

Person Authentication Using Typing Pattern Analysis

การระบุตัวบุคคลโดยการวิเคราะห์รูปแบบที่ได้จากการพิมพ์

สมมติฐาน

เราสามารถระบุตัวบุคคลได้จากรูปแบบการพิมพ์ โดยการเก็บข้อมูลและนำมาวิเคราะห์เพื่อหารูปแบบ

วัตถุประสงค์

1. เพื่อระบุตัวบุคคล จากรูปแบบการพิมพ์
2. เพื่อศึกษาว่า มีวิธีใดบ้างที่จะสามารถนำมาใช้ในการระบุตัวบุคคลได้ และวิธีใดแม่นยำที่สุด

เริ่มทำการเก็บข้อมูล

- เตรียมบทความทั้งภาษาอังกฤษ และภาษาไทย ที่มีความยาวพอประมาณ
- ใช้โปรแกรมเก็บข้อมูลที่เขียนขึ้น
- ต้องการกลุ่มตัวอย่างประมาณ 100 คน

บทความที่ใช้: ภาษาอังกฤษ (en_puma)

Pumas are large, cat-like animals which are found in America. When reports came into London Zoo that a wild puma had been spotted forty-five miles south of London, they were not taken seriously. However, as the evidence began to accumulate, experts from the Zoo felt obliged to investigate, for the descriptions given by people who claimed to have seen the puma were *extraordinarily* similar. The hunt for the puma began in *a* small village where a woman picking strawberries saw 'a large cat' only five yards away from her. It immediately ran away when she saw it, and experts confirmed that a puma will not attack a human being unless it is cornered.

- จากหนังสือแบบเรียนสำหรับผู้ใช้ภาษาอังกฤษเป็นภาษาที่สอง
- มีตัวอักษรทั้งหมด 650 ตัวอักษร (รวมวรรคตอน)
- คำที่สั้นที่สุด มี 1 ตัวอักษร (a)
- คำที่ยาวที่สุด มี 15 ตัวอักษร (extraordinarily)

บทความที่ใช้: ภาษาไทย 1 (th_font_test)

เป็นมนุษย์สุดประเสริฐเลิศคุณค่า
กว่าบรรดาฝูงสัตว์เดรัจฉาน
จงฝ่าฟันพัฒนาวิชาการ
อย่าล้างผลาญตาเข่นฆ่าบีฑาใคร
ไม่ถือโทษโกรธแข่งชดฮึดฮัดด่า
หัดอภัยเหมือนกีฬาอัชฌาสัย
ปฏิบัติประพฤติกฎกำหนดใจ
พูดจาให้จ๊ะ ๆ จ๋า ๆ น่าฟังเอยฯ

- กลอนแปดที่นิยมใช้ทดสอบฟอนต์ (เพราะมีพยัญชนะไทยครบทุกตัว)
- มีตัวอักษรทั้งหมด 218 ตัวอักษร (รวมวรรณยุกต์)

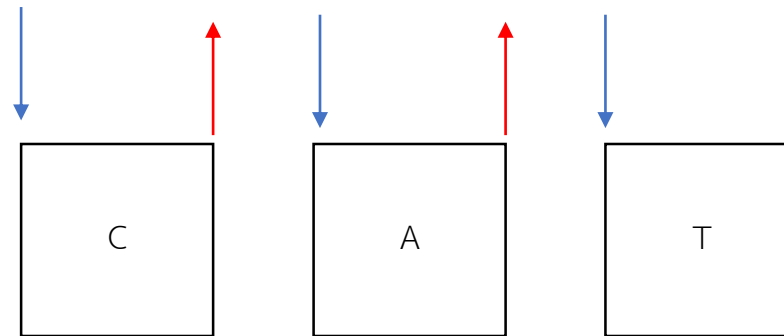
บทความที่ใช้: ภาษาไทย 2 (th_breakfast)

การที่เราไม่รับประทานอาหารเช้าจะทำให้ระบบการทำงานของร่างกายผิดปกติ กล่าวคือ เมื่อเราตื่นขึ้นมา ระดับน้ำตาลในเลือดที่มีหน้าที่กระตุ้นการทำงานของกล้ามเนื้อและสมองจะต่ำ อาหารเช้าจะทำให้ระดับน้ำตาลในเลือดเป็นปกติ เนื่องจากอาหารเช้ามีสารอาหารที่จำเป็นต่อร่างกาย เมื่อเรารับประทานอาหารเช้า เราจะได้รับวิตามินและสารอาหารจากอาหารที่เรารับประทานเข้าไป เช่น นม ธัญพืช หรือผลไม้ แต่ถ้าเราข้ามอาหารเช้าไป ร่างกายจะไม่ได้รับพลังงาน ทำให้ไม่มีพลังงานเพื่อนำไปใช้เลี้ยงส่วนต่าง ๆ ของร่างกาย นอกจากนี้เราจะกินอาหารในมื้อต่อไปมากขึ้น เมื่อเราไม่ได้รับประทานอาหารเช้า ซึ่งนำไปสู่ภาวะน้ำหนักเกินได้

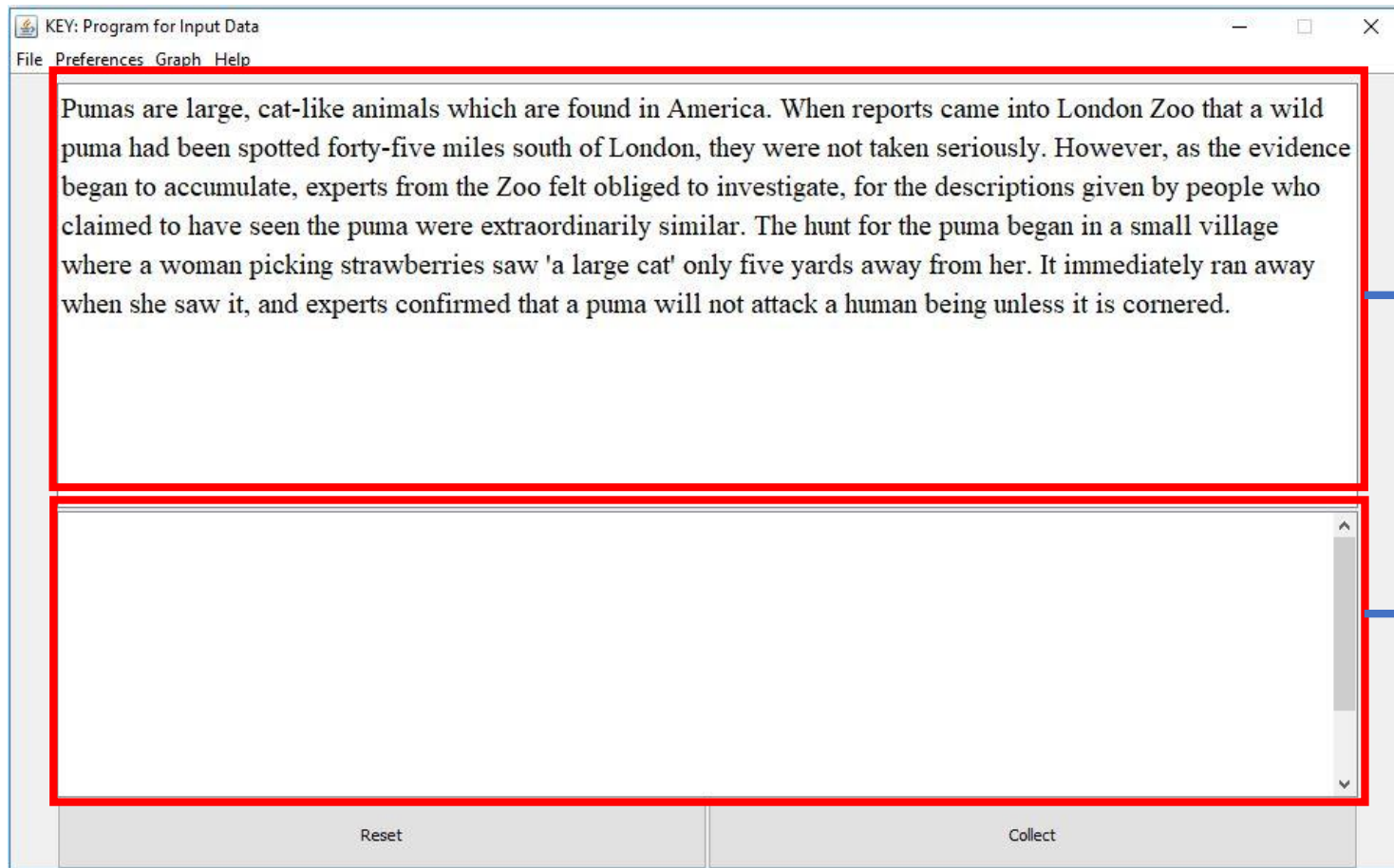
- บทความที่สุ่มเลือกมา
- มีตัวอักษรทั้งหมด 583 ตัวอักษร (รวมวรรคตอน)

ข้อมูลอะไรบ้างที่สามารถเก็บได้จากการพิมพ์?

- ค่าของแป้นที่กดลงไป (Key Code)
- เวลาขณะที่กดแป้นนั้น (Key Pressed)
- เวลาขณะที่ปล่อยแป้นนั้น (Key Released)



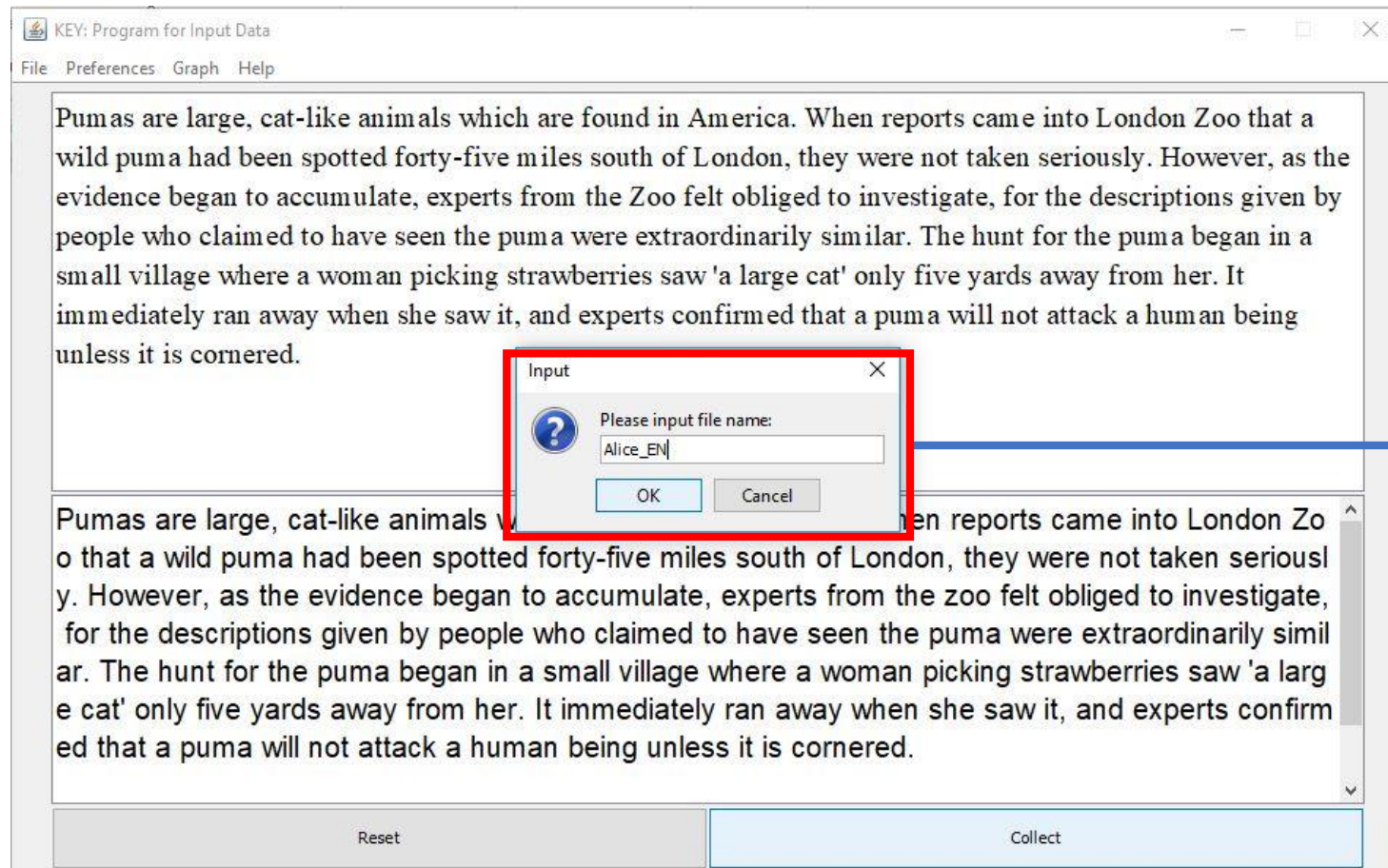
โปรแกรมสำหรับเก็บข้อมูล: หน้าจอหลัก



ส่วนแสดงตัวอย่างบทความ

พื้นที่สำหรับพิมพ์บทความ

โปรแกรมสำหรับเก็บข้อมูล: บันทึกข้อมูล



ตั้งชื่อไฟล์เพื่อบันทึก

ข้อมูลที่เก็บได้

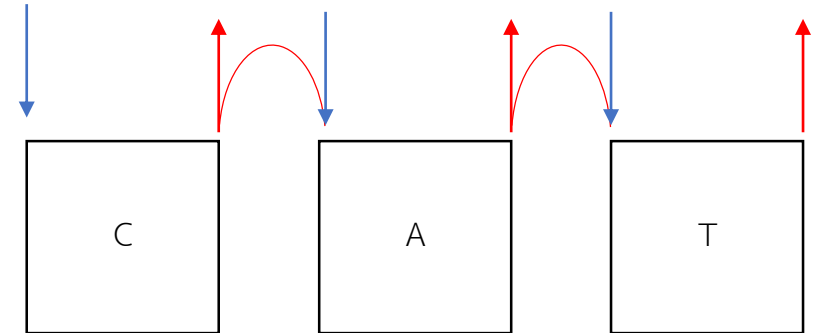
1_THANAPHAT_EN.json	2/2/2562 2:42	JSON File	18 KB
1_THANAPHAT_TH.json	2/2/2562 2:42	JSON File	6 KB
2_NATTAPONG_EN.json	2/2/2562 2:42	JSON File	15 KB
2_NATTAPONG_TH.json	2/2/2562 2:42	JSON File	6 KB
3_WIPADA_EN.json	2/2/2562 2:42	JSON File	15 KB
3_WIPADA_TH.json	2/2/2562 2:42	JSON File	6 KB
4_SURACHAI_EN.json	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_JANJIRA_EN.json	2/2/2562 2:42	JSON File	16 KB
12_JANJIRA_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

- เก็บในรูปแบบ .json ไฟล์
- กำหนดชื่อและลำดับไว้ชัดเจน

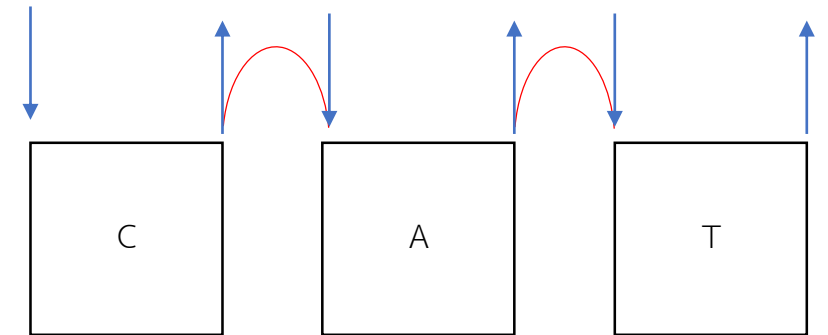
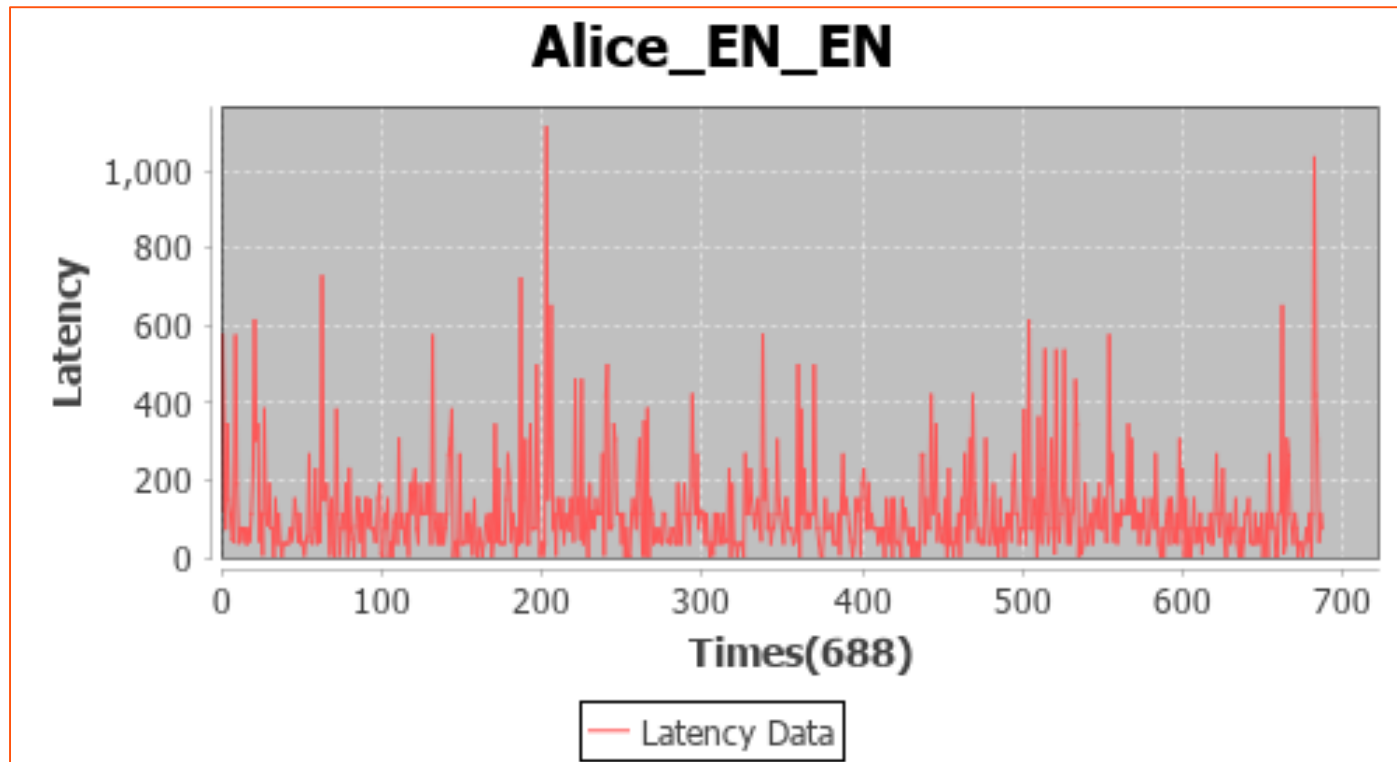
ตัวอย่างโครงสร้างไฟล์

```
{  
  "name" : "TEERAWAT" ,  
  "article" : "th_puma" ,  
  "keyCode" : [16,80,85,...,78] ,  
  "keyPressed" : [12765,12824,13083,...,353067] ,  
  "keyReleased" : [12593,12723,13003,...,349176] ,  
  "1-gram" : [52, 746, 402,...,80]  
  "2-gram" : [786, 839, 855,...,1620]  
  "3-gram" : [1250, 1300, 784,...,2016]  
  "4-gram" : [1687, 1512, 1001,...,1252]  
  "5-gram" : 2401, 1145, 1514,...,1374]  
  "duration" : [129, 92, 62,...,69]  
}
```

หน่วย: มิลลิวินาที



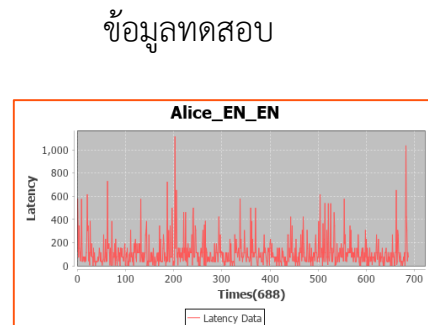
กราฟแท่งตัวบุคคล



ขั้นตอนการทดลอง

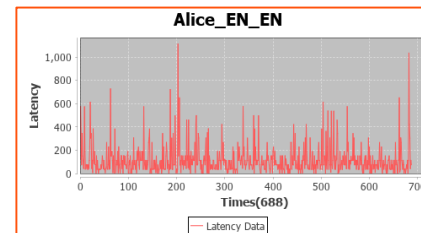
- ใช้วิธีการหาค่าความคล้ายกันของกราฟทั้ง Dynamic Time Warping, Euclidian และ Pearson
- หากกราฟที่มีความคล้ายกับข้อมูลทดสอบมากที่สุด ในการระบุตัวบุคคล
- วัดประสิทธิภาพด้วย accuracy

เปรียบเทียบเพื่อระบุตัวบุคคล

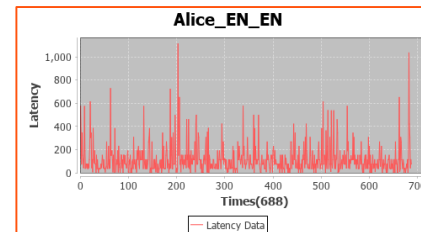


เปรียบเทียบความคล้าย

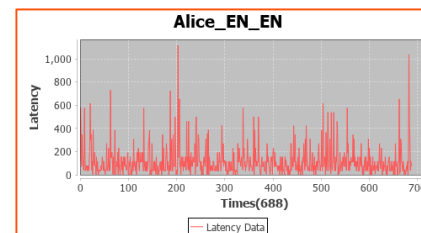
ข้อมูลเรียนรู้



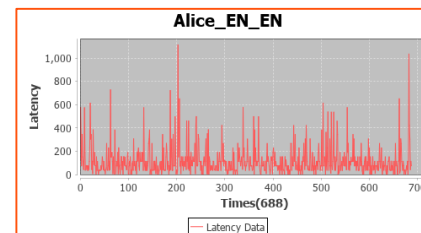
บุคคลที่ 1



บุคคลที่ 2



บุคคลที่ 3



บุคคลที่ 4

ดำเนินการทดลอง

```
D:\Projects\keystroke-biometrics>py experiment.py
-----
dataset: cut_non_typo_data_set\en_puma
attribute: Sgram
experiment type: euclid-non-fourier_non-mapping
training set number: 100
test set number: 20
[('THANAWUT', 9411.815977801522), ('SIRIKANYA', 9598.734239471369), ('JIRAPORN', 9686.623199030713)]
> predicted=('THANAWUT', 9411.815977801522), actual='THANAWUT'
[('NICHAPAT', 8780.912822708127), ('THANAKORN', 9427.57513892093), ('KANNIKA', 9583.182613307543)]
> predicted=('NICHAPAT', 8780.912822708127), actual='THANAKORN'
[('PIMOLPAN', 8082.542916681606), ('BENYAPHA', 8129.066121025219), ('SUTHASINEE', 8169.436883898425)]
> predicted=('PIMOLPAN', 8082.542916681606), actual='PAWINEE'
[('BENYAPHA', 8537.403645137087), ('THANAWUT', 8709.921469221177), ('SIRIKANYA', 9203.59815506957)]
> predicted=('BENYAPHA', 8537.403645137087), actual='SIRIKANYA'
[('THANAWUT', 6159.069329046394), ('JIRAPORN', 7404.606944328645), ('SIRIKANYA', 7491.57460084327)]
> predicted=('THANAWUT', 6159.069329046394), actual='NICHANAN'
[('JIRAPORN', 10666.860503447113), ('PANUPONG', 10989.0975971642), ('NICHAPAT', 11096.26878729963)]
> predicted=('JIRAPORN', 10666.860503447113), actual='JIRAPORN'
[('THANAKORN', 7308.749824696423), ('NICHAPAT', 7420.71714863193), ('METIDA', 7797.272856582614)]
> predicted=('THANAKORN', 7308.749824696423), actual='WORAPOL'
[('TEERAWAT', 7368.905346114849), ('PANUPONG', 8162.355603623258), ('BENYAPHA', 8170.465470216492)]
> predicted=('TEERAWAT', 7368.905346114849), actual='METIDA'
[('SIRIKANYA', 134374.8441486573), ('KEAWALIN', 134794.73244875582), ('SUPUTSORN', 135994.83375060174)]
> predicted=('SIRIKANYA', 134374.8441486573), actual='NUNTIYA'
[('SAKDA', 64236.84437398203), ('SIRIKANYA', 392140.790640863), ('KEAWALIN', 392489.230879539)]
> predicted=('SAKDA', 64236.84437398203), actual='PANUPONG'
[('THANAWUT', 7248.640079352816), ('JIRAPORN', 7653.804152184716), ('BENYAPHA', 7921.966548780675)]
> predicted=('THANAWUT', 7248.640079352816), actual='TEERAWAT'
[('NICHAPAT', 8548.954263534224), ('WORAPOL', 9372.85074030308), ('METIDA', 9780.295394311974)]
> predicted=('NICHAPAT', 8548.954263534224), actual='TIPPAWAN'
[('SIRIKANYA', 104477.90704880223), ('KEAWALIN', 105419.101559553), ('SUPUTSORN', 106634.26139923676)]
> predicted=('SIRIKANYA', 104477.90704880223), actual='BENYAPHA'
[('NICHAPAT', 9895.040575965315), ('PANUPONG', 10301.532264668203), ('THANAKORN', 10545.148220864417)]
> predicted=('NICHAPAT', 9895.040575965315), actual='NICHAPAT'
[('SUTHASINEE', 18537.653357423642), ('ATHIWAT', 19495.77243917255), ('KANNIKA', 19652.51475002627)]
> predicted=('SUTHASINEE', 18537.653357423642), actual='NATTAPOL'
[('NICHAPAT', 8063.539421370742), ('JIRAPORN', 8139.961179268609), ('METIDA', 8224.963161011725)]
> predicted=('NICHAPAT', 8063.539421370742), actual='JULARAT'
[('BENYAPHA', 11799.893389348905), ('NICHANAN', 11883.554392520784), ('SUTHASINEE', 12059.678312459251)]
> predicted=('BENYAPHA', 11799.893389348905), actual='SUKANYA'
[('NICHANAN', 7885.847449703805), ('PANUPONG', 8361.056153381582), ('THANAWUT', 8448.335812454427)]
> predicted=('NICHANAN', 7885.847449703805), actual='WILAWAN'
[('SUTHASINEE', 10503.665074629902), ('THANAWUT', 10573.36313572933), ('BENYAPHA', 10602.8423076079)]
> predicted=('SUTHASINEE', 10503.665074629902), actual='SUTHASINEE'
[('PIMOLPAN', 8740.503303586127), ('TEERAWAT', 8929.882306055326), ('NICHAPAT', 9678.340766887679)]
> predicted=('PIMOLPAN', 8740.503303586127), actual='PIMOLPAN'
Accuracy: 25.0%
Time used: 0.2510061264038086 seconds.
-----
```

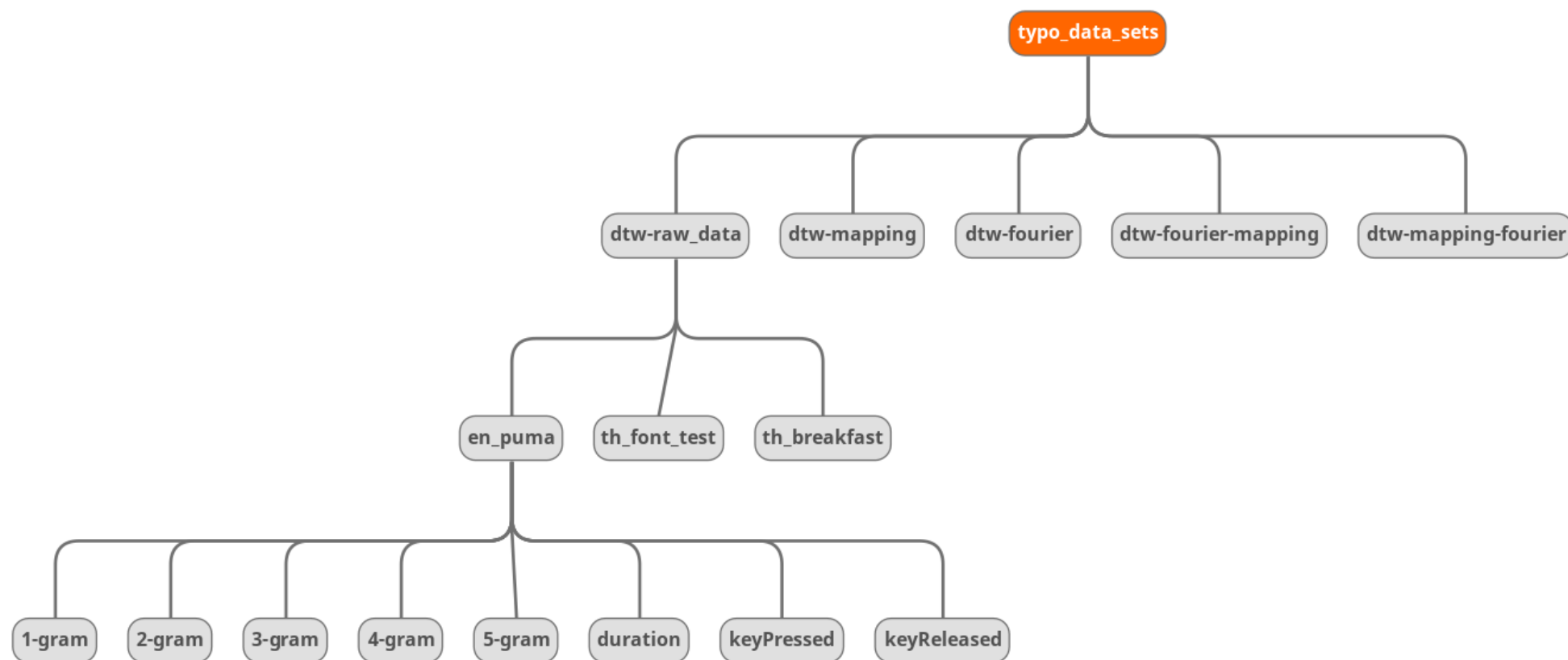
แบ่งชุดข้อมูลออกเป็น 3 ชุด

- typo_data_sets: ชุดข้อมูลดิบ ที่ไม่ได้ทำการเปลี่ยนแปลง มีความยาวไม่เท่ากัน
- non-typo_data_sets: ชุดข้อมูลที่ตัดตัวอักษรที่พิมพ์ผิดออกไป
- trim_data_sets: ชุดข้อมูลที่ถูกปรับแต่งให้เท่ากับจำนวน keyCode ของบทความต้นฉบับ

typo_data_sets

ชุดข้อมูลดิบ ที่ไม่ได้ทำการเปลี่ยนแปลง มีความยาวไม่เท่ากัน

typo_data_sets: แผนการทดลอง



typo_data_sets: ผลการทดลอง

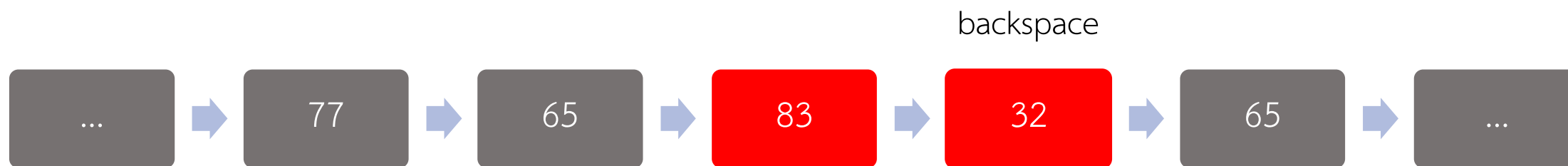
typo_data_sets	average
dtw-raw	14.92%
dtw-mapping	6.27%
dtw-fourier	9.32%
dtw-fourier_mapping	3.58%
dtw-mapping_fourier	4.75%

	en_puma	th_font_test	th_breakfast	average
dtw-raw				
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%

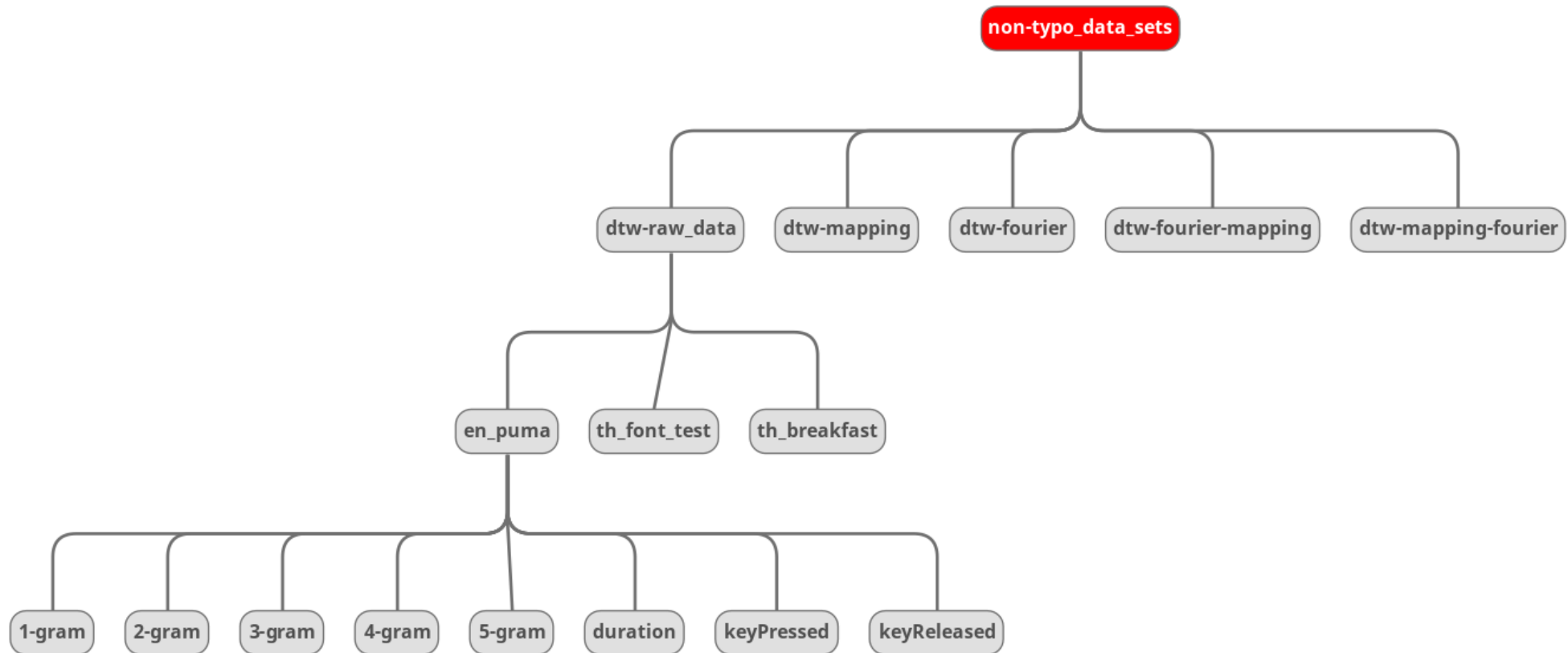
non-typo_data_sets

ชุดข้อมูลที่ตัดตัวอักษรที่พิมพ์ผิดออกไป

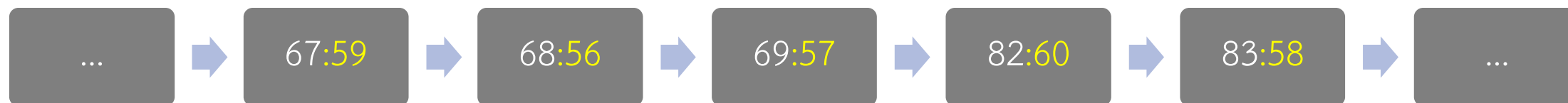
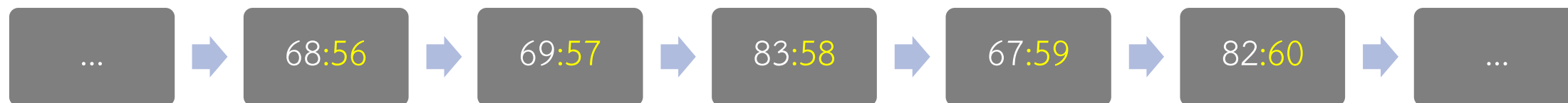
non-typo_data_sets: ตัดตัวอักษรที่พิมพ์ผิดออกไป



non-typo_data_sets: แผนการทดลอง



การ mapping-index



non-typo_data_sets: ผลการทดลอง

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%

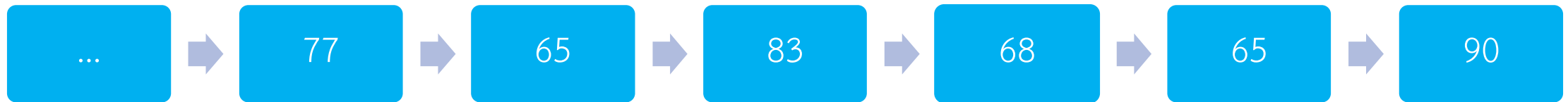
	en_puma	th_font_test	th_breakfast	average
dtw-raw				
5-gram	10.00%	9.09%	10.00%	9.70%
4-gram	10.00%	13.64%	10.00%	11.21%
3-gram	15.00%	18.18%	15.00%	16.06%
2-gram	30.00%	13.64%	5.00%	16.21%
1-gram	25.00%	18.18%	10.00%	17.73%
duration	0.00%	22.73%	10.00%	10.91%
keyPressed	25.00%	9.09%	0.00%	11.36%
keyReleased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

trim_data_sets

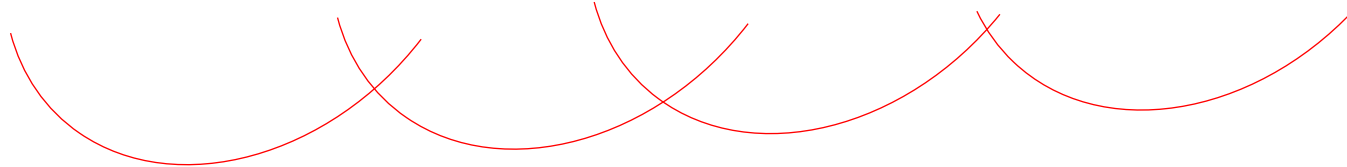
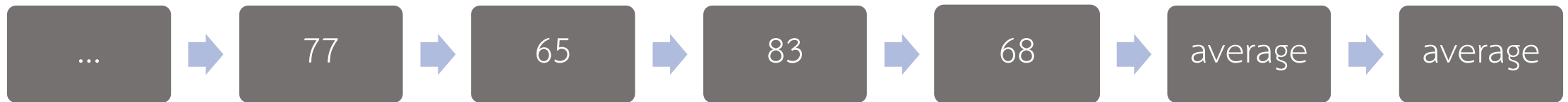
ชุดข้อมูลที่ถูกปรับแต่งให้เท่ากับจำนวน keyCode ของบทความต้นฉบับ

trim_data_sets: ปรับแต่งให้เท่ากับจำนวน keyCode ของบทความต้นฉบับ

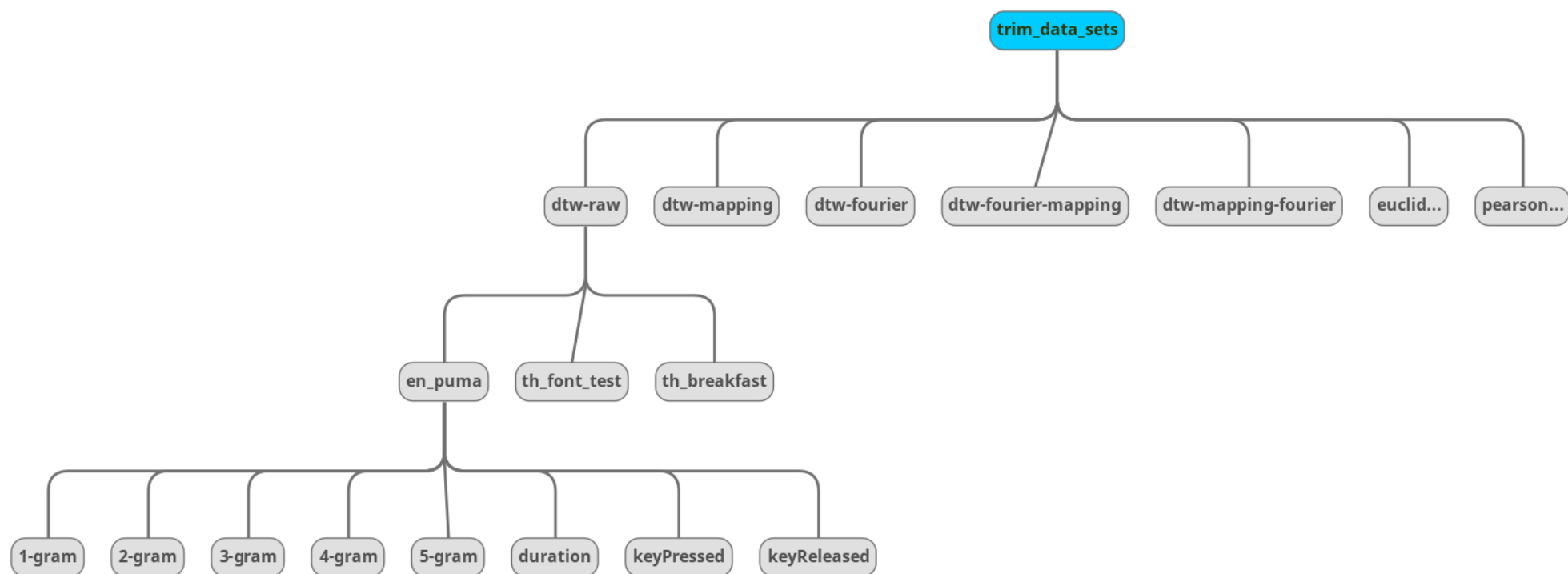
ต้นฉบับ



ข้อมูล



trim_data_sets: แผนการทดลอง



trim_data_sets: ผลการทดลอง

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%

trim_data_sets	average
pearson-raw	1.25%
pearson-mapping	1.38%
pearson-fourier	1.25%
pearson-fourier_mapping	1.38%
pearson-mapping_fourier	2.18%

trim_data_sets: ผลการทดลอง

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%

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%

trim_data_sets: ผลการทดลอง

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%

จบการนำเสนอ

ขอบคุณครับ