เว็บแอปพลิเคชันสำหรับวิเคราะห์เครือข่ายผู้ก่อความไม่สงบโดยใช้ข้อมูลจากข่าวภาษาไทย

A Web Application for Insurgent Network Analysis using Thai News

นายคชาภัฏ บัวขาว

Kachapat Buakhaw

นายรติภาพ ชินราช

Ratiphab Chinnarath

นายพงศกร คำนึง

Pongsakorn Khamneueng

โครงงานนี้เป็นส่วนหนึ่งของการศึกษาหลักสูตรวิทยาศาสตรบัณฑิต

ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์

มหาวิทยาลัยศรีนครินทรวิโรฒ ปีการศึกษา พ.ศ. 2562

คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ

ชื่อหัวข้อโครงงาน	เว็บแอปพลิเคชันสำหรับวิเคราะห์เครือข่ายผู้ก่อความไม่สงบ โดยใช้ข้อมูลจากข่าวภาษาไทย A Web Application for Insurgent Network Analysis using Thai News		
นิสิต	นายคชาภัฏ	บัวขาว	59102010240
	นายรติภาพ	ชินราช	59102010278
	นายพงศกร	คำนึง	59102010907
ปริญญา	วิทยาศาสตรบัณฑิต (วท.บ.)		
ภาควิชา	วิทยาการคอมพิวเตอร์		
อาจารย์ที่ปรึกษาโครงงาน	ผู้ช่วยศาสตราจารย์ ดร.ศุภชัย ไทยเจริญ		

ลงชื่อ.....

(ผู้ช่วยศาสตราจารย์ ดร.ศุภชัย ไทยเจริญ)

อาจารย์ที่ปรึกษาโครงงาน

บทคัดย่อ

การสร้างกราฟเพื่อแสดงความสัมพันธ์ระหว่างสมาชิกในกลุ่มก่อการร้ายโดยใช้ social graph ร่วมกับการวิเคราะห์เครือข่ายทางสังคม (social network analysis) สามารถช่วยเจ้าหน้าที่ทางกฎหมายให้สามารถเข้าใจความสัมพันธ์ภายในขบวนการก่อการร้ายในเชิงลึกได้ดียิ่งขึ้น นอกเหนือจากข้อมูลทางกระบวนการสอบสวนคดีอาชญากรรมของหน่วยงานราชการแล้ว แหล่งข้อมูลที่มีประโยชน์อีกแหล่งหนึ่งสำหรับพัฒนากราฟความสัมพันธ์ของเครือข่ายอาชญากรรมนั่นคือข้อมูลจากข่าวสาธารณะ ในบทความนี้ได้พัฒนาเว็บแอปพลิเคชันสำหรับการวิเคราะห์เครือข่ายอาชญากรรมจากข่าวด้วย social graph โดยแอปพลิเคชันประกอบด้วยสองส่วน คือ การวิเคราะห์ข้อความและการวิเคราะห์กราฟ ในส่วนการวิเคราะห์ข้อความ มีการรวบรวมข่าวอาชญากรรม โดยในข่าวแต่ละฉบับจะมีการแท็กชื่อกลุ่มก่อการร้าย ชื่อบุคคลและเหตุการณ์ที่สำคัญด้วยตนเองโดยใช้ Doccano ซึ่งข่าวที่ติดแท็กจะถูกใช้เป็นคลังข้อมูลสำหรับการ train เพื่อสร้างโมเดล Conditional Random Fields (CRFs) สำหรับการติดแท็กให้แก่คำที่เราสนใจ โดยผลลัพธ์ที่ได้คือ ความสัมพันธ์ที่เกี่ยวข้องกันระหว่างคำที่ถูกแท็ก ถูกสกัดและขุดหา นอกจากนี้ในส่วนการวิเคราะห์กราฟความสัมพันธ์ของคำเหล่านั้นจะถูกป้อนเข้าสู่ NetworkX ซึ่งใช้ในการสร้าง social graph จากนั้นกราฟจะถูกนำมาแสดงบนเว็บแอปพลิเคชันสำหรับการสร้างภาพและวิเคราะห์โดยใช้ Django จากผลการสำรวจพบว่าความสัมพันธ์ระหว่างคำเหล่านี้สามารถสกัดออกมาได้จากข่าวสารที่เป็นข้อความและสามารถนำไปสร้างกราฟความสัมพันธ์ของเครือข่ายของอาชญากรรมได้

Abstract

Building and visualizing relationships among members in a criminal community using a social graph in combination with the principle of social network analysis could facilitate law enforcement officers to gain better insight of criminal social network structures. In addition to official information from a criminal investigation process by government agencies, another potential source of data for developing a criminal social graph is from public news data. To explore this area of studies, in this paper, a Web application for criminal social network analyses from news stories using social graphs is developed. The application consists of two modules, text mining and graph analysis. In the text mining module, a collection of criminal news stories was collected. In each news document, names of organization, person, and event are manually tagged by using Doccano. The tagged news collection was next used as a training corpus to build a Conditional Random Fields (CRF) model for tagging more news documents. Finally, possible relationships among those tagged criminal-related entities were extracted and mined. In the graph analysis module, those mined relationships are fed into NetworkX by which social network graphs are constructed. Then, the social graphs are displayed on a Web application for visualizations and analyses using Django. The exploratory result confirms that relationships among the entities can be extracted from textual news data from which criminal social networks can be constructed.

กิตติกรรมประกาศ

การทำเว็บแอปพลิเคชันสำหรับวิเคราะห์เครือข่ายผู้ก่อความไม่สงบโดยใช้ข้อมูลจากข่าวภาษาไทย (A Web Application for Insurgent Network Analysis using Thai News) สามารถดำเนินการได้สำเร็จลุล่วง โดยได้รับความช่วยเหลือ และความอนุเคราะห์จากบุคคล ดังต่อไปนี้

ขอขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.ศุภชัย ไทยเจริญ อาจารย์ที่ปรึกษาโครงงานที่กรุณาเสียสละเวลาเพื่อให้ความรู้ คำแนะนำเกี่ยวกับแหล่งค้นคว้าข้อมูลที่เป็นประโยชน์ รวมถึงช่วยแนะนำแนวทางในการแก้ไขปัญหาและแก้ไขข้อบกพร่องที่เกิดขึ้นระหว่างการจัดทำโครงงาน

ขอขอบคุณภาควิชาวิทยาการคอมพิวเตอร์ที่สนับสนุนสถานที่และอุปกรณ์ในการจัดทำโครงงาน รวมถึงขอขอบพระคุณอาจารย์ทุกๆท่านในภาควิชาวิทยาการคอมพิวเตอร์ที่กรุณาให้ความรู้ คำปรึกษาและคำแนะนำที่เป็นประโยชน์ ทำให้คณะผู้จัดทำสามารถแก้ไขโครงงานได้สมบูรณ์มากยิ่งขึ้น

ขอขอบพระคุณครอบครัวของคณะผู้จัดทำ รวมถึงเพื่อนๆ ที่ได้ให้การช่วยเหลือและสนับสนุนมาโดยตลอด รวมถึงให้ความห่วงใยและเป็นกำลังใจสำคัญในการทำโครงงานเสมอมา

สุดท้ายขอขอบพระคุณผู้ที่มีส่วนเกี่ยวข้องทุกๆ ท่านที่ทำให้โครงงานนี้สำเร็จลุล่วงไปได้ด้วยดี คณะผู้จัดทำหวังเป็นอย่างยิ่งว่าโครงงานชิ้นนี้จะเป็นประโยชน์ต่อผู้ที่สนใจศึกษา และสามารถเป็นส่วนหนึ่งสำหรับผู้ที่สนใจจะนำไปพัฒนาต่อยอดในอนาคตต่อไป

คณะผู้จัดทำ

สารบัญ

บทคัดย่อ ..... **ก**

Abstract ..... **ข**

กิตติกรรมประกาศ ..... **ค**

สารบัญ ..... **ง**

สารบัญรูปภาพ ..... **ช**

สารบัญตาราง ..... **ฌ**

บทที่ 1 บทนำ **1**

1.1 ที่มาและความสำคัญ 1

1.2 วัตถุประสงค์ของงานวิจัย 2

1.3 ขอบเขตของงานวิจัย 2

1.4 ประโยชน์ที่คาดว่าจะได้รับ 3

บทที่ 2 องค์ความรู้ที่เกี่ยวข้อง **4**

2.1 แบบจำลองทางสถิติคอนดิชันนอลแรนดอมฟิลด์ส 4

2.2 การสกัดชื่อเฉพาะจากข้อความ 5

2.3 หลักไวยกรณ์ภาษาไทยที่เกี่ยวข้อง 5

2.4 การหาความสัมพันธ์ของข้อมูลจากข้อมูลชื่อเฉพาะโดยใช้กราฟ 7

2.5 งานวิจัยที่เกี่ยวข้อง 8

2.5.1 Thai Named Entity Recognition Using Bi-LSTM-CRF with Word and Character Representation 8

2.5.2 Named Entity Recognition Modeling for the Thai Language from a Disjoinedly Labeled Corpus 8

2.5.3 Trends in deqi research: a text mining and network analysis 8

บทที่ 3 วิธีการดำเนินโครงงาน 10

3.1 วิธีการดำเนินงาน 10

3.2 แผนการดำเนินงานตลอดโครงงาน....11

3.3 อุปกรณ์และเครื่องมือที่ใช้ 12

3.4 ชุดข้อมูล (Data Set) ที่ใช้งานและเครื่องมือในการทำงาน 13

3.5 การออกแบบระบบ 15

3.6 ขั้นตอนการทำงานของระบบ 15

บทที่ 4 ผลการดำเนินงาน 25

4.1 ผลการทดสอบประสิทธิภาพโมเดลการรู้จำชื่อเฉพาะ....25

4.2 ผลการคำนวณค่าต่างๆของกราฟตามหลักการวิเคราะห์เครือข่ายทางสังคม....25

บทที่ 5 สรุปผลการทดลอง....27

5.1 สรุปผลการดำเนินโครงการ 27

5.2 ปัญหาและอุปสรรคในการดำเนินการ 27

5.3 ข้อเสนอแนะ....27

บรรณานุกรม 28

ภาคผนวก ก 32

ภาคผนวก ข 36

การสร้างโมเดล 36

การดึงข้อมูลจากไฟล์ html 48

การสร้างข้อมูลความสัมพันธ์จากข่าว.....51

การสร้างกราฟจากข้อมูลความสัมพันธ์จากข่าว 53

สารบัญรูปภาพ

รูปที่ 1 กราฟแสดงการเปรียบเทียบแบบจำลอง HMMs, MEMMs และ CRFs 4
รูปที่ 2 ตัวอย่างการทำ Named Entity Recognition 5
รูปที่ 3 ตัวอย่างกราฟ 7
รูปที่ 4 ตัวอย่างการทำกราฟแสดงความสัมพันธ์ระหว่างชื่อเฉพาะ 7
รูปที่ 5 ตัวอย่าง Social Network Analysis แสดงความสัมพันธ์ของ De-qi 9
รูปที่ 6 ตัวอย่างของข้อมูลข่าว 13
รูปที่ 7 ตัวอย่างการใช้ NetworkX ในการสร้างกราฟแสดงความสัมพันธ์ 13
รูปที่ 8 Bokeh เครื่องมือที่ใช้แสดงผลกราฟ 14
รูปที่ 9 เครื่องมือที่ช่วยในการแทรคคำ 14
รูปที่ 10 ข้อมูลข่าวที่ดึงมาจากเว็บไซต์หนึ่ง 15
รูปที่ 11 เครื่องมือที่ช่วยในการแทรคคำ Doccano 16
รูปที่ 12 ผลลัพธ์การทำงานของโปรแกรม Doccano 16
รูปที่ 13 ผลลัพธ์การแทรคคำ 17
รูปที่ 14 ภาพการแทรคคำหน้าและหลังของคำ 17
รูปที่ 15 ผลลัพทธ์ของการเพิ่ม tag ลงในข้อมูล 18
รูปที่ 16 รูปตัวอย่างข้อมูลไฟล์ conll2002 18
รูปที่ 17 ตารางแสดงความสัมพันธ์ของข้อมูลในรูปแบบ 1:1 19
รูปที่ 18 ตัวอย่างกราฟที่ได้จากกข้อมูล 20
รูปที่ 19 การแสดงผลกราฟความสัมพันธ์ของข้อมูลผ่าน Bokeh 22
รูปที่ 20 หน้าหลักของเว็บไซต์ 23
รูปที่ 21 หน้าอธิบายค่าตามหลักการวิเคราะห์เครือข่ายทางสังคม 23
รูปที่ 22 หน้า About 24
รูปที่ 23 หน้าการแสดงผลกราฟ 24

รูปที่ 24 ตัวอย่างการเปิดใช้งานโปรแกรม Docker.....32
รูปที่ 25 หน้าเว็บไซต์.....54
รูปที่ 26 การสร้าง Project.....54

สารบัญตาราง

ตารางที่ 1 แผนการดำเนินโครงการวิจัย 12
ตารางที่ 2 ตารางแสดงประสิทธิภาพโมเดลการรู้จำชื่อเฉพาะ 25

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญ

ในปัจจุบันคอมพิวเตอร์ได้เข้ามามีบทบาทในชีวิตประจำวันของเราอย่างมาก โดยเฉพาะเรื่องการใช้อินเตอร์เน็ตที่มีผู้ใช้งานเป็นจำนวนมาก และมีแนวโน้มว่าจะมากขึ้นเรื่อยๆ สิ่งหนึ่งที่คนนิยมใช้งานอินเตอร์เน็ต คือ การค้นหาข้อมูล เพราะมีความสะดวก และรวดเร็ว อย่างไรก็ตาม ความสำคัญของการใช้ประโยชน์จากคอมพิวเตอร์ไม่ได้จำกัดอยู่แค่การค้นหาข้อมูลต่างๆภายในอินเตอร์เน็ต แต่ยังครอบคลุมไปถึงการใช้งานตามห้องสมุดต่างๆเช่น หากเราต้องการค้นหาหนังสือก็จะมีระบบสำหรับ สืบค้นเพื่ออำนวยความสะดวก ซึ่งทั้งการค้นหาข้อมูลบนอินเตอร์เน็ต หรือการค้นหาหนังสือที่เราต้องการ ในห้องสมุด เราต้องใส่คำที่ต้องการจะค้นหาเข้าไป เพื่อให้ระบบสามารถค้นหาและดึงข้อมูลหรือ เอกสารที่เกี่ยวข้องกับสิ่งที่เราต้องการออกมาได้ โดยส่วนใหญ่ผลการค้นที่ได้จะต้องมีเอกสารที่ไม่ตรงตาม ความต้องการของเรารวมอยู่ด้วยไม่มากก็น้อย โดยที่การค้นหานั้น ไม่สามารถแยกความแตกต่างระหว่าง ชื่อเฉพาะ (named entity) และคำนามทั่วไป (common noun) ในปัจจุบันนี้ความต้องการในการเข้า ถึงข้อมูลเริ่มมีเพิ่มมากขึ้น เราจึงจำเป็นต้องพัฒนาความสามารถของคอมพิวเตอร์ให้ดียิ่งขึ้นในการที่จะเข้าใจภาษามนุษย์ได้ ด้วยการประมวลผลภาษาธรรมชาติหรือ NLP (Natural Language Processing) จึงเข้ามามีบทบาทสำคัญในการจัดการกับข้อมูลทางภาษาเพื่อทำการแยกระหว่าง ชื่อเฉพาะ (named entity) และคำนามทั่วไป (common noun)

เริ่มแรกนั้นการทำ Name Entity Recognition (NER) หรือการรู้จำชื่อเฉพาะมีการพัฒนามา อย่างต่อเนื่องมานานกว่า 15 ปีโดยได้มีการริเริ่มโดยเป็นการทำให้คอมพิวเตอร์นั้นสามารถแยกแยะได้ว่า ข้อมูลที่พบนั้นข้อมูลใดเป็นชื่อเฉพาะ(named entity) และแยกประเภทของคำ โดยมีการใช้กันอย่าง แพร่หลายในเรื่องของ NLP ซึ่งคือการจำแนกประเภทของคำออกจากประโยคที่ได้รับการแยกคำ ออกมาจากประโยค เช่น ชื่อบุคคล (PERSON) เหตุการณ์ (EVENT)และชื่อสถานที่ (LOCATION) ในการสกัดข้อมูล (Information Extraction) การรู้จำชื่อเฉพาะ ที่ถูกต้องนั้นสามารถทำให้เราสามารถกรองข้อมูล ที่มีอยู่จำนวนมากได้อย่างถูกต้อง โดยไม่ต้องใช้เวลามาก ในการเข้าถึงเนื้อหาของข้อมูลนั้นๆเพราะฉะนั้น

การศึกษาวิธีเพื่อให้คอมพิวเตอร์ รู้จำชื่อเฉพาะ (Name Entity Recognition)จะช่วยทำให้ การประมวลผล ของภาษาธรรมชาติมีประสิทธิภาพมากยิ่งขึ้น โดยงานวิจัยเกี่ยวกับการรู้จำชื่อเฉพาะ ในภาษาต่างประเทศ ได้มีการทำการวิจัยกันมานานแล้วสำหรับ ภาษาตะวันตก โดยเฉพาะภาษาอังกฤษ ทำให้ค่าความถูกต้อง ในการรู้จำชื่อเฉพาะของภาษาอังกฤษ ค่อนข้างสูง ในขณะที่ภาษา ในแถบเอเชีย เพิ่งเริ่มมีการศึกษาเมื่อไม่นานมานี้ เช่น ภาษาญี่ปุ่น ภาษาจีน สำหรับปริมาณงานวิจัยเรื่อง การรู้จำชื่อเฉพาะของภาษาไทย ในปัจจุบันนั้น พบว่ายังมีไม่มากและประสิทธิภาพยังไม่ดีเท่า ที่ควรที่จะนำไป ประยุกต์ใช้ได้จริง

ดังนั้นผู้ทำวิจัยจึงมีความสนใจที่จะศึกษาและพัฒนาระบบวิธีการที่ช่วยในการสกัดชื่อเฉพาะ ภาษาไทย เพื่อเป็นประโยชน์ต่อการพัฒนาระบบงานที่เกี่ยวข้องต่อไป รวมไปถึงการนำชื่อเฉพาะต่างๆ ที่ได้มาจากการทำ Name Entity Recognition (NER) มาทำการสร้าง Social Network Analysis (SNA) และนำไปใช้กับข้อมูลที่มีความหลากหลายและกว้างขวาง เพื่อให้ได้ผลลัพธ์ที่สามารถแสดงให้เห็นถึง ความสัมพันธ์ระหว่างชื่อเฉพาะได้

1.2 วัตถุประสงค์ของงานวิจัย

1. ศึกษาและวิเคราะห์รูปแบบและโครงสร้างทางไวยากรณ์ของภาษาไทยออกมาในรูปแบบของประโยค
2. ศึกษาหลักการในการสกัดชื่อเฉพาะจากข้อความสำหรับสกัดคำประเภท บุคคล เหตุการณ์ และสถานที่
3. เพื่อประยุกต์ใช้หลักการวิเคราะห์เครือข่ายความสัมพันธ์ของข้อมูล และสามารถสร้างแอปพลิเคชันสำหรับการวิเคราะห์เครือข่ายอาชญากรรมได้สะดวกยิ่งขึ้น
4. เป็นแนวทางให้ผู้ที่สนใจมาพัฒนาต่อยอดได้

1.3 ขอบเขตของงานวิจัย

1. ทำการศึกษาและพัฒนาระบบในการสกัดชื่อเฉพาะตามประเภท ได้แก่ ชื่อบุคคล เหตุการณ์ และชื่อสถานที่จากชุดข้อมูลข่าว
2. เน้นการแยกชื่อเฉพาะบุคคล องค์กร สถานที่ และเหตุการณ์
3. ศึกษาการหาความสัมพันธ์ของข้อมูล (Social Network Analysis) ระหว่างชื่อเฉพาะ (Named Entity)

4. ใช้ข้อมูลข่าวภาษาไทยเกี่ยวกับปัญหาการก่ออาชญากรรม
5. ทำเว็บแอปพลิเคชันสำหรับแยกชื่อเฉพาะจากข้อมูลข่าว และนำมาแสดงเป็น Social Network Analysis

1.4 ประโยชน์ที่คาดว่าจะได้รับ

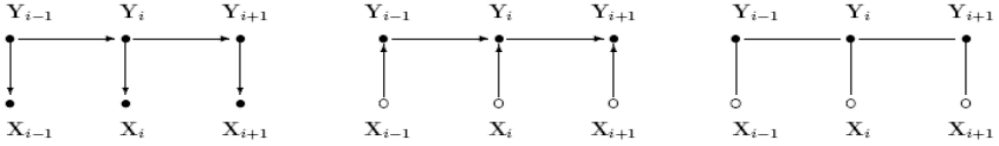
1. สามารถแยกประเภทของคำเฉพาะออกมาได้อย่างถูกต้องและมีความแม่นยำ
2. ผู้ทำวิจัยสามารถเข้าใจความสัมพันธ์ของการรู้จำชื่อเฉพาะในข้อมูลได้อย่างรวดเร็วและมีประสิทธิภาพ
3. สามรถหาความสัมพันธ์ระหว่างบุคคล องค์กร เหตุการณ์ และสถานที่ในข้อมูลออกมาเป็นกราฟที่สามารถแสดงให้เห็นความสัมพันธ์ได้
4. เป็นแนวทางในการศึกษาการรู้จำชื่อเฉพาะในภาษาไทย

บทที่ 2
องค์ความรู้ที่เกี่ยวข้อง

2.1 แบบจำลองทางสถิติคอนดิชันนอลแรนดอมฟิลด์ส [1]

แบบจำลองทางสถิติคอนดิชันนอลแรนดอมฟิลด์ส์ (CRFs) เป็นแบบจำลองที่ยอมรับกันในปัจจุบันว่ามีประสิทธิภาพมากกว่าแบบจำลอง Hidden Markov Models (HMMs) ซึ่งมีลักษณะเป็น Generative models ที่อาศัยค่าความน่าจะเป็นรวม (joint probability) ระหว่างข้อมูลเข้ากับผลหรือเลเบลที่ออกมาจึงมีปัญหาว่า ไม่สามารถจับความสัมพันธ์ระหว่างคุณสมบัติต่างๆที่เกี่ยวข้องกันในข้อมูลเข้าได้ เพราะคุณสมบัติต่างๆ เป็นอิสระต่อกัน แบบจำลองที่ลดปัญหา ดังกล่าวได้คือแบบจำลอง Maximum Entropy Markov Models (MEMMs) ที่มีลักษณะเป็น Discriminative models ที่อาศัยค่าความน่า จะเป็นแบบเงื่อนไข (conditional probability) ของผลหรือสายของเลเบล (label sequence) แบบต่างๆ เมื่อพบสายข้อมูลเข้า (observation sequence) ซึ่งทำให้จับความสัมพันธ์ระหว่างคุณสมบัติต่างๆ ในข้อมูลที่พบได้แต่แบบจำลอง MEMMs ก็ยังพบปัญหาที่เรียกว่า label biased เพราะการตัดสินผล ณ สภาวะใดขึ้นกับสภาวะปัจจุบันและ สายข้อมูลที่พบ (observation sequence) เท่านั้น สภาวะอื่นๆ ทั้งหมดในแบบจำลองไม่มีผลต่อการคำนวณ ค่าความน่าจะเป็น แบบจำลอง CRFs ที่เสนอโดย Lafferty et al. (2001) เป็นแบบจำลองที่ลดปัญหาที่ว่านี้

Conditional Random Fields (CRFs) คือ Discriminative models รูปแบบหนึ่งที่ถูกยอมรับกันในปัจจุบันว่ามีประสิทธิภาพมาก ที่ใช้พยากรณ์เหตุการณ์ใดๆ (state) มีลักษณะเป็นแบบจำลองกราฟไม่ระบุทิศทาง (undirected graphical model) โดยที่ CRFs หาคาความน่าจะเป็นของ เลเบลถัดไป โดยนำเลเบลก่อนหน้าทั้งหมดที่มีลำดับเหตุการณ์เป็นเงื่อนไขมาคำนวณด้วย น้ำหนักของคุณสมบัติต่างๆ จากสภาวะที่ต่างกันจึงมีการปรับสมดุลให้ค่าไม่เอนเอียงไปสภาวะใดสภาวะหนึ่ง



รูปที่ *1* กราฟแสดงการเปรียบเทียบแบบจำลอง *HMMs, MEMMs* และ *CRFs*

(ที่มา : http://www.arts.chula.ac.th/~ling/thesis/2553MA-Ling-Nutcha.pdf หน้าที่ 14)

2.2 การสกัดชื่อเฉพาะจากข้อความ [2]

คำว่า Named Entity ถูกใช้กันอย่างแพร่หลายในเรื่องของ Natural Language Processing (NLP) ซึ่งเป็นศาสตร์ที่เกี่ยวกับการ Extract ข้อมูลออกจาก Unstructed text ยกตัวอย่างเช่น เอาชื่อของ คน, องค์กร ออกมาจากข่าวในหนังสือพิมพ์ เป็นต้น การสกัดชื่อเฉพาะจากข้อความ (Named Entity Recognition หรือ NER) คือ การสกัดนิพจน์เฉพาะหรือ ชื่อเฉพาะในประโยคออกมา ยกตัวอย่างเช่น มีประโยค "นายAจะไปเดินเล่นที่หนองคาย พร้อมกับนั่งเรือข้ามไปประเทศลาว" จากประโยคข้างต้นจะมีคำเฉพาะ คือ นาย A อยู่ในหมวดหมู่ของบุคคล หนองคาย และ ประเทศลาว อยู่ในหมวดหมู่ของสถานที่

รูปที่ *2* ตัวอย่างการทำ *Named Entity Recognition*

(ที่มา : https://confusedcoders.com/wp-content/uploads/2019/11/image2.png)

2.3 หลักไวยกรณ์ภาษาไทยที่เกี่ยวข้อง [3]

การทำ Named Entity Recognition ขึ้นมาใหม่สามารถใช้ sklearn-crfsuite ซึ่งเป็นโมดูล สำหรับทำ CRF model โดยเชื่อมกับ CRFsuite อีกที และนำคลังข้อมูล CoNLL2002 ไป train กับโมเดล Conditional Random Fields (CRFs) ซึ่งเป็นโมเดลยอดนิยมในฝั่ง machine learning

โครงสร้างของภาษาไทยมีลักษณะหลายประการที่ทำให้ยากต่อการรู้จำชื่อเฉพาะ (Chanlekha and Kawtrakul, 2004) ได้แก่

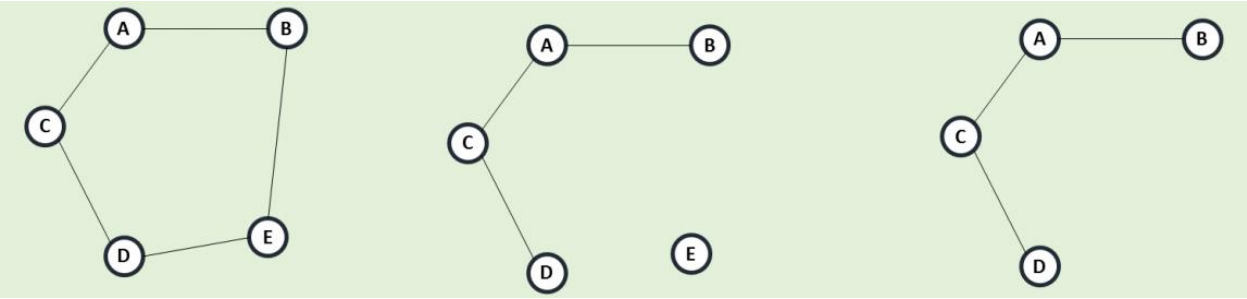
1. ภาษาไทยไม่มีข้อมูลบ่งบอกถึงชื่อเฉพาะ เช่นในภาษาอังกฤษที่ใช้อักษรตัวพิมพ์ใหญ่เสมอเมื่อกล่าวถึงชื่อเฉพาะ ทำให้ยากต่อการแยกแยะระหว่างชื่อเฉพาะและคำทั่วไป เช่น ฉัน อยากเลยไปที่เลย “เลย” คำแรกเป็นคำกริยาหมายถึงเกินจุดที่กำหนด ในขณะที่ “เลย” คำที่สอง เป็นชื่อของจังหวัด หรือ

สมชาย นี่ช่างสมชายสมชื่อจริง ๆ “สมชาย” คำแรกเป็นชื่อบุคคล ในขณะ ที่ “สมชาย” ที่สองเป็นวลี มาจาก “สม” **ซึ่งเป็น**คำวิเศษณ์ หมายถึง เหมาะ และ “ชาย” เป็นคำนาม หมายถึง ผู้ชาย เป็นต้น นอกจากนี้ภาษาไทยไม่มี การใช้อักขระพิเศษสำหรับชื่อเฉพาะที่ถ่ายถอด เสียงมาจากภาษาต่างประเทศ เช่น จอห์น ไมเคิล เป็นชื่อที่ถ่ายถอดเสียงโดยใช้ตัวอักษรภาษาไทย ปกติซึ่งต่างจากภาษาญี่ปุ่น ที่มีการใช้อักษรคาตาคานะ (Katakana) เพื่อบ่งบอกวาเป็นชื่อ เฉพาะที่ถ่ายถอดเสียงมา

2. ภาษาไทยไม่มีการเว้นวรรคหรือใช้อักษรพิเศษในการแบ่งคำ ทำให้มีปัญหาในการตัด แบ่งคำ ซึ่งถ้าหากตัดแบ่งคำผิดก็จะส่งผลถึงการรู้จำชื่อเฉพาะด้วย เช่น ประโยค “คุณตูนบอก” ที่ ตัดคำถูกคือ คุณ-ตูน-บอก โดย “ตูน” เป็นชื่อเฉพาะประเภทบุคคล แต่หากตัดคำเป็น คุณ-ตู-นบอก ชื่อเฉพาะจะกลายเป็น “ตู” ซึ่งผิด
3. ลักษณะการสร้างชื่อเฉพาะไม่มีหลักเกณฑ์ที่แน่นอน สามารถสร้างขึ้นใหม่ด้วยคำใด ก็ได้ทำให้ยากต่อการสร้างกฎ เช่น บริษัท กระเรียนทอง จำกัด “กระเรียนทอง” เป็นชื่อของนกชนิด หนึ่ง และได้นำมาใช้เป็นชื่อขององค์กรหรือชื่อสิ่งต่าง ๆ เช่น ส้ม นุ้น ฟิล์ม ก็สามารถนำมาตั้งเป็นชื่อบุคคลได้เช่นกัน นอกจากนี้ชื่อเฉพาะแต่ละประเภทยังอาจซ้ำกันได้ทำให้ยากต่อการระบุชนิดของชื่อเฉพาะ เช่น “อ่างทอง” เป็นได้ทั้งชื่อจังหวัด และชื่อบุคคล เป็นต้น
4. ลักษณะงานเขียนของภาษาไทยที่จะกล่าวถึงชื่อเต็มของชื่อเฉพาะในครั้งแรกแล้ว จากนั้นเมื่อจะกล่าวถึงชื่อเฉพาะนั้นอีกจะใช้ชื่อย่อ หรือกล่าวโดยไม่มีคำบ่งชี้ทำใหเกิดความ กำกวมระหว่างชื่อเฉพาะและคำนามทั่วไปได้เช่น “แหล่งข่าวจากบริษัท ห้างสรรพสินค้าโรบินสัน จำกัด (มหาชน) เปิดเผยกับ “ฐานเศรษฐกิจ” ว่า ห้างสรรพสินค้าโรบินสัน ได้ตัดสินใจ... สำหรับโรบินสัน สาขาสีลมเปิดให้บริการมากว่า 24 ปี” เป็นต้น

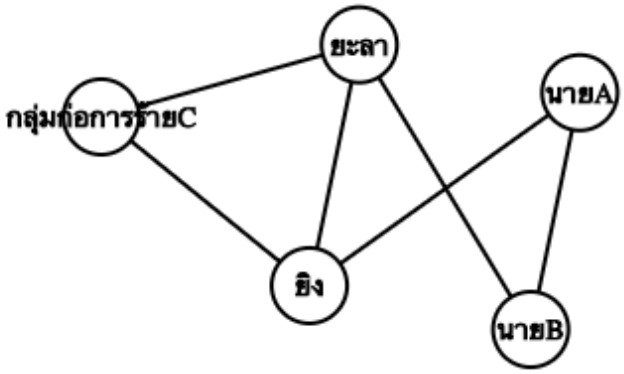
2.4 การหาความสัมพันธ์ของข้อมูลจากข้อมูลชื่อเฉพาะโดยใช้กราฟ [4][5]

กราฟ (Graph) เป็นโครงสร้างที่แทนความสัมพันธ์ระหว่างข้อมูลที่มีข้อมูลจำกัดน้อยความสัมพันธ์ที่ไม่มีจำกัดว่าต้องเป็นตามลำดับชั้น หรือข้อมูลต้องเรียงจากซ้ายไปขวา โดยกราฟจะประกอบไปด้วยโหนด (nodes) และเส้นที่ใช้เชื่อมระหว่างโหนด (edges)



รูปที่ *3* ตัวอย่างกราฟ

การหาความสัมพันธ์ของข้อมูลจากข้อมูลชื่อเฉพาะหรือ Social Network Analysis คือ การหารูปแบบความสัมพันธ์ระหว่างข้อมูลชื่อเฉพาะ (named entity) ว่าข้อมูลชื่อเฉพาะแต่ละตัวนั้นมีความสัมพันธ์กับมูลชื่อเฉพาะตัวอื่นๆอย่างไรบ้างในรูปแบบของกราฟ ด้วยการกำหนดให้ Nodes แทนด้วยชื่อเฉพาะ และเส้น Edges แทนความสัมพันธ์ของข้อมูลระหว่างชื่อเฉพาะ



รูปที่ *4* ตัวอย่างการทำกราฟแสดงความสัมพันธ์ระหว่างชื่อเฉพาะ

2.5 งานวิจัยที่เกี่ยวข้อง

2.5.1 Thai Named Entity Recognition Using Bi-LSTM-CRF with Word and Character Representation [6]

ทดสอบแนวคิดโดยการประยุกต์ใช้ Bi-LSTM-CRF ด้วยระดับของคำ เพื่อนำมาแก้ปัญหาของ Thai Named Entity Recognition รวมถึงมีการสร้าง Neural Network รวมกับ CRF เพื่อเรียนรู้ลำดับของคำ และเรียนรู้การแยกคำเพื่อให้ทราบ Named Entity Recognition ซึ่งโมเดลได้ผลลัพธ์นี้สามารถให้ precision, recall and F1 อยู่ที่ 91.79%, 91.51% และ 91.65% ตามลำดับ

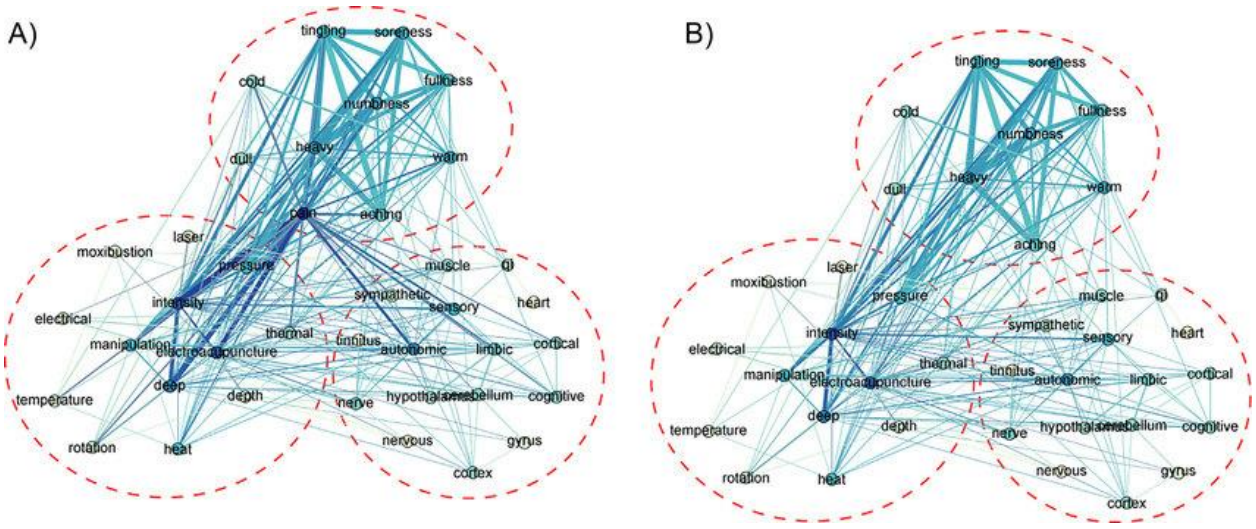
2.5.2 Named Entity Recognition Modeling for the Thai Language from a Disjoinedly Labeled Corpus [7]

เป็นแนวคิดที่จะทำ Named Entity Recognition (NER) โดยการทำโมเดล Bi-LSTM มาร่วมกับ part-of-speech (POS) เพื่อเพิ่มความถูกต้องในการทำ NER เพราะทางผู้ทำคิดว่า POS จะมีผลต่อการทำ NER โดยจะนำ มาเปรียบเทียบกับ model CRF โดยผลลัพธ์ที่ได้นั้น model Bi-LSTM นั้นได้คะแนนที่ดีกว่า โดยถ้าเพิ่มการทำ POS เข้าไปทำให้model ได้ผลลัพธ์ดีขึ้น

2.5.3 Trends in deqi research: a text mining and network analysis [8]

งานวิจัยชิ้นนี้ทำเพื่อตรวจสอบหาปัจจัยที่เกี่ยวข้องในการเกิดสภาวะ deqi โดยมีการใช้ keyword ต่างๆ เช่น De-qi, Deqi หรือ acupuncture sensation(ความรู้สึกจากการฝังเข็ม) ที่เอามาจากบทคัดย่อของ 148 บทความ ข้อมูลที่ได้มาจากบทคัดย่อถูกนำมา Text Mining และ Network Analysis โดยมีการใช้เทคนิค ต่างๆ อันได้แก่

1. Tf-idf เพื่อหาความถี่ของคำซึ่งจะนำมาซึ่งความสำคัญของคำ
2. Network construction เพื่อหาแหล่งที่มาของการเกิด Deqi โดยจะขึ้นอยู่กับ Tf-idf ใน เงื่อนไข tf-idf(a|b) หรือการหาจำนวน คำศัพท์ a อยู่จำนวนเท่าใดในขณะที่มีคำศัพท์ b อยู่ ด้วย โดยให้ a แทนเป็น ‘ความรู้สึก’ และ b แทนเป็น ‘สิ่งที่ใช้ในการกระตุ้น’
3. Clustering โดยใช้ machine learning ด้วย software Weka และ EM Algorithm



รูปที่ *5* ตัวอย่าง *Social Network Analysis* แสดงความสัมพันธ์ของ *De-qi*

โดยแยกได้ 37 รายการ 3 กลุ่ม

1. ความรู้สึกของการฝังเข็ม(ARS) 10 รายการ (ปวด, น่าเบื่อ, หนัก, รู้สึกเสียวซ่า, ความรุนแรง, ความ แน่น, อบอุ่น, ชา, ความเจ็บปวด และความหนาวเย็น)

2. การแทรกแซง 13 รายการ (หมุน, เลเซอร์, ความเข้ม, ความลึก, ความร้อน, ความดัน, การจัดการ, การฝังเข็มด้วยไฟฟ้า (EA), อุณหภูมิ, ความร้อน, การรมควัน, ความลึก และไฟฟ้า)

3. กลไกทางชีวภาพ 15 รายการ (cognitive, cortical, cortex, sympathetic, limbic, hypothalamus, muscle, sensory, cerebellum, autonomic nerve system, nerve, gyrus, tinnitus and heart)

บทที่ 3

วิธีการดำเนินโครงงาน

3.1 วิธีการดำเนินงาน

1. เริ่มต้นและวางแผนโครงการ
    - ประชุมเพื่อเลือกหัวข้อโครงงาน
    - หาข้อมูลศึกษาแนวคิดและทฤษฎีที่เกี่ยวข้อง
    - ศึกษาเทคโนโลยีที่จำเป็นมาใช้
2. รวบรวมและจัดการข้อมูล
    - หา Dataset และทำความเข้าใจข้อมูล
    - ทำการ Clean ข้อมูล
    - จัดรูปแบบข้อมูลให้เหมาะสมต่อการใช้งานในขั้นถัดไป
3. การออกแบบการทำงาน
    - กำหนดวัตถุประสงค์และขอบเขตของระบบ
    - ใช้งานโมเดลและปรับ**ค่า**พารามิเตอร์
    - นำโมเดลที่สร้างขึ้นไปทำ NER เพื่อให้มีข้อมูลที่จะถูกนำไปใช้มากขึ้น
    - ตรวจสอบข้อมูลและนำไปใช้ต่อในขั้นต่อไป
4. การพัฒนาและปรับปรุงประสิทธิภาพ
    - ทำการพัฒนาความแม่นยำ
    - ทำการทดสอบและปรับปรุง
5. การสร้างกราฟเพื่อแสดงความสัมพันธ์
    - นำข้อมูลชื่อเฉพาะมาทำการจัดการให้อยู่ในรูปแบบที่เหมาะสม
    - นำข้อมูลมาสร้างกราฟโดยใช้ NetworkX
    - นำกราฟที่ได้ไปสร้างด้วย Bokeh เพื่อช่วนเพิ่มประสิทธิภาพในการแสดงผล
6. สรุปผลและเผยแพร่งานวิจัย
    - ทำการปรับปรุงระบบ
    - จัดทำคู่มือและเอกสารที่เกี่ยวข้อง
    - สรุปผลการดำเนินงาน
    - นำเสนอระบบที่สมบูรณ์

3.2 แผนการดำเนินงานตลอดโครงงาน

แผนการดำเนินการวิจัย	พ.ศ. 2562							พ.ศ. 2563				
	มิ.ย.	ก.ค.	ส.ค.	ก.ย.	ต.ค.	พ.ย.	ธ.ค.	ม.ค.	ก.พ.	มี.ค.	เม.ย.	พ.ค.
1. วางแผนโครงการ												
เลือกหัวข้อโครงงาน												
ศึกษาแนวคิดและทฤษฎี												
ศึกษาเทคโนโลยีที่นำมาใช้												
2. รวบรวมและจัดการข้อมูล												
ทำการจัดหา dataset และทำการจัดเตรียมข้อมูลให้พร้อมใช้งานในขั้นต่อไป												
3. การแยกประเภทของคำ												
สร้าง Model ที่ทำการจำแนกคำ												
ปรับปรุงประสิทธิภาพการทำงาน												
4. จับความสัมพันธ์ของข้อมูล												
สร้างความสัมพันธ์ของข้อมูล												
สร้างกราฟแสดงความสัมพันธ์ของข้อมูล												
5. แอพลิเคชั่นรองรับข้อมูล												
สร้างแอพลิเคชั่นรองรับข้อมูล												
6. สรุปและประมวลผล												

ทำการปรับปรุงระบบ												
จัดทำข้อมูลและเอกสารที่เกี่ยวข้อง												
สรุปผลการดำเนินงาน												
นำเสนอระบบที่สมบูรณ์												

ตารางที่ *1* แผนการดำเนินโครงการวิจัย

3.3 อุปกรณ์และเครื่องมือที่ใช้

1. ฮาร์ดแวร์
    - คอมพิวเตอร์ intel® CORETM i5-3470 @ 3.20GHz 8.00 GB of RAM
    - iMac intel® CORE i5 @2.8GHZ 8.00 GB of RAM
2. ซอฟต์แวร์และเครื่องมือที่ใช้
    - Jupyter Notebook
    - Google Collaboratory
    - Pycharm
    - Pythainlp และ Deepcut เครื่องมือที่ใช้ในการตัดคำ
    - NetworkX [9] เป็น Python Package ทีใช้ในการสร้างกราฟจากข้อมูล
    - Doccano [10] เครื่องมือ open source ที่มีคุณสมบัติในการจำแนกชนิดของคำ
    - Bokeh [11] เครื่องมือสำหรับการ plot ข้อมูลในรูปแบบ interactive
3. ภาษาที่ใช้
    - Python
    - Django

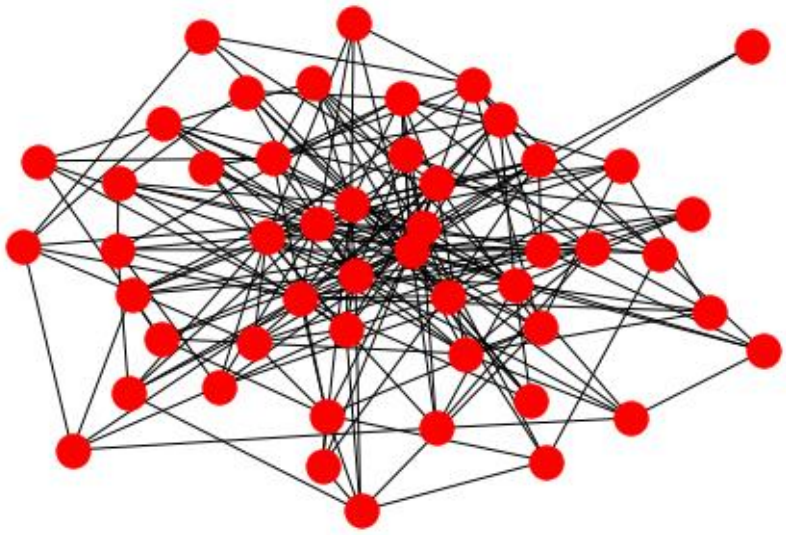
3.4 ชุดข้อมูล (Data Set) ที่ใช้งานและเครื่องมือในการทำงาน

1. ชุดข้อมูลที่ใช้คือ ข่าวจากเว็บไซต์หนึ่ง

```
news_id: 113621.txt

title: ยะลาตรวจเข้ม 4 มุมเมืองหวั่นคนร้ายก่อเหตุครบ 8 ปีปล้นปืน

content: ยะลา – ตำรวจ ทหาร ยะลา เตรียมเข้ม หวั่นการก่อเหตุรับครบรอบ 8 ปี ปล้นปืน 413
กระบอก ค่ายปิเหล็ง อ.เจาะไอร้อง จ.นราธิวาส (วันเสียงปืนแตก) วันนี้
(4 ม.ค.) กำลังเจ้าหน้าที่ตำรวจ ทหาร และ ฝ่ายปกครอง ในพื้นที่ 3 จังหวัดชายแดนใต้
ได้สนธิกำลังคุมเข้มมาตรการรักษาความปลอดภัยในพื้นที่ หวั่นกลุ่มผู้ก่อความไม่สงบ
อาศัยจังหวะในช่วงครบรอบ 8 ปี ปล้นปืน 413 กระบอก ค่ายปิเหล็ง อ.เจาะไอร้อง
จ.นราธิวาส (วันเสียงปืนแตก) ก่อเหตุรายในพื้นที่ ในส่วนของ จ.ยะลา
เจ้าหน้าที่ได้สนธิกำลังตั้งจุดตรวจจุดสกัดตามถนนสายหลัก และสายรอง
พร้อมกับมีการลาดตระเวนในพื้นที่เป้าหมาย และการดูแลความปลอดภัยในย่านชุมชน
และสถานที่ราชการ โดยผู้บังคับบัญชาของทุกหน่วยได้เน้นย้ำให้มีการตรวจรถทุกชนิด
อาวุธต่างๆ และ บุคคลเป้าหมายอย่างเข้มงวดที่สุด นายเดชรัฐ สิมศิริ
ผู้วาราชการจังหวัดยะลา ได้กำชับไปยังฝ่ายความมั่นคงอำเภอทุกอำเภอ ให้จัดกำลัง
อส.และประสานกำนัน ผู้ใหญ่บ้าน ดูแลสถานที่ราชการอย่างใกล้ชิด ในขณะที่ พล.ต.ต.พีระ
บุญเลี้ยง ผู้บังคับการตำรวจภูธรจังหวัดยะลา ได้สั่งการให้ทุก สภ.เพิ่มความเข้ม
จัดกำลังตำรวจ สายตรวจออกลาดตระเวนดูแลพื้นที่เป้าหมาย ย่านชุมชน สถานที่ราชการ
และให้จัดตั้งด่านตรวจลอย เพื่อกดดันกลุ่มคนร้ายที่อาจจะลักลอบเข้ามาก่อเหตุในพื้นท
อย่างเข้มข้น สวนพื้นที่ในเขตเทศบาลนครยะลา ได้ให้ สภ.เมืองฯ
จัดกำลังเสริมประจำด่านตรวจ 4 มุมเมืองที่จะเข้ามายังในเขตเทศบาล ให้ตรวจคนรถยนต์
และรถจักรยานยนต์ที่จะเข้ามาทุกคัน ไม่มีการยกเว้น
```

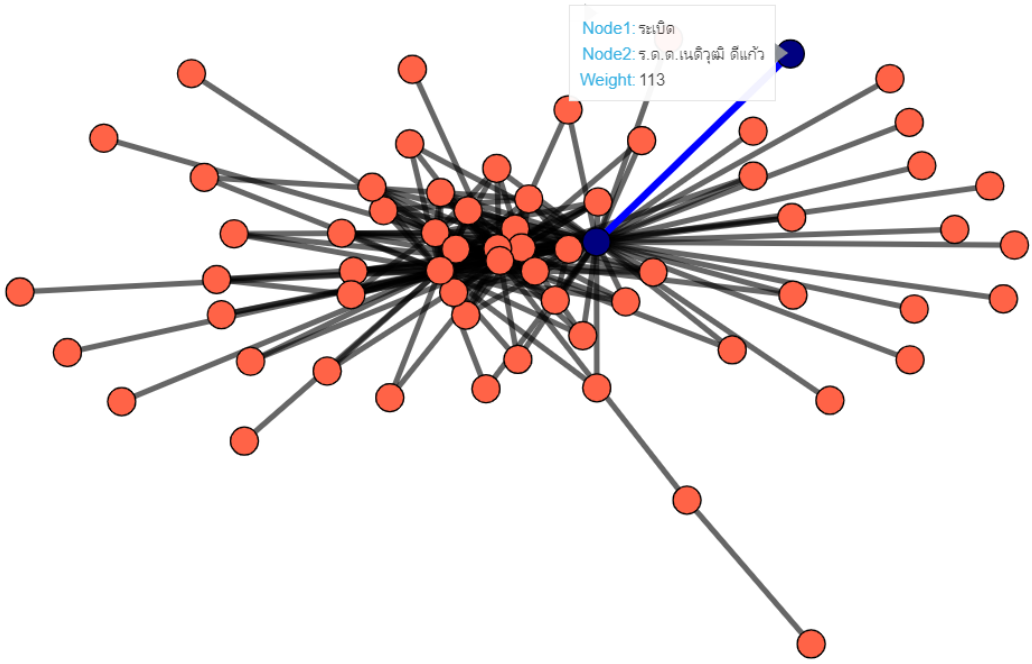
รูปที่ *6* ตัวอย่างของข้อมูลข่าว

2. ใช้ Pythainlp และ Deepcut เป็นเครื่องมือในการตัดคำ
3. ใช้ NetworkX ในการสร้างกราฟความสัมพันธ์ของข้อมูล



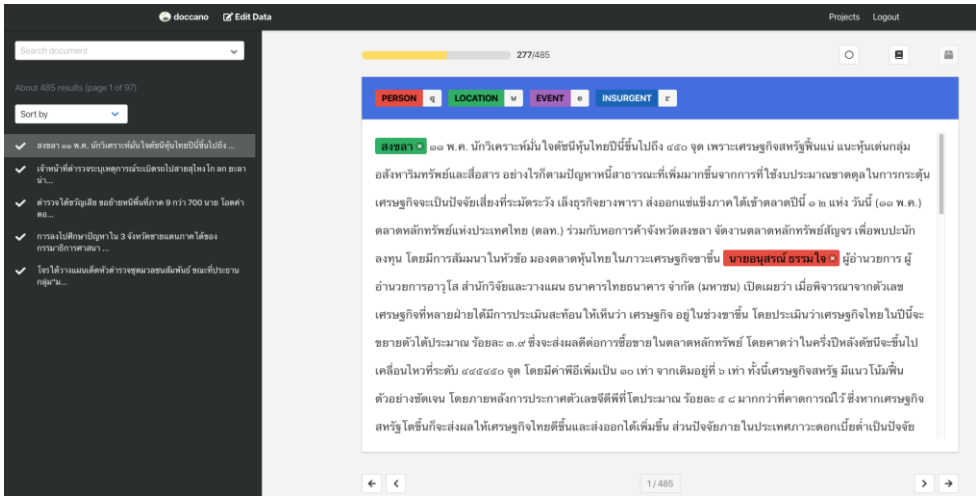
รูปที่ *7* ตัวอย่างการใช้ *NetworkX* ในการสร้างกราฟแสดงความสัมพันธ์

4. ใช้ Bokeh ในการ plot กราฟที่ได้มาจาก NetworkX ให้อยู่ในรูปแบบ interactive ที่เพิ่มความสวยงามและการแสดงผลที่หลากหลายมากขึ้น



รูปที่ *8 Bokeh* เครื่องมือที่ใช้แสดงผลกราฟ

5. ใช้ Doccano ในการสร้าง dataset เพื่อนำไปสร้างโมเดล



รูปที่ *9* เครื่องมือที่ช่วยในการแทรคคำ

3.5 การออกแบบระบบ

ขั้นตอนการออกแบบระบบมีดังนี้

1. นำเข้าข้อมูลที่ดึงจากเว็บไซต์ข่าวหนึ่งมาทำการรวมข่าวและคัดเฉพาะเนื้อหาข่าวเท่านั้น
2. ใช้โปรแกรม Doccano ในการแทรคคำว่าเป็นชื่อเฉพาะประเภทใด
3. ใช้ Model CRF ในการแยกประเภทคำของข่าวที่ใส่มาของข่าวที่ใส่มาเพื่อเพิ่มข้อมูลให้มีมากขึ้น
4. ใช้ Network X ในการสร้างความสัมพันธ์ที่ได้จากข่าวออกมาในรูปแบบของกราฟ
5. นำกราฟที่ได้จาก NetworkX ไปสร้างกราฟอีกรอบด้วย Bokeh เพื่อช่วยเพิ่มประสิทธิภาพในการแสดงผลของกราฟ
6. ใช้ django เพื่อรองรับ Data ที่เข้ามาจากเว็บไซต์

3.6 ขั้นตอนการทำงานของระบบ

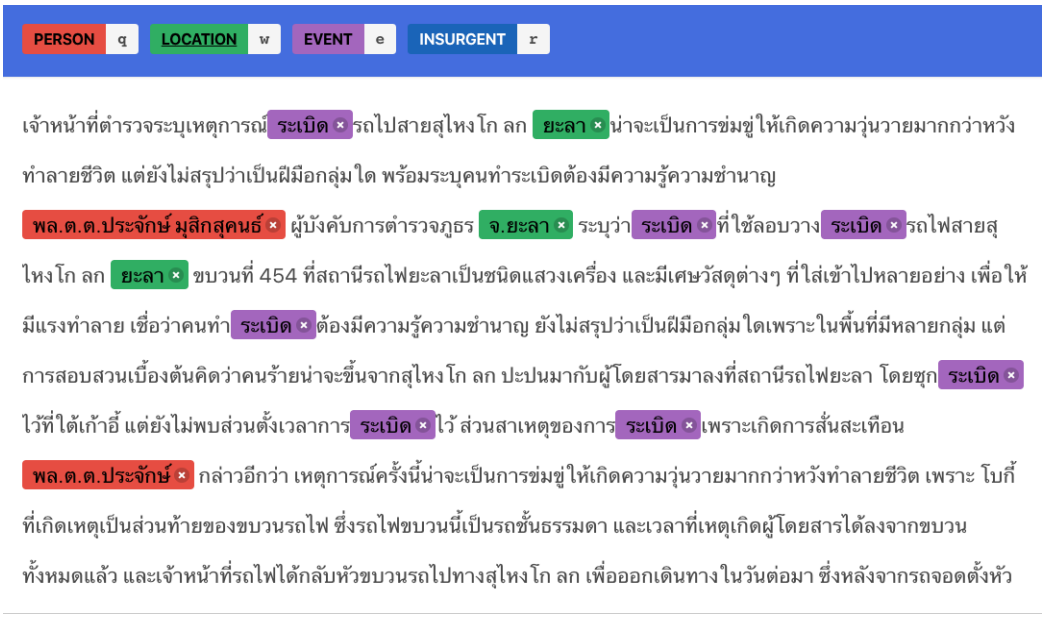
การเตรียมข้อมูลข่าวสำหรับใช้ในการฝึกโมเดล

1. โหลดข้อมูลข่าวมาจากเว็บไซต์ข่าว โดยจะอยู่ในรูปแบบของ plain text เพื่อทำเป็น Dataset

```
news_id: 113621.txt

title: ยะลาตรวจเข้ม 4 มุมเมืองหวั่นคนร้ายก่อเหตุครบ 8 ปีปล้นปืน

content: ยะลา – ตำรวจ ทหาร ยะลา เตรียมเข้ม หวั่นการก่อเหตุรับครบรอบ 8 ปี ปล้นปืน 413
กระบอก ค่ายปิเหล็ง อ.เจาะไอร้อง จ.นราธิวาส (วันเสียงปืนแตก) วันนี้
(4 ม.ค.) กำลังเจ้าหน้าที่ตำรวจ ทหาร และ ฝ่ายปกครอง ในพื้นที่ 3 จังหวัดชายแดนใต้
ได้สนธิกำลังคุมเข้มมาตรการรักษาความปลอดภัยในพื้นที่ หวั่นกลุ่มผู้ก่อความไม่สงบ
อาศัยจังหวะในช่วงครบรอบ 8 ปี ปล้นปืน 413 กระบอก ค่ายปิเหล็ง อ.เจาะไอร้อง
จ.นราธิวาส (วันเสียงปืนแตก) ก่อเหตุร้ายในพื้นที่ ในส่วนของ จ.ยะลา
เจ้าหน้าที่ได้สนธิกำลังตั้งจุดตรวจจุดสกัดตามถนนสายหลัก และสายรอง
พร้อมกับมีการลาดตระเวนในพื้นที่เป้าหมาย และการดูแลความปลอดภัยในย่านชุมชน
และสถานที่ราชการ โดยผู้บังคับบัญชาของทุกหน่วยได้เน้นย้ำให้มีการตรวจรถทุกชนิด
อาวุธต่างๆ และ บุคคลเป้าหมายอย่างเข้มงวดที่สุด นายเดชรัฐ สิมศิริ
ผู้ว่าราชการจังหวัดยะลา ได้กำชับไปยังฝ่ายความมั่นคงอำเภอทุกอำเภอ ให้จัดกำลัง
อส.และประสานกำนัน ผู้ใหญ่บ้าน ดูแลสถานที่ราชการอย่างใกล้ชิด ในขณะที่ พล.ต.ต.พีระ
บุญเลี้ยง ผู้บังคับการตำรวจภูธรจังหวัดยะลา ได้สั่งการให้ทุก สภ.เพิ่มความเข้ม
จัดกำลังตำรวจ สายตรวจออกลาดตระเวนดูแลพื้นที่เป้าหมาย ย่านชุมชน สถานที่ราชการ
และให้จัดตั้งด่านตรวจลอย เพื่อกดดันกลุ่มคนร้ายที่อาจจะลักลอบเข้ามาก่อเหตุในพื้นที่
อย่างเข้มข้น ส่วนพื้นที่ในเขตเทศบาลนครยะลา ได้ให้ สภ.เมือง
จัดกำลังเสริมประจำด่านตรวจ 4 มุมเมืองที่จะเข้ามายังในเขตเทศบาล ให้ตรวจคนรถยนต์
และรถจักรยานยนต์ที่จะเข้ามาทุกคัน ไม่มีการยกเว้น
```

รูปที่ *10* ข้อมูลข่าวที่ดึงมาจากเว็บไซต์หนึ่ง

2. ทำการรวบรวมข่าวทั้งหมดให้เป็น plain text ในไฟล์เดียว โดยทำการตัดเฉพาะส่วนของ content ของข่าวออกมา
3. ทำการแทรคคำที่เราต้องการให้เรียนรู้โดยใช้ Doccano โดยแบ่งแป็น 3 ประเภทคำคือ
    - PERSON แทนบุคคล
    - LOCATION แทนสถานที่
    - EVENT แทนเหตุการณ์



รูปที่ 11 เครื่องมือที่ช่วยในการแท็กคำ Doccano

4. เมื่อทำการแท็กคำเสร็จสิ้น Doccano จะให้ผลลัพธ์ของข้อมูลออกมาในรูปแบบของ Json

รูปที่ *12* ผลลัพธ์การทำงานของโปรแกรม *Doccano*

5. นำไฟล์ Json ที่ได้มาจาก Doccano มาเปลี่ยนให้เป็น Plain Text

"meta": {}, "annotation_approver": null, "labels": [[502, 516, "PERSON"], [1424, 1431, "PERSON"], [1825, 1847, "PERSON"], [2905, 2924, "PERSON"], [3210, 3222, "PERSON"]]}
{"id": 972, "text": "เจ้าหน้าที่ตำรวจระบุเหตุการณ์ระเบิดรถไปสายสุไหงโก ลก ยะลาน่าจะเป็นการข่มขู่ให้เกิดความวุ่นวายมากกว่าหวังทำลายชีวิต แต่ยังไม่สรุปว่าเป็นฝีมือกลุ่มใด พร้อมระบุคนทำระเบิดต้องมีความรู้ความชำนาญ พล.ต.ต.ประจักษ์ มุสิกสุคนธ์ ผู้บังคับการตำรวจภูธร จ.ยะลา ระบุว่าระเบิดที่ใช้ลอบวางระเบิดรถไฟสายสุไหงโก ลก ยะลา ขบวนที่ 454 ที่สถานีรถไฟยะลาเป็นชนิดแสวงเครื่อง และมีเศษวัสดุต่างๆ ที่ใส่เข้าไปหลายอย่าง เพื่อให้มีแรงทำลาย เชื่อว่าคนทำระเบิดต้องมีความรู้ความชำนาญ ยังไม่สรุปว่าเป็นฝีมือกลุ่มใดเพราะในพื้นที่มีหลายกลุ่ม แต่การสอบสวนเบื้องต้นคิดว่าคนร้ายน่าจะขึ้นจากสุไหงโก ลก ปะปนมากับผู้โดยสารมาลงที่สถานีรถไฟยะลา โดยชุกระเบิดไว้ที่ใต้เก้าอี้ แต่ยังไม่พบส่วนตั้งเวลาการระเบิดไว้ ส่วนสาเหตุของการระเบิดเพราะเกิดการสั่นสะเทือน พล.ต.ต.ประจักษ์ กล่าวอีกว่า เหตุการณ์ครั้งนี้น่าจะเป็นการข่มขู่ให้เกิดความวุ่นวายมากกว่าหวังทำลายชีวิต เพราะ โบกี้ที่เกิดเหตุเป็นส่วนท้ายของขบวนรถไฟ ซึ่งรถไฟขบวนนี้เป็นรถชั้นธรรมดา และเวลาที่เหตุเกิดผู้โดยสารได้ลงจากขบวนทั้งหมดแล้ว และเจ้าหน้าที่รถไฟได้กลับหัวขบวนรถไปทางสุไหงโก ลก เพื่อออกเดินทางในวันต่อมา ซึ่งหลังจากรถจอดตั้งหัวขบวนเสร็จ เจ้าหน้าที่ตำรวจรถไฟและเจ้าหน้าที่รักษาความปลอดภัยขึ้นไปตรวจสอบโบกี้ตามปกติ เห็นกล่องกระดาษใต้เบาะที่นั่ง ตำรวจได้หยิบขึ้นมาเปิดจึงเกิดระเบิดขึ้น ทำให้ทั้งสองได้รับบาดเจ็บ", "meta": {}, "annotation_approver": null, "labels": [[197, 217, "PERSON"], [718, 726, "PERSON"], [53, 57, "LOCATION"], [242, 246, "LOCATION"], [296, 300, "LOCATION"], [29, 35, "EVENT"], [595, 599, "LOCATION"]]}
{"id": 973, "text": "ตำรวจใต้ขวัญเสีย ขอย้ายหนีพื้นที่ภาค 9 กว่า 700 นาย โอดค่าตอบแทนต่ำ ตั้งศูนย์เฉพาะกิจระดมตำรวจทหารกว่า 500 นายลงพื้นที่ 3 จังหวัดชายแดน พล.ต.ท.ธวัชชัย จุลสุคนธ์ ผู้บัญชาการตำรวจภูธรภาค 9 แถลงแนวทางการแก้ไขปัญหาจังหวัดชายแดนภาคใต้ว่า ได้จัดตั้งศูนย์เฉพาะ กิจรักษาความปลอดภัยจังหวัดชายแดนภาคใต้ขึ้นที่ปัตตานี ตั้งแต่วันที่ 1 ก.ค. ที่ผ่านมา โดยเป็นกองบัญชาการผสม พลเรือน ตำรวจ ทหารที่ 43 (พตท.43) ซึ่งมี พล.ต.ต.ประเสริฐ จันทร์อินทร์ ผู้ช่วย ผบช.ภ.9 เป็นประธาน มีกำลังหลักในศูนย์ 62 และชุดชุมชนสัมพันธ์ ตำรวจตระเวน

รูปที่ *13* ผลลัพธ์การแทรคคำ

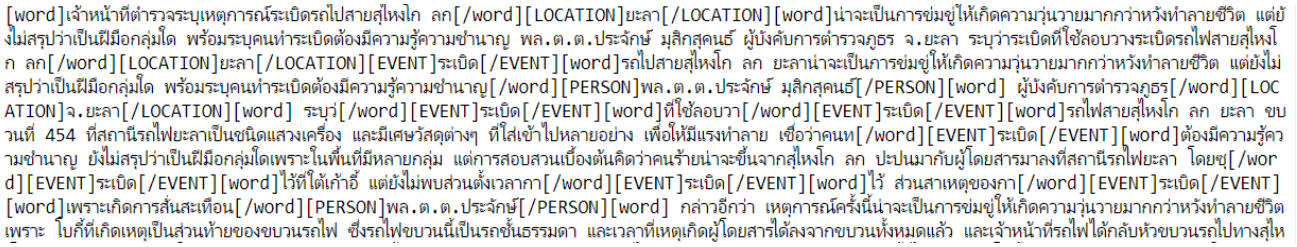
6. นำ Plain text มาแท็กคำที่เป็นชื่อเฉพาะ ที่ด้านหน้าและหลังของคำตามรูปแบบที่แท็กในโปรแกรม Doccano เช่น [LOCATION]สงขลา[/LOCATION] เป็นต้น

สงขลา ๑๑ พ.ค. นักวิเคราะห์มั่นใจดัชนีหุ้นไทยปีนี้ขึ้นไปถึง ๔๕๐ จุด เพราะเศรษฐกิจสหรัฐฟื้นแน่ แนะหุ้นเด่นกลุ่มอสังหาริมทรัพย์และสื่อสาร อย่างไรก็ตามปัญหาหนี้สาธารณะที่เพิ่มมากขึ้นจากการที่ใช้งบประมาณขาดดุลในการกระตุ้นเศรษฐกิจจะเป็นปัจจัยเสี่ยงที่ระมัดระวัง เล็งธุรกิจยางพารา ส่งออกแข็งภาคใต้เข้าตลาดปีนี้ ๑ ๒ แห่ง วันนี้ (๑๑ พ.ค.) ตลาดหลักทรัพย์แห่งประเทศไทย (ตลท.) ร่วมกับหอการค้าจังหวัดสงขลา จัดงานตลาดหลักทรัพย์สัญจร เพื่อพบปะนักลงทุน โดยมีการสัมมนาในหัวข้อ มองตลาดหุ้นไทยในภาวะเศรษฐกิจขาขึ้น [PERSON]นายอนุสรณ์ ธรรมใจ[/PERSON]สงขลา ๑๑ พ.ค. นักวิเคราะห์มั่นใจดัชนีหุ้นไทยปีนี้ขึ้นไปถึง ๔๕๐ จุด เพราะเศรษฐกิจสหรัฐฟื้นแน่ แนะหุ้นเด่นกลุ่มอสังหาริมทรัพย์และสื่อสาร อย่างไรก็ตามปัญหาหนี้สาธารณะที่เพิ่มมากขึ้นจากการที่ใช้งบประมาณขาดดุลในการกระตุ้นเศรษฐกิจจะเป็นปัจจัยเสี่ยงที่ระมัดระวัง เล็งธุรกิจยางพารา ส่งออกแข็งภาคใต้เข้าตลาดปีนี้ ๑ ๒ แห่ง วันนี้ (๑๑ พ.ค.) ตลาดหลักทรัพย์แห่งประเทศไทย (ตลท.) รวมกับหอการค้าจังหวัดสงขลา จัดงานตลาดหลักทรัพย์สัญจร เพื่อพบปะนักลงทุน โดยมีการสัมมนาในหัวข้อ มองตลาดหุ้นไทยในภาวะเศรษฐกิจขาขึ้น นายอนุสรณ์ ธรรมใจ ผู้อำนวยการ ผู้อำนวยการอาวุโส สำนักวิจัยและวางแผน ธนาคารไทยธนาคาร จำกัด (มหาชน) เปิดเผยว่า เมื่อพิจารณาจากตัวเลขเศรษฐกิจที่หลายฝ่ายได้มีการประเมินสะท้อนให้เห็นว่า เศรษฐกิจ อยู่ในช่วงขาขึ้น โดยประเมินว่าเศรษฐกิจไทยในปีนี้จะขยายตัวได้ประมาณ ร้อยละ ๓.๙ ซึ่งจะส่งผลดีต่อการซื้อขายในตลาดหลักทรัพย์ โดยคาดว่าในครึ่งปีหลังดัชนีจะขึ้นไปเคลื่อนไหวที่ระดับ ๔๔๔๔๐ จุด โดยมีค่าพีอีเพิ่มเป็น ๑๐ เท่า จากเดิมอยู่ที่ ๖ เท่า ทั้งนี้เศรษฐกิจสหรัฐ มีแนวโน้มฟื้นตัวอย่างชัดเจน โดยภายหลังการประกาศตัวเลขจีดีพีที่โตประมาณ ร้อยละ ๕ ๘ มากกว่าที่คาดการณ์ไว้ ซึ่งหากเศรษฐกิจสหรัฐโตขึ้นก็จะส่งผลให้เศรษฐกิจไทยดีขึ้นและส่งออกได้เพิ่มขึ้น ส่วนปัจจัยภายในประเทศภาวะดอกเบี้ยต่ำเป็นปัจจัยเสริม ซึ่งหากทางการมีการยกเลิกการค้ำประกันเงินฝากจะช่วยให้มีเงินไหลเข้ามาลงทุนในตลาดหุ้นประมาณ ร้อยละ ๓๐๕๐ ซึ่งจะช่วยเสริมให้ตลาดทุนมีขนาดใหญ่ขึ้นและเป็นแหล่งระดมทุนของภาคธุรกิจ เพราะที่ผ่านมาภาคธุรกิจส่วนใหญ่จะใช้การระดมทุนจาก ภาคธนาคารเป็นหลัก [PERSON]นายอนุสรณ์[/PERSON]สงขลา ๑๑ พ.ค. นักวิเคราะห์มั่นใจดัชนีหุ้นไทยปีนี้ขึ้นไปถึง ๔๕๐ จุด เพราะเศรษฐกิจสหรัฐฟื้นแน่ แนะหุ้นเด่นกลุ่มอสังหาริมทรัพย์และสื่อสาร อย่างไรก็ตามปัญหาหนี้สาธารณะที่เพิ่มมากขึ้นจากการที่ใช้งบประมาณขาดดุลในการกระตุ้นเศรษฐกิจจะเป็นปัจจัยเสี่ยงที่ระมัดระวัง เล็งธุรกิจยางพารา ส่งออกแข็งภาคใต้เข้าตลาดปีนี้ ๑ ๒ แห่ง วันนี้ (๑๑ พ.ค.) ตลาดหลักทรัพย์แห่งประเทศไทย (ตลท.) รวมกับหอการค้าจังหวัดสงขลา จัดงาน

รูปที่ *14* ภาพการแท็กคำหน้าและหลังของคำ

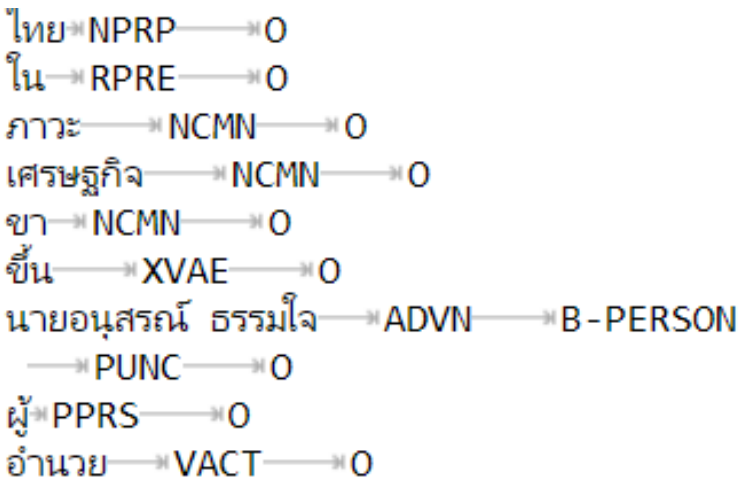
ขั้นตอนการสร้างและฝึกโมเดล

1. นำ Dataset ที่ได้จากส่วนการเตรียมข้อมูลมาทำให้อยู่ในรูปแบบที่จะไปใช้ฝึกโมเดล โดยการแปลงส่วนที่มี tag เช่น [EVENT]ยิง[/EVENT] ให้เป็น ('EVENT',ยิง,'EVENT') และทำการจัดการกับคำที่ไม่ได้ tag ให้เป็น [WORD][/WORD] โดยในการตัดคำเราใช้ library ของ nltk [12]


[word]เจ้าหน้าที่ตำรวจระบุเหตุการณ์ระเบิดรถไปสายสุไหงโก ลก[/word][LOCATION]ยะลา[/LOCATION][word]น่าจะเป็นการข่มขู่ให้เกิดความวุ่นวายมากกว่าหวังทำลายชีวิต แต่ยังไม่สรุปว่าเป็นฝีมือกลุ่มใด พร้อมระบุคนทำระเบิดต้องมีความรู้ความชำนาญ พล.ต.ต.ประจักษ์ มุสิกสุคนธ์ ผู้บังคับการตำรวจภูธร จ.ยะลา ระบุว่าระเบิดที่ใช้ลอบวางระเบิดรถไฟสายสุไหงโก ลก[/word][LOCATION]ยะลา[/LOCATION][EVENT]ระเบิด[/EVENT][word]รถไปสายสุไหงโก ลก ยะลาน่าจะเป็นการข่มขู่ให้เกิดความวุ่นวายมากกว่าหวังทำลายชีวิต แต่ยังไม่สรุปว่าเป็นฝีมือกลุ่มใด พร้อมระบุคนทำระเบิดต้องมีความรู้ความชำนาญ[/word][PERSON]พล.ต.ต.ประจักษ์ มุสิกสุคนธ์[/PERSON][word] ผู้บังคับการตำรวจภูธร[/word][LOCATION]จ.ยะลา[/LOCATION][word] ระบุว่า[/word][EVENT]ระเบิด[/EVENT][word]ที่ใช้ลอบวา[/word][EVENT]ระเบิด[/EVENT][word]รถไฟสายสุไหงโก ลก ยะลา ขบวนที่ 454 ที่สถานีรถไฟยะลาเป็นชนิดแสวงเครื่อง และมีเศษวัสดุต่างๆ ที่ใส่เข้าไปหลายอย่าง เพื่อให้มีแรงทำลาย เชื่อว่าคนท[/word][EVENT]ระเบิด[/EVENT][word]ต้องมีความรู้ความชำนาญ ยังไม่สรุปว่าเป็นฝีมือกลุ่มใดเพราะในพื้นที่มีหลายกลุ่ม แต่การสอบสวนเบื้องต้นคิดว่าคนร้ายน่าจะขึ้นจากสุไหงโก ลก ปะปนมากับผู้โดยสารมาลงที่สถานีรถไฟยะลา โดยซ[/word][EVENT]ระเบิด[/EVENT][word]ไว้ที่ใต้เก้าอี้ แต่ยังไม่พบส่วนตั้งเวลาก[/word][EVENT]ระเบิด[/EVENT][word]ไว้ ส่วนสาเหตุของก[/word][EVENT]ระเบิด[/EVENT][word]เพราะเกิดการสั่นสะเทือน[/word][PERSON]พล.ต.ต.ประจักษ์[/PERSON][word] กล่าวอีกว่า เหตุการณ์ครั้งนี้น่าจะเป็นการข่มขู่ให้เกิดความวุ่นวายมากกว่าหวังทำลายชีวิต เพราะ ใบกี่ที่เกิดเหตุเป็นส่วนท้ายของขบวนรถไฟ ซึ่งรถไฟขบวนนี้เป็นรถชั้นธรรมดา และเวลาที่เหตุเกิดผู้โดยสารได้ลงจากขบวนทั้งหมดแล้ว และเจ้าหน้าที่รถไฟได้กลับหัวขบวนรถไปทางสุไห

รูปที่ *15* ผลลัพธ์ของการเพิ่ม *tag* ลงในข้อมูล

2. นำ Data ที่ได้จากข้อที่ 1 มาทำให้ข้อมูลอยู่ในรูปของ conll2002 และทำการเพิ่ม part of speech (POS) [13] ให้กับข้อมูลโดยใช้ pythainlp



รูปที่ *16* รูปตัวอย่างข้อมูลไฟล์ *conll2002*

3. นำข้อมูล conll2002 ที่ได้ มาทำ feature word เพื่อใช้ฝึกโมเดล โดย feature word นั้นจะเป็นการบ่งบอกถึงลักษณะเฉพาะของคำนั้นๆ และใช้หาคำที่มีลักษณะของคำตามที่ต้องการ เช่น POS, การมีเว้นวรรค, การมีจุดในคำนั้นกี่ตัว และนำคำก่อนหน้าและคำหลัง เป็นต้น โดยโมเดล

จะให้ความต่างของ feature word เหล่านี้ในการคำนวณว่าคำนี้เป็น PERSON, EVENT, LOCATION หรือไม่ใช่คำในชนิดเหล่านี้

4. นำ feature word เป็นตัวฝึกและนำ tag เป็นคำตอบในการฝึกโมเดล conditional random fields (CRFs)
5. ทำการแบ่ง training set และ test set เป็นอัตราส่วน 0.8 และ 0.2 ตามลำดับ
6. ทำการฝึกโมเดล
7. นำข้อมูลที่ต้องการสกัดคำจากเว็บมาสกัดข่าวเพิ่มเติมเพื่อให้มีข้อมูลมากขึ้นโดยใช้โมเดล CRFs ที่ได้ทำการฝึกมา

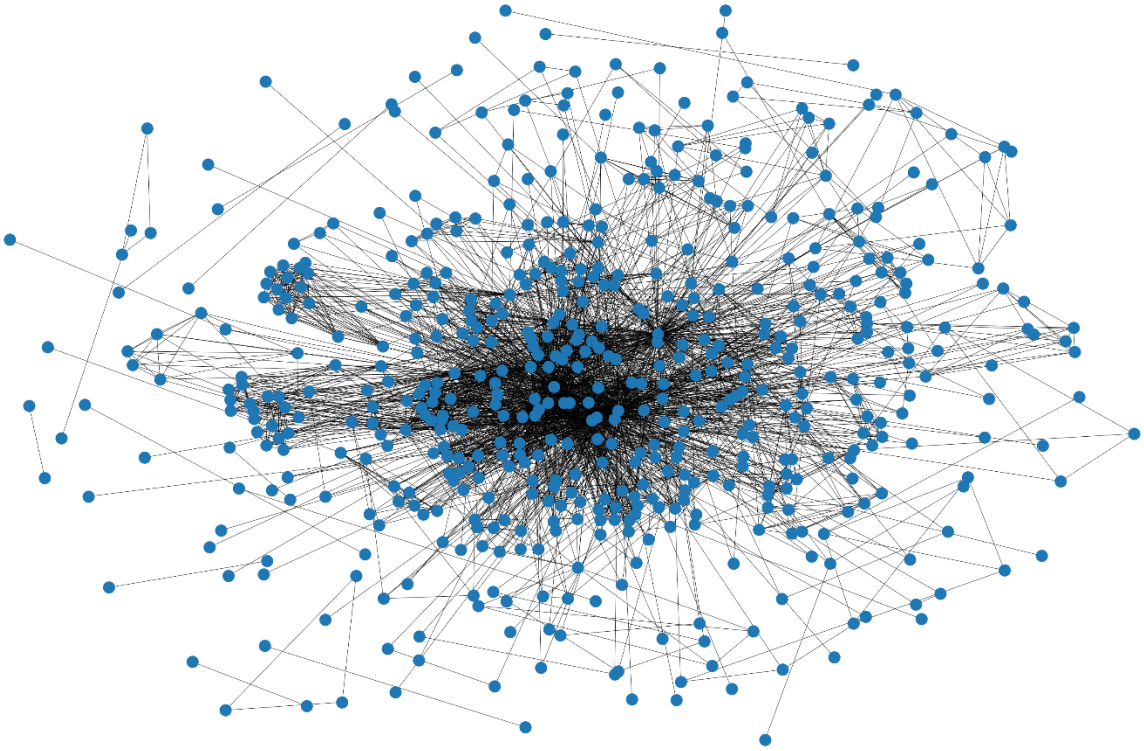
ขั้นตอนการสร้างกราฟความสัมพันธ์จากข้อมูล

1. นำข้อมูลที่ได้จากการสกัดคำจากโมเดลมาสร้างตารางความสัมพันธ์ของข้อมูล โดยต้องมีการกำหนดความสัมพันธ์ของคำ ซึ่งในที่นี้จะทำการกำหนดความสัมพันธ์ของคำที่ได้จากข่าวเดียวกันมีความสัมพันธ์ต่อกัน

	word1	word2	nameagenvy	date	url
0	ส.ต.อ.ธวัช บัวศรี	จ.นราธิวาส	กรุงเทพธุรกิจ	19 ต.ค. 2545	http://www.bangkokbiznews.com/2002/10/19/pol/i...
1	นายอายุ สาเมาะ	จ.นราธิวาส	กรุงเทพธุรกิจ	19 ต.ค. 2545	http://www.bangkokbiznews.com/2002/10/19/pol/i...
2	นายอายุ สาเมาะ	ส.ต.อ.ธวัช บัวศรี	กรุงเทพธุรกิจ	19 ต.ค. 2545	http://www.bangkokbiznews.com/2002/10/19/pol/i...
3	พ.ต.อ.ปริญญา ขวัญยืน	จ.นราธิวาส	กรุงเทพธุรกิจ	19 ต.ค. 2545	http://www.bangkokbiznews.com/2002/10/19/pol/i...
4	พ.ต.อ.ปริญญา ขวัญยืน	ส.ต.อ.ธวัช บัวศรี	กรุงเทพธุรกิจ	19 ต.ค. 2545	http://www.bangkokbiznews.com/2002/10/19/pol/i...
...	...	...	...	...	...
253550	พล.อ.ชวลิต	พล.ต.อ. สุนทร ซ้ายขวัญ	ไทยรัฐ	25 พ.ค. 2547	http://www.thairath.co.th/thairath1/2547/page1...
253551	พล.อ.ชวลิต	พ.ต.ท.ทักษิณ ชินวัตร	ไทยรัฐ	25 พ.ค. 2547	http://www.thairath.co.th/thairath1/2547/page1...
253552	พล.อ.ชวลิต	พ.ต.ท.ทักษิณ	ไทยรัฐ	25 พ.ค. 2547	http://www.thairath.co.th/thairath1/2547/page1...
253553	พล.อ.ชวลิต	พ.ต.ท. ทักษิณ	ไทยรัฐ	25 พ.ค. 2547	http://www.thairath.co.th/thairath1/2547/page1...
253554	พล.อ.ชวลิต	พล.อ.ชาลิต ยงใจยุทธ	ไทยรัฐ	25 พ.ค. 2547	http://www.thairath.co.th/thairath1/2547/page1...

รูปที่ *17* ตารางแสดงความสัมพันธ์ของข้อมูลในรูปแบบ 1:1

2. นำข้อมูลความสัมพันธ์ที่ได้จากขั้นตอนก่อนหน้ามาใช้ในการสร้างกราฟแสดงความสัมพันธ์ด้วย NetworkX โดยขั้นแรกทำการเพิ่ม Node จากชื่อเฉพาะทุกคำให้กับกราฟก่อน จากนั้นทำการเพิ่ม Edge ให้กับกราฟโดยการใส่ค่า Node 2 ตัวที่มีความสัมพันธ์ต่อกันเข้าไป เพื่อให้ได้กราฟที่แสดงความสัมพันธ์ของข้อมูลออกมา



รูปที่ *18* ตัวอย่างกราฟที่ได้จากกข้อมูล

3. หลังจากได้กราฟที่เราต้องการแล้ว จะทำการคำนวณหาค่าต่างๆ ตามหลักการวิเคราะห์เครือข่ายทางสังคม **[**14**]** โดยค่าที่เราเลือกคำนวณออกมา ได้แก่
    - Degree จำนวนของ Edge ที่ Node นั้นมีความสัมพันธ์
    - Betweenness Centrality วัดได้จาก shortest paths เส้นทางที่สั้นที่สุดที่ผ่านแต่ละ

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

node ถ้ามีค่าการเชื่อมโยงมาก แสดงว่า Node นั้นมีความสำคัญต่อระบบเครือข่ายมาก

เมื่อ $\boldsymbol{\sigma_{st}}$ เป็นจำนวน shortest paths จาก node s ไปถึง t และ

$\boldsymbol{\sigma_{st}(v)}$ เป็นจำนวน shortest paths ทั้งหมดที่ผ่านจุด v

- **Degree Centrality** วัดได้จากจำนวน edges ที่ต่อเข้ากับ node แต่ละ node ยิ่งมีค่ามากแสดงว่า Node นั้นมีความสำคัญต่อระบบเครือข่ายมาก

$$\boldsymbol{C_D(v) = \deg(v)}$$

เมื่อ $\boldsymbol{deg(v)}$ คือ degree ของ node v

$$C_D(G) = \frac{\sum_{i=1}^{|V|}[C_D(v*) - C_D(v_i)]}{H}$$

เมื่อ $\boldsymbol{H}$ เป็นคือค่า $\boldsymbol{maximized}$ เมื่อ กราฟ $\boldsymbol{x}$ กระกอบไปด้วย central node ไปยัง node อื่นในกรณีนี้

$$H = (n-1)\cdot((n-1)-1) = n^2 - 3n + 2.$$

- **Closeness Centrality** วัดได้จากการวัดระยะทางที่ใกล้กับ node อื่นๆในกราฟ ยิ่งมีค่ามากแสดงว่ามีความสามารถในการติดต่อกับ node อื่นๆมาก

$$C(x) = \frac{1}{\sum_y d(y,x)}$$

เมื่อ $\boldsymbol{d(y,x)}$ เป็น ระยะห่างระหว่างจุด $\boldsymbol{y}$ และ $\boldsymbol{x}$

- **Eigenvector Centrality** วัดว่า node แต่ละ node เชื่อมต่อกับ node ที่มีความสำคัญมากน้อยแค่ไหนโดยการใช้ adjacency matrix ในการหาค่า Eigenvector Centrality

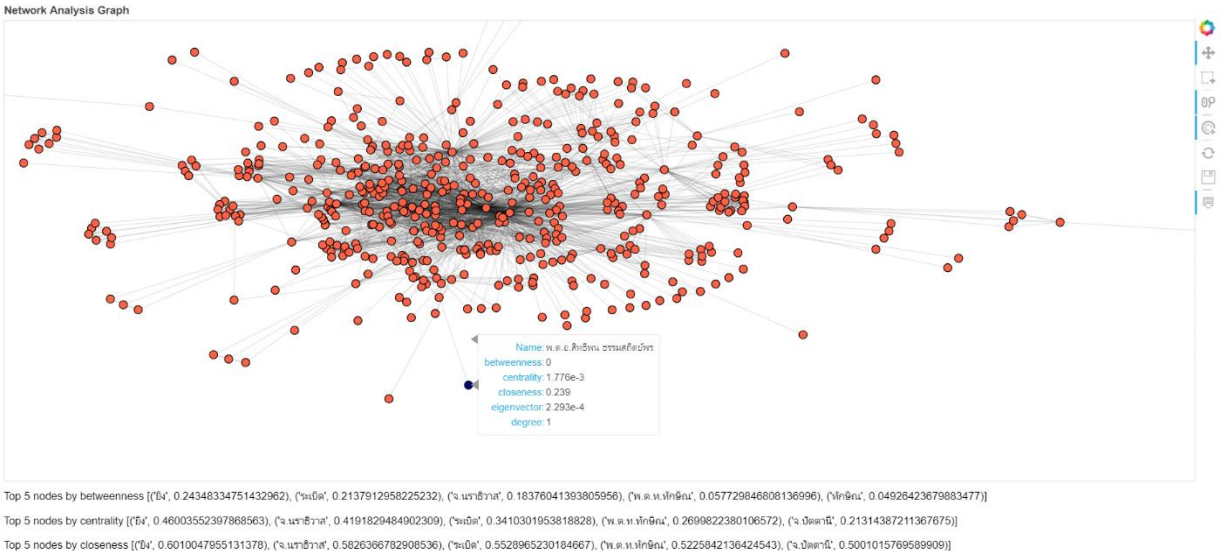
ให้ $\boldsymbol{A = (a_{v,t})}$ เป็น adjacency matrix

นั้นคือ $\boldsymbol{(a_{v,t}) = 1}$ ถ้า จุด $\boldsymbol{v}$ เชื่อมกับ จุด $\boldsymbol{t}$ และ $\boldsymbol{(a_{v,t}) = 0}$

$$x_v = \frac{1}{\lambda}\sum_{t\in M(v)} x_t = \frac{1}{\lambda}\sum_{t\in G} a_{v,t}x_t$$

เมื่อ $\boldsymbol{M(v)}$ เป็น set ของ neighbors ของ $\boldsymbol{v}$ และ $\boldsymbol{\lambda}$ เป็นค่าคงที่

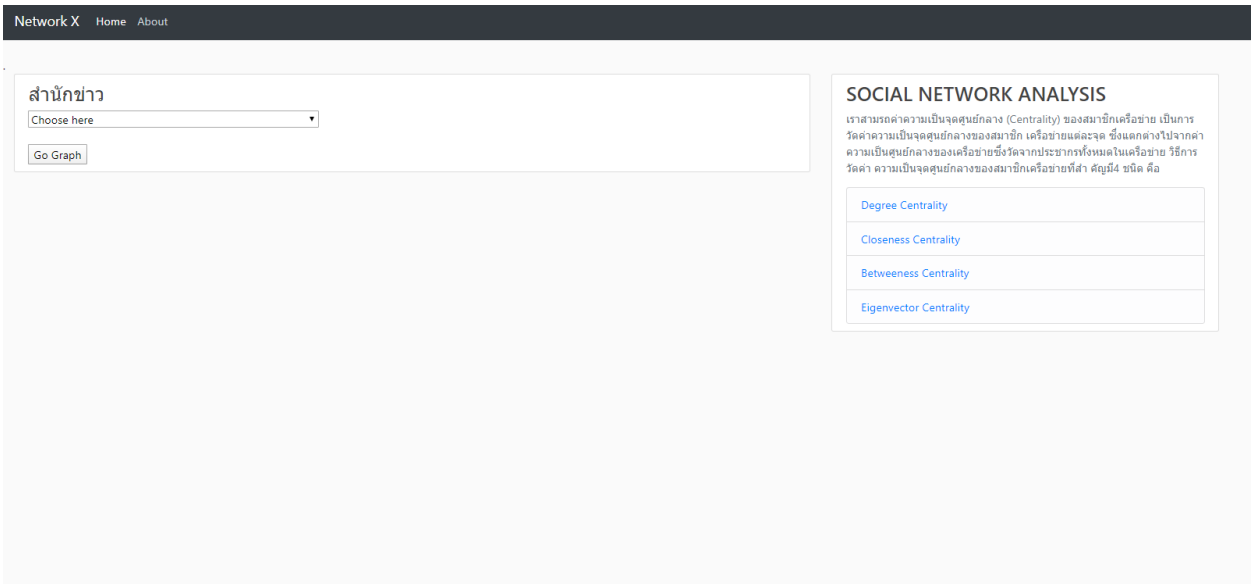
4. จากนั้นเพิ่มค่าต่างๆ ที่ได้จากรการคำนวณเป็น attribute ของ Node เพื่อนำมาแสดงผลโดยใช้ Bokeh รวมถึงแสดงโหนด **5** อันกับแรกของหลักการวิเคราะห์เครือข่ายทางสังคม และทำการเพิ่ม Tools ของ Bokeh ลงไปเพื่อความสะดวกในการดูข้อมูล



รูปที่ *19* การแสดงผลกราฟความสัมพันธ์ของข้อมูลผ่าน *Bokeh*

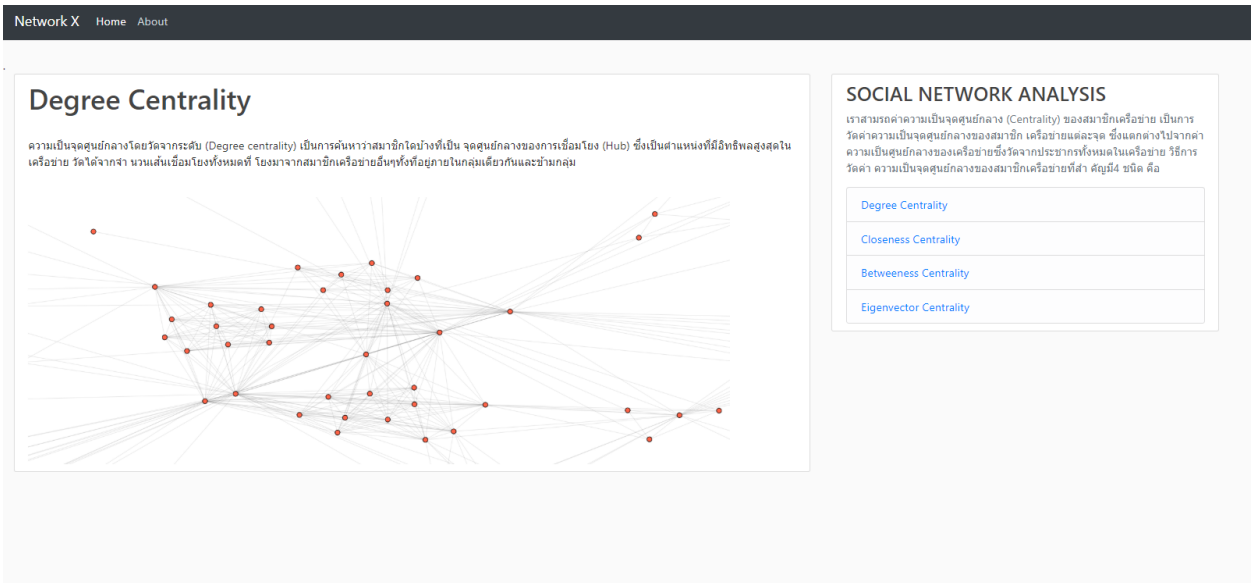
ขั้นตอนการแสดงผลโดย Django

1. เมื่อเปิดหน้าเว็บไซต์ขึ้นมาจะมีสำนักข่าวให้เลือก โดยสามารถเลือกได้ตามลิสต์ที่กำหนดไว้ให้



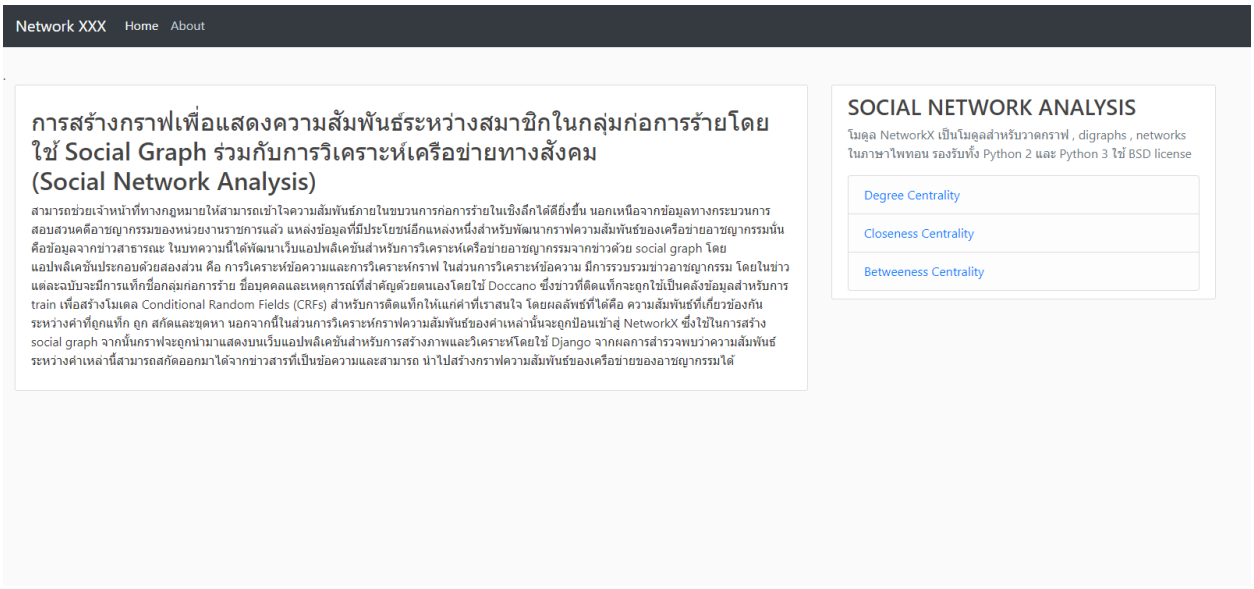
รูปที่ *20* หน้าหลักของเว็บไซต์

2. ทางด้านขวาของของเว็บไซต์ จะเป็นคำอธิบายเกี่ยวกับค่าที่เราได้นำมาแสดงในกราฟ ซึ่งจะประกอบด้วย 4 ค่า ดังรูปที่ 21



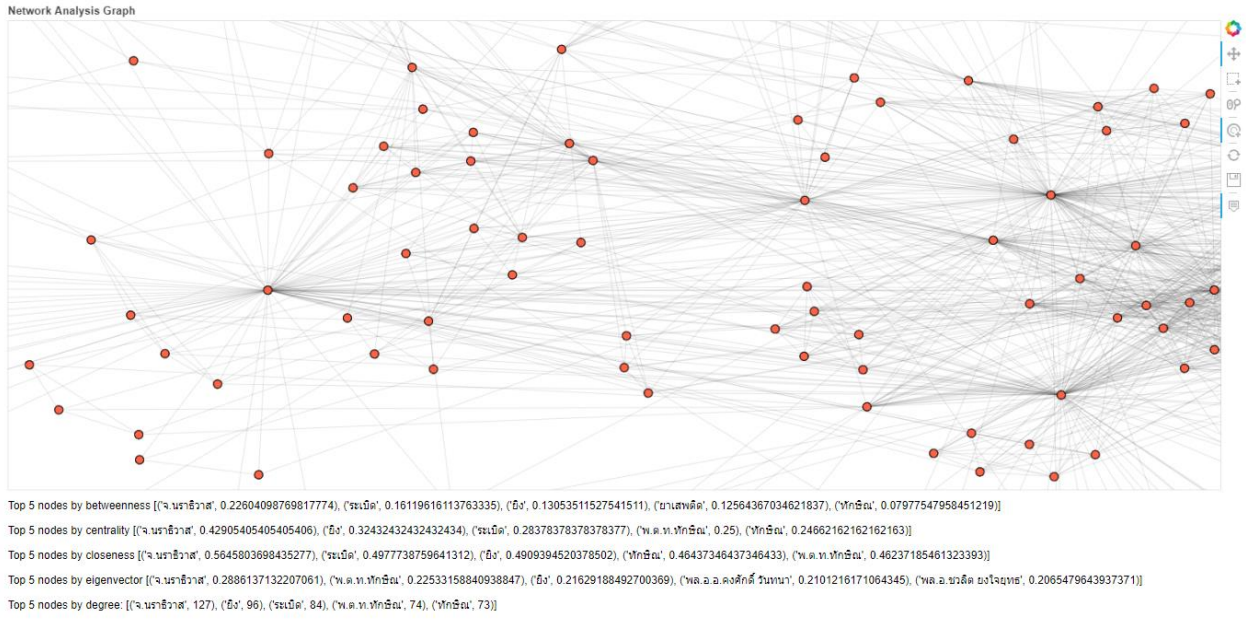
รูปที่ *21* หน้าอธิบายค่าตามหลักการวิเคราะห์เครือข่ายทางสังคม

3. ในหน้า About จะแสดงเกี่ยวกับข้อมูลต่างๆ เช่น ขั้นตอนในการสร้างกราฟ ไปจนถึงประโยชน์ของกราฟ ซึ่งจะแสดงใน รูปที่ 22



รูปที่ *22* หน้า *About*

4. จากหน้า Home เมื่อเราทำการกดเลือกสำนักข่าวที่เราต้องการ ระบบจะทำการสร้างกราฟออกมาตามสำนักข่าวนั้น รวมถึงจะมีการแสดง Top5 ของค่าตามหลักการวิเคราะห์เครือข่ายทางสังคม ซึ่งมีผลลัพธ์ดังรูปที่ 23



รูปที่ *23* หน้าการแสดงผลกราฟ

บทที่ 4

ผลการดำเนินงาน

ในงานวิจัยชิ้นนี้ คณะผู้จัดทำได้ทำการทดสอบประสิทธิภาพของ CRF โมเดล โดยการเปรียบเทียบประสิทธิภาพโมเดล โดยใช้ตัวชี้วัด ได้แก่ F1 Score, Precision, Recall และการแสดงค่า Betweenness, Centrality, Degree Centrality, Closeness Centrality, Eigenvector Centrality, Degree ของกราฟที่ได้จากข้อมูลทั้งหมด

4.1 ผลการทดสอบประสิทธิภาพโมเดลการรู้จำชื่อเฉพาะ

Nametag	Precision	Recall	F1-score
B-EVENT	0.808	0.737	0.771
I-EVENT	1.000	0.759	0.863
B-LOCATION	0.951	0.706	0.811
I-LOCATION	0.00	0.00	0.00
B-PERSON	0.911	0.793	0.848
I-PERSON	0.900	0.429	0.581
Micro avg	0.910	0.737	0.851
Macro avg	0.762	0.571	0.645
Weighted avg	0.908	0.737	0.810

ตารางที่ *2* ตารางแสดงประสิทธิภาพโมเดลการรู้จำชื่อเฉพาะ

โดยผลการทดสอบประสิทธิภาพของโมเดลในการู้จำชื่อเฉพาะนั้นได้ค่า Precision = 0.908, Recall = 0.737 และ F1-score = 0.810 โดยข้อมูลที่ใช้ 200 ข่าว

4.2 ผลการคำนวณค่าต่างๆของกราฟตามหลักการวิเคราะห์เครือข่ายทางสังคม [14]

- Degree Centrality 5 อันดับได้แก่
    1. จ.นราธิวาส 0.574
    2. ยิง 0.524

3. ระเบิด 0.482
4. พ.ต.ท.ทักษิณ 0.383
5. จ.ปัตตานี 0.365

- Betweenness Centrality 5 อันดับได้แก่
  1. จ.นราธิวาส 0.197
  2. ยิง 0.173
  3. ระเบิด 0.149
  4. พ.ต.ท.ทักษิณ 0.073
  5. จ.ปัตตานี 0.072
- Closeness Centrality คือ 5 อันดับได้แก่
  1. จ.นราธิวาส 0.689
  2. ยิง 0.665
  3. ระเบิด 0.645
  4. พ.ต.ท.ทักษิณ 0.604
  5. จ.ปัตตานี 0.597
- Eigenvector Centrality 5 อันดับได้แก่
  1. จ.นราธิวาส 0.221
  2. ยิง 0.206
  3. ระเบิด 0.192
  4. พ.ต.ท.ทักษิณ 0.178
  5. จ.ปัตตานี 0.169

บทที่ 5

สรุปผลการทดลอง

5.1 สรุปผลการดำเนินโครงการ

ในการทดลองครั้งนี้เราได้นำ CRFs โมเดลมาใช้ในการสกัดประเภทของคำจากข่าวภาษาไทยเพื่อนำข้อมูลที่ได้นั้นมาใช้ในการวิเคราะห์โดยใช้หลักทฤษฎีกราฟ โดย CRFs โมเดลที่เราได้สร้างขึ้นมานั้นมีค่า Precision = 0.908 , Recall = 0.737 และ F1-score = 0.810 และได้คำนวณค่าต่างๆที่บ่งบอกถึงความสำคัญของ คน สถานที่ และ เหตุการณ์ ว่ามีผลต่อการเหตุการณ์ความไม่สงบ โดยผลที่ได้คือ 5 อันดับของคน สถานที่ และ เหตุการณ์ที่มีผลต่อเหตุการณ์ความไม่สงบ ที่เกิดขึ้นโดยค่า Degree Centrality, Betweenness Centrality, Closeness Centrality และ Eigenvector Centrality ได้บ่งบอกไปในทิศทางเดี่ยวกันและจัดอับดับออกมากได้แก่

1. จ.นราธิวาส
2. ยิง
3. ระเบิด
4. พ.ต.ท.ทักษิณ
5. จ.ปัตตานี

5.2 ปัญหาและอุปสรรคในการดำเนินการ

- Bokeh ไม่สามารแสดงผลตัว hover ของ node และ edge พร้อมกันได้
- เนื่องจากสถานการณ์ covid-19 นั้นทำให้ไม่สามารถสร้างตัวฝึก data เพราะในการใช้งาน doccano ต้องใช้ docker และ docker ต้องใช้ iMac หรือ windows 10 pro ทำให้ข้อมูลที่ใช้ในการฝึกมีน้อย
- การจำลอง vpn ของมหาลัยเพื่อเข้าใช้งานเครื่องเซิฟเวอร์ที่ชั้นภาคมีปัญหา การเชื่อมต่อไม่ได้เป็นบางเวลา

5.3 ข้อเสนอแนะ

ในอนาคตถ้ามีโมเดลในการสร้างความสัมพันธ์ของคำจากการสกัดคำจะทำให้ข้อมูลที่ออกมาจากการวิเคราะห์มีความแม่นยำมากขึ้น

บรรณานุกรม

[1] นางสาวนัชชา ถิระสาโรช. (2553). **การรู้จำชื่อเฉพาะภาษาไทย: การใช้แบบจำลองคอนดิชั่นนอลแรนดอมฟิลด์ส.** สืบค้นเมื่อ 11 มกราคม 2563. จากhttp://www.arts.chula.ac.th/~ling/thesis/2553MA-Ling-Nutcha.pdf: 13-15.

[2] **สรุป Survey of Named Entity Recognition and Classification (NERC).** สืบค้นเมื่อ 9 ธันวาคม 2562. จาก https://lukkiddd.com/%E0%B8%AA%E0%B8%A3%E0%B8%B8%E0%B8%9B-survey-of-named-entity-recognition-and-classification-nerc-4a6597856ad3.

[3] นางสาวนัชชา ถิระสาโรช. (2553). **การรู้จำชื่อเฉพาะภาษาไทย: การใช้แบบจำลองคอนดิชั่นนอลแรนดอมฟิลด์ส.** สืบค้นเมื่อ 12 มกราคม 2563. จากhttp://www.arts.chula.ac.th/~ling/thesis/2553MA-Ling-Nutcha.pdf: 18.

[4] **GRAPH โครงสร้างข้อมูลแบบ กราฟ.** สืบค้นเมื่อ 10 ธันวาคม 2562. จาก https://sites.google.com/site/graphg6/kraf-khux-1.

[5] Supphachai Thaicharoen. (2019, April 27). **Text Mining with Word Embedded and Social Network Analysis.** Retrieved January 29, 2020, from https://drive.google.com/file/d/1lHxKewUddDz0z2Fc7hqtta_3piGP8hSt/view.

[6] Suphanut Thattinaphanich; Santitham Prom-on. (2019). **Thai Named Entity Recognition Using Bi-LSTM-CRF with Word and Character Representation.** Retrieved November 15, 2019, from https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8912091&tag=1.

[7] Kitiya Suriyachay; Virach Sornlertlamvanich. (2018). **Named Entity Recognition Modeling for the Thai Language from a Disjointedly Labeled Corpus.** Retrieved November 13, 2019, from https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8541344.

[8] O Sang Kwon; Junbeom Kim; Kwang-Ho Choi; Yeonhee Ryu; Ji-Eun Park. (2018). **Trends in deqi research: a text mining and networkanalysis.** Retrieved November 11, 2019, from https://www.researchgate.net/publication/323652076_Trends_in_deqi_research_a_text_mining_and_network_analysis.

[9] **Overview of NetworkX.** Retrieved October 30, 2019, from https://networkx.github.io/documentation/stable/.

[10] Aaron (Ari) Bornstein. (2019). **What is Doccano?.** Retrieved March 19, 2020, from https://towardsdatascience.com/text-annotation-on-a-budget-with-azure-web-apps-doccano-b29f479c0c54.

[11] **Bokeh.** Retrieved March 19, 2020, from https://docs.bokeh.org/en/latest/.

[12] **ทำ Named Entity Recognition ภาษาไทย : เบื้องหลังการทำ NER ให้ PyThaiNLP.** สืบค้นเมื่อ 8 พฤศจิกายน 2562. จาก https://python3.wannaphong.com/2018/12/named-entity-recognition-ner-pythainlp.html.

[13] **pythainlp.tag POS tags.** Retrieved December 23, 2019, from https://www.thainlp.org/pythainlp/docs/2.0/api/tag.html.

[14] **การวิเคราะห์เครือข่ายทางสังคม.** สืบค้นเมื่อ 29 พฤศจิกายน 2562. จาก https://medium.com/data-growing/%E0%B8%81%E0%B8%B2%E0%B8%A3%E0%B8%A7%E0%B8%B4%E0%B9%80%E0%B8%84%E0%B8%A3%E0%B8%B2%E0%B8%B0%E0%B8%AB%E0%B9%8C%E0%B9%80%E0%B8%84%E0%B8%A3%E0%B8%B7%E0%B8%AD%E0%B8%82%E0%B9%88%E0%B8%B2%E0%B8%A2%E0%B8%97%E0%B8%B2%E0%B8%87%E0%B8%AA%E0%B8%B1%E0%B8%87%E0%B8%84%E0%B8%A1-ep-2-15f9451ce116.

ภาคผนวก

ภาคผนวก ก

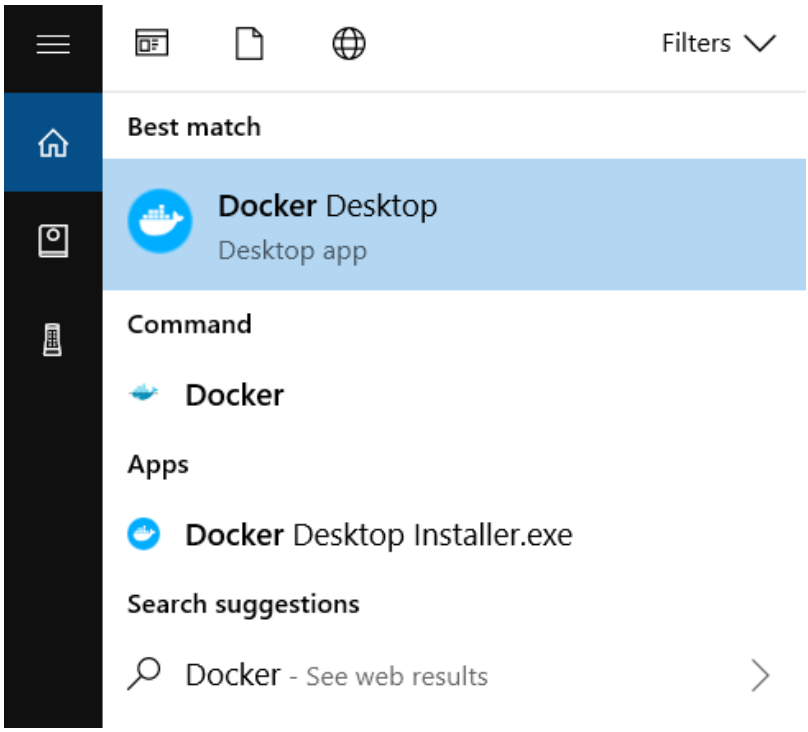
สอนการใช้งาน Doccano

1. ทำการติดตั้งและใช้งาน Docker

- ทำการดาวโหลด Docker เวอร์ชั่นล่าสุดจาก :

  https://www.docker.com/products/docker-desktop

- ทำการเปิดไฟล์ Docker Desktop Installer.exe เพื่อติดตั้ง
- เมื่อติดตั้งสำเร็จ ให้ทำการเปิดโปรแกรม Docker ดังรูปด้านล่าง



รูปที่ *24* ตัวอย่างการเปิดใช้งานโปรแกรม *Docker*

2. **ทำการเปิดใช้งาน Doccano**

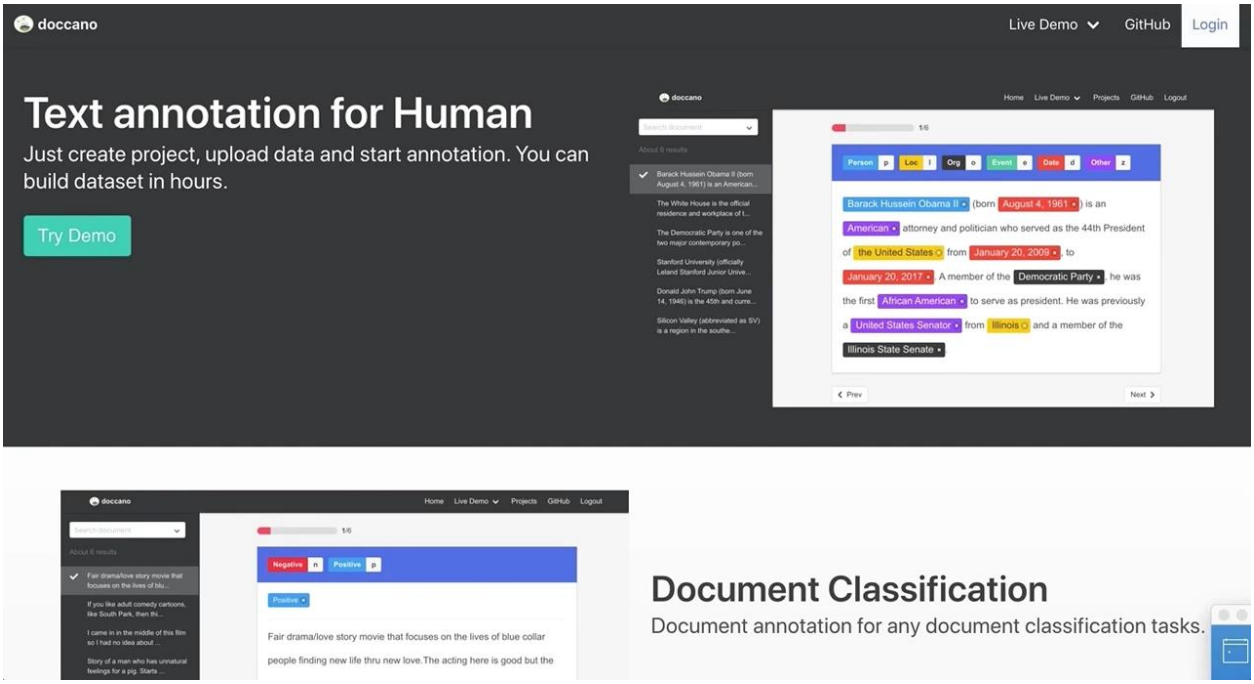
- ทำการเปิดโปรแกรม Command Prompt หรือ Powershell และใส่ข้อความตามด้านล่าง

```
git clone https://github.com/doccano/doccano.git
cd doccano
```

- เข้าโปรแกรม Docker และใส่ข้อความ ดังนี้

```
$ docker-compose -f docker-compose.prod.yml up
```

- เข้าเว็บเบราเซอร์ และใส่พลอต http://0.0.0.0/ หากไม่ได้ให้ใส่ http://127.0.0.1:3000/
- จากนั้นจะเข้าสู่ตัวโปรแกรม Doccano

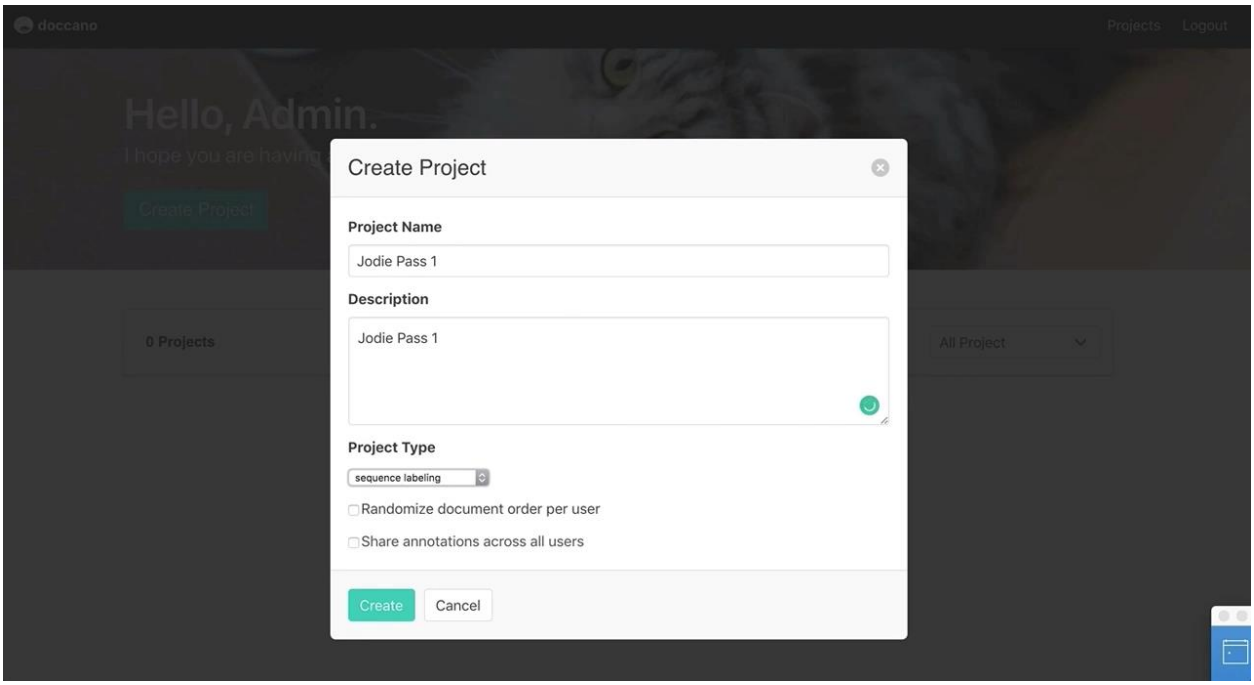


รูปที่ *25* ตัวเว็บไซต์ *Doccano*

- ทำการ login โดยใส่ username และ password ตามด้านล่าง

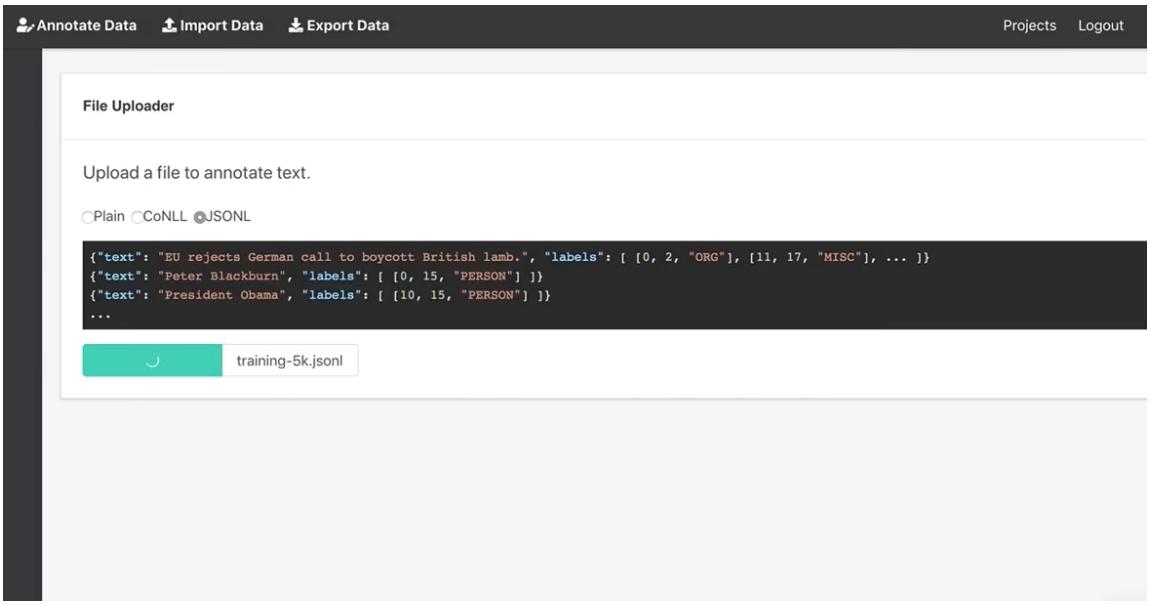
```
USERNAME: "admin"
PASSWORD: "password"
```

- เมื่อทำการเข้าสู่ระบบสำเร็จให้ทำการคลิกปุ่ม Create Project และกรอกข้อมูล ดังรูป



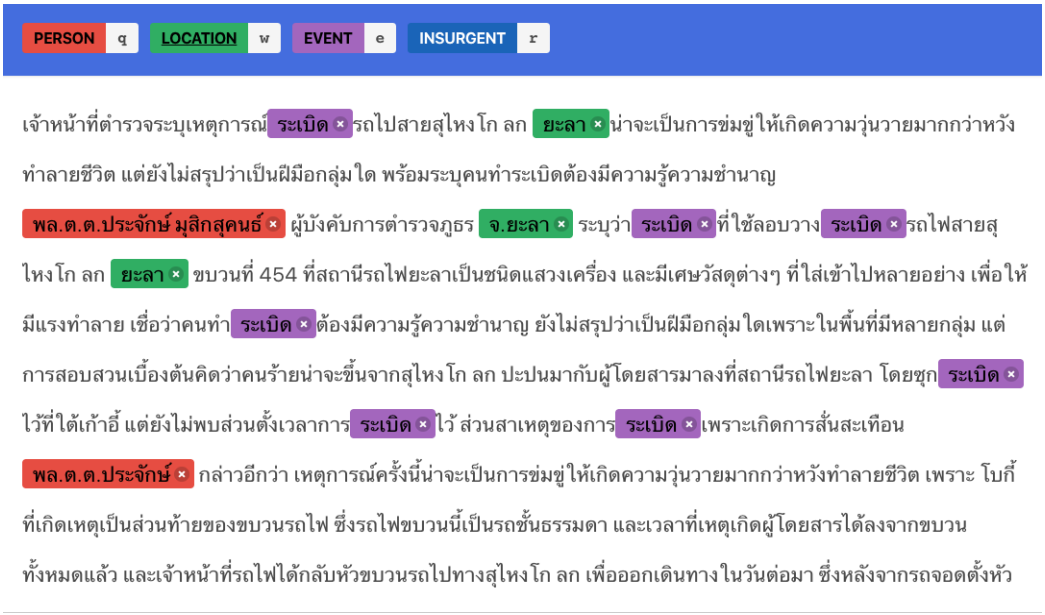
รูปที่ *26* การสร้าง *New Project* ใน *Doccano*

- เมื่อเข้าสู่ตัวโปรเจคที่สร้างขึ้นมา จากนั้นทำการเข้าเมนู Import Data เพื่อทำการอัพโหลดไฟล์ข้อมูลที่ต้องการนำไปแทรค



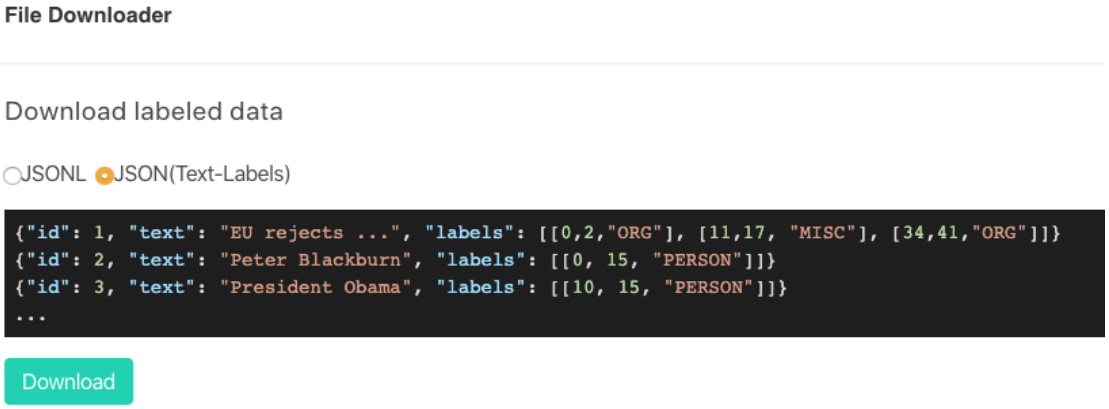
รูปที่ *27* การอัพโหลดไฟล์ข้อมูลเข้าสู่ *Doccano*

- จากนั้นทำการแทรคคำตามที่ต้องการ



รูปที่ *28* ทำการแทรคคำที่เป็นชื่อเฉพาะ

- เมื่อทำการแทรคข้อมูลสำเร็จ ให้ทำการเข้าเมนู Export Data และทำการ Export ไฟล์ออกมาในรูปแบบของ json



รูปที่ *29* การ *export* ข้อมูลของ *Doccano*

ภาคผนวก ข

โค้ดที่ใช้งาน

การสร้างโมเดล

```
!pip install pythainlp
!pip install sklearn_crfsuite
!pip install deepcut
!pip install attacut
!pip install --upgrade pandas==0.25.1
# Apache License 2.0
file_name="200full" # ชื่อไฟล์คลังข้อมูล
import codecs
from pythainlp.tokenize import word_tokenize
from pythainlp.tag import pos_tag
from nltk.tokenize import RegexpTokenizer
from pythainlp.corpus import thai_stopwords
from pythainlp.util import isthai
import glob
import nltk
import re
#จัดการประโยคซ้ำ
data_not=[]
def Unique(p):
 text=re.sub("<[^>]*>","",p)
 text=re.sub("\[(.*?)\]","",text)
 text=re.sub("\[\/(.*?)\]","",text)
 if text not in data_not:
  data_not.append(text)
  return True
```

```
 else:
  return False
# เตรียมตัวตัด tag ด้วย re
pattern = r'\[(.*?)\](.*?)\[\/(.*?)\]'
tokenizer = RegexpTokenizer(pattern) # ใช้ nltk.tokenize.RegexpTokenizer เพื่อ
ตัด [TIME]8.00[/TIME] ให้เป็น ('TIME','ไง','TIME')
# จัดการกับ tag ที่ไม่ได้ tag
def toolner_to_tag(text):
 text=text.strip()
 text=re.sub("<[^>]*>","",text)
 text=re.sub("(\[\/(.*?)\])","\\1***",text)#.replace('(\[(.*?)\])','***\\1')#  ตัดการกับพวกไม่มี tag word
 text=re.sub("(\[\w+\])","***\\1",text)
 text2=[]
 for i in text.split('***'):
  if "[" in i:
   text2.append(i)
  else:
   text2.append("[word]"+i+"[/word]")
 text="".join(text2)#re.sub("[word][/word]","","".join(text2))
 return text.replace("[word][/word]","")
# แปลง text ให้เป็น conll2002
def text2conll2002(text,pos=True):
    """
    ใช้แปลงข้อความให้กลายเป็น conll2002
    """
    text=toolner_to_tag(text) # นำไปใส่ tag [word]
    text=text.replace("""","")
    text=text.replace("’","").replace("‘","")#.replace("","")
    tag=tokenizer.tokenize(text) # แยก tag ออกมาจากข้อความ
    j=0
```

```
    conll2002="" # ประกาศตัวแปรเก็บ conll2002
    for tagopen,text,tagclose in tag: # ลูปใน tag โดยเป็น (tagopen,text,tagclose)
        word_cut=word_tokenize(text,engine="deepcut") # ใช้ตัวตัดคำ newmm ของ PyThaiNLP
        i=0
        txt5=""
        while i<len(word_cut): #ลูปตามจำนวน token ที่ตัดในtag
            if word_cut[i]=="" or word_cut[i]=="":pass
            elif i==0 and tagopen!='word': # ไม่เป็น tag [word] และเป็น i หรือตัวเริ่มต้น tag
                txt5+=word_cut[i]
                txt5+='\t'+'B-'+tagopen
            elif tagopen!='word':
                txt5+=word_cut[i]
                txt5+='\t'+'I-'+tagopen
            else: # เป็น [word]
                txt5+=word_cut[i]
                txt5+='\t'+'O'
            txt5+='\n'
            #j+=1
            i+=1
        conll2002+=txt5
    if pos==False:
        return conll2002
    return postag(conll2002) # เพิ่ม postag ใส่
# ใช้สำหรับกำกับ pos tag เพื่อใช้กับ NER
# print(text2conll2002(t,pos=False))
def postag(text):
    listtxt=[i for i in text.split('\n') if i!='']
    list_word=[]
    for data in listtxt:
        list_word.append(data.split('\t')[0])
```

```
    #print(text)
    list_word=pos_tag(list_word,engine="perceptron")
    text=""
    i=0
    for data in listtxt:
        text+=data.split('\t')[0]+'\t'+list_word[i][1]+'\t'+data.split('\t')[1]+'\n'
        i+=1
    return text
# อ่านข้อมูลจากไฟล์
def get_data(fileopen):
  """
    สำหรับใช้อ่านทั้งหมดทั้งในไฟล์ทีละรรทัดออกมาเป็น list
    """
  with codecs.open(fileopen, 'r',encoding='utf-8-sig') as f:
    lines = f.read().splitlines()
  return [a for a in lines if Unique(a)] # เอาไม่ซ้ำกัน

def alldata(lists):
    text=""
    for data in lists:
        text+=text2conll2002(data)
        text+='\n'
    return text

def alldata_list(lists):
    data_all=[]
    for data in lists:
        data_num=[]
        try:
            txt=text2conll2002(data,pos=True).split('\n') # นำไปแปลงเป็น conll2002
```

```
            for d in txt:
                tt=d.split('\t')
                if d!="":
                    if len(tt)==3:
                        data_num.append((tt[0],tt[1],tt[2]))
                    else:
                        data_num.append((tt[0],tt[1]))
            #print(data_num)
            data_all.append(data_num)
        except:
            print(data)
    #print(data_all)
    return data_all

def alldata_list_str(lists):
  string=""
  for data in lists:
    string1=""
    for j in data:
      string1+=j[0]+" "+j[1]+"  "+j[2]+"\n"
    string1+="\n"
    string+=string1
  return string

def get_data_tag(listd):
  list_all=[]
  c=[]
  for i in listd:
    if i !='':
      c.append((i.split("\t")[0],i.split("\t")[1],i.split("\t")[2]))
```

```
    else:
      list_all.append(c)
      c=[]
  return list_all
def getall(lista):
    ll=[]
    for i in lista:
        o=True
        for j in ll:
            if re.sub("\[(.*?)\]","",i)==re.sub("\[(.*?)\]","",j):
                o=False
                break
        if o==True:
            ll.append(i)
    return ll

data1=getall(get_data(file_name+".txt")) # นำคลังเข้าไป แยกออกเป็น list ละบรรทัด
datatofile=alldata_list(data1) # นำไปผ่านขั้นตอน 1 2 3 4
tt=[]
with open(file_name+"-pos.conll","w") as f:
    i=0
    while i<len(datatofile):
        for j in datatofile[i]:
            f.write(j[0]+"\t"+j[1]+"\t"+j[2]+"\n")
        if i+1<len(datatofile):
            f.write("\n")
        i+=1

with open(file_name+".conll","w") as f:
    i=0
```

```
    while i<len(datatofile):
        for j in datatofile[i]:
            f.write(j[0]+"\t"+j[2]+"\n")
        if i+1<len(datatofile):
            f.write("\n")
        i+=1


def _is_stopword(word: str) -> bool:  # เช็คว่าเป็นคำฟุ่มเฟือย
    return word in thai_stopwords()


def _count_dot(word: str) -> int:
    countnum = 0
    for i in word:
        if i == ".":
            countnum += 1
    return countnum
def _have_dot(word: str) -> bool:
    return "." in word
    from sklearn_crfsuite import scorers,metrics
from sklearn.metrics import make_scorer
from sklearn.model_selection import cross_validate,train_test_split
import sklearn_crfsuite
def doc2features(doc, i):
    word = doc[i][0]
    postag = doc[i][1]
    # Features from current word
    features={
        'word.word': word,
        'word.stopword': _is_stopword(word),
        'word.isspace':word.isspace(),
```

```
        'postag':postag,
        'word.isdigit': word.isdigit(),
        'word.isthai' : isthai(word),
        'word.havedot' : _count_dot(word)
    }
    if i > 0:
        prevword = doc[i-1][0]
        postag1 = doc[i-1][1]
        features['word.prevword'] = prevword
        features['word.previsspace']=prevword.isspace()
        features['word.prevstopword']=_is_stopword(prevword)
        features['word.prevpostag'] = postag1
        features['word.prevwordisdigit'] = prevword.isdigit()
        features['word.previsthai'] = isthai(prevword)
        features['word.prevhavedot'] = _count_dot(prevword)
    else:
        features['BOS'] = True # Special \"Beginning of Sequence\" tag
    if i > 0:
        prevword2 = doc[i-2][0]
        postag12 = doc[i-2][1]
        features['word.2prevword'] = prevword2
        features['word.2previsspace']=prevword2.isspace()
        features['word.2prevstopword']=_is_stopword(prevword2)
        features['word.2prevpostag'] = postag12
        features['word.2prevwordisdigit'] = prevword2.isdigit()
        features['word.2previsthai'] = isthai(prevword2)
        features['word.2prevhavedot'] = _count_dot(prevword2)
    else:
        features['BOS'] = True # Special \"Beginning of Sequence\" tag
    # Features from next word
```

```
    if i < len(doc)-1:
        nextword = doc[i+1][0]
        postag1 = doc[i+1][1]
        features['word.nextword'] = nextword
        features['word.nextisspace']=nextword.isspace()
        features['word.nextstopword']=_is_stopword(nextword)
        features['word.nextpostag'] = postag1
        features['word.nextwordisdigit'] = nextword.isdigit()
        features['word.nextisthai'] = isthai(nextword)
        features['word.nexthavedot'] = _count_dot(nextword)
    else:
        features['EOS'] = True # Special \"End of Sequence\" tag
    if i < len(doc)-2:
        nextword2 = doc[i+2][0]
        postag12 = doc[i+2][1]
        features['word.2nextword'] = nextword2
        features['word.2nextisspace']=nextword2.isspace()
        features['word.2nextstopword']=_is_stopword(nextword2)
        features['word.2nextpostag'] = postag12
        features['word.2nextwordisdigit'] = nextword2.isdigit()
        features['word.2nextisthai'] = isthai(nextword2)
        features['word.2nexthavedot'] = _count_dot(nextword2)
    else:
        features['EOS'] = True # Special \"End of Sequence\" tag
    return features
def extract_features(doc):
    return [doc2features(doc, i) for i in range(len(doc))]

    def get_labels(doc):
    return [tag for (token,postag,tag) in doc]
```

```
X_data = [extract_features(doc) for doc in datatofile] # เอา คำ แยกออกมา
y_data = [get_labels(doc) for doc in datatofile] # เอา tag แยกออกมา

X, X_test, y, y_test = train_test_split(X_data, y_data, test_size=0.2) # แบ่ง 0.1 หรือ 10%
crf = sklearn_crfsuite.CRF(
    algorithm='lbfgs',
    c1=0.1,
    c2=0.1,
    max_iterations=500,
    all_possible_transitions=True,
    model_filename=file_name+"-pos.model0" # ตั้งชื่อโมเดล
)
crf.fit(X, y); # train

labels = list(crf.classes_)
labels.remove('O')
y_pred = crf.predict(X_test)
e=metrics.flat_f1_score(y_test, y_pred,
                        average='weighted', labels=labels)
print(e) # โชว์ประสิทธิภาพ
sorted_labels = sorted(
    labels,
    key=lambda name: (name[1:], name[0])
)
print(metrics.flat_classification_report(
    y_test, y_pred, labels=sorted_labels, digits=3
))
def get_ner(text):
    word_cut=word_tokenize(text,engine="deepcut")
```

```
    list_word=pos_tag(word_cut,engine='perceptron')
    X_test = extract_features([(data,list_word[i][1]) for i,data in enumerate(word_cut)])
    y_=crf.predict_single(X_test)
    return [(word_cut[i],list_word[i][1],data) for i,data in enumerate(y_)]

while True:
    t=input("Text : ")
    if t == "stop":
        break
    print(get_ner(t))

alldata = []
newsagency = []
day = []
url = []
seb2 = ["---"]
seb = ["---","---","---"]
for x in range(len(data)):
    textfromdf = get_ner(data.get_value(x,"newstext"))
    for y in textfromdf:
        if y[2] != "O":
            if y[0].strip() != "":
                alldata.append(y)
                newsagency.append(data.get_value(x,"newsagency"))
                day.append(data.get_value(x,"day"))
                url.append(data.get_value(x,"url"))
    alldata.append(seb)
    newsagency.append(seb2)
    day.append(seb2)
    url.append(seb2)
```

```
dfalldata = pd.DataFrame(alldata,columns = ['word','pos','NER'])
dfalldatacomponent = pd.DataFrame(list(zip(newsagency,day,url)),columns = ['newsagency','day','
url'])
resultdf = pd.concat([dfalldata,dfalldatacomponent], axis=1, ignore_index=True)
resultdf.to_csv(r'datanewaddcomponent.csv', encoding='utf-8-sig',index=None)
#dfallpred = pd.DataFrame(list(zip(alldata,newsagency,day,url)),columns = ['word','pos','NER','new
sagency','day','url'])
dfalldata = pd.DataFrame(alldata,columns = ['word','pos','NER'])
dfalldatacomponent = pd.DataFrame(list(zip(newsagency,day,url)),columns = ['newsagency','day','
url'])
print(len(alldata))
resultdf = pd.concat([dfalldata,dfalldatacomponent], axis=1, ignore_index=True)
resultdf.to_csv(r'datanewaddcomponent.csv', encoding='utf-8-sig',index=None)
```

การดึงข้อมูลจากไฟล์ html

```
!pip install beautifulsoup4
import zipfile
import re
from bs4 import BeautifulSoup
import requests
import zipfile
archive = zipfile.ZipFile('FINAL PROJECT/Thai-News-Dataset.zip', 'r')
keyword = ["ยิง","ระเบิด","ฆ่า","ยาเสพติด","สังหาร"]
namelist = archive.namelist()
mistakelist = []
correct = 0
mistake = 0
testtext = ""
textre = ""
newsagencylist = []
urllist = []
newstextlist = []
daylist = []
for i in range(len(namelist)):
  cnt = 0
  val = False
  if i > 6 :
    if i % 2 == 1:
      news = archive.open(namelist[i])
      text = news.read()
      try :
        soup = BeautifulSoup(text.decode("UTF-8"),"lxml")
        sizenews = len(soup.findAll("td"))
        for texts in soup.findAll("td"):
```

```
            cnt = cnt +1
            if cnt == 8 and (sizenews == 18 or sizenews == 20):
              newsagency = texts.text
            if cnt == 12 and (sizenews == 18 or sizenews == 20) :
              url = texts.text
            if cnt == 14 and (sizenews == 18 or sizenews == 20) :
              newstext = texts.text.strip()
              for x in keyword:
                if x in newstext:
                  val = True
            if cnt == 16 and (sizenews == 18):
              day=texts.text
            if cnt == 18 and sizenews == 20:
              day=texts.text
          if val:
            newsagencylist.append(newsagency)
            urllist.append(url)
            newstextlist.append(newstext)
            daylist.append(day)
        except:
          print("An exception occurred")
          pass
print(len(newsagencylist))
print(len(urllist))
print(len(newstextlist))
print(len(daylist))
namelist = archive.namelist()

#159 problem
```

```
news = archive.open(namelist[79])
text = news.read()
cnt = 0
title = ""
writer = ""
newsagency = []
url = []
newstext = []
day = []
soup = BeautifulSoup(text.decode("UTF-8","lxml"))

for texts in soup.findAll("td"):
  sizenews = len(soup.findAll("td"))
  cnt = cnt +1
  if cnt == 8:
    newsagency.append(texts.text)
  if cnt == 12:
    url.append(texts.text)
  if cnt == 14:
    newstext.append(texts.text)
  if cnt == 16 and sizenews == 18:
    day.append(texts.text)
  if cnt == 18 and sizenews == 20:
    day.append(texts.text)

import pandas as pd
dfpred = pd.DataFrame(list(zip(newsagencylist,urllist,newstextlist,daylist)),columns = ['newsagenc
y','url','newstext','day'])
dfpred.to_csv(r'FINAL PROJECT/Datanewsaddcomponent.csv', encoding='utf-8-sig',index=None)
```

การสร้างข้อมูลความสัมพันธ์จากข่าว

```
!pip install pythainlp
import pandas as pd
import re
import pythainlp
import numpy as np
data = pd.read_csv("datanewaddcomponent.csv")
data = data.rename(columns={"0":"word","1":"pos","2":"tag","3":"nameagenvy","4":"date","5":"url"})
def createNetwork(news):
  result = pd.DataFrame({"word1":[],"word2":[],"nameagenvy":[],"date":[],"url":[]})
  word = news["word"].array
  tag = news["tag"].array
  nameagenvy = news["nameagenvy"].array
  data = news["date"].array
  url = news["url"].array

  lindextag = []
  for a in range(len(tag)):
    if "I-" in tag[a]:
      lindextag.append(a)



  if len(lindextag) > 0:
    for b in range(len(word)-1,-1,-1):
      for c in range(len(lindextag)-1,-1,-1):
        if len(lindextag) > 0:
          if lindextag[c] == b:
            word[b-1] = str(word[b-1]) + str(word[b])
            word[b] = np.NaN
```

```
  uword = word.unique()
  uword = uword.dropna()
  for i in range(len(uword)):
    for j in range(len(uword)):
      if i != j:
        if j < i:
          row = pd.DataFrame({"word1":[uword[i]],"word2":[uword[j]],"nameagenvy":[nameagenvy[0]],
"date":[data[0]],"url":[url[0]]})
          result = result.append(row)
  return result
test = ""
result = pd.DataFrame({"word1":[],"word2":[],"nameagenvy":[],"date":[],"url":[]})
temp = pd.DataFrame({"word":[],"pos":[],"tag":[],"nameagenvy":[],"date":[],"url":[]})
for index, row in data.iterrows():
    if row["word"] == "---":
      if temp.size > 6:
        result = result.append(createNetwork(temp))
        temp = pd.DataFrame({"word":[],"pos":[],"tag":[],"nameagenvy":[],"date":[],"url":[]})
    else :
        if pythainlp.util.isthai(row["word"].replace(" ","")):
          temp = temp.append(row)
result.to_csv(r'datatonas.csv', encoding='utf-8-sig',index=None)
```

การสร้างกราฟจากข้อมูลความสัมพันธ์จากข่าว

```
import os
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.colors as mcolors
import networkx as nx

filename = 'datatonasnew.csv'
df = pd.read_csv(filename, sep=',', index_col=False , encoding='utf-8')
#df=df.rename(columns={"word1": "NER1", "word2": "NER2"})
#new=df.loc[df['nameagenvy'] == "\xa0ประชาไท"]

df.replace(u'\xa0',u'', regex=True, inplace=True)
df
new=df.loc[df['nameagenvy'] == "ประชาไท"]
new

ner1=list(new['word1'])
ner2=list(new['word2'])
nameagenvy = list(new['nameagenvy'])

G = nx.Graph()
for i in range(len(ner1)):
    G.add_node(ner1[i],name=ner1[i])
    G.add_node(ner2[i],name=ner2[i])

for i in range(len(ner1)):
    G.add_edge(ner1[i],ner2[i],source=ner1[i].encode('utf-8').decode('utf-
8'),destination=ner2[i].encode('utf-8').decode('utf-8'))
```

```
# plt.figure(figsize=(60,40))
# pos = nx.spring_layout(G,k=0.15,iterations=20)
# nx.draw_networkx(G,pos,with_labels=False,node_size=1000,width=1,alpha=1)
# plt.axis('off')
# plt.show()

import networkx as nx
from bokeh.io import show, output_file
from bokeh.models.graphs import from_networkx
from bokeh.io import output_notebook
from bokeh.models import HoverTool, ColumnDataSource
from bokeh.resources import CDN
from bokeh.embed import file_html
from bokeh.models import ColumnDataSource, Label, LabelSet, Range1d
from bokeh.plotting import figure, output_file, show, ColumnDataSource
from bokeh.models import Plot, Range1d, MultiLine, Circle, HoverTool, BoxZoomTool, ResetTool,WheelZoomTool,SaveTool,PanTool
from bokeh.palettes import Spectral4
from bokeh.models import Plot, Range1d, MultiLine, Circle, HoverTool, TapTool, BoxSelectTool, Select,Paragraph
from bokeh.models.graphs import from_networkx, NodesAndLinkedEdges, EdgesAndLinkedNodes
from bokeh.layouts import column

import operator

betweenness_dict = nx.betweenness_centrality(G) # Run betweenness centrality
centrality_dict = nx.degree_centrality(G) # Run degree centrality
closeness_dict = nx.closeness_centrality(G) # Run closeness centrality
```

```
eigenvector_dict = nx.eigenvector_centrality(G) # Run eigenvector centrality

sorted_betweenness = sorted(betweenness_dict.items(), key=operator.itemgetter(1), reverse=Tru
e)
sorted_centrality = sorted(centrality_dict.items(), key=operator.itemgetter(1), reverse=True)
sorted_closeness = sorted(closeness_dict.items(), key=operator.itemgetter(1), reverse=True)
sorted_eigenvector = sorted(eigenvector_dict.items(), key=operator.itemgetter(1), reverse=True)
sorted_degree_lst = sorted(G.degree, key=lambda x: x[1], reverse=True)

data=(str(sorted_betweenness[0:5]))
data1=(str(sorted_centrality[0:5]))
data2=(str(sorted_closeness[0:5]))
data3=(str(sorted_eigenvector[0:5]))
data4=(str(sorted_degree_lst[0:5]))

for i in range(len(sorted_betweenness)):
    G.add_node(sorted_betweenness[i][0],betweenness=sorted_betweenness[i][1])
for i in range(len(sorted_betweenness)):
    G.add_node(sorted_centrality[i][0],centrality=sorted_centrality[i][1])
for i in range(len(sorted_betweenness)):
    G.add_node(sorted_closeness[i][0],closeness=sorted_closeness[i][1])
for i in range(len(sorted_betweenness)):
    G.add_node(sorted_eigenvector[i][0],eigenvector=sorted_eigenvector[i][1])
for i in range(len(sorted_betweenness)):
    G.add_node(sorted_degree_lst[i][0],degree=sorted_degree_lst[i][1])

TOOLTIPS = [("Name","@name"),("betweenness","@betweenness"),("centrality","@centrality"),("clos
eness","@closeness"),("eigenvector","@eigenvector"),("degree","@degree")]
plot = Plot(plot_width=1500, plot_height=600,x_range=Range1d(-1.1, 1.1), y_range=Range1d(-
1.1, 1.1))
```

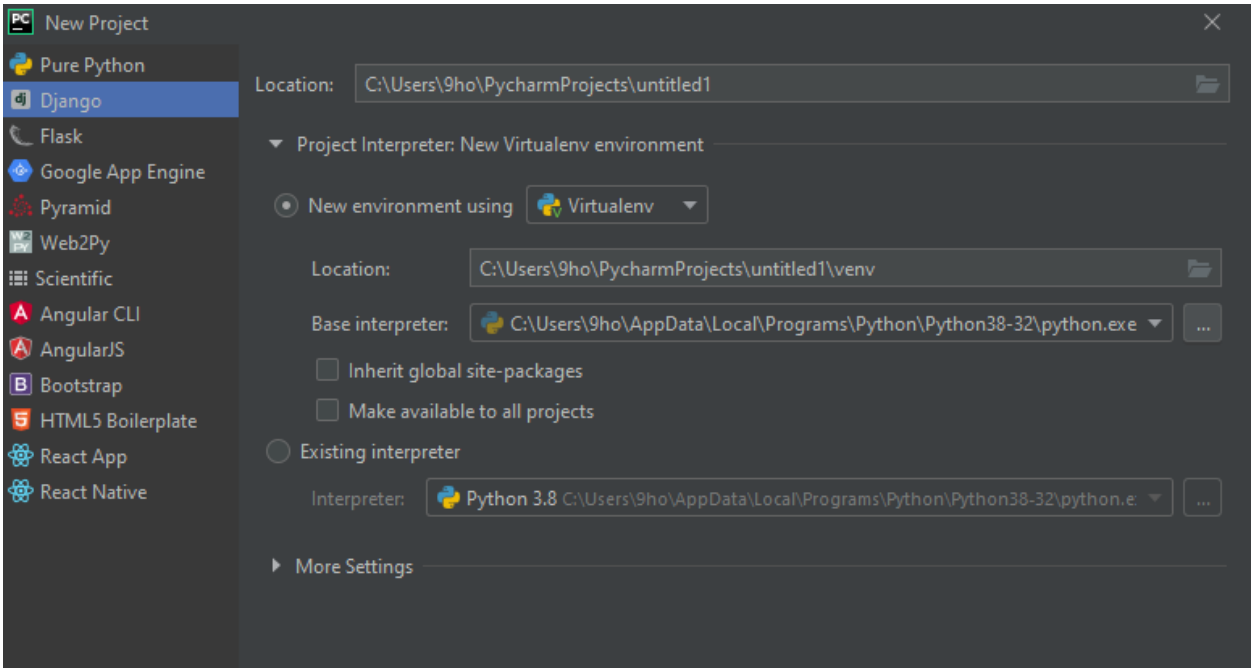
```
plot.title.text = "Network Analysis Graph"
node_hover_tool = HoverTool(tooltips=TOOLTIPS)
graph_renderer = from_networkx(G, nx.spring_layout, scale=20, center=(0, 0))
plot.add_tools(node_hover_tool,WheelZoomTool(), TapTool(), BoxSelectTool(), ResetTool(),Save
Tool(),PanTool())
graph_renderer.node_renderer.glyph = Circle(size=10, fill_color="tomato")
graph_renderer.edge_renderer.glyph = MultiLine(line_color="black", line_alpha=0.1, line_width=
1)
graph_renderer.node_renderer.selection_glyph = Circle(size=15, fill_color="green")
graph_renderer.node_renderer.hover_glyph = Circle(size=15, fill_color="navy")
graph_renderer.edge_renderer.selection_glyph = MultiLine(line_color="green", line_width=1)
graph_renderer.edge_renderer.hover_glyph = MultiLine(line_color="blue", line_width=1)
plot.renderers.append(graph_renderer)

textsorted_betweenness = "Top 5 nodes by betweenness "+ data
textsorted_centrality = "Top 5 nodes by centrality "+ data1
textsorted_closeness = "Top 5 nodes by closeness "+ data2
textsorted_eigenvector = "Top 5 nodes by eigenvector "+ data3
textsorted_degree_lst = "Top 5 nodes by degree: "+ data4

p = Paragraph(text = textsorted_betweenness,width=2000, height=20)
p1 = Paragraph(text = textsorted_centrality,width=2000, height=20)
p2 = Paragraph(text = textsorted_closeness,width=2000, height=20)
p3 = Paragraph(text = textsorted_eigenvector,width=2000, height=20)
p4 = Paragraph(text = textsorted_degree_lst,width=2000, height=20)

show(column(plot,p,p1,p2,p3,p4))
output_file("testNetwork_prototype.html")
```

การติดตั้ง Django

1.Install Pycharm จากเว็บไซต์ https://www.jetbrains.com/pycharm/

รูปที่ *30* หน้าเว็บไซต์

2.เมื่อติดตั้งเสร็จให้เลือก Django และเลือก path python ที่ได้ติดตั้งเอาไว้



รูปที่ *31* การสร้าง *Project*

การสร้างลิ้งค์ข้อมูลเพื่อไปยังหน้าเว็บไซต์ Project/urls.py

```
from django.contrib import admin
from django.urls import path, include

from django_netjsongraph.api import urls as netjsongraph_api
from django_netjsongraph.visualizer import urls as netjsongraph_visualizer

urlpatterns = [
    path('admin/', admin.site.urls),
    path('', include('blog.urls')),
    path('api/', include(netjsongraph_api)),
    path('', include(netjsongraph_visualizer)),
]
```

การสร้างหน้า View เพื่อกำหนดการ Redirect web/views.py

```
import networkx as nx
from django.shortcuts import render
from django.http import HttpResponse
import sys
from subprocess import run, PIPE
from django import template
from blog.code.test import graph_drawing, G, pos
# Create your views here.

def home(request):
    return render(request, 'blog/home.html', {'title': 'Home'})


def about(request):
    return render(request, 'blog/about.html', {'title': 'About'})


def information(request):
    return render(request, 'blog/Information.html', {'title': 'Information'})


def degree(request):
    return render(request, 'blog/degree.html', {'title': 'Degree'})


def closeness(request):
    return render(request, 'blog/closeness.html', {'title': 'Closeness'})


def betweeness(request):
    return render(request, 'blog/betweeness.html', {'title': 'Betweeness'})
```

```
def external(request):
    text_inp = request.POST.get('text')
    print(text_inp)

    out = run([sys.executable, 'blog/templates/blog/test_CNA_3.py', text_inp], shell=False, stdout
=PIPE)
    #print(out)
    return render(request, 'blog/test_CNA_3.html')

from .models import solution


def showscore(request):
    score = solution.objects.all()

    context = {'score': score}

    return render(request, 'blog/showscore.html', context)
```

การกำหนด Path ของตัวเว็บไซต์ Web/urls.py

```
from django.urls import path
from . import views

urlpatterns = [
    path('home/', views.home, name='blog-home'),
    path('about/', views.about, name='blog-about'),
    path('information/', views.information, name='blog-information'),
    path('Degree/', views.degree, name='blog-degree'),
    path('Closeness/', views.closeness, name='blog-closeness'),
    path('Betweeness/', views.betweeness, name='blog-betweeness'),
    path('external/', views.external),
    path('score/', views.showscore,name='blog-score'),
]
```

การสร้าง Base template ที่เป็นหลักของทุกหน้า

```
{% load static %}
<!DOCTYPE html>
<html>
<head>

    <!-- Required meta tags -->
    <meta charset="utf-8">
    <meta name="viewport" content="width=device-width, initial-scale=1, shrink-to-fit=no">

    <!-- Bootstrap CSS -->
    <link rel="stylesheet" href="https://maxcdn.bootstrapcdn.com/bootstrap/4.0.0/css/bootstrap.min.css"
          integrity="sha384-Gn5384xqQ1aoWXA+058RXPxPg6fy4IWvTNh0E263XmFcJlSAwiGgFAW/dAiS6JXm" crossorigin="anonymous">
    <link rel="stylesheet" type="text/css" href="{% static 'blog/bootstrap-4.3.1.css' %}">
    <link rel="stylesheet" type="text/css" href="{% static 'blog/main.css' %}">
    <link rel="stylesheet" type="text/javascript" href="{% static 'blog/jsnetworkx.js' %}">

    <script type="text/javascript" src="https://cdn.pydata.org/bokeh/release/bokeh-1.4.0.min.js"></script>
    <script type="text/javascript">
        Bokeh.set_log_level("info");
    </script>

    <script type="text/javascript" href="{% static 'blog/jsnetworkx.js' %}"></script>
    .
    <script src="https://d3js.org/d3.v5.min.js"></script>
```

```
    {% if title %}
        <title>SWU - {{ title }}</title>
    {% else %}
        <title>SWU</title>
    {% endif %}
</head>
<body>
<header class="site-header container">
    <nav class="navbar fixed-top navbar-expand-lg navbar-dark bg-dark"><a class="navbar-
brand" href="{% url 'blog-home' %}">Network
        X</a>
        <button class="navbar-toggler" type="button" data-toggle="collapse" data-
target="#navbarSupportedContent1" aria-controls="navbarSupportedContent1" aria-
expanded="false" aria-label="Toggle navigation"><span class="navbar-toggler-
icon"></span></button>
        <div class="collapse navbar-collapse" id="navbarSupportedContent1">
            <ul class="navbar-nav mr-auto">
                <li class="nav-item active"><a class="nav-link" href="{% url 'blog-
home' %}">Home <span
                        class="sr-only">(current)</span></a>
                </li>
                <!--<li class="nav-item"><a class="nav-link" href="{% url 'blog-
home' %}">Information</a></li>-->
                <li class="nav-item"><a class="nav-link" href="{% url 'blog-about' %}">About</a></li>
            </ul>
        </div>
    </nav>
</header>
```

```
<div class="row container-fluid">
    <div class="col-md-8">
        {% block content %}{% endblock %}
    </div>
    <div class="col-md-4">
        <div class="content-section">
            <h3>SOCIAL NETWORK ANALYSIS</h3>
            <p class='text-muted'>เราสามรถค่าความเป็นจุดศูนย์กลาง (Centrality) ของสมาชิกเครือข่าย
                เป็นการวัดค่าความเป็นจุดศูนย์กลางของสมาชิก
                เครือข่ายแต่ละจุด ซึ่งแตกต่างไปจากค่าความเป็นศูนย์กลางของเครือข่ายซึ่งวัดจากประชากรทั้งหมด
ในเครือข่าย วิธีการวัดค่าความเป็นจุดศูนย์กลางของสมาชิกเครือข่ายที่สำคัญมี 4 ชนิด คือ </p>
            <ul class="list-group">
                <li class="list-group-item list-group-item-light"><a class="btn-link" href="{% url 'blog-
degree' %}">Degree
                    Centrality</a></li>
                <li class="list-group-item list-group-item-light"><a class="btn-link" href="{% url 'blog-
closeness' %}">Closeness
                    Centrality</a></li>
                <li class="list-group-item list-group-item-light"><a class="btn-link"
                                                    href="{% url 'blog-
betweeness' %}">Betweeness
                    Centrality</a></li>
                <li class="list-group-item list-group-item-light"><a class="btn-link"
                                                    href="{% url 'blog-information' %}">Eigenvector
                    Centrality</a></li>
            </ul>
        </div>

    </div>
</div>
```

```
<div class="row container">
    <div class="col-md-7">
        {% block content3 %}{% endblock %}
    </div>
</div>

<script src="https://code.jquery.com/jquery-3.2.1.slim.min.js"
        integrity="sha384-
KJ3o2DKtIkvYIK3UENzmM7KCkRr/rE9/Qpg6aAZGJwFDMVNA/GpGFF93hXpG5KkN"
        crossorigin="anonymous"></script>
<script src="https://cdnjs.cloudflare.com/ajax/libs/popper.js/1.12.9/umd/popper.min.js"
        integrity="sha384-
ApNbgh9B+Y1QKtv3Rn7W3mgPxhU9K/ScQsAP7hUibX39j7fakFPskvXusvfa0b4Q"
        crossorigin="anonymous"></script>
<script src="https://maxcdn.bootstrapcdn.com/bootstrap/4.0.0/js/bootstrap.min.js"
        integrity="sha384-
JZR6Spejh4U02d8jOt6vLEHfe/JQGiRRSQQxSfFWpi1MquVdAyjUar5+76PVCmYl"
        crossorigin="anonymous"></script>
<script src="../../static/blog/jsnetworkx.js"></script>
</body>
</html>
```

หน้า Home.html

```
{% extends "blog/base.html" %}
{% load static %}
{% block content %}
    <div class="content-section">
        <form action="/external/" method="post">
            {% csrf_token %}
            <h3>สำนักข่าว</h3>
            <select id="cars" name="text" required>
                <option value="" selected disabled hidden>Choose here</option>
                <option value="กรุงเทพธุรกิจ">กรุงเทพธุรกิจ</option>
                <option value="ข่าวสด">ข่าวสด</option>
                <option value="บ้านเมือง">บ้านเมือง</option>
                <option value="โทรทัศน์ไอทีวี">โทรทัศน์ไอทีวี</option>
                <option value="ผู้จัดการ">ผู้จัดการ</option>
                <option value="ประชาไท">ประชาไท</option>
                <option value="เอ็มเว็บนิวส์">เอ็มเว็บนิวส์</option>
                <option value="ไทยรัฐ">ไทยรัฐ</option>
                <option value="มติชน">มติชน</option>
                <option value="Bangkok Post">Bangkok Post</option>
                <option value="โทรทัศน์ช่อง 9">โทรทัศน์ช่อง 9</option>
                <option value="เดลินิวส์">เดลินิวส์</option>
                <option value="คมชัดลึก">คมชัดลึก</option>
                <option value="สยามรัฐ">สยามรัฐ</option>
                <option value="แนวหน้า">แนวหน้า</option>
                <option value="โพสต์ทูเดย์">โพสต์ทูเดย์</option>
                <option value="พิมพ์ไทย-ซิงหว่า">พิมพ์ไทย-ซิงหว่า</option>
                <option value="โทรทัศน์ช่อง 3">โทรทัศน์ช่อง 3</option>
                <option value="ประชาชาติอิสลามออนไลน์">ประชาชาติอิสลามออนไลน์</option>
                <option value="โทรทัศน์ช่อง 7">โทรทัศน์ช่อง 7</option>
```

```
            <option value="ไทยโพสต์">ไทยโพสต์</option>
            <option value="INN">INN</option>
            <option value="ไฮไลท์การเมือง">ไฮไลท์การเมือง</option>
            <option value="โทรทัศน์ช่อง 11">โทรทัศน์ช่อง 11</option>
            <option value="โทรทัศน์ช่อง 5">โทรทัศน์ช่อง 5</option>
            <option value="ผู้จัดการรายวัน">ผู้จัดการรายวัน</option>
            <option value="สยามธุรกิจ">สยามธุรกิจ</option>
            <option value="โฟกัสภาคใต้">โฟกัสภาคใต้</option>
            <option value="ประชาชาติธุรกิจ">ประชาชาติธุรกิจ</option>
            <option value="ผู้จัดการรายสัปดาห์">ผู้จัดการรายสัปดาห์</option>
            <option value="ผู้จัดการออนไลน์">ผู้จัดการออนไลน์</option>
            <option value="สถาบันข่าวอิศรา สมาคมนักข่าวหนังสือพิมพ์แห่งประเทศไทย">สถาบันข่าวอิศรา
             สมาคมนักข่าวหนังสือพิมพ์แห่งประเทศไทย
            </option>
        </select><br><br>
        <input type="submit" value="Go Graph">
    </form>

</div>

{% endblock content %}
```

หน้า about

```
{% extends "blog/base.html" %}
{% load static %}
{% block content %}
    <div class="content-section">
    <p>
    <h2>
การสร้างกราฟเพื่อแสดงความสัมพันธ์ระหว่างสมาชิกในกลุ่มก่อการร้ายโดยใช้ Social Graph ร่วมกับการ
วิเคราะห์เครือข่ายทางสังคม <br>(Social Network Analysis)</h2>
        สามารถช่วยเจ้าหน้าที่ทางกฎหมายให้สามารถเข้าใจความสัมพันธ์ภายในขบวนการก่อการร้ายในเชิงลึกได้ดี
ยิ่งขึ้น นอกเหนือจากข้อมูลทางกระบวนการสอบสวนคดีอาชญากรรมของหน่วยงานราชการแล้ว แหล่งข้อมูลที่มี
ประโยชน์อีกแหล่งหนึ่งสำหรับพัฒนากราฟความสัมพันธ์ของเครือข่ายอาชญากรรมนั่นคือข้อมูลจากข่าวสาธารณะ
        ในบทความนี้ได้พัฒนาเว็บแอปพลิเคชันสำหรับการวิเคราะห์เครือข่ายอาชญากรรมจากข่าว
ด้วย social graph โดยแอปพลิเคชันประกอบด้วยสองส่วน คือ การวิเคราะห์ข้อความและการวิเคราะห์กราฟ ใน
ส่วนการวิเคราะห์ข้อความ มีการรวบรวมข่าวอาชญากรรม โดยในข่าวแต่ละฉบับจะมีการแท็กชื่อกลุ่มก่อการ
ร้าย ชื่อบุคคลและเหตุการณ์ที่สำคัญด้วยตนเองโดยใช้ Doccano ซึ่งข่าวที่ติดแท็กจะถูกใช้เป็นคลังข้อมูลสำหรับ
การ train เพื่อสร้างโมเดล Conditional Random Fields (CRFs) สำหรับการติดแท็กให้แก่คำที่เราสนใจ โดย
ผลลัพธ์ที่ได้คือ ความสัมพันธ์ที่เกี่ยวข้องกันระหว่างคำที่ถูกแท็ก ถูกสกัดและขุดหา นอกจากนี้ในส่วนการวิเคราะห์
กราฟความสัมพันธ์ของคำเหล่านั้นจะถูกป้อนเข้าสู่ NetworkX
        ซึ่งใช้ในการสร้าง social graph จากนั้นกราฟจะถูกนำมาแสดงบนเว็บแอปพลิเคชันสำหรับการสร้างภาพ
และวิเคราะห์โดยใช้ Django จากผลการสำรวจพบว่าความสัมพันธ์ระหว่างคำเหล่านี้สามารถสกัดออกมาได้จาก
ข่าวสารที่เป็นข้อความและสามารถนำไปสร้างกราฟความสัมพันธ์ของเครือข่ายของอาชญากรรมได้
</p>
</div>
{% endblock content %}
```

หน้า degree , betweenness, closeness, eigenvector

```
{% extends "blog/base.html" %}
{% load static %}
{% block content %}
    <div class="content-section">
        <h1>Degree Centrality</h1><br>
        <p class="center-block">ความเป็นจุดศูนย์กลางโดยวัดจากระดับ (Degree centrality) เป็นการค้นหาว่าสมาชิกใดบ้างที่เป็น จุดศูนย์กลางของการเชื่อมโยง (Hub) ซึ่งเป็นตำแหน่งที่มีอิทธิพลสูงสุดในเครือข่าย วัดได้จากจำ นวนเส้นเชื่อมโยงทั้งหมดที่โยงมาจากสมาชิกเครือข่ายอื่นๆทั้งที่อยู่ภายในกลุ่มเดียวกันและข้ามกลุ่ม
</p><br>
        <div class=center>
            <img src="{% static 'images/graph.PNG' %}" alt="Placeholder image" width="1000" height
="1000"
                 class="img-fluid center">

        </div>
    </div>


{% endblock content %}
{% extends "blog/base.html" %}
{% load static %}
{% block content %}
    <div class="content-section">
        <h1>Betweeness Centrality</h1><br>
        <p>ความเป็นจุดศูนย์กลางโดยวัดจากการคั่นกลาง (Betweenness centrality) เป็นการค้นหาว่าสมาชิกใดบ้างที่มีตำแหน่งเป็นสะพาน (Bridgers) เชื่อมกลุ่มต่าง ๆ ที่อยู่ห่างกันให้เข้าหากัน ทำ หน้าที่เป็นนายหน้าหรือตัวกลางในการติดต่อเชื่อมโยงระหว่างสมาชิกอื่นๆ และมีบทบาทสำ คัญในการนำ นวัตกรรม ความเจริญ รวมทั้งผลกระทบต่าง ๆ เข้ามาในเครือข่าย ใช้วิธีคำ นวณจากสัดส่วนของระยะทางที่สั้นที่สุด (Geodesic path) ในการเชื่อมโยงระหว่างสมาชิกแต่ละคู่ ยกตัวอย่างเช่น A เชื่อมโยงกับ B และ C ในขณะที่ B เชื่อมโยงต่อไป
```

ยัง D และ E ส่วน C เชื่อมโยงต่อไปยัง F และ G ดังนั้น A จะมีค่าคั่นกลางสูงที่สุดเนื่องจากมีเส้นทางเชื่อมโยงไปได้ทั้งสองทางและมีระยะทางที่สั้นที่สุด ส่วน B และ C มีค่าคั่นกลางด้วยเช่นกันเพราะอยู่ระหว่าง A กับสมาชิกอื่น แต่ D,E,F,G ซึ่งเป็นปลายทางจะมีค่าคั่นกลางเป็นศูนย์

```
</p>
<br>
        <div class=center>
            <img src="{% static 'images/graph.PNG' %}" alt="Placeholder image" width="1000"
                height="1000"
                class="img-fluid center">


        </div>
    </div>
{% endblock content %}
{% extends "blog/base.html" %}
{% load static %}
{% block content %}
<div class="content-section">
        <h1>Closeness Centrality</h1><br>
        <p>ความเป็นจุดศูนย์กลางโดยวัดจากความใกล้ชิด (Closeness centrality) เป็นการค้นหาว่าสมาชิกใดบ้างที่มีตำแหน่งเป็นจุดศูนย์กลางของเครือข่าย มีความใกล้ชิดกับสมาชิกอื่นและใช้ระยะทางที่สั้นที่สุดในการเข้าถึง (Geodesic path distance) วัดได้จากจำ นวนเส้นเชื่อมโยงทั้งหมดที่ใช้ในการเดินทางจากสมาชิกหนึ่งไปยังอีกสมาชิกหนึ่ง โดยการลากผ่านสมาชิกอื่น ๆ ภายในเครือข่ายด้วยเส้นทางที่สั้นที่สุด
            </p><br>
        <div class = center>
            <img src="{% static 'images/graph.PNG' %}" alt="Placeholder image" width="1000" height
="1000" class="img-fluid center">
        </div>
    </div>
```

```
{% endblock content %}
{% extends "blog/base.html" %}
{% load static %}
{% block content %}
    <div class="content-section">
        <h1>Eigenvector Centrality</h1><br>
        <p class="center-block">ความเป็นจุดศูนย์กลางโดยวัดจากเวกเตอร์
ลักษณะเฉพาะ (Eigenvector centrality) เป็นการวัดค่าอิทธิพล
ของสมาชิกเครือข่ายโดยใช้หลักการคือ สมาชิกที่เชื่อมโยงกับสมาชิกอื่นที่มีค่าอิทธิพลสูงอยู่แล้วจะมีค่าเวกเตอร์
ลักษณะเฉพาะ
ที่สูงกว่าสมาชิกที่เชื่อมโยงกับสมาชิกอื่นที่มีค่าอิทธิพลต่ำ ตัวอย่างของการวัดด้วยเวกเตอร์ลักษณะเฉพาะ
ดังกล่าว เช่น การ
วัดและจัดอันดับความสำ คัญของเว็บเพจของกูเกิ้ล (Google PageRank) เป็นต้น</p><br>
        <div class=center>
            <img src="{% static 'images/graph.PNG' %}" alt="Placeholder image" width="1000" height
="1000" class="img-fluid center">
        </div>
    </div>
{% endblock content %}
```