

การระบุตัวบุคคลโดยการวิเคราะห์รูปแบบที่ได้จากการพิมพ์

Person Authentication Using Typing Pattern Analysis



ปัญหาพิเศษนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรบัณฑิต (วิทยาการคอมพิวเตอร์)
ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2561

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Person Authentication Using Typing Pattern Analysis

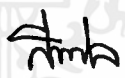
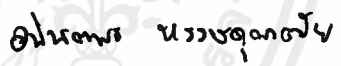


A SPECIAL PROBLEM SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENT FOR
THE DEGREE OF BACHELOR OF SCIENCE (COMPUTER SCIENCE)
DEPARTMENT OF COMPUTER SCIENCE, FACULTY OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
ACADEMIC YEAR 2018

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อปัญหาพิเศษ	การระบุตัวบุคคลโดยวิเคราะห์รูปแบบที่ได้จากการพิมพ์ Person Authentication Using Typing Pattern Analysis
ชื่อนักศึกษา	นายอรรถชัย มาลาอูตม์ รหัสนักศึกษา 58050417 นายสิริวิชญ์ ไพโรสณฑ์วัฒน รหัสนักศึกษา 58050401
ปริญญา	วิทยาศาสตรบัณฑิต (วิทยาการคอมพิวเตอร์)
ภาควิชา	วิทยาการคอมพิวเตอร์
คณะ	วิทยาศาสตร์
มหาวิทยาลัย	สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (สจล.)
ปีการศึกษา	2561
อาจารย์ที่ปรึกษา	ดร.อัคเดช อุดมชัยพร

คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (สจล.)
อนุมัติให้ปัญหาพิเศษนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต
(วิทยาการคอมพิวเตอร์) ประจำปีการศึกษา 2561

คณะกรรมการสอบ	ลายมือชื่อ
ผศ.ดร.สายชล ใจเย็น ประธานกรรมการ	
ผศ.ดร.อนันตพร ทรราชคุณาฒย์ กรรมการ	
ดร.อัคเดช อุดมชัยพร กรรมการและอาจารย์ที่ปรึกษา	

ลิขสิทธิ์ของคณะวิทยาศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อปัญหาพิเศษ	การระบุตัวบุคคลโดยวิเคราะห์รูปแบบที่ได้จากการพิมพ์
ชื่อนักศึกษา	นายอรรถชัย มาลาอุตม์ รหัสนักศึกษา 58050417 นายสิริวิษณุ ไพโรสณฑ์วัฒน รหัสนักศึกษา 58050401
ปริญญา	วิทยาศาสตรบัณฑิต (วิทยาการคอมพิวเตอร์)
ภาควิชา	วิทยาการคอมพิวเตอร์
คณะ	วิทยาศาสตร์
มหาวิทยาลัย	สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (สจล.)
ปีการศึกษา	2561
อาจารย์ที่ปรึกษา	ดร.อัคเดช อุดมชัยพร

บทคัดย่อ

ปัญหาพิเศษนี้นำเสนอวิธีการระบุตัวบุคคลโดยการวิเคราะห์รูปแบบที่ได้จากการพิมพ์ของแต่ละบุคคลโดยใช้วิธีการ Dynamic Time Warping ในการเปรียบเทียบความคล้ายกันของรูปแบบการพิมพ์ซึ่งถูกแปลงมาอยู่ในรูปแบบของกราฟ การทดลองนี้ใช้ชุดข้อมูลที่เก็บมาจากบุคคลทั่วไปประมาณ 100 คน และมีการเปรียบเทียบผลลัพธ์ที่ได้กับโมเดลมาตรฐานแบบอื่น ๆ ผลการทดลองที่ได้บ่งชี้ว่าวิธีการที่นำเสนอให้ผลลัพธ์ที่ดีกว่าวิธีการมาตรฐานอื่น ๆ ทั้งในแง่ของประสิทธิภาพและประสิทธิผล

คำสำคัญ : การระบุตัวบุคคล ไดนามิกไทม์วอร์ปิง รูปแบบที่ได้จากการพิมพ์

Title	Person Authentication Using Typing Pattern Analysis
Students	Mr.Athachai Malaut Student ID 58050417 Mr.Sirawit Paisonwattana Student ID 58050401
Degree	Bachelor of Science (Computer Science)
Department	Computer Science
Faculty	Science
University	King Mongkut's Institute of Technology Ladkrabang (KMITL)
Academic Year	2018
Advisor	Dr.Akadej Udomchaiporn

Abstract

This special problem presents a method to identify people by analyzing their typing patterns using Dynamic Time Warping method in order to compare similarity between the graphs extracted from the patterns. The experiments were conducted using typing datasets collected from 100 people. The results obtained from the proposed approach were compared to some standard existing models. The experimental results indicated that the proposed approach outperformed any other classical models in terms of both effectiveness and efficiency.

Keywords : Authentication, Dynamic Time Warping, Typing Patterns.

กิตติกรรมประกาศ

การจัดทำปัญหาพิเศษเรื่องการระบุตัวบุคคลโดยการวิเคราะห์รูปแบบที่ได้จากการพิมพ์นี้สำเร็จ ลุล่วงไปได้ด้วยดี คณะผู้จัดทำต้องขอขอบพระคุณอาจารย์ที่ปรึกษา ดร.อัคเดช อุดมชัยพร ที่ช่วยให้ คำปรึกษาและคำแนะนำที่ดีแก่คณะผู้จัดทำในการปรับปรุงปัญหาพิเศษนี้

ขอขอบพระคุณอาจารย์ผู้ควบคุมการสอบปัญหาพิเศษ ผศ.ดร.สายชล ใจเย็น และ ผศ.ดร. อนันตพร ทรรษคุณาฒย ที่มีส่วนช่วยในการตรวจสอบ และให้คำแนะนำ ทำให้ปัญหาพิเศษนี้มีความ สมบูรณ์มากยิ่งขึ้น

ขอขอบพระคุณบุคคลต่าง ๆ ที่เกี่ยวข้องที่ได้ให้การช่วยเหลือในการทำปัญหาพิเศษนี้ และ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบังที่มอบโอกาสให้ได้เข้าศึกษาในสถาบันแห่งนี้

อรรถชัย มาลาอุตม์
สิริวิษญ์ ไพรสณห์วัฒน์

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ก
บทคัดย่อภาษาอังกฤษ.....	ข
กิตติกรรมประกาศ.....	ค
สารบัญ.....	ง
สารบัญตาราง.....	ฉ
สารบัญภาพ.....	ช
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญ	1
1.2 วัตถุประสงค์	1
1.3 ขอบเขตการศึกษา.....	2
1.4 ประโยชน์ที่คาดว่าจะได้รับ.....	2
1.5 ขั้นตอนการดำเนินงาน.....	2
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	3
2.1 การยืนยันตัวตน (Authentication).....	3
2.2 ไบโอเมตริกซ์ (Biometric).....	4
2.2.1 ความหมายของไบโอเมตริกซ์.....	4
2.2.2 ประโยชน์ของไบโอเมตริกซ์.....	4
2.2.3 การยืนยันตัวตนด้วยไบโอเมตริกซ์.....	5
2.3 ไดนามิกไทม์วอร์ปิง (Dynamic Time Warping).....	6
2.4 สหสัมพันธ์แบบเพียร์สัน (Pearson Correlation).....	7
2.5 งานวิจัยที่เกี่ยวข้อง.....	8
2.5.1 Learning User Keystroke Patterns for Authentication.....	8
2.5.2 Keystroke-Based Authentication by Key Press Intervals as a Complementary Behavioral Biometric.....	10
บทที่ 3 วิธีการดำเนินงานวิจัย	12
3.1 ระเบียบวิธีการระบุตัวบุคคล.....	12
3.1.1 ศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....	13
3.1.2 เก็บข้อมูลที่ให้นำมาทดลอง.....	13
3.1.3 ทดลองวิเคราะห์ข้อมูลด้วยวิธีการต่าง ๆ.....	13

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

	หน้า
3.1.4 ประเมินประสิทธิภาพของวิธีการต่าง ๆ	13
3.1.5 สรุปผลการดำเนินงาน.....	13
3.2 การเก็บข้อมูลที่จะนำมาทดลอง.....	13
3.2.1 พัฒนาโปรแกรมเพื่อเก็บข้อมูลจากผู้ใช้.....	14
3.2.1.1 กำหนดความต้องการของโปรแกรม.....	14
3.2.1.2 เขียนลำดับการทำงานของโปรแกรม.....	15
3.2.1.3 ออกแบบส่วนติดต่อผู้ใช้.....	16
3.2.1.4 เลือกเครื่องมือที่ใช้เขียนโปรแกรม.....	16
3.2.1.5 เริ่มพัฒนาโปรแกรม.....	16
3.2.1.6 ทดสอบการทำงานของโปรแกรม.....	16
3.2.2 ให้ผู้เข้าร่วมการทดลองพิมพ์ข้อความตามที่ได้เตรียมไว้.....	17
3.2.3 ตั้งชื่อไฟล์ตามชื่อผู้เข้าร่วมการทดลอง.....	17
3.2.4 จัดเก็บข้อมูลที่ได้ลงในเท็กซ์ไฟล์.....	18
3.3 ทดลองวิเคราะห์ข้อมูลด้วยวิธีการต่าง ๆ.....	20
3.3.1 สร้างกราฟเพื่อแทนตัวบุคคล.....	20
3.3.2 ทดลองวิเคราะห์ข้อมูลด้วยวิธี KNN.....	20
3.3.3 ทดลองวิเคราะห์ข้อมูลของ typo data sets.....	21
3.3.4 ทดลองวิเคราะห์ข้อมูลของ non typo data sets.....	22
3.3.5 ทดลองวิเคราะห์ข้อมูลของ trim data sets.....	23
บทที่ 4 ผลการวิจัยและการอภิปรายผล	24
4.1 ผลการดำเนินงานของ typo data sets.....	24
4.2 ผลการดำเนินงานของ non typo data sets.....	28
4.3 ผลการดำเนินงานของ trim data sets.....	33
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ.....	52
5.1 สรุปผลการดำเนิน.....	52
5.2 ข้อเสนอแนะ.....	52

สารบัญตาราง

ตารางที่	หน้า
2.1 สิ่งที่เป็นลักษณะทางกายภาพ และสิ่งที่เป็นลักษณะทางพฤติกรรม.....	4
2.2 ตารางผลการคำนวณ.....	7
2.3 ตารางเปรียบเทียบประสิทธิภาพของแต่ละโมเดล.....	8
2.4 ตารางเปรียบเทียบประสิทธิภาพของโมเดล เมื่อใช้จำนวนแอททริบิวต์ที่ต่างกัน.....	9
2.5 ตารางเปรียบเทียบประสิทธิภาพของโมเดล เมื่อใช้จำนวนของค่าที่ต่างกัน.....	9
4.1 ค่าความถูกต้องโดยเฉลี่ยของข้อมูลทุกชุด.....	48
4.2 ค่าความถูกต้องของ typo data sets ด้วยวิธี dtw-raw.....	49
4.3 ค่าความถูกต้องของ non typo data sets ด้วยวิธี dtw-raw.....	50
4.4 ค่าความถูกต้องของ trim data sets ด้วยวิธี dtw-raw.....	50
4.5 ค่าความถูกต้องของ trim data sets ด้วยวิธี euclid-raw.....	51
4.6 ค่าความถูกต้องของ trim data sets ด้วยวิธี euclid-fourier.....	51

สารบัญภาพ

ภาพที่	หน้า
2.1 ระบบยืนยันตัวตนด้วยไบโอเมตริกซ์	5
2.2 ภาพเปรียบเทียบระหว่างวิธีการเปรียบเทียบแบบยูคลิเดียน และวิธีการเปรียบเทียบแบบไดนามิกไทม์วอร์ปิง	6
3.1 ระเบียบวิธีการทำงาน	12
3.2 ขั้นตอนการเก็บข้อมูลจากผู้เข้าร่วมการทดลอง.....	13
3.3 ลำดับขั้นตอนการพัฒนาโปรแกรมเพื่อเก็บข้อมูลจากผู้ใช้	14
3.4 ลำดับขั้นตอนการทำงานของโปรแกรมเก็บข้อมูล	15
3.5 ภาพต้นแบบของการออกแบบส่วนติดต่อผู้ใช้.....	16
3.6 หน้าจอส่วนติดต่อกับผู้ใช้งานขณะพิมพ์ข้อมูล.....	17
3.7 การตั้งชื่อไฟล์ก่อนบันทึกข้อมูล.....	17
3.8 ภาพไฟล์ข้อมูลที่ได้จากผู้เข้าร่วมการทดลอง.....	18
3.9 ภาพไฟล์ตัวอย่างโครงสร้างข้อมูลที่ได้มาจากผู้ใช้.....	19
3.10 ภาพค่า latency เมื่อนำมาแสดงผลในรูปแบบกราฟ	20
3.11 ภาพกระบวนการวิเคราะห์ข้อมูล typo data sets	21
3.12 ภาพกระบวนการวิเคราะห์ข้อมูล non typo data sets.....	22
3.13 ภาพกระบวนการวิเคราะห์ข้อมูล trim data sets	23
4.1 กราฟเปรียบเทียบค่าความถูกต้องของ typo data sets โดยใช้วิธี dtw-raw	24
4.2 กราฟเปรียบเทียบค่าความถูกต้องของ typo data sets โดยใช้วิธี dtw-mapping.....	25
4.3 กราฟเปรียบเทียบค่าความถูกต้องของ typo data sets โดยใช้วิธี dtw-fourier.....	25
4.4 กราฟเปรียบเทียบค่าความถูกต้องของ typo data sets โดยใช้วิธี dtw-fourier-mapping.....	26
4.5 กราฟเปรียบเทียบค่าความถูกต้องของ typo data sets โดยใช้วิธี dtw-mapping-fourier.....	27
4.6 กราฟเปรียบเทียบค่าความถูกต้องของ non typo data sets โดยใช้วิธี dtw-raw.....	28
4.7 กราฟเปรียบเทียบค่าความถูกต้องของ non typo data sets โดยใช้วิธี dtw-mapping	29
4.8 กราฟเปรียบเทียบค่าความถูกต้องของ non typo data sets โดยใช้วิธี dtw-fourier ...	30
4.9 กราฟเปรียบเทียบค่าความถูกต้องของ non typo data sets โดยใช้วิธี dtw-fourier-mapping.....	31
4.10 กราฟเปรียบเทียบค่าความถูกต้องของ non typo data sets โดยใช้วิธี dtw-mapping-fourier.....	32

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญภาพ (ต่อ)

ภาพที่	หน้า
4.11 กราฟเปรียบเทียบค่าความถูกต้องของ trim data sets โดยใช้วิธี dtw-raw.....	33
4.12 กราฟเปรียบเทียบค่าความถูกต้องของ trim data sets โดยใช้วิธี dtw-mapping	34
4.13 กราฟเปรียบเทียบค่าความถูกต้องของ trim data sets โดยใช้วิธี dtw-fourier	35
4.14 กราฟเปรียบเทียบค่าความถูกต้องของ trim data sets โดยใช้วิธี dtw-fourier-mapping.....	36
4.15 กราฟเปรียบเทียบค่าความถูกต้องของ trim data sets โดยใช้วิธี dtw-mapping-fourier.....	37
4.16 กราฟเปรียบเทียบค่าความถูกต้องของ trim data sets โดยใช้วิธี euclid-raw	38
4.17 กราฟเปรียบเทียบค่าความถูกต้องของ trim data sets โดยใช้วิธี euclid-mapping	39
4.18 กราฟเปรียบเทียบค่าความถูกต้องของ trim data sets โดยใช้วิธี euclid-fourier	40
4.19 กราฟเปรียบเทียบค่าความถูกต้องของ trim data sets โดยใช้วิธี euclid-fourier-mapping.....	41
4.20 กราฟเปรียบเทียบค่าความถูกต้องของ trim data sets โดยใช้วิธี pearson-raw.....	42
4.21 กราฟเปรียบเทียบค่าความถูกต้องของ trim data sets โดยใช้วิธี pearson-mapping..	43
4.22 กราฟเปรียบเทียบค่าความถูกต้องของ trim data sets โดยใช้วิธี pearson-fourier.....	44
4.23 กราฟเปรียบเทียบค่าความถูกต้องของ trim data sets โดยใช้วิธี pearson-fourier-mapping	45
4.24 กราฟเปรียบเทียบค่าความถูกต้องของ trim data sets โดยใช้วิธี pearson-mapping-fourier	46
4.25 กราฟเปรียบเทียบค่าความถูกต้องของ trim data sets โดยใช้วิธี euclid-mapping-fourier.....	47

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

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

พฤติกรรมกรรมการพิมพ์เป็นสิ่งที่ผู้ทำปัญหาพิเศษให้ความสนใจ โดยผู้ทำปัญหาพิเศษได้ตั้งสมมติฐานว่า พฤติกรรมการพิมพ์ของแต่ละบุคคลนั้นมีความแตกต่างกัน ถ้าหากพิจารณารูปแบบการพิมพ์อย่างถี่ถ้วนแล้ว เราจะสามารถระบุตัวบุคคลออกมาได้ เพราะพฤติกรรมกรรมการพิมพ์เป็นเอกลักษณ์เฉพาะบุคคล 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) เพื่อสร้างซอฟต์แวร์ที่ช่วยระบุตัวบุคคลจากรูปแบบการพิมพ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.3 ขอบเขตการศึกษา

- 1) ข้อมูลรูปแบบการพิมพ์ที่นำมาวิเคราะห์ในการทดลอง มาจากนักศึกษา อาจารย์ และบุคคลทั่วไป ในบริเวณสถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง มีจำนวนประมาณ 100 คน
- 2) เป็นพิมพ์ที่นำมาให้ผู้เข้าร่วมการทดลองใช้พิมพ์ข้อมูล เป็นแบบพิมพ์ที่ผู้ทำการทดลองจัดเตรียมไว้ให้
- 3) บทความที่นำมาให้ผู้เข้าร่วมการทดลองใช้พิมพ์ข้อมูล เป็นบทความภาษาอังกฤษเป็นบางส่วนจากแบบเรียนสำหรับผู้ใช้ภาษาอังกฤษเป็นภาษาที่ 2 ชื่อว่า “A Puma at large” รวมไปถึงบทความภาษาไทยมาจากข้อความที่เป็นที่นิยมสำหรับทดสอบฟอนต์ ชื่อว่า “ประโยคที่มีพยัญชนะภาษาไทยครบทุกตัว” และ “อาหารเข้ามีความสำคัญอย่างไร”

1.4 ขั้นตอนการดำเนินงาน

- 1) ศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้อง
- 2) เก็บข้อมูลที่จะนำมาทดลอง
- 3) วิเคราะห์ข้อมูลด้วยวิธีการต่าง ๆ ที่ได้ศึกษามา
- 4) วัดประสิทธิภาพของวิธีการต่าง ๆ
- 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 ไบโอเมตริกซ์ คีย์สโตรคไดนามิก วิธีการ Dynamic Time Warping โดยจะแสดงรายละเอียดในหัวข้อที่ 2.1 ถึง 2.5

2.1 การยืนยันตัวตน (Authentication)

การยืนยันตัวตน คือการตรวจสอบผู้ใช้ ก่อนที่จะอนุญาตให้ใช้งานในระบบ เพื่อพิสูจน์ว่าผู้ที่ต้องการจะใช้งานในระบบนี้คือใคร และผู้ใช้คนนี้มีสิทธิ์เข้าถึงทรัพยากรได้แค่ไหน โดยมีข้อมูลอยู่ 3 ประเภทหลัก ๆ ที่จะสามารถนำมาพิสูจน์ตัวบุคคลได้ คือ สิ่งที่ใช้รู้ สิ่งที่ใช้มี สิ่งที่ใช้เป็น

สิ่งที่ใช้รู้ (Something you know) คือการสมมติชุดข้อมูลขึ้นมาชุดหนึ่ง และเป็นที่ยอมรับกันระหว่างผู้ใช้กับระบบเท่านั้น เป็นวิธีที่ได้รับความนิยมมากที่สุดในปัจจุบัน เช่น การใช้รหัสผ่าน ชุดตัวเลข PINs การที่จะสร้างรหัสผ่านชุดหนึ่งให้มีความปลอดภัยมากที่สุดเพื่อป้องกันการถูกสวมรอยนั้น จำเป็นต้องมีกฎมากมาย เช่น ต้องมีความยาวมากกว่า 8 ตัวอักษร มีตัวอักษรพิมพ์เล็กและตัวอักษรพิมพ์ใหญ่ ประกอบกัน และต้องไม่ใช้รหัสผ่านนี้กับระบบอื่น ๆ เป็นต้น และเมื่อทำตามกฎนี้อย่างครบถ้วน ชุดรหัสผ่านที่ได้มาก็ยากที่จะจำเสียแล้ว

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

สิ่งที่ใช้เป็น (Something you are) คือการยืนยันตัวตนด้วยข้อมูลทางชีวภาพของผู้ใช้เองหรือเรียกอีกอย่างว่าไบโอเมตริกซ์ เช่น ใบหน้า ม่านตา ลายนิ้วมือ น้ำเสียง การพิมพ์ เป็นต้น ซึ่งเป็นวิธีที่สะดวก และปลอดภัยมากที่สุด เมื่อเทียบกับวิธีที่ผ่านมา เนื่องจากผู้ใช้ไม่ต้องจำชุดข้อมูลใหม่ และไม่ต้องพกสิ่งของเพิ่มเติม แต่ข้อจำกัดต่าง ๆ ก็แตกต่างกันออกไป ขึ้นอยู่กับว่าระบบเลือกที่จะใช้ข้อมูลทางชีวภาพแบบใด

2.2 ไบโอมेटริกซ์ (Biometric)

หัวข้อนี้จะกล่าวถึงไบโอมेटริกซ์ ซึ่งเป็นเทคโนโลยีหลัก และเป็นหัวใจสำคัญของปัญหาพิเศษนี้

2.2.1 ความหมายของไบโอมेटริกซ์

ไบโอมेटริก คือ กระบวนการตรวจสอบและจดจำลักษณะเฉพาะบุคคล โดยอิงจากลักษณะทางชีวภาพ

ลักษณะทางชีวภาพ สามารถจำแนกได้ 2 ลักษณะ ดังตารางที่ 2.1

ตารางที่ 2.1 สิ่งที่เป็นลักษณะทางกายภาพ และสิ่งที่เป็นลักษณะทางพฤติกรรม

ลักษณะทางกายภาพ	ลักษณะทางพฤติกรรม
<ul style="list-style-type: none"> • ลายนิ้วมือ • ใบหน้า ใบหู • ม่านตา จอประสาทตา • รูปร่าง รูปนิ้ว • การไหลเวียนของเลือด 	<ul style="list-style-type: none"> • การออกเสียง • ลายเซ็นต์ • การพิมพ์

2.2.2 ประโยชน์ของไบโอมेटริกซ์

ไบโอมेटริกซ์ สามารถนำมาใช้ในการยืนยันตัวตนบุคคล ก่อนที่จะมีการอนุญาตให้เข้าใช้ระบบ ซึ่งมีประสิทธิภาพมากกว่าการยืนยันตัวตนที่เป็นที่นิยมในปัจจุบัน เช่น บัตรเอทีเอ็ม กุญแจ รหัสผ่าน เหล่านี้เป็นการใช้สิ่งสมมติขึ้นมาแทนตัวตนจริง ๆ ของผู้ใช้ และมอบให้ผู้ใช้แต่ละคนเพื่อใช้ในการยืนยันตัวตน มีข้อเสียคือ สามารถถูกขโมย ทำซ้ำ และหลงลืมได้ แต่ลักษณะทางชีวภาพเป็นสิ่งที่ติดตัวเราอยู่เสมอ การสวมรอยจากผู้ไม่หวังดีนั้นทำได้ยากกว่า มีความแม่นยำสูงจึงเชื่อถือได้ ผู้ใช้มีความสะดวกในการใช้งาน และอาจมีราคาถูกลงกว่าวิธีการแบบดั้งเดิม

2.2.3 การยืนยันตัวตนด้วยไบโอเมตริกซ์

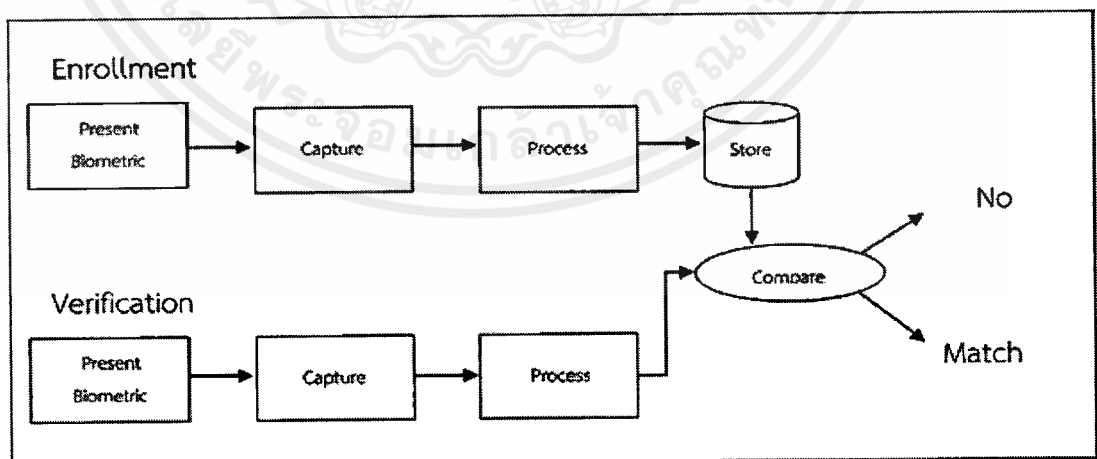
ระบบจะมีข้อมูลอยู่ 2 ชุด คือ 1. ข้อมูลลักษณะทางชีวภาพที่ได้มาจากการลงทะเบียน และนำไปประมวลผลเพื่อเก็บลงฐานข้อมูล และ 2. ข้อมูลลักษณะทางชีวภาพที่ได้มาจากการขอเข้าใช้ระบบ ซึ่งเป็นข้อมูลชุดใหม่ ข้อมูลชุดนี้ก็จะถูกนำไปประมวลผล และนำมาเปรียบเทียบกับข้อมูลที่มีในฐานข้อมูล หากไม่ตรงกับชุดข้อมูลใด ๆ ในฐานข้อมูล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

ความแตกต่างของระบบการยืนยันตัวตนด้วยไบโอเมตริกซ์กับระบบแบบดั้งเดิมคือ การแทนที่ชื่อผู้ใช้และรหัสผ่าน ด้วยชุดข้อมูลลักษณะทางชีวภาพ ซึ่งสะดวกกว่ากันมาก

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

ในส่วนยืนยันตัวตนก็จะทำเช่นเดียวกันคือ รับข้อมูลเข้ามา แปลงข้อมูลให้อยู่ในรูปแบบเดียวกัน นำไปประมวลผล และเก็บลงฐานข้อมูล เพื่อนำไปใช้เปรียบเทียบกับข้อมูลที่ลงทะเบียนไว้แล้วในฐานข้อมูล

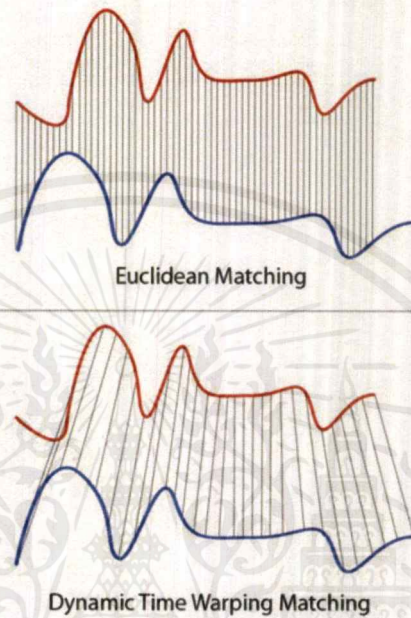
เมื่อมีการเปรียบเทียบเกิดขึ้น ระบบก็ต้องตัดสินใจให้ได้ว่า ลักษณะทางชีวภาพที่รับเข้ามานั้น ตรงกับลักษณะทางชีวภาพใดในชุดข้อมูลหรือไม่ การทำงานของระบบยืนยันตัวตนด้วยไบโอเมตริกซ์แสดงตามภาพที่ 2.2



ภาพที่ 2.1 ระบบยืนยันตัวตนด้วยไบโอเมตริกซ์

2.3 ไดนามิกไทม์วอร์ปิง (Dynamic Time Warping)

ไดนามิกไทม์วอร์ปิง เป็นขั้นตอนวิธีที่ใช้สำหรับการเปรียบเทียบความคล้ายกันระหว่างอนุกรมเวลา 2 ชุด ซึ่งจะให้ผลลัพธ์ออกมาเป็นค่าระยะทางและวิธีปรับแนว (Alignment) ที่ดีที่สุดระหว่างข้อมูลทั้งสอง ดังแสดงในภาพที่ 2.3



ภาพที่ 2.2 ภาพการเปรียบเทียบระหว่างวิธีการเปรียบเทียบแบบยูคลิเดียน และวิธีการเปรียบเทียบแบบไดนามิกไทม์วอร์ปิง

ขั้นตอนวิธีของไดนามิกไทม์วอร์ปิง

สร้างเมทริกซ์ของข้อมูลอนุกรมเวลา 2 ชุด ที่จะเปรียบเทียบกัน

$$X : (x_1, x_2, \dots, x_M)$$

$$Y : (y_1, y_2, \dots, y_N)$$

Matrix $A[i, j]$

บันทึกผลต่างระหว่าง x_1 กับ x_2 ไปจนถึง x_M กับ y_N 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 สหสัมพันธ์แบบเพียร์สัน (Pearson Correlation)

สหสัมพันธ์แบบเพียร์สัน เป็นเทคนิคทางสถิติที่ใช้หาขนาดและทิศทางของความสัมพันธ์เชิงเส้นตรงระหว่างตัวแปรที่มีมาตรวัดแบบช่วง (Interval scale) หรืออัตราส่วน (Ratio scale) สองตัว สัญลักษณ์ที่ใช้แทนค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน คือ r

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

ค่าสัมประสิทธิ์สหสัมพันธ์

- 0.81 – 1.00 : มีความสัมพันธ์กันมาก
- 0.51 – 0.80 : มีความสัมพันธ์กันปานกลาง
- 0.21 – 0.50 : มีความสัมพันธ์กันน้อยหรือต่ำ
- 0.01 – 0.20 : มีความสัมพันธ์กันน้อยมาก
- 0.00 : ไม่มีความสัมพันธ์กัน

2.5 งานวิจัยที่เกี่ยวข้อง

หัวข้อนี้จะกล่าวถึงงานวิจัยที่เกี่ยวข้อง ซึ่งประกอบไปด้วยงานวิจัยเรื่อง Learning User Keystroke Patterns for Authentication ซึ่งแต่ละงานวิจัยจะถูกกล่าวถึงในหัวข้อที่ 2.5.1 และ 2.5.2

2.5.1 Learning User Keystroke Patterns for Authentication

Keystroke authentication คือ การเข้าสู่ระบบด้วยการยืนยันตัวตนโดยวิเคราะห์จากพฤติกรรมพิมพ์ของผู้ใช้ 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%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

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

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

เมื่อพยายามที่จะเข้าถึงระบบคอมพิวเตอร์ โดยปกติจะมีการรับรองความถูกต้องผ่านชื่อผู้ใช้และรหัสผ่าน ซึ่งได้รับการพิสูจน์แล้วว่า การตรวจสอบแบบนี้ไม่ได้ปลอดภัยอย่างสมบูรณ์ แต่การวิเคราะห์รูปแบบการกดแป้นพิมพ์ สามารถใช้เป็นมาตรวัดเสริมเพื่อป้องกันการบุกรุกได้ การวิเคราะห์รูปแบบการกดแป้นพิมพ์แตกต่างจากวิธีอื่น เพราะเป็นการปฏิสัมพันธ์ระหว่างคอมพิวเตอร์กับมนุษย์โดยตรง การพิสูจน์ตัวตนด้วยการกดแป้นพิมพ์เป็นสาขาย่อยของการรักษาความปลอดภัยของระบบคอมพิวเตอร์ การศึกษาจำนวนมากได้แสดงให้เห็นถึงเอกลักษณ์ของรูปแบบการกดแป้นพิมพ์ของแต่ละบุคคล ในระบบที่ใช้รูปแบบการกดแป้นเป็นส่วนหนึ่งในการรับรองความถูกต้องของกระบวนการเข้าสู่ระบบไม่เพียง แต่ต้องชื่อผู้ใช้และรหัสผ่านที่ถูกต้อง แต่ยังจับคู่เฉพาะรูปแบบการกดแป้นพิมพ์ที่ได้รับการระบุและจัดเก็บสำหรับแต่ละรายบุคคล กลไกดังกล่าวสามารถเพิ่มความยากของผู้บุกรุกได้ เนื่องจากรูปแบบการพิมพ์ที่ของแต่ละคนเป็นเอกลักษณ์จึงเป็นการยากที่จะทำซ้ำ

ใช้ช่วงเวลากดแป้นอาจมีการสร้างตัวระบุลายเซ็นที่มีประสิทธิภาพมากขึ้น การใช้งานไม่จำเป็นต้องมีการคำนวณมากเกินไปและส่วนใหญ่เป็นแพลตฟอร์มอิสระ เป้าหมายหลักของการพิสูจน์ตัวตนด้วยช่วงเวลาการกดแป้นพิมพ์การกำหนดเวลาคือการป้องกันผู้บุกรุกโดย False Rejection Rate ควรมีค่ามากและ False Accept Rate ควรมีค่าน้อย

ในบทความนี้มุ่งเน้นไปที่กลุ่มผู้ใช้คอมพิวเตอร์โดยเฉพาะ ผู้ที่ใช้แป้นพิมพ์เป็นประจำ
 นี้คือปัญหาที่สำคัญเพราะโดยปกติมนุษย์จะมีการพิมพ์ที่ไม่เหมือนเดิม ถ้าคีย์บอร์ด
 เปลี่ยนแปลง การคุ้นเคยกับแป้นพิมพ์ที่ไม่รู้จักต้องใช้เวลา ผลที่ตามมา การกำหนดเวลา
 ช่วงเวลา keystroke เหมาะสมที่สุดสำหรับผู้ใช้คอมพิวเตอร์ที่มีการพิมพ์ได้อย่างมี
 ประสิทธิภาพในคีย์บอร์ดที่คุ้นเคย ระบบการรับรองความถูกต้องโดยใช้ keystroke ความ
 เป็นไปได้ของการโจมตีด้วยรหัสผ่านโดยผู้บุกรุก (เช่น Dictionary Attack) เพราะถึงแม้จะมี
 ข้อมูลล็อกอินและรหัสผ่านก็ตามเป็นเรื่องยากมากสำหรับผู้บุกรุกที่จะทำซ้ำรูปแบบการกด
 แป้น การโจมตีเกี่ยวกับพฤติกรรมทางชีวภาพเป็นเรื่องยากเพราะพฤติกรรมเชิงชีวภาพเป็น
 เอกลักษณะของแต่ละคน เป็นเรื่องยากที่จะทำซ้ำรูปแบบการกดแป้นพิมพ์แม้จะมีการฝึกมากก็
 ตา



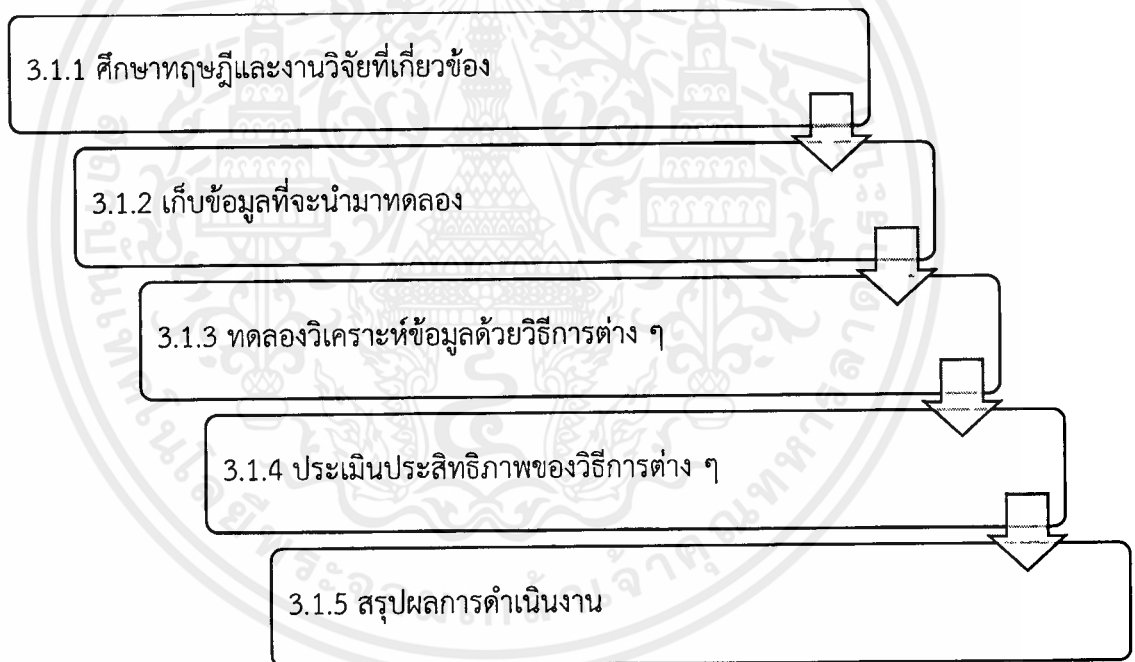
บทที่ 3

วิธีการดำเนินงานวิจัย

ในบทนี้จะกล่าวถึงการดำเนินการวิจัยเรื่องการระบุตัวบุคคลโดยการวิเคราะห์รูปแบบที่ได้จากการพิมพ์ โดยเนื้อหาประกอบด้วย ระเบียบวิธีการระบุตัวบุคคล เก็บข้อมูลที่จะนำมาทดลอง ทดลอง วิเคราะห์ข้อมูลด้วยวิธีการต่าง 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

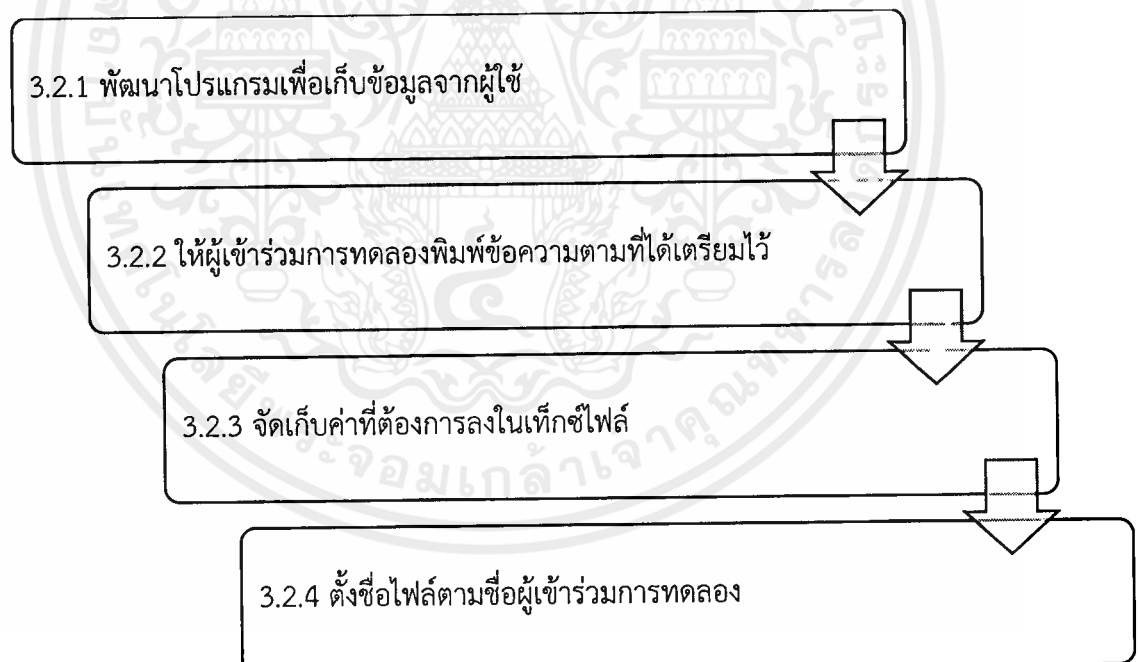
ประเมินประสิทธิภาพด้วยวิธีการที่เหมาะสม และทำการทดสอบทางสถิติ

3.1.5 สรุปผลการดำเนินงาน

สรุปผลการดำเนินงาน และเปรียบเทียบผลลัพธ์ที่ได้กับงานวิจัยอื่น ๆ

3.2 การเก็บข้อมูลที่จะนำมาทดลอง

การเก็บข้อมูลที่จะนำมาทดลอง มีลำดับขั้นตอนการทำงานดังภาพที่ 3.2

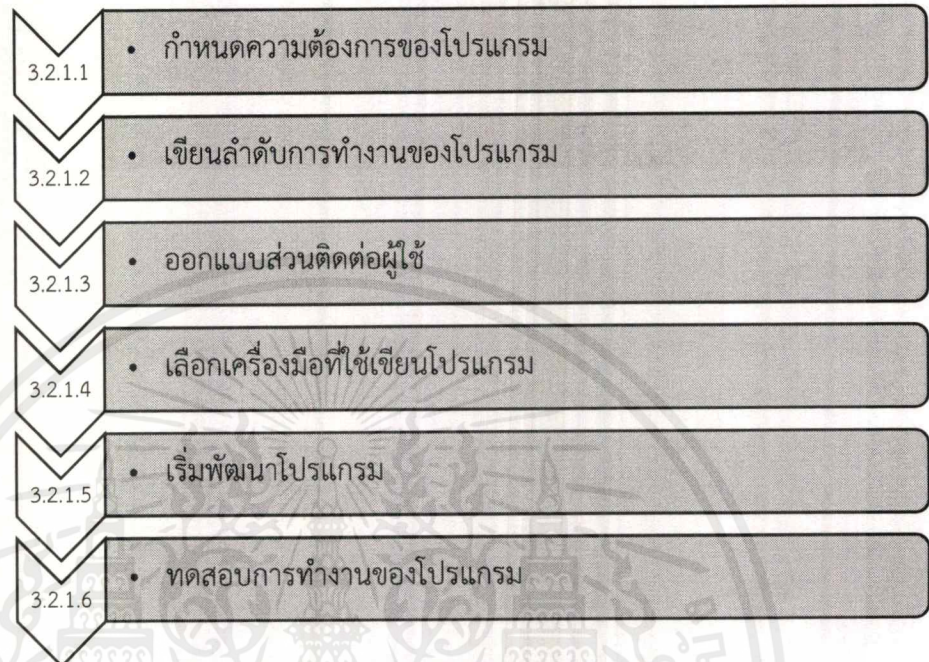


ภาพที่ 3.2 ขั้นตอนการเก็บข้อมูลจากผู้เข้าร่วมการทดลอง

3.2.1 พัฒนาโปรแกรมเพื่อเก็บข้อมูลจากผู้ใช้

การพัฒนาโปรแกรมเพื่อเก็บข้อมูลจากผู้ใช้ มีลำดับขั้นตอนการทำงานตามภาพที่

3.3



ภาพที่ 3.3 ลำดับขั้นตอนการพัฒนาโปรแกรมเพื่อเก็บข้อมูลจากผู้ใช้

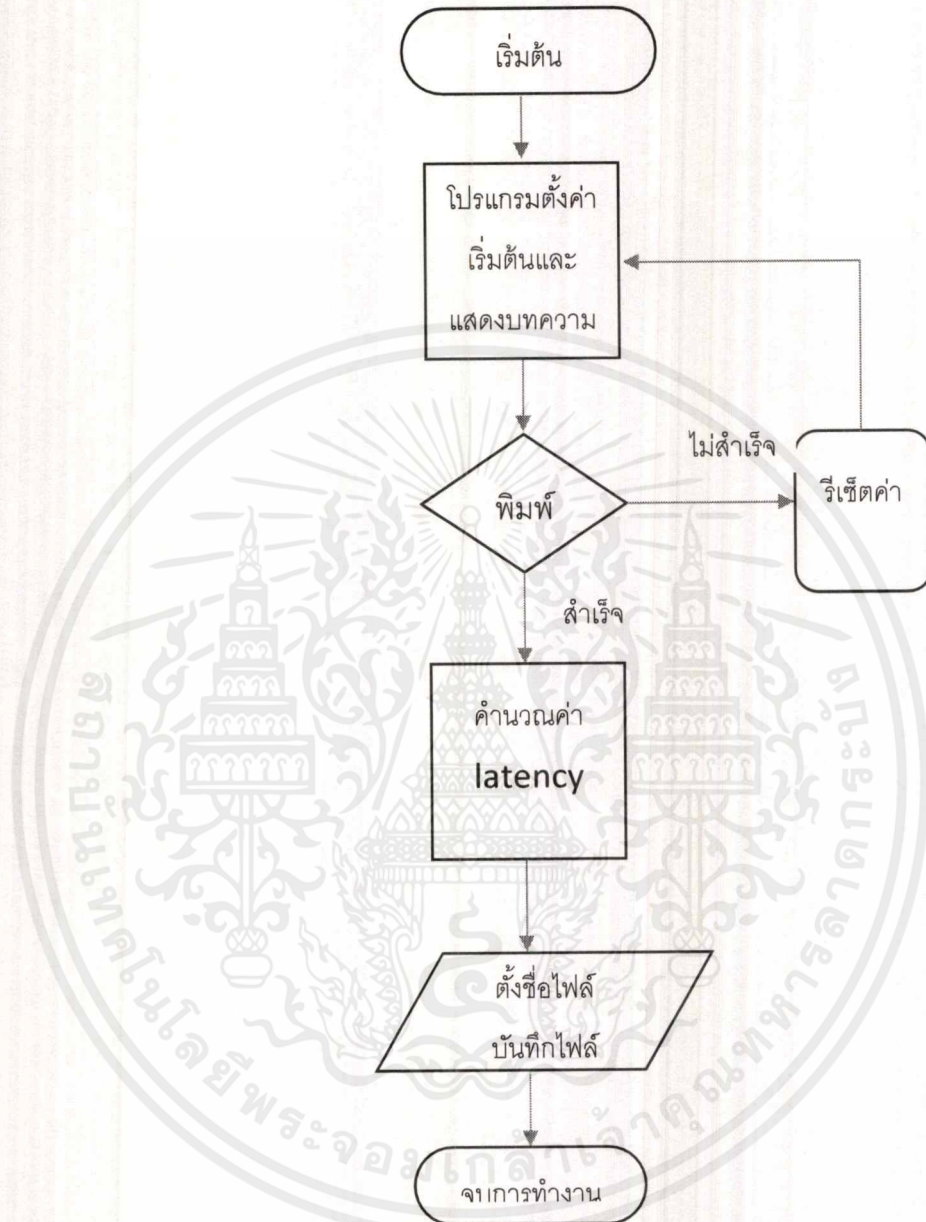
3.2.1.1 กำหนดความต้องการของโปรแกรม

- 1) โปรแกรมต้องการเก็บค่า 3 ค่า ในระหว่างที่ผู้เข้าร่วมการทดลองทำการพิมพ์อยู่ โดยค่าดังกล่าวคือ
 - a. ตำแหน่งของแป้นบนคีย์บอร์ดที่ผู้เข้าร่วมการทดลองกด
 - b. เวลาที่ผู้เข้าร่วมการทดลองกดแป้นนั้น
 - c. เวลาที่ผู้เข้าร่วมการทดลองปล่อยแป้นนั้น
- 2) หน่วยของเวลาต้องมีความละเอียดพอที่จะเห็นความเปลี่ยนแปลง
- 3) มีบทความให้ผู้เข้าร่วมการทดลองดู ขณะที่ทำการพิมพ์
- 4) เมื่อผู้เข้าร่วมการทดลองพิมพ์บทความเสร็จ ใช้ข้อมูลที่ได้รับมาคำนวณค่า latency
- 5) นำค่าทั้งหมดที่เก็บได้ ไปเขียนลงในไฟล์ ในฟอร์มเมตที่เหมาะสม เพื่อนำไปวิเคราะห์ต่อไป
- 6) มีการบันทึกชื่อผู้เข้าร่วมการทดลอง เพื่อเป็นประโยชน์ในการทดสอบผลการวิเคราะห์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.1.2 เขียนลำดับการทำงานของโปรแกรม

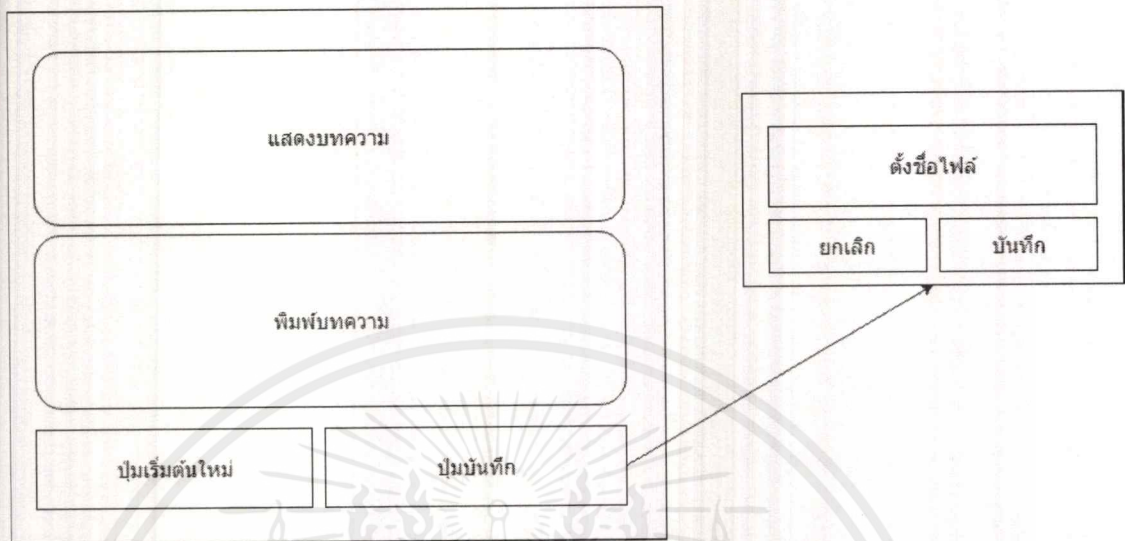
ลำดับการทำงานของโปรแกรมสามารถเขียนออกมาได้ดังภาพที่ 3.4



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 เวอร์ชัน 1.8.0 ในการคอมไพล์ซอร์สโค้ด
- 3) ใช้ JRE เวอร์ชัน 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

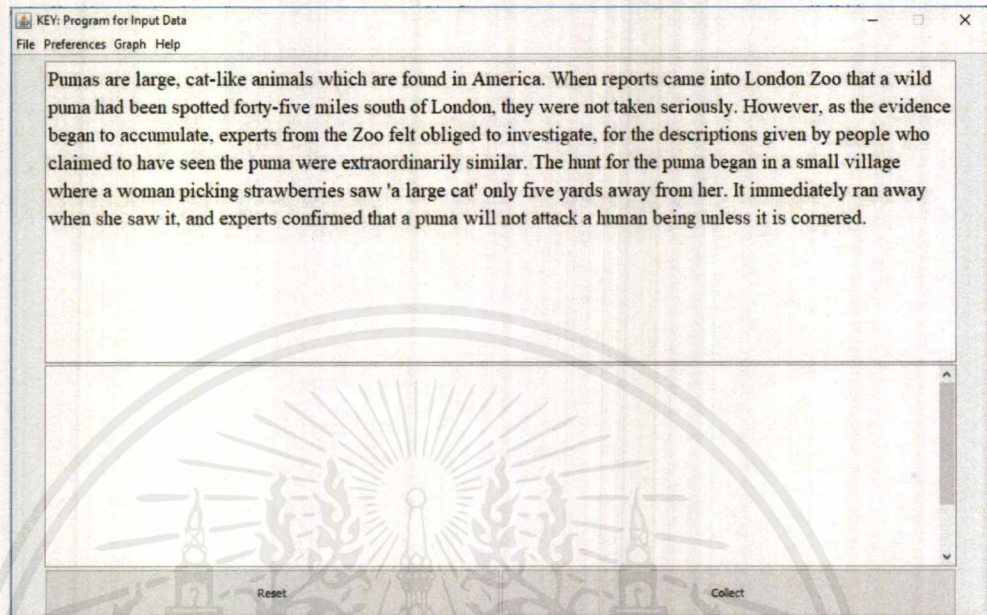
- 1) สร้างสภาพแวดล้อมเพื่อให้ง่ายต่อการพัฒนาโปรแกรมด้วย Maven
- 2) เริ่มพัฒนาโปรแกรมให้เป็นไปตามลำดับการทำงานที่วางไว้
- 3) แปลงไฟล์ทั้งหมดที่จำเป็นต่อการทำงานของโปรแกรมให้อยู่ในแพ็คเกจ JAR เพื่อง่ายต่อการเรียกใช้ในเครื่องอื่น ๆ

3.2.1.6 ทดสอบการทำงานของโปรแกรม

ทดสอบการทำงานของโปรแกรมโดยผู้ทำการทดลอง

3.2.2 ให้ผู้เข้าร่วมการทดลองพิมพ์ข้อความที่ได้เตรียมไว้

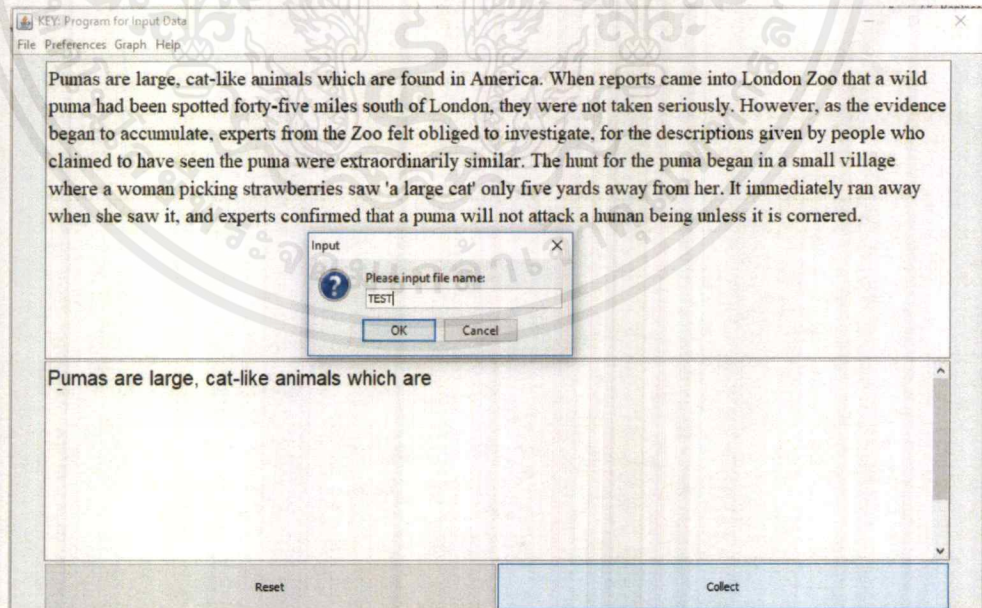
เก็บข้อมูลโดยการให้ผู้เข้าร่วมการทดลองพิมพ์ข้อความที่แสดงไว้ด้านบน ลงในช่องเก็บข้อมูลด้านล่าง ดังภาพที่ 3.6



ภาพที่ 3.6 หน้าจอส่วนติดต่อกับผู้ใช้งานขณะพิมพ์ข้อมูล

3.2.3 ตั้งชื่อไฟล์ตามชื่อผู้เข้าร่วมการทดลอง

ทำการตั้งชื่อไฟล์ ก่อนที่จะจัดเก็บข้อมูล ดังภาพที่ 3.7



ภาพที่ 3.7 การตั้งชื่อไฟล์ก่อนบันทึกข้อมูล

3.2.4 จัดเก็บข้อมูลที่ไต้ลงในเท็กซ์ไฟล์

จัดเก็บข้อมูลที่ไต้ลงในเท็กซ์ไฟล์นามสกุล .json ดังภาพที่ 3.8 ซึ่งโครงสร้างข้อมูลของไฟล์ .json จะมีหน้าตาดังภาพที่ 3.9



1_THANAPHAT_EN.json	2/2/2562 2:42	JSON File	18 KB
1_THANAPHAT_TH.json	2/2/2562 2:42	JSON File	6 KB
2_NATTAPONG_EN.json	2/2/2562 2:42	JSON File	15 KB
2_NATTAPONG_TH.json	2/2/2562 2:42	JSON File	6 KB
3_WIPADA_EN.json	2/2/2562 2:42	JSON File	15 KB
3_WIPADA_TH.json	2/2/2562 2:42	JSON File	6 KB
4_SURACHAI_EN.json	2/2/2562 2:42	JSON File	15 KB
4_SURACHAI_TH.json	2/2/2562 2:42	JSON File	6 KB
5_SIRATTAYA_EN.json	2/2/2562 2:42	JSON File	16 KB
5_SIRATTAYA_TH.json	2/2/2562 2:42	JSON File	6 KB
6_NISARA_EN.json	2/2/2562 2:42	JSON File	18 KB
6_NISARA_TH.json	2/2/2562 2:42	JSON File	6 KB
7_PATCHAREEPORN_EN.json	2/2/2562 2:42	JSON File	14 KB
7_PATCHAREEPORN_TH.json	2/2/2562 2:42	JSON File	6 KB
8_WARISARA_EN.json	2/2/2562 2:42	JSON File	15 KB
8_WARISARA_TH.json	2/2/2562 2:42	JSON File	7 KB
9_METIDA_EN.json	2/2/2562 2:42	JSON File	17 KB
9_METIDA_TH.json	2/2/2562 2:42	JSON File	9 KB
10_YANEE_EN.json	2/2/2562 2:42	JSON File	15 KB
10_YANEE_TH.json	2/2/2562 2:42	JSON File	6 KB
11_AEKKASIT_EN.json	2/2/2562 2:42	JSON File	18 KB
11_AEKKASIT_TH.json	2/2/2562 2:42	JSON File	7 KB
12_JANIIRA_EN.json	2/2/2562 2:42	JSON File	16 KB
12_JANIIRA_TH.json	2/2/2562 2:42	JSON File	9 KB
13_SUPUTSORN_EN.json	2/2/2562 2:42	JSON File	15 KB
13_SUPUTSORN_TH.json	2/2/2562 2:42	JSON File	6 KB
14_SAKDA_EN.json	2/2/2562 2:42	JSON File	15 KB
14_SAKDA_TH.json	2/2/2562 2:42	JSON File	7 KB
15_TEERAPONG_EN.json	2/2/2562 2:42	JSON File	15 KB

ภาพที่ 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]
}
```

ภาพที่ 3.9 ภาพตัวอย่างโครงสร้างข้อมูลที่ได้มาจากผู้ใช้

จากภาพที่ 3.9 คือตัวอย่างโครงสร้างข้อมูลที่ถูกทดลองทำการเก็บจากผู้ใช้ ประกอบไปด้วย

‘Name’ หมายถึง ชื่อของผู้ที่ทำการพิมพ์

‘Language’ หมายถึง ภาษาที่ใช้เป็นตัวอย่างการพิมพ์

‘KeyPressed’ หมายถึง ค่าเวลาในขณะที่ผู้ใช้กดแป้นหนึ่งแป้น

‘KeyReleased’ หมายถึง ค่าเวลาในขณะที่ผู้ใช้ปล่อยแป้นหนึ่งแป้น

‘KeyCode’ หมายถึง รหัสของแป้นที่ผู้ใช้กด

‘isEqual’ หมายถึง ตัวแปรที่คอยตรวจสอบว่า จำนวนค่าของ KeyPressed และ KeyReleased เท่ากันหรือไม่

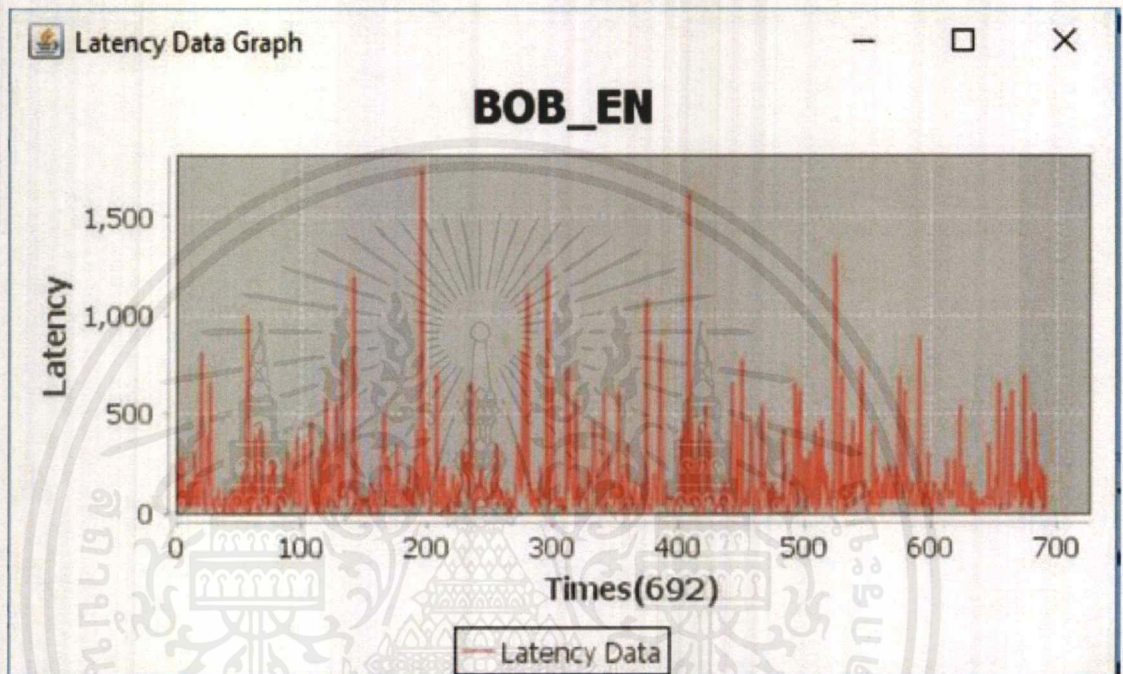
‘Latency’ หมายถึง เป็นค่าที่คำนวณได้จาก เวลาของการปล่อยแป้นที่แล้วลบกับเวลาของการกดแป้นถัดไป

3.3 ทดลองวิเคราะห์ข้อมูลด้วยวิธีการต่าง ๆ

การทดลองวิเคราะห์ข้อมูลด้วยวิธีการต่าง ๆ มีลำดับขั้นตอนการทำงานดังนี้

3.3.1 สร้างกราฟเพื่อแทนตัวบุคคล

- นำค่า latency ที่คำนวณได้มาพล็อตเป็นกราฟ ซึ่งเราตั้งสมมติฐานว่ากราฟเป็นตัวแทนของรูปแบบการพิมพ์ของแต่ละบุคคล ก็จะได้ดังภาพที่ 3.10



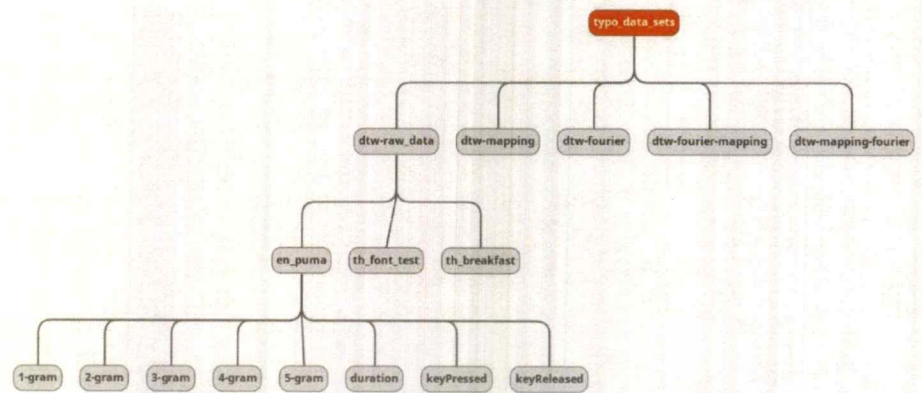
ภาพที่ 3.10 ภาพค่า latency เมื่อนำมาแสดงผลในรูปแบบกราฟ

แกน y (Latency) หมายถึง ช่วงระยะเวลาระหว่างที่ปล่อยแป้นไปจนถึงการกดแป้นถัดไป ซึ่งคำนวณได้จาก $|เวลาของการปล่อยแป้นที่แล้ว - เวลาของการกดแป้น|$ เช่น เมื่อผู้ใช้พิมพ์คำว่า cat โปรแกรมจะคำนวณค่า latency ออกมาได้ 2 ค่าคือค่าที่เกิดขึ้นระหว่างที่ปล่อยแป้น 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 และ Pearson

3.3.3 ทดลองวิเคราะห์ข้อมูลของ typo data sets



ภาพที่ 3.11 ภาพกระบวนการวิเคราะห์ข้อมูล typo data sets

จากภาพที่ 3.11 คือ ภาพขั้นตอนกระบวนการวิเคราะห์ตั้งแต่เริ่มจนเสร็จสิ้นของข้อมูล typo data sets

typo data sets คือ วิธีการที่ผู้ทำการทดลองไม่ได้ทำการเปลี่ยนแปลงใด ๆ กับข้อมูล เป็นข้อมูลดิบ ๆ ที่เก็บมาได้จากผู้เข้าร่วมการทดลอง

ต่อไปนี้เป็นวิธีการทดลองในแบบต่าง ๆ

dtw-raw-data คือ การใช้วิธีการ dynamic time warping ในการคำนวณระยะทางของกราฟ ก่อนที่จะนำไปทำนายผล ทำแบบนี้กับทุกชุดข้อมูล en_puma, th_font_test และ th_breakfast

dtw-mapping คือ การใช้วิธีการ mapping ในการทำ normalization ก่อนที่จะนำไปคำนวณระยะทางด้วยวิธีการ dynamic time warping แล้วนำไปทำนายผลในขั้นตอนสุดท้าย ทำแบบนี้กับทุกชุดข้อมูล en_puma, th_font_test และ th_breakfast

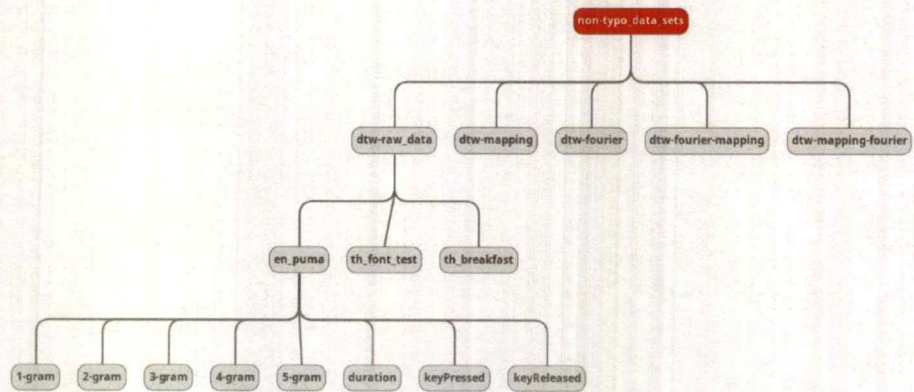
dtw-fourier คือ การใช้วิธีการ fourier ก่อนที่จะนำไปคำนวณระยะทางด้วยวิธีการ dynamic time warping แล้วนำไปทำนายผลในขั้นตอนสุดท้าย ทำแบบนี้กับทุกชุดข้อมูล en_puma, th_font_test และ th_breakfast

dtw-fourier-mapping คือ การใช้วิธีการ fourier ก่อนที่จะนำไปคำนวณระยะทางด้วยวิธีการอีกครั้ง แล้วจึงใช้วิธีการ dynamic time warping ในการคำนวณระยะทาง แล้วนำไปทำนายผลในขั้นตอนสุดท้าย ทำแบบนี้กับทุกชุดข้อมูล en_puma, th_font_test และ th_breakfast

dtw-mapping-fourier คือ การใช้วิธีการ mapping ก่อนที่จะใช้วิธีการ fourier แล้วจึงใช้วิธีการ dynamic time warping ในการคำนวณระยะทาง แล้วนำไปทำนายผลในขั้นตอนสุดท้าย ทำแบบนี้กับทุกชุดข้อมูล en_puma, th_font_test และ th_breakfast

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3.4 ทดลองวิเคราะห์ข้อมูลของ non typo data sets

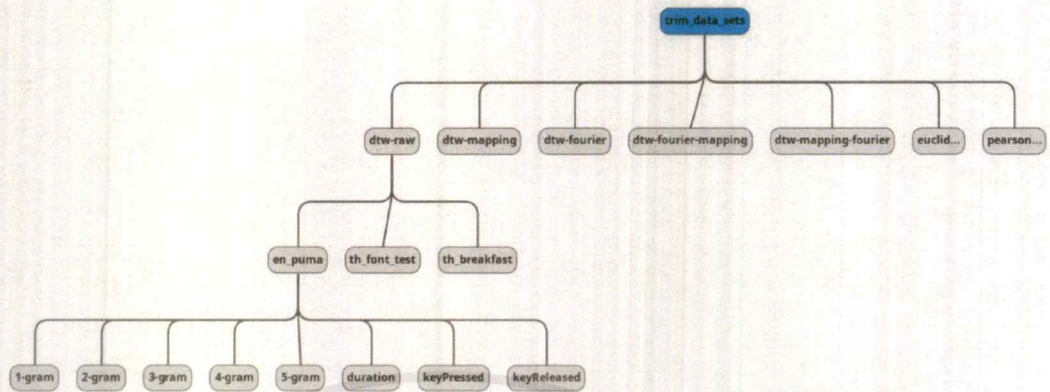


ภาพที่ 3.12 ภาพกระบวนการวิเคราะห์ข้อมูล non typo data sets

จากภาพที่ 3.12 คือ ภาพขั้นตอนกระบวนการวิเคราะห์ตั้งแต่เริ่มจนเสร็จสิ้นของข้อมูล non typo data sets

`non-typo-data_sets` คือ ชุดข้อมูลที่ถูกตัดข้อมูลบางส่วนออก โดยจะเป็นในส่วนของช่วงข้อมูลที่มีค่า `backspace` และตัวอักษรก่อนหน้า ก่อนที่จะนำไปผ่านกระบวนการทดลองเหมือนเดิมอีกครั้ง ตามวิธีการข้างต้น

3.3.5 ทดลองวิเคราะห์ข้อมูลของ trim data sets



ภาพที่ 3.13 ภาพกระบวนการวิเคราะห์ข้อมูล trim data sets

จากภาพที่ 3.13 คือ ภาพขั้นตอนกระบวนการวิเคราะห์ตั้งแต่เริ่มจนเสร็จสิ้นของข้อมูล trim data sets

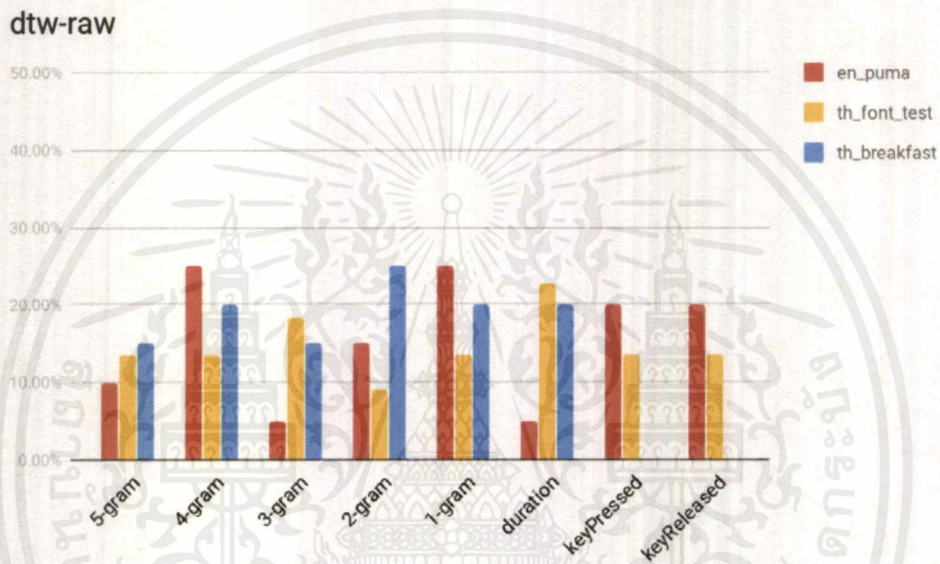
`trim_data_sets` คือ วิธีการตัดข้อมูลจากวิธีการที่แล้ว ให้มีจำนวนเท่ากับชุดข้อมูลที่ควรจะเป็น ก่อนที่จะนำไปผ่านการวิเคราะห์ข้อมูลอีกครั้ง ตามแบบเดิมที่ได้ทำมา และเพิ่มวิธีการคำนวณระยะทางของกราฟทั้ง euclid และ pearson จากเดิมที่มีแค่ fourier อย่างเดียว

บทที่ 4

ผลการวิจัยและการอภิปรายผล

ในบทนี้จะกล่าวถึง ผลการดำเนินงาน และการทดสอบประสิทธิภาพ

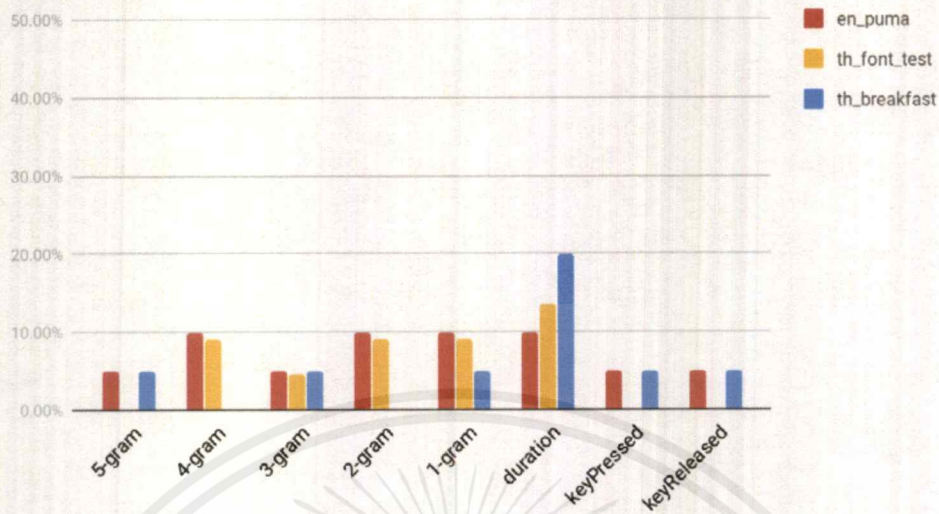
4.1 ผลการดำเนินงานของ typo data sets



ภาพที่ 4.1 กราฟเปรียบเทียบค่าความถูกต้องของ typo data sets โดยใช้วิธี dtw-raw

จากภาพที่ 4.1 คือ ภาพผลการดำเนินงานของ typo data sets ที่ใช้วิธี dtw-raw ผลปรากฏว่า 4-gram, 2-gram และ 1-gram ให้ผลลัพธ์ที่ดีที่สุด 25% ส่วน 3-gram และ duration ให้ผลลัพธ์แย่งที่สุด 5%

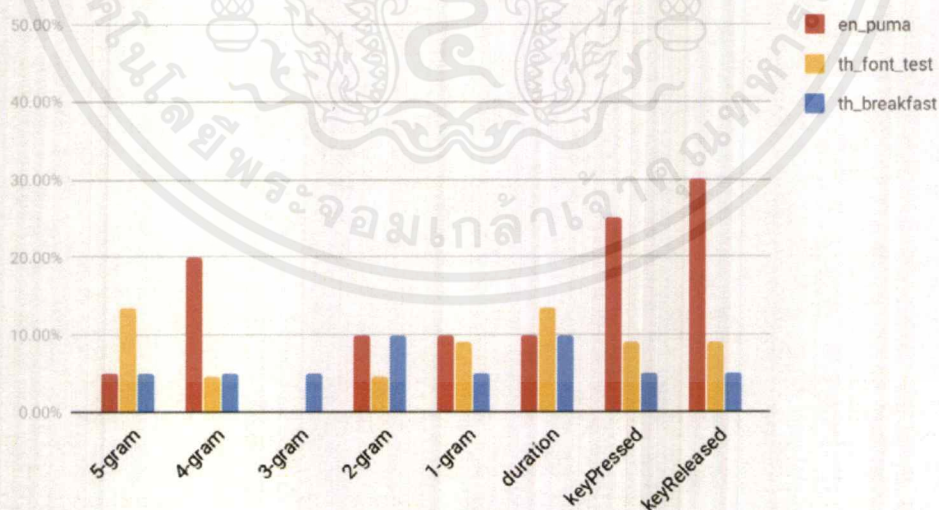
dtw-mapping



ภาพที่ 4.2 กราฟเปรียบเทียบค่าความถูกต้องของ typo data sets โดยใช้วิธี dtw-mapping

จากภาพที่ 4.2 คือ ภาพผลการดำเนินงานของ typo data sets ที่ใช้วิธี dtw-mapping ผลปรากฏว่า duration ให้ผลลัพธ์ที่ดีที่สุด 20% ส่วน 5-gram, 3-gram, 1-gram, keyPressed และ keyReleased ให้ผลลัพธ์แย่ที่สุด 5%

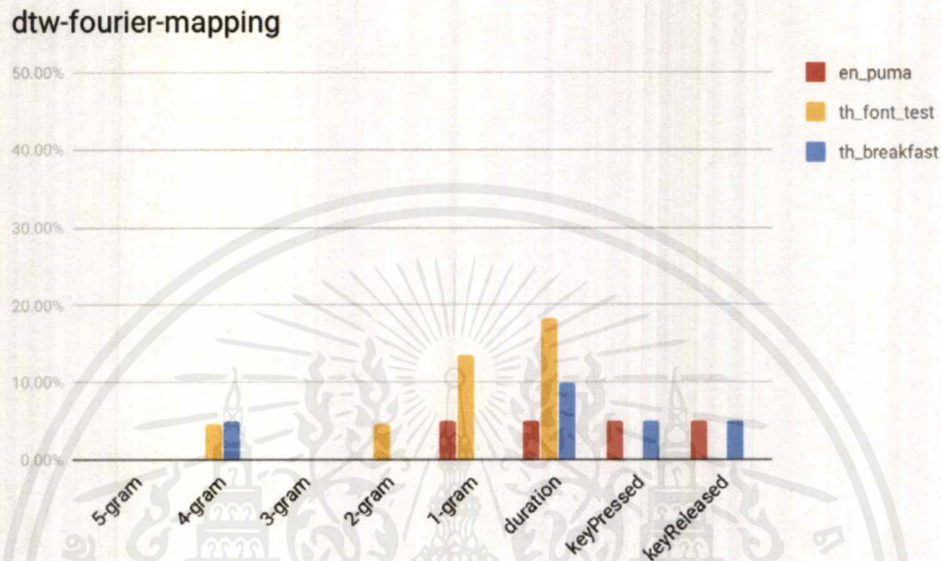
dtw-fourier



ภาพที่ 4.3 กราฟเปรียบเทียบค่าความถูกต้องของ typo data sets โดยใช้วิธี dtw-fourier

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

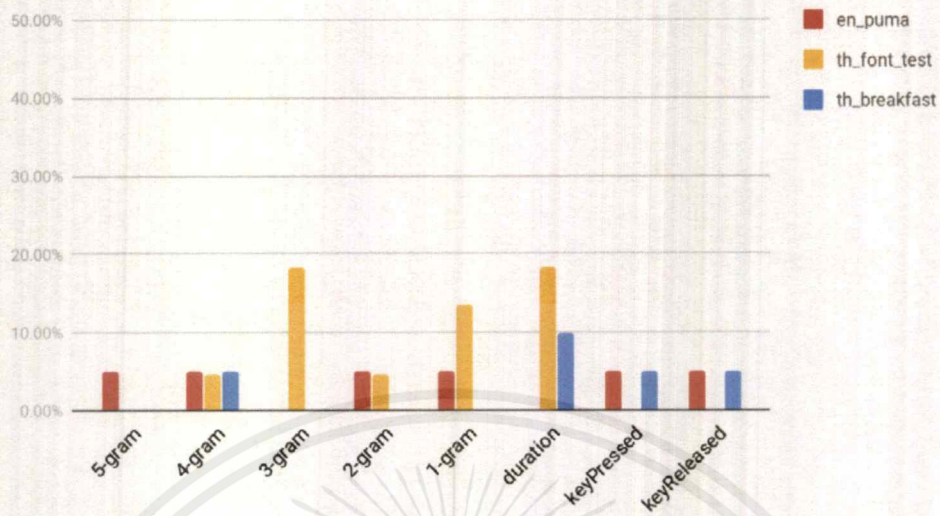
จากภาพที่ 4.3 คือ ภาพผลการดำเนินงานของ typo data sets ที่ใช้วิธี dtw-fourier ผลปรากฏว่า keyReleased ให้ผลลัพธ์ที่ดีที่สุด 30% ส่วน 5-gram, 3-gram, 2-gram, 1-gram ให้ผลลัพธ์แย่งที่สุด 5%



ภาพที่ 4.4 กราฟเปรียบเทียบค่าความถูกต้องของ typo data sets โดยใช้วิธี dtw-fourier-mapping

จากภาพที่ 4.4 คือ ภาพผลการดำเนินงานของ typo data sets ที่ใช้วิธี dtw-fourier-mapping ผลปรากฏว่า duration ให้ผลลัพธ์ที่ดีที่สุด 18% ส่วน 5-gram, 3-gram ให้ผลลัพธ์แย่งที่สุด 0%

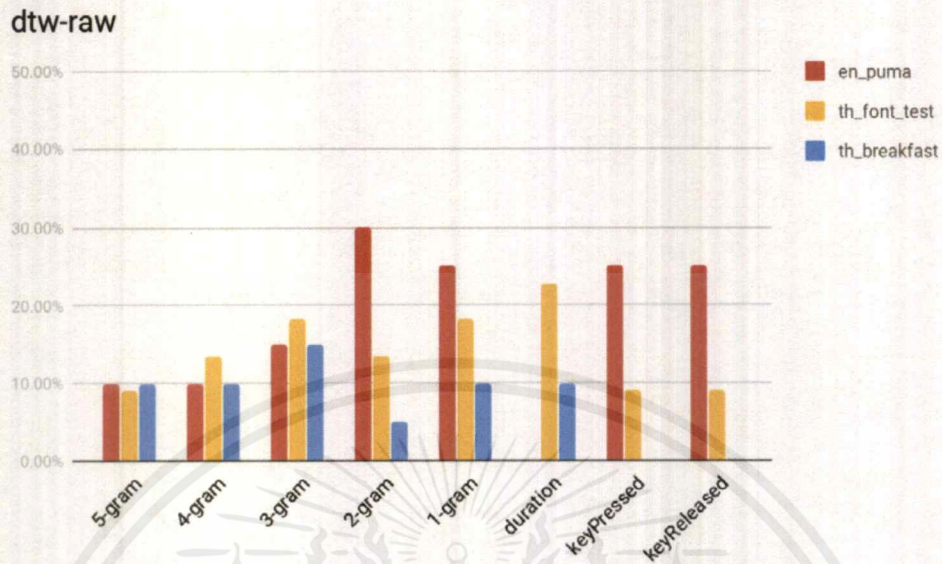
dtw-mapping-fourier



ภาพที่ 4.5 กราฟเปรียบเทียบค่าความถูกต้องของ typo data sets โดยใช้วิธี dtw-mapping-fourier

จากภาพที่ 4.5 คือ ภาพผลการดำเนินงานของ typo data sets ที่ใช้วิธี dtw-mapping-fourier ผลปรากฏว่า 3-gram และ duration ให้ผลลัพธ์ที่ดีที่สุด 18% ส่วน 5-gram, 4-gram, 2-gram, keyPressed, keyReleased ให้ผลลัพธ์แย่ที่สุด 5%

4.2 ผลการดำเนินงานของ non typo data sets

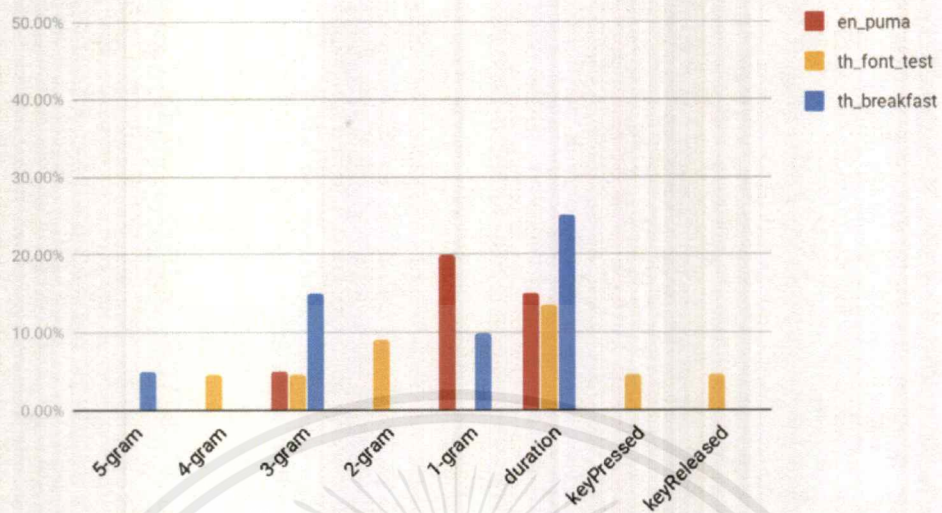


ภาพที่ 4.6 กราฟเปรียบเทียบค่าความถูกต้องของ non typo data sets โดยใช้วิธี

dtw-raw

จากภาพที่ 4.6 คือ ภาพผลการดำเนินงานของ non typo data sets ที่ใช้วิธี dtw-raw ผลปรากฏว่า 2-gram (en-puma) ให้ผลลัพธ์ที่ดีที่สุด 30% ส่วน 2-gram (th_breakfast) ให้ผลลัพธ์แย่มากที่สุด 5%

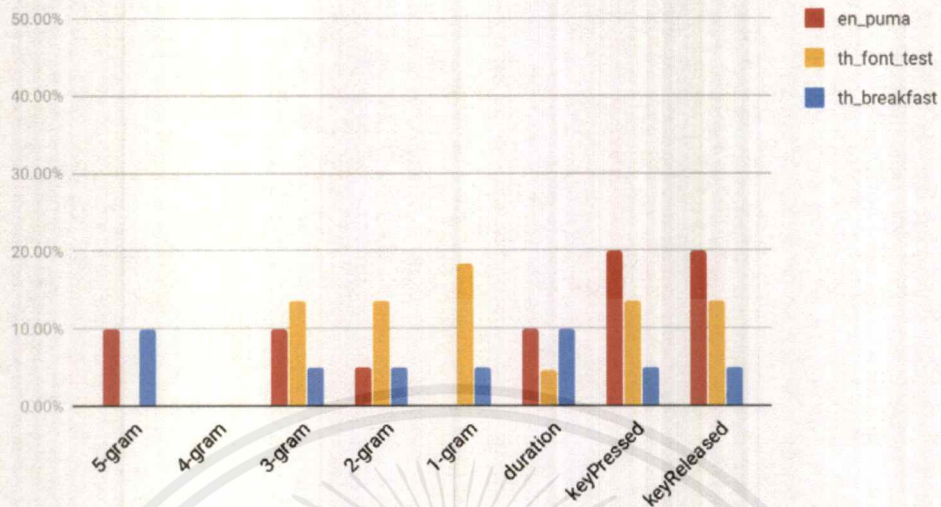
dtw-mapping



ภาพที่ 4.7 กราฟเปรียบเทียบค่าความถูกต้องของ non typo data sets โดยใช้วิธี dtw-mapping

จากภาพที่ 4.7 คือ ภาพผลการดำเนินงานของ non typo data sets ที่ใช้วิธี dtw-mapping ผลปรากฏว่า duration ให้ผลลัพธ์ดีที่สุดที่สุด 25% ส่วน 5-gram, 4gram, 3-gram, keyPressed, keyReleased ให้ผลลัพธ์แย่ที่สุด 5%

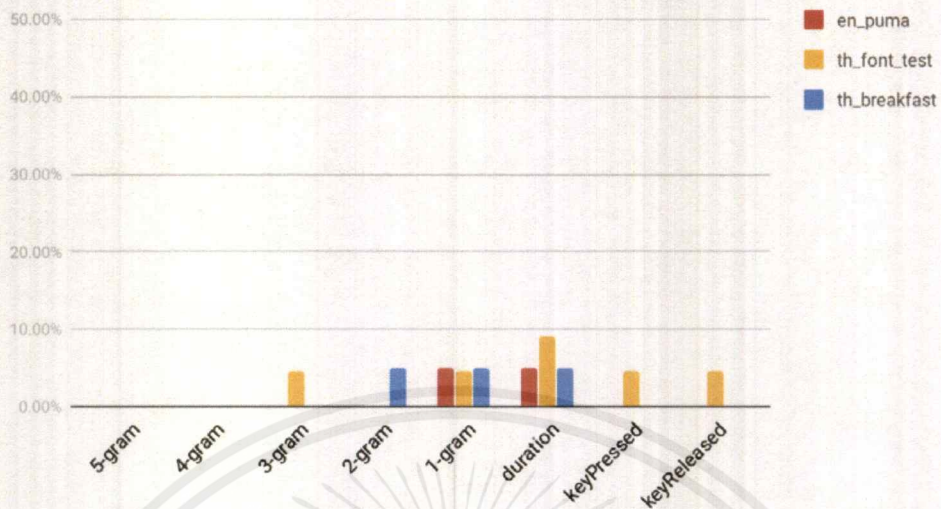
dtw-fourier



ภาพที่ 4.8 กราฟเปรียบเทียบค่าความถูกต้องของ non typo data sets โดยใช้วิธี dtw-fourier

จากภาพที่ 4.8 คือ ภาพผลการดำเนินงานของ non typo data sets ที่ใช้วิธี dtw-fourier ผลปรากฏว่า keyPressed, keyReleased ให้ผลลัพธ์ดีที่สุดที่ 20% ส่วน 3-gram, 2-gram, 1-gram, duration ให้ผลลัพธ์แย่ที่สุด 5%

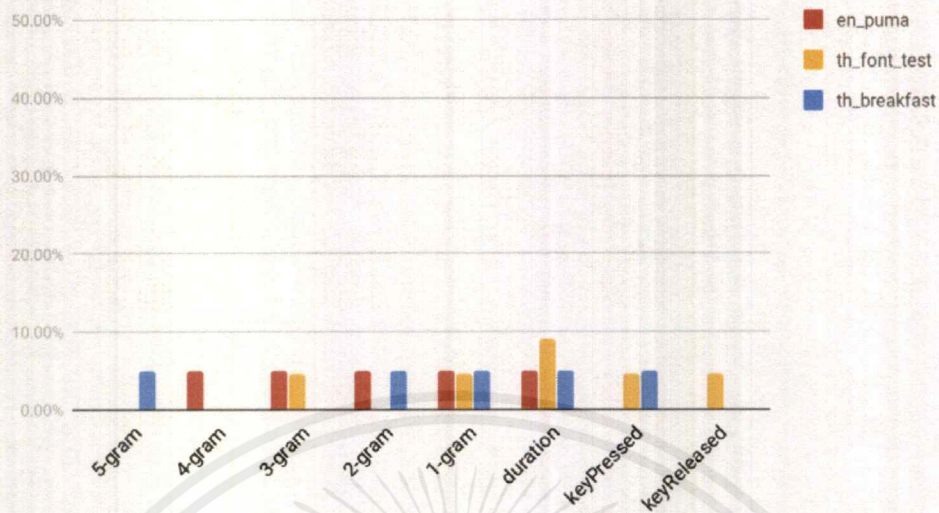
dtw-fourier-mapping



ภาพที่ 4.9 กราฟเปรียบเทียบค่าความถูกต้องของ non typo data sets โดยใช้วิธี dtw-fourier-mapping

จากภาพที่ 4.9 คือ ภาพผลการดำเนินงานของ non typo data sets ที่ใช้วิธี dtw-fourier-mapping ผลปรากฏว่า duration ให้ผลลัพธ์ดีที่สุดที่ 9% ส่วน 5-gram, 4-gram ให้ผลลัพธ์แย่มากที่สุด 0%

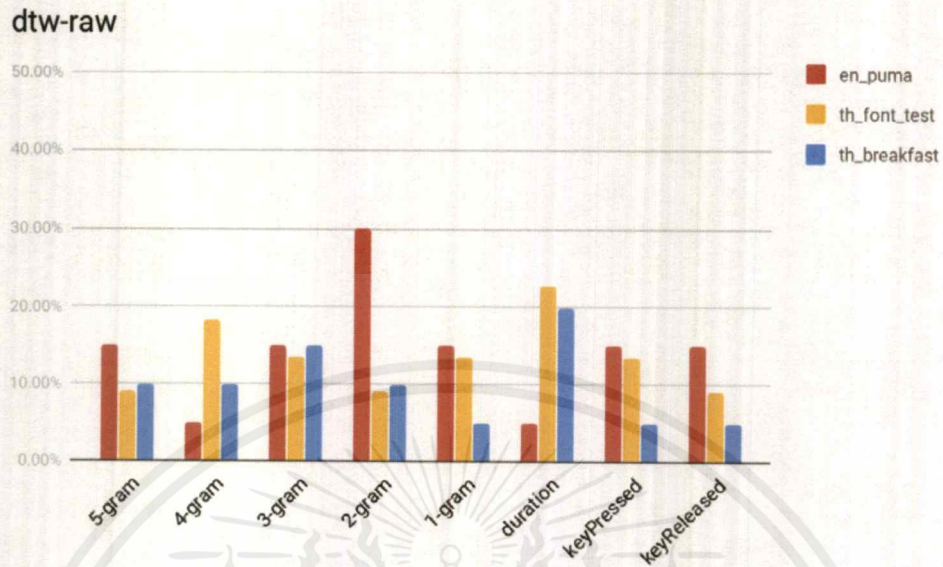
dtw-mapping-fourier



ภาพที่ 4.10 กราฟเปรียบเทียบค่าความถูกต้องของ non typo data sets โดยใช้วิธี dtw-mapping-fourier

จากภาพที่ 4.10 คือ ภาพผลการดำเนินงานของ non typo data sets ที่ใช้วิธี dtw-mapping-fourier ผลปรากฏว่า duration ให้ผลลัพธ์ดีที่สุดที่ 9% ส่วน 5-gram, 4-gram, 3-gram, 2-gram, 1-gram, keyPressed, keyReleased ให้ผลลัพธ์แย่งที่สุด 5%

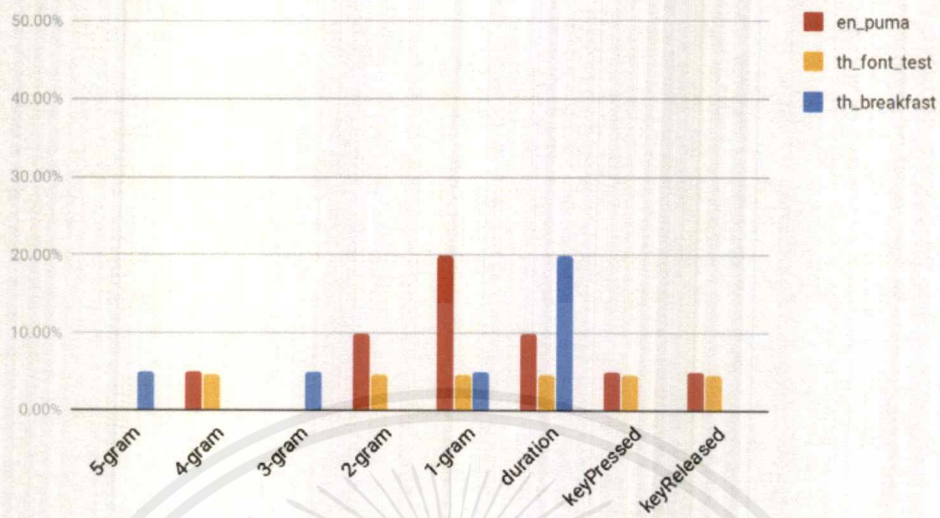
4.3 ผลการดำเนินงานของ trim data sets



ภาพที่ 4.11 กราฟเปรียบเทียบค่าความถูกต้องของ trim data sets โดยใช้วิธี dtw-raw

จากภาพที่ 4.11 คือ ภาพผลการดำเนินงานของ trim data sets ที่ใช้วิธี dtw-raw ผลปรากฏว่า 2-gram ให้ผลลัพธ์ดีที่สุดที่ 30% ส่วน 4-gram, 1-gram, duration, keyPressed, keyReleased ให้ผลลัพธ์แย่ที่สุด 5%

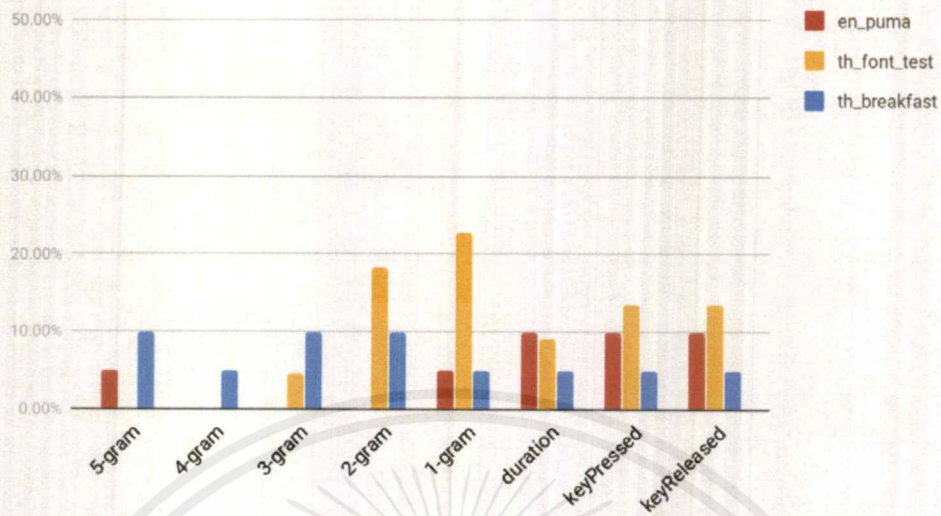
dtw-mapping



ภาพที่ 4.12 กราฟเปรียบเทียบค่าความถูกต้องของ trim data sets โดยใช้วิธี dtw-mapping

จากภาพที่ 4.12 คือ ภาพผลการดำเนินงานของ trim data sets ที่ใช้วิธี dtw-mapping ผลปรากฏว่า 1-gram, duration ให้ผลลัพธ์ดีที่สุดที่ 20% ส่วน 5-gram, 4-gram, 3-gram, 2-gram, keyPressed, keyReleased ให้ผลลัพธ์แย่ที่สุด 5%

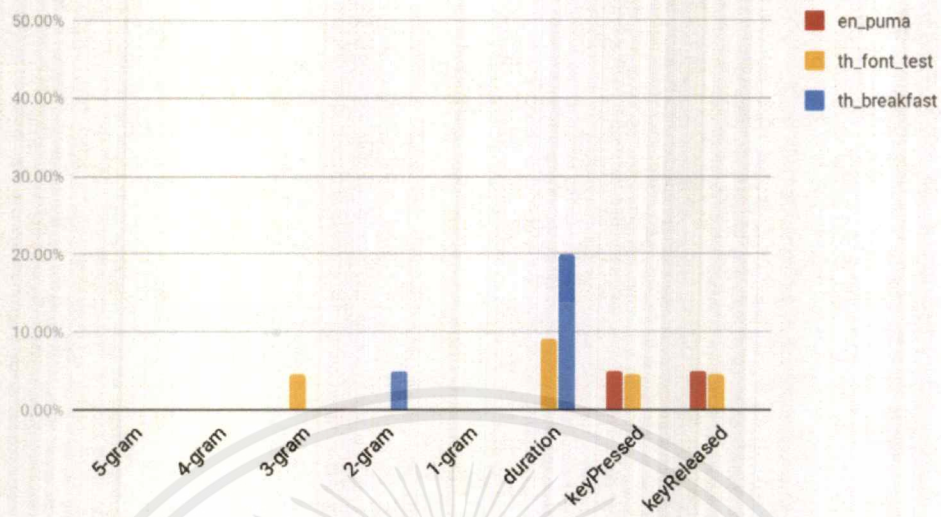
dtw-fourier



ภาพที่ 4.13 กราฟเปรียบเทียบค่าความถูกต้องของ trim data sets โดยใช้วิธี dtw-fourier

จากภาพที่ 4.13 คือ ภาพผลการดำเนินงานของ trim data sets ที่ใช้วิธี dtw-fourier ผลปรากฏว่า 1-gram ให้ผลลัพธ์ดีที่สุดที่ 22% ส่วน 5-gram, 4-gram, 3-gram, duration, keyPressed, keyReleased ให้ผลลัพธ์แย่ที่สุด 5%

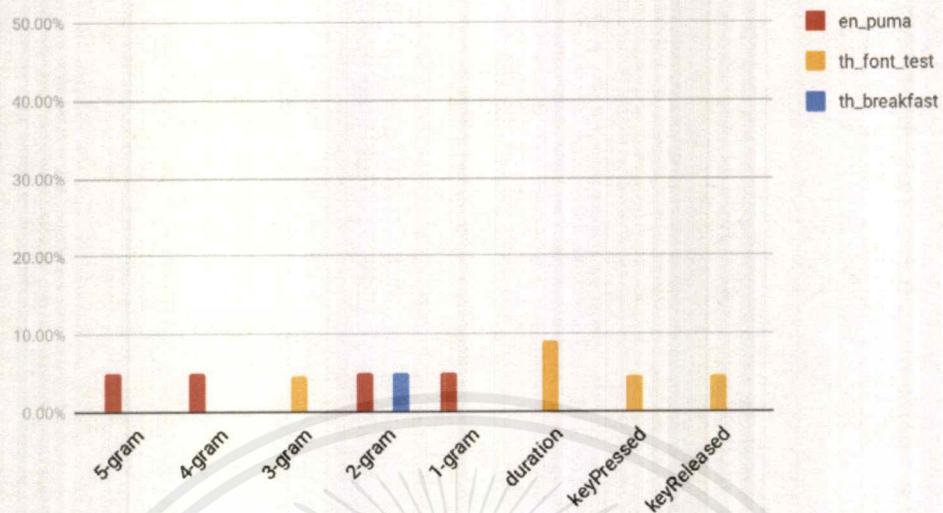
dtw-fourier-mapping



ภาพที่ 4.14 กราฟเปรียบเทียบค่าความถูกต้องของ trim data sets โดยใช้วิธี dtw-fourier-mapping

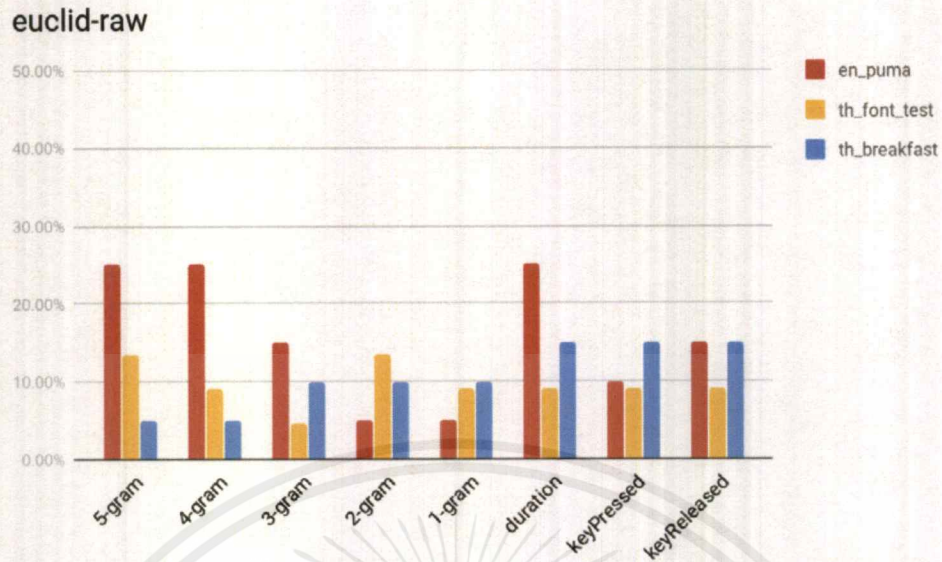
จากภาพที่ 4.14 คือ ภาพผลการดำเนินงานของ trim data sets ที่ใช้วิธี dtw-fourier-mapping ผลปรากฏว่า duration ให้ผลลัพธ์ที่ดีที่สุด 20% ส่วน 5-gram, 4-gram, 1-gram ให้ผลลัพธ์แย่ที่สุด 0%

dtw-mapping-fourier



ภาพที่ 4.15 กราฟเปรียบเทียบค่าความถูกต้องของ trim data sets โดยใช้วิธี dtw-mapping-fourier

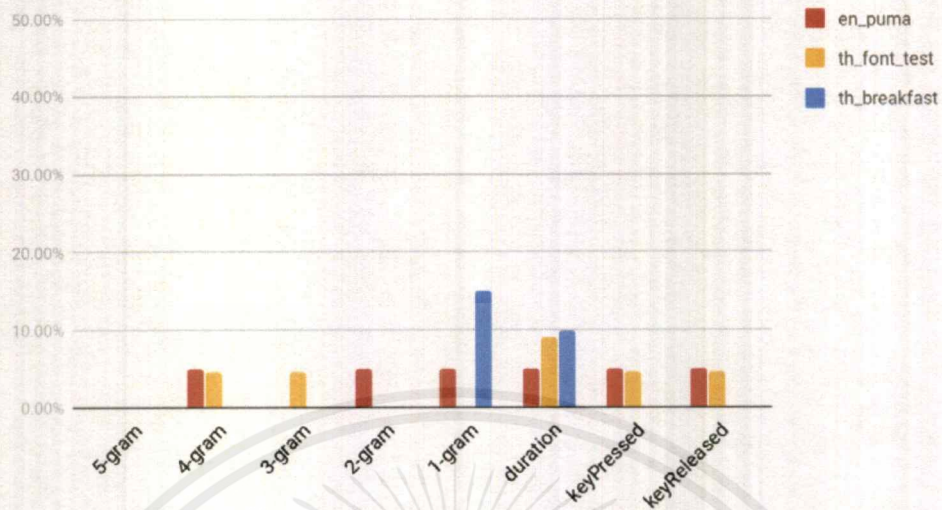
จากภาพที่ 4.15 คือ ภาพผลการดำเนินงานของ trim data sets ที่ใช้วิธี dtw-mapping-fourier ผลปรากฏว่า duration ให้ผลลัพธ์ที่ดีที่สุด 9% ส่วน 5-gram, 4-gram, 3-gram, 2-gram, 1-gram, keyPressed, keyReleased ให้ผลลัพธ์แย่ที่สุด 5%



ภาพที่ 4.16 กราฟเปรียบเทียบค่าความถูกต้องของ trim data sets โดยใช้วิธี euclid-raw

จากภาพที่ 4.16 คือ ภาพผลการดำเนินงานของ trim data sets ที่ใช้วิธี euclid-raw ผลปรากฏว่า 5-gram, 4-gram, duration ให้ผลลัพธ์ที่ดีที่สุด 25% ส่วน 3-gram, 2-gram, 1-gram ให้ผลลัพธ์แย่งที่สุด 5%

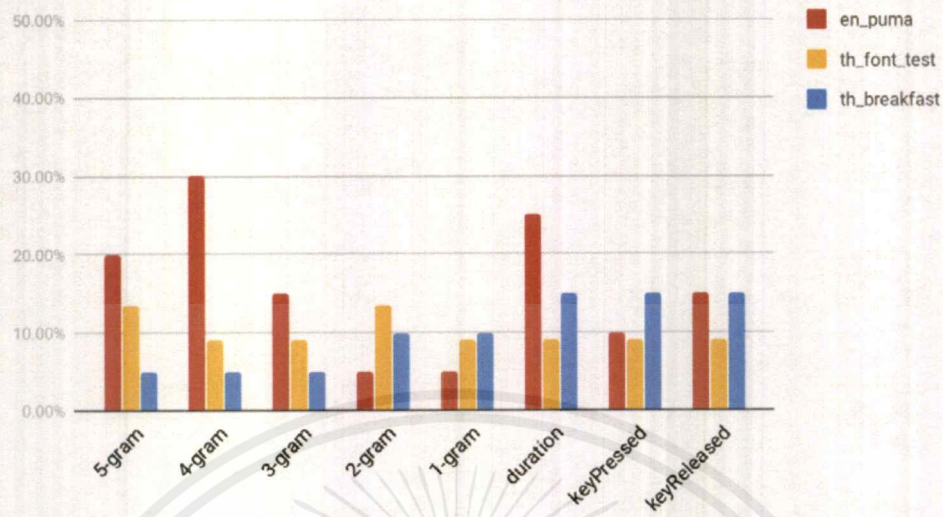
euclid-mapping



ภาพที่ 4.17 กราฟเปรียบเทียบค่าความถูกต้องของ trim data sets โดยใช้วิธี euclid-mapping

จากภาพที่ 4.17 คือ ภาพผลการดำเนินงานของ trim data sets ที่ใช้วิธี euclid-mapping ผลปรากฏว่า 1-gram ให้ผลลัพธ์ที่ดีที่สุด 15% ส่วน 5-gram ให้ผลลัพธ์แย่ที่สุด 0%

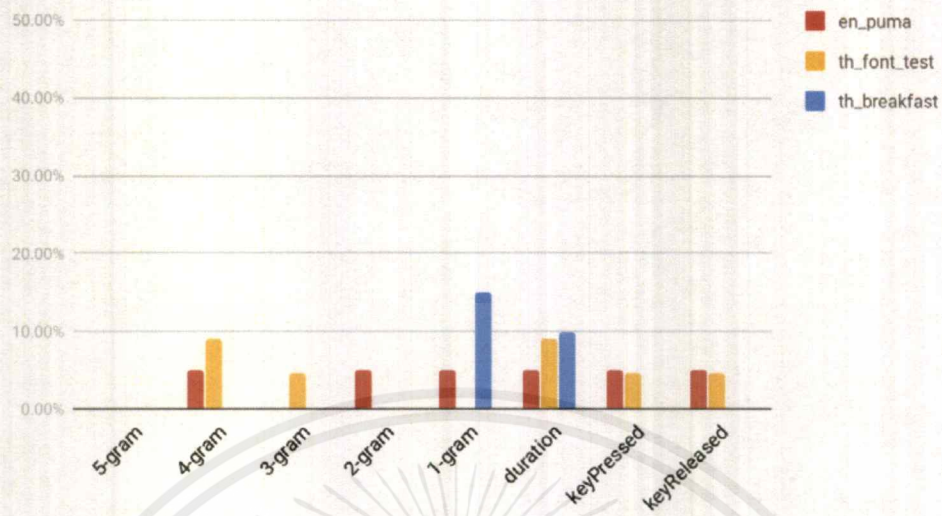
euclid-fourier



ภาพที่ 4.18 กราฟเปรียบเทียบค่าความถูกต้องของ trim data sets โดยใช้วิธี euclid-fourier

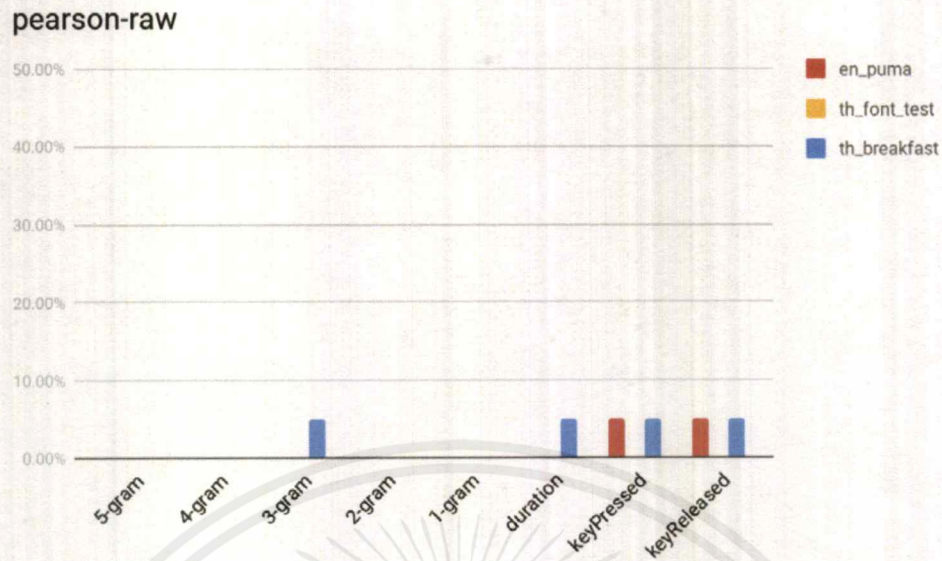
จากภาพที่ 4.18 คือ ภาพผลการดำเนินงานของ trim data sets ที่ใช้วิธี euclid-fourier ผลปรากฏว่า 4-gram ให้ผลลัพธ์ดีที่สุด 30% ส่วน 5-gram, 3-gram, 2-gram, 1-gram ให้ผลลัพธ์แย่ที่สุด 5%

euclid-fourier-mapping



ภาพที่ 4.19 กราฟเปรียบเทียบค่าความถูกต้องของ trim data sets โดยใช้วิธี euclid-fourier-mapping

จากภาพที่ 4.19 คือ ภาพผลการดำเนินงานของ trim data sets ที่ใช้วิธี euclid-fourier-mapping ผลปรากฏว่า 1-gram ให้ผลลัพธ์ดีที่สุดที่ 15% ส่วน 5-gram ให้ผลลัพธ์แย่มากที่สุดที่ 0%

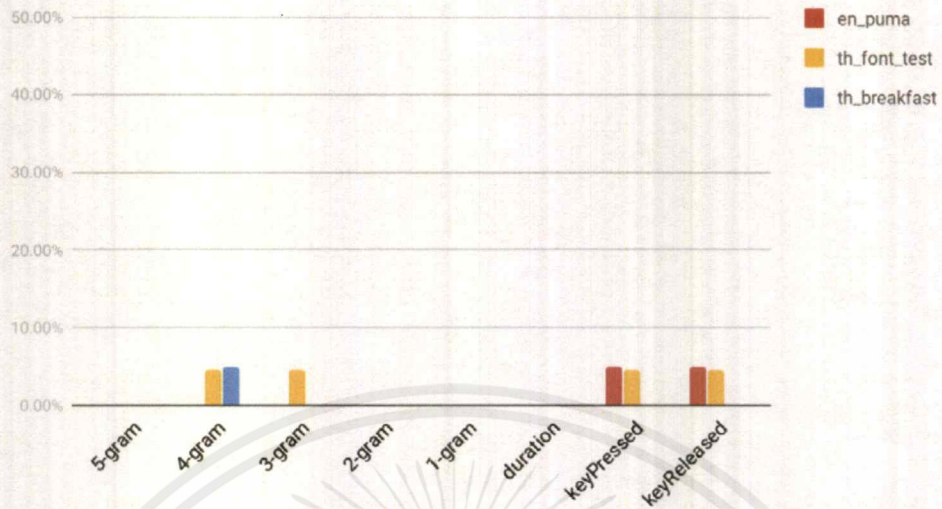


ภาพที่ 4.20 กราฟเปรียบเทียบค่าความถูกต้องของ trim data sets โดยใช้วิธี

pearson-raw

จากภาพที่ 4.20 คือ ภาพผลการดำเนินงานของ trim data sets ที่ใช้วิธี pearson-raw ผลปรากฏว่า 3-gram, duration, keyPressed, keyReleased ให้ผลลัพธ์ที่ดีที่สุด 5% ส่วน 5-gram, 4-gram, 2-gram, 1-gram ให้ผลลัพธ์แย่มากที่สุด 0%

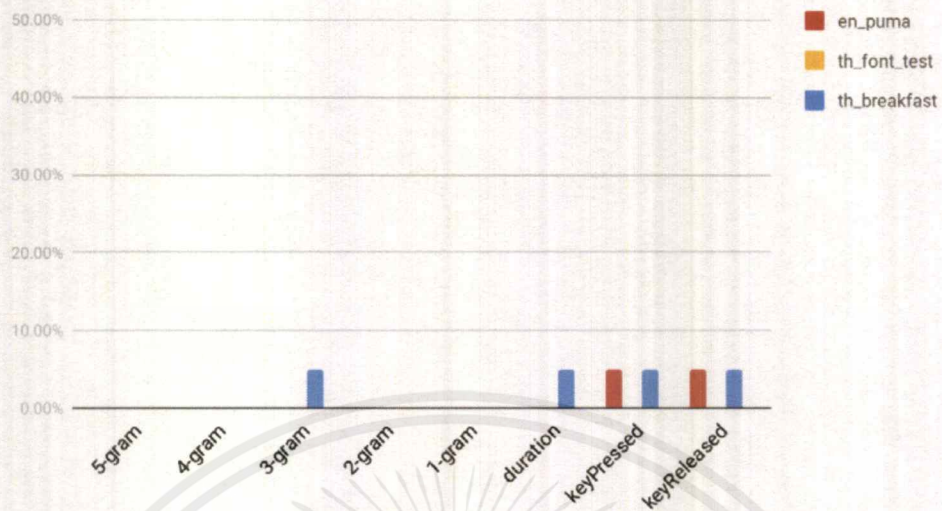
pearson-mapping



ภาพที่ 4.21 กราฟเปรียบเทียบค่าความถูกต้องของ trim data sets โดยใช้วิธี pearson-mapping

จากภาพที่ 4.21 คือ ภาพผลการดำเนินงานของ trim data sets ที่ใช้วิธี pearson-mapping ผลปรากฏว่า 4-gram, 3-gram, keyPressed, keyReleased ให้ผลลัพธ์ดีที่สุดที่ 5% ส่วน 5-gram, 2-gram, 1-gram, duration ให้ผลลัพธ์แย่มากที่สุด 0%

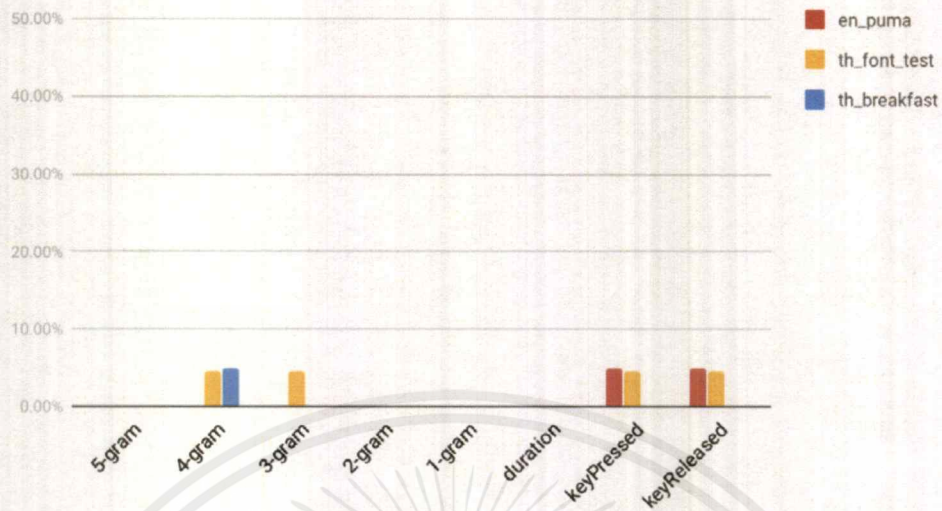
pearson-fourier



ภาพที่ 4.22 กราฟเปรียบเทียบค่าความถูกต้องของ trim data sets โดยใช้วิธี pearson-fourier

จากภาพที่ 4.22 คือ ภาพผลการดำเนินงานของ trim data sets ที่ใช้วิธี pearson-fourier ผลปรากฏว่า 3-gram, duration, keyPressed, keyReleased ให้ผลลัพธ์ที่ดีที่สุด 5% ส่วน 5-gram, 4-gram, 2-gram, 1-gram ให้ผลลัพธ์แย่งที่สุด 0%

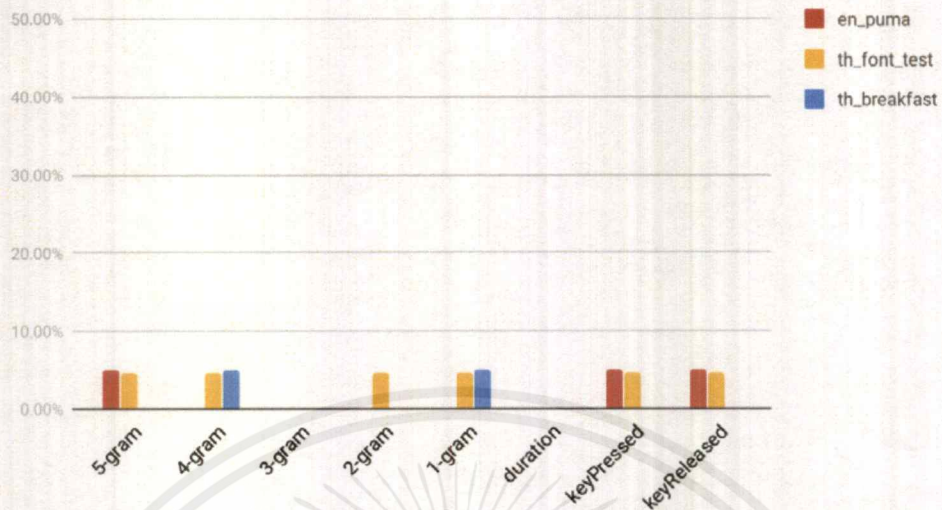
pearson-fourier-mapping



ภาพที่ 4.23 กราฟเปรียบเทียบค่าความถูกต้องของ trim data sets โดยใช้วิธี pearson-fourier-mapping

จากภาพที่ 4.23 คือ ภาพผลการดำเนินงานของ trim data sets ที่ใช้วิธี euclid-fourier-mapping ผลปรากฏว่า 4-gram, 3-gram, keyPressed, keyReleased ให้ผลลัพธ์ที่ดีที่สุด 5% ส่วน 5-gram, 2-gram, 1-gram, duration ให้ผลลัพธ์แย่มากที่สุด 0%

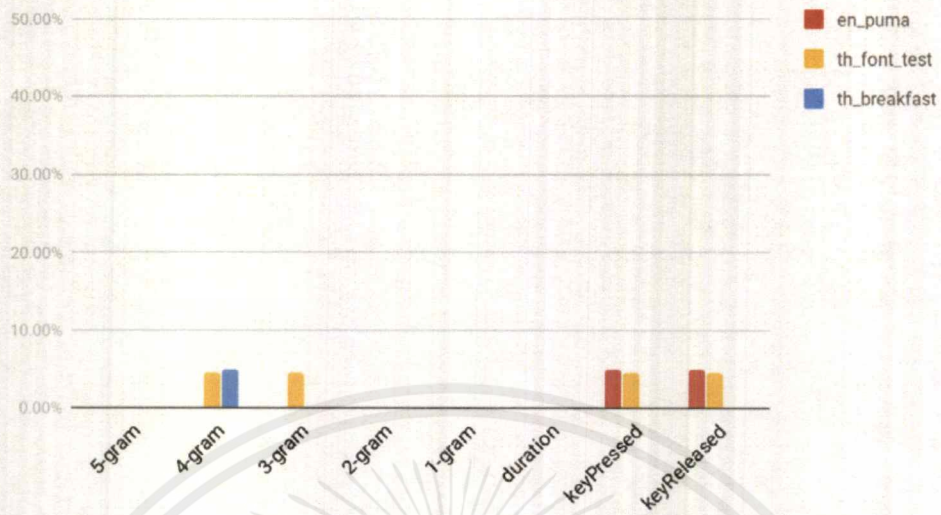
pearson-mapping-fourier



ภาพที่ 4.24 กราฟเปรียบเทียบค่าความถูกต้องของ trim data sets โดยใช้วิธี pearson-mapping-fourier

จากภาพที่ 4.24 คือ ภาพผลการดำเนินงานของ trim data sets ที่ใช้วิธี pearson-mapping-fourier ผลปรากฏว่า 5-gram, 4-gram, 2-gram, 1-gram, keyPressed, keyReleased ให้ผลลัพธ์ดีที่ 5% ส่วน 3-gram, duration ให้ผลลัพธ์แย่ที่สุด 0%

euclid-mapping-fourier



ภาพที่ 4.25 กราฟเปรียบเทียบค่าความถูกต้องของ trim data sets โดยใช้วิธี euclid-mapping-fourier

จากภาพที่ 4.25 คือ ภาพผลการดำเนินงานของ typo data sets ที่ใช้วิธี euclid-mapping-fourier ผลปรากฏว่า 4-gram, 3-gram, keyPressed, keyReleased ให้ผลลัพธ์ที่ดีที่สุด 5% ส่วน 5-gram, 2-gram, 1-gram, duration ให้ผลลัพธ์แย่มากที่สุด 0%

ตารางที่ 4.1 ค่าความถูกต้องโดยเฉลี่ยของข้อมูลทุกชุด

typo_data_sets	average
dtw-raw	14.92%
dtw-mapping	6.27%
dtw-fourier	9.32%
dtw-fourier_mapping	3.58%
dtw-mapping_fourier	4.75%
non-typo_data_sets	average
dtw-raw	13.07%
dtw-mapping	5.66%
dtw-fourier	8.22%
dtw-fourier_mapping	2.18%
dtw-mapping_fourier	3.22%
trim_data_sets	average
dtw-raw	12.67%
dtw-mapping	4.89%
dtw-fourier	7.37%
dtw-fourier_mapping	2.41%
dtw-mapping_fourier	1.99%
euclid-raw	11.97%
euclid-mapping	3.43%
euclid-fourier	11.95%
euclid-fourier_mapping	3.62%
euclid-mapping_fourier	1.38%
pearson-raw	1.25%
pearson-mapping	1.38%
pearson-fourier	1.25%
pearson-fourier_mapping	1.38%
pearson-mapping_fourier	2.18%

จากตารางที่ 4.1 แสดงให้เห็นว่า typo data sets มีค่าความถูกต้องเฉลี่ยสูงสุด 14.92% ข้อมูล non typo data sets มีค่าความถูกต้องเฉลี่ยสูงสุด 12.67% และ trim data sets มีค่าความถูกต้องเฉลี่ยสูงสุด 12.67%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

typo_data_sets

dtw-raw

ตารางที่ 4.2 ค่าความถูกต้องของ typo data sets ด้วยวิธี dtw-raw

	en_puma	th_font_test	th_breakfast	average
5-gram	10.00%	13.64%	15.00%	12.88%
4-gram	25.00%	13.64%	20.00%	19.55%
3-gram	5.00%	18.18%	15.00%	12.73%
2-gram	15.00%	9.09%	25.00%	16.36%
1-gram	25.00%	13.64%	20.00%	19.55%
duration	5.00%	22.73%	20.00%	15.91%
keyPressed	20.00%	13.64%	0.00%	11.21%
keyReleased	20.00%	13.64%	0.00%	11.21%

จากตารางที่ 4.2 แสดงให้เห็นว่าวิธี 4-gram, 1-gram มีค่าความถูกต้องสูงสุดที่ 25% ส่วน keyPressed, keyReleased มีค่าความถูกต้องต่ำสุดที่ 0%

non-typo_data_sets

dtw-raw

ตารางที่ 4.3 ค่าความถูกต้องของ non-typo data sets ด้วยวิธี dtw-raw

	en_puma	th_font_test	th_breakfast	average
5-gram	10.00%	9.09%	10.00%	9.70%
4-gram	10.00%	13.64%	10.00%	11.21%
3-gram	15.00%	18.18%	15.00%	16.06%
2-gram	30.00%	13.64%	5.00%	16.21%
1-gram	25.00%	18.18%	10.00%	17.73%
duration	0.00%	22.73%	10.00%	10.91%
keyPressed	25.00%	9.09%	0.00%	11.36%
keyReleased	25.00%	9.09%	0.00%	11.36%

จากตารางที่ 4.3 แสดงให้เห็นว่าวิธี 2-gram มีค่าความถูกต้องสูงสุดที่ 30% ส่วน duration, keyPressed, keyReleased มีค่าความถูกต้องต่ำสุดที่ 0%

trim_data_sets

dtw-raw

ตารางที่ 4.4 ค่าความถูกต้องของ trim data sets ด้วยวิธี dtw-raw

	en_puma	th_font_test	th_breakfast	average
5-gram	15.00%	9.09%	10.00%	11.36%
4-gram	5.00%	18.18%	10.00%	11.06%
3-gram	15.00%	13.64%	15.00%	14.55%
2-gram	30.00%	9.09%	10.00%	16.36%
1-gram	15.00%	13.64%	5.00%	11.21%
duration	5.00%	22.73%	20.00%	15.91%
keyPressed	15.00%	13.64%	5.00%	11.21%
keyReleased	15.00%	9.09%	5.00%	9.70%

จากตารางที่ 4.4 แสดงให้เห็นว่าวิธี 2-gram มีค่าความถูกต้องสูงสุดที่ 30% ส่วน 4-gram, 1-gram, keyPressed, keyReleased มีค่าความถูกต้องต่ำสุดที่ 5%

euclid-raw

ตารางที่ 4.5 ค่าความถูกต้องของ trim data sets ด้วยวิธี euclid-raw

	en_puma	th_font_test	th_breakfast	average
5-gram	25.00%	13.64%	5.00%	14.55%
4-gram	25.00%	9.09%	5.00%	13.03%
3-gram	15.00%	4.55%	10.00%	9.85%
2-gram	5.00%	13.64%	10.00%	9.55%
1-gram	5.00%	9.09%	10.00%	8.03%
duration	25.00%	9.09%	15.00%	16.36%
keyPressed	10.00%	9.09%	15.00%	11.36%
keyReleased	15.00%	9.09%	15.00%	13.03%

จากตารางที่ 4.5 แสดงให้เห็นว่าวิธี 5-gram, 4-gram, duration มีค่าความถูกต้องสูงสุดที่ 25% ส่วน 3-gram มีค่าความถูกต้องต่ำสุดที่ 4.55%

euclid-fourier

ตารางที่ 4.6 ค่าความถูกต้องของ trim data sets ด้วยวิธี euclid-fourier

	en_puma	th_font_test	th_breakfast	average
5-gram	20.00%	13.64%	5.00%	12.88%
4-gram	30.00%	9.09%	5.00%	14.70%
3-gram	15.00%	9.09%	5.00%	9.70%
2-gram	5.00%	13.64%	10.00%	9.55%
1-gram	5.00%	9.09%	10.00%	8.03%
duration	25.00%	9.09%	15.00%	16.36%
keyPressed	10.00%	9.09%	15.00%	11.36%
keyReleased	15.00%	9.09%	15.00%	13.03%

จากตารางที่ 4.6 แสดงให้เห็นว่าวิธี 4-gram มีค่าความถูกต้องสูงสุดที่ 30% ส่วน 5-gram, 3-gram, 2-gram, 1-gram มีค่าความถูกต้องต่ำสุดที่ 5%

สรุปผลการดำเนินงานและข้อเสนอแนะ

ในบทนี้จะกล่าวถึงการสรุปผลการดำเนินงานและข้อเสนอแนะ ซึ่งการสรุปผลการดำเนินงาน จะถูกนำเสนอในหัวข้อ 5.1 และข้อเสนอแนะจะถูกนำเสนอในหัวข้อ 5.2

5.1 สรุปผลการวิจัย

ปัญหาพิเศษนี้ได้นำเสนอวิธีการระบุตัวบุคคลโดยวิเคราะห์รูปแบบที่ได้จากการพิมพ์ โดยใช้วิธีการไดนามิกไทม์วอร์ปปีงร่วมกับวิธีการปรับแต่งข้อมูลต่าง ๆ ซึ่งผลการทดลองการระบุตัวบุคคล ด้วยวิธีที่นำเสนอกับข้อมูลรูปแบบการพิมพ์ต้นฉบับไม่มีการเปลี่ยนแปลงให้ผลลัพธ์ค่อนข้างดีพอสมควร แตกต่างจากการใช้วิธีการไดนามิกไทม์วอร์ปปีงร่วมกับ mapping และ fourier ให้ผลลัพธ์ไม่ดีเท่าที่ควร ทั้งนี้อาจเป็นเพราะว่าการ mapping และ fourier เป็นการเปลี่ยนแปลงโครงสร้างของข้อมูลก่อนที่จะนำไป normalization ทำให้การทำนายผิดพลาดไปเป็นอย่างมาก

5.2 ข้อเสนอแนะ

จากการดำเนินงานของปัญหาพิเศษนี้ทำให้ทราบทิศทางการวิจัยหรือสามารถเป็นแนวทางในการพัฒนาซอฟต์แวร์ได้ในอนาคต

อาจจะเพิ่มพินิจในการเก็บข้อมูลมากกว่านี้ เนื่องจากยังมีอีกหลายปัจจัยที่อยู่นอกเหนือการควบคุมในการทำการทดลองชุดนี้ ในขั้นตอนการเก็บข้อมูล ยกตัวอย่างปัจจัยที่อยู่เหนือการควบคุมเช่น จำนวนครั้งการพิมพ์ของผู้เข้าร่วมการทดลอง ความตั้งใจในการพิมพ์ของผู้เข้าร่วมการทดลอง และปัจจัยแวดล้อมอื่น ๆ อย่างความถนัดในการใช้แป้นคีย์บอร์ดของผู้เข้าร่วมการทดลอง เป็นต้น เหล่านี้อาจทำให้ผลการทดลองมีความคลาดเคลื่อนไปได้

ในส่วนของการปรับแต่งข้อมูลและการวิเคราะห์ข้อมูลรูปแบบการพิมพ์สามารถใช้เทคนิคอื่น ๆ ซึ่งอาจช่วยเพิ่มประสิทธิภาพให้มีความแม่นยำในการทำนายมากยิ่งขึ้น

บรรณานุกรม

จุฬาลงกรณ์มหาวิทยาลัย. (ม.ป.ป.). “ความสัมพันธ์ระหว่างตัวแปรและการทำนายตัวแปร.”

[Online]. <http://pioneer.netsew.chula.ac.th/~jaimorn/b4st.htm> (3 July 2019).

อัคเดช อุดมชัยพร. 2560. “Dynamic Time Warping” เอกสารประกอบการสอนวิชาการ
ทำเหมืองภาพ, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.

“อาหารเข้ามีความสำคัญอย่างไร.” [Online].

<https://www.trueplookpanya.com/knowledge/content/70734/-blo-scihea-sci->
(5 April 2019).

“ประโยคที่มีอักษรครบทุกตัว ไทย-อังกฤษ.” [Online].

<https://teen.mthai.com/variety/65760.html> (10 March 2019).

“A Puma at Large.” [Online].

<http://fictionchannel.tripod.com/ipx/mis/t-puma.htm> (10 March 2019).

Ying Zhao “Learning User Keystroke Patterns for Authentication”, International
Journal of Computer and Information Engineering Volume 2, No. 2, 2008.

Shallen Giroux, Renata Wachowiak-Smolikova, Member, IEEE, and Mark P.

Wachowiak*, Member, IEEE “Keystroke-Based Authentication by Key Press
Intervals as a Complementary Behavioral Biometric”, 2009 IEEE International
Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009.