละเอียดเกี่ยวกับแพ้มและแหล่งที่มาของข้อมูล
4	นำแพ้มที่ได้มาบันทึกลงแผ่นซีดี-อาร์ พร้อมทั้งเมตาเดตาที่บอกรายละเอียดเกี่ยวกับข้อมูลที่อยู่ในแผ่นซีดี-อาร์

การรวบรวมศึกษาและวิเคราะห์การลงเมตาเดตา

การกำหนดโครงสร้างของระเบียบบรรณานุกรมมีขึ้นเพื่อให้ครอบคลุมหน้าที่หลักของการควบคุมทางบรรณานุกรมซึ่งมีอยู่ 2 ประการ [44] คือ ประการที่หนึ่งเพื่อการจัดหาช่องทางในการเข้าถึงข้อมูลได้อย่างรวดเร็วและมีประสิทธิภาพ และประการที่สองเพื่อช่วยในการอนุรักษ์ การบำรุงรักษา รวมทั้งเป็นการสืบทอดข้อมูลที่มีการบันทึกไว้

การกำหนดระเบียบบรรณานุกรมใน 2 ลักษณะดังกล่าวต้องมีตัวเชื่อมระหว่าง chunk กับคอลเลกชัน เพื่อให้สามารถกำหนดเขตข้อมูลที่สำคัญที่เกี่ยวข้องกับการอนุรักษ์ (preservation) การทำซ้ำ (duplication) และการถ่ายโอนข้อมูล (migration) เหตุผลที่สำคัญประการหนึ่งของการวางตัวเชื่อมเหล่านี้ คือ การทำระบบบันทึกถาวรต้องใช้หลักการ pre-active คือ ทำการ refresh ก่อนที่บันทึกนั้นจะเกิดการเสียหาย หรือเกิดการเปลี่ยนแปลงจนใช้ไม่ได้ วิธีการเก็บจึงแตกต่างจากวิธีการของการใช้สื่อที่เป็นกระดาษ เพราะในระบบกระดาษนั้น การชำรุดเป็นเพียงแต่ ขาดหรือหลุด ไม่มีผลหรือมีผลแต่น้อยต่อเนื้อหา สามารถซ่อม หรือบำรุงรักษาแบบ retroactive โดยไม่ทำให้เกิดการสูญหายของเนื้อหา

ในการวิจัยนี้ ได้ศึกษาและเทียบเคียงกับการลงเมตาดาตาข้อมูลดิจิทัลโดยเลือกเฉพาะโครงการของห้องสมุดรวม 5 โครงการ ได้แก่ โครงการ PANDORA Cedars NEDLIB OCLC/RLG และ LC (ตามรายละเอียดในภาคผนวก ค) เนื่องจากเป็นโครงการที่ค่อนข้างมีสภาพใกล้เคียงกับงานวิจัยนี้ ทั้งในด้านระบบจัดการ การจัดการ และผู้ปฏิบัติงาน ดังตารางที่ 10

ตารางที่ 10 เปรียบเทียบความสอดคล้องของเมตาดาตาระหว่างโครงการต่าง ๆ

โครงการ	PANDORA	Cedars	NEDLIB	OCLC/ORG	LC
เมตาดาตา					
1. วันที่สร้างข้อมูล	✓	✓	✓	✓	✓
2. ผู้สร้างข้อมูล / ผู้ผลิต / ผู้รับผิดชอบ	✓	✓		✓	
3. ฮาร์ดแวร์ / ซอฟต์แวร์ที่ใช้	✓		✓		
4. การเปลี่ยนแปลงของข้อมูลดิจิทัล	✓	✓	✓	✓	✓
5. สิทธิในการเข้าใช้	✓	✓		✓	✓
6. ลักษณะของแฟ้มข้อมูล	✓	✓	✓	✓	✓
7. การสงวนรักษาข้อมูล	✓	✓	✓		✓

จากการศึกษาเปรียบเทียบเมตาดาตาของข้อมูลดิจิทัลของโครงการข้างต้น พบว่า เป็นการกำหนดเมตาดาตาของข้อมูลดิจิทัลที่มีลักษณะต่าง ๆ กัน แต่โดยรวมแล้ว สามารถนำมาปรับใช้กับข้อมูลดิจิทัลประเภทอื่น ๆ ได้ แต่ทั้งนี้จะต้องมีการปรับเปลี่ยนให้เหมาะสมกับข้อมูลดิจิทัลที่ต้องการกำหนด เมตาดาตาของโครงการทั้ง 5 โครงการข้างต้นมีการกำหนดเมตาดาตาที่สอดคล้องกันส่วนใหญ่ในด้านวันที่สร้างข้อมูล ผู้สร้างข้อมูล / ผู้ผลิต / ผู้รับผิดชอบ ลักษณะของฮาร์ดแวร์ที่ใช้ การเปลี่ยนแปลงของข้อมูลดิจิทัล สิทธิในการเข้าใช้ ลักษณะของแฟ้มข้อมูล เช่น ลักษณะของข้อมูล การบีบอัดข้อมูล ขนาด เวลา และหมายเลขแฟ้ม เป็นต้น รวมทั้งในด้านการสงวนรักษาข้อมูล

การกำหนดเมตาดาตาสำหรับข้อมูลที่บันทึก

หลังจากที่ได้ศึกษาสื่อที่ใช้บันทึก ลักษณะของข้อมูล กำหนดขั้นตอนการบันทึกลงสื่อบันทึก และรูปแบบการลงรายการบรรณานุกรมต่าง ๆ แล้ว การวิจัยนี้ได้พิจารณารายการบรรณานุกรมแบบมาร์ก เพื่อเป็นตัวกำหนดเมตาดาตา เนื่องจาก

1. เป็นการลงรายการระเบียบบรรณานุกรมของทรัพยากรสารสนเทศที่อ่านได้โดยคอมพิวเตอร์ ที่เป็นมาตรฐานสำหรับแสดง จัดเก็บ และสื่อสารข้อมูลทางบรรณานุกรมระหว่างผู้ใช้กับผู้ให้บริการระหว่างคอมพิวเตอร์กับคอมพิวเตอร์

2. เป็นระบบการลงรายการระเบียบบรรณานุกรมที่นิยมใช้ในห้องสมุดในประเทศไทยและต่างประเทศมากที่สุด รวมทั้งห้องสมุดในเครือข่ายจุฬาลงกรณ์มหาวิทยาลัยได้ใช้ระบบนี้เช่นกัน ในการลงรายการระเบียบบรรณานุกรมของสารสนเทศแต่ละรายการ สามารถให้ผู้ใช้งานเข้าถึงเมตาดาตาของสารสนเทศของข้อมูลสถิติการใช้งานอินเทอร์เน็ตได้ดี เป็นที่คุ้นเคยของผู้ปฏิบัติงานและผู้ใช้

3. สามารถเก็บข้อมูลที่เป็นพรรณนาข้อมูล (descriptive metadata) ข้อมูลเชิงโครงสร้าง (structural metadata) และข้อมูลเชิงการบริหาร (administrative metadata) [45]

4. เป็นการนำข้อมูลเข้าไปเป็นส่วนหนึ่งของระบบปัจจุบัน (conventional system) ที่มีเสถียรภาพและใช้กันแพร่หลาย เนื่องจากมีทั้งโปรแกรมและเซิร์ฟเวอร์ที่เอื้ออำนวยในการสืบค้น

แม้ระบบมาร์ก จะมีความละเอียดครอบคลุม แต่เนื่องจากข้อมูลในงานวิจัยนี้ไม่อยู่ในรูปของสื่อปกติ จึงต้องวิเคราะห์ถึงคุณลักษณะพิเศษ (special attributes) ของข้อมูลการใช้อินเทอร์เน็ตเพื่อให้เหมาะสมกับการลงรายการในมาร์ก เช่น ลักษณะของข้อมูล ลักษณะทางกายภาพของสื่อ อายุ ลักษณะการบันทึกข้อมูล และอุปกรณ์ที่ดักเก็บ โปรแกรมที่ใช้ในการอ่านข้อมูล เป็นต้น เพื่อให้ผู้ใช้สามารถเข้าถึงเมตาดาตาของข้อมูลสถิติการใช้งานอินเทอร์เน็ตและใช้ประโยชน์ในการบำรุงรักษาข้อมูลด้วย

คณะผู้วิจัย ได้นำข้อมูลเมตาดาตาของโครงการห้องสมุดทั้งห้าโครงการเบื้องต้นเปรียบเทียบกับเมตาดาตาที่ได้บันทึกไว้ในซีดี-อาร์ มาทดลองกำหนดข้อมูลที่ต้องจัดเก็บ โดยต้องวิเคราะห์จากเมตาดาตา ที่ผู้บันทึกข้อมูลสถิติการใช้อินเทอร์เน็ตบันทึกไว้ ดังรูปที่ 7

1. CD Label: CD TR 2000-001, 554.0 MB.
2. ISSN
3. Chunk
3.1 Chunk1: CU-TR-2000-0001.zip, Fri 25 Aug 2000 17:50 - Fri 25 Aug 2000 23:50, 46.0 MB.
3.2 Chunk2: CU-TR-2000-0002.zip, Fri 25 Aug 2000 23:50 - Sat 26 Aug 2000 23:51, 166.0 MB.
3.3 Chunk3: CU-TR-2000-0003.zip, Sat 26 Aug 2000 23:51 - Sun 27 Aug 2000 23:51, 153.0 MB.
3.4 Chunk4: CU-TR-2000-0004.zip, Sun 27 Aug 2000 23:51 - Mon 28 Aug 2000 23:51, 188.0 MB.
4. Chunk characteristics
4.1 Meta-data consists of data and equipment files that contain data and equipment configurations, Respectively.
4.2 Data and Meta-data packaged and compressed by program "zip 2.1" on Solaris 2.5.1.
5. Write date: Sat 23 Sep 2000.
6. Expected expiry date: Fri 23 Sep 2005.
7. Hardware writer configuration: Pentium(r) Processor, HP SureStore CD-Writer Internal 6020i.
8. Software writer configuration: Microsoft Windows98 Second Edition, Adaptec Easy CD Creator 4.02.

รูปที่ 7 เมตาดาตาของข้อมูลในซีดี-อาร์ที่บันทึกข้อมูลแทรฟฟิก หมายเลข CD TR 2000-001

จากรูปที่ 7 เมตาดาตาของข้อมูลแทรฟฟิกในซีดี-อาร์จะประกอบด้วยเขตข้อมูล 8 ส่วน คือ

1. หมายเลขของซีดี-อาร์และขนาดของข้อมูลทั้งหมด
2. หมายเลขของ ISSN ในที่นี้หมายถึง International Standard Serial Number (ขณะนี้ยังมีได้ดำเนินการ)
3. จำนวน chunk หมายเลขของ chunk วันและเวลาที่เริ่มต้นและสิ้นสุดของข้อมูล และขนาดข้อมูลในแต่ละ chunk ซึ่งเขตข้อมูลนี้มีเขตข้อมูลย่อยแยกตาม chunk แต่ละชิ้น
4. ลักษณะของสภาพแวดล้อมในการเก็บข้อมูลของ chunk ได้แก่ องค์ประกอบของ อุปกรณ์ฮาร์ดแวร์ และรายละเอียดของซอฟต์แวร์ที่เกี่ยวข้อง
5. วันและเวลาที่บันทึกข้อมูลลงซีดี-อาร์
6. วันและเวลาที่คาดว่าซีดี-อาร์แผ่นนี้จะหมดอายุ
7. ลักษณะของฮาร์ดแวร์ที่เป็นตัวบันทึกข้อมูล
8. ลักษณะของซอฟต์แวร์ที่เป็นตัวบันทึกข้อมูล

จากเมตาดาตาที่เก็บไว้ในซีดี-อาร์ที่บันทึกข้อมูลสถิติการใช้อินเทอร์เน็ตนั้น ได้ให้ข้อมูลทางกายภาพ ข้อมูลของการเกิดข้อมูลในแต่ละวัน แต่ละช่วงเวลา การหมดอายุ ลักษณะของอุปกรณ์ที่เป็นตัวบันทึกและอ่าน ซึ่งสอดคล้องกับโครงสร้างของเมตาดาตาที่ได้จากโครงการต่าง ๆ ดังตารางเปรียบเทียบที่ 10

เนื่องจาก คณะผู้วิจัยประกอบด้วยผู้บันทึกข้อมูล ผู้ใช้ข้อมูล และบรรณารักษ์ จึงได้มีการปรึกษาหารือในลักษณะที่เป็น multi-disciplinary discussion ในการกำหนดเมตาดาตาร่วมกันเพื่อการสงวนรักษาข้อมูลระยะยาว และสอดคล้องกับความเป็นไปได้ในการลงข้อมูลระเบียบบรรณานุกรมของมาร์ก โดยเริ่มตั้งแต่การกำหนดชื่อเรื่อง การกำหนดการแสดงผลของข้อมูลเพื่อให้ข้อมูลมีการเรียงลำดับตามวันที่เก็บข้อมูล และสามารถดูจากการแสดงผลได้ว่ามีข้อมูลเรียงลำดับกันอย่างไร มีข้อมูลในช่วงเวลาใดที่หายไป ตลอดจนการสืบค้นข้อมูลเมื่อนำมาเทียบเคียงกับการเก็บข้อมูลดิจิทัลจากโครงการต่าง ๆ ตามตารางเปรียบเทียบที่ 10 กับเขตข้อมูลของมาร์ก 21 [46] สามารถแจกแจงเพื่อเปรียบเทียบระหว่างเขตข้อมูลและเมตาดาตาได้ดังตารางที่ 11

ตารางที่ 11 รายการเขตข้อมูลที่กำหนดในระเบียบบรรณานุกรม

รายการ	ชื่อของเขตข้อมูล	หมายเลขเขตข้อมูล
เลขหมู่ของแผ่นซีดี-อาร์	Local Call Number	099 * **
ชื่อเรื่อง	Title	245 # a * **
ชื่อผู้รับผิดชอบ	Statement of responsibility, etc	245 # c *
สถานที่ของผู้รับผิดชอบ	Place of publication, distribution, etc	260 # a *
ชื่อผู้จัดทำ	Name of publisher, distributor, etc	260 # b *
ปีที่บันทึกซีดี-อาร์ / ปีที่เกิดข้อมูล	Date of publication, distribution, etc	260 # c * **
จำนวนแผ่นซีดี-อาร์ / จำนวนของแฟ้ม	Physical description (Extent)	300 # a * **
ขนาดของแฟ้มข้อมูล	Dimensions	300 # c *
คาบการเกิดข้อมูล	Current publication frequency	310 # a *
ปีที่ผลิตข้อมูล และ/หรือลำดับของการเกิดข้อมูล	Dates of publication and/or sequential designation	362 # a *
หมายเหตุ	General note	500 # a * **
ข้อจำกัดในการเข้าถึง	Terms governing access	506 # a *
การเข้าถึงแผ่นบันทึก	Physical access provisions	506 # c *
ข้อกำหนดในการใช้	Authorized users	506 # d *
แหล่งที่เกิดข้อมูลและผู้บันทึก	Creation/production credits note	508 # a *
ขอบเขตของข้อมูล	Numbering peculiarities note	515 # a *
ลักษณะของข้อมูล	Type of computer file or data note	516 # a *
วัน เดือน ปี ที่บันทึกข้อมูลลงซีดี	Date/time and place of an event note	518 # a *
เนื้อเรื่องย่อ	Summary, etc. note	520 # a *
การอ้างถึงทางบรรณานุกรม	Preferred citation of described materials note	524 # a *
ผู้ดูแลข้อมูล	Custodian	535 # a *

ตารางที่ 11 รายการเขตข้อมูลที่กำหนดในระเบียบบรรณานุกรม (ต่อ)

รายการ	ชื่อของเขตข้อมูล	หมายเลขเขตข้อมูล
ที่อยู่ของผู้ดูแลข้อมูล	Postal address	535 # b *
ทุนที่ได้รับ	Text of note (Funding Information Note)	536 # a *
ข้อกำหนดของระบบ	System details note	538 # a *
เงื่อนไขการใช้งานและการทำสำเนาออก	Terms governing use and reproduction Note (Jurisdiction)	540 # b *
วิธีการได้ข้อมูล	Method of acquisition	541 # c *
หน่วยของข้อมูล	Type of unit	541 # o *
กรรมวิธีการประมวลข้อมูล	Methodology note	567 # a *
วัน เดือน ปี ที่คาดว่าซีดี-อาร์ หมดอายุ	Action	583 # a *
หัวเรื่องสำหรับหน่วยงานที่ เชื่อมโยงไปถึง	Subject Added Entry Corporate Name	610 # a *
หัวเรื่องสำหรับหมายเลขไอพีของ เครื่องและหัวเรื่องทั่วไป	Subject Added Entry – Topical Term	650 # a *
หน่วยงานที่ผลิตข้อมูล	Added Entry - Corporate Name	710 # a * **
ชื่อของ chunk	Host Item Entry: Title	773 # t **
หมายเลขแฟ้มข้อมูลของ chunk	Host Item Entry: Relationship information	773 # g **

* หมายถึง เขตข้อมูลที่ลงใน Bibliographical record of collection

** หมายถึง เขตข้อมูลที่ลงใน Bibliographical record of content analysis

นอกจากนี้ยังมีเขตข้อมูลที่ไม่เกี่ยวกับการลงรายการทางบรรณานุกรมในขณะนี้ แต่อาจมีเกี่ยวข้องกับในอนาคต เช่น เขตข้อมูล 533 คือ Reproduction Note และเขตข้อมูล 534 Original Version Note เนื่องจากคณะผู้วิจัยมีความเห็นว่า ในกรณีที่มีการขอทำสำเนาข้อมูลสถิติการใช้อินเทอร์เน็ตไปใช้โดยห้องสมุดสถาบันอุดมศึกษาอื่น ๆ ต้องมีการลงรายการในเขตข้อมูลนี้ด้วย ทั้งนี้แม้ว่าการทำสำเนาในระบบดิจิทัล จะสามารถประกันความถูกต้องตรงกันระหว่างตัวจริงกับสำเนาถึงขั้นแยกแยะจากกันไม่ได้ แต่สภาพของสื่อที่ใช้บรรณจะมีอายุต่างกัน เขตข้อมูลเหล่านี้จะบ่งบอกถึงตัวจริงและสำเนาเพื่อประโยชน์ในการ refresh สื่อต่อไป

ประเด็นสำคัญเกี่ยวเนื่องกับเขตข้อมูลหลักและสอดคล้องกับเมตาตาตาของโครงการทั้งห้า ได้แก่

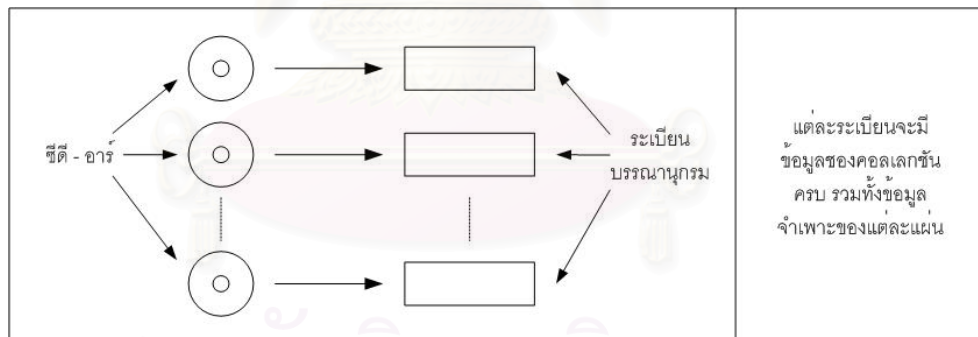
1. ผู้สร้างข้อมูล / ผู้ผลิต / ผู้รับผิดชอบข้อมูล คือ เขตข้อมูล 245, 508 และ 710
2. วันที่สร้างข้อมูล คือ เขตข้อมูล 260 # c
3. ฮาร์ดแวร์ / ซอฟต์แวร์ที่ใช้ คือ เขตข้อมูล 520, 538 และ 541
4. การเปลี่ยนแปลงของข้อมูลดิจิทัล คือ เขตข้อมูล 518 และ 583
5. สิทธิในการเข้าใช้ คือ เขตข้อมูล 506 และ 540
6. ลักษณะของแฟ้มข้อมูล คือ เขตข้อมูล 300, 515 และ 567

7. การสงวนรักษาข้อมูล คือ เขตข้อมูล 518 และ 583

ประเด็นที่สำคัญอีกประการหนึ่ง คือ การที่ข้อมูลสถิติมีความต่อเนื่องในระยะยาว จึงมีลักษณะคล้ายกับเอกสารในรูปของสิ่งพิมพ์ต่อเนื่อง (periodical หรือ series) อยู่บ้าง ดังนั้น ในการกำหนดเมตาดาตาเพื่อลงรายการระเบียบทางบรรณานุกรมในมาร์ก จึงต้องพิจารณาการลงรายการออกเป็น 2 ลักษณะ คือ

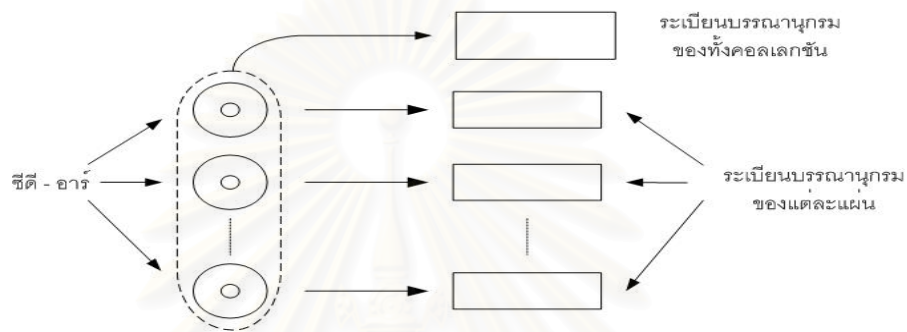
1. Bibliographic record of collection หมายถึง การลงรายการทางบรรณานุกรมของข้อมูลสถิติทั้งหมดที่เก็บไว้
2. Bibliographic record of content analysis หมายถึง การลงรายการทางบรรณานุกรมของข้อมูลในแต่ละช่วงเวลา

เหตุผลที่ต้องกำหนดออกเป็น 2 ลักษณะข้างต้น เนื่องจากข้อมูลสถิติการใช้อินเตอร์เน็ตที่บ้านที่กินนั้นเป็นข้อมูลวันต่อวัน ถ้ากำหนดการลงรายการบรรณานุกรมจากซีดี-อาร์ 1 แผ่น เป็น 1 ชื่อเรื่อง จะทำให้ไม่สามารถพรรณนาข้อมูลทั้งในด้านการพรรณนาข้อมูล ข้อมูลเชิงโครงสร้าง และข้อมูลเชิงการบริหาร และไม่สามารถอำนวยความสะดวกให้ผู้ใช้ในการลำดับเหตุการณ์ที่เกิดข้อมูลได้เลย อีกทั้งเป็นการเปลืองเนื้อที่ในกรบันทึก ข้อมูลของระบบอินโนแพค ซึ่งมีข้อจำกัดในการบันทึกข้อมูล 500,000 อักขระ ต่อ 1 ระเบียบ เนื่องจากต้องเพิ่มรายการชื่อเรื่องให้ตามวันและเวลาที่เกิดข้อมูล เพื่อให้ข้อมูลแสดงผลการเกิดข้อมูลได้ตามลำดับ และไม่สามารถระบุวันที่บันทึกข้อมูลลงซีดี-อาร์ รวมทั้งการหมดอายุของแต่ละแผ่นด้วย ซึ่งจะมีสภาพการหมดอายุลดหลั่นกันไปตามวันที่บันทึกข้อมูล ดังแสดงในรูปที่ 8 ซึ่งแต่ละระเบียบบรรณานุกรมในรูปจะเก็บข้อมูลไว้เกือบจะเหมือนกัน ทำให้เกิดความซ้ำซ้อนและเปลืองเนื้อที่



รูปที่ 8 การลงรายการแบบหนึ่งแผ่นต่อหนึ่งระเบียบ

จากข้อจำกัดนี้ ทำให้เห็นว่าการลงรายการระเบียบบรรณานุกรม ในลักษณะหนึ่งแผ่นต่อหนึ่งระเบียบบรรณานุกรม ไม่เหมาะสมกับการเก็บเอกสารดิจิทัลในระยะยาว จึงได้ทดลองการลงรายการเป็นลักษณะที่เป็นการทำกฤตภาค (clipping) บทความของวารสาร หนังสือพิมพ์ เพื่อให้ผู้ใช้สามารถเข้าถึงข้อมูลได้มากที่สุด โดยมีแนวความคิดที่ว่า ต้องมีระเบียบบรรณานุกรม 1 ระเบียบ ซึ่งเป็นระเบียบที่คุมทรัพยากรสารสนเทศของข้อมูลสถิติการใช้อินเตอร์เน็ตทั้งหมด เรียกว่าเป็นบรรณานุกรมของทั้งคอลเลกชัน และมีระเบียบบรรณานุกรมย่อยหรือเป็นการทำกฤตภาคของซีดี-อาร์ ที่บันทึกข้อมูลในแต่ละวันนั่นเอง ดังแสดงในรูปที่ 9



รูปที่ 9 การลงรายการแบบมีระเบียบหลักของคอลเลกชัน

การลงรายการระเบียบบรรณานุกรมในระดับคอลเลกชัน

ในการวิเคราะห์และลงรายการทางบรรณานุกรมของข้อมูลสถิติการใช้อินเตอร์เน็ตนั้น คณะผู้วิจัยได้วางแนวทางให้สร้าง bibliographical record of collection ใหม่ทุกครั้ง เมื่อมีการทำคอลเลกชันขึ้นใหม่ หรือเมื่อมีการเปลี่ยนแปลงการเชื่อมโยงของสายการสื่อสาร เพื่อใช้เป็นระเบียบเก็บข้อมูลของระบบการสื่อสารในแต่ละระบบ ในขณะนี้มีการทำ bibliographical record of collection ของข้อมูลสถิติการใช้อินเตอร์เน็ต 3 ระเบียบ เป็นระเบียบของข้อมูลแทรฟฟิก 2 ระเบียบ เนื่องจากมีการเปลี่ยนแปลงสายการสื่อสาร จากเดิมที่มีการเชื่อมโยงกับบริษัทเทเลโกลป และต่อมามีการยกเลิก จึงต้องมีกำหนดการระเบียบบรรณานุกรมของคอลเลกชันขึ้นใหม่ เพราะการทำ configuration ของระบบหรือของเครื่องไม่เหมือนกัน ดังแสดงในตารางที่ 12 และ 13 ส่วนอีก คอลเลกชันหนึ่งนั้น คือ ระเบียบของข้อมูลเว็บแคช 1 ระเบียบ ดังแสดงไว้ในตารางที่ 14

ตารางที่ 12 การลงรายการแบบ bibliographical record of collection ของข้อมูลแทรฟฟิกที่เชื่อมโยงกับบริษัท
เทเลโกลป

เขตข้อมูล	ตัวบ่งชี้	ลักษณะสมบัติ
099	2	Data Traffic Log
245	00	Internet traffic log h[Computer file] : bUninet gateway of CUNET, Teleglobe, Southern and Northern hub / cSystem Department, Center of Academic Resources, Chulalongkorn University
260		Bangkok : bSystem Department, Center of Academic Resources, Chulalongkorn University, c2000-2001
300		Computer optical laser discs (CD-R) ; c4 3/4 in
310		34 discs per year (approx.)
362	1	2000-2001
500		Location: IP address 202.28.0.1.125/25 - 10 Mbps connect to KMUTT's Router; IP address 202.28.18.35/27 - 155 Mbps connect to MUA ATM backbone; IP address 202.28.0.253/25 - 100 Mbps connect to CUNET; IP address 64.86.84.30/30 - 1984 Kbps connect to Teleglobe; Ipaddress 64.86.34/30 - 1984 Kbps connect to Teleglobe
506		Not available for publication or commercialization ; aRestricted online access through appointment and with password ; cPhysical access prohibited ; dFor use in researches granted by official bodies
508		Data stream originator, Office of Information Technology, Chulalongkorn University; Chunks and CD-ROMs produced from 2000-2001 , Nitass Sutaveepromochanon
515		Data covers August 2000 – April 2001
515		Missing data / Very irregular data (see chunk for details)
516		Numeric (Computer log file)
520		Contains data on the Internet traffic from router: Cisco 7206/IOS version 11.1 (17.3) CA. Aggregates data from program Netflowcollector (version 2), under CUNET administration and generated by account.car.chula.ac.th

ตารางที่ 12 การลงรายการแบบ bibliographical record of collection ของข้อมูลแทรฟฟิกที่เชื่อมโยงกับบริษัท
เทเลโกลป (ต่อ)

เขตข้อมูล	ตัวบ่งชี้	ลักษณะสมบัติ
524		Creator. "Name of Chunk Title." Name of CD-R title, Number of CD-R, Chunk ID. ตัวอย่าง Chulalongkorn University. Center of Academic Resources. "Internet traffic log 20011231 23:53 - 20020101 11:53." Traffic Traffic Log, CD TR 2002-001, File CU-TR-2002-0001.zip
535		Chulalongkorn University. bCenter of Academic Resources. bSystem Department
536		Original collection up to April 2002 supported by Rachadapiseksompoj Research Fund, Research Affairs, Chulalongkorn University
538		Data and metadata packaged and compressed by program "zip 2.1" on Solaris 2.5.1
540		bRelated privacy information must be removed or masked before publication of research result except by specific authorization from the owner
541		cContains the Internet traffic from router: Cisco 7206/IOS version: 11.1 (17.3) CA ; oChunk of 30-minute log data from router
567		30-minute chunk data stream
610	20	KMUTT
610	20	MUA
610	20	CUNET
610	20	Teleglobe
650	7	202.28.0.125/25
650	7	202.28.18.35/27
650	7	202.28.0.253/25
650	7	64.86.84.30/30
650	7	64.86.84.34/30
650	0	Internet
650	7	Internet traffic log file

ตารางที่ 12 การลงรายการแบบ bibliographical record of collection ของข้อมูลแทรฟฟิกที่เชื่อมโยงกับบริษัท เทเลโกลป (ต่อ)

เขตข้อมูล	ตัวบ่งชี้	ลักษณะสมบัติ
710	2	Chulalongkorn University. bCenter of Academic Resources. bSystem Department

ตารางที่ 13 เขตข้อมูลที่เพิ่มหรือเปลี่ยนแปลงการลงรายการแบบ bibliographical record of collection ของข้อมูลแทรฟฟิกที่เมื่อเลิกเชื่อมโยงกับบริษัทเทเลโกลป

เขตข้อมูล	ตัวบ่งชี้	ลักษณะสมบัติ
245	00	Internet traffic log h[Computer file] ; bUninet gateway of CUNET / cSystem Department, Center of Academic Resources, Chulalongkorn University
500		Location: IP address 202.28.0.1.125/25 - 10 Mbps connect to KMUTT's Router; IP address 202.28.18.35/27 - 155 Mbps connect to MUA ATM backbone; IP address 202.28.0.253/25 - 100 Mbps connect to CUNET
520		Contains data on the Internet traffic from router: Cisco 7206/IOS version 11.1 (17.3) CA; Changed to (35) CC2 in September 1st 2001. Aggregates data from program Netflowcollector (version 2), under CUNET administration and generated by account.car.chula.ac.th
541		cContains the Internet traffic from router: Cisco 7206/IOS version: 11.1 (17.3) CA and (35) CC2 since 1st September 2001; cChunk of 30-minute log data from router

ตารางที่ 14 การลงรายการแบบ bibliographical record of collection ของข้อมูลเว็บแคช

เขตข้อมูล	ตัวบ่งชี้	ลักษณะสมบัติ
099	2	Web Traffic Log
245	00	Web traffic log h[Computer file] ; bWeb campus network usage / cSystem Department, Center of Academic Resources, Chulalongkorn University
260		Bangkok ; bSystem Department, Center of Academic Resources, Chulalongkorn University, c2001-
300		Computer optical laser disc (CD-R) ; c4 3/4 in
310		disc per year (approx.)
362	1	2001-

ตารางที่ 14 การลงรายการแบบ bibliographical record of collection ของข้อมูลเว็บแคช (ต่อ)

เขตข้อมูล	ตัวบ่งชี้	ลักษณะสมบัติ
500		Location: IP address 161.200.255.162 (proxy 2); 161.200.255.163 (proxy 3)
506		Not available for publication or commercialization ; aRestricted online access through appointment and with password ; cPhysical access prohibited ; dFor use in researches granted by official bodies
508		Data stream originator, Office of Information Technology, Chulalongkorn University; Chunks and CD-ROMs produced from 2001- , Nitass Sutaveepramochanon
515		Data covers Oct 2001-
515		Missing data / very irregular data (see chunk for details)
516		Numeric (Computer log file)
520		Contains log data from Web cache engine under CUNET administration and generated by proxy2.chula.ac.th and proxy3.chula.ac.th
524		Creator. "Name of chunk title." Name of CD-R title, Number of CD-R, Chunk ID. ตัวอย่าง Chulalongkorn University. Center of Academic Resources. "Web traffic log 20011231 00:00 - 20020101 00:00 proxy 2, CD WT 2002-001, File CU-WT-2002-0001.zip.
535		Chulalongkorn University. bCenter of Academic Resources. bSystem Department
536		Original collection up to April 2002 supported by Rachadapiseksompoj Research Fund, Research Affairs, Chulalongkorn University
538		Data and metadata packaged and compressed by program zip v.2.3 on Linux v. 2.4.3
540		bRelated privacy information must be removed or masked before publication of research result except by specific authorization from the owner
541		cContains log data from Web cache engine: Cisco 550/CE 2.30 (7); oChunk of 6-hour log data from cache engine
567		6-hour chunk data stream
610	2	CUNET
650	7	161.200.255.162
650	7	161.200.255.163

ตารางที่ 14 การลงรายการแบบ bibliographical record of collection ของข้อมูลเว็บแคช (ต่อ)

เขตข้อมูล	ตัวบ่งชี้	ลักษณะสมบัติ
650	0	Internet
650	0	World Wide Web
650	0	Web servers
650	7	Web cache
710	2	Chulalongkorn University. bCenter of Academic Resources. bSystem Department

เมื่อดูค่าที่ลงในรายการทางบรรณานุกรมในระดับคอลเลกชันนั้น จะลงเขตข้อมูลเกือบทุกเขตข้อมูลตามตารางที่ 11 มีจุดสำคัญที่เป็นความแตกต่างระหว่างการลงรายการทางบรรณานุกรมในแต่ละช่วงเวลา คือ

เขตข้อมูล 260 ลงรายการปีที่บันทึกข้อมูล

เขตข้อมูล 300 ลงรายการลักษณะทางกายภาพของแผ่นบันทึกข้อมูล

เขตข้อมูล 310 ลงรายการจำนวนแผ่นที่เกิดขึ้นในแต่ละช่วงของปี หรือของการใช้เครื่องนั้น ซึ่งเป็นเสมือนตัวคุมคอลเลกชัน

เขตข้อมูล 362 ลงรายการปีที่เริ่มต้นของการบันทึกข้อมูล

เขตข้อมูล 5xx ลงรายการเกี่ยวกับระบบต่าง ๆ ของการบันทึกข้อมูล ข้อกำหนดในการเข้าใช้ตลอดจนการจัดการแฟ้มข้อมูล ซึ่งเขตข้อมูลเหล่านี้จะไม่มีปรากฏในรายการทางบรรณานุกรมของข้อมูลแต่ละช่วงเวลา

การลงรายการทางบรรณานุกรมของข้อมูลแต่ละช่วงเวลา

การลงรายการทางบรรณานุกรมของข้อมูลแต่ละช่วงเวลา (bibliographical record of content analysis) เป็นการลงรายการข้อมูลที่เก็บไว้ในแต่ละวัน แต่ละเวลา เรียงลำดับตามวันที่เก็บข้อมูล โดยมีลักษณะดังตารางที่ 15

ตารางที่ 15 การลงรายการแบบ bibliographical record of content analysis ของข้อมูลแทรฟฟิก

เขตข้อมูล	ตัวบ่งชี้	ลักษณะสมบัติ
099	2	CD TR 2002-001
245	00	Internet traffic log 20011231 23:53 – 20020101 11:53
260		c2002
300		1 log file (1.9 MB)
500		Log for 31 Dec 2001 23:53 - 01 Jan 2002 11:53
500		Usage through online access on appointment with staff
518		CD published on Jan 25, 2002
583		Expected expiry date; c20070125

ตารางที่ 15 การลงรายการแบบ bibliographical record of content analysis ของข้อมูลแทรฟฟิก (ต่อ)

เขตข้อมูล	ตัวบ่งชี้	ลักษณะสมบัติ
650	0	Internet
650	7	Internet traffic log file
710	2	Chulalongkorn University. bCenter of Academic Resources. bSystem Department
773	0	tInternet Traffic Log,lgCD TR 2002-001, File CU-TR-2002-0001.zip

การลงรายการทางบรรณานุกรมของข้อมูลแต่ละช่วงเวลานั้น มีการลงรายการเขตข้อมูลที่แตกต่างจากการลงรายการทางบรรณานุกรมในระดับคอลเลกชัน คือ

เขตข้อมูล 245 ลงรายการเป็นปี เดือน วัน เวลาที่เกิดข้อมูล - ปี เดือน วัน เวลาที่สิ้นสุดข้อมูล ในลักษณะที่เป็นตัวเลข เพื่อให้ระบบเรียงตามลำดับวันที่เกิดข้อมูลได้

เขตข้อมูล 260 ลงรายการเพียงปีที่บันทึกข้อมูล

เขตข้อมูล 300 ลงรายการเป็นลักษณะข้อมูลที่เกิดเป็น 1 แฟ้มข้อมูล และขนาดของแฟ้มข้อมูล

เขตข้อมูล 500 ลงรายการการเกิดข้อมูลและสิ้นสุดข้อมูลเหมือนเขตข้อมูล 245 แต่เป็นเชิงข้อความ

เขตข้อมูล 500 ลงรายการลักษณะการเข้าใช้

เขตข้อมูล 518 ลงรายการวัน เดือน ปี ที่บันทึกข้อมูล

เขตข้อมูล 583 ลงรายการวัน เดือน ปี ที่คาดว่าแผ่นบันทึกนี้จะหมดอายุ

เขตข้อมูล 6xx ลงรายการหัวเรื่อง

เขตข้อมูล 710 ลงรายการผู้รับผิดชอบการผลิตแผ่นบันทึกนี้

เขตข้อมูล 773 ลงรายการชื่อของแผ่นบันทึกข้อมูล ชื่อของ chunk และหมายเลขของ

chunk

จะเห็นได้จากตารางที่ 15 ว่าปริมาณข้อมูลในระเบียบบรรณานุกรมของแต่ละแผ่นของซีดี-อาร์วี มีจำนวนน้อยมากเปรียบเทียบกับข้อมูลในระเบียบบรรณานุกรมของคอลเลกชัน ในตารางที่ 12 และ 14 ทั้งนี้เป็นการสนับสนุนประโยชน์ของการแยกระเบียบบรรณานุกรมออกเป็นสองระดับ ดังแสดงไว้ในรูปที่ 9 ที่กล่าวแล้ว

อนึ่ง คณะผู้วิจัยมีความคิดในการจัดทำระเบียบบรรณานุกรมของโปรแกรมบีบอัดข้อมูลที่ใช้ในการวิจัยนี้ขึ้นอีก 1 ระเบียบ โดยขณะนี้เก็บเพียงโปรแกรมไว้ในซีดี-อาร์วีเท่านั้น เนื่องจากโปรแกรมมีการพัฒนาเปลี่ยนแปลงอย่างรวดเร็ว จึงจำเป็นต้องมีการเก็บโปรแกรมที่ใช้ในช่วงของการเก็บบันทึกข้อมูลนี้ไว้ด้วย และเพื่อสามารถนำกลับมาใช้กับข้อมูลนี้ได้เหมือนเดิม

ทดสอบการบันทึกข้อมูล

เมื่อสามารถกำหนดเมตาตาตาของข้อมูลสถิติการใช้อินเทอร์เน็ตทั้งข้อมูลที่เป็นแทรฟฟิก และเว็บแคชกับการลงรายการบรรณานุกรมได้แล้วนั้น ต้องนำข้อมูลมาบันทึกในระบบห้องสมุดอัตโนมัติอินโนแพค เพื่อใช้

ผ่านโอแพค (OPAC-On-line Public Access Catalog) [47] ซึ่งเป็นระบบที่ห้องสมุดในเครือข่ายจุฬาลงกรณ์มหาวิทยาลัยใช้ในการลงข้อมูลเพื่อให้บริการข้อมูลสารสนเทศ โดยบันทึกตามวิธีการบันทึกข้อมูลของระบบอินโนแพค แต่เรียกใช้เซตข้อมูลเฉพาะที่เกี่ยวข้องกับการลงรายการนี้เท่านั้น และเมื่อบันทึกแล้วจะต้องมีการตรวจสอบความถูกต้องของการบันทึกข้อมูล ซึ่งจุดที่มีการบันทึกผิดพลาดมาก ได้แก่

1. เซตข้อมูลที่เป็นชื่อเรื่อง (tag 245) ในส่วนที่เป็นช่วงของวันที่และเวลา ซึ่งเป็นตำแหน่งที่สำคัญมากในการบันทึกข้อมูลและสามารถทำให้ผู้ใช้ข้อมูลเข้าใจผิดได้ เนื่องจากรูปแบบเฉพาะของวันที่ และเวลาในการเก็บข้อมูล เป็น ปี ค.ศ. ตามด้วยวันและเวลา เช่น Internet traffic log 20020102 22:13 – 20020102 23:30 ทำให้มีโอกาสผิดพลาดในการบันทึกข้อมูลได้ง่าย จึงจำเป็นต้องมีความละเอียดรอบคอบ และมีการตรวจสอบซ้ำ ทั้งนี้การบันทึกข้อมูลในลักษณะดังกล่าว มีประโยชน์สำหรับการเรียงลำดับของการเกิดข้อมูลในระบบเพื่อเอื้ออำนวยต่อการค้นหา

2. เซตข้อมูลที่เป็นหมายเลขของซีดี-อาร์ (tag 099) ซึ่งถ้ามีการบันทึกผิดพลาดจะทำให้ผู้ใช้ ข้อมูลหาแผ่นบันทึกข้อมูลที่ต้องการไม่พบ

3. เซตข้อมูลหมายเลขของ chunk (tag 773) มักพบว่ามีการพิมพ์ผิดพลาด เป็นผลทำให้ผู้ใช้ไม่สามารถเรียกใช้ข้อมูลที่ต้องการได้ถูกต้อง

การเรียกใช้ข้อมูลเพื่อทดสอบการบันทึกข้อมูล สามารถทำได้โดยการสืบค้นจากชื่อเรื่อง ตามตารางที่ 16 ทั้งนี้ เมื่อสืบค้นด้วยชื่อเรื่องแล้ว เช่น Internet traffic log จะได้ผลลัพธ์จากโอแพค เช่นดังรูปที่ 10

ตารางที่ 16 การสืบค้นที่เหมาะสมในระดับคอลเลกชัน

Collection	วลีที่ใช้ค้นจาก OPAC ของ Chulalinet
Traffic	Internet traffic log
Web cache	Web traffic log

- 1 Internet traffic log 20000825 17:50 - 20000825 23:50 .. 1 entry
- 2 Internet traffic log 20000825 23:50 - 20000826 23:51 .. 1 entry
- 3 Internet traffic log 20000826 23:51 - 20000827 23:51 .. 1 entry
- 4 Internet traffic log 20000827 23:51 - 20000828 23:51 .. 1 entry
- 5 Internet traffic log 20000828 23:51 - 20000829 23:52 .. 1 entry
- 6 Internet traffic log 20000829 23:52 - 20000830 23:52 .. 1 entry
- 7 Internet traffic log 20000830 23:52 - 20000831 23:53 .. 1 entry
- 8 Internet traffic log 20000831 23:53 - 20000901 23:53 .. 1 entry

รูปที่ 10 แสดงผลการสืบค้นจากชื่อเรื่อง

จากรายการแสดงผลในรูปที่ 10 จะเห็นได้ว่า มีการเรียกข้อมูลขึ้นมาทั้งคอลเลกชัน เรียงลำดับของการเกิดข้อมูล เช่น จากรูปที่ 10 เป็นการเกิดข้อมูลแทรกฟลัก ตั้งแต่วันที่ 25 สิงหาคม 2000 เวลา 17:50 ถึง วันที่ 1 กันยายน 2000 เวลา 23:53 การเรียกข้อมูลในช่วงเวลาใดให้กำหนดหมายเลขหน้ารายการนั้น ๆ จะปรากฏรายละเอียดตามตารางที่ 15 ส่วนการที่ข้อมูลบางช่วงหายไปนั้น เกิดขึ้นเนื่องจากข้อจำกัดในการปฏิบัติงานทำให้การเก็บข้อมูลอาจเกิดการขาดตอนเป็นครั้งคราวได้ เช่น เราเตอร์ ต้องทำการบำรุงรักษา พีซีที่มีความเสียหาย ไฟดับ หรือสายสื่อสารหยุดทำงานเหล่านี้ เป็นต้น ช่วงของข้อมูลที่หายไปนี้ มีผลต่อคุณภาพพื้นฐานของข้อมูล เนื่องจากข้อมูลที่เก็บเป็นข้อมูลต่อเนื่อง เพราะฉะนั้น คุณภาพที่ดีที่สุด คือ ข้อมูลมีความต่อเนื่องไม่ขาดตอน ดังนั้น จึงได้ออกแบบการบันทึกข้อมูลในลักษณะข้างต้น เพื่อให้ผู้ใช้ข้อมูลสามารถเห็นได้ด้วยตนเองว่ามีข้อมูลขาดหายไป และเนื่องจากระบบอินโนแพคมีข้อจำกัดในการแสดงผลบางประการ จึงดำเนินการในส่วนของ Web interface ช่วยผู้ใช้ด้วยอีกประการหนึ่ง ซึ่งจะกล่าวโดยละเอียดในหัวข้อถัดไป

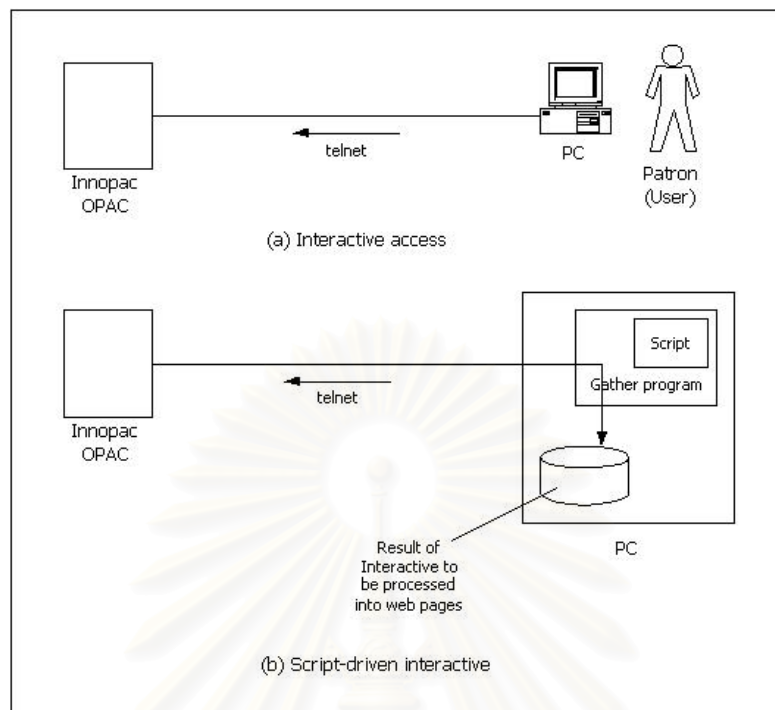
การศึกษา ออกแบบและสร้าง Web interface

เนื่องจากการแสดงผลและการสืบค้นลักษณะบางอย่างของระบบอินโนแพคของบริษัท Innovative ไม่เอื้ออำนวยต่อผู้ใช้ข้อมูล [48] โดยไม่สามารถสืบค้นเกี่ยวข้องกับข้อมูลที่เป็นช่วงเวลาได้ตรง ๆ และการแสดงผลเป็นเนื้อความไม่แสดงผลเป็นภาพกราฟิกส์ คณะผู้วิจัย จึงได้ออกแบบและสร้าง Web interface เพื่อเป็นการให้ข้อมูลเพิ่มเติมจากระบบของห้องสมุด เช่น ข้อมูลที่ขาดช่วงไป การสรุปข้อมูลในลักษณะกราฟเป็นรายปี การสรุปปริมาณข้อมูลเป็นรายเดือน และสามารถให้ผู้สืบค้นตามช่วงเวลาที่ต้องการได้

ระบบสารสนเทศกับอินโนแพค

จากการที่อินโนแพคเป็นโปรแกรมสำเร็จรูปที่ราคาสูง การแก้ไข ปรับเปลี่ยน หรือเพิ่มเติมโดยตรงเข้าไปที่ตัวโปรแกรมจะกระทำได้ลำบาก หรือเสียค่าใช้จ่ายสูง ดังนั้นในการวิจัยครั้งนี้ได้พัฒนาระบบเชื่อมต่อผ่านเว็บเข้าสู่ระบบอินโนแพค โดยการใช้ script ในการติดต่อสนทนากับอินโนแพค เสมือนเป็นการเรียกใช้จากผู้ใช้ปกติ ทำให้ไม่ต้องแก้ไขในส่วนของโปรแกรมอินโนแพคแต่อย่างใด ดังแสดงในรูปที่ 11

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 11 การสนทนากับอินโนแพค

จากรูปที่ 11 จะเห็นได้ว่าโปรแกรม Gather ที่เป็น scripting utility จะทำหน้าที่เสมือนเป็นผู้ใช้คนหนึ่งทำการติดต่อสนทนากับระบบโอแพค (OPAC: On-line Public Access Catalog) ภายใต้โปรแกรมอินโนแพค โดยที่ระบบอินโนแพคจะไม่ทราบว่ากำลังติดต่อกับโปรแกรมแทนที่จะเป็นคน ทำให้ไม่ต้องแก้ไขเปลี่ยนแปลงการทำงานของอินโนแพคแต่อย่างใด

สิ่งที่ต้องการจากอินโนแพคเพื่อมาสร้างเป็นเว็บเพจจะถูกกำหนดเป็น “บท” (script) ไว้ใน script file ของโปรแกรม Gather (ดังรายละเอียดที่ปรากฏในภาคผนวก จ) โปรแกรม Gather จะทำการสนทนา (Interact) กับอินโนแพคเหมือนเป็นคนคนหนึ่งโดยอาศัยตามบทที่ระบุไว้ ตัวอย่างของบทแสดงไว้ในรูปที่ 12

1. spawn telnet library.car.chula.ac.th
2. expect {
3. timeout { puts "timed out";exit }
4. "connection refused" exit
5. "login:" { send "myaccount\r" }
6. }
7. expect "Password:" { send "mypassword\r" }

รูปที่ 12 ตัวอย่างบทในการสนทนากับอินโนแพค

บทสนทนาตามรูปที่ 12 สำหรับการสนทนา หรือการติดต่อระหว่างโปรแกรม Gather กับอินโนแพค แปลความหมายได้ว่า

บรรทัดที่ 1 เริ่มการติดต่อกับอินโนแพคโดยเทลเน็ต (telnet) ไปยังเครื่องเซิร์ฟเวอร์ของสถาบันวิทยบริการ

บรรทัดที่ 2 รอดผลลัพธ์ โดยจะทำงานตามเงื่อนไขที่กำหนดไว้ในบรรทัดที่ 3 ถึง 5

บรรทัดที่ 3 ถ้าผลลัพธ์ที่ได้รับไม่ตรงกับเงื่อนไขใด ๆ เลยในระยะเวลาที่กำหนด (timeout) จะทำการแสดงข้อความ “timed out” และยกเลิกการติดต่อ

บรรทัดที่ 4 ถ้าผลลัพธ์ที่ได้เป็นข้อความ “connection refused” จะยกเลิกการติดต่อ

บรรทัดที่ 5 ถ้าผลลัพธ์ที่ได้เป็นข้อความ “login:” จะส่งชื่อผู้ใช้ให้แก่ระบบ

บรรทัดที่ 6 ปิดเงื่อนไขในการรอดผลลัพธ์จากบรรทัดที่ 2

บรรทัดที่ 7 ถ้าผลลัพธ์ต่อไปเป็นข้อความ “Password:” จะส่งรหัสผ่านของผู้ใช้ให้แก่ระบบ

บทที่วางไว้ให้ใช้กับโปรแกรม Gather จะถูกทำตามทุก 30 วันเพื่อดึงข้อมูลที่จำเป็นต่าง ๆ จากอินโนแพคมารวบรวมไว้เป็นฐานข้อมูลเล็ก ๆ ชื่อ db.gather โดยมีลักษณะเป็นแฟ้มข้อความธรรมดา (text file) เพื่อนำไปใช้เป็นข้อมูลสำหรับการสร้างเว็บเพจในระบบการเชื่อมต่อด้วยเว็บต่อไป โครงสร้างของฐานข้อมูลนี้แสดงไว้ในรูปที่ 13 ส่วนรูปที่ 14 แสดงรายการข้อมูลต่าง ๆ ที่ต้องการจากอินโนแพค

TITLE	Internet traffic log 20020227 11:43 - 20020227 23:44
DESCRIPT	1 log file (137.0 MB)
LINK	Internet Traffic Log, CD TR 2002-021, File CU-TR-2002-0116.zip
TITLE	Internet traffic log 20020227 23:44 - 20020227 23:44
DESCRIPT	1 log file (86.2 MB)
LINK	Internet Traffic Log, CD TR 2002-021, File CU-TR-2002-0117.zip

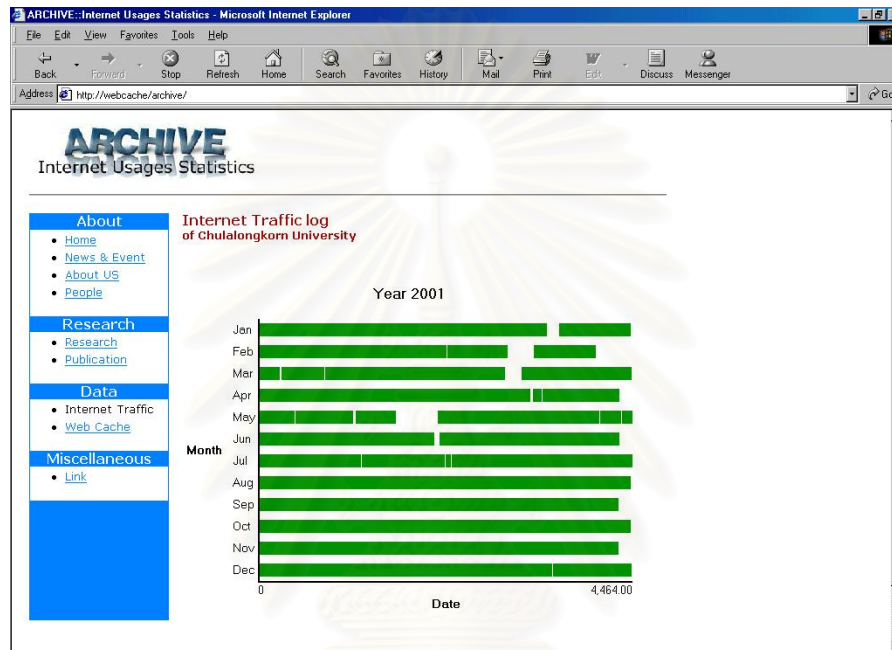
รูปที่ 13 โครงสร้างฐานข้อมูลของการสนทนากับอินโนแพค

ข้อมูล	คำอธิบาย
ชื่อแผ่นซีดี-อาร์	ชื่อของแผ่นซีดี-อาร์ที่บรรจุ Chunk อยู่
ชื่อ chunk	ชื่อไฟล์ chunk
ระยะเวลาของ chunk	วัน เวลาเริ่มต้น และสิ้นสุดของข้อมูลที่อยู่ภายใน chunk
ขนาดของ chunk	ขนาดของไฟล์ chunk

รูปที่ 14 รายการข้อมูลที่ต้องการจากอินโนแพค

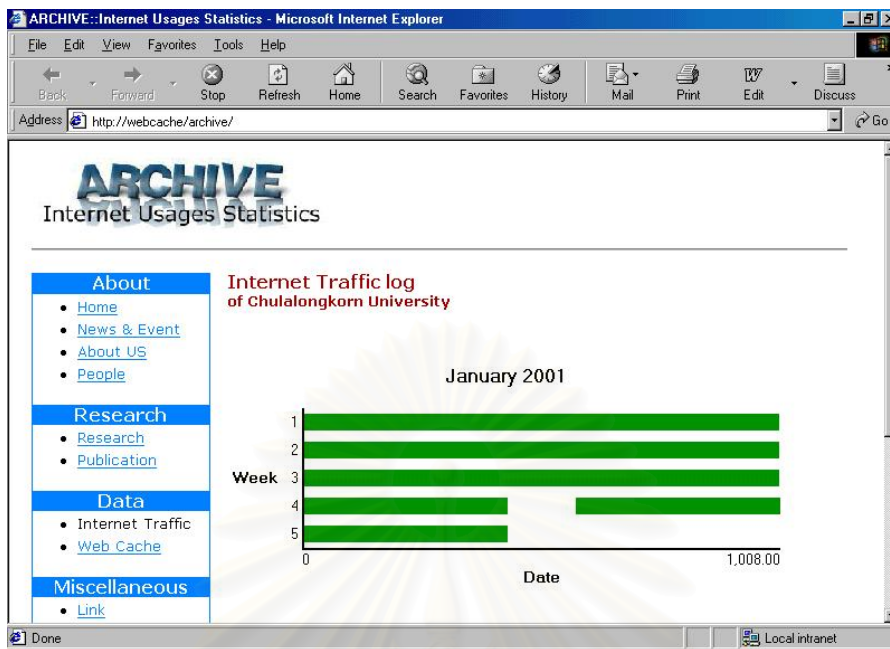
การแสดงผลในรูปของเว็บ

จากการที่ระบบเว็บสามารถแสดงผลเป็นภาพกราฟิกส์ได้สะดวก ทำให้คณะผู้วิจัยนำข้อมูลดิบทางบรรณานุกรมที่ได้จากการสอบถามอินโนแพคผ่านทางระบบการสนทนาตามที่ได้กล่าวแล้ว มาจัดแสดงเป็นภาพกราฟิกส์ให้สามารถระบุชี้ถึงคุณภาพของข้อมูลได้ในแต่ละช่วงเวลา เช่นแต่ละวัน มีข้อมูลหรือไม่ ซึ่งแม้จะไม่ได้ให้ตัวเลขปริมาณที่ชัดเจน แต่สามารถเห็นภาพรวมของคุณภาพของข้อมูลได้อย่างสะดวกรวดเร็ว โดยใช้เว็บเพจเพียงไม่กี่หน้าเท่านั้น ดังแสดงในรูปที่ 15 รูปที่ 16 และรูปที่ 17

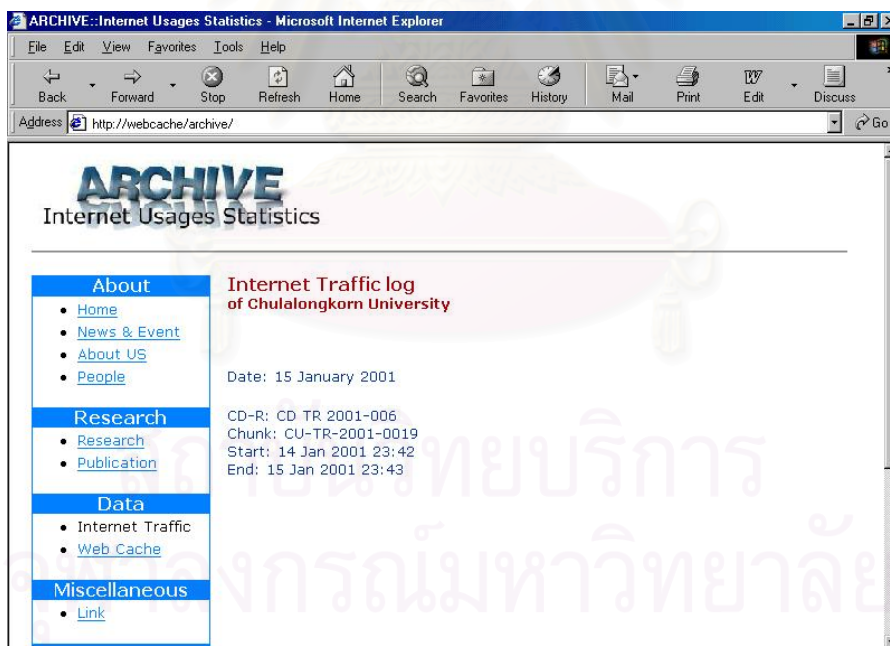


รูปที่ 15 สรุปคุณภาพข้อมูลเป็นรายเดือน

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 16 สรุปคุณภาพข้อมูลเป็นรายสัปดาห์



รูปที่ 17 สรุปข้อมูลในแต่ละ Chunk

การกำหนดรูปแบบบรรณานุกรม และการวางแผนทางการจัดเก็บแบบถาวรสำหรับข้อมูลต่อเนื่องที่ตัดมาเป็นกลุ่มให้สืบค้นผ่านห้องสมุดอัตโนมัติโดยอาศัยกรณีศึกษาเป็นข้อมูลสถิติการใช้งานอินเทอร์เน็ตของจุฬาลงกรณ์มหาวิทยาลัยนั้น จึงเริ่มตั้งแต่การศึกษาลักษณะของข้อมูลที่ต้องจัดเก็บ การตัดแบ่งข้อมูลเพื่อให้สามารถนำลงแผ่นบันทึกข้อมูลได้เหมาะสม การมีความร่วมมืออย่างดีในการให้บรรณารักษ์เข้ามามีส่วนร่วมในการกำหนดระบบของข้อมูล หรือเมตาดาตาที่ต้องการ การทดสอบข้อมูลที่นำลงสู่ระบบ ตลอดจนการทดลองใช้ผลของข้อมูล และเสริมการทำงานขึ้นมาในส่วนที่ระบบเดิมไม่สามารถรองรับได้ โดยส่วนของผลการวิจัยจะได้นำเสนอในบทที่ 4



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 4 ผลการวิจัย

จากการศึกษาการกำหนดรูปแบบบรรณานุกรมและการวางแผนทางการจัดเก็บแบบถาวรสำหรับข้อมูลต่อเนื่องที่ตัดมาเป็นกลุ่มโดยใช้ข้อมูลสถิติการใช้อินเทอร์เน็ตของจุฬาลงกรณ์มหาวิทยาลัยเป็นกรณีศึกษานั้น สรุปผลการวิจัย ได้ดังนี้

สื่อบันทึก

จากการวิจัยพบว่า สื่อบันทึกที่มีคุณลักษณะ และคุณสมบัติเหมาะสมกับการบันทึกข้อมูลถาวรในระยะยาว คือ ซีดี-อาร์ เนื่องจากสะดวกต่อการปฏิบัติงาน มีขนาดไม่ใหญ่มากเกินไป สามารถหาซื้อได้ทั่วไป ราคาไม่แพง ไม่จำเป็นต้องให้การดูแลรักษามากเกินไป อีกทั้งผู้ปฏิบัติงานยังมีความคุ้นเคยกับสื่อประเภทนี้ อีกด้วย

การที่เลือกซีดี-อาร์ นั้น เพราะต้องสามารถบันทึกข้อมูลได้เอง เนื่องจากเป็นการบันทึกเพียงสำเนาเดียว ไม่ได้ผลิตจำนวนมาก จึงไม่เหมาะที่จะใช้ซีดี-รอม ที่ผลิตแบบแผ่นเสียงเป็นคราวละมาก ๆ (500 แผ่นขึ้นไป) แต่การบันทึกจะเป็นการทำครั้งเดียวแล้วไม่เปลี่ยนแปลงอีกในลักษณะที่เป็นบันทึกถาวร จึงไม่ได้ใช้ซีดี-อาร์ ดับเบิลวิ ซึ่งสามารถบันทึกซ้ำหรือแก้ไขได้

ปัญหาที่สำคัญประการหนึ่ง คือ การเปลี่ยนแปลงพัฒนาเทคโนโลยีของสื่ออย่างค่อยเป็นค่อยไปเพื่อผลทางการค้า ดังนั้น ในช่วงระยะเวลาประมาณ 1 ปีของการวิจัย พบว่า ซีดี-อาร์ ขนาดความจุ 650 MB ได้เลิกผลิตหมดไปจากท้องตลาด เปลี่ยนเป็นขนาด 700 MB ออกมาแทนที่ อีกทั้งยังเป็นการยากที่จะระบุยี่ห้อและรุ่นของซีดี-อาร์ เหล่านี้ เพราะมีการเปลี่ยนแปลงตามพลวัตรของตลาดอย่างมาก แม้เทคโนโลยีพื้นฐานและ compatibility จะไม่เสียหายหรือเปลี่ยนแปลงแต่อย่างใด การเปลี่ยนแปลงนี้แสดงให้เห็นได้ในตารางที่ 17

ตารางที่ 17 แสดงพัฒนาการของแผ่นสื่อบันทึกซีดี-อาร์

ยี่ห้อ	ประเภท	ความจุ	จำนวน
VERBATIM	แผ่นเงิน	650 MB	100 แผ่น
KODAK	แผ่นเงิน	650 MB	100 แผ่น
KODAK	แผ่นทอง	650 MB	100 แผ่น
SONY	แผ่นเงิน	700 MB	100 แผ่น

การตัดข้อมูลต่อเนื่องเป็น chunk

จากการวิจัยพบว่า ข้อมูลสถิติการใช้อินเทอร์เน็ตของจุฬาลงกรณ์มหาวิทยาลัยนั้น มีลักษณะที่เข้ามาอย่างต่อเนื่อง จึงต้องหาวิธีการจัดการเพื่อให้สามารถบันทึกลงในแผ่นบันทึกข้อมูลได้อย่างเหมาะสม โดยการตัดข้อมูลออกเป็นส่วน ๆ หรือเป็นกลุ่ม ซึ่งในงานวิจัยนี้เรียกว่า chunk ให้เป็นการเก็บข้อมูลในลักษณะที่เป็นอนุกรมตามเวลา (series by time) การตัดตอนข้อมูลนี้อาศัยหลักเน้นความสะดวกในการปฏิบัติงานและการใช้

งานโดยมีข้อกำหนดในการเก็บ คือ เก็บเป็นวันโดยประมาณ และมีขนาดที่เหมาะสม โดยกำหนดว่า chunk ควร มีขนาดเล็กพอที่จะบรรจุได้ในเนื้อที่ไม่เกิน 60% ของแผ่นสื่อบันทึก มีขนาดประมาณ 200 MB ทำให้บรรจุได้ 2 chunk ในหนึ่งแผ่นของซีดี-อาร์ขนาด 650 MB ทั้งนี้การที่เหลือเนื้อที่ไว้ถึง 40% ของแผ่น เนื่องจากข้อมูลการใช้งานในแต่ละคาบไม่แน่นอน สามารถเพิ่มลดได้ค่อนข้างมาก และในการทำกำหนดขนาดของ chunk ไว้ที่ 200 MB ทำให้สะดวกต่อการจัดการกับ chunk ในการถ่ายโอนข้อมูลระหว่างเครื่อง และการบีบอัดข้อมูล

การบันทึกข้อมูลลงแผ่นบันทึกข้อมูล (ซีดี-อาร์)

ในการวิจัยครั้งนี้ สามารถกำหนดขั้นตอนการบันทึกข้อมูลลงแผ่นบันทึกข้อมูลที่มีระบบ และเป็น ขั้นตอนมากขึ้น ลดความเสียหายและการสูญหายการจัดกระจายของข้อมูลในระหว่างการผลิตซีดี-อาร์ รวมทั้ง ได้กำหนดเมตาเดตาของข้อมูลที่บันทึกลงในแผ่นบันทึกด้วย เพื่อที่สามารถสืบค้นและนำข้อมูลที่บันทึกไปใช้ได้ พร้อมกับมีรายละเอียดของอุปกรณ์การอ่านและอุปกรณ์การบันทึกเพื่อการเก็บถาวรต่อไป

คณะผู้วิจัยได้คำนึงถึงการนำผลการวิจัยไปใช้ในการปฏิบัติงานจึงได้มีการบันทึกกระบวนการขั้นตอน เหล่านี้จัดรวบรวมเป็นคู่มือการปฏิบัติงานการบันทึกข้อมูลลงซีดี-อาร์ ดังแสดงในภาคผนวก ง

ขั้นตอนการจัดเก็บข้อมูลแทรฟฟิก

ขั้นตอน	การดำเนินงาน
1	Export ข้อมูล packet ที่ผ่านเข้า / ออก router 7206
2	ทำ aggregation ในรูปของแฟ้มข้อความโดยโปรแกรม NetFlow flowcollector version 2 บน เครื่อง Sun Sparc Ultra-1 ที่ใช้ระบบปฏิบัติการ Solaris 2.5.1 ทุก 30 นาที
3	นำแฟ้มมาต่อกันเป็น chunk เพื่อความสะดวกในการจัดการ โดยจำนวนแฟ้ม ที่ใช้ขึ้นกับขนาดของ แฟ้มและความสะดวกในการใช้งาน
4	นำ chunk มาห่อและบีบอัดลดขนาดกับเมตาเดตาที่บอกรายละเอียดเกี่ยวกับ chunk และ แหล่งที่มาของข้อมูล
5	นำแฟ้มที่ได้มาบันทึกลงแผ่นซีดี-อาร์ พร้อมกับเมตาเดตาที่บอกรายละเอียดเกี่ยวกับข้อมูลที่อยู่ ภายในแผ่นซีดี-อาร์

ขั้นตอนการเก็บข้อมูลเว็บแคช

ขั้นตอน	การดำเนินการ
1	โอนย้ายข้อมูล log ของเว็บแคชมาที่เครื่อง PC ที่ใช้บนระบบปฏิบัติการ Linux วันละ 1 ครั้ง โดย แบ่งเป็น 4 แฟ้ม ๆ ละ 6 ชั่วโมง โดยอยู่ในรูปของแฟ้มข้อความ
2	นำแฟ้มมาต่อกันเป็น 1 chunk ต่อวัน
3	นำมาห่อและบีบอัดลดขนาดกับเมตาเดตาที่บอกรายละเอียดเกี่ยวกับแฟ้มและแหล่งที่มาของ ข้อมูล
4	นำแฟ้มที่ได้มาบันทึกลงแผ่นซีดี-อาร์ พร้อมกับเมตาเดตาที่บอกรายละเอียดเกี่ยวกับข้อมูลที่อยู่ ภายในแผ่นซีดี-อาร์

เมตาดาตาหรือรูปแบบระเบียบบรรณานุกรม

จากการวิจัยในครั้งนี้ สามารถกำหนดเมตาดาตาหรือรูปแบบระเบียบบรรณานุกรมของการลงรายการข้อมูลสถิติการใช้อินเทอร์เน็ต โดยมีทั้งหมด 25 เขตข้อมูลหลัก ซึ่งหน่วยงานหรือห้องสมุดอื่นที่มีการเก็บข้อมูลสถิติการใช้อินเทอร์เน็ต หรือข้อมูลที่มีลักษณะเป็นกลุ่มแบบนี้ สามารถนำเมตาดาตาหรือรูปแบบลงรายการระเบียบบรรณานุกรมนี้เป็นต้นแบบได้

คณะผู้วิจัยได้คำนึงถึงการนำผลการวิจัยไปใช้ในการปฏิบัติงาน จึงได้มีการบันทึกกระบวนการขั้นตอนเหล่านี้ จัดรวบรวมเป็นคู่มือการปฏิบัติงาน ดังแสดงในภาคผนวก ก และภาคผนวก ข

เขตข้อมูลในการลงรายการระเบียบบรรณานุกรม

หมายเลขเขตข้อมูล	ชื่อของเขตข้อมูล	รายการ
099 * **	Local Call Number	เลขหมู่ของแผ่นซีดี-อาร์
245 # a * **	Title	ชื่อเรื่อง
245 # c *	Statement of responsibility, etc	ชื่อผู้รับผิดชอบ
260 # a *	Place of publication, distribution, etc	สถานที่ของผู้รับผิดชอบ
260 # b *	Name of publisher, distributor, etc	ชื่อผู้จัดทำ
260 # c * **	Date of publication, distribution, etc	ปีที่บันทึกซีดี-อาร์/ ปีที่เกิดข้อมูล
300 # a * **	Physical description (Extent)	จำนวนแผ่นซีดี-อาร์/จำนวนของแฟ้ม
300 # c *	Dimensions	ขนาดของแฟ้มข้อมูล
310 # a *	Current publication frequency	คาบการเกิดข้อมูล
362 # a *	Dates of publication and / or sequential designation	ปีที่ผลิตข้อมูลและ/หรือลำดับของการเกิดข้อมูล
500 # a * **	General Note	หมายเหตุ
506 # a *	Terms governing access	ข้อจำกัดในการเข้าใช้
506 # c *	Physical access provisions	การเข้าใช้แผ่นบันทึก
506 # d *	Authorized users	ข้อกำหนดในการใช้
508 # a *	Creation/production credits note	แหล่งที่เกิดข้อมูลและผู้บันทึก
515 # a *	Numbering peculiarities note	ขอบเขตของข้อมูล
516 # a *	Type of computer file or data note	ลักษณะของข้อมูล
518 # a *	Date/time and place of an event note	วัน เดือน ปี ที่บันทึกข้อมูลลงซีดี-อาร์

เขตข้อมูลในการลงรายการระเบียบบรรณานุกรม (ต่อ)

หมายเลขเขตข้อมูล	ชื่อของเขตข้อมูล	รายการ
520 # a *	Summary, etc. note	เนื้อเรื่องย่อ
524 # a *	Preferred citation of described materials note	การอ้างถึงทางบรรณานุกรม
535 # a *	Custodian	ผู้ดูแลข้อมูล
535 # b *	Postal address	ที่อยู่ของผู้ดูแลข้อมูล
536 # a *	Text of note (Funding Information Note)	ทุนที่ได้รับ
538 # a *	System details note	ข้อกำหนดของระบบ
540 # b *	Terms governing use and reproduction note (Jurisdiction)	เงื่อนไขการใช้งานและการทำสำเนาออก
541 # c *	Method of acquisition	วิธีการได้ซึ่งข้อมูล
541 # o *	Type of unit	หน่วยของข้อมูล
567 # a *	Methodology note	กรรมวิธีการประมวลข้อมูล
583 # a *	Action	วัน เดือน ปี ที่คาดว่าซีดี-อาร์ หมดอายุ
610 # a *	Subject Added Entry Corporate Name	หัวเรื่องสำหรับหน่วยงานที่เชื่อมโยง ไปถึง
650 # a *	Subject Added Entry - Topical Term	หัวเรื่องสำหรับหมายเลขไอพีของ เครื่องและหัวเรื่องทั่วไป
710 # a * **	Added Entry - Corporate Name	หน่วยงานที่ผลิตข้อมูล
773 # t **	Host Item Entry: Title	ชื่อของ chunk
773 # g **	Host Item Entry: Relationship information	หมายเลขเพิ่มข้อมูลของ chunk

* หมายถึง เขตข้อมูลที่ลงใน bibliographical record of collection

** หมายถึง เขตข้อมูลที่ลงใน bibliographical record of content analysis

ข้อมูลตัวอย่าง

ผลจากการศึกษานี้ ทำให้ได้ข้อมูลตัวอย่างของการใช้อินเทอร์เน็ตทั้งลักษณะที่เป็นแทรฟฟิก และเว็บ-
แคช ทั้งที่เป็นระดับคอลเลกชัน และระดับช่วงเวลา ดังตัวอย่าง

ตัวอย่างข้อมูลแทรฟฟิก

เขตข้อมูล	ตัวบ่งชี้	ลักษณะสมบัติ
099	2	CD TR 2002-001
245	00	Internet traffic log[h[Computer file] ; bUninet gateway of CUNET / cSystem Department, Center of Academic Resources, Chulalongkorn University
260		Bangkok ; bSystem Department, Center of Academic Resources, Chulalongkorn University, c2001-
300		Computer optical laser discs (CD-R) ; c4 3/4 in
310		119 discs per year (approx.)
362	1	2001-
500		Location: IP address 202.28.0.1.125/25 - 10 Mbps connect to KMUTT's Router; IP address 202.28.18.35/27 - 155 Mbps connect to MUA ATM backbone; IP address 202.28.0.253/25 - 100 Mbps connect to CUNET
506		Not available for publication or commercialization ; aRestricted online access through appointment and with password ; cPhysical access prohibited ; dFor use in researches granted by official bodies
508		Data stream originator, Office of Information Technology, Chulalongkorn University;Chunks and CD-ROMs produced from 2001- , Nitass Sutaveepramochochanon
515		Data covers May 2001-
515		Missing data / very irregular data (see chunk for details)
516		Numeric (Computer log file)
520		Contains data on the Internet traffic from router: Cisco 7206/IOS version 11.1 (17.3) CA; Changed to (35) CC2 in September 1st 2001. Aggregates data from program Netflowcollector (version 2), under CUNET administration and generated by account.car.chula.ac.th
524		Chulalongkorn University. Center of Academic Resources. "Internet traffic log 20011231 23:53 - 20020101 11:53." Traffic Traffic Log, CD TR 2002-001, File CU-TR-2002-0001.zip
535		Chulalongkorn University. bCenter of Academic Resources. bSystem Department
536		Original collection up to April 2002 supported by Rachadapiseksompoj Research Fund, Research Affairs, Chulalongkorn University

ตัวอย่างข้อมูลแพรฟฟิก (ต่อ)

เขตข้อมูล	ตัวบ่งชี้	ลักษณะสมบัติ
538		Data and metadata packaged and compressed by program zip v.2.1 on Solaris 2.5.1
540		bRelated privacy information must be removed or masked before publication of research result except by specific authorization from the owner
541		cContains the Internet traffic from router: Cisco 7206/IOS version: 11.1 (17.3) CA and (35) CC2 since 1 st September 2001; oChunk of 30 minute log data from router
567		30-minute chunk data stream
610	20	KMUTT
610	20	MUA
610	20	CUNET
650	7	202.28.0.125/25
650	7	202.28.18.35/27
650	7	202.28.0.253/25
650	0	Internet
650	7	Internet traffic log file
710	2	Chulalongkorn University. bCenter of Academic Resources. bSystem Department

ตัวอย่างข้อมูลในแต่ละช่วงเวลา

เขตข้อมูล	ตัวบ่งชี้	ลักษณะสมบัติ
099	2	CD TR 2000-001
245	00	Internet traffic log 20000825 17:50 – 20000825 23:50
260		c2000
300		1 log file (46.0 MB)
500		Log for 25 Aug 2000 17:50 - 25 Aug 2000 23:50
500		Usage through online access on appointment with staff
518		CD published on Sep 23, 2000
583		Expected expiry date; c20050923
650	0	Internet
650	7	Internet traffic log file

ตัวอย่างข้อมูลในแต่ละช่วงเวลา (ต่อ)

เขตข้อมูล	ตัวบ่งชี้	ลักษณะสมบัติ
710	2	Chulalongkorn University. bCenter of Academic Resources. bSystem Department
773	0	tInternet Traffic Log, gCD TR 2000-001, File CU-TR-2000-0001.zip

ผู้บริหารหรือผู้วิจัยที่ต้องการใช้ข้อมูลดังกล่าว เพื่อนำมาประกอบในการบริหารจัดการหรือเพื่อการวิจัย สามารถสืบค้นข้อมูลจากฐานข้อมูลของห้องสมุดในเครือข่ายจุฬาลงกรณ์มหาวิทยาลัย ทั้งนี้ จากการศึกษาวิจัยนี้ได้พยายามเพิ่มศักยภาพเพื่อสนองตอบต่อการสืบค้นจากระบบเดิมในห้องสมุดโดยผ่านทาง Web interface โดย

1. สรุปคุณภาพของข้อมูลในลักษณะกราฟเป็นรายปี
2. กราฟสรุปปริมาณข้อมูล ตลอดปีเป็นรายเดือน
3. สามารถสืบค้นตามช่วงเวลาที่ต้องการได้

ข้อมูลเชิงปริมาณ

การวิจัยนี้ได้ดำเนินการเก็บข้อมูลและบันทึกลงซีดี-อาร์ เป็นเวลาประมาณ 20 เดือน ได้เป็นซีดี-อาร์ทั้งสิ้น 265 แผ่น ดังแสดงไว้ในรูปที่ 18, 19 และ 20 โดยแบ่งเป็นข้อมูลแทรฟฟิก และข้อมูลเว็บแคช ดังนี้

1. บันทึกข้อมูลที่เป็นแทรฟฟิกลงซีดี-อาร์ จำนวน 188 แผ่น (เป็นข้อมูลตั้งแต่วันที่ 25 สิงหาคม 2543 เวลา 17:50 ถึง 31 เมษายน 2545 เวลา 23:57)

1.1 รวมจำนวนวันที่เก็บข้อมูล 583 วัน 25 ชั่วโมง

1.2 ข้อมูลหายคิดรวมเป็น 635 ชั่วโมง

1.3 ขนาดของข้อมูลทั้งหมดที่เก็บได้ 99,865.71 MB หรือ 97.53 GB

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 18 แผ่นซีดี-อาร์ของข้อมูลแทรฟฟิกในปี 2000



รูปที่ 19 แผ่นซีดี-อาร์ของข้อมูลแทรฟฟิกในปี 2001



รูปที่ 20 แผ่นซีดี-อาร์ของข้อมูลแทรฟฟิกในปี 2002

2. บันทึกข้อมูลที่เป็นเว็บแคชของซีดี-อาร์ จำนวน 77 แผ่น ดังแสดงไว้ในรูปที่ 21 และ 22 (เป็นข้อมูลตั้งแต่วันที่ 5 ตุลาคม 2544 เวลา 00:00 ถึง 1 เมษายน 2545 00:00)

2.1 รวมจำนวนวันที่เก็บข้อมูล 178 วัน

2.2 ข้อมูลหายจาก proxy 2 ขาดไป 571 ชั่วโมง และ proxy 3 ขาดไป 469

ชั่วโมง

2.3 ขนาดของข้อมูลทั้งหมดที่เก็บได้ 41,837.50 MB หรือ 40.86 GB

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 21 แผ่นซีดี-อาร์ของข้อมูลเว็บแคชในปี 2001



รูปที่ 22 แผ่นซีดี-อาร์ของข้อมูลเว็บแคชในปี 2002

3. บันทึกข้อมูลลงโปรแกรมห้องสมุดอัตโนมัติอินโนแพคและตรวจสอบความถูกต้องโดย
แบ่งเป็น

3.1 ระเบียบข้อมูลทางบรรณานุกรมของแทรฟฟิก จำนวน 709 ระเบียบ

1. Bibliographical record of collection จำนวน 2 ระเบียบ
2. Bibliographical record of content analysis จำนวน 707 ระเบียบ

3.2 ระเบียบข้อมูลทางบรรณานุกรมของเว็บแคช จำนวน 295 ระเบียบ

1. Bibliographical record of collection จำนวน 1 ระเบียบ
2. Bibliographical record of content analysis จำนวน 294 ระเบียบ

จากผลของการวิจัยข้างต้น กล่าวได้ว่าเป็นไปตามวัตถุประสงค์ที่ตั้งไว้ โดยสามารถกำหนดรูปแบบ
บรรณานุกรมสำหรับข้อมูลต่อเนื่องที่ตัดมาเป็นกลุ่ม รวมถึงสามารถวางแผนทางจัดเก็บข้อมูลโดยกำหนดสื่อ
บันทึกที่สามารถเก็บข้อมูลในระยะยาว อีกทั้งผู้ใช้สามารถเข้าถึงข้อมูลที่เป็นระบบได้อีกด้วย

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 5

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

สรุปผลการวิจัย

การศึกษาการกำหนดรูปแบบบรรณานุกรมและการวางแผนทางการจัดเก็บแบบถาวรสำหรับข้อมูลต่อเนื่องที่ตัดมาเป็นกลุ่ม โดยใช้ข้อมูลสถิติการใช้อินเทอร์เน็ตในจุฬาลงกรณ์มหาวิทยาลัยนั้น สามารถสรุปผลการวิจัยโดยแบ่งเป็นประเด็นได้ดังนี้

1. การวิจัยได้ผลว่าใช้สื่อบันทึกที่เป็นซีดี-อาร์ จะเหมาะสมต่อการเก็บข้อมูลเมื่อเป็นบันทึกถาวร
2. สำหรับข้อมูลที่มีลักษณะต่อเนื่องนั้นใช้หลักการของ chunk เป็นหน่วยของการตัดแบ่งเป็นกลุ่ม โดยเน้นความสะดวกในการปฏิบัติงานต่อ chunk เหล่านั้น ได้แก่ การตัดแบ่งตามวันทำการ และการประมาณขนาดให้สอดคล้องกับความจุของสื่อบันทึก
3. การระบุบรรณานุกรมของสาระอาศัยระบบมาร์ก และอิงระบบการปฏิบัติงานปกติของห้องสมุด คือ เซิร์ฟเวอร์โปรแกรมอินโนแพค สำหรับห้องสมุดอัตโนมัติ และกระบวนการจัดหาและให้บริการในเชิงบรรณารักษ์
4. อาศัยเทคโนโลยีของเว็บขั้นพื้นฐานเพื่อเสริมข้อจำกัดของระบบห้องสมุดอัตโนมัติที่แสดงผลเป็นข้อความ โดยให้แสดงผลผ่านเบราว์เซอร์ (browser) เป็นภาพกราฟิกส์ ซึ่งเน้นคุณภาพและปริมาณของข้อมูลให้สะดวกต่อการเลือกใช้งานได้

อภิปรายผลการวิจัย

ผลการวิจัยที่ได้จากคณะผู้วิจัยในครั้งนี้ มีข้อสังเกตในประเด็นต่างๆ ดังต่อไปนี้

1. สื่อบันทึกข้อมูล จากการวิจัยพบว่า ซีดี-อาร์ ที่เลือกเหมาะสมกับการใช้ประโยชน์มาก สะดวกต่อการปฏิบัติงาน สามารถหาซื้อได้ทั่วไป เป็นสื่อที่ไม่ต้องการการดูแลรักษามากจนเกินไป มีขนาดไม่ใหญ่มาก เป็นสื่อที่เจ้าหน้าที่ระดับปฏิบัติงานและบุคคลทั่วไปมีความคุ้นเคยทัดเทียมกับดิสเก็ตต์ หรือซีดี-รอม สำหรับเพลงและภาพยนตร์ ซึ่งแตกต่างจากสื่อบันทึกอื่น ๆ เช่น เทปแม่เหล็ก
2. ระบบการลงรายการทางบรรณานุกรมเอกสารดิจิทัลในลักษณะที่เป็นกฤตภาค ซึ่งมีลักษณะเหมือนกับการทำกฤตภาคของหนังสือพิมพ์ เพื่อให้ผู้ใช้ได้เจาะลึกถึงข้อมูลได้มากที่สุด ซึ่งพบว่าเป็นการจัดระบบที่เหมาะสมต่อการจัดเก็บข้อมูลที่เป็นอนุกรมตามเวลา ระบบการลงรายการระเบียบบรรณานุกรมของมาร์ก ก็สามารถหาเขตข้อมูลครอบคลุมข้อมูลพวกนี้ได้
3. การจัดเก็บและตัดตอนข้อมูลออกเป็น chunk ให้สอดคล้องกับวิธีการปฏิบัติงาน คือ วันต่อวัน และขนาดของข้อมูลให้สอดคล้องกับขนาดของสื่อที่บันทึก จากการวิจัยพบว่า หลักการของ chunk แม้จะเป็นหลักการง่าย ๆ แต่ทำให้เกิดความชัดเจนและมีหลักการที่แน่นอนในการตัดตอนข้อมูลอย่างเป็นระบบ ซึ่งนอกจากจะมีผลทำให้เกิดความสะดวกต่อการปฏิบัติกรยังมีผลข้างเคียง คือ สามารถนำโครงสร้างของ chunk มาเป็นหน่วยของข้อมูล สามารถแสดงคุณภาพของข้อมูลได้อย่างเป็นระบบ

4. ระเบียบบรรณานุกรม จากการวิจัยพบว่า การกำหนดเขตข้อมูลในการลงรายการ ระเบียบบรรณานุกรมนั้น ถ้าบรรณารักษ์ซึ่งเป็นผู้ดูแลการลงรายการได้มีส่วนร่วมในการกำหนดเมตาตาตาด้วย จะส่งผลดีต่อการลงรายการเป็นอย่างมาก เนื่องจากการกำหนดเมตาตาตาของเอกสารดิจิทัลไม่เคยมีใครทำ มาก่อนในประเทศไทย การกำหนดเมตาตาตาของเอกสารดิจิทัลจึงเป็นเรื่องใหม่และมีความซับซ้อนมาก นัก คอมพิวเตอร์ซึ่งมีความรู้ ความเข้าใจในเรื่องธรรมชาติของข้อมูลหรือเอกสารดิจิทัล จะเป็นผู้ที่ให้ข้อมูลได้ดี แต่ ไม่ทราบวิธีการหรือรูปแบบที่ควรจะเป็นเมตาตาตา เมื่อมีการทำงานร่วมกันระหว่างบรรณารักษ์และผู้ใช้ข้อมูล น่าจะเป็นการส่งผลดีให้แก่งานนั้น ๆ

5. การเก็บข้อมูล โดยปกติในการเก็บข้อมูลทั่วไปจะไม่ทราบถึงรายละเอียดของการนำ ข้อมูลไปใช้ จะทราบเฉพาะการใช้ข้อมูลเฉพาะกิจเท่านั้น ทำให้ไม่มีการจัดกระบวนการจัดเก็บและรูปแบบของ ข้อมูลที่ชัดเจน รวมถึงการจัดเก็บเมตาตาตาที่เหมาะสม เมื่อนำข้อมูลที่เก็บไว้แล้วมาทำระเบียบรายการทาง บรรณานุกรม ทำให้เกิดปัญหาไม่มีเมตาตาตาเพียงพอ ดังนั้น ควรทำการกำหนดโครงสร้างของระเบียบ บรรณานุกรมขึ้นก่อนการเก็บข้อมูล ซึ่งพบว่าในทางปฏิบัติสามารถทำได้ยากเนื่องจากการใช้งานไม่ได้เกิดขึ้น ภายในห้องสมุดที่เป็นผู้กำหนดโครงสร้างของระเบียบบรรณานุกรม จึงควรประชาสัมพันธ์ให้แก่นักวิจัยว่าถ้า ต้องการเก็บข้อมูลขนาดใหญ่ให้ปรึกษากับทางห้องสมุดเพื่อกำหนดโครงสร้างของระเบียบบรรณานุกรมให้ เหมาะสมก่อนเกิดกระบวนการเก็บข้อมูล

6. สื่อบันทึกมีการเปลี่ยนแปลงรุ่นและยี่ห้อ ทำให้ยากต่อการเปรียบเทียบคุณสมบัติต่างๆ เช่น อายุ และความทนทานต่อการใช้งานหรืออัตราการเสื่อมของข้อมูลที่เก็บ เป็นต้น

7. จากการที่บรรณารักษ์ห้องสมุดเข้ามามีบทบาทอย่างสูงในการวิจัยนี้ ทำให้มองเห็น แนวทางที่ชัดเจนว่า สำหรับข้อมูลที่มีลักษณะเป็นสาระพื้นถิ่น โดยเฉพาะในสถาบันอุดมศึกษาอย่าง จุฬาลงกรณ์มหาวิทยาลัย จำต้องอาศัยห้องสมุดให้ทำหน้าที่เป็นเสมือนผู้ตีพิมพ์ (publisher) เนื่องจาก กระบวนการของการบันทึกข้อมูลและการสร้างระเบียบบรรณานุกรม เป็นกระบวนการที่ซับซ้อนไม่เหมาะสมที่ จะผลักดันให้เป็นภาระของแหล่งกำเนิดข้อมูล ไม่ว่าจะแก่นักวิจัย หรือระบบอุปกรณ์ต่าง ๆ อีกทั้งสาระเหล่านี้ก็ ไม่ได้มีการตีพิมพ์เผยแพร่ในเชิงการค้ายากที่จะอาศัยกระบวนการตีพิมพ์เชิงอิเล็กทรอนิกส์ (eletronic publishing) ของบริษัทเอกชนได้ จากข้อสังเกตนี้ จะเห็นได้ว่าห้องสมุดของหน่วยงานที่มีการสร้างสาระพื้นถิ่น เป็นจำนวนมาก เช่น สถาบันอุดมศึกษา ควรจะได้มีการเตรียมความพร้อมในการทำหน้าที่ เป็นผู้ตีพิมพ์ควบคู่ ไปกับหน้าที่การลงรายการบรรณานุกรมตามปกติ เพื่อให้เกิดเป็นระบบ สามารถให้บริการแก่สาระพื้นถิ่นใน ลักษณะที่เป็นทางการและส่งผลต่อความถาวรของสาระมากขึ้น

8. จุดบกพร่องประการหนึ่งของการทำวิจัยในครั้งนี้ สืบเนื่องมาจากข้อมูลมีลักษณะ ต่อเนื่องเข้ามาตลอดเวลา ความต้องการที่จะดำเนินการให้ทันต่อการไหลเข้ามาของข้อมูล เพื่อป้องกันการ เสียหายหรือการขาดช่วงของการเก็บรักษาข้อมูล ทำให้เกิดความเร่งรีบในการดำเนินการวิจัย เกิดสภาวะของ การนำขั้นตอนการปฏิบัติงานเข้ามามีอิทธิพลต่อการดำเนินงานวิจัย กล่าวคือ ได้ดำเนินการบันทึกข้อมูลที่ เพิ่มขึ้นมากจนล้นฮาร์ดดิสก์ลงสู่ซีดี-อาร์ก่อนเวลาอันควร โดยยังมีได้กำหนดโครงสร้างของเมตาตาตาและ บรรณานุกรมให้เป็นที่ยอมรับอยู่ อย่างไรก็ตาม ในการดำเนินการนี้ได้กำหนดเมตาตาตาขึ้นจำนวนหนึ่งเพื่อ ประกอบเสริมกับการบันทึกเข้าซีดี-อาร์ (ซึ่งกระทำได้ครั้งเดียว) ทำให้มีเมตาตาตามากพอเพียงในการบรรจุใน

ระเบียบบรรณานุกรมในภายหลัง ดังนั้น กระบวนการที่ต้องจึงควรแยกกระบวนการปฏิบัติกรออกจากกระบวนการวิจัย และต้องยอมให้มีข้อมูลเสียหายหรือขาดช่วงไป เนื่องจากสันฮาร์ดดิสก์ โดยมิได้ทำการบันทึกลงซีดี-อาร์ เพื่อให้การเตรียมทำในส่วเมตาตาตาและเขตข้อมูลต่าง ๆ ของบรรณานุกรม มีความครบถ้วนสมบูรณ์เป็นที่แน่ชัด ก่อนกระบวนการบันทึกจะเริ่มต้นขึ้น

9. เนื่องจากข้อจำกัดของเวลาและขอบเขตของการวิจัย งานวิจัยไม่ได้ครอบคลุมถึงการวางระบบการให้บริการของผู้ที่จะนำข้อมูลไปใช้ คงหยุดแต่เพียงการตีพิมพ์ข้อมูลให้อยู่ในบันทึกถาวรในซีดี-อาร์ ซึ่งคณะผู้วิจัยยังมีความกังวลว่า กระบวนการให้บริการ อาจต้องการเมตาตาตาหรือเขตข้อมูลเพิ่มเติมในระบบบรรณานุกรม เขตข้อมูลหรือเมตาตาตาที่อาจหายไปนี้ จะส่งผลกระทบต่อกรบันทึกถาวรในระยะยาวได้ แต่จากประสบการณ์ของบรรณารักษ์ที่ร่วมโครงการ กระบวนการให้บริการที่ยังไม่ได้ทำไม่น่าจะมีผลกระทบที่สำคัญ

ข้อเสนอแนะที่ได้จากการวิจัย

จากผลการศึกษาค้นคว้า พบว่า มีข้อเสนอแนะต่อการกำหนดรูปแบบบรรณานุกรมและแนวทางการจัดเก็บในระยะยาวสำหรับข้อมูลต่อเนื่องที่ตัดมาเป็นกลุ่ม ดังนี้

1. จากประเด็นความเร่งรีบในการเก็บข้อมูล ควรมีการวิเคราะห์ข้อมูลหรือสาระเพื่อวางทั้งเมตาตาตาและระเบียบบรรณานุกรมก่อนที่จะเริ่มดำเนินการเก็บบันทึกข้อมูล ไม่ควรเป็นกังวลกับความเสียหายของข้อมูลที่เก็บไม่ทัน

2. เนื่องจากในการวิจัยครั้งนี้ มีกระบวนการในการเก็บและบันทึกข้อมูลในเชิงปฏิบัติการผสมอยู่ด้วย ซึ่งการดำเนินการนี้จะสิ้นสุดลงตามการสิ้นสุดของการวิจัย ดังนั้น มหาวิทยาลัยควรมีการเตรียมการปฏิบัติการ ซึ่งประกอบด้วย บุคลากร กระบวนการ และงบประมาณเข้าสนับสนุน เพื่อมิให้เกิดการขาดตอนในการเก็บข้อมูล ซึ่งเป็นการคำนึงต่อวัตถุประสงค์ในการทำบันทึกถาวร

3. ควรมีการตรวจสอบข้อมูลที่เก็บ เพราะเมื่อข้อมูลมีปริมาณมากขึ้นอย่างมากในเชิงสถิติ อาจจะมีโอกาสเกิดข้อมูลผิดพลาด (possibility of error) เช่น อ่านสาระไม่ได้ ควรมีการตรวจสอบเป็นระยะโดยการสุ่มตรวจจากแผ่นซีดี-อาร์

4. จากการศึกษาที่มีการเปลี่ยนแปลงเทคโนโลยีของสื่ออย่างต่อเนื่อง ทำให้สื่อที่ซื้อในแต่ละคราวต่างชนิดต่างยี่ห้อกัน (มีการเพิ่มความจุจากรุ่น 650 MB เป็น 700 MB) ดังนั้น ควรซื้อสื่อคราวละมาก ๆ เช่น 500-700 แผ่น เพื่อให้เกิดความสม่ำเสมอของสื่อและจะได้ราคาที่ถูก แต่ทั้งนี้ ไม่ควรซื้อในปริมาณเกินความต้องการมากเกินไป ไม่เช่นนั้นจะเหลือแผ่นตกค้างในแต่ละคราว ซึ่งมีผลต่ออายุของสื่อได้ ตัวอย่างเช่น ซื้อพร้อมกัน 5,000 แผ่น แต่ใช้หมดภายใน 2 ปี ดังนั้น แผ่นสุดท้ายที่ใช้จะมีอายุเท่ากับแผ่นแรกในกลุ่ม และข้อมูลที่เก็บในแผ่นสุดท้ายจะมีอายุน้อยกว่าข้อมูลที่เก็บในแผ่นแรก 2 ปีโดยข้อมูล 2 แผ่นจะหมดอายุพร้อมกัน

5. ในการจัดทำบันทึกถาวร จะต้องมีการวางแผนเตรียมการโดยกลุ่มคนจากหลายสาขาซึ่งประกอบด้วย บุคลากร 3 ประเภท ได้แก่

5.1. ผู้ที่มีความรู้ ความเชี่ยวชาญ ในสื่อสาระที่จะเก็บ ซึ่งจะให้คำแนะนำไม่เฉพาะ

คุณลักษณะของสาระ แต่รวมถึงการนำสาระไปใช้ประโยชน์ด้วย เช่น สาระเป็นข้อมูลตำแหน่งการโคจรของดาวเคราะห์ บุคลากรที่เกี่ยวข้องควรเป็นนักดาราศาสตร์ หากได้ผู้ที่เป็นเจ้าของหรือนักวิจัยที่ใช้ข้อมูลนั้นโดยตรง จะยิ่งได้คำแนะนำที่ตรงต่อความต้องการและลักษณะสมบัติของสาระยิ่งขึ้น

5.2. ผู้ที่มีความรู้ ความเชี่ยวชาญในการกำหนดระเบียบบรรณานุกรมและเมตา-ดาตา บุคคลในกลุ่มนี้ ได้แก่ บรรณารักษ์หรือนักเอกสารสนเทศ ที่มีความรู้ ความเข้าใจในการวิเคราะห์และกำหนดเขตข้อมูลทางบรรณานุกรม

5.3. ผู้เชี่ยวชาญทางด้านสื่อการบันทึกและเทคโนโลยีคอมพิวเตอร์ เพื่อช่วยกำหนดกระบวนการ รูปแบบและเมตาดาตา บางส่วนให้เหมาะสม

แต่ในการวิจัยคราวนี้ เนื่องจากสาระที่จะเก็บเป็นข้อมูลทางคอมพิวเตอร์ เพราะฉะนั้น บุคลากรประเภทที่ 1 และประเภทที่ 3 จึงเป็นคน ๆ เดียวกัน ทำให้สะดวกต่อการดำเนินการมากขึ้น

6. ควรคำนึงถึงการนำไปใช้งานมากกว่าที่เป็นอยู่ เพราะในงานวิจัยนี้เป็นเพียงสืบค้นได้ มีระบบบรรณานุกรม แต่ไม่ได้เน้นในแง่ของการให้บริการ คือ ผู้ใช้บริการสามารถนำข้อมูลไปใช้ได้ เพราะการให้ยืมสื่อออกไปจะเป็นการสร้างความเสี่ยงให้กับระบบข้อมูลมากกว่าสื่อประเภทกระดาษ เพราะผู้ใช้บริการโดยทั่วไปจะมีวัฒนธรรมการใช้กระดาษค่อนข้างดีอยู่แล้วในสังคม เอกสารที่จะเกิดความเสียหายต่อกระดาษจะต่ำและสามารถซ่อมแซมรูปเล่มได้ โดยมีผลแต่น้อยกับข้อมูล แต่ถ้าเป็นสื่ออิเล็กทรอนิกส์ ถ้าเกิดความเสียหาย แม้แต่เสียหายเพียงเล็กน้อย เช่น รอยขีดข่วน ๆ อาจทำให้สาระที่แผ่นนั้นใช้ประโยชน์ไม่ได้อีกต่อไป

7. ควรซื้อเครื่องพิมพ์อิงค์เจ็ตต์ สำหรับการพิมพ์ปกและบนแผ่นของซีดี-อาร์ ซึ่งแม้จะมีราคาสูง แต่ในระยะยาวทำให้การตีพิมพ์สื่อสาระเหล่านี้เป็นระเบียบ มีความชัดเจนมากขึ้น เนื่องจากในปัจจุบัน การเขียนแผ่นและปกกระทำด้วยมือ เกิดความไม่เป็นระเบียบเรียบร้อย

ข้อเสนอแนะเพื่องานวิจัยในอนาคต

ผลสืบเนื่องจากการวิจัยในครั้งนี้ ทำให้มีหัวข้อเรื่องที่เหมาะสมจะดำเนินการวิจัยเพื่อการศึกษาต่อเนื่อง ดังนี้

1. การทำ pre-processing ข้อมูลดิบ
ข้อมูลประเภทที่นำมาเป็นตัวอย่งศึกษาดังในวิจัยนี้ มีคุณสมบัติที่จะพอกพูนขึ้นอย่างรวดเร็ว เนื่องจากเป็นข้อมูลต่อเนื่องเกิดขึ้นตลอดเวลา ทำให้ปริมาณข้อมูลมีจำนวนมาก หากเป็นระบบเครือข่ายขนาดใหญ่อย่างยูนิเน็ต ซึ่งเชื่อมโยงมหาวิทยาลัยและวิทยาลัยต่าง ๆ ทั้งประเทศ มีสายสัญญาณขนาด 34 Mbps ไปยังต่างประเทศเพื่อเชื่อมต่ออินเทอร์เน็ต เปรียบเทียบกับขนาด 8 Mbps ของ จุฬาลงกรณ์มหาวิทยาลัยปริมาณข้อมูลเกี่ยวกับแทรฟฟิกจะสูงมหาศาล จำเป็นต้องมีการทำ pre-processing เพื่อลดปริมาณข้อมูลลงอย่างมาก เช่น มากกว่าร้อยละ 90 โดยวิธีการต่าง ๆ เช่น การบีบอัดข้อมูลที่เหมาะสมกว่าที่ทำอยู่ในวิจัยนี้ การกรองข้อมูลส่วนที่ไม่จำเป็นออก และการทำสรุปข้อมูลบางส่วน เหล่านี้เป็นต้น การทำ pre-processing ยังมีความจำเป็นอีกประการหนึ่ง คือ การกลั่นกรองข้อมูลที่เป็นความลับ เพื่อไม่ให้ข้อมูลส่วนบุคคลของการใช้อินเทอร์เน็ตแพร่หลายออกไปโดยไม่จำเป็น ซึ่งต้องมีการกลั่นกรอง มีวิธีแปลงจากข้อมูล

ดั้งเดิมไปเป็นข้อมูลที่ไม่เปิดเผยบุคคล รวมทั้งการแปลงกลับเป็นข้อมูลดั้งเดิม เพื่อประโยชน์ในการตรวจสอบต่าง ๆ ด้วย

2. การจัดเก็บทางกายภาพ

แนวทางการวิจัยในปัจจุบันมีความจำเป็นในการศึกษาหาแนวทางที่เหมาะสม เพื่อการเก็บบันทึกถาวรในรูปของห้องหรือบริเวณ โดยต้องพิจารณาประเด็นต่างๆ เช่น ด้านกายภาพของความชื้น อุณหภูมิ แสงสว่าง และการสำรองไฟ ความสะดวกในการเข้าถึงสื่อบันทึกเพื่อการปฏิบัติงาน การป้องกันภัยจากโจรกรรม และภัยพิบัติต่างๆ เช่น น้ำรั่วซึมหรือไฟไหม้ รวมถึงการแบ่งส่วนระหว่างการเตรียมสื่อที่เข้ามาใหม่ การจัดเก็บ และการนำออกให้บริการ นอกจากนี้ยังต้องคำนึงถึงการขยายตัวของปริมาณสื่อในระยะยาวอีกด้วย ซึ่งข้อคำนึงเหล่านี้จำเป็นต้องพิจารณาการลงทุนและความคุ้มค่าด้วยไม่มากนัก

3. การให้บริการอย่างครบวงจร

เนื่องจากข้อมูลที่เก็บมีความสำคัญในแง่มูลค่าของความต่อเนื่อง การให้ผู้ใช้ยืมข้อมูลออกไปเป็นแผ่นซีดี-อาร์ จะมีโอกาสเกิดความเสียหายต่อข้อมูลได้โดยง่าย ซึ่งทำให้ข้อมูลเกิดการขาดตอน ลดคุณภาพของข้อมูลลง จึงต้องมีมาตรการที่เหมาะสมในการให้บริการ เพื่อให้สอดคล้องกับข้อจำกัดของการเก็บบันทึกถาวร เช่น การสำเนาบันทึกถาวรเพื่อแก้ปัญหาจะไม่เหมาะสม เนื่องจากปริมาณมหาศาลของบันทึก อย่างไรก็ตามเป็นต้น

4. การทำ technology refreshment

ในการดำเนินการจัดเก็บแล้วเมื่อนานไปมีจำนวนสื่อมาก ผลกระทบในด้านนี้จะมีอย่างมหาศาลในเรื่องการย้ายข้อมูล เวลา และทรัพยากรที่ใช้ ซึ่งต้องคำนึงถึงสื่อบันทึก ซอฟต์แวร์และฮาร์ดแวร์ที่เกี่ยวข้อง รวมถึงองค์ประกอบข้างเคียงที่ทำให้สื่อบันทึกหมดอายุเร็วกว่าที่กำหนดไว้

งานวิจัยนี้ แม้จะมีประเด็นที่ต้องคำนึงและดำเนินการต่อเนื่อง ทั้งในด้านการวิจัย และการพัฒนาจัดทำระบบปฏิบัติการ แต่ผลลัพธ์ที่ได้ คือ การเลือกใช้สื่อบันทึกแบบซีดี-อาร์ การตัดส่วนข้อมูลแบบ chunk และการใช้เมตาตาตามงกกับระบบบรรณานุกรมแบบมาร์ก จะเป็นแนวทางพื้นฐานในการเพิ่มคุณค่า (value-add) ให้กับสาระพื้นถิ่นในหน่วยงานอุดมศึกษาและวิจัยอย่างจุฬาลงกรณ์มหาวิทยาลัยต่อไปในระยะยาวได้เป็นอย่างดี

จุฬาลงกรณ์มหาวิทยาลัย

รายการอ้างอิง

- [1] ศิริพงษ์ วิทย์วิโรจน์. (2544). “เอกสารประวัติศาสตร์ : ปัญหาในโลกดิจิทัล.” มติชนรายสัปดาห์ 21, 1078 (16 เมษายน) : 26.
- [2] Task Force on Archiving of Digital Information. (1996). "Preserving Digital Information." Report of the Task Force on Archiving of Digital Information Commissioned by the Commission on Preservation and Access and the Research Libraries Group, version 1 May 1996.
- [3] NLANR (2000). <http://ftp.ircache.net/Traces/README>. Visited : 30/5/2001. 10:30.
- [4] McCormick, John. (1996). DC-Recordable Drives : CD-R Finally Approaches the Mainstream for Archiving and Volume Storage. Government Computer News. Vol.15 No.11 (May 27) : 63(3).
- [5] Tangle, Laura. (1998). Storing Information on Disk and Tape Is Convenient. But How Long Will It Last? [On-line]. Available: <http://www.usnews.com/usnews/issue/980216/16gidi.htm>. Visited : 20/8/1999. 11:20.
- [6] Pace, Andrew K. (2000). "Digital Preservation : Everything New Is Old Again." Computer in Libraries. (February) : 55-58.
- [7] Chen, Su-Shing. (2001). "The Paradox of Digital Preservation." Computer. (March) : 24-28.
- [8] Solbakk, Svein Arne. (1995). Long Term Preservation of Electronic Material : Preliminary Experiences from the Norwegian National Library. [On-line]. Available : <http://www.nb.no/rapporter/preserv.html>. Visited : 1/3/2002. 18:31.
- [9] Granger, Stewart. (2000). Metadata and Digital Preservation : a Plea for Cross-interest Collaboration. [On-line]. Available : <http://ds.dial.pipex.com/stewartg/metpres.html>. Visited : 23/5/2001. 10:30.
- [10] Lehmann-Haupt, John. (2000). "The Well-Preserved." In Think Research. No.2 : 30-31.
- [11] The Making of America II Tested Project White Paper Version 2.0 (September 15, 1998). [On-line]. Available: <http://sunsite.berkeley.edu/MOA2/wp-v2.html>. Visited: 9/04/2002. 17:30.
- [12] McKemish, S., Perer, D. (1998). Towards Frameworks for Standardising Recordkeeping Metadata. Archives and Manuscripts 26 : 24-25.
- [13] Bearman, D., Sochats, K. (1996). Metadata Requirements for Evidence. University of Pittsburgh School of Information Science, Pittsburgh, Pa. [On-line]. Available : <http://www.lis.pitt.edu/~nhprc/BACartic.html>. Visited : 7/3/2002. 19:12.
- [14] Duranti, Luciana, Eastwood, Terry and MacNeil, Heather. The Preservation of the Integrity of Electronic Records. [On-line]. Available: <http://www.interpares.org/UBCProject/>. Visited: 7/3/2002. 20:10.

- [15] Gilliland-Swetland, A. J., Eppard, P.B. (2000). Preserving the Authenticity of Contingent Digital Objects : the InterPARES project. D-Lib Magazine, 6. [On-line]. Available : <http://www.dlib.org/dlib/july00/eppard/07eppard.html>. Visited : 7/3/2002. 19:20.
- [16] National Archives of Australia. (1999). Recordkeeping Metadata Standard for Commonwealth Agencies, version 1.0. National Archives of Australia, Canberra. [On-line]. Available : <http://www.naa.gov.au/recordkeeping/control/rkms/summary.htm>. Visited : 15/3/2002. 20:15.
- [17] Consultative Committee for Space Data Systems. (2001). Reference Model for an Open Archival Information System (OAIS), Red Book, Issue. 1. CCSDS 650.0-R-2. CCSDS Secretariat, National Aeronautics and Space Administration, Washington, D.C. [On-line]. Available : <http://ssdoo.gsfc.nost.isoas/ref model.html>. Visited : 5/3/2002. 9:10.
- [18] Metadata for Digital Preservation the Cedars Project Outline Specification Draft for Public Consultation. (2000). [On-line]. Available : <http://www.leeds.ac.uk/cedars/documents/Metadata/cedars/html>. Visited : 14/3/2002. 19:20.
- [19] Werf-Davelaar, Titia van der. (1999). "Long-term Preservation of Electronic Publications : the NEDLIB Project." D-Lib Magazine, 5, (9), September.
- [20] National Library of Australia. (1999). Preservation Metadata for Digital Collection Exposure Draft. [On-line]. Available : <http://www.nla.gov.au/preserve/pmeta.html>. Visited: 14/3/2002. 19:50.
- [21] RLG Working Group on Preservation Issues of Metadata : Final Report. (1998). [On-line]. Available : <http://www.rlg.org/preserv/presmeta.html>. Visited : 14/3/2002. 20:25.
- [22] Library of Congress. (2000). Library of Congress Digital Repository Development : Core Metadata Elements. [On-line]. Available : <http://lcweb.loc.gov/standards/metable.html>. Visited : 4/4/2002. 18:30.
- [23] ประจักษ์ พุ่มวิเศษ. (2538). "เครือข่ายคอมพิวเตอร์ห้องสมุดในจุฬาฯ : ความเป็นมาของระบบห้องสมุดอัตโนมัติ." ใน ระบบห้องสมุดอัตโนมัติ : แนวคิดและประสบการณ์. กรุงเทพฯ : สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย : 1-18.
- [24] Sparks, Michael, Neisser, George and Hanby, Richard. (1999). "An Initial Statistical Analysis of the Performance of the UK National JANET Cache." In the 4th International Web Caching Workshop 1999. San Diego, California. Mar 31-Apr 2, 1999. [On-line]. Available : <http://www.cache.ja.net/>. Visited : 6/7/2000. 17:20.
- [25] ภาณุพันธ์ สุวรรณมาตร. (2541). การวิเคราะห์สถิติการใช้งานอินเทอร์เน็ตในระบบเครือข่ายระดับสถาบันอุดมศึกษา. วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ จุฬาลงกรณ์มหาวิทยาลัย.
- [26] นพรัตน์ พรหมหาไชย. (2543). การวิเคราะห์ความสัมพันธ์ระหว่างลักษณะของยูอาร์แอลกับการใช้งานบริการเว็บแคช. วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ จุฬาลงกรณ์มหาวิทยาลัย.

- [27] พรทวี วัฒนวิฑูกร. (2542). การวิเคราะห์เปรียบเทียบประสิทธิภาพของขั้นตอนวิธีการแทนที่ในพรีอ็อกซีแคช. วิทยานิพนธ์ปริญญาามหาบัณฑิต จุฬาลงกรณ์มหาวิทยาลัย.
- [28] อัมพูนี พัฒนกำจร. (2544). การศึกษาวิธีการคงความต้องการกันของเอกสารในเว็บแคชโดยใช้วิธีส่งข้อความการเรียกขอแบบไอเอ็มล่องหน้า. วิทยานิพนธ์ปริญญาามหาบัณฑิต จุฬาลงกรณ์มหาวิทยาลัย.
- [29] รชา โรจนตรีคุณ. (2544). การวิเคราะห์ผลกระทบของจำนวนแคชและวิธีการประสานงานที่มีประสิทธิภาพของแคชแบบกระจาย. วิทยานิพนธ์ปริญญาามหาบัณฑิต จุฬาลงกรณ์มหาวิทยาลัย.
- [30] "The Future's ... Blue?." (2002). Tomorrow's Technology Today (T3). May : 44-48.
- [31] Longevity (n.d.). [On-line]. Available. <http://www.maxell.com/cdr/osta/long.htm>. Visited : 20/8/1999. 9:20.
- [32] "9 เหตุผลที่ CD-ROM Drive ต้องคู่กับคอมพิวเตอร์." (2538). คอมพิวเตอร์ทูเดย์ 5, 1 (กรกฎาคม) : 116-122.
- [33] พันธุ์ศักดิ์ ศรีทรัพย์. (2542). "DVD เทคโนโลยีซีดีรอมแห่งอนาคต." E-Commerce. 16, 177 (มิถุนายน) : 57-60.
- [34] พัวร์, อัลเฟรด. (2540). "DVD และ CD-ROM แหล่งเก็บข้อมูลในศตวรรษที่ 21." PC Magazine Thailand. 5, 52 (กรกฎาคม) : 108-119.
- [35] A Comparison of Various Mass Storage Media. (2001). [On-line]. Available : <http://www.3dham.com/storage>. Visited : 6/6/2001. 11:29.
- [36] Dowling, Mark. (2000). Secondary Storage Media. [On-line]. Available : <http://www.medicinenu.net.au/computing/columns/gpcomputing/gpcomputing5.html>. Visited : 6/6/2001. 11:50.
- [37] Traub, David C. (1994). "A Historical Perspective of CD-ROM." In CD-ROM Handbook. 2nd ed. New York : Intertext Publications.
- [38] TDK CD-R Technology. (1999). [On-line]. Available : <http://www.cd-info.com/CDIC/Technology/CD-R/Media/TDK.html>. Visited : 20/8/2001. 13:00.
- [39] Stinson, Douglas, Ameli, Fred and Zaino, Nick. (1995). Lifetime of KODAK Writable CD and Photo CD Media. [On-line]. Available : <http://www.cd-info.com/CDIC/Technology/CD-R/Media/Kodak.html>. Visited : 20/8/2001. 11:50.
- [40] White, Stephen. (2001). A Brief History of Computing-Operating Systems. [On-line]. Available: <http://www.ox.compsoc.net/~swhite/history/timeline-OS.html>. Visited: 10/4/2002. 11:30.
- [41] Lloyd, Eddie. Windows Timeline. (2000). [On-line]. Available: <http://www.geocities.com/the cyberprice/wintimeline.htm>. Visited: 10/4/2002. 12:10.
- [42] Webster's Third New International Dictionary Unabridged and Seven Language Dictionary. (1986). Chicago : Encyclopedia Britannica.

[43] SETI Australia Centre. (2001). Updates from the Parkes Radio Telescope on Australian SETI Effort. [On-line]. Available : <http://seti.uws.edu.au/main/livedata.htm>. Visited : 21/5/2001. 15:03.

[44] Gorman, Michael. (1999). "Metadata or Cataloging? A False Choice." Journal of Internet Cataloging. Vol.2, No. 1 : 1-22.

[45] Tennant, Roy. (1998). "21st-Century Cataloging." Library Journal. (April 15) : 30.

[46] The Library of Congress, Network Development and MARC Standard Office and National Library of Canada Standards and Support. (1999). MARC21 Formats for Bibliographic Data : Including Guideline for Content Designation. Washington : Cataloging Distribution Service, Library of Congress.

[47] Innovative Interfaces. [On-line]. Available: <http://www.iii.com>. Visited: 20/1/2002. 9:10.

[48] Clark, Judith. (2001). Subject: Re: Follow up Chula C529110. Email: III Help Desk helpdesk@iii.com.



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ประวัติผู้วิจัย

อ. ดร. ยรรยง เต็งอำนวย

จบการศึกษาระดับปริญญาตรี ทางวิศวกรรมไฟฟ้า
และปริญญาโทด้านคอมพิวเตอร์จากจุฬาลงกรณ์
มหาวิทยาลัยและจบปริญญาเอกด้านคอมพิวเตอร์
จากมหาวิทยาลัยไอโอวาสเตท
ปัจจุบันทำงานเป็นอาจารย์ที่ภาควิศวกรรมคอมพิวเตอร์
จุฬาลงกรณ์มหาวิทยาลัย เชี่ยวชาญและศึกษาค้นคว้า
ด้านวิทยาการสารสนเทศและระบบปฏิบัติการ
ระบบความมั่นคงเครือข่ายอินเทอร์เน็ตและระบบกระจาย



น.ส. สุภาพร ชัยธัมมะปกรณ์

จบการศึกษาระดับปริญญาตรี (ภาษาอังกฤษ) จาก
มหาวิทยาลัยศิลปากร และปริญญาโททาง
บรรณารักษศาสตร์ จากจุฬาลงกรณ์มหาวิทยาลัย
ปัจจุบันเป็นบรรณารักษ์ฝ่ายวิเคราะห์ทรัพยากร
สารสนเทศ สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย



นายนิทัศน์ สุทวีปราโมชานนท์

นิสิตบัณฑิตศึกษา ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์
จุฬาลงกรณ์มหาวิทยาลัย



ภาคผนวก

- ภาคผนวก ก. คู่มือเขตข้อมูลในการลงรายการระเบียบบรรณานุกรมข้อมูลสถิติการใช้อินเทอร์เน็ตใน
จุฬาลงกรณ์มหาวิทยาลัย
- ภาคผนวก ข. คู่มือการปฏิบัติงานการลงรายการระเบียบบรรณานุกรมข้อมูลสถิติการใช้งาน
อินเทอร์เน็ต
- ภาคผนวก ค. เมตาตาตาข้อมูลดิจิทัลของโครงการของห้องสมุดทั้งห้า
- ภาคผนวก ง. คู่มือการปฏิบัติงานการบันทึกข้อมูลลงซีดี-อาร์
- ภาคผนวก จ. Script สำหรับสนทนากับอินโนแพค



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ก

คู่มือเขตข้อมูลการลงรายการระเบียบบรรณานุกรมข้อมูลสถิติการใช้อินเทอร์เน็ตใน
จุฬาลงกรณ์มหาวิทยาลัย



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

คู่มือเขตข้อมูลในการลงรายการระเบียบบรรณานุกรมข้อมูลสถิติการใช้อินเทอร์เน็ต
ในจุฬาลงกรณ์มหาวิทยาลัย

เป็นส่วนหนึ่งของการวิจัย

เรื่อง

การกำหนดรูปแบบบรรณานุกรมและการวางแนวทางการจัดเก็บแบบถาวรสำหรับข้อมูลต่อเนื่องที่ตัดมาเป็นกลุ่ม

ให้สืบค้นผ่านห้องสมุดอัตโนมัติ : กรณีศึกษาข้อมูลสถิติการใช้งานอินเทอร์เน็ตของ

จุฬาลงกรณ์มหาวิทยาลัย

สนับสนุนโดย

กองทุนรัชดาภิเษกสมโภช

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สถาบันวิทยบริการ

และ

ห้องปฏิบัติการวิศวกรรมระบบสารสนเทศ ภาควิชาวิศวกรรมคอมพิวเตอร์

จุฬาลงกรณ์มหาวิทยาลัย

30 เมษายน 2545

**คู่มือเขตข้อมูลในการลงรายการระเบียบบรรณานุกรมข้อมูลสถิติการใช้อินเทอร์เน็ต
ในจุฬาลงกรณ์มหาวิทยาลัย**

บทนำ

เมตาาดาหรือเขตข้อมูลในการลงรายการระเบียบบรรณานุกรมข้อมูลสถิติการใช้อินเทอร์เน็ตในจุฬาลงกรณ์มหาวิทยาลัย เป็นข้อกำหนดที่เกิดจากผลการศึกษาเรื่อง “การกำหนดรูปแบบบรรณานุกรมและการวางแนวทางการจัดเก็บแบบถาวร สำหรับข้อมูลต่อเนื่องถัดมาเป็นกลุ่มให้สืบค้นผ่านห้องสมุดอัตโนมัติ : การศึกษา ข้อมูลสถิติการใช้งานอินเทอร์เน็ตของจุฬาลงกรณ์มหาวิทยาลัย” ในระหว่างเดือน พฤษภาคม พ.ศ. 2544 – พฤษภาคม พ.ศ. 2545 โดยคณะผู้วิจัย ซึ่งประกอบด้วย อ.ดร.ยรรยง เต็งอำนวย อาจารย์ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย และในฐานะรองผู้อำนวยการสถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย รวมถึงในฐานะผู้เชี่ยวชาญทางด้านเครือข่ายคอมพิวเตอร์ นางสาวสุภาพร ชัยธัมมะปกรณ์ บรรณารักษ์หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย และนายนิทัศน์ สุทวีปราโมชานนท์ นิสิตบัณฑิตศึกษา ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย เล็งเห็นว่าข้อมูลสถิติการใช้อินเทอร์เน็ตในจุฬาลงกรณ์มหาวิทยาลัย มีประโยชน์ต่อการบริหารจัดการ และต่อการวิจัย แต่ไม่มีการจัดเก็บที่เป็นระบบ ซึ่งมีแต่จะสูญหายไป และข้อมูลประเภทนี้ถือว่าเป็นสาระพื้นถิ่น (local content) อันมีค่าของมหาวิทยาลัย ซึ่งมหาวิทยาลัยควรจะต้องตระหนักถึงคุณค่า หาแนวทางในการรวบรวมและจัดเก็บแบบถาวรเพื่อการใช้งานในระยะยาว

จากการศึกษาวิจัยนี้ ทำให้สามารถกำหนดเมตาาดาหรือการลงรายการระเบียบบรรณานุกรมข้อมูลสถิติการใช้อินเทอร์เน็ต เพื่อเป็นข้อมูลในการเข้าถึงข้อมูลสถิติ ตลอดจนข้อมูลในการสงวนรักษาข้อมูลสำหรับผู้ดูแลระบบอีกด้วย โดยใช้รูปแบบการลงรายการระเบียบบรรณานุกรมแบบมาร์ก 21 ซึ่งห้องสมุดสถาบันอุดมศึกษาอื่นๆ ใช้ระบบนี้กันเป็นส่วนใหญ่ สามารถใช้เมตาาดาที่กำหนดขึ้นนี้เป็นต้นแบบในการลงรายการระเบียบบรรณานุกรมข้อมูลสถิติการใช้อินเทอร์เน็ต หรือข้อมูลที่มีลักษณะต่อเนื่องที่ถัดมาเป็นกลุ่ม เมตาาดาหรือเขตข้อมูลที่ใช้ในการลงรายการระเบียบบรรณานุกรมประกอบด้วยเขตข้อมูลดังนี้

ชื่อ	1. Local Call Number
คำอธิบาย	ใช้สำหรับการลงรายการหมายเลขซีดี-อาร์
การซ้ำของข้อมูล	-
หมายเลขเขตข้อมูล	099
ตัวอย่าง	CD TR 2000-001
	CD WT 2000-001
หมายเหตุ	เนื่องจากข้อมูลสถิติการใช้อินเทอร์เน็ตบันทึกลงในซีดี-อาร์

ชื่อ	2. Title Statement
คำอธิบาย	ใช้สำหรับลงรายการชื่อเรื่อง
การซ้ำของข้อมูล	-
หมายเลขเขตข้อมูล	245
เขตข้อมูลย่อย	#a, #b, #c, #h
หมายเหตุ	Internet traffic log h[Computer file] : bUninet gateway of CUNET/ cSystem Department, Center of Academic Resources, Chulalongkorn University

ชื่อ	3. Publication, Distribution, etc.
คำอธิบาย	ใช้สำหรับลงรายการสถานที่ ผู้รับผิดชอบงานและปีที่บันทึก
การซ้ำของข้อมูล	-
หมายเลขเขตข้อมูล	260
เขตข้อมูลย่อย	#a, #b, #c
ตัวอย่าง	Bangkok : bSystem Department, Center of Academic Resources, Chulalongkorn University, c2001

ชื่อ	4. Physical Description
คำอธิบาย	ใช้สำหรับลงรายการลักษณะของซีดี-อาร์ เช่น จำนวนแผ่น ขนาดของแฟ้มข้อมูล
การซ้ำของข้อมูล	-
หมายเลขเขตข้อมูล	300
เขตข้อมูลย่อย	#a, #c
ตัวอย่าง	Computer optical discs (CD-R) : c4 ¾ in 1 log file (145 MB)

ชื่อ	5. Current Publication Frequency
คำอธิบาย	ใช้สำหรับลงรายการจำนวนแผ่นซีดี-อาร์ที่บันทึกโดยประมาณต่อปี
การซ้ำของข้อมูล	-
หมายเลขเขตข้อมูล	310
ตัวอย่าง	119 discs per year (approx.)

ชื่อ	6. Dates of Publication and/or Volume Designation
คำอธิบาย	ใช้สำหรับลงรายการปีที่เริ่มต้นบันทึกข้อมูล
การซ้ำของข้อมูล	ซ้ำ
หมายเลขเขตข้อมูล	362
ตัวอย่าง	2001 -

ชื่อ	7. General Note
คำอธิบาย	ใช้สำหรับลงรายการหมายเหตุทั่วไป ในที่นี้ลงรายการหมายเลขไอพีที่เชื่อมต่อ กัน
การซ้ำของข้อมูล	ซ้ำ
หมายเลขเขตข้อมูล	500
ตัวอย่าง	Location : IP address 202.28.0.1.125/25- 10 Mbps connect to KMUTT's Router : IP address 202.28.18.35/27- 155 Mbps connect to MUA ATM Backbone : IP address 202.28.0.253/25- 100 Mbps connect to CUNET

ชื่อ	8. Restrictions on Access Note
คำอธิบาย	ใช้สำหรับลงรายการข้อจำกัดในการเข้าใช้ข้อมูล
การซ้ำของข้อมูล	ซ้ำ
หมายเลขเขตข้อมูล	506
เขตข้อมูลย่อย	#a, #c, #d
ตัวอย่าง	Not available for publication or commercialization ; aRestricted online access through appointment and with password ; cPhysical access prohibited ; dFor use in researches granted by official bodies

ชื่อ	9. Creation / Production Credits Note
คำอธิบาย	ใช้สำหรับลงรายการผู้สร้างข้อมูล ผู้บันทึกข้อมูลลงซีดี-อาร์
การซ้ำของข้อมูล	-
หมายเลขเขตข้อมูล	508
ตัวอย่าง	Data stream originator, Office of Information Technology, Chulalongkorn University ; Chunks and CD-ROMs produced from 2001-, Nitass Sutaveepramochanon

ชื่อ	10. Numbering Peculiarities Note
คำอธิบาย	ใช้สำหรับลงรายการจำนวนข้อมูลที่มีอยู่ / ข้อมูลที่หายไป
การเข้าของข้อมูล	ซ้ำ
หมายเลขเขตข้อมูล	515
ตัวอย่าง	Data covers May 2001- Missing data / very irregular data (see chunk for details)

ชื่อ	11. Type of Computer File or Data Note
คำอธิบาย	ใช้สำหรับลงรายการลักษณะของข้อมูล
การเข้าของข้อมูล	ซ้ำ
หมายเลขเขตข้อมูล	516
ตัวอย่าง	Numeric (Computer log file)

ชื่อ	12. Date / Time and Place of an Event Note
คำอธิบาย	ใช้สำหรับลงรายการวัน เดือน ปีที่บันทึกข้อมูลซีดี-อาร์
การเข้าของข้อมูล	ซ้ำ
หมายเลขเขตข้อมูล	518
ตัวอย่าง	CD published on Mar 05, 2002

ชื่อ	13. Summary
คำอธิบาย	ใช้สำหรับลงรายการเนื้อเรื่องย่อ
การเข้าของข้อมูล	ซ้ำ
หมายเลขเขตข้อมูล	520
ตัวอย่าง	Contains data on the Internet traffic from router : Cisco 7206/IOS version 11.1 (17.3) CA ; Changed to (35) CC2 in September 1 st 2001. Aggregates data from program Netflowcollector (Version 2), under CUNET Administration and generated by account.car.chula.ac.th

ชื่อ	14. Preferred Citation of Described Material Note
คำอธิบาย	ใช้สำหรับลงรายการการอ้างอิงทางบรรณานุกรม
การเข้าของข้อมูล	ซ้ำ
หมายเลขเขตข้อมูล	524
ตัวอย่าง	Chulalongkorn University. Center of Academic Resources. System Department. "Internet traffic log 20000825 17:50 – 20000825 23:50" Internet Traffic log, CD TR 2000-001, File CU-TR-2000-0001.zip

ชื่อ	15. Location of Originals / Duplicates Note
คำอธิบาย	ใช้สำหรับลงรายการสถานที่ แหล่งกำเนิดข้อมูล / ผู้ดูแลข้อมูล
การเข้าของข้อมูล	ซ้ำ
หมายเลขเขตข้อมูล	535
ตัวอย่าง	Chulalongkorn University. Center of Academic Resources. System Department

ชื่อ	16. Funding Information Note
คำอธิบาย	ใช้สำหรับลงรายการชื่อทุนที่ได้รับ
การเข้าของข้อมูล	ซ้ำ
หมายเลขเขตข้อมูล	536
ตัวอย่าง	Original collection up to April 2002 supported by Rachadapiseksompoj Research Fund, Research Affairs, Chulalongkorn University

ชื่อ	17. System Details Note
คำอธิบาย	ใช้สำหรับลงรายการข้อกำหนดของระบบคือ โปรแกรมที่ใช้บีบอัดข้อมูล
การเข้าของข้อมูล	ซ้ำ
หมายเลขเขตข้อมูล	538
ตัวอย่าง	Data and metadata packaged and compressed by program zip v.2.1 on Solaris 2.5.1

ชื่อ	18. Terms Governing Use and Reproduction Note
คำอธิบาย	ใช้สำหรับลงรายการเงื่อนไขการใช้งานและการสำเนาออก
การซ้ำของข้อมูล	ซ้ำ
หมายเลขเขตข้อมูล	540
เขตข้อมูลย่อย	#b
ตัวอย่าง	bRelated privacy information must be removed or masked before publication of research result except by specific authorization from the owner

ชื่อ	19. Immediate Source of Acquisition Note
คำอธิบาย	ใช้สำหรับลงรายการวิธีการได้มาซึ่งข้อมูล และหน่วยของข้อมูล
การซ้ำของข้อมูล	ซ้ำ
หมายเลขเขตข้อมูล	541
เขตข้อมูลย่อย	#c, #o
ตัวอย่าง	cContains the Internet traffic from router : Cisco 7206/IOS version : 11.1 (17.3) CA and (35) CC2 since 1 st September 2001 ; oChunk of 30 minute log data from router

ชื่อ	20. Methodology
คำอธิบาย	ใช้สำหรับลงรายการกรรมวิธีการประมวลข้อมูล
การซ้ำของข้อมูล	ซ้ำ
หมายเลขเขตข้อมูล	567
ตัวอย่าง	30 - minute chunk data stream

ชื่อ	21. Action Note
คำอธิบาย	ใช้สำหรับลงรายการวัน เดือน ปีที่คาดว่าซีดี-อาร์จะหมดอายุ
การซ้ำของข้อมูล	ซ้ำ
หมายเลขเขตข้อมูล	583
เขตข้อมูลย่อย	#c
ตัวอย่าง	Expected expiry date ; c20070305

ชื่อ	22. Subject Added Entry – Corporate Name
คำอธิบาย	ใช้สำหรับลงรายการหัวเรื่องที่เป็นองค์กร / หน่วยงานที่เชื่อมโยงในเครือข่าย
การเข้าของข้อมูล	ซ้ำ
หมายเลขเขตข้อมูล	610
ตัวอย่าง	KMUTT MUA CUNET

ชื่อ	23. Subject Added Entry – Topical Term
คำอธิบาย	ใช้สำหรับลงรายการหัวเรื่องทั่วไป ในที่นี้ลงรายการหมายเลขไอพีของเครื่องที่
	เชื่อมโยง
การเข้าของข้อมูล	ซ้ำ
หมายเลขเขตข้อมูล	650
ตัวอย่าง	202.28.0.125/25 202.28.18.35/27

ชื่อ	24. Added Entry – Corporate Name
คำอธิบาย	ใช้สำหรับลงรายการเพิ่มของหน่วยงานจากเขตข้อมูล 245
การเข้าของข้อมูล	ซ้ำ
หมายเลขเขตข้อมูล	710
เขตข้อมูลย่อย	#a, #b
ตัวอย่าง	Chulalongkorn University. bCenter of Academic Resources. bSystem Department

ชื่อ	25. Host Item Entry
คำอธิบาย	ใช้สำหรับลงรายการชื่อเรื่องของ chunk หมายเลขของซีดี-อาร์ และชื่อเพิ่มข้อมูล ของแต่ละ chunk
การเข้าของข้อมูล	ซ้ำ
หมายเลขเขตข้อมูล	773
เขตข้อมูลย่อย	#t, #g
ตัวอย่าง	tInternet Traffic Log, gCD TR 2000-001, File CU-TR-2000-0001.zip

ภาคผนวก ข

คู่มือการปฏิบัติงานการลงรายการระเบียบบริหารนุกรมข้อมูลสถิติการใช้อินเทอร์เน็ตใน
จุฬาลงกรณ์มหาวิทยาลัย



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

คู่มือการปฏิบัติงาน

การลงรายการระเบียบบรรณานุกรมข้อมูลสถิติการใช้อินเทอร์เน็ตในจุฬาลงกรณ์มหาวิทยาลัย

เป็นส่วนหนึ่งของการวิจัย

เรื่อง

การกำหนดรูปแบบบรรณานุกรมและการวางแนวทางการจัดเก็บแบบถาวรสำหรับข้อมูลต่อเนื่องที่ตัดมาเป็นกลุ่ม

ให้สืบค้นผ่านห้องสมุดอัตโนมัติ : กรณีศึกษาข้อมูลสถิติการใช้งานอินเทอร์เน็ตของ

จุฬาลงกรณ์มหาวิทยาลัย

สนับสนุนโดย

กองทุนรัชดาภิเษกสมโภช

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สถาบันวิทยบริการ

และ

ห้องปฏิบัติการวิศวกรรมระบบสารสนเทศ ภาควิชาวิศวกรรมคอมพิวเตอร์

จุฬาลงกรณ์มหาวิทยาลัย

30 เมษายน 2545

**คู่มือการปฏิบัติงานการลงรายการระเบียบบรรณานุกรมข้อมูลสถิติการใช้อินเทอร์เน็ต
ในจุฬาลงกรณ์มหาวิทยาลัย**

การลงรายการระเบียบบรรณานุกรมข้อมูลสถิติการใช้อินเทอร์เน็ตในจุฬาลงกรณ์มหาวิทยาลัย เป็นการลงรายการผ่านระบบห้องสมุดอัตโนมัติในเครื่องข่ายจุฬาลงกรณ์มหาวิทยาลัยที่ชื่อว่าอิน โนแพค โดยใช้มาตรฐานในการลงรายการของมาร์ก 21 ประกอบด้วย การลงรายการระเบียบบรรณานุกรมของข้อมูล 2 ประเภท คือ ข้อมูลแทรฟฟิก และข้อมูลเว็บแคช โดยมีเขตข้อมูลในการลงรายการ ดังนี้

เขตข้อมูลในการลงรายการระเบียบบรรณานุกรม

หมายเลขเขตข้อมูล	ชื่อของเขตข้อมูล	รายการ
099 * **	Local Call Number	เลขหมู่ของแผ่นซีดี-อาร์
245 # a * **	Title	ชื่อเรื่อง
245 # c *	Statement of responsibility, etc	ชื่อผู้รับผิดชอบ
260 # a *	Place of publication, distribution, etc	สถานที่ของผู้รับผิดชอบ
260 # b *	Name of publisher, distributor, etc	ชื่อผู้จัดทำ
260 # c * **	Date of publication, distribution, etc	ปีที่บันทึกซีดี-อาร์/ปีที่เกิดข้อมูล
300 # a * **	Physical description (Extent)	จำนวนแผ่นซีดี-อาร์ / จำนวนของ แฟ้ม
300 # c *	Dimensions	ขนาดของแฟ้มข้อมูล
310 # a *	Current publication frequency	คาบการเกิดข้อมูล
362 # a *	Dates of publication and/or sequential designation	ปีที่ผลิตข้อมูล และ/หรือลำดับของ การเกิดข้อมูล
500 # a * **	General Note	หมายเหตุ
506 # a *	Terms governing access	ข้อจำกัดในการเข้าใช้
506 # c *	Physical access provisions	การเข้าใช้แผ่นบันทึก
506 # d *	Authorized users	ข้อกำหนดในการใช้
508 # a *	Creation/production credits note	แหล่งที่เกิดข้อมูลและผู้บันทึก
515 # a *	Numbering peculiarities note	ขอบเขตของข้อมูล

เขตข้อมูลในการลงรายการระเบียบบรรณานุกรม (ต่อ)

หมายเลขเขตข้อมูล	ชื่อของเขตข้อมูล	รายการ
516 # a *	Type of computer file or data note	ลักษณะของข้อมูล
518 # a *	Date/time and place of an event note	วัน เดือน ปี ที่บันทึกข้อมูลลงซีดี-อาร์
520 # a *	Summary, etc. note	เนื้อเรื่องย่อ
524 # a *	Preferred citation of described materials note	การอ้างถึงทางบรรณานุกรม
535 # a *	Custodian	ผู้ดูแลข้อมูล
535 # b *	Postal address	ที่อยู่ของผู้ดูแลข้อมูล
536 # a *	Text of note (Funding Information Note)	ทุนที่ได้รับ
538 # a *	System details note	ข้อกำหนดของระบบ
540 # b *	Terms governing use and reproduction Note (Jurisdiction)	เงื่อนไขการใช้งานและการทำสำเนาออก
541 # c *	Method of acquisition	วิธีการได้มาซึ่งข้อมูล
541 # o *	Type of unit	หน่วยของข้อมูล
567 # a *	Methodology note	กรรมวิธีการประมวลข้อมูล
583 # a *	Action	วัน เดือน ปี ที่คาดว่าซีดี-อาร์หมดอายุ
610 # a *	Subject Added Entry Corporate Name	หัวเรื่องสำหรับหน่วยงานที่เชื่อมโยงไปถึง
650 # a *	Subject Added Entry - Topical Term	หัวเรื่องสำหรับหมายเลขไอพีของเครื่องและหัวเรื่องทั่วไป
710 # a * **	Added Entry - Corporate Name	หน่วยงานที่ผลิตข้อมูล
773 # t **	Host Item Entry: Title	ชื่อของ chunk
773 # g **	Host Item Entry: Relationship information	หมายเลขเพิ่มข้อมูลของ chunk

* หมายถึง เขตข้อมูลที่ลงใน bibliographical record of collection

** หมายถึง เขตข้อมูลที่ลงใน bibliographicla record of content analysis

จากลักษณะของข้อมูลแทรฟฟิกและเว็บแคม ทำให้ต้องแบ่งการลงรายการระเบียบนบรรณานุกรมออกเป็น 2 ระดับ คือ

1. การลงรายการระเบียบนบรรณานุกรมในระดับคอลเลกชัน (bibliographical record of collection)
2. การลงรายการระเบียบนบรรณานุกรมในระดับช่วงเวลา (bibliographical record of content analysis)

การลงรายการระเบียบนบรรณานุกรมในระดับคอลเลกชัน

ในขณะนี้มีการลงรายการระเบียบนบรรณานุกรมในระดับคอลเลกชันของแทรฟฟิก 2 ระเบียบ และของเว็บแคม 1 ระเบียบ ในการลงรายการประเภทนี้ จะลงเมื่อมีการเปลี่ยน configuration ของเครื่องใหม่ การลงรายการให้ลงเฉพาะเขตข้อมูลที่มีเครื่องหมายดอกจัน (*) 1 อัน ที่กำกับข้างท้ายเขตข้อมูลตามตารางข้างต้น โดยใช้คำพรรณนาในการลงรายการตามตัวอย่าง ดังต่อไปนี้

เขตข้อมูล	ตัวบ่งชี้	ลักษณะสมบัติ
099	2	CD TR 2002-001
245	00	Internet traffic log h[Computer file] : bUninet gateway of CUNET / cSystem Department, Center of Academic Resources, Chulalongkorn University
260		Bangkok : bSystem Department, Center of Academic Resources, Chulalongkorn University, c2001-
300		Computer optical laser discs (CD-R) ; c4 3/4 in
310		119 discs per year (approx.)
362	1	2001-
500		Location: IP address 202.28.0.1.125/25 - 10 Mbps connect to KMUTT's Router; IP address 202.28.18.35/27 - 155 Mbps connect to MUA ATM backbone; IP address 202.28.0.253/25 - 100 Mbps connect to CUNET
506		Not available for publication or commercialization ; aRestricted online access through appointment and with password ; cPhysical access prohibited ; dFor use in researches granted by official bodies
508		Data stream originator, Office of Information Technology, Chulalongkorn University; Chunks and CD-ROMs produced from 2001- , Nitass Sutaveepromochanon
515		Data covers May 2001-
515		Missing data / very irregular data (see chunk for details)
516		Numeric (Computer log file)
520		Contains data on the Internet traffic from router: Cisco 7206/IOS version 11.1 (17.3) CA; Changed to (35) CC2 in September 1st 2001. Aggregates data from program Netflowcollector (version 2), under CUNET administration and

การลงรายการระเบียบนบรรณานุกรมในระดับ collection (ต่อ)

เขตข้อมูล	ตัวบ่งชี้	ลักษณะสมบัติ
		Generated by account.car.chula.ac.th
524		Creator. "Name of Chunk Title." Name of CD-R title, Number of CD-R, Chunk ID. ตัวอย่าง: Chulalongkorn University. Center of Academic Resources. "Internet traffic log 20011231 23:53 - 20020101 11:53." Traffic Traffic Log, CD TR 2002-001, File CU-TR-2002-0001.zip
535		Chulalongkorn University. bCenter of Academic Resources. bSystem Department
536		Original collection up to April 2002 supported by Rachadapiseksompoj Research Fund, Research Affairs, Chulalongkorn University
538		Data and metadata packaged and compressed by program zip v.2.1 on Solaris 2.5.1
540		bRelated privacy information must be removed or masked before publication of research result except by specific authorization from the owner
541		cContains the Internet traffic from router: Cisco 7206/IOS version: 11.1 (17.3) CA and (35) CC2 since 1 st September 2001; oChunk of 30-minute log data from router
567		30-minute chunk data stream
610	20	KMUTT
610	20	MUA
610	20	CUNET
650	7	202.28.0.125/25
650	7	202.28.18.35/27
650	7	202.28.0.253/25
650	0	Internet
650	7	Internet traffic log file
710	2	Chulalongkorn University. bCenter of Academic Resources. bSystem Department

เขตข้อมูล 099 ลงรายการหมายเลขของซีดี-อาร์

เขตข้อมูล 245 ลงรายการชื่อเรื่อง โดยกำหนดเป็น Internet traffic log (สำหรับข้อมูลแทรฟฟิก) และ กำหนดเป็น Web traffic log (สำหรับข้อมูลเว็บเพจ)

- เขตข้อมูล 260 ลงรายการโดยให้ผู้รับผิดชอบ คือฝ่ายระบบสารสนเทศ ส่วนปีคือ ปีที่บันทึกข้อมูลลงซีดี-อาร์ และเปิดปีไว้
- เขตข้อมูล 300 ลงรายการตามข้อความนี้ โดยไม่ต้องระบุจำนวนแผ่น
- เขตข้อมูล 310 ลงรายการโดยระบุปีที่เริ่มเก็บข้อมูลบันทึก
- เขตข้อมูล 362 ลงรายการโดยระบุปีที่เริ่มต้นของการบันทึกข้อมูล
- เขตข้อมูล 500 ลงรายการโดยขึ้นต้นด้วยคำว่า location ตามตัวอย่างและตามด้วยหมายเลขไอพีของเครื่องที่ส่ง และรับข้อมูล
- เขตข้อมูล 506 ลงรายการโดยคงข้อความตามตัวอย่าง
- เขตข้อมูล 508 ลงรายการโดยคงข้อความข้างต้นไว้ตามตัวอย่าง แต่ต้องระวังในเรื่องการเปลี่ยนข้อมูลในเรื่อง การบันทึกข้อมูลตั้งแต่ปีใด และใครเป็นผู้บันทึกในกรณีที่เปลี่ยนผู้รับผิดชอบ
- เขตข้อมูล 515 ลงรายการโดยระบุข้อมูลที่บันทึกครอบคลุมตั้งแต่เมื่อใด
- เขตข้อมูล 515 ลงรายการลักษณะของข้อมูลตามข้อความตัวอย่าง
- เขตข้อมูล 516 ลงรายการลักษณะของข้อมูลตามข้อความตัวอย่าง
- เขตข้อมูล 520 ลงรายการสรุปย่อของเนื้อหาที่บันทึก ให้ลงตามลักษณะนี้ แต่ให้ปรับเปลี่ยนเมื่อมีการเปลี่ยน ระบบการจัดเก็บ หรือ configuration
- เขตข้อมูล 535 ลงรายการโดยคงตามข้อความนี้
- เขตข้อมูล 536 ลงรายการในกรณีที่ขังได้รับทุนในการดำเนินงานนี้ต่อ ถ้าไม่มีให้ข้ามเขตข้อมูลนี้ไป
- เขตข้อมูล 538 ลงรายการเกี่ยวกับโปรแกรมการจัดเก็บและบีบอัดข้อมูล ให้ระบุโปรแกรมและรุ่นที่ใช้ใหม่ถ้ามีการเปลี่ยนแปลง
- เขตข้อมูล 540 ลงรายการตามข้อความในตัวอย่าง
- เขตข้อมูล 541 ลงรายการโดยระบุการเก็บข้อมูลจากเครื่องใด ด้วยโปรแกรมใด รวมทั้งการระบุจำนวนการตัด chunk ต่อ 30 นาที ถ้ามีการปรับเปลี่ยน configuration หรือการตัด chunk ให้ระบุในเขตข้อมูล นี้ด้วย
- เขตข้อมูล 576 ลงรายการจำนวนการตัด chunk ต่อ 30 นาที เมื่อเปลี่ยนในเขตข้อมูล 541 ให้เปลี่ยนในเขต ข้อมูลนี้ด้วย
- เขตข้อมูล 610 ลงรายการหน่วยงาน/องค์กรที่เชื่อมโยงอินเทอร์เน็ต
- เขตข้อมูล 650 ลงรายการหมายเลขไอพีของเครื่องที่เชื่อมโยง
- เขตข้อมูล 650 ลงรายการหัวเรื่องทั่วไปให้ลงตามรายการที่ให้ไว้ตามตัวอย่างหรือเพิ่มเติมแล้วแต่เหมาะสม
- เขตข้อมูล 710 ลงรายการตามตัวอย่าง

การลงรายการระเบียบบรรณานุกรมในระดับช่วงเวลา

การลงรายการระเบียบบรรณานุกรมในระดับช่วงเวลา ให้เปิดดูเพิ่มข้อมูลเมตาดาตาของซีดี-อาร์จะปรากฏดังตัวอย่าง

1. CD Label: CD TR 2000-001, 554.0 MB.
2. ISSN
3. Chunk
3.1 Chunk1: CU-TR-2000-0001.zip, Fri 25 Aug 2000 17:50 - Fri 25 Aug 2000 23:50, 46.0 MB.
3.2 Chunk2: CU-TR-2000-0002.zip, Fri 25 Aug 2000 23:50 - Sat 26 Aug 2000 23:51, 166.0 MB.
3.3 Chunk3: CU-TR-2000-0003.zip, Sat 26 Aug 2000 23:51 - Sun 27 Aug 2000 23:51, 153.0 MB.
3.4 Chunk4: CU-TR-2000-0004.zip, Sun 27 Aug 2000 23:51 - Mon 28 Aug 2000 23:51, 188.0 MB.
4. Chunk characteristics
4.1 Meta-data consists of data and equipment files that contain data and equipment configurations, Respectively.
4.2 Data and Meta-data packaged and compressed by program "zip 2.1" on Solaris 2.5.1.
5. Write date: Sat 23 Sep 2000.
6. Expected expiry date: Fri 23 Sep 2005.
7. Hardware writer configuration: Pentium(r) Processor, HP SureStore CD-Writer Internal 6020i.
8. Software writer configuration: Microsoft Windows98 Second Edition, Adaptec Easy CD Creator 4.02.

ในการลงรายการเลือกเฉพาะเขตข้อมูล 099, 245, 260, 300, 500, 518, 583, 6xx, 710 และ 773 หรือลงเฉพาะเขตข้อมูลที่มีเครื่องหมายดอกจัน (*) 2 อัน ที่กำกับท้ายเขตข้อมูลตามตารางข้างต้น โดยนำข้อมูลเมตาดาตาของซีดี-อาร์ทั้งหมด 8 ส่วน (จากตัวอย่างเป็นข้อมูลของซีดี-อาร์หมายเลข CD TR 2000-001 จะต้องทำระเบียบบรรณานุกรมในระดับช่วงเวลา 4 ระเบียบ) มาลงรายการดังตัวอย่าง

เขตข้อมูล	ตัวบ่งชี้	ลักษณะสมบัติ
099	2	CD TR 2000-001
245	00	Internet traffic log 20000825 17:50 – 20000825 23:50
260		c2000
300		1 log file (46.0 MB)
500		Log for 25 Aug 2000 17:50 - 25 Aug 2000 23:50
500		Usage through online access on appointment with staff
518		CD published on Sep 23, 2000
583		Expected expiry date; c20050923
650	0	Internet
650	7	Internet traffic log file
710	2	Chulalongkorn University. bCenter of Academic Resources. bSystem Department
773	0	tInternet Traffic Log, gCD TR 2000-001, File CU-TR-2000-0001.zip

- เขตข้อมูล 099 ลงรายการหมายเลขของซีดี-อาร์ (คู่มือในส่วนที่ 1)
- เขตข้อมูล 245 ลงรายการชื่อเรื่องโดยกำหนดเป็น Internet traffic log (สำหรับข้อมูลแทรฟฟิก) และกำหนดเป็น Web traffic log (สำหรับข้อมูลเว็บเพจ) ตามด้วยปี เดือน วัน และ เวลา ของวันที่เกิดข้อมูล - ปี เดือน วัน เวลา ของวันที่สิ้นสุดข้อมูล (คู่มือในส่วนที่ 3)
- เขตข้อมูล 260 ลงรายการเฉพาะปีที่บันทึก
- เขตข้อมูล 300 ลงรายการ โดยคงตามข้อความตามตัวอย่าง เปลี่ยนเฉพาะขนาดของข้อมูล (คู่มือในส่วนที่ 3 ในที่นี้ คือ 3.1)
- เขตข้อมูล 500 ลงรายการตามข้อความในตัวอย่าง เปลี่ยนเฉพาะช่วงเวลาที่เกิดข้อมูลและสิ้นสุดข้อมูลซึ่งสัมพันธ์กับเขตข้อมูล 245 (คู่มือในส่วนที่ 3 ในที่นี้ คือ 3.1)
- เขตข้อมูล 500 ลงรายการตามข้อความในตัวอย่าง คือ Usage through online access on appointment with staff
- เขตข้อมูล 518 ลงรายการตามข้อความในตัวอย่าง โดยให้ดูวัน เดือน ปีที่บันทึกข้อมูล (คู่มือในส่วนที่ 5)
- เขตข้อมูล 583 ลงรายการตามข้อความในตัวอย่าง (คู่มือในส่วนที่ 6)
- เขตข้อมูล 6xx ลงรายการตามข้อความในตัวอย่าง
- เขตข้อมูล 710 ลงรายการตามข้อความในตัวอย่าง
- เขตข้อมูล 773 ลงรายการตามข้อความในตัวอย่าง แต่เปลี่ยนเฉพาะเขตข้อมูลย่อย g ตามหมายเลขของซีดี-อาร์ และชื่อเพิ่มข้อมูลของ chunk (คู่มือในส่วนที่ 3 ในที่นี้ คือ 3.1)

เมื่อลงรายการระเบียบบรรณานุกรมในแต่ละช่วงเวลาตามตัวอย่างข้างต้นเสร็จแล้ว จะได้เพียง 1 ระเบียบเท่านั้น ต้องลงรายการระเบียบบรรณานุกรมจากข้อมูลในซีดี-อาร์ที่บันทึกตามตัวอย่างข้างต้นอีก 3 ระเบียบ โดยเปลี่ยนเฉพาะเขตข้อมูล 245, 300, 500 (Log for...), และ 773 เท่านั้น โดยดูรายการที่เปลี่ยนจากข้อ 3.2, 3.3 และ 3.4 ตามลำดับ

ภาคผนวก ค

เมตาดาตาข้อมูลดิจิทัลของโครงการของห้องสมุดทั้งห้า



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

เมตาดาตาข้อมูลดิจิทัลของโครงการของห้องสมุดทั้งห้า

เป็นส่วนหนึ่งของการวิจัย
เรื่อง

การกำหนดรูปแบบบรรณานุกรมและการวางแนวทางการจัดเก็บแบบถาวรสำหรับข้อมูลต่อเนื่องที่ตัดมาเป็นกลุ่ม
ให้สืบค้นผ่านห้องสมุดอัตโนมัติ : กรณีศึกษาข้อมูลสถิติการใช้งานอินเทอร์เน็ตของ

จุฬาลงกรณ์มหาวิทยาลัย

สนับสนุนโดย
กองทุนรัชดาภิเษกสมโภช

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สถาบันวิทยบริการ

และ

ห้องปฏิบัติการวิศวกรรมระบบสารสนเทศ ภาควิชาวิศวกรรมคอมพิวเตอร์

จุฬาลงกรณ์มหาวิทยาลัย

30 เมษายน 2545

เมตาดาตาข้อมูลดิจิทัลของโครงการของห้องสมุดทั้งห้า

ในวงการห้องสมุดได้มีการศึกษาวิจัยเกี่ยวกับการลงเมตาดาตาสำหรับการบำรุงรักษาข้อมูลดิจิทัลโดยหอสมุดแห่งชาติ และโครงการศึกษาอีกหลายโครงการ ซึ่งแต่ละโครงการประกอบด้วยเมตาดาตา ดังนี้

1. PANDORA ได้พัฒนาข้อกำหนดภายใต้ชื่อ "Preservation Metadata for Digital Collections" โดยพิมพ์ฉบับร่างออกมาในเดือนตุลาคม ค.ศ. 1999 ประกอบด้วย 25 องค์ประกอบ ได้แก่

1. Persistent identifier	13. Relationships
2. Date of creation	14. Quirks
3. Structure type	15. Archiving decision (work)
4. Technical infrastructure of complex objects	16. Decision reason (work)
5. File description	17. Institution responsible for archiving (work)
6. Know system requirements	18. Archiving decision (manifestation)
7. Installation requirements	19. Decision reason (manifestation)
8. Storage information	20. Institution responsible for archiving
9. Access inhibitors	21. Intention type
10. Finding and searching aids, and accession facilitators	22. Institution with preservation responsible
11. Preservation action permission	23. Process
12. Validation	24. Record creator
	25. Other

2. The Cedars Project กำหนดเมตาดาตาฉบับร่างโดยมีโครงสร้างจากโครงการของ OAIS ประกอบด้วย

1. Information package	26. Rights information
2. Preservation description	27. Copyright statement
3. Reference information	28. Name of publisher
4. Resource description	29. Date of publication
5. Existing metadata	30. Place of publication
6. Existing record	31. Rights warning
7. Context information	32. Contacts of rights holders
8. Related information object	33. Actors
9. Provenance information	34. Actions
10. History of origin	35. Permitted by statute
11. Reason for creation	36. Permitted by licence
12. Custody history	37. Fixity information
13. Change history before archiving	38. Authentication indicator
14. Original technical environments	39. Content information
15. Prerequisites	40. Representation information
16. Procedures	41. Structure information
17. Documentation	42. Underlying abstract form description
18. Reason for preservation	43. Transformer objects
19. Management history	44. Platform
20. Ingest process history	45. Render/analyse engines
21. Administration history	46. Output format
22. Action history	47. Input format
23. Policy history	48. Render/analyse/convert objects
24. Rights management	49. Semantic information
25. Negotiation history	50. Render/analysis objects
	51. Primary digital objects

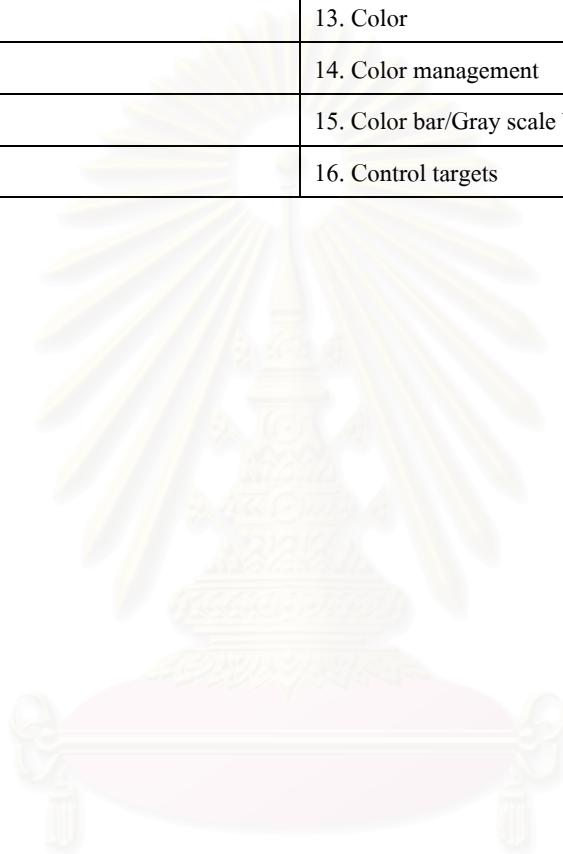
3. The NEDLIB Project ได้พัฒนาระบบการจัดเก็บสิ่งพิมพ์อิเล็กทรอนิกส์ หรือ Deposit System for Electronic Publications (DSEP) ซึ่งมีพื้นฐานมาจาก OAIS เช่นเดียวกับโครงการของ Cedars ประกอบด้วย

1. Metadata for representation information	2. Metadata for preservation and description information
1.1 Specific hardware requirements	2.1 Reference information
1.2 Specific microprocessor requirements	2.2 Assigned identifier
1.3 Specific multimedia requirements	2.3 Fixity information
1.4 Specific peripheral requirements	2.4 Checksum
1.5 Operating system	2.5 Change history
1.6 Interpreter and compiler	2.6 Main metadata concerned
1.7 Object format	2.7 Tool
1.8 Application	2.8 Reverse
	2.9 Other metadata concerned

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

4. The OCLC/RLG Working Group on Preservation Metadata ศึกษาเฉพาะเพิ่มข้อมูลภาพ
ดิจิทัล โดยกำหนดองค์ประกอบของเมตาดาตา 16 องค์ประกอบ ดังนี้

1. Date	9. Watermark
2. Transcriber	10. Resolution
3. Producer	11. Compression
4. Capture device	12. Source
5. Capture details	13. Color
6. Change history	14. Color management
7. Validation key	15. Color bar/Gray scale bar
8. Encryption	16. Control targets



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

5. Library of Congress Digital Repository Development ประกอบด้วยเมตาดาตาเชิงพรรณนาเชิงบริหาร และเชิงโครงสร้าง สำหรับการอธิบายถึงแกนหลักขององค์ประกอบของเมตาดาตาที่ต้องใช้ในการพัฒนา การทดสอบ และการนำไปใช้สำหรับการจัดเก็บที่หลากหลาย รวม 76 องค์ประกอบ ได้แก่

1. Access_category	26. Deposit_date_time	52.Original_content_type
2. Access_display_message	27. Description_content_list	53. Parent_object_ID
3. Access_expiration_date	28. Description_coverage	54. Preservation_profile
4. Access_information	29. Description_creation	55.Preservation_information
5. Access_rights	30. Description_subject	56. Preservation_master_ID
6. Archive_date_time	31. Description_summary	57. Preservation_original_ID
7. Archive_history	32. Description_title	58. Quantity_of_intermediate_objects
8. Archive_ID	33. Dimension_horizontal	59. Quantity_of_terminal_objects
9. Archive_next_date_time	34. Dimension_vertical	60. Reformatted_original_objects
10. Archiving_profile	35. Duration	61. Reformatting_guidelines
11. Associated_file_name	36. External_descriptive_information	62. Reformatting_information
12. Associated_file_type	37. External_migration_ID	63. Reformatting_method
13. Audio_bits_per_sample	38. External_record_ID	64. Relationship_type
14. Audio_channel_configuration	39. External_record_type	65. Relationship_value
15. Audio_channel_information	40. External_reproduction_ID	66. Responsibility_entity
16. Audio_sampling_frequency	41. External_reproduction_procedure	
17. Capture_device_ID	42. Feature_label	67. Responsibility_information
18. Capture_device_settings	43. File_extension	68. Revision_date_time
19. Capture_entity_corporate	44. Handle	69. Segment_type
20. Capture_entity_individual	45. Image_bit_depth	70. Segment_value
21. Capture_production_ID	46. Image_color_space	71. Serial_part
22. Checksum_creation_date_time	47. Image_orientation	72. Serial_relationship
23. Checksum_value	48. Image_resolution	73. Size
24. Creation_date_time	49. Intermediate_object_ID	74. Use
25. Datastream_compression	50. Intermediate_object_use	75. Video_data_rate
	51. Internal_media_type	76. Video_frame_rate

ภาคผนวก ง
คู่มือการปฏิบัติงานการบันทึกข้อมูลลงซีดี-อาร์



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

คู่มือการปฏิบัติงานการบันทึกข้อมูลลงซีดี-อาร์

เป็นส่วนหนึ่งของการวิจัย
เรื่อง

การกำหนดรูปแบบบรรณานุกรมและการวางแนวทางการจัดเก็บแบบถาวรสำหรับข้อมูลต่อเนื่องที่ตัดมาเป็นกลุ่ม
ให้สืบค้นผ่านห้องสมุดอัตโนมัติ : กรณีศึกษาข้อมูลสถิติการใช้งานอินเทอร์เน็ตของ
จุฬาลงกรณ์มหาวิทยาลัย

สนับสนุนโดย
กองทุนรัชดาภิเษกสมโภช

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สถาบันวิทยบริการ

และ

ห้องปฏิบัติการวิศวกรรมระบบสารสนเทศ ภาควิชาวิศวกรรมคอมพิวเตอร์

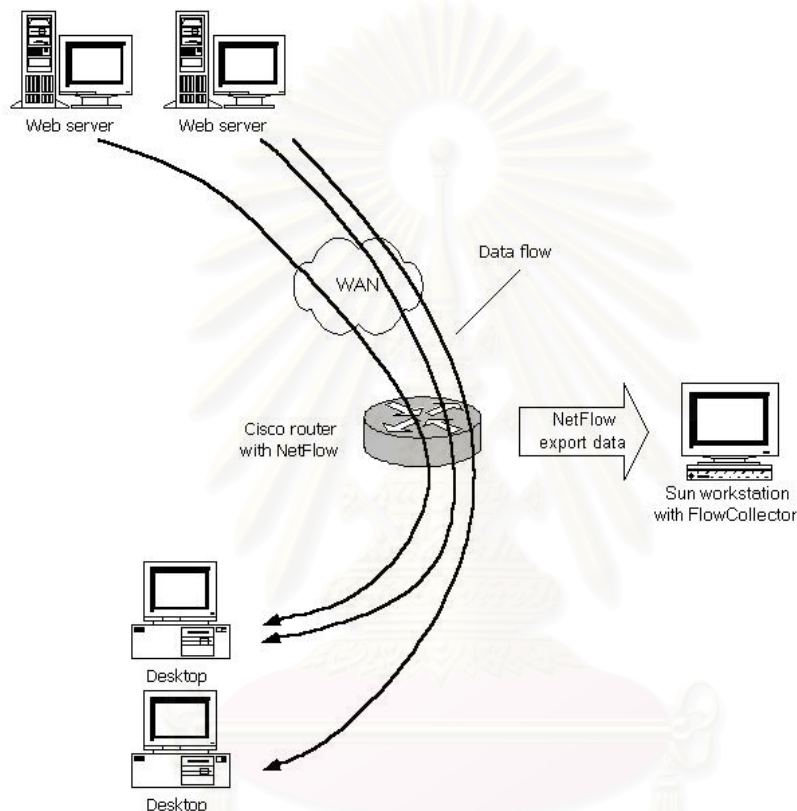
จุฬาลงกรณ์มหาวิทยาลัย

30 เมษายน 2545

การ Publish Internet traffic CD

การทำงานของ NetFlow

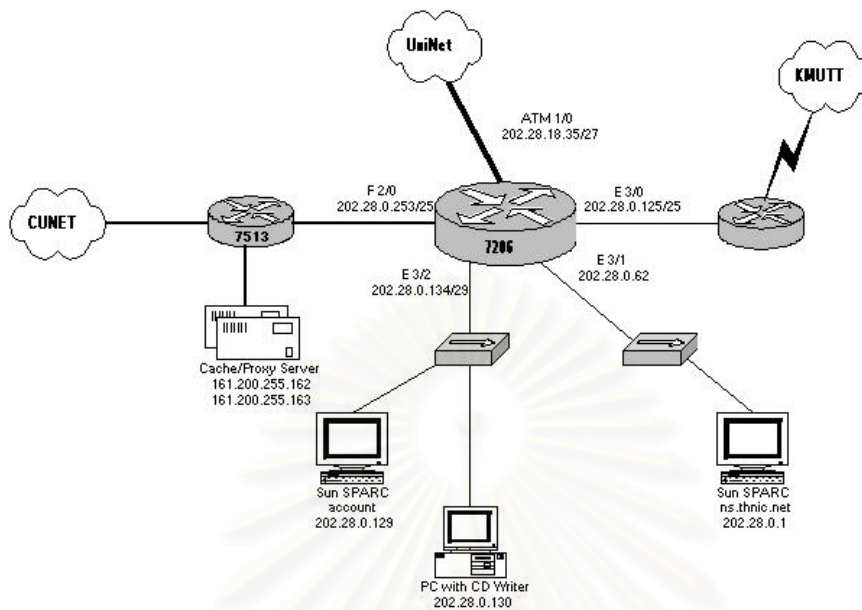
NetFlow เป็นคุณสมบัติที่มีอยู่ใน router และ switch ของ Cisco โดยจะทำการเก็บข้อมูลส่วนหัวของ packet ที่ผ่านเข้า/ออก router หรือ switch ซึ่งข้อมูลนี้สามารถนำไปใช้ทำสถิติเพื่อวิเคราะห์การใช้งานระบบเครือข่ายได้



รูปที่ 1 แสดงการทำงานของ NetFlow

จากรูปแสดงการทำงานของ NetFlow โดย router หรือ switch จะทำการ export data flow ให้แก่โปรแกรม FlowCollector ซึ่งทำงานอยู่บนเครื่อง Sun workstation เพื่อทำการสรุปข้อมูลตามแบบที่กำหนดไว้

ขั้นตอนการ Publish Internet traffic CD



รูปที่ 2 แสดงโครงสร้างระบบเครือข่ายในการเก็บข้อมูล
NetFlow ของจุฬาลงกรณ์มหาวิทยาลัย

จากรูปแสดงโครงสร้างของระบบเครือข่ายที่ใช้ในการเก็บข้อมูล NetFlow ของจุฬาลงกรณ์มหาวิทยาลัย โดย Cisco 7206 router ซึ่งเป็น Internet gateway ของจุฬาลงกรณ์มหาวิทยาลัยจะทำการ export data flow ไปยังเครื่อง account ซึ่งมีโปรแกรม FlowCollector ทำงานอยู่ โดยจะทำการสรุปข้อมูลแบบ CallRecord ทุก ๆ ครั้ง ชั่วโมง รูปที่ 3 แสดงข้อมูลที่ได้จากการโปรแกรม FlowCollector โดยข้อมูลจะถูกเก็บอยู่ที่ directory /opt/CSCOnfc/Data/<วันที่>/<ชื่อ หรือ ip ของ router หรือ switch ที่ export ข้อมูล>/รูปแบบการสรุปข้อมูล> และชื่อไฟล์จะอยู่ในรูปแบบ <ชื่อหรือ ip ของ router หรือ switch ที่ export ข้อมูล>_ปี_เดือน_วัน.ชั่วโมงนาที

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

```

account - SecureCRT
File Edit View Options Transfer Script Window Help
[initass@account /opt/CSC0nFc]*
[initass@account /opt/CSC0nFc]*ls -l
total 22
lrwxrwxrwx 1 root other 10 Jun 18 2001 Data -> /usr2/data
drwxr-xr-x 3 nitass bin 512 Jul 19 2000 bin
drwxr-xr-x 2 nitass bin 512 Jul 21 2000 config
drwxr-xr-x 2 nitass bin 512 Nov 9 1998 doc
drwxr-xr-x 2 nitass bin 4608 Apr 18 00:04 logs
drwxr-xr-x 2 nitass bin 512 Nov 9 1998 tmp
drwxr-xr-x 2 nitass bin 512 Nov 9 1998 tools
[initass@account /opt/CSC0nFc]*cd Data
[initass@account /opt/CSC0nFc/Data]*ls -l
total 36
drwxr-xr-x 3 nitass root 512 Apr 1 00:28 2002_04_01
drwxr-xr-x 3 nitass root 512 Apr 2 00:28 2002_04_02
drwxr-xr-x 3 nitass root 512 Apr 3 00:28 2002_04_03
drwxr-xr-x 3 nitass root 512 Apr 4 00:29 2002_04_04
drwxr-xr-x 3 nitass root 512 Apr 5 00:29 2002_04_05
drwxr-xr-x 3 nitass root 512 Apr 6 00:29 2002_04_06
drwxr-xr-x 3 nitass root 512 Apr 7 00:00 2002_04_07
drwxr-xr-x 3 nitass root 512 Apr 8 00:00 2002_04_08
drwxr-xr-x 3 nitass root 512 Apr 9 00:00 2002_04_09
drwxr-xr-x 3 nitass root 512 Apr 10 00:01 2002_04_10
drwxr-xr-x 3 nitass root 512 Apr 11 00:01 2002_04_11
drwxr-xr-x 3 nitass root 512 Apr 12 00:01 2002_04_12
drwxr-xr-x 3 nitass root 512 Apr 13 00:02 2002_04_13
drwxr-xr-x 3 nitass root 512 Apr 14 00:02 2002_04_14
drwxr-xr-x 3 nitass root 512 Apr 15 00:03 2002_04_15
drwxr-xr-x 3 nitass root 512 Apr 16 00:03 2002_04_16
drwxr-xr-x 3 nitass root 512 Apr 17 00:03 2002_04_17
drwxr-xr-x 3 nitass root 512 Apr 18 00:04 2002_04_18
[initass@account /opt/CSC0nFc/Data]*
Ready Telnet 32, 36 38 Rows, 91 Cols VT100 NUM

```

```

account - SecureCRT
File Edit View Options Transfer Script Window Help
[initass@account /opt/CSC0nFc/Data]*
[initass@account /opt/CSC0nFc/Data]*cd 2002_04_17/202.28.0.134/CallRecord
[initass@account /opt/CSC0nFc/Data/2002_04_17/202.28.0.134/CallRecord]*ls
202.28.0.134_2002_04_17.0003.gz 202.28.0.134_2002_04_17.1203.gz
202.28.0.134_2002_04_17.0033.gz 202.28.0.134_2002_04_17.1233.gz
202.28.0.134_2002_04_17.0103.gz 202.28.0.134_2002_04_17.1303.gz
202.28.0.134_2002_04_17.0133.gz 202.28.0.134_2002_04_17.1333.gz
202.28.0.134_2002_04_17.0203.gz 202.28.0.134_2002_04_17.1403.gz
202.28.0.134_2002_04_17.0233.gz 202.28.0.134_2002_04_17.1433.gz
202.28.0.134_2002_04_17.0303.gz 202.28.0.134_2002_04_17.1503.gz
202.28.0.134_2002_04_17.0333.gz 202.28.0.134_2002_04_17.1533.gz
202.28.0.134_2002_04_17.0403.gz 202.28.0.134_2002_04_17.1603.gz
202.28.0.134_2002_04_17.0433.gz 202.28.0.134_2002_04_17.1633.gz
202.28.0.134_2002_04_17.0503.gz 202.28.0.134_2002_04_17.1703.gz
202.28.0.134_2002_04_17.0533.gz 202.28.0.134_2002_04_17.1733.gz
202.28.0.134_2002_04_17.0603.gz 202.28.0.134_2002_04_17.1803.gz
202.28.0.134_2002_04_17.0633.gz 202.28.0.134_2002_04_17.1833.gz
202.28.0.134_2002_04_17.0703.gz 202.28.0.134_2002_04_17.1903.gz
202.28.0.134_2002_04_17.0733.gz 202.28.0.134_2002_04_17.1933.gz
202.28.0.134_2002_04_17.0803.gz 202.28.0.134_2002_04_17.2003.gz
202.28.0.134_2002_04_17.0833.gz 202.28.0.134_2002_04_17.2034.gz
202.28.0.134_2002_04_17.0903.gz 202.28.0.134_2002_04_17.2104.gz
202.28.0.134_2002_04_17.0933.gz 202.28.0.134_2002_04_17.2134.gz
202.28.0.134_2002_04_17.1003.gz 202.28.0.134_2002_04_17.2204.gz
202.28.0.134_2002_04_17.1033.gz 202.28.0.134_2002_04_17.2234.gz
202.28.0.134_2002_04_17.1103.gz 202.28.0.134_2002_04_17.2304.gz
202.28.0.134_2002_04_17.1133.gz 202.28.0.134_2002_04_17.2334.gz
[initass@account /opt/CSC0nFc/Data/2002_04_17/202.28.0.134/CallRecord]*
Ready Telnet 28, 71 38 Rows, 91 Cols VT100 NUM

```

รูปที่ 3 แสดงข้อมูลที่ได้จากการสรุปข้อมูลแบบ CallRecord ของโปรแกรม FlowCollector

จากนั้นจะทำการสร้าง chunk โดยนำข้อมูลที่ได้มาต่อกัน โดยชื่อไฟล์ chunk จะอยู่ในรูปแบบ CU-TR-<ปีของข้อมูล>-<หมายเลข 4 หลักแสดงลำดับการสร้าง chunk> เสร็จแล้วจะนำมา compress รวมกับ meta data ของ chunk ซึ่งประกอบด้วยไฟล์ data.txt และ equipment.txt เป็นรายละเอียดของข้อมูล และอุปกรณ์ที่ export ข้อมูลของ chunk นั้น ๆ ตามลำดับ โดยใช้ชื่อไฟล์เดียวกับชื่อ chunk รูปที่ 4 แสดงการสร้าง chunk สำหรับจำนวนข้อมูลที่จะนำมาต่อกันเป็น chunk นั้นจะขึ้นกับความสะดวกในการทำงาน และขนาดของไฟล์ ในงานวิจัยนี้กำหนดให้จำนวนข้อมูลที่จะนำมาต่อกันเป็นหลักของวัน และขนาดของไฟล์ที่ผ่านการ compress แล้วมีขนาดไม่เกิน 200 MB

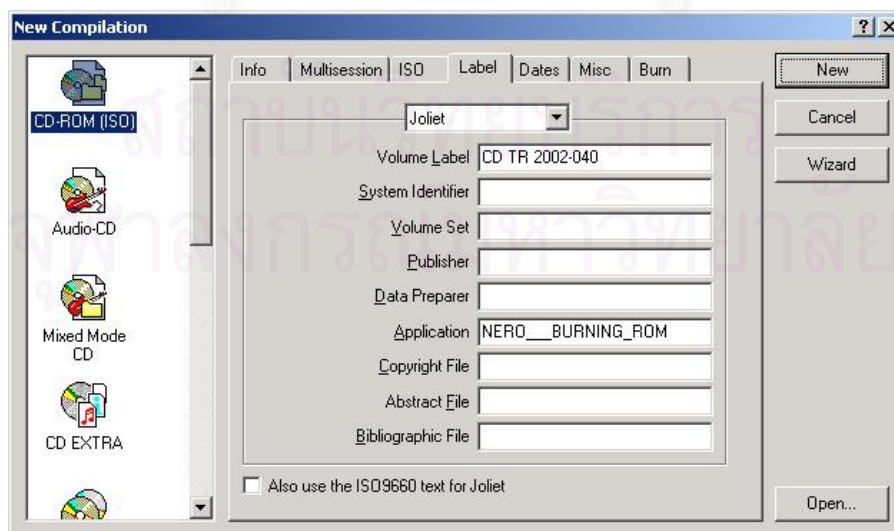
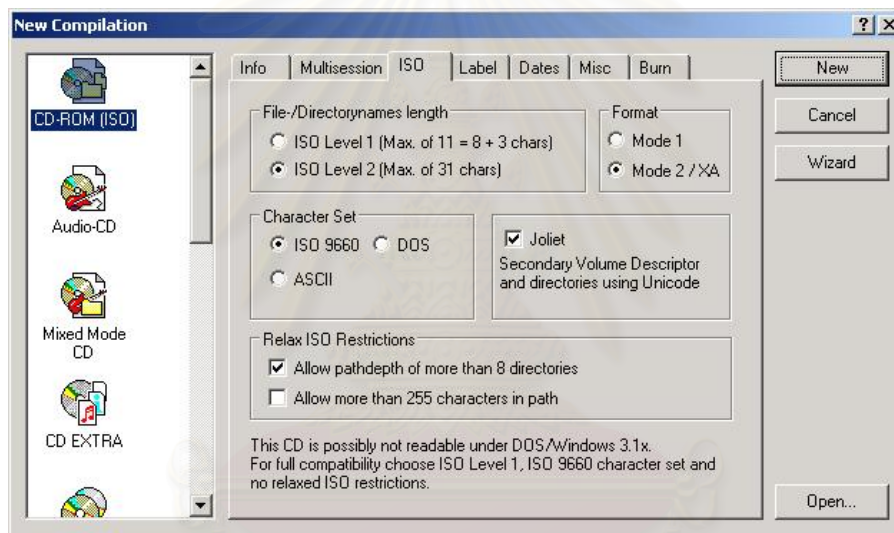
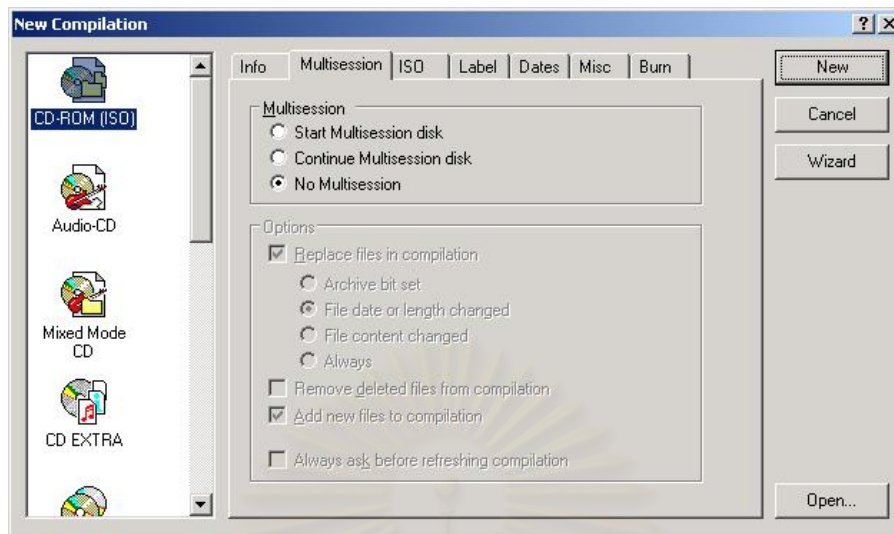
```

account - SecureCRT
File Edit View Options Transfer Script Window Help
[initass@account /usr2/chunk]*
[initass@account /usr2/chunk]*ls -l
total 503046
-rw-r--r--  1 nitass  staff   88308853 Apr 18 15:47 CU-TR-2002-0213.zip
-rw-r--r--  1 nitass  staff  169084724 Apr 18 16:14 CU-TR-2002-0214.zip
-rw-r--r--  1 nitass  staff   1555 Apr 18 16:23 data.txt
-rw-r--r--  1 nitass  staff   492 Apr 18 16:23 equipment.txt
[initass@account /usr2/chunk]*
[initass@account /usr2/chunk]*zip -u CU-TR-2002-0213.zip data.txt equipment.txt
adding: data.txt (deflated 53%)
adding: equipment.txt (deflated 37%)
[initass@account /usr2/chunk]*
[initass@account /usr2/chunk]*unzip -l CU-TR-2002-0213.zip
Archive:  CU-TR-2002-0213.zip
Length  Date       Time      Name
-----  -
425013119  04-18-02  15:28    CU-TR-2002-0213
1555  04-18-02  16:23    data.txt
492  04-18-02  16:23    equipment.txt
-----
425015166                               3 files
[initass@account /usr2/chunk]*

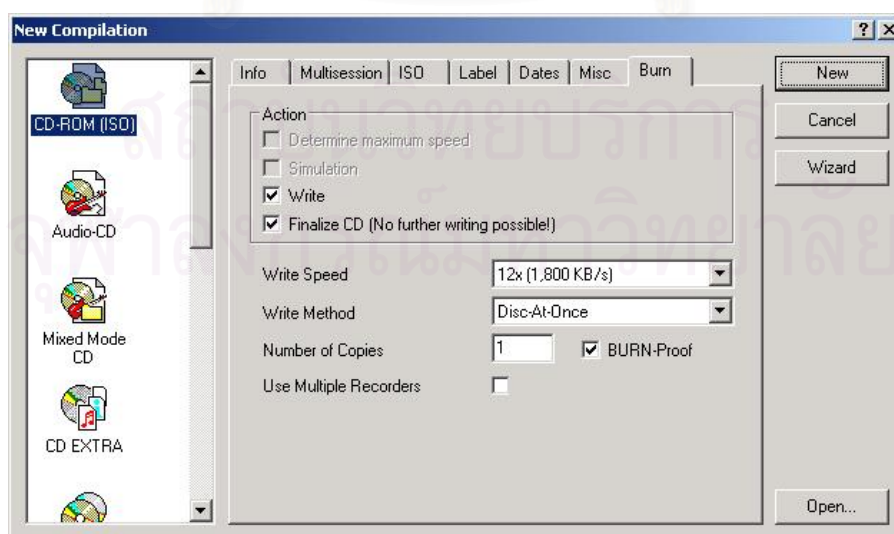
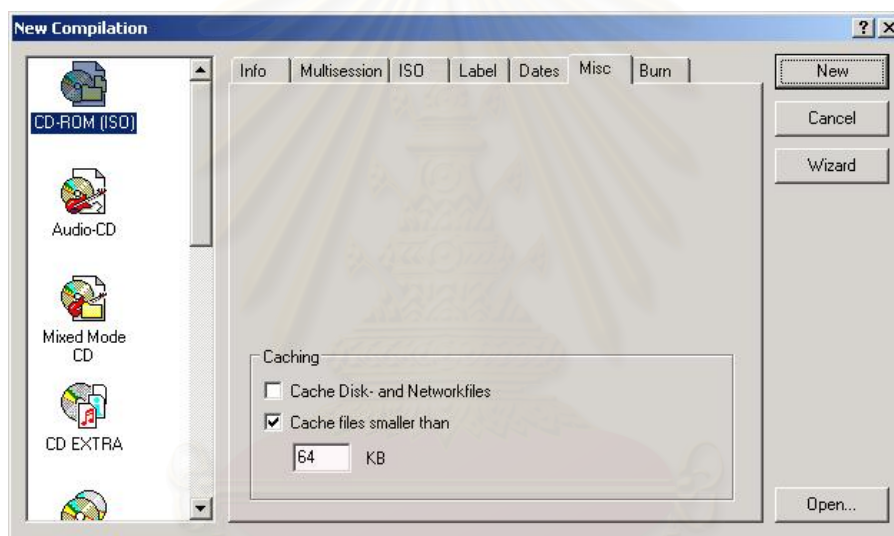
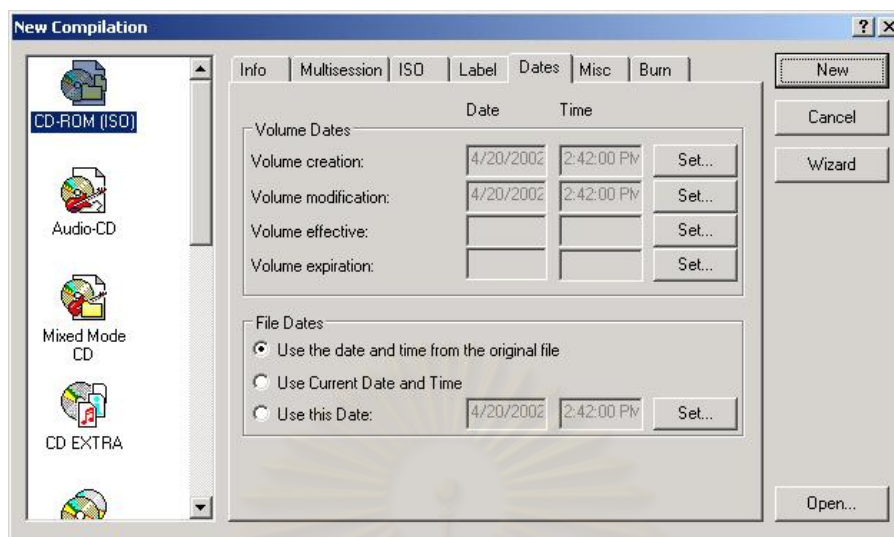
```

รูปที่ 4 แสดงการสร้าง chunk

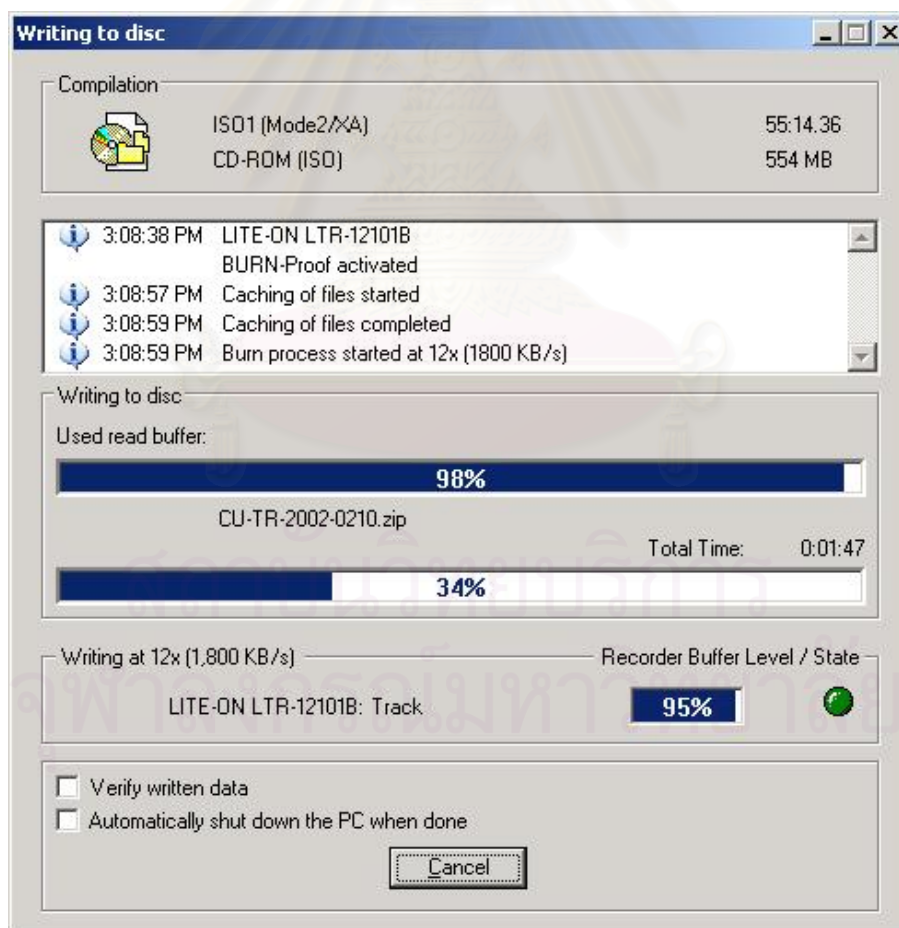
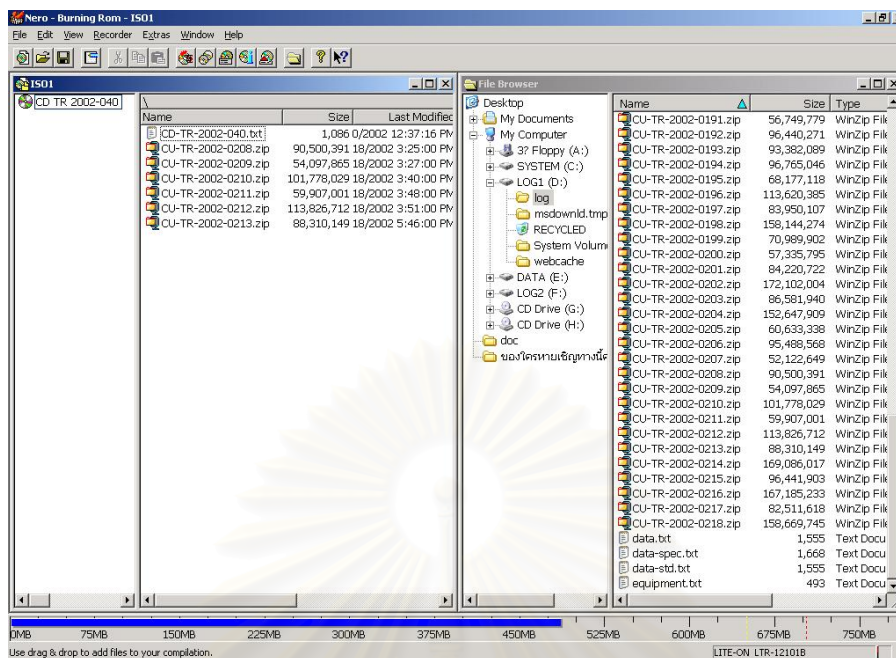
ต่อจากนั้นจะทำการ FTP ไฟล์ chunk จากเครื่อง account ไปยังเครื่อง PC ที่มีเครื่องบันทึก CD เพื่อบันทึกลง CD พร้อมกับ meta data ของ CD ซึ่งเป็นรายละเอียดของข้อมูลที่อยู่ในแผ่น CD นั้น ๆ ต่อไป โดยชื่อแผ่น CD จะอยู่ในรูปแบบ CD-TR-<ปีของข้อมูล>-<หมายเลข 3 หลักแสดงลำดับการบันทึก CD> รูปที่ 5 แสดงการบันทึก chunk ลงแผ่น CD



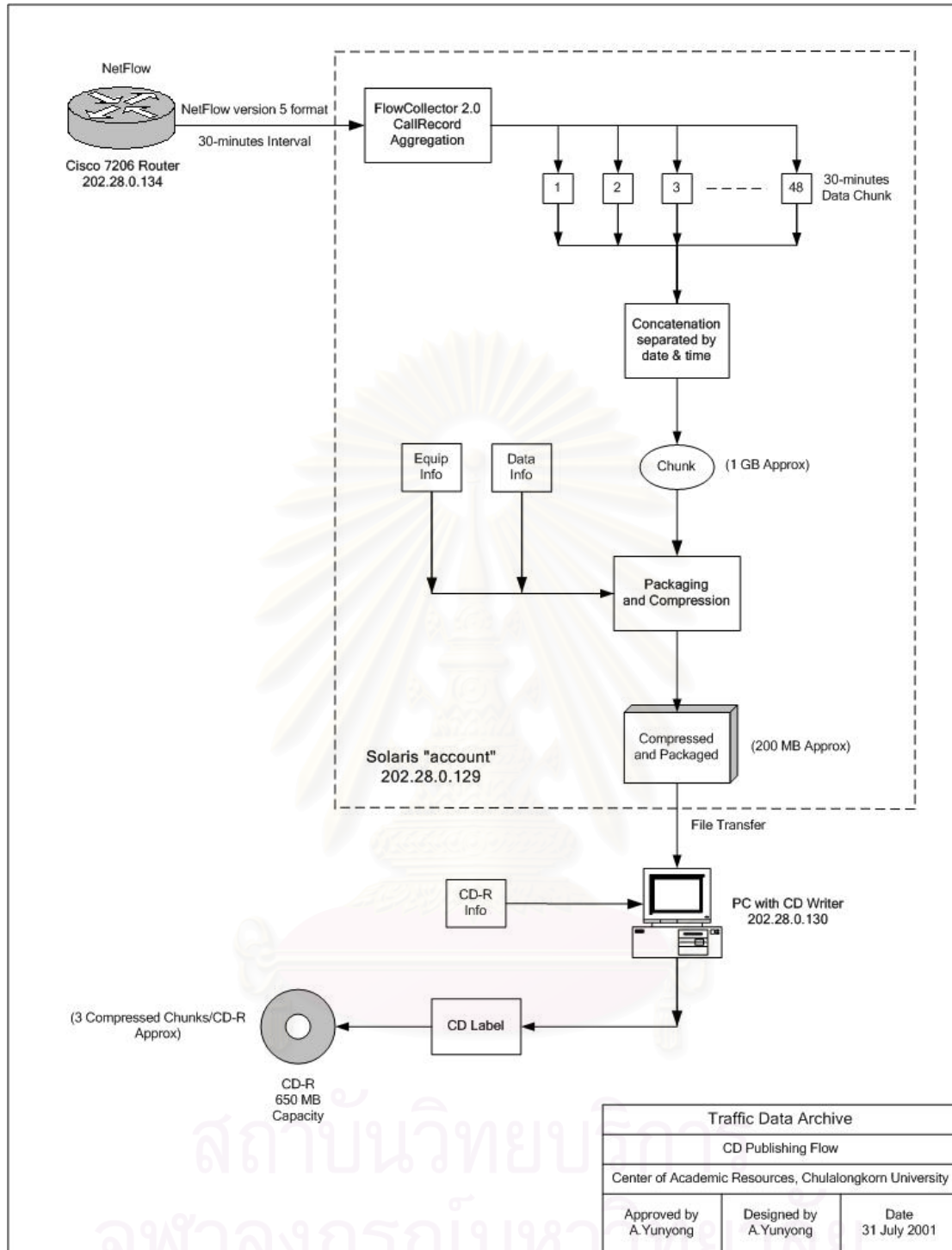
รูปที่ 5 แสดงการบันทึก chunk ลงแผ่นซีดี-อาร์



รูปที่ 5 แสดงการบันทึก chunk ลงแผ่นซีดี-อาร์ (ต่อ)



รูปที่ 5 แสดงการบันทึก chunk ลงแผ่นซีดี-อาร์ (ต่อ)



รูปที่ 6 ขั้นตอนการผลิตซีดี-อาร์ของบันทึกถาวร

ตัวอย่างไฟล์ *data.txt* เป็น *meta data* ของข้อมูลใน *chunk*

1. File ID : data.txt for Chunk CU-TR-2002-0213
2. Period : 12 hours
3. Start : Tue 16 Apr 2002 23:33
4. End : Wed 17 Apr 2002 11:33
5. Used by : CUNET
6. Generated by : account.car.chula.ac.th
7. Type : Aggregate data from program netflowcollector (version 2)

8. Data Dictionary of file. This file consists of 24 groups of data. Each group represents 30 minutes of data and its structure is as follows:

8.1 Header of flow aggregation.

In each header, there are 9 fields separated by "|".

- Field description

- a. Source router/switch
- b. Type of format
- c. Type of aggregation
- d. Period of aggregation
- e. Start time (UTC time)
- f. End time (UTC time)
- g. Total flow
- h. Missed flow
- i. Total number of records

8.2 Records of flow.

In each line, there are 12 fields separated by "|".

- Field description

- a. Srcaddr : Source IP address
- b. Dstaddr : Destination IP address
- c. Srcport : TCP/UDP source port number
- d. Dstaddr : TCP/UDP destination port number
- e. Protocol : IP protocol type (example TCP=6, UDP=17)
- f. ToS : IP type of service
- g. Packet count : Packet count as part of this record
- h. Byte count : Total number of Layer 3 bytes counted as part of this record
- i. Flow count : Total number of flows aggregated into this record
- j. First Time Stamp : The time, in UTC seconds, of the first packet summarized into this record.
- k. Last Time Stamp : The time, in UTC seconds, of the last packet summarized into this record.

1. Total Active Time : The sum of individual active time for all the flows summarized into the current record.



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตัวอย่างไฟล์ *equipment.txt* เป็น *meta data* ของอุปกรณ์ที่เป็น *export* ข้อมูลใน *chunk*

1. File ID: equipment.txt for Chunk CU-TR-2002-0213
2. Type : Router
3. Model : Cisco 7206 / IOS version: 11.1(35)CC2
4. Description : UniNet gateway for CUNET, southern and northern hubs
5. Location : IP address 202.28.0.125/25 - 10 Mbps connecting to KMUTT's router
: IP address 202.28.18.35/27 - 155 Mbps connecting to MUA ATM backbone
: IP address 202.28.0.253/25 - 100 Mbps connecting to CUNET
6. Reference : export data in Netflow version 5 format
: ntp server 161.200.192.13



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตัวอย่างไฟล์ *CD-TR-2002-040.txt* เป็น *meta data* ของข้อมูลในแผ่น CD

1. CD Label: CD TR 2002-040, 484.0 MB.

2. ISSN

3. Chunk

3.1 Chunk1: CU-TR-2002-0208.zip, Sun 14 Apr 2002 11:32 - Sun 14 Apr 2002 23:32, 86.3 MB.

3.2 Chunk2: CU-TR-2002-0209.zip, Sun 14 Apr 2002 23:32 - Mon 15 Apr 2002 11:33, 51.5 MB.

3.3 Chunk3: CU-TR-2002-0210.zip, Mon 15 Apr 2002 11:33 - Mon 15 Apr 2002 23:33, 97.0 MB.

3.4 Chunk4: CU-TR-2002-0211.zip, Mon 15 Apr 2002 23:33 - Tue 16 Apr 2002 11:33, 57.1 MB.

3.5 Chunk5: CU-TR-2002-0212.zip, Tue 16 Apr 2002 11:33 - Tue 16 Apr 2002 23:33, 108.0 MB.

3.6 Chunk6: CU-TR-2002-0213.zip, Tue 16 Apr 2002 23:33 - Wed 17 Apr 2002 11:33, 84.2 MB.

4. Chunk characteristics

4.1 Meta-data consists of data and equipment files that contain data and equipment configurations, respectively.

4.2 Data and Meta-data packaged and compressed by program "zip 2.1" on Solaris 2.5.1.

5. Write date: Sat 20 Apr 2002.

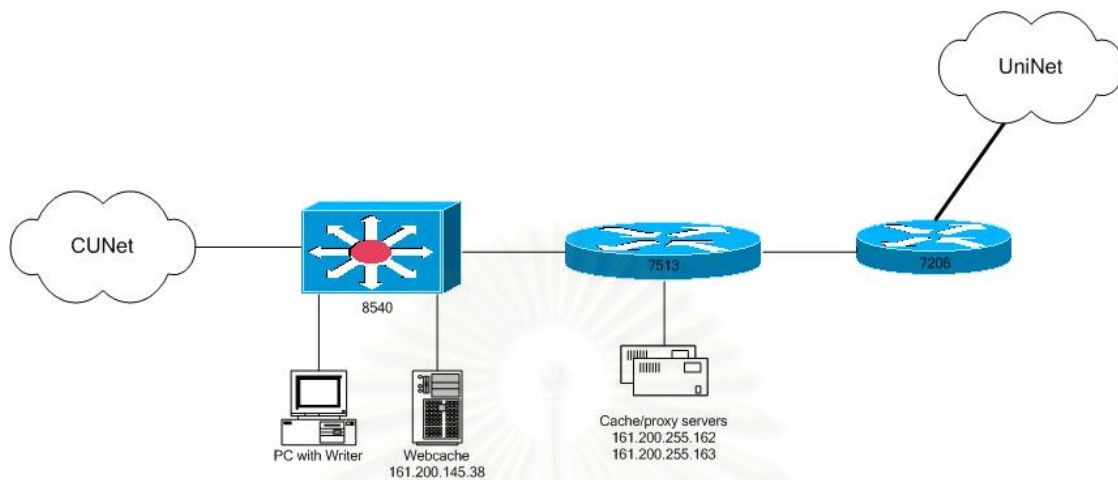
6. Expected expiry date: Fri 20 Apr 2007.

7. Hardware writer configuration: Celeron(r) Processor, CD-ReWritable LiteOn LTR-12101B.

8. Software writer configuration: Microsoft WindowsXP Professional, Nero-Burning Rom 5.5.5.1.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

การ Publish web cache traffic CD

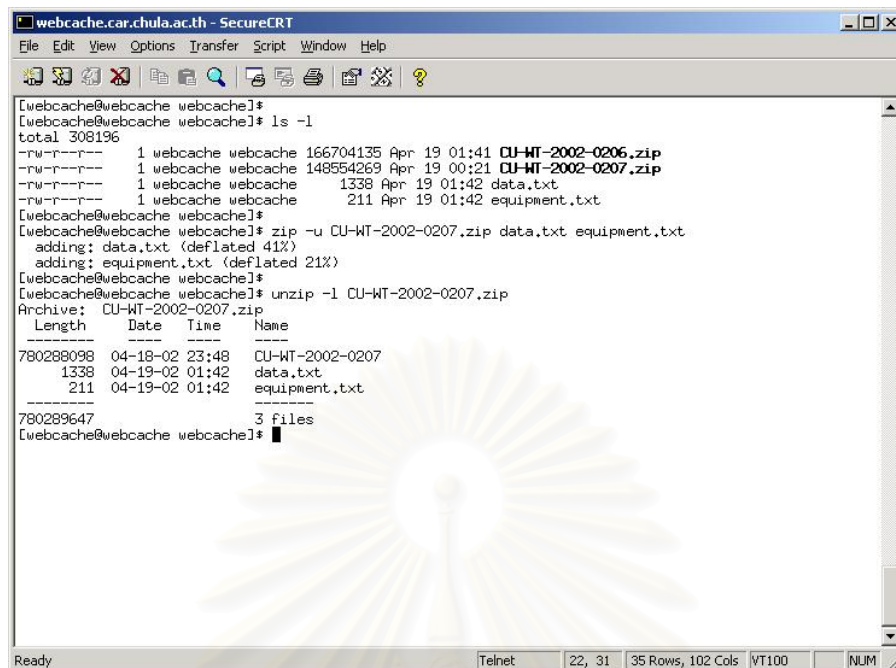


รูปที่ 1 แสดงโครงสร้างระบบเครือข่ายในการเก็บข้อมูล

web cache traffic ของจุฬาลงกรณ์มหาวิทยาลัย

จากรูปที่ 1 cache/proxy servers จะทำหน้าที่เป็น transparent proxy ของระบบเครือข่าย โดยทุก ๆ http หรือ ftp packet จะถูกทำการ cache ที่ cache/proxy servers ก่อนจะส่งไปให้ client ดังนั้น log ของ cache/proxy servers จะเก็บข้อมูลของ web ทั้งหมดที่ถูกเรียกจากเครื่องคอมพิวเตอร์ที่อยู่ภายในเครือข่ายจุฬาลงกรณ์มหาวิทยาลัย ข้อมูล log เหล่านี้จะถูก FTP ไปยังเครื่อง webcache วันละ 1 ครั้ง โดยแต่ละ log file จะเก็บข้อมูลเป็นระยะเวลา 6 ชั่วโมง

จากนั้นจะทำการสร้าง chunk โดยนำ log file ที่ได้มาต่อกัน โดยชื่อไฟล์ chunk จะอยู่ในรูปแบบ CU-WT-<ปีของข้อมูล>-<หมายเลข 4 หลักแสดงลำดับการสร้าง chunk> เสร็จแล้วจะนำมา compress รวมกับ meta data ของ chunk ซึ่งประกอบด้วยไฟล์ data.txt และ equipment.txt เป็นรายละเอียดของข้อมูล และอุปกรณ์ที่เป็น source ของข้อมูลใน chunk นั้น ๆ ตามลำดับ โดยใช้ชื่อไฟล์เดียวกับชื่อ chunk รูปที่ 2 แสดงการสร้าง chunk สำหรับจำนวนข้อมูลที่นำมาต่อกันเป็น chunk นั้นจะขึ้นกับความสะดวกในการทำงาน และขนาดของไฟล์ ในงานวิจัยนี้กำหนดให้จำนวนข้อมูลที่นำมาต่อกันเป็นหลักของวัน และขนาดของไฟล์ที่ผ่านการ compress แล้วไม่เกิน 200 MB



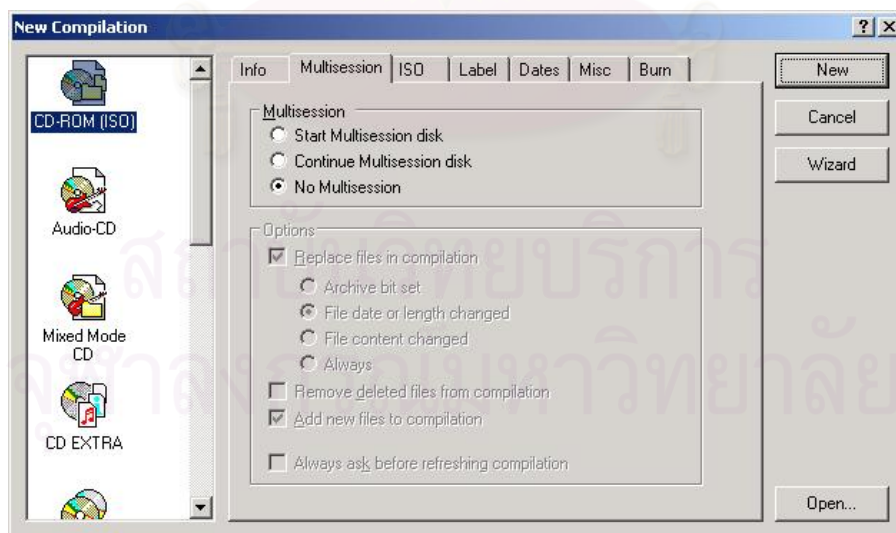
```

[webcache@webcache webcache]$
[webcache@webcache webcache]$ ls -l
total 308196
-rw-r--r--  1 webcache webcache 166704135 Apr 19 01:41 CU-WT-2002-0206.zip
-rw-r--r--  1 webcache webcache 148554269 Apr 19 00:21 CU-WT-2002-0207.zip
-rw-r--r--  1 webcache webcache    1338 Apr 19 01:42 data.txt
-rw-r--r--  1 webcache webcache    211 Apr 19 01:42 equipment.txt
[webcache@webcache webcache]$
[webcache@webcache webcache]$ zip -u CU-WT-2002-0207.zip data.txt equipment.txt
  adding: data.txt (deflated 41%)
  adding: equipment.txt (deflated 21%)
[webcache@webcache webcache]$
[webcache@webcache webcache]$ unzip -l CU-WT-2002-0207.zip
Archive:  CU-WT-2002-0207.zip
  Length   Date   Time    Name
  -----  -
780288098  04-18-02 23:48  CU-WT-2002-0207
 1338      04-19-02 01:42  data.txt
  211      04-19-02 01:42  equipment.txt
  -----  -
780289647                   3 files
[webcache@webcache webcache]$ █

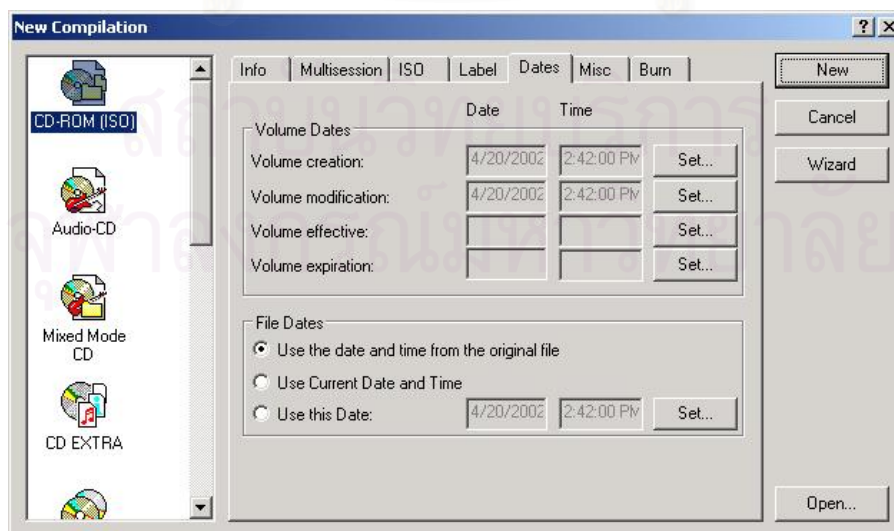
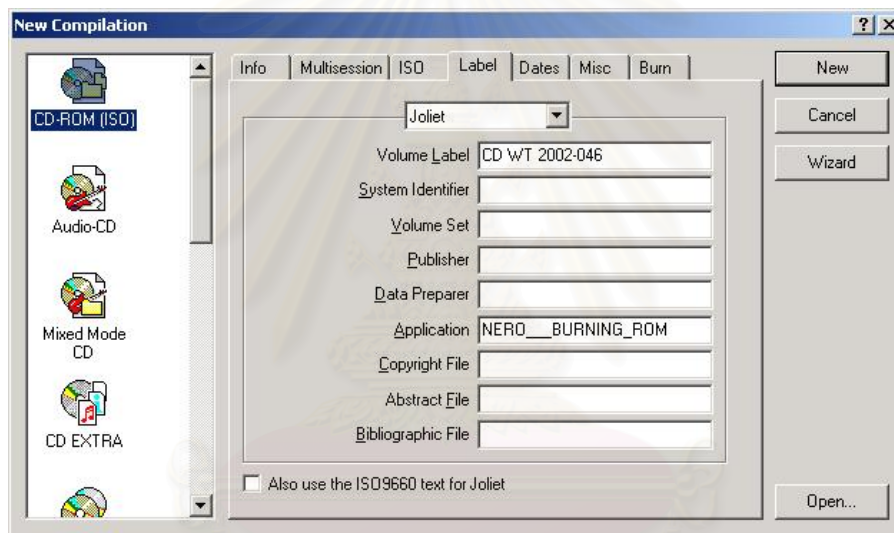
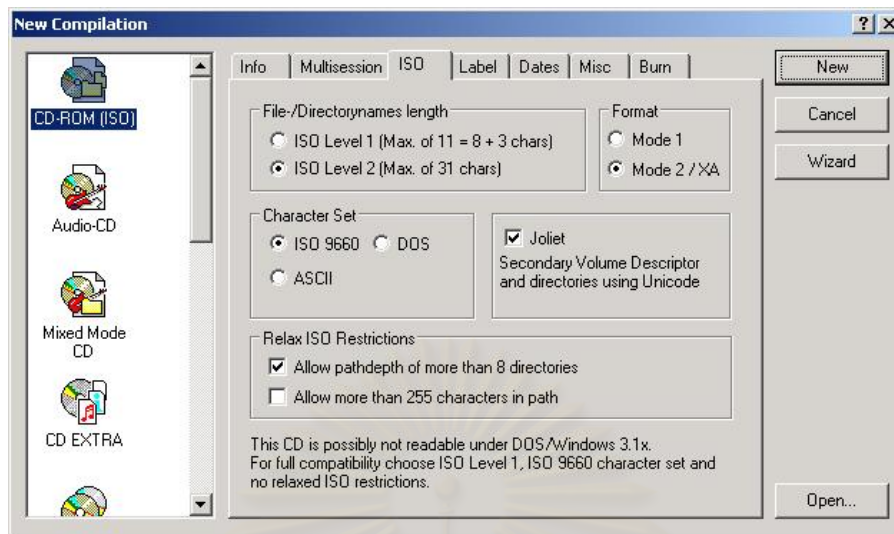
```

รูปที่ 2 แสดงการสร้าง chunk

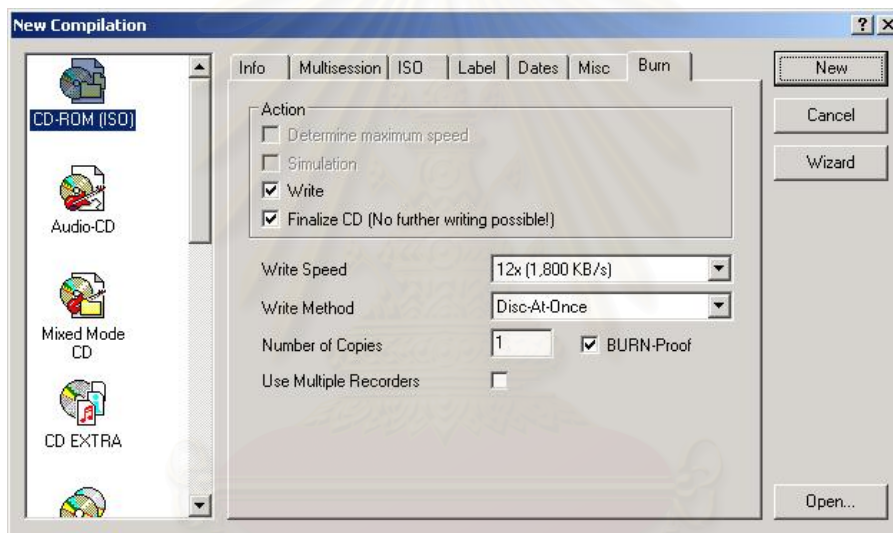
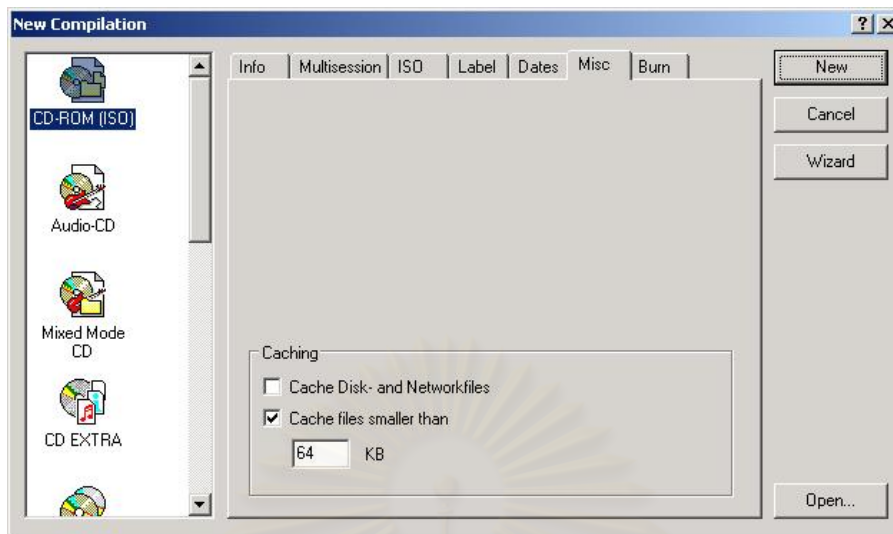
ต่อจากนั้นจะทำการ FTP chunk จากเครื่อง webcache ไปยังเครื่อง PC ที่มีเครื่องบันทึก CD เพื่อบันทึกลง CD พร้อมกับ meta data ของ CD ต่อไป โดยชื่อแผ่น CD จะอยู่ในรูปแบบ CD-WT-ปีของข้อมูล-หมายเลข 3 หลักแสดงลำดับการบันทึก CD รูปที่ 3 แสดงการบันทึก chunk ลงแผ่น CD



รูปที่ 3 แสดงการบันทึก chunk ลงแผ่นซีดีอาร์

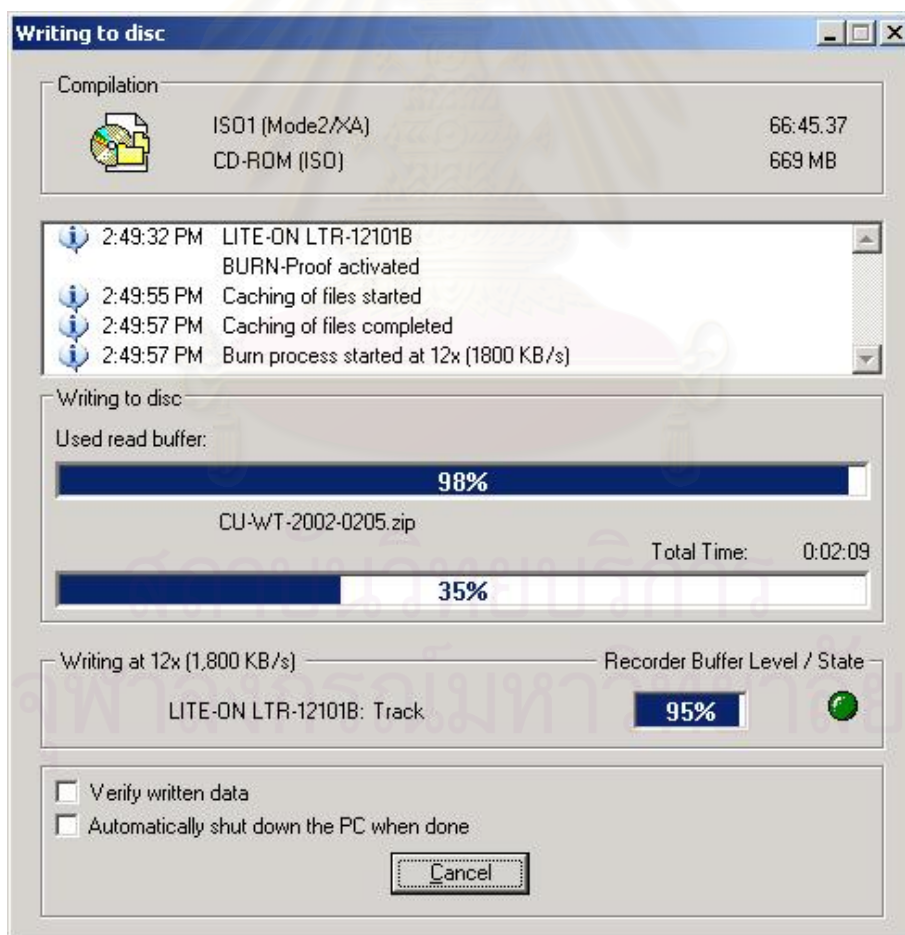
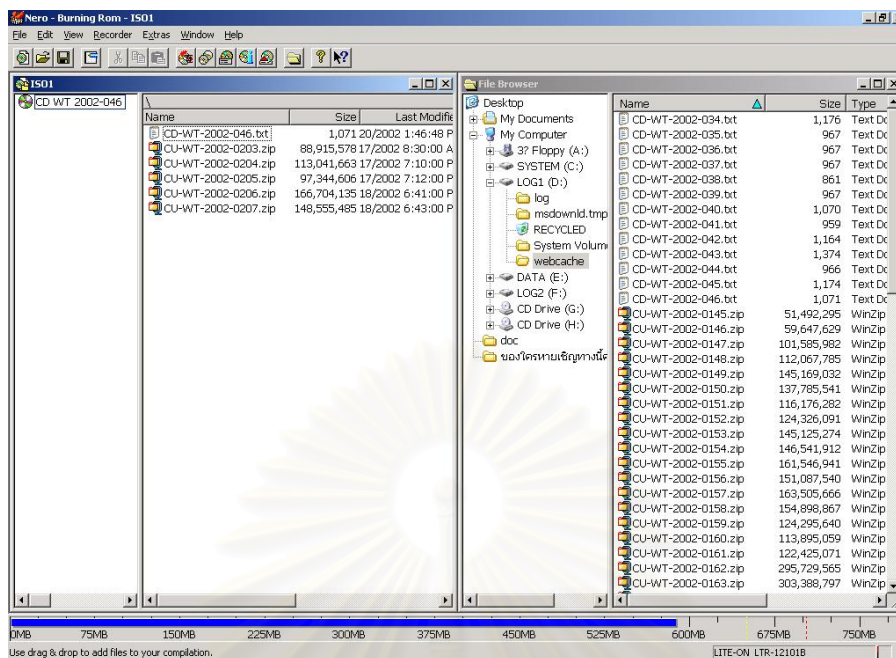


รูปที่ 3 แสดงการบันทึก chunk ลงแผ่นซีดี-อาร์ (ต่อ)



รูปที่ 3 แสดงการบันทึก chunk ลงแผ่นซีดี-อาร์ (ต่อ)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 3 แสดงการบันทึก chunk ลงแผ่นซีดี-อาร์ (ต่อ)

ตัวอย่างไฟล์ *data.txt* เป็น *meta data* ของข้อมูลใน *chunk*

1. File ID : data.txt for Chunk CU-WT-2002-0207
2. Period : 24 hours
3. Start : Wed 17 Apr 2002 00:00:00
4. End : Thu 18 Apr 2002 00:00:00
5. Used by : CUNET
6. Generated by : proxy3.chula.ac.th
7. Type : Web cache traffic log file.
8. Data Dictionary of file.

In each line, there are 10 fields separated by space.

See details in Squid manual pages.

- Field description

- a. Time : Unix timestamp in UTC with millisecond resolution.
- b. Duration : Elapse time of transaction (HTTP, ICP, ...) in milliseconds.
- c. Client address : IP address of the requesting machine.
- d. Result codes : Two entries separated by a slash.
 - Request type
 - Status
- e. Bytes : Amount of data delivered to the client.
- f. Request method : The request method to obtain an object.
- g. URL : This column contains the URL requested.
- h. RFC931 : The ident lookups for the requesting client.
- i. Hierarchy code : The hierarchy information consists of three items:
 - `_TIMEOUT` prefix for ICP replies from neighbours (optional).
 - A code that explains how the request was handled.
 - The name of the host the object was requested from.
- j. Type : The content type of the object in the HTTP reply header.

9. Example

```
1002711612.844 3983 161.200.131.208 TCP_REFRESH_HIT/200 22466 GET
http://www.cnn.com/virtual/2000/code/main.js - DIRECT/www.cnn.com binary/*
```

ตัวอย่างไฟล์ *equipment.txt* เป็น *meta data* ของอุปกรณ์ที่เป็น *source* ของข้อมูลใน *chunk*

1. File ID : equipment.txt for Chunk CU-WT-2002-0207
2. Type : Cache Engine
3. Model : Cisco Cache Engine 550 / CE version: 2.30(7)
4. Description : Web Cache for CUNET
5. Location : IP address 161.200.255.163



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตัวอย่างไฟล์ *CD-WT-2002-046.txt* เป็น *meta data* ของข้อมูล ในแผ่น *CD*

1. CD Label: CD WT 2002-046, 586.0 MB.
2. ISSN
3. Chunk
 - 3.1 Chunk1: CU-WT-2002-0203.zip, Mon 15 Apr 2002 00:00:00 - Tue 16 Apr 2002 00:00:00, 84.7 MB.
(proxy3)
 - 3.2 Chunk2: CU-WT-2002-0204.zip, Tue 16 Apr 2002 00:00:00 - Wed 17 Apr 2002 00:00:00, 107.0 MB.
(proxy2)
 - 3.3 Chunk3: CU-WT-2002-0205.zip, Tue 16 Apr 2002 00:00:00 - Wed 17 Apr 2002 00:00:00, 92.8 MB.
(proxy3)
 - 3.4 Chunk4: CU-WT-2002-0206.zip, Wed 17 Apr 2002 00:00:00 - Thu 18 Apr 2002 00:00:00, 158.0 MB.
(proxy2)
 - 3.5 Chunk5: CU-WT-2002-0207.zip, Wed 17 Apr 2002 00:00:00 - Thu 18 Apr 2002 00:00:00, 141.0 MB.
(proxy3)
4. Chunk characteristics
 - 4.1 Meta-data consists of data and equipment files that contain data and equipment configurations, respectively.
 - 4.2 Data and Meta-data packaged and compressed by program "zip 2.3" on Linux 2.4.3.
5. Write date: Sat 20 Apr 2002.
6. Expected expiry date: Fri 20 Apr 2007.
7. Hardware writer configuration: Celeron(r) Processor, CD-ReWritable LiteOn LTR-12101B.
8. Software writer configuration: Microsoft WindowsXP Professional, Nero-Burning Rom 5.5.5.1.

ภาคผนวก จ
Script สำหรับสนทนากับอินโนแพด



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Script สำหรับสนทนากับอินโนแพค

เป็นส่วนหนึ่งของการวิจัย
เรื่อง

การกำหนดรูปแบบบรรณานุกรมและการวางแนวทางการจัดเก็บแบบถาวรสำหรับข้อมูลต่อเนื่องที่ตัดมาเป็นกลุ่ม
ให้สืบค้นผ่านห้องสมุดอัตโนมัติ : กรณีศึกษาข้อมูลสถิติการใช้งานอินเทอร์เน็ตของ
จุฬาลงกรณ์มหาวิทยาลัย

สนับสนุนโดย
กองทุนรัชดาภิเษกสมโภช

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย
สถาบันวิทยบริการ

และ

ห้องปฏิบัติการวิศวกรรมระบบสารสนเทศ ภาควิชาวิศวกรรมคอมพิวเตอร์
จุฬาลงกรณ์มหาวิทยาลัย

30 เมษายน 2545

Script สำหรับสนทนากับอินโนแพค

โปรแกรม Gather

```
#!/usr/bin/expect --
```

```
set timeout 5
```

```
match_max -d 60000 ;#maximum size of match
```

```
#read current record number
```

```
proc rec {} {
```

```
    set frec [open ./rec.curr r]
```

```
    while {[gets $frec buf] != -1} {
```

```
        if {[scan $buf "rec\t%d" number] == 1} {
```

```
            close $frec
```

```
            return $number
```

```
        }
```

```
    }
```

```
    close $frec
```

```
    return 1
```

```
}
```

```
set count [rec]
```

```
log_user 0 ;#stop output of the spawn process
```

```
#telnet to library server
```

```
spawn telnet library.car.chula.ac.th
```

```
expect {
```

```
    timeout { puts "timed out";exit }
```

```
    "connection refused" exit
```

```
    "login:" { send "myaccount\r" }
```

```
}
```

```
expect "Password:" { send "mypassword\r" }
```

```

#search data
expect "Choose one (L,S,D,C,O,M,A,X)" { send "S\r" }
expect "Choose one (B,O,C,A,I,P,R,S,Q)" { send "B\r" }
expect "Choose one (A,T,S,W,C,I,R,P,Y,B,L,Q)" { send "S\r" }
expect "SUBJECT : " { send "Internet traffic log file\r" }
expect "Choose one (1-8,F,W,N,A,D,L,J,E,X)"
if {$count == 1} {
    send "I\r"
} else {
    send "J\r"
    expect "Advance how many entries or what title?" { send "[expr $count - 1]\r" }
    send "$count\r"
}

#open database file
set fdb [open ./db.gather a]

#write data to database file
while {1} {
    expect {
        "TITLE*IMPRINT" { puts $fdb [string range "$expect_out(0,string)" 0 [expr [string length
$expect_out(0,string)] - 14]] }
        "CALL*Data Traffic Log" { break }
    }
    expect "DESCRIPT*)" { puts $fdb "$expect_out(0,string)" }
    expect "LINK*.zip" { puts $fdb "$expect_out(0,string)\n" }
    incr count
    expect {
        "Choose one (R,F,N,A,S,Z,I,T,E)" { send "F\r" }
        "Choose one (R,F,B,N,A,S,Z,I,T,E)" { send "F\r" }
        "Choose one (M,R,B,N,A,Z,I,T,E,+)" { break }
        timeout { send "F\r" }
    }
}
}

```

```
#write current record number
set frec [open ./rec.curr w]
puts $frec "rec\t$count"
close $frec

#close database file
close $fdb

log_user 1      ;#restore output of the spawn process
```



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย