

การศึกษาการสืบค้นและวิเคราะห์ข้อมูลข่าวกรองแบบเปิดผ่าน อินเทอร์เน็ต

Feasibility Study on Search and Analysis of Open-Sources Web Intelligence

วัชรภูมิ ไหวว่อง¹ อำนวย ขาวเน²

¹สาขาวิศวกรรมป้องกันประเทศ คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

E-mail : 55613951@kmitl.ac.th

²ภาควิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

E-mail : kkamnach@kmitl.ac.th

บทคัดย่อ

งานวิจัยนี้ศึกษาถึงความเป็นไปได้ของการนำข้อมูลข่าวเปิดในอินเทอร์เน็ต มารวบรวมประมวลผล และสืบค้นหาความเชื่อมโยงผ่านการใช้งานของซอฟต์แวร์โอเพ่นซอร์ส เนื่องจากปัจจุบัน ข้อมูลข่าวสารมีอยู่มากมายบนรูปแบบเว็บไซต์ โดยจะทำการเก็บข้อมูลต่างๆ ในรูปแบบของข้อมูลข่าวกรองที่เป็นประโยชน์ เพื่อนำไปสู่การประมวลผลแบบหาเครือข่ายความเชื่อมโยง ของข้อมูลดิบที่มีอยู่ และนำไปสู่การสอบสวนเพื่อขยายผล หาเครือข่ายของผู้ต้องสงสัย หรือบุคคลเป้าหมาย เพื่อทำการคาดการณ์เหตุการณ์ หรือความรุนแรงต่างๆ ที่จะเกิดขึ้นในอนาคตได้ และพัฒนาไปสู่การสร้างเป็นระบบสืบค้นข้อมูลที่มีประสิทธิภาพยิ่งขึ้นในอนาคต

คำสำคัญ : ข่าวกรองแบบเปิด, ข่าวเปิด, การสืบค้น

Abstracts

This research investigates the possibility of introducing open news information on the Internet. To compile and find the link through use of open-source software. Because of at the present the information is abundant on the website layout. It will store data in the form of metadata to lead to the processing of network links of raw data available. And lead to the investigation to expand the results. Find a network of suspects or target person to make an event forecast or violence before it appeared in the future and will developed into a more efficient search system tools in the future.

Keywords: Open-source Intelligence, Intelligent, Search

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. บทนำ

ข่าวกรองมีชื่อเรียกกันหลายชนิด หรือหลายประเภท โดยเรียกตามลักษณะของเทคนิค หรือวิธีการรวบรวมข่าวสาร และความมุ่งหมายในการใช้ และเพื่อประโยชน์ในการแบ่งความรับผิดชอบในการปฏิบัติการใช้ข่าวสาร ข่าวกรอง และการประสานงานด้านการข่าวกรอง

โดยทั่วไปจึงแบ่งข่าวกรองออกเป็น 2 ประเภทใหญ่ๆ คือ ข่าวกรองแห่งชาติ ข่าวกรองทางทหาร และข่าวกรองประเภทอื่น ๆ โดยมีรายละเอียด ดังนี้

1. ข่าวกรองแห่งชาติ เป็นข่าวกรองในระดับกระทรวง ทบวง กรม ตั้งแต่หนึ่งหน่วยขึ้นไปเป็นผู้รวบรวมและดำเนินการวิธี หรือเป็นผู้ใช้เพื่อรักษาผลประโยชน์และความมั่นคงของชาติ

2. ข่าวกรองทางทหาร เป็นข่าวกรองสำหรับนำมาใช้ในการวางแผนการปฏิบัติการตามแผน นโยบาย โครงการ และกำหนดการทางทหาร ซึ่งจะแบ่งได้ดังนี้

2.1. ข่าวกรองทางการรบ หรือข่าวกรองทางยุทธวิธี คือความรู้เกี่ยวกับฝ่ายตรงข้าม ลมฟ้าอากาศ และลักษณะทางภูมิศาสตร์ ที่ผู้บังคับบัญชาต้องการในการวางแผนและการปฏิบัติการทางยุทธวิธี ข่าวกรองทางการรบ ได้มาจากการดำเนินการวิธีต่อข่าวสารที่เกี่ยวข้องกับองค์ประกอบฝ่ายตรงข้าม ลมฟ้าอากาศ และภูมิประเทศที่ได้รับมาจากเจ้าหน้าที่รวบรวมข่าวสารของหน่วยเหนือ หน่วยรอง หน่วยข้างเคียง และแหล่งข่าวอื่น ๆ ซึ่งจะให้ข่าวสารและข้อสรุปที่เกี่ยวกับพื้นที่ปฏิบัติการ จัดความสามารถ จุดล่อแหลม และหนทางปฏิบัติของฝ่ายตรงข้าม

2.2 ข่าวกรองทางยุทธศาสตร์ คือ ความรู้เกี่ยวกับพื้นฐานในการกำหนดนโยบายและแผนการทางของต่างชาติใดชาติหนึ่ง หรือหลาย ๆ ชาติ โดยมุ่งพิจารณาถึงวัตถุประสงค์ของชาติ การวางแผน และวิธีการปฏิบัติ เพื่อให้บรรลุวัตถุประสงค์ความต้องการข่าวกรองทางยุทธศาสตร์ ได้แก่ จัดความสามารถ จุดล่อแหลม และหนทางปฏิบัติที่น่าจะเป็นไปได้ของต่างชาติทั้งที่เป็นพันธมิตร เป็นกลาง และเป็นศัตรูหรือที่น่าจะเป็นศัตรู กล่าวโดยสรุปข่าวกรองทางยุทธศาสตร์ คือ การพิจารณาถึงพลัง

อำนาจของต่างชาติ ซึ่งพลังอำนาจดังกล่าว คือความสามารถของชาติหนึ่งในการที่จะดำเนินการตามวัตถุประสงค์ของชาตินั้น ข่าวกรองชนิดนี้มีลักษณะและจุดมุ่งหมาย ในการรวบรวมข่าวสารที่มุ่งเน้นในด้านการเตรียมการป้องกันประเทศและการทำสงคราม ซึ่งแตกต่างจากข่าวกรองการต่างประเทศ ที่เน้นในด้านของการกำหนดนโยบายต่างประเทศ และการดำเนินความสัมพันธ์กับต่างประเทศ อย่างไรก็ตามข่าวกรองทั้งสองชนิดนี้ต่างมีความสำคัญเท่าเทียมกัน เนื่องจากจะใช้เป็นมูลฐานในการกำหนดนโยบายของชาติ โดยเฉพาะนโยบายทางการเมืองทั้งภายในและต่างประเทศ นโยบายทางเศรษฐกิจ

2.3 ข่าวกรองประเภทอื่นๆ เป็นข่าวกรองที่กำหนดขึ้นตามหน้าที่ ตามลักษณะของการปฏิบัติการและความมุ่งหมายในการใช้อีกหลายชนิด ซึ่งข่าวกรองชนิดต่างๆ มีลักษณะเป็นส่วนประกอบทั้งข่าวกรองแห่งชาติ และข่าวกรองทาง โดยอาจเป็นข่าวกรองทางยุทธศาสตร์ หรือข่าวกรองทางยุทธวิธีก็ได้ เช่น ข่าวกรองทำเนียบ กาลังรบ ข่าวกรองแบบเปิด (Open-Source Intelligence – OSINT) ข่าวกรองทางเทคนิค (Technical Intelligence – TECHINT) ข่าวกรองภาพถ่าย (Imagery Intelligence – IMINT) ข่าวกรองเป้าหมาย (Target Intelligence) ข่าวกรองทางการติดต่อสื่อสาร (Communications Intelligence – COMINT) ข่าวกรองภูมิประเทศ ข่าวกรองลมฟ้าอากาศ ข่าวกรองปัจจุบัน ข่าวกรองมูลฐาน ข่าวกรองเพื่อความมั่นคง ข่าวกรองเรือ ข่าวกรองทางอากาศ เป็นต้น

ในปัจจุบันข่าวกรองแบบเปิดผ่านทางเว็บไซต์เป็นที่นิยมและแพร่หลายในหลายๆ รูปแบบ เนื่องจากมีรูปแบบการใช้งานที่ง่าย และไม่ยุ่งยากจนเกินไป ประกอบกับการที่มีในส่วนของโซเชียลมีเดีย เครือข่ายทางสังคมออนไลน์ ทำให้การเผยแพร่ข่าวเปิด เป็นไปอย่างรวดเร็ว และยังมีความแม่นยำในระดับหนึ่ง โดยการเผยแพร่นี้ยังรวมไปถึงการสร้างข่าวสำหรับเชิงลบ และข่าวลือที่เป็นประโยชน์ต่อกลุ่มใดกลุ่มหนึ่ง ดังนั้น ในบทความนี้จะเน้นทางด้านข่าวเปิดบนโลกอินเทอร์เน็ตเป็นหลัก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 1: ส่วนประกอบ และองค์ประกอบทางสื่อไอเฟนซอร์ส อินเทอร์เน็ตเว็บไซต์

Media	Components	Elements
Internet Web Sites	Communications	- Chat - Web cam
		- E-mail - Web cast
		- News - Web log
	Databases	- Commerce - Government - Education - Military Orgs.
Information (Web page content)		- Commerce - Government - Education - Military Orgs.
		Services

2. ทฤษฎีที่เกี่ยวข้อง

การศึกษาในครั้งนี้เป็นการศึกษาความเป็นไปได้ในการสืบค้นหาข้อมูลข่าวเปิดจากอินเทอร์เน็ต เพื่อนำไปสู่การสืบสวนข้อมูลเพื่อการทำงานหากลุ่มเป้าหมายที่ต้องสงสัยต่อไป

ทฤษฎีที่ได้กล่าวถึงในการศึกษานี้ จะได้กล่าวถึงการใช้วีรอนเน็ตเวิร์ก ในการทำการศึกษานี้ และจำแนกประเภทข้อมูล และการใช้การสืบสวนสอบสวนในการแยกเป้าหมาย บ่งชี้ลักษณะ พฤติกรรมของเป้าหมาย และการขยายผลการสืบค้นต่อไป

สำหรับในคอมพิวเตอร์ Neurons ประกอบด้วย input และ output เหมือนกัน โดยจำลองให้ input แต่ละ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

อันมี weight เป็นตัวกำหนดน้ำหนักของ input โดย neuron แต่ละหน่วยจะมีค่า threshold เป็นตัวกำหนดว่าน้ำหนักรวมของ input ต้องมากขนาดไหนจึงจะสามารถส่ง output ไปยัง neurons ตัวอื่นได้ เมื่อนำ neuron แต่ละหน่วยมาต่อกันให้ทำงานร่วมกันการทำงานนี้ในทางตรรกแล้วก็จะเหมือนกับปฏิกิริยาเคมีที่เกิดในสมอง เพียงแต่ในคอมพิวเตอร์ทุกอย่างเป็นตัวเลขเท่านั้นเอง

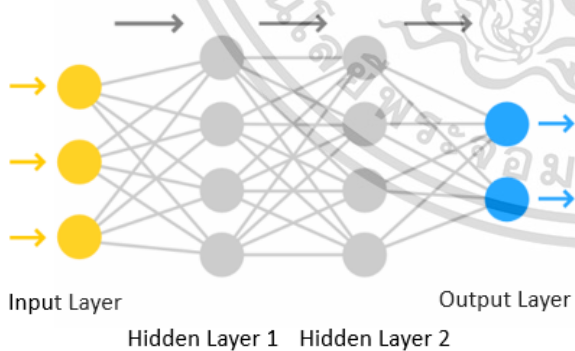
การทำงานของ Neural Networks คือเมื่อมี input เข้ามายัง network ก็เอา input มาคูณกับ weight ของแต่ละขา ผลที่ได้จาก input ทุก ๆ ขาของ neuron จะเอามารวมกันแล้วก็เอามาเทียบกับ threshold ที่กำหนดไว้ ถ้าผลรวมมีค่ามากกว่า threshold แล้ว neuron ก็จะส่ง output ออกไป output นี้ก็จะถูกส่งไปยัง input ของ neuron อื่น ๆ ที่เชื่อมกันใน network ถ้าค่าน้อยกว่า threshold ก็จะไม่เกิด output สิ่งสำคัญคือเราต้องทราบค่า weight และ threshold สำหรับสิ่งที่เราต้องการเพื่อให้คอมพิวเตอร์รู้จัก ซึ่งเป็นค่าที่ไม่แน่นอน แต่สามารถกำหนดให้คอมพิวเตอร์ปรับค่าเหล่านั้นได้โดยการสอนให้รู้จัก pattern ของสิ่งที่เราต้องการให้รู้จัก เรียกว่า "back propagation" ซึ่งเป็นกระบวนการย้อนกลับของการรู้จัก ในการฝึก feed-forward Neural Networks จะมีการใช้อัลกอริทึมแบบ back-propagation เพื่อใช้ในการปรับปรุงน้ำหนักคะแนนของเครือข่าย (Network Weight) หลังจากใส่รูปแบบข้อมูลสำหรับฝึกให้แก่เครือข่ายในแต่ละครั้งแล้ว ค่าที่ได้รับ (output) จากเครือข่ายจะถูกนำไปเปรียบเทียบกับผลที่คาดหวัง แล้วทำการคำนวณหาค่าความผิดพลาด ซึ่งค่าความผิดพลาดนี้จะถูกส่งกลับเข้าสู่เครือข่ายเพื่อใช้แก้ไขค่าน้ำหนักคะแนนต่อไป

การประยุกต์ใช้งาน Neural Networks ในบทความนี้ จะใช้ในการประมาณค่าฟังก์ชันหรือการประมาณความสัมพันธ์ (โดยมี inputs และ outputs แต่ไม่ทราบว่า inputs กับ outputs มีความสัมพันธ์กันอย่างไร)

ส่วนที่เล็กที่สุดของ Neural Network ก็คือ Neuron ซึ่งทำหน้าที่คำนวณ input ที่เข้ามาเพื่อให้ได้ผลลัพธ์ออกไป โดยมีส่วนประกอบสำคัญดังนี้

1. Input หรือค่าที่ส่งเข้ามาที่ Neuron โดยจะมีค่าที่เข้ามาได้หลายค่า ขึ้นอยู่กับสิ่งที่เราต้องการจะสร้าง
2. Weight เป็นการให้น้ำหนักของ input แต่ละที่ส่งเข้ามา โดยมีค่าระหว่าง 0-1 เมื่อเริ่มต้นจะเป็นการ Random ขึ้นมา จากนั้นตัว Neuron เมื่อทำการเรียนรู้แบบต่อเนื่อง ก็จะทำการปรับ weight ให้ได้คำตอบที่ใกล้เคียงกับที่เราต้องการให้มากที่สุด
3. Bias คือค่าที่จะช่วยเข้ามาทำให้ค่าที่เข้ามาอยู่ในระหว่าง 0 - 1 ได้ โดยจะเป็นเลข random และปรับไปเรื่อยๆ ทุกครั้งที่ได้เรียนรู้
4. Output คือผลลัพธ์
5. Back Propagation คือการที่ Neuron นำค่า Error ของ Output ที่ได้ กับ Output ที่เราสั่งให้มันเรียนรู้ นำไปปรับ Weight และ Bias ให้เกิดผลลัพธ์ที่ถูกต้องตามที่ได้เรียนรู้มา

รูปแบบของ Neural Network ที่ง่ายที่สุดคือ Feed-Forward Neural Network โดยจะแบ่ง Neuron ออกเป็นกลุ่มๆ โดยแต่ละกลุ่มจะเรียกเป็น Layer โดยข้อมูลที่เข้ามาจะไหลไปในทิศทางเดียว ไม่ไหลย้อนกลับจาก Layer หนึ่งสู่อีก Layer หนึ่ง ดังรูปที่ 1



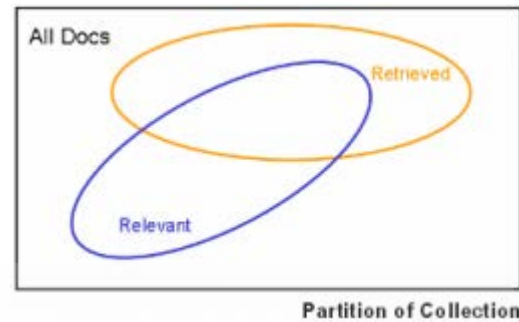
รูปที่ 1 : ข้อมูลที่เข้ามาจะไหลไปในทิศทางเดียว ไม่ไหลย้อนกลับ จาก Layer หนึ่งสู่อีก Layer หนึ่ง

เพื่อที่จะทำการวัดผลข้อมูล เราจะใช้หลักการ Precision Recall ดังนี้

$$Recall = \frac{\text{จำนวนของเอกสารที่เกี่ยวข้องทั้งหมดและถูกดึงออกมา}}{\text{จำนวนเอกสารที่เกี่ยวข้องทั้งหมด}} \quad (1)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$Precision = \frac{\text{จำนวนของเอกสารที่เกี่ยวข้องและถูกดึงออกมา}}{\text{จำนวนเอกสารที่ถูกดึงออกมาทั้งหมด}} \quad (2)$$

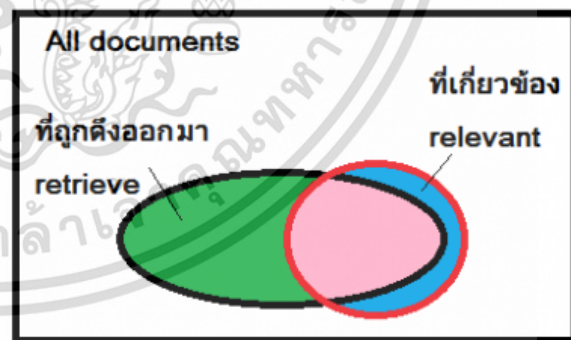


รูปที่ 2 : แสดงตามทฤษฎี Precision และ Recall

Precision คือ การวัดความสามารถในการที่จะขจัดเอกสารที่ไม่เกี่ยวข้องออกไป โดย precision เป็นอัตราส่วนของจำนวนเอกสารที่เกี่ยวข้องและถูกดึงออกมากับจำนวนเอกสารที่ถูกดึงออกมาทั้งหมด

recall คือ การวัดความสามารถของระบบในการดึงเอกสารที่เกี่ยวข้องออกมา โดย recall เป็น อัตราส่วนของจำนวนเอกสารที่เกี่ยวข้องและถูกดึงออกมากับจำนวนเอกสารที่เกี่ยวข้องทั้งหมด

โดย Relevant คือ จำนวนเอกสารที่เกี่ยวข้อง Retrieved คือ จำนวนเอกสารที่ถูกดึงออกมา



รูปที่ 3 : แสดงค่า Precision จะได้อัตราส่วนต่ำ แสดงว่าขจัดไม่เก่ง, ค่า Recall จะได้อัตราส่วนสูง แสดงว่าดึงข้อมูลเก่ง

จากหลักการนี้ นำไปสู่การเก็บข้อมูลข่าวเปิดที่มีอยู่ทั่วไปในโลกของอินเทอร์เน็ต สื่อสังคมที่ออนไลน์ และข้อมูลข่าวเปิดต่างๆ ที่เป็นประโยชน์ เพื่อนำข้อมูลต่างๆ ที่ได้มาวิเคราะห์ สืบสวนต่อในกระบวนการต่อไป

ในส่วนของการสืบสวนนั้น มีหลักในการปฏิบัติคือการรวบรวมข้อมูลที่เกี่ยวข้องกับ สิ่งที่ต้องสงสัยให้มากที่สุดเท่าที่จะมากได้ แล้วนำข้อมูลนั้นมาวิเคราะห์ และคาดเดา ด้วยการสร้างจินตนาการตามสามัญสำนึกแห่งบุคคลบนพื้นฐานของเหตุและผล โดยตั้งเป็นสมมุติฐานแล้ว พิสูจน์สมมุติฐานนั้น เพราะฉะนั้นข้อมูลของสิ่งที่ต้องสงสัย จึงเป็นปัจจัยอันสำคัญที่สุดในการสืบสวนโดยทั่วไปแล้ว ข้อมูล จะแบ่งออกเป็น 2 ชนิด คือ

1. ข้อเท็จจริง คือข้อมูลที่เป็นความจริง ได้รับการพิสูจน์ว่าเป็นความจริงแล้ว
2. ข้ออนุมาน คือ ข้อมูลที่คาดเดาเอง ยังไม่ได้รับการพิสูจน์ว่าเป็นความจริง

ดังนั้น ข้อมูลที่เป็นจริงจึงเป็นปัจจัยอันสำคัญที่สุดที่ต้องใช้ในการสืบสวน

3. การทดลอง

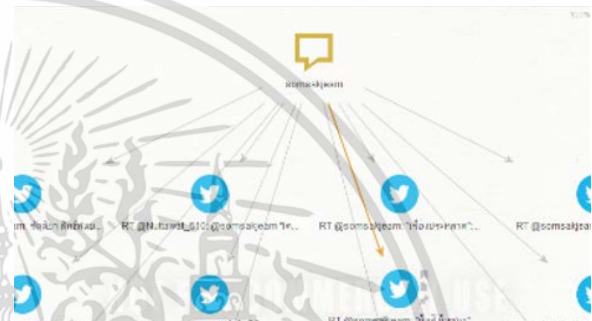
ในงานวิจัยนี้จะทำการทดลองโดยการใช้โปรแกรมประยุกต์ในการรวบรวมหาข้อมูลข่าวสารจากแหล่งข่าวเปิด เพื่อใช้ในการวิเคราะห์ผล โดยจะแตกต่างกับข้อมูลที่ได้รับจากทางที่ได้รับ จากเดิมที่ต้องทำแบบไม่อัตโนมัติซึ่งทำให้เสียเวลาในการวิเคราะห์ข้อมูล แต่วิธีการนี้ จะช่วยให้ทำการวิเคราะห์ข้อมูลแบบอัตโนมัติ และมีความรวดเร็วในการทำงานเพิ่มขึ้น และไม่เสียเวลาในการวิเคราะห์ข้อมูลแบบ One-by-one แต่จะได้ข้อมูลมาช่วยในการวิเคราะห์มากขึ้น

โปรแกรม Maltego เป็นเครื่องมือที่สามารถใช้งานได้ฟรี และมีแบบจ่ายเพิ่ม เครื่องมือนี้มีความแตกต่างจากเครื่องมือประเภทอื่นๆ คือ จะทำหน้าที่เป็น Digital forensics โดยอาศัยหลักการของ Data mining ในการติดตาม แกระอยข้อมูลที่เรากำลังสนใจ ซึ่งสามารถวิเคราะห์ข้อมูลได้แบบ Real time และทำงานได้หลากหลาย platform เพื่อให้เราสามารถเรียกดูข้อมูลได้ง่ายและสะดวก

ในการทดลองนี้ จะใช้เป้าหมายที่เป็นบุคคลที่เป็นที่ต้องการตัวของทางการไทย รศ.สมศักดิ์ เจียมธีรสกุล ผู้ต้องหาคดีหมิ่นพระบรมเดชานุภาพ ตามประมวล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กฎหมายอาญา มาตรา 112 ปัจจุบันเป้าหมายได้ทำการหลบหนีไปอยู่ต่างประเทศ แต่ยังคงมีการเคลื่อนไหวทางการเมือง เพื่อแสดงความคิดเห็น และแนวคิดต่างๆ ผ่านทางโซเชียลมีเดียต่างๆ ไม่ว่าจะเป็น Facebook หรือ Twitter แต่ในการศึกษานี้จะทำการ Monitor เป้าหมายผ่าน Twitter เป็นหลัก เนื่องจากข้อจำกัดของโปรแกรมที่ยังไม่สามารถ Monitor ผ่าน Facebook ได้ โดยทำการค้นหา URL เป้าหมายผ่านทางเว็บไซต์ Twitter และได้ URL เป้าหมายมาดังนี้ <https://twitter.com/somsakjeam> ดังรูปที่ 3



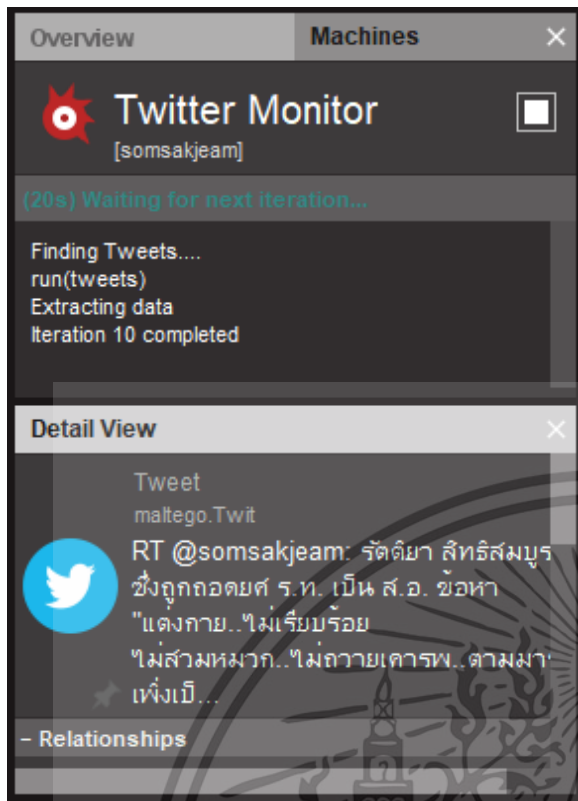
รูปที่ 4 : แสดงการค้นหาเครือข่ายความเชื่อมโยงของเป้าหมาย

โดยอาศัยสมมุติฐานของการใช้ทฤษฎีของนิเวศน์เน็ตเวิร์ก เพื่อทำการเก็บข้อมูล และหาเครือข่ายความเชื่อมโยงของเป้าหมายได้ ดังรูปที่ 5



รูปที่ 5 : แสดงเครือข่ายความเชื่อมโยงของเป้าหมายที่ได้ทำการทดลอง

สามารถติดตามดูข้อความจริงได้จาก URL จริงจากเครือข่ายความเชื่อมโยง ดังรูปที่ 6 และรูปที่ 7



รูปที่ 6 : แสดงผลการเก็บข้อมูลที่ได้



รูปที่ 7 : แสดงผลอ้างอิงจาก URL จริง

จากการทดลอง ได้ทำการเก็บผลเพื่ออ้างอิงความแม่นยำของการค้นหาจากเป้าหมายทั้งหมด เพื่อให้เห็นถึงคุณค่าของข่าวกรองที่ได้ ในเชิงปริมาณว่ามีความสำคัญและแม่นยำแค่ไหน โดยใช้ทฤษฎีของ Precision & Recall เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยจะทำการทดสอบในระบบ twitter โดยงานศึกษานี้มีการทดสอบประสิทธิภาพด้วยค่าความแม่นยำ ค่าความครบถ้วน และ F-Measure ซึ่งเป็นที่ยอมรับและแพร่หลายเพื่อทดสอบความถูกต้องของข้อมูล โดยผู้ทำการศึกษาได้เลือกคำสำคัญจำนวน 5 คำ จาก 105 คำ เพื่อใช้ทำการสืบค้น โดยทำการทดสอบกับข้อมูลจริงจำนวน 2000 ดังตารางที่ 2

ตารางที่ 2 : แสดงผลการทดลองเทียบกับข้อมูลที่เก็บได้และความแม่นยำ

คำสำคัญ	ค้นข้อมูลแบบธรรมดา		ค้นข้อมูลแบบมี Entity	
	Precision	Recall	Precision	Recall
ศันผวน	1	0.1	0.89	1
ถวายความเคารพ	1	0.4	0.89	1
สำคัญ	1	0.1	0.59	1
หูน	1	0.3	0.41	1
SCB	0.71	0.2	0.55	1
คำเกลียด	0.94	0.22	0.66	1
ค่า F-Measure	0.36		0.79	

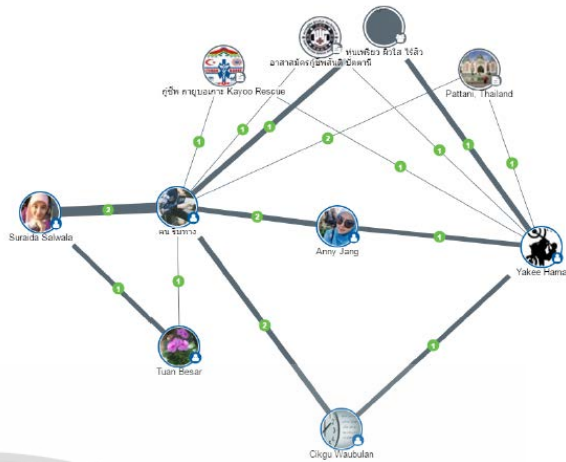
จากผลการทดลอง การค้นข้อมูลแบบธรรมดาพบว่าได้ค่า Precision = 0.94 ค่า Recall = 1 และค่า F-Measure = 0.36

จากผลการทดลอง การค้นข้อมูลแบบมี Entityพบว่าได้ค่า Precision = 0.79 ค่า Recall = 1 และค่า F-Measure = 0.79

การทดลองถัดมา ได้ทำการเลือกบุคคลต้องสงสัยผ่านโซเชียลเน็ตเวิร์ค จำนวน 3 เป้าหมาย โดยเป้าหมายนั้นมีความเกี่ยวข้องกับความรุนแรงในสถานการณ์ทางภาคใต้ โดยมีจำนวนเป้าหมายดังต่อไปนี้

1. Suraida Salwala
2. Combeng Cahya
3. Yakee Hama

โดยเป้าหมายที่ชื่อ Combeng Cahya ไม่เปิดเผยจำนวนของเพื่อนใน Facebook แต่ทางผู้ทำการทดสอบใช้งานได้ทำการทดสอบด้วยการใช้การ Reconstruction เพื่อทำการรวบรวมข้อมูลทางเพื่อนของเพื่อน เพื่อที่ทำการเปิดเผยจำนวนเพื่อนทั้งหมดของเป้าหมาย โดยหลังจากทำการทดสอบการใช้งานจริงแล้ว พบเพื่อนทั้งหมด 272 ราย กับเป้าหมาย



รูปที่ 9 : แสดงความเชื่อมโยงของ 3 เป้าหมายที่มีความเกี่ยวข้องกัน

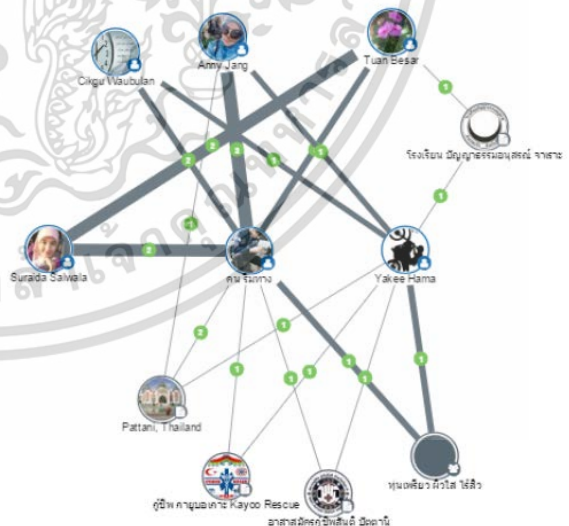


รูปที่ 8 : แสดงจำนวนเพื่อนที่ไม่เปิดเผยตัวตนของ Combeng Cahya

จากการทำการวิเคราะห์ในเชิงลึก เราจะเห็น 3 เป้าหมายที่เพิ่มเข้ามาใหม่ ที่มีความสัมพันธ์กันกับ 3 เป้าหมายแรก ดังนี้

1. Anny Jang (FB username: nurainee.salah)
2. Tuan Besar (FB username: amrin deelak)
3. Cikgu Waubulan (FB username: samsudin.dama)

จากนั้น ได้ทำการเก็บบันทึกข้อมูลบุคคลทั้งสามเพิ่มเติม ได้ผลสรุปออกมาดังรูปความสัมพันธ์ ดังต่อไปนี้



รูปที่ 10 : แสดงความสัมพันธ์เพิ่มเติมของกลุ่มเป้าหมายใหม่และเก่า

ข้อมูลจากความสัมพันธ์ระหว่าง Suraida Salwala กับ Combeng Cahya (คน ริมทาง) นั้นเป็นบุคคลที่มีปฏิสัมพันธ์กัน โดย Yakee Hama นั้น มีความสัมพันธ์กัน เอกสารนี้เป็นเอกสารที่ส่งมอบไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผ่านกับเป้าหมายใหม่ Anny Jang ไปยัง Combeng Cahya (คน ริมทาง) นั่นเอง

4. ผลการทดลอง

จากผลการทดลอง พบว่าข้อมูลข่าวกรองที่ได้ เป็นประโยชน์ต่อการนำไปวิเคราะห์ผล และสามารถหา ความเชื่อมโยงต่อไปของบุคคลต้องสงสัย เพื่อที่จะได้ทำ การพยากรณ์ คาดการณ์เป้าหมาย ต่อไปในอนาคต

5.สรุป

จากการศึกษางานวิจัยนี้พบว่า ข้อมูลที่ได้เป็น การนำข้อมูลข่าวกรองเปิดในโลกออนไลน์ เพื่อนำข้อมูล ต่างๆ มาประมวลผล และเพื่อทำการวิเคราะห์ข้อมูลเชิงลึก โดยทำการเปรียบเทียบความคล้ายกันของข้อมูล ทำการ เปรียบเทียบกับความคล้ายกันของคำ มุ่งเน้นไปที่ผลลัพธ์ การขยายผลต่อไป เพื่อตรวจสอบเครือข่ายในการ แพร่กระจายข้อมูลข่าวเปิดออกไป จากผลการทดลอง ประสิทธิภาพในตารางที่ 2 ได้ค่า Precision = 0.66 ค่า Recall = 1 และค่า F-Measure = 0.79 แสดงให้เห็นว่าสิ่งที่ พัฒนาขึ้นให้ผลประเมินในระดับที่ดี และยอมรับได้ โดย ในอนาคต สามารถนำข้อมูล แนวคิด ทฤษฎีต่างๆ เหล่านี้ ไปประยุกต์ใช้ สำหรับหน่วยงานความมั่นคง สำนักงาน ตำรวจแห่งชาติ กองอำนาจการรักษาความมั่นคงภายใน กองบัญชาการกองทัพบก เป็นต้น

6.กิตติกรรมประกาศ

ขอขอบคุณ ดร. อำนาจ ขาวเน ที่ได้ให้การ เสียสละ เวลา ความรู้ และคำแนะนำที่เป็นประโยชน์ต่อ งานวิจัยนี้ ขอขอบคุณ เจ้าหน้าที่ทหารศูนย์รักษาความ ปลอดภัย กองบัญชาการกองทัพไทย และเจ้าหน้าที่ทหาร กรมยุทธการทหาร กองบัญชาการกองทัพไทย ที่ให้ความ อนุเคราะห์ข้อมูลต่างๆ ที่ใช้ในการทำกรณีศึกษา และการ วัตถุประสงค์

7. เอกสารอ้างอิง

- [1] Raymond T. Ordierno, Joyce E. Morrow, "Techniques Publication, Designation: ATP 2-22.9 "Open-Source Intelligence".," Headquarters, Department of the Army, Army, pp 20-22, 1996
- [2] Christian Nünlist, "CSS Analyses in Security Policy. Open Source Intelligence: A Strategic Enabler of National Security, Center for Security Studies (CSS)," ET H Zurich, vol.22, pp.-51-57, 2010
- [3] Fogelmann-Soulié et al.(Eds.), "Mining Massive Data Sets for Security: Advances in Data Mining, Search, Social Networks and Text Mining, and Their Applications to Security 19," Amsterdam, IOS Press, pp. 331-344, 2008
- [4] Darren Bradbury, "In plain view: Open Source Intelligence," Computer Fraud & Security 4, pp. 5-9, 2011
- [5] Loch Johnson, "R.D: Open Source Intelligence," Handbook of Intelligence Studies, pp. 129-147, 2007
- [6] Sutida Numak, Maleerat Sodanin, "Development of Query Expansion System using Topic Model," The 9th National Conference on Computing and Information Technology, pp.367-372, 2013

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้