สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

รายงานโครงการวิจัย
ฉบับสมบูรณ์

การวัดและศึกษาคุณลักษณะด้านอายุและขนาดของเว็บออบเจ็กต์

Measurement and Study of Web Object's Life and Size Characteristics

อัครินทร์ คุณกิตติ
AKHARIN KHUNKITTI

เงินรายได้ประจำปีงบประมาณ 2545
คณะเทคโนโลยีสารสนเทศ
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

RCH
QA
76.64
อ479ก

เลขหมู่.....
เลขทะเบียน 131073
วัน,เดือน,ปี 22 พ.ค. 2557

b.12603958
i.....

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชื่อโครงการวิจัย (ภาษาไทย) **การวัดและศึกษาคุณลักษณะด้านอายุและขนาดของเว็บออบเจ็กต์**

(ภาษาอังกฤษ) **Measurement and Study of Web Object's Life and Size Characteristics**

ชื่อหัวหน้าโครงการวิจัย นาย อัครินทร์ คุณกิตติ

ชื่อหน่วยงานที่สังกัด คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

บทคัดย่อ

เว็บพร็อกซี่และแคชมีส่วนสำคัญในการช่วยลดความซ้ำซ้อนและเวลาที่ใช้ในการค้นหาข้อมูลจากเครือข่ายอินเตอร์เน็ต เนื่องจากเว็บพร็อกซี่และแคชจะมีการเก็บสำเนาออบเจ็กต์ที่เคยมีการเรียกใช้แล้วไว้ในแคช เมื่อมีการเรียกใช้ออบเจ็กต์เดิมอีกพร็อกซี่และแคชเซิร์ฟเวอร์จะนำออบเจ็กต์ที่อยู่ในแคชส่งไปให้ ในขณะเดียวกันก็จะมีการบันทึกการเข้าถึงออบเจ็กต์และเก็บสถานะต่างๆ ไว้ในล็อกไฟล์ งานวิจัยนี้จะเป็นการพัฒนาอัลกอริธึมและโปรแกรมเพื่อหาอายุและขนาดของออบเจ็กต์จากโปรแกรม Squid บนเว็บ การพัฒนาแบ่งได้เป็น 3 ส่วนคือ ส่วนแรกเป็นส่วนของการแก้ไขปรับปรุงโปรแกรม Squid เพื่อสร้าง Log File ใหม่ขึ้นมา Log File นี้จะถูกนำไปคำนวณในส่วนต่อไป ส่วนที่สองคือส่วนของการพัฒนาโปรแกรมสำหรับคำนวณหาอายุและขนาดของออบเจ็กต์ โดยจะมีการปรับปรุงอัลกอริธึมที่ใช้ในการคำนวณซึ่งจะใช้วิธีการคำนวณแบบใช้ค่า Last-Modified เข้ามาเพิ่ม เพื่อลดค่าความคลาดเคลื่อนให้น้อยลงส่วนที่สามคือส่วนที่ติดต่อกับผู้ใช้เพื่อการแสดงผลและสร้างรายงานซึ่งจะอยู่ในสภาพแวดล้อมแบบเว็บ โดยโปรแกรมจะนำข้อมูลสารสนเทศที่ผ่านการคำนวณหาอายุและขนาดแล้วจากฐานข้อมูล มาสร้างเป็นรายงานและแสดงผล การทำงานของโปรแกรมจะเริ่มต้นจากการตรวจสอบความถูกต้องของข้อมูลเกี่ยวกับออบเจ็กต์ที่ได้จาก Log File ใหม่ จากนั้นนำข้อมูลนั้นมาคำนวณหาอายุและขนาด แล้วนำข้อมูลเหล่านั้นไปเก็บไว้ยังฐานข้อมูลก่อนที่จะส่งไปแสดงผลยังผู้ใช้ เพื่อให้ผู้ใช้ทำการวิเคราะห์ข้อมูลต่อไป จากผลการศึกษาและทดลองพบว่า หาค่าความผิดพลาดและความแม่นยำของการหาอายุของเว็บออบเจ็กต์ อยู่ภายในขอบเขตที่กำหนด และจำนวนของ Log ของเว็บออบเจ็กต์นั้นๆ มีผลต่อความแม่นยำของการวัด โดยที่ถ้ามีการเรียกใช้เว็บออบเจ็กต์นั้นๆ มาก จะทำให้มีความแม่นยำของการวัดมากขึ้นด้วย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ABSTRACT

Web Proxy and cache help to reduce repetitive work and time searching information on the internet. Since Web Proxy and cache will copy called object and keep in cache, when there is a call to the same object the proxy and cache server would retrieve object from cache and forward to call function.Meanwhile, accession to the object and state would be stored in log file.

This research's objective is to develop algorithm and tool to calculate ages and size of object from "Squid" program on website, by research and make some change on parameters and algorithm which used in calculate age and size of objects from existing techniques about web proxy and cache. Development of tool has been divided into 3 main parts. First part is modifying "Squid" program to generate new log file which would be used in calculation in second part. The second part is development tool to calculate age and size of object by using c++ and add last modified value in the existing algorithm to reduce errors. Third part is API to display and generate report in web form. User could be able to use tool from any workstation which have internet connection. Tool would collect calculated information from database to generate report that display result of age and size of object from new log file.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

บทคัดย่อภาษาไทย 2
บทคัดย่อภาษาอังกฤษ 3
สารบัญ 4
บทที่ 1 บทนำ 5
บทที่ 2 การหาขนาดและอายุของเว็บออบเจ็กต์ 6
บทที่ 3 การวิเคราะห์และออกแบบระบบโปรแกรม 20
บทที่ 4 การพัฒนาระบบโปรแกรม 34
บทที่ 5 บทสรุป 47
บรรณานุกรม 49

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ที่มาของงานวิจัย

ในปัจจุบันเครือข่ายอินเตอร์เน็ตได้เข้ามามีบทบาทที่สำคัญในการสื่อสาร แลกเปลี่ยน เผยแพร่และค้นหาข้อมูลมากขึ้น โดยเฉพาะอย่างยิ่งการใช้งานเว็บ (Web Application) ทำให้ปริมาณการใช้งานเว็บเพิ่มขึ้นอย่างรวดเร็ว และจากลักษณะการใช้งานเว็บจะเป็นการเรียกดูข้อมูลจากแหล่งข้อมูลหรือ Web Server เป็นส่วนใหญ่ จึงมีการนำเอาระบบเว็บแคช (Web Cache) มาใช้ เพื่อให้การใช้งานเครือข่ายมีประสิทธิภาพมากขึ้น ทั้งในด้านการลดเวลาในการใช้งานเรียกข้อมูลของผู้ใช้ และช่วยให้การใช้งาน Bandwidth ของการเชื่อมต่อเครือข่ายลดความซ้ำซ้อนของการใช้ข้อมูลลงได้ ทำให้สามารถใช้งานการเชื่อมต่อนั้นได้อย่างคุ้มค่ามากขึ้น แต่การใช้ระบบเว็บแคชยังไม่สามารถที่จะกำหนดได้ว่าจะต้องมีการใช้ทรัพยากรสำหรับระบบเว็บแคชในขนาดใดจึงจะเหมาะสมกับการใช้งานของเครือข่ายนั้นๆ ดังนั้นจึงได้มีการศึกษาและวิจัยหาวิธีการจัดการและประเมินขนาดของระบบเว็บแคชให้เหมาะสมกับปริมาณการใช้งาน เช่น การกำหนดขนาดของเนื้อที่ของแคชเพื่อเก็บข้อมูล ขนาดความสามารถในการประมวลผล ความเร็วของการรับส่งข้อมูลภายในเครื่อง เป็นต้น

ปัจจัยหนึ่งที่มีผลโดยตรงต่อระบบเว็บแคชก็คือ คุณลักษณะของข้อมูลเว็บหรือเว็บออบเจ็กต์ที่ถูกใช้งาน ซึ่งเป็นสิ่งที่ระบบจะต้องเก็บไว้เพื่อใช้งานในการเรียกใช้ครั้งต่อๆไป และคุณลักษณะของเว็บออบเจ็กต์ที่สำคัญที่มีผลมากก็คือ อายุและขนาด ของเว็บออบเจ็กต์ที่มีการใช้งานผ่านระบบเว็บแคช โดยที่ผ่านมาได้มีการศึกษาและวิจัยการหาและวิเคราะห์คุณลักษณะของอายุและขนาดของข้อมูลเว็บอยู่หลายโครงการ แต่ส่วนใหญ่จะเป็นการเฝ้าตรวจสอบ ณ แหล่งเก็บข้อมูลโดยตรง ทำให้ต้องใช้ทรัพยากรของระบบมาก เนื่องจากต้องคอยเฝ้าดูการเปลี่ยนแปลงอยู่ตลอดเวลา รวมทั้งข้อมูลนั้นอาจจะไม่ได้ใช้งานผ่านระบบเว็บแคชก็ได้ เพราะผู้ใช้ไม่ได้ใช้งานข้อมูลดังกล่าว ทำให้เป็นการสิ้นเปลืองและยากในการศึกษาลักษณะข้อมูลด้านอายุและขนาดของเว็บ

ดังนั้นเพื่อเป็นการใช้งานระบบเว็บแคชได้อย่างมีประสิทธิภาพ ข้อมูลการใช้งานของผู้ใช้ผ่านเว็บแคช ได้แก่ Log Files ต่างๆ สามารถที่จะนำเอามาวิเคราะห์หาคุณลักษณะของอายุและขนาดของเว็บออบเจ็กต์ได้ และยังสะท้อนถึงคุณลักษณะของข้อมูลเว็บที่มีการใช้งานจริงในระบบอีกด้วย ทำให้สามารถนำเอาคุณลักษณะดังกล่าวไปใช้ในการประเมินและจัดการระบบเว็บแคชได้เป็นอย่างดี

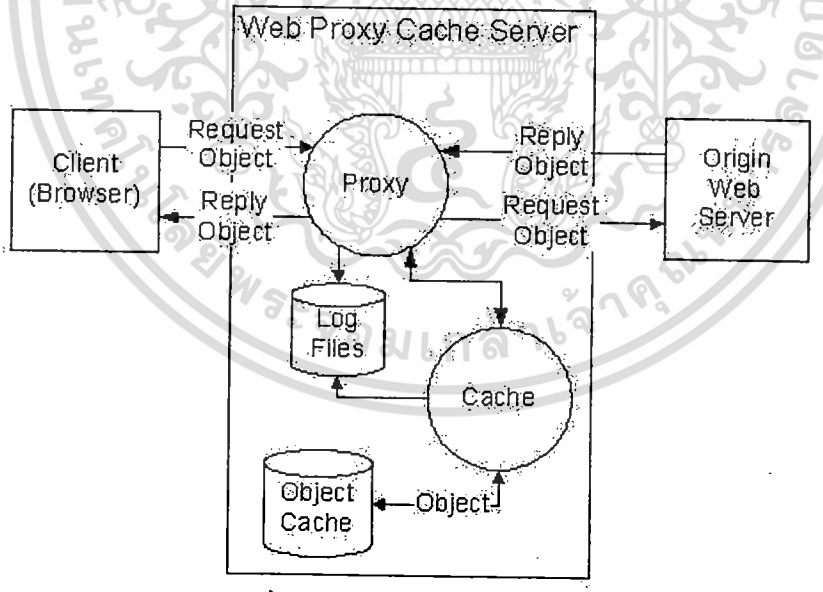
การศึกษาวิจัยในโครงการนี้จะเป็นการศึกษาและปรับปรุงการหาอายุของเว็บออบเจ็กต์ที่มีการเรียกใช้งานผ่านเว็บแคช ที่ชื่อ Squid โดยใช้ Log File ที่ชื่อ Access.log ในการประมวลหาอายุและขนาดของเว็บออบเจ็กต์จะเริ่มจากการคัดเลือก Log ที่เกี่ยวข้องกับการหาอายุและขนาด พบว่าควรจะเลือกเฉพาะ HTTP GET Method หลังจากนั้นจะทำการจัดแบ่งประเภทเป็น LIFE_HIT, LIFE_MISS, LIFE_UNKNOWN และ LIFE_IGNORED และเข้าสู่ขั้นตอนการคำนวณอายุและขนาดของเว็บออบเจ็กต์สำหรับ LOG นั้น ต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
การหาขนาดและอายุของเว็บออบเจ็กต์

2.1 หลักการทำงานของเว็บพร็อกซี่และแคช

ในการทำงานของเว็บพร็อกซี่และแคชโดยทั่ว ๆ ไป (รวมทั้ง Squid) จะทำงานเป็นแบบ Demand Driven คือ จะให้บริการก็ต่อเมื่อมีการร้องขอจากไคลเอนต์ (Client) ซึ่งมีขั้นตอนการทำงานดังภาพที่ 2.1 เมื่อไคลเอนต์ร้องขอข้อมูลจากพร็อกซี่เซิร์ฟเวอร์ (Proxy Server) ก็จะมีการตรวจดูว่ามีออบเจ็กต์ที่ต้องการอยู่ในแคชหรือไม่ ถ้ามีก็จะนำออบเจ็กต์นั้นส่งไปให้ไคลเอนต์ แต่ถ้าไม่มีพร็อกซี่เซิร์ฟเวอร์จะเป็นตัวแทนไปติดต่อกับเซิร์ฟเวอร์ปลายทาง (Origin Server) และนำออบเจ็กต์มาเก็บสำเนา (Caching) ไว้ที่เซิร์ฟเวอร์ชั่วคราว แล้วจึงนำส่งต่อไปให้ไคลเอนต์ เมื่อมีการร้องขอออบเจ็กต์นี้อีกในครั้งต่อไป ก็จะนำสำเนาออบเจ็กต์ที่อยู่ในแคชส่งไปให้ ทำให้สามารถลดการใช้งานช่องทางการสื่อสารในเครือข่ายอินเตอร์เน็ตได้เป็นอย่างมาก และสามารถตอบสนองต่อผู้ใช้ได้อย่างมีประสิทธิภาพ เพราะส่วนใหญ่พร็อกซี่และแคชจะติดตั้งอยู่ในเครือข่ายภายในองค์กรที่มีความเร็วสูงอยู่แล้ว



รูปที่ 2.1 แสดงการทำงานของพร็อกซี่และแคช

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยทั่วไปแล้วโปรแกรมพร็อกซี่และแคชเซิร์ฟเวอร์จะสามารถรองรับโพรโทคอล (Protocol)ได้หลายโพรโทคอล โดยเราจะแบ่งโพรโทคอลที่สามารถทำงานร่วมกับโปรแกรมพร็อกซี่และแคชออกได้เป็น 3 ลักษณะ คือ

- โพรโทคอลที่ให้บริการกับผู้ใช้ทั่ว ๆ ไป เช่น HTTP, FTP, Gopher, WAIS, SSL
- โพรโทคอลที่ใช้สื่อสารและจัดการระหว่างพร็อกซี่และแคชด้วยกันเอง โพรโทคอลที่มีลักษณะนี้เช่น HTTP (ใช้รับสำเนาของออบเจ็กต์จากแคชในการทำงานแบบ Cooperative), ICP, Cache Digests, CARP, HTCP เป็นต้น
- โพรโทคอลที่ใช้ในการควบคุมและจัดการแคช เช่นโพรโทคอล SNMP เป็นต้น

ลักษณะการเก็บข้อมูลหรือออบเจ็กต์จะเก็บในรูปแบบของไฟล์ ซึ่งในไฟล์นั้นจะเก็บทั้งตัวออบเจ็กต์และส่วนของ Header โดยที่ชื่อไฟล์มีการเข้ารหัส Hash Function เก็บไว้ใน path สองลำดับชั้น ซึ่งสามารถปรับเปลี่ยนขนาดได้เมื่อที่เก็บข้อมูลเต็มก็จะทำการลบหรือแทนที่ข้อมูลเก่า ๆ ตามนโยบายการแทนที่ข้อมูล (Replacement policy)ที่กำหนดไว้ เพื่อให้มีเนื้อที่เหลือพอสำหรับเก็บข้อมูลใหม่ต่อไป โดยการเก็บข้อมูลภายในแคชจะมีการเก็บไว้อย่างเป็นระเบียบ เพื่อความรวดเร็วและสะดวกในการค้นหาเพื่อส่งสำเนาของข้อมูลดังกล่าวเมื่อมีไคลเอนต์ร้องขอข้อมูลในรูปแบบของออบเจ็กต์ เพราะในอินเตอร์เน็ตมีออบเจ็กต์หลากหลายประเภทไม่ว่าจะเป็นข้อมูล ภาพ เสียง หรือ วิดีโอ และจะมีการเก็บข้อมูลที่จำเป็นไว้ เช่น URL ของออบเจ็กต์เพื่อเปรียบเทียบกับ URL ที่ไคลเอนต์ร้องขอ, Time to live (TTL) คือ อายุของออบเจ็กต์ เมื่อออบเจ็กต์หมดอายุ แม้ว่าจะมีผู้ร้องขอข้อมูลที่มีอยู่ในเว็บ (Web) แคชก็จะยังไม่ส่งสำเนาที่หมดอายุไปให้ แต่จะติดต่อไปยังเครื่องต้นทางที่แท้จริงเพื่อร้องขอออบเจ็กต์ใหม่ โดยถ้าออบเจ็กต์นั้นหมดอายุก็จะทำการร้องขอแบบ IMS GET (If Modify Since GET) ส่วนในกรณีไม่พบในแคชก็จะทำการร้องขอแบบ GET

ในปัจจุบันนิยมใช้พร็อกซี่และแคชเซิร์ฟเวอร์ ในลักษณะที่เรียกว่า Transparency เพื่อให้สะดวกต่อการใช้งาน เพราะผู้ใช้ไม่จำเป็นต้องรับรู้ว่าต้องมีการใช้พร็อกซี่เซิร์ฟเวอร์และผู้ใช้ก็ไม่จำเป็นต้องตั้งค่าใด ๆ เลยก็สามารถใช้พร็อกซี่และแคชได้โดยอัตโนมัติ นอกจากนี้ในองค์กรบางแห่งที่มีผู้ใช้งานอินเตอร์เน็ตเป็นจำนวนมากจึงอาจจะต้องมีการติดตั้งพร็อกซี่และแคชเซิร์ฟเวอร์มากกว่าหนึ่งเครื่องทำงานร่วมกัน เพื่อรองรับความต้องการการใช้งานที่เกิดขึ้น ซึ่งเป็นการทำงานของพร็อกซี่ และแคชเซิร์ฟเวอร์ในแบบ Cooperative

2.2 โปรแกรม Squid

โปรแกรม Squid เป็นโปรแกรมสำหรับทำ Web Caching มีจุดเริ่มต้นมาจากการพัฒนาโปรแกรมของ Harvest Project ซึ่งไม่ได้มีจุดมุ่งหมายในการพัฒนาแคชโดยตรง แต่เป็นโครงการที่ได้รับเงินทุนสนับสนุนจาก National Laboratory of Network Research (NLANR)

โปรแกรม Squid เป็นแอพพลิเคชั่นที่เผยแพร่ Source Code โดยไม่คิดค่าใช้จ่ายจึงสามารถนำโปรแกรม Squid ไปพัฒนาเพิ่มเติมความสามารถที่ต้องการ และแก้ไขข้อผิดพลาดที่เกิดขึ้นได้ ทำให้โปรแกรม Squid เป็นที่นิยมนำไปใช้เพราะผู้ที่มีความรู้ด้านการเขียนโปรแกรมสามารถปรับปรุงประสิทธิภาพและความสามารถของโปรแกรม Squid ได้ตามที่ตนเองต้องการ

โปรแกรม Squid สามารถทำงานร่วมกับระบบปฏิบัติการ UNIX ได้เกือบทุก Platforms เช่น FreeBSD, Linux, OpenBSD, NetBSD เป็นต้น ปกติแล้วจะสามารถรองรับ Features ได้ใกล้เคียงกัน เนื่องจากโปรแกรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Squid ถูกพัฒนาบน Digital UNIX ที่ทำงานกับ GNU C compiler จึงสามารถนำโปรแกรม Squid ไปติดตั้งได้บนระบบปฏิบัติการที่มี Compiler ดังกล่าวติดตั้งอยู่ได้ โดยการติดตั้งจะเลือกนำ Source Code มา คอมไพล์ (Compile) หรือนำ Binary Versions มาติดตั้ง ซึ่งสามารถกระทำได้ง่ายกว่า แต่อาจจะมีปัญหาเรื่องของความปลอดภัยที่เกิดจากการคอมไพล์ Source Code ที่ถูกดัดแปลงการทำงานเพื่อเปิดช่องโหว่ในการโจมตีจากผู้ไม่หวังดีได้

ในการทำงานของโปรแกรม Squid จะมีการบันทึกผลการร้องขอและผลของการทำงานที่เกิดขึ้นไว้ในล็อกไฟล์ ซึ่งจะมีชื่อไฟล์แตกต่างกันตามประเภทของข้อมูลที่ถูกบันทึกไว้ โดยล็อกไฟล์จะเป็นตัวบันทึกการทำงานและบอกถึงประสิทธิภาพของการทำงาน ในโปรแกรม Squid จะประกอบด้วยล็อกไฟล์ต่าง ๆ เช่น

- Store.log ทำหน้าที่บันทึกการเก็บหรือนำข้อมูลออกจากแคช บอกถึงลักษณะของออบเจ็กต์ว่าถูกนำเก็บลงหรือนำออกจากแคชหรือดิสก์
- Access.log จะทำหน้าที่เก็บผลของการร้องขอที่เกิดจากการร้องขอของผู้รับบริการ รวมถึงการแลกเปลี่ยน ICP Message ที่เกิดขึ้นในการทำงานของโปรแกรม
- Cache.log ประกอบด้วยข้อมูลที่เกี่ยวข้องกับความผิดพลาดที่เกิดขึ้นจากการทำงานของโปรแกรม Squid เอง

โดยล็อกไฟล์ Access.log และล็อกไฟล์ Store.log เป็นไฟล์ที่นำมาพัฒนาสร้างเป็นล็อกไฟล์ตัวใหม่เพื่อนำข้อมูลมาคำนวณค่าอายุและขนาดของเว็บออบเจ็กต์ที่ถูกร้องขอผ่านโปรแกรม Squid โดยในไฟล์ Access.log นี้จะประกอบไปด้วยฟิลด์ต่าง ๆ ทั้งหมด 10 ฟิลด์ [1] ซึ่งมีรายละเอียดดังนี้

ตาราง 2.1 ฟิลด์ต่างๆของ Access.log ที่เกิดจากการประมวลผลการร้องขอของ Squid

Timestamp	Elapsed	Client Address	Log Tag / HTTP Code	Byte	Request method	URL	Rfc931	Hierarchy Hostname	Content Type

- Timestamp จากการร้องขอที่เกิดขึ้น อยู่ในรูปแบบ UNIX Time โดยจะบันทึกเมื่อปิดการเชื่อมต่อกับไคลเอนต์
- Elapsed Time เป็นฟิลด์เก็บระยะเวลาที่ใช้ในการทำงานของโปรแกรมต่อการร้องขอนั้น ซึ่งมีหน่วยเป็น milliseconds
- Client Address เก็บค่า IP address ของไคลเอนต์ที่ร้องขอออบเจ็กต์
- Log Tag/HTTP-Code เก็บผลการทำงานของการร้องแต่ละครั้ง เช่น ประเภทของการร้องขอ ความผิดพลาดที่เกิดขึ้น ซึ่งก็คือผลการทำงานของ Squid ที่มีต่อการร้องขอนั้น/ผลการทำงานจากโพรโทคอล HTTP
- Byte เก็บขนาดของข้อมูลทั้งหมดที่ถูกส่งไปยังผู้รับบริการ แต่ข้อมูลส่วนนี้ไม่ใช่ขนาดของออบเจ็กต์ที่แท้จริงเพราะจะนับรวมถึงส่วนของ Header ด้วย
- Request Method เก็บข้อมูลรูปแบบของการร้องขอที่เกิดขึ้นกับออบเจ็กต์ ซึ่งก็คือวิธีการของการร้องขอเพื่อจะกระทำกับออบเจ็กต์เช่น GET, POST เป็นต้น
- URL เก็บที่อยู่ของออบเจ็กต์ที่ถูกไคลเอนต์ร้องขอ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- Rfc931 ทำหน้าที่เก็บ Ident ของไคลเอนต์ที่ร้องขอออบเจ็กต์
- Hierarchy Code เก็บข้อมูลที่เกี่ยวข้องกับการทำงานเป็นลำดับชั้นของโปรแกรมเช่น ออบเจ็กต์ถูกร้องขอมาจาก Neighbor Cache หรือ Origin Server เป็นต้น
- Type บอกถึงประเภทของออบเจ็กต์ที่ถูกส่งมาใน HTTP header หากเป็นการแลกเปลี่ยน ICP จะไม่มีข้อมูลในส่วนนี้

จากข้อมูลทั้งหมดที่ถูกเก็บไว้ในล็อกไฟล์ Access.log ที่กล่าวมานี้ จะสามารถนำข้อมูลที่ถูกบันทึกไว้มาใช้ประกอบเพื่อคำนวณหาอายุและขนาดของออบเจ็กต์ ที่ถูกร้องขอผ่านโปรแกรม Squid ได้

ส่วนในไฟล์ Store.log นี้จะประกอบไปด้วยฟิลด์ต่าง ๆ ทั้งหมด 12 ฟิลด์ ซึ่งมีรายละเอียดดังนี้

ตารางที่2.2 ฟิลด์ต่างๆของ Store.log ที่เกิดจากการประมวลผลของ Squid

Timestamp	Action	Dir Numer	File Number	Status	Datehdr	Lastmode	Expires	Type	Size	Method	Key

- Timestamp จากการร้องขอที่เกิดขึ้น อยู่ในรูปแบบ UNIX Time โดยจะบันทึกเมื่อปิดการเชื่อมต่อกับไคลเอนต์
- Action ชนิดการกระทำของออบเจ็กต์
  - Create เป็นออบเจ็กต์ใหม่ ไม่เคยมีการใช้งานมาก่อน
  - Release เป็นออบเจ็กต์ที่มีอยู่แล้ว นำออกมาจากแคชเพื่อไปใช้งาน
  - Swapout มีการบันทึกออบเจ็กต์ไว้แล้ว สามารถนำออกไปใช้ได้เลย
  - Swapin หมดอายุของออบเจ็กต์ในดิสก์แล้ว ต้องอ่านจากหน่วยความจำเข้ามา
- Dir numer เป็นค่าของออบเจ็กต์ที่ถูกเก็บลงในแคช ถ้าเป็นครั้งแรกก็มีค่าเป็น 0
- File number เก็บ file number ของการเก็บออบเจ็กต์ลง storage file
- Status เก็บรหัสบอกสถานะของ HTTP
- Datehdr เก็บค่า Dateของออบเจ็กต์ที่ถูกส่งมาใน HTTP header
- Lastmod เก็บค่า Last-Modifiedของออบเจ็กต์ที่ถูกส่งมาใน HTTP header
- Expires เก็บค่าการสิ้นอายุของออบเจ็กต์ที่ถูกส่งมาใน HTTP header
- Type บอกถึงประเภทของออบเจ็กต์ที่ถูกส่งมาใน HTTP header
- Size เก็บขนาดของข้อมูลทั้งหมดที่ถูกนำเก็บลงหรือนำออกจากแคชหรือดิสก์แต่ข้อมูลส่วนนี้ไม่ใช่ขนาดของออบเจ็กต์ที่แท้จริงเพราะจะนับรวมถึงส่วนของ Header ด้วย
- Method เก็บข้อมูลรูปแบบของการร้องขอที่เกิดขึ้นกับออบเจ็กต์ ซึ่งก็คือวิธีการของการร้องขอเพื่อจะกระทำกับออบเจ็กต์เช่น GET, POST เป็นต้น
- Key เก็บURL ของออบเจ็กต์ที่ถูกนำเก็บลงหรือนำออกจากแคชหรือดิสก์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3 หลักการหาอายุและขนาดของออบเจ็กต์

ในการคำนวณหาอายุและขนาดของออบเจ็กต์จากไฟล์ mix.log ที่ได้พัฒนาขึ้นจากโปรแกรม Squid จะเลือกใช้เฉพาะข้อมูลส่วนที่เกี่ยวข้องกับการหาอายุและขนาดของเว็บออบเจ็กต์จากไฟล์ access.log จำนวน 5 ฟิลด์จากทั้งหมด 10 ฟิลด์ ดังต่อไปนี้

- Time: เวลาที่เกิดการร้องขอในรูปแบบของ UTC กับรายละเอียดเป็น millisecond โดยเวลานี้จะถูกบันทึกลงล็อกไฟล์เมื่อมีการปิดการเชื่อมต่อกับไคลเอนต์
- Result Code: บอกผลการทำงานของโปรแกรม Squid ที่มีต่อการร้องขอที่เกิดขึ้น เป็นตัวบ่งบอกถึงประเภทการร้องขอ
- Byte: เป็นฟิลด์ที่ใช้บอกขนาดของออบเจ็กต์ทั้งหมด(รวม Header) ที่ส่งไปให้กับไคลเอนต์มีหน่วยเป็น Byte
- Request Method: คือวิธีการที่ใช้ในการร้องขอออบเจ็กต์ เช่น GET, POST เป็นต้น ซึ่งในระบบงานนี้จะสนใจเฉพาะ Request Method ที่เป็น GET เท่านั้น
- URL: เป็นที่อยู่ของออบเจ็กต์ที่ไคลเอนต์ได้ร้องขอ ซึ่งในที่นี้จะใช้เป็นตัวระบุชื่อ เพื่อแยกแยะออบเจ็กต์แต่ละอัน

และจากไฟล์ Store.log จำนวน 2 ฟิลด์จากทั้งหมด 12 ฟิลด์ ดังต่อไปนี้

- Expires เก็บค่าการสิ้นอายุของออบเจ็กต์ที่ถูกส่งมาใน HTTP header
- Lastmod เก็บค่า Last-Modifiedของออบเจ็กต์ที่ถูกส่งมาใน HTTP header

2.3.1 วิธีการคำนวณหาขนาดของเว็บออบเจ็กต์

เราสามารถนำข้อมูลในฟิลด์ Byte ซึ่งเป็นฟิลด์ที่ใช้ในการเก็บขนาดของเว็บออบเจ็กต์มาใช้ได้เลย โดยขนาดที่เก็บอยู่ในฟิลด์นี้ เป็นขนาดของเว็บออบเจ็กต์ที่รวมส่วนของ Header ของข้อมูลนั้นไว้ด้วย ดังนั้น ค่าที่ได้จึงไม่ใช่ขนาดของออบเจ็กต์จริงที่อยู่ในอินเตอร์เน็ต เพราะเป็นค่าขนาดของออบเจ็กต์รวมกับ Header และจะบันทึกค่า ก็ต่อเมื่อพบว่าเว็บออบเจ็กต์นั้นเกิดการเปลี่ยนแปลงขึ้นเท่านั้น

ในการหาขนาดของเว็บออบเจ็กต์นี้ มีเงื่อนไขอยู่ว่าเราไม่ได้เก็บค่าขนาดของเว็บออบเจ็กต์ทุกๆ ค่าที่ปรากฏอยู่ในล็อกไฟล์นั้น แต่จะเก็บค่าขนาดของเว็บออบเจ็กต์ก็ต่อเมื่อพบว่าเว็บออบเจ็กต์นั้น ๆ มีการเปลี่ยนแปลงเกิดขึ้นเท่านั้น โดยจะนับจำนวนของอายุของเว็บออบเจ็กต์ไปด้วย

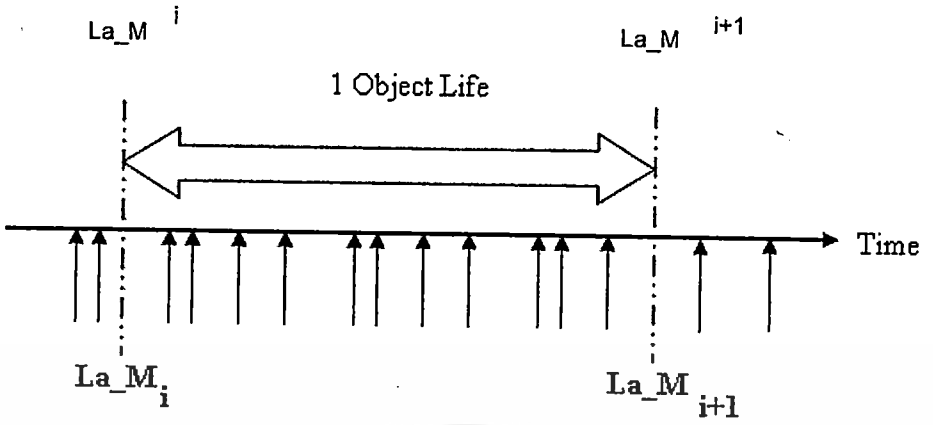
2.3.2 วิธีการคำนวณหาอายุของเว็บออบเจ็กต์

วิธีการคำนวณหาอายุของเว็บออบเจ็กต์สามารถแบ่งได้ 2 วิธีดังต่อไปนี้คือ

2.3.2.1 การคำนวณหาอายุของเว็บออบเจ็กต์แบบเว็บออบเจ็กต์มีค่า Last-Modified

นำข้อมูลที่ได้จาก mix.log ที่ได้ทำการพัฒนาจากล็อกไฟล์ของโปรแกรม squid ซึ่งเว็บออบเจ็กต์มีค่า Last-Modified มาคำนวณหาอายุของออบเจ็กต์ เมื่อนำค่า Last-Modifiedมาเขียนแทนด้วยลูกศรบนแกนเวลาก็จะได้แผนภาพดังแสดงใน รูปที่ 2.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.2 **แสดงการหาอายุออบเจ็กต์จากค่า Last-Modifiedในพร็อกซี่เซิร์ฟเวอร์**

นิยามตัวแปร

La_M_i คือ ค่า Timestamp ของ Last-Modified ค่าแรกสุด ที่ยังไม่เคยมีการ access เข้ามา

La_M_{i+1} คือ ค่า Timestampของ Last-Modified ที่ค่าได้เกิดการเปลี่ยนแปลงไป

$LifeCal_i$ คือ ค่าอายุของเว็บออบเจ็กต์ที่ได้จากการคำนวณ

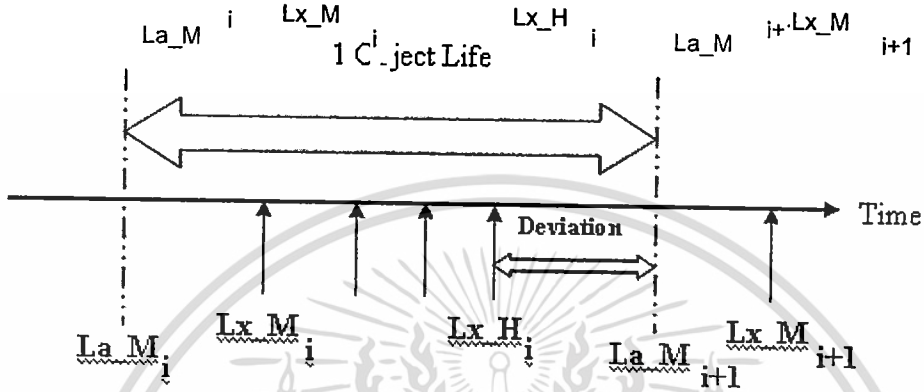
จากรูปที่ 2.2 เมื่อมีการร้องขอเว็บออบเจ็กต์เข้ามาทางซ้ายสุดของแกนเวลา ซึ่งอาจจะเป็นการร้องขอครั้งแรกออบเจ็กต์จะยังไม่มีค่าของ Last-Modified จึงทำการบันทึกค่าไว้ในล็อกไฟล์ เมื่อมีการร้องขอเว็บออบเจ็กต์นี้เข้ามาอีกและพร็อกซี่เซิร์ฟเวอร์พบเว็บออบเจ็กต์ที่ต้องการในแคชแล้วได้ทำการตรวจสอบแล้วว่าเว็บออบเจ็กต์ปัจจุบันที่อยู่ในอินเทอร์เน็ตยังไม่มีการเปลี่ยนแปลงขนาด และ/หรือ วันที่เว็บออบเจ็กต์นั้นถูกสร้าง(ค่า Last-Modified ไม่มีการเปลี่ยนแปลง) จึงทำการส่งเว็บออบเจ็กต์ที่เก็บไว้ในแคชไปให้ไคลเอนต์ที่ร้องขอมา โดยจะกำหนดให้การร้องขอดังกล่าวนี้อยู่ในประเภทของการร้องขอแบบ La_M_i จากนั้น การร้องขอครั้งถัดไปดังแสดงในรูป เมื่อมีการร้องขอเว็บออบเจ็กต์นี้เข้ามาอีก แต่เว็บออบเจ็กต์ปัจจุบันที่อยู่ในอินเทอร์เน็ตมีการเปลี่ยนแปลงไป พร็อกซี่เซิร์ฟเวอร์ก็จะต้องติดต่อเพื่อไปเอาออบเจ็กต์นี้มาเก็บลงในแคชแทนออบเจ็กต์ตัวเดิมที่มีอยู่ก่อนหน้านี้ และทำการส่งออบเจ็กต์นั้นให้กับไคลเอนต์(ค่าของ Last-Modified มีการเปลี่ยนแปลงไป) ในกรณีนี้จะกำหนดให้การร้องขอดังกล่าวอยู่ในประเภทของการร้องขอแบบ La_M_{i+1} ในความเป็นจริงการร้องขอที่เข้ามายังพร็อกซี่เซิร์ฟเวอร์นั้น อาจมีการร้องขอเว็บออบเจ็กต์เดียวกันจากหลายไคลเอนต์ในเวลาไล่เลี่ยกัน จึงมีโอกาสของการเกิดกรณีของ La_M_{i+1} มากกว่า La_M_i นอกจากนี้ ยังมีผลของการร้องขอที่ออบเจ็กต์ไม่ค่า Last-Modified จะจัดให้อยู่ในประเภทของการร้องขอแบบเว็บออบเจ็กต์ไม่มีค่า Last-Modified จะนำไปแบ่งแยกและทำการคำนวณในอีกรูปแบบหนึ่ง

ดังนั้นเราสามารถประมาณอายุของเว็บออบเจ็กต์ที่มีค่า Last-Modified ในช่วงที่ i ($LifeCal_i$) ดังแสดงในสมการที่ 2.1

$$LifeCal_i = La_M_{i+1} - La_M_i \qquad \ldots\ldots\ldots\ldots (2.1)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อนำข้อมูลที่ได้จาก mix.log ที่ได้ทำการพัฒนาจากล็อกไฟล์ของโปรแกรม squid ซึ่งเว็บออบเจ็กต์มีค่า Last-Modified มาคำนวณหาอายุของออบเจ็กต์ จะพบว่ามีความคลาดเคลื่อนเกิดขึ้นได้ค่าหนึ่ง เมื่อนำค่า Last-Modified และช่วงของความคลาดเคลื่อนมาเขียนแทนด้วยลูกศรบนแกนเวลา ก็จะได้แผนภาพดังแสดงในรูปที่ 2.3



รูปที่ 2.3 แสดงการหาความคลาดเคลื่อนจากวิธีการหาอายุของเว็บออบเจ็กต์จากค่า Last-Modified ในพร็อกซี่เซิร์ฟเวอร์

นิยามตัวแปร

La_M_i คือ ค่า Timestamp ของ Last-Modified ค่าแรกสุด ที่ยังไม่เคยมีการ access เข้ามาเลย

La_M_{i+1} คือ ค่า Timestamp ของ Last-Modified ที่ค่าได้เกิดการเปลี่ยนแปลงไป

Lx_M_i คือ ค่า Timestamp ของ การร้องขอเข้ามาครั้งแรกสุดยังซึ่งไม่เคยมีการร้องขอใดก่อนหน้านี้

Lx_M_{i+1} คือ ค่า Timestamp ของการร้องขอครั้งแรกสุดหลังจากค่า Last-Modified ได้เกิดการเปลี่ยนแปลงไป ซึ่งจะไม่มีการร้องขอใดๆเข้ามาก่อนหน้านี้อย่างแน่นอน

Lx_H_i คือ ค่า Timestamp ของการร้องขอครั้งสุดท้ายก่อนที่ค่า Last-Modified จะเกิดการเปลี่ยนแปลงไป

จากรูปที่ 2.3 มีการร้องขอเว็บออบเจ็กต์เข้ามาทางซ้ายสุดของแกนเวลา ซึ่งอาจจะเป็นการร้องขอครั้งแรกสุด (ค่าการร้องขอที่ Lx_M_i) เว็บออบเจ็กต์จะยังไม่มีค่าของ Last-Modified จึงทำการบันทึกค่าไว้ในล็อกไฟล์ และค่าการร้องขอแรกสุดที่ Lx_M_iจะมีค่า Last-Modified เท่ากับค่า Last-Modified ครั้งแรกสุดด้วยคือ La_M_i เมื่อต่อมามีการร้องขอเว็บออบเจ็กต์นี้เข้ามาอีกและพร็อกซี่เซิร์ฟเวอร์พบเว็บออบเจ็กต์ที่ต้องการในแคชแล้วได้ทำการตรวจสอบแล้วว่าเว็บออบเจ็กต์ปัจจุบันที่อยู่ในอินเทอร์เน็ตยังไม่มีการเปลี่ยนแปลงขนาด และ/หรือค่า Last-Modified(ค่าการร้องขอที่ Lx_H_i) จึงทำการส่งเว็บออบเจ็กต์ที่เก็บไว้ในแคชไปให้ไคลเอนต์ที่ร้องขอมา

เมื่อมีการร้องขอเว็บออบเจ็กต์นี้เข้ามาอีก และพบว่าเว็บออบเจ็กต์ปัจจุบันที่อยู่ในอินเทอร์เน็ตมีการเปลี่ยนแปลงไป พร็อกซี่เซิร์ฟเวอร์ก็จะต้องติดต่อเพื่อไปเอาออบเจ็กต์นี้มาเก็บลงในแคชแทนออบเจ็กต์ตัวเดิมที่มีอยู่ก่อนหน้านี้ และทำการส่งออบเจ็กต์นั้นให้กับไคลเอนต์ (ค่าการร้องขอที่ Lx_M_{i+1}) ในกรณีนี้ จากรูปจะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สามารถบอกได้ว่าการร้องขอที่ Lx_M_{i+1} จะเป็นการร้องขอครั้งแรกสุดหลังจากที่ค่า Last-Modified ได้เกิดการเปลี่ยนแปลงไป ซึ่งจะไม่มีการร้องขอใดๆเข้ามาก่อนหน้านี้อย่างแน่นอนและค่า Last-Modified ของการร้องขอ Lx_M_{i+1} ที่บันทึกลงไปในล็อกไฟล์จะมีค่าเท่ากับค่า Last-Modified ที่ La_M_{i+1} ซึ่งเป็นค่า Last-Modified ล่าสุด ซึ่งจากรูปจะเห็นได้ว่าค่าความคลาดเคลื่อนจะสามารถเกิดขึ้นได้ในช่วงระหว่างค่าการร้องขอที่ Lx_H_i ถึง ค่า Last-Modified ที่ La_M_{i+1} เนื่องจากในช่วงดังกล่าวอาจจะมีเกิดการเปลี่ยนแปลงค่า Last-Modified ขึ้นซึ่งเราไม่สามารถทราบได้

ดังนั้นเราสามารถหาค่าความคลาดเคลื่อนที่เกิดจากการหาอายุของเว็บออบเจ็กต์ที่ได้จากการคำนวณได้ดังสมการต่อไปนี้

$$a_i = La_M_{i+1} - Lx_H_i \quad \text{..... (2.2)}$$

นิยามตัวแปร

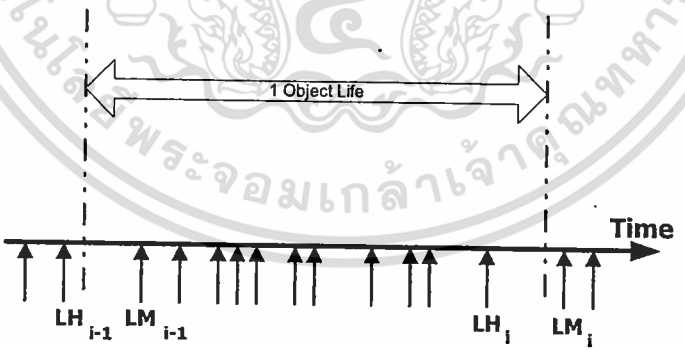
a_i คือ ค่าความคลาดเคลื่อนจากการหาอายุของเว็บออบเจ็กต์

La_M_{i+1} คือ ค่า timestamp ของ Last-Modified ที่ค่าได้เกิดการเปลี่ยนแปลงไป

Lx_H_i คือ ค่า timestamp ของการร้องขอครั้งสุดท้ายก่อนที่ค่า Last-Modified จะเกิดการเปลี่ยนแปลงไป

2.3.2.2 วิธีการคำนวณหาอายุของออบเจ็กต์แบบเว็บออบเจ็กต์ไม่มีค่า Last-Modified

เมื่อนำข้อมูลที่มีค่า Last-Modified มาคำนวณด้วยวิธีแรกแล้ว ส่วนเว็บออบเจ็กต์ที่ไม่มีค่าLast-Modified จะนำมาคำนวณด้วยวิธีที่ได้จากการวิจัย[2] ซึ่งเมื่อนำข้อมูลจากล็อกไฟล์มาแยกประเภทและนำ มาเขียนแทนด้วยลูกศรบนแกนเวลา ก็จะได้แผนภาพดังแสดงในรูปที่ 2.4



รูปที่ 2.4 แสดงการหาอายุออบเจ็กต์แบบวิธีไม่มีค่า Last-Modified

นิยามตัวแปร

LH_{i-1} คือ ค่า Timestamp ของการร้องขอครั้งสุดท้ายซึ่งเกิดขึ้นในขณะที่ออบเจ็กต์ไม่มีการเปลี่ยนแปลงทั้งขนาดและ/หรือวันที่ถูกสร้าง

LM_{i-1} คือ ค่าTimestamp ของการร้องขอเข้ามาครั้งแรกสุดหลังจากขณะที่ออบเจ็กต์เกิดการเปลี่ยนแปลงทั้งขนาดและ/หรือวันที่ถูกสร้าง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

LH_i คือ ค่า Timestamp ของการร้องขอครั้งสุดท้ายซึ่งเกิดขึ้นในขณะที่ออบเจ็กต์ไม่มีการเปลี่ยนแปลงทั้งขนาดและ/หรือวันที่ถูกสร้าง ของครั้งต่อมา

LM_i คือ ค่าTimestamp ของการร้องขอเข้ามาครั้งแรกสุดหลังจากขณะที่ออบเจ็กต์เกิดการเปลี่ยนแปลงทั้งขนาดและ/หรือวันที่ถูกสร้าง ของครั้งแต่มา

จากรูปที่ 2.4 เมื่อมีการร้องขอเว็บออบเจ็กต์เข้ามาทางซ้ายสุดของแกนเวลา ซึ่งอาจจะเป็นการร้องขอครั้งแรกที่บันทึกไว้ในล็อกไฟล์ เมื่อมีการร้องขอเว็บออบเจ็กต์นี้เข้ามาอีกและพร็อกซี่เซิร์ฟเวอร์พบเว็บออบเจ็กต์ที่ต้องการในแคช แล้วได้ทำการตรวจสอบแล้วว่าเว็บออบเจ็กต์ปัจจุบันที่อยู่ในอินเทอร์เน็ตยังไม่มีการเปลี่ยนแปลงขนาด และ/หรือ วันที่เว็บออบเจ็กต์นั้นถูกสร้าง จึงทำการส่งเว็บออบเจ็กต์ที่เก็บไว้ในแคชไปให้ไคลเอนต์ที่ร้องขอมา โดยจะกำหนดให้การร้องขอดังกล่าวนี้อยู่ในประเภทของการร้องขอแบบ Life_Hit (LH) จากนั้น การร้องขอครั้งถัดไปดังแสดงในรูป เมื่อมีการร้องขอเว็บออบเจ็กต์นี้เข้ามาอีก แต่เว็บออบเจ็กต์ปัจจุบันที่อยู่ในอินเทอร์เน็ตมีการเปลี่ยนแปลงไป พร็อกซี่เซิร์ฟเวอร์ก็จะต้องติดต่อเพื่อไปเอาเว็บออบเจ็กต์นี้มาเก็บลงในแคชแทนเว็บออบเจ็กต์ตัวเดิมที่มีอยู่ก่อนหน้านี้ และทำการส่งเว็บออบเจ็กต์นั้นให้กับไคลเอนต์ ในกรณีนี้จะกำหนดให้การร้องขอดังกล่าวอยู่ในประเภทของการร้องขอแบบ Life_Miss (LM) ในความเป็นจริงการร้องขอที่เข้ามายังพร็อกซี่เซิร์ฟเวอร์นั้น อาจมีการร้องขอเว็บออบเจ็กต์เดียวกันจากหลายไคลเอนต์ในเวลาไล่เลี่ยกัน จึงมีโอกาสของการเกิดกรณีของ Life_Hit มากกว่า Life_Miss นอกจากนี้ ยังอาจจะมีผลของการร้องขอที่ยังไม่สามารถระบุได้แน่ชัดว่าเว็บออบเจ็กต์มีการเปลี่ยนแปลงหรือไม่ ซึ่งจะจัดให้อยู่ในประเภทของการร้องขอแบบ Life_Unknown (LU) ส่วนกรณีของการร้องขอที่ไม่สำเร็จหรือไม่มีผลต่อการหาอายุและขนาดของออบเจ็กต์จะจัดอยู่ในการร้องขอแบบ Life_Ignore (LI) ซึ่งจะไม่นำมาใช้ในการคำนวณหาอายุและขนาดของเว็บออบเจ็กต์

2.3.3 ประเภทของการร้องขอที่ใช้คำนวณอายุของเว็บออบเจ็กต์

ในการที่จะแยกประเภทของการร้องขอแต่ละครั้งเพื่อนำไปใช้ในการคำนวณหาอายุของเว็บออบเจ็กต์ โดยในงานวิจัย[2] ที่ได้ศึกษาได้มีการกำหนดการแบ่งประเภทของการร้องขอออกตามผลลัพธ์ที่ได้จากการร้องขอ ซึ่งจะแสดงประเภทของการร้องขอที่มีผลต่อการคำนวณอายุ (Request method) ในส่วนของ Result Code ของแต่ละการร้องขอที่เกิดขึ้น ที่นำมาเป็นเกณฑ์ในการจัดแบ่งประเภทของการร้องขอดังแสดงในตารางที่ 2.3

ตารางที่ 2.3 ประเภทของการร้องขอที่มีผลต่อการหาอายุของออบเจ็กต์

ประเภท	ผลของการร้องขอที่ปรากฏใน Result Code ของ Squid
Life_Miss (LM)	TCP_REFRESH_MISS
Life_Hit (LH)	TCP_REFRESH_HIT, TCP_IMS_HIT, TCP_IMS_MISS
Life_Unknown (LU)	TCP_HIT,TCP_MISS, TCP_MEM_HIT, TCP_CLIENT_REFRESH_MISS, UDP_HIT, UDP_MISS
Life_Ignore (LI)	NONE, ERR_*, TCP_DENIDE, TCP_NEGATIVE_HIT, TCP_SWAPFAIL

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การร้องขอแต่ละประเภทมีความหมายดังนี้

- Life_Miss (LM) หมายถึง ผลของการร้องขอที่เกิดขึ้นหลังจากออบเจ็กต์ปัจจุบันที่อยู่ในอินเทอร์เน็ตมีการเปลี่ยนแปลงขนาดและ/หรือวัน ที่ออบเจ็กต์นั้น ๆ ถูกสร้างขึ้นมา (Created Date)
- Life_Hit (LH) หมายถึง ผลของการร้องขอที่เกิดขึ้นในขณะที่ออบเจ็กต์ปัจจุบันที่อยู่ในอินเทอร์เน็ตไม่มีการเปลี่ยนแปลงทั้งขนาดและ/หรือวันที่ถูกสร้าง
- Life_Unknown (LU) หมายถึง ผลของการร้องขอที่ยังไม่สามารถชี้ชัดเจนได้ว่าออบเจ็กต์ปัจจุบันที่อยู่ในอินเทอร์เน็ตมีการเปลี่ยนแปลงขนาดและ/หรือวันที่ถูกสร้างหรือไม่
- Life_Ignore (LI) หมายถึง ผลของการร้องขอที่เกิดขึ้นไม่เกี่ยวข้องกับการเปลี่ยนแปลงทั้งอายุและขนาดของออบเจ็กต์ในอินเตอร์เน็ต ซึ่งจะไม่นำผลการร้องขอประเภทนี้มาประมวลผลแต่อย่างใด

การแบ่งประเภทของการร้องขอที่จะนำมาใช้ในการคำนวณหาอายุของออบเจ็กต์นั้น จะถูกแบ่งออกเป็น 3 ประเภทเท่านั้น คือ Life_Miss, Life_Hit และ Life_Unknown ซึ่งในสองประเภทแรกสามารถนำไปคำนวณหาอายุได้เลย โดยจะจัดให้มีความแม่นยำของอายุที่หาให้การหาแบบที่มีค่า Last-Modified ความแม่นยำอยู่ในระดับที่ 1(Accuracy Level 1) วิธีที่ไม่ได้คำนวณด้วยค่า Last-Modified จะให้ความแม่นยำอยู่ในระดับที่ 2 (Accuracy Level 2) ทั้งนี้จะกำหนดให้ค่าอายุจริงของออบเจ็กต์มีความแม่นยำอยู่ในระดับ 0 (Accuracy Level 0) ส่วนในกรณีของผลของการร้องขอแบบ Life_Unknown นั้นจะต้องมีการตรวจสอบก่อนว่าการร้องขอออบเจ็กต์ที่ทำให้เกิดกรณีของ Life_Unknown นั้นตัวออบเจ็กต์ที่ถูกร้องขอมีการเปลี่ยนแปลงขนาดหรือไม่ ถ้ามีการเปลี่ยนแปลงขนาดจึงจะถือว่าออบเจ็กต์นั้นมีการเปลี่ยนแปลงและจะอนุโลมให้การร้องขอนี้จัดอยู่ในประเภท Life_Miss เพื่อนำไปใช้ในการคำนวณต่อไป และจะจัดให้ความแม่นยำของค่าอายุที่หาได้จากกรณี Life_Unknown นี้อยู่ในระดับที่ 3 (Accuracy Level 3)

ดังนั้นเราสามารถประมาณอายุของออบเจ็กต์ ในช่วงที่ i ($LifeCal_i$)ได้จากการนำเวลาของการเกิด Life_Miss ครั้งล่าสุด (LM_i) และ Life_Hit ครั้งล่าสุด (LH_i) มาหาค่าเฉลี่ยแล้วลบด้วยค่าเฉลี่ยของเวลาที่เกิดการเปลี่ยนแปลงของออบเจ็กต์ในช่วงก่อนหน้านั้น (LM_i-1 , LH_i-1) ดังแสดงในสมการที่ 2.3

$$Lifecal_i = \frac{(LM_i + LH_i)}{2} - \frac{(LM_{i-1} + LH_{i-1})}{2} \qquad \ldots\ldots\ldots\ldots\ldots\ldots (2.3)$$

จากรูปที่ 2.4 ข้างต้น เราสามารถคำนวณหาอายุของเว็บออบจ็กต์ได้ตามสมการ (2.3) ซึ่งเป็นการคำนวณหาอายุจากค่าเฉลี่ย โดยที่เราไม่สามารถที่จะวัดค่าอายุจริง (Lifeactual) ของเว็บออบเจ็กต์นั้นได้โดยทราบแต่เพียงว่าจะมีค่าอยู่ระหว่าง ($LH_i - LM_{i-1}$) และ ($LM_i - LH_{i-1}$)

ดังนั้นเราสามารถหาค่าความคลาดเคลื่อนที่เกิดจากการหาอายุของเว็บออบเจ็กต์แบบเว็บออบเจ็กต์ไม่มีค่า Last-Modified ที่ได้จากงานวิจัย[2] ซึ่งสามารถคำนวณได้ดังสมการที่ (2.4)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$a_i = \frac{1}{2}\left[\left(LM_{i-1} - LH_{i-1}\right) + \left(LM_i - LH_i\right)\right] \qquad \ldots\ldots\ldots\ldots\ldots\ldots \quad (2.4)$$

2.3.4 ขั้นตอนการหาอายุของออบเจ็กต์

จากวิธีการคำนวณหาอายุและขนาดที่ได้กล่าวมาแล้วข้างต้น ในการทำงานเมื่อนำมาเขียนเป็นอัลกอริธึม (Algorithm) จะเริ่มต้นจากการอ่านค่าของข้อมูลแต่ละบรรทัดที่เก็บอยู่ใน mix.log แล้วนำค่าของ Result Code มาพิจารณา

2.3.4.1 การหาอายุของออบเจ็กต์โดยที่ข้อมูลมีค่า Last-Modified

ให้นำ URL ที่ถูกร้องขอนั้นไปหาในฐานข้อมูลที่เก็บค่าระหว่างการทำงานว่ามี URL นั้นเก็บอยู่ในฐานข้อมูลหรือไม่ ถ้าไม่พบแสดงว่าเป็น URL ใหม่ที่ไม่เคยมีการเก็บข้อมูลมาก่อน ให้ทำการเก็บบันทึกข้อมูล URL และเวลาที่เกิดการร้องขอนี้ ซึ่งเป็นเวลาที่ถือว่าออบเจ็กต์นี้เกิดในช่วงอายุใหม่ (Birth) ลงในฐานข้อมูลระหว่างการทำงานและบันทึกค่า Last-Modified ของเว็บออบเจ็กต์นี้ไว้ด้วย

ถ้าพบว่ามี URL อยู่แล้วให้ทำการเปรียบเทียบค่า Last-Modified เวลาที่ออบเจ็กต์ที่เก็บไว้ในฐานข้อมูลว่าเท่ากับค่า Last-Modified ปัจจุบันหรือไม่ ถ้าเท่าแสดงว่าเว็บออบเจ็กต์เดิมยังไม่มีการเปลี่ยนแปลงค่า Last-Modified ให้ทำการเปรียบเทียบว่าขนาดของออบเจ็กต์ปัจจุบันมีค่าเท่ากับขนาดของออบเจ็กต์ที่พบในฐานข้อมูลหรือไม่ ถ้าขนาดเท่ากัน ให้ทำการเพิ่มค่าจำนวนของการร้องขอที่มีเข้ามาในช่วงอายุนี้ (Count) ในฐานข้อมูลระหว่างการทำงาน และเก็บค่าเวลาที่เกิดการร้องขอนี้ไว้เป็นเวลาล่าสุดที่มีการร้องขอออบเจ็กต์เพื่อใช้ในการคำนวณเมื่อมีการร้องขออันถัดไปเข้ามา ถ้าพบว่าขนาดของออบเจ็กต์ที่พบไม่เท่ากับขนาดของออบเจ็กต์ปัจจุบันให้บันทึกลงไฟล์ Error.log

แต่เมื่อค่า Last-Modified ปัจจุบันไม่เท่ากับค่าที่เก็บไว้ในฐานข้อมูล แสดงว่าค่า Last-Modified ได้เกิดการเปลี่ยนแปลงไปแล้ว ก็จะสามารถทำการคำนวณหาอายุของเว็บออบเจ็กต์ได้จากการนำค่า Last-Modified ที่เก็บไว้ในฐานข้อมูลลบกับค่า Last-Modified ปัจจุบัน และทำการคำนวณหาความคลาดเคลื่อนได้จากการนำค่าเวลาที่เกิดการร้องขอครั้งก่อนหน้าลบกับเวลาการร้องขอปัจจุบันที่ค่า Last-Modified ได้เปลี่ยนแปลงไป ก็จะได้ค่าความเคลื่อนเคลื่อนและทำการบันทึกลงในตารางสรุปผลอายุและขนาดของออบเจ็กต์ในส่วนของค่าความแม่นยำจะบันทึกค่าเป็น 1 และทำการปรับฐานข้อมูลระหว่างการทำงานให้มีค่า Last-Modified และค่าเวลาการร้องขอให้เปลี่ยนเป็นค่าปัจจุบัน

2.3.4.2 การหาอายุของออบเจ็กต์โดยที่ข้อมูลไม่มีค่า Last-Modified

เมื่อข้อมูลไม่เข้าวิธีการคำนวณในแบบแรกก็จะถูกนำมาแบ่งเป็นอีก 3 ประเภท โดยในแต่ละประเภทสามารถนำมาคำนวณหาอายุของออบเจ็กต์ได้ดังนี้คือ

- Life Hit Case (LH)

  ให้นำ URL ที่ถูกร้องขอนั้นไปหาในฐานข้อมูลว่ามี URL นั้นอยู่ในฐานข้อมูลระหว่างการทำงานหรือไม่ ถ้าไม่พบ แสดงว่าเป็น URL ใหม่ที่ไม่เคยมีการเก็บข้อมูลมาก่อน ให้ทำการเก็บบันทึกข้อมูล URL และเวลาลงในฐานข้อมูลที่เก็บค่าระหว่างการทำงาน ถ้าพบว่ามี URL อยู่แล้วให้ทำการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

เปรียบเทียบว่าขนาดของออบเจ็กต์ปัจจุบันมีค่าเท่ากับขนาดของออบเจ็กต์ที่พบในฐานข้อมูลหรือไม่ ถ้าขนาดเท่ากัน ให้ทำการเพิ่มค่าจำนวนของการร้องขอที่มีเข้ามาในช่วงอายุนี้ (Count) ในฐานข้อมูลระหว่างการทำงาน และเก็บค่าเวลาที่เกิดการร้องขอนี้ไว้เป็นเวลาล่าสุดที่มีการร้องขอออบเจ็กต์เพื่อใช้ในการคำนวณเมื่อมีการร้องขออันถัดไปเข้ามา ถ้าพบว่าขนาดของออบเจ็กต์ที่พบไม่เท่ากับขนาดของออบเจ็กต์ปัจจุบันให้บันทึกลงไฟล์ Error.log

- Life Miss Case (LM)

ให้นำ URL ที่ถูกร้องขอนั้นไปหาในฐานข้อมูลที่เก็บค่าระหว่างการทำงานว่ามี URL นั้นเก็บอยู่ในฐานข้อมูลหรือไม่ ถ้าไม่พบแสดงว่าเป็น URL ใหม่ที่ไม่เคยมีการเก็บข้อมูลมาก่อน ให้ทำการเก็บบันทึกข้อมูล URL และเวลาที่เกิดการร้องขอนี้ ซึ่งเป็นเวลาที่ถือว่าออบเจ็กต์นี้เกิดในช่วงอายุใหม่ (o Birth) ลงในฐานข้อมูลระหว่างการทำงาน ถ้าพบว่ามี URL อยู่แล้วให้ทำการเปรียบเทียบค่า เวลาที่ออบเจ็กต์เกิดในช่วงอายุล่าสุด

ถ้าเวลาที่ออบเจ็กต์เกิดในช่วงอายุล่าสุดมีค่าเท่ากับ 0 ซึ่งหมายความว่ายังไม่เคยบันทึกการร้องขอที่ออบเจ็กต์มีการเกิดมาก่อน ให้ทำการตรวจสอบเวลาในการร้องขอครั้งล่าสุดของออบเจ็กต์ ถ้าเวลาในการร้องขอครั้งล่าสุดของออบเจ็กต์นั้นเท่ากับ 0 (ไม่เคยมีการร้องขอออบเจ็กต์นี้มาก่อน) ให้กำหนดค่าเวลาที่เกิดการร้องขอเป็นเวลาในการเกิดของออบเจ็กต์นี้ ถ้าเวลาในการร้องขอครั้งล่าสุดของออบเจ็กต์นั้นไม่เท่ากับ 0 ให้กำหนดค่าเวลาในการเกิดของออบเจ็กต์นี้จากค่าเฉลี่ยระหว่างเวลาปัจจุบันที่มีการร้องขอนี้กับเวลาในการร้องขอครั้งล่าสุดก่อนหน้าการร้องขอปัจจุบัน

จากนั้นให้ทำการบันทึกข้อมูลของ URL นี้ในฐานข้อมูล โดยกำหนดให้ค่าเวลาในการเกิดของออบเจ็กต์ให้มีค่าเท่ากับที่คำนวณได้จากข้างต้น และกำหนดให้ค่าการร้องขอครั้งล่าสุดของออบเจ็กต์ในอายุใหม่นี้มีค่าเท่ากับเวลาปัจจุบันของการร้องขอออบเจ็กต์นี้

ถ้าเวลาที่ออบเจ็กต์เกิดในช่วงอายุล่าสุดมีค่าไม่เท่ากับ 0 ซึ่งหมายความว่าออบเจ็กต์นี้เคยเกิดมาแล้ว การเปลี่ยนแปลงครั้งนี้แสดงให้เห็นว่าออบเจ็กต์ตัวเก่าได้หมดช่วงอายุแล้ว จึงทำการคำนวณหาค่าเวลาที่ออบเจ็กต์หมดช่วงอายุ โดยหาจากค่าเฉลี่ยระหว่างเวลาปัจจุบันที่มีการร้องขอนี้กับเวลาในการร้องขอครั้งล่าสุดก่อนหน้าการร้องขอปัจจุบัน เมื่อได้เวลาในการหมดช่วงอายุของออบเจ็กต์แล้ว ก็สามารถจะนำค่าดังกล่าวไปคำนวณอายุของออบเจ็กต์ในช่วงนี้ได้ จากการนำค่าเวลาในการสิ้นอายุที่เพิ่งจะคำนวณได้ในครั้งใหม่นี้ลบด้วยเวลาในการเกิดครั้งล่าสุด ก็จะได้เป็นค่าอายุของออบเจ็กต์ตลอดช่วงอายุนี้และคำนวณหาความคลาดเคลื่อนได้จากการนำค่าเวลาการร้องขอปัจจุบันลบด้วยเวลาของการร้องขอก่อนหน้าแล้วนำมาหารด้วยสองก็จะได้เวลาของความคลาดเคลื่อนในช่วงปัจจุบัน และทำการบันทึกค่าความคลาดเคลื่อนปัจจุบันที่คำนวณได้ลงในฐานข้อมูลระหว่างการทำงาน แล้วจึงนำเอาเวลาของความคลาดเคลื่อนปัจจุบันบวกกับเวลาของความคลาดเคลื่อนก่อนหน้าในฐานข้อมูลก็จะได้เวลาของความคลาดเคลื่อนของการหาอายุของเว็บออบเจ็กต์จากการคำนวณ แล้วจึงนำค่าที่ได้ทั้งหมดบันทึกลงในฐานข้อมูล สรุปผลอายุและขนาดของออบเจ็กต์ ในส่วนของค่าความแม่นยำจะบันทึกค่าเป็น 2 ยกเว้นในกรณีที่ค่าความแม่นยำถูกกำหนดให้เป็น 3 จากกรณีที่เกิดจากการร้องขอแบบ Life Unknown

131073

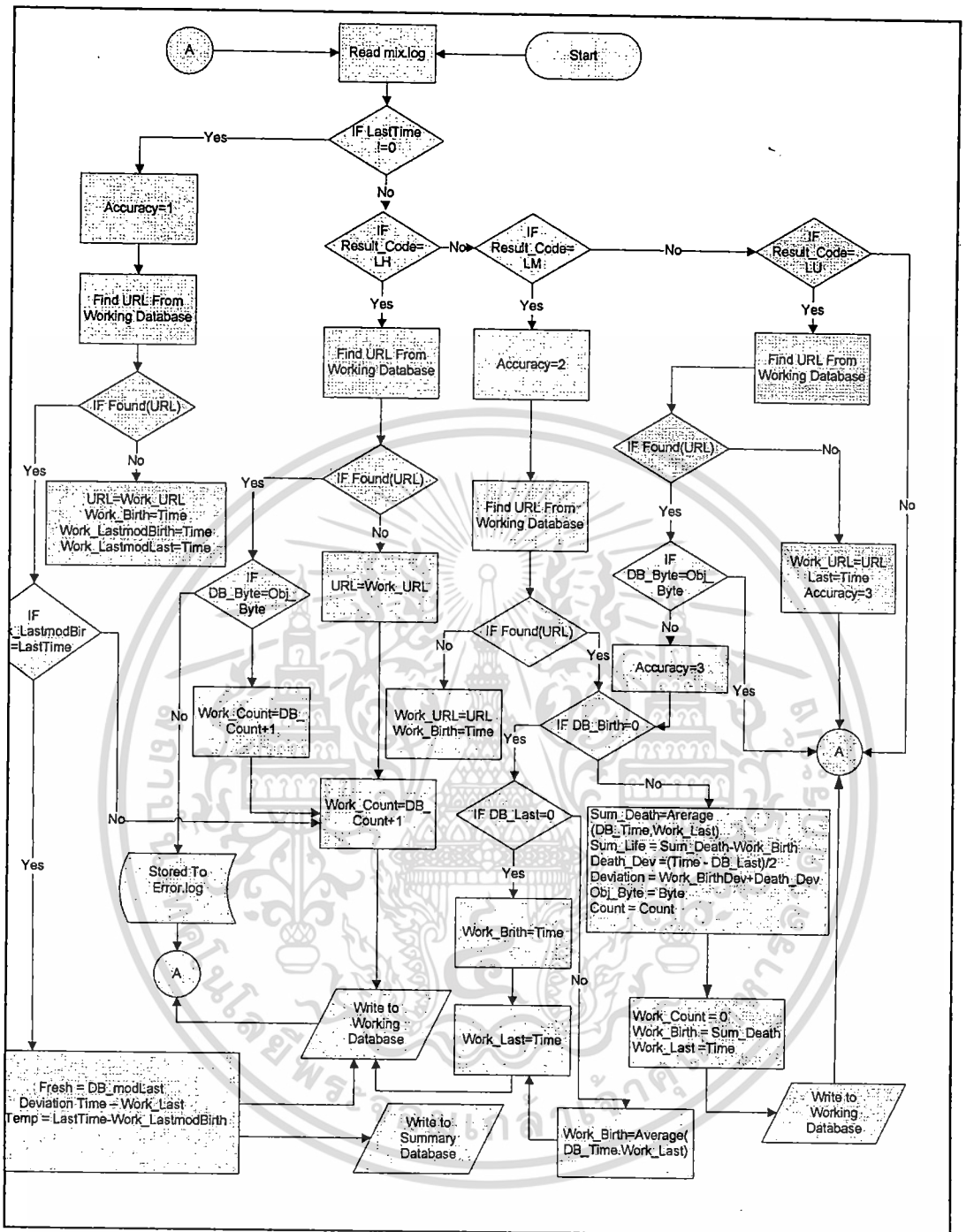
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากนั้นให้กำหนดให้เวลาในการเกิดของออบเจ็กต์ ให้มีค่าเท่ากับเวลาที่ออบเจ็กต์หมดช่วงอายุที่คำนวณได้จากข้างต้น และกำหนดให้ค่าการร้องขอครั้งล่าสุดของออบเจ็กต์ในอายุใหม่นี้มีค่าเท่ากับเวลาปัจจุบันของการร้องขอออบเจ็กต์นี้ในช่วงอายุที่เพิ่งจบไป

- Life Unknown (LU)

ให้นำ URL ที่ถูกร้องขอนั้นไปตรวจสอบว่ามี URL นั้นอยู่ในฐานข้อมูลหรือไม่ ถ้าไม่พบ แสดงว่าเป็น URL ใหม่ที่ไม่เคยบันทึกข้อมูลมาก่อน ให้ทำการบันทึกข้อมูล URL และเวลาลงฐานข้อมูล ระหว่างการทำงานและกำหนดระดับความแม่นยำเท่ากับ 3 (Accuracy Level =3) ถ้าพบว่ามี URL อยู่แล้วให้ทำการเปรียบเทียบขนาดของ URL ปัจจุบันมีค่าเท่ากับขนาดของ URL ในฐานข้อมูลหรือไม่ ถ้าขนาดเท่ากัน ให้ไปอ่านข้อมูลบรรทัดต่อไปจาก Mixlog.log และเริ่มต้นการทำงานที่จัดประเภทต่อไป ถ้าขนาดไม่เท่ากัน (มีการเปลี่ยนช่วงอายุเกิดขึ้น) ให้กำหนดระดับความแม่นยำเท่ากับ 3 (Accuracy Level =3) แล้วไปทำงานต่อเหมือนในกรณีของ Life Miss Case (LM)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.5 แผนผังแสดงกระบวนการการคำนวณหาอายุและขนาดของเว็บออบเจ็กต์ รวมแบบมีค่า Last-Modified และไม่มีค่า Last-Modified

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3
การวิเคราะห์และออกแบบระบบโปรแกรม

3.1 ความต้องการของระบบงานใหม่

3.1.1 ระบบสร้างล็อกไฟล์ใหม่ จากโปรแกรม squid

3.1.2 ระบบสามารถตรวจสอบรูปแบบและความถูกต้องของข้อมูลที่จะนำมาใช้ในการคำนวณได้

3.1.3 ระบบสามารถทำการคัดแยกประเภทของล็อกไฟล์เพื่อส่งไปทำการคำนวณด้วยวิธีที่แตกต่างกัน

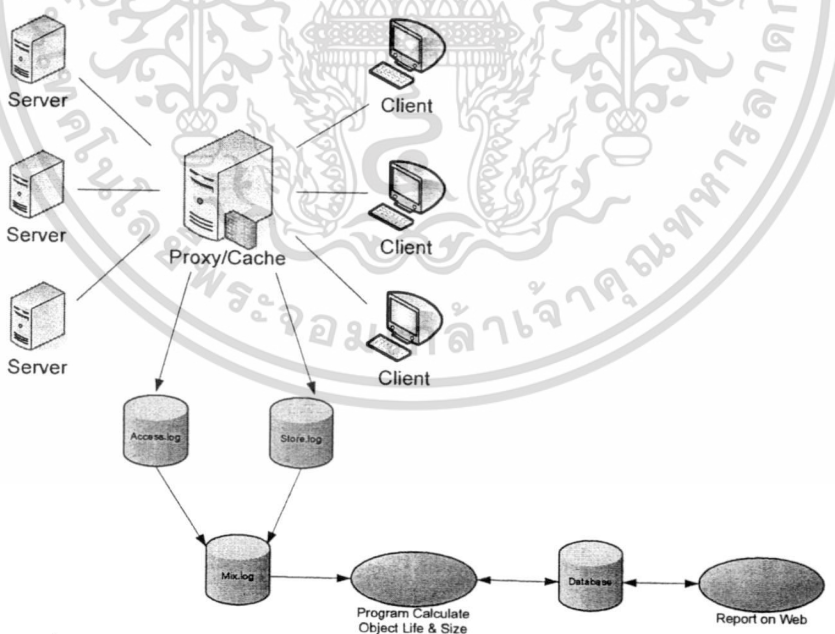
3.1.4 ระบบสามารถบอกค่าความคลาดเคลื่อนและความแม่นยำจากการคำนวณได้

3.1.5 ผู้ใช้สามารถขอดูผลการคำนวณหาอายุและขนาดของเว็บออบเจ็กต์ได้ในรูปแบบของกราฟหรือตาราง

3.1.6 ผู้ดูแลระบบสามารถสั่งให้ระบบทำการคำนวณหาขนาดและอายุของเว็บออบเจ็กต์ได้เมื่อต้องการ

3.2 ภาพรวมของระบบการหาขนาดและอายุของเว็บออบเจ็กต์จากโปรแกรม squid บนเว็บ

แสดง ดังรูปที่ 3.1



รูปที่ 3.1 แสดงภาพรวมของระบบการหาขนาดและอายุของเว็บออบเจ็กต์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3 กระบวนการในการทำงานของระบบ

จากการศึกษาถึงความต้องการของระบบ ทำให้สามารถแยกระบบได้เป็น 2 ส่วน คือ

3.3.1 การพัฒนาการสร้าง LogFile ใน Squid

โดยส่วนนี้ ได้ทำการพัฒนาโปรแกรม squid-2.6.STABLE13 ในส่วนของการสร้าง Logfile โดยให้โปรแกรมสร้าง Logfile ใหม่ขึ้นมา ชื่อ mix.log แล้วดึงข้อมูลบางส่วนจาก ไฟล์ access.log และ store.log มาใส่ในไฟล์ mix.log เพื่อนำข้อมูลไปใช้ในการคำนวณต่อไป ซึ่งในการเขียนโปรแกรมด้วยภาษา C++ จะมีการเขียน code เพิ่มเข้ามาโดยจะกำหนดให้ไฟล์ mix.log ทำการเขียนทุกครั้งที่มีการเขียนไฟล์ access.log ซึ่งจะมีการประกาศชื่อไฟล์และตัวแปรที่สร้างขึ้นใหม่ไว้ในไฟล์ต่าง ๆ และมีการแก้ไขไฟล์หลัก ๆ ด้วยกันสองไฟล์คือ ในไฟล์ store.log โดยจะสร้างฟังก์ชัน GetReplay ขึ้นมาเพื่อส่งผ่านค่า HttpReply ซึ่งภายในมีค่า Last-Modified และค่า Expires ที่เราต้องการไปยังไฟล์ access.log ดังแสดงด้วย Code Program ในรูปที่ 3.2

```
        %%%%%%%%%%%%%%%% new method. %%%%%%%%%%%%%%%%
HttpReply *
    GetReplay(MemObject *mem){
  HttpReply *reply;
#if UNUSED_CODE
  if (EBIT_TEST(e->flags, ENTRY_DONT_LOG))
    return;
#endif
  if (mem != NULL) {
            if (mem->log_url == NULL) {
                    debug(20, 1) ("storeLog: NULL log_url for %s\n", mem->url);
                    storeMemObjectDump(mem);
                    mem->log_url = xstrdup(mem->url);
            }
            reply = mem->reply;
  }
    return reply;
}
      %%%%%%%%%%%%%%%%End  new method. %%%%%%%%%%%%%%%%
```

รูปที่ 3.2 Code Program ที่เพิ่มในไฟล์ store.log

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ทำการเพิ่ม code ในไฟล์ access.log โดยเพิ่มใน method ชื่อ accessLogSquid ให้มีการส่งผ่านค่า StoreEntry e ซึ่งมีค่า Last-Modified และค่า Expires ที่เราต้องการมาเขียนในไฟล์ mix.log ซึ่งแสดงด้วย Code Program ดังแสดงในรูปที่ 3.3

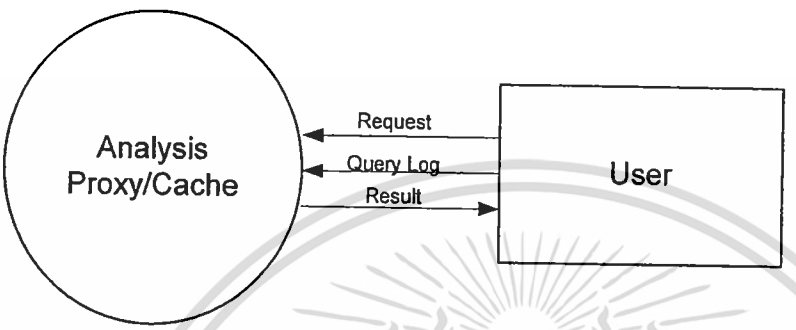
```
========== Write Mix Log ================
static void accessLogSquid(AccessLogEntry * al, Logfile * logfile,const StoreEntry * e)
.
.           // Write Access.log
.
if(e != NULL){
        reply=GetReplay(e->mem_obj);
        if(reply !=NULL){
                logfilePrintf(mixlog, "%9ld.%03d %s/%03d %" PRINTF_OFF_T " %s %s %s
%s%s/%s %s %9ld %9ld\n",
                        (long int) current_time.tv_sec,
                        (int) current_time.tv_usec / 1000,
                        log_tags[al->cache.code],
                        al->http.code,
                        al->cache.size,
                        al->private.method_str,
                        al->url,
                        user ? user : dash_str,
                        al->hier.ping.timedout ? "TIMEOUT_" : "",
                        hier_strings[al->hier.code],
                        al->hier.host,
                        al->http.content_type,
                        (long int) reply->last_modified,
                        (long int) reply->expires);
        }
} // end e not null.
        ========== End Write Mix Log ================
```

รูปที่ 3.3 Code Program ที่เพิ่มใน method ชื่อ accessLogSquid ในไฟล์ access.log

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3.2 โปรแกรมหาอายุและขนาดของเว็บออบเจ็กต์

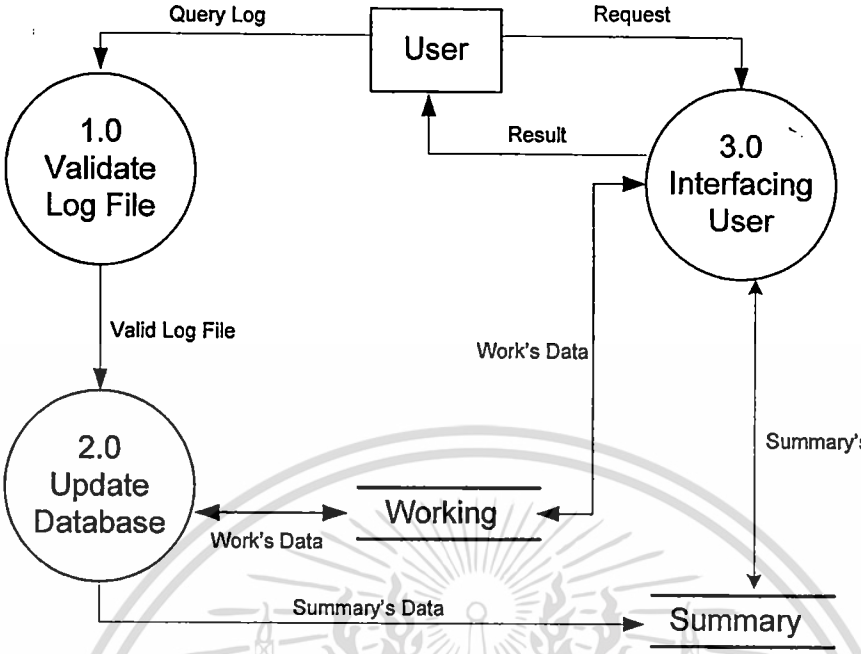
จากความต้องการของระบบโปรแกรมการหาอายุและขนาดของเว็บออบเจ็กต์ เราสามารถเขียนเป็น Context Diagram ได้ดังรูปที่ 3.4



รูปที่ 3.4 Context Diagram ของระบบ

จากแผนภาพ Context Diagram จะเห็นได้ว่าระบบติดต่อกับส่วนภายนอกระบบอยู่ 2 ส่วนหลัก ๆ คือ ส่วนของพร็อกซี่และแคชเซิร์ฟเวอร์กับส่วนของผู้ใช้ โดยในส่วนการติดต่อกับพร็อกซี่และแคชเซิร์ฟเวอร์จะเป็นการนำข้อมูลล็อกไฟล์จากพร็อกซี่และแคชเซิร์ฟเวอร์คือไฟล์ mix.log ที่ได้ทำการพัฒนาขึ้นใหม่เข้ามาเพื่อทำการวิเคราะห์และประมวลผลเพื่อหาอายุและขนาดของเว็บออบเจ็กต์ ในส่วนของผู้ใช้จะเป็นการที่ระบบติดต่อกับผู้ใช้ผ่านทางเว็บ โดยผู้ใช้จะทำการร้องขอข้อมูลที่ต้องการเข้ามาที่ระบบก่อน จากนั้นส่วนของระบบจึงทำการประมวลผลการร้องขอและส่งผลลัพธ์ที่ผ่านการวิเคราะห์และประมวลผลแล้วกลับไปให้ผู้ใช้ เพื่อให้เห็นภาพการทำงานของระบบได้ชัดเจนมากยิ่งขึ้น จึงจะทำการแสดงรายละเอียดของ Context Diagram ออกเป็นกระบวนการย่อยในรูปของแผนภาพแสดงการไหลของข้อมูล (Data Flow Diagram Level-1) ได้ดังรูปที่ 3.5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

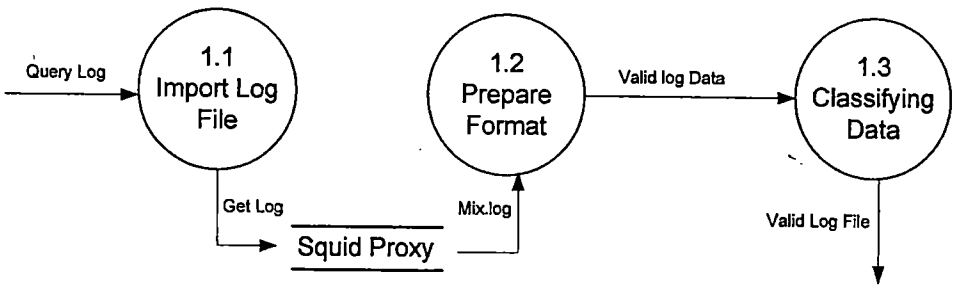


รูปที่ 3.5 Data Flow Diagram Level-1

จาก Data Flow Diagram Level-1 ในรูปที่ 3.5 จะเห็นว่าระบบงานมีกระบวนการการทำงานหลักๆ อยู่ด้วยกัน 3 กระบวนการ คือ

- กระบวนการที่ 1 เป็นกระบวนการตรวจสอบรูปแบบของข้อมูลในล็อกไฟล์และการจัดแยกประเภทข้อมูลของออบเจ็กต์ โดยในการตรวจสอบรูปแบบของข้อมูลในล็อกไฟล์จะมีการตรวจสอบว่ามีรูปแบบที่ถูกต้องตามที่กำหนดไว้หรือไม่ ถ้ามีรูปแบบที่ถูกต้องสมบูรณ์จึงจะส่งไปทำการแยกประเภทของเว็บออบเจ็กต์ตามที่ได้กำหนดไว้ ข้อมูลที่ถูกจัดแยกประเภทเสร็จแล้ว จะถูกส่งไปประมวลผลตามกระบวนการของเว็บออบเจ็กต์แต่ละประเภทยังกระบวนการต่อไป
- กระบวนการที่ 2 กระบวนการนี้จะเป็นการนำข้อมูลจากล็อกไฟล์ที่ถูกจัดแยกประเภทของเว็บออบเจ็กต์แล้ว มาทำการตรวจสอบและค้นหาข้อมูลของเว็บออบเจ็กต์ในฐานข้อมูล จากนั้นจึงเป็นการนำข้อมูลของเว็บออบเจ็กต์ที่จัดแบ่งแยกประเภทแล้วมาทำการคำนวณหาอายุและขนาด นำผลการคำนวณที่ได้ไปเก็บไว้ในฐานข้อมูล
- กระบวนการที่ 3 ส่วนใหญ่จะเป็นการทำงานบนสภาพแวดล้อมที่เป็นเว็บ ผู้ใช้จะร้องขอการแสดงรายงานเข้ามาในระบบ โปรแกรมจะจัดเตรียมและสร้างคำสั่งเพื่อไปสอบถามข้อมูลจากฐานข้อมูล นำข้อมูลที่ได้มาประมวลผลและสร้างเป็นรายงาน ก่อนที่นำรายงานที่ได้ไปแสดงให้ผู้ใช้วิเคราะห์ผลตามที่ผู้ใช้ร้องขอเข้ามา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

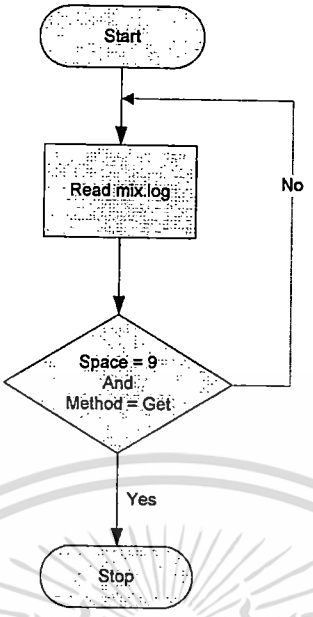


รูปที่ 3.6 Data Flow Diagram Level-2 ของกระบวนการที่ 1

จากรูปที่ 3.6 เป็นการแสดงภาพการทำงานโดยละเอียดของกระบวนการการตรวจสอบความถูกต้องของออบเจ็กต์ โดยมีลำดับกระบวนการการทำงานดังน

- กระบวนการที่ 1.1 เป็นกระบวนการการนำเข้าล็อกไฟล์จากโปรแกรม squid โดยเริ่มจากเมื่อผู้ใช้ทำการสั่งให้โปรแกรมเริ่มทำการคำนวณ จะมีการส่งคำสั่งไปเพื่อดึงข้อมูลล็อกไฟล์จากโปรแกรม squid โปรแกรมsquid ก็จะส่งล็อกไฟล์ mix.log เข้ามาเพื่อนำไปใช้ในการคำนวณต่อไป
- กระบวนการที่ 1.2 เป็นกระบวนการการจัดเตรียมรูปแบบของล็อกไฟล์ เริ่มจากการอ่านข้อมูลจากล็อกไฟล์ในแต่ละบรรทัดมาทีละตัวอักษร และนำไปเก็บไว้ในตัวแปรชนิดสตริง โดยในล็อกไฟล์จะใช้ช่องว่างเป็นตัวแยกฟิลด์แต่ละฟิลด์ ดังนั้นในการตรวจสอบรูปแบบของล็อกไฟล์จะใช้วิธีตรวจสอบจากจำนวนช่องว่างในล็อกไฟล์ว่ามีครบทั้ง 9 ช่องหรือไม่ นอกจากนี้ยังใช้วิธีตรวจสอบจากฟิลด์ Request Method ว่าเป็น GET หรือไม่ ถ้าข้อมูลในบรรทัดนั้นมีช่องว่างไม่ครบหรือเกิน 9 ช่อง ก็จะถือว่ามีรูปแบบไม่ถูกต้อง ส่วนถ้า Request Method ไม่ใช่ GET ก็ไม่มีผลต่อการหาอายุก็ไม่นำมาคำนวณเช่นกัน ให้ทำการอ่านข้อมูลของล็อกไฟล์ในบรรทัดถัดไปเข้ามาตรวจสอบต่อไป แต่ถ้าล็อกไฟล์ในบรรทัดนั้นมีรูปแบบถูกต้องและมีข้อมูลครบถ้วนตามที่กำหนดไว้ ข้อมูลที่เกี่ยวข้องกับการหาอายุและขนาดของออบเจ็กต์ ซึ่งประกอบไปด้วย TimeStamp, Result Code, Byte, Request Method, URL, Lastmod และ Expires ก็จะถูกส่งไปยังกระบวนการต่อไป ซึ่งสามารถเขียนกระบวนการในการจัดเตรียมรูปแบบของล็อกไฟล์เป็นแผนผังได้ดัง รูปที่ 3.7

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

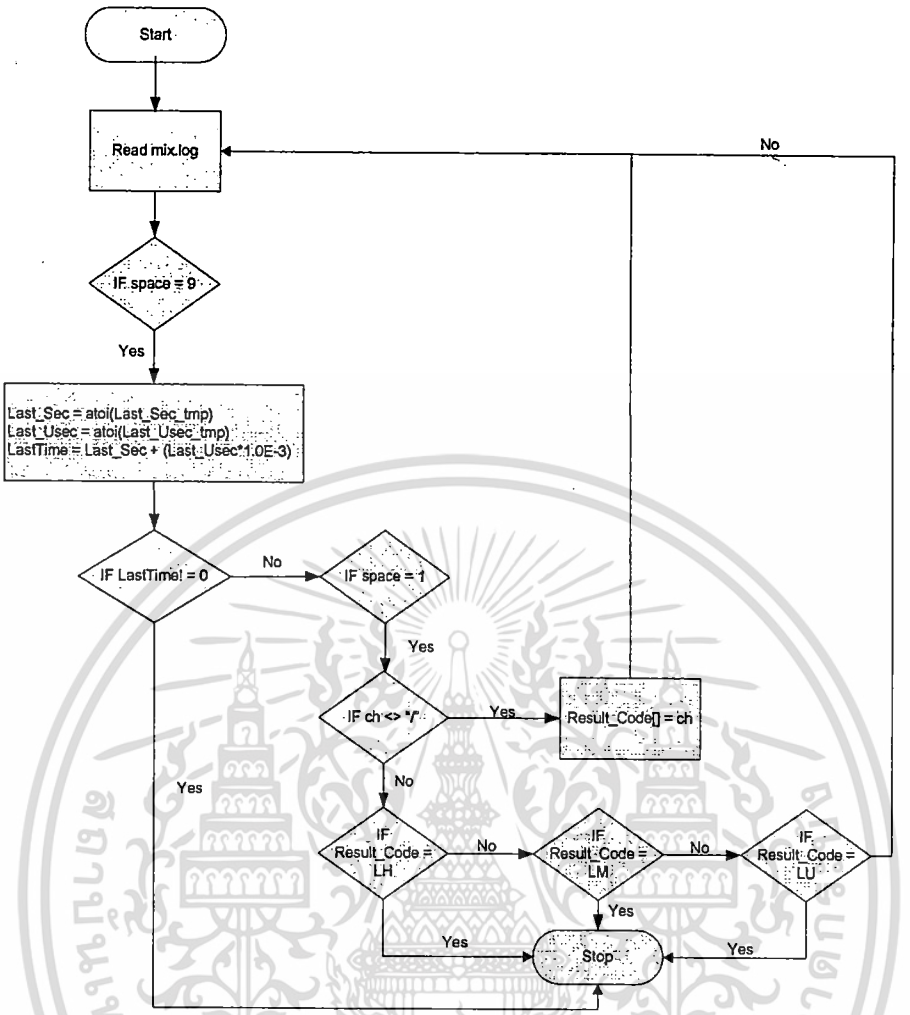


รูปที่ 3.7 แผนผังในการอ่านข้อมูลจากล็อกไฟล์เข้ามาตรวจสอบ

จากรูปที่ 3.7 ในการอ่านล็อกไฟล์ จะมีการอ่านข้อมูลมาทีละตัวอักษรและแยกฟิลด์ต่าง ๆ ไปเก็บไว้ในตัวแปรประเภทสตริง ดังนั้นเมื่ออ่านข้อมูลจบในแต่ละฟิลด์ก็จะมีเครื่องหมาย "\0" ปิดท้ายตัวแปรทุกตัว ซึ่งการแยกฟิลด์จะตรวจสอบจากช่องว่างที่คั่นในแต่ละฟิลด์โดยจะคิดเครื่องหมายช่องว่างที่คั่นอยู่เป็นเพียง 1 ช่องเท่านั้น ถึงแม้ว่าจะมีจำนวนช่องว่างที่คั่นฟิลด์นั้นมากกว่าหนึ่งช่องก็ตาม เมื่อเสร็จสิ้นกระบวนการอ่านล็อกไฟล์แล้วตัวแปรใดที่ต้องใช้ในการคำนวณ จะถูกแปลงจากสตริงให้เป็นตัวแปรชนิดตัวเลขที่สามารถใช้ในการคำนวณได้ ข้อมูลที่เกี่ยวข้องทั้งหมดจะถูกส่งต่อไปยังกระบวนการต่อไป

- กระบวนการที่ 1.3 เป็นกระบวนการของการแยกประเภทของเว็บออบเจ็กต์โดยจะทำการตรวจสอบก่อนว่าในฟิลด์ที่ส่งเข้ามามีค่า Last-Modified อยู่หรือไม่ ถ้ามีจะทำการแยกไว้เป็นประเภทมีค่า Last-Modified ส่วนที่ไม่มีจะนำมาพิจารณาแบ่งรายละเอียดเป็นอีก 3 ประเภทเท่านั้น คือ Life Hit , Life Miss และ Life Unknow ซึ่งรายละเอียดในส่วนนี้ได้มีการกล่าวไว้แล้วในบทที่ 2 ซึ่งสามารถเขียนกระบวนการในการแยกประเภทของออบเจ็กต์เป็นแผนผังได้ดังรูปที่ 3.8

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

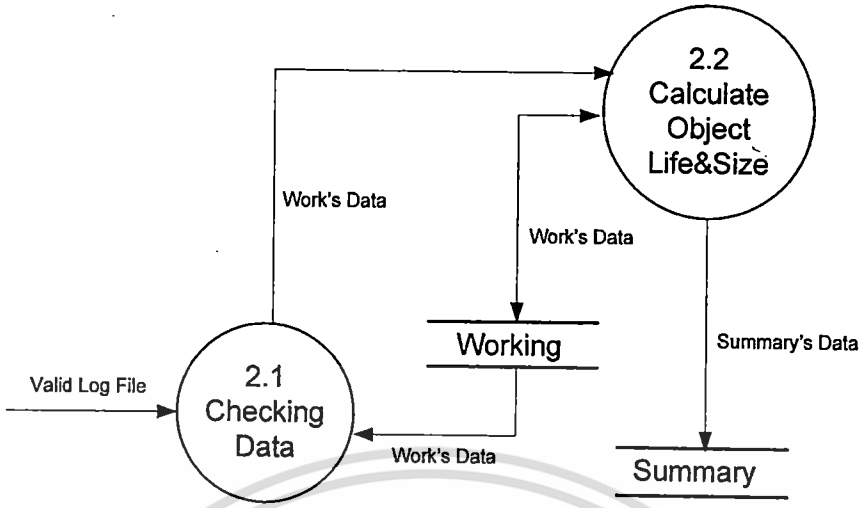


รูปที่ 3.8 แผนผังในการอ่านข้อมูลจากล็อกไฟล์เข้ามาแยกประเภท

จากรูปที่ 3.8 เป็นการแสดงรายละเอียดการแยกประเภท โดยการอ่านข้อมูลจาก ล็อกไฟล์เข้ามาทีละตัวอักษรตรวจสอบว่าเป็นฟิลด์ที่ 8 ของล็อกไฟล์หรือไม่ (space = 9 ในโปรแกรมเริ่มนับจากฟิลด์ที่ 0) ซึ่งในฟิลด์นี้เป็นฟิลด์ที่เก็บค่าของ Lastmod โดยจะถูกนำไปเก็บไว้ในตัวแปรชนิดChar ชื่อ Last_Sec_tmp และ Last_Usec_tmpแล้วถูกนำไปแปลงเป็นตัวแปรชนิด int แล้วนำไปเก็บไว้ในตัวแปรชนิด double ชื่อ LastTime และนำค่าที่ได้ไปเช็คเงื่อนไข ถ้าค่าไม่เท่ากับ 0 แสดงว่ามีค่า Lastmod ก็จะถูกแยกไปคำนวณตามวิธีดังกล่าว แต่ถ้าไม่ใช่ แสดงว่าไม่มีค่า Lastmod ก็จะนำมาพิจารณาจากการอ่านข้อมูลของฟิลด์ที่ 2 ต่อไป

ฟิลด์ที่ 2 (space = 1 ในโปรแกรมเริ่มนับจากฟิลด์ที่ 0)ของล็อกไฟล์ ซึ่งในฟิลด์นี้มีสองส่วนแยกกันด้วยเครื่องหมาย "/" ส่วน Result Code ที่ต้องการคือส่วนที่อยู่ก่อนหน้าเครื่องหมาย "/" ดังนั้นข้อมูลฟิลด์ที่ 4 ที่อยู่ก่อนเครื่องหมาย "/" จึงถูกนำไปเก็บไว้ในตัวแปรชนิดสตริงชื่อ Result_Code ซึ่งข้อมูลที่ได้จะนำมาใช้ในการแยกประเภทของออบเจ็กต์ โดยจะเป็นการตรวจสอบจากค่าตัวแปร Result_Code เปรียบเทียบกับค่าที่กำหนดไว้ว่าออบเจ็กต์นี้จัดอยู่ในประเภทใด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

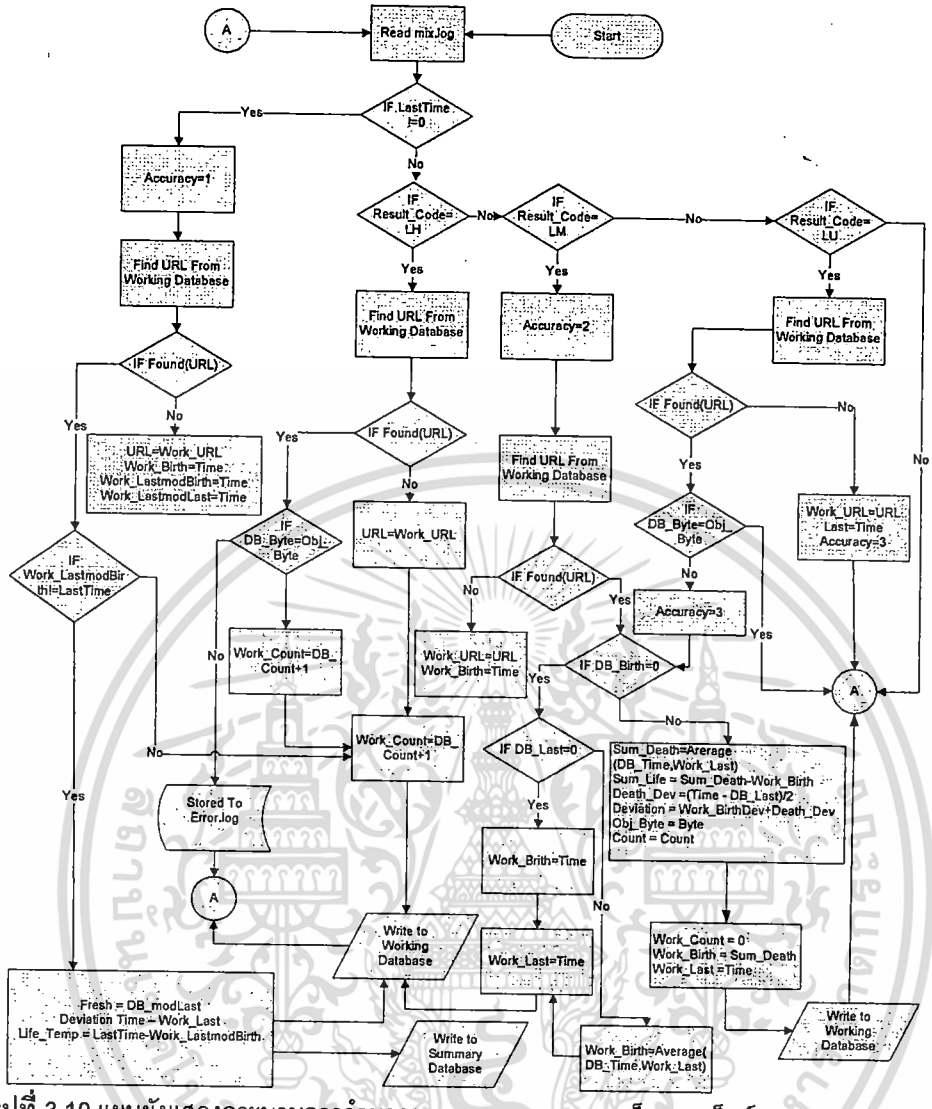


รูปที่ 3.9 Data Flow Diagram Level-2 ของกระบวนการที่ 2

จากรูปที่ 3.9 แสดง Data Flow Diagram Level-2 ของกระบวนการที่ 2 ซึ่งแสดงการทำงานโดยละเอียดในการคำนวณหาอายุและขนาดของออบเจ็กต์โดยมีกระบวนการหลัก ๆ ดังนี้

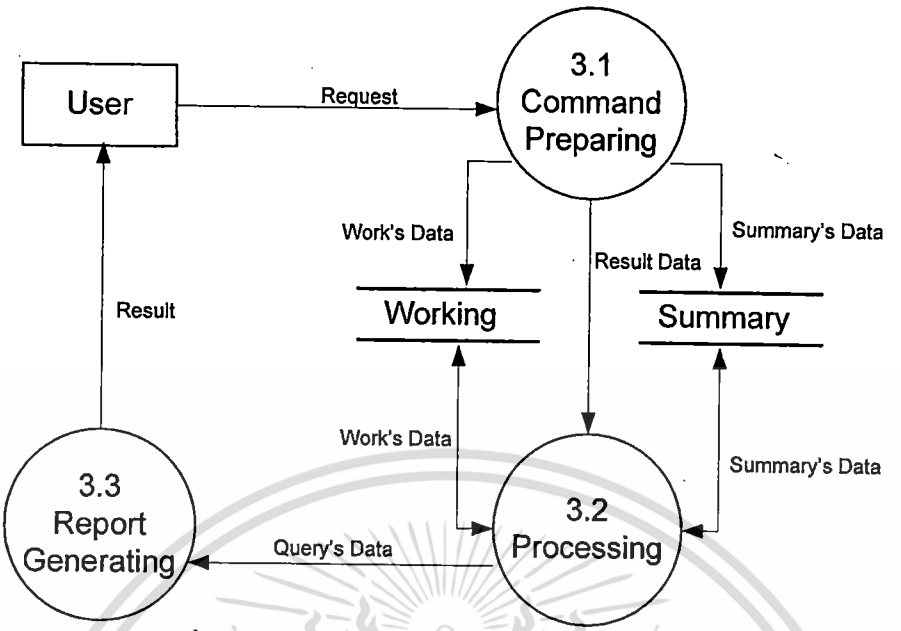
- กระบวนการที่ 2.1 เป็นการนำข้อมูลที่ผ่านการแยกประเภทแล้วมาทำการตรวจสอบและค้นหาว่ามีเว็บออบเจ็กต์นี้ในฐานข้อมูลแล้วหรือยัง โดยจะใช้ข้อมูลชื่อ URL ในการค้นหา หากค้นหาแล้วไม่พบข้อมูล URL นี้ จะทำการจัดเก็บข้อมูลของเว็บออบเจ็กต์นี้ลงในฐานข้อมูลในตารางระหว่างการทำงาน (Working) แต่ถ้าตรวจสอบแล้วพบว่ามีอยู่แล้ว จะทำการประมวลผลตามเงื่อนไขในแต่ละประเภทของเว็บออบเจ็กต์
- กระบวนการที่ 2.2 เป็นกระบวนการคำนวณหาอายุและขนาดของเว็บออบเจ็กต์ เมื่อตรวจสอบแล้วพบว่ามีข้อมูลในฐานข้อมูลอยู่แล้ว และข้อมูลเว็บออบเจ็กต์นั้นได้มีการเปลี่ยนแปลงเวลาที่ถูกสร้างข้อมูลที่ถูกจัดเก็บในตารางระหว่างการทำงาน จะถูกนำไปใช้คำนวณหาอายุและขนาดของเว็บออบเจ็กต์ตามวิธีการที่ได้กำหนดไว้โดยประเภทที่ถูกแบ่งว่ามีค่า Last-Modified จะนำมาคำนวณหาอายุด้วยสมการแบบที่ 2.1 ส่วนประเภทที่ไม่มีจะถูกนำมาคำนวณด้วยสมการแบบที่ 2.3 ตามที่ได้กล่าวไว้ในบทที่ 2 เมื่อคำนวณหาอายุและขนาดเสร็จแล้วจะนำข้อมูลบางส่วนไปปรับปรุงและจัดเก็บที่ฐานข้อมูลในตารางแสดงการทำงาน(Working) และข้อมูลของอายุและขนาดของแต่ละออบเจ็กต์ที่คำนวณได้จะถูกนำไปเก็บไว้ในตารางสรุปผล (Summary) ดังแสดงกระบวนการการคำนวณหาอายุและขนาดของออบเจ็กต์ได้ในรูปผังงานในรูปที่ 3.10

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.10 แผนผังแสดงกระบวนการคำนวณหาอายุและขนาดของเว็บออบเจ็กต์รวมแบบมีค่า Last-Modified และไม่มีค่า Last-Modified

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.11 Data Flow Diagram Level-2 ของกระบวนการที่ 3

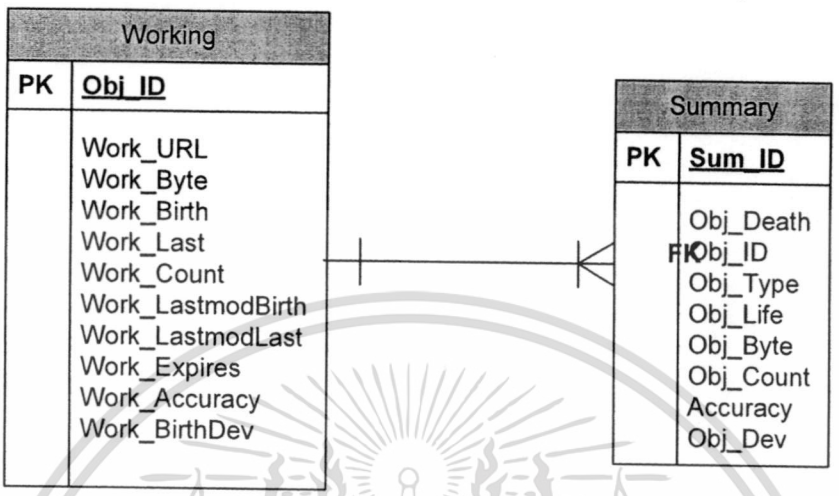
จากรูปที่ 3.11 เป็นการแสดงภาพการทำงานโดยละเอียดของกระบวนการการติดต่อกับผู้ใช้ โดยสามารถอธิบายการทำงานได้ดังนี้

- กระบวนการ 3.1 เป็นกระบวนการการตรวจสอบและจัดเตรียมคำสั่งจากการร้องขอของผู้ใช้ เมื่อผู้ใช้ติดต่อกับระบบโดยเรียกใช้โปรแกรมผ่านทางเว็บ และผู้ใช้ทำการเลือกและกำหนดข้อมูลต่าง ๆ ที่ต้องการสอบถาม จากนั้นส่งการร้องขอเข้าสู่ระบบ โปรแกรมระบบจะทำการจัดเตรียมคำสั่งและจัดเตรียมค่าต่าง ๆ เพื่อนำข้อมูลการร้องขอนั้นมาสร้างเป็นคำสั่ง SQL สำหรับติดต่อและใช้ในการสอบถามข้อมูลจากระบบจัดการฐานข้อมูล เมื่อสร้างคำสั่งเสร็จแล้วโปรแกรมจะส่งคำสั่ง SQL นี้ ไปสอบถามข้อมูลในฐานข้อมูลที่เก็บอยู่ในโปรแกรม MySQL Server เพื่อนำผลลัพธ์จากการสอบถามข้อมูลนี้ส่งไปประมวลผลยังกระบวนการต่อไป
- กระบวนการ 3.2 เป็นกระบวนการที่ใช้ในการประมวลผลข้อมูลที่ได้จากการสอบถามข้อมูล เพื่อให้ได้ข้อมูลต่าง ๆ ตามเงื่อนไขที่กำหนด โดยจะเป็นการนำข้อมูลของออบเจ็กต์มาทำการประมวลผล เช่น การประมวลผลเพื่อหาความถี่ของข้อมูลอายุและขนาดของออบเจ็กต์ ก่อนจะทำการส่งข้อมูลที่ได้ไปยังกระบวนการของการสร้างรายงาน
- กระบวนการ 3.3 เป็นกระบวนการในการสร้างและจัดทำรายงาน ข้อมูลสารสนเทศต่าง ๆ ที่ผ่านการประมวลผลในกระบวนการที่ 3.2 จะถูกส่งเข้ามายังกระบวนการนี้ และนำผลลัพธ์ที่ได้มาจัดรูปแบบที่กำหนดไว้ ทั้งในรูปแบบของกราฟ เท็กซ์ไฟล์และตาราง เช่นในส่วนของการสร้างกราฟเมื่อได้ข้อมูลมาแล้วก็จะส่งไปให้โปรแกรม JpGraph ทำการสร้างกราฟตามข้อมูลที่กำหนด ในส่วนของตารางจะนำข้อมูลที่ได้มาสร้างเป็นคำสั่ง HTML เพื่อสร้างเป็นรายงานในรูปของตารางตามที่ผู้ใช้กำหนด จากนั้นจึงส่งข้อมูลรายงานนี้ไปแสดงผลให้กับผู้ใช้ เพื่อให้ผู้ใช้ทำการวิเคราะห์ข้อมูลต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.4 การออกแบบฐานข้อมูล

Entity Relationship Diagram



รูปที่ 3.12 Entity Relationship Diagram (ER)

จากรูปที่ 3.12 ตารางการทำงาน (Working) ใช้ในการเก็บข้อมูลต่าง ๆ ของเว็บออบเจ็กต์ที่ใช้ระหว่างการทำงานโดยมีการนำ Obj_ID เข้ามาเพื่อใช้เป็นคีย์หลัก แทนการเลือกใช้ชื่อ URL เนื่องจากชื่อ URL ไม่เหมาะสมที่จะนำมาใช้เป็นคีย์หลัก เพราะชื่อ URL มีความยาวค่อนข้างมาก ทำให้การเข้าถึงข้อมูลในแต่ละบรรทัดยังทำได้ช้าลง มีการเพิ่ม Work_LastmodBirth และ Work_LastmodLast เพื่อใช้เก็บค่า Last-Modified เพื่อนำไปใช้ในการคำนวณ และมีการเพิ่ม Work_BirthDev และ Work_LastDev เข้ามาเพื่อใช้เก็บค่าความคลาดเคลื่อนเพื่อนำไปใช้ในการคำนวณ

และในตารางสรุปผล (Summary) จะใช้ Obj_IDเพื่อทำให้ข้อมูลทั้งสองตารางดังกล่าวสามารถสร้างความความสัมพันธ์ระหว่างกันได้ และมี Obj_Type เพื่อเก็บข้อมูลชนิดของออบเจ็กต์ว่าเป็นชนิด Dynamic หรือชนิด Static ทั้งนี้เพื่อเป็นการลดเวลาในการไปดึงข้อมูลจาก Work_URL ในตาราง Working มาตรวจสอบหาชนิดของออบเจ็กต์ นอกจากนี้ยังได้เพิ่ม Obj_Dev ขึ้นมา เพื่อใช้เก็บค่าความคลาดเคลื่อนที่ได้จากคำนวณได้

Data Dictionary

โดยการออกแบบฐานข้อมูล จะประกอบด้วยตารางสองตาราง คือตารางที่ใช้เก็บข้อมูลในระหว่างการทำงาน (Working Table) และ ตารางที่ใช้เก็บค่าผลสรุปอายุของออบเจ็กต์ (Summary Table) ซึ่งจะมีชื่อของข้อมูล รูปแบบการเก็บข้อมูล และ ความหมาย ดังที่ได้แสดงในตารางที่ 3.1 และตารางที่ 3.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.1 Data Dictionary ของตาราง Working

Attribute	Type	Width	Decimal	Comment	Null (Yes/No)
Obj_ID	Counter	Auto	-	Primary Key , เก็บหมายเลขเพื่อใช้ในการอ้างอิงถึงออบเจ็กต์	No
Work_URL	VarChar	255	-	เก็บค่าที่อยู่ ของออบเจ็กต์	No
Work_Byte	Unsigned	7	-	เก็บค่าขนาดของออบเจ็กต์ (Byte)	No
Work_Birth	Double	14	3	เก็บค่าเวลาที่ออบเจ็กต์เกิดครั้งล่าสุด (UTC)	No
Work_Last	Double	14	3	เก็บค่าเวลาที่ออบเจ็กต์ถูกร้องขอครั้งล่าสุด(UTC)	No
Work_Count	Unsigned	5	-	เก็บค่าจำนวนการร้องขอที่มีเข้ามาในแต่ละช่วงอายุของออบเจ็กต์	No
Work_LastmodBirth	Double	14	3	เก็บค่าLast-Modifiedที่ออบเจ็กต์เกิดการร้องขอครั้งล่าสุด	No
Work_LastmodLast	Double	14	3	เก็บค่าLast-Modifiedที่ออบเจ็กต์ถูกร้องขอครั้งล่าสุด	No
Work_Expires	Double	14	3	เก็บค่าเวลาที่ออบเจ็กต์หมดอายุ	No
Work_Accuracy	Unsigned	1	-	เก็บค่าความแม่นยำในการคำนวณหาอายุซึ่งขึ้นอยู่กับประเภทของการร้องขอ (ค่าเท่ากับ0 คือ อายุจริงของออบเจ็กต์, 1 คือ อายุที่ได้จากการคำนวนแบบมีค่า Last-Modified , 2 คือ อายุที่ได้จากการคำนวณแบบไม่มีค่าLast-	No

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

				Modified)	

ตารางที่ 3.1 Data Dictionary ของตาราง Working (ต่อ)

Attribute	Type	Width	Decimal	Comment	Null (Yes/No)
Work_BirthDev	Double	14	3	เก็บเวลาความคลาดเคลื่อนที่ออบเจ็กต์เกิดครั้งล่าสุด	No

ตารางที่ 3.2 Data Dictionary ของตาราง Summary

Attribute	Type	Width	Decimal	Comment	Null (Yes/No)
Sum_ID	Counter	Auto	-	Primary Key	No
Obj_Death	Date/ Time	16	-	เก็บค่าเวลาที่ออบเจ็กต์หมดช่วงอายุ โดยมีรูปแบบเป็นดังนี้ (YYYY-MM-DD HH:MM:SS)	No
Obj_ID	Double	8	-	Foreign Key	No
Obj_Type	Byte	1	-	เก็บชนิดของออบเจ็กต์	No
Obj_Byte	Unsigned	7	-	เก็บขนาดของออบเจ็กต์ (Byte)	No
Obj_Life	Double	7	-	เก็บอายุของออบเจ็กต์ (Second)	No
Obj_Count	Unsigned	5	-	เก็บจำนวนการร้องขอที่มีเข้ามาตลอดช่วงอายุของออบเจ็กต์	No
Accurary	Unsigned	1	-	เก็บค่าสรุปความแม่นยำของอายุที่คำนวณได้ (ระดับ 1, 2,3)	No
Obj_Dev	Double	7	-	เก็บค่าความคลาดเคลื่อนของการคำนวณ	No

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4
การพัฒนาระบบโปรแกรม

ในบทนี้จะกล่าวถึงรายละเอียดของการออกแบบโปรแกรม ขั้นตอนการพัฒนา เครื่องมือต่าง ๆ ที่ใช้ในการพัฒนาโปรแกรม รวมถึงการทดสอบโปรแกรมด้วย

4.1 การออกแบบโปรแกรม

การทำงานแบ่งได้เป็น 3 ส่วนหลักๆ ส่วนแรกเป็นส่วนของการแก้ไขปรับปรุงโปรแกรม Squid เพื่อสร้าง Log File ใหม่ขึ้นมาโดยระบบงานเดิมไม่ได้มีการทำในส่วนนี้ไว้ จึงต้องทำการพัฒนาขึ้นมาใหม่ทั้งหมด ส่วนที่สอง คือส่วนของการพัฒนาโปรแกรมสำหรับคำนวณหาอายุและขนาดของออบเจ็กต์ จะต้องทำหน้าที่อ่านข้อมูลที่ต้องใช้จาก Log File ที่สร้างใหม่จากโปรแกรม Squid ดังนั้นระบบงานเดิมจึงออกแบบให้โปรแกรมส่วนนี้ทำงานบน UNIX โดยจะรับข้อมูลเข้ามาจากทาง Standard Input ของ UNIX ซึ่งจะใช้ภาษา C++ ในการพัฒนา เนื่องจากโปรแกรมที่ได้จะมีความเร็วในการประมวลผลที่ดี เพื่อรองรับกับข้อมูลจำนวนมหาศาลที่โปรแกรมต้องทำการคำนวณ

โดยการทำงาน จะต้องทำงานในเครื่องเดียวกันกับโปรแกรมพร็อกซี่และแคชเซิร์ฟเวอร์ หรือเครื่องที่มีล็อกไฟล์ เพื่อทำการ Redirection ล็อกไฟล์ mix.log เข้าสู่โปรแกรม โดยในการทำงานจะทำการรับข้อมูล ผ่านทาง Standard Input ของระบบปฏิบัติการ UNIX เพื่อทำการประมวลผลก่อนที่จะนำข้อมูลที่ได้ส่งไปให้กับระบบจัดการฐานข้อมูล MySQL ซึ่งทำหน้าที่จัดเก็บและจัดการส่งค่าที่ต้องการใช้ในการประมวลผลให้กับโปรแกรมผ่านทาง MySQL API ซึ่งงานในส่วนนี้ทำงานได้ดีในระดับหนึ่งแล้ว ดังนั้นโปรแกรมส่วนนี้จะยังใช้ระบบเดิมเป็นหลัก แต่อาจต้องมีการปรับปรุงแก้ไขบางส่วน โดยในเรื่องการอ่านข้อมูลจากล็อกไฟล์ที่ต้องมีการแก้ไขเพิ่มเติมสำหรับรูปแบบของ LogFile ที่ได้สร้างขึ้นใหม่ เพื่อให้การตรวจสอบรูปแบบของล็อกไฟล์มีความถูกต้องก่อนจะนำมาประมวลผลทุกครั้ง

ในส่วนของการคำนวณได้มีการเขียนโปรแกรมเพิ่มเติมให้มีวิธีการคำนวณจากค่าLast-Modified เพิ่มเติมจากเดิม และมีการคำนวณหาความคลาดเคลื่อนซึ่งระบบเดิมไม่ได้มีการคำนวณในส่วนนี้ และส่วนสุดท้าย ส่วนที่สามคือส่วนที่ติดต่อกับผู้ใช้เพื่อการแสดงผลและสร้างรายงานซึ่งจะอยู่ในสภาพแวดล้อมแบบเว็บ โดยผู้ใช้สามารถทำงานได้จากเครื่องใดก็ได้ที่มีการเชื่อมต่อเครือข่ายอินเทอร์เน็ต ทำให้ผู้ใช้สามารถสอบถามข้อมูลต่าง ๆ จากฐานข้อมูลได้สะดวกและรวดเร็วยิ่งขึ้น

ซึ่งในการพัฒนาโปรแกรมในส่วนของการออกแบบและพัฒนาระบบได้ทำการพัฒนาไม่ต่างจากระบบเดิมมากนักเพราะระบบเดิมก็ทำได้ดีอยู่แล้ว แต่ก็จะมีการเปลี่ยนแปลงบางเว็บเพจให้สวยงามและสะดวกในการใช้งานมากยิ่งขึ้น และ ในส่วนของเครื่องมือที่ใช้ในการพัฒนาในส่วนนี้ ได้เลือกใช้ภาษาสคริปต์ PHP ในการพัฒนาเพราะสามารถพัฒนาบนเว็บได้ง่าย มีความรวดเร็วในการเข้าถึงข้อมูลในฐานข้อมูล MySQL มี API ติดต่อกับ MySQL ได้โดยตรงโดยไม่จำเป็นต้องติดตั้ง API เพิ่มเติมและติดต่อกับ MySQL โดยไม่ต้องผ่านตัวกลาง (ODBC) จากการที่พัฒนาโปรแกรมส่วนสร้างรายงานบนสภาพแวดล้อมที่เป็นเว็บนี้ ยังเป็นข้อดีที่ทำให้ผู้ใช้สามารถทำงานบนคอมพิวเตอร์เครื่องเดียวกันกับโปรแกรมส่วนที่ใช้คำนวณหาอายุและขนาดได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยการทำงานจะต้องมีการติดตั้งเว็บเซิร์ฟเวอร์ให้เรียบร้อยก่อน จากนั้นติดตั้งโปรแกรมที่ใช้ในการสร้างและแสดงรายงานไว้ในไดเรกทอรีใดก็ได้ที่สามารถเรียกใช้งานผ่านทางเว็บได้ จากนั้นเมื่อผู้ใช้เรียกโปรแกรมผ่านทางเว็บบราวเซอร์จะปรากฏหน้าจอหลักของโปรแกรม โดยผู้ใช้จะทำการเลือกและกำหนดข้อมูลต่าง ๆ ตามต้องการ และเมื่อผู้ใช้ส่งการร้องขอนี้ผ่านทางโปรแกรม โปรแกรมจะนำข้อมูลต่าง ๆ มาสร้างเป็นชุดคำสั่งเพื่อไปติดต่อกับฐานข้อมูล MySQL หลังจากนั้นข้อมูลจากฐานข้อมูลจะถูกส่งกลับมาที่โปรแกรมเพื่อนำข้อมูลมาสร้างเป็นรายงานในรูปแบบต่าง ๆ ตามที่ผู้ใช้ต้องการ

4.2 เครื่องมือที่ใช้ในการพัฒนาโปรแกรม

เครื่องมือที่สำคัญและมีส่วนช่วยให้การพัฒนาโปรแกรมระบบงานนี้ มีประสิทธิภาพและได้รับความสะดวกในการพัฒนามากยิ่งขึ้น มีดังต่อไปนี้

4.2.1 ระบบปฏิบัติการ FreeBSD เวอร์ชั่น 5.5

ระบบปฏิบัติการ FreeBSD เป็นฟรีเวอร์ชั่น ของ Berkeley UNIX เป็นระบบปฏิบัติการ UNIX ที่มีประสิทธิภาพสูงเหมาะสำหรับใช้เป็น Internet หรือ Intranet เซิร์ฟเวอร์ เนื่องจากมีความเสถียรภาพในการให้บริการทางด้านเครือข่ายสูง มีการจัดสรรการใช้หน่วยความจำที่ดี มีระบบรักษาความปลอดภัยสูง ให้เวลาในการตอบสนองต่อผู้ใช้ได้ดี นอกจากนี้ยังสามารถใช้ได้กับคอมพิวเตอร์ในหลายตระกูลเช่น x86 ,DEC Alpha, IA-64, PC-98 และ UltraSPARC เป็นต้น

4.2.2 GNU C Compiler เวอร์ชั่น 2.95.4

ภาษาซี เป็นภาษาคอมพิวเตอร์ที่คู่กับระบบปฏิบัติการ UNIX มานาน โดยมีจุดเด่นที่สามารถติดต่อกับฮาร์ดแวร์ได้ดี ส่งผลให้ภาษาซีทำงานได้เร็วใกล้เคียงกับภาษาระดับต่ำ ซึ่ง GNU C Compiler เป็นคอมไพล์เลอร์ในระบบปฏิบัติการ UNIX และเสมือน UNIX ที่แจกจ่ายให้ใช้ฟรีตามข้อตกลง GNU (General Public License) และยังคอมแพทิเบิลกับมาตรฐาน ANSI C อีกด้วย

4.2.3 เว็บเซิร์ฟเวอร์ Apache เวอร์ชั่น 2.2.0.7

มีต้นกำเนิดมาจากโปรแกรม NCSA httpd1.3 ได้รับการปรับปรุงและพัฒนาอย่างต่อเนื่อง จนอาจถือได้ว่าเป็นเว็บเซิร์ฟเวอร์ที่ดีที่สุดบน UNIX ในปัจจุบัน Apache เป็นเว็บเซิร์ฟเวอร์ที่ทำงานได้เร็ว มีความน่าเชื่อถือได้สูง และมีความสามารถอื่น ๆ ที่หลากหลาย เช่นรองรับโปรโตคอลมาตรฐาน HTTP/1.1 มีระบบโมดูลให้ผู้ใช้เขียนโปรแกรมเพื่อเพิ่มเติมความสามารถให้กับเว็บเซิร์ฟเวอร์ได้ นอกจากนี้ยังมีระบบ Database for authentication โดยสามารถกำหนดรหัสผ่านสำหรับอนุญาตเฉพาะผู้ใช้ที่มีสิทธิ์เข้าชมเพจต่าง ๆ ของโฮมเพจแต่ละหน้าได้

4.2.4 ระบบจัดการฐานข้อมูล MySQL Server เวอร์ชั่น 5

เป็น DBMS ที่ใช้ในการจัดการฐานข้อมูลเชิงสัมพันธ์ (Relational Database) ทำงานในรูปแบบไคลเอนต์/เซิร์ฟเวอร์ ซึ่งมีขนาดเล็กแต่มีประสิทธิภาพและความเร็วในการประมวลผลสูง ความสามารถโดยทั่วๆ ไปจะครอบคลุมความต้องการของโปรแกรมระบบงานนี้อย่างเพียงพอ จัดเป็นระบบฐานข้อมูลประเภท SQL-Base

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยที่ผู้ใช้สามารถใช้คำสั่ง SQL ในการสั่ง หรือใช้งานได้โดยไม่ต้องศึกษาคำสั่งเพิ่มเติมแต่อย่างใด นอกจากนี้ยังสนับสนุน API เพื่อใช้งานกับโปรแกรมอื่น ๆ มากมาย และยังสามารถรองรับข้อมูลขนาดใหญ่ได้อีกด้วย ในการทำงานกับระบบงานนี้จำเป็นต้องมีการติดตั้งในส่วนของโปรแกรมที่เป็นเซิร์ฟเวอร์ก่อน หลังจากนั้นจึงทำการสร้างฐานข้อมูล, ตาราง และ กำหนดรูปแบบของข้อมูลที่ต้องการจัดเก็บในตารางรวมถึงคีย์หลักตามที่ได้ออกแบบไว้ โดยในการสร้างนี้จะใช้วิธีสร้างจากสคริปต์ที่ได้เตรียมไว้ เมื่อสร้างฐานข้อมูลต่าง ๆ เสร็จแล้ว ในการที่จะสั่งงานด้วยภาษาซีเพื่อจัดเก็บข้อมูลในฐานข้อมูล จำเป็นจะต้องติดตั้งชุดคำสั่งมาตรฐานที่ใช้ในการทำงานพัฒนาโปรแกรมภาษาซีร่วมกับฐานข้อมูล MySQL ด้วย

4.2.5 โปรแกรม MySQL++ API เวอร์ชั่น 2.1.1

เป็นโปรแกรมที่ประกอบด้วยชุดคำสั่งมาตรฐาน ที่ใช้ในการติดต่อส่งคำสั่งที่ต้องการไปยัง MySQL Server เพื่อให้การพัฒนาโปรแกรมด้วย C++ ทำได้ง่ายขึ้นและใช้เวลาน้อยลง เพราะเพียงแต่เรียกใช้คำสั่งมาตรฐานที่ถูกพัฒนาไว้แล้วเท่านั้น ก็ทำให้ C++ สามารถติดต่อกับ MySQL Server ได้แล้ว โดยในโปรแกรมระบบงานนี้จะใช้ MySQL++ เฉพาะในส่วนของโปรแกรมที่ใช้ในการคำนวณหาอายุและขนาดของออบเจ็กต์ เพื่อการติดต่อกับภาษา C++ กับMySQL Server เท่านั้น ส่วนโปรแกรมในส่วนการสร้างรายงาน PHP มี API ในการติดต่อกับ MySQL Server ในตัวอยู่แล้วไม่ต้องติดตั้ง API เพิ่มแต่อย่างใด

ในการใช้งาน MySQL++ ในระบบปฏิบัติการ FreeBSD จำเป็นจะต้องมีการคอมไพล์โปรแกรมภาษาซีด้วยพารามิเตอร์บางอย่างเพิ่มเติม ซึ่งแตกต่างจากการคอมไพล์โปรแกรมภาษาซีทั่วๆ ไป เพื่อเป็นการแก้ปัญหาการคอมไพล์โปรแกรมที่เกิดขึ้นระหว่าง MySQL++ กับ ระบบปฏิบัติการ FreeBSD ซึ่งในตัวอย่างข้างล่างนี้เป็นตัวอย่างการคอมไพล์โปรแกรมที่ชื่อว่า objcal.cpp ที่ใช้ MySQL++ ในระบบปฏิบัติการ FreeBSD

```
# c++ -D_FIX_FOR_BSD_ -I/usr/local/include/mysql -L/usr/local/lib -lsqlplus objcal.cpp -o
objcal
```

4.2.6 โปรแกรมภาษาสคริปต์ PHP เวอร์ชั่น 5 และ GD Library เวอร์ชั่น 2.0.1

PHP เป็นโปรแกรมภาษาสคริปต์ที่ทำงานบน Web Server เพื่อใช้ในการจัดการข้อมูล ประมวลผลข้อมูลสำหรับเว็บไซต์ PHP ทำให้การเขียนโปรแกรมบนเว็บไซต์ง่ายกว่าการใช้ CGI มีการ Complie และ Execute ได้อย่างรวดเร็ว ทำงานได้หลาย Platform สามารถดาวน์โหลดได้ฟรีและยังเข้าถึงข้อมูลได้หลายประเภทรวมทั้ง MySQL Server ที่โปรแกรมระบบงานนี้ใช้ด้วย นอกจากนี้ PHP ยังมีความสามารถในการสร้างภาพกราฟิกได้อีกด้วย โดยเป็นการสร้างแบบ On the Fly คือ การสร้างแบบลอย ๆ การที่เราจะใช้ภาษาสคริปต์ PHP สร้างภาพกราฟิกได้นั้นต้องใช้ Graphic Library หรือ GD ร่วมด้วยโดยในปัจจุบัน GD Library มีความสามารถในการรองรับภาพแบบ jpeg ได้แล้ว ซึ่งเราจะนำความสามารถส่วนนี้มาสร้างรายงานในรูปแบบกราฟต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.7 โปรแกรม JpGraph เวอร์ชั่น 2.1.4

ซึ่งในการสร้างกราฟจะใช้โปรแกรม JpGraph ซึ่งพัฒนาจาก PHP มาเป็นเครื่องมือที่ช่วยในการสร้างกราฟได้อย่างมีประสิทธิภาพ โปรแกรม JpGraph มีความสามารถสร้างกราฟที่กำหนดสเกลหลายแบบ สนับสนุนการสร้างรูปภาพหลายรูปแบบไม่ว่าจะเป็น jpeg, gif หรือ png นอกจากนี้ยังมีความสามารถอื่น ๆ อีกมากมายที่เป็นประโยชน์ต่อการใช้สร้างกราฟ ซึ่งโปรแกรม JpGraph ต้องการ PHP เวอร์ชั่น 4.02 ขึ้นไปเท่านั้น ดังนั้นควรตรวจสอบด้วยว่า PHP ที่ติดตั้งไว้ในระบบเป็นเวอร์ชั่นใด ในส่วนการติดตั้งผู้ใช้ไม่ต้องติดตั้งเพิ่มเติมเพราะ JpGraph จะถูกรวมเข้ากับส่วนโปรแกรมสร้างรายงานอยู่แล้ว

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การพัฒนาโปรแกรมคำนวณอายุและขนาดของเว็บออบเจ็กต์

ในการพัฒนาโปรแกรมคำนวณอายุและขนาดของเว็บออบเจ็กต์นี้ จะทำการพัฒนาบนระบบปฏิบัติการ FreeBSD และใช้ C Compiler GNU ที่มีแพ็กเกจให้ใช้ในระบบปฏิบัติการ FreeBSD อยู่แล้ว

4.2.8 รูปแบบการทำงานของโปรแกรมหาอายุและขนาดของออบเจ็กต์

โปรแกรมหาอายุและขนาดของออบเจ็กต์นี้จะทำงานบนระบบปฏิบัติการ FreeBSD โดยทำหน้าที่คำนวณอายุและขนาดของเว็บออบเจ็กต์ที่ถูกร้องขอผ่านโปรแกรม Squid โดยอาศัยการอ่านข้อมูลประเภทตัวอักษรจากไฟล์ mix.log ซึ่งแสดงให้เห็นถึงผลของการร้องขอออบเจ็กต์ที่เกิดขึ้นจากการใช้งานของผู้ใช้ โดยในการพัฒนาจะเริ่มต้นจากการติดตั้งเครื่องมือที่ต้องใช้บน UNIX ทั้งหมดให้เรียบร้อย โดยประกอบด้วยโปรแกรมระบบจัดการฐานข้อมูล MySQL และโปรแกรม MySQL ++ ซึ่งเป็น API ที่ใช้ร่วมกับตัวแปรภาษา C++ หลังจากติดตั้งเรียบร้อยแล้วจึงเริ่มทำการเขียนโปรแกรมโดยใช้ภาษาซี ซึ่งในการทำงานของโปรแกรมจะทำการอ่านไฟล์ mix.log โดยใช้กลไกการรับข้อมูลผ่านทาง Standard Input ของระบบปฏิบัติการที่จะส่งข้อมูลแบบตัวอักษรในรูปแบบสายของข้อมูล (Streaming Data) ไปให้โปรแกรมทำการตรวจสอบความผิดพลาดของข้อมูลที่รับมา หลังจากนั้น จึงทำการแยกประเภทของการร้องขอจากข้อมูลในส่วนของ Result Code เพื่อจัดการตามประเภทของการร้องขอนั้นตามวิธีการวัดอายุและขนาดของออบเจ็กต์ที่ได้กล่าวถึงในบทที่ 2

โดยในการติดต่อกับระบบจัดการฐานข้อมูลจะอาศัยคำสั่งจาก MySQL++ ในการทำงาน หลังจากการทำงานทั้งหมดเสร็จสิ้น ข้อมูลทั้งเวลาที่คำนวณอายุ, ที่อยู่ของออบเจ็กต์, อายุ, ขนาด, จำนวนครั้งที่ออบเจ็กต์ถูกร้องขอ, ค่าความแม่นยำ และ ค่าความคลาดเคลื่อน จะถูกบันทึกในฐานข้อมูลเพื่อรอการนำไปใช้งานต่อไป

โดยหลังจากทำการสร้างโปรแกรมเสร็จแล้วและสามารถนำไปใช้งานได้ เริ่มจากการนำ Source Code ไปคอมไพล์ ก่อนที่จะสั่งให้โปรแกรมทำงานต้องทำการเริ่มต้นเดมอนโปรเซสของ MySQL ที่มีฐานข้อมูลและโครงสร้างตารางตามที่ออกแบบไว้อยู่ด้วย หากไม่มีจะทำให้โปรแกรมทำงานผิดพลาด ตัวอย่างคำสั่งที่ใช้ในการสั่งให้โปรแกรมทำงานมีดังต่อไปนี้

```
# ./objcal < /[Squid home directory]/mix.log
```

4.2.9 โปรแกรมสร้างและแสดงรายงานการหาอายุและขนาดของออบเจ็กต์

เป็นโปรแกรมที่ทำหน้าที่สร้างรายงานในรูปแบบของ กราฟ ตารางและเท็กซ์ไฟล์ เพื่อใช้ในการวิเคราะห์ผลที่ได้จากการคำนวณเพื่อนำไปใช้ในการศึกษาถึงลักษณะเฉพาะของออบเจ็กต์ในเครือข่าย นอกจากนี้ค่าอายุและขนาดของออบเจ็กต์ที่หาได้สามารถนำไปเปลี่ยนแปลงการตั้งค่าปฏิบัติการของโปรแกรมได้ เช่น ขนาดของแคช การหมดเวลาของออบเจ็กต์ในแคช และค่าอื่นที่เกี่ยวข้อง โดยในการทำงานสามารถจะกำหนดขอบเขตของข้อมูลที่ต้องการได้ ทั้งในด้านอายุ ขนาด เวลา และ ความแม่นยำ ในการพัฒนาโปรแกรมในส่วนนี้จะใช้ภาษาสคริปต์ PHP เพื่อสร้างเป็นส่วนติดต่อระหว่างผู้ใช้กับระบบ ซึ่งในการทำงานภาษาสคริปต์ PHP จะเป็นเครื่องมือที่ใช้สร้างคำสั่ง SQL เพื่อสอบถามข้อมูลไปยังระบบจัดการฐานข้อมูล MySQL เมื่อได้รับข้อมูลที่สอบถามกลับมา PHP ก็จะสร้างเป็นรายงานตามรูปแบบและเงื่อนไขต่าง ๆ ตามที่ผู้ใช้ได้กำหนดไว้ ซึ่งในโปรแกรมจะมีเมนูหลักอยู่ 3 เมนู คือ Home เป็นหน้าจอหลักที่ใช้ในการสอบถามข้อมูล Downloads เป็นเมนูที่ใช้ในการ Downloads เท็กซ์ไฟล์ที่ได้สร้างไว้ Help เป็นเมนูแสดงวิธีใช้โปรแกรมโดยสังเขป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

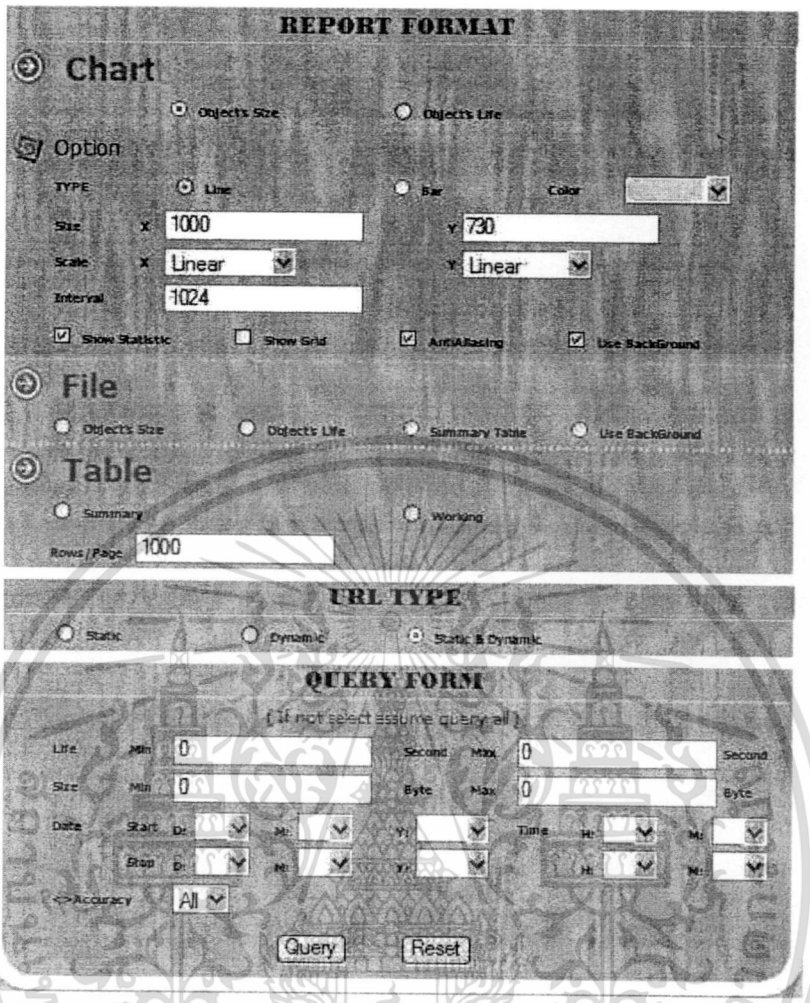
เมนู Home ซึ่งเป็นหน้าจอหลักซึ่งถือว่าเป็นส่วนสำคัญของโปรแกรมสร้างและแสดงรายงานการหาอายุและขนาดของออบเจ็กต์เลย จะประกอบไปด้วย 3 ส่วนหลัก ๆ คือ

- Report Format ในการพัฒนาโปรแกรมในส่วนนี้ จะทำเป็นตัวเลือกให้ผู้ใช้เลือกรูปแบบของรายงาน โดยมีให้เลือก 3 รูปแบบคือ
  - Chart รายงานในรูปแบบกราฟ โปรแกรมในส่วนนี้ใช้โปรแกรม JpGraph เป็นเครื่องมือในการสร้างกราฟ คือหลังจากผู้ใช้เลือกรูปแบบและเงื่อนไขอื่นแล้วทำการส่งข้อมูลเข้ามาให้ PHP ประมวลผล และ PHP จะสร้างเป็นคำสั่ง SQL ไปสอบถามข้อมูลที่ MySQL Server จากนั้นเมื่อได้ข้อมูลกลับมาข้อมูลนั้นจะถูกเก็บไว้ในอาร์เรย์ก่อนที่จะนำข้อมูลในส่วนนี้ไปสร้างกราฟด้วยโปรแกรม JpGraph ตามเงื่อนไขที่ผู้ใช้กำหนดไว้
  - Table รายงานในรูปแบบของตาราง หลังจากผู้ใช้เลือกรูปแบบและเงื่อนไขอื่นแล้วทำการส่งข้อมูลเข้ามาให้ PHP ประมวลผล และ PHP จะสร้างเป็นคำสั่ง SQLไปสอบถามข้อมูลที่ MySQL Server เมื่อได้ข้อมูลกลับมาจะถูกเก็บไว้ในอาร์เรย์จากนั้นที่จะนำข้อมูลในส่วนนี้ไปสร้างเป็นคำสั่ง HTML เพื่อแสดงในรูปแบบของตาราง ตามเงื่อนไขที่ผู้ใช้กำหนดไว้
  - File รายงานในรูปแบบของเท็กซ์ไฟล์ โดยโปรแกรมในส่วนการสร้างเท็กซ์ไฟล์นี้ ในการทำงานจะใช้โปรแกรมภาษาสคริปต์ PHP ไปสร้างไฟล์สคริปต์เป็นคำสั่ง SQL ไว้ จากนั้นให้โปรแกรมไปเรียกใช้คำสั่ง UNIX และทำงานในแบบเบื้องหลัง ให้ทำการรีไดเร็กสคริปต์ที่สร้างนี้ เข้าไปประมวลผลใน MySQL และรีไดเร็กผลลัพธ์ออกเป็นเท็กซ์ไฟล์โดยคำสั่งที่ใช้มีรูปแบบดังนี้

    # mysql < Script Sql Command > Output file

- URL TYPE โปรแกรมในส่วนนี้ จะสร้างตัวเลือกให้ผู้ใช้สามารถเลือกว่าต้องการรายงาน URL ที่เป็นประเภท Static หรือ Dynamic ในการแยกประเภทนี้ทำได้โดยวิธีการ ตรวจสอบจาก URL ที่มีเครื่องหมาย "?" ถ้า URL ใดมีเครื่องหมาย "?" แสดงว่าเป็นแบบไดนามิกส่วน URL ใดไม่มีเครื่องหมาย "?" แสดงว่าเป็นแบบสแตติก ซึ่งในการตรวจสอบจะตรวจสอบตั้งแต่ขั้นตอนการอ่านล็อกไฟล์แล้ว เพราะจะต้องนำข้อมูลประเภทของออบเจ็กต์ไปเก็บในตาราง Summary เพราะในตารางนี้ไม่มีการเก็บ URL ไว้ เพื่อความรวดเร็วในการประมวลผลโดยไม่ต้องไปดึงข้อมูลจากตารางอื่นเข้ามาอีก
- QUERY FORM คือส่วนที่เป็นข้อมูลที่ผู้ใช้กำหนดเงื่อนไขในการสร้างรายงานเองได้ ในการพัฒนาโปรแกรม ก็จะนำข้อมูลที่ผู้ใช้กำหนด ไปสร้างเป็นคำสั่ง SQL ตามที่ผู้ใช้กำหนด หากผู้ใช้ไม่มีการกำหนดข้อมูลในส่วนนี้ โปรแกรมก็จะนำค่า Default ที่ตั้งไว้ไปสร้างเป็นคำสั่ง SQL แทน

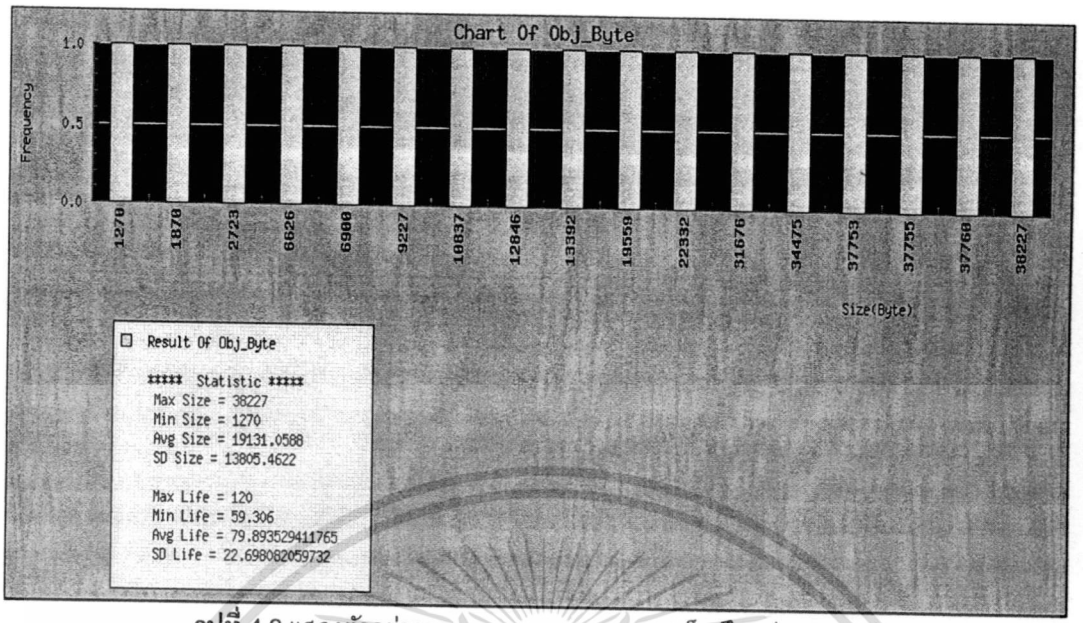
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



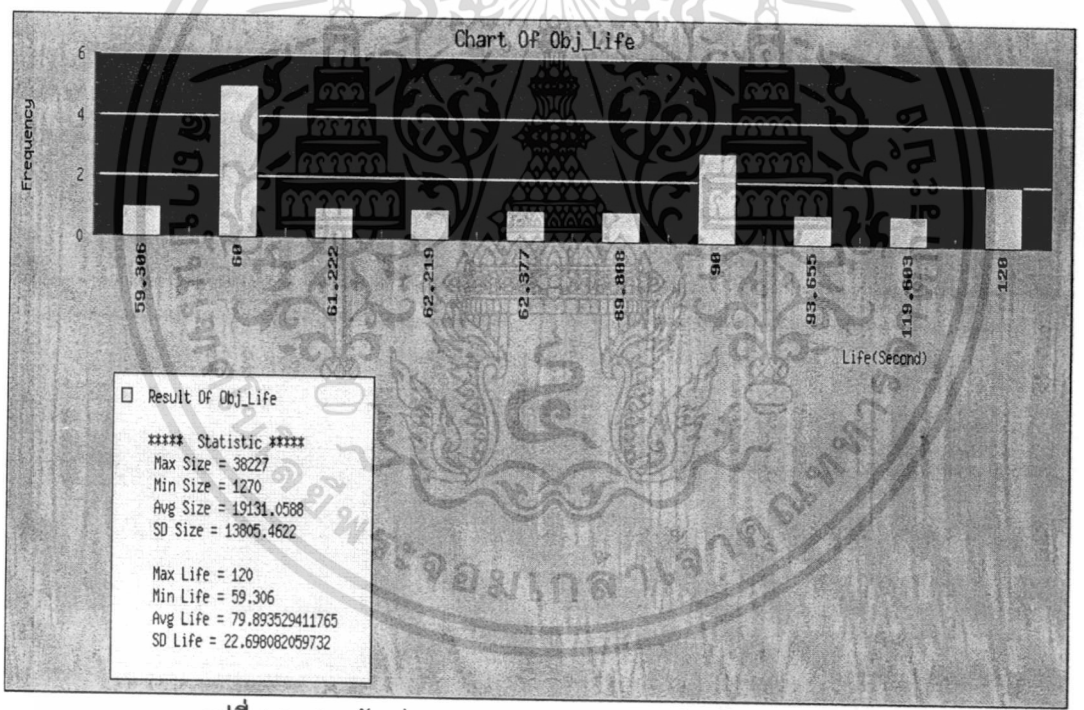
รูปที่ 4.1 แสดงหน้าจอหลักของโปรแกรมสร้างรายงานผ่านเว็บ

โดยเมื่อผู้ใช้เรียกโปรแกรมผ่านทางเว็บบราวเซอร์ตามที่อยู่ที่ได้ติดตั้งโปรแกรมไว้ก็จะปรากฏหน้าจอหลักดังในรูปที่ 4.1 ในการใช้งาน เมื่อผู้ใช้เลือกและกำหนดรูปแบบของรายงานเสร็จแล้ว หลังจากนั้นผู้ใช้จึงทำการเลือกชนิดของเว็บออบเจ็กต์ที่ต้องการสอบถามข้อมูลจากฐานข้อมูลว่า ต้องการให้แสดงออบเจ็กต์ที่เป็นแบบ Static หรือแบบ Dynamic หรือถ้าต้องการจะสอบถามข้อมูลทั้ง 2 ประเภท ก็สามารถทำได้ โปรแกรมจะแยกชนิดโดยจะทำการตรวจสอบจากเครื่องหมาย "? " ใน URL จากนั้นให้ผู้ใช้ป้อนข้อมูลต่าง ๆ เข้าไปเพื่อกำหนดช่วงของข้อมูลตามแบบต่างๆ ได้ตามผู้ใช้ต้องการ โปรแกรมจะสร้างเงื่อนไขในการสอบถามข้อมูลจากฐานข้อมูล โดยข้อมูลทั้ง 4 รูปแบบจะถูกนำมาสร้างเงื่อนไขร่วมกัน (AND) เมื่อผู้ใช้เลือกข้อมูลต่าง ๆ ครบถ้วนแล้ว ทำการกดปุ่ม Query เพื่อสอบถามข้อมูลจากฐานข้อมูลผ่านโปรแกรม รายงานที่ได้จะเป็นการเปิด windows ใหม่ขึ้นมาแสดงรายงานในรูปแบบต่าง ๆ ที่ได้เลือกไว้ ดังตัวอย่างในรูปที่ 4.2 , 4.3 , 4.4 , 4.5 และ 4.6 ตามลำดับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

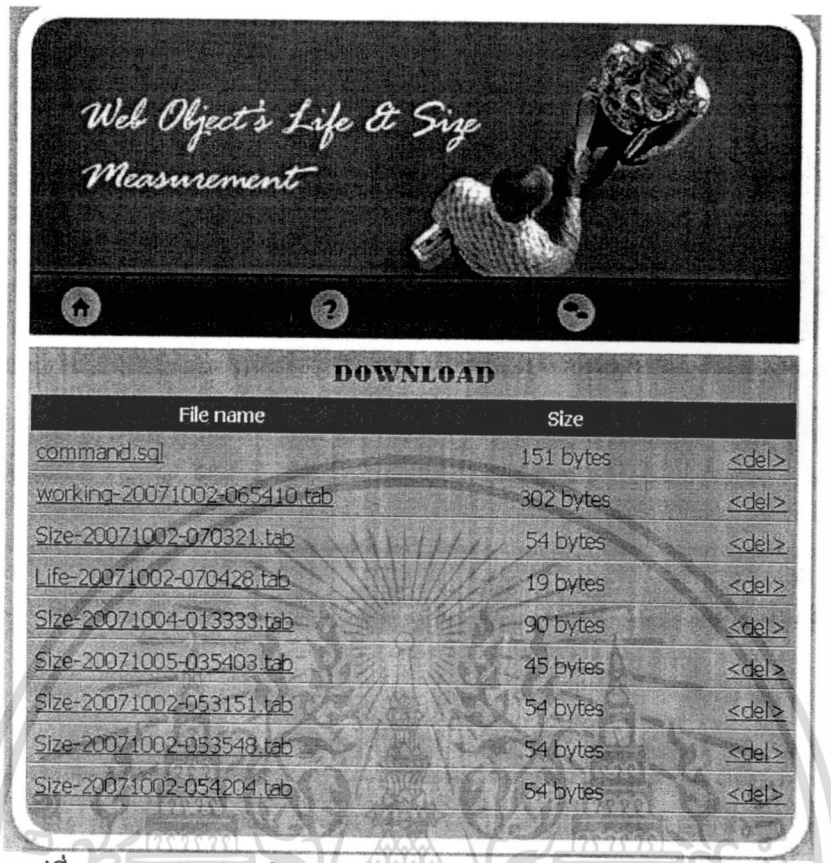


รูปที่ 4.2 แสดงตัวอย่างการรายงานขนาดของออบเจ็กต์ในรูปแบบกราฟ



รูปที่ 4.3 แสดงตัวอย่างการรายงานอายุของออบเจ็กต์ในรูปแบบกราฟ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.4 แสดงตัวอย่างหน้าจอการรายงานอายุของออบเจ็กต์ในรูปแบบเท็กซ์ไฟล์

Table of Summary

First Page | Prev Page | Next Page | Last Page

ID	Death	Obj ID							
1734	2007-09-18 03:28:37	7236	http://161.246.49.38:9090/w108d3652.1ecb059b:00000006/t01/_00000001	2.706	4580	0	3		3.29
1735	2007-09-17 20:25:06	7237	http://161.246.49.38:9090/w108d3652.1ecb059b:00000006/t01/_00000002	120	6009	0	1		8.397
1736	2007-09-17 20:27:06	7237	http://161.246.49.38:9090/w108d3652.1ecb059b:00000006/t01/_00000002	120	6020	0	1		1.731
1737	2007-09-17 20:25:18	7243	http://161.246.49.38:9090/w108d3652.1ecb059b:00000006/t01/_00000008	120	1195	0	1		22.168
1738	2007-09-17 20:25:40	7238	http://161.246.49.38:9090/w108d3652.1ecb059b:00000006/t01/_00000003	120	1141	0	1		48.735
1739	2007-09-17 20:27:40	7238	http://161.246.49.38:9090/w108d3652.1ecb059b:00000006/t01/_00000003	120	1152	0	1		2.49
1740	2007-09-17 20:25:50	7245	http://161.246.49.38:9090/w108d3652.1ecb059b:00000006/t01/_0000000a	120	6778	0	1		48.56
1741	2007-09-18 03:29:25	7236	http://161.246.49.38:9090/w108d3652.1ecb059b:00000006/t01/_00000001	48.633	4582	0	3		48.634
1742	2007-09-17 20:26:43	7257	http://161.246.49.38:9090/w108d3652.1ecb059b:00000006/t01/_00000016	120	1080	0	1		48.956
1743	2007-09-17 20:26:47	7242	http://161.246.49.38:9090/w108d3652.1ecb059b:00000006/t01/_00000007	120	8271	0	1		92.56
1744	2007-09-17 20:25:45	7244	http://161.246.49.38:9090/w108d3652.1ecb059b:00000006/t01/_00000009	120	1264	0	1		88.799
1745	2007-09-17 20:26:35	7270	http://161.246.49.38:9090/w108d3652.1ecb059b:00000006/t01/_00000023	120	9271	0	1		39.25
1746	2007-09-17 20:26:44	7272	http://161.246.49.38:9090/w108d3652.1ecb059b:00000006/t01/_00000025	120	422210		1		36.613
1747	2007-09-17 20:25:58	7256	http://161.246.49.38:9090/w108d3652.1ecb059b:00000006/t01/_00000015	120	251420		1		63.665
1748	2007-09-17 20:28:44	7272	http://161.246.49.38:9090/w108d3652.1ecb059b:00000006/t01/_00000025	120	422320		1		2.96
1749	2007-09-17 20:26:12	7264	http://161.246.49.38:9090/w108d3652.1ecb059b:00000006/t01/_0000001d	120	6829	0	1		49.994
1750	2007-09-17 20:28:43	7257	http://161.246.49.38:9090/w108d3652.1ecb059b:00000006/t01/_00000016	120	1091	0	1		12.644

รูปที่ 4.5 แสดงตัวอย่างการรายงานข้อมูลในตาราง Summary

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Table of Working

First Page | Prev Page | Next Page | Last Page

7277	http://161.246.49.38:9090/w108d3652.1ecb059b:00000006/t01/_0000002a	29507	0	1190060998.223	0	0
7278	http://161.246.49.38:9090/w108d3652.1ecb059b:00000006/t01/_0000002b	12631	1190061121.508	1190061146.913	0	1190035826
7279	http://161.246.49.38:9090/w108d3652.1ecb059b:00000006/t01/_0000002c	3405	1190061010.377	1190061127.327	0	1190035626
7280	http://161.246.49.38:9090/w108d3652.1ecb059b:00000006/t01/_0000002d	1463	0	1190061010.803	0	1190035692
7281	http://161.246.49.38:9090/w108d3652.1ecb059b:00000006/t01/_0000002e	2723	1190061041.525	1190061156.368	0	1190035744
7282	http://161.246.49.38:9090/w108d3652.1ecb059b:00000006/t01/_0000002f	20370	0	1190061017.388	0	1190035654
7283	http://161.246.49.38:9090/w108d3652.1ecb059b:00000006/t01/_00000030	2851	1190061109.74	1190061137.516	0	1190035759
7284	http://161.246.49.38:9090/w108d3652.1ecb059b:00000006/t01/_00000031	7540	1190061090.878	1190061133.307	0	1190035753
7285	http://161.246.49.38:9090/w108d3652.1ecb059b:00000006/t01/_00000032	19032	1190061019.493	1190061073.14	0	1190035631
7286	http://161.246.49.38:9090/w108d3652.1ecb059b:00000006/t01/_00000033	4797	0	1190061019.775	0	1190035708
7287	http://161.246.49.38:9090/w108d3652.1ecb059b:00000006/t01/_00000034	4126	0	1190061020.499	0	1190035639
7288	http://161.246.49.38:9090/w108d3652.1ecb059b:00000006/t01/_00000035	9505	0	1190061035.414	0	1190035657
7289	http://161.246.49.38:9090/w108d3652.1ecb059b:00000006/t01/_00000036	13102	1190061043.309	1190061081.356	0	1190035656
7290	http://161.246.49.38:9090/w108d3652.1ecb059b:00000006/t01/_00000037	18767	1190061044.101	1190061123.228	0	1190035692
7291	http://161.246.49.38:9090/w108d3652.1ecb059b:00000006/t01/_00000038	855	1190061065.652	1190061082.16	0	0
7292	http://161.246.49.38:9090/w108d3652.1ecb059b:00000006/t01/_00000039	14612	1190061101.868	1190061119.739	0	1190035779

รูปที่ 4.6 แสดงตัวอย่างการรายงานข้อมูลระหว่างการทำงานในรูปของตาราง

4.3 การตรวจสอบความถูกต้องของโปรแกรม

เพื่อความถูกต้องในการทำงานของโปรแกรม จึงได้มีการตรวจสอบความถูกต้องของโปรแกรมตามวิธีการต่าง ๆ โดยเครื่องมือที่ใช้ในการทดสอบครั้งนี้ มีสภาพแวดล้อมต่าง ๆ ดังนี้

CPU : intel Core 2 CPU 1.8 GHz

RAM : DDR2 1.5 GB

HardDisk : 120 GB

Operating System : FreeBSD 5.5

Web Server : Apache 2.2

Web Browser: IE 7

Proxy/Cache Program : Squid 2.6 STABLE 13

4.3.1 การตรวจสอบการคำนวณหาอายุของเว็บออบเจ็กต์

4.4.1.1 แสดงตัวอย่างวิธีคิดแบบไม่ใช้ค่า Last-Modified

จากอัลกอริธึมในการวัดอายุของออบเจ็กต์แบบไม่ใช้ค่า Last-Modified ในบทที่ 2 จึงนำมาทำการทดลองเพื่อทดสอบการทำงานของโปรแกรม โดยจะทำการจำลองอายุของ Web Object LifeCycle ให้เป็นแบบ Constant โดยจะทดสอบด้วยโปรแกรม polygraph [4] เวอร์ชัน 2.8.1 ให้ทำการสร้างการร้องขอเว็บออบเจ็กต์ผ่านทางล็อกไฟล์ของพร็อกซี่ของโปรแกรม squid ที่ได้ทำการพัฒนาไว้ในตอนต้น โดยในขณะที่ทำการสร้างการร้องขอจากไคลเอนท์ไปยังพร็อกซี่เซิร์ฟเวอร์นั้นจะมีการควบคุมค่าพารามิเตอร์ต่างๆ ที่ป้อนให้โปรแกรม polygraph ดังต่อไปนี้

1. ขนาดของออบเจกต์โดยเฉลี่ยที่ส่งมาจากเซิร์ฟเวอร์ประมาณ 13 กิโลไบต์
2. จำนวนของออบเจกต์ที่สามารถแคชได้จากเซิร์ฟเวอร์คิดเป็นร้อยละ 80
3. เวลาที่ใช้ในการทดสอบ 5 นาที

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4. อัตราการร้องขอคงที่ (Request Rate) 1/ Sec

การทดลองที่ 4.1 ทำการทดลอง โดยกำหนดค่าพารามิเตอร์ข้างต้นให้คงที่ และทำการกำหนดค่าอายุของเว็บออบเจ็กต์เป็นอายุคงที่ดังนี้ คือ 60 , 90 , 120 วินาที เมื่อนำล็อกไฟล์ดังกล่าวไปทดสอบด้วยโปรแกรมหาอายุและขนาดของเว็บออบเจ็กต์ที่ได้ทำการพัฒนา ได้ผลการทดลองดัง รูปที่ 4.7 , 4.8 และ 4.9 ต่อไปนี้

Sum_ID	Obj_Death	Obj_ID	Obj_Type	Obj_Life	Obj_Byte	Obj_Count	Accuracy	Obj_Dev
1945	2007-09-27 14:17:28	7393	S	59.306	38227	125	3	1.293
1946	2007-09-27 14:18:29	7393	S	61.222	6626	184	3	0.442
1947	2007-09-27 14:19:32	7393	S	62.377	22332	240	3	1.37
1948	2007-09-27 14:20:34	7393	S	62.219	19559	298	3	0.566

รูปที่ 4.7 แสดงตาราง Summary ซึ่งเก็บผลการหาอายุและค่าความคลาดเคลื่อน จากตัวอย่างการทดลองที่ Web Object Life Cycle ที่ 60 วินาที

Sum_ID	Obj_Death	Obj_ID	Obj_Type	Obj_Life	Obj_Byte	Obj_Count	Accuracy	Obj_Dev
1949	2007-09-27 14:32:25	7394	S	93.655	10837	187	3	3.1
1950	2007-09-27 14:33:55	7394	S	89.808	1870	267	3	1.801

รูปที่ 4.8 แสดงตาราง Summary ซึ่งเก็บผลการหาอายุและค่าความคลาดเคลื่อน จากตัวอย่างการทดลองที่ Web Object Life Cycle ที่ 90 วินาที

Sum_ID	Obj_Death	Obj_ID	Obj_Type	Obj_Life	Obj_Byte	Obj_Count	Accuracy	Obj_Dev
1951	2007-09-27 14:45:36	7395	S	119.603	12846	243	3	0.963

รูปที่ 4.9 แสดงตาราง Summary ซึ่งเก็บผลการหาอายุและค่าความคลาดเคลื่อน จากตัวอย่างการทดลองที่ Web Object Life Cycle ที่ 120 วินาที

วิเคราะห์ผลการทดลอง

ผลจากการทดลอง แสดงให้เห็นว่าโปรแกรมคำนวณหาอายุและขนาดของเว็บออบเจ็กต์นั้นสามารถทำการคำนวณอายุเฉลี่ยได้เท่ากับ 61.281 , 91.7315 และ 119.603 วินาทีตามลำดับ ซึ่งไม่เท่ากับค่าอายุที่ได้กำหนดไว้ในตอนต้น แต่จะได้แค่ค่าที่ใกล้เคียงเท่านั้น

ในส่วนของความคลาดเคลื่อนที่วัดได้จากการทดลองก็มีค่าค่อนข้างมาก โดยค่าของความคลาดเคลื่อนจะขึ้นอยู่กับระยะเวลาระหว่างค่าการร้องขอก่อนหน้าที่ออบเจ็กต์จะเกิดการเปลี่ยนแปลงกับค่าการร้องขอที่พบว่าออบเจ็กต์เกิดการเปลี่ยนแปลงนั้น ถ้าระยะห่างของเวลาทั้งสองมีค่าห่างกันมาก ค่าความคลาดเคลื่อนก็จะมากตามไปด้วย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4.1.2 **แสดงตัวอย่างวิธีคิดแบบใช้ค่า** Last-Modified

จากอัลกอริธึมในการวัดอายุของออบเจ็กต์แบบใช้ค่า Last-Modified ในบทที่ 2 จึงนำมาทำการทดลองเพื่อทดสอบการทำงานของโปรแกรม โดยจะทำการจำลองอายุของ Web Object LifeCycle ให้เป็นแบบ Constant โดยจะทดสอบด้วยโปรแกรม polygraph [4] เวอร์ชัน 2.8.1 ให้ทำการสร้างการร้องขอเว็บออบเจ็กต์ผ่านทางล็อกไฟล์ของพร็อกซี่ของโปรแกรม squid ที่ได้ทำการพัฒนาไว้ในตอนต้น โดยในขณะที่ทำการสร้างการร้องขอจากไคลเอนท์ไปยังพร็อกซี่เซิร์ฟเวอร์นั้นจะมีการควบคุมค่าพารามิเตอร์ต่างๆ ที่ป้อนให้โปรแกรม polygraph ดังต่อไปนี้

1. ขนาดของออบเจกต์โดยเฉลี่ยที่ส่งมาจากเซิร์ฟเวอร์ประมาณ 13 กิโลไบต์
2. จำนวนของออบเจกต์ที่สามารถแคชได้จากเซิร์ฟเวอร์คิดเป็นร้อยละ 80
3. เวลาที่ใช้ในการทดสอบ 5 นาที
4. อัตราการร้องขอคงที่ (Request Rate) 1/ Sec

การทดลองที่ 4.2 ทำการทดลอง โดยกำหนดพารามิเตอร์ข้างต้นให้คงที่ และทำการกำหนดค่าอายุของเว็บออบเจ็กต์เป็นอายุคงที่ดังนี้คือ 60 , 90 , 120 วินาที เมื่อนำล็อกไฟล์ดังกล่าวไปทดสอบด้วยโปรแกรมหาอายุและขนาดของเว็บออบเจ็กต์ที่ได้ทำการพัฒนา ได้ผลการทดลองดัง รูปที่ 4.10 , 4.11 และ 4.12 ต่อไปนี้

Sum_ID	Obj_Death	Obj_ID	Obj_Type	Obj_Life	Obj_Byte	Obj_Count	Accuracy	Obj_Dev
1938	2007-09-26 18:46:50	7391	S	60	37753	46	1	1.578
1939	2007-09-26 18:47:50	7391	S	60	37760	105	1	2.2
1940	2007-09-26 18:48:50	7391	S	60	37755	163	1	0.549
1941	2007-09-26 18:49:50	7391	S	60	34475	226	1	1.184
1942	2007-09-26 18:50:50	7391	S	60	31676	277	1	0.319

รูปที่ 4.10 แสดงตาราง Summary ซึ่งเก็บผลการหาอายุและค่าความคลาดเคลื่อน
จากตัวอย่างการทดลองที่ Web Object Life Cycle ที่ 60 วินาที

Sum_ID	Obj_Death	Obj_ID	Obj_Type	Obj_Life	Obj_Byte	Obj_Count	Accuracy	Obj_Dev
1935	2007-09-27 13:54:56	7390	S	90	2723	81	1	0.142
1936	2007-09-27 13:56:26	7390	S	90	1270	166	1	0.178
1937	2007-09-27 13:57:56	7390	S	90	13392	250	1	2.311

รูปที่ 4.11 แสดงตาราง Summary ซึ่งเก็บผลการหาอายุและค่าความคลาดเคลื่อน
จากตัวอย่างการทดลองที่ Web Object Life Cycle ที่ 90 วินาที

Sum_ID	Obj_Death	Obj_ID	Obj_Type	Obj_Life	Obj_Byte	Obj_Count	Accuracy	Obj_Dev
1943	2007-09-26 19:22:30	7392	S	120	9227	71	1	0.365
1944	2007-09-26 19:24:30	7392	S	120	6900	189	1	0.288

รูปที่ 4.12 แสดงตาราง Summary ซึ่งเก็บผลการหาอายุและค่าความคลาดเคลื่อน
จากตัวอย่างการทดลองที่ Web Object Life Cycle ที่ 120 วินาที

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

วิเคราะห์ผลการทดลอง

จากผลการทดลองแสดงให้เห็นว่า เมื่อทำการจำลองอายุ Web Object Life Cycle เป็นค่าคงตัวค่าหนึ่ง โดยให้โปรแกรม Polygraph ทำการสร้างการติดต่อระหว่าง Client และ Server ผ่านเว็บพร็อกซี่ เมื่อนำล็อกไฟล์ของเว็บพร็อกซี่ดังกล่าวมาทำการคำนวณหาอายุของเว็บออบเจ็กต์ผลการคำนวณควรจะได้ค่าอายุของเว็บออบเจ็กต์ดังกล่าว เท่ากับค่าคงที่ที่ได้ทำการกำหนดไว้ในตอนต้น และค่าความคลาดเคลื่อนจากการคำนวณจะสามารถเกิดขึ้นได้ในช่วงระหว่างค่าการร้องขอ ครั้งสุดท้ายก่อนที่ค่า Last-Modified จะเกิดการเปลี่ยนแปลงไป ถึงค่าการร้องขอที่ค่า Last-Modified ที่ได้เกิดการเปลี่ยนแปลงไป เพราะในช่วงดังกล่าวอาจจะมีเกิดการเปลี่ยนแปลงค่า Last-Modified ขึ้นซึ่งเราไม่สามารถทราบได้ ดังที่ได้อธิบายไว้ในบทที่ 2

เมื่อนำ mix.log ที่ได้จากการทดลองข้างต้น ไปทำการคำนวณหาอายุของเว็บออบเจ็กต์ ได้ผลการทดลองดังรูปที่ 4.10 , 4.11 และ 4.12 ข้างต้น คือ 60 , 90 และ 120 แสดงให้เห็นว่าโปรแกรมสามารถทำการหาอายุของเว็บออบเจ็กต์ได้ตรงกับค่าอายุที่ได้ทำการกำหนดไว้ในตอนต้น

ส่วนค่าความคลาดเคลื่อนที่ได้จากการทดลองดังกล่าวคือ 0.319 , 2.311 และ 0.288 ซึ่งจะเห็นว่าค่าความคลาดเคลื่อนนี้อยู่ในช่วงระหว่างค่าการร้องขอครั้งสุดท้ายก่อนที่ค่า Last-Modified จะเกิดการเปลี่ยนแปลงไปถึงค่าการร้องขอที่ค่า Last-Modified ที่ค่าได้เกิดการเปลี่ยนแปลงไป และไม่เกินค่าอายุของเว็บออบเจ็กต์จึงสามารถสรุปได้ว่า โปรแกรมคำนวณหาอายุของเว็บออบเจ็กต์สามารถทำการหาค่าความคลาดเคลื่อนด้วยวิธีใช้ค่า Last-Modified ได้อย่างถูกต้องตรงกับสมการที่ได้กำหนดไว้ในบทที่ 2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
บทสรุป

5.1 บทสรุปผลการทดลอง

จากผลการทดลองในบทที่ 4 แสดงให้เห็นว่า ค่าความแม่นยำที่ได้จากการคำนวณหาอายุของเว็บออบเจ็กต์ด้วยวิธีใช้ค่า Last-Modified จะได้ค่าอายุที่ตรงกับค่าอายุจริงที่กำหนดไว้มากกว่าการใช้วิธีแบบไม่ใช้ค่า Last-Modified และค่าความคลาดเคลื่อนจากวิธีที่ไม่ใช้ค่า Last-Modified จะมีค่าที่สูงกว่าค่าความคลาดเคลื่อนจากวิธีใช้ค่า Last-Modified

5.2 ผลจากการพัฒนาระบบ

ผลการศึกษาพบว่าสามารถนำโปรแกรมไปใช้ในการศึกษาลักษณะของข้อมูลในเครือข่ายทั้งเว็บออบเจ็กต์ที่มีค่า Last-Modified และไม่มีค่า Last-Modified โดยข้อมูลที่ได้จากโปรแกรมนี้จะมีค่าคลาดเคลื่อนคลาดที่น้อยลง จากวิธีที่ไม่มีค่า Last-Modified เพียงอย่างเดียวและยังสามารถนำข้อมูลที่ได้นี้ไปใช้ศึกษาและวิจัยถึงการลดปริมาณการใช้งานเครือข่ายที่จะติดต่อออกไปในอินเตอร์เน็ต โดยเฉพาะในส่วนของข้อมูลที่มีผู้ใช้ร้องขอเป็นจำนวนมาก

ซึ่งในการทำงานของโปรแกรมส่วนที่พัฒนาล็อกไฟล์ของโปรแกรม squid จะได้ทำการเก็บข้อมูลการทำงานที่ถูกร้องขอผ่านโปรแกรม squid ไว้ เพื่อส่งต่อข้อมูลให้ส่วนของโปรแกรมคำนวณหาค่าอายุและขนาดของเว็บออบเจ็กต์ทำการคำนวณต่อไป โดยในส่วนของการคำนวณได้เพิ่มวิธีการคำนวณด้วยวิธีการใช้ค่า Last-Modified เพิ่มเข้ามาจากหลักการหาอายุและขนาดของเว็บออบเจ็กต์เดิม และในส่วนท้ายสุดคือส่วนของการแสดงรายงานผลที่คำนวณได้ให้กับผู้ใช้ ที่ใช้รูปแบบการติดต่อผู้ใช้ผ่านทางเว็บ ทำให้ผู้ใช้สามารถนำผลที่ได้จากโปรแกรมไปศึกษาได้อย่างสะดวก และ ทำความเข้าใจได้ง่ายยิ่งขึ้น

5.3 ประโยชน์ที่ได้รับ

โปรแกรมที่ได้จากการพัฒนาจะสามารถใช้เป็นเครื่องมือ เพื่อศึกษาลักษณะเฉพาะของออบเจ็กต์ที่วิ่งผ่านเครือข่ายที่ทำงานได้อย่างมีประสิทธิภาพและสะดวกต่อการใช้งาน สามารถแก้ปัญหาต่างๆ ที่เกิดขึ้นในระบบงานเดิมได้ ทำให้โปรแกรมสามารถคำนวณหาค่าอายุและขนาดของเว็บออบเจ็กต์ได้ถูกต้องและใกล้เคียงกับความเป็นจริงและมีค่าความคลาดเคลื่อนลดลง โดยข้อมูลในส่วนนี้จะบอกได้ถึงลักษณะเฉพาะของออบเจ็กต์ที่ถูกร้องขอผ่านเครือข่าย ดังนั้นจึงสามารถนำข้อมูลที่ได้ไปใช้วิเคราะห์ในเชิงสถิติ เพื่อใช้ในงานวิจัยอื่น ๆ และยังสามารถนำข้อมูลที่ได้มาใช้ในการปรับปรุงประสิทธิภาพของเว็บพร็อกซี่และแคชเพื่อรองรับการใช้งานที่เกิดขึ้นจริงในเครือข่าย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.4 ข้อเสนอแนะและแนวทางในการพัฒนาต่อ

เนื่องจากจำนวนข้อมูลที่มีอย่างมากมายมหาศาล การทำงานผ่านทางเว็บอาจไม่เหมาะกับการทำงานในลักษณะโต้ตอบกับผู้ใช้แบบทันที เนื่องจากการทำงานผ่านทางเว็บจะมีข้อจำกัดในด้านเวลาที่เชื่อมต่อกับเว็บเซิร์ฟเวอร์ (อาจเกิด Time Out ได้) ดังนั้นควรจะมีการปรับปรุงโปรแกรมในส่วนการสร้างรายงานและแสดงผลให้มีการทำงานในลักษณะทำงานแบบเบื้องหลัง (Back Ground) โดยเฉพาะอย่างยิ่งในระหว่างการสร้างกราฟ ควรปรับปรุงส่วนที่ใช้ในการสร้างกราฟให้เป็นการทำงานในแบบเบื้องหลังและสร้างเป็นไฟล์รูปภาพเก็บไว้ เมื่อประมวลผลเสร็จจึงนำข้อมูลที่ได้มาแสดงผลผ่านทางเว็บ ซึ่งจะทำให้ผู้ใช้ไม่รู้สึกว่าต้องรอนาน และเพื่อให้ผู้ใช้โปรแกรมทำงานได้สะดวกขึ้น ควรมีการปรับปรุงและพัฒนาโปรแกรมให้สามารถรองรับการวิเคราะห์ข้อมูลจากหลายพร็อกซี่เซิร์ฟเวอร์ รวมถึงการพัฒนาโปรแกรมให้สามารถนำข้อมูลในส่วนที่เก็บอยู่ในฐานข้อมูลไปใช้ปรับปรุงค่าปฏิบัติการของโปรแกรม Squid แบบอัตโนมัติ

จากอัลกอริธึมในการวัดอายุของออบเจ็กต์แบบใช้ค่า Last-Modified ก็ยังมีข้อจำกัดคือ ถ้าเวลาที่ Server และ Client ไม่ได้ถูกกำหนดค่าให้ตรงกัน ซึ่งสามารถพบได้ ก็จะไม่สามารถใช้อัลกอริธึมในการวัดอายุของออบเจ็กต์แบบใช้ค่า Last-Modified หาความคลาดเคลื่อนได้ แต่ยังคงสามารถใช้วิธีหาอายุแบบใช้ค่า Last-Modified ทำการหาอายุได้ ในกรณีที่เกิดปัญหาดังกล่าวสามารถใช้การคำนวณดังสมการต่อไปนี้เพื่อหาค่าความคลาดเคลื่อน แต่ค่าความคลาดเคลื่อนที่ได้จะมีค่าความคลาดเคลื่อนที่เพิ่มขึ้นจากเดิม

$$a_i = La_M_{i+1} - Lx_H_i$$

a_i คือ ค่าความคลาดเคลื่อนจากการหาอายุของเว็บออบเจ็กต์

Lx_M_i คือ ค่าTimestampของการร้องขอเข้ามาครั้งแรกสุดหลังจากที่ค่า Last-Modified ได้เกิดการเปลี่ยนแปลงไป ซึ่งยังไม่เคยมีการร้องขอใดก่อนหน้านี้

Lx_H_i คือ ค่า Timestampของการร้องขอครั้งสุดท้ายก่อนที่ค่า Last-Modified จะเกิดการเปลี่ยนแปลงไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

[1] ณัฐพล แสนคำ. **การพัฒนาโปรแกรมสำหรับหาอายุและขนาดของออบเจ็กต์จาก Squid's Log File.** กรุงเทพฯ: โครงการพัฒนาระบบงาน คณะเทคโนโลยีสารสนเทศ สถาบัน เทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. 2545.

[2] วิฑิต อินทรหะและอัครินทร์ คุณกิตติ. **การวัดอายุและขนาดของออบเจ็กต์ใน World Wide Web จาก Log File.** กรุงเทพฯ : วิทยานิพนธ์ คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง, 2544.

[3] Kevin Atkinson, Mysql++ A C++ API for Mysql. [Online]. Available:http://www.mysql.com, 2001.

[4] Web Polygraph. [Online]. Available: http://www.web-polygraph.org

[5] Kinkie, Squid Faq/ SquidLogs. [Online]. Available:http://wiki.Squid-cache.org/SquidFag/Squidlogs#/head-d4ce65324ebaacc7fa87cc152edb17fc3342ecfbl, 2006.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้