การจัดหมวดหมู่ของการเปลี่ยนแปลงวัฏจักรชีวีอีบนพื้นฐานของข้อมูลสาธารณะ

นายอรรถพล พวงพุ่ม

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2555

CLASSIFICATION OF CVE LIFECYCLE BASED ON PUBLIC INFORMATION

Mr. Atthapon Poungpum

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science Program in Computer Science

Department of Computer Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2012

หัวข้อวิทยานิพนธ์	การจัดหมวดหมู่ของการเปลี่ยนแปลงวัฏจักรซีวีอีบนพื้นฐานของข้อมูลสาธารณะ
โดย	นายอรรถพล พวงพุ่ม
สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	อาจารย์ ดร. ยรรยง เต็งอำนวย

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

…….. คณบดีคณะวิศวกรรมศาสตร์
(รองศาสตราจารย์ ดร.บุญสม เลิศหิรัญวงศ์)

คณะกรรมการสอบวิทยานิพนธ์

……. ประธานกรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.เกริก ภิรมย์โสภา)

……... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(อาจารย์ ดร.ยรรยง เต็งอำนวย)

……. กรรมการภายนอกมหาวิทยาลัย
(ผู้ช่วยศาสตราจารย์ ดร.ชวลิต ศรีสถาพรพัฒน์)

อรรถพล พวงพุ่ม : การจัดหมวดหมู่ของการเปลี่ยนแปลงวัฏจักรซีวีอีบนพื้นฐานของข้อมูลสาธารณะ. (Classification of CVE Lifecycle Based on Public Information) อ.ที่ปรึกษาวิทยานิพนธ์หลัก : อ.ดร.ยรรยง เต็งอำนวย, 32 หน้า.

หน้าที่ในการจัดการปิดจุดอ่อนโดยผู้ดูแลระบบเป็นงานที่ใช้เวลาและทรัพยากรมาก เนื่องจากจำนวนจุดอ่อนมีจำนวนมากและจุดอ่อนมีความแตกต่างกันทั้งในความรุนแรง ช่วงเวลาของการค้นพบ การพัฒนาการของวงจรชีพที่ต่างกัน งานวิจัยนี้ต้องการกำหนดกรอบการทำงานเพื่อลำดับความสำคัญของจุดอ่อน ศึกษาความเปลี่ยนแปลงของปริมาณข้อมูลสาธารณะที่เกี่ยวข้องกับจุดอ่อนแต่ละตัว เพื่อพิจารณาลักษณะของการเปลี่ยนแปลงของวงจรชีพจุดอ่อนโดยสามารถแบ่งสถานะของวงจรชีพของจุดอ่อนออกเป็น 6 แบบและแยกจุดอ่อนเป็นกลุ่มที่มีวงจรชีพที่มีความเสถียรและกลุ่มที่ยังมีการเคลื่อนไหว เพื่อแนะนำให้ผู้ดูแลระบบใส่ใจกับกลุ่มหลังที่ยังมีความสำคัญ

ภาควิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่อนิสิต .....

สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์ ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก .....

ปีการศึกษา 2555

5271470521 : MAJOR COMPUTER SCIENCE
KEYWORDS : COMPUTER SECURITY / VULNERABILITY LIFECYCLE / WEB DATA MINING / ONTOLOGY.

ATTHAPON POUNGPUM : CLASSIFICATION OF CVE LIFECYCLE BASED ON PUBLIC INFORMATION. ADVISOR : YUNYONG TENG-AMNUAY, Ph.D., 32 pp.

System administrators spend much time and resource in protecting system from vulnerabilities. Vulnerabilities differ in violence, discovery time, and development of lifecycle. This research defines a framework for vulnerability prioritization by tracking the variation of amount of public information for each specific vulnerability. The lifecycles of vulnerabilities are classified into six types with stable and recurring lifecycles. The latter are the ones to be monitored by system administrators.

Department : Computer Engineering Student's Signature .....
Field of Study : Computer Science Advisor's Signature .....
Academic Year : 2012 .....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปด้วยดีด้วยความช่วยเหลือจาก อาจารย์ ดร. ยรรยง เต็งอำนวย อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ ได้มอบความรู้ คำแนะนำ ตรวจทานเพื่อแก้ไขส่วนบกพร่องของงานวิจัย ตลอดจนการตรวจทานแก้ไขวิทยานิพนธ์ให้มีความสมบูรณ์ นางสาว รัศมีทิพย์ วิตา ผู้ที่ให้ความช่วยเหลือในการทำงานวิจัย ทั้งในด้านของงานวิจัยพื้นฐานที่นำมาให้ผู้เขียนได้ใช้เป็นแนวทางในการวิจัย คำปรึกษา ข้อมูลสำหรับทำการวิจัย นอกจากนี้ผู้เขียนยังได้รับความกรุณาจากผู้ช่วยศาตราจารย์ ดร.เกริก ภิรมย์โสภา ประธานกรรมการสอบวิทยานิพนธ์ รวมถึงผู้ช่วยศาตราจารย์ ดร.ชวลิต ศรีสถาพรพัฒน์ กรรมการสอบวิทยานิพนธ์ผู้ทรงคุณวุฒิจากภายนอก ที่ได้ให้คำแนะนำ รวมทั้งข้อเสนอแนะต่างๆ ที่เป็นประโยชน์เพื่อนำมาใช้ปรับปรุงวิทยานิพนธ์ให้เกิดความสมบูรณ์มากยิ่งขึ้น

ผู้เขียนขอกราบขอบพระคุณบิดา มารดา ที่ได้สนับสนุนด้านทุนทรัพย์ในการศึกษาและคอยเป็นกำลังใจให้เสมอมา รวมทั้งนางสาวประภาวดี เอกวงศ์ นายกิตติศักดิ์ สะอาดเอี่ยม และนายธนชนม์ ชีพบริสุทธิกุล เพื่อนร่วมห้องปฏิบัติการวิศวกรรมระบบสารสนเทศของข้าพเจ้าที่คอยให้กำลังใจและเป็นแรงบันดาลใจให้ข้าพเจ้าเสมอมา

สารบัญ

หน้า

หน้า

สารบัญตาราง

หน้า

สารบัญภาพ

หน้า

บทที่ 1

บทนำ

1.1. ความเป็นมาและความสำคัญของปัญหา

จุดอ่อนในซอฟต์แวร์ได้ถูกรวบรวมไว้เพื่อให้เป็นมาตรฐานในการประสานงานกันในระหว่างผู้ทำงานในการรักษาความปลอดภัยด้านคอมพิวเตอร์ ในชื่อ ซีวีอี (CVE) [1] แต่เนื่องจากจุดอ่อนเหล่านี้มีจำนวนมาก [2] ทำให้ผู้ดูแลระบบทำงานได้ลำบากในการจัดลำดับความสำคัญในการดูแลรักษาความปลอดภัยจากจุดอ่อนเหล่านี้

งานวิจัยเกี่ยวกับข้อมูลจุดอ่อน ได้มีการจัดกลุ่มของจุดอ่อน ตามประเภทของการเกิด หรือประเภทของผลกระทบเพื่อให้สามารถจัดการป้องกันจุดอ่อนได้ตามกลุ่มที่ระบบนั้น ๆ ให้ความสนใจเป็นพิเศษ ต่อมา ได้มีการวิเคราะห์ถึงความสัมพันธ์ระหว่างจำนวนของการโจมตี และวงจรชีพของจุดอ่อน ทำให้พบว่าวงจรชีพมีส่วนกำหนดการกระจายและจำนวนของการโจมตี และเนื่องจากจุดอ่อนแต่ละตัว มีวงจรชีพที่ต่างกัน ทำให้ลักษณะของจำนวนของการโจมตีต่างกันขึ้นอยู่กับพัฒนาการของวงจรชีพ

ในงานวิจัยนี้ได้นำเสนอวิธีการติดตามการเปลี่ยนแปลงของวงจรชีพของจุดอ่อนโดยใช้ข้อมูลสาธารณะจากการค้นคืนผ่านเว็บไซต์ เพื่อจัดลำดับความสำคัญของจุดอ่อนตามสถานะของวงจรชีพเพื่อช่วยผู้ดูแลระบบในการตัดสินใจติดตามหรือเพิกเฉยจุดอ่อนเหล่านั้น

1.2. วัตถุประสงค์ของการวิจัย

1.2.1 เพื่อจัดหมวดหมู่ของการเปลี่ยนแปลงวัฏจักรซีวีอีในการให้ลำดับความสำคัญของซีวีอี

1.2.2 เพื่อนำเสนอวิธีการติดตามหรือเพิกเฉยซีวีอี โดยอาศัยข้อมูลสาธารณะในการพิจารณา

1.3. ขอบเขตของการวิจัย

1.3.1. พัฒนาระบบด้วยภาษาจาวา

1.3.2. ทำการค้นหาข้อมูลสาธารณะด้วย Google Search API

1.3.3. การทำเหมืองข้อมูลโดยใช้เทคนิคการจำแนกประเภทข้อมูลแบบเบย์ (Bayesian Classification)

1.3.4. ประมวลผลการทำเหมืองข้อมูลโดยใช้ Weka Tools

1.4. คำจำกัดความที่ใช้ในการวิจัย

1.4.1. วีแอลโอ (Vulnerability Lifecycle Ontology) คือ ออนโทโลยีของวงจรชีพของจุดอ่อน

1.4.2. เอ็นวีดี (National Vulnerability Database) คือ ฐานข้อมูลจุดอ่อนที่เผยแพร่โดยกระทรวงกลาโหมของสหรัฐอเมริกา

1.5. ประโยชน์ที่คาดว่าจะได้รับ

งานวิจัยนี้นำเสนอการจัดหมวดหมู่ของซีวีอีและความสำคัญเพื่อช่วยผู้ดูแลระบบในการติดตามหรือเพิกเฉยต่อซีวีอีที่เกิดขึ้นโดยใช้ข้อมูลการเปลี่ยนแปลงของซีวีอีที่ได้รับมาจากเว็บไซต์ของเอ็นวีดี เป็นแนวทางในการช่วยเหลือผู้ดูแลระบบในการให้ความสำคัญในการจัดการแก้ไขปัญหาที่เกิดจากซีวีอีที่ต้องทำการแก้ไขหรือสามารถเพิกเฉยได้

1.6. **วิธีดำเนินการวิจัย**

1.6.1. ศึกษาและทดสอบแนวความคิดและทฤษฎีของงานวิจัยที่เกี่ยวข้อง

1.6.2. ทดลองแนวทางและวิธีการของงานวิจัยต่างๆ เกี่ยวกับข้อมูลจุดอ่อนและการเปลี่ยนแปลงข้อมูลจากสาธารณะ

1.6.3. ทดสอบ วิเคราะห์ ปรับปรุงแก้ไขงานวิจัย ให้อยู่ในแนวทางที่ต้องการ

1.6.4. นำเสนอผลลัพธ์ของงานวิจัยที่ได้ในรูปแบบต่างๆ เช่น การแสดงผลในรูปแบบกราฟ

1.6.5. สรุปผลงานวิจัย

1.7. **ลำดับขั้นตอนในการเสนอผลการวิจัย**

วิทยานิพนธ์นี้แบ่งเนื้อหาทั้งหมด 5 บท โดยแต่ละบทประกอบไปด้วยเนื้อหาดังต่อไปนี้

บทที่ 1 นำเสนอ ความเป็นมาและความสำคัญของปัญหา, วัตถุประสงค์ของการวิจัย, ขอบเขตของการวิจัย, คำจำกัดความที่ใช้ในการวิจัย, ประโยชน์ที่คาดว่าจะได้รับ และวิธีดำเนินการวิจัย

บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง เช่น ทฤษฎีที่เกี่ยวข้อง เช่น ซีวีอี, การทำเหมืองข้อมูลเว็บ เอกสารและงานวิจัยที่เกี่ยวข้อง เช่น Vulnerability Relevancy Framework

บทที่ 3 ประกอบด้วย ภาพรวมและขั้นตอนการดำเนินงานวิจัย แหล่งที่มาของข้อมูล การเก็บรวบรวบข้อมูล การนอมัลไลซ์ข้อมูลเพื่อการสรุปผล วิธีการประเมินผลการทดลอง

บทที่ 4 นำเสนอ ตัวอย่างและผลลัพธ์การทดลอง จากการเก็บข้อมูลการเปลี่ยนแปลงของคะแนนซีวีอีที่ได้นำมาวิเคราะห์ออกเป็นกลุ่มของซีวีอีที่ต้องติดตามและเพิกเฉยได้

บทที่ 5 สรุปผลการวิจัย

1.8. ผลงานที่ตีพิมพ์จากวิทยานิพนธ์

ส่วนหนึ่งของวิทยานิพนธ์ฉบับนี้ได้รับการตีพิมพ์และนำเสนอบทความทางวิชาการชื่อ Classification of CVE Lifecycle Based on Public Information โดย Atthapon Pounpum, Ratsameetip Wita และ Yunyong Teng-amnuay ในงานประชุมวิชาการ The 9th National Conference on Computing and Information Technology (NCCIT 2013) ซึ่งจัดขึ้นในวันที่ 9-10 พฤษภาคม 2556 ณ. กรุงเทพมหานคร ประเทศไทย

บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

2.1. ทฤษฎีที่เกี่ยวข้อง

2.1.1. CVE

ซีวีอี คือมาตรฐานสำหรับการกำหนดชื่อจุดอ่อนและคำอธิบายซึ่งประกอบไปด้วยข้อมูลทั่วไป ผลกระทบ เวอร์ชั่นของซอฟท์แวร์ วิธีการแก้ไข เพื่ออ้างอิงถึงจุดอ่อนเดียวกันและง่ายต่อการอ้างอิง [1] ตัวอย่างดังภาพที่ 1

Name	CVE-2011-1302
Summary	Heap-based buffer overflow in the GPU process in Google Chrome before 10.0.648.205 allows remote attackers to execute arbitrary code via unknown vectors.
Published	4/5/2011

ภาพที่ 2.1 ลักษณะตัวอย่างของซีวีอี

แหล่งข้อมูลหมายเลขซีวีอีถูกรวบรวมไว้ในฐานข้อมูลเอ็นวีดี (NVD: National Vulnerability Database) ซึ่งพัฒนาโดยสถาบันมาตรฐานและเทคโนโลยีแห่งประเทศสหรัฐอเมริกาหรือเอ็นไอเอสที (NIST: National Institute of Standard and Technology) [3]

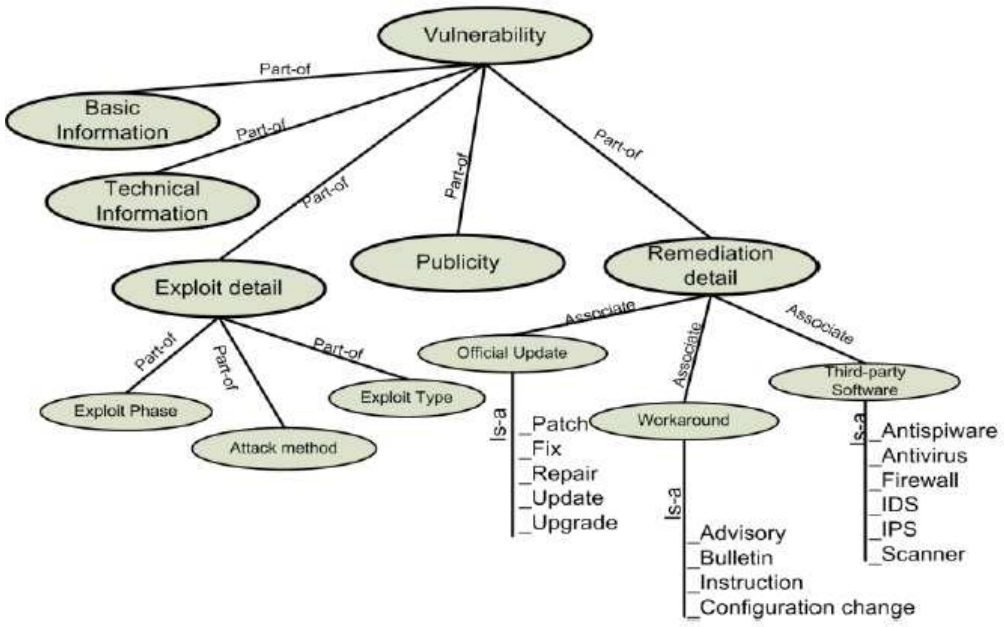
2.1.2. การทำเหมืองข้อมูลโดยการจำแนกประเภทข้อมูล

กระบวนการทำเหมืองข้อมูลโดยการจำแนกประเภทข้อมูลเป็นกระบวนการสร้างแบบจำลองจัดการข้อมูลให้อยู่ในกลุ่มที่กำหนด โดยนำข้อมูลส่วนหนึ่งสอนให้ระบบเรียนรู้ (Training Data) เพื่อจำแนกข้อมูลออกเป็นกลุ่ม ผลลัพธ์ที่ได้จากการเรียนรู้คือ แบบจำลองจำแนกประเภทข้อมูล (Classifier Model) และนำข้อมูลส่วนที่เหลือจากการสอนระบบมาใช้เป็นข้อมูลทดสอบ (Testing Data) ซึ่งกลุ่มที่แท้จริงของข้อมูลที่ใช้ทดสอบนี้นำมาเปรียบเทียบกับกลุ่มที่หามาได้จากแบบจำลองเพื่อทดสอบความถูกต้องและปรับปรุงแบบจำลองจนได้ค่าความถูกต้องในระดับที่ต้องการ เพื่อนำแบบจำลองที่ได้มาใช้ในการทำนายกลุ่มของข้อมูลใหม่ [4]

2.2. เอกสารและงานวิจัยที่เกี่ยวข้อง

2.2.1. Vulnerability Relevancy Framework

งานวิจัยนี้กำหนดกรอบการทำงานสำหรับการจัดลำดับความสำคัญของจุดอ่อนจากความเกี่ยวข้องกับข้อมูลออนไลน์ที่รวบรวมได้และใช้ความสัมพันธ์ระหว่างวงจรชีพและลักษณะของข้อมูลที่เกี่ยวข้องกับจุดอ่อนโดยสร้างเป็นวีแอลโอ (VLO หรือ Vulnerability Lifecycle Ontology) [5] ซึ่งวีแอลโอถูกสร้างจากข้อมูลความปลอดภัยของระบบคอมพิวเตอร์และข้อมูลพื้นฐานของจุดอ่อน สกัดข้อมูลจากเว็บไซต์ความปลอดภัย เพื่อใช้ในการอธิบาย Basic Information, Technical Detail, Exploit Detail, Publicity Detail และ Remediation Detail ของออนโทโลจีตามภาพที่ 2.2



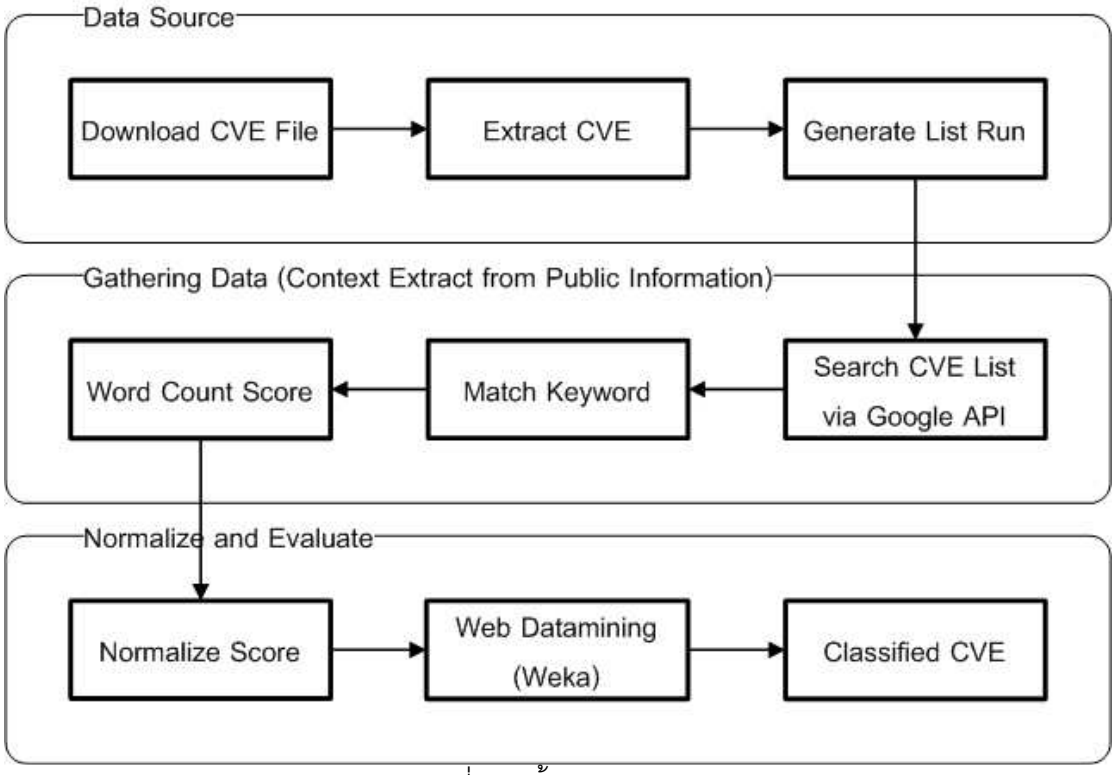
ภาพที่ 2.2 ลักษณะของวีแอลโอ

ในงานวิจัยนี้ใช้การคำนวณของ Subcontext ของหัวข้อที่สนใจที่ปรากฏในเอกสารเพื่อนำเสนอออกมาให้รูปของเวคเตอร์ของ Subcontext เพื่อเปรียบเทียบกับ เวคเตอร์ของข้อมูลเดิมโดยใช้วีธีเอสวีเอ็ม (SVM: Support Vector Machine) ในการแสดงผลของข้อมูลในวีแอลโอ [6]

บทที่ 3

วิธีดำเนินการวิจัย

งานวิจัยนี้แบ่งขั้นตอนการทำงานออกเป็น 3 ส่วน ส่วนแรกเป็นการจัดเตรียมข้อมูลซีวีอีจากแหล่งข้อมูลที่ดาวน์โหลดมาจากเว็บไซต์เอ็นวีดี ส่วนที่สองคือการเก็บข้อมูลจากซีวีอีที่เราสนใจโดยใช้ Vulnerablity Relavancy Framework ในการให้คะแนนผลการค้นหา ส่วนสุดท้ายคือการนำข้อมูลที่ได้จากการเก็บข้อมูลมาปรับปรุงเพื่อให้ง่ายต่อการวิเคราะห์ผลการเปลี่ยนแปลงข้อมูล และการประเมินผลการทดลองด้วยกระบวนการของเหมืองข้อมูลโดยการจำแนกประเภทข้อมูล เพื่อประเมินผลการทดลองตามวงจรชีพที่กำหนดดังภาพที่ 3.1

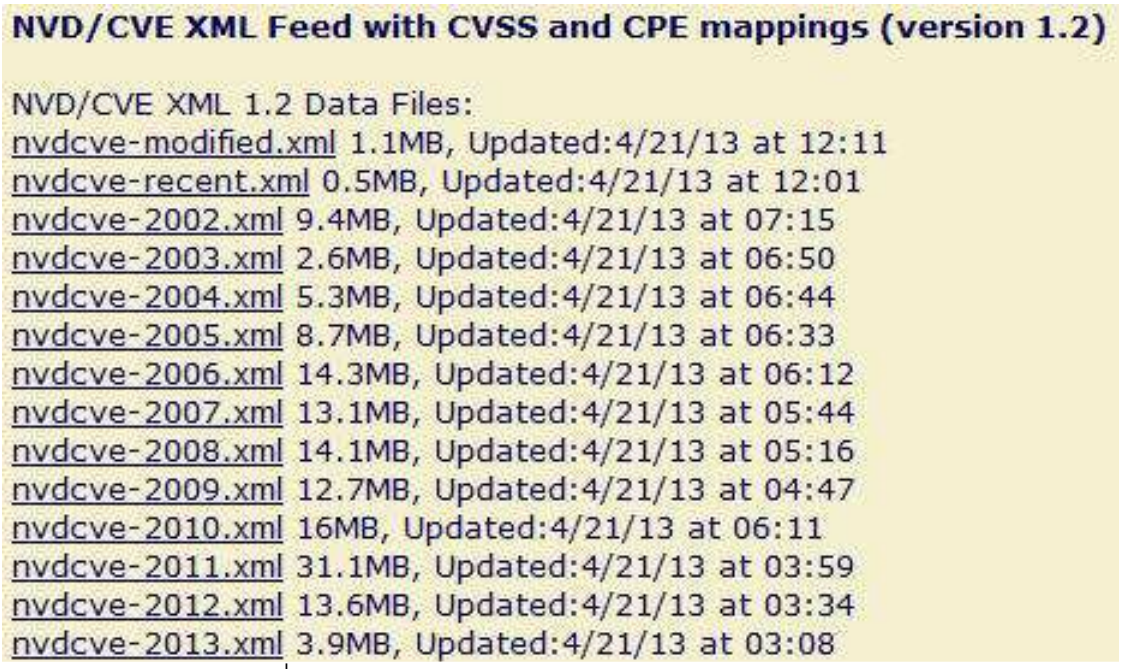


ภาพที่ 3.1 ขั้นตอนการวิจัย

โดยงานวิจัยนี้การนำวิธีการเก็บข้อมูลโดยฐานข้อมูลคะแนนวีอาร์ (VR-Score) [6] ในการเก็บรวบรวมข้อมูลซีวีอีจากข้อมูลสาธารณะ เพื่อใช้พิจารณาผลการติดตามการเปลี่ยนแปลงวงจรชีพของซีวีอี

3.1. แหล่งที่มาของข้อมูล

งานวิจัยนี้ใช้ข้อมูลของซีวีอีที่ได้เผยแพร่จากเว็บไซต์เอ็นวีดี [3] เพื่อศึกษาการเปลี่ยนแปลงที่เกิดขึ้นในแต่ละวัน ข้อมูลซีวีอีที่ใช้เป็นข้อมูลตั้งต้นสำหรับการทดลองแบ่งออกเป็น 2 กลุ่มคือ กลุ่มของปัจจุบัน (nvdcve-2013.xml) ที่จำแนกตามปีที่เกิด และกลุ่มของข้อมูลทั้งหมดที่เปลี่ยนแปลงของปีก่อนหน้า (nvdcve-modified.xml) การแยกกลุ่มข้อมูลตั้งต้นออกเป็น 2 กลุ่ม เพื่อศึกษาถึงการเปลี่ยนแปลงระหว่างซีวีอีที่เกิดขึ้นใหม่ กับซีวีอีเก่าที่กลับมาได้รับความสนใจอีกครั้งตามภาพที่ 3.2

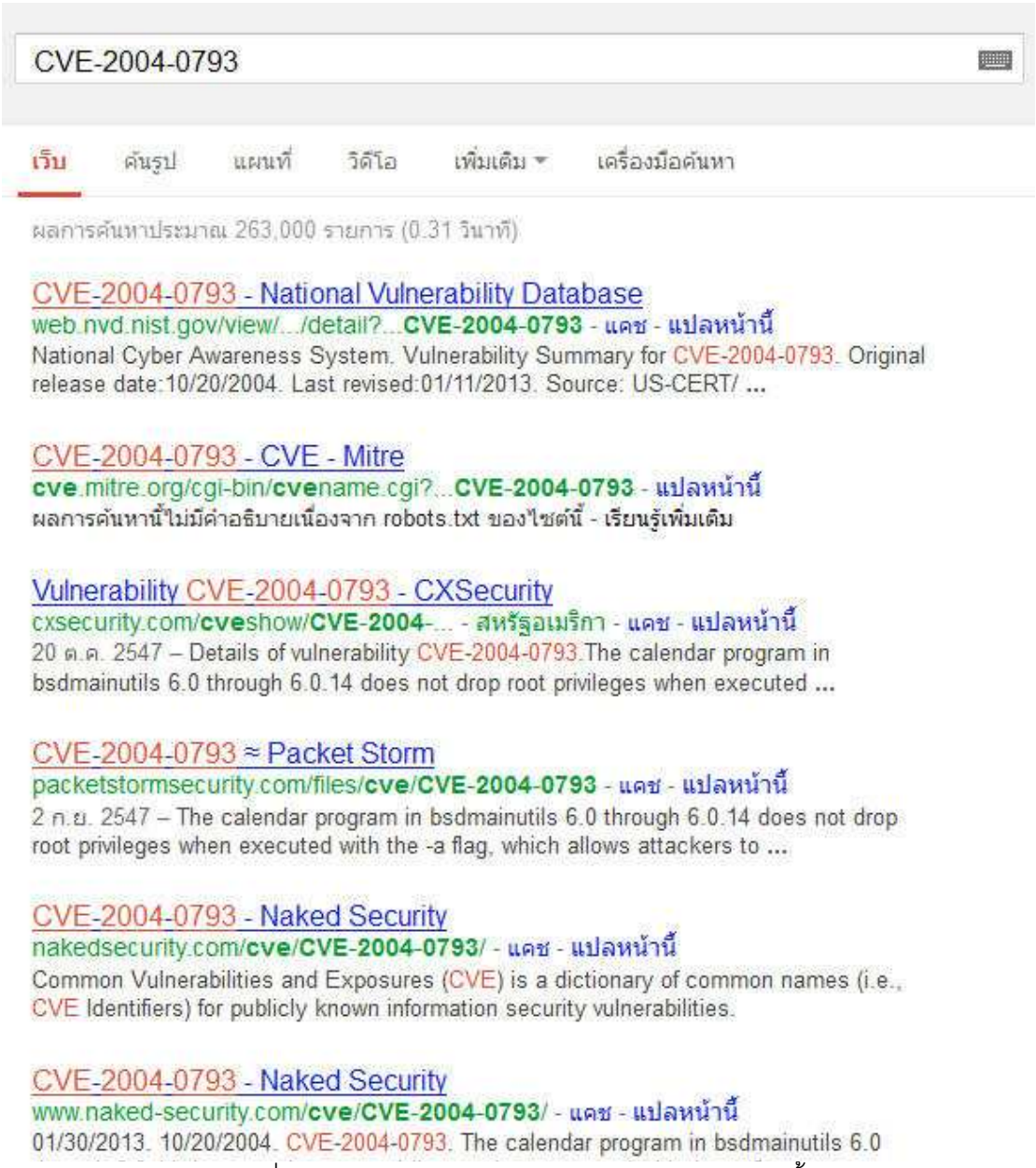


ภาพที่ 3.2 ข้อมูลซีวีอีสำหรับดาวน์โหลดจากเว็บไซต์เอ็นวีดี

การเก็บข้อมูลของซีวีอีในปีปัจจุบันจาก nvdcve-2013.xml ซึ่งเป็นข้อมูลของซีวีอีของปี ค.ศ. 2013 เพียงปีเดียว ส่วนข้อมูลของซีวีอีเก่านำมาจาก nvdcve-modified.xml ซึ่งเป็นข้อมูลของซีวีอีทั้งหมดทุกปีที่มีการแก้ไข มาตัดข้อมูลของซีวีอีของปี ค.ศ. 2013 ออกแล้วจึงนำมาเป็นคำค้นในการเก็บข้อมูลจากข่าวสาธารณะ โดยในข้อมูลที่ดาวน์โหลดจากเว็บไซต์เอ็นวีดีจะมีลักษณะตามภาพที่ 3.3

ภาพที่ 3.3 ตัวอย่างข้อมูลซีวีอีจากเว็บไซต์เอ็นวีดี

อีกแหล่งข้อมูลของงานวิจัยนี้คือข้อมูลสาธารณะ (Public Information) เป็นข้อมูลในอินเทอร์เน็ตในหน้าเว็บต่างๆ ซึ่งในการทดลองนี้คือข้อมูลที่มีความเกี่ยวข้องกับซีวีอีที่เราสนใจ ผ่านทาง Google Search Service ซึ่งเป็นบริการค้นหาข้อมูลทางอินเตอร์เน็ตที่ได้รับความนิยมสูงสุด [8] โดยให้ความสนใจผลการค้นหา 10 อันดับแรกเพื่อนำมาใช้ในการทดลอง ซึ่งลักษณะข้อมูลของการค้นหาผ่านทาง Google จะมีลักษณะดังภาพที่ 3.4



ภาพที่ 3.4 ตัวอย่างการค้นหาข้อมูลผ่านทางเว็บไซต์กูเกิ้ล

3.2. การเก็บรวบรวมข้อมูล

ในงานวิจัยนี้ได้ทำการเก็บข้อมูลของซีวีอีจากแหล่งสาธารณะทุกวัน โดยเริ่มตั้งแต่วันที่ 25 ธันวาคม 2012 เก็บสะสมข้อมูลของซีวีอีที่ได้จาก nvdcve-modified.xml ซึ่งทำให้เราได้ซีวีอีทั้งหมดที่มีการเปลี่ยนแปลงในแต่ละวันโดยเก็บสะสมจนถึงวันที่ 20 กุมภาพันธ์ 2013

ข้อมูลหมายเลขซีวีอีที่เก็บรวบรวมมาได้จะถูกนำมาสร้างเป็นรายการของซีวีอีที่ใช้ในการกรองหาความเปลี่ยนแปลง โดยใช้ซีวีอีเหล่านั้นเป็นคำค้นในการค้นหาข้อมูลผ่าน Google Search API [9]

ตารางที่ 3.1 ตัวอย่างผลการค้นหาข้อมูลจากกูเกิ้ล

CVENAME	DATE	WORD COUNTS
CVE-2003-4696	2013-02-12 04:23:28	455
CVE-2003-4696	2013-02-10 15:14:25	531
CVE-2003-4696	2013-02-08 08:11:35	530
CVE-2003-4696	2013-02-07 07:33:58	540
CVE-2003-4696	2013-02-06 08:33:32	548
CVE-2003-4696	2013-02-05 03:46:36	626
CVE-2003-4696	2013-02-04 09:26:04	626
CVE-2003-4696	2013-02-03 06:41:30	634
CVE-2003-4696	2013-02-02 10:51:08	815
CVE-2003-4696	2013-02-01 01:23:05	850
CVE-2003-4696	2013-01-31 08:19:39	814
CVE-2003-4696	2013-01-30 03:24:02	414

ผลลัพธ์ที่ได้จากการค้นหาข้อมูลผ่านทาง Google Search API มีลักษณะเป็นคำ ซึ่งเรานำผลดังกล่าวมาทำการเปรียบเทียบกับคำสำคัญที่เกี่ยวข้องกับความปลอดภัย (Security Keywords) โดยทำการนับเฉพาะคำสำคัญที่เกี่ยวข้องกับความปลอดภัย เก็บเป็นข้อมูลของการค้นหาซีวีอีในแต่ละวันดังตารางที่ 3.1

ค่าในคอลัมน์ word counts จากตารางที่ 3.1 สามารถบ่งบอกถึงความสนใจของสาธารณะที่มีต่อซีวีอี ในช่วงเวลาที่ทำการค้นคืนข้อมูล ค่า word counts ที่มากขึ้นแสดงถึงซีวีอีกำลังเป็นปัญหามีการกล่าวถึงบนเว็บไซต์มากขึ้น

3.3. การนอมัลไลซ์ข้อมูล

ข้อมูลที่รวบรวมมาได้ ยังไม่อาจบอกได้ว่าการเปลี่ยนแปลงของข้อมูลที่ค้นหาผ่านทาง Google Search API เพียงเล็กน้อยหมายถึงการเปลี่ยนแปลงของหน้าเว็บมากน้อยเพียงใด ดังนั้นจึงทำการนอมัลไลซ์ตามแนวทางของ Wita [6] เพื่อใช้ในการคำนวณเปรียบเทียบค่าตามสมการที่ 1 ดังนี้

$$\omega_{hit} = \begin{cases} \frac{\log_{10}(\text{result})}{\log_{10}(1000)}, if\ \text{result} < 1000 \\ 1, if\ \text{result} > 1000 \end{cases} \quad (1)$$

นำผลการคำนวณด้วยสมการมาทำการนอมัลไลส์ข้อมูลเพื่อความสะดวกในการเปรียบเทียบลำดับการเปลี่ยนแปลงของข้อมูล และเตรียมข้อมูลสำหรับการทำเหมืองข้อมูล โดยค่าน้ำหนักของผลการค้นหา (ω_hit) มาจากลอการึทึ่มของผลการค้นหาเทียบกับจำนวนสูงสุดของผลที่ได้ ซึ่ง Google Search API มีค่าสูงสุดของผลการค้นหาหนึ่งครั้งเท่ากับ 1000 [9]

ในตารางที่ 3.2 จะเห็นว่า การนำผลการเก็บข้อมูลมาคำนวณผ่านสมการที่ 1 ตามคอลัมน์ ω_hit มีค่าเป็นทศนิยม ซึ่งในการทำเหมืองข้อมูลของการทดลองเพื่อตรวจผล เราให้ความสนใจในลักษณะการเปลี่ยนแปลงของข้อมูลในลักษณะของกราฟและการใช้ข้อมูลที่มีหน่วยเป็นทศนิยมในการสร้างแบบจำลองสำหรับการฝึกหมายถึงมีจำนวนความเป็นน่าจะเป็นของข้อมูลที่เป็นไปได้ถึง 1000^X ซึ่งค่า X เป็นจำนวนของวันที่เราใช้ในการประเมินผล ดังนั้นเราได้ทำการนำข้อมูลมาทำการปรับปรุงให้อยู่ในรูปแบบของลำดับ ซึ่งจำนวนสูงสุดที่เป็นไปได้จะมีค่าเป็น X^X ซึ่งมีค่าน้อยกว่า เพื่อใช้ในการเตรียมข้อมูลฝึกสอนในการทำเหมืองข้อมูลต่อไป

ตารางที่ 3.2 ตัวอย่างผลการคำนวณผ่านสมการ

cvename	last_update	log10(word count)	ω_hit	Normalized
CVE-2004-0793	2013-01-12	2.02	0.672	3
CVE-2004-0793	2013-01-13	2.02	0.672	3
CVE-2004-0793	2013-01-14	2.01	0.670	1
CVE-2004-0793	2013-01-15	2.05	0.682	6
CVE-2004-0793	2013-01-16	2.05	0.684	7
CVE-2004-0793	2013-01-17	2.01	0.670	1
CVE-2004-0793	2013-01-19	2.02	0.672	3
CVE-2004-0793	2013-01-21	2.03	0.676	5
CVE-2004-0793	2013-01-23	2.03	0.675	4
CVE-2004-0793	2013-01-26	2.01	0.671	2
CVE-2004-0793	2013-01-27	2.01	0.670	1
CVE-2004-0793	2013-01-29	2.01	0.671	2
CVE-2004-0793	2013-01-30	2.01	0.671	2
CVE-2004-0793	2013-01-31	2.03	0.676	5
CVE-2004-0793	2013-02-07	2.08	0.692	9
CVE-2004-0793	2013-02-10	2.06	0.688	8

3.4. การประเมินผล

ผลจากการปรับปรุงข้อมูลตามตารางที่ 3.2 เรานำข้อมูลนอมัลไลซ์ ไปผ่านกระบวนการของการทำเหมืองข้อมูลโดยใช้โปรแกรม Weka [10] โดยสร้างแบบจำลองของลักษณะการเปลี่ยนแปลงของวงจรชีพซีวีอีจากรูปแบบการเปลี่ยนแปลงของข้อมูลตามตารางที่ 3.1 แยกเป็นผลของซีวีอีผ่านการคำนวณจากสูตรในแต่ละวัน แบบจำลองข้อมูลที่สร้างมีลักษณะดังภาพที่ 3.5

```
@Relation change_in_cve_7_day

@attribute cve string
@attribute day1 NUMERIC
@attribute day2 NUMERIC
@attribute day3 NUMERIC
@attribute day4 NUMERIC
@attribute day5 NUMERIC
@attribute day6 NUMERIC
@attribute day7 NUMERIC
@attribute class{increase,decrease,swing,full_cycle,stable,recurrent}
@data
CVE-2004-0793,5,5,6,3,3,6,5,recurrent
CVE-2006-0175,2,2,2,1,2,2,2,stable
CVE-2006-0218,4,4,4,3,2,2,2,decrease
CVE-2006-0434,2,3,4,4,3,3,3,full_cycle
CVE-2006-0487,6,5,2,1,3,3,4,recurrent
CVE-2006-0533,6,5,4,4,2,2,2,decrease
CVE-2006-0585,4,5,4,4,2,2,2,decrease
CVE-2006-0633,5,5,5,6,5,5,5,stable
CVE-2006-0697,6,6,6,4,4,3,4,decrease
CVE-2006-0707,1,2,3,3,2,2,3,stable
CVE-2006-0857,13,13,5,2,7,7,8,stable
CVE-2006-1423,8,7,6,6,6,6,6,stable
CVE-2006-1898,2,2,2,1,1,1,1,decrease
CVE-2006-1918,1,1,1,1,1,1,1,stable
CVE-2006-1978,3,3,3,2,1,1,1,decrease
CVE-2006-2084,4,3,4,1,2,2,2,swing
```

ภาพที่ 3.5 ตัวอย่างแบบจำลองข้อมูลสำหรับฝึกสอน

จากภาพที่ 3.5 ชุดข้อมูลสำหรับแบบจำลองจะมีพารามิเตอร์ 9 ตัว คือ หมายเลขซีวีอี และข้อมูลผลการค้นคืนที่ผ่านการปรับปรุงข้อมูลจากข้อ 3.3 จากข้อมูล 7 วันล่าสุดที่ทำการเก็บผลการทดลอง และส่วนสุดท้ายเป็นข้อมูลของกลุ่มสถานะที่ทำการแบ่งกลุ่ม

จากนั้นใช้โปรแกรม Weka ในการตรวจสอบผลการทดลองและความถูกต้องของชุดข้อมูลโดยใช้เทคนิคการจำแนกประเภทข้อมูลแบบเบย์ (Bayesian Classification) โดยใช้ข้อมูลจำนวนฝึกสอน 70 เปอร์เซ็นต์ต่อข้อมูลการทดสอบ 30 เปอร์เซ็นต์ ซึ่งจำแนกวงจรชีพออกได้เป็น 6 ประเภทดังภาพที่ 3.6 ดังนี้

1. สถานะเสถียร (Stable) ลักษณะของข้อมูลไม่มีการเปลี่ยนแปลง

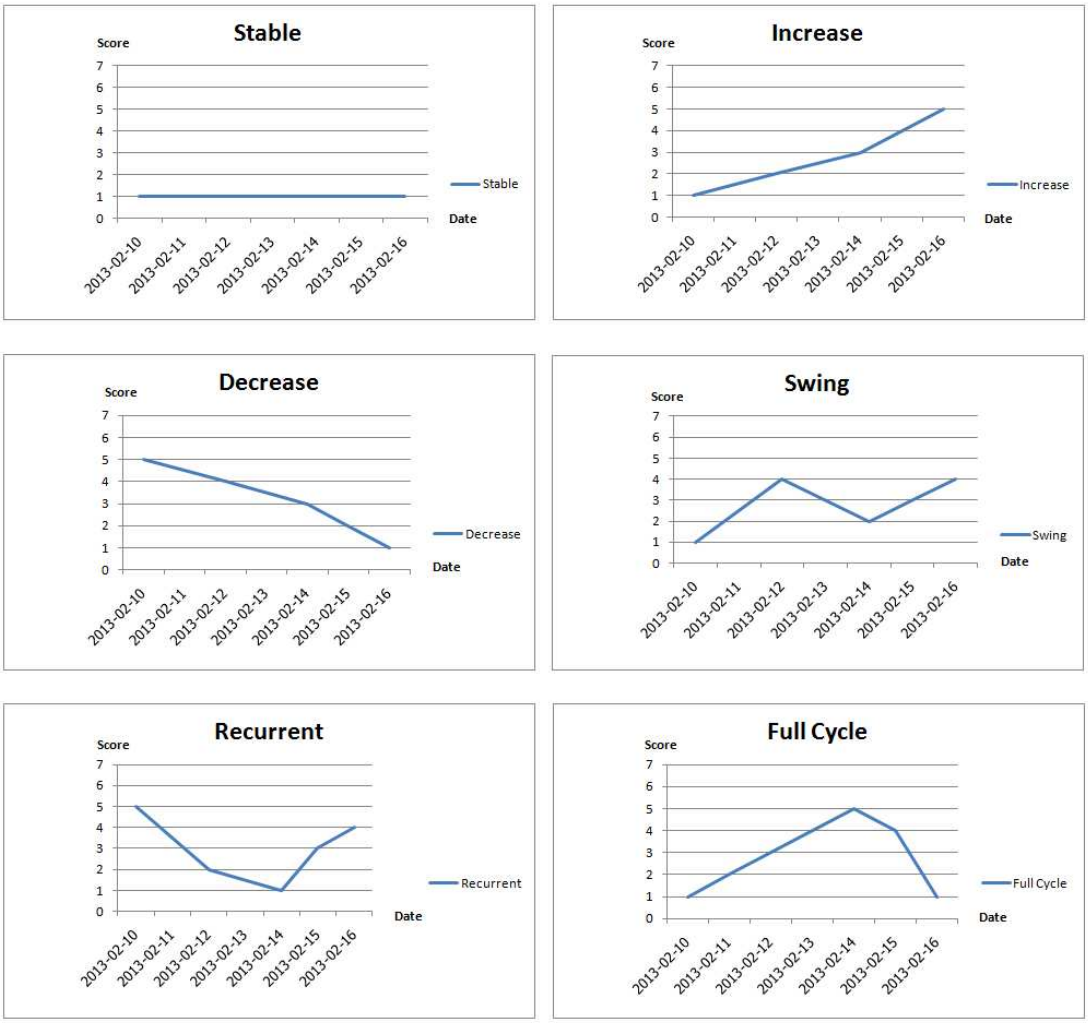
2. สถานะเพิ่มขึ้น (Increase) ข้อมูลมีแนวโน้มเพิ่มขึ้น

3. สถานะลดลง (Decrease) ข้อมูลมีแนวโน้มลดลง

4. สถานะแกว่ง (Swing) ข้อมูลมีการเปลี่ยนแปลงขึ้นลงสลับกัน

5. สถานะฟื้นคืน (Recurrent) ข้อมูลลดลงแล้วกลับมาเพิ่มขึ้น

6. สถานะครบวงจร (Full cycle) ข้อมูลเพิ่มขึ้น คงตัว และลดกลับลงมาเป็นปกติ วงจรชีพ



ภาพที่ 3.6 การจำแนกกลุ่มข้อมูลตามการเปลี่ยนแปลงของสถานะ

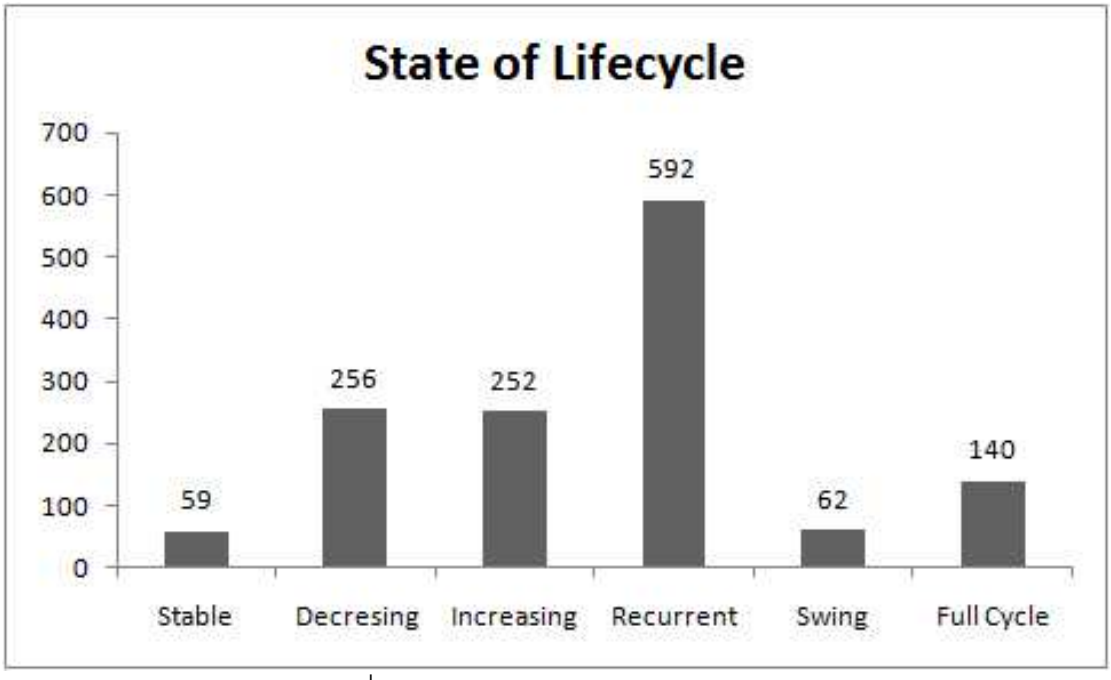
บทที่ 4

ผลการวิจัย

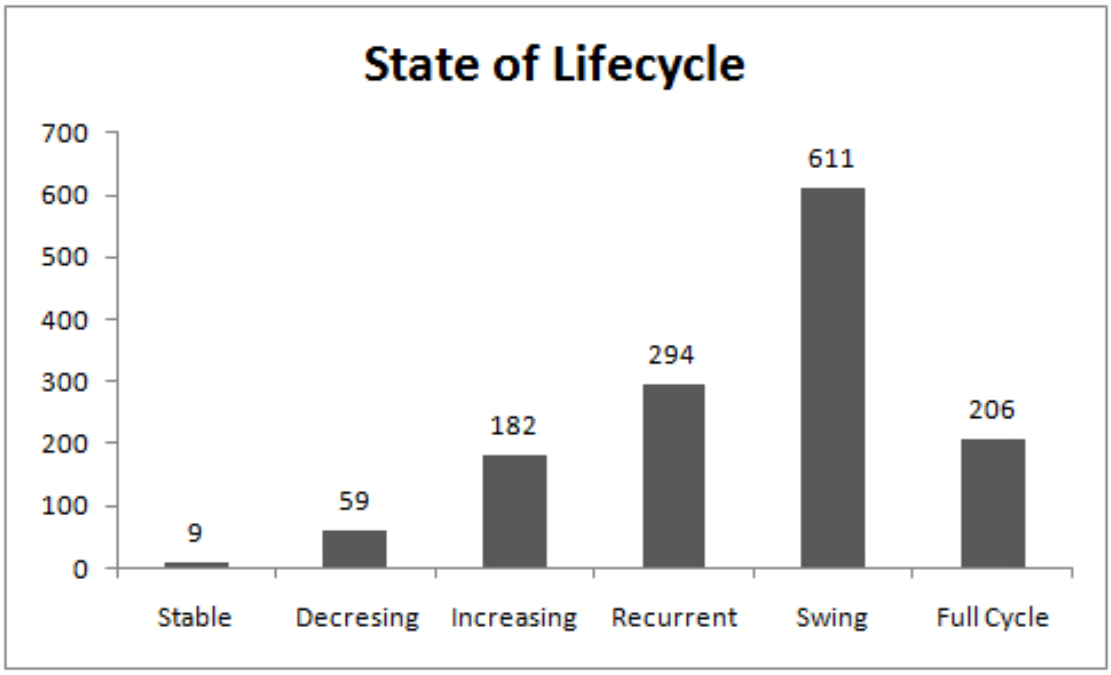
งานวิจัยนี้มีวัตถุประสงค์เพื่อจัดหมวดหมู่ของการเปลี่ยนแปลงวงจรชีพของซีวีอี ในการให้ลำดับความสำคัญของซีวีอีและแยกประเภทเพื่อการติดตามหรือเพิกเฉยซีวีอี โดยอาศัยข้อมูลสาธารณะในการพิจารณาสถานะของซีวีอีเพื่อช่วยเหลือผู้ดูแลระบบในการตัดสินใจติดตามแก้ไขปัญหาของซีวีอี

4.1. การประเมินผลการทดลองจากการเก็บข้อมูล

จากการทดลองติดตามการเปลี่ยนแปลงของซีวีอีจากแหล่งข้อมูลซีวีอีที่ใช้ทดสอบเป็นเวลา 30 วัน ผลการทดลองที่ได้สามารถแยกสถานะวงจรชีพของซีวีอีแยกตามปี ค.ศ. ที่เกิด โดยใช้ข้อมูลย้อนหลัง 7 วันและ 15 วันในการพิจารณา โดยตามเอสแอลเอเป็นข้อตกลงของการแก้ไขปัญหาที่เกิดกับระบบตามระดับความสำคัญโดยมีระยะเวลารับเรื่องและแก้ไขปัญหารวม 7 วัน [11], [12] ในการทดลองได้ทำการทดลองเก็บข้อมูลโดยใช้ช่วงเวลาดังกล่าวเป็นหลัก ผลการทดลองดังภาพที่ 4.1 และภาพที่ 4.2



ภาพที่ 4.1 สัดส่วนของสถานะวงจรชีพ 7 วัน



ภาพที่ 4.2 สัดส่วนของสถานะวงจรชีพ 15 วัน

ตารางที่ 4.1 สถานะของซีวีอีย้อนหลัง 7 วัน

CVE-Year	Stable	Decrease	Increase	Recurrent	Swing	Full Cycle	Total
2004	0	0	0	0	0	1	1
2005	0	2	0	0	0	0	2
2006	3	10	3	5	2	2	25
2007	1	3	3	3	1	1	12
2008	0	9	2	7	1	0	19
2009	3	13	4	5	0	2	27
2010	4	6	3	11	3	2	29
2011	27	27	15	51	5	9	134
2012	19	38	214	447	36	65	819
2013	2	148	8	63	15	57	293
Total	59	256	252	592	62	140	1361

ตารางที่ 4.2 สถานะของซีวีอีย้อนหลัง 15 วัน

CVE-Year	Stable	Decrease	Increase	Recurrent	Swing	Full Cycle	Total
2004	0	0	0	0	1	0	1
2005	0	2	0	0	0	0	2
2006	2	3	0	8	9	3	25
2007	0	0	1	2	7	2	12
2008	0	2	1	5	8	3	19
2009	2	2	2	9	9	3	27
2010	0	1	1	6	16	5	29
2011	4	10	18	23	58	21	134
2012	1	8	152	116	378	164	819
2013	0	31	7	125	125	5	293
Total	9	59	182	294	611	206	1361

สถานะวงจรชีพของซีวีอีที่ทำการติดตามจากข้อมูลสาธารณะในการทดลองนี้ เมื่อพิจารณาจากลักษณะกราฟการเปลี่ยนแปลงของวงจรชีพตามที่กำหนดไว้แบ่งได้ 2 กลุ่มคือ กลุ่มของข้อมูลวงจรชีพของซีวีอีที่ต้องติดตามความเปลี่ยนแปลงโดยในกลุ่มนี้ลักษณะกราฟของข้อมูลออกมาในรูปแบบที่มีแนวโน้มเพิ่มขึ้นได้แก่สถานะเพิ่มขึ้น สถานะแกว่ง และสถานะฟื้นคืน ซึ่งสถานะเหล่านี้บอกได้ว่ากลุ่มซีวีอีเหล่านี้การได้รับความสนใจจากสาธารณะ ส่วนของซีวีอีอีกกลุ่มที่สามารถเพิกเฉยได้เนื่องจากลักษณะกราฟของข้อมูลออกมาในรูปแบบที่มีแนวโน้มลดลงหรือไม่เปลี่ยนแปลงได้แก่สถานะคงที่ สถานะลดลง และสถานะครบวงจร ซึ่งแสดงให้ถึงการได้รับความสนใจจากสาธารณะที่น้อยลง ตามตารางที่ 4.3 และตารางที่ 4.4

ตารางที่ 4.3 ผลการแยกแยะซีวีอีที่ต้องติดตามและเพิกเฉยจากข้อมูลย้อนหลัง 7 วัน

CVE-Year	Tracking	Ignore	Total
2004	0	1	1
2005	0	2	2
2006	10	15	25
2007	7	5	12
2008	10	9	19
2009	9	18	27
2010	17	12	29
2011	71	63	134
2012	697	122	819
2013	86	207	293
Total	907	454	1361

ตารางที่ 4.4 ผลการแยกแยะซีวีอีที่ต้องติดตามและเพิกเฉยจากข้อมูลย้อนหลัง 15 วัน

CVE-Year	Tracking	Ignore	Total
2004	1	0	1
2005	0	2	2
2006	17	8	25
2007	10	2	12
2008	14	5	19
2009	20	7	27
2010	23	6	29
2011	99	35	134
2012	646	173	819
2013	257	36	293
Total	1087	274	1361

จากการทดลองโดยนำผลการทดลองเปรียบเทียบระหว่างการประเมินข้อมูลย้อนหลัง 7 วันกับ 15 วัน จะเห็นว่าเมื่อใช้ช่วงเวลาในการประเมินมากขึ้นทำให้ผลการจัดกลุ่มซีวีอี

มีการเปลี่ยนแปลง ในตารางที่ 4.3 และตารางที่ 4.4 กลุ่มของซีวีอีที่ต้องติดตามมีจำนวนเพิ่มขึ้น เป็นผลมาจากจำนวนซีวีอีที่อยู่ในกลุ่มของซีวีอีในสถานะฟื้นคืนและสถานะแกว่งเพิ่มขึ้น เนื่องจากเมื่อช่วงการพิจารณาข้อมูลมากขึ้นทำให้ลักษณะกราฟของข้อมูลมีการเปลี่ยนแปลง ส่งผลให้การจัดกลุ่มสถานะของซีวีอีมีการเปลี่ยนแปลง เช่น ซีวีอีที่มีสถานะเพิ่มขึ้นเมื่อพิจารณาจากข้อมูล 7 วันแต่มีสถานะของการฟื้นคืนเมื่อพิจารณาจากข้อมูล 15 วันย้อนหลัง หรือสถานะลดลงเมื่อพิจารณาจากข้อมูล 7 วันเปลี่ยนเป็นสถานะครบวงจรเมื่อพิจารณาจากข้อมูล 15 วัน ดังนั้นสถานะของซีวีอีขึ้นอยู่กับช่วงเวลาที่ในการพิจารณาข้อมูลของซีวีอี

เมื่อเปรียบเทียบกลุ่มของซีวีอีที่มีความเสถียรและกลุ่มที่ยังมีการเคลื่อนไหวจากผลการทดลองในตารางที่ 4.3 และ 4.4 จะเห็นว่าในการพิจารณาซีวีอีด้วยข้อมูลย้อนหลัง 7 วัน มีซีวีอีที่ยังมีการเคลื่อนไหวน้อยกว่า

บทที่ 5
สรุปผลการวิจัย และข้อเสนอแนะ

5.1. สรุปผลการวิจัย

งานวิจัยนี้ได้นำเสนอแนวทางในการช่วยเหลือผู้ดูแลระบบในการดูแลแก้ไขปัญหาที่เกิดจากซีวีอี โดยใช้การเปลี่ยนแปลงวงจรชีพของซีวีอีตามการเปลี่ยนแปลงของข้อมูลสาธารณะ ซึ่งแบ่งออกได้เป็น 2 กลุ่มของซีวีอี กลุ่มแรกคือกลุ่มที่สามารถเพิกเฉยได้เพราะพฤติกรรมของการเปลี่ยนแปลงอยู่ในลักษณะที่นิ่งหรือมีแนวโน้มลดลง ส่วนกลุ่มที่สองคือกลุ่มที่ยังคงมีการเคลื่อนไหวและผู้ดูแลระบบต้องทำการติดตามการเปลี่ยนแปลงต่อไป จากซีวีอีที่ติดตามจำนวน 1,361 พบว่ามี 454 ตัวที่สามารถเพิกเฉยได้และอีก 907 ตัวต้องติดตามดูแลต่อ โดยซีวีอีบางส่วนที่มีอายุมาหลายปีแต่ยังคงสถานะที่ต้องติดตามเนื่องมาจากยังคงมีการเคลื่อนไหวของข้อมูลเพราะส่วนโปรแกรมหรือคอมโพเนนท์ที่มีจุดอ่อนยังคงถูกใช้งานและยังคงสร้างปัญหาให้กับความปลอดภัยของระบบ

5.2. ข้อเสนอแนะ

เนื่องด้วยซีวีอีมีเป็นจำนวนมากกว่าหนึ่งหมื่นตัว การติดตามซีวีอีทั้งหมดจำเป็นต้องมีการปรับปรุงระบบที่ใช้ในการเก็บข้อมูล อีกทั้งข้อจำกัดของการค้นหาข้อมูลผ่านทาง Google Search API ที่ไม่สามารถค้นหาข้อมูลซีวีอีทั้งหมดในเวลาต่อเนื่องกันเพราะพฤติกรรมดังกล่าวเสมือนเป็นบอทที่มีจุดประสงค์ในการโจมตีการเว็บไซต์ จึงควรทำการแบ่งกลุ่มของการติดตามซีวีอีและแบ่งแยกไคลเอนต์ที่ใช้เรียก Google Search API เพื่อให้ระบบมีความสามารถในการติดตามซีวีอีทุกตัวทุกวันและหลีกเลี่ยงพฤติกรรมที่มีลักษณะคล้ายการโจมตีเว็บไซต์

5.3. แนวทางการวิจัยต่อ

ปรับปรุงกระบวนการทั้งหมดในงานวิจัยให้อยู่ในรูปแบบของการให้บริการอัตโนมัติ โดยใช้ข้อมูลซีวีอีเป็นข้อมูลสำหรับระบบทำการคิดคำนวณและจัดกลุ่มของซีวีอีดังกล่าวอยู่ในกลุ่มเพิกเฉยหรือติดตาม เพื่อให้เกิดความสะดวกในการประยุกต์ใช้งานจริง

ควรใช้เวลาในการเก็บรวบรวมข้อมูลในการทดลองเป็นระยะเวลาที่ยาวนานกว่านี้เพื่อนำมาเปรียบเทียบผลการทดลองที่ได้ตามช่วงเวลาที่เก็บข้อมูล

รายการอ้างอิง

[1] Mitre Corp. Common Vulnerability and Exposure [Online]. 2013. Available from : http://cve.mitre.org/ [2013, January 18].

[2] CVE Details [Online]. 2013. Available from : http://www.cvedetails.com [2013, January 9].

[3] NIST. National Vulnerability Database [Online]. 2013. Available from : http://nvd.nist.gov/home.cfm [2013, January 11].

[4] P. Ozer, Data Mining Algorithms for Classification, BSc. Thesis, Radboud University Nijmegen Netherland, January 2008.

[5] R. Wita, N. Jiamnapanon, and Y. Teng-amnuay, An ontology for vulnerability lifecycle, Los Alamitos, CA, USA, pp.553–557, IEEE Computer Society, 2010.

[6] R. Wita, N. Jiamnapanon, and Y. Teng-Amnuay, Ontology-based document profile for vulnerability relevancy analysis, World Scientific and Engineering Academy and Society (WSEAS), 2010, pp.210–215.

[7] H.L. Yu, L. Bingwu, and Y. Fang, Similarity Computation of Web Pages of Focused Crawler, International Forum on Information Technology and Applications, 2010, pp. 70-72.

[8] Hitwise. 2009. Top Search Engine Volume [Online]. 2013. Available from : http://www.hitwise.com/uk/datacentre/main/dashboard-7323.html [2013, March 4].

[9] Google Corp. Google AJAX Search API [Online]. 2013. Available from : https://developers.google.com/web-search/ [2013, January 10].

[10] The University of Waikato. Weka Software [Online]. 2013. Available from : http://www.cs.waikato.ac.nz/ml/weka/index.html [2013, February 5].

[11] Qualys, Inc. Service Level Agreement [Online]. 2013. Available from : http://www.qualys.com/support/sla/ [2013, January 10].

[12] SANS. Internal SLA [Online]. 2013. Available from : http://www.sans.org/reading_room/whitepapers/standards/internal-sla-service-level-agreements-information-security_548 [2013, February 5].

ประวัติผู้เขียนวิทยานิพนธ์

นายอรรถพล พวงพุ่ม เกิดเมื่อวันที่ 24 พฤศจิกายน พ.ศ. 2526 ที่จังหวัดตรัง เป็นบุตรชายคนที่สอง ของนายไพฑูรย์ พวงพุ่ม และนางฉวีวรรณ พวงพุ่ม สำเร็จการศึกษาระดับปริญญาวิทยาศาสตรบัณฑิต สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์จากมหาวิทยาลัยสงขลานครินทร์ ในปี พ.ศ. 2548 และได้เข้าศึกษาต่อในระดับปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในภาคการศึกษาต้น ปีการศึกษา 2552