|  |  |
| --- | --- |
| Image result for ตราสถาบัน | **ข้อเสนอโครงงานวิศวกรรมคอมพิวเตอร์**  **วิชา 01076014 การเตรียมโครงงานวิศวกรรมคอมพิวเตอร์**  **ภาคเรียนที่ 2 ปีการศึกษา 2562** |

1. ชื่อหัวข้อโครงงาน (ไทย) การตรวจสอบความถูกต้องของหนังสือเสียงโดยใช้ machine

learning และ AI

1. ชื่อหัวข้อโครงงาน (อังกฤษ) Assessing Accuracy of Audiobooks by Using Machine

Learning and AI

1. คำสำคัญ 3 คำ (3 keywords) Audiobooks, Machine Learning Accuracy, AI
2. รายชื่อผู้ทำโครงงาน
   1. นาย พรเทพ หล่มแสง รหัส 60010659
   2. นาย ภัคพงษ์ ปุรสาชิต รหัส 60010760
3. อาจารย์ที่ปรึกษา
   1. อาจารย์ที่ปรึกษาหลัก ผศ.อัครเดช วัชระภูพงษ์
4. ที่มาและความสำคัญของปัญหา (Motivation)

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

เนื่องจากในปัจจุบันนี้มีการจัดทำหนังสือเสียง (Audiobooks) สำหรับผู้บกพร่องทางสายตาและผู้สูงอายุ หรือคนปกติที่ต้องการรับรู้เนื้อหาของหนังสือผ่านทางการฟังได้มีโอกาสรับรู้ข่าวสารข้อมูลจากสื่อสิ่งพิมพ์เพิ่มขึ้น 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

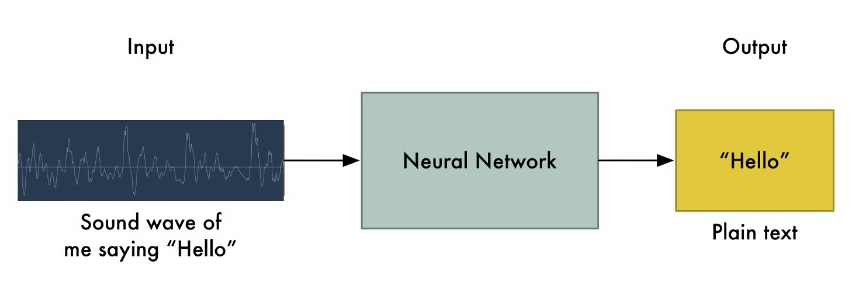
1. วัตถุประสงค์ (Objectives)

เพื่อวัดความถูกต้องของหนังสือเสียงด้วย machine learning และ AI

1. ทฤษฎีที่เกี่ยวข้อง (Theoretical Background)

**Speech Recognition**

กระบวนการ Speech Recognition เป็นกระบวนการที่รับเอาเสียงพูดของมนุษย์ นำไปตีความและถอดความให้กลายเป็นตัวอักษร โดยวิธีการแบ่งเสียงพูดของมนุษย์ออกเป็นส่วนๆ และทำการวิเคราะห์เสียงแต่ละส่วน โดยการใช้ algorithms เพื่อหาคำที่เหมาะสมที่สุดกับภาษานั้นและทำการแปลงภาษาเหล่านั้นเป็นตัวอักษร โดยอาศัยกระบวนการดังนี้

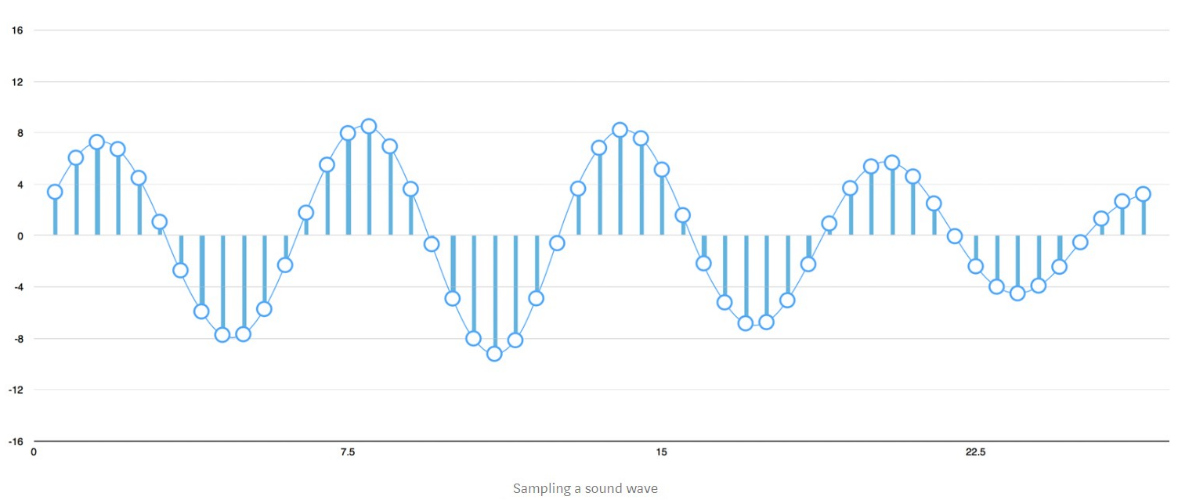
* ป้อนเสียงพูดที่บันทึก ลงใน neural network เพื่อแปลงเสียงพูดที่ถูกบันทึกไปเป็นข้อความ

simply feed sound recordings into a neural network and train to produce text

กระบวนการดังกล่าวมีปัญหาใหญ่อยู่นั่นก็คือ ความเร็วในการออกเสียงพูดของแต่ละคนไม่เท่ากัน บางคนพูดคำว่า “Hello” แบบเร็วๆ บางคนก็พูดคำว่า “Hello” แบบช้าๆ ซึ่งทำให้ขนาดของข้อมูลที่ถูกจัดเก็บมีขนาดไม่เท่ากัน ซึ่งตามหลักการแล้ว ไฟล์เสียงทั้งสองควรต้องจัดการให้เป็นข้อความเดียวกัน “Hello” ซึ่งการจัดการไฟล์เสียงทั้ง 2 ที่มีความยาวหลากหลายโดยอัตโนมัติให้เป็นข้อความที่มีความยาวคงที่นั้นทำได้ค่อนข้างยาก

เพื่อหลีกเลี่ยงปัญหาดังกล่าวจำเป็นต้องใช้เทคนิคพิเศษ และ processing เพิ่มเติมนอกเหนือจาก neural network กระบวนการที่แก้ไขปัญหามีขั้นตอนดังนี้

1. **Turning Sounds into Bits** ทำการป้อนคลื่นเสียงลงไปในคอมพิวเตอร์เพื่อเปลี่ยนคลื่นเสียงให้เป็นตัวเลข โดยวิธีการ sampling ทำการอ่านค่าเป็นพันๆ ครั้งต่อวินาที และบันทึกตัวเลขที่แสดงถึงความสูงของคลื่นเสียง ณ เวลานั้น



Sampled Digital Signal

2**. Pre-processing our Sampled Sound Data** ทำการจัดการกับข้อมูลที่ sampling มาและทำการแบ่งข้อมูลออกเป็นช่วงสั้นๆ แต่เนื่องจากข้อมูลเสียงนั้น มีความซับซ้อน เนื่องจากเป็นเสียงพูดของมนุษย์ซึ่งจะมีทั้งเสียงสูง เสียงกลาง เสียงต่ำ จึงต้องใช้เทคนิคพิเศษในการจำแนกเสียงและทำการแมพผลลัพธ์ตามความถี่ของเสียง (ต่ำ - สูง) และนำโมเดลที่ได้เข้าสู่ neural network ทำให้ neural network สามารถค้นหารูปแบบของข้อมูลประเภทนี้ได้ง่ายกว่าคลื่นเสียงแบบดังเดิม

3. **Recognizing Characters from Short Sounds** นำเอาท์พุทธ์ที่ได้จาก neural network ไปเป็นอินพุทธ์ให้กับ recurrent neural network (มีหน่วยความจำที่สามารถทำนายอนาคตได้) เนื่องจากตัวอักษรแต่ละตัวที่ทำนายควรมีผลต่อโอกาสของตัวอักษรถัดไปที่จะเกิดขึ้น เช่นหากพูดว่า "HEL" ก็จะมีโอกาสมากที่คำถัดไปจะพูดว่า "LO" และมีโอกาสน้อยที่จะพูดสิ่งที่ไม่สามารถออกเสียงได้ถัดไป เช่น "XYZ" ดังนั้นการมีหน่วยความจำที่เก็บค่าก่อนหน้านี้จะช่วยให้ neural network ทำการคาดการณ์ที่แม่นยำขึ้น

1. งานวิจัยที่เกี่ยวข้อง (Related Works)

* **A Study on Automatic Speech Recognition**

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

* **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

1. ขอบเขตของโครงงาน (Scope)

* เพื่อตรวจสอบความถูกต้องของหนังสือเสียงว่าตรงตามข้อความในหนังสือมากน้อยเพียงใด
* เนื่องจากในการตรวจสอบความถูกต้องของหนังสือเสียง ใช้ไฟล์ text ของไฟล์หนังสือต้นฉบับมาเปรียบเทียบความเหมือนกับไฟล์ที่ได้จากการแปลงหนังสือเสียงเป็นไฟล์ text ดังนั้นหนังสือเสียงจำเป็นต้องมีไฟล์ text ต้นฉบับด้วย
* ไฟล์หนังสือเสียงที่ใช้ในการตรวจสอบความถูกต้อง ต้องไม่มี noise มากเกินไป
* เนื่องจากใช้งานบริการ API ของ speech recognition ตัวโปรแกรมจึงจำเป็นต้องเชื่อมต่อกับอินเทอร์เน็ต

1. การพัฒนาโครงงาน (Project Development)
   1. ขั้นตอนการพัฒนา (Methodology)
2. กำหนดขอบเขต เป้าหมาย และวัตถุประสงค์ของโครงงาน
3. วางแผนการดำเนินงาน
4. แบ่งการดำเนินงานออกเป็น 2 ส่วนหลักๆ ดังนี้
   * ส่วน Speech recognition
   * ส่วน word matching
5. ศึกษาและพัฒนาโปรแกรมในแต่ละส่วน
6. รวมระบบย่อยทั้ง 2 เข้าด้วยกัน
   1. การออกแบบ (Design)

**Speech recognition**

**Word matching**

* 1. แนวทางการทดสอบและการวัดประสิทธิภาพ (Test and Performance Evaluation Approaches )
* ทดสอบตัวโปรแกรมกับไฟล์หนังสือเสียงที่สร้างจากหนังสือเรื่องสั้นหนึ่งเรื่อง จำนวน 2 ชุด โดยแต่ละชุดผู้จัดทำคนละคนกัน
* ทดสอบประสทธิภาพของตัวโปรแกรม โดยการตรวจสอบความถูกต้องของหนังสือเสียงเรื่องเดิม จำนวน 2–3 ครั้ง เพื่อดูผลลัพธ์ที่ได้ในแต่ละครั้ง ว่าแตกต่างกันหรือไม่

1. แผนการดำเนินโครงงาน (Timeline)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ID | Task Name | Aug2020 | | | | Sep2020 | | | | Oct2020 | | | | Nov2020 | | | | Dec2020 | | | | Jan2021 | | | | |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 1 | ศึกษาทฤษฎีและผลงานทางวิชาการที่เกี่ยวข้อง |  | | | | | | | | | | | | | | | | | | | | | | | | |
| 2 | **ส่วน Speech Recognition** |  | | | | | | | | | | | | | | | | | | | | | | | | |
| 3 | ศึกษาการใช้งาน Speech recognition API |  | | | | | | | | | | | | | | | | | | | | | | | | |
| 4 | ศึกษาและพัฒนาส่วน File Handling |  | | | | | | | | | | | | | | | | | | | | | | | | |
| 5 | พัฒนาส่วน Audiobooks to text | (ช่องละ 1 สัปดาห์) -> | | | | | | | | | | | | | | | | | | | | | | | | |
| 6 | **ส่วน Word matching** |  | | | | | | | | | | | | | | | | | | | | | | | | |
| 7 | ออกแบบเกณฑ์การวัดค่าความถูกต้องหนังสือเสียง |  | | | | | | | | | | | | | | | | | | | | | | | | |
| 8 | พัฒนาส่วน Word matching |  | | | | | | | | | | | | | | | | | | | | | | | | |
| 9 | **รวมทั้ง 2 Module เข้าด้วยกัน** |  | | | | | | | | | | | | | | | | | | | | | | | | |
| 10 | พัฒนาระบบวัดความถูกต้องของหนังสือเสียง |  | | | | | | | | | | | | | | | | | | | | | | | | |
| 11 | ทดสอบและวัดประสิทธิภาพตัวโปรแกรม |  | | | | | | | | | | | | | | | | | | | | | | | | |

ตารางแสดงแผนการดำเนินโครงงาน “การตรวจสอบความถูกต้องของหนังสือเสียงโดยใช้ machine learning และ AI”

1. ประโยชน์ที่คาดว่าจะได้รับ (Expected Benefits)

เนื่องจากในปัจจุบันนี้มีการจัดทำหนังสือเสียง สำหรับผู้บกพร่องทางสายตาและผู้สูงอายุ หรือคนปกติที่ต้องการรับรู้เนื้อหาของหนังสือผ่านทางการฟัง ซึ่งหนังสือเสียงที่จัดทำอาจมีความผิดพลาดของการออกเสียงหรือการสะกดคำที่ทำให้สารที่ได้รับนั้นมีความหมายผิดไป โดยปัญหาที่พบสำหรับการตรวจสอบความถูกต้องหนังสือเสียงที่เสร็จแล้วด้วยบุคคลนั้นมีข้อจำกัดด้านเวลาและความแม่นยำในการตรวจสอบ

โครงงานนี้จึงจัดทำขึ้นเพื่อมุ่งเน้นที่จะแก้ปัญหาดังกล่าว โดยการใช้เทคโนโลยี speech recognition มาใช้เพื่อแปลงเสียงพูดเป็นข้อความและนำข้อความที่ได้ไปเปรียบเทียบกับข้อความในหนังสือต้นฉบับ เพื่อทำการตรวจสอบความถูกต้องของหนังสือเสียง และนอกจากโครงงานนี้จะเป็นประโยชน์ต่อผู้จัดทำหนังสือเสียงในการช่วยลดเวลาในการตรวจสอบความถูกต้องแล้วยังเป็นประโยชน์ต่อผู้อ่านในทางอ้อม โดยหนังสือเสียงที่ผ่านการตรวจสอบความถูกต้องแล้วนั้นจะมีคุณภาพมากยิ่งขึ้นและสามารถนำโครงงานนี้ไปใช้เพื่อพัฒนาต่อได้

1. เอกสารอ้างอิง (Reference)
2. Lit2Go, <https://etc.usf.edu/lit2go/>
3. Machine Learning is Fun Part 5, “Language Translation with Deep Learning and the Magic of Sequences”,

<https://medium.com/@ageitgey/machine-learning-is-fun-part-5-language-translation-with-deep-learning-and-the-magic-of-sequences-2ace0acca0aa>

1. Machine Learning is Fun Part 6, “How to do Speech Recognition with Deep Learning”, <https://medium.com/@ageitgey/machine-learning-is-fun-part-6-how-to-do-speech-recognition-with-deep-learning-28293c162f7a>
2. Gamal Bohouta, Veton Z Këpuska, “Comparing Speech Recognition Systems (Microsoft API, Google API And CMU Sphinx)”, Florida Institute of Technology, 2017, <https://www.researchgate.net/publication/314938892_Comparing_Speech_Recognition_Systems_Microsoft_API_Google_API_And_CMU_Sphinx>
3. Nenny Anggraini, Angga Kuniawan, Luh Kesuma Wardhani, Nashrul Hakiem, “Speech recognition application for the speech impaired using the android-based google cloud speech API”, Syarif Hidayatullah State Islamic University Jakarta, 2018, <https://www.researchgate.net/publication/329865231_Speech_recognition_application_for_the_speech_impaired_using_the_android-based_google_cloud_speech_API>
4. Convert an AUDIO FILE into TEXT using Google Speech Recognition in Python, <https://www.youtube.com/watch?v=l-eD12KDmRc>
5. Multi-Lingual Speech To Text Conversion | Google Speech Recognition API | Python, <https://www.youtube.com/watch?v=vGV-6UeUNk8&feature=youtu.be>