

**A Text-to-speech Mobile Application for visually impaired people (READ2U)**

**แอปพลิเคชันในการช่วยอ่านข้อความเป็นเสียงสำหรับช่วยเหลือผู้พิการทางสายตา**

**นางสาวเจษฎาภรณ์ พุทธโกศัย**

**Jedsadaporn Puttakosai**

**นางสาวต่วนอะห์ลัม ต่วนสนิ**

**Tuanahlam Tuansani**

**โครงงานนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรบัณฑิต**

**ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์**

**มหาวิทยาลัยศรีนครินทรวิโรฒ ปีการศึกษา พ.ศ.2564**



**คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ**

**ชื่อหัวข้อโครงงาน A Text-to-speech Mobile Application for visually impaired people (READ2U)**

**แอปพลิเคชันในการช่วยอ่านข้อความเป็นเสียงสำหรับช่วยเหลือ**

**ผู้พิการทางสายตา**

**คณะผู้จัดทำ** นางสาวเจษฎาภรณ์ พุทธโกศัย 61102010139

นางสาวต่วนอะห์ลัม ต่วนสนิ 61102010424

**ปริญญา** วิทยาศาสตรบัณฑิต (วท.บ.)

**ภาควิชา** วิทยาการคอมพิวเตอร์

**อาจารย์ที่ปรึกษา** ผศ.ดร. ศุภชัย ไทยเจริญ

ลงชื่อ............................................................

(ผศ.ดร. ศุภชัย ไทยเจริญ)

อาจารย์ที่ปรึกษาโครงงาน

**บทคัดย่อ**

โดยทั่วไป ผู้พิการทางสายตาจะได้รับการเข้าถึงข้อมูลต่างๆ การอ่าน เขียน ผ่านอักษรเบรลล์ หรือเรียนรู้จากเสียง สื่ออิเล็กทรอนิกส์ เช่น โทรศัพท์ ดังนั้นเพื่อช่วยผู้พิการให้สามารถรับรู้ข้อมูลต่างๆได้อย่างง่ายดาย สะดวกและรวดเร็วและมีประสิทธิภาพโครงงานนี้จึงภูมิใจนำเสนอแอปพลิเคชั่น A TEXT TO SPEECH MOBILE APPLICATION FOR VISUALLY IMPAIRED ซึ่งแอปพลิเคชั่นนี้ได้ใช้อัลกอริทึม Tesseract ซึ่งเป็นเครื่องมือในการแปลงข้อความออกจากรูปภาพ (optical character recognition: OCR) โดยเมื่อผู้ใช้งานถ่ายภาพข้อความในหนังสือ ภาพที่ได้จะถูกประมวลผลเพื่อวิเคราะห์หาข้อความที่อยู่บนรูปภาพจากนั้น text-to-speech engine จะทำการอ่านข้อความนั้นออกมา

**Abstract**

In general, the visually impaired have access to information through Braille, electronic media such as telephones, computer, and sound. Therefore, to help those with disabilities to be able to receive information fast, conveniently, and efficiently, this project is proud to present A-TEXT-TO-SPEECH MOBILE APPLICATION FOR THEVISUALLY IMPAIRED. This app apply the Tesseract algorithm, which is a tool to convert image to text (optical character recognition: OCR). When the user takes a picture of the text in a book the resulting image will be processed to analyze the text on the image. The text-to-speech engine then reads the text.

**กิตติกรรมประกาศ**

ขอขอบคุณภาควิชาวิทยาการคอมพิวเตอร์ สำหรับเครื่องมือ เครื่องใช้ และอุปกรณ์ต่าง ๆ ที่ใช้ในการ ดำเนินงานในครั้งนี้ รวมถึงสถานที่ที่ใช้ในการทำงาน ใช้ในการทดสอบอุปกรณ์ต่าง ๆ ขอขอบพระคุณอาจารย์ที่ปรึกษาโครงงานอาจารย์ศุภชัย ไทยเจริญและอาจารย์ประจำภาควิชาวิทยาการคอมพิวเตอร์ทุกท่าน ที่ได้ให้ความรู้ คำปรึกษา และความช่วยเหลือในการสนับสนุนการดำเนินโครงงานในครั้งนี้จนสามารถจัดทำโครงงานจนประสบความสำเร็จลุล่วงไปได้ด้วยดีและมีประสิทธิภาพ

ขอขอบคุณมหาวิทยาลัยศรีนครินทรวิโรฒ ที่อำนวยความสะดวกในเรื่องของสถานที่สำหรับการ ดำเนินโครงงาน และเป็นแหล่งศึกษาหาความรู้เพิ่มเติม เพื่อนำความรู้ที่ได้จากการศึกษาไปพัฒนาระบบการ ทำงานภายในโครงงานและปรับปรุงระบบการทำงานให้ดียิ่งขึ้น

ขอขอบคุณเพื่อน คณะผู้จัดทำในสาขาวิทยาการคอมพิวเตอร์ทุกท่านและขอขอบพระคุณทุกท่านที่มีส่วนเกี่ยวข้องในการดำเนินโครงงานในครั้งนี้ ทั้งที่คอยให้คำแนะนำ ให้คำปรึกษา และให้การช่วยเหลือในเรื่องต่าง ๆ  ซึ่งทำให้การจัดทำโครงงานนี้สำเร็จลุล่วงไปได้ด้วยดี

สุดท้ายนี้ทางคณะผู้จัดทำหวังว่าโครงงานแอปพลิเคชันช่วยในการอ่านข้อความเป็นเสียงเพื่อช่วยเหลือผู้พิการทางสายตา (A Text-to-speech Mobile Application for visually impaired people : READ2U) จะเป็นประโยชน์อย่างมากต่อผู้พิการทางสายตาในการรับรู้ถึงข้อมูล และสำหรับผู้ที่สนใจเพื่อที่จะสามารถนำไปพัฒนาต่อยอดในอนาคตได้ดียิ่ง ๆ ขึ้นไป หากมีข้อผิดพลาดประการใด คณะผู้จัดทำต้องขออภัย และน้อมรับไว้ ณ ที่นี้ด้วย

**สารบัญ**

เรื่อง หน้า

**บทคัดย่อ**

**Abstract**

**กิตติกรรมประกาศ**

**สารบัญ**

**สารบัญรูปภาพ**

**สารบัญตาราง**

**บทที่** 1 **บทนำ**

* 1. ที่มาและความสำคัญของโครงงาน
  2. วัตถุประสงค์ของโครงงาน
  3. สมมติฐาน
  4. ขอบเขตของโครงงาน
  5. ประโยชน์ที่คาดว่าจะได้รับ

**บทที่** 2 **องค์ความรู้ที่เกี่ยวข้อง**

2.1. ทฤษฎีที่เกี่ยวข้อง

2.1.1. ผู้พิการทางสายตา

2.1.2. Optical Character recognition (OCR)

2.1.3. Tessearct OCR

2.1.4. Text To Speech (TTS)

2.1.5. Open Source Computer Vision (OpenCV)

2.1.6. Ngrok

2.1.7. Flask Framework

2.1.8. API

2.1.9. **REST และ RESTful**

2.2. งานวิจัยที่เกี่ยวข้อง

2.2.1. ชื่อเรื่อง : Document Image Skew Adjusting Based on the Feedback I Information Recognized byOCR

2.2.2. ชื่อเรื่อง : Smart Specs: Voice Assisted Text Reading system for Visually Impaired Persons Using TTS Method

2.2.3. ชื่อเรื่อง : Text Reader for Visually Impaires People :ANY READER

2.2.4. ชื่อเรื่อง : OCR Based Image Text To Speech Conversion Using MATLAB

2.2.5. ชื่อเรื่อง : Challenges in input preprocessing for mobile OCR applications: A realistic testing scenario

2.2.6. ชื่อเรื่อง : การพัฒนาแอปพลิเคชันระบุธนบัตรไทยด้วยเสียงสำหรับผู้พิการทางสายตาผ่านสมาร์ทโฟน

2.2.7. ชื่อเรื่อง : AI based Reading System for Blind using OCR

2.2.8. ชื่อเรื่อง : Efficient Portable Camera Based Text to Speech Converter for Blind Person

**บทที่** 3  **วิธีดำเนินโครงงาน**

3.1. วิธีดำเนินงานวิจัย

3.1.1.เริ่มต้นและวางแผนโครงการ

3.1.2.การวิเคราะห์เครื่องมือที่ใช้

3.1.3.การพัฒนา

3.1.4.การทดลอง

3.1.5.สรุปผล

3.2. อุปกรณ์และเครื่องมือ

3.2.1 ฮาร์ดแวร์ (Hardware)

3.2.2 ซอฟต์แวร์ (Software)

3.2.3. ภาษาที่ใช้ในการพัฒนา (Language)

3.3. การออกแบบระบบ

3.4.

3.5.

3.6.

**บทที่** 4 **ผลการดำเนินโครงงาน**

4.1.

4.2.

4.3.

**บทที่** 5 **สรุปผล อภิปรายผล และ ข้อเสนอแนะ**

5.1. สรุปผลการดำเนินงาน

5.2 ปัญหาและอุปสรรคในการดำเนินงาน

**บรรณานุกรม**

**ภาคผนวก**

**สารบัญรูปภาพ**

รูปภาพประกอบ 1

รูปภาพประกอบ 2

รูปภาพประกอบ 3

รูปภาพประกอบ 4

รูปภาพประกอบ 5

รูปภาพประกอบ 1

รูปภาพประกอบ 2

รูปภาพประกอบ 3

รูปภาพประกอบ 4

รูปภาพประกอบ 5

รูปภาพประกอบ 1

รูปภาพประกอบ 2

รูปภาพประกอบ 3

รูปภาพประกอบ 4

รูปภาพประกอบ 5

**สารบัญตาราง**

ตาราง 1

ตาราง 2

ตาราง 3

**บทที่ 1**

**บทนำ**

* 1. ที่มาและความสำคัญของโครงงาน

ในปัจจุบันผู้พิการทางสายตา หรือผู้ที่การมองเห็นไม่ชัดเจน ส่งผลต่อการดำรงชีวิต การมองไม่เห็นทำให้ขาดการรับรู้ทางภาพ การอ่าน แต่ในเมื่อพุทธศักราช 2364 ได้ประดิษฐ์คิดค้นอักษรเบรล์ และจุดกำเนิดในไทย ในปีพุทธศักราช 2482 ผู้พิการทางสายตา จะได้รับการสอนให้ใช้อักษรเบรลล์เพื่อเป็นสื่อในการอ่าน เขียนหนังสือ ในการเข้าถึงข้อมูล และยังสามารถเรียนรู้ผ่านสื่ออื่นๆ โดยจากเสียง สื่ออิเล็กทรอนิกส์ หรือคอมพิวเตอร์ และ อักษรที่ขยายใหญ่สำหรับคนที่เห็นเลือนรางได้ การมีอักษรเบรลล์ หรือสื่อต่างๆ ที่เข้ามาช่วยให้ผู้พิการทางสายตา ให้สามารถรับรู้และเข้าถึงข้อมูล แต่ก็ต้องใช้ระยะเวลาในการสอน และทำความเข้าใจในตัวอักษร อีกทั้งไม่ใช่หนังสือทุกเล่มจะมีอักษรเบรลล์ให้ผู้พิการทางสายตาอ่านเสมอไป

เทคโนโลยีที่ชื่อว่า Optical Character Regcognition หรือ OCR คือเทคโนโลยีการอ่านอักขระด้วยแสง ที่จะช่วยในการวิเคราะห์ข้อความที่อยู่บนรูปภาพ ให้ได้เป็นข้อความที่สามารถนำไปใช้งานในด้านต่างๆได้ เช่น การตรวจเช็คป้ายทะเบียนรถเข้าออกอาคาร หรือเช็คข้อมูลในหนังสือเดินทางเป็นต้น

ประกอบกับสมาร์ทโฟนในปัจจุบันเป็นสิ่งหนึ่งในการดำรงชีวิตที่สามรถช่วยตอบสนองความต้องการได้หลายอย่าง ทั้งการติดต่อสื่อสาร การชำระสินค้าต่างๆ การช่วยนำทางหรือGPS หรือแม้กระทั่งการรับรู้ข่าวสาร ซึ่งข้อมูลต่างๆที่อยู่บนโลกอินเทอร์เน็ตนั้น แค่มีเพียงสามาร์ทโฟน หรืออุปกรณ์สื่อสารอิเล็กทรอนิกส์ ก็สามารถให้เข้าถึงข้อมูลได้อย่าง ง่าย สะดวกและรวดเร็ว

คณะผู้จัดทำจึงได้มีการพัฒนาแอปพลิเคชันในการช่วยอ่านข้อความเป็นเสียงสำหรับผู้พิการทางสายตา (A Text-to-speech Mobile Application for visually impaired people :READ2U ) ที่ช่วยในการวิเคราะห์ แปลงข้อความที่อยู่บนรูปภาพ และสามารถอ่านให้ผู้พิการทางสายตาฟัง รับรู้ถึงข้อมูลที่อยู่ตรงหน้า

* 1. วัตถุประสงค์ของโครงงาน
* เพื่อศึกษาหลักการของ Computer Vision และระบบรู้จำตัวอักขระด้วยแสง (Optical character recognition : OCR) และการแปลงข้อความเป็นเสียง (Text-to-Speech : TTS)
* เพื่อประยุกต์ใช้หลักการข้อแรกสำหรับการแปลงข้อความออกมาเป็นเสียงและสามารถนำข้อความที่ได้ไปใช้งานได้
* เพื่อศึกษาหลักการของ Tesseract OCR เครื่องมือ Open Source ในการรู้จำตัวอักษร
* เพื่อพัฒนาแอปพลิเคชันสำหรับการแปลงข้อความออกมาเป็นเสียงเพื่อช่วยในการอ่านข้อมูลให้สำหรับผู้พิการทางสายตา
* เพื่อศึกษาและพัฒนาการสร้างระบบ API ในการรับส่งข้อมูลจากโทรศัพท์ไปวิเคราะห์ประมวลผลได้และตอบกลับข้อความได้
  1. สมมติฐาน
* คาดว่าแอปพลิเคชันจะสามารถวิเคราะห์ข้อความที่อยู่บนรูปภาพออกมาได้
* คาดว่าแอปพลิเคชันจะสามารถอ่านข้อความที่วิเคราะห์ออกมาได้
* คาดว่าแอปพลิเคชันจะสามารถช่วยผู้พิการทางสายตาสามารถรับรู้ถึงข้อมูลที่ต้องการได้
  1. ขอบเขตของโครงงาน
* พัฒนาแอปพลิเคชันโดยมีการนำ Tesseracr OCR มาทำการวิเคราะห์ข้อความบนรูปภาพ
* พัฒนาแอปพลิเคชันที่รองรับข้อความบนรูปภาพที่เป็นภาษาไทยและภาษาอังกฤษได้
* พัฒนาApiในการรับส่งข้อมูลเพื่อทำการวิเคราะห์ภาพตอบกลับมายังแอปพลิเคชันได้
  1. ประโยชน์ที่ควาดว่าจะได้รับ
* แอปพลิเคชันสามารถวิเคราะห์แปลงข้อความที่อยู่บนรูปภาพได้
* แอปพลิเคชันสามารถอ่านข้อความออกมาเป็นเสียงได้
* แอปพลิเคชันสามารช่วยผู้พิการทางสายตาสามารถรับรู้ถึงข้อมูลได้

**บทที่ 2**

**องค์ความรู้ที่เกี่ยวข้อง**

2.1. ทฤษฎีที่เกี่ยวข้อง

2.1.1. ผู้พิการทางสายตา(Visually impaired people)

คือผู้ที่บกพร่องทางด้านการมองเห็น ดวงตาไม่สามารถมองเห็นสิ่งต่างๆหรือรับรู้ความสว่างและความมืดได้ ไม่สามารถใช้ชีวิตประจำวันได้อย่างคนทั่วไป ซี่งความพิการทางการมองเห็นสามารถแบ่งได้เป็น 2 ประเภท ได้แก่ 1) ตาบอดเลือนราง(Low Vision) หมายถึงการที่มีความบกพร่องทาง การเห็นสามารถมองเห็นได้บ้างแต่ไม่เท่ากับคนปกติ และ 2) ตาบอดสนิท (Blindness) หมายถึงการมองเห็นได้น้อยมากจนไม่สามารถมองเห็นอะไรได้เลย

|  |  |  |
| --- | --- | --- |
| **ระดับความพิการทางสายตา** | | **ระดับความสมบูรณ์ของการมองเห็น** |
| สายตาเลือนราง  (Low Vsion) | ระดับ 1 | สามารถมองเห็นได้ 100.0% |
| สามารถมองเห็นได้ 87.5 |
| สามารถมองเห็นได้ 64.0 |
| สามารถมองเห็นได้ 48.9 |
| ระดับ 2 | สามารถมองเห็นได้ 20.0 |
| สายตาบอด  (Blindness) | ระดับ 3 | สามารถมองเห็นได้ 10.0 |
| ระดับ 4 | สามารถมองเห็นได้ 2.0 |
| ระดับ 5 | มองไม่เห็นแม้แสงสว่าง |

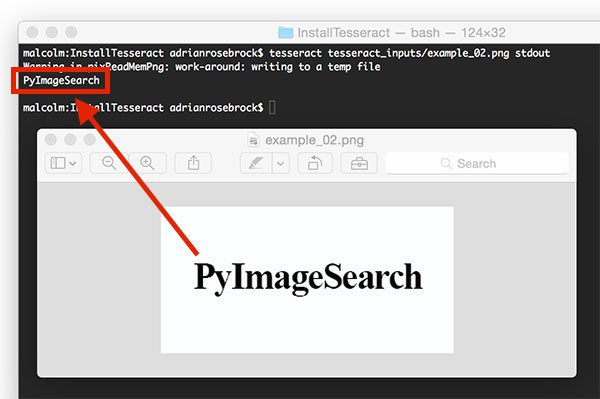
ตารางที่ แสดงระดับความพิการทางสายตาที่กำหนดโดยองค์การอนามัยโลก

2.1.2. Optical Character recognition (OCR)

คือกระบวนการในการเปลี่ยนข้อความที่อยู่ในรูป ให้อยู่ในรูปแบบของข้อความ หรือเป็นการสร้างข้อความ ไฟล์เอกสารจากไฟล์สแกนเอกสารทั้งในรูปแบบการพิมพ์ หรือลายมือ

2.1.3. Tessearct OCR

เครื่องมือที่ใช้ในการรู้จำตัวอักษร Tesseract มีความสามารถในการเรียนรู้จดจำได้มากกว่า 100 ภาษา และเป็น Open Source ที่มี Community ที่ค่อนข้างใหญ่ โดยปัจจุบัน Google ได้ใช้ Tesseract ในงาน Text Detection บนมือถือ,วีดีโอ และการตรวจจับสแปมรูปภาพ Gmail



Tesseract ข้างในก็คือ Machine Learning โดยผู้พัฒนามี Model ให้ผู้ใช้เลือกทั้งหมด 3 โมเดล ซึ่งแต่ละโมเดลมี Weights ที่แตกต่างกันที่ส่งผลให้ได้ผลลัพธ์และความเร็วที่แตกต่างกันด้วย โดยอัลกอริทึมข้างในมีให้ใช้อยู่ 2 แบบ LSTM (อัลกอริทึมใหม่) และ Legacy สามารถสรุปความแตกต่างของแต่ละโมเดลได้ดังนี้

Table

Description automatically generated

2.1.4. Text To Speech (TTS)

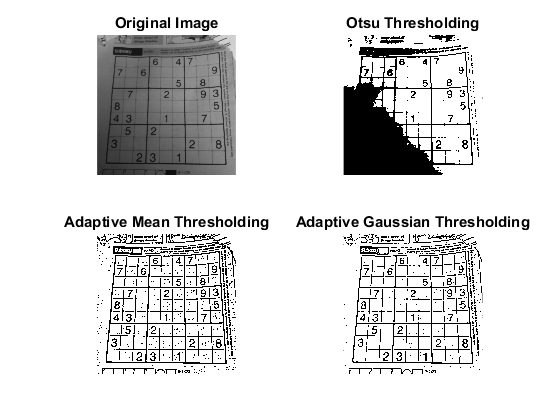
เป็นการส่งข้อความให้ระบบแปลงข้อความออกมาให้กลายเป็นเสียง โดยอุปกรณ์แอนดรอยด์แต่ละยี่ห้อก็อาจจะมี TTS Engine แตกต่างกันออกไป ซึ่งรองรับภาษาส่วนใหญ่ทั่วโลก รวมไปถึงภาษาไทย



2.1.5. Open Source Computer Vision (OpenCV)

เป็นไลบารีสำหรับผู้พัฒนาซอฟต์แวร์ให้สามารถประมวลผลภาพ หรือ วิดีโอได้โดยจะมุ่งเป้าไปที่การแสดงผลด้วยคอมพิวเตอร์แบบเรียลไทม์ สามารถใช้ในการพัฒนาระบบOpen Source ที่นำมาใช้ สร้าง Machine Learning หรือ AI ในการจำเเนกวัตถุ หรือการจดจำใบหน้าได้

* Image Processing
  + Adaptive Threshoding คือการกำหนดจุดแบ่งให้สามารถปรับเปลี่ยนหรือยืดหยุ่นค่าได้โดยอ้างอิงตามบริเวณพื้นที่ใกล้เคียงกัน ปรับรูปแบบของภาพให้เหมาะสม กำหนดค่า Adaptive method รูปแบบ cv.ADAPTIVE\_THRESH\_GAUSSIAN\_C เป็นการคำนวณค่าเฉลี่ยรอบๆBlockSize และอ้างอิงทำการทำงานกับฟังก์ชัน Gaussian และค่า C



* + Dilation คือ การขยายภาพเป็นลักษณะของการเพิ่มข้อมูลภาพตามลำดับตลอดทั้งภาพ โดยจะเป็นเพิ่มส่วนสีขาวหรือขนาดของวัตถุเบื้องหน้า (Foreground)

A picture containing text, crossword puzzle

Description automatically generated

* + Erosion คือ การย่อภาพเป็นลักษณะของการลบข้อมูลภาพบริเวณขอบของภาพ โดยทั้งหมดที่อยู่ใกล้เคียงนั้นจะโดนลบออกไปตามขนาดของ kernel ดังนั้นความหนาหรือขนาดของวัตถุข้างหน้าจะลดลง(หรือพื้นที่ที่มีสีขาวก็จะลดลง) และสิ่งนั้นจะเป็นประโยชน์ด้านการลด noises ออก

A picture containing text, crossword puzzle, colors

Description automatically generated

2.1.6. Ngrok

เป็น Tool Open Source พัฒนาโดย GitHub ซึ่งอำนวยความสะดวกให้บุคคลอื่นสามารถเข้าใช้งาน Website หรือ Application ที่กำลังทำงานอยู่บนเครื่อง Localhost โดยบุคคลอื่นสามารถเข้าใช้งาน Website หรือ Application กำลังทำงานอยู่บนเครื่อง Localhost ผ่านทาง URL ของทาง Ngrok โดยที่ทาง Ngrok จะทำการสุ่มสร้าง URL ขึ้นมา และ URL ที่ได้มานั้น จะทำการเปลี่ยนไปทุกครั้งเมื่อมีการปิดหรือเปิดใช้งาน Ngrok

Text

Description automatically generated

2.1.7. Flask Framework

Framework ที่ใช้ในการสร้าง Web Application ที่รองรับ Web Server Gateway Interface (WSGI) Flask เป็น Framework ที่ออกแบบให้ Lightweight มีขนาดเล็ก ไม่ซับซ้อน มีฟังก์ชันเท่าที่จำเป็นเท่านั้น

**Flask-RESTful**

Flask-RESTful เป็นส่วนขยายของ Flask สำหรับสร้าง REST API

2.1.8. API

API ย่อมาจาก Application Programming Interface คือการเชื่อมต่อจากระบบหนึ่งไปสู่อีกระบบหนึ่ง เพื่อให้ซอฟต์แวร์ภายนอกเข้าถึงและอัพเดทข้อมูลนั้นๆได้ แต่ยังอยู่ในขอบเขตที่ถูกกำหนดไว้ หรือจะบอกให้ง่ายขึ้นก็คือ API เป็นตัวกลางที่จะทำให้คอยรับคำสั่งต่าง ๆ ประมวลผลและกระทำข้อมูลส่งกลับคืนไปยังคนสั่งโดยอัตโนมัติ

2.1.9. **REST และ RESTful**

**REST ย่อมาจาก Representational state transfer**  เป็นรูปแบบสถาปัตยกรรมซอฟต์แวร์(architecture) ที่ใช้ประโยชน์จากเทคโนโลยี Web Protocol เพื่อใช้ในการสร้าง Web Service

RESTful – RESTful Web Service(RWS) คือ Web Service ที่ใช้สถาปัตยกรรม Rest ซึ่งเจ้าตัว RWS อนุญาต ให้ระบบ Request และเข้าถึง **Resource** บนเว็บโดยใช้ชุดคำสั่งที่กำหนดเอาไว้ล่วงหน้า โดยที่การโต้ตอบของระบบที่ใช้ **REST** จะอยู่บนพื้นฐานของ **Hypertext Transfer Protocol (HTTP). Request** จะส่งคำขอไปยัง URI ที่กำหนด และล้วงเอา response กลับมาเป็น Payload ในแบบ HTML, XML, JSON หรือ format อื่น ๆ โดย RESTful จะประกอบไปด้วย

Client – ผู้ที่เข้ามาเป็น Request Resource  
Server – ผู้ที่ให้บริการ Resource

HTTP Method ที่สำคัญ

1. Get เป็นการร้องขอข้อมูลจาก resource
2. Post เป็นการสร้างข้อมูลใหม่ใน resource
3. Put เป็นการอัพเดตข้อมูลที่มีอยู่แล้ว หรือสร้างใหม่ resource
4. Delete เป็นการลบข้อมูลที่มีอยู่แล้วใน resource

2.2. งานวิจัยที่เกี่ยวข้อง

2.2.1. ชื่อเรื่อง : Document Image Skew Adjusting Based on the Feedback Information Recognized byOCR

ผู้แต่ง : Xiaoling Fu, Yazhuo Xu, Lijing Tong

ปีที่พิมพ์ : 2011

บทความนี้นำเสนอการตรวจจับภาพเอียงให้ได้ภาพที่ตรง การปรับภาพเอียง(skew) ถือเป็นขั้นตอนที่สำคัญมากในการประมวลผลภาพเพื่อทำ OCR เนื่องจากภาพเอียงอาจทำให้อักขระผิดรูป ปัจจุบันการจัดเก็บเอกสารแบบดิจิทัลนั้นแทบจะกลายเป็นเรื่องปกติธรรมดา อย่างไรก็ตามเมื่อเราสแกนหรือถ่ายภาพเอกสารเอกสารหรือผลลัพธ์ของรูปภาพนั้นจะเกิดการเบ้อย่างหลีกเลี่ยงไม่ได้เลย ดังนั้นเมื่อภาพที่ได้มาดำเนินการโดยกระบวนการ OCR เลยจะทำให้เกิดข้อผิดพลาดขนาดใหญ่และลดอัตราการจดจำอักขระลง ดังนั้นจึงจำเป็นต้องปรับภาพเอกสารเอียงให้ได้ผลลัพธ์ที่ดีที่สุดเสียก่อน ในบทความนี้ได้แนะนำอัลกอริทึมการแก้ไขภาพเอียงของเอกสาร 2แบบ วิธีการแรกวิธีการเหมาะสมกับภาพตามพารามิเตอร์ที่ป้อนกลับเข้าสู่กระบวนการ OCR และใช้วิธีการแก้ไขเชิงเส้นสองครั้งครั้งเพื่อให้แน่ใจว่ามีการชดเชยพิกเซลในขณะหมุมการแก้ไข ผลการทดลองระบุว่าอัลกอริทึมในบทความนี้มีความแม่นยำสูงเมื่อใช้เพื่อแก้ไขภาพเอกสารที่บิดเบี้ยว ข้อได้เปรียบสูงสุดของอัลกอริทึมคือสามารถแก้ไขรูปภาพในเอกสารที่มีคุณสมบัติแตกต่างกันได้อย่างเหมาะสมที่สุด ซึ่งทำให้ภาพมมีประสิทธิภาพสูงขึ้น

Text, letter

Description automatically generated

รูปภาพประกอบ ผลลัพธ์จากการแก้ไขภาพโดยใช้อัลกอริทึมในบทความนี้

2.2.2. ชื่อเรื่อง : Smart Specs: Voice Assisted Text Reading system for Visually Impaired Persons Using TTS Method

ผู้แต่ง : Ani R, Effy Maria, J Jameema Joyce, Sakkaravarthy V, Dr.M.A.Raja

ปีที่พิมพ์ : 2017

บทความนี้ได้นำเสนอการพัฒนาระบบอ่านข้อความสำหรับผู้พิการทางสายตา โดยใช้ราสเบอร์รี่พาย(Raspberry Pi) ซึ่งสามารถจับภาพแบบเรียไทม์อย่างรวดเร็วและแปลงเข้ารหัสเป็นไฟล์ JPEG อย่างไรก็ตามการจับภาพจากวิดีโอลักษณะแบบนี้ภาพอาจจะไม่ได้มีความละเอียดและขนาดที่ต้องการ การถ่ายภาพด้วยวิดีโอมักจะได้ภาพแตกมากกว่าภาพที่ถ่ายภาพนิ่ง ดังนั้นจำเป็นต้องใช้กระบวนการประมวลภาพมากขึ้น บทความนี้เลือกใช้ราสเบอรรี่พาย( Raspberry Pi) เนื่องจากเป็นการจับภาพที่เร็วที่สุดที่เคยมีมา ปัญหาอีกประการหนึ่งของราสเบอรรี่พาย( Raspberry Pi) นั้นจำกัด รูปแบบการดึงรูปภาพทำให้กระบวนนี้มีประสิทธิภาพน้อยลง นอกจากนี้ หากขนาดการ์ด SD ไม่ใหญ่พอการ์ดจะไม่สามารถเก็บภาพทั้งหมดที่จับโดยกล้องได้ บทความได้นำเสนอการรู้จำอักขระด้วยแสงโดยใช้ Tesseract ซึ่งเป็นเครื่องมือ open source หนึ่งทำหน้าที่แปลงข้อความจากภาพเป็นข้อความแก้ไขได้ที่รองรับภาษาต่างๆทั่วโลกมากกว่า 60 ภาษา และใช้ eSpeak สำหรับการแปลงข้อความออกมาเป็นเสียง eSpeak เป็นอีกหนึ่งเครื่องมือ open source ขนาดกะทัดรัดที่รองรับภาษาอังกฤษและอีกหลายภาษาข้อได้เปรียบหลักของการใช้ eSpeak คือคำพูดมีความชัดเจนและสามารถใช้ความเร็วได้

2.2.3. ชื่อเรื่อง : Text Reader for Visually Impaires People :ANY READER

ผู้แต่ง : AVM Manikandan, Shouham Choudhury, Souptik Majumder

ปีที่พิมพ์ : 2017

บทความนี้ได้นำเสนอแอปพลิเคชันแอนดรอยอ่านข้อความสำหรับผู้พิการทางสายตา เพื่อให้ผู้พิการทางสายตามีสิทธิเข้าถึงข้อมูลได้สะดวกและรวดเร็วมากขึ้นอีกทั้งยังประหยัดค่าใช้จ่ายได้อีกด้วย บทความได้นำเสนอวิธีการโดยใช้ OpenCV ไลบารี่สำหรับการประมวลภาพ โดยทำการประมวลภาพขั้นละเอียด ได้แก่ Gray Scaling, Adaptive Thresholding, Detecting Hough lines, De-skew images, Morphological Operations, Contour detection เพื่อให้ได้ภาพที่มีประสิทธิภาพสูงสุดก่อนเข้าสู่กระบวนการรู้จำอักขระด้วยแสง( Optical Character Recognition) โดยการใช้ Tesseract OCR ซึ่งเป็นเครื่องมือ open source นึงทำหน้าที่แปลงข้อความจากภาพเป็นข้อความแก้ไขได้ที่รองรับภาษาต่างๆทั่วโลกมากกว่า 60 ภาษา จากนั้นใช้ Google Cloud Speech API เป็น API text to speech ที่มีสเถียรภาพและรองรับมากกว่าถึง 80 ภาษา สำหรับอ่านข้อความที่ได้จากภาพอ่านให้กับผู้ใช้

Diagram

Description automatically generated

รูปภาพประกอบที่ การทำงานของแอปพลิเคชัน

A page of a book

Description automatically generated with medium confidence

รูปภาพประกอบที่ ภาพตัวอย่างที่ได้จาการทำ image preprocessing

2.2.4. ชื่อเรื่อง : OCR Based Image Text To Speech Conversion Using MATLAB

ผู้แต่ง : Sneha.C.Madre. , Prof.S.B.Gundre.

ปีที่พิมพ์ : 2018

บทความนี้ได้นำเสนอวิธีการแปลงข้อความบนรูปภาพให้ออกเป็นรูปแบบเสียงด้วย MATLAB ทั้งหมด โดยขั้นตอนแรกคือการจับภาพโดยใช้เว็บแคม/กล้อง ตามด้วยการประมวลผลข้อความล่วงหน้า(pre-processing) ลำดับของงานที่เสนอจะเป็นดังนี้ ไบนารี(binarization), การสแกนด้วยแสง (optical scanning) การแยกคุณลักษณะ(feature extraction), การแบ่งส่วน(segmentation) และการรู้จำอักขระ(character recognition) บทความนี้นำเสนอถึงการดึงข้อความออกจากภาพโดยใช้การรู้จำอักขระด้วยแสง และอักขระที่รู้จักจะถูกแปลงเป็นเสียงโดยใช้ MATLAB และระบบเรียลไทม์ จุดประสงค์หลักของระบบนี้คือเพื่อตอบสนองความต้องการของผู้พิการทางเสียงและสามารถสื่อสารกับผู้ที่ไม่เข้าใจภาษามือได้ ความช่วยเหลืออีกประการหนึ่งคือสามารถเรียนรู้การออกเสียงคำศัพท์ได้อย่างถูกต้องโดย ระบบ TTS ใน MATLAB ยังสามารถใช้ในแอปพลิเคชันนี้โดยเฉพาะ เช่น ประกาศเกี่ยวกับรถไฟ ระบบสามารถใช้สร้างการท่องข้อมูลสำหรับผู้ที่ไม่สามารถเขียนและอ่านได้ บทความนี้ทั้งช่วยให้คนตาบอดอ่านเอกสารได้ และยังให้คนหูหนวกจะได้รับประโยชน์ในการแบ่งปันความรู้สึก/ความคิดเห็นจากหนังสือ

Diagram

Description automatically generated

รูปภาพประกอบ Block Diagram

A screenshot of a computer

Description automatically generated with medium confidence

รูปภาพประกอบ ผลลัพธ์การทำงาน

2.2.5. ชื่อเรื่อง : : Challenges in input preprocessing for mobile OCR applications: A realistic testing scenario

ผู้แต่ง : Nikolaos Sourvanos, Georgios Tsatiris

ปีที่พิมพ์ : 2018

บทความนี้นำเสนอการศึกษาปัญหาการทำงานของ OCR บนสมาร์ทโฟน แอปพลิเคชันบน สมาร์ทโฟนที่ใช้ OCR มีหลากหลายบทบาทแตกต่างกันไป มีตั้งแต่การป้อนข้อความอัตโนมัติจากรูปภาพและวิดีโอ และการแปลงข้อความเป็นคำพูด ไปจนถึงการแปลข้อความอัตโนมัติแบบเรียลไทม์ อย่างไรก็ตามในการออกแบบระบบ OCR แบบ end-to-end สำหรับสมาร์ทโฟนนั้นยังมีความยุ่งยากซับซ้อนเนื่องจากมีขั้นตอนและความท้าทายมากมายซึ่งอยู่ในตัว OCR เอง ซึ่งอาจพิสูจน์ได้ว่ายากและไม่สะดวกสำหรับอุปกรณ์มือถือทั่วไป ก่อนเข้าสู่กระบวน OCR จริง input จะต้องได้รับการประมวลผลล่วงหน้าก่อน เพื่อปรับปรุงข้อมูลที่เป็นข้อความที่มีอยู่ และลดเสียงรบกวนและค่าผิดปกติให้เหลือน้อยที่สุด

สิ่งที่เห็นได้ชัดจากการทดลองครั้งนี้ มีขั้นตอนที่ต้องปฏิบัติตามเพื่อให้ได้ผลลัพธ์ของ OCR ที่น่าพึงพอใจสำหรับแอปพลิเคชัน โดย input จะต้องทำการประมวลผลล่วงหน้าก่อน (pre-processing) ได้แก่ถูกแปลงเป็นภาพระดับสีเทา(grayscale) ปรับปรุงดีนอยส์(denoising) การเพิ่มคอนทราสต์(contrast) การเบลอ(blurring) และการกำหนดเกณฑ์(thresholding) อย่างไรก็ตามยังไม่มีแนวทางที่เป็นชัดเจนว่าการกระบวนการทำประมวลผล่วงหน้าก่อน (pre-processing)แบบไหนถึงจะได้ผลลัพธ์ที่ดีที่สุดเนื่องจากความแตกต่างของการคำนวณภาพบนสมาร์ทโฟน สภาพแวดล้อมและการใช้เทคนิคการประมวลผลล่วงหน้า(pre-processing)และจำเป็นที่จะต้องกำหนดขอบเขตของ input นั้น ในทางกลับกัน เราไม่สามารถจัดการ OCR แบบวิดีโอได้ผลลัพธ์อย่างพอใจ แต่ยังกำหนดรูปแบบและทิศทางดังกล่าวเพื่อให้ใช้งานได้อย่างกว้างขวางกับอุปกรณ์พกพาที่หลากหลาย

2.2.6. ชื่อเรื่อง : การพัฒนาแอปพลิเคชันระบุธนบัตรไทยด้วยเสียงสำหรับผู้พิการทางสายตาผ่านสมาร์ทโฟน

ผู้แต่ง : ณัฐวดี หงส์บุญมี และ กาญจนาแสงตาล

ปีที่พิมพ์ : 2019

งานวิจัยนี้มีวัตถุประสงค์เพื่อนำเสนอการพัฒนาแอปพลิเคชันช่วยระบุธนบัตรไทยด้วยเสียงสำหรับผู้พิการทางสายตาผ่านสมาร์ทโฟนระบบปฏิบัติการแอนดรอยด์ด้วยเทคโนโลยีการเรียนรู้เชิงลึกสำหรับการจำแนกภาพธนบัตรและเทคโนโลยี Text to Speech สำหรับการแปลงข้อความเป็นเสียงผ่านสมาร์ทโฟน การดำเนินงานเริ่มจากการรวบรวมกลุ่มตัวอย่างธนบัตรไทยจำนวน 2,700 ภาพ โดยการนำภาพมาวิเคราะห์และสร้างโมเดลจำแนกภาพด้วยโครงข่ายประสาทเทียมแบบคอนโวลูชันที่มีการใช้อัลกอริธึม MobileNet ผ่านไลบรารีTensorflow จากนั้นนำโมเดลไปพัฒนาเป็นแอปพลิเคชันบนสมาร์ทโฟน แอปพลิเคชันนี้พัฒนาด้วยโปรแกรม Android Studio ภาษา JAVA และไลบรารี่ Text to Speech ผลการวัดประสิทธิภาพโมเดลได้ค่าความถูกต้องเท่ากับ 95.00% ผลการทดสอบแอปพลิเคชันพบว่าสามารถจำแนกธนบัตรได้ถูกต้อง 84.00%ส่วนผลประเมินความพึงพอใจจากผู้ใช้งานพบว่ามคีวามพึงพอใจเฉลี่ยเท่ากับ4.33 ส่วนเบี่ยงเบนมาตรฐาน 0.56อยู่ ในระดับดี สามารถสรุปได้ว่าแอปพลิเคชันนี้มีประสิทธิภาพสามารถช่วยให้ผู้พิการทางสายตาทราบถึงชนิดธนบัตรได้ง่าย ขึ้น เนื่องจากสามารถระบุชนิดธนบัตรด้วยเสียงและอำนวยความสะดวกแก่ผู้พิการทางสายตาในการใช้งานได้ทุกที่ทุกเวลา ทั้งยังัช่วยส่งเสริมการดำเนินชีวิตสำหรับผู้พิการทางสายตาให้สามารถดำรงชีวิตขั้นพื้นฐานได้สะดวกเพิ่มมากขึ้น

2.2.7. ชื่อเรื่อง : AI based Reading System for Blind using OCR

ผู้แต่ง : Abhishek Mathur, Akshada Pathare, Prerna Sharma

ปีที่พิมพ์ : 2019

บทความนี้ได้นำเสนอแอปพลิเคชันช่วยในการถ่ายภาพหรือสแกนเอกสารโดยใช้กล้องของโทรศัพท์  รูปภาพจะถูกสแกนและแอปพลิเคชันจะอ่านข้อความที่เขียนเป็นภาษาอังกฤษทั้งหมดและแปลงข้อความออกมาเป็นรูปแบบคำพูด วัตถุประสงค์ของการส่งออกในรูปแบบของเสียง/คำพูดคือการให้บริการข้อมูลที่มีอยู่ในเอกสารไปยังผู้พิการทางสายตา โดยบทความนี้ได้แนะนำ MATLAB, LabVIEW ใช้ในการประมวลผลภาพล่วงหน้าจากนั้นบทความนี้ได้เสนอกระบวนการ OCR โดยใช้อุปกรณ์ ราสเบอร์รี่ พาย และใช้กล้องเว็บที่เชื่อมต่อกับไมโครคอนโทรลเลอร์ ARM ผ่าน USB โดย Tesseract สำหรับโมดูล OCR และใช้ Flite เป็นไลบรารี่สำหรับการแปลงข้อมูลเป็นเสียง

A picture containing text, newspaper

Description automatically generated

รูปภาพประกอบ ผลลัพธ์การรู้จำอักขระด้วยแสง

2.2.8. ชื่อเรื่อง : Efficient Portable Camera Based Text to Speech Converter for Blind Person

ผู้แต่ง : Trupti Shah, Sangeeta Pashionikar

ปีที่พิมพ์ : 2019

บทความนี้ได้นำเสนอการจดจำลำดับของรูปภาพในแบบเรียลไทม์ทั่วไป เช่นข้อความฉากธรรมชาติ การเขียนด้วยลายมือ จะเกิดขึ้นตามลำดับไม่ใช่ในรูปแบบที่แยกออกมา ผลงานหลักของบทความนี้คือการจดจำข้อความที่มีอยู่ในวัตถุใดๆ โดยใช้ OCR และ CRNN ร่วมกัน มีการเสนอ Raspberry pi รุ่น 3b ซึ่งใช้ประโยชน์จากการประมวลผลหลายตัว เช่น จับภาพ ประมวลผล และสุดท้ายเป็นเอาต์พุตเสียง ซึ่งทำให้ตาบอดในการอ่านภาพด้วยลายมือและภาพที่เป็นธรรมชาติ บทความนี้เสนอแนวทางที่ดีกว่าสำหรับการแปลงข้อความและการแยกข้อความบนรูปภาพ ขนาดข้อความเป็นปัจจัยสำคัญที่ควรเลือกมิติข้อมูลอย่างเหมาะสม เพื่อทำให้วิธีการเป็นแบบทั่วไปมากขึ้นและไม่คำนึงถึงรูปร่างและขนาดแบบอักษรต่างๆ วิธีการที่นำเสนอนี้เกี่ยวข้องกับการตรวจจับวัตถุสี่ขั้นตอน การแปลข้อความด้วยเทคนิคก่อนการประมวลผลข้อความ การดึงข้อความจากภาพและการแปลงข้อความเป็นคำพูด ขอบเขตที่สนใจถูกดึงออกมาจากพื้นหลังที่รก จากนั้นจึงนำอัลกอริธึมการแปลข้อความไปใช้เพื่อค้นหาและแยกข้อความ หลังจากแยกข้อความแล้ว ข้อความนั้นจะถูกแปลงเป็นคำพูด ทำงานได้อย่างมีประสิทธิภาพมากขึ้นด้วยการรู้จำอักขระด้วยแสง มีการเสนอโครงข่ายประสาทเทียมแบบ Convolutional สำหรับฝึกคำศัพท์แยกกัน การทดลองและการฝึกอบรมดำเนินการกับชุดข้อมูล Synth 90k word ชุดของรูปภาพได้รับการทดสอบและใช้วิธีเส้นขอบตามที่กำหนดไว้ใน OCR แต่ละคำจะถูกดึงออกมา มีการฝึกอบรมคำศัพท์แต่ละคำ ในที่สุดการใช้ OCR และ CRNN ก็ได้พัฒนาแบบจำลองรวมกัน

**บทที่ 3**

**วิธีดำเนินโครงงาน**

3.1. วิธีดำเนินงานวิจัย

3.1.1. เริ่มต้นและวางแผนโครงงาน

* ศึกษาหาข้อมูลแนวคิดและทฤษฎีที่ใช้ในการทำโครงงาน
* เลือกเครื่องมือ
* กำหนดรายละเอียดของโครงงาน

3.1.2. การวิเคราะห์เครื่องมือที่ใช้

* ศึกษาไลบารีและอุปกรณ์ที่นำมาใข้งาน
* ศึกษาการทำงานของไลบารี

3.1.3. การพัฒนา

* ทดสอบการทำงานของไลบารีเพื่อนำมาวิเคราะห์
* ทดสอบการทำงานของอุปกรณ์ที่นำใช้งาน
* หาข้อผิดพลาดและแก้ไขปรับปรุง
* พัฒนาระบบให้สามารถใช้งานร่วมกันได้
* พัฒนาระบบให้มีประสิทธิภาพมากขึ้น

3.1.4. การทดลอง

* ทดสอบการวิเคราะห์การทำงานของรูปภาพ
* ทดสอบการทำงานของระบบAPI ในการรับส่งข้อมูล
* ทดสอบการเรียกใช้งานTTS Engine บนอุปกรณ์โทรศัพท์ภายในเครื่อง
* ทดสอบการทำงานร่วมกันของระบบที่สร้างขึ้น
* หาข้อผิดพลาดและปรับปรุงแก้ไข

3.1.5. สรุปผล

* สรุปผลและเผยแพร่งานวิจัย

3.2. อุปกรณ์และเครื่องมือ

3.2.1. ฮาร์ดแวร์ (Hardware)

* LAPTOP
  + ASUS TUF GAMING FX504 SERIES
* ระบบปฎิบัติการ: Windoews 11
* Processor Intel® Core™ i5-8300H CPU @ 2.30GHz
* RAM 8.00 GB
  + DELL G3 15
    - ระบบปฎิบัติการ: Windoews 11
    - Processor Intel® Core™ i5-9300H CPU @ 2.40GHz
    - RAM 8.00 GB
* โทรศัพท์
  + HUAWEI NOVA 3
    - RAM 6 GB

3.2.2. ซอฟต์แวร์ (Software)

* Visual Studio Code
* Tesseract

3.2.3. ภาษาที่ใช้ในการพัฒนา (Language)

* Python
* Java

3.3. การออกแบบระบบ

ในการทำโครงงานคณะผู้จัดทำได้มีการรวมแนวคิดและองค์ความรู้หลายๆ อย่างในการทำโครงงานครั้งนี้ ซึ่งได้ออกแบบแอพพลิเคชันไว้ว่า ให้แอพพลิเคชันทำหน้าที่ในการส่งรูปภาพ และ รับข้อมูลข้อความกลับมาแสดงผล และทำการอ่านออกเสียง ในส่วนการวิเคราะห์ข้อความบนรูปภาพเป็นการทำงานของระบบTesseract OCR ทำการสกัดข้อความออกจากภาพ และ สร้างระบบApi เพื่อให้ผู้ที่มาใช้บริการสามารถมาเรียกใช้บริการผ่าน Api ได้ ซึ่งจะใช้ Ngrok เพื่อให้เข้าระบบที่ทำงานอยู่บน locallhost เป็น Public url

Text

Description automatically generated

รูปภาพประกอบ แอพพลิเคชันเชื่อมต่อกับ Url ของ Server ที่ให้บริการ

Graphical user interface, text, application

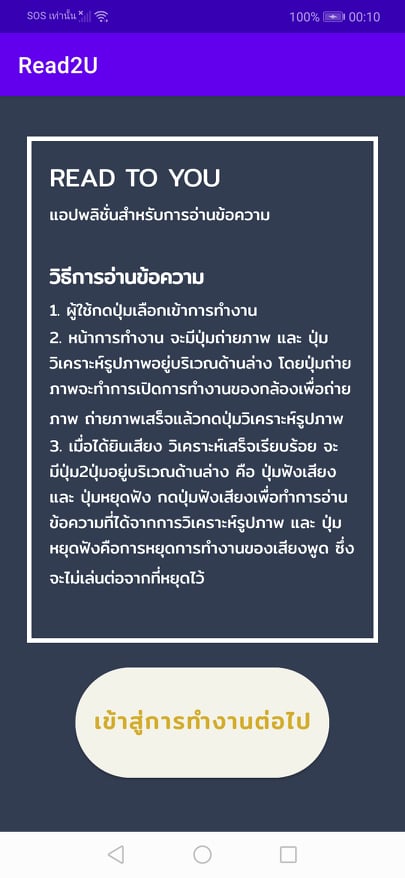
Description automatically generated

รูปภาพประกอบ การทดสอบUrl ผ่าน POSTMAN

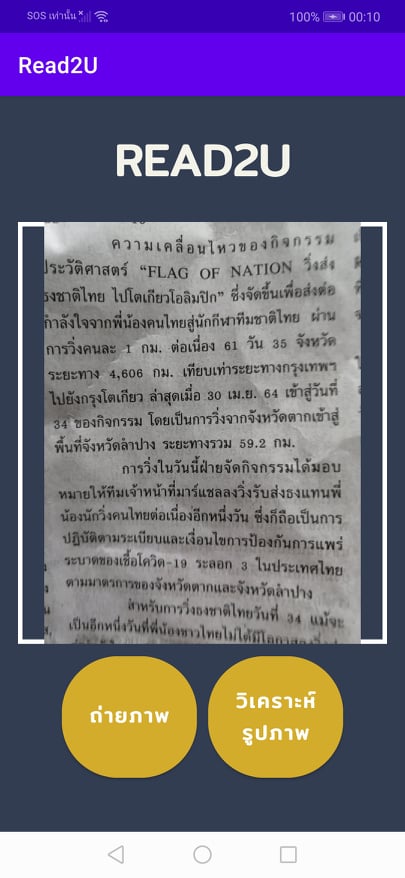
* การออกแบบแอพพลิเคชัน

ทำการออกแบบแอพพลิเคชันบนโปรแกรม Android Studio โดยรูปแบบของแอพลิเคชันจะออกแบบให้เรียบง่าย และการใช้งานไม่ซับซ้อน และ แอพพลิเคชันจะทำการเชื่อมต่อกับAPI ผ่าน Url โดยส่งรูปภาพผ่าน Url ไปสกัดหาข้อความบนรูปภาพ และรับข้อความที่ตอบกลับมาแสดงผลบนแอพพลิเคชัน

A screenshot of a phone

Description automatically generated with medium confidence

รูปภาพประกอบ หน้าตาของแอพพลิเคชัน READ2U

 Text

Description automatically generated

รูปภาพประกอบ หน้าตาของแอพพลิเคชัน READ2U

A person holding a phone in front of a computer

Description automatically generated with medium confidence

รูปภาพประกอบ ทำการติดตั้งแอพลิเคชันลงบนโทรศัพท์

* การออกแบบระบบ REST API

Text

Description automatically generated with low confidence

Text

Description automatically generated

รูปภาพประกอบ เปิดการทำงานบน localhost บนเครื่องคอมพิวเตอร์

A screenshot of a computer screen

Description automatically generated with medium confidence

รูปภาพประกอบ Ngrok Command

จากรูปภาพประกอบที่ การสร้างAPI และการทำงานของโปรแกรมผ่าน Localhost และรูปภาพประกอบที่ Ngrok ทำการสร้างUrl และแสดงผลการเรียกใช้งาน โดยUrl เป็น Public Url ของ Ngrokจะช่วยสามารถช่วยการเข้าถึงการทำงานบนเครื่องคอมพิวเตอร์ได้

3.4. ขั้นตอนการทำงานของระบบ

Diagram

Description automatically generated

รูปภาพประกอบ การทำงานของแอพพลิเคชัน

จากรูปภาพประกอบ แสดงการทำงานของแอพพลิเคชัน READ2U ซึ่งโทรศัพท์จะทำการถ่ายภาพและ ส่งรูปภาพผ่านUrl การเรียกใช้งานผ่าน API รูปภาพนั้นจะถูกส่งไปประมวลผลและทำการวิเคราะห์ข้อความที่อยู่บนรูปภาพ และตอบกลับข้อความที่ได้ทำการวิเคราะห์ หลังจากนนั้นข้อความที่ได้จะถูกทำการส่งกลับมายังแอพพลิเคชัน และทำการฟังเสียงโดยการให้แอพพลิเคชันทำการเรียกใช้งานของ TTS engine ในการอ่านข้อความ

ตาราง 2 ขั้นตอนการดำเนินงานและช่วงระยะเวลาปี 2564-2565

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ขั้นตอนการดำเนินงาน** | **2564-2565** | | | | | | | | | |
| **ส.ค.** | **ก.ย.** | **ต.ค.** | **พ.ย.** | **ธ.ค.** | **ม.ค.** | **ก.พ.** | **มี.ค.** | **เม.ย.** | **พ.ค.** |
| **เริ่มต้นวางแผนโครงการ** | | | | | | | | | | |
| เลือกหัวข้อโครงงาน |  |  |  |  |  |  |  |  |  |  |
| ศึกษาหาข้อมูลแนวคิดและทฤษฎี |  |  |  |  |  |  |  |  |  |  |
| เลือกเครื่องมือ |  |  |  |  |  |  |  |  |  |  |
| กำหนดรายละเอียดของโครงงาน |  |  |  |  |  |  |  |  |  |  |
| **การวิเคราะห์เครื่องมือที่ใช้** | | | | | | | | | | |
| ศึกษาการทำงานของไลบารี่ |  |  |  |  |  |  |  |  |  |  |
| ศึกษาการทำงานของอุปกรณ์ |  |  |  |  |  |  |  |  |  |  |
| จัดหาวัสดุ อุปกรณ์ สำหรัยใช้ในการทำงาน |  |  |  |  |  |  |  |  |  |  |
| **การออกแบบระบบ** | | | | | | | | | | |
| ออกแบบรูปแบบแอพพลิเคชัน |  |  |  |  |  |  |  |  |  |  |
| ออกแบบรูปแบบการทำงานของการวิเคราะห์รูปภาพ |  |  |  |  |  |  |  |  |  |  |
| ออกแบบโครงสร้างของAPI |  |  |  |  |  |  |  |  |  |  |
| **การพัฒนาระบบ** | | | | | | | | | | |
| จัดทำแอพพลิเคชัน |  |  |  |  |  |  |  |  |  |  |
| จัดทำระบบและสร้างAPI |  |  |  |  |  |  |  |  |  |  |
| **การทดลอง** | | | | | | | | | | |
| ทดสอบการทำงานของแอพพลิเคชัน |  |  |  |  |  |  |  |  |  |  |
| ทดสอบการทำงานของระบบAPI |  |  |  |  |  |  |  |  |  |  |
| ทดสอบการทำงานของระบบแอพพลิเคชันร่วมกับการทำงานของระบบAPI |  |  |  |  |  |  |  |  |  |  |
| ประเมินวัดความถูกต้องของระบบและข้อความ |  |  |  |  |  |  |  |  |  |  |
| หาข้อผิดพลาดและปรับปรุงแก้ไขระบบ |  |  |  |  |  |  |  |  |  |  |
| ปรับปรุงประสิทธิภาพการทำงานของระบบ |  |  |  |  |  |  |  |  |  |  |
| ทดสอบการทำงานของระบบแอพพลิเคชันร่วมกับการทำงานของระบบAPI ครั้งที่ 2 |  |  |  |  |  |  |  |  |  |  |
| ประเมินวัดความถูกต้องของระบบและข้อความ |  |  |  |  |  |  |  |  |  |  |
| **สรุปผลการทดลอง** | | | | | | | | | | |
| สรุปผลการทดลองและเผยแพร่โครงงาน |  |  |  |  |  |  |  |  |  |  |

3.5. Dataset

* ข้อความภาษาไทย 10 ภาพ
* ข้อความภาษาอังกฤษ 10 ภาพ
* ข้อความภาษาไทยผสมภาษาอังกฤษ 10 ภาพ

**บทที่ 4**

**ผลการดำเนินโครงงาน**

* 1. การทำงานของแอพพลิเคชัน

ผลการดำเนินงานในส่วนการทำงานของแอพพลิเคชัน แอพพลิเคชันสามารถ่ายภาพและส่งรูปภาพไปทำการประมวลผลวิเคราะห์หาข้อความบนรูปภาพได้ โดยจะรองรับภาษาทั้งภาษาอังกฤษ ภาษาไทย และรูปภาพที่ผสมระหว่างภาษาอังกฤษและภาษาไทย และระบบApi สามารถนำรูปภาพทีได้รับการส่งเข้ามา สามารถนำไปวิเคราะห์ข้อความที่อยู่บนรูปภาพ โดยการใช้ Tesseract OCR ในการวิเคราะห์ข้อความออกมาและสามารถตอบกลับข้อความออกมาได้ และแอพพลิเคชันสามารถได้รับข้อความตอบกลับ เพื่อแสดงผลให้ผู้ใช้สามารถฟังเสียงได้ โดยผ่านการทำงานของ TTS Engine ที่ช่วยในการอ่านข้อความ

* 1. การเลือกโมเดลที่ใช้ในการวิเคราะห์

ภายในTesseract มี Machine Learning ในการทำงานซึ่งผู้พัฒนามี Model ให้เลือกใช้ 3 โมเดล ซึ่งโครงงานได้เลือกใช้โมเดล tessdata-best ที่ให้ความแม่นยำมากที่สุด

Table

Description automatically generated

* 1. ผลการประเมินความถูกต้องของข้อความ

|  |  |  |  |
| --- | --- | --- | --- |
| ลำดับรูป | Thai(100) | Eng(100) | Thai+Eng(100) |
| 1 | 60 | 89 | 94 |
| 2 | 57 | 37 | 82 |
| 3 | 80 | 90 | 91 |
| 4 | 3 | 85 | 22 |
| 5 | 92 | 19 | 22 |
| 6 | 76 | 80 | 0 |
| 7 | 0 | 93 | 76 |
| 8 | 8 | 86 | 56 |
| 9 | 94 | 98 | 97 |
| 10 | 0 | 100 | 20 |
|  | จะทำการแปลงข้อความจากด้านบนลงล่างทำให้การประมวลผลข้อความด้านล่างได้ผลลัพธ์ที่ไม่มีประสิทธิภาพหรือไม่ได้เลย | การประมวลผลดีขึ้นอยู่กับภาพถ่าย | การประมวลผลดีสามารถแปลงออกมาทั้ง2ภาษา และมีข้อผิดพลาดบ้าง  ข้อความจากด้านบนหรือด้านล่างไม่มีผลต่อการประมวลผลโมเดลนี้ |

ThaiEng คำผิดส่วนใหญ่จะเป็นการตกหล่นวรรณยุกต์ แต่ยังไม่สามารถฟังแล้วเข้าใจได้ มีบางรูปที่ประมวลภาพได้ไม่ดีอยู่อาจจะเกิดจากรูปภาพที่ถ่ายเพราะทำการประมวลไป2ครั้งได้ผลลัพธ์เหทือนเดิม

Thai แปลงคำผิดเลย แปลงวนไปวนมาข้ามบรรทัดบ้าง แปลงในบรรทัดเดียวกันแต่ข้อความอยู่คนละที่กันบ้าง

Eng แปลงดี บางภาพแปลงออกมาไม่ดีเลย อาจจะเกิดจากภาพที่ถ่าย

วิเคราะห์ค่าที่ได้ 0 เนื่องจากข้อความที่จะเอามาวิเคราะห์อยู่ย่อหน้าล่างๆ tesseract อาจจะจำกัดจำนวนคำประมวลผลทำให้ไม่สามารถประมวลออกมาย่อหน้าสุดท้าย

**บทที่ 5**

**สรุปผล อภิปรายผล และ ช้อเสนอแนะ**