การตรวจจับไวรัสข้อความมัลติมีเดียโดยใช้เทคนิคการทำเหมืองข้อมูล

ปริญญานิพนธ์

ของ

อัณศยา เกิดคล้าย

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา

ตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาการจัดการทางวิศวกรรม

พฤษภาคม 2553

การตรวจจับไวรัสข้อความมัลติมีเดียโดยใช้เทคนิคการทำเหมืองข้อมูล

ปริญญานิพนธ์

ของ

อัณศยา เกิดคล้าย

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา

ตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาการจัดการทางวิศวกรรม

พฤษภาคม 2553

การตรวจจับไวรัสข้อความมัลติมีเดียโดยใช้เทคนิคการทำเหมืองข้อมูล

บทคัดย่อ

ของ

อัณศยา เกิดคล้าย

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา

ตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาการจัดการทางวิศวกรรม

พฤษภาคม 2553

อัณศยา เกิดคล้าย. (2553). *การตรวจจับไวรัสข้อความมัลติมีเดียโดยใช้เทคนิคการทำเหมืองข้อมูล*. ปริญญานิพนธ์ วศ.ม.(การจัดการทางวิศวกรรม). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. คณะกรรมการควบคุม: ผู้ช่วยศาสตราจารย์ ดร. นำคุณ ศรีสนิท, ดร. พรศักดิ์ เรียบร้อยเจริญ.

งานวิจัยนี้ได้นำเทคนิคกฎการค้นหาความสัมพันธ์ ซึ่งเป็นเทคนิคหนึ่งในเทคนิคเหมืองข้อมูลมาประยุกต์ใช้ออกแบบโปรแกรมสำหรับใช้ในการตรวจจับข้อความมัลติมีเดียที่มีไวรัสชนิด CommWarrior โดยผู้วิจัยได้จัดเก็บข้อมูล 4 ส่วนคือ 1.การส่งข้อความมัลติมีเดียของผู้ใช้งานภายในระบบบริการรับส่งข้อความมัลติมีเดีย 2.ลักษณะของข้อมูลที่ส่งแนบมากับข้อความมัลติมีเดีย 3.ขนาดข้อมูลที่ส่งแนบมากับข้อความมัลติมีเดีย 4.รุ่นโทรศัพท์ของผู้ส่งข้อความมัลติมีเดีย เพื่อจัดหมวดหมู่ของข้อมูล แล้วนำข้อมูลทั้ง 4 ส่วนมาสร้างความสัมพันธ์ โดยงานวิจัยนี้นำข้อมูลลักษณะของข้อมูล, ขนาดข้อมูลที่ส่งแนบมากับข้อความมัลติมีเดีย และรุ่นโทรศัพท์ของผู้ส่งข้อความมัลติมีเดีย มาค้นหาความสัมพันธ์เพื่อออกแบบโปรแกรม โดยพิจารณาจากค่าความเชื่อมั่นและค่าสนับสนุน ผู้วิจัยได้ทำการทดลองศึกษาโปรแกรม โดยผลของการศึกษาโปรแกรมพบว่าโปรแกรมที่สร้างขึ้นสามารถตรวจจับข้อความมัลติมีเดียที่มีไวรัสได้ ผลของการใช้เทคนิคเหมืองข้อมูลพบว่าโปรแกรมที่สร้างขึ้นสามารถตรวจจับข้อความมัลติมีเดียที่มีไวรัสได้โดยมีความถูกต้องร้อยละ 99.89 หรือเทียบเท่า Viral Marketing Tool ซึ่งเป็นโปรแกรมเชิงพาณิชย์

งานวิจัยนี้สามารถตรวจจับข้อความมัลติมีเดียที่มีไวรัสชนิด CommWarrirr ได้ จึงช่วยให้ผู้ให้บริการโครงข่ายโทรศัพท์มือถือไม่จำเป็นต้องเสียค่าใช้จ่ายในการเก็บรวบรวมข้อมูลการส่งข้อความมัลติมีเดียถึงระดับ Packet เหมือนเช่น Viral Marketing Tool ที่อุปกรณ์เชื่อมต่อเพื่อทำการ Mirror Traffic ต้องมีขนาดความจุที่ใหญ่เพียงพอกับการเก็บรวบรวม Traffic ในแต่ละวัน เนื่องจากข้อมูลที่นำมาค้นหาความสัมพันธ์ในงานวิจัยเป็นข้อมูลจากระบบบริการรับส่งข้อความมัลติมีเดียที่มีการเก็บรวบรวมไว้ในระบบบริการรับส่งข้อความมัลติมีเดียอยู่แล้ว และเพียงพอสำหรับการวิเคราะห์ตรวจจับไวรัส

DETECTING MMS VIRUS USING DATA MINING TECHNIQUE

AN ABSTRACT

BY

UNSAYA KERDKLAI

Presented in Partial Fulfillment of the Requirements for the

Master of Engineering Degree in Engineering Management

at Srinakharinwirot University

May 2010

Unsaya Kerdklai. (2010). *Detecting MMS Virus Using Data Mining Technique.* Master thesis, M.Eng. (Engineering Management). Bangkok: Graduate School, Srinakharinwirot University. Advisor Committee: Asst. Prof. Dr. Namkhun Srisanit, Dr. Pronsak Raibroycharoen.

This research applies one of data mining techniques called 'association rule discovery'. The program was originally created by a concept to resolve unprecedented cost from third party, who is neither user nor mobile network operator. The objective for this research is to detect MMS containing CommWarrior-type virus. Researchers have collected four groups of data: 1.User-submitted MMS in multimedia message service server, 2.Content-type database, 3.Content-size database and 4.Phone-model database; in order to categorize data. Then, these 4 groups of data will be applied to analyze data relationship, according to confident and support values. Researchers practically examine the program and achieve the results that the created program is able to detect virus-contained MMS. The result from applying data mining technique is apparently shown that the program is able to detect virus-contained MMS corrected 99.89% or have the same accuracy as Viral Marketing Tool, commercially available program.

Thus, mobile network operators do not have to spend any cost to collect user-submitted-MMS traffic at packet level as Viral Marketing Tool, which requires massive portable device to mirror traffic. On the other hand, the researched program analyzes relationship from the database created by multimedia message service servers, which daily collect these categories of data and are sufficient for virus detecting analysis.

ปริญญานิพนธ์

เรื่อง

การตรวจจับไวรัสข้อความมัลติมีเดียโดยใช้เทคนิคการทำเหมืองข้อมูล

ของ

อัณศยา เกิดคล้าย

ได้รับอนุมัติจากบัณฑิตวิทยาลัยให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาการจัดการทางวิศวกรรม

ของมหาวิทยาลัยศรีนครินทรวิโรฒ

..... คณบดีบัณฑิตวิทยาลัย

(รองศาสตราจารย์ ดร.สมชาย สันติวัฒนากุล)

วันที่ ..... เดือน ..... พ.ศ. 2553

คณะกรรมการควบคุมปริญญานิพนธ์

..... ประธาน

(ผู้ช่วยศาสตราจารย์ ดร.นำคุณ ศรีสนิท)

..... กรรมการ

(ดร.พรศักดิ์ เรียบร้อยเจริญ)

คณะกรรมการสอบปากเปล่า

..... ประธาน

(รองศาสตราจารย์ ธนรัตน์ แต้วัฒนา)

..... กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.วชิระ จงบุรี)

..... กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.นำคุณ ศรีสนิท)

..... กรรมการ

(ดร.พรศักดิ์ เรียบร้อยเจริญ)

ประกาศคุณูปการ

ปริญญานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยดี โดยได้รับความเมตตาช่วยเหลือ และเอาใจใส่อย่างดียิ่งจากอาจารย์ที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร.นำคุณ ศรีสนิท และ ดร.พรศักดิ์ เรียบร้อยเจริญ ที่คอยให้คำแนะนำและข้อคิดเห็นสำหรับการแก้ไขข้อบกพร่องต่างๆ อันเป็นประโยชน์อย่างยิ่งแก่ผู้วิจัย รวมทั้ง รองศาสตราจารย์ ธนรัตน์ แต้วัฒนา และ ผู้ช่วยศาสตราจารย์ ดร.วชิระ จงบุรี กรรมการสอบปริญญานิพนธ์ ที่กรุณาให้ข้อเสนอแนะต่างๆ เพิ่มเติมแก่ผู้วิจัย จนทำให้ปริญญานิพนธ์ฉบับนี้มีความสมบูรณ์ยิ่งขึ้น ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้ด้วย

ขอกราบขอบพระคุณคณาจารย์และกรรมการบริหารหลักสูตรสาขาการจัดการทางวิศวกรรม คณะวิศวกรรมศาสตร์ มหาวิทยาลัยศรีนคนิทรวิโรฒทุกท่าน ที่ได้กรุณาประสิทธิ์ประสาทความรู้ต่างๆ ให้แก่ผู้วิจัย ตลอดจนให้ความช่วยเหลือในการทำวิจัยครั้งนี้

ขอกราบขอบพระคุณเจ้าหน้าที่ประจำหลักสูตรสาขาการจัดการทางวิศวกรรม คณะวิศวกรรมศาสตร์และบัณฑิตวิทยาลัยทุกท่าน ตลอดจนผู้มีส่วนสำเร็จต่องานวิจัยชินนี้

ขอกราบขอบพระคุณบิดา มารดา สมาชิกทุกคนในครอบครัว พี่และเพื่อนๆ รวมถึงบุคคลอีกหลายท่านที่ไม่ได้กล่าวนามไว้ ณ ที่นี้ที่ได้ให้ความช่วยเหลือ คอยให้คำปรึกษาและให้กำลังใจแก่ผู้วิจัยด้วยดีเสมอมา จนทำให้ปริญญานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยดีในที่สุด

อัณศยา เกิดคล้าย

สารบัญ

สารบัญ(ต่อ)

บัญชีตาราง

บัญชีตาราง(ต่อ)

ตาราง **หน้า**

บัญชีภาพประกอบ

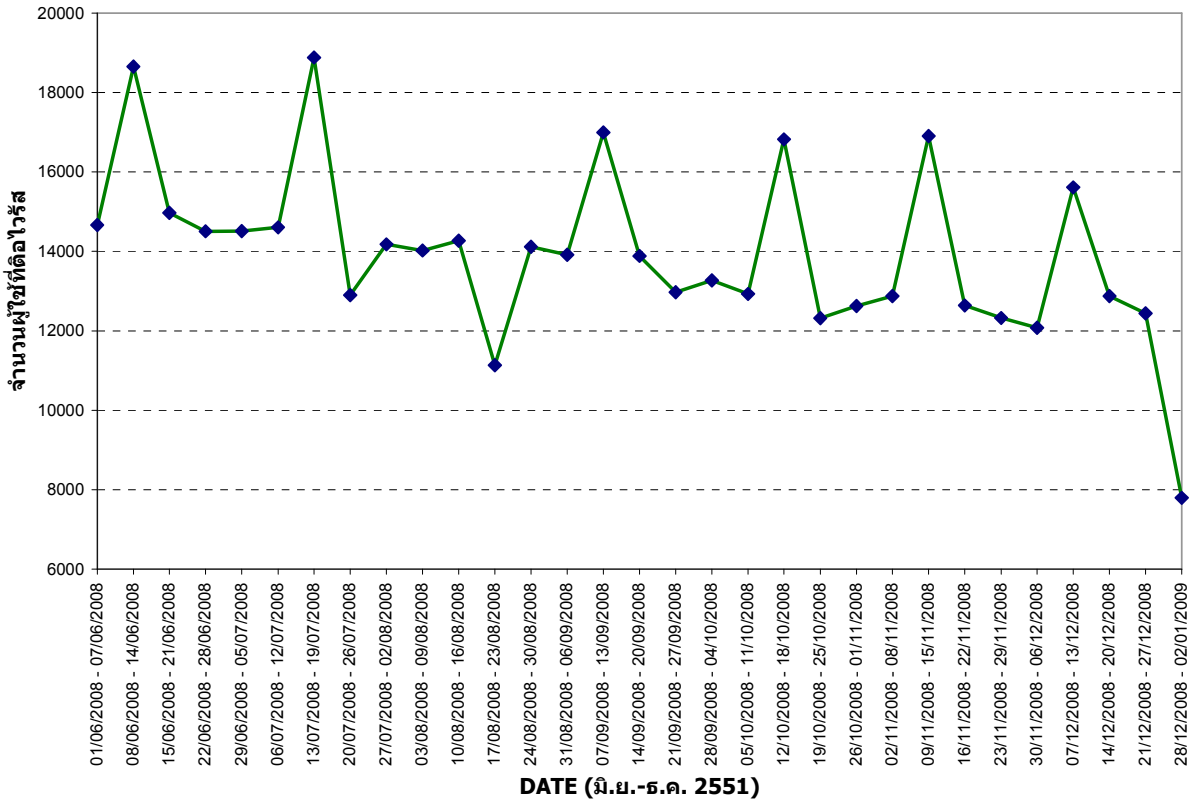
บทที่ 1
บทนำ

1. ความเป็นมาและความสำคัญของปัญหา

บริการเสริมในโครงข่ายโทรศัพท์เคลื่อนที่มีอยู่หลากหลายประเภททั้งด้านเสียง (Voice) และด้านข้อมูล (Data) การรับ-ส่งข้อความมัลติมีเดียเป็นบริการเสริมประเภทหนึ่ง ที่ผู้ใช้บริการสามารถรับและส่งข้อความที่เป็นทั้งตัวอักษร (Text) รูปภาพ วีดีโอ และเสียง เมื่อดูจากสถิติของการส่งข้อความมัลติมีเดียของผู้ให้บริการโครงข่ายโทรศัพท์เคลื่อนที่เบอร์ 1 และเบอร์ 2 อย่างเอไอเอสและดีแทค ซึ่งมีฐานลูกค้ารวมกันแล้วมากกว่าครึ่งหนึ่งของตลาดโทรศัพท์เคลื่อนที่ทั้งหมดของประเทศไทยก็พบว่าตัวเลขการส่งข้อความมัลติมีเดียของทั้งสองค่ายสะท้อนให้เห็นได้เป็นอย่างดีว่า บริการรับ-ส่งข้อความมัลติมีเดียได้รับความนิยมอย่างมาก และยังเป็นความนิยมอย่างก้าวกระโดด โดยผู้ใช้บริการจะมีพฤติกรรมการใช้บริการรับ-ส่งข้อความมัลติมีเดียแตกต่างจากบริการอื่นๆ กล่าวคือ ผู้ใช้บริการจะนิยมใช้บริการรับ-ส่งข้อความมัลติมีเดียมากเป็นพิเศษในช่วงเทศกาล ด้วยลักษณะการส่งภาพข้อความมัลติมีเดียคอนเทนต์น่ารักๆ แทนความรู้สึกดีๆ ให้แก่กันแทนการ์ดที่เป็นกระดาษ ซึ่งการใช้งานรับ-ส่งข้อความมัลติมีเดียในลักษณะนี้มีแนวโน้มเพิ่มสูงขึ้นเรื่อยๆ ทุกปี ดังเช่นเทศกาลปีใหม่ที่เพิ่งผ่านไประหว่างวันที่ 31 ธ.ค. – 1 ม.ค. เพียงแค่ 2 วัน ผู้ใช้บริการของเอไอเอสมียอดส่งข้อความมัลติมีเดียสูงถึง 1 ล้านข้อความ ซึ่งแตะยอด 1 ล้านเป็นครั้งแรก จากปกติที่มียอดการส่งข้อความมัลติมีเดียอยู่ที่ 4 ล้านข้อความต่อเดือนเท่านั้น

และปัญหาหนึ่งที่ผู้ให้บริการโครงข่ายโทรศัพท์เคลื่อนที่กำลังประสบอยู่ ก็คือ ไวรัสโทรศัพท์เคลื่อนที่ ผู้ใช้โทรศัพท์เคลื่อนที่โดยส่วนใหญ่จะติดไวรัสมาจากการดาวน์โหลดโปรแกรม หรือคอนเทนต์จากทางอินเตอร์เน็ตที่โหลดเข้าสู่มือถือ หรือได้รับข้อความมัลติมีเดียที่ติดไวรัสจากผู้อื่น วิธีการแพร่ระบาดของไวรัสข้างต้น สะท้อนให้เห็นว่ายิ่งเทคโนโลยีโทรศัพท์เคลื่อนที่ มีความก้าวหน้ามากยิ่งขึ้นเท่าไหร่ ก็ดูเหมือนว่าช่องทางสำหรับผู้ที่ไม่ประสงค์ดี ก็จะเพิ่มมากขึ้นด้วย โดยเฉพาะเรื่องไวรัสโทรศัพท์เคลื่อนที่ที่ผู้ให้บริการโครงข่ายโทรศัพท์เคลื่อนที่กำลังประสบอยู่ แม้ในช่วงแรกจะติดต่อกันผ่านทาง Bluetooth ในโทรศัพท์เคลื่อนที่ Symbian Smart Phone Series 60 และล่าสุดไวรัสโทรศัพท์เคลื่อนที่ยังสามารถติดต่อกันผ่านทางข้อความมัลติมีเดียได้อีกด้วย โดยไวรัสมือถือที่ว่านี้มีชื่อว่า CommWarrior ซึ่งก็คือ ไวรัสประเภท Worm ที่ทำงานอยู่บนโทรศัพท์เคลื่อนที่ Symbian Series 60 โดยที่ Worm นี้สามารถแพร่กระจายผ่านทาง Bluetooth และข้อความมัลติมีเดียเมื่อ CommWarrior เข้าไปฝังตัวอยู่ในโทรศัพท์เคลื่อนที่แล้ว มันจะเริ่มทำการค้นหาโทรศัพท์เคลื่อนที่เครื่องอื่นที่เปิด

Bluetooth เอาไว้ แล้วทำการส่งไฟล์ไวรัสในรูปแบบของ SIS ไปยังโทรศัพท์เคลื่อนที่เครื่องที่หาเจอ โดยไฟล์ SIS ของไวรัส CommWarrior นี้ จะส่งไปด้วยชื่อของไฟล์ที่ทำการสุ่มขึ้นมา ทำให้ผู้ที่ได้รับไม่สามารถแน่ใจได้ว่าไฟล์นี้เป็นไฟล์ไวรัสหรือไม่ นอกจากที่ CommWarrior จะแพร่กระจายผ่านทาง Bluetooth ไปยังเครื่อง โทรศัพท์เคลื่อนที่อื่นได้แล้ว มันก็ยังจะทำการอ่านข้อมูลหมายเลขโทรศัพท์ภายในเครื่องที่มันไปฝังตัวอยู่ แล้วทำการส่งข้อความมัลติมีเดียไปยังหมายเลขโทรศัพท์เหล่านั้น โดยที่จะมีไฟล์ SIS ของไวรัส CommWarrior แนบไปด้วย ผู้ใช้โทรศัพท์เคลื่อนที่จะไม่รู้ตัวว่าเครื่องของตนเองกำลังทำการส่งข้อความมัลติมีเดียไปยังโทรศัพท์เคลื่อนที่เครื่องอื่นอยู่เรื่อยๆ ซึ่งทำให้ต้องเสียค่าบริการส่งข้อความมัลติมีเดียเพิ่มขึ้นมากกว่าปกติอย่างมาก ซึ่งปัจจุบันมีผู้ใช้มือถือสมาร์ทโฟนประมาณ 10% ที่ตกเป็นเหยื่อไวรัส โดยข้อมูลสถิติผู้ใช้บริการรับ-ส่งข้อความมัลติมีเดียที่ติดไวรัส สามารถดูได้จากภาพประกอบ 1



ภาพประกอบ 1 สถิติผู้ใช้บริการรับ-ส่งข้อความมัลติมีเดียที่ติดไวรัสของผู้ให้บริการโครงข่ายโทรศัพท์เคลื่อนที่เจ้าหนึ่ง

เนื่องด้วยจำนวนผู้ใช้สมาร์ทโฟนเพิ่มขึ้นอย่างรวดเร็ว นั่นหมายความว่า จะมีผู้ใช้ที่ตกเป็นเหยื่อไวรัสมือถือมากขึ้นไปด้วย รวมทั้งยังสร้างปัญหาให้แก่ผู้ให้บริการ จากพฤติกรรมของไวรัสโทรศัพท์เคลื่อนที่ที่ได้กล่าวไปข้างต้น ยกตัวอย่างเช่น โทรศัพท์เคลื่อนที่ที่ติดไวรัสจะส่งข้อความมัลติมีเดียไปให้ผู้อื่นเอง โดยที่ผู้ใช้บริการไม่ได้ส่ง หรือไม่รู้ตัวว่ามีการส่งข้อความมัลติมีเดียออกจากโทรศัพท์ ทำให้ผู้ใช้บริการเสียค่าใช้จ่าย ทั้งที่ตนเองไม่ได้ใช้งาน ในด้านของผู้ให้บริการก็จะได้รับการร้องเรียนปัญหาจากผู้ใช้งานเป็นจำนวนมาก โดยไม่สามารถช่วยเหลือผู้ใช้บริการได้แต่อย่างใด เนื่องจากผู้ให้บริการสามารถรู้แต่เพียงว่าผู้ใช้บริการมีพฤติกรรมการส่งข้อความมัลติมีเดียที่ผิดปกติเท่านั้น และยังเพิ่มปริมาณทราฟฟิกให้กับเครือข่ายโดยไม่จำเป็น รวมทั้งยังต้องสูญเสียค่าใช้จ่ายให้กับบริษัทที่เก็บเงินค่า License ของทราฟฟิกที่วิ่งผ่านระบบบริการรับส่งข้อความมัลติมีเดีย (MMSC System) โดยไร้ประโยชน์ เพราะผู้ใช้บริการส่วนใหญ่ที่ร้องเรียนปัญหาจะไม่ยอมที่จะชำระค่าบริการ และผู้ใช้บริการบางรายที่ไม่เข้าใจปัญหาของไวรัสที่เกิดขึ้น ก็จะยกเลิกบริการทั้งหมดกับผู้ให้บริการโครงข่ายโทรศัพท์เคลื่อนที่เจ้านั้นๆ ซึ่งการสูญเสียผู้ใช้บริการโครงข่ายไปนั้น ถือเป็นเรื่องที่ร้ายแรงมากสำหรับผู้ให้บริการเอง และสิ่งที่ต้องสูญเสียไปโดยไร้ประโยชน์อีกอย่างก็คือ ทรัพยากรบุคคลของผู้ให้บริการที่ทำหน้าที่รับปัญหาร้องเรียนจากผู้ใช้บริการที่ประสบปัญหาไวรัส

ผู้ให้บริการโครงข่ายโทรศัพท์เคลื่อนที่รายหนึ่ง พบปัญหาไวรัสโทรศัพท์เคลื่อนที่ โดยการใช้ Viral Marketing Tool ซึ่งเป็นเครื่องมือที่บริษัทเอกชนรายหนึ่งผลิตขึ้นเพื่อขายให้กับผู้ให้บริการที่ต้องการจะนำมาใช้วิเคราะห์ MMS traffic โดยวัตถุประสงค์ของการใช้เครื่องมือนี้คือ เพื่อหา Content ที่มีการส่งกันมากๆ และผู้ใช้งานที่มีการส่งข้อความมัลติมีเดียมากๆ ซึ่งผลลัพท์ของการทดลองใช้ Viral Marketing Tool วิเคราะห์ทราฟฟิกข้อความมัลติมีเดียพบว่า Content ที่มีการส่งต่อกันมากที่สุดเป็นอันดับ 1-3 คือ ไวรัส CommWarrior โดยจำนวนทั้งหมดของไวรัสที่ Viral Marketing Tool ตรวจพบตั้งแต่เดือนมิถุนายนถึงเดือนธันวาคม พ.ศ.2551 แสดงดังตาราง 1

ตาราง 1 จำนวน Content ที่มีการส่งต่อกันมากที่สุดเป็นอันดับ 1-3

Rank	Content	Frequency Jun-08	Frequency Jul-08	Frequency Aug-08	Total Frequency
1	CommWarrior.A Virus	188,901	205,880	212,000	606,781
2	CommWarrior.B Virus	147,814	145,521	148,752	442,087
3	CommWarrior.C-2 Virus	144,801	142,766	136,799	424,366

จากผลลัพธ์ของการทดลองใช้ Viral Marketing Tool วิเคราะห์ทราฟฟิกข้อความมัลติมีเดียและผลกระทบของปัญหาไวรัสโทรศัพท์เคลื่อนที่ที่ผู้วิจัยได้กล่าวไปข้างต้นทั้งหมดนั้น ทำให้ผู้วิจัยเกิดความสนใจที่จะจัดทำโครงการวิจัยขึ้น เพื่อออกแบบโปรแกรมไว้ใช้ในการตรวจจับข้อความมัลติมีเดียที่มีไวรัสชนิด CommWarrior และผลลัพธ์ของโปรแกรมมีประสิทธิภาพทัดเทียมกับ Viral Marketing Tool โดยที่ผู้ให้บริการโครงข่ายโทรศัพท์เคลื่อนที่ไม่จำเป็นต้องเสียค่าใช้จ่ายในการเก็บรวบรวม Traffic เพราะข้อมูลที่นำมาใช้ในการวิเคราะห์เป็นข้อมูลจากระบบบริการรับส่งข้อความมัลติมีเดียที่มีการเก็บรวบรวมไว้อยู่แล้ว ซึ่งผู้วิจัยมีความคิดเห็นว่าเป็นข้อมูลที่มีความละเอียดเพียงพอ สำหรับการนำไปใช้วิเคราะห์ตรวจจับไวรัส ไม่จำเป็นต้องมีความละเอียดถึงระดับ Packet เนื่องจากผู้ให้บริการโครงข่ายโทรศัพท์เคลื่อนที่ต้องการผลลัพธ์ของการวิเคราะห์ เพียงเพื่อนำไปหาแนวทางในการปรับปรุงบริการเท่านั้น ซึ่งต่างจาก Viral Marketing Tool ที่สามารถตรวจสอบได้ลึกถึงระดับ Packet จึงทำให้สามารถที่จะวิเคราะห์ข้อมูลได้ถึง Content ภายในของข้อความมัลติมีเดีย ซึ่งส่งผลทำให้อัตราการตรวจจับไวรัสมีความถูกต้องสูงมาก แต่ในทางกลับกันด้วยข้อมูลที่ Viral Marketing Tool ต้องเก็บเพื่อนำไปเข้ากระบวนการวิเคราะห์นั้น ทำให้อุปกรณ์ที่ไปเชื่อมต่อเพื่อทำการ Mirror traffic ต้องมีขนาดความจุที่ใหญ่เพียงพอกับการเก็บรวบรวม Traffic ในแต่ละวัน ซึ่งผู้ให้บริการต้องเสียค่าใช้จ่ายเพิ่มขึ้น อีกทั้งยังต้องสูญเสียเวลาในการดึงและอ่านข้อมูลจาก Packet ที่มีขนาดใหญ่ เพื่อนำมาใช้ในการวิเคราะห์อีกด้วย

ในงานวิจัยนี้ยังสนใจศึกษาการนำคลังข้อมูลและเทคนิคการทำเหมืองข้อมูล (Data mining) มาใช้ในการออกแบบโปรแกรม โดยเหตุผลหลักสำหรับการเลือกใช้เทคนิคนี้ เนื่องจากข้อมูลของระบบบริการรับส่งข้อความมัลติมีเดียที่ต้องการการประมวลผลมีอยู่เป็นจำนวนมากทั้งที่มีอยู่แล้วและที่เกิดขึ้นใหม่ จำนวนของข้อมูลที่ถูกเก็บรวบรวมในแต่ละวันของระบบมีขนาดใหญ่ การทำเหมืองข้อมูล (Data mining) จึงเป็นวิธีหนึ่งที่ช่วยในการค้นหา วิเคราะห์ หรือสร้างองค์ความรู้ใหม่จากข้อมูลที่มีขนาดใหญ่ ซึ่งอาจจะเป็นการค้นหารูปแบบหรือกฎ โดยการใช้เทคนิคทางสถิติ ทางคณิตศาสตร์ หรือเทคนิคทางวิทยาการคอมพิวเตอร์ เพื่อเป็นการนำข้อมูลออกมาใช้งานให้เกิดประโยชน์สูงสุด

2. วัตถุประสงค์การวิจัย

2.1 เพื่อออกแบบโปรแกรมสำหรับใช้ในการตรวจจับข้อความมัลติมีเดียที่มีไวรัส โดยการประยุกต์ใช้เทคนิคการทำเหมืองข้อมูล และนำโปรแกรมที่ได้ไปใช้ทดแทน Viral Marketing Tool

2.2 เพื่อนำเทคนิคการทำเหมืองข้อมูล แบบกฎการวิเคราะห์ความสัมพันธ์ (Association Rule) มาประยุกต์ใช้ในการออกแบบโปรแกรม

3. ขอบเขตของการวิจัย

ปริญญานิพนธ์ฉบับนี้ มีเป้าหมายในการออกแบบโปรแกรมเพื่อใช้ในการตรวจจับข้อความมัลติมีเดียที่มีไวรัส การศึกษารายงานที่เคยนำเสนอพบว่ามีความเป็นไปได้ในการใช้เทคนิคการทำเหมืองข้อมูลมาพัฒนาโมเดลต้นแบบ โดยขอบเขตปริญญานิพนธ์เป็นดังนี้

3.1ออกแบบโปรแกรมสำหรับใช้ในการตรวจจับข้อความมัลติมีเดียที่มีไวรัสชนิด CommWarrior โดยใช้เทคนิคการทำเหมืองข้อมูล แบบกฎการวิเคราะห์ความสัมพันธ์ (Association Rule)

3.2 ทดสอบความถูกต้องของโปรแกรม โดยการนำไปใช้วิเคราะห์ข้อมูลกลุ่มตัวอย่างจากระบบบริการรับส่งข้อความมัลติมีเดีย (MMSC System) บนคอมพิวเตอร์ที่มีระบบฐานข้อมูล (Database System) รองรับการทำงานของโปรแกรม

3.3 เปรียบเทียบและวิเคราะห์ผลการตรวจจับจากการใช้โปรแกรมที่ออกแบบขึ้น และผลจากการใช้ Viral Marketing Tool กับข้อมูลในระบบบริการรับส่งข้อความมัลติมีเดีย (MMSC)

4. นิยามศัพท์เฉพาะ

4.1 โทรศัพท์เคลื่อนที่ หมายถึง เครื่องและระบบโทรศัพท์ระบบเซลลูล่าที่อาศัยสัญญาณวิทยุความถี่ใดความถี่หนึ่งในย่านให้บริการโดยทำให้ผู้ใช้บริการเคลื่อนที่ขณะใช้งานในพื้นที่เขตการให้บริการ

4.2 ผู้ให้บริการ หมายถึง ผู้ให้บริการโทรคมนาคมที่ได้รับอนุญาตจากคณะกรรมการกิจการโทรคมนาคมแห่งชาติ รวมถึงผู้ได้รับอนุญาต สัมปทาน หรือสัญญาจากบริษัท กสท โทรคมนาคม จำกัด (มหาชน) และ บริษัท ทีโอที จำกัด (มหาชน) หรือหน่วยงานของผู้ได้รับอนุญาต สัมปทาน หรือสัญญา ที่ได้รับสิทธิหน้าที่ และความรับผิดตามนัยมาตรา ๘๐ วรรค ๒

4.3 ผู้ใช้บริการ หมายถึง ผู้ใช้บริการที่ใช้และเคยใช้บริการรับ-ส่งข้อความมัลติมีเดียผ่านโทรศัพท์เคลื่นที่

4.4 การทำเหมืองข้อมูล หมายถึง กระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูล

4.5 ไวรัสชนิด CommWarrior หมายถึง ไวรัสประเภท Worm ที่ทำงานอยู่บนโทรศัพท์เคลื่อนที่ ระบบปฏิบัติการ Symbian series 60 โดยที่ Worm นี้สามารถแพร่กระจายผ่านทาง Bluetooth และข้อความมัลติมีเดีย (MMS)

5. กรอบแนวคิดในการวิจัย

ในการวิจัยครั้งนี้ ผู้วิจัยมุ่งเน้นพัฒนาโปรแกรมตรวจจับข้อความมัลติมีเดียที่มีไวรัส โดยการประยุกต์ใช้เทคนิคการทำเหมืองข้อมูล แบบกฎการวิเคราะห์ความสัมพันธ์ (Association Rule) เพื่อสามารถนำโปรแกรมที่ได้ไปใช้ทดแทน Viral Marketing Tool โดยอาศัยกรอบแนวคิด ดังนี้

5.1 ศึกษาเทคนิคการทำเหมืองข้อมูล

5.2 ดำเนินการออกแบบและพัฒนาโปรแกรมเพื่อใช้ในการตรวจจับข้อความมัลติมีเดียที่มีไวรัส

5.3 การทดสอบโปรแกรมด้วยข้อมูลกลุ่มตัวอย่างจากระบบบริการรับส่งข้อความมัลติมีเดีย (MMSC System)

6. สมมุติฐานการวิจัย

การออกแบบและพัฒนาโปรแกรมตรวจจับข้อความมัลติมีเดียที่มีไวรัส โดยใช้เทคนิคการทำเหมืองข้อมูล แบบกฎการวิเคราะห์ความสัมพันธ์ (Association Rule) สามารถตรวจจับข้อความมัลติมีเดียที่มีไวรัสได้มีประสิทธิภาพเทียบเท่า Viral Marketing Tool

7. วิธีดำเนินการวิจัย

ในงานวิจัยนี้สามารถแบ่งขั้นตอนในการดำเนินการวิจัยออกเป็น 7 ขั้นตอน ดังนี้

7.1 ศึกษาการทำงานของเหมืองข้อมูล เพื่อให้ทราบว่าวิธีใดเหมาะสมกับข้อมูลในระบบบริการรับส่งข้อความมัลติมีเดีย

7.2 เก็บรวบรวมข้อมูลจากแหล่งข้อมูลที่เกี่ยวข้อง

7.3 ออกแบบและพัฒนาโปรแกรม

7.4 ทดสอบความถูกต้องของโปรแกรมกับข้อมูลกลุ่มตัวอย่าง

7.5 วิเคราะห์ผลการทดสอบโปรแกรม

7.6 เปรียบเทียบและวิเคราะห์ผลการตรวจจับจากการใช้โปรแกรมที่ออกแบบขึ้น และผลจากการใช้ Viral Marketing Tool กับข้อมูลในระบบบริการรับส่งข้อความมัลติมีเดีย

7.7 สรุปผลการดำเนินการวิจัย

8. ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย

8.1 เพื่อทราบรูปแบบของโปรแกรมในการตรวจจับข้อความมัลติมีเดียที่มีไวรัส โดยใช้เทคนิคการทำเหมืองข้อมูล

8.2 สามารถนำโปรแกรมที่ออกแบบขึ้นมาตรวจหาเบอร์โทรศัพท์ที่ติดไวรัส และนำข้อมูลที่ได้ไปใช้แก้ปัญหาไวรัสข้อความมัลติมีเดียให้กับผู้ใช้บริการโทรศัพท์เคลื่อนที่ เช่น ส่ง Antivirus ให้โดยอัตโนมัติ เพื่อลดปัญหาการร้องเรียนจากผู้ใช้งาน

8.3 เพื่อเป็นเครื่องมือให้ผู้ดูแลระบบบริการรับส่งข้อความมัลติมีเดีย (MMSC System) ตรวจจับข้อความมัลติมีเดียที่มีไวรัส

8.4 เพื่อเป็นประโยชน์แก่ผู้ให้บริการโครงข่ายโทรศัพท์เคลื่อนที่ในการปรับปรุงคุณภาพของบริการรับส่งข้อความมัลติมีเดียให้ดีขึ้น

บทที่ 2
ทฤษฎีและผลงานวิจัยที่เกี่ยวข้อง

การศึกษาทฤษฎีและผลงานวิจัยที่เกี่ยวข้องในบทนี้ผู้วิจัยได้ศึกษาจากเอกสาร แนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง ประกอบด้วย

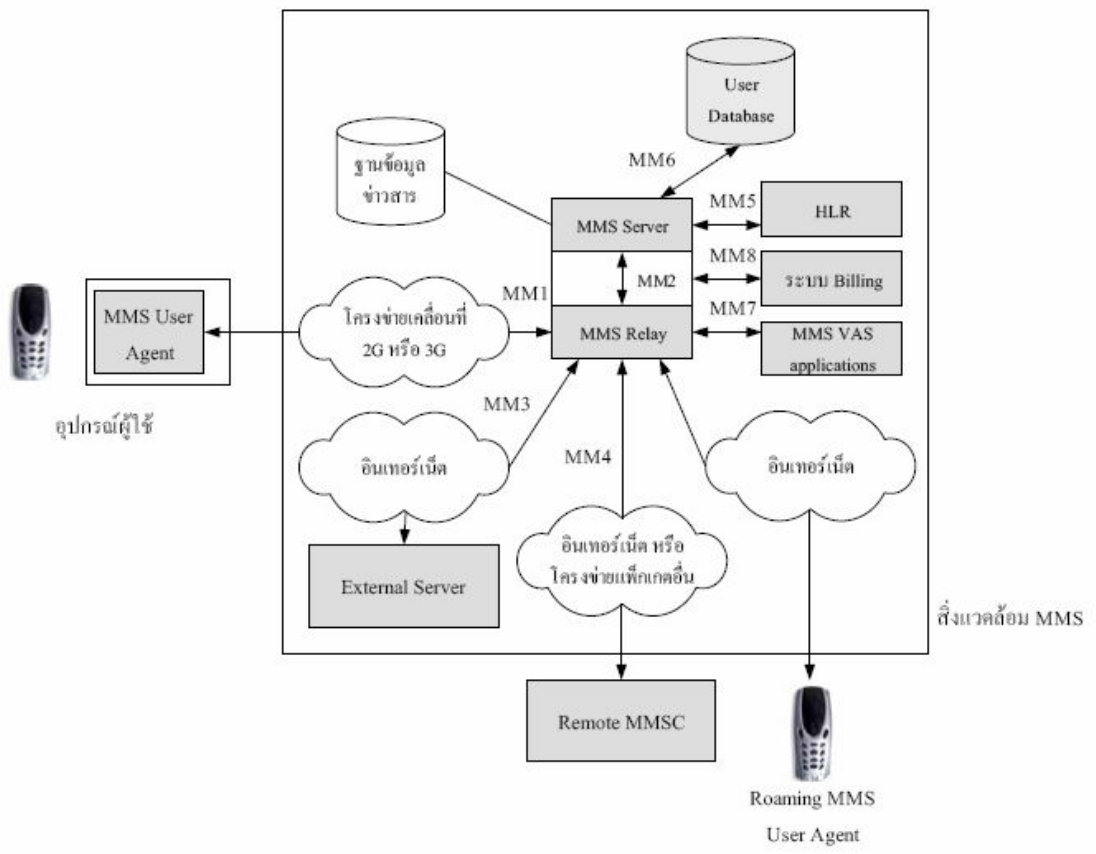
1. สถาปัตยกรรมของระบบบริการรับส่งข้อความมัลติมีเดีย (MMSC)
2. การทำเหมืองข้อมูล (Data mining)
3. วิวัฒนาการของเหมืองข้อมูล
4. รูปแบบในการทำเหมืองข้อมูล
5. เทคนิคของการทำเหมืองข้อมูล
6. ขั้นตอนการทำเหมืองข้อมูล
7. ฐานข้อมูล (Database)
8. งานวิจัยที่เกี่ยวข้อง

1. สถาปัตยกรรมของระบบบริการรับส่งข้อความมัลติมีเดีย (MMSC)

SMS (Short Messaging Service) นับว่าประสบความสำเร็จระดับหนึ่งในโครงข่ายโทรศัพท์เคลื่อนที่ยุคที่ 2 อย่างเช่น ระบบ GSM นอกจากนี้การใช้แบนด์วิดท์ที่สูงขึ้นในโครงข่ายยุคที่ 2.5 และยุคที่ 3 ทำให้เกิดการทางเลือกและบริการใหม่ๆ การบริการข่าวสารชนิดหนึ่งเป็นที่รู้จักกันในนามของ MMS (Multimedia Messaging Service)

MMS ทำให้เกิดการแลกเปลี่ยนข่าวสารระหว่างผู้ใช้และระหว่างอุปกรณ์กับผู้ใช้ MMS ได้เปรียบ SMS ในแง่ที่มีการส่งทั้งข่าวสาร เสียง และภาพได้พร้อมๆ กัน การออกแบบใช้งานสำหรับ MMS ยังคิดไปถึงความสามารถรองรับโปรโตคอลที่มีอยู่ในปัจจุบันและรูปแบบเนื้อหาที่จะใช้กันอย่างกว้างขวางในโลกอินเทอร์เน็ต

3GPP (3rd Generation Partnership Project) และ WAP (Wireless Application Protocol) Forum เป็นกลุ่มที่กำหนดมาตรฐานขั้นตอนต่าง ๆ สำหรับ MMS โดย 3GPP เกี่ยวข้องกับความต้องการบริการขั้นสูง สถาปัตยกรรม MMS โครงสร้างข่าวสาร และรูปแบบเนื้อหา รวมถึงการเลือกใช้ชนิดของการเชื่อมต่อระหว่างอุปกรณ์ในโครงข่ายสื่อสาร ส่วน WAP Forum เกี่ยวข้องกับการเชื่อมต่อระหว่างโทรศัพท์เคลื่อนที่และโครงข่ายบนพื้นฐานของ WAP และโปรโตคอลอินเทอร์เน็ต



ภาพประกอบ 2 สถาปัตยกรรมของระบบบริการรับส่งข้อความมัลติมีเดีย (MMSC)

สถาปัตยกรรม MMS นั้นรวมไปถึง Software Messaging Application ในอุปกรณ์ MMS เพื่อใช้ในการแต่งข้อความ ส่งและรับข่าวสาร นอกจากนั้นยังมีองค์ประกอบอื่นในโครงข่ายที่ใช้ในการให้เส้นทางข่าวสารเพื่อทำการปรับเนื้อหาของข่าวสารให้เหมาะก่อนที่จะทำให้อุปกรณ์สื่อสารรับได้ ในรูปแสดงสถาปัตยกรรมโดยทั่วไปของสิ่งแวดล้อม MMS ซึ่งล้อมรอบด้วยกรอบสี่เหลี่ยมใหญ่ โดยหน่วยนี้จะให้บริการแก่ผู้ใช้ MMS ทั้งนี้หน้าที่ขององค์ประกอบต่าง ๆ แสดงโดยสังเขปได้ดังนี้

- MMS User Agent คือ Software Application ที่อยู่ในโทรศัพท์เคลื่อนที่ซึ่งจะทำให้ผู้ใช้สามารถแต่ง ดู ส่ง และรับข่าวสารมัลติมีเดียได้

- MMS Relay/Server รับผิดชอบต่อการให้เส้นทางข่าวสารภายในและภายนอกสิ่งแวดล้อม MMS ในขณะที่ MMS Server เกี่ยวข้องกับการบันทึกข่าวสารที่รอคอยการรับของผู้ใช้ ปกติการรวมกันของ MMS Relay และ MMS Server เรียกว่า MMSC (MMS Centre)

- การเชื่อมต่อ MM1 ถือว่าเป็นหัวใจของการเชื่อมต่อในสิ่งแวดล้อม MMS เนื่องจากทำให้เกิดการแลกเปลี่ยนข่าวสารระหว่าง MMS User Agent ที่อยู่ในเครื่องของผู้ใช้และ MMSC นอกจากนี้

ปฏิบัติการต่าง ๆ เช่น การลบข่าวสารหรือการรับข่าวสาร สามารถกระทำผ่านการเชื่อมต่อนี้ 3GPP ได้กำหนดความต้องการของการเชื่อมต่อนี้ ส่วน WAP Forum ได้ออกแบบองค์ประกอบที่เกี่ยวข้องกับ MM1

- การเชื่อมต่อ MM2 เป็นการเชื่อมต่อระหว่าง MMS Relay กับ MMS Server ซึ่งการเชื่อมต่อนี้ ถือว่ายังไม่มีมาตรฐาน ดังนั้นการออกแบบจึงขึ้นกับผู้ผลิตแต่ละราย
- การเชื่อมต่อ MM3 เป็นการเชื่อมต่อระหว่าง MMSC กับ External Sever ทำให้เกิดการแลกเปลี่ยนของข่าวสารระหว่าง MMSC และ External Server เช่น Email Server และศูนย์ SMS
- การเชื่อมต่อ MM4 เป็นการเชื่อมต่อระหว่าง MMSC กับ MMSC อื่น เพื่อแลกเปลี่ยนข่าวสารมัลติมีเดียซึ่งกันและกัน
- การเชื่อมต่อ MM5 เป็นการเชื่อมต่อระหว่าง MMSC และองค์ประกอบโครงข่าย เช่น HLR (Home Register) โดยเป็นการส่งผ่านข้อมูลของผู้ใช้แต่ละราย
- การเชื่อมต่อ MM6 เป็นการเชื่อมต่อระหว่าง MMSC และ User Database แต่ ณ ขณะนี้ยังไม่มีมาตรฐาน
- การเชื่อมต่อ MM7 เป็นการเชื่อมต่อระหว่าง MMSC และ VAS (Value Added Service) Application ซึ่งมีผลทำให้ VAS Application สามารถร้องขอบริการจาก MMSC และเพื่อได้รับข่าวสารจาก Remote MMS User Agent ได้
- การเชื่อมต่อ MM8 จำเป็นสำหรับการเชื่อมต่อระหว่าง MMSC และระบบการคิดเงิน แต่ ณ ขณะนี้ก็ยังไม่มีการกำหนดมาตรฐานแต่อย่างใด

2. **การทำเหมืองข้อมูล** (Data mining)

มีผู้ให้คำจำกัดความของการทำเหมืองข้อมูลไว้หลายคำจำกัดความ ดังนี้

บุญเสริม กิจศิริกุล (2545) กล่าวว่า การทำเหมืองข้อมูล หมายถึง กระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูล

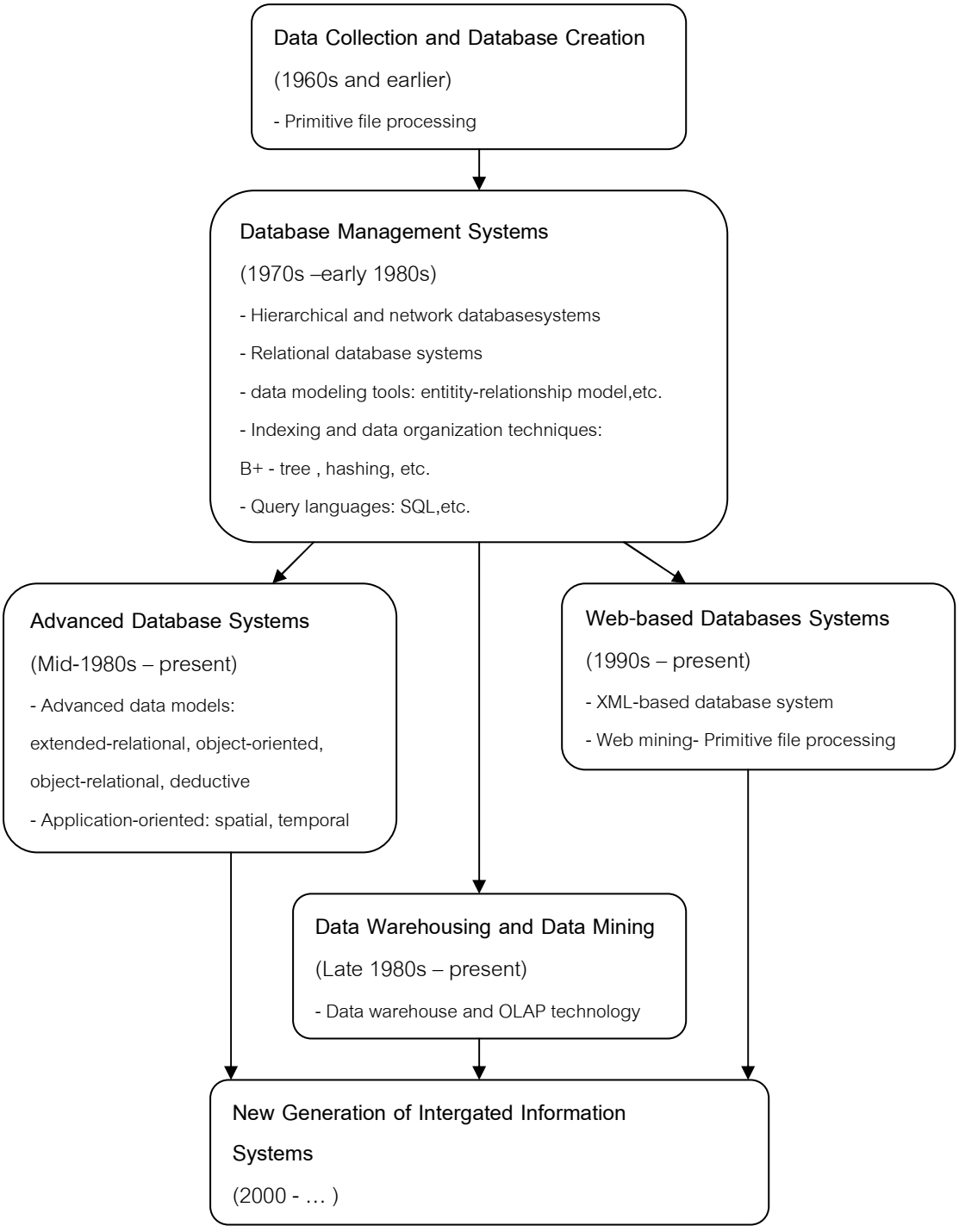
เบอร์รี่ ลินนอฟ (Berry; & Linnoff. 2004: Online) ให้คำจำกัดความของการทำเหมืองข้อมูลไว้ว่าหมายถึง การสำรวจและวิเคราะห์ข้อมูลที่มีขนาดใหญ่เพื่อค้นหารูปแบบหรือกฎที่ซ่อนอยู่ในข้อมูลนั้น และนำความรู้ที่ค้นพบไปใช้ประโยชน์ในการพัฒนาองค์กร

Data Mining & Data Exploration Lab คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (2548) ให้คำจำกัดความของการทำเหมืองข้อมูล ว่าหมายถึง ขบวนการทำงาน (process) ที่สกัดข้อมูล (Extract data) จากฐานข้อมูลขนาดใหญ่ (Large

Information) เพื่อให้ได้สารสนเทศ (Usefull Information) ที่เรายังไม่รู้ (Unknown data) โดยเป็นสารสนเทศที่มีเหตุผล (Valid) และสามารถนำไปใช้ได้ (Actionable) ซึ่งเป็นสิ่งสำคัญในการที่จะช่วยการตัดสินใจในการทำธุรกิจ

การทำเหมืองข้อมูล (Data mining) หรืออาจจะเรียกว่า การค้นหาความรู้ในฐานข้อมูล (Knowledge Discovery in Databases - KDD) เป็นเทคนิคเพื่อค้นหารูปแบบ (pattern) ของจากข้อมูลจำนวนมหาศาลโดยอัตโนมัติ โดยใช้ขั้นตอนวิธีจากวิชาสถิติ การเรียนรู้ของเครื่อง และ การรู้จำแบบ หรือในอีกนิยามหนึ่ง การทำเหมืองข้อมูล คือ กระบวนการที่กระทำกับข้อมูล (โดยส่วนใหญ่จะมีจำนวนมาก) เพื่อค้นหารูปแบบ แนวทาง และความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น โดยอาศัยหลักสถิติ การรู้จำ การเรียนรู้ของเครื่อง และหลักคณิตศาสตร์ (วิกิพีเดีย สารานุกรมเสรี)

3. วิวัฒนาการของเหมืองข้อมูล



ภาพประกอบ 3 วิวัฒนาการของเหมืองข้อมูล

ที่มา: Han, Jiawei and Micheline Kamber, Data Mining Concepts and Techniques. (USA : Morgan Kaufman,2001), 2.

จากภาพประกอบ 3 แสดงให้เห็นถึงวิวัฒนาการของการจัดเก็บและวิเคราะห์ข้อมูล ตั้งแต่ ปี ค.ศ.1960 มีการนำข้อมูลมาจัดเก็บอย่างเหมาะสมในอุปกรณ์ที่น่าเชื่อถือ และป้องกันการสูญหาย (Data Collection and Database Creation) ต่อมามีการพัฒนาระบบฐานข้อมูลให้มีความสามารถในการจัดการข้อมูลด้านต่าง ๆ เช่น การกำหนดความสัมพันธ์ระหว่างฟิลด์ต่าง ๆ ในเรคอร์ด , การจัดการประมวลผล ปรับเปลี่ยน แก้ไขข้อมูล และจัดการกำหนดควบคุมการใช้ข้อมูลที่มีอยู่ได้อย่างเป็นระบบ(Database Management Systems) ในปี ค.ศ. 1970 และในปีช่วงกลางของ ค.ศ. 1980 มีการเพิ่มประสิทธิภาพของระบบฐานข้อมูลโดยการนำข้อมูลที่จัดเก็บมาสร้างความสัมพันธ์เชิงวัตถุเพื่อใช้ในการวิเคราะห์และตัดสินใจที่มีคุณภาพ (Advanced Database Systems) ในช่วงหลังปี ค.ศ. 1980 จนถึงปัจจุบัน มีการรวมรวมข้อมูลมาจัดเก็บลงในฐานข้อมูลขนาดใหญ่ เพื่อช่วยสนับสนุนการตัดสินใจ และนำข้อมูลจากฐานข้อมูลมาวิเคราะห์และประมวลผลโดยการสร้างแบบจำลองและความสัมพันธ์ทางสถิติ ปี ค.ศ.1990 ได้เริ่มมีการจัดทำระบบฐานข้อมูลบนเว็บ (Web-base Databases Systems)

4. รูปแบบในการทำเหมืองข้อมูล

เราสามารถแบ่งรูปแบบการสร้างแบบจำลองในการทำเหมืองข้อมูล ออกเป็น 2 ประเภท คือ

4.1 การสร้างแบบจำลองเพื่อการทำนาย (Predictive Modeling หรือ Supervised Learning)

คือ การนำข้อมูลที่มีอยู่มาใช้ในการทำนายผลข้อมูลในอนาคต ซึ่งการสร้างแบบจำลองรูปแบบนี้จะเน้นการแบ่งข้อมูลออกเป็นกลุ่มตามคุณสมบัติของข้อมูล ในกรณีที่ข้อมูลไม่ต่อเนื่อง จะใช้เทคนิค การจำแนกประเภทข้อมูล (Classification) และในกรณีที่ข้อมูลมีความต่อเนื่องจะใช้เทคนิคการถดถอย (Regression)

4.2 การสร้างแบบจำลองในการบรรยาย (Descriptive Modeling หรือ Unsupervised Learning)

คือ การนำข้อมูลที่มีอยู่มาศึกษา เป็นการเรียนรู้จากข้อมูลที่มีอยู่และอธิบายให้เห็นภาพชัดเจน ซึ่งอาจใช้เทคนิคการหาความสัมพันธ์ (Association) หรือ เทคนิคการจัดกลุ่ม (Clustering)

5. เทคนิคของการทำเหมืองข้อมูล

การแก้ปัญหาของงานชนิดต่างๆ โดยใช้วิธีการทำเหมืองข้อมูล ในแต่ละงานก็จะมีเทคนิคของการทำเหมืองข้อมูลที่จะนำมาใช้ได้อย่างเหมาะสม โดยเทคนิคของการทำเหมืองข้อมูลนั้นมีมากมาย ซึ่งในที่นี้ผู้วิจัยจะขอกล่าวถึงทฤษฎีที่เกี่ยวข้องกับงานวิจัยที่จัดทำขึ้นเท่านั้น

กฎการวิเคราะห์ความสัมพันธ์ (Association rule)

เป็นการหาความสัมพันธ์ระหว่างข้อมูลด้วยกันเองซึ่งมีพื้นฐานมาจากการเกิดขึ้นร่วมกันหรือพร้อมกันในฐานข้อมูล

1. **รูปแบบของการค้นหากฎความสัมพันธ์**

รูปแบบทั่วไปของการค้นหากฎความสัมพันธ์ คือ A → B

โดยที่ A: เป็นเงื่อนไข หรือ LHS (Left - Hand Side)

และ B: เป็นผลลัพธ์ที่เกิดขึ้น หรือ RHS (Right – Hand Side)

หรืออยู่ในรูปของ “ถ้า…….แล้ว” (If……Then….) เช่น

A → B; if A Then B เป็นกฎที่ 1

B → A; if B Then A เป็นกฎที่ 2

การประเมินค่าของกฎจะใช้ค่าสนับสนุน (Support) และค่าความเชื่อมั่น (Confidence)

โดยที่

ค่าสนับสนุน คือ เปอร์เซ็นต์ของข้อมูลที่มีเงื่อนไข และผลลัพธ์สอดคล้องตามกฎต่อจำนวนข้อมูลทั้งหมด สามารถเขียนเป็นสมการดังนี้

$$\text{ค่าสนับสนุน}(A,B) = \frac{\text{จำนวนของ Transaction } (A,B)}{\text{จำนวน Transaction ทั้งหมด}}$$

โดยที่ A หมายถึง เหตุการณ์ที่ใช้เป็นเงื่อนไขในการหาผลลัพธ์

B หมายถึง เหตุการณ์ที่เป็นผลลัพธ์

Transaction (A, B) หมายถึง เหตุการณ์ที่ประกอบด้วยเหตุการณ์ A และ B

ค่าความเชื่อมั่น คือเปอร์เซ็นต์ของข้อมูลที่มีเงื่อนไข และผลลัพธ์สอดคล้องตามกฎต่อจำนวนข้อมูลทั้งหมดที่เป็นเงื่อนไข สามารถเขียนเป็นสมการดังนี้

$$\text{ค่าความเชื่อมั่น (A,B)} = \frac{\text{จำนวนของ Transaction (A,B)}}{\text{จำนวน Transaction (A)}}$$

โดยที่ Transaction (A) หมายถึง เหตุการณ์ที่ประกอบด้วยเหตุการณ์ A อย่างเดียว

ในการเลือกว่าจะกฎใดนั้นจะต้องพิจารณาค่าสนับสนุน และค่าความเชื่อมั่นที่มีค่าสูงกว่าค่า Threshold ที่ตั้งไว้ นอกจากนี้จะต้องกำหนดค่าสนับสนุนต่ำสุด (Minimum Support) และค่าความเชื่อมั่นต่ำสุด (Minimum Confidence) โดยทั่วไปจะกำหนดค่าสนับสนุนต่ำสุดเป็น 5-10% และค่าความเชื่อมั่นต่ำสุดเป็น 50-100%

2. ประเภทของการค้นหากฎความสัมพันธ์

2.1 Boolean Association Rule : เป็นการแสดงความสัมพันธ์ระหว่างสิ่งที่มีอยู่หรือไม่อยู่

buys(X,"computer")→buys(X, "financial_management_software")

หมายความว่า ถ้า X ซื้อ หนังสือ computer แล้ว X จะซื้อซอฟต์แวร์การจัดการการเงิน

2.2 Quantitative Association Rule : เป็นการแสดงความสัมพันธ์ของข้อมูลที่เป็นตัวเลข(Interval) เช่น

age(X,"30…39")∧income(X,"42K…48K")→buys(X, high resolution TV)

หมายความว่า ถ้า X อายุระหว่าง 30-39 และมีรายได้ระหว่าง 42K-48K บาท แล้ว X จะซื้อทีวีที่มีความละเอียดสูง

2.3 Single-Dimension Association Rule : ซึ่งจะอยู่ในรูปแบบดังนี้

buys(X,"computer")→buys(X, "financial_management_software")

2.4 Multi dimension : ซึ่งจะอยู่ในรูปแบบดังนี้

age(X,"30…39")∧income(X,"42K…48K")→buys(X, high resolution TV)

2.5 Multilevel Association Rule : เป็นกฎที่สร้างจากแอทริบิวต์ ที่มีระดับต่างกัน

Age(X,"30-39") → (X,"laptop_computer")

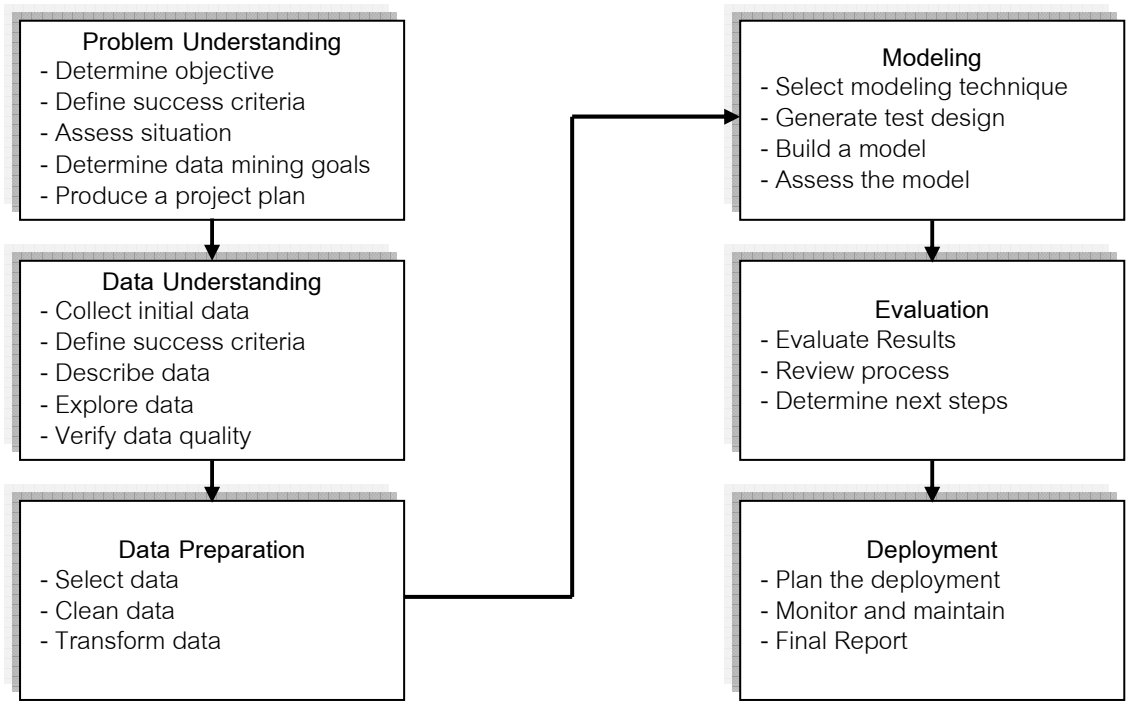
Age(X,"30-39") → (X,"Computer")

6. ขั้นตอนการทำเหมืองข้อมูล

ในการทำเหมืองข้อมูลนั้นต้องใช้ความรู้จากศาสตร์หลายแขนง ซึ่งประกอบด้วยความรู้ทางด้านต่างๆ ต่อไปนี้

- ฐานข้อมูล (Database technology) เป็นแหล่งในการจัดเก็บ และรวบรวมข้อมูล
- สถิติ (Statistics) ใช้สำหรับวิเคราะห์ข้อมูลทางสถิติเบื้องต้น
- การเรียนรู้ของเครื่อง (Machine learning) เป็นอัลกอริทึมที่ใช้ในการค้นหารูปแบบและความสัมพันธ์ของข้อมูล
- การมองเห็น (Visualization) เป็นการแสดงผลลัพธ์ และ ความสัมพันธ์เพื่อให้ผู้ใช้เข้าใจได้ง่าย
- ความรู้ด้านวิทยาศาสตร์ (Information science) เป็นความรู้ทางด้านวิทยาศาตร์อื่นๆ ที่เกี่ยวข้อง

ขั้นตอนการทำเหมืองข้อมูลประกอบด้วยขั้นตอนหลัก 6 ขั้นตอน ดังนี้



ภาพประกอบ 4 ขั้นตอนการทำเหมืองข้อมูล

ที่มา : บุญเสริม กิจศิริกุล, “รายงานวิจัยฉบับสมบูรณ์,” โครงการวิจัยร่วมภาครัฐและเอกชน ปีงบประมาณ 2545 โครงการย่อยที่ 7 อัลกอริทึมการทำเหมืองข้อมูล ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, 2546.

6.1 การทำความเข้าใจกับปัญหา (Problem Understanding)

ขั้นตอนการทำความเข้าใจปัญหาประกอบด้วย การตั้งวัตถุประสงค์ของการทำเหมืองข้อมูล (Determine objective) ตั้งเกณฑ์วัดความสำเร็จ (Define success criteria) ประเมินสถานการณ์ในด้านต่างๆ (Assess situation) ระบุเป้าหมายที่ใช้ในการตัดสินใจทำเหมืองข้อมูล (Determine data mining goals) วางแผนการทำเหมืองข้อมูล (Produce a project plan) ว่าจะเก็บข้อมูลด้วยวิธีใด และใช้อัลกอริทึมไหน

6.2 การทำความเข้าใจข้อมูล (Data Understanding)

ขั้นตอนการทำความเข้าใจข้อมูล ประกอบด้วยรวบรวมข้อมูล (Collect initial data) กำหนดคุณสมบัติของข้อมูล (Define success criteria) อธิบายรายละเอียดของข้อมูล (Describe data) การสำรวจข้อมูล (Explore data) การตรวจสอบความถูกต้องและความสมบูรณ์ของข้อมูล (Verify data quality)

6.3 การเตรียมข้อมูล (Data Preparation)

ขั้นตอนการเตรียมข้อมูล ประกอบด้วยการคัดเลือกข้อมูลที่จะนำมาใช้ (Select data)

- การทำความสะอาดข้อมูล (Clean data) ซึ่งเป็นกระบวนการเตรียมข้อมูลให้เหมาะสมที่สุดเพื่อนำไปใช้ในขั้นตอนต่อไป
- แก้ไขข้อมูลให้ถูกต้องสมบูรณ์ เช่น การแก้ไขค่าว่างของข้อมูลโดยใส่ค่า 9 (เก้า)
- ปรับเปลี่ยนข้อมูลให้มีค่าที่เหมาะสมในการตัดสินใจ เช่น ข้อมูลที่มีค่า “มาม่า” และ “ไวไว” อาจเปลี่ยนค่าเป็น “บะหมี่กึ่งสำเร็จรูป”
- เลือกข้อมูลเฉพาะที่สนใจ เช่น ต้องการหาลักษณะลูกค้าที่ซื้อรถเก๋ง ไม่ควรนำรายชื่อพนักงานขายเข้ามาเกี่ยวข้อง
- คอลัมน์ที่มีค่าสำหรับทุกแถวเป็นค่าเดียว เช่น “สัญชาติไทย” หรือ คอลัมน์ที่มีค่าไม่ซ้ำกันเลย เช่น “หมายเลขสมาชิก” ไม่ควรนำมาใช้ เนื่องจากไม่สามารถบอกรูปแบบของข้อมูลได้

การปรับเปลี่ยนรูปแบบข้อมูล (Transform data) เช่น นำตารางในฐานข้อมูลมาเชื่อมต่อกัน ขั้นตอนนี้เป็นขั้นตอนที่สำคัญมากเนื่องจากความถูกต้อง และสมบูรณ์ของผลลัพธ์สุดท้ายซึ่งขึ้นอยู่กับว่านักวิเคราะห์ข้อมูลนั้นตัดสินใจกำหนดโครงสร้างและเสนอลักษณะของข้อมูลที่จะใช้ในการ

ประมวลผลอย่างไรอย่างไร กรรมวิธีนี้ รวมไปถึงการทำ Data recording (การจัดเก็บข้อมูล) และ Data Format Conversion (รูปแบบในการแปลงข้อมูล) เช่น การแปลงเวลา UNIX timestamp เป็นเวลาปัจจุบันเป็นต้น

6.4 การสร้างแบบจำลอง (Modeling)

ขั้นตอนการสร้างแบบจำลอง ประกอบด้วย การเลือกเทคนิคที่เหมาะสมในการทำเหมืองข้อมูล (Select modeling technique) กำหนดรูปแบบการทดสอบผลลัพธ์ (Generate test design) สร้างแบบจำลองตามเทคนิคที่เลือก (Model Building) ทดสอบความถูกต้องและความน่าเชื่อถือของแบบจำลองที่สร้างขึ้น (Model Assessing)

6.5 การประเมินผล (Evaluation)

ขั้นตอนการประเมิน ประกอบด้วย การประเมินผลที่ได้จากการทดลอง (Evaluate Results) อาจจะเป็นการประเมินแบบจำลองที่สร้างขึ้นด้วยการลองนำไปใช้กับสถานการณ์จริงเพื่อตรวจสอบประสิทธิภาพของแบบจำลอง การทบทวนกระบวนการ (review process) ใช้เป็นขั้นตอนถัดไปในการตัดสินใจ (Determine next steps)

6.6 การนำไปใช้ (Deployment)

ขั้นตอนการนำไปใช้ ประกอบด้วย แผนการในการนำไปใช้ (Plan the deployment) และสรุปผลของการทดลอง

7. ฐานข้อมูล (Database)

ในปัจจุบันการจัดโครงสร้างข้อมูลให้เป็นแบบฐานข้อมูลกำลังเป็นที่นิยม เกือบทุกหน่วยงานที่มีการใช้ระบบสารสนเทศจะจัดทำข้อมูลให้เป็นแบบฐานข้อมูล เนื่องจากปริมาณข้อมูลมีมากถ้าจัดข้อมูลเป็นแบบแฟ้มข้อมูลจะทำให้มีแฟ้มข้อมูลเป็นจำนวนมาก ซึ่งจะทำให้เกิดข้อมูลที่ซ้ำซ้อนกันได้ ข้อมูลที่ซ้ำซ้อนนี้จะก่อให้เกิดปัญหามากมาย

7.1 ความหมายของระบบฐานข้อมูล

ฐานข้อมูล (database) หมายถึง กลุ่มของข้อมูลที่ถูกเก็บรวบรวมไว้ โดยมีความสัมพันธ์ซึ่งกันและกัน โดยไม่ได้บังคับว่าข้อมูลทั้งหมดนี้จะต้องเก็บไว้ในแฟ้มข้อมูลเดียวกันหรือแยกเก็บหลาย ๆ แฟ้มข้อมูล นั่นก็คือการเก็บข้อมูลในฐานข้อมูลนั้นเราอาจจะเก็บทั้งฐานข้อมูล โดยใช้แฟ้มข้อมูลเพียงแฟ้มข้อมูลเดียวกันได้ หรือจะเก็บไว้ในหลาย ๆ แฟ้มข้อมูล ที่สำคัญคือจะต้องสร้างความสัมพันธ์ระหว่างระเบียนและเรียกใช้ความสัมพันธ์นั้นได้ มีการกำจัดความซ้ำซ้อนของข้อมูลออกและเก็บแฟ้มข้อมูลเหล่านี้ไว้ที่ศูนย์กลาง เพื่อที่จะนำข้อมูลเหล่านี้มาใช้ร่วมกัน ควบคุมดูแลรักษาเมื่อผู้ต้องการใช้งานและผู้มีสิทธิ์จะใช้ข้อมูลนั้นสามารถดึงข้อมูลที่ต้องการออกไปใช้ได้ ข้อมูลบางส่วนอาจใช้ร่วมกับ

ผู้อื่นได้ แต่บางส่วนผู้มีสิทธิ์เท่านั้นจึงจะสามารถใช้ได้ โดยทั่วไปองค์กรต่าง ๆ จะสร้างฐานข้อมูลไว้เพื่อเก็บข้อมูลต่าง ๆ ของตัวองค์กร โดยเฉพาะอย่างยิ่งข้อมูลในเชิงธุรกิจ เช่น ข้อมูลของลูกค้า ข้อมูลของสินค้า ข้อมูลของลูกจ้าง และการจ้างงาน เป็นต้น การควบคุมดูแลการใช้ฐานข้อมูลนั้น เป็นเรื่องที่ยุ่งยากกว่าการใช้แฟ้มข้อมูลมาก เพราะเราจะต้องตัดสินใจว่าโครงสร้างในการจัดเก็บข้อมูลควรจะเป็นเช่นไร การเขียนโปรแกรมเพื่อสร้างและเรียกใช้ข้อมูลจากโครงสร้างเหล่านี้ ถ้าโปรแกรมเหล่านี้เกิดทำงานผิดพลาดขึ้นมา ก็จะเกิดความเสียหายต่อโครงสร้างของข้อมูลทั้งหมดได้ เพื่อเป็นการลดภาวะการทำงานของผู้ใช้ จึงได้มีส่วนของฮาร์ดแวร์และโปรแกรมต่าง ๆ ที่สามารถเข้าถึงและจัดการข้อมูลในฐานข้อมูลนั้น เรียกว่า ระบบจัดการฐานข้อมูล หรือDBMS (data base management system) ระบบจัดการฐานข้อมูล คือ ซอฟต์แวร์ที่เปรียบเสมือนสื่อกลางระหว่างผู้ใช้และโปรแกรมต่าง ๆ ที่เกี่ยวข้องกับการใช้ฐานข้อมูล ซึ่งมีหน้าที่ช่วยให้ผู้ใช้เข้าถึงข้อมูลได้ง่ายสะดวกและมีประสิทธิภาพ การเข้าถึงข้อมูลของผู้ใช้อาจเป็นการสร้างฐานข้อมูล การแก้ไขฐานข้อมูล หรือการตั้งคำถามเพื่อให้ข้อมูลมา โดยผู้ใช้ไม่จำเป็นต้องรับรู้เกี่ยวกับรายละเอียดภายในโครงสร้างของฐานข้อมูล เปรียบเสมือนเป็นสื่อกลางระหว่างผู้ใช้และโปรแกรมต่าง ๆ ที่เกี่ยวข้องกับการใช้ฐานข้อมูล

7.2 ความสำคัญของระบบฐานข้อมูล

การจัดข้อมูลให้เป็นระบบฐานข้อมูลทำให้ข้อมูลมีส่วนดีกว่าการเก็บข้อมูลในรูปของแฟ้มข้อมูล เพราะการจัดเก็บข้อมูลในระบบฐานข้อมูล จะมีส่วนที่สำคัญกว่าการจัดเก็บข้อมูลในรูปของแฟ้มข้อมูลดังนี้

7.2.1 ลดการเก็บข้อมูลที่ซ้ำซ้อน ข้อมูลบางชุดที่อยู่ในรูปของแฟ้มข้อมูลอาจมีปรากฏอยู่หลาย ๆ แห่ง เพราะมีผู้ใช้ข้อมูลชุดนี้หลายคน เมื่อใช้ระบบฐานข้อมูลแล้วจะช่วยให้ความซ้ำซ้อนของข้อมูลลดน้อยลง เช่น ข้อมูลอยู่ในแฟ้มข้อมูลของผู้ใช้หลายคน ผู้ใช้แต่ละคนจะมีแฟ้มข้อมูลเป็นของตนเอง ระบบฐานข้อมูลจะลดการซ้ำซ้อนของข้อมูลเหล่านี้ให้มากที่สุด โดยจัดเก็บในฐานข้อมูลไว้ที่เดียวกัน ผู้ใช้ทุกคนที่ต้องการใช้ข้อมูลชุดนี้จะใช้โดยผ่านระบบฐานข้อมูล ทำให้ไม่เปลืองเนื้อที่ในการเก็บข้อมูลและลดความซ้ำซ้อนลงได้

7.2.2 รักษาความถูกต้องของข้อมูล เนื่องจากฐานข้อมูลมีเพียงฐานข้อมูลเดียว ในกรณีที่มีข้อมูลชุดเดียวกันปรากฏอยู่หลายแห่งในฐานข้อมูล ข้อมูลเหล่านี้จะต้องตรงกัน ถ้ามีการแก้ไขข้อมูลนี้ทุก ๆ แห่งที่ข้อมูลปรากฏอยู่จะแก้ไขให้ถูกต้องตามกันหมดโดยอัตโนมัติด้วยระบบจัดการฐานข้อมูล

7.2.3 การป้องกันและรักษาความปลอดภัยให้กับข้อมูลทำได้อย่างสะดวก การป้องกันและรักษาความปลอดภัยกับข้อมูลระบบฐานข้อมูลจะให้เฉพาะผู้ที่ เกี่ยวข้องเท่านั้นจึงจะมีสิทธิ์เข้าไปใช้ฐานข้อมูลได้เรียกว่ามีสิทธิส่วน บุคคล (privacy) ซึ่งก่อให้เกิดความปลอดภัย (security) ของข้อมูล

ด้วย ฉะนั้นผู้ใดจะมีสิทธิ์ที่จะเข้าถึงข้อมูลได้จะต้องมีการกำหนดสิทธิ์กันไว้ ก่อนและเมื่อเข้าไปใช้ข้อมูลนั้น ๆ ผู้ใช้จะเห็นข้อมูลที่ถูกเก็บไว้ในฐานข้อมูลในรูปแบบที่ผู้ใช้ออกแบบไว้
ตัวอย่างเช่น ผู้ใช้สร้างตารางข้อมูลขึ้นมาและเก็บลงในระบบฐานข้อมูล ระบบจัดการฐานข้อมูลจะเก็บข้อมูลเหล่านี้ลงในอุปกรณ์เก็บข้อมูลในรูปแบบของ ระบบจัดการฐานข้อมูลซึ่งอาจเก็บข้อมูลเหล่านี้ลงในแผ่นจานบันทึกแม่เหล็ก เป็นระเบียน บล็อกหรืออื่น ๆ ผู้ใช้ไม่จำเป็นต้องรับรู้ว่าโครงสร้างของแฟ้มข้อมูลนั้นเป็นอย่างไร ปล่อยให้เป็นหน้าที่ของระบบจัดการฐานข้อมูล

ดังนั้นถ้าผู้ใช้เปลี่ยนแปลงลักษณะการเก็บข้อมูล เช่น เปลี่ยนแปลงรูปแบบของตารางเสียใหม่ ผู้ใช้ก็ไม่ต้องกังวลว่าข้อมูลของเขาจะถูกเก็บลงในแผ่นจานบันทึกแม่เหล็กในลักษณะใด ระบบการจัดการฐานข้อมูลจะจัดการให้ทั้งหมด ในทำนองเดียวกันถ้าผู้ออกแบบระบบฐานข้อมูลเปลี่ยนวิธีการเก็บข้อมูลลงบนอุปกรณ์จัดเก็บข้อมูล ผู้ใช้ก็ไม่ต้องแก้ไขฐานข้อมูลที่เขาออกแบบไว้แล้ว ระบบการจัดการฐานข้อมูลจะจัดการให้ ลักษณะเช่นนี้เรียกว่า ความไม่เกี่ยวข้องกันของข้อมูล (Data independent)

7.2.4 สามารถใช้ข้อมูลร่วมกันได้ เนื่องจากในระบบฐานข้อมูลจะเป็นที่เก็บรวบรวมข้อมูลทุกอย่างไว้ ผู้ใช้แต่ละคนจึงสามารถที่จะใช้ข้อมูลในระบบได้ทุกข้อมูล ซึ่งถ้าข้อมูลไม่ได้ถูกจัดให้เป็นระบบฐานข้อมูลแล้ว ผู้ใช้ก็จะใช้ได้เพียงข้อมูลของตนเองเท่านั้น เช่น ดังภาพที่ 4.9 ข้อมูลของระบบเงินเดือน ข้อมูลของระบบงานบุคคลถูกจัดไว้ในระบบแฟ้มข้อมูลผู้ใช้ที่ใช้ข้อมูลระบบเงินเดือนจะใช้ข้อมูลได้ระบบเดียว แต่ถ้าข้อมูลทั้ง 2 ถูกเก็บไว้เป็นฐานข้อมูลซึ่งถูกเก็บไว้ในที่ที่เดียวกัน ผู้ใช้ทั้ง 2 ระบบก็จะสามารถเรียกใช้ฐานข้อมูลเดียวกันได้ ไม่เพียงแต่ข้อมูลเท่านั้นสำหรับโปรแกรมต่าง ๆ ถ้าเก็บไว้ในฐานข้อมูลก็จะสามารถใช้ร่วมกันได้

7.2.5 มีความเป็นอิสระของข้อมูล เมื่อผู้ใช้ต้องการเปลี่ยนแปลงข้อมูลหรือนำข้อมูลมาประยุกต์ใช้ให้เหมาะสมกับโปรแกรมที่เขียนขึ้นมา จะสามารถสร้างข้อมูลนั้นขึ้นมาใช้ใหม่ได้ โดยไม่มีผลกระทบต่อระบบฐานข้อมูล เพราะข้อมูลที่ผู้ใช้นำมาประยุกต์ใช้ใหม่นั้นจะไม่กระทบต่อโครงสร้างที่แท้จริงของการจัดเก็บข้อมูล นั่นคือ การใช้ระบบฐานข้อมูลจะทำให้เกิดความเป็นอิสระระหว่างการจัดเก็บข้อมูลและการประยุกต์ใช้

7.2.6 สามารถขยายงานได้ง่าย เมื่อต้องการจัดเพิ่มเติมข้อมูลที่เกี่ยวข้องจะสามารถเพิ่มได้อย่างง่ายไม่ซับซ้อน เนื่องจากมีความเป็นอิสระของข้อมูล จึงไม่มีผลกระทบต่อข้อมูลเดิมที่มีอยู่

7.2.7 ทำให้ข้อมูลบูรณะกลับสู่สภาพปกติได้เร็วและมีมาตรฐาน เนื่องจากการจัดพิมพ์ข้อมูลในระบบที่ไม่ได้ใช้ฐานข้อมูล ผู้เขียนโปรแกรมแต่ละคนมีแฟ้มข้อมูลของตนเองเฉพาะ ฉะนั้นแต่ละคนจึงต่างก็สร้างระบบการบูรณะข้อมูลให้กลับสู่สภาพปกติในกรณีที่ข้อมูลเสียหายด้วยตนเองและด้วยวิธีการของตนเอง จึงขาดประสิทธิภาพและมาตรฐาน แต่เมื่อมาเป็นระบบฐานข้อมูลแล้ว การ

บูรณะข้อมูลให้กลับคืนสู่สภาพปกติจะมีโปรแกรมชุดเดียวและมีผู้ดูแลเพียงคนเดียวที่ดูแลทั้งระบบ ซึ่งย่อมต้องมีประสิทธิภาพและเป็นมาตรฐานเดียวกันแน่นอน

7.3 การบริหารฐานข้อมูล

ในระบบฐานข้อมูลนอกจากจะมีระบบการจัดการฐานข้อมูล ซึ่งเป็นซอฟต์แวร์ที่สร้างขึ้นเพื่อจัดการกับข้อมูลให้เป็นระบบ จะได้นำไปเก็บรักษา เรียกใช้ หรือนำมาปรับปรุงให้ทันสมัยได้ง่ายแล้ว ในระบบฐานข้อมูลยังต้องประกอบด้วยบุคคลที่มีหน้าที่ควบคุมดูแลระบบฐานข้อมูล คือ ผู้บริหารฐานข้อมูล

เหตุผลสำหรับประการหนึ่งของการจัดทำระบบจัดการฐานข้อมูล คือ การมีศูนย์กลางควบคุมทั้งข้อมูลและโปรแกรมที่เข้าถึงข้อมูลเหล่านั้น บุคคลที่มีอำนาจหน้าที่ดูแลการควบคุมนี้เรียกว่า ผู้บริหารฐานข้อมูล หรือ DBA (Database administer) คือ ผู้มีหน้าที่ควบคุมการบริหารงานของฐานข้อมูลทั้งหมด

7.4 หน้าที่ของผู้บริหารฐานข้อมูล

7.4.1 กำหนดโครงสร้างหรือรูปแบบของฐานข้อมูล โดยทำการวิเคราะห์และตัดสินใจว่าจะรวมข้อมูลใดเข้าไว้ในระบบใดบ้าง ควรจะจัดเก็บข้อมูลด้วยวิธีใด และใช้เทคนิคใดในการเรียกใช้ข้อมูลอย่างไร

7.4.2 กำหนดโครงสร้างของอุปกรณ์เก็บข้อมูลและวิธีการเข้าถึงข้อมูล โดยกำหนดโครงสร้างของอุปกรณ์เก็บข้อมูลและวิธีการเข้าถึงข้อมูล พร้อมทั้งกำหนดแผนการในการสร้างระบบข้อมูลสำรองและการฟื้นสภาพ โดยการจัดเก็บข้อมูลสำรองไว้ทุกระยะ และจะต้องเตรียมการไว้ว่าถ้าเกิดความผิดพลาดขึ้นแล้วจะทำการฟื้นสภาพได้อย่างไร

7.4.3 มอบหมายขอบเขตอำนาจหน้าที่ของการเข้าถึงข้อมูลของผู้ใช้ โดยการประสานงานกับผู้ใช้ ให้คำปรึกษา ให้ความช่วยเหลือแก่ผู้ใช้ และตรวจตราความต้องการของผู้ใช้

7.5 ระบบการจัดการฐานข้อมูล (Database management system, DBMS)

หน้าที่ของระบบการจัดการฐานข้อมูล

7.5.1 ระบบจัดการฐานข้อมูลเป็นซอฟต์แวร์ที่ทำหน้าที่ดังต่อไปนี้ ดูแลการใช้งานให้กับผู้ใช้ในการติดต่อกับตัวจัดการระบบแฟ้มข้อมูลได้ ในระบบฐานข้อมูลนี้ข้อมูลจะมีขนาดใหญ่ ซึ่งจะถูกจัดเก็บไว้ในหน่วยความจำสำรองเมื่อผู้ใช้ต้องการจะใช้ฐานข้อมูล ระบบการจัดการฐานข้อมูลจะทำหน้าที่ติดต่อกับระบบแฟ้มข้อมูลซึ่งเสมือนเป็นผู้จัดการแฟ้มข้อมูล (File manager) นำข้อมูลจากหน่วยความจำสำรองเข้าสู่หน่วยความจำหลักเฉพาะส่วนที่ต้องการใช้งาน และทำหน้าที่ประสานกับตัวจัดการระบบแฟ้มข้อมูลในการจัดเก็บ เรียกใช้ และแก้ไขข้อมูล

7.5.2 ควบคุมระบบความปลอดภัยของข้อมูลโดยป้องกันไม่ให้ผู้ที่ไม่ได้รับอนุญาตเข้ามาเรียกใช้หรือแก้ไขข้อมูลในส่วนป้องกันเอาไว้ พร้อมทั้งสร้างฟังก์ชันในการจัดทำข้อมูลสำรอง โดยเมื่อเกิดมีความขัดข้องของระบบแฟ้มข้อมูลหรือของเครื่องคอมพิวเตอร์เกิดการเสียหายนั้น ฟังก์ชันนี้จะสามารถทำการฟื้นสภาพของระบบข้อมูลกลับเข้าสู่สภาพที่ถูกต้องสมบูรณ์ได้

7.5.3 ควบคุมการใช้ข้อมูลในสภาพที่มีผู้ใช้พร้อม ๆ กันหลายคน โดยจัดการเมื่อมีข้อผิดพลาดของข้อมูลเกิดขึ้น

8. งานวิจัยที่เกี่ยวข้อง

8.1 งานวิจัยในประเทศ

กฤษณะ ไวยมัย และ ธีระวัฒน์ พงษ์ศิริปรีดา (กฤษณะ ไวยมัย; ธีระวัฒน์ พงษ์ศิริปรีดา. 2546: บทคัดย่อ) นำเทคนิคการจำแนกประเภทข้อมูลมาสร้างตัวจำแนกข้อมูล ในการวิจัยได้นำเทคนิคการจำแนกประเภท และเทคนิคการค้นหากฎความสัมพันธ์ มาประยุกต์ใช้ในการจัดสรรกฎหมายที่เหมาะสมกับคดีความ โดยนำเทคนิคเทคนิคการจำแนกประเภทข้อมูลมาสร้างตัวจำแนกข้อมูลจากกฎเกณฑ์ที่ได้จากเทคนิคการค้นหากฎความสัมพันธ์อีกทีหนึ่ง ซึ่งตัวจำแนกข้อมูลที่ได้นี้จะสามารถนำไปใช้ทำนายคดีความแต่ละคดีว่าควรใช้กฎหมายฉบับใดในการพิจารณา ผลการวิจัยที่ได้แสดงให้เห็นว่าการสร้างตัวจำแนกตามวิธีที่เสนอนี้ได้ประสิทธิภาพดีกว่าการสร้างตัวจำแนกตามเทคนิค Data classification แบบปกติ นอกจากนี้วิธีดังกล่าวยังช่วยแก้ปัญหาต่างๆที่เกิดขึ้นจากการใช้เทคนิคการจำแนกประเภทข้อมูลโดยทั่วไปได้อีกด้วย

พิจิตรา จอมศรี และ ปานใจ ธารทัศนวงศ์(พิจิตรา จอมศรี; ปานใจ ธารทัศนวงศ์. 2549: บทคัดย่อ) ได้นำเอาเทคนิคการทำเหมืองข้อมูลด้วยกฎการค้นหาความสัมพันธ์มาใช้ในการทำนายข้อมูลการเรียกใช้เว็บในอนาคต โดยการนำเอาข้อมูลการใช้เว็บภายในมหาวิทยาลัยศิลปากร มาค้นหาความสัมพันธ์ที่ได้จากข้อมูลการใช้เว็บเพื่อใช้ทำนายเว็บที่ถูกเรียกใช้ในอนาคต ผลของการใช้เทคนิคเหมืองข้อมูลพบว่า โมเดลที่สร้างขึ้นสามารถทำนายเนื้อหาเว็บที่จะถูกเรียกใช้ได้และสามารถเพิ่มอัตราการพบในระบบพร็อกซี ได้จากเดิม 38.15% เป็น 54.54% ซึ่งถ้าอัตราการพบในระบบพร็อกซีเพิ่มขึ้น อาจทำให้ประสิทธภาพการเรียกใช้เว็ปเพิ่มขึ้นและสามารถลดปริมาณข้อมูลในระบบเครือข่ายได้ อย่างไรก็ตามการใช้เทคนิคนี้ยังไม่สามารถครอบคลุมการทำงานในช่วงเหตุการณ์ที่ไม่เป็นปกติได้ เช่น อุบัติภัย และเทศกาลต่างๆ เป็นต้น

ดุลย์วิทย์ ปรางชุมพล และ ปานใจ ธารทัศนวงศ์(ดุลย์วิทย์ ปรางชุมพล; ปานใจ ธารทัศนวงศ์. 2549: บทคัดย่อ) ได้ทำการศึกษาการทำนายปริมาณการจราจรในเครือข่ายโดยใช้เทคนิค

เหมืองข้อมูลมาวิเคราะห์ปริมาณการจราจรในระบบเครือข่ายและทำการสร้างตัวแบบ ซึ่งสามารถทำนายแนวโน้มของปริมาณการจราจร เพื่อเป็นแนวทางในการตัดสินใจที่จะเลือกใช้เส้นทางในการจัดส่งข้อมูลต่อไปได้ เทคนิคเหมืองข้อมูลที่นำมาประยุกต์ใช้ในงานวิจัยนี้ คือ เทคนิคการค้นหากฎความสัมพันธ์ (Association Rule Discovery) ในงานวิจัยนี้ได้เก็บข้อมูลในการเข้าใช้อินเตอร์เน็ตจากเครื่องเม่ข่ายพร็อกซี ปริมาณการจราจรจะถูกเก็บรวบรวมไว้ทุกๆ 1 ชั่วโมง และมีระดับ (Rate) แสดงไว้ว่าการจราจรดังกล่าวอยู่ในระดับใด ระดับปริมาณการจราจรข้างต้นสามารถแสดงความสัมพันธ์ระหว่างช่วงเวลากับระดับปริมาณการจราจรได้ ในงานวิจัยนี้ได้แบ่งข้อมูลออกเป็น 2 ส่วนคือ ข้อมูลการเรียนรู้และข้อมูลตรวจสอบ โดยที่การแบ่งข้อมูลนั้นใช้วิธีการแลกตัวอย่างแบบสุ่ม และนำข้อมูลที่ได้มาหาความสัมพันธ์ประสิทธิภาพและความเร็วของการใช้งานในระบบเครือข่ายขึ้นอยู่กับปัจจัยต่างๆ มากมาย ปริมาณการจราจรที่ใช้เส้นทางต่างๆ ก็เป็นปัจจัยหนึ่งที่จะส่งผลกระทบต่อการใช้งาน ผลสรุปจากการศึกษาหาความสัมพันธ์ของปริมาณการจราจรบนเส้นทางของระบบเครือข่ายพบว่าการใช้เทคนิคการค้นหากฎความสัมพันธ์สามารถทำนายแนวโน้มปริมาณการจราจรที่จะใช้ในเส้นทางต่างๆ ได้ การใช้เทคนิคการค้นหากฎความสัมพันธ์เพื่อทำนายปริมาณการจราจรนี้สามารถนำไปเป็นแนวทางพิจารณาจัดการบริหารการเลือกใช้เส้นทางให้เกิดประสิทธิภาพสูงสุดและสามารถนำไปเป็นปัจจัยหนึ่งในการพัฒนาประสิทธิภาพการรับส่งข้อมูลให้มีประสิทธิภาพสูงขึ้นต่อไปได้

8.2 งานวิจัยต่างประเทศ

วิเวอรอส, นีรอส และรอทแมน (Marisa S. Viveros; John P. Nearhos; & Michael J. Rothman. 1996: Online) ได้ทำการประยุกต์เทคนิคเหมืองข้อมูลไปใช้ในระบบสารสนเทศประกันสุขภาพ โดยใช้กฎความสัมพันธ์ (Association Rule) ในการประยุกต์ใช้ในการหาความสัมพันธ์ของฐานข้อมูลจำแนกกลุ่มของผู้ฝึกหัดเพื่อจะทำให้การตรวจสอบรูปแบบการฝึกงานของบริษัทและการบริการของผู้ฝึกงานให้มีประสิทธิภาพยิ่งขึ้น งานวิจัยนี้ได้นำเทคนิคเหมืองข้อมูลสองวิธีมาใช้โดยรวมให้อยู่ในรูปของฐานข้อมูลขนาดใหญ่ ซึ่งอัลกอริทึมนี้สามารถจัดการกับข้อมูลขนาดใหญ่ ข้อมูลดิบที่มากกว่านั้นสามารถแสดงปริมาณของผลกำไร และสิ่งที่น่าสนใจภายในองค์กรได้

ยี่ฮัง วู และปีเตอร์ เฉิน (Yi-Hung Wu; & P. Chen. 2002: Online) ได้ทำการศึกษาพฤติกรรมการใช้เว็บโดยการเรียงลำดับหน้าเว็บที่มีการเรียกใช้งาน จากข้อมูลจริงใน access log ที่พร็อกซี เซิร์ฟเวอร์โดยมี วัตถุประสงค์เพื่อทำนายความต้องการของผู้ใช้ และ โครงสร้างของพร็อกซีสำหรับทำการดึงหน้าเว็บ ซึ่งงานวิจัยนี้ใช้ข้อมูลจริงในการทดลอง นี้แบ่งขั้นตอนในการทดลองเป็น 4 ขั้นตอน คือ ขั้นตอนการเลือกข้อมูล, ขั้นตอนการจัดการกับข้อมูลเป็นจัดทำข้อมูลให้เหมาะสมโดยใช้รูปแบบของต้นไม้ (Tree) ก่อนการทำเหมืองข้อมูล, ขั้นตอนการทำเหมืองข้อมูลซึ่งใช้อัลกอริทึมต้นไม้

ตัดสินใจ และขั้นตอนสุดท้ายคือ ขั้นตอนของการทำนาย ซึ่งผลของการทำนายที่ได้จะเก็บรวบรวมไว้เพื่อใช้ในการ Prefetching การทดลองนี้มีข้อจำกัดคือพิจารณาเฉพาะกรณีที่มีผู้ใช้เพียงคนเดียว ผลการทดลองพบว่าการทำนายมีความถูกต้องสูง โดย hit ratio เหมาะสมที่สมที่สุดในการทำนายคือ 75.69% ซึ่งใช้เวลาในการทำนายเพียง 1.9 ms ซึ่งถือว่าใช้เวลาในการบริการโดยเฉลี่ยน้อยมาก

ชางเทียน ลู่, อาร์โนล พี โบดี้ฮาร์ดโจ้ และปรัชวาล มานาลวอร์ (Chang-Tien Lu; Arnold P. Boedihardjo; & Prajwal Manalwar. 2005: Online) ทำการศึกษาวิจัยเรื่อง ""Exploiting Efficient Data Mining Techniques to Enhance Intrusion Detection Systems" งานวิจัยนี้อธิบายเทคนิค Data mining ซึ่งแตกต่างกันและความมีประโยชน์ของพวกมันในสถานการณ์ต่างๆ ของระบบและอธิบายเกี่ยวกับระบบตรวจจับการบุกรุกในปัจจุบันที่ทำการใช้ Data mining สำหรับการตรวจจับการบุกรุก ในงานวิจัยนี้กล่าวถึงเทคนิค Data mining ต่างๆ เช่น การจัดกลุ่ม (Clustering), การจำแนกประเภท (Classification) และกฎความสัมพันธ์ (Association rule) เพื่อวิเคราะห์ข้อมูลเครือข่ายให้ได้รับข้อมูลที่เกี่ยวข้องกับการบุกรุก

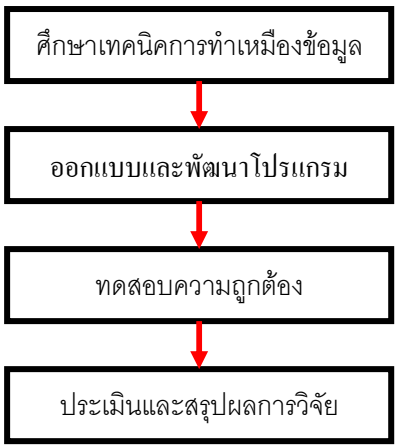
Viral Marketing Tool เป็นเครื่องมือที่บริษัทเอกชนรายหนึ่งผลิตขึ้นเพื่อขายให้กับผู้ให้บริการรับส่งข้อความมัลติมีเดียเจ้าหนึ่งที่ต้องการจะนำมาใช้ในการวิเคราะห์ MMS traffic โดยวัตถุประสงค์ของการใช้เครื่องมือนี้คือ เพื่อหา Content ที่มีการส่งกันมากๆ และผู้ใช้งานที่มีการส่งข้อความมัลติมีเดียมากๆ โดยหลักการทำงานของเครื่องมือนี้คือ จะเก็บข้อมูลของ Traffic โดยการ Mirror port จากอุปกรณ์ Switch ที่ต่ออยู่กับระบบรับส่งข้อความมัลติมีเดีย ข้อมูล Traffic ที่เก็บมาได้นั้นจะเป็นข้อมูลระดับ Packet หลังจากนั้นทางบริษัทที่ผลิตเครื่องมือจะนำข้อมูลที่เก็บได้ไปเข้ากระบวนการวิเคราะห์ ซึ่งส่วนหนึ่งของกระบวนการจะมีการนำเทคนิคเหมืองข้อมูลมาใช้ ผลลัพธ์ของการใช้ Viral Marketing Tool โดยการหา Content ที่มีการส่งกันมากที่สุดพบว่า อันดับที่1 เป็นไวรัส CommWarrior

จากการศึกษางานวิจัยที่เกี่ยวข้องพบว่า ส่วนของการวิจัยภายในประเทศนั้น มีการนำเทคนิคการค้นหากฎความสัมพันธ์ มาประยุกต์ใช้ในการจัดสรรกฎหมายที่เหมาะสมกับคดีความ, ทำนายข้อมูลการเรียกใช้เว็บในอนาคต และทำนายปริมาณการจราจรในเครือข่าย ในส่วนของงานวิจัยต่างประเทศนั้น ได้มีการนำไปใช้ในระบบสารสนเทศประกันสุขภาพ, การศึกษาพฤติกรรมการใช้เว็บโดยการเรียงลำดับหน้าเว็บที่มีการเรียกใช้งาน, พัฒนาระบบตรวจจับการบุกรุก ซึ่งงานวิจัยทั้งหมดที่ผู้วิจัยได้ศึกษามานั้นไม่ได้ประยุกต์ใช้บนระบบระบบบริการรับส่งข้อความมัลติมีเดีย แต่ยังคงสามารถนำมาใช้เป็นแนวทางในการออกแบบและพัฒนาโปรแกรมเพื่อใช้ในการตรวจจับข้อความมัลติมีเดียที่มีไวรัส ส่วน Viral Marketing Tool นั้น มีความสามารถในการตรวจจับข้อความมัลติมีเดียที่มีไวรัสเหมือนโปรแกรมที่ผู้วิจัยต้องการจะออกแบบและพัฒนาขึ้น โดย Viral Marketing Tool สามารถที่จะ

ตรวจสอบได้ลึกถึงระดับ Packet ตามที่ได้กล่าวไปข้างต้น จึงทำให้สามารถที่จะวิเคราะห์ข้อมูลได้ถึง Content ภายในของข้อความมัลติมีเดีย ซึ่งส่งผลทำให้อัตราการตรวจจับไวรัสมีความถูกต้องสูงมาก แต่ในทางกลับกันด้วยข้อมูลที่ Viral Marketing Tool ต้องเก็บเพื่อนำไปเข้ากระบวนการวิเคราะห์นั้น ทำให้อุปกรณ์ที่ไปเชื่อมต่อเพื่อทำการ Mirror traffic ต้องมีขนาดความจุที่ใหญ่เพียงพอกับการเก็บรวบรวม Traffic ในแต่ละวัน อีกทั้งในกระบวนการยังต้องเสียเวลาในการดึงข้อมูลจาก Packet เพื่อนำมาใช้ในการวิเคราะห์อีกด้วย ทางผู้วิจัยจึงได้ออกแบบและพัฒนาโปรแกรมในการตรวจจับไวรัส ที่ไม่ต้องเสียค่าใช้จ่ายในการเก็บรวบรวม Traffic เพราะข้อมูลที่นำมาใช้ในการวิเคราะห์เป็นข้อมูลจากระบบบริการรับส่งข้อความมัลติมีเดียที่มีการเก็บรวบรวมไว้อยู่แล้ว ซึ่งผู้วิจัยมีความคิดเห็นว่าเป็นข้อมูลที่มีความละเอียดเพียงพอแล้ว สำหรับการนำไปใช้วิเคราะห์ตรวจจับไวรัส ไม่จำเป็นต้องมีความละเอียดถึงระดับ Packet เนื่องจากผู้ให้บริการโครงข่ายโทรศัพท์มือถือต้องการผลลัพธ์ของการวิเคราะห์ เพียงเพื่อนำไปหาแนวทางในการปรับปรุงบริการเท่านั้น ซึ่งโปรแกรมที่ผู้วิจัยพัฒนาขึ้นสามารถนำไปประยุกต์ใช้กับข้อมูลในระบบบริการรับส่งข้อความมัลติมีเดีย เพื่อช่วยให้ผู้ให้บริการโครงข่ายโทรศัพท์มือถือสามารถทราบจำนวน และพฤติกรรมของผู้ใช้งานข้อความมัลติมีเดียที่ติดไวรัส เพื่อนำข้อมูลไปวิเคราะห์ ปรับปรุง และดูแลระบบให้มีประสิทธิภาพมากยิ่งขึ้น

บทที่ 3
วิธีการดำเนินการวิจัย

งานวิจัยนี้ได้นำเทคนิคการทำเหมืองข้อมูล (Data mining) มาออกแบบและพัฒนาโปรแกรมเพื่อใช้ในการตรวจจับข้อความมัลติมีเดียที่มีไวรัส โดยข้อมูลที่ใช้ในงานวิจัยนี้คือข้อมูลจากระบบบริการรับส่งข้อความมัลติมีเดีย (MMSC system) โดยได้กำหนดขั้นตอนการดำเนินการวิจัยไว้ดังนี้



ภาพประกอบ 5 แผนผังขั้นตอนการดำเนินการวิจัย

1. การศึกษาการทำงานของเหมืองข้อมูล

ศึกษาเทคนิคการทำเหมืองข้อมูล (Data mining) จากหลักทฤษฎีทั่วไปที่มีการตีพิมพ์เผยแพร่ทางหนังสือและทางอินเตอร์เน็ต รวมทั้งจากงานวิจัยที่เคยมีผู้นำเสนอมาแล้ว เพื่อทำความเข้าใจกระบวนการทำเหมืองข้อมูล รวมทั้งหาวิธีการทำเหมืองข้อมูล (Data mining) ที่เหมาะสมกับข้อมูลในระบบบริการรับส่งข้อความมัลติมีเดีย ซึ่งพบว่าการค้นหากฎความสัมพันธ์ (Association Rule) แบบวิธีค้นหารูปแบบ (Frequent Pattern Mining) ระหว่างข้อมูล ทำให้สามารถแบ่งกลุ่ม หรือคัดแยกข้อมูลเป็นรูปแบบเพื่อใช้เป็นข้อมูลความรู้ต่อไป โดยใช้ค่าสนับสนุน (Support) และค่าความเชื่อมั่น (Confidence) เป็นเกณฑ์ในการตัดสินว่ารูปแบบใดสมควรกำหนดเป็นกฎขึ้น

2. การออกแบบและพัฒนาโปรแกรมสำหรับใช้ในการตรวจจับข้อความมัลติมีเดียที่มีไวรัส

ในงานวิจัยบทที่ 2 ได้กล่าวถึงวิธีการทำเหมืองข้อมูลเบื้องต้น (Basic of Data Mining) ผู้วิจัยได้ศึกษากระบวนการในการขุดหาความรู้จากข้อมูล (Knowledge Discovery from data) และได้ทำตามกระบวนการดังนี้

2.1 ทำความสะอาดข้อมูล (Data Cleaning)

ข้อมูลการส่งข้อความมัลติมีเดีย ของผู้ใช้งานภายในระบบบริการรับส่งข้อความมัลติมีเดีย (MMSC System) มีความหลากหลายและละเอียดมาก เนื่องจาก Log แต่ละรายการ เป็น log ที่เกิดขึ้นระดับแอพพลิเคชั่น คือ ทุกๆ การทำงานของแอพพลิเคชั่นบนระบบบริการรับส่งข้อความมัลติมีเดีย (MMSC System) จะมี Log เกิดขึ้น ดังนั้นผู้วิจัยจึงเลือกเก็บเฉพาะ Log ส่วนที่จำเป็นในการวิจัยเท่านั้น และกำจัดข้อมูลส่วนที่ไม่ต้องการออกไป ซึ่งจะเหลือ Log ส่วนที่ต้องรวบรวมและจัดเก็บ ดังภาพประกอบ 6

2009.07.01 00:00:18.879 vendor.services.mms.relay.waphdlr.WapHandlerManager TrafficLog: [info] ip-address-vendor.services.mms
.relay.util.MsendReq,info,"M-Send.req received. Status: Ok","year='2009' month='07' day='01' hour='00' min='00' sec='18' mil='879'",
"1214845218879:64925760","+668996xxxxx/TYPE=PLMN","08925xxxxx/TYPE=PLMN",m-send-req,80,-127,"NokiaN72/2.0617.1.0.3 Series60/2.8 Prof
ile/MIDP-2.0 Configuration/CLDC-1.1",,"Relay","E12E98400DD705","text/plain;application/vnd.symbian.install","NO_MEDIA_TYPE_INFO",Personal,30712,1,NO_PARTY_CHARGE_INFO,"m-send-req","MM1","NO_INTERFACE_ID_INFO","0.0",-1,"Success",false,"muqeckgz1enzl7fp00"

ภาพประกอบ 6 ตัวอย่างข้อมูลของระบบบริการรับส่งข้อความมัลติมีเดีย

2.2 การรวบรวมข้อมูล (Data Integration)

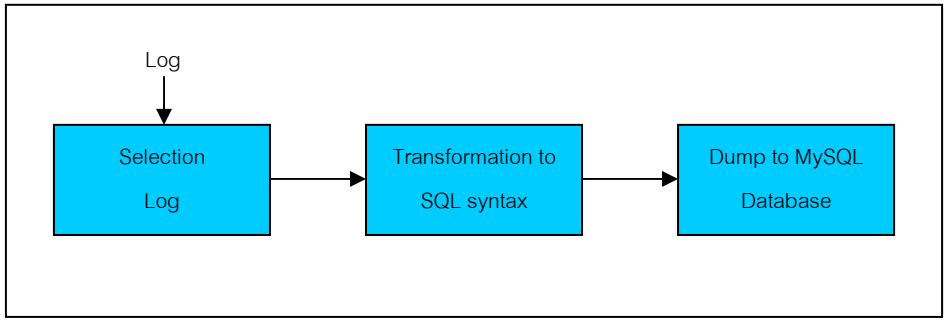
ผู้วิจัยได้ดำเนินการรวมรวมข้อมูล Log การส่งข้อความมัลติมีเดีย ของผู้ใช้งานภายในระบบบริการรับส่งข้อความมัลติมีเดีย (MMSC System) จำนวน 4 Server เป็นเวลา 2 เดือนตั้งแต่วันที่ 1 มิถุนายน 2552 ถึงวันที่ 31 กรกฎาคม 2552 และแปลงข้อมูลจาก log ซึ่งเป็น Text ไฟล์ ให้เป็น

ข้อมูลที่สามารถประมวลผลได้ โดยแปลงลงคลังข้อมูล (Data Warehouse) ซึ่งใช้ฐานข้อมูลมายเอสคิวแอล (MySQL) และแบ่งข้อมูลออกเป็น 2 ชุด คือ ชุดฝึกสอน (Training set) และชุดทดสอบ (Test set) โดยชุดฝึกสอนจะใช้ในการสร้างโปรแกรม และชุดทดสอบใช้ในการตรวจสอบความถูกต้อง

ข้อมูลที่ทำการรวบรวมจากระบบบริการรับส่งข้อความมัลติมีเดีย (MMSC System) เป็นข้อมูลของ Log ซึ่งข้อมูลเป็นชนิด Text ไฟล์ ดังภาพประกอบ 6 โดยประกอบด้วยข้อมูล 9 ข้อมูล ดังนี้

Date	วันที่ผู้ใช้งานส่งข้อความมัลติมีเดีย
Time	เวลาที่ผู้ใช้งานส่งข้อความมัลติมีเดีย
Sender msisdn	เบอร์โทรศัพท์ของผู้ส่งข้อความมัลติมีเดีย
Recipient msisdn	เบอร์โทรศัพท์ของผู้รับข้อความมัลติมีเดีย
Phone model	รุ่นโทรศัพท์ของผู้ส่งข้อความมัลติมีเดีย
Content type	ประเภทของเนื้อหาที่ส่งมากับข้อความมัลติมีเดีย
Content size	ขนาดของเนื้อหาที่ส่งมากับข้อความมัลติมีเดีย
Request status	ระบุสถานะของการร้องขอใช้งานบริการรับ-ส่งข้อความมัลติมีเดีย
Message ID	หมายเลขของข้อความมัลติมีเดียที่ผู้ใช้งานส่ง

ในขั้นตอนการแปลงข้อมูล Log ผู้วิจัยได้สร้างโปรแกรมสำหรับใช้โหลดข้อมูล Log เข้าสู่ฐานข้อมูล ซึ่งโปรแกรมนี้ต้องทำการดึงข้อมูล 9 ข้อมูลข้างต้น ออกมาจากแต่ละรายการ และแปลงข้อมูลทั้ง 9 ข้อมูลให้อยู่ในรูปแบบของภาษาเอสคิวแอล (SQL) เพื่อที่จะสามารถโหลดข้อมูลเข้าสู่ฐานข้อมูลได้ กระบวนการทำงานของโปรแกรมดังภาพประกอบ 7



ภาพประกอบ 7 กระบวนการทำงานของโปรแกรมโหลดข้อมูล Log สู่ฐานข้อมูล

2.3 การคัดเลือกข้อมูล (Data Selection)

งานวิจัยนี้นำข้อมูลการรับส่งข้อความมัลติมีเดีย ของผู้ใช้งานภายในระบบบริการรับส่งข้อความมัลติมีเดีย (MMSC System) มาใช้ การเตรียมข้อมูลเพื่อนำไปสร้างโมเดลจะคัดเลือกเฉพาะข้อมูลที่มีความสัมพันธ์กับคุณสมบัติของไวรัสชนิด CommWarrior คือ Phone model, Content type, Content size เนื่องมาจากไวรัสชนิดนี้จะทำงานบนโทรศัพท์ที่ใช้ระบบปฏิบัติการ Symbian series 60 และมีค่าเฉลี่ยของ Content size อยู่ที่ 27 – 30 KB รวมทั้งเนื้อหาที่ไวรัสส่งมาพร้อมกับข้อความมัลติมีเดียจะเป็น Symbian archive (*.SIS)

Phone model	รุ่นโทรศัพท์ของผู้ส่งข้อความมัลติมีเดีย
Content size	ขนาดของเนื้อหาที่ส่งมากับข้อความมัลติมีเดีย
Content type	ประเภทของเนื้อหาที่ส่งมากับข้อความมัลติมีเดีย

2.4 การแปลงข้อมูล (Data Transformation)

เป็นการแปลงข้อมูลที่ต้องนำมาใช้ เพื่อให้เหมาะสมกับอัลกอริธึมที่เลือกใช้ในการทำเหมืองข้อมูล (Data mining) โดยผู้วิจัยได้ทำการแปลงข้อมูล Phone model, Content type, Content size ที่อยู่ในฐานข้อมูลให้เป็นหมวดหมู่ เพื่อเอื้อต่อขั้นตอนในการประเมินรูปแบบของการทำเหมืองข้อมูล (Data mining) ที่จะไม่เกิดการกระจายรูปแบบออกเป็นพันๆ รูปแบบขึ้น

2.5 การทำเหมืองข้อมูล (Data mining)

งานวิจัยนี้เลือกใช้เทคนิคการค้นหากฎความสัมพันธ์ (Association Rule) แบบวิธีค้นหารูปแบบ (Frequent Pattern Mining) เพื่อใช้ในการตรวจจับข้อความมัลติมีเดียที่มีไวรัส โดยประเมินรูปแบบ (Pattern Evaluation) คือระบุความรู้ที่สกัดได้ มีความน่าเชื่อถือเพียงใด โดยใช้วิธีหาค่าความเชื่อมั่น (Confidence) และค่าสนับสนุน (Support) ที่เกิดขึ้น สามารถหาค่าความเชื่อมั่น และค่าสนับสนุนจากสมการดังนี้

$$\text{ค่าความเชื่อมั่น (A,B)} = \frac{\text{จำนวนของ Transaction (A,B)}}{\text{จำนวน Transaction (A)}}$$

$$\text{ค่าสนับสนุน (A,B)} = \frac{\text{จำนวนของ Transaction (A,B)}}{\text{จำนวน Transaction ทั้งหมด}}$$

โดยที่ A หมายถึง เหตุการณ์ที่ใช้เป็นเงื่อนไขในการตรวจจับข้อความมัลติมีเดียที่มีไวรัส

B หมายถึง เหตุการณ์ที่พบข้อความมัลติมีเดียที่มีไวรัส

Transaction (A, B) หมายถึง เหตุการณ์ที่ประกอบด้วยเหตุการณ์ A และ B

Transaction (A) หมายถึง เหตุการณ์ที่ประกอบด้วยเหตุการณ์ A อย่างเดียว

โดยในงานวิจัยนี้ผู้วิจัยต้องการค้นหาความสัมพันธ์ของ Phone model, Content type, Content size ซึ่งกำหนดให้เป็นเงื่อนไข และผลลัพธ์ที่เกิดขึ้น คือพบรายการข้อความมัลติมีเดียที่มีไวรัส

ขั้นตอนของการหากฎความสัมพันธ์ทั้งหมด สามารถสรุปได้เป็น 2 ขั้นตอน คือ การหารายการข้อมูลทั้งหมดที่มีค่าสนับสนุนมากกว่าค่าที่น้อยที่สุดของค่าสนับสนุน (Minimum Support) ขั้นตอนต่อจากนั้นหากฎทั้งหมดที่มีค่าความเชื่อมั่นมากกว่าค่าที่น้อยที่สุดของค่าความเชื่อมั่น (Minimum Confidence) จากรายการข้อมูลในขั้นตอนแรก

หลักเกณฑ์การจำแนกกฎความสัมพันธ์

1. ผลลัพธ์ของเงื่อนไขต้องมีค่าสนับสนุน (Support) มากกว่าหรือเท่ากับค่าน้อยที่สุดของค่าสนับสนุน (Minimum Support) ที่กำหนดไว้ด้วย โดยในงานวิจัยนี้กำหนดไว้ที่ 1% เนื่องจากจำนวนของไวรัสมีอยู่ประมาณ 10% ของข้อมูลทั้งหมดเท่านั้น
2. ถ้าเงื่อนไขใดมีค่าความเชื่อมั่น (Confidence) มากกว่า 80% ขึ้นไป จะถือว่าเงื่อนไขนั้นเป็นกฎความสัมพันธ์

3. การทดสอบโปรแกรมในการวิจัย

ทดสอบความถูกต้องของโปรแกรม โดยการนำไปใช้วิเคราะห์ข้อมูลกลุ่มตัวอย่างจากระบบบริการรับส่งข้อความมัลติมีเดีย (MMSC System) บนคอมพิวเตอร์ที่มีระบบฐานข้อมูล (Database System) รองรับการทำงานของโปรแกรม ดังภาพประกอบ 8

ภาพประกอบ 8 แผนผังขั้นตอนการทดสอบโมเดล

3.1 ตรวจจับข้อความมัลติมีเดียที่มีไวรัส

เป็นการนำผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยเทคนิคการทำเหมืองข้อมูล กฎการวิเคราะห์ความสัมพันธ์ (Association Rule) แบบวิธีค้นหารูปแบบ (Frequent Pattern Mining) ไปตรวจจับไวรัส โดยใช้ข้อมูลกลุ่มตัวอย่างในการตรวจจับ

3.2 ตรวจสอบผลการตรวจจับ

นำผลที่ได้จากการตรวจจับข้อความมัลติมีเดียที่มีไวรัสไปเปรียบเทียบกับผลที่ได้จากการตรวจจับโดยใช้ Viral Marketing Tool หากข้อความมัลติมีเดียที่ตรวจจับได้ว่ามีไวรัสสอดคล้องกันในแต่ละวัน ถือว่าผลของการตรวจจับถูกต้อง

4. การประเมินผลจากการทดสอบโปรแกรมในการวิจัย

เป็นขั้นตอนของการประเมินผลจากการทดสอบโปรแกรมที่ออกแบบขึ้น รวมทั้งยังเป็นการทดสอบสมมติฐานการวิจัยด้วย โดยความสำเร็จของงานวิจัยนั้นวัดผลได้จาก ความถูกต้องของโปรแกรม คือ สามารถตรวจจับข้อความมัลติมีเดีย (MMS) ที่มีไวรัสได้ผลสอดคล้องกับการตรวจจับโดยใช้ Viral Marketing Tool

ตัวชี้วัดและสถิติที่ใช้ในการวิจัย

ตัวชี้วัดที่ใช้ในการวิจัย คือ อัตราการตรวจจับ (Detection Rate) ระหว่างโปรแกรมที่พัฒนาขึ้น กับ Viral Marketing Tool ภายใต้เงื่อนไขเดียวกัน ซึ่งอัตราการตรวจจับ คือ อัตราส่วนระหว่างจำนวนของการตรวจจับข้อความมัลติมีเดีย (MMS) ที่มีไวรัสได้ถูกต้อง กับจำนวนของข้อความมัลติมีเดีย (MMS) ที่มีไวรัสทั้งหมด

สถิติที่ใช้ในการวิจัย คือ จำนวนของข้อความมัลติมีเดีย (MMS) ที่มีไวรัส เปรียบเทียบกันระหว่างการใช้โปรแกรมที่ออกแบบขึ้นตรวจจับ กับการตรวจจับโดยใช้ Viral Marketing Tool โดยให้การตรวจจับนั้นอยู่ภายใต้เงื่อนไขเดียวกัน

บทที่ 4
ผลการดำเนินงานวิจัย

ในการดำเนินงานวิจัย ผู้วิจัยได้พัฒนาโปรแกรมขึ้นด้วย ภาษา PHP, ฐานข้อมูลมายเอสคิวแอล (MySQL), อะแพชี เว็บเซิร์ฟเวอร์ (Apache Web Server) และ จาวาสคริปต์ (JavaScript) เพื่อให้งานวิจัยบรรลุตามวัตถุประสงค์ คือออกแบบโปรแกรมสำหรับใช้ในการตรวจจับข้อความมัลติมีเดีย (MMS) ที่มีไวรัส โดยการประยุกต์ใช้เทคนิคการทำเหมืองข้อมูล แบบกฎการวิเคราะห์ความสัมพันธ์ (Association Rule) และนำโปรแกรมที่ได้ไปใช้ทดแทน Viral Marketing Tool
ขั้นตอนการดำเนินวิจัยแบ่งขั้นตอนออกเป็นดังนี้

1. การเตรียมข้อมูล เพื่อใช้สำหรับออกแบบและพัฒนาโปรแกรม
2. ออกแบบโปรแกรมเพื่อใช้ตรวจจับข้อความมัลติมีเดียที่มีไวรัสโดยใช้เทคนิคเหมืองข้อมูล
3. พัฒนาโปรแกรมเพื่อใช้ตรวจจับข้อความมัลติมีเดียที่มีไวรัสโดยใช้เทคนิคเหมืองข้อมูล
4. ทดสอบความถูกต้องของโปรแกรมที่พัฒนาขึ้น โดยใช้ข้อมูลชุดทดสอบ
5. ประเมินผลการตรวจจับข้อความมัลติมีเดีย (MMS) ที่มีไวรัสจากโปรแกรมที่ออกแบบขึ้นและจากการใช้ Viral Marketing Tool

ผลการดำเนินงานวิจัยตามขั้นตอนต่างๆ สามารถอธิบายได้ดังนี้

1. การเตรียมข้อมูล

1.1 การรวบรวมข้อมูล (Data Integration)

หลังจากขั้นตอนการทำความสะอาดข้อมูล (Data Cleaning) ที่เลือกเฉพาะ Log ส่วนที่จำเป็นในการวิจัยเท่านั้น และกำจัดข้อมูลส่วนที่ไม่ต้องการออกไป จะเหลือ Log ส่วนที่ต้องรวบรวมและจัดเก็บ ดังภาพประกอบ 6 ในงานวิจัยบทที่ 3 ซึ่งผู้วิจัยได้แปลงข้อมูลจาก Text ไฟล์ ลงคลังข้อมูล (Data Warehouse) ที่ใช้ฐานข้อมูลมายเอสคิวแอล (MySQL) และแบ่งการจัดเก็บข้อมูลเป็น 2 ชุด คือ

1.1.1 ชุดฝึกสอน (Training set) คือข้อมูลการส่งข้อความมัลติมีเดียของผู้ใช้งานภายในระบบบริการรับส่งข้อความมัลติมีเดีย (MMSC System) จำนวน 4 Server ระหว่างวันที่ 1-30 มิถุนายน 2552 ที่ใช้ในการสร้างโปรแกรม

1.1.2 ชุดทดสอบ (Test set) คือข้อมูลการส่งข้อความมัลติมีเดียของผู้ใช้งานภายในระบบบริการรับส่งข้อความมัลติมีเดีย (MMSC System) จำนวน 4 Server ระหว่างวันที่ 1-31 กรกฎาคม 2552 ที่ใช้ในการตรวจสอบความถูกต้องของโปรแกรมที่สร้างขึ้น

ข้อมูลที่จัดเก็บลงคลังข้อมูล (Data Warehouse) จากระบบบริการรับส่งข้อความมัลติมีเดีย (MMSC System) ประกอบด้วยข้อมูล 9 ข้อมูล ดังภาพประกอบ 9

Field	Type	Null	Key	Default	Extra
year	int(11)	YES		NULL	
month	int(11)	YES		NULL	
day	int(11)	YES		NULL	
hour	int(11)	YES		NULL	
min	int(11)	YES		NULL	
sec	int(11)	YES		NULL	
mil	int(11)	YES		NULL	
sender	text	YES		NULL	
recipient	text	YES		NULL	
phoneModel	text	YES		NULL	
contentType	text	YES		NULL	
contentSize	int(11)	YES		NULL	
triggerPoint	varchar(20)	YES		NULL	
status	text	YES		NULL	
messageID	varchar(25)	YES		NULL	

ภาพประกอบ 9 ข้อมูลที่จัดเก็บลงคลังข้อมูล (Data Warehouse)

1.2 การคัดเลือกข้อมูล (Data Selection)

ในขั้นตอนการคัดเลือกข้อมูลเพื่อนำไปสร้างโมเดลของเทคนิคเหมืองข้อมูล ผู้วิจัยเลือกเฉพาะข้อมูลที่มีความสัมพันธ์กับคุณสมบัติของไวรัสชนิด CommWarrior คือ

1.2.1 รุ่นโทรศัพท์ของผู้ส่งข้อความมัลติมีเดีย (Phone Model) ไวรัสชนิด CommWarrior จะทำงานบนโทรศัพท์เคลื่อนที่ที่ใช้ระบบปฏิบัติการ Symbian series 60 เท่านั้น ด้วยคุณสมบัตินี้ ช่วยให้ผู้วิจัยสามารถนำ Phone Model มาค้นหาความสัมพันธ์ได้ เนื่องจากข้อมูลการส่งข้อความมัลติมีเดียของผู้ใช้งานภายในระบบบริการรับส่งข้อความมัลติมีเดีย (MMSC System) มีข้อมูลทั้ง Phone Model และระบบปฏิบัติการรวมอยู่ด้วย ดังตัวอย่างข้อมูลตามภาพประกอบ 10

phoneModel
Nokia6681/2.0 (3.10.6) SymbianOS/8.0 Series60/2.6 Profile/MIDP-2.0 Configuration/CLDC-1.1
Nokia7610/2.0 (6.0522.0) SymbianOS/7.0s Series60/2.1 Profile/MIDP-2.0 Configuration/CLDC-1.0
Nokia2630/2.0 (04.21) Profile/MIDP-2.1 Configuration/CLDC-1.1
NokiaN72/5.0706.4.0.1 Series60/2.8 Profile/MIDP-2.0 Configuration/CLDC-1.1
Nokia2630/2.0 (04.21) Profile/MIDP-2.1 Configuration/CLDC-1.1
NokiaN70/ 5.0741.4.0.1 Series60/2.8 Profile/MIDP-2.0 Configuration/CLDC-1.1
NokiaN80-1/3.0 (3.0617.0.5) Series60/3.0 Profile/MIDP-2.0 Configuration/CLDC-1.1

ภาพประกอบ 10 ตัวอย่างข้อมูล Phone Model และระบบปฏิบัติการที่จัดเก็บลงคลังข้อมูล (Data Warehouse)

1.2.2 ขนาดของเนื้อหาที่ส่งมากับข้อความมัลติมีเดีย (Content size) ข้อมูลของไวรัสชนิดนี้ระบุไว้ว่า ค่าเฉลี่ยของขนาดข้อความมัลติมีเดียที่มีไวรัสจะอยู่ที่ประมาณ 27 – 30 KB หรือ 27729 – 30720 Byte ซึ่งข้อมูลการส่งข้อความมัลติมีเดียของผู้ใช้งานภายในระบบบริการรับส่งข้อความมัลติมีเดีย (MMSC System) มีการเก็บรวบรวมอยู่แล้ว ผู้วิจัยจึงเลือกนำ Content size มาหาความสัมพันธ์เพื่อนำไปสู่ข้อความมัลติมีเดียที่มีไวรัส โดยข้อความมัลติมีเดียของผู้ใช้งานภายในระบบบริการรับส่งข้อความมัลติมีเดีย (MMSC System) จะจัดเก็บ Content size หน่วยเป็น Byte ตามตัวอย่าง Content size ภาพประกอบ 11

contentSize
17085
9124
9859
27300
38
61767
5705
43811
27300
27301
17354
40898
38861

ภาพประกอบ 11 ตัวอย่างข้อมูล Content size หน่วยเป็น Byte

1.2.3 ประเภทของเนื้อหาที่ส่งมากับข้อความมัลติมีเดีย (Content Type) ข้อความมัลติมีเดียที่มีไวรัสประเภทนี้จะประกอบไปด้วย Symbian archive (*.SIS) และข้อความ Symbian archive จัดอยู่ในเนื้อหาประเภท application/vnd.symbian.install ซึ่งเป็นประเภทของเนื้อหามาตรฐานที่ระบบบริการรับส่งข้อความมัลติมีเดีย (MMSC System) จะรู้จัก และผู้ใช้บริการไม่สามารถส่งเนื้อหาประเภท Symbian archive ได้ นอกจากมีโปรแกรมที่สร้างขึ้นมาโดยเฉพาะ

contentType
text/plain;application/vnd.symbian.install
application/vnd.wap.multipart.related ; Start= <220089036> ;Type= application/vnd.symbian.install
application/vnd.symbian.install;text/plain
application/vnd.symbian.install;text/plain
application/vnd.symbian.install;text/plain
application/vnd.wap.multipart.related ; Start= <1428020583> ;Type= application/vnd.symbian.install
application/vnd.symbian.install;text/plain
application/vnd.symbian.install;text/plain
application/vnd.symbian.install;text/plain
application/vnd.symbian.install;text/plain
application/vnd.symbian.install;text/plain
application/vnd.symbian.install;text/plain

ภาพประกอบ 12 ตัวอย่างข้อมูลการส่งข้อความมัลติมีเดีย ที่ประกอบด้วยเนื้อหาประเภท application/vnd.symbian.install

1.3 การแปลงข้อมูล (Data Transformation)

ผู้วิจัยได้ทำการแปลงข้อมูล Phone model, Content size, Content type ที่อยู่ในฐานข้อมูลให้เป็นหมวดหมู่ เพื่อเอื้อต่อขั้นตอนในการประเมินรูปแบบของการทำเหมืองข้อมูล (Data mining) ที่จะไม่เกิดการกระจายรูปแบบออกเป็นพันๆ รูปแบบขึ้น อีกทั้งผู้วิจัยเลือกที่จะจัดหมวดหมู่ให้สอดคล้องตามคุณสมบัติของไวรัส เพื่อลดเวลาในการค้นหาความสัมพันธ์ของข้อมูลการส่งข้อความมัลติมีเดียทั้งหมดที่จัดเก็บ โดย

1.3.1 รุ่นโทรศัพท์ของผู้ส่งข้อความมัลติมีเดีย (Phone Model) จัดหมวดหมู่เป็น 2 ชุด โดยชุดที่ 1 จะใช้ค้นหาความสัมพันธ์ตามหลักเทคนิคการทำเหมืองข้อมูล เพื่อพิสูจน์คุณสมบัติของไวรัสชนิด CommWarrior ว่าทำงานบนโทรศัพท์เคลื่อนที่ที่ใช้ระบบปฏิบัติการ Symbian series 60 เท่านั้นจริง ส่วนชุดที่ 2 จะใช้สำหรับค้นหาความสัมพันธ์ตามหลักเทคนิคการทำเหมืองข้อมูลเพื่อให้ได้ข้อมูลการส่งข้อความมัลติมีเดียที่มีไวรัสของผู้ใช้งานภายในระบบบริการรับส่งข้อความมัลติมีเดีย (MMSC System) โดยแบ่งตามเวอร์ชั่นของระบบปฏิบัติการ Symbian series 60

ตาราง 2 รุ่นโทรศัพท์ของผู้ส่งข้อความมัลติมีเดีย (Phone Model) ชุดที่ 1

ลำดับ	รุ่นโทรศัพท์ของผู้ส่งข้อความมัลติมีเดีย (Phone Model)
1	โทรศัพท์ระบบปฏิบัติการ Symbian series 60
2	โทรศัพท์ที่ไม่ใช่ระบบปฏิบัติการ Symbian series 60

ตาราง 3 รุ่นโทรศัพท์ของผู้ส่งข้อความมัลติมีเดีย (Phone Model) ชุดที่ 2

ลำดับ	รุ่นโทรศัพท์ของผู้ส่งข้อความมัลติมีเดีย (Phone Model)
1	โทรศัพท์ระบบปฏิบัติการ Symbian series 60 เวอร์ชั่น 0.9
2	โทรศัพท์ระบบปฏิบัติการ Symbian series 60 เวอร์ชั่น 1.2
3	โทรศัพท์ระบบปฏิบัติการ Symbian series 60 เวอร์ชั่น 2.0
4	โทรศัพท์ระบบปฏิบัติการ Symbian series 60 เวอร์ชั่น 2.1
5	โทรศัพท์ระบบปฏิบัติการ Symbian series 60 เวอร์ชั่น 2.6
6	โทรศัพท์ระบบปฏิบัติการ Symbian series 60 เวอร์ชั่น 2.8
7	โทรศัพท์ระบบปฏิบัติการ Symbian series 60 เวอร์ชั่น 3.0
8	โทรศัพท์ระบบปฏิบัติการ Symbian series 60 เวอร์ชั่น 3.1
9	โทรศัพท์ระบบปฏิบัติการ Symbian series 60 เวอร์ชั่น 3.2

1.3.2 ขนาดของเนื้อหาที่ส่งมากับข้อความมัลติมีเดีย (Content size) จัดหมวดหมู่เป็น 2 ชุด โดยชุดที่ 1 แบ่งช่วงของขนาดเนื้อหาเป็น 5 KB หรือ 5120 Byte เนื่องจากผู้วิจัยต้องการให้ช่วงของขนาดที่แบ่งตกค่าเฉลี่ยของขนาดข้อความมัลติมีเดียที่มีไวรัสจะอยู่ที่ประมาณ 27 – 30 KB หรือ 27729 – 30720 Byte และชุดที่ 2 แบ่งช่วงของขนาดเนื้อหาเป็น 1 KB หรือ 1024 Byte โดยเลือกขนาดเนื้อหาช่วง 25601 – 40960 Byte จากช่วงของขนาดเนื้อหาชุดที่ 1 เนื่องจากผลการค้นหาความสัมพันธ์ตามหลักเทคนิคการทำเหมืองข้อมูลของข้อมูลชุดที่ 1 แสดงให้เห็นว่าข้อมูลในชุดที่ 1

แบ่งช่วงของขนาดเนื้อหากว้างเกินไป สามารถดูผลการค้นหาความสัมพันธ์โดยละเอียดจากขั้นตอนการออกแบบโปรแกรม

ตาราง 4 ขนาดของเนื้อหาที่ส่งมากับข้อความมัลติมีเดีย (Content size) ชุดที่ 1

ลำดับ	ขนาดของเนื้อหาที่ส่งมากับข้อความมัลติมีเดีย (Byte)
1	0 - 5120
2	5121 - 10240
3	10241 - 15360
4	15361 - 20480
5	20481 - 25600
6	25601 - 30720
7	30721 - 35840
8	35840 - 40960
9	40961 - 46080
10	46081 - 51200
11	51201 - 56320
12	56321 - 61440
13	61441 - 66560
14	66561 - 71680
15	71681 - 76800
16	76801 - 81920
17	81921 - 87040
18	87041 - 92160
19	92161 - 97280
20	> หรือเท่ากับ 97281

ตาราง 5 ขนาดของเนื้อหาที่ส่งมากับข้อความมัลติมีเดีย (Content size) ชุดที่ 2

ลำดับ	ขนาดของเนื้อหาที่ส่งมากับข้อความมัลติมีเดีย (Byte)
1	25600 - 26624
2	26625 - 27648
3	27649 - 28672
4	28673 - 29696
5	29697 - 30720
6	30721 - 31744
7	31745 - 32768
8	32769 - 33792
9	33793 - 34816
10	34817 - 35840
11	35841 - 36864
12	36865 - 37888
13	37889 - 38912
14	38913 - 39936
15	39937 - 40960

1.3.3 ประเภทของเนื้อหาที่ส่งมากับข้อความมัลติมีเดีย (Content Type) เนื่องด้วยเนื้อหาของข้อความมัลติมีเดียที่มีไวรัสประเภทนี้จะประกอบไปด้วย Symbian archive (*.SIS) และข้อความ ผู้วิจัยจึงได้จัดหมวดหมู่โดยศึกษาจากข้อมูลการส่งข้อความมัลติมีเดียของผู้ใช้งานภายในระบบบริการรับส่งข้อความมัลติมีเดีย (MMSC System) ที่มีเนื้อหาประกอบด้วย Symbian archive เป็นหลักส่วนนอกเหนือจากนี้ จะจัดรวมอยู่ในข้อความมัลติมีเดียที่ไม่ได้ประกอบด้วย Symbian archive

ตาราง 6 ประเภทของเนื้อหาที่ส่งมากับข้อความมัลติมีเดีย (Content Type)

ลำดับ	ประเภทของเนื้อหาที่ส่งมากับข้อความมัลติมีเดีย (Content Type)
1	application/vnd.symbian.install
2	application/vnd.symbian.install และ text/plain
3	application/vnd.symbian.install และ application/vnd.wap.multipart.related
4	Content type ประเภทอื่นๆ

2. ออกแบบโปรแกรมเพื่อใช้ตรวจจับข้อความมัลติมีเดียที่มีไวรัสโดยใช้เทคนิคเหมืองข้อมูล

ผู้วิจัยได้เลือกนำเทคนิคการทำเหมืองข้อมูล กฎการวิเคราะห์ความสัมพันธ์ (Association Rule) แบบวิธีค้นหารูปแบบ (Frequent Pattern Mining) มาประยุกต์ใช้ในการออกแบบโปรแกรม โดยใช้ข้อมูลชุดฝึกสอน (Training set) ตั้งแต่วันที่ 1 มิถุนายน 2552 ถึงวันที่ 30 มิถุนายน 2552 และได้ศึกษากฎการวิเคราะห์ความสัมพันธ์ ดังนี้

รูปแบบของการค้นหากฎความสัมพันธ์

$$A \rightarrow B$$

โดยที่ A เป็นเงื่อนไข และ B เป็นผลลัพธ์ที่เกิดขึ้น

ในงานวิจัยนี้ผู้วิจัยต้องการค้นหาความสัมพันธ์ระหว่าง Phone model, Content size, Content type และข้อความมัลติมีเดีย (MMS) ที่มีไวรัสชนิด CommWarrior ซึ่งกำหนดเงื่อนไขทั้งหมด 3 เงื่อนไข ดังนี้ Phone model เป็นเงื่อนไข ข้อความมัลติมีเดีย (MMS) ที่มีไวรัสเป็นผลลัพธ์ที่เกิดขึ้น, Content size เป็นเงื่อนไข ข้อความมัลติมีเดีย (MMS) ที่มีไวรัสเป็นผลลัพธ์ที่เกิดขึ้น, Content type เป็นเงื่อนไข ข้อความมัลติมีเดีย (MMS) ที่มีไวรัสเป็นผลลัพธ์ที่เกิดขึ้น สามารถเขียนกฎความสัมพันธ์ได้ดังนี้

$$\text{Phone model} \rightarrow \text{Virus MMS}$$

Content size → Virus MMS

Content type → Virus MMS

ตัวอย่างกฎความสัมพันธ์

Content size ช่วง 38913 – 39936 Byte → Virus MMS

จากกฎความสัมพันธ์นี้สามารถอธิบายได้ว่า ข้อความมัลติมีเดียที่มีขนาดของเนื้อหาช่วงระหว่าง 38913 – 39936 Byte มีแนวโน้มที่จะเป็นข้อความมัลติมีเดีย (MMS) ที่มีไวรัสชนิด CommWarrior

การหากฎความสัมพันธ์ของเงื่อนไข Phone model, Content type, Content size ที่ทำให้เกิดผลลัพธ์ คือพบรายการข้อความมัลติมีเดียที่มีไวรัส จะใช้ค่าสนับสนุน (Support) และค่าความเชื่อมั่น (Confidence) เป็นเกณฑ์ในการตัดสินใจว่ารูปแบบใดสมควรกำหนดเป็นกฎ โดยที่

ค่าสนับสนุน (Support) คือร้อยละของข้อมูลที่มีเงื่อนไขและผลลัพธ์สอดคล้องตามกฎต่อจำนวนข้อมูลทั้งหมด สามารถเขียนเป็นสมการได้ดังนี้

$$\frac{\text{จำนวนรายการข้อมูลที่มีเงื่อนไขและผลลัพธ์สอดคล้องกันตามกฎ}}{\text{จำนวนรายการข้อมูลทั้งหมด}} \quad \text{.....สมการที่ 1}$$

ค่าความเชื่อมั่น (Confidence) คือร้อยละของข้อมูลที่มีเงื่อนไขและผลลัพธ์สอดคล้องตามกฎต่อจำนวนรายการข้อมูลที่เป็นเงื่อนไข สามารถเขียนเป็นสมการได้ดังนี้

$$\frac{\text{จำนวนรายการข้อมูลที่มีเงื่อนไขและผลลัพธ์สอดคล้องกันตามกฎ}}{\text{จำนวนรายการข้อมูลที่เป็นเงื่อนไข}} \quad \text{.....สมการที่ 2}$$

ดังนั้นงานวิจัยนี้สามารถหาค่าสนับสนุนและค่าความเชื่อมั่นจากสมการ 1 และ 2 ได้ดังนี้

1. Phone model → Virus MMS

$$\text{ค่าสนับสนุน} = \frac{\text{จำนวนรายการข้อความมัลติมีเดียที่มี Phone model และไวรัสตามกฎ}}{\text{จำนวนรายการข้อความมัลติมีเดียทั้งหมด}}$$

$$\text{ค่าความเชื่อมั่น} = \frac{\text{จำนวนรายการข้อความมัลติมีเดียที่มี Phone model และไวรัสตามกฎ}}{\text{จำนวนรายการข้อความมัลติมีเดียที่มี Phone model เดียวกัน}}$$

2. Content size → Virus MMS

$$\text{ค่าสนับสนุน} = \frac{\text{จำนวนรายการข้อความมัลติมีเดียที่มี Content size และไวรัสตามกฎ}}{\text{จำนวนรายการข้อความมัลติมีเดียทั้งหมด}}$$

$$\text{ค่าความเชื่อมั่น} = \frac{\text{จำนวนรายการข้อความมัลติมีเดียที่มี Content size และไวรัสตามกฎ}}{\text{จำนวนรายการข้อความมัลติมีเดียที่มี Content size เดียวกัน}}$$

3. Content type → Virus MMS

$$\text{ค่าสนับสนุน} = \frac{\text{จำนวนรายการข้อความมัลติมีเดียที่มี Content type และไวรัสตามกฎ}}{\text{จำนวนรายการข้อความมัลติมีเดียทั้งหมด}}$$

$$\text{ค่าความเชื่อมั่น} = \frac{\text{จำนวนรายการข้อความมัลติมีเดียที่มี Content type และไวรัสตามกฎ}}{\text{จำนวนรายการข้อความมัลติมีเดียที่มี Content type เดียวกัน}}$$

จากสมการผู้วิจัยพบว่าในการหากฎความสัมพันธ์ของเงื่อนไข Phone model, Content type, Content size จะต้องคำนวณหาค่าพารามิเตอร์ เพื่อใช้ในการหากฎความสัมพันธ์ดังต่อไปนี้

1. จำนวนรายการข้อความมัลติมีเดียทั้งหมด
2. จำนวนรายการข้อความมัลติมีเดียที่มี Phone model เดียวกัน
3. จำนวนรายการข้อความมัลติมีเดียที่มี Content size เดียวกัน
4. จำนวนรายการข้อความมัลติมีเดียที่มี Content type เดียวกัน
5. จำนวนรายการข้อความมัลติมีเดียที่มี Phone model และไวรัสตามกฎ
6. จำนวนรายการข้อความมัลติมีเดียที่มี Content size และไวรัสตามกฎ
7. จำนวนรายการข้อความมัลติมีเดียที่มี Content type และไวรัสตามกฎ
8. คำนวณหาค่าสนับสนุน (Phone model → Virus MMS)
9. คำนวณหาค่าสนับสนุน (Content size → Virus MMS)
10. คำนวณหาค่าสนับสนุน (Content type → Virus MMS)
11. คำนวณหาค่าความเชื่อมั่น (Phone model → Virus MMS)
12. คำนวณหาค่าความเชื่อมั่น (Content size → Virus MMS)
13. คำนวณหาค่าความเชื่อมั่น (Content type → Virus MMS)

ตาราง 7 ตัวอย่างข้อมูลที่จัดเก็บลงคลังข้อมูล (Data Warehouse) จากระบบบริการรับส่งข้อความมัลติมีเดีย (MMSC System) วันที่ 1 มิถุนายน 2552 ระหว่างเวลา 00:00:10 – 00:00:19 น.

ลำดับ	Phone model	Content type	Content size
1	โทรศัพท์ระบบปฏิบัติการ Symbian series 60 เวอร์ชั่น 2.6	application/vnd.symbian.install และ text/plain	30720
2	โทรศัพท์ที่ไม่ใช่ระบบปฏิบัติการ Symbian series 60	Content type ประเภทอื่นๆ	5577
3	โทรศัพท์ระบบปฏิบัติการ Symbian series 60 เวอร์ชั่น 2.6	application/vnd.symbian.install และ text/plain	30720
4	โทรศัพท์ระบบปฏิบัติการ Symbian series 60 เวอร์ชั่น 1.2	application/vnd.symbian.install และ text/plain	30745
5	โทรศัพท์ระบบปฏิบัติการ Symbian series 60 เวอร์ชั่น 3.0	Content type ประเภทอื่นๆ	45296
6	โทรศัพท์ระบบปฏิบัติการ Symbian series 60 เวอร์ชั่น 2.0	application/vnd.symbian.install และ text/plain	27305
7	โทรศัพท์ที่ไม่ใช่ระบบปฏิบัติการ Symbian series 60	Content type ประเภทอื่นๆ	22990
8	โทรศัพท์ระบบปฏิบัติการ Symbian series 60 เวอร์ชั่น 2.1	application/vnd.symbian.install และ application/vnd.wap. multipart.related	27311
9	โทรศัพท์ระบบปฏิบัติการ Symbian series 60 เวอร์ชั่น 2.1	application/vnd.symbian.install และ text/plain	30720
10	โทรศัพท์ที่ไม่ใช่ระบบปฏิบัติการ Symbian series 60	Content type ประเภทอื่นๆ	8469

จากตาราง 7 สามารถคำนวณหาค่าพารามิเตอร์ ค่าสนับสนุน ค่าความเชื่อมั่น และกฎความสัมพันธ์ โดยใช้ตัวอย่างข้อมูลชุดฝึกสอน (Training set) วันที่ 1 มิถุนายน 2552 ได้ดังนี้ โดยกำหนดให้

จำนวนรายการข้อมูลที่เป็นเงื่อนไข = CNTCOND

จำนวนรายการข้อมูลที่มีเงื่อนไขและผลลัพธ์สอดคล้องกันตามกฎ = CNTASSO

และงานวิจัยนี้มีหลักเกณฑ์การจำแนกกฎความสัมพันธ์ดังนี้

1. ผลลัพธ์ของเงื่อนไขต้องมีค่าสนับสนุน (Support) มากกว่าหรือเท่ากับค่าน้อยที่สุดของค่าสนับสนุน (Minimum Support) ที่กำหนดไว้ด้วย โดยในงานวิจัยนี้กำหนดไว้ที่ 1% เนื่องจากจำนวนของไวรัสมีอยู่ประมาณ 10% ของข้อมูลทั้งหมดเท่านั้น
2. ถ้าเงื่อนไขใดมีค่าความเชื่อมั่น (Confidence) มากกว่า 80% ขึ้นไป จะถือว่าเงื่อนไขนั้นเป็นกฎความสัมพันธ์

2.1 Phone model → Virus MMS

ผลลัพธ์จากขั้นตอนการแปลงข้อมูล Phone model จะได้หมวดข้อมูล 2 ชุด ดังตาราง 2 และ 3 ผู้วิจัยต้องคำนวณค่าพารามิเตอร์ 2 ครั้งตามชุดของข้อมูล เนื่องจากข้อมูลที่ได้จัดหมวดหมู่ไว้มีความแตกต่างกัน ส่งผลให้การคำนวณจำนวนของแต่ละพารามิเตอร์แตกต่างกัน

2.1.1 คำนวณหาค่าพารามิเตอร์ ค่าสนับสนุน ค่าความเชื่อมั่น และกฎความสัมพันธ์ โดยใช้หมวด Phone model ชุดที่ 1 ตามตาราง 2 ซึ่งได้ผลการคำนวณดังตาราง 8

จำนวนรายการข้อความมัลติมีเดียทั้งหมด = 184,629

ตาราง 8 ผลการค้นหากฎความสัมพันธ์และคำนวณหาค่าพารามิเตอร์ ข้อมูลชุดฝึกสอน (Training set) วันที่ 1 มิถุนายน 2552 ของหมวด Phone model ชุดที่ 1

กฎ	CNTCOND	CNTASSO	ค่าความเชื่อมั่น	ค่าสนับสนุน
Symbian series 60 → Virus	55124	17802	32.29%	9.64%
ไม่ใช่ Symbian series 60 → Virus	129505	3	0%	0%

จากตาราง 8 เมื่อพิจารณาตามหลักเกณฑ์การจำแนกกฎความสัมพันธ์พบว่ากฎความสัมพันธ์ที่ค้นหาได้จากเงื่อนไขของหมวด Phone model ชุดที่ 1 ไม่ผ่านหลักเกณฑ์

2.1.2 คำนวณหาค่าพารามิเตอร์ ค่าสนับสนุน ค่าความเชื่อมั่น และกฎความสัมพันธ์ โดยใช้หมวด Phone model ชุดที่ 2 ตามตาราง 3 ซึ่งได้ผลการคำนวณดังตาราง 9

จำนวนรายการข้อความมัลติมีเดียทั้งหมด = 55,124

ตาราง 9 ผลการค้นหากฎความสัมพันธ์และคำนวณหาค่าพารามิเตอร์ ข้อมูลชุดฝึกสอน (Training set) วันที่ 1 มิถุนายน 2552 ของหมวด Phone model ชุดที่ 2

กฎ	CNTCOND	CNTASSO	ค่าความเชื่อมั่น	ค่าสนับสนุน
Symbian series 60 เวอร์ชั่น 0.9 → Virus	965	0	0.00%	0.00%
Symbian series 60 เวอร์ชั่น 1.2 → Virus	1462	977	66.83%	1.77%
Symbian series 60 เวอร์ชั่น 2.0 → Virus	3911	2479	63.39%	4.50%
Symbian series 60 เวอร์ชั่น 2.1 → Virus	6330	2750	43.44%	4.99%
Symbian series 60 เวอร์ชั่น 2.6 → Virus	11543	5273	45.68%	9.57%
Symbian series 60 เวอร์ชั่น 2.8 → Virus	23693	6323	26.69%	11.47%
Symbian series 60 เวอร์ชั่น 3.0 → Virus	2048	0	0.00%	0.00%
Symbian series 60 เวอร์ชั่น 3.1 → Virus	5172	0	0.00%	0.00%
Symbian series 60 เวอร์ชั่น 3.2 → Virus	0	0	-	0.00%

จากตาราง 9 เมื่อพิจารณาตามหลักเกณฑ์การจำแนกกฎความสัมพันธ์พบว่ากฎความสัมพันธ์ที่ค้นหาได้จากเงื่อนไขของหมวด Phone model ชุดที่ 2 ไม่ผ่านหลักเกณฑ์เช่นกัน

2.2 Content size → Virus MMS

ผลลัพธ์จากขั้นตอนการแปลงข้อมูล Content size จะได้หมวดข้อมูล 2 ชุด ดังตาราง 4 และ 5 ผู้วิจัยต้องคำนวณค่าพารามิเตอร์ 2 ครั้งตามชุดของข้อมูล เนื่องจากข้อมูลที่ได้จัดหมวดหมู่ไว้มีความแตกต่างกัน ส่งผลให้การคำนวณจำนวนของแต่ละพารามิเตอร์แตกต่างกัน

2.2.1 คำนวณหาค่าพารามิเตอร์ ค่าสนับสนุน ค่าความเชื่อมั่น และกฎความสัมพันธ์ โดยใช้หมวด Content size ชุดที่ 1 ตามตาราง 4 ซึ่งได้ผลการคำนวณดังตาราง 10

จำนวนรายการข้อความมัลติมีเดียทั้งหมด = 184,629

ตาราง 10 ผลการค้นหากฎความสัมพันธ์และคำนวณหาค่าพารามิเตอร์ ข้อมูลชุดฝึกสอน (Training set) วันที่ 1 มิถุนายน 2552 ของหมวด Content size ชุดที่ 1

กฎ	CNTCOND	CNTASSO	ค่าความเชื่อมั่น	ค่าสนับสนุน
0 – 5120 byte → Virus	25478	0	0.00%	0.00%
5121 – 10240 byte → Virus	23155	0	0.00%	0.00%
10241 - 15360 byte → Virus	10921	18	0.16%	0.01%
15361 - 20480 byte → Virus	10473	0	0.00%	0.00%
20481 - 25600 byte → Virus	12973	0	0.00%	0.00%
25601 - 30720 byte → Virus	28234	12920	45.76%	7.00%
30721 - 35840 byte → Virus	12179	208	1.71%	0.11%
35840 - 40960 byte → Virus	14833	4557	30.72%	2.47%
40961 - 46080 byte → Virus	6647	102	1.53%	0.06%
46081 - 51200 byte → Virus	4815	0	0.00%	0.00%
51201 - 56320 byte → Virus	3667	0	0.00%	0.00%
56321 - 61440 byte → Virus	3018	0	0.00%	0.00%
61441 - 66560 byte → Virus	3117	0	0.00%	0.00%
66561 - 71680 byte → Virus	2693	0	0.00%	0.00%
71681 - 76800 byte → Virus	2776	0	0.00%	0.00%
76801 - 81920 byte → Virus	2571	0	0.00%	0.00%
81921 - 87040 byte → Virus	1926	0	0.00%	0.00%
87041 - 92160 byte → Virus	1652	0	0.00%	0.00%
92161 - 97280 byte → Virus	1629	0	0.00%	0.00%

ตาราง 10 (ต่อ)

กฎ	CNTCOND	CNTASSO	ค่าความเชื่อมั่น	ค่าสนับสนุน
>= 97281 byte → Virus	11872	0	0.00%	0.00%

จากตาราง 10 เมื่อพิจารณาตามหลักเกณฑ์การจำแนกกฎความสัมพันธ์พบว่ากฎความสัมพันธ์ที่ค้นหาได้จากเงื่อนไขของหมวด Content size ชุดที่ 1 ไม่ผ่านหลักเกณฑ์เช่นกัน

2.2.2 คำนวณหาค่าพารามิเตอร์ ค่าสนับสนุน ค่าความเชื่อมั่น และกฎความสัมพันธ์ โดยใช้หมวด Content size ชุดที่ 2 ตามตาราง 5 ซึ่งได้ผลการคำนวณดังตาราง 11

จำนวนรายการข้อความมัลติมีเดียทั้งหมด = 81,981

ตาราง 11 ผลการค้นหากฎความสัมพันธ์และคำนวณหาค่าพารามิเตอร์ ข้อมูลชุดฝึกสอน (Training set) วันที่ 1 มิถุนายน 2552 ของหมวด Content size ชุดที่ 2

กฎ	CNTCOND	CNTASSO	ค่าความเชื่อมั่น	ค่าสนับสนุน
25600 - 26624 byte → Virus	2572	0	0.00%	0.00%
26625 - 27648 byte → Virus	11299	8713	77.11%	10.63%
27649 - 28672 byte → Virus	3543	0	0.00%	0.00%
28673 - 29696 byte → Virus	2589	0	0.00%	0.00%
29697 - 30720 byte → Virus	8233	4207	51.10%	5.13%
30721 – 31744 byte → Virus	3040	119	3.91%	0.15%
31745 – 32768 byte → Virus	2700	68	2.52%	0.08%
32769 – 33792 byte → Virus	2365	21	0.89%	0.03%

ตาราง 11 (ต่อ)

กฎ	CNTCOND	CNTASSO	ค่าความเชื่อมั่น	ค่าสนับสนุน
33793 – 34816 byte → Virus	2082	0	0.00%	0.00%
34817 – 35840 byte → Virus	1992	0	0.00%	0.00%
35841 – 36864 byte → Virus	0	0	0.00%	0.00%
36865 - 37888 byte → Virus	0	0	0.00%	0.00%
37889 - 38912 byte → Virus	0	0	0.00%	0.00%
38913 - 39936 byte → Virus	1652	0	0.00%	0.00%
39937 - 40960 byte → Virus	34172	4557	13.34%	5.56%

จากตาราง 11 เมื่อพิจารณาตามหลักเกณฑ์การจำแนกกฎความสัมพันธ์พบว่ากฎความสัมพันธ์ที่ค้นหาได้จากเงื่อนไขของหมวด Content size ชุดที่ 2 ไม่ผ่านหลักเกณฑ์เช่นกัน

2.3 Content type → Virus MMS

ผลลัพธ์จากขั้นตอนการแปลงข้อมูล Content type จะได้หมวดข้อมูลดังตาราง 6 เมื่อนำมาคำนวณหาค่าพารามิเตอร์ ค่าสนับสนุน ค่าความเชื่อมั่น และกฎความสัมพันธ์ ได้ผลการคำนวณดังตาราง 12

จำนวนรายการข้อความมัลติมีเดียทั้งหมด = 184,629

ตาราง 12 ผลการค้นหากฎความสัมพันธ์และคำนวณหาค่าพารามิเตอร์ ข้อมูลชุดฝึกสอน (Training set) วันที่ 1 มิถุนายน 2552 ของหมวด Content type

กฎ	CNTCOND	CNTASSO	ค่าความเชื่อมั่น	ค่าสนับสนุน
application/vnd.symbian.install → Virus	15	15	100.00%	0.01%
application/vnd.symbian.install และ text/plain → Virus	15001	15001	100.00%	8.12%
application/vnd.symbian.install และ application/vnd.wap.multipart.related → Virus	2789	2789	100.00%	1.51%
Content type ประเภทอื่นๆ → Virus	166824	0	0.00%	0.00%

จากตาราง 12 เมื่อพิจารณาตามหลักเกณฑ์การจำแนกกฎความสัมพันธ์พบว่ากฎความสัมพันธ์ที่ค้นหาได้จากเงื่อนไขของหมวด Content type มีค่าสนับสนุน (Support) มากกว่าค่าน้อยที่สุดของค่าสนับสนุน (Minimum Support) ที่กำหนดไว้ คือ 1% และมีค่าความเชื่อมั่น (Confidence) มากกว่า 80% อยู่ 2 กฎ ดังตาราง 13

ตาราง 13 ผลการจำแนกกฎความสัมพันธ์จากข้อมูลตาราง 12

กฎ	ค่าความเชื่อมั่น	ค่าสนับสนุน
application/vnd.symbian.install และ text/plain → Virus	100.00%	8.12%
application/vnd.symbian.install และ application/vnd.wap.multipart.related → Virus	100.00%	1.51%

ผลการจำแนกกฎความสัมพันธ์ของข้อมูลชุดฝึกสอน (Training set) ตั้งแต่วันที่ 1 มิถุนายน 2552 ถึงวันที่ 30 มิถุนายน 2552 ของเงื่อนไข Content type มีอยู่ 2 กฎความสัมพันธ์ที่มีค่าสนับสนุนมากกว่าค่าที่น้อยที่สุดของค่าสนับสนุน (Minimum Support) และมีค่าความเชื่อมั่นมากกว่าค่าที่น้อยที่สุดของค่าความเชื่อมั่น (Minimum Confidence) ที่ผู้วิจัยกำหนด ตามตาราง 13

ส่วนเงื่อนไข Phone model และ Content size เมื่อพิจารณาตามหลักเกณฑ์การจำแนกกฎความสัมพันธ์ พบว่ากฎความสัมพันธ์ที่ค้นหาได้จากเงื่อนไขทั้ง 2 ไม่ผ่านหลักเกณฑ์การจำแนกเลยแม้แต่วันเดียว ผู้วิจัยจึงไม่นำเสนอผลการจำแนกกฎความสัมพันธ์ตลอดทั้งเดือนของเงื่อนไขทั้ง 2 และจะนำเสนอเฉพาะเงื่อนไข Content type ดังตาราง 14 เท่านั้น

ตาราง 14 ผลการจำแนกกฎความสัมพันธ์ของข้อมูลชุดฝึกสอน (Training set) ตั้งแต่วันที่ 1-30 มิถุนายน 2552 ของเงื่อนไข Content type

วันที่	กฎ	ค่าความเชื่อมั่น	ค่าสนับสนุน
01/06/2552	application/vnd.symbian.install และ text/plain→ Virus	100.00%	8.12%
	application/vnd.symbian.install และ application/vnd.wap.multipart.related→ Virus	100.00%	1.51%
02/06/2552	application/vnd.symbian.install และ text/plain→ Virus	100.00%	9.13%
	application/vnd.symbian.install และ application/vnd.wap.multipart.related→ Virus	100.00%	1.67%
03/06/2552	application/vnd.symbian.install และ text/plain→ Virus	100.00%	8.55%
	application/vnd.symbian.install และ application/vnd.wap.multipart.related→ Virus	100.00%	1.64%
04/06/2552	application/vnd.symbian.install และ text/plain→ Virus	100.00%	8.61%
	application/vnd.symbian.install และ application/vnd.wap.multipart.related→ Virus	100.00%	1.43%

ตาราง 14 (ต่อ)

วันที่	กฎ	ค่าความเชื่อมั่น	ค่าสนับสนุน
05/06/2552	application/vnd.symbian.install และ text/plain→ Virus	100.00%	8.80%
	application/vnd.symbian.install และ application/vnd.wap.multipart.related→ Virus	100.00%	1.43%
06/06/2552	application/vnd.symbian.install และ text/plain→ Virus	100.00%	8.99%
	application/vnd.symbian.install และ application/vnd.wap.multipart.related→ Virus	100.00%	1.58%
07/06/2552	application/vnd.symbian.install และ text/plain→ Virus	100.00%	8.47%
	application/vnd.symbian.install และ application/vnd.wap.multipart.related→ Virus	100.00%	1.41%
08/06/2552	application/vnd.symbian.install และ text/plain→ Virus	100.00%	8.48%
	application/vnd.symbian.install และ application/vnd.wap.multipart.related→ Virus	100.00%	1.39%
09/06/2552	application/vnd.symbian.install และ text/plain→ Virus	100.00%	8.40%
	application/vnd.symbian.install และ application/vnd.wap.multipart.related→ Virus	100.00%	1.49%
10/06/2552	application/vnd.symbian.install และ text/plain→ Virus	100.00%	8.69%
	application/vnd.symbian.install และ application/vnd.wap.multipart.related→ Virus	100.00%	1.34%
11/06/2552	application/vnd.symbian.install และ text/plain→ Virus	100.00%	8.55%
	application/vnd.symbian.install และ application/vnd.wap.multipart.related→ Virus	100.00%	1.39%
12/06/2552	application/vnd.symbian.install และ text/plain→ Virus	100.00%	8.89%
	application/vnd.symbian.install และ application/vnd.wap.multipart.related→ Virus	100.00%	1.44%

ตาราง 14 (ต่อ)

วันที่	กฎ	ค่าความเชื่อมั่น	ค่าสนับสนุน
13/06/2552	application/vnd.symbian.install และ text/plain→ Virus	100.00%	10.96%
	application/vnd.symbian.install และ application/vnd.wap.multipart.related→ Virus	100.00%	1.45%
14/06/2552	application/vnd.symbian.install และ text/plain→ Virus	100.00%	21.06%
	application/vnd.symbian.install และ application/vnd.wap.multipart.related→ Virus	100.00%	2.59%
15/06/2552	application/vnd.symbian.install และ text/plain→ Virus	100.00%	8.70%
	application/vnd.symbian.install และ application/vnd.wap.multipart.related→ Virus	100.00%	1.53%
16/06/2552	application/vnd.symbian.install และ text/plain→ Virus	100.00%	9.04%
	application/vnd.symbian.install และ application/vnd.wap.multipart.related→ Virus	100.00%	1.59%
17/06/2552	application/vnd.symbian.install และ text/plain→ Virus	100.00%	9.07%
	application/vnd.symbian.install และ application/vnd.wap.multipart.related→ Virus	100.00%	1.48%
18/06/2552	application/vnd.symbian.install และ text/plain→ Virus	100.00%	9.13%
	application/vnd.symbian.install และ application/vnd.wap.multipart.related→ Virus	100.00%	1.44%
19/06/2552	application/vnd.symbian.install และ text/plain→ Virus	100.00%	9.25%
	application/vnd.symbian.install และ application/vnd.wap.multipart.related→ Virus	100.00%	1.50%
20/06/2552	application/vnd.symbian.install และ text/plain→ Virus	100.00%	9.12%
	application/vnd.symbian.install และ application/vnd.wap.multipart.related→ Virus	100.00%	1.50%

ตาราง 14 (ต่อ)

วันที่	กฎ	ค่าความเชื่อมั่น	ค่าสนับสนุน
21/06/2552	application/vnd.symbian.install และ text/plain→ Virus	100.00%	8.15%
	application/vnd.symbian.install และ application/vnd.wap.multipart.related→ Virus	100.00%	1.35%
22/06/2552	application/vnd.symbian.install และ text/plain→ Virus	100.00%	8.30%
	application/vnd.symbian.install และ application/vnd.wap.multipart.related→ Virus	100.00%	1.42%
23/06/2552	application/vnd.symbian.install และ text/plain→ Virus	100.00%	8.08%
	application/vnd.symbian.install และ application/vnd.wap.multipart.related→ Virus	100.00%	1.43%
24/06/2552	application/vnd.symbian.install และ text/plain→ Virus	100.00%	8.23%
	application/vnd.symbian.install และ application/vnd.wap.multipart.related→ Virus	100.00%	1.48%
25/06/2552	application/vnd.symbian.install และ text/plain→ Virus	100.00%	8.68%
	application/vnd.symbian.install และ application/vnd.wap.multipart.related→ Virus	100.00%	1.32%
26/06/2552	application/vnd.symbian.install และ text/plain→ Virus	100.00%	8.88%
	application/vnd.symbian.install และ application/vnd.wap.multipart.related→ Virus	100.00%	1.49%
27/06/2552	application/vnd.symbian.install และ text/plain→ Virus	100.00%	9.14%
	application/vnd.symbian.install และ application/vnd.wap.multipart.related→ Virus	100.00%	1.61%
28/06/2552	application/vnd.symbian.install และ text/plain→ Virus	100.00%	8.43%
	application/vnd.symbian.install และ application/vnd.wap.multipart.related→ Virus	100.00%	1.45

ตาราง 14 (ต่อ)

วันที่	กฎ	ค่าความเชื่อมั่น	ค่าสนับสนุน
29/06/2552	application/vnd.symbian.install และ text/plain→ Virus	100.00%	7.80%
	application/vnd.symbian.install และ application/vnd.wap.multipart.related→ Virus	100.00%	1.28%
30/06/2552	application/vnd.symbian.install และ text/plain→ Virus	100.00%	8.91%
	application/vnd.symbian.install และ application/vnd.wap.multipart.related→ Virus	100.00%	1.53%

3. พัฒนาโปรแกรมเพื่อใช้ตรวจจับข้อความมัลติมีเดียที่มีไวรัสโดยใช้เทคนิคเหมืองข้อมูล

เมื่อได้ศึกษาขั้นตอนการออกแบบโปรแกรม โดยการค้นหากฎความสัมพันธ์แล้ว จึงดำเนินการพัฒนาโปรแกรมเพื่อใช้ในการตรวจจับข้อความมัลติมีเดียที่มีไวรัส โดยผู้วิจัยนำกฎความสัมพันธ์ที่ผ่านการจำแนกกฎความสัมพันธ์ทั้ง 2 กฎ ของเงื่อนไขประเภทของเนื้อหาที่ส่งมากับข้อความมัลติมีเดีย (Content Type) มาเป็นกระบวนการทำงานหลัก ซึ่งต่อไปนี้จะอธิบายรายละเอียดขั้นตอนการพัฒนาโปรแกรมดังนี้

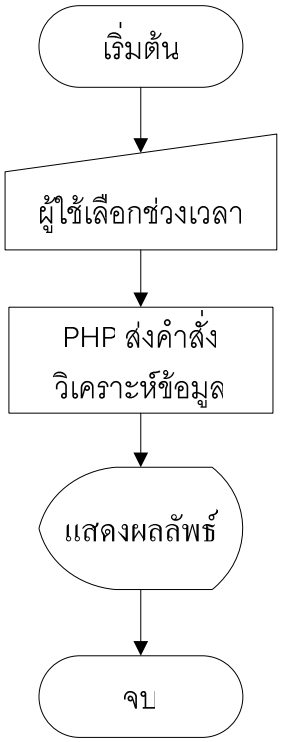
3.1 ขั้นตอนการทำงานของโปรแกรม

กระบวนการทำงานหลักของโปรแกรม มีดังนี้

3.1.1 ผู้ใช้เลือกช่วงเวลาที่ต้องการตรวจจับข้อความมัลติมีเดียที่มีไวรัส

3.1.2 โปรแกรมภาษา PHP ส่งคำสั่งไปทำการวิเคราะห์ข้อมูลการส่งข้อความมัลติมีเดียของผู้ใช้งานภายในระบบบริการรับส่งข้อความมัลติมีเดีย (MMSC System) ที่เก็บอยู่ในฐานข้อมูลมายเอสคิวแอล (MySQL) คำสั่งที่โปรแกรมภาษา PHP ส่งไปมาจากการนำกฎความสัมพันธ์ทั้ง 2 กฎ ของเงื่อนไขประเภทของเนื้อหาที่ส่งมากับข้อความมัลติมีเดีย (Content Type) มาแปลงเป็นภาษาโปรแกรม PHP และมายเอสคิวแอล (MySQL)

3.1.3 โปรแกรมภาษา PHP แสดงผลลัพธ์ที่ได้จากการวิเคราะห์เป็นข้อมูลที่เข้าใจง่าย คือ กราฟ ตัวเลขสรุปจำนวนข้อความมัลติมีเดีย จำนวนข้อความมัลติมีเดียที่มีไวรัสรายชั่วโมง และรายการข้อความมัลติมีเดียที่มีไวรัส

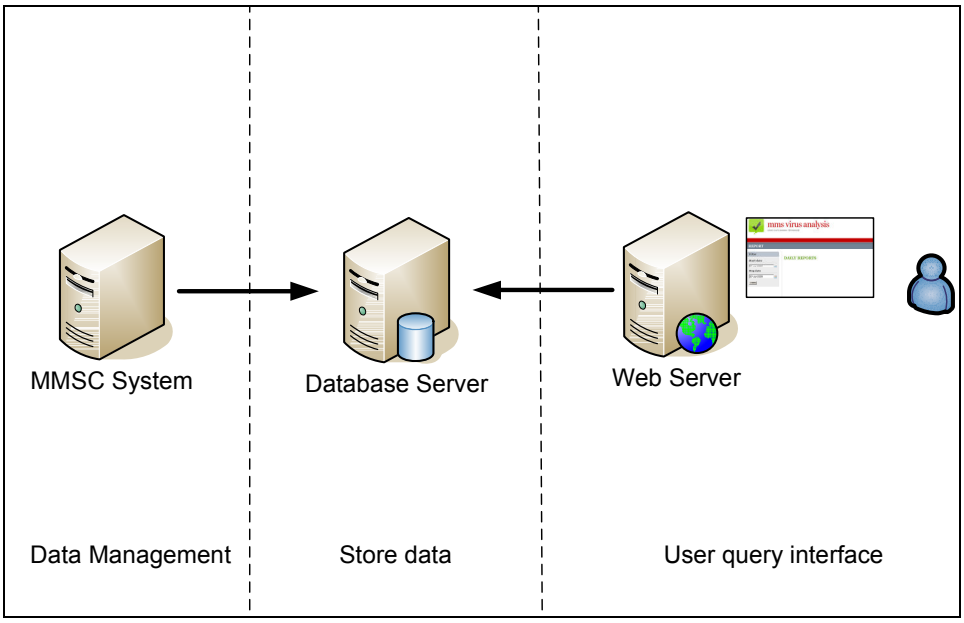


ภาพประกอบ 13 ผังงาน (Flowchart) ขั้นตอนการทำงานของโปรแกรม

3.2 การออกแบบระบบ

3.2.1 สถาปัตยกรรมระบบ

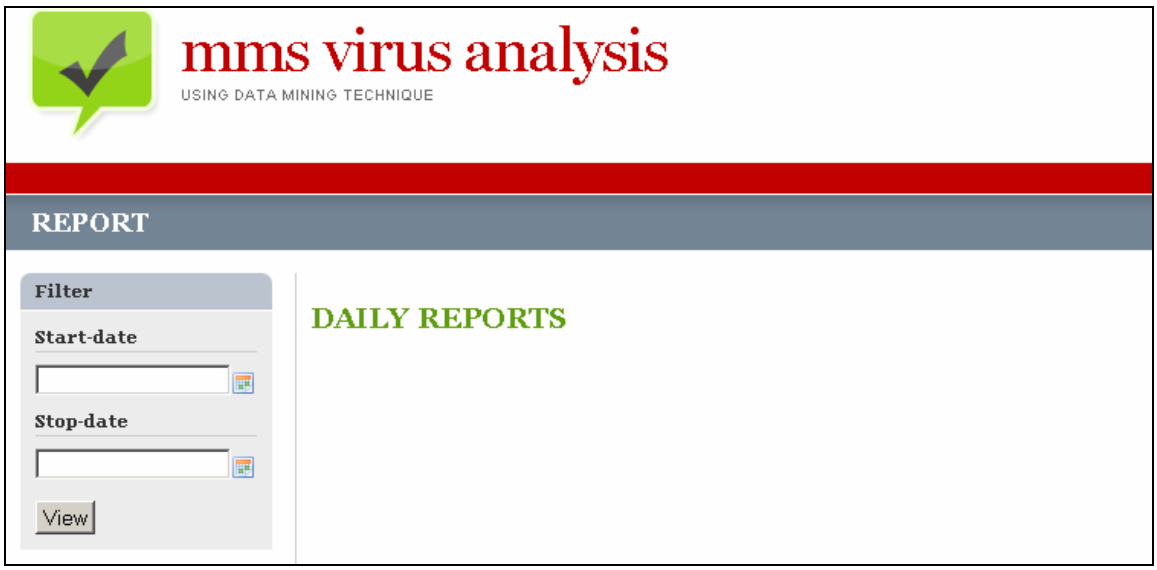
สถาปัตยกรรมระบบจะเป็นลักษณะเว็บแอพพลิเคชั่น (Web application) ซึ่งพัฒนาขึ้นด้วย ภาษา PHP, จาวาสคริปต์ (JavaScript), ฐานข้อมูลมายเอสคิวแอล (MySQL) และอะแพชี เว็บเซิร์ฟเวอร์ (Apache Web Server)



ภาพประกอบ 14 สถาปัตยกรรมระบบ

3.2.2 การออกแบบหน้าจอ (User Interface Design)

หน้าจอการใช้งานจะประกอบด้วยเมนูการทำงานดังนี้

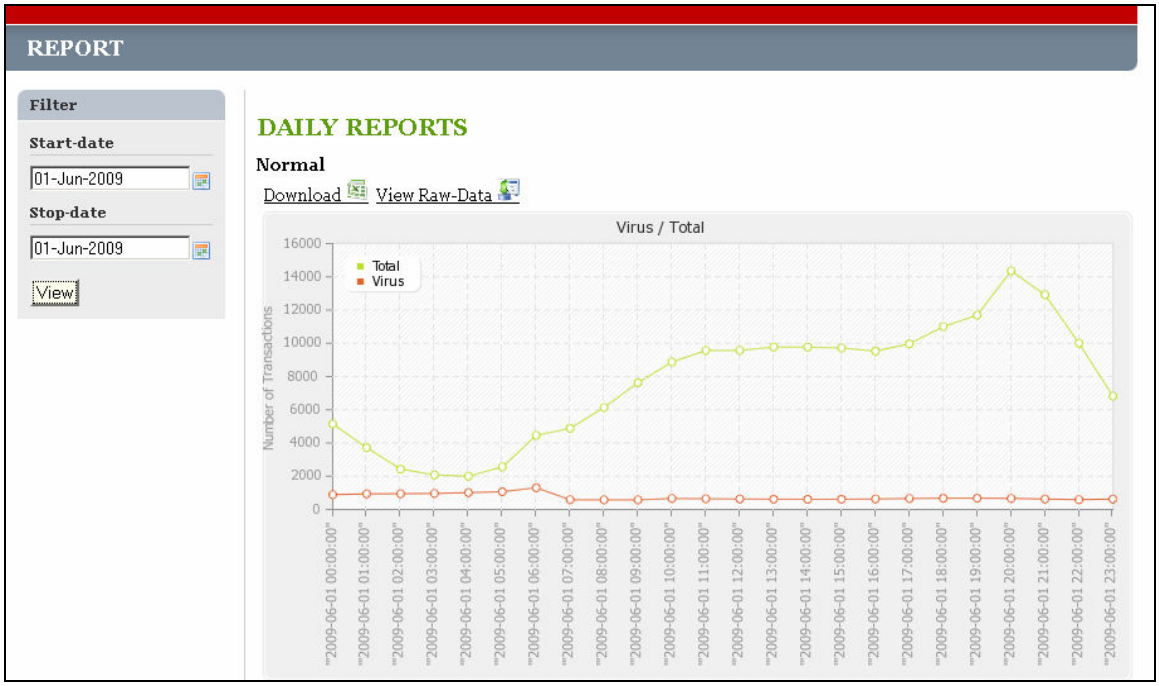


ภาพประกอบ 15 เมนูการใช้งานเว็บแอพพลิเคชั่น (Web application)

ในส่วนของการใช้งาน Web Application จะกล่าวโดยละเอียดในภาคผนวก

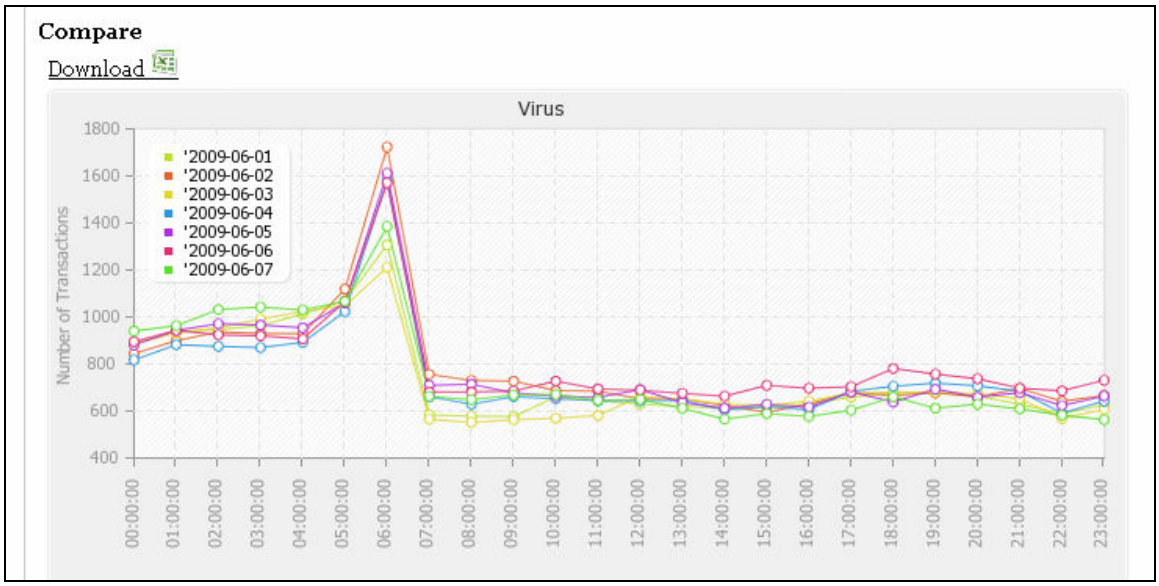
3.3 ผลลัพธ์ของการใช้เว็บแอพพลิเคชั่น (Web application) **วิเคราะห์ข้อมูลชุดฝึกสอน** (Training set) **วันที่** 1-7 **มิถุนายน** 2552

3.3.1 ผลลัพธ์ของการวิเคราะห์ข้อมูลการส่งข้อความมัลติมีเดียของผู้ใช้งานภายในระบบบริการรับส่งข้อความมัลติมีเดีย (MMSC System) รายชั่วโมง แสดงดังภาพประกอบ 16 ซึ่งหน้าจอของเว็บแอพพลิเคชั่นจะแสดงช่วงเวลาที่เลือกวิเคราะห์ข้อมูล กราฟสรุปจำนวนข้อความมัลติมีเดียทั้งหมดเป็นรายชั่วโมงตามเส้นกราฟสีเหลืองและจำนวนข้อความมัลติมีเดียที่มีไวรัสเป็นรายชั่วโมงตามเส้นกราฟสีแดง ไฟล์เอ็กเซลแสดงจำนวนข้อความมัลติมีเดียทั้งหมดเป็นรายชั่วโมงและจำนวนข้อความมัลติมีเดียที่มีไวรัสเป็นรายชั่วโมง รวมทั้งไฟล์เอ็กเซลแสดงรายการข้อความมัลติมีเดียที่มีไวรัส ซึ่งทั้งหมดสามารถดาวน์โหลดได้จากหน้าเว็บแอพพลิเคชั่น



ภาพประกอบ 16 ผลลัพธ์ของการใช้เว็บแอพพลิเคชั่น (Web application) วิเคราะห์ข้อมูลการส่งข้อความมัลติมีเดียของผู้ใช้งานรายชั่วโมง วันที่ 1 มิถุนายน 2552

3.3.2 ผลลัพธ์ของการวิเคราะห์รายการข้อความมัลติมีเดียที่มีไวรัสเปรียบเทียบรายชั่วโมง แสดงดังภาพประกอบ 17 ซึ่งหน้าจอของเว็บแอพพลิเคชั่นจะแสดงช่วงเวลาที่เลือกวิเคราะห์ข้อมูล กราฟสรุปจำนวนข้อความมัลติมีเดียที่มีไวรัสเป็นรายชั่วโมงตามเส้นกราฟแต่ละวัน ไฟล์เอ็กเซลแสดงจำนวนข้อความมัลติมีเดียที่มีไวรัสเป็นรายชั่วโมงของแต่ละวัน ซึ่งทั้งหมดสามารถดาวน์โหลดได้จากหน้าเว็บแอพพลิเคชั่น



ภาพประกอบ 17 ผลลัพธ์ของการใช้เว็บแอพพลิเคชั่น (Web application) วิเคราะห์รายการข้อความมัลติมีเดียที่มีไวรัสเปรียบเทียบรายชั่วโมง วันที่ 1-7 มิถุนายน 2552

4. ทดสอบความถูกต้องของโปรแกรมที่พัฒนาขึ้น โดยใช้ข้อมูลชุดทดสอบ

ทดสอบความถูกต้องของโปรแกรมที่พัฒนาขึ้นกับข้อมูลการส่งข้อความมัลติมีเดียของผู้ใช้งานภายในระบบบริการรับส่งข้อความมัลติมีเดีย (MMSC System) จำนวน 4 Server ระหว่างวันที่ 1-31 กรกฎาคม 2552 ซึ่งเป็นข้อมูลชุดทดสอบ (Test set) ตามกระบวนการทดสอบดังภาพประกอบ 8 ในงานวิจัยบทที่ 3

4.1 ผลลัพธ์ของการตรวจจับข้อความมัลติมีเดียที่มีไวรัส

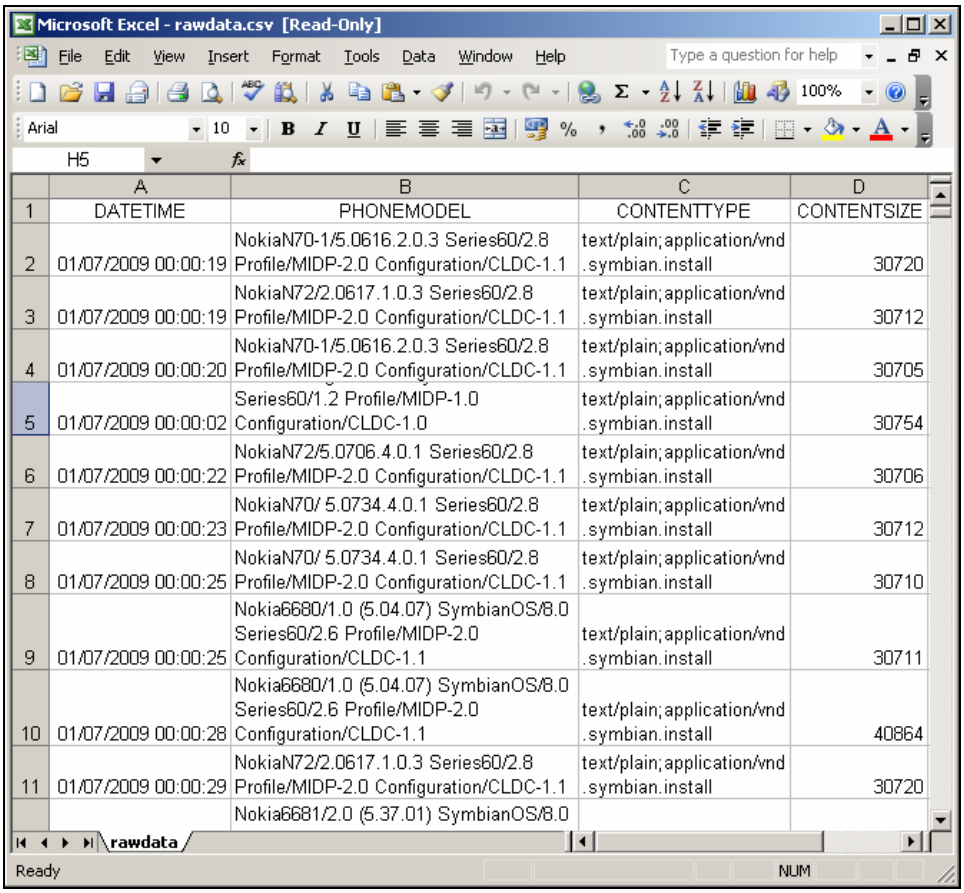
ผลลัพธ์ของการตรวจจับข้อความมัลติมีเดียที่มีไวรัสโดยโปรแกรมที่พัฒนาขึ้นได้จำนวนข้อความมัลติมีเดียที่มีไวรัสดังต่อไปนี้

ตาราง 15 จำนวนข้อความมัลติมีเดียที่มีไวรัสที่ตรวจจับโดยโปรแกรมที่พัฒนาขึ้น กับข้อมูลชุดทดสอบ (Test set) ตั้งแต่วันที่ 1-31 กรกฎาคม 2552 รายวัน

วันที่	จำนวนข้อความมัลติมีเดียที่มีไวรัส
01/07/2552	19350
02/07/2552	18338
03/07/2552	18232
04/07/2552	18937
05/07/2552	18360
06/07/2552	17656
07/07/2552	19029
08/07/2552	18506
09/07/2552	17718
10/07/2552	10841
11/07/2552	2148
12/07/2552	6538
13/07/2552	7497
14/07/2552	43183
15/07/2552	19833
16/07/2552	18836
17/07/2552	17781
18/07/2552	18454
19/07/2552	18559
20/07/2552	17531
21/07/2552	18655
22/07/2552	18052
23/07/2552	18727
24/07/2552	18458
25/07/2552	18871

ตาราง 15 (ต่อ)

วันที่	จำนวนข้อความมัลติมีเดียที่มีไวรัส
26/07/2552	18505
27/07/2552	17754
28/07/2552	10309
29/07/2552	18673
30/07/2552	19012
31/07/2552	18416



	A	B	C	D
1	DATETIME	PHONEMODEL	CONTENTTYPE	CONTENTSIZE
2	01/07/2009 00:00:19	NokiaN70-1/5.0616.2.0.3 Series60/2.8 Profile/MIDP-2.0 Configuration/CLDC-1.1	text/plain;application/vnd.symbian.install	30720
3	01/07/2009 00:00:19	NokiaN72/2.0617.1.0.3 Series60/2.8 Profile/MIDP-2.0 Configuration/CLDC-1.1	text/plain;application/vnd.symbian.install	30712
4	01/07/2009 00:00:20	NokiaN70-1/5.0616.2.0.3 Series60/2.8 Profile/MIDP-2.0 Configuration/CLDC-1.1	text/plain;application/vnd.symbian.install	30705
5	01/07/2009 00:00:02	Series60/1.2 Profile/MIDP-1.0 Configuration/CLDC-1.0	text/plain;application/vnd.symbian.install	30754
6	01/07/2009 00:00:22	NokiaN72/5.0706.4.0.1 Series60/2.8 Profile/MIDP-2.0 Configuration/CLDC-1.1	text/plain;application/vnd.symbian.install	30706
7	01/07/2009 00:00:23	NokiaN70/ 5.0734.4.0.1 Series60/2.8 Profile/MIDP-2.0 Configuration/CLDC-1.1	text/plain;application/vnd.symbian.install	30712
8	01/07/2009 00:00:25	NokiaN70/ 5.0734.4.0.1 Series60/2.8 Profile/MIDP-2.0 Configuration/CLDC-1.1	text/plain;application/vnd.symbian.install	30710
9	01/07/2009 00:00:25	Nokia6680/1.0 (5.04.07) SymbianOS/8.0 Series60/2.6 Profile/MIDP-2.0 Configuration/CLDC-1.1	text/plain;application/vnd.symbian.install	30711
10	01/07/2009 00:00:28	Nokia6680/1.0 (5.04.07) SymbianOS/8.0 Series60/2.6 Profile/MIDP-2.0 Configuration/CLDC-1.1	text/plain;application/vnd.symbian.install	40864
11	01/07/2009 00:00:29	NokiaN72/2.0617.1.0.3 Series60/2.8 Profile/MIDP-2.0 Configuration/CLDC-1.1	text/plain;application/vnd.symbian.install	30720
		Nokia6681/2.0 (5.37.01) SymbianOS/8.0		

ภาพประกอบ 18 ตัวอย่างรายการข้อความมัลติมีเดียที่มีไวรัสที่ตรวจจับโดยโปรแกรมที่พัฒนาขึ้น กับข้อมูลชุดทดสอบ (Test set) ตั้งแต่วันที่ 1-31 กรกฎาคม 2552

4.2 ผลลัพธ์ของการตรวจสอบผลการตรวจจับ

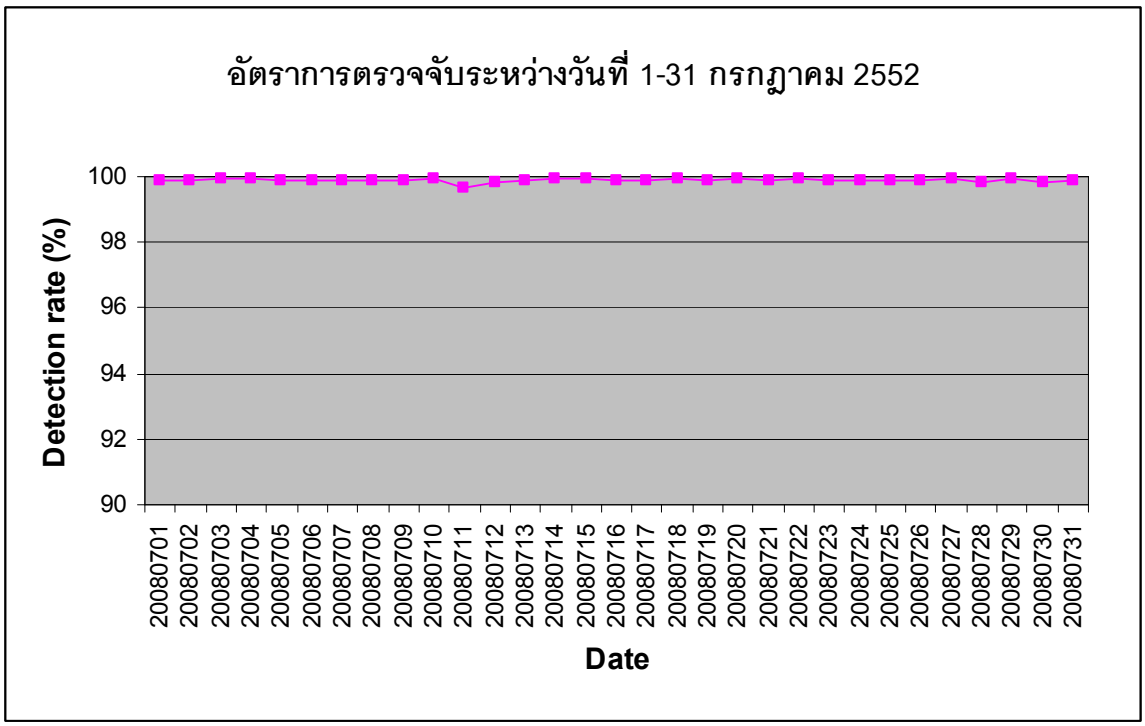
นำผลลัพธ์ที่ได้จากการตรวจจับข้อความมัลติมีเดียที่มีไวรัสโดยโปรแกรมที่พัฒนาขึ้นไปเปรียบเทียบกับผลที่ได้จากการตรวจจับโดยใช้ Viral Marketing Tool พบว่ารายการข้อความมัลติมีเดียที่ตรวจจับได้ตรงกับรายการที่ Viral Marketing Tool มีในแต่ละวัน แสดงว่าผลลัพธ์ของการตรวจจับข้อความมัลติมีเดียที่มีไวรัสโดยโปรแกรมที่พัฒนาขึ้นมีความถูกต้อง

5. ประเมินผลการตรวจจับข้อความมัลติมีเดีย (MMS) ที่มีไวรัสจากโปรแกรมที่ออกแบบขึ้นและจากการใช้ Viral Marketing Tool

จากตัวชี้วัดและสถิติที่ใช้ในการวิจัยที่ผู้วิจัยได้กำหนดไว้ในงานวิจัยบทที่ 3 คือ อัตราการตรวจจับ (Detection Rate) ระหว่างโปรแกรมที่พัฒนาขึ้น กับ Viral Marketing Tool และ จำนวนของข้อความมัลติมีเดีย (MMS) ที่มีไวรัส เปรียบเทียบกันระหว่างการใช้โปรแกรมที่ออกแบบขึ้นตรวจจับ กับการตรวจจับโดยใช้ Viral Marketing Tool โดยให้การตรวจจับนั้นอยู่ภายใต้เงื่อนไขเดียวกัน โดยทำการประเมินดังนี้

5.1 การประเมินประสิทธิภาพของอัตราการตรวจจับ

อัตราการตรวจจับ (Detection Rate) เป็นการวัดค่าความถูกต้องของการตรวจจับข้อความมัลติมีเดีย (MMS) ที่มีไวรัส ผู้วิจัยใช้ผลที่ได้จากการทดสอบความถูกต้องของโปรแกรมที่พัฒนาขึ้นโดยใช้ข้อมูลชุดทดสอบระหว่างวันที่ 1-31 กรกฎาคม 2552 ในหัวข้อที่ 4 มาเป็นค่าตั้งต้นในการหาอัตราการตรวจจับ โดยอัตราการตรวจจับที่ได้จากการใช้ Viral Marketing Tool เป็น 100% เนื่องจาก Tool ตัวนี้สามารถตรวจสอบข้อมูลการส่งข้อความมัลติมีเดียของผู้ใช้งานภายในระบบบริการรับส่งข้อความมัลติมีเดีย (MMSC System) ได้ถึงระดับ Packet ซึ่งสามารถที่จะเห็นคุณสมบัติของไวรัสชนิดนี้ครบทุกด้าน จึงส่งผลให้มีประสิทธิภาพ 100% เต็ม ส่วนอัตราการตรวจจับของโปรแกรมที่พัฒนาขึ้นได้ดังนี้



ภาพประกอบ 19 อัตราการตรวจจับข้อความมัลติมีเดียที่มีไวรัสโดยโปรแกรมที่พัฒนาขึ้นกับข้อมูลชุดทดสอบ (Test set) ตั้งแต่วันที่ 1-31 กรกฎาคม 2552

จากภาพประกอบ 19 ผลของการตรวจจับข้อความมัลติมีเดียที่มีไวรัสจากข้อมูลการส่งข้อความมัลติมีเดียของผู้ใช้งานภายในระบบบริการรับส่งข้อความมัลติมีเดีย (MMSC System) ระหว่างวันที่ 1-31 กรกฎาคม 2552 จำนวน 5,445,348 รายการ ผู้วิจัยพบว่าอัตราการตรวจจับข้อความมัลติมีเดียที่มีไวรัสตลอดทั้งเดือนกรกฎาคม มีประสิทธิภาพใกล้เคียงกัน นั่นคือมีประสิทธิภาพเฉลี่ยอยู่ที่ 99.89%

5.2 การเปรียบเทียบจำนวนของข้อความมัลติมีเดีย (MMS) ที่มีไวรัส ระหว่างการใช้โปรแกรมที่ออกแบบขึ้นตรวจจับ กับการใช้ Viral Marketing Tool ตรวจจับ

ตาราง 16 แสดงการเปรียบเทียบจำนวนข้อความมัลติมีเดียที่มีไวรัสที่ตรวจจับโดยโปรแกรมที่พัฒนาขึ้นและ Viral Marketing Tool กับข้อมูลชุดทดสอบ (Test set) ตั้งแต่วันที่ 1-31 กรกฎาคม 2552 รายวัน

วันที่	จำนวนข้อความมัลติมีเดียที่มีไวรัสตรวจจับโดยโปรแกรมที่พัฒนาขึ้น	จำนวนข้อความมัลติมีเดียที่มีไวรัสตรวจจับโดย Viral Marketing Tool	ร้อยละความแตกต่างของผลที่ได้จากโปรแกรมทั้งสอง
01/07/2552	19350	19370	0.10
02/07/2552	18338	18357	0.10
03/07/2552	18232	18243	0.06
04/07/2552	18937	18951	0.07
05/07/2552	18360	18376	0.09
06/07/2552	17656	17672	0.09
07/07/2552	19029	19046	0.09
08/07/2552	18506	18523	0.09
09/07/2552	17718	17739	0.12
10/07/2552	10841	10849	0.07
11/07/2552	2148	2155	0.32
12/07/2552	6538	6548	0.15
13/07/2552	7497	7505	0.11
14/07/2552	43183	43199	0.04
15/07/2552	19833	19845	0.06
16/07/2552	18836	18861	0.13
17/07/2552	17781	17800	0.11
18/07/2552	18454	18467	0.07
19/07/2552	18559	18584	0.13

ตาราง 16 (ต่อ)

วันที่	จำนวนข้อความมัลติมีเดียที่มีไวรัสตรวจจับโดยโปรแกรมที่พัฒนาขึ้น	จำนวนข้อความมัลติมีเดียที่มีไวรัสตรวจจับโดย Viral Marketing Tool	ร้อยละความแตกต่างของผลที่ได้จากโปรแกรมทั้งสอง
20/07/2552	17531	17544	0.07
21/07/2552	18655	18677	0.12
22/07/2552	18052	18067	0.08
23/07/2552	18727	18743	0.09
24/07/2552	18458	18481	0.12
25/07/2552	18871	18895	0.13
26/07/2552	18505	18522	0.09
27/07/2552	17754	17767	0.07
28/07/2552	10309	10328	0.18
29/07/2552	18673	18686	0.07
30/07/2552	19012	19040	0.15
31/07/2552	18416	18433	0.09
รวม	542,759	543,273	0.11

บทที่ 5
สรุปผลการวิจัย อธิปรายผล และข้อเสนอแนะ

การวิจัยเรื่องการตรวจจับไวรัสข้อความมัลติมีเดียโดยใช้เทคนิคการทำเหมืองข้อมูล สามารถสรุปผลโดยมีรายละเอียดตามหัวข้อต่างๆ ดังนี้

1. สรุปผลการวิจัย
2. อภิปรายผลการวิจัย
3. ข้อเสนอแนะ

1. สรุปผลการวิจัย

ผลการดำเนินงานวิจัยนี้แบ่งออกเป็น 4 ส่วน ซึ่งแต่ละส่วนสามารถสรุปผลได้ดังนี้

1.1 ออกแบบโปรแกรมเพื่อใช้ตรวจจับข้อความมัลติมีเดียที่มีไวรัสโดยใช้เทคนิคเหมืองข้อมูล

การออกแบบโปรแกรมในงานวิจัยนี้นำเทคนิคการทำเหมืองข้อมูล กฎการวิเคราะห์ความสัมพันธ์ (Association Rule) แบบวิธีค้นหารูปแบบ (Frequent Pattern Mining) มาใช้ค้นหาความสัมพันธ์ระหว่าง Phone model, Content size, Content type และข้อความมัลติมีเดีย (MMS) ที่มีไวรัสชนิด CommWarrior ซึ่งกำหนดเงื่อนไขทั้งหมด 3 เงื่อนไข ดังนี้ Phone model เป็นเงื่อนไข ข้อความมัลติมีเดีย (MMS) ที่มีไวรัสเป็นผลลัพธ์ที่เกิดขึ้น, Content size เป็นเงื่อนไข ข้อความมัลติมีเดีย (MMS) ที่มีไวรัสเป็นผลลัพธ์ที่เกิดขึ้น, Content type เป็นเงื่อนไข ข้อความมัลติมีเดีย (MMS) ที่มีไวรัสเป็นผลลัพธ์ที่เกิดขึ้น

ผลการจำแนกกฎความสัมพันธ์พบว่ากฎความสัมพันธ์ที่ค้นหาได้จากเงื่อนไขของหมวด Content type มีค่าสนับสนุน (Support) มากกว่าค่าน้อยที่สุดของค่าสนับสนุน (Minimum Support) ที่กำหนดไว้ คือ 1% และมีค่าความเชื่อมั่น (Confidence) มากกว่า 80% อยู่ 2 กฎ ดังนี้

application/vnd.symbian.install และ text/plain → Virus

application/vnd.symbian.install และapplication/vnd.wap.multipart.related → Virus

1.2 พัฒนาโปรแกรมเพื่อใช้ตรวจจับข้อความมัลติมีเดียที่มีไวรัสโดยใช้เทคนิคเหมืองข้อมูล

นำกฎความสัมพันธ์ที่ผ่านการจำแนกกฎความสัมพันธ์ของเงื่อนไขประเภทของเนื้อหาที่ส่งมากับข้อความมัลติมีเดีย (Content Type) ในขั้นตอนการออกแบบโปรแกรมมาเป็นกระบวนการทำงานหลัก โดยพัฒนาเป็นลักษณะเว็บแอพพลิเคชั่น (Web application) ซึ่งพัฒนาขึ้นด้วย ภาษา PHP, จาวาสคริปต์ (JavaScript), ฐานข้อมูลมายเอสคิวแอล (MySQL) และอะแพชี เว็บเซิร์ฟเวอร์ (Apache Web Server) โดยผลลัพธ์ของการวิเคราะห์ข้อมูลโดยใช้เว็บแอพพลิเคชั่นที่พัฒนาขึ้นจะสามารถแสดงช่วงเวลาที่เลือกวิเคราะห์ข้อมูล กราฟสรุปจำนวนข้อความมัลติมีเดียทั้งหมดเป็นรายชั่วโมง จำนวนข้อความมัลติมีเดียที่มีไวรัสเป็นรายชั่วโมง ไฟล์เอ็กเซลแสดงจำนวนข้อความมัลติมีเดียทั้งหมดเป็นรายชั่วโมงและจำนวนข้อความมัลติมีเดียที่มีไวรัสเป็นรายชั่วโมง รวมทั้งไฟล์เอ็กเซลแสดงรายการข้อความมัลติมีเดียที่มีไวรัส ซึ่งผู้ใช้งานโปรแกรมสามารถดาวน์โหลดได้จากหน้าเว็บแอพพลิเคชั่น

1.3 ทดสอบความถูกต้องของโปรแกรมที่พัฒนาขึ้น โดยใช้ข้อมูลชุดทดสอบ

รายการข้อความมัลติมีเดียที่มีไวรัสที่ได้จากการตรวจจับโดยโปรแกรมที่พัฒนาขึ้นจำนวน 542,759 ข้อความ เป็นรายการข้อความมัลติมีเดียที่ Viral Marketing Tool ก็ตรวจจับได้เช่นกัน ซึ่งไม่มีรายการใดเลยที่โปรแกรมที่พัฒนาขึ้นตรวจจับผิดพลาด แสดงว่าผลลัพธ์ของการตรวจจับข้อความมัลติมีเดียที่มีไวรัสโดยโปรแกรมที่พัฒนาขึ้นมีความถูกต้อง

1.4 ประเมินผลการตรวจจับข้อความมัลติมีเดีย (MMS) ที่มีไวรัสจากโปรแกรมที่ออกแบบขึ้นและจากการใช้ Viral Marketing Tool

ผลการตรวจจับข้อความมัลติมีเดียที่มีไวรัสจากข้อมูลการส่งข้อความมัลติมีเดียของผู้ใช้งานภายในระบบบริการรับส่งข้อความมัลติมีเดีย (MMSC System) ระหว่างวันที่ 1-31 กรกฎาคม 2552 จำนวน 5,445,348 รายการ พบข้อความมัลติมีเดียที่มีไวรัสจำนวน 542,759 ข้อความ จากข้อความมัลติมีเดียที่มีไวรัสทั้งหมด 543,273 ข้อความ ซึ่งคิดเป็นอัตราการตรวจจับ (Detection Rate) ที่มีประสิทธิภาพเฉลี่ยอยู่ที่ 99.89% นั่นคือมีประสิทธิภาพไม่แตกต่างจาก Viral Marketing Tool

2. อภิปรายผลการวิจัย

จากสมมติฐานที่ว่า “การออกแบบและพัฒนาโปรแกรมตรวจจับข้อความมัลติมีเดีย (MMS) ที่มีไวรัส โดยใช้เทคนิคการทำเหมืองข้อมูล แบบกฎการวิเคราะห์ความสัมพันธ์ (Association Rule) สามารถตรวจจับข้อความมัลติมีเดีย (MMS) ที่มีไวรัสได้มีประสิทธิภาพเทียบเท่า Viral Marketing Tool”

จากการทดสอบสมมติฐานโดยการวัดผลจาก ความถูกต้องของโปรแกรม คือต้องสามารถตรวจจับข้อความมัลติมีเดีย (MMS) ที่มีไวรัสได้ผลสอดคล้องกับการตรวจจับโดยใช้ Viral Marketing Tool ด้วยวิธีการคำนวณอัตราการตรวจจับ (Detection Rate) ระหว่างโปรแกรมที่พัฒนาขึ้น กับ Viral Marketing Tool พบว่าโปรแกรมที่พัฒนาขึ้นมีอัตราการตรวจจับ (Detection Rate) ที่ 99.89% ซึ่งสอดคล้องกับสมมติฐานการวิจัยที่ตั้งไว้ ทั้งนี้เนื่องจากข้อมูลการส่งข้อความมัลติมีเดียของผู้ใช้งานภายในระบบบริการรับส่งข้อความมัลติมีเดีย (MMSC System) ที่โปรแกรมที่พัฒนาขึ้นใช้ในการออกแบบและพัฒนา เป็นข้อมูลที่มีการเก็บรวบรวมไว้อยู่แล้ว และมีความละเอียดเพียงพอสำหรับการนำไปใช้วิเคราะห์ตรวจจับไวรัส ซึ่งไม่จำเป็นต้องมีความละเอียดถึงระดับ Packet เหมือนเช่นการตรวจจับโดยใช้ Viral Marketing Tool เนื่องจากผู้ให้บริการโครงข่ายโทรศัพท์มือถือต้องการผลลัพธ์ของการวิเคราะห์เพียงเพื่อนำไปหาแนวทางในการปรับปรุงคุณภาพของบริการเท่านั้น ซึ่งผลงานวิจัยนี้สามารถตรวจจับข้อความมัลติมีเดียที่มีไวรัสได้สอดคล้องกับการตรวจจับโดยใช้ Viral Marketing Tool ที่สามารถตรวจจับได้ในอัตราการตรวจจับ (Detection Rate) ที่ 100%

3. ข้อเสนอแนะ

3.1 ข้อจำกัดของการวิจัย

งานวิจัยนี้ออกแบบมาสำหรับใช้ในการตรวจจับข้อความมัลติมีเดีย (MMS) ที่มีไวรัสชนิด CommWarrior เท่านั้น ไม่สามารถนำไปตรวจจับไวรัสข้อความมัลติมีเดียที่มีคุณสมบัตินอกเหลือจากไวรัสชนิดนี้ได้

3.2 ข้อเสนอแนะสำหรับการนำผลการวิจัยไปใช้งาน

การนำโปรแกรมที่พัฒนาขึ้นไปใช้งานกับข้อมูลการส่งข้อความมัลติมีเดียของผู้ใช้งานภายในระบบบริการรับส่งข้อความมัลติมีเดีย (MMSC System) จริง จะต้องมีการดำเนินการเพิ่มเติมดังนี้

3.2.1 ถ้านำไปใช้งานกับข้อมูลการส่งข้อความมัลติมีเดียของผู้ใช้งานภายในระบบบริการรับส่งข้อความมัลติมีเดีย (MMSC System) จริงที่ไม่ได้มีการเขียน Log เช่นเดียวกับในงานวิจัยนี้ จะต้องแก้ไขโปรแกรมสำหรับใช้โหลดข้อมูล Log เข้าสู่ฐานข้อมูลใหม่ ให้เหมาะสมกับ Log

3.2.2 ต้องปรับค่าเวลาในการตอบกลับ (Response time) ของฐานข้อมูลกับเว็บแอพพลิเคชั่น (Web application) ให้เหมาะสมกัน ไม่เช่นนั้นเมื่อใช้งานเว็บแอพพลิเคชั่นวิเคราะห์ข้อมูลต่อเนื่องกันหลายๆ วัน จะไม่สามารถแสดงผลได้ เนื่องจากฐานข้อมูลตอบกลับมาหาเว็บแอพพลิเคชั่นช้าเกินเวลาที่กำหนดไว้

3.3 ข้อเสนอแนะสำหรับการวิจัยต่อไป

การวิจัยนี้มีข้อเสนอแนะอันจะเป็นประโยชน์ต่อผู้วิจัยที่ต้องการนำผลของการวิจัยไปพัฒนาหรือนำไปใช้ต่อ ดังนี้

3.3.1 ควรออกแบบโปรแกรมตรวจจับไวรัสข้อความมัลติมีเดียโดยใช้เทคนิคการทำเหมืองข้อมูลแบบอื่น ที่น่าจะช่วยให้อัตราการตรวจจับ (Detection Rate) เป็น 100% หรือดีขึ้นกว่างานวิจัยนี้

3.3.2 ควรนำวิธีการออกแบบโปรแกรมตรวจจับไวรัสข้อความมัลติมีเดียในงานวิจัยนี้ไปใช้พัฒนาโปรแกรมแอนตี้ไวรัส (Anti-Virus Software) เพื่อเป็นการต่อยอดงานวิจัยให้สามารถใช้แก้ปัญหาโดยตรงให้แก่ผู้ใช้งานบริการรับส่งข้อความมัลติมีเดียที่ประสบปัญหาไวรัสดังกล่าว

3.3.3 ควรนำเทคนิคการทำเหมืองข้อมูล แบบกฎการวิเคราะห์ความสัมพันธ์ (Association Rule) ไปประยุกต์ใช้ในการออกแบบโปรแกรมตรวจจับไวรัสประเภทอื่นที่มีคุณสมบัติใกล้เคียงกับงานวิจัยนี้ หรือไวรัสประเภทอื่นๆ ที่มีพฤติกรรมของการทำงานสามารถจับรูปแบบได้

บรรณานุกรม

บรรณานุกรม

เกียรติศักดิ์ โยชะนัง. (2550). *การการระบุและการตรวจจับรายการเปลี่ยนแปลงที่มุ่งร้ายสำหรับฐานข้อมูลโดยใช้วิธีการทำเหมืองข้อมูล*. วิทยานิพนธ์ วศ.ม. (วิศวกรรมไฟฟ้า). กรุงเทพฯ: บัณฑิตมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ. ถ่ายเอกสาร

กฤษณะ ไวยมัย; และ ธีระวัฒน์ พงษ์ศิริปรีดา. (2546). *การใช้เทคนิค Association Rule Discovery เพื่อการจัดสรรกฎหมายในการพิจารณาคดีความ*. สืบค้นเมื่อ 15 เมษายน 2552 จาก http://www.nectec.or.th/NTJ/No11/No11_full_6.php

จาแกต แบนดู พันดาร. (2551). *การเลือกลักษณะพิเศษโดยการใช้การทำเหมืองข้อมูล สำหรับการตรวจหาอีเมลขยะ*. วิทยานิพนธ์ วท.ม. (เทคโนโลยีสารสนเทศ). กรุงเทพฯ: บัณฑิตมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ. ถ่ายเอกสาร

ดุลย์วิทย์ ปรางชุมพล; และ ปานใจ ธารทัศนวงศ์. (2549). *การทำนายปริมาณการจราจรในเครือข่ายโดยใช้เทคนิคเหมืองข้อมูล*. ในการประชุมทางวิชาการระดับชาติด้านเทคโนโลยีสารสนเทศครั้งที่ 1. (NCIT2009). 2-3 พฤศจิกายน 2549 กรุงเทพฯ. หน้า 313-317

บุญเสริม กิจศิริกุล. (2546). *อัลกอริทึมการทำเหมืองข้อมูล*. กรุงเทพฯ: ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

พิจิตรา จอมศรี; และ ปานใจ ธารทัศนวงศ์. (2549). *การเพิ่มอัตราการพบในระบบพร๊อกซี่โดยใช้เทคนิคเหมืองข้อมูล*. วิทยานิพนธ์ วท.ม. (วิทยาการคอมพิวเตอร์). กรุงเทพฯ: บัณฑิตมหาวิทยาลัยศิลปากร. ถ่ายเอกสาร.

Berry; Michael J.A.; & Gordon S. Linnoff. (2004). Data Mining Techniques For Marketing, Sale and Customer Relationship Management. New York: Wiley Publishing.

Chang-Tien Lu; Arnold P. Boedihardjo; & Prajwal Manalwar. (2005). Exploiting Efficient Data Mining Techniques to Enhance Intrusion Detection Systems. In IEEE Proceedings of the 15th International Conference on Advances in Knowledge Discovery and Data Mining.:512-517. Retrieved February 25, 2009, from http://www.cacs.louisiana.edu/~jyoon/grad/adb/Presentation/Stu_Dm2EnhanceIntruDet.pdf

บรรณานุกรม (ต่อ)

Elovici,A.Y.; Kandel; M.Last; B.Shapira; & O.Zaafrany. (2008). *Using data mining Techniques for Detecting Terror-Related Activities on the Web*. Retrieved February 23, 2009, from http://www.zdnet.co.uk/white-papers/data-tools/2008/09/20/using-data-mining-techniques-for-detecting-terror-related-activities-on-the-web-260499737/

Frederic Perriot; & Peter Ferrie. (2007). *SymbOS.Commwarrior.A*. Retrieved February 18, 2009, from http://www.symantec.com/security_response/writeup.jsp?docid=2005-030721-2716-99

Jiawei Han; & Micheline Kamber. (2006). *Data Mining Concepts and Techniques*. Retrieved February 22, 2009, from http://www.google.com/books?hl=th&lr=&id=AfL0t-YzOrEC&oi=fnd&pg=PR19&dq=Data+Mining+Concepts+and+Techniques&ots=UuZSqSdrA5&sig=gvIAkMFSevN3UW2SKb6wNCPuBas#v=onepage&q&f=false

Kaspersky Lab ZAO. (2005). *Worm.SymbOS.Comwar.a*. Retrieved February 18, 2009, from http://www.securelist.com/en/descriptions/old75541

Marisa S. Viveros; John P. Nearhos; & Michael J. Rothman. (1996). Applying data mining techniques to a health insurance information system. In *proceedings of the 22th International Conference on Very Large Data Bases*. Retrieved February 25, 2009, from http://portal.acm.org/citation.cfm?id=673464

Panda Security. (2008). *ComWar.A.* Retrieved February 18, 2009, from http://www.panda security.com/homeusers/security-info/61929/ComWar.A

Sujaa Rani Mohan; E. K. Park; & Yijie Han. (2005). Associatoin Rule Based Data Mining Agent for Personalized Web Caching. *In proceeding of the COMPSAC 05.* Retrieved March 23, 2009, from http://www.computer.org/portal/web/csdl/doi/10.1109/COMPSAC.2005.46

Yi Hu; Brajendra Panda. (2004). A Data Mining Approach for Database Intrusion Detection. *In proceeding of ACM Applies Computing*. Retrieved March 23, 2009, from http://portal.acm.org/citation.cfm?id=968048

บรรณานุกรม (ต่อ)

Yi-Hung Wu; & P. Chen. (2002). Prediction of Web Page Accesses by Proxy Server Log. *In World Wide Web Journ.* Volume 5:67-68. Retrieved February 25, 2009, from http://www.springerlink.com/content/72gpmbqthk02q2pa/

ภาคผนวก

คู่มือการใช้งานโปรแกรมตรวจจับข้อความมัลติมีเดียที่มีไวรัส โดยใช้เทคนิคเหมืองข้อมูล

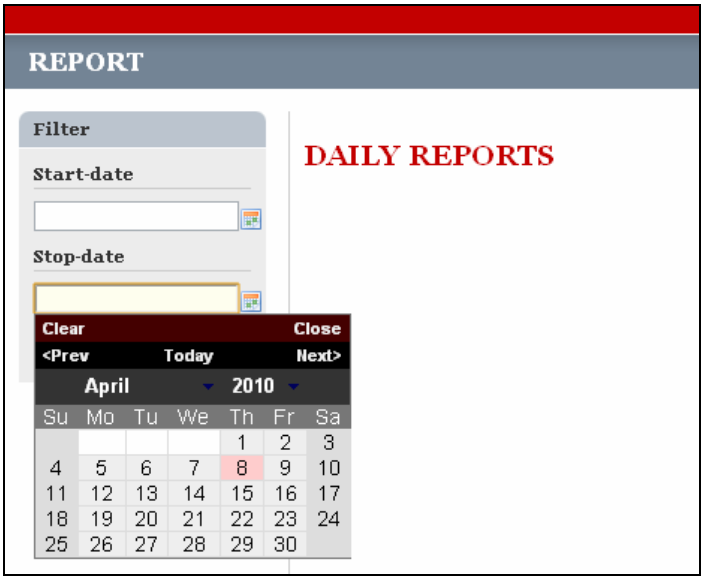
คู่มือการใช้งานโปรแกรมตรวจจับข้อความมัลติมีเดียที่มีไวรัสโดยใช้เทคนิคเหมืองข้อมูล ประกอบด้วยการอธิบายการใช้งานของโปรแกรมในแต่ละส่วน ดังนี้

1. การเรียกใช้โปรแกรม ผู้ใช้งานต้องเปิดเบราว์เซอร์ด้วย URL: http://x.x.x.x/mining โดย x.x.x.x หมายถึง IP address ของเครื่องที่ทำการติดตั้งโปรแกรมตรวจจับข้อความมัลติมีเดียที่มีไวรัสไว้ โดยเบราว์เซอร์จะแสดงผลลัพธ์ออกมาดังภาพประกอบแสดงส่วนติดต่อผู้ใช้โปรแกรม

ภาพประกอบแสดงส่วนติดต่อผู้ใช้โปรแกรม (User interface)

2. ผู้ใช้งานสามารถเลือกช่วงเวลาที่ต้องการตรวจจับข้อความมัลติมีเดียที่มีไวรัสจากส่วนติดต่อผู้ใช้โปรแกรม (User interface) โดยการเลือกช่วงเวลาที่เมนู Filter ดังภาพประกอบแสดงส่วนติดต่อผู้ใช้ในการเลือก Start-date และภาพประกอบแสดงส่วนติดต่อผู้ใช้ในการเลือก Stop-date โดย Start-date หมายถึงเวลาเริ่มต้น และ Stop-date หมายถึงเวลาสิ้นสุด

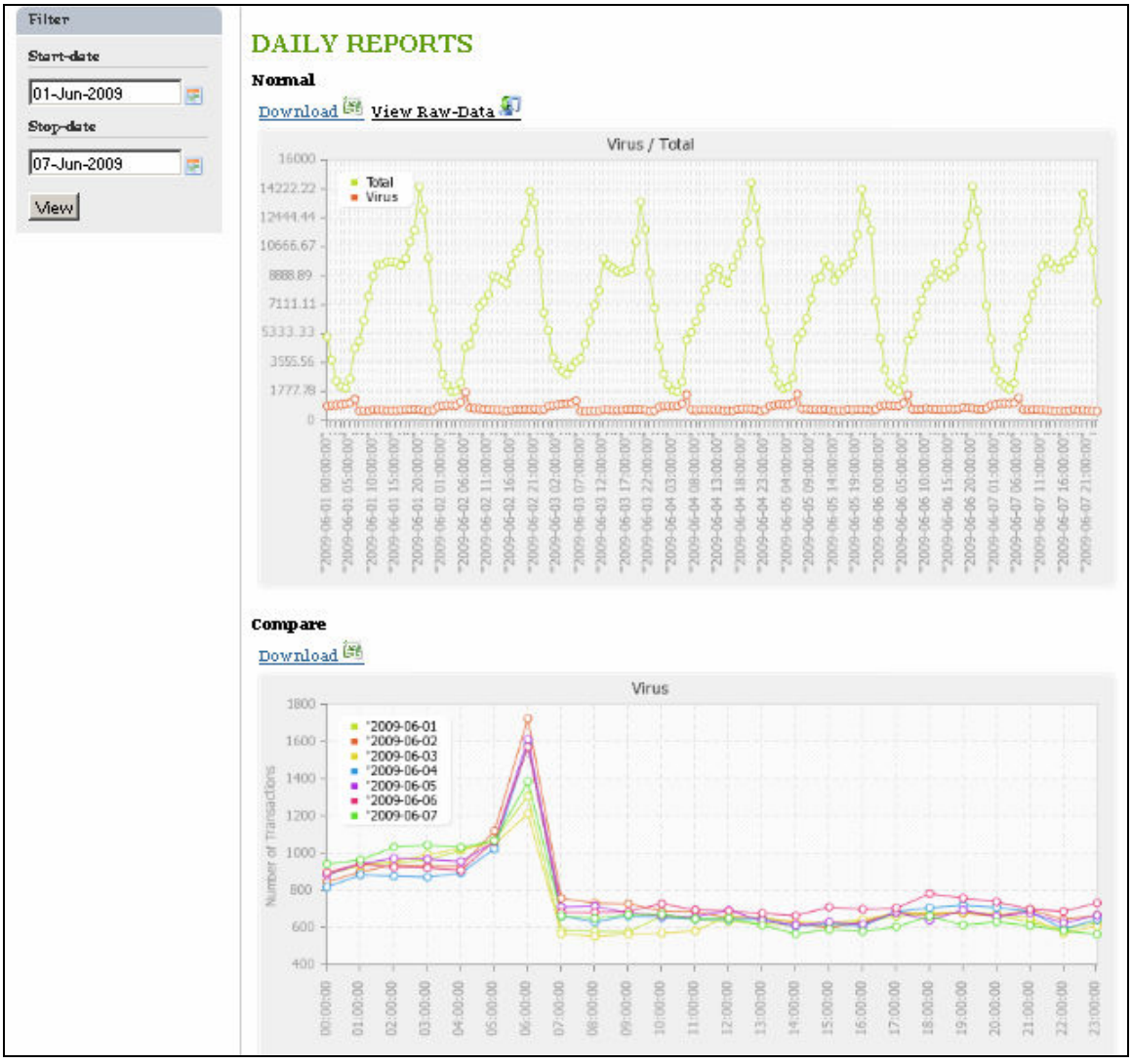
ภาพประกอบแสดงส่วนติดต่อผู้ใช้ในการเลือก Start-date



ภาพประกอบแสดงส่วนติดต่อผู้ใช้ในการเลือก Stop-date

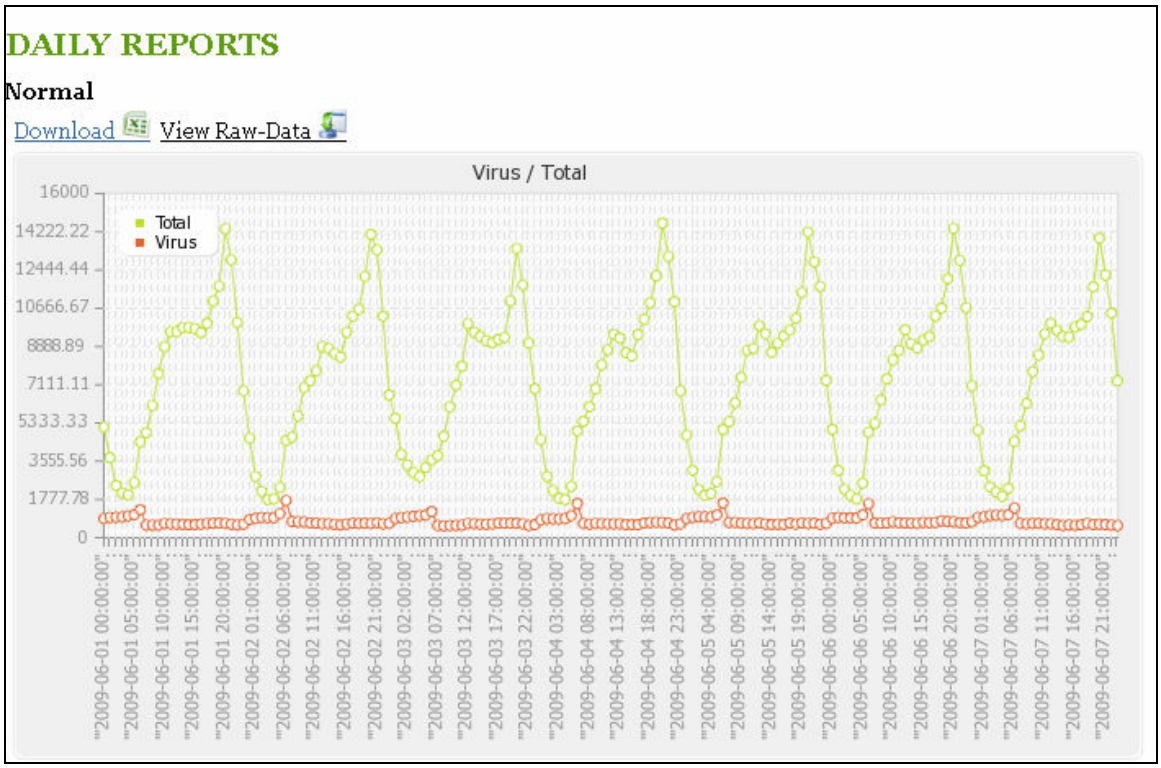
3. เมื่อผู้ใช้งานเลือกช่วงเวลาเรียบร้อยแล้วให้กดปุ่ม View ดังภาพประกอบแสดงปุ่ม View ที่ใช้สั่งให้โปรแกรมเริ่มประมวลผล เพื่อให้โปรแกรมประมวลผล และแสดงผลลัพธ์ตามภาพประกอบแสดงตัวอย่างผลลัพธ์

ภาพประกอบแสดงปุ่ม View ที่ใช้สั่งให้โปรแกรมเริ่มประมวลผล

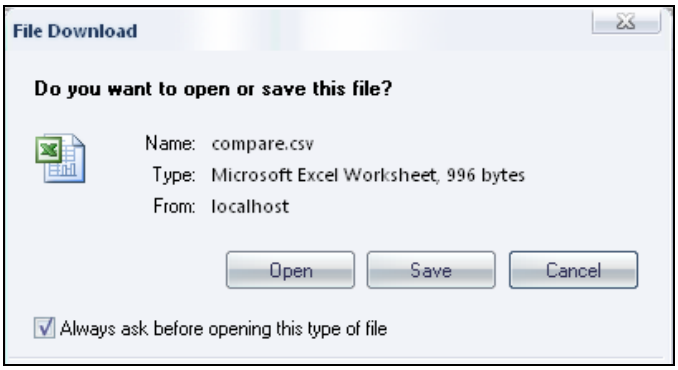


ภาพประกอบแสดงตัวอย่างผลลัพธ์เมื่อโปรแกรมประมวลผลเสร็จ

4. ผลลัพธ์ของโปรแกรมแบ่งออกเป็น 2 ส่วน โดยส่วนที่ 1 แสดงจำนวนข้อความมัลติมีเดียและจำนวนข้อความมัลติมีเดียที่มีไวรัสทั้งหมดเป็นกราฟแบบรายชั่วโมงตามช่วงเวลาที่ผู้ใช้งานเลือกประมวลผลตามภาพประกอบแสดงตัวอย่างกราฟแสดงจำนวนข้อความมัลติมีเดียและจำนวนข้อความมัลติมีเดียที่มีไวรัสทั้งหมดเป็นรายชั่วโมง และผู้ใช้งานสามารถดาวน์โหลดไฟล์ไมโครซอฟต์เอ็กเซล เพื่อดูสรุปตัวเลขเป็นรายชั่วโมงได้โดยการกดที่ Download หลังจากที่กดเบราว์เซอร์จะแสดงข้อความถามผู้ใช้งานว่าต้องการจะดาวน์โหลดหรือไม่ ดังภาพประกอบแสดงข้อความถามผู้ใช้งานว่าต้องการจะดาวน์โหลดหรือไม่

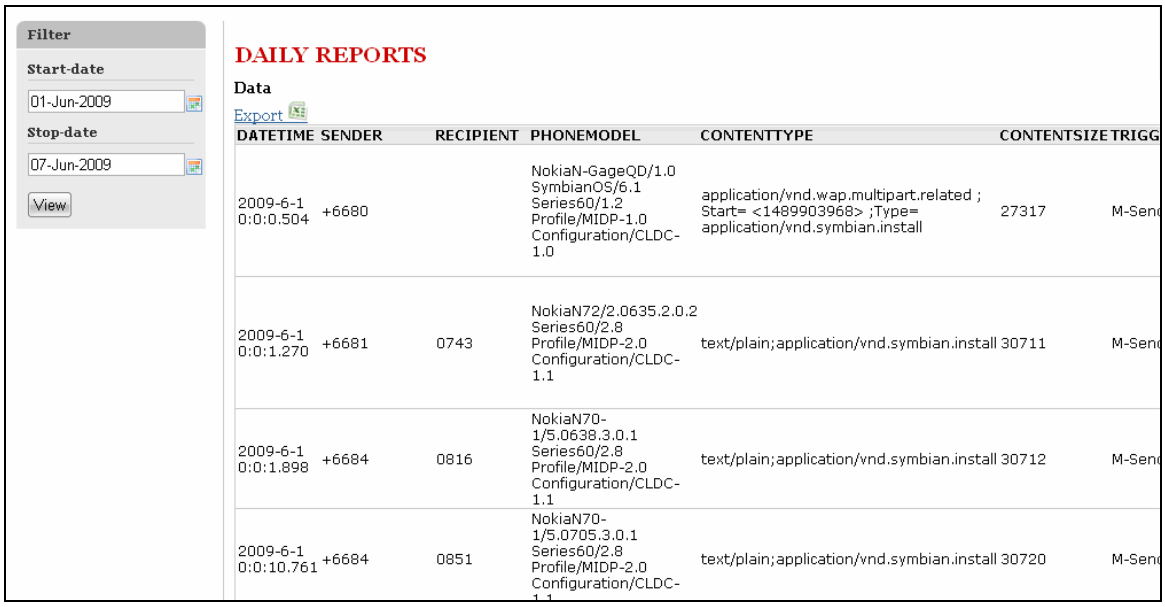


ภาพประกอบแสดงตัวอย่างกราฟแสดงจำนวนข้อความมัลติมีเดียและจำนวนข้อความมัลติมีเดียที่มีไวรัสทั้งหมดเป็นรายชั่วโมง



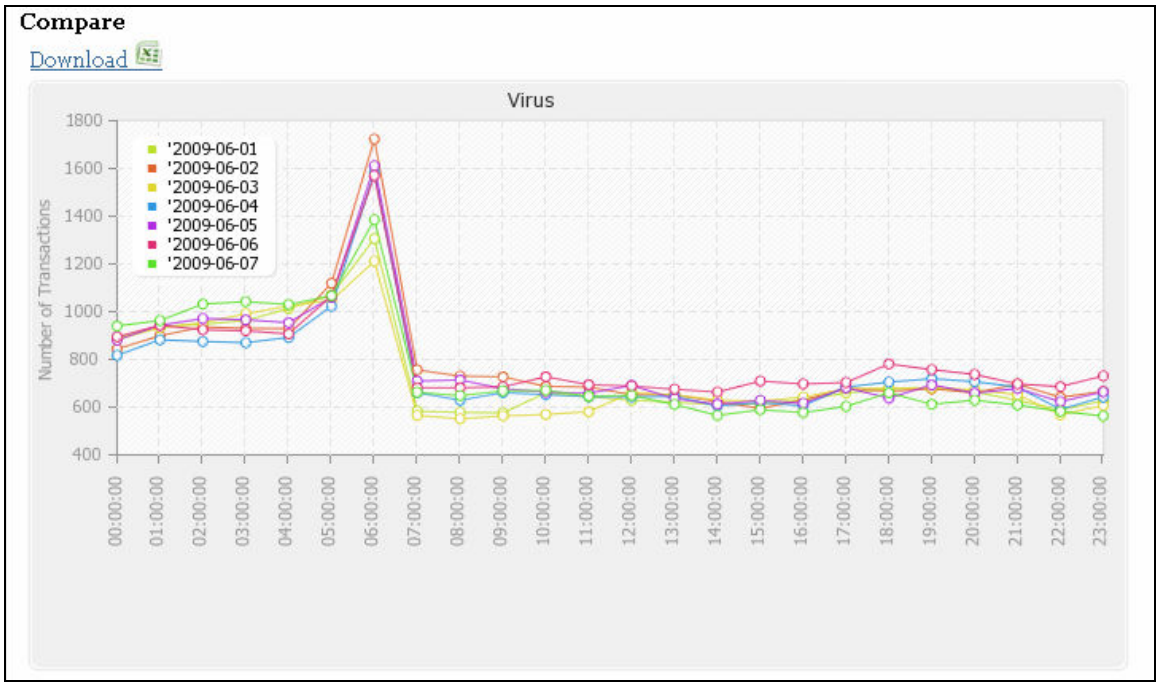
ภาพประกอบแสดงข้อความถามผู้ใช้งานว่าต้องการจะดาวน์โหลดหรือไม่

และผู้ใช้งานยังสามารถกดที่ View Raw-Data เพื่อเข้าไปดูรายละเอียดข้อความมัลติมีเดียที่มีไวรัส โดยหน้าเว็บของ View Raw-Data แสดงรายละเอียดดังภาพประกอบแสดงรายละเอียดข้อความมัลติมีเดียที่มีไวรัส และจะแสดงเพียง 20 รายการแรกเท่านั้น ซึ่งผู้ใช้งานสามารถดาวน์โหลดไฟล์ไมโครซอฟต์เอ็กเซล เพื่อดูรายการทั้งหมดได้ โดยกดที่ Export



ภาพประกอบแสดงรายละเอียดข้อความมัลติมีเดียที่มีไวรัส เมื่อกด View Raw-Data

4. ผลลัพธ์ของโปรแกรมส่วนที่ 2 แสดงจำนวนข้อความมัลติมีเดียที่มีไวรัสทั้งหมด โดยแสดงผลเปรียบเทียบเป็นรายชั่วโมงของแต่ละวัน ตามช่วงเวลาที่ผู้ใช้งานเลือกประมวลผลในรูปแบบกราฟ ตามภาพประกอบแสดงผลเปรียบเทียบเป็นรายชั่วโมงของแต่ละวัน



ภาพประกอบแสดงผลเปรียบเทียบเป็นรายชั่วโมงของแต่ละวัน

ประวัติย่อผู้วิจัย

ประวัติผู้วิจัย

ชื่อ - ชื่อสกุล	อัณศยา เกิดคล้าย
วันเดือนปีเกิด	22 พฤศจิกายน 2527
สถานที่เกิด	สุพรรณบุรี
สถานที่อยู่ปัจจุบัน	ร.1/42 ถ.เตชะวณิช แขวงบางซื่อ เขตบางซื่อ กรุงเทพมหานคร 10800
ตำแหน่งหน้าที่การงานปัจจุบัน	วิศวกร
สถานที่ทำงานปัจจุบัน	บริษัท แอดวานซ์ อินโฟร์ เซอร์วิส จำกัด (มหาชน)

ประวัติการศึกษา

พ.ศ. 2539	ชั้นประถมศึกษาปีที่ 6 จากโรงเรียนสุวรรณภูมิ สุพรรณบุรี
พ.ศ. 2542	ชั้นมัธยมศึกษาปีที่ 3 จากโรงเรียนสงวนหญิง สุพรรณบุรี
พ.ศ. 2545	ชั้นมัธยมศึกษาปีที่ 6 จากโรงเรียนสงวนหญิง สุพรรณบุรี
พ.ศ. 2550	วศ.บ. (วิศวกรรมคอมพิวเตอร์) จากมหาวิทยาพระจอมเกล้าธนบุรี