การระบุตัวบุคคลโดยการวิเคราะห์รูปแบบที่ได้จากการพิมพ์
Person Authentication Using Typing Pattern Analysis

นายอรรถชัย มาลาอุตม์
นายสิรวิชญ์ ไพรสณฑ์วัฒน

ปัญหาพิเศษนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรบัณฑิต (วิทยาการคอมพิวเตอร์)
ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ปีการศึกษา 2561

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การยืนยันตัวตนในการใช้งานคอมพิวเตอร์สามารถทำได้หลายรูปแบบ ทั้งการให้ผู้ใช้จดจำข้อมูล เช่น การใช้ชื่อผู้ใช้และรหัสผ่าน คำถามกันลืม หรือให้ผู้ใช้พกสิ่งของบางอย่าง เช่น บัตรเอทีเอ็ม เหล่านี้เป็นการใช้สิ่งที่สร้างขึ้นมาเพื่อใช้ยืนยันตัวตนโดยเฉพาะ แล้วมอบให้ผู้ใช้แต่ละคน ซึ่งเป็นวิธีการที่ได้รับความนิยมในปัจจุบัน แต่ปัญหาจากวิธีการดังกล่าวคือ ผู้ใช้งานอาจมีการหลงลืม และสามารถถูกขโมยได้หากไม่ระวัง เพราะรหัสผ่านที่ถูกตั้งขึ้นจากผู้ใช้ส่วนใหญ่มักมาจากชุดข้อมูลใกล้ตัว เช่น วันเกิด เบอร์โทรศัพท์ ฯลฯ วิธีการระบุตัวบุคคลอีกวิธีหนึ่ง คือการใช้ข้อมูลทางชีวภาพของแต่ละบุคคลเข้ามาช่วยในการยืนยันตัวตน โดยข้อมูลทางชีวภาพนี้สามารถแบ่งออกเป็น 2 ประเภทหลัก ๆ คือ ข้อมูลทางกายภาพ และข้อมูลทางพฤติกรรม ตัวอย่างข้อมูลทางกายภาพที่นำมาช่วยในการยืนยันตัวตน เช่น การสแกนม่านตา การสแกนลายนิ้วมือ ส่วนข้อมูลทางพฤติกรรมที่สามารถนำมาช่วยในการยืนยันตัวตน เช่น วิธีการออกเสียง พฤติกรรมการพิมพ์ของแต่ละบุคคล ซึ่งข้อมูลทางชีวภาพนี้เป็นสิ่งที่ขโมยไปจากตัวบุคคลได้ยาก และค่อนข้างมีความแม่นยำ

พฤติกรรมการพิมพ์เป็นสิ่งที่ผู้ทำปัญหาพิเศษให้ความสนใจ โดยผู้ทำปัญหาพิเศษได้ตั้งสมมติฐานว่า พฤติกรรมการพิมพ์ของแต่ละบุคคลนั้นมีความแตกต่างกัน ถ้าหากพิจารณารูปแบบการพิมพ์อย่างถี่ถ้วนแล้ว เราอาจสามารถระบุตัวบุคคลออกมาได้ เพราะพฤติกรรมการพิมพ์เป็นเอกลักษณ์เฉพาะบุคคล และวิธีการนี้ไม่ต้องการอุปกรณ์ที่เฉพาะ เมื่อเทียบกับวิธีการสแกนม่านตาหรือวิธีการสแกนลายนิ้วมือ ซึ่งต้องมีเครื่องสแกนโดยเฉพาะ จึงใช้งานได้ง่ายขึ้น

ปัญหาพิเศษนี้ จึงได้นำเสนอวิธีการระบุตัวบุคคล โดยการวิเคราะห์รูปแบบที่ได้จากการพิมพ์ โดยใช้เทคนิคการทำเหมืองข้อมูล และเทคนิคการเปรียบเทียบกันของกราฟ ที่เป็นตัวแทนของรูปแบบการพิมพ์ของแต่ละบุคคล ซึ่งวิธีการที่นำเสนอ สามารถนำไปใช้ประโยชน์ในการระบุตัวบุคคลจากรูปแบบการพิมพ์ เช่น การทำข้อสอบออนไลน์ เป็นต้น

1.2 วัตถุประสงค์ของการดำเนินงาน

1) เพื่อระบุตัวบุคคลจากรูปแบบการพิมพ์
2) เพื่อศึกษาว่าวิธีใดบ้าง ที่สามารถนำมาช่วยในการวิเคราะห์รูปแบบการพิมพ์ของผู้ใช้ได้ และวิธีใดที่มีความแม่นยำมากที่สุด
3) เพื่อสร้างซอฟต์แวร์ที่ช่วยระบุตัวบุคคลจากรูปการพิมพ์

1.3 ขอบเขตการศึกษา

1) ข้อมูลรูปแบบการพิมพ์ที่นำมาวิเคราะห์ในการทดลอง มาจากนักศึกษา อาจารย์ และบุคคลทั่วไป ในบริเวณสถาบันเทคโนโลยีพระจอมเกล้า เจ้าคุณทหาร ลาดกระบัง มีจำนวนประมาณ 100 คน
2) แป้นพิมพ์ที่นำมาให้ผู้เข้าร่วมการทดลองใช้พิมพ์ข้อมูล เป็นแป้นพิมพ์ที่ผู้ทำการทดลองจัดเตรียมไว้ให้
3) บทความที่นำมาให้ผู้เข้าร่วมการทดลองใช้พิมพ์ข้อมูล เป็นบทความภาษาอังกฤษเป็นบางส่วนจากแบบเรียนสำหรับผู้ใช้ภาษาอังกฤษเป็นภาษาที่ 2 ชื่อว่า “A Puma at large” และบทความภาษาไทยมาจากข้อความที่เป็นที่นิยมสำหรับทดสอบฟอนต์

1.4 ขั้นตอนการดำเนินงาน

1) ศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้อง
2) เก็บข้อมูลที่จะนำมาทดลอง
3) วิเคราะห์ข้อมูลด้วยวิธีการต่าง ๆ ที่ได้ศึกษามา
4) วัดประสิทธิภาพของวิธีการต่าง ๆ และทดสอบทางสถิติ
5) สรุปผลการดำเนินงาน

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1) สามารถนำองค์ความรู้ที่ได้ ไปช่วยระบุตัวบุคคลของผู้ใช้งาน นอกเหนือจากการป้อนชื่อผู้ใช้และรหัสผ่าน (ซึ่งเป็นวิธีการที่นิยมมากที่สุดในปัจจุบัน) ถือเป็น 2-Method Authentication
2) สามารถนำองค์ความรู้ที่ได้ ไปใช้งานระบุตัวบุคคลในระบบที่เกี่ยวข้องกับการพิมพ์ เช่น เว็บไซต์ที่ทำการเรียนการสอนแบบออนไลน์ เพื่อยืนยันผู้ใช้ตัวจริง ในขณะที่มีการเข้าใช้งานระบบอยู่
3) ได้ทราบถึงวิธีการที่เหมาะสม และชุดของค่าพารามิเตอร์ต่าง ๆ ในการระบุตัวบุคคล โดยการวิเคราะห์รูปแบบที่ได้จากการพิมพ์

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงทฤษฎีและงานวิจัยที่เกี่ยวข้อง โดยเนื้อหาประกอบด้วย การยืนยันตัวตน ไบโอเม็ตตริก คีย์สโตรคไดนามิค วิธีการ Dynamic Time Warping โดยจะแสดงรายละเอียดในหัวข้อที่ 2.1 ถึง 2.5

2.1 การยืนยันตัวตน (Authentication)

การยืนยันตัวตน คือการตรวจสอบผู้ใช้ ก่อนที่จะอนุญาตให้เข้าใช้งานในระบบ เพื่อพิสูจน์ว่าผู้ที่ต้องการจะเข้าใช้งานในระบบนี้คือใคร และผู้ใช้คนนี้มีสิทธิ์เข้าถึงทรัพยากรได้แค่ไหน โดยมีข้อมูลอยู่ 3 ประเภทหลัก ๆ ที่จะสามารถนำมาพิสูจน์ตัวบุคคลได้ คือ สิ่งที่ผู้ใช้รู้ สิ่งที่ผู้ใช้มี สิ่งที่ผู้ใช้เป็น

สิ่งที่ผู้ใช้รู้ (Something you know) คือการสมมติชุดข้อมูลขึ้นมาชุดหนึ่ง และเป็นที่รู้กันระหว่างผู้ใช้กับระบบเท่านั้น เป็นวิธีที่ได้รับความนิยมมากที่สุดในปัจจุบัน เช่น การใช้รหัสผ่าน ชุดตัวเลข PINs การที่จะสร้างรหัสผ่านชุดหนึ่งให้มีความปลอดภัยมากที่สุดเพื่อป้องกันการถูกสวมรอยนั้น จำเป็นต้องมีกฎมากมาย เช่น ต้องมีความยาวมากกว่า 8 ตัวอักษร มีตัวอักษรพิมพ์เล็กและตัวอักษรพิมพ์ใหญ่ประกอบกัน และต้องไม่ใช้รหัสผ่านนี้กับระบบอื่น ๆ เป็นต้น และเมื่อทำตามกฎนี้อย่างครบถ้วน ชุดรหัสผ่านที่ได้มาก็ยากที่จะจำเสียแล้ว

สิ่งที่ผู้ใช้มี (Something you have) คือการสร้างสิ่งของบางอย่างขึ้นมาเพื่อมอบให้ผู้ใช้ เช่น บัตรเครดิต มักจะใช้คู่กับชุดข้อมูลด้านบนเพื่อความปลอดภัยที่มากขึ้น แต่ผู้ใช้ก็ต้องพกสิ่งนี้ติดตัวอยู่ตลอดเวลา ซึ่งบางครั้งก็อาจมีการหลงลืมไปบ้าง

สิ่งที่ผู้ใช้เป็น (Something you are) คือการยืนยันตัวตนด้วยข้อมูลทางชีวภาพของผู้ใช้เองหรือเรียกอีกอย่างว่าไบโอเม็ตตริก เช่น ใบหน้า ม่านตา ลายนิ้วมือ น้ำเสียง การพิมพ์ เป็นต้น ซึ่งเป็นวิธีที่สะดวก และปลอดภัยมากที่สุด เมื่อเทียบกับวิธีที่ผ่านมา เนื่องจากผู้ใช้ไม่ต้องจำชุดข้อมูลใหม่ และไม่ต้องพกสิ่งของเพิ่มเติม แต่ข้อจำกัดต่าง ๆ ก็จะแตกต่างกันออกไป ขึ้นอยู่กับว่าระบบเลือกที่จะใช้ข้อมูลทางชีวภาพแบบใด

2.2 ไบโอเม็ตตริก (Biometric)

หัวข้อนี้จะกล่าวถึงไบโอเม็ตตริก ซึ่งเป็นเทคโนโลยีหลัก และเป็นหัวใจสำคัญของปัญหาพิเศษนี้

2.2.1 ความหมายของไบโอเม็ตตริก

ไบโอเม็ตตริก คือ กระบวนการตรวจสอบและจดจำลักษณะเฉพาะบุคคล โดยอิงจากลักษณะทางชีวภาพ

ลักษณะทางชีวภาพ สามารถจำแนกได้ 2 ลักษณะ ดังตารางที่ 2.1

ตารางที่ 2.1 สิ่งที่เป็นลักษณะทางกายภาพ และสิ่งที่เป็นลักษณะทางพฤติกรรม

ลักษณะทางกายภาพ	ลักษณะทางพฤติกรรม
• ลายนิ้วมือ • ใบหน้า ใบหู • ม่านตา จอประสาทตา • รูปมือ รูปนิ้ว • การไหลเวียนของเลือด	• การออกเสียง • ลายเซ็นต์ • การพิมพ์

2.2.2 ประโยชน์ของไบโอเม็ตตริก

ไบโอเม็ตตริก สามารถนำมาใช้ในการยืนยันตัวบุคคล ก่อนที่จะมีการอนุญาตให้เข้าใช้ระบบ ซึ่งมีประสิทธิภาพมากกว่าการยืนยันตัวตนที่เป็นที่นิยมในปัจจุบัน เช่น บัตรเอทีเอ็ม กุญแจ รหัสผ่าน เหล่านี้เป็นการใช้สิ่งสมมติขึ้นมาแทนตัวตนจริง ๆ ของผู้ใช้ และมอบให้ผู้ใช้แต่ละคนเพื่อใช้ในการยืนยันตัวตน มีข้อเสียคือ สามารถถูกขโมย ทำซ้ำ และหลงลืมได้ แต่ลักษณะทางชีวภาพเป็นสิ่งที่ติดตัวเราอยู่เสมอ การสวมรอยจากผู้ไม่หวังดีนั้นทำได้ยากกว่า มีความแม่นยำสูงจึงเชื่อถือได้ ผู้ใช้มีความสะดวกในการใช้งาน และอาจมีราคาถูกกว่าวิธีการแบบดั้งเดิม

2.2.3 การยืนยันตัวตนด้วยไบโอเม็ตตริก

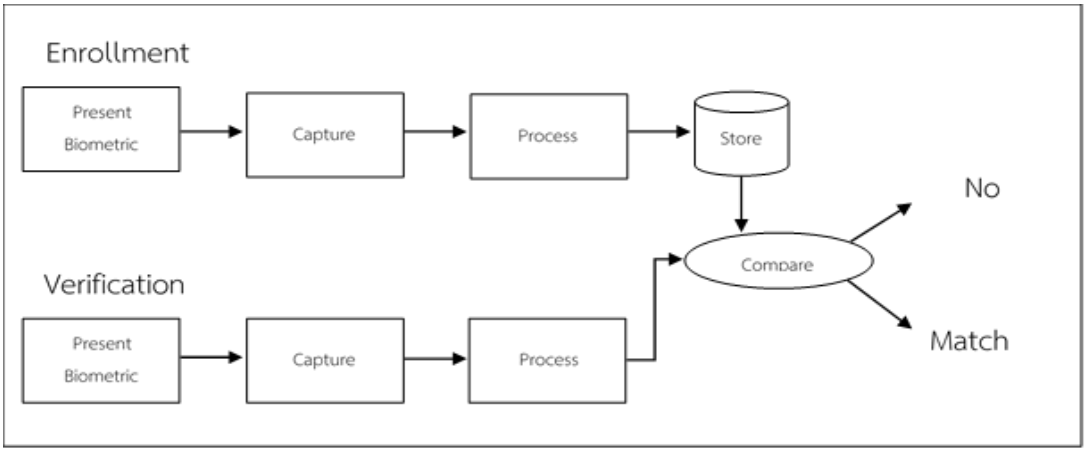
ระบบจะมีข้อมูลอยู่ 2 ชุด คือ 1. ข้อมูลลักษณะทางชีวภาพที่ได้มาจากการลงทะเบียนและนำไปประมวลผลเพื่อเก็บลงฐานข้อมูล และ 2. ข้อมูลลักษณะทางชีวภาพที่ได้มาจากการขอเข้าใช้ระบบ ซึ่งเป็นข้อมูลชุดใหม่ ข้อมูลชุดนี้ก็จะถูกนำไปประมวลผล และนำมาเปรียบเทียบกับข้อมูลที่มีในฐานข้อมูล หากไม่ตรงกับชุดข้อมูลใด ๆ ในฐานข้อมูล ระบบก็จะไม่อนุญาตให้มีการเข้าใช้งาน

ความแตกต่างของระบบการยืนยันตัวตนด้วยไบโอเม็ตตริกกับระบบแบบดั้งเดิมคือ การแทนที่ชื่อผู้ใช้และรหัสผ่าน ด้วยชุดข้อมูลลักษณะทางชีวภาพ ซึ่งสะดวกกว่ากันมาก

ระบบไบโอเม็ตตริกจะแบ่งการทำงานออกเป็น 2 ส่วน คือส่วนลงทะเบียน (Enrollment) และส่วนยืนยันตัวตน (Verification) โดยส่วนลงทะเบียนจะรับข้อมูลเข้ามา (Present Biometric) ซึ่งขึ้นอยู่กับระบบไบโมเม็ตตริกที่ใช้ เช่น การทาบลายนิ้วมือ ก็จะได้ข้อมูลนำเข้าเป็นภาพลายนิ้วมือที่มีร่องเล็ก ๆ ต่าง ๆ มากมาย จากนั้นก็จะทำการแปลงข้อมูลให้อยู่ในรูปเดียวกัน (Capture) เพื่อให้ง่ายต่อการนำไปประมวลผล (Process) และเก็บลงฐานข้อมูล (Store) เพื่อนำไปใช้ในการเปรียบเทียบข้อมูลใหม่ที่จะรับเข้ามา (Compare)

ในส่วนยืนยันตัวตนก็จะทำเช่นเดียวกันคือ รับข้อมูลเข้ามา แปลงข้อมูลให้อยู่ในรูปแบบเดียวกัน นำไปประมวลผล และเก็บลงฐานข้อมูล เพื่อนำไปใช้เปรียบเทียบกับข้อมูลที่ลงทะเบียนไว้แล้วในฐานข้อมูล

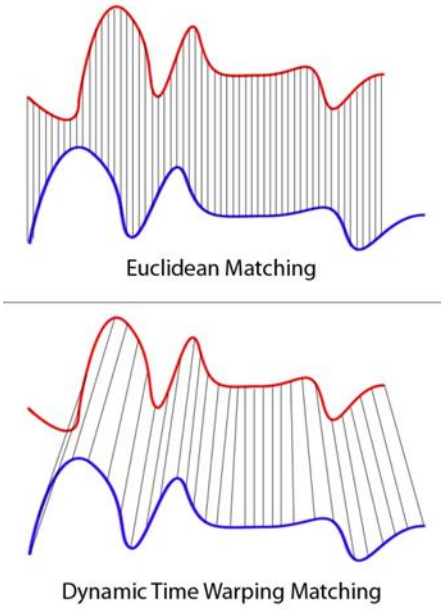
เมื่อมีการเปรียบเทียบเกิดขึ้น ระบบก็จะต้องตัดสินใจให้ได้ว่า ลักษณะทางชีวภาพที่รับเข้ามานั้น ตรงกับลักษณะทางชีวภาพใดในชุดข้อมูลหรือไม่ การทำงานของระบบยืนยันตัวตนด้วยไบโอเม็ตตริกแสดงตามภาพที่ 2.2



ภาพที่ 2.1 ระบบยืนยันตัวตนด้วยไบโอเมต็ตตริก

2.3 ไดนามิกไทม์วอร์ปปิง (Dynamic Time Warping)

ไดนามิกไทม์วอร์ปปิง เป็นขั้นตอนวิธีที่ใช้สำหรับการเปรียบเทียบความคล้ายกันระหว่างอนุกรมเวลา 2 ชุด ซึ่งจะให้ผลลัพธ์ออกมาเป็นค่าระยะทางและวิธีปรับแนว (Alignment) ที่ดีที่สุดระหว่างข้อมูลทั้งสอง ดังแสดงในภาพที่ 2.3



ภาพที่ 2.2 ภาพการเปรียบเทียบระหว่างวิธีการเปรียบเทียบแบบยูคลิเดียน และวิธีการเปรียบเทียบแบบไดนามิกไทม์วอร์ปปิ้ง

ขั้นตอนวิธีของไดนามิกไทม์วอร์ปปิง

สร้างเมทริกซ์ของข้อมูลอนุกรมเวลา 2 ชุด ที่จะเปรียบเทียบกัน

X : (x1,x2,...,xM)

Y : (y1,y2,...,yN)

Matrix A[i,j]

บันทึกผลต่างระหว่าง x1 กับ x2 ไปจนถึง xM กับ yN เก็บลงในเมทริกซ์

ตัวอย่างเช่น เมื่อกำหนดจุดสองจุด X = [1, 1, 2, 3, 2, 0] และ Y = [0, 1, 1, 2, 3, 2, 1] เมทริกซ์ A จะถูกสร้างขึ้น ดังตารางที่ 2.2

ตารางที่ 2.2 ตารางผลการคำนวณ

1	0	0	1	2	1	1
2	1	1	0	1	0	2
3	2	2	1	0	1	3
2	1	1	0	1	0	2
1	0	0	1	2	1	1
1	0	0	1	2	1	1
0	1	1	2	3	2	0
	1	1	2	3	2	0

ค่าในแต่ละช่อง คือ ผลต่างระหว่าง X กับ Y absolute (| |) เริ่มที่ 1,1 ไปจนถึง M,N หาเส้นทางที่ผลรวมน้อยที่สุด ยิ่งผลรวมระยะทางน้อยเท่าไร ความคล้ายกันก็มากเท่านั้น

2.4 งานวิจัยที่เกี่ยวข้อง

หัวข้อนี้จะกล่าวถึงงานวิจัยที่เกี่ยวข้อง ซึ่งประกอบไปด้วยงานวิจัยเรื่อง Learning User Keystroke Patterns for Authentication ซึ่งแต่ละงานวิจัยจะถูกกล่าวถึงในหัวข้อที่ 2.4.1

2.4.1 Learning User Keystroke Patterns for Authentication

Keystroke authentication คือ การเข้าสู่ระบบด้วยการยืนยันตัวตนโดยวิเคราะห์จากพฤติกรรมการพิมพ์ของผู้ใช้ โดยการทดลองนี้จะนำเทคนิค Machine learning เช่น Decision Tree, Naive Bayesian, Instance Base Learning, Decision Table, One Rule, Random Tree, K-star เข้ามาปรับใช้ ทั้งหมดนี้เป็นการแก้ปัญหาแบบ Classification และจะมีแค่ 3 เทคนิคเท่านั้นที่จะลงลึกถึงรายละเอียด ผลลัพธ์แสดงให้เห็นว่า การใช้ Machine learning เข้ามาช่วยนั้นเป็นตัวเลือกที่เหมาะสมและมีความแม่นยำกว่า เมื่อเทียบกับการใช้เทคนิค Nearest Neighbor โดยเฉพาะเทคนิค Decision Tree นั้นให้ผลลัพธ์ที่ดีกว่ามาก ผลลัพธ์ที่ได้ยังแสดงให้เห็นอีกว่าสำหรับการทำ Feature Extractions นั้น ในบรรดา N-Gram(s) 3-Gram แม่นยำที่สุด และการมี Attribute มากก็มีแนวโน้มที่จะเพิ่มค่าความแม่นยำ (Accuracy) ได้ สรุปผลการทดลองแสดงในตารางที่ 2.3-2.5

ตารางที่ 2.3 ตารางเปรียบเทียบประสิทธิภาพของแต่ละโมเดล

TABLE I
COMPARISON OF LEARNING METHODS (AVERAGE OF 5,000 SPLITS, 3-GRAMS)

Learning Methods	Training	Test Accuracy
C4.5 Decision Tree	95.6%	93.3%
Naive Bayesian	93.3%	90.8%
K-star	100%	85.6%
Decision table	95.6%	81.1%
Random Tree	100%	77.8%
OneR	91.3%	75.2%
IB KNN		
$(k = 8)$	90.2%	87.4%
$(k = 7)$	91.1%	89.4%
$(k = 5)$	93.3%	91.1%
$(k = 1)$	100%	81.5%

ตารางที่ 2.4 ตารางเปรียบเทียบประสิทธิภาพของโมเดล เมื่อใช้จำนวนแอททริบิวต์ที่ต่างกัน

TABLE II
COMPARISON OF THE NUMBER OF ATTRIBUTES

Algorithm	4-Attributes.	$M_e + E_e$	$M_d + E_d$	$M_e + M_d$	$E_e + E_d$
Naive Bayesian	90.8%	39.6%	72.9%	81.1%	84.6%
J48 Decision Tree	93.3%	43.8%	70.8%	82.2%	87.9%
IB KNN					
$(k = 7)$	89.4%	39.6%	77.3%	82.9%	83.3%
$(k = 5)$	91.1%	45.8%	77.1%	85.7%	86.9%
$(k = 1)$	81.5%	43.8%	62.5%	77.7%	80.1%

ตารางที่ 2.5 ตารางเปรียบเทียบประสิทธิภาพของโมเดล เมื่อใช้จำนวนของคำที่ต่างกัน

TABLE III
COMPARISON OF N-GRAMS

Algorithm	2-Grams	3-Grams	4-Grams
Naive Bayesian	82.8%	90.8%	75.6%
J48 Decision Tree	83.7%	93.3%	80.3%
IB KNN			
$(k = 7)$	79.8%	89.4%	75.6%
$(k = 5)$	84.2%	91.1%	82.1%
$(k = 1)$	73.1%	81.5%	71.1%

2.4.2 Keystroke-Based Authentication by Key Press Intervals as a Complementary Behavioral Biometric

ไบโอเม็ตตริกเป็นการศึกษาการใช้งานสรีรวิทยาและลักษณะพฤติกรรมที่ระบุตัวบุคคล เช่น ลักษณะทางสรีระวิทยา, ลายนิ้วมือ, รูปแบบหลอดเลือดจอประสาทตา, ลักษณะใบหน้า เป็นลักษณะทางชีวภาพที่มีลักษณะเฉพาะและ ไม่เปลี่ยนแปลงเว้นแต่ว่าจะได้รับการเปลี่ยนแปลงจากอันตรายทางร่างกาย แต่ลักษณะพฤติกรรมเช่นลายเซ็นที่เขียนด้วยลายมือ รูปแบบเสียงและการกดแป้น เป็นรูปแบบพฤติกรรมที่ถือได้ว่าเป็นเอกลักษณ์ ซึ่งลักษณะเหล่านี้เกิดขึ้นจากลักษณะทางสรีรวิทยาบางอย่าง (เช่น ขนาดมือและกล้ามเนื้อนิ้วมีผลต่อรูปแบบการกดแป้นพิมพ์) รวมทั้งปัจจัยทางจิตวิทยาและสิ่งแวดล้อม แต่พฤติกรรมอาจเปลี่ยนแปลงได้ตลอดชีวิต

เมื่อพยายามที่จะเข้าถึงระบบคอมพิวเตอร์ โดยปกติจะมีการรับรองความถูกต้องผ่านชื่อผู้ใช้และรหัสผ่าน ซึ่งได้รับการพิสูจน์แล้วว่าการตรวจสอบแบบนี้ไม่ได้ปลอดภัยอย่างสมบูรณ์ แต่การวิเคราะห์รูปแบบการกดแป้นพิมพ์ สามารถใช้เป็นมาตราวัดเสริมเพื่อป้องกันการบุกรุกได้ การวิเคราะห์รูปแบบการกดแป้นพิมพ์แตกต่างจากวิธีอื่น เพราะเป็นการปฏิสัมพันธ์ระหว่างคอมพิวเตอร์กับมนุษย์โดยตรง การพิสูจน์ตัวตนด้วยการกดแป้นพิมพ์เป็นสาขาย่อยของการรักษาความปลอดภัยของระบบคอมพิวเตอร์ การศึกษาจำนวนมากได้แสดงให้เห็นถึงเอกลักษณ์ของรูปแบบการกดแป้นพิมพ์ของแต่ละบุคคล ในระบบที่ใช้รูปแบบการกดแป้นเป็นส่วนหนึ่งในการรับรองความถูกต้องของกระบวนการเข้าสู่ระบบไม่เพียง แต่ต้องชื่อผู้ใช้และรหัสผ่านที่ถูกต้อง แต่ยังจับคู่เฉพาะรูปแบบการกดแป้นพิมพ์ที่ได้รับการระบุและจัดเก็บสำหรับแต่รายบุคคล กลไกดังกล่าวสามารถเพิ่มความยากของผู้บุกรุกได้ เนื่องจากรูปแบบการพิมพ์ที่ของแต่ละคนเป็นเอกลักษณ์จึงเป็นการยากที่จะทำซ้ำ

ใช้ช่วงเวลากดแป้นอาจมีการสร้างตัวระบุลายเซ็นที่มีประสิทธิภาพมากขึ้น การใช้งานไม่จำเป็นต้องมีการคำนวณมากเกินไปและส่วนใหญ่เป็นแพลตฟอร์มอิสระ เป้าหมายหลักของการพิสูจน์ตัวตนด้วยช่วงเวลาการกดแป้นพิมพ์การกำหนดเวลาคือการป้องกันผู้บุกรุก โดย False Rejection Rate ควรมีค่ามากและ False Accept Rate ควรมีค่าน้อย

ในบทความนี้มุ่งเน้นไปที่กลุ่มผู้ใช้คอมพิวเตอร์โดยเฉพาะ ผู้ที่ใช้แป้นพิมพ์เป็นประจำนี่คือปัญหาที่สำคัญเพราะโดยปกติมนุษย์จะมีการพิมพ์ที่ไม่เหมือนเดิม ถ้าคีย์บอร์ดเปลี่ยนแปลง การคุ้นเคยกับแป้นพิมพ์ที่ไม่รู้จักต้องใช้เวลา ผลที่ตามมา การกำหนดเวลาช่วงเวลา keystroke เหมาะสมที่สุดสำหรับผู้ใช้คอมพิวเตอร์ที่มีการพิมพ์ได้อย่างมีประสิทธิภาพในคีย์บอร์ดที่คุ้นเคย ระบบการรับรองความถูกต้องโดยใช้ keystroke ความเป็นไปได้ของการโจมตีด้วยรหัสผ่านโดยผู้บุกรุก (เช่น Dictionary Attack) เพราะถึงแม้จะมี

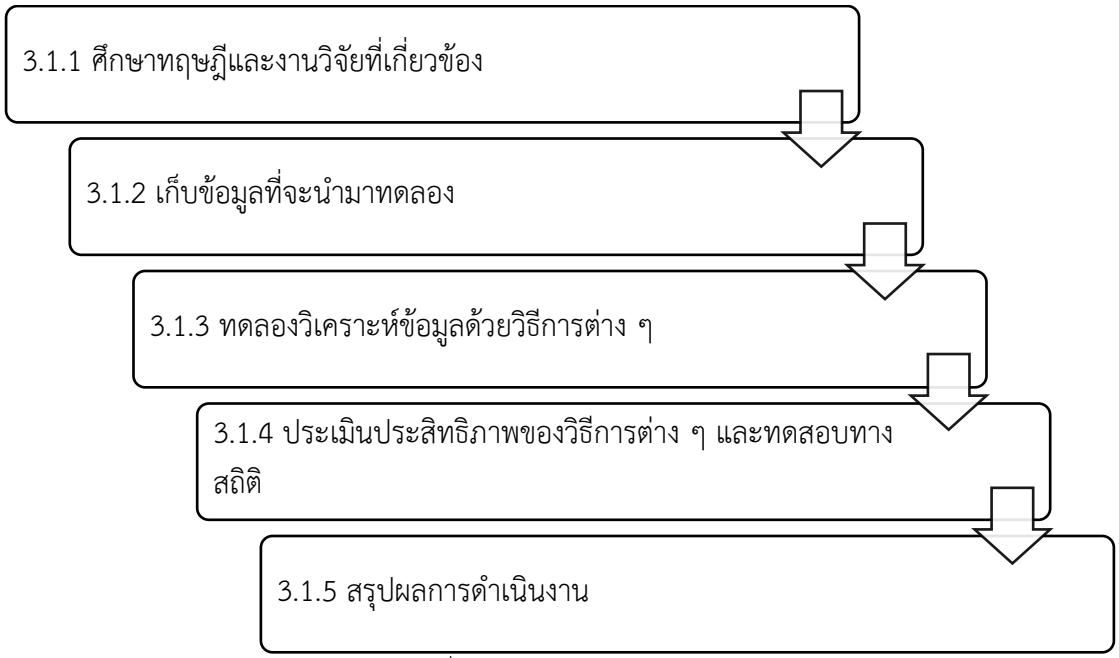
ข้อมูลล็อกอินและรหัสผ่านก็ตามเป็นเรื่องยากมากสำหรับผู้บุกรุกที่จะทำซ้ำรูปแบบการกดแป้น การโจมตีเกี่ยวกับพฤติกรรมทางชีวภาพเป็นเรื่องยากเพราะพฤติกรรมเชิงชีวภาพเป็นเอกลักษณ์ของแต่ละคน เป็นเรื่องยากที่จะทำซ้ำรูปแบบการกดแป้นพิมพ์แม้จะมีการฝึกมากก็ตาม

บทที่ 3
วิธีการดำเนินงาน

ในบทนี้จะกล่าวถึงการดำเนินการวิจัยเรื่องการระบุตัวบุคคลโดยการวิเคราะห์รูปแบบที่ได้จากการพิมพ์ โดยเนื้อหาประกอบด้วย ระเบียบวิธีการระบุตัวบุคคล เก็บข้อมูลที่จะนำมาทดลอง ทดลองวิเคราะห์ข้อมูลด้วยวิธีการต่าง ๆ วัดประสิทธิภาพของวิธีการต่าง ๆ และทดสอบทางสถิติ สรุปผลการดำเนินงาน รายละเอียดแสดงในหัวข้อ 3.1 และ 3.5 ดังนี้

3.1 ระเบียบวิธีการระบุตัวบุคคล

ระเบียบวิธีการระบุตัวบุคคล มีระเบียบวิธีการดำเนินงานดังภาพที่ 3.1 และได้ถูกอธิบายในหัวข้อที่ 3.1.1 ถึง 3.1.5



ภาพที่ 3.1 ระเบียบวิธีการทำงาน

3.1.1 ศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ศึกษาวิธีการที่เกี่ยวข้อง ได้แก่ วิธี KNN, Dynamic Time Warping

3.1.2 เก็บข้อมูลที่จะนำมาทดลอง

ทำการเก็บข้อมูล โดยใช้โปรแกรมที่เขียนขึ้นเพื่อเก็บค่าการพิมพ์จากผู้ใช้

3.1.3 ทดลองวิเคราะห์ข้อมูลด้วยวิธีการต่าง ๆ

ทดลองวิธีการหาความคล้ายกันของข้อมูลด้วย Dynamic Time Warping และใช้วิธี KNN ในการวิเคราะห์ข้อมูล

3.1.4 ประเมินประสิทธิภาพของวิธีการต่าง ๆ และทดสอบทางสถิติ

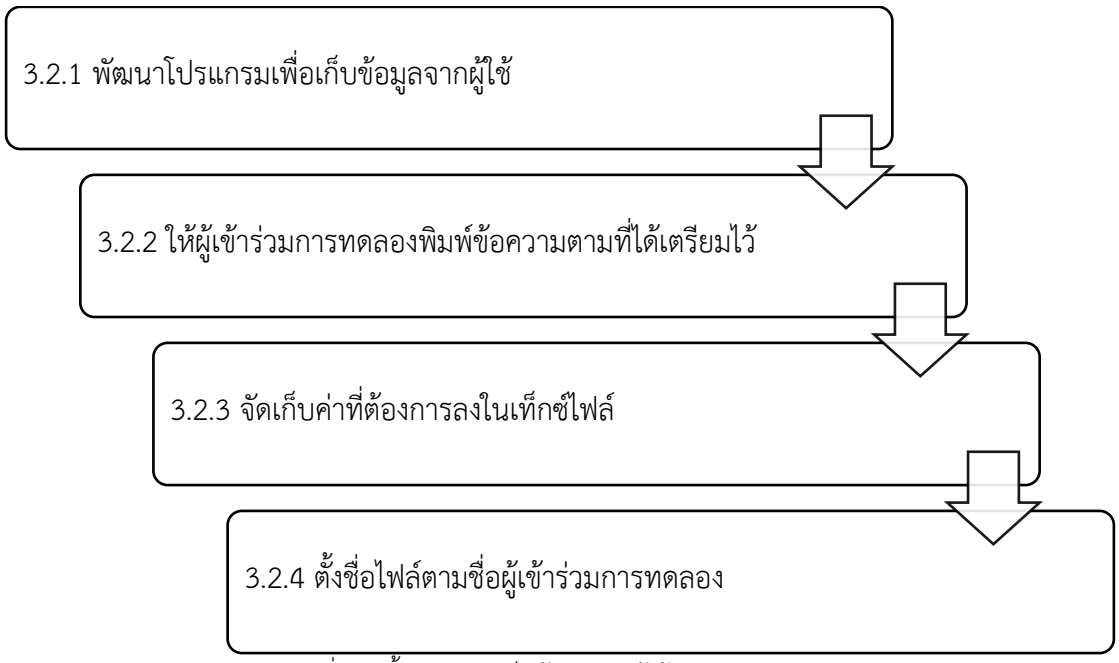
ประเมินประสิทธิภาพด้วยวิธีการที่เหมาะสม และทำการทดสอบทางสถิติ

3.1.5 สรุปผลการดำเนินงาน

สรุปผลการดำเนินงาน และเปรียบเทียบผลลัพธ์ที่ได้กับงานวิจัยอื่น ๆ

3.2 การเก็บข้อมูลที่จะนำมาทดลอง

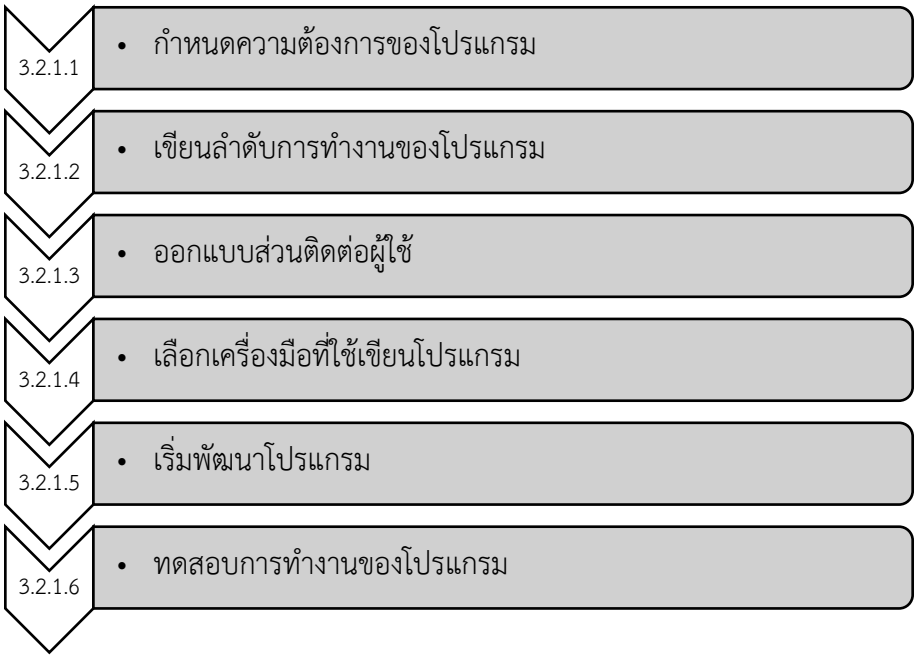
การเก็บข้อมูลที่จะนำมาทดลอง มีลำดับขั้นตอนการทำงานดังภาพที่ 3.2



ภาพที่ 3.2 ขั้นตอนการเก็บข้อมูลจากผู้เข้าร่วมการทดลอง

3.2.1. พัฒนาโปรแกรมเพื่อเก็บข้อมูลจากผู้ใช้

การพัฒนาโปรแกรมเพื่อเก็บข้อมูลจากผู้ใช้ มีลำดับขั้นตอนการทำงานตามภาพที่ 3.3



ภาพที่ 3.3 ลำดับขั้นตอนการพัฒนาโปรแกรมเพื่อเก็บข้อมูลจากผู้ใช้

3.2.1.1. กำหนดความต้องการของโปรแกรม

1) โปรแกรมต้องการเก็บค่า 3 ค่า ในระหว่างที่ผู้เข้าร่วมการทดลองทำการพิมพ์อยู่ โดยค่าดังกล่าวคือ

a. ตำแหน่งของแป้นบนคีย์บอร์ดที่ผู้เข้าร่วมการทดลองกด

b. เวลาที่ผู้เข้าร่วมการทดลองกดแป้นนั้น

c. เวลาที่ผู้เข้าร่วมการทดลองปล่อยแป้นนั้น

2) หน่วยของเวลาต้องมีความละเอียดพอที่จะเห็นความเปลี่ยนแปลง

3) มีบทความให้ผู้เข้าร่วมการทดลองดู ขณะที่ทำการพิมพ์

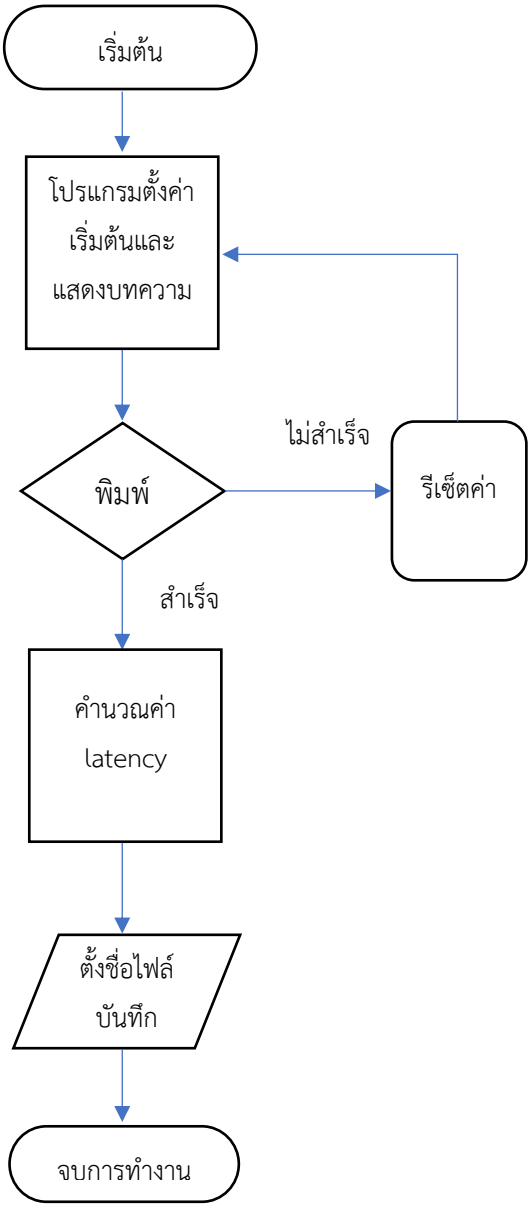
4) เมื่อผู้เข้าร่วมการทดลองพิมพ์บทความเสร็จ ใช้ข้อมูลที่ได้รับมาคำนวณค่า latency

5) นำค่าทั้งหมดที่เก็บได้ ไปเขียนลงในไฟล์ ในฟอร์แมตที่เหมาะสม เพื่อนำไปวิเคราะห์ต่อไป

6) มีการบันทึกชื่อผู้เข้าร่วมการทดลอง เพื่อเป็นประโยชน์ในการทดสอบผลการวิเคราะห์

3.2.1.2. เขียนลำดับการทำงานของโปรแกรม

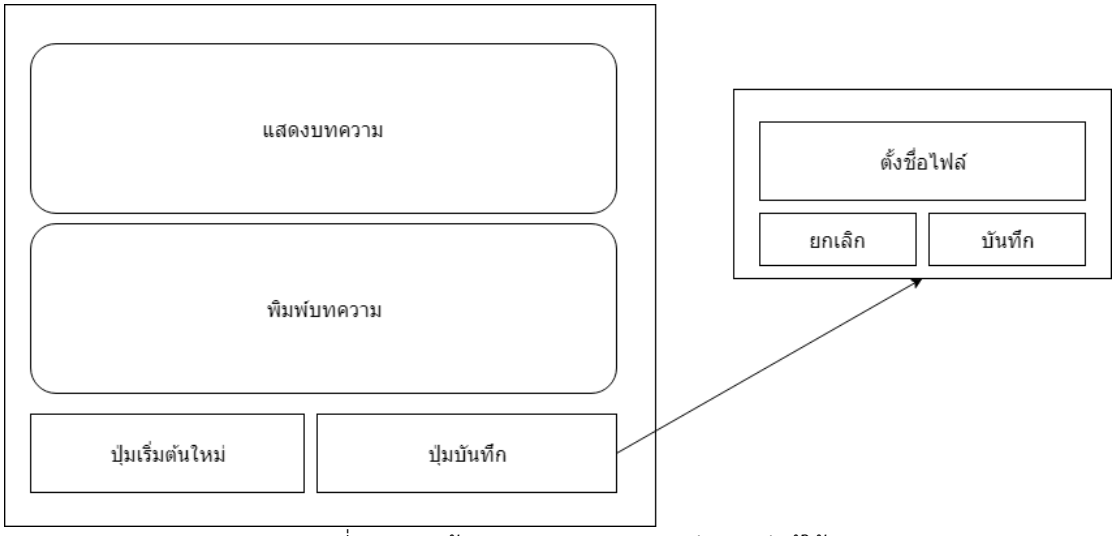
ลำดับการทำงานของโปรแกรมสามารถเขียนออกมาได้ดังภาพที่ 3.4



ภาพที่ 3.4 ลำดับการทำงานของโปรแกรมเก็บข้อมูล

3.2.1.3. ออกแบบส่วนติดต่อผู้ใช้

การออกแบบส่วนติดต่อผู้ใช้ มีการออกแบบดังภาพที่ 3.5



ภาพที่ 3.5 ภาพต้นแบบของการออกแบบส่วนติดต่อผู้ใช้

3.2.1.4. เลือกเครื่องมือที่ใช้เขียนโปรแกรม

1) ใช้ภาษา Java ในการพัฒนาโปรแกรม
2) ใช้ JDK เวอร์ชั่น 1.8.0 ในการคอมไพล์ซอร์สโค้ด
3) ใช้ JRE เวอร์ชั่น 1.8.0 ในการเรียกใช้โปรแกรม
4) ใช้ Text Editor ชื่อ Sublime Text 3
5) ใช้ Build Tool ชื่อ Apache Maven 3.6.0
6) ใช้ Library JSON Simple 1.1.1 เพื่อจัดการเกี่ยวกับไฟล์
7) ใช้ Library JFreeChart 1.0.19 เพื่อจัดการเกี่ยวกับกราฟ

3.2.1.5. เริ่มพัฒนาโปรแกรม

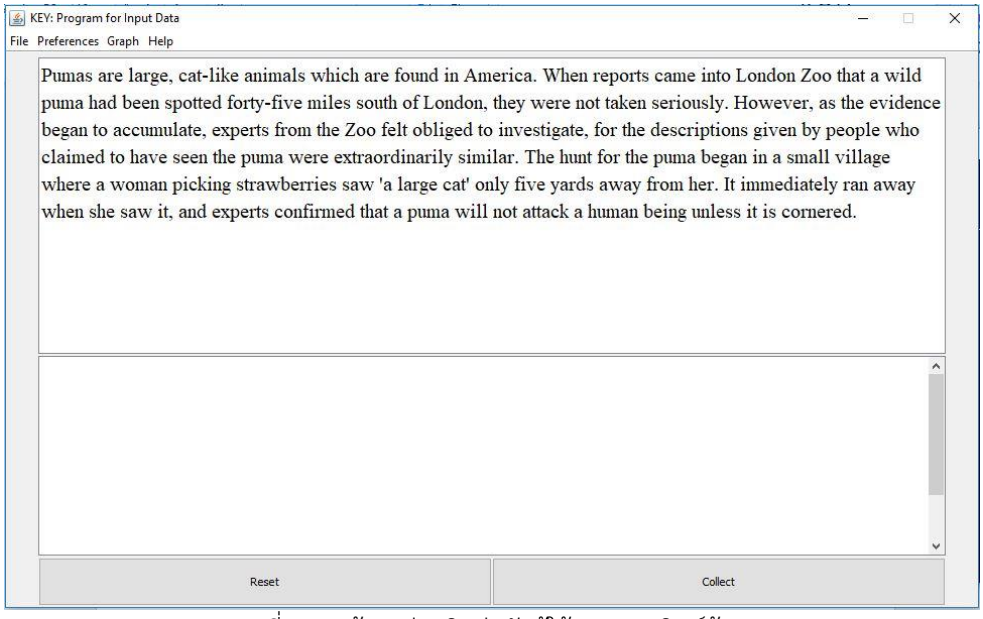
1) สร้างสภาพแวดล้อมเพื่อให้ง่ายต่อการพัฒนาโปรแกรมด้วย Maven
2) เริ่มพัฒนาโปรแกรมให้เป็นไปตามลำดับการทำงานที่วางไว้
3) แปลงไฟล์ทั้งหมดที่จำเป็นต่อการทำงานของโปรแกรมให้อยู่ในแพ็คเกจ JAR เพื่อง่ายต่อการเรียกใช้ในเครื่องอื่น ๆ

3.2.1.6. ทดสอบการทำงานของโปรแกรม

ทดสอบการทำงานของโปรแกรมโดยผู้ทำการทดลอง

3.2.2. ให้ผู้เข้าร่วมการทดลองพิมพ์ข้อความตามที่ได้เตรียมไว้

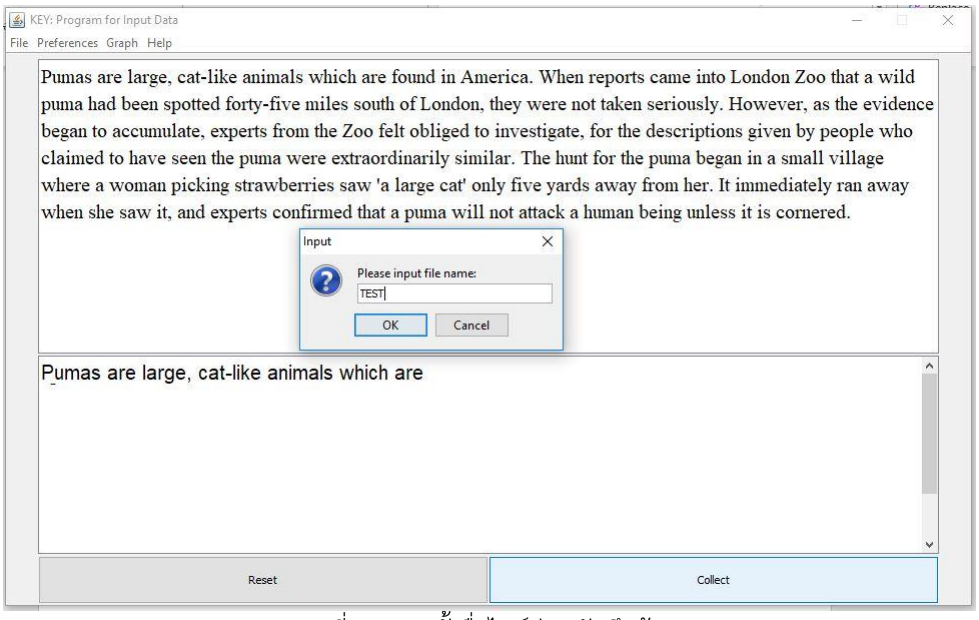
เก็บข้อมูลโดยการให้ผู้เข้าร่วมการทดลองพิมพ์ข้อความตามที่แสดงไว้ด้านบน ลงในช่องเก็บข้อมูลด้านล่าง ดังภาพที่ 3.6



ภาพที่ 3.6 หน้าจอส่วนติดต่อกับผู้ใช้งานขณะพิมพ์ข้อมูล

3.2.3. ตั้งชื่อไฟล์ตามชื่อผู้เข้าร่วมการทดลอง

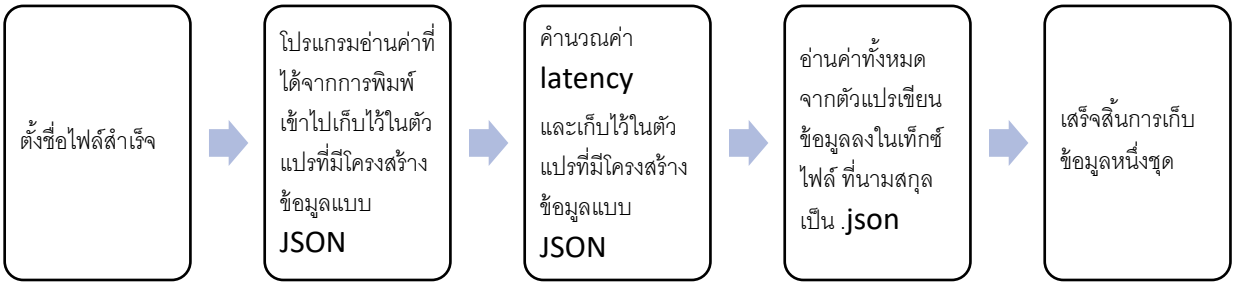
ทำการตั้งชื่อไฟล์ ก่อนที่จะจัดเก็บข้อมูล ดังภาพที่ 3.7

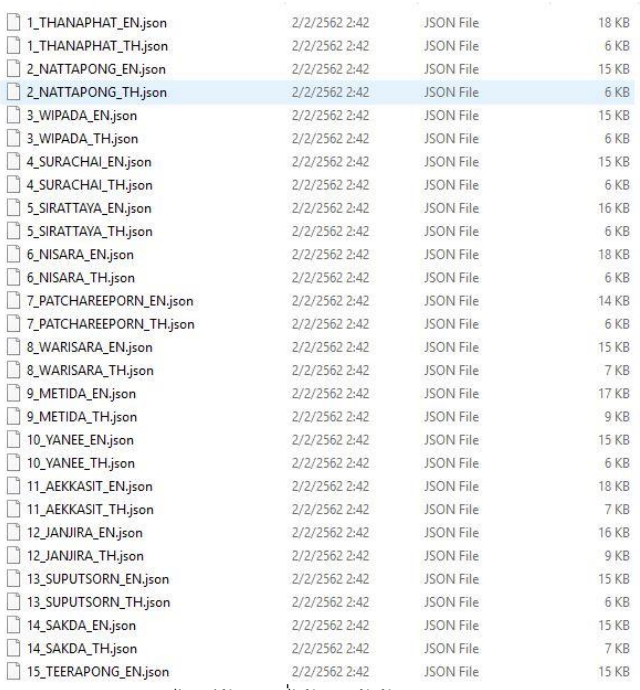


ภาพที่ 3.7 การตั้งชื่อไฟล์ก่อนบันทึกข้อมูล

3.2.4. จัดเก็บข้อมูลที่ได้ลงในเท็กซ์ไฟล์

จัดเก็บข้อมูลที่ได้ลงในเท็กซ์ไฟล์นามสกุล .json ดังภาพที่ 3.8 ซึ่งโครงสร้างข้อมูลของไฟล์ .json จะมีหน้าตาดังภาพที่ 3.9





3.8 ภาพไฟล์ข้อมูลที่ได้จากผู้เข้าร่วมการทดลอง

```
{
        "Name" : "NATTAPONG" ,
        "Language" : "EN" ,
        "KeyPressed" : [12765,12824,13083,...,353067] ,
        "KeyReleased" : [12593,12723,13003,...,349176] ,
        "KeyCode" : [16,80,85,...,78] ,
        "isEqual" : true ,
        "Latency" : [231,360,351,...,4294]
}
```


3.9 ภาพตัวอย่างโครงสร้างข้อมูลที่ได้มาจากผู้ใช้

จากภาพที่ 3.9 คือตัวอย่างโครงสร้างข้อมูลที่ผู้ทำการทดลองทำการเก็บจากผู้ใช้ประกอบไปด้วย

‘Name’ หมายถึง ชื่อของผู้ที่ทำการพิมพ์

‘Language’ หมายถึง ภาษาที่ใช้เป็นตัวอย่างการพิมพ์

‘KeyPressed’ หมายถึง ค่าเวลาในขณะที่ผู้ใช้กดแป้นหนึ่งแป้น

‘KeyReleased’ หมายถึง ค่าเวลาในขณะที่ผู้ใช้ปล่อยแป้นหนึ่งแป้น

‘KeyCode’ หมายถึง รหัสของแป้นที่ผู้ใช้กด

‘isEqual’ หมายถึง ตัวแปรที่คอยตรวจสอบว่า จำนวนค่าของ KeyPressed และ KeyReleased เท่ากันหรือไม่

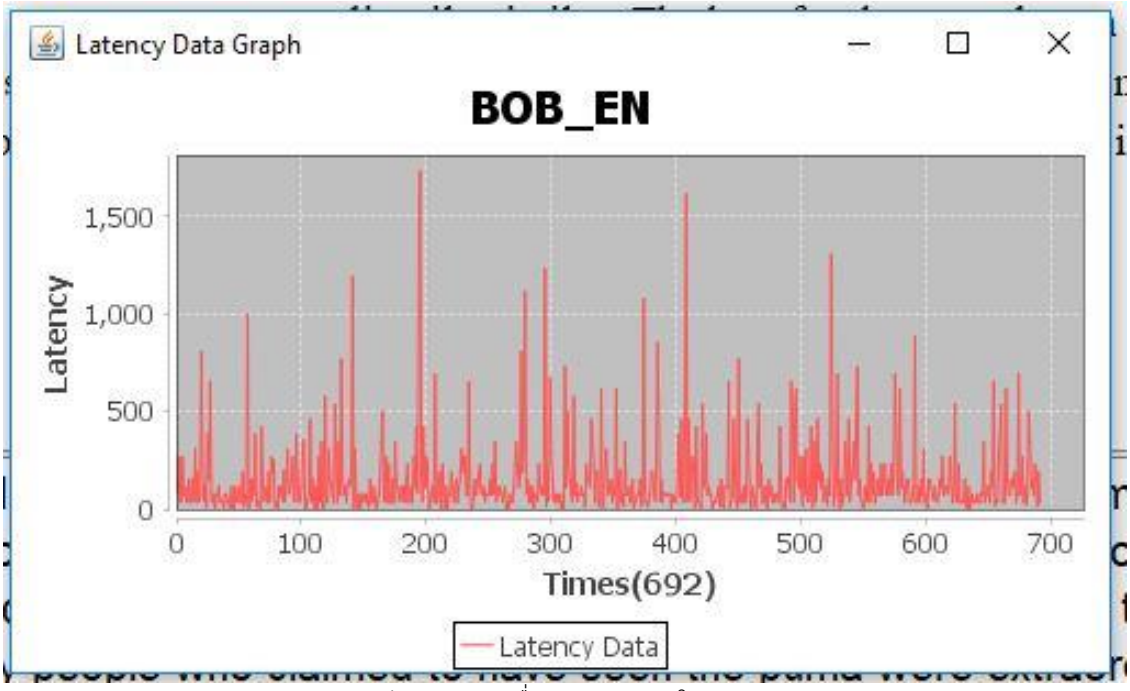
‘Latency’ หมายถึง เป็นค่าที่คำนวณได้จาก เวลาของการปล่อยแป้นที่แล้วลบกับเวลาของการกดแป้นถัดไป

3.3 ทดลองวิเคราะห์ข้อมูลด้วยวิธีการต่าง ๆ

การทดลองวิเคราะห์ข้อมูลด้วยวิธีการต่าง ๆ มีลำดับขั้นตอนการทำงานดังนี้

3.3.1. การวิเคราะห์ข้อมูลโดยการเปรียบเทียบ

1) นำค่า latency ที่คำนวณได้มาพล็อตเป็นกราฟ ซึ่งเราตั้งสมมติฐานว่ากราฟเป็นตัวแทนของรูปแบบการพิมพ์ของแต่ละบุคคล ก็จะได้ดังภาพที่ 3.11



3.11 ภาพค่า latency เมื่อนำมาแสดงผลในรูปแบบกราฟ

แกน y (Latency) หมายถึง ช่วงระยะเวลาระหว่างที่ปล่อยแป้นไปจนถึงการกดแป้นถัดไป ซึ่งคำนวณได้จาก |เวลาของการปล่อยแป้นที่แล้ว – เวลาของการกดแป้น| เช่น เมื่อผู้ใช้พิมพ์คำว่า cat โปรแกรมจะคำนวณค่า latency ออกมาได้ 2 ค่า คือค่าที่เกิดขึ้นระหว่างที่ปล่อยแป้น c ไปกดแป้น a และค่าที่เกิดขึ้นระหว่างที่ปล่อยแป้น a ไปกดแป้น t ในส่วนของแกน x (Times) หมายถึง ดัชนีของแป้นที่ผู้ใช้ได้ทำการกด ส่วนตัวเลขในวงเล็บ หมายถึง จำนวนแป้นทั้งหมดที่ผู้ใช้ได้ทำการกด

3.3.2. ทดลองวิเคราะห์ข้อมูลด้วยวิธี KNN

เมื่อเราได้กราฟ ซึ่งเป็นตัวแทนของแต่ละบุคคลแล้ว ก็จะเริ่มทำการวิเคราะห์โดยใช้เทคนิค KNN และคำนวณความคล้ายกันของกราฟด้วยวิธีต่าง ๆ เช่น Dynamic Time warping, Euclidian, Pearson

3.3.3. วัดประสิทธิภาพ

วัดประสิทธิภาพโดยใช้ accuracy, precision และอื่น ๆ