การตรวจสอบความถูกต้องของหนังสือเสียงโดยใช้ Machine Learning และ AI

ASSESSING ACCURACY OF AUDIOBOOKS BY USING MACHINE LEARNING AND AI

ภัคพงษ์ ปุรสาชิต

รณกฤต วงษ์อินทร์

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2563

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

ปริญญานิพนธ์ปีการศึกษา 2563

ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เรื่อง การตรวจสอบความถูกต้องของหนังสือเสียงโดยใช้ Machine Learning และ AI

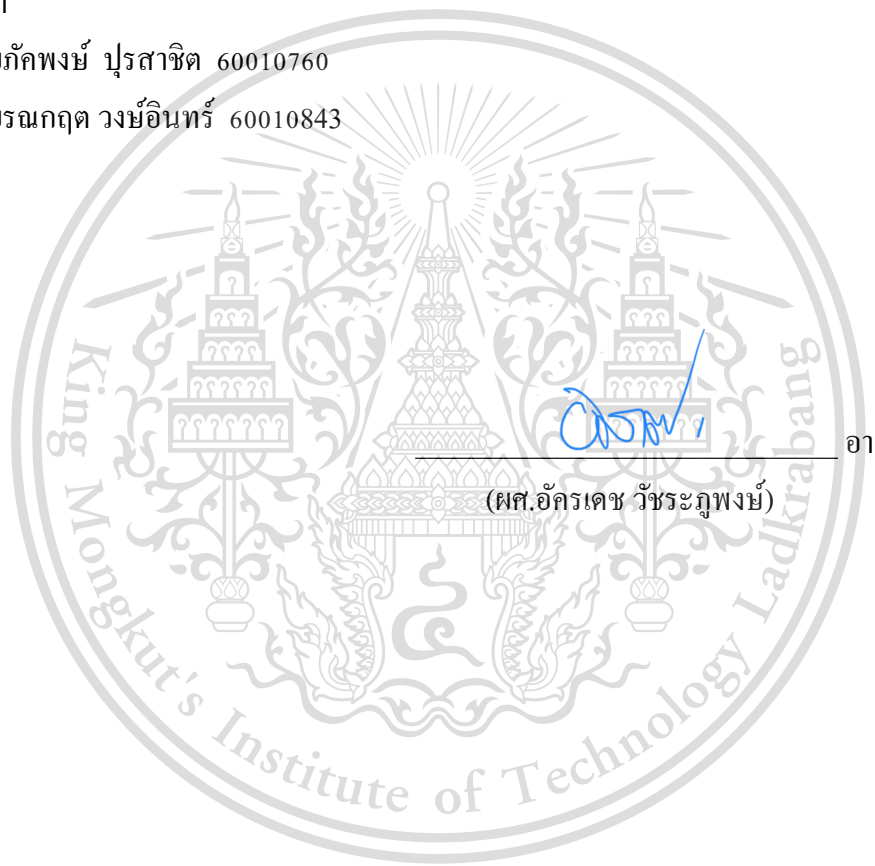
ASSESSING ACCURACY OF AUDIOBOOKS BY USING

MACHINE LEARNING AND AI

ผู้จัดทำ

1. นายภัคพงษ์ ปุรสาชิต 60010760

2. นายรณกฤต วงษ์อินทร์ 60010843



_____ อาจารย์ที่ปรึกษา

(ผศ.อัครเดช วัชระภูพงษ์)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

การตรวจสอบความถูกต้องของหนังสือเสียงโดยใช้ Machine Learning และ AI

นายภัคพงษ์ ปุรสาชิต 60010760
นายรณกฤต วงษ์อินทร์ 60010814
ผศ.อัครเดช วัชระภูพงษ์ อาจารย์ที่ปรึกษา
ปีการศึกษา 2563

บทคัดย่อ

โครงงานฉบับนี้นำเสนอการออกแบบและพัฒนาระบบการตรวจสอบความถูกต้องของหนังสือเสียงโดยใช้ Machine Learning และ AI โดยการศึกษาในด้าน Speech Recognition เพื่อนำมาพัฒนาโปรแกรมโดยตัวซอฟต์แวร์จะใช้ภาษา Python ในการพัฒนา AI และ Machine Learning บน Google Cloud Platform เพื่อให้เกิดประโยชน์แก่เป้าหมายที่เป็นกลุ่มเด็กซึ่งมีความบกพร่องทางสายตา

โครงงานนี้จะเป็นเครื่องมือซึ่งช่วยเสริมสร้างพัฒนาการด้านการฟังและการพูดออกเสียงภาษาไทย โดยมีเกณฑ์การตรวจสอบและวัดผลเพื่อให้ได้ผลลัพธ์ซึ่งสอดคล้องไปกับทฤษฎีและผลงานวิจัยที่เกี่ยวข้องตั้งแต่ด้านการสะกดคำที่ถูกต้อง ความเร็วในการพูด จนไปถึงการวรรคตอนที่เหมาะสมแก่การเรียนรู้สำหรับกลุ่มเป้าหมาย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ASSESSING ACCURACY OF AUDIOBOOKS BY USING MACHINE LEARNING AND AI

Pakkapong	Purasachit	60010760
Ronnakrit	Wongin	60010814
Asst.Prof.Akkradach	Watcharapupong	Advisor

Academic Year 2020

ABSTRACT

This project presents designs and development of Audiobooks accuracy assessing by Machine Learning and AI with research about Speech Recognition for software development. The Software uses Python to develop AI and Machine Learning on Google Cloud Platform for benefits to visually impaired children.

This Project will be a tool to develop listening skills in Thai and also have testing measurement to conform the results to the related theories and researches including spelling correction, speaking speed, and properly length of space to the target group of children.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

สารบัญ

หน้า

บทคัดย่อภาษาไทย I

บทคัดย่อภาษาอังกฤษ II

คำนำ III

สารบัญ IV

สารบัญรูป V

สารบัญตาราง VI

บทที่ 1 1

1.1 ความเป็นมาของปัญหา 1

1.2 วัตถุประสงค์ของการศึกษา 2

1.3 เป้าหมาย 2

1.4 ขอบเขตของโครงงาน 2

1.5 ขั้นตอนการดำเนินงาน 2

บทที่ 2 2

2.1 Speech Recognition 2

2.2 API 5

2.4 RESTful หรือ REST 7

2.5 Google API Speech to Text 8

2.6 ประเภทของไฟล์เสียง 10

2.7 งานวิจัยที่เกี่ยวข้อง (Related Works) 10

บทที่ 3 19

3.1 การจัดหา Dataset สำหรับทำการทดสอบ 19

3.2 การสร้างเกณฑ์สำหรับการตรวจสอบคุณภาพของหนังสือเสียง 19

3.3 ภาพรวมการทำงานของระบบ 19

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

สารบัญ (ต่อ)

หน้า

3.4 การออกแบบการทำงานและโครงสร้างของสถาปัตยกรรม 20

บทที่ 4 21

4.1 ทดสอบ Demo กับไฟล์ตัวอย่างประเภท wav 21

บทที่ 5 25

5.1 ข้อสรุป 25

5.2 ข้อเสนอแนะ 25

5.3 ปัญหาและอุปสรรคที่พบ 25

บรรณานุกรม 26

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

สารบัญรูป

รูป 2.1 แผนภาพแสดงการติดต่อของ API 5

รูป 2.2 รูปภาพแสดงการนำเอา API ไปใช้งาน 6

รูป 2.3 แผนภาพแสดง RESTFul 7

รูป 2.4 กราฟแสดง regression 13

รูป 2.5 รูปแสดงการจับกลุ่มของข้อมูล 14

รูป 2.6 ภาพแสดงการบีบอัดข้อมูล 14

รูป 2.7 ภาพตัวอย่างการเรียนรู้แบบเสริมแรง 16

รูป 2.8 แผนภาพรวมการเรียนรู้แบบเสริมแรง 16

รูป 2.9 ตารางตัวอย่างข้อมูล 18

รูป 2.10 ภาพแสดงกราฟ AUC 28

รูป 3.1 แผนภาพแสดงการทำงานของโปรแกรม 20

รูป 4.1 demo files 21

รูป 4.2 File Selector 22

รูป 4.3 File Selector: ป้อน input ‘hello world’ 22

รูป 4.4 File Selector: ป้อน input เกินลำดับที่กำหนดให้ 22

รูป 4.5 File Selector: ป้อน input 0 22

รูป 4.6 Script Selector: พิมพ์ลำดับเป็น ‘hello world’ 23

รูป 4.7 Script Selector: พิมพ์ลำดับเกินที่กำหนดให้ 23

รูป 4.8 Script Selector: พิมพ์ลำดับที่ 0 23

รูป 4.9 Word-Matching Result 24

รูป 4.10 Report 24

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

สารบัญตาราง

ตาราง 1.1 ตารางแสดงระยะเวลาการดำเนินงาน 3
ตาราง 4.2 เปรียบเทียบความแม่นยำของโมเดลโดยเฉลี่ย 26
ในนิทานเรื่องที่ 1 ถึง 4 โดยใช้ความเร็วในการอ่านแตกต่างกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

บทที่ 1
บทนำ

1.1 ความเป็นมาของปัญหา

ในปัจจุบันเทคโนโลยีแปลงเสียงพูดเป็นข้อความหรือที่เราเรียกกันว่า Speech Recognition นั้น มีบทบาทอย่างมากต่อการดำรงชีวิตประจำวันของเรา ทั้งใน Smart phone ที่นำเทคโนโลยีนี้เข้ามาเพื่ออำนวยความสะดวกแก่ผู้ใช้งาน เช่น การพิมพ์ข้อความด้วยเสียง ตลอดจนการใช้งานระบบสั่งการด้วยเสียง ในสมาร์ทโฟน หรือ สมาร์ทโฮม เป็นต้น นอกจากนั้นยังมีการใช้เทคโนโลยีนี้บนอุปกรณ์อื่น เช่น ระบบคำสั่งเสียงสำหรับหุ่นยนตร์ เป็นต้น

เนื่องจากในปัจจุบันนี้มีการจัดทำหนังสือเสียงสำหรับผู้บกพร่องทางสายตาและผู้สูงอายุหรือคนปกติที่ต้องการรับรู้เนื้อหาของหนังสือผ่านทางการฟังได้มีโอกาสรับรู้ข่าวสารข้อมูลจากสื่อสิ่งพิมพ์เพิ่มขึ้น ทั้งในด้านของการศึกษาหรือเพื่อความบันเทิง แต่หนังสือเสียงนั้นก็อาจจะมีความผิดพลาดอยู่บ้าง ไม่ว่าจะเป็นการออกเสียงหรือการสะกดคำที่ทำให้สารที่ได้รับนั้นมีความหมายผิดไป ทางกลุ่มเราจึงเล็งเห็นถึงความสำคัญในการพัฒนาคุณภาพของหนังสือเสียงให้ดียิ่งขึ้นด้วยการตรวจสอบคุณภาพของหนังสือเสียง โดยการนำเทคโนโลยี Speech Recognition ที่แปลงจากเสียงพูดไปเป็นตัวอักษรและนำเอาตัวอักษรเหล่านั้นไปเปรียบเทียบกับหนังสือต้นฉบับ เพื่อวัดความถูกต้องของหนังสือเสียงว่ามีความถูกต้องมากน้อยเพียงใด เพื่อให้เป็นประโยชน์ต่อทั้งผู้ที่ใช้งานหนังสือเสียงตลอดจนอาสาสมัครที่เป็นผู้จัดทำ

โครงงาน **“การตรวจสอบความถูกต้องของหนังสือเสียง โดยใช้ machine learning และ AI” (Assessing Accuracy of Audiobooks by Using Machine Learning and AI)** นี้จึงถูกจัดทำขึ้น เพื่อทำการศึกษาเทคโนโลยี Speech Recognition โดยการนำเทคโนโลยีนี้มาใช้ในการวัดความถูกต้องของหนังสือเสียง (Audiobooks) เพื่อบอกค่าความถูกต้องของหนังสือเสียงเหล่านั้น โดยผลสำเร็จจะเป็นโปรแกรมระบบวัดความถูกต้องของหนังสือเสียงโดยใช้เทคโนโลยี machine learning และ AI

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

1.2 **วัตถุประสงค์ของการศึกษา**

1.2.1 เพื่อพัฒนาคุณภาพของหนังสือเสียงให้ดียิ่งขึ้น ทั้งในแง่ของความถูกต้อง ความชัดเจน การวรรคตอน ตลอดจนความเร็วที่เหมาะสม

1.2.2 เพื่อให้ผู้ใช้งานโปรแกรมสามารถทราบถึงข้อผิดพลาดและแก้ไขปรับปรุง เพื่อพัฒนาทางด้านการอ่านให้ดีขึ้นได้

1.2.3 มุ้งเน้นเพื่อให้เกิดประโยชน์แก่กลุ่มเป้าหมายที่เป็นเด็กซึ่งมีความบกพร่องทางสายตา เพื่อเป็นหนึ่งในเครื่องมือซึ่งสามารถช่วยเสริมสร้างพัฒนาการทางด้านการฟังและการพูดออกเสียง

1.3 เป้าหมาย

เพื่อวัดความถูกต้องของหนังสือเสียงด้วย machine learning และ AI

1.4 ขอบเขตของโครงงาน

1.4.1 ตัวโปรแกรมรองรับเฉพาะภาษาไทย

1.4.2 ใช้กลุ่มข้อมูลเฉพาะหนังสือนิทานสำหรับเด็ก

1.4.3 ผู้อ่านจะต้องอ่านออกเสียงไม่เร็วเกินไปและชัดถ้อยชัดคำ

1.5 ขั้นตอนการดำเนินงาน

1.5.1 ศึกษาทฤษฎีและผลงานที่เกี่ยวข้อง

1.5.2 พัฒนาส่วนของ Dataset

1.5.3 พัฒนาส่วนของ Word Matching

1.5.4 พัฒนาส่วน Speech Recognition

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตาราง 1.1 ตารางแสดงระยะเวลาการดำเนินงาน

ขั้นตอนการดำเนินงาน	พ.ศ.2563																				พ.ศ.2564			
	สิงหาคม				กันยายน				ตุลาคม				พฤศจิกายน				ธันวาคม				มกราคม			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
ศึกษาทฤษฎีและผลงานทางวิชาการที่เกี่ยวข้อง																								
1.ศึกษาทฤษฎีและผลงานทางวิชาการที่เกี่ยวข้องกับการรับรู้ของเด็กตาบอด																								
2.ศึกษาทฤษฎีและผลงานทางวิชาการที่เกี่ยวกับ Speech Recognition API																								
3.ศึกษาทฤษฎีและผลงานทางวิชาการที่เกี่ยวกับการสอนเด็กปฐมวัย																								
ส่วน Speech Recognition																								
4.ศึกษาการใช้งาน Speech Recognition API																								
5.ศึกษาและพัฒนาส่วน File Handling																								
6.พัฒนาส่วน Audiobooks to text																								
7.พัฒนาส่วนตรวจสอบน้ำเสียงการอ่าน																								
8.ทดสอบและแก้ไขส่วน Speech Recognition																								
ส่วน Dataset																								
9.จัดหาและพัฒนา dataset สำหรับใช้ทดสอบ																								
10.จัดหาและพัฒนา dataset																								
ส่วน Word matching																								
11.ศึกษาเพื่อออกแบบเกณฑ์การวัดค่าความถูกต้องของหนังสือเสียง																								
12.ออกแบบเกณฑ์การวัดค่าความถูกต้องของหนังสือเสียง																								
13.ศึกษาเกี่ยวกับ Machine Learning และ AI																								
14.ศึกษาเกี่ยวกับการเขียน code เพื่อเรียกใช้ API ต่างๆ																								
15.พัฒนาส่วน Word matching																								
16.พัฒนาส่วนตรวจสอบความเร็วในการอ่าน																								
17.พัฒนาส่วนตรวจสอบวรรคตอนในการอ่าน																								
18.ทดสอบและแก้ไขส่วน Word matching																								
รวมทั้ง 2 Module เข้าด้วยกัน																								
19.ออกแบบระบบวัดความถูกต้องของหนังสือเสียง																								
20.พัฒนาระบบวัดความถูกต้องของหนังสือเสียง																								

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

บทที่ 2
ทฤษฎีและหลักการที่เกี่ยวข้อง

2.1 Speech Recognition

กระบวนการ Speech Recognition คือ กระบวนการที่เอาเสียงพูดของมนุษย์นำไปตีความและถอดความให้กลายเป็นตัวอักษร โดยวิธีการแบ่งเสียงพูดของมนุษย์ออกเป็นส่วนๆ และทำการวิเคราะห์เสียงแต่ละส่วน โดยการใช้อัลกอริทึมเพื่อหาคำที่เหมาะสมที่สุดกับภาษานั้นและทำการแปลงภาษาเหล่านั้นเป็นตัวอักษร โดยอาศัยกระบวนการดังนี้

การนำเสียงพูดที่บันทึกลงใน neural network ไปแปลงเสียงพูดที่ถูกบันทึกเป็นข้อความ กระบวนการดังกล่าวมีปัญหาคือ ความเร็วในการออกเสียงพูดของแต่ละคนไม่เท่ากัน ซึ่งทำให้ขนาดของข้อมูลที่ถูกเก็บมีขนาดต่างกัน ซึ่งตามหลักการแล้ว ไฟล์เสียงทั้งสองควรต้องจัดการให้เป็นข้อความเดียวกัน ซึ่งการจัดการไฟล์เสียงทั้ง 2 ที่มีความยาวไม่เท่ากัน โดยอัตโนมัติให้เป็นข้อความที่มีความยาวคงที่นั้นทำได้ค่อนข้างยาก

เพื่อแก้ปัญหาดังกล่าวจำเป็นต้องใช้การ processing เพิ่มเติม นอกเหนือจาก neural network กระบวนการที่แก้ไขปัญหามีขั้นตอนดังนี้

2.1.1 Turning Sounds into Bits นำคลื่นเสียงลงไปในคอมพิวเตอร์เพื่อแปลงคลื่นเสียงให้เป็นตัวเลข โดยทำการ sampling และบันทึกตัวเลขที่แสดงถึงแอมพลิจูดของคลื่นเสียงในเวลานั้น

2.1.2. Pre-processing our Sampled Sound Data ทำการจัดการกับข้อมูลที่ sampling มาและทำการแบ่งข้อมูลออกเป็นช่วงสั้นๆ แต่เนื่องจากข้อมูลเสียงนั้น มีความซับซ้อน เนื่องจากเป็นเสียงพูดของมนุษย์ซึ่งจะมีความหลากหลายของโทนเสียง จึงต้องใช้เทคนิคในการจำแนกเสียงและทำการ mapped ผลลัพธ์ตามความถี่ของเสียง (ต่ำ - สูง) และนำโมเดลที่ได้เข้าสู่ neural network ทำให้ neural network สามารถค้นหารูปแบบของข้อมูลประเภทนี้ได้ง่ายกว่าคลื่นเสียงแบบดังเดิม

2.1.3. Recognizing Characters from Short Sounds นำผลข้อมูลที่ได้จาก neural network ไปป้อนให้กับ recurrent neural network (มีหน่วยความจำที่ทำหน้าที่คาดการณ์) เนื่องจากตัวอักษรแต่ละตัวที่จะคาดการณ์ควรมีผลต่อโอกาสของตัวอักษรถัดไปที่จะเกิดขึ้น เช่นหากพูดว่า "HEL" ก็จะมีโอกาสสูงที่คำถัดไปจะพูดว่า "LO" และมีโอกาสน้อยที่จะพูดสิ่งที่ไม่สามารถออกเสียงได้ถัดไป เช่น "WXY" ดังนั้นการมีหน่วยความจำที่เก็บค่าก่อนหน้าจะช่วยให้ neural network ทำการคาดการณ์ที่แม่นยำขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

2.2 API

Application Program Interface (API) ซึ่งคือ คำสั่งที่อนุญาตให้ software program สามารถสื่อสารระหว่างกันได้ API เป็นช่องทางสำหรับขอใช้บริการคำสั่ง จาก operation system หรือ แอพพลิเคชัน อื่น ซึ่งมันใช้งานโดยติดตั้ง function และเรียกใช้งานตาม document ที่เขียนไว้

รูปที่ 2.1 แผนภาพแสดงการติดต่อของ API

2.2.1 ส่วนประกอบของ API

API สร้างขึ้นจากส่วนสำคัญ 2 อย่าง คือ

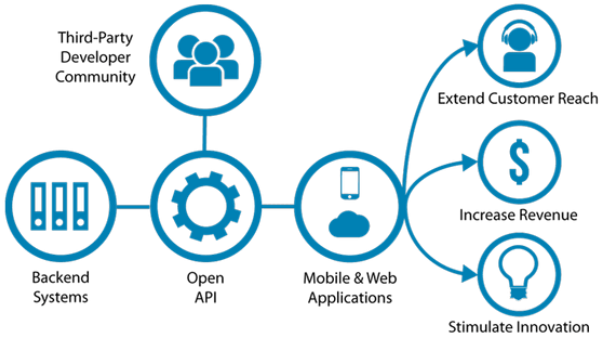
2.2.1.1 ข้อกำหนดที่จะอธิบายการแลกเปลี่ยนข้อมูลระหว่าง program ซึ่งทำออกมาในลักษณะ document เพื่อบอกว่า request/response ต้องเป็นอย่างไร Software ที่เขียนขึ้นตามข้อกำหนด และทำการเผยแพร่ออกไปให้ใช้งาน

2.2.1.2 โดยปกติแล้ว แอปพลิเคชันที่มี API จะต้องถูกเขียนเป็นภาษา programming และ พัฒนาเพิ่มได้ง่าย จึงจำเป็นต้องมีการตรวจสอบโครงสร้าง API เพราะฉะนั้น API ที่ดี ผู้ที่ออกแบบต้องให้ความสำคัญในการทดสอบเพื่อตรวจสอบ logic ที่สามารถเกิดขึ้นได้จากการใช้งาน

2.2.2 การใช้งาน API

ปัจจุบันนี้ API ถูกใช้งานใน แอปพลิเคชัน เพื่อสื่อสารกับ user โดยไม่จำเป็นต้องมีความรู้ บริษัทใหญ่ๆหลายบริษัทมีการเปิด API ให้ภายนอกเข้ามาใช้งาน เช่น facebook, google, twitter ผู้พัฒนาระบบที่สนใจ สามารถนำเอา API เหล่านี้ไปไปต่อยอด ซึ่งทางบริษัทก็สามารถขยายฐานลูกค้าออกไปได้อีก รูปแบบการนำเอา API ไปใช้งานมีดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.



รูปที่ 2.2 รูปภาพแสดงการนำเอา API ไปใช้งาน

2.2.2.1 Libraries and frameworks

API มักจะเอาไปใช้เป็น software library ซึ่งเขียนขึ้นตาม document ในรูปบบภาษา program ที่ต่างกันออกไป ตามความเหมาะสมกับงาน เพื่อเอาไปทำเป็น framework ให้กับระบบใช้ในการสื่อสารหากัน

2.2.2.2 Operating Systems

API สามารถใช้งานในการสื่อสารระหว่าง แอปพลิเคชัน และ operating system เช่น POSIX หรือ มาตรฐานการสื่อสารของ OS เองก็มี API เป็น command line เพื่อควบคุมการทำงานของ OS

2.2.2.3 Remote APIs

Remote APIs ทำไว้ให้ developer สามารถเข้าควบคุมทรัพยากรผ่านทาง protocol เพื่อให้มีมาตราฐานการสื่อสารเดียวกัน ถึงแม้ว่าจะเป็นคนละ technology เช่น Database API สามารถอนุญาตให้ developer เข้ามาดึงข้อมูลใน database หลากหลายชนิดได้ ผ่าน function เดียวกัน เพราะฉะนั้น remote API จึงถูกใช้บ่อยในงาน maintenance ด้วยทำทำงานที่ฝั่ง client ให้ไปดึงข้อมูลจาก server กลับลงมาทำงาน

2.2.2.4. Web APIs

นิยมใช้กันมากในปัจจุบัน เพราะอยู่ในกลุ่มของ HTTP และขยายออกไปสู่

รูปแบบ XML และ JSON ซึ่งโดนรวมแล้วก็คืออยู่บน web service เช่น

SOAP (Simple Object Access Protocol) ใช้ XML format ส่งข้อมูล

REST (Representational State Transfer) สามารถใช้ XML หรือ JSON format ส่งข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

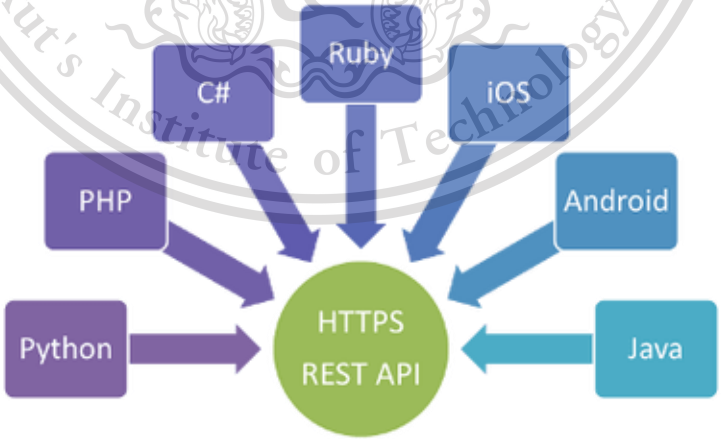
2.3 gRPC

RPC (gRPC Remote Procedure Calls) เป็นระบบการเรียกใช้ระยะไกลแบบ open source (RPC) ซึ่งพัฒนาครั้งแรกที่ Google ในปี 2558 ใช้ HTTP / 2 สำหรับส่ง Protocol Buffers เป็นภาษาของ Interface และมีความสามารถในการตรวจสอบสิทธิ์สตรีมมิงแบบสองทิศทาง การควบคุมโฟลว์การบล็อก การยกเลิก การกำหนดระยะเวลาสิ้นสุดในกระบวนการ และสร้างการเชื่อมโยงไคลเอนต์ข้ามแพลตฟอร์มและเซิร์ฟเวอร์สำหรับหลายภาษา

2.4 RESTful หรือ REST

Representational state transfer หรือ REST คือ การสร้าง Webservice ชนิดหนึ่งที่ใช้สื่อสารกันบน Internet ใช้หลักการแบบ stateless คือไม่มี session ซึ่งต่างจาก webservice แบบอื่น เช่น WSDL และ SOAP การทำงานของ RESTful Webservice จะอาศัย URI/URL ของ request เพื่อค้นหาและประมวลผลแล้วตอบกลับไปในรูป XML, HTML, JSON โดย response ที่ตอบกลับจะเป็นการยืนยันผลของคำสั่งที่ส่งมา และสามารถพัฒนาด้วยภาษา programming ได้หลากหลาย คำสั่งก็จะมีตาม HTTP verbs ซึ่งก็คือ

- GET ทำกการดึงข้อมูลภายใน URI ที่กำหนด
- POST สำหรับสร้างข้อมูล
- PUT ใช้แก้ไขข้อมูล
- DELETE สำหรับลบข้อมูล



รูปที่ 2.3 แผนภาพแสดง RESTFul

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

2.4.1 ประวัติของ REST

REST ถูกตั้งขึ้นโดย Roy Fielding ในปี 2000 ที่ University of California, Irvine ซึ่งได้ทำการพัฒนาขึ้นมาควบคู่กับ HTTP1.1 และ Uniform Resource Identifiers (URI)

2.4.2 คุณสมบัติของ REST

เป็น API อย่างหนึ่ง ซึ่งทุกๆ system ต่างใช้ resource ซึ่งเป็นได้ทั้ง image, video, web page หรือข้อมูลทางธรุกิจ ก็ได้ที่สามารถแสดงบนระบบ computer วัตถุประสงค์เพื่อให้ user สามารถเข้าถึง, ติดตั้ง, ปรับแต่ง, ขยาย resource เหล่านี้ได้ง่าย ซึ่งทาง RESTful ได้ออกแบบมาให้มีคุณสมบัติต่อไปนี้

- แสดงผล
- เก็บข้อมูล
- มี URIs
- Stateless ทำงานโดยไม่ต้องมี session
- เชื่อมต่อระหว่าง web service
- Caching

2.5 Google API Speech to Text

Speech to Text มี 3 วิธีหลักในการจดจำเสียงพูด คือ

2.5.1 Speech requests

1) Synchronous Recognition (REST และ gRPC) ส่งข้อมูลเสียงไปยัง Speech-to-Text API ทำการจดจำข้อมูลนั้นและส่งกลับผลลัพธ์หลังจากประมวลผลเสียงทั้งหมดแล้ว สามารถใช้ได้กับข้อมูลเสียงที่มีความยาว 1 นาทีหรือน้อยกว่า

2) Asynchronous Recognition (REST และ gRPC) ส่งข้อมูลเสียงไปยัง Speech-to-Text API และเริ่มการทำงานระยะยาว เมื่อใช้การดำเนินการนี้สามารถดูผลการจดจำได้เป็นระยะ ๆ โดยใช้ request แบบ Asynchronous สำหรับข้อมูลเสียงในช่วงเวลาใดก็ได้ถึง 480 นาที

3) Streaming Recognition (gRPC เท่านั้น) ทำการจดจำข้อมูลเสียงที่ให้ไว้ภายในสตรีมแบบสองทิศทางของ gRPC request สตรีมมิ่งได้รับการออกแบบมาเพื่อวัตถุประสงค์ในการจดจำแบบเรียลไทม์เช่นการบันทึกเสียงสดจากไมโครโฟน การจดจำการสตรีมให้ผลลัพธ์ระหว่างกาลในขณะที่กำลังบันทึกเสียงทำให้ผลลัพธ์ปรากฏขึ้นเช่นในขณะที่ผู้ใช้ยังคงพูดอยู่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

request มีพารามิเตอร์การกำหนดค่าเช่นเดียวกับข้อมูลเสียง ส่วนต่อไปนี้จะอธิบายประเภทของ request การรับรู้คำตอบที่สร้างขึ้นและวิธีจัดการกับคำตอบเหล่านั้นโดยละเอียด

Speech-to-Text API recognition

request recognition Synchronous Speech-to-Text API เป็นวิธีที่ง่ายที่สุดในการดำเนินการจดจำข้อมูลเสียงพูด Speech-to-Text สามารถประมวลผลข้อมูลเสียงพูดได้ถึง 1 นาทีที่ส่งในrequest แบบ Synchronous หลังจากประมวลผล Speech-to-Text และจดจำเสียงทั้งหมดแล้วระบบจะส่งกลับการตอบสนอง

request recognition แบบ Synchronous ถูกบล็อกหมายความว่า Speech-to-Text ต้องส่งคืนการตอบกลับก่อนที่จะประมวลผล request ถัดไป โดยทั่วไป Speech-to-Text จะประมวลผลเสียงได้เร็วกว่าแบบเรียลไทม์โดยจะประมวลผลเสียง 30 วินาทีใน 15 วินาทีโดยเฉลี่ย ในกรณีที่คุณภาพเสียงไม่ดี request อาจใช้เวลานานกว่ามาก

Speech-to-Text มีทั้งวิธี REST และ gRPC สำหรับการเรียกใช้ Speech-to-Text API แบบ Synchronous และ Asynchronous บทความนี้สาธิต REST API เนื่องจากแสดงและอธิบายการใช้งานพื้นฐานของ API ได้ง่ายกว่า

request API คำพูดเป็นข้อความแบบ Synchronous ประกอบด้วยการกำหนดค่าการรู้จำเสียงและข้อมูลเสียง request ตัวอย่างแสดงอยู่ด้านล่าง:

```
{
  "config": {
    "encoding": "LINEAR16",
    "sampleRateHertz": 16000,
    "languageCode": "en-US",
  },
  "audio": {
    "uri": "gs://bucket-name/path_to_audio_file"
  }
}
```

request การรู้จำแบบ Synchronous Speech-to-Text API ทั้งหมดต้องมีฟิลด์กำหนดค่าการรู้จำเสียง (ประเภท RecognitionConfig) RecognitionConfig มีฟิลด์ย่อยต่อไปนี้:

- encoding (จำเป็น) ระบุรูปแบบการเข้ารหัสของเสียงที่ให้มา (ประเภท AudioEncoding) หากมีทางเลือกในตัวแปลงสัญญาณให้เลือกเข้ารหัสแบบไม่สูญเสีย เช่น FLAC หรือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

LINEAR16 เพื่อประสิทธิรูปที่ดีที่สุด ฟิลด์การเข้ารหัสเป็นทางเลือกสำหรับไฟล์ FLAC และ WAV ที่การเข้ารหัสรวมอยู่ในส่วน file header

- sampleRateHertz (จำเป็น) ระบุ samplerate (Hz) ของเสียงที่ให้มา ฟิลด์ sampleRateHertz เป็นทางเลือกสำหรับไฟล์ FLAC และ WAV ซึ่ง samplerate จะรวมอยู่ในส่วน file header
- languageCode (จำเป็น) มีภาษา + ภูมิภาค / สถานที่ที่จะใช้สำหรับการรู้จำเสียงของเสียงที่ให้มา รหัสภาษาต้องเป็นตัวระบุ BCP-47
- maxAlternatives (ไม่บังคับค่าเริ่มต้นคือ 1) ระบุจำนวนการถอดเสียงทางเลือกที่จะให้ในการตอบกลับ ตามค่าเริ่มต้น Speech-to-Text API จะมีการถอดเสียงหลักหนึ่งรายการ

2.6 ประเภทของไฟล์เสียง

โดยทั่วไปแล้วประเภทของไฟล์เสียงแบ่งออกเป็น 2 กลุ่มใหญ่ๆ ได้แก่ LOSSY Audio และ LOSSLESS Audio

2.6.1 LOSSY Audio

Lossy Audio คือไฟล์ที่มีการบีบอัดแล้วเสียข้อมูลไปบ้าง ยิ่งบีบอัดให้จำนวน Bitrate น้อยลงเท่าไหร่ คุณภาพเสียงก็จะลดทอนลงไปเท่านั้น เสียงเพลงจากที่ได้ยินรายละเอียดครบถ้วนก็อาจจะลดน้อยถอยลงไปตามลำดับครับ กลุ่มนี้จะค่อนข้างพบเห็นได้ทั่วไป ได้แก่ ไฟล์ประเภท MP3, AAC, OGG, WMA เป็นต้น

2.6.2 LOSSLESS Audio

สำหรับ Lossless Audio แปลตามตรงก็หมายถึง ไฟล์ที่ปราศจากการสูญเสียนั่นเอง ซึ่งจุดเด่นของไฟล์ประเภทนี้อยู่ที่คุณภาพของเสียงที่ค่อนข้างทัดเทียมกับแผ่น CD เลยก็ว่าได้ จะสังเกตได้ว่าขนาดของไฟล์ค่อนข้างใหญ่ กลุ่มนี้จะตามหากันยากหน่อย แต่หากพบไฟล์นามสกุลจำพวกนี้รับรองได้ในระดับหนึ่งเลยว่าเป็น Lossless แน่นอน ได้แก่ FLAC, M4A, Wav, Cda, THM, Aiff, APE, TTA, WavPack เป็นต้น

2.7 งานวิจัยที่เกี่ยวข้อง (Related Works)

2.7.1 A Study on Automatic Speech Recognition

งานวิจัยเกี่ยวกับการอธิบายการสื่อสารกันระหว่างคอมพิวเตอร์กับมนุษย์ที่เรียกว่า Human Computer Interaction (HCI) โดยการใช้เครื่องมือและภาพรวมของคำจำกัดความหลักของ Automatic Speech Recognition (ASR) ซึ่งเป็นส่วนหลักที่สำคัญของ Artificial Intelligence (AI) และงานวิจัยนี้ยังให้บทสรุปของงานวิจัยที่สำคัญที่เกี่ยวข้องกับ speech processing

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งไม่อนุญาตให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

2.7.2 Comparing Speech Recognition Systems (Microsoft API, Google API And CMU Sphinx)

งานวิจัยเกี่ยวกับการออกแบบเครื่องมือที่จะทำการทดสอบและเปรียบเทียบบริการ speech recognition ของแต่ละบริการ เช่น Microsoft Speech API, Google Speech API และ open-source speech recognition เช่น Sphinx-4 งานวิจัยค้นพบวิธีการที่ดีที่สุดในการเปรียบเทียบ automatic speech recognition systems ในแต่ละบริการ โดยการใช้เสียงที่บันทึก โดยเลือกจากแหล่งที่มาที่แตกต่างกันและทำการคำนวณ word error (WER) ถึงแม้ว่าค่า WER ทั้ง 3 บริการดังกล่าวนั้นจะเป็นที่ยอมรับ แต่พบว่า Google API นั้นเป็นบริการที่ดีที่สุด

2.7.3 Communication Skills for Children Who Are Blind or Visually Impaired

เป็นบทความเกี่ยวกับทักษะในการสื่อสารและการพัฒนาทักษะในการสื่อสารของเด็กตาบอดหรือเด็กที่มีความพิการทางสายตาโดยจะกล่าวถึงการสื่อสารในแบบต่างๆได้แก่ การรับรู้ การแสดงออก การสื่อสาร และการพูด ซึ่งทักษะเหล่านี้เป็นทักษะที่เด็กทุกคนจะมีเหมือนกันไม่ว่าจะตาบอด หูหนวกหรือมีความพิการทางด้านต่างๆอย่างใด เด็กบางกลุ่มอาจจะมีการพัฒนาทักษะบางอย่างช้ากว่ากลุ่มอื่นอย่างเช่น การรับรู้ เด็กที่พิการทางสายตาจะไม่สามารถหรือมีความสามารถรับรู้ด้วยการมองเห็นได้น้อยกว่าเด็กที่ตาปกติ จึงจะมีการพัฒนาทักษะการรับรู้ด้วยการฟังเสียงได้มากขึ้นและรองลงมาด้วยการสัมผัส เป็นต้น

2.7.4 Repeated Interactive Read Alouds in Preschool and Kindergarten

เป็นบทความที่อ้างอิงถึงงานวิจัยเกี่ยวกับการศึกษาเทคนิคในการอ่านออกเสียงของเด็กปฐมวัยในช่วงก่อนอนุบาลหรือช่วงอนุบาลหรือในช่างอายุ 3-6 ขวบที่ได้ผลดีที่สุด ซึ่งได้ข้อสรุปว่าการอ่านออกเสียงที่ดีที่สุดคือการอ่านออกเสียงโต้ตอบซ้ำๆ (Repeated Interactive Read Alouds) ที่มีความเกี่ยวข้องกับการถามและตอบคำถามและการให้เด็กคาดคะเนเหตุการณ์ล่วงหน้า การที่ให้เด็กอ่านออกเสียงน้อยนั้นจะส่งผลต่อการพัฒนาทักษะการพูด การจดจำคำศัพท์ การรับรู้โดยการฟังการอ่านที่ให้เด็กได้มีโอกาสโต้ตอบและคาดคะเนเหตุการณ์ล่วงหน้าหรือมีการอธิบายแรงผลักดันของตัวละครหรือเหตุหารณ์ต่างๆจะช่วยพัฒนาทักษะการคิดวิเคราะห์ของเด็กในการให้ความคิดเห็นและการตั้งคำถามได้

2.8 Machine Learning

Machine Learning เป็นศาสตร์ที่ว่าด้วยอัลกอริทึมที่ทำให้เครื่องกลเรียนรู้และเข้าใจในประเด็นที่เราสนใจจากข้อมูล

ปัจจุบันอัลกอริทึมที่ใช้ในการแก้ปัญหานั้นมีจำนวนมาก (มากกว่า 100) โดยเราจะแบ่งอัลกอริทึมออกเป็น 3 ประเภท

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น มีไว้เพื่อไม่ให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

1.การเรียนรู้แบบมีผู้สอน (Supervised Learning)

2.การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning)

3.การเรียนรู้แบบเสริมแรง (Reinforcement Learning)

2.8.1. การเรียนรู้แบบมีผู้สอน (Supervised Learning)

รูปแบบการเรียนรู้แบบมีผู้สอนโดยทั่วไป จะอยู่ในลักษณะการทำนายผลลัพธ์ เช่น จะทำนายว่าตัวอักษรในภาพคืออะไร หรือรายได้ที่ควรคาดหวังจากการลงทุนเป็นเท่าใด "การมีผู้สอน" ในที่นี้หมายความว่า ในข้อมูลที่ใช้ฝึก (Training Data) เราจะมีมนุษย์มาคอยแยกประเภทหรือบอกผลลัพธ์ (Label) ที่ควรจะเป็นไปไว้ด้วย จากนั้นเราจะนำข้อมูลที่ใช้ฝึก (ที่มี Label) ไปผ่านอัลกอริทึมสำหรับสร้างโมเดลที่ไว้ทำนายผลลัพธ์
เมื่อเราได้โมเดลที่ไว้ทำนายแล้ว เราจะนำข้อมูลใหม่ที่เครื่องไม่เคยเห็น กล่าวคือไม่ใช่ข้อมูลชุดเดียวกันกับข้อมูลฝึกหัด เครื่องจะต้องทำนาย (Predict) ว่าคำตอบที่ได้ควรจะเป็นอะไร
ประเภทของการเรียนรู้แบบมีผู้สอนจะแยกออกเป็นสองประเภท

1.การแยกประเภท (Classification)

2.การถดถอยเชิงสถิติ (Statistical Regression)

2.8.1.1 การแยกประเภท Classification

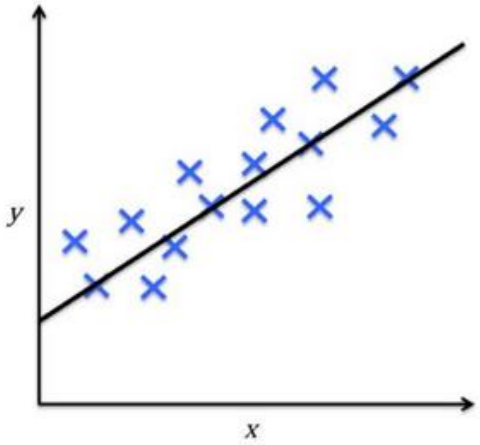
การแยกประเภทเป็นหนึ่งในการเรียนรู้แบบมีผู้สอน และเป็นประเภทที่พบบ่อย เช่น การจดจำใบหน้าของคน การจดจำใบหน้าคนนั้นแต่ละส่วนในรูปแบบพื้นฐานประกอบด้วยอะไรบ้าง

คำว่า แยก"ประเภท" (Class) ในที่นี้ก็คือชื่อของคนนั้นเอง
Class เป็นกลุ่มของชุดคำตอบที่เป็นไปได้ เช่น เรามี 100,000 ภาพก็จริง แต่ใน 100,000 ภาพนั้นเป็นภาพของคน 10 คน จะได้ว่า Class ของเรานั้นจะเท่ากับ 10

2.8.1.2การถดถอยเชิงสถิติ (Statistical Regression)

การถดถอยเชิงสถิติเป็นการหาความสัมพันธ์ระหว่างตัวแปร เช่น พื้นที่ของที่ดิน (x) กับราคาที่ดิน (y) โมเดลที่ใช้ในการหาความสัมพันธ์มีด้วยกันอยู่หลายแบบ แต่ที่นิยมกันคือการถดถอยเชิงเส้น (Linear Regression)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.



รูปที่ 2.4 กราฟแสดง regression

ตัวอย่างการถดถอยเชิงเส้น ในที่นี้เราจะใส่ข้อมูลเป็น x และ ระบุค่า y ที่ควรจะเป็น จากนั้นเครื่องจะประมาณเส้นใดๆ ที่จะสามารถประมาณค่าของ y จาก x ที่กำหนดได้ดีที่สุด

2.8.2 การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning)

ในการเรียนรู้แบบมีผู้สอน เราจะรู้ว่าคำตอบที่ควรจะเป็นคืออะไร และเราก็ระบุคำตอบที่ถูกต้องลงไปใน Label แต่ในการเรียนรู้แบบไม่มีผู้สอน เราจะไม่รู้คำตอบที่แน่ชัด แต่เราต้องการให้เครื่องตามหาโครงสร้างข้อมูลที่เราไม่รู้จัก (Unknown Structure)

เช่น ในโครงสร้าง DNA เราอยากตามหายีนที่ส่งผลให้คนมีตาสีฟ้าหรือสีดำ เราไม่รู้ว่าตรงไหนของ DNA ที่ส่งผลต่อสีตา

เรารู้แต่ว่าข้อมูล DNA ของคนที่เรานำมาทดสอบมีสองกลุ่ม กลุ่มหนึ่งคือ คนที่มีตาสีฟ้า อีกกลุ่มคือคนที่มีตาสีดำ

เราอยากให้เครื่องตรวจหาว่าส่วนไหนของ DNA ที่แสดงความแตกต่างออกมาอย่างเด่นชัดระหว่างกลุ่ม และเหมือนกันมากในกลุ่มเดียวกัน

การเรียนรู้แบบไม่มีผู้สอนแบ่งออกเป็น 2 ประเภท

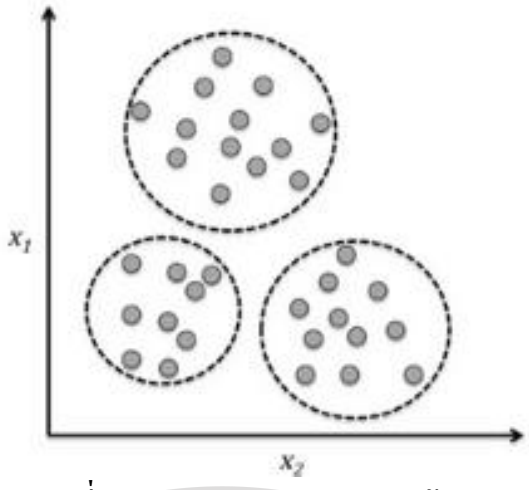
1.การจับกลุ่มของข้อมูล (Clustering)

2.การลดจำนวนมิติเพื่อบีบอัดข้อมูล (Dimensionality Reduction)

2.8.2.1 การจับกลุ่มของข้อมูล (Clustering)

ในบางที่เรามีข้อมูลจำนวนมากที่เราไม่สามารถจะใส่ Label ได้ไหว เรารู้แต่ว่าข้อมูลของเรานั้นประกอบด้วยจำนวน Class เท่าใด เราอาจจะระบุให้เครื่องแบ่งข้อมูลออกเป็นจำนวน Class ที่เราต้องการ จากนั้นเราค่อยมาระบุเองว่าแต่ละกลุ่มที่เครื่องพบคืออะไร จากที่เราจะต้องระบุ Label จำนวนมากจะลดลงเหลือแค่เท่ากับจำนวน Class

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

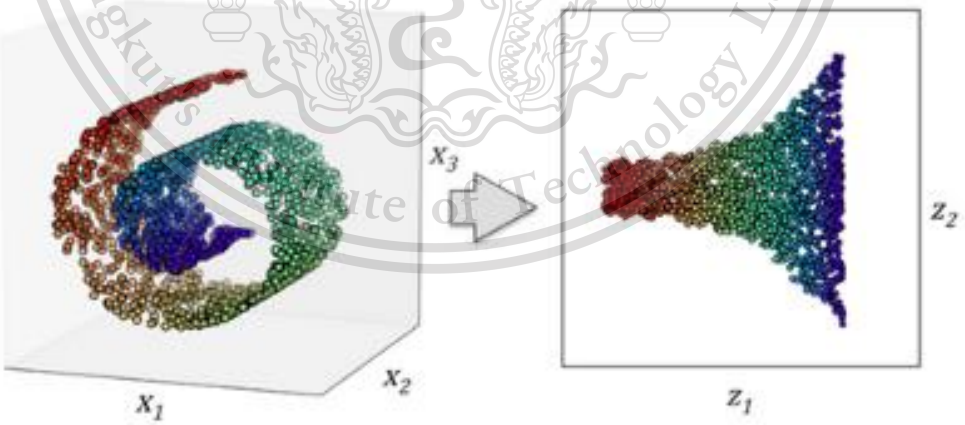


รูปที่ 2.5 รูปแสดงการจับกลุ่มของข้อมูล

ภาพตัวอย่างการจับกลุ่มข้อมูลจากข้อมูลเข้าสองมิติ (ข้อมูลเข้าเป็นตัวเลขสองค่าคือ x1 และ x2) เราไม่รู้ว่าตัวไหนอยู่ในกลุ่มใด (จุดสีเทา) แต่เรากำาหนดให้เครื่องหาวิธีแบ่งกลุ่มข้อมูลออกเป็นสามกลุ่ม และ ได้ผลออกเป็นบริเวณขอบเขตของแต่ละกลุ่ม

2.8.2.2 การลดจำนวนมิติเพื่อบีบอัดข้อมูล (Dimensionality Reduction)

การลดจำนวนมิติเพื่อบีบอัดข้อมูล เป็นกลไกที่ทำให้เราไม่จำเป็นต้องเก็บข้อมูลไว้ครบ แต่ก็ยังสามารถจำแนกข้อมูลได้ เช่นเรามีข้อมูลสามมิติของข้อมูลหลายๆคลาส สิ่งที่เราต้องการคือ การลดข้อมูลให้เหลือสองมิติและยังสามารถนำไปแยกประเภทคลาสได้ดีเท่าเดิม (อาจดีขึ้นเป็นไปได้)



รูปที่ 2.6 ภาพแสดงการบีบอัดข้อมูล

เพราะเราไม่ทราบว่าการจะแปลงข้อมูลสามมิติให้เป็นสองมิติที่ดีได้อย่างไร โดยการจัดเรียงข้อมูลนั้นมีโครงสร้างแบบ Unknown Structure (โดยเฉพาะตอนเป็นสามมิติเราไม่มีวิธีการแสดงผล) เราอาจจะกำหนดไปเพียงแค่ว่าโมเดลของการเปลี่ยนแปลงข้อมูลเป็นอย่างไร เช่น

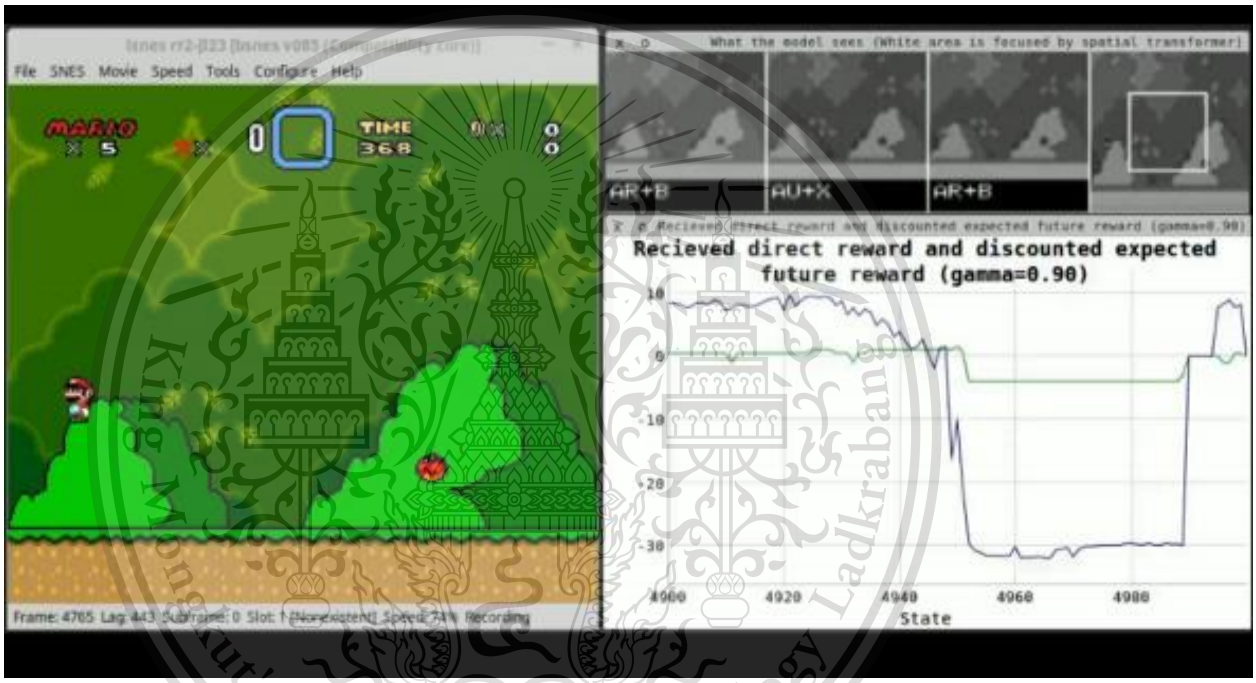
$$Z_1 = k_1x_1 + k_2x_2 + k_3x_3 \quad (2.1)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

หมายความว่าเราจะสร้างค่า Z ใดๆมาจากการรวมกันแบบเชิงเส้นของค่า x ทั้งสามแกน แต่เราไม่รู้ว่าค่า k แต่ละอันควรเป็นเท่าใด และเราต้องการให้เครื่องหาให้

2.8.3 การเรียนรู้แบบเสริมแรง (Reinforcement Learning)

หลายคนอาจเคยคิดอยากจะให้คอมพิวเตอร์เล่นเกมแทนเรา หรือที่เราเคยชินกับคำว่า “Bot” ถ้าเป็นเกมเก็บเลเวลแบบปกติ เราอาจจะเขียนโปรแกรมให้มันดำเนินตามกฎอะไรบางอย่างไปเรื่อยๆ แต่ถ้าเป็นเกมอย่าง Mario ล่ะ

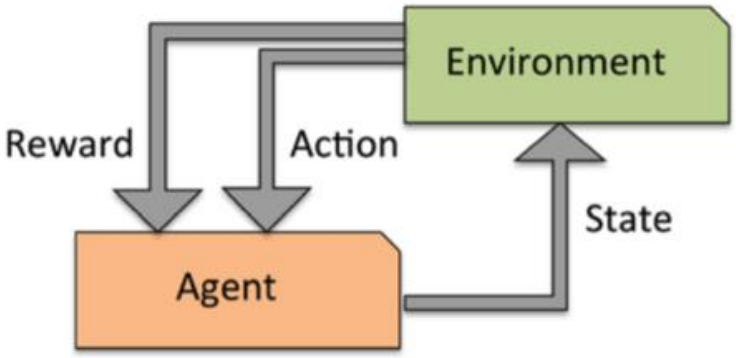


ภาพ 2.7 ภาพตัวอย่างการเรียนรู้แบบเสริมแรง

เราอาจจะใช้โมเดลการให้รางวัลเมื่อบอทเราสามารถทำแต้มได้ดี และให้มันคอยสังเกตว่าการทำ Action แต่ละอย่างแล้วจะได้คะแนนสุดท้ายออกมาเป็นเท่าไร

ถ้าเราทำ Action ตามลำดับต่างๆ ออกมาแล้วคะแนนดี เราก็จะได้รางวัล (Reward) เครื่องจะเริ่มจดจำลำดับการทำ Action แล้วได้คะแนนดี และพยายามจะทำ Action นั้นเรื่อยๆ การที่มันได้รางวัลเท่ากับเป็นการสนับสนุนให้โปรแกรมทำ Action นั้นซ้ำๆ ยิ่ง Action นั้นได้รางวัลมาก ก็จะทำแบบนั้นบ่อยๆ สิ่งที่สำคัญคือ เราต้องคิดโมเดลการให้คะแนนที่ดี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.



รูป 2.8 แผนภาพรวมการเรียนรู้แบบเสริมแรง

ภาพด้านบนจะเป็นภาพรวมของการเรียนรู้แบบเสริมแรง โดย Agent จะทำหน้าที่คอยใส่ Action ให้กับระบบ จากนั้น Action เข้าสู่สถานะหนึ่งในระบบและถูกประเมินออกมาเป็นคะแนน โดยจะทำแบบนี้ไปเรื่อยๆจนกว่าจะสามารถสรุปผลลัพธ์ของคะแนน (เกมจบ หรือตาย)

2.9 วิธีการทดสอบประสิทธิภาพโมเดลแบบ Classification

		Actual	
		Spam	Ham
Prediction	Spam	20	12
	Ham	18	50

รูปที่ 2.9 ตารางตัวอย่างข้อมูล

ตัวอย่างการคำนวณ metrics ต่างๆจะใช้ตัวเลขในตารางนี้เป็นหลัก และผลการทำนายอีเมล์ทั้งหมดนี้มี N=100

2.9.1 Accuracy

Accuracy คือ metric ที่ใช้งานง่ายที่สุด บอกว่าโมเดลเราทำนายถูกทั้งหมดกี่ % จากตารางด้านบน accuracy = (20 + 50) / 100 = 70% คำนวณจากเส้นทแยงมุมของ confusion matrix ได้เลย ถ้าแบบทางการหน่อย เราจะเขียนสูตรว่า accuracy = (TP + TN)/ N โดยค่า accuracy จะมีค่าอยู่ระหว่าง 0–1 ยิ่งเข้าใกล้ 1 แปลว่าโมเดลเราทำนายผลได้ดีมาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

2.9.2 Precision

นิยามของ Precision คือความน่าจะเป็นที่โมเดลทำนาย spam ถูกต้องจากการทำนาย spam ทั้งหมด 32 ครั้ง (20 + 12 ผลรวมแถวบนของ confusion matrix)

แทนค่าในสมการ precision = TP / (TP + FP) = 20 / (20 + 12) = 62.5%

2.9.3 Recall

นิยามของ Recall คือความน่าจะเป็นที่โมเดลสามารถตรวจจับ spam จากจำนวน spam email ทั้งหมดในข้อมูลของเรา 38 ฉบับ (20 + 18 ผลรวมคอลัมน์แรกของ confusion matrix)

แทนค่าในสมการ recall = TP / (TP + FN) = 20 / (20 + 18) = 52.6%

2.9.4 F1-Score

F1-Score คือค่าเฉลี่ยแบบ harmonic mean ระหว่าง precision และ recall นักวิจัยสร้าง F1 ขึ้นมาเพื่อเป็น single metric ที่วัดความสามารถของโมเดล (ไม่ต้องเลือกระหว่าง precision, recall เพราะเฉลี่ยให้แล้ว)

$$F1 = 2 * \left(\frac{precision * recall}{precision + recall}\right) \quad (2.2)$$

แทนค่าในสมการ F1 = 2 * ((0.625 * 0.526) / (0.625 + 0.526)) = 57.1%

Accuracy ไม่ใช่ metric เดียวที่เราต้องดู

2.9.5 F-Beta Score

F Score มีสูตรทั่วไปที่เราสามารถกำหนดค่า Beta ได้เอง (เช่น F1 คือการกำหนด Beta = 1) ถ้าเราอยากให้น้ำหนักไปทาง precision ให้กำหนดค่า Beta < 1 แต่ถ้าอยากเน้นที่ recall ให้กำหนดค่า Beta > 1 .

$$F_\beta = (1 + \beta^2) * \left(\frac{precision * recall}{\beta^2 precision + recall}\right) \quad (2.3)$$

ถ้าเรากำหนด Beta=0 จะได้ค่า precision เพียวๆเลย (prove สูตรให้ดูด้านล่าง) การใช้ F-Beta Score ช่วยให้เราปรับ metric ได้เหมาะสมตามสถานการณ์ เช่น งานที่เน้น recall ก็ให้ใช้ Beta = 2, 3, 5 เป็นต้น

example beta=0

F0 = (1 + 0) * (precision * recall) / (0 + recall)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

F0 = precision * recall / recall

F0 = precision

2.9.6 AUC

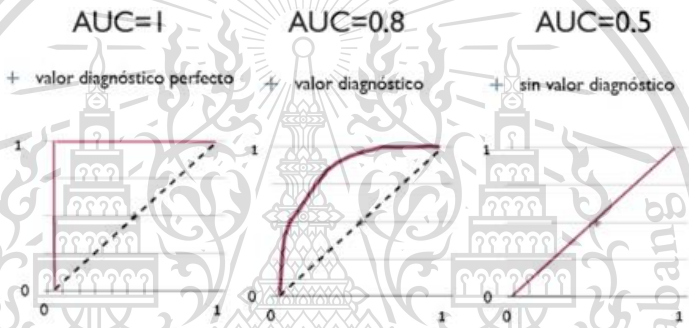
AUC ย่อมาจาก “Area Under <ROC> Curve” เป็นอีกหนึ่ง metric ยอดนิยมที่ใช้กันแทบทุกงานเลย AUC มีค่าอยู่ระหว่าง 0–1 ยิ่งเข้าใกล้ 1 แปลว่าโมเดลในภาพรวมสามารถทำนาย y ได้ดีมาก

AUC = 0.50 ไม่ต่างอะไรกับการเดาสุ่มเลย

AUC > 0.70 คือเกณฑ์มาตรฐานสำหรับโมเดลส่วนใหญ่

AUC > 0.80 โมเดลทำงานได้ดี

AUC > 0.90 โมเดลทำงานได้ดีมาก



รูป 2.10 ภาพแสดงกราฟ AUC

2.10 Linear SVC Model

วัตถุประสงค์ของ Linear SVC (Support Vector Classifier) คือเพื่อให้เหมาะสมกับข้อมูลที่ใส่เข้าไปโดยจะส่งคืนไฮเปอร์เพลนที่ "เหมาะสมที่สุด" ที่แบ่งหรือจัดหมวดหมู่ข้อมูล จากนั้นหลังจากได้รับไฮเปอร์เพลนแล้วสามารถป้อนคุณสมบัติบางอย่างไปยังตัวจำแนกเพื่อดูว่าคลาส "ทำนาย" คืออะไร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

บทที่ 3
วิธีการดำเนินงาน

3.1 การจัดหา Dataset สำหรับทำการทดสอบ

3.1.1 รายละเอียดของ Dataset

เป็นไฟล์เสียงของนิทานเรื่องลูกหมูสามตัวมีความยาวประมาณ 4 หน้ากระดาษ A4

3.1.2 การแปลงไฟล์

ไฟล์เสียงที่ได้มาจะเป็น ไฟล์ MP3/M4A ซึ่งตัว API ที่ใช้ไม่รองรับจะต้องใช้เป็นไฟล์ประเภท WAV หรือ RAW เท่านั้น จึงต้องอาศัยโปรแกรมภายนอกในการแปลงประเภทไฟล์

3.2 การสร้างเกณฑ์สำหรับการตรวจสอบคุณภาพของหนังสือเสียง

3.2.1 ความถูกต้อง

3.3.1.1 อธิบายวิธีการออกแบบและใช้

มีการตรวจสอบว่าเสียงพูดที่ได้เมื่อมีการแปลงข้อมูลออกมาเป็น text แล้วนั้นตรงกับ Script ที่ใส่เข้าไป ไม่ผิดเพี้ยนหรือได้ความหมายที่ต่างออกไป

3.2.2 ความเร็ว

3.3.1.1 อธิบายวิธีการออกแบบและใช้

มีการตรวจสอบว่าเสียงพูดที่ได้ไม่เร็วเกินไปจนจับความหมายไม่ได้ หรือช้าเกินไปจนทำให้ฟังดูน่าเบื่อและไม่เหมาะสมต่อพัฒนาการทางด้านการเรียนรู้ของกลุ่มเป้าหมายซึ่งสอดคล้องต่อทฤษฎีอ้างอิง

3.2.3 การวรรคตอน

3.3.1.1 อธิบายวิธีการออกแบบและใช้

มีการตรวจสอบว่า วรรคตอนแต่ละแบบมีความยาวที่เหมาะสมต่อพัฒนาการทางด้านการเรียนรู้ของกลุ่มเป้าหมายซึ่งสอดคล้องต่อทฤษฎีอ้างอิง หรือการวรรคตอนที่ยาวนานเกินไปอาจจะทำให้ความหมายผิดเพี้ยนไปได้เช่นกัน

3.3 ภาพรวมการทำงานของระบบ

3.3.1 รับไฟล์เสียงของหนังสือเสียงและบทของไฟล์เสียง

3.3.2 นำไฟล์เสียงไปแปลงเป็นข้อความ

3.3.3 นำข้อความที่ได้ไปเปรียบเทียบกับบทของไฟล์เสียงและบันทึกผล

3.3.4 นำไฟล์เสียงไปตรวจสอบความเร็ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

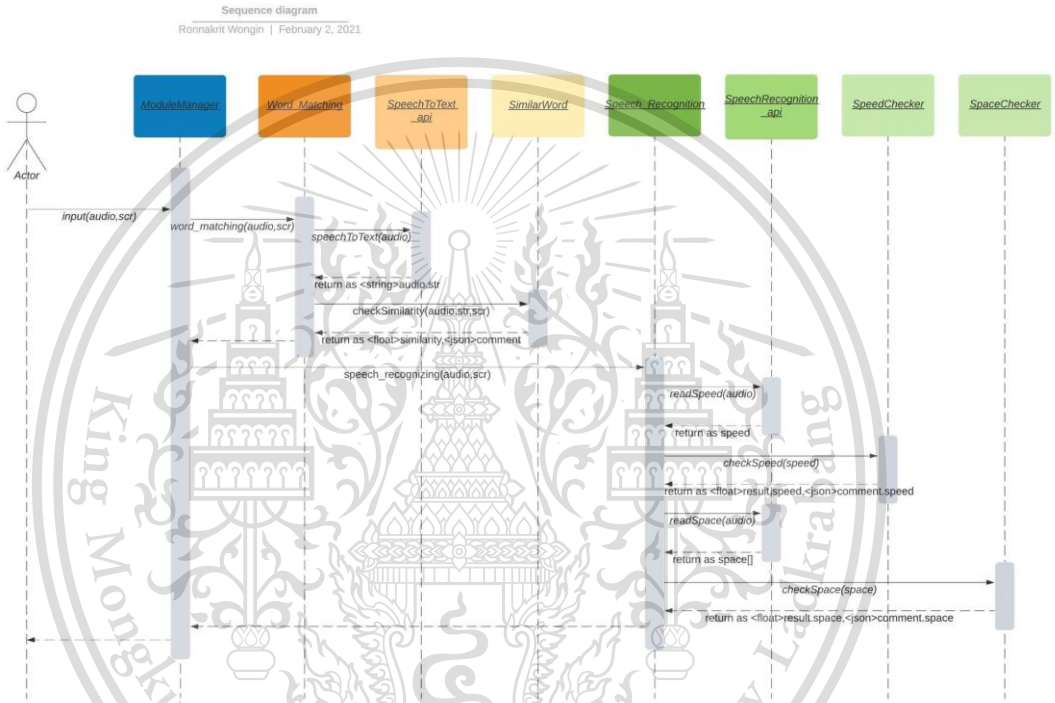
3.3.5 นำผลลัพธ์ที่ได้ไปเปรียบเทียบกับความเร็วตามเกณฑ์ที่เราต้องการและบันทึกผล

3.3.6 นำไฟล์เสียงที่ได้ไปตรวจสอบการเว้นวรรค

3.3.7 นำผลลัพธ์ที่ได้ไปเปรียบเทียบระยะเว้นวรรคตามเกณฑ์ที่เราต้องการและบันทึกผล

3.3.8 แสดงผลลัพธ์ที่ได้ทั้งหมด

3.4 การออกแบบการทำงานและโครงสร้างของสถาปัตยกรรม



รูป 3.1 แผนภาพแสดงการทำงานของโปรแกรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

บทที่ 4
การทดลอง

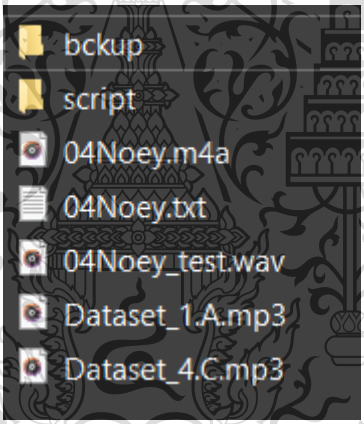
4.1 ทดสอบ Demo กับไฟล์ตัวอย่างประเภท wav

4.1.1 วัตถุประสงค์

ทดสอบ demo ของโมดูล Word-Matching ว่าใช้งานได้กับไฟล์ตัวอย่างประเภท wav หรือไม่

4.1.2 วิธีการทดสอบ

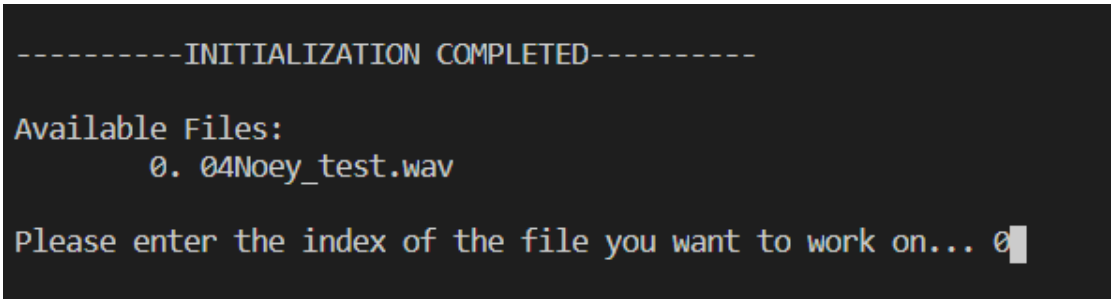
4.1.2.1 นำไฟล์เสียงตัวอย่างที่จะใช้ทดสอบ จากรูปที่ 4.1 จะใช้ไฟล์ 04Noey.wav มาใส่ไว้ในโฟลเดอร์ resource ในdirectory ของโปรแกรม



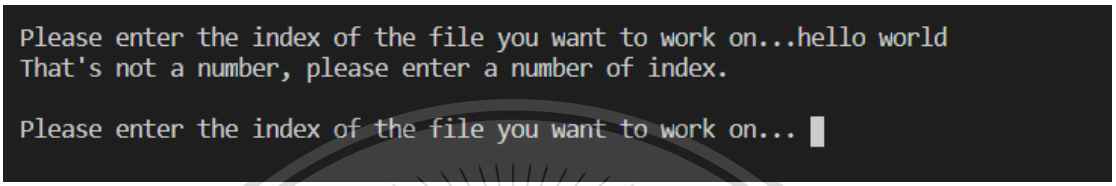
รูป 4.1 demo files

4.1.2.2 เมื่อเริ่มทำงาน main.py โปรแกรมจะใช้เวลาสักพักเพื่อเริ่มต้นการทำงาน เมื่อเสร็จแล้วจะแสดงชื่อไฟล์ที่เป็นประเภท wav ขึ้นมาให้เลือกพร้อมลำดับของไฟล์นั้นๆ ตามภาพ 4.2 ในส่วนนี้จะทำการทดสอบ input แบบต่างๆตามภาพ 4.3 ทดสอบโดยการป้อน input ‘hello world’ ภาพ 4.4 ทดสอบโดยการป้อน input ตัวเลขเกินลำดับที่ให้มา และภาพ 4.5 ทดสอบโดยการป้อน input ตัวเลข 0

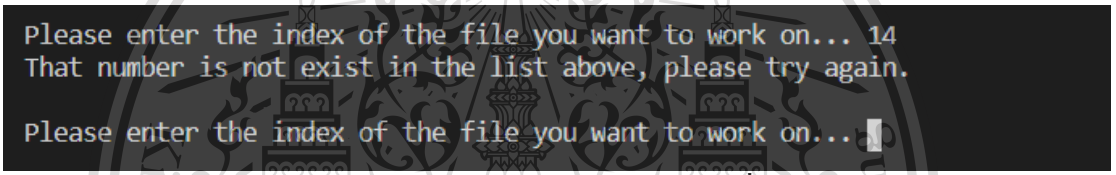
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.



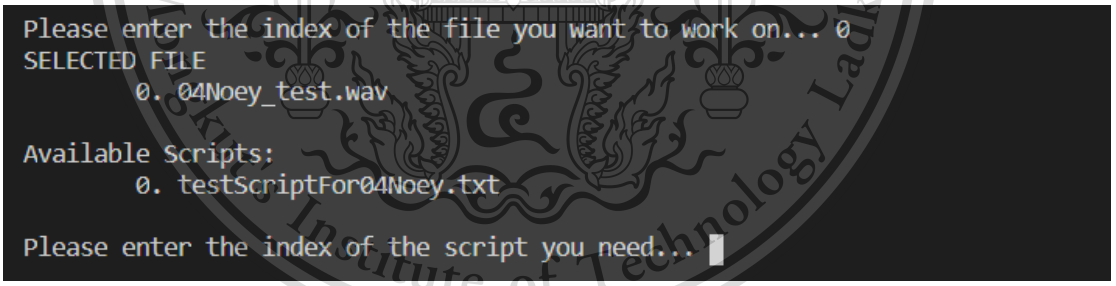
รูป 4.2 File Selector



รูป 4.3 File Selector: ป้อน input 'hello world'



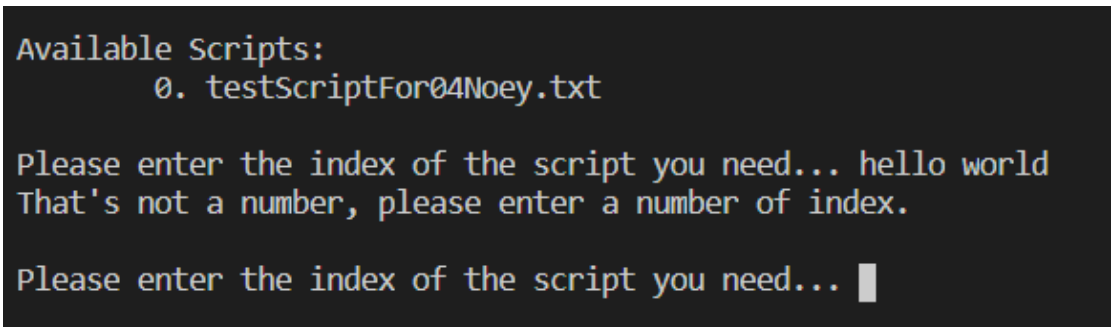
รูป 4.4 File Selector: ป้อน input เกินลำดับที่กำหนดให้



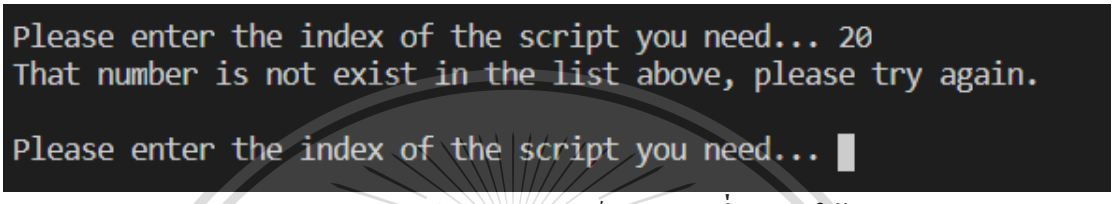
รูป 4.5 File Selector: ป้อน input 0

4.1.2.3 หลังจากภาพ 4.5 ที่พิมพ์ลำดับที่ 0 แล้วโปรแกรมจะแสดงชื่อไฟล์สคริปต์ประเภท txt เช่นเดียวกันกับใน 4.1.2.2 ในส่วนนี้จะทำการทดสอบป้อน input แบบต่างๆ เช่นเดียวกันกับใน 4.1.2.2 ตามภาพ 4.6 ทดสอบโดยการป้อน input 'hello world' ภาพ 4.7 ทดสอบโดยการป้อน input ตัวเลขเกินลำดับที่ให้มา และภาพ 4.8 ทดสอบโดยการป้อน input ตัวเลข 0

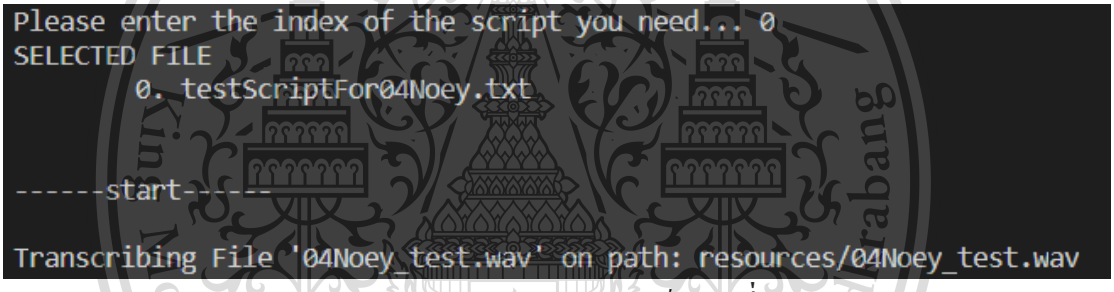
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.



รูป 4.6 Script Selector: พิมพ์ลำดับเป็น 'hello world'

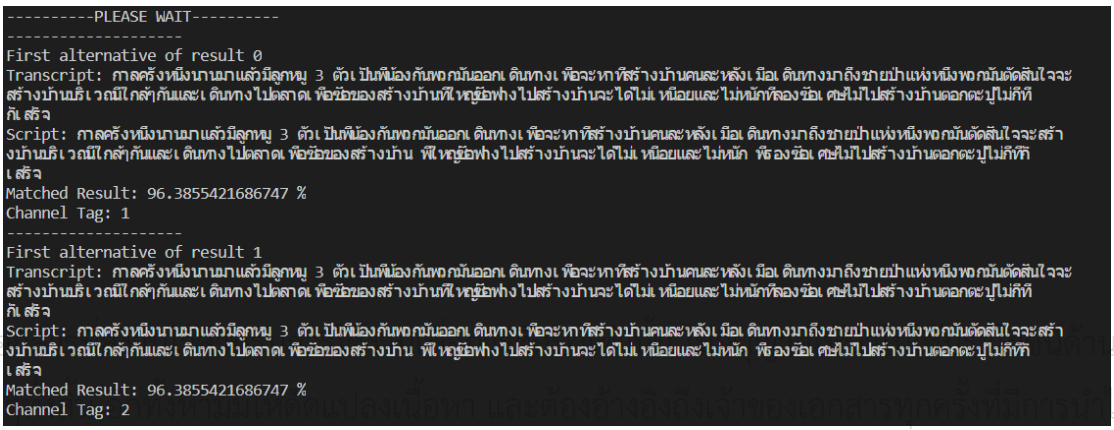


รูป 4.7 Script Selector: พิมพ์ลำดับเกินที่กำหนดให้



รูป 4.8 Script Selector: พิมพ์ลำดับที่ 0

4.1.2.4 ตามภาพที่ 4.8 โปรแกรมจะทำงาน รอสักพักจนโปรแกรมทำงานเสร็จจะแสดงผล Word-Matching ตามภาพที่ 4.9 และรายงานต่างๆตามภาพที่ 4.10 ซึ่งจะแสดงค่าความเหมาะสมของส่วนต่างๆในช่วง 0-1



รูป 4.9 Word-Matching Result

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น ไม่อนุญาตให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

```
---------report---------

word matching result = 0.963855421686747
That's Great!

speed = 166.38731268898155 words per minute
speed checker result = 0.9936126873110185
A little too fast but ok

------report ended------


-------end-------
```

รูป 4.10 Report

4.1.3 ผลการทดสอบ

เสียงที่ใส่เข้าไปมีความถูกต้องตรงกับบท 0.9639 หรือ 96.39 % ซึ่งเป็นผลการทดสอบที่ดีมาก และความเร็วของเสียงมีความเหมาะสม 0.9936 หรือ 99.36 % ซึ่งเป็นผลการทดสอบที่ดีมากเช่นกัน ตามภาพ 4.10

ตาราง 4.2 เปรียบเทียบความแม่นยำของโมเดลโดยเฉลี่ย ในนิทานเรื่องที่ 1 ถึง 4 โดยใช้ความเร็วในการอ่านแตกต่างกัน

ความเร็วในการอ่าน	ลูกหมูสามตัว	นมสดกับหม้อดิน	ราชสีห์กับหมูป่า	นกกระยางเรื่องมาก
100 คำ/วินาที	**89.23%**	**86.01%**	**82.06%**	**77.63%**
140 คำ/วินาที	**96.44%**	**92.98%**	**88.71%**	**83.69%**
160 คำ/วินาที	**98.13%**	**97.30%**	**91.53%**	**87.32%**
180 คำ/วินาที	**92.55%**	**89.91%**	**85.85%**	**80.10%**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

บทที่ 5
ข้อสรุปและข้อเสนอแนะ

5.1 ข้อสรุป

5.1.1 ออกแบบเกณฑ์การวัดค่าความถูกต้องของหนังสือเสียงที่สอดคล้องกับผลงานวิจัยอ้างอิง

5.1.2 สามารถพัฒนาโปรแกรมตรวจสอบหนังสือเสียงได้จนสมบูรณ์ สามารถใช้กับภาษาไทยได้และได้ผลลัพธ์ที่มีความถูกต้องในการตรวจสอบความถูกต้องของการสะกดคำ

5.2 ข้อเสนอแนะ

สำหรับไฟล์เสียงที่ใช้หากมีขนาดที่ใหญ่เกินไปสามารถลดขนาดของไฟล์ลงได้โดยไม่มีผลเป็นนัยยะสำคัญกับการพัฒนาเอไอ เนื่องจากในการตรวจสอบความถูกต้องของหนังสือเสียงนั้นเราสนใจที่เนื้อหาในชุดข้อมูลนั้น ในแต่ละคำและประโยคมีความถูกต้องหรือไม่ โดยไม่ได้พิจารณาลงลึกไปถึงความละเอียดของระดับเสียง ซึ่งจะช่วยให้ประหยัดพื้นที่ในการประมวลผลลง

5.3 ปัญหาและอุปสรรคที่พบ

ปัญหาด้านชุดข้อมูลที่จะต้องจัดทำเองให้ทำให้ใช้เวลาในการทำค่อนข้างมาก เนื่องจากมีขั้นตอนในการจัดเตรียมให้สามารถใช้ได้ค่อนข้างมาก ไม่ว่าจะเป็นการลบเสียงรบกวน การระบุแท็กซึ่งระบุพารามิเตอร์ในการนำไปใช้สำหรับเป็นชุดข้อมูลสำหรับการฝึกฝนของเอไอ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

บรรณานุกรม

Adam Geitgey,(2016),Machine Learning is Fun Part 5, “Language Translation with Deep Learning and the Magic of Sequences”,[Online],เข้าถึงได้จาก: https://medium.com/@ageitgey/machine-learning-is-fun-part-5-language-translation-with-deep-learning-and-the-magic-of-sequences-2ace0acca0aa

Adam Geitgey,(2016), Machine Learning is Fun Part 6, “How to do Speech Recognition with Deep Learning”,[Online],เข้าถึงได้จาก: https://medium.com/@ageitgey/machine-learning-is-fun-part-6-how-to-do-speech-recognition-with-deep-learning-28293c162f7a

Gamal Bohouta, Veton Z Këpuska (2017), “Comparing Speech Recognition Systems (Microsoft API, Google API And CMU Sphinx)”, Florida Institute of Technology,[Online],เข้าถึงได้จาก:https://www.researchgate.net/publication/314938892_Comparing_Speech_Recognition_Systems_Microsoft_API_Google_API_And_CMU_Sphinx

Nenny Anggraini, Angga Kuniawan, Luh Kesuma Wardhani, Nashrul Hakiem (2018),“Speech recognition application for the speech impaired using the android-based google cloud speech API”, Syarif Hidayatullah State Islamic University ,[Online],เข้าถึงได้จาก:Jakarta,https://www.researchgate.net/publication/329865231_Speech_recognition_application_for_the_speech_impaired_using_the_android-based_google_cloud_speech_API

Carla A. Brooks,Communication Skills for Children Who Are Blind or Visually Impaired,[Online],เข้าถึงได้จาก:https://familyconnect.org/browse-by-age/infants-and-toddlers/growth-and-development-iandt/communication-skills-for-children-who-are-blind-or-visually-impaired/1235/

Lea M. McGee, Judith Schickedanz,Repeated Interactive Read Alouds in Preschool and Kindergarten,[Online],เข้าถึงได้จาก:https://www.readingrockets.org/article/repeated-interactive-

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

read-alouds-preschool-and-kindergarten?fbclid=IwAR32_bWDN09YzAYdqHCs0IMM9uLHBYmniPUMH5LsU-oUtB_DxYZQSg87-4o

ไม่ปรากฏชื่อผู้เขียน,ชนิดของไฟล์เสียงบอกอะไรได้บ้าง,[Online],เข้าถึงได้จาก:https://www.mercular.com/review-article/audio-file-type?fbclid=IwAR3m2Z1j-6A15cgJfkulogoLypQBvzLASjiVtXrwSJc0Af3qUXGoobTDFOE

ไม่ปรากฏชื่อผู้เขียน,gRPC,[Online],เข้าถึงได้จาก:https://en.wikipedia.org/wiki/GRPC?fbclid=IwAR0maXSHRnYmq4xPqkImwF1sYGN7ZnT8A4yPO9skdu587aX-mcflsrl9jGk

saixiii,(2017),RESTs,[Online],เข้าถึงได้จาก:https://en.wikipedia.org/wiki/GRPC?fbclid=IwAR0maXSHRnYmq4xPqkImwF1sYGN7ZnT8A4yPO9skdu587aX-mcflsrl9jGk

saixiii,(2017),API คืออะไร,[Online],เข้าถึงได้จาก:https://saixiii.com/what-is-api/

Google,Google API Speech to Text,[Online],เข้าถึงได้จาก:shorturl.at/diJQ1

มาทำความรู้จัก Machine Learning เบื้องต้น,[online],เข้าถึงได้จาก: https://medium.com/@natthawatphongchit/machine-learning-basics-2b38700cb10b

Metrics พื้นฐานสำหรับวัดประสิทธิภาพของโมเดล Machine Learning,[online],เข้าถึงได้จาก: shorturl.at/fkmFV

Average Speaking Rate and Words per Minute,[online],เข้าถึงได้จาก: https://virtualspeech.com/blog/average-speaking-rate-words-per-minute?fbclid=IwAR1567rK64inzyOxuIIjgjWLmYhi_3zY29Pw1fLOFEOdsgnaPyevt6ICfWc

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.