

แบบเสนอหัวข้อโครงการวิศวกรรมคอมพิวเตอร์

ภาคเรียนที่ 1 / 2567

1. ชื่อโครงการ

ระบบถอดข้อความจากทรานสคริปต์

Transcript text data extractor

2. ชื่อผู้เสนอโครงการ

- 2.1. นาย โอภาสพันธ์ กลิ่นชื่นจิต รหัส 65543206039-1 ลายเซ็น..... หัวหน้ากลุ่ม
- 2.2. นาย ญัฐภัทร เจริญกิจหัตถกร รหัส 65543206011-0 ลายเซ็น.....
- 2.3. นาย พุฒิพงศ์ หลีแก้วสาย รหัส 65543206028-4 ลายเซ็น.....

3. อาจารย์ที่ปรึกษาโครงการ

- 3.1. ดร.จักรภาพ ใหม่เสน ลายเซ็น.....

4. ที่มาและความสำคัญของปัญหา

ในปัจจุบันโลกได้มีการเปลี่ยนแปลงไปเข้าสู่ยุคของดิจิทัลมากขึ้น ทำให้เกิดเทคโนโลยีใหม่ ๆ ขึ้นมามากมายในช่วงเวลาอันสั้น โดยเทคโนโลยีเหล่านั้นเข้ามาบดบังบทบาทสำคัญในชีวิตประจำวัน เพื่อให้สามารถทำงานและดำเนินธุรกิจได้อย่างสะดวกสบาย ยกตัวอย่างนวัตกรรม OCR (Optical Character Recognition) คือนวัตกรรมที่เปลี่ยนโลกดิจิทัล ในยุคดิจิทัลที่ข้อมูลเป็นสิ่งสำคัญจึงไม่สามารถมองข้ามได้ เพราะ OCR หรือการรู้จำอักขระด้วยแสง เป็นเทคโนโลยีสำหรับการเปลี่ยนข้อมูลจากเอกสารซึ่งรูปแบบที่เข้าถึงได้ยาก ให้กลายเป็นข้อมูลดิจิทัลที่เข้าถึงได้ง่าย ลดความผิดพลาด ประหยัดเวลา และสะดวกต่อการจัดการกับข้อมูล โดยจะได้ข้อมูลจากเอกสารประเภทต่าง ๆ เช่น ใบเสนอราคา ใบกำกับภาษี ภาพถ่ายเอกสาร หรือ ภาพที่ถูกสแกนเนอร์มา เป็นต้น ซึ่ง OCR ทำให้สามารถดึงข้อมูลที่มีค่าเหล่านั้นออกมาใช้ประโยชน์ได้อย่างเต็มที่ โดยจะถูกนำมาใช้ในงานที่หลากหลาย ได้แก่ งานจัดเก็บเอกสาร การเงินและธนาคาร การศึกษา ธุรกิจและองค์กรต่าง ๆ ยังมีอีกหลายงานที่ไม่ได้กล่าวมาข้างต้นที่น่า OCR เข้ามาช่วยในการทำงานหรือดำเนินธุรกิจ ดังนั้น OCR เป็นเทคโนโลยีที่มีประโยชน์และมีการใช้งานในหลากหลายวงการ ช่วยเพิ่มประสิทธิภาพในการทำงานและการจัดการข้อมูลในรูปแบบดิจิทัล

เทคโนโลยี OCR เป็นเทคโนโลยีที่มีประโยชน์ในหลายๆ ด้านของการทำงาน โดยเฉพาะอย่างยิ่งในการจัดการกับเอกสารและข้อมูลที่เป็นตัวอักษร การใช้ OCR สามารถช่วยในการสแกนและแปลงเอกสารกระดาษเป็นไฟล์ดิจิทัลที่สามารถค้นหาและแก้ไขได้ ช่วยลดพื้นที่การจัดเก็บเอกสารและเพิ่มความสะดวกในการเข้าถึงข้อมูล สำหรับองค์กรที่มีปริมาณข้อมูลและเอกสารจำนวนมาก OCR สามารถช่วยแปลงข้อมูลเหล่านี้ให้เป็นรูปแบบดิจิทัลที่สามารถจัดการและวิเคราะห์ได้ง่ายขึ้น จึงมีการนำเทคโนโลยี OCR เข้ามาช่วยในการจัดการเอกสารการเรียนของนักเรียนนักศึกษา รวมถึงเอกสารสำคัญอย่างใบทรานสคริปต์ เพื่อความสะดวกสบายและประหยัดเวลาต่อการจัดการข้อมูล โดยเทคโนโลยี OCR ในยุคปัจจุบันนี้มีหลากหลายเวอร์ชันให้เลือกใช้

งานมากมาย แต่คุณภาพของเทคโนโลยี OCR นั้นต่างกันไป อาจส่งผลทำให้ไม่สามารถใช้งานตามความต้องการของผู้ใช้งาน โดยเมื่อนำมาใช้ในการสกัดคำจากใบทรานสคริปต์ยังมีการสกัดคำออกมาไม่ถูกต้องตามแบบหรือการประมวลผลใช้เวลานานจนเกินไป จึงเป็นปัญหาที่จะต้องได้รับการแก้ไข

จากปัญหาดังกล่าวจึงทำให้ทราบว่าเทคโนโลยี OCR ในปัจจุบันมีมากมายแต่ยังไม่สามารถใช้งานตามความต้องการได้ จึงทำให้คณะผู้จัดทำได้เล็งเห็นปัญหานี้จึงอยากสร้างเทคโนโลยี OCR ขึ้นมาเอง โดยเริ่มจากการให้ OCR สามารถอ่านตัวอักษรจากภาพให้ถูกต้องก่อน โดยใช้ภาษา Python ในการสร้างและพัฒนา เครื่องมือการประมวลผลภาพที่ใช้ ได้แก่ OpenCV ใช้ในงาน Image Reading and Writing, Image Transformations , Image Processing และ Image Segmentation การทำ Data Augmentation ใช้ TensorFlow/Keras ImageDataGenerator สำหรับการทำ Data Augmentation เช่น การหมุนภาพ การปรับแสง การยืดหรือหดภาพ การสร้างและฝึกโมเดล CNN และ Metrics ที่นิยมใช้สำหรับการประเมินโมเดล OCR ได้แก่ 1) Accuracy วัดความถูกต้องของการทำนายทั้งหมด 2) Precision วัดความแม่นยำของโมเดลในกรณีที่ทำนายว่าเป็นคลาสบวก (Positive) 3) Recall (Sensitivity) วัดความสามารถของโมเดลในการค้นหาลาสบวกทั้งหมด 4) F1 Score เป็นค่าเฉลี่ยแบบถ่วงน้ำหนักระหว่าง Precision และ Recall 5) Confusion Matrix ช่วยให้เห็นภาพรวมของการทำนายที่ถูกต้องและผิดพลาดในแต่ละคลาส

5. วัตถุประสงค์ของโครงการ

- 5.1 เพื่อศึกษาการใช้เครื่องมือในการสร้างและพัฒนาระบบปัญญาประดิษฐ์ โดยใช้ภาษา python
- 5.2 เพื่อสร้างชุดข้อมูลตัวอักษรไทย อังกฤษ และตัวเลข สำหรับใช้ในงาน OCR
- 5.3 เพื่อสร้างและพัฒนาระบบ OCR เฉพาะทางในการอ่านข้อความจากใบทรานสคริปต์

6. ทบทวนวรรณกรรม (งานวิจัยที่เกี่ยวข้อง)

ทศนัย ชุ่มวัฒนะ, วรัมพร รัตนอำนวยชัย, ภัทธานุช ช่วยชู (2565) ได้ศึกษาการสกัดข้อมูลผู้ป่วยด้วยเทคนิคตัวช่วยแปลงไฟล์เอกสาร ปัญหาหลักที่สำคัญคือการทำที่หน่วยงานยังคงมีการจัดเก็บข้อมูลผู้ป่วย บางส่วนในรูปแบบของเอกสาร (hard copies) เอกสารสแกน รูปภาพ และไฟล์ PDF ซึ่งจำเป็นต้องแปลงเอกสารดังกล่าวให้อยู่ในรูปแบบดิจิทัลเพื่อการวิเคราะห์ และการใช้ประโยชน์ในด้านต่าง ๆ งานวิจัยนี้จึงมีวัตถุประสงค์เพื่อนำเสนอวิธีในการ แปลงข้อความจากเอกสาร (physical document) ให้อยู่ในรูปแบบดิจิทัล โดยใช้ Optical Character Recognition หรือที่เรียกว่า OCR ซึ่งเป็นเทคโนโลยีที่จะช่วยสกัดข้อความทั้งหมดจากเอกสาร ลงสู่รูปแบบข้อมูลที่มีโครงสร้าง (Structure Data Format) ผลลัพธ์จากการวิจัยแสดงให้เห็นว่า การใช้เทคนิค OCR สามารถ ช่วยสกัดข้อความจากภาพหรือเอกสารให้อยู่ในรูปแบบดิจิทัล ที่สามารถค้นหา แก้ไข และทำการวิเคราะห์ได้ โดยมีอัตราความ แม่นยำในการประมวลผลประมาณ 74.62% โดยเฉลี่ย สำหรับการสกัดข้อมูลแอททริบิวต์ (Attributes) และ 68.46% สำหรับการสกัดข้อมูลที่เป็นค่า (Values) จากเอกสาร ดังนั้น ประโยชน์จากการใช้เทคโนโลยี OCR จะช่วยให้หน่วยงานด้านสาธารณสุข สามารถเปลี่ยนข้อมูลในรูปแบบเอกสารให้เป็นรูปแบบดิจิทัล เพื่อนำไปสู่การใช้ประโยชน์จากข้อมูลได้อย่างมีประสิทธิภาพ และ ช่วยลดปริมาณเอกสารที่มีการจัดเก็บภายในองค์กร

นายวิรุฬห์ ศรีบริรักษ์ (2562) ได้ศึกษาเทคโนโลยีเภสัชสารสนเทศแสดงข้อมูลฉลากยาเอกสารกำกับยาแบบอัตโนมัติสำหรับบริหารจัดการคลังยาปฏิชีวนะเพื่อความปลอดภัย โดยระบบที่พัฒนาขึ้นนี้ถูกแบ่งออกเป็น 4 ส่วน โดยส่วนแรกคือส่วนของการเก็บข้อมูล ซึ่งข้อมูลยาที่ใช้ในการเก็บมาจากเอกสารกำกับยา โดยใช้เทคโนโลยี OCR (Optical Character Recognition) เป็นการแปลงรูปภาพเป็นตัวอักษร ส่วนที่สองการจัดการข้อมูลยาซึ่งเป็นส่วนที่ใช้ในการสร้าง แก๊ส ไลบ และแสดง ข้อมูลยาทั้งหมดภายในระบบ ส่วนที่สามการค้นหาข้อมูลยาจะเป็นการค้นหาข้อมูลยาได้จากชื่อสรรพคุณของยา รวมถึงอาการของผู้ป่วย ส่วนที่สี่ส่วนการจัดการผู้ใช้จะควบคุมการเข้าถึงระบบในส่วนต่างๆ ของผู้ใช้แต่ละประเภท ระบบเทคโนโลยีเภสัชสารสนเทศแสดงข้อมูลฉลากยาเอกสารกำกับยาแบบอัตโนมัตินี้มีประสิทธิภาพ ในการแปลงรูปภาพเป็นตัวอักษรด้วยเทคโนโลยี OCR ที่ความแม่นยำได้ถึงร้อยละ 96.61 และสามารถช่วยลด เวลาของการเก็บข้อมูลลงฐานข้อมูลยา และช่วยให้ผู้บริโภคมุ่งความเข้าใจในการใช้ยา มีความรู้ความถูกต้องเกี่ยวกับข้อมูล รู้วิธีการเก็บรักษาของยาเพื่อให้ได้รับประโยชน์ช่วยให้ปลอดภัยและลดปัญหาของการใช้ยาที่ผิดกับโรค

สรวิทย์ โพธิ์พยอม (2562) ได้ศึกษาการตรวจจับและรู้จำข้อความเพื่อการทำงานแบบอัตโนมัติในภาคอุตสาหกรรม การตรวจจับตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนบนจอภาพแอลอีดีนั้นมีความท้าทาย โดยเฉพาะอย่างยิ่งในการค้นหาตัวเลขท่ามกลางวัตถุที่หลากหลายบนภาพซึ่งอาจมีส่วนประกอบหรือรูปทรงที่คล้ายคลึงกัน ดังนั้นในงานวิจัยนี้จะนำเสนอวิธีการนำตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วนออกมาจากภาพ โดยการตัดแบบกำหนดช่วงบนแบบจำลองสี HSV (Predefined HSV color slicing technique) เพื่อการตรวจจับและประยุกต์ใช้โครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional neural network) ในการรู้จำ ทางผู้วิจัยนำเสนอวิธีการเพิ่มข้อมูลโดยการแปลงเรดอนหลายขนาด (Multiscale Radon transform) เพื่อลดปัญหาด้านข้อจำกัดของข้อมูลภาพสอนสำหรับโครงข่ายประสาทเทียม รวมถึงการลดปัญหาพฤติกรรมการถ่ายภาพซึ่งส่งผลให้ภาพเอียง มีขนาดภาพที่หลากหลาย มุมมองของภาพ และผลจากการตกกระทบของแสงที่แตกต่างกัน ซึ่งอาจส่งผลต่อการรู้จำภาพตัวเลขที่แสดงผลในรูปแบบเจ็ดส่วน โดยเครื่องอ่านอักขระด้วยแสง (Optical character recognition) มีความสามารถในการรู้จำตัวอักษรแล้วนำไปเก็บไว้ในหน่วยความจำมีค่าใช้จ่ายที่น้อยกว่าเครื่องจักรขนาดใหญ่ด้วยลักษณะที่เป็นเจ็ดส่วนนั้นทำให้ตัวเลขมีโครงสร้างที่ไม่ต่อเนื่อง และได้ผลการทดลองแสดงให้เห็นว่าวิธีการ ตรวจจับที่ผู้วิจัยนำเสนอมีความแม่นยำและค่าการเรียกคืนตัวอักษรที่สูงที่สุดอยู่ที่ 94.46% และ 92.24% ตามลำดับ รวมไปถึงวิธีการรู้จำที่ผู้วิจัยนำเสนอมีค่าเฉลี่ยของอัตราความถูกต้องอยู่ที่ 87.05%

นางสาวนริศรา เหม้ง , นายปณชัย เทพรักษ์ (2564) ได้ศึกษาการตรวจสอบผลิตภัณฑ์นมโดยใช้เทคนิค Deep Learning มีการพัฒนาขึ้นอย่างต่อเนื่องโดยเฉพาะภาษาอังกฤษ ภาษาไทย เป็นต้น แต่ในส่วนการรู้จำตัวอักษรจากรูปภาพนั้นยังมีการพัฒนาที่ไม่มากพอ อีกทั้งประสิทธิภาพในการรู้จำยังไม่ดีเท่าที่ควร เนื่องจากงานวิจัยทางด้านการรู้จำตัวอักษรจากรูปภาพยังไม่แพร่หลาย ประกอบด้วย Algorithm ที่ใช้ในการรู้จำมีความซับซ้อนเป็นอย่างมาก ดังนั้น จึงมีแนวคิดที่อยากจะศึกษาและพัฒนางานทางด้านการรู้จำตัวอักษรจากรูปภาพเพื่อเรียนรู้ Algorithm ตลอดจนทดสอบผลทำนายของโมเดลรู้จำที่ออกแบบ โดยในงานนี้จะเป็นลักษณะการตรวจสอบผลิตภัณฑ์ของกล่องนม ซึ่งจะใช้หลักการ Image Processing มาใช้ในขั้นตอน Pre-process เป็น

ขั้นตอนการเตรียมข้อมูลก่อนเข้าไปประมวลผลในโมเดล จากนั้นจะใช้ Convolutional Neural Network (CNN) และใช้ Tesseract OCR (Optical Character Recognition) ในการรู้จำตัวอักษรและทดสอบอัลกอริทึมที่ใช้ในการรู้จำตัวอักษร โดยใช้การประมวลผลภาพจากการทดลองเพื่อประเมินประสิทธิภาพของโมเดล สถาปัตยกรรม CNN บนพื้นฐานของ Convolutional ที่แตกต่างกันและการเลือกใช้ Optimizer โดยจะเลือกใช้ SGD ซึ่งทำการทดสอบประสิทธิภาพกับข้อมูล จากการทดลองวัดประสิทธิภาพการรู้จำตัวอักษรโดยใช้ชุดข้อมูลรูปภาพโลโก้ 4000 ภาพ รูปภาพอย. 200 ภาพ และรูปภาพวันหมดอายุ 200 ภาพ นำรูปภาพโลโก้มา Generate data และใช้ Model จะพบว่าค่า Accuracy เท่ากับ 99.875 และค่า Loss เท่ากับ 0.011 ส่วนรูปภาพอย.และวันหมดอายุใช้ Tesseract OCR ในการรู้จำโดยมีความผิดพลาดค่า Character Error Rate ของรูปภาพอย.ต่ำที่สุดที่ 0.00% และรูปภาพวันหมดอายุต่ำที่สุดอยู่ที่ 12.5%

จิรโรจน์ โสอุตร (2562) ได้ศึกษาและพัฒนาระบบส่งเสริมการท่องเที่ยวจากข้อมูลการเข้าพักของนักท่องเที่ยว งานวิจัยนี้จึงได้นำเสนอวิธีในการบันทึกข้อมูลนักท่องเที่ยวโดยอัตโนมัติเพื่อแก้ไขปัญหาบันทึกข้อมูลนักท่องเที่ยวมีหลายขั้นตอน ทำให้เกิดความล่าช้า ชั้ชซ้อน และอาจเกิดความผิดพลาด ด้วยการพัฒนาระบบโมบายแอปพลิเคชันที่ใช้จัดเก็บข้อมูลนักท่องเที่ยว โดยใช้กล้องจากสมาร์ตโฟนสแกนจากหน้าหนังสือพาสปอร์ตและแปลงข้อมูลเป็นรูปแบบดิจิทัลโดยเทคโนโลยีการรู้จำตัวอักษรด้วยแสง (Optical Character Recognition: OCR) เพื่อจัดเก็บ และสามารถนำข้อมูลไปวิเคราะห์เพื่อออกรายงานการเข้าพัก รวมถึงการนำส่งข้อมูลให้ทางสำนักงานตรวจคนเข้าเมืองได้อย่างถูกต้องแม่นยำ และช่วยลดเวลาในการเช็คอิน ผลที่ได้คือสามารถทดแทนวิธีการลงทะเบียนนักท่องเที่ยวรูปแบบเดิมได้มี ประสิทธิภาพที่ดีขึ้น อีกทั้งช่วยสนับสนุนการทำงานของทางโรงแรม ไม่ว่าจะเป็นการลงทะเบียนผู้เข้าพัก การทำสรุปรายงานประจำวันในรูปแบบกราฟ หรือตารางการออกรายงานนำส่งสำนักงานตรวจคนเข้าเมืองในรูปแบบ Excel File การยืนยันตัวบุคคลของนักท่องเที่ยวที่มีความถูกต้องแม่นยำมากขึ้น ช่วยลดการทำงานของพนักงาน ลดเวลา ลดขั้นตอนที่เกินความจำเป็น แต่ยังคงสอดคล้องกับหลัก กฎหมาย พ.ร.บ. ตรวจคนเข้าเมือง พ.ศ. 2522 และ พ.ร.บ. คุ้มครองข้อมูลส่วนบุคคล พ.ศ. 2562

วิศรุต เหล่าดารา (2565) ศึกษาและพัฒนาระบบเซ็นเซอร์ชื่อบุคคลออกจากเอกสารสแกนคำพิพากษา ด้วยปัญญาประดิษฐ์ การศึกษานี้จึงมีวัตถุประสงค์เพื่อจัดทำระบบกระบวนการที่นำเทคโนโลยีมาใช้ตรวจสอบชื่อบนเอกสารคำพิพากษานำมาช่วยผู้ปฏิบัติงานในการดำเนินการย่อคำพิพากษาและลดระยะเวลาและแบ่งเบาภาระ ในการดำเนินการของเจ้าหน้าที่ผู้ปฏิบัติงาน ในงานวิจัยนี้ จึงนำไฟล์คำพิพากษาที่มีลักษณะเป็นรูปภาพมาเข้ากระบวนการ Optical Character Recognition ทำการอ่านข้อความในเอกสารออกมาเป็นข้อความ ที่สามารถนำมาคัดแยกได้ด้วยรูปแบบของ เอกสารและการจับคำนำหน้าชื่อและใช้ Named entity recognition เพื่อช่วยแยกชื่อบุคคลออกจากข้อความ ธรรมดา ซึ่งผลการวิจัยสามารถช่วยให้ตำแหน่งของชื่อบุคคลในคำพิพากษาและปกปิดข้อความส่วนใหญ่ได้ จากผลการวิจัยการเซ็นเซอร์ชื่อบุคคลจากเอกสารสแกนคำพิพากษาดังด้วยเทคโนโลยีปัญญาประดิษฐ์ สรุปได้ดังนี้ 1) ระบบใช้เรียกไฟล์คำพิพากษาเพื่อเซ็นเซอร์ชื่อบุคคลที่ปรากฏบนเอกสาร 2) จำนวนข้อความที่พบคิดเป็น 89 % จากข้อความทั้งหมด 3) การใช้เทคนิค OCR ช่วยแยกคำจากเอกสารสแกนคำพิพากษาสามารถแยกชื่อบุคคลได้บางส่วน จากรูปแบบเอกสารหรือคำนำ

หน้าที่พบข้อ 4) การนำ NER ตรวจสอบคำที่แยกมาแล้วเพื่อช่วยยืนยันการ โดยในการวิจัยจากตัวอย่างพบว่า ตรวจสอบชื่อบุคคลถูกต้องเพิ่มขึ้น 5) การใช้ NER ช่วยตรวจสอบประโยคส่วนเกินที่ติดมาจากการขั้นตอนแยก หรือข้อความที่อยู่หลังชื่อได้

7. ทฤษฎีและหลักการ

7.1 หลักการ แนวความคิด

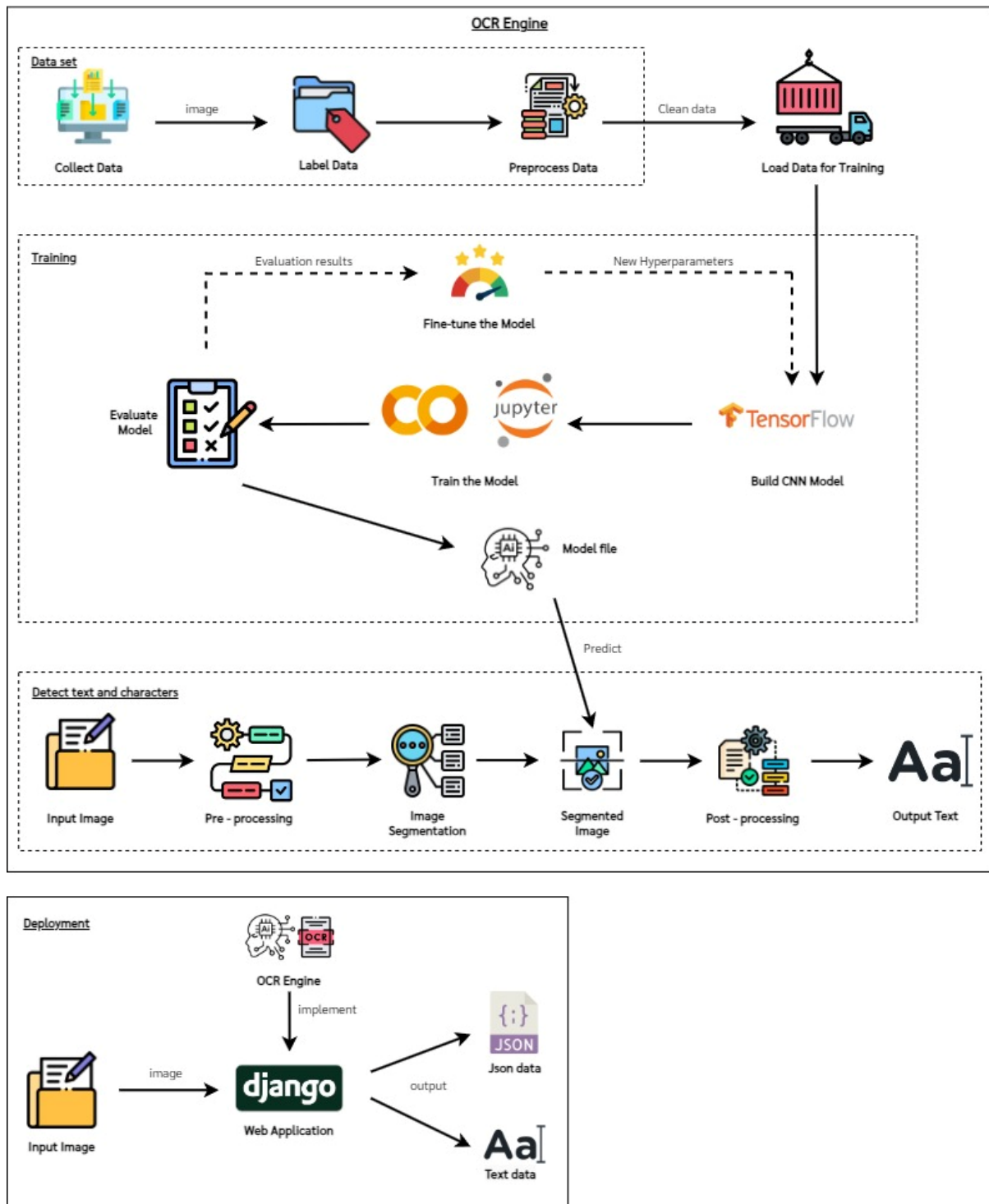
ระบบถอดข้อความจากทรานสคริปต์ ระบบนี้จะใช้เทคโนโลยีของ Deep learning โดยใช้สถาปัตยกรรม Convolutional Neural Network (CNN) ในการตรวจจับข้อความหรือตัวอักษรจากรูปภาพ โดยเราจะทำการสร้าง Data set และ Model ด้วยตนเอง

Data set เราใช้เครื่องมือหรือไลบรารีของภาษา Python ด้วย OpenCV สำหรับ การประมวลผลภาพ เช่น Clean data, Crop images, Convert to grayscale เป็นต้น

Training เราใช้ TensorFlow/Keras ซึ่งเป็นไลบรารีและเฟรมเวิร์กในการสร้างและกำหนดสถาปัตยกรรมของโมเดล โดย TensorFlow/Keras จะไปเขียนและพัฒนามบน Google Colab และ Jupyter Notebook หลังจาก Training เสร็จ เราทำการ Evaluate model ประเมินผลลัพธ์ของโมเดลโดยการวัดค่า accuracy, precision, recall, F1 Score และ confusion matrix เพื่อประเมินความถูกต้องและสมบูรณ์ของโมเดล

Detect text and characters เราใช้ OpenCV ในการทำ Image processing และ Image Segmentation รูปภาพใบทรานสคริปต์ เช่น GrayScale image, Gaussian blur, Binary image, Morphological, Canny method, Contour เป็นต้น

Deployment เราใช้ Django ในการสร้าง Web Application เพื่อที่จะนำ OCR Engine ที่เราสร้างและพัฒนาไปรันบน Web Application เพื่อให้ผู้ใช้สามารถอัปโหลดรูปภาพใบทรานสคริปต์และใช้งานตัว OCR ได้ ผลลัพธ์ที่แสดงออกมามี 2 อย่างคือ ไฟล์ json กับ text ใบทรานสคริปต์ที่อ่านได้จากระบบ OCR

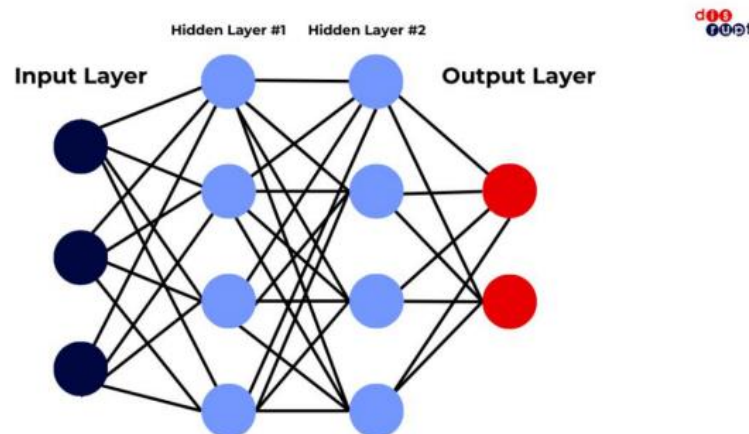


รูปภาพที่ 1 Block Diagram ของระบบถอดข้อความจากทรานสคริปต์

7.2 ทฤษฎีที่เกี่ยวข้อง

Deep Learning

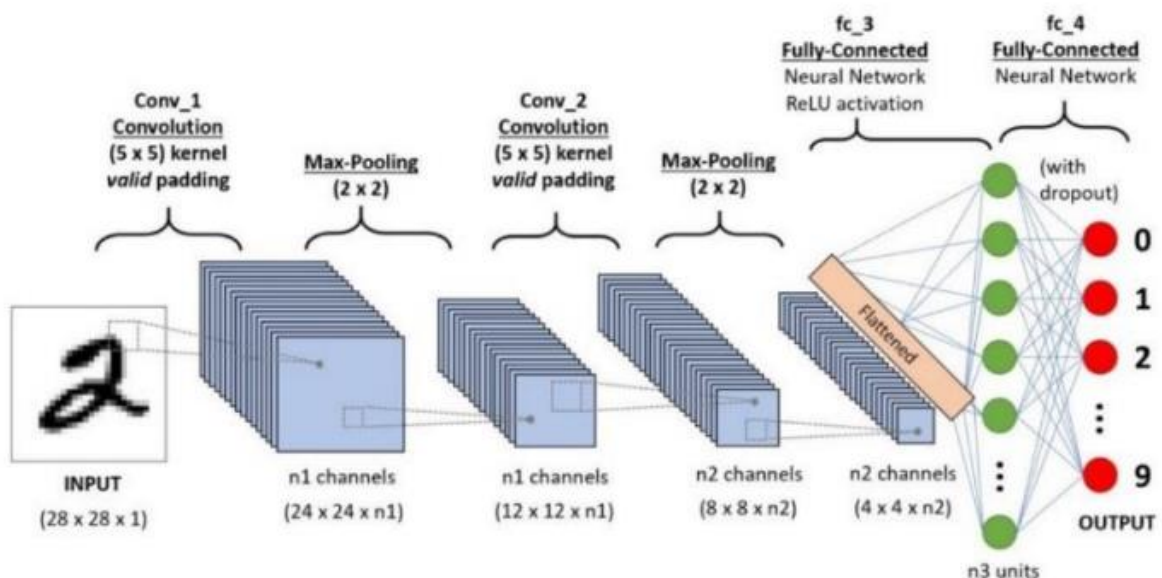
Deep Learning เป็นหนึ่งในสาขาของ Machine Learning ซึ่งใช้ในการสอนปัญญาประดิษฐ์หรือ AI ที่ถูกออกแบบมาเพื่อจำลองการทำงานของสมองของมนุษย์ในการรับข้อมูลและเรียนรู้ผ่านประสบการณ์ที่เน้นการใช้โครงข่ายประสาทเทียม (Artificial Neural Networks) ซึ่งประกอบไปด้วยหลายชั้น (layers) เพื่อทำการเรียนรู้และทำความเข้าใจรูปแบบหรือความสัมพันธ์ในข้อมูลที่ซับซ้อนมากขึ้น



รูปภาพที่ 2 โครงข่ายประสาทเทียม

Convolutional Neural Network

Convolutional Neural Network (CNN) เป็นโครงข่ายประสาทเทียมชนิดหนึ่งที่มีโครงสร้างเฉพาะเจาะจงสำหรับการประมวลผลข้อมูลภาพและข้อมูลที่มีลักษณะเป็นตาราง (grid-like) CNN ได้รับ การออกแบบมาเพื่อจัดการกับการจำแนกภาพ, การตรวจจับวัตถุ และงานด้านการมองเห็นของ คอมพิวเตอร์อื่น ๆ



รูปภาพที่ 3 การประมวลผลข้อมูลภาพ

TensorFlow

TensorFlow คือตัวโปรแกรมช่วยในการเขียนและพัฒนาโมเดล Machine Learning และ Deep Learning ที่พัฒนาโดย Google Brain Team โดยเป็น Open Source Library ที่ใช้ภาษา Python ในการเขียนโค้ด โดย TensorFlow สามารถใช้งานได้บนหลายแพลตฟอร์มรวมถึง Windows, macOS, Linux, Android และ iOS โดย TensorFlow มีการเปิดให้นักพัฒนาสามารถใช้งานได้ฟรีและมีชุดคำสั่งที่ครอบคลุมมากพอสมควรเพื่อการพัฒนาโมเดล Machine Learning และ Deep Learning ในการตัดสินใจ ทางธุรกิจ



รูปภาพที่ 4 TensorFlow

Transfer Learning

Transfer Learning คือ การนำโมเดลของผู้อื่น ที่เทรนกับข้อมูลอื่นเรียบร้อยแล้ว ตัดเอาบางส่วนเฉพาะส่วนที่เราสนใจ มา Reuse ใช้ต่อ มาประกอบในการสร้างโมเดลใหม่ของเรา เช่น เรานำโมเดล Convolution Neural Network (CNN) ที่มี 100 Layer มาตัด Layer สุดท้ายทิ้งไป เปลี่ยนเป็น Layer ที่เหมาะกับงานของเราแทน แล้วเวลาเทรน ก็เทรนแค่ Layer ใหม่ที่อยู่ท้ายสุดอย่างเดียว เพื่อลดเวลา และข้อมูล ที่ใช้ในการเทรน เพราะมี 99 Layer ที่ทำงานได้ถูกต้องอยู่แล้ว ไม่ต้องเทรนตั้งแต่ต้น เรื่อง Transfer Learning จะอธิบายเพิ่มเติมต่อไป

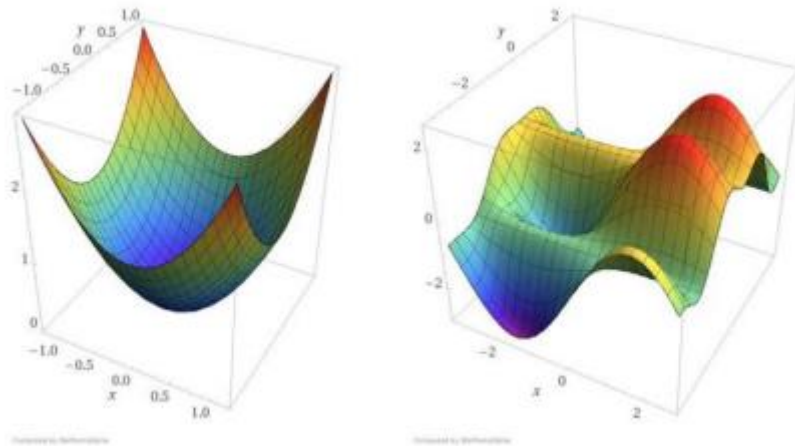
Loss Function

Loss Function คือค่าหนึ่งในกระบวนการฝึกสอนโมเดล (Machine Learning, Deep Learning) โดย เป้าหมายคือการทำให้ค่านี้ลดต่ำลงให้ได้มากที่สุด เพราะยิ่งค่านี้ต่ำลงโมเดลก็จะยิ่งทำงานได้ดียิ่งขึ้น ซึ่งจริง ๆ แล้วค่า Loss Function นี้เป็นฟังก์ชันคณิตศาสตร์ที่มีไว้ใช้สำหรับการทำ Optimization เป็นหลัก โดยจะทำ หน้าทีรับเอาค่าของพีเจอร์ที่มีไปคำนวณออกมาเป็นจำนวนจริง และจำนวนจริงที่ได้มานี้โดยพื้นฐานแล้วจะสื่อ ถึงสิ่งที่สูญเสียไปสำหรับงานนั้น ๆ ดังนั้นแล้วยังมีสิ่งที่สูญเสียหรือต้องจ่ายน้อยเท่าไรก็จะหมายความว่าโมเดล นั้นมีประสิทธิภาพยิ่งขึ้น

Gradient Descent

Gradient Descent เป็น optimization algorithm ชนิดหนึ่งที่ใช้ในการเทรนโมเดลสำหรับการทำ Machine Learning และ Deep Learning โดยสามารถทำงานได้ทั้งในโจทย์ประเภท regression และ classification แค่ปรับเปลี่ยน Loss function ตามประเภทของโจทย์นั้น ๆ Gradient Descent จะทำการ

ปรับค่า parameter ในโมเดล เพื่อลดค่าของ Loss function ให้มีค่าน้อยที่สุด หลักการของ Gradient Descent คือเหมือนกับที่เรายืนอยู่บนยอดเขาแล้วเราค่อย ๆ ลงมาจากเขาเพื่อลงมาถึงจุดที่ต่ำที่สุดของเขานั้น หากนำไปใช้กับโมเดลของเรา มันก็คือการปรับค่า parameter ไปเรื่อย ๆ จนได้ค่า parameter ที่ทำให้ Loss function มีค่าต่ำที่สุด

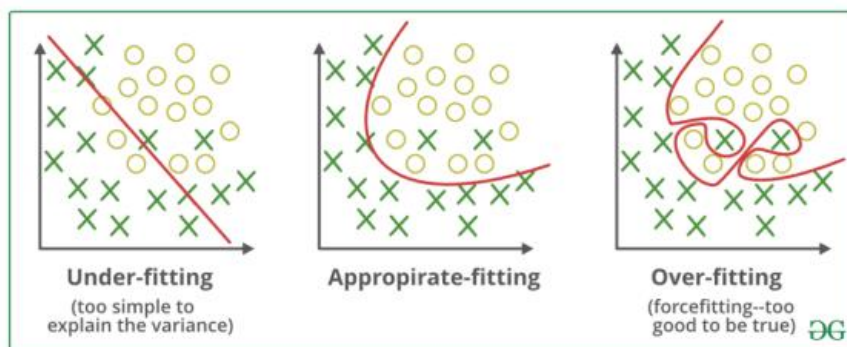


รูปภาพที่ 5 กราฟ Gradient Descent

Overfitting And Underfitting

Overfitting เป็นปัญหาที่เกิดขึ้นเมื่อสถานการณ์ที่เกิดขึ้นเมื่อแบบจำลองทางสถิติหรือ Machine learning ทำงานได้ดี มีความแม่นยำสูง ทำนายแม่นยำกับ training data แต่ทำงานได้แย่ ไม่แม่นยำ เมื่อเจอข้อมูล ใหม่ ๆ หรือ ข้อมูลที่ไม่เคยเห็นมาก่อน (testing data) เรือง่าย ๆ ว่าโมเดลสามารถปรับตัวให้ตรงกับข้อมูล ฝึกสอนได้ดีมากเกินไป จนทำให้ไม่สามารถทำนายข้อมูลใหม่ๆ ได้เท่าที่ควร

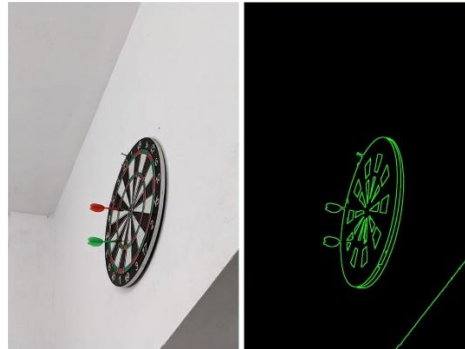
Underfitting เป็นปัญหาที่เกิดขึ้นเมื่อโมเดลไม่สามารถเรียนรู้ข้อมูลฝึกสอนในระดับพอเพียง ผลลัพธ์คือโมเดลทำนายข้อมูลฝึกสอนและข้อมูลที่ไม่เคยเห็นมาก่อนได้ด้วยความแม่นยำที่ต่ำ นั่นหมายความว่าโมเดลไม่สามารถทำนายข้อมูลจริงในสถานการณ์จริงได้อย่างเหมาะสม กล่าวคือ เป็นสถานการณ์ที่เกิดขึ้น เมื่อแบบจำลองทางสถิติหรือ Machine learning ทำงานได้แย่ ทำนาย หาค่าความสัมพันธ์ของชุดข้อมูลไม่ถูกต้อง ไม่ใกล้เคียงกับค่าที่เฉลยและไม่ตรงกับคำตอบที่แท้จริง



รูปภาพที่ 6 Overfitting And Underfitting

Contour Detection

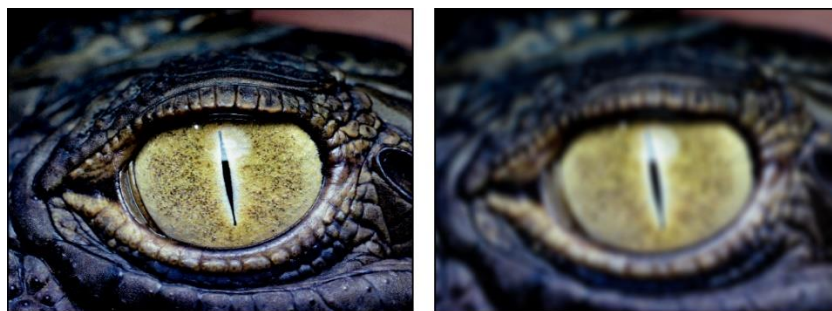
Contour Detection คือการตรวจจับและแสดงขอบเขต (Contour) ของวัตถุในภาพ โดย Contours เป็นเส้นที่เชื่อมต่อกันที่มีค่าเหมือนหรือใกล้เคียงกันซึ่งมักใช้ในการระบุขอบเขตของวัตถุในภาพดิจิทัล Contour จะถูกใช้ในการวิเคราะห์รูปร่างและโครงสร้างของวัตถุ ซึ่งเป็นกระบวนการสำคัญใน Computer Vision และการประมวลผลภาพ (Image Processing) สรุป Contour Image เป็นภาพที่แสดงขอบเขตหรือโครงร่างของวัตถุที่ตรวจจับได้ในภาพ ช่วยในการวิเคราะห์และรู้จำรูปร่างของวัตถุในงานประมวลผลภาพ



รูปภาพที่ 7 Contour Detection

Gaussian Blur

Gaussian Blur คือเทคนิคหนึ่งในการทำให้ภาพเบลอ (blur) โดยการใช้ฟังก์ชันคณิตศาสตร์แบบ Gaussian function ซึ่งเป็นฟังก์ชันที่มีการกระจายค่าตามรูปแบบปกติ (normal distribution) การเบลอด้วย Gaussian Blur จะทำให้พิกเซลในภาพถูกปรับให้เรียบและกลืนกัน โดยพิกเซลใกล้เคียงจะถูกเฉลี่ยกัน ทำให้ภาพดูเบลอขึ้น Gaussian Blur นั้นเป็นเทคนิคพื้นฐานที่มีความสำคัญในการประมวลผลภาพและมีการใช้งานอย่างกว้างขวางในงานด้าน Computer Vision และ Image Processing



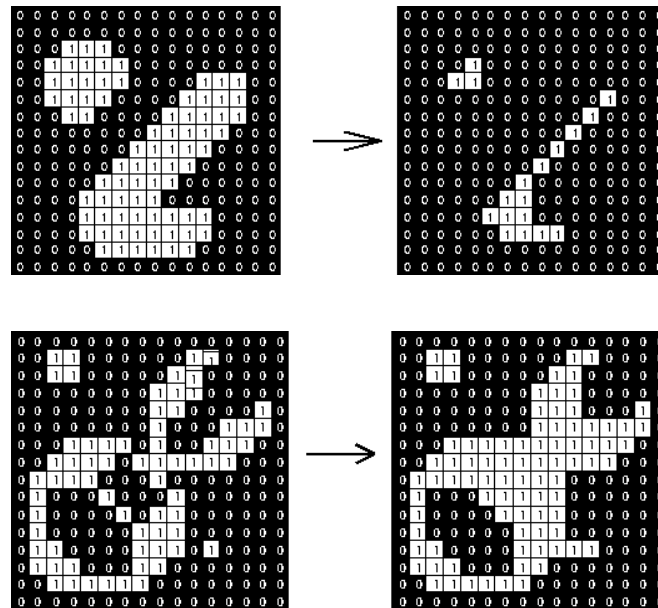
Original image

Gaussian Blur filter applied

รูปภาพที่ 8 Gaussian Blur

Morphological Image

Morphological เป็นเทคนิคหนึ่งในงาน Image Processing ที่ใช้ในการปรับเปลี่ยนหรือวิเคราะห์โครงสร้างทางรูปทรงเรขาคณิต (shape) ของวัตถุในภาพ เทคนิคนี้มักถูกใช้ในภาพแบบ binary images (ภาพที่มีเพียงสองสี เช่น ขาว-ดำ) และใช้ในการจัดการกับรูปร่างของวัตถุในภาพ เช่น การขยายวัตถุ การลดขนาด หรือการลบสัญญาณรบกวนที่เกิดขึ้นในภาพ Morphological มักจะใช้นงานต่าง ๆ เช่น การแยกแยะวัตถุ การเชื่อมต่อจุดต่าง ๆ ในภาพ และการลบสัญญาณรบกวน โดยมีการใช้โครงสร้างที่เรียกว่า structuring element ในการประมวลผล



รูปภาพที่ 9 Morphological Image

8. ขอบเขตของโครงการ

การจัดทำโครงการนี้เพื่อสร้างและนำเทคโนโลยี OCR ที่สร้างขึ้นมามีงานจัดการข้อมูลตรวจจับข้อมูลสำคัญจากใบทรานสคริปต์ โดยมีองค์ประกอบในการสร้างอยู่ 4 ส่วน ดังนี้

- 1) จัดทำชุดข้อมูลภาษาไทย ภาษาอังกฤษ และตัวเลขอารบิก สำหรับการฝึกโมเดล
 - 1.1 ชุดข้อมูลตัวอักษรภาษาไทย
 - 1.2 ชุดข้อมูลตัวอักษรภาษาอังกฤษ
 - 1.3 ชุดข้อมูลตัวเลขอารบิก 0 – 9
- 2) OCR Engine
 - 2.1) สามารถรองรับได้ 2 ภาษาและตัวเลขอารบิก
 - 1) ภาษาไทย
 - 2) ภาษาอังกฤษ
 - 3) ตัวเลขอารบิก

2.2) ความสามารถของ OCR Engine

- 1) สามารถตรวจจับตัวอักษรแต่ละตัวอักษรได้
- 2) สามารถตรวจจับเป็นข้อความในใบทรานสคริปต์
- 3) สามารถบอกพิกัดของตัวอักษรหรือข้อความที่ตรวจจับได้

3) OCR Transcript

3.1) ใบทรานสคริปต์ของสำนักงานคณะกรรมการการอาชีวศึกษาระดับปวช. และระดับปวส.

- 1) สามารถตรวจจับชื่อ-นามสกุล
- 2) สามารถตรวจจับวุฒิการศึกษา
- 3) สามารถตรวจจับสาขาวิชา
- 4) สามารถตรวจจับรหัสวิชา
- 5) สามารถตรวจจับผลการเรียน
- 6) สามารถตรวจจับจำนวนหน่วยกิตแต่ละวิชา
- 7) สามารถตรวจจับระดับคะแนนเฉลี่ยสะสม

3.2) ใบทรานสคริปต์ของกระทรวงศึกษาธิการระดับมัธยมศึกษาปีที่ 6

- 1) สามารถตรวจจับชื่อ-นามสกุล
- 2) สามารถตรวจจับวุฒิการศึกษา
- 3) สามารถตรวจจับรหัสวิชา
- 4) สามารถตรวจจับผลการเรียน
- 5) สามารถตรวจจับจำนวนหน่วยกิตแต่ละวิชา
- 6) สามารถตรวจจับหน่วยกิตและผลการเรียนเฉลี่ยของวิชาคณิตศาสตร์และวิชาวิทยาศาสตร์
- 7) สามารถตรวจจับระดับคะแนนเฉลี่ยสะสม

4) Web Application

- 4.1) เพื่อนำ OCR engine ที่สร้างมาใช้ใน platform Web application
- 4.2) สามารถอัปโหลดรูปภาพเข้าไปใน Web application
- 4.3) แสดงผลลัพธ์ข้อมูลทรานสคริปต์ออกมาในรูปแบบชุดข้อมูล Json
- 4.4) แสดงผลลัพธ์ข้อมูลทรานสคริปต์ออกมาในรูปแบบชุดข้อมูล Text

9. วิธีการดำเนินโครงการ

การพัฒนาระบบถอดข้อความจากทรานสคริปต์ เกิดขึ้นจากความต้องการถอดข้อความจากทรานสคริปต์ เพื่อนำข้อมูลในแบบรูปภาพถอดออกมาเป็นข้อมูลดิจิทัลได้อย่างสะดวก สามารถนำข้อมูลที่ได้ไปใช้ในการเทียบโอนรายวิชาต่าง ๆ ตามความต้องการได้ง่าย ทางคณะผู้จัดทำจึงได้ทำการวางแผนการดำเนินงาน เพื่อที่ทางคณะผู้จัดทำจะสามารถดำเนินงานได้ทันกำหนดเวลาตามระยะเวลาที่มีจำกัดและสามารถแบ่งแผนการดำเนินงานได้ เป็นดังนี้

9.1 การวางแผนโครงการ

วัตถุประสงค์ ระบุเป้าหมายหลักของระบบถอดข้อความจากทรานสคริปต์ เช่น การแปลงข้อความจากเอกสารสแกนหรือรูปภาพให้เป็นข้อความที่แก้ไขได้

การศึกษาความเป็นไปได้ (Feasibility Study) วิเคราะห์ความเป็นไปได้ทางเทคนิคและธุรกิจ โดยพิจารณาทรัพยากรที่ต้องใช้ เช่น ฮาร์ดแวร์, ซอฟต์แวร์, งบประมาณ และเวลาในการพัฒนา

การกำหนดขอบเขตของโปรเจกต์ วางขอบเขตของระบบ เช่น ระบบจะรองรับไฟล์รูปแบบใดบ้าง (เช่น PDF, PNG, JPG) และภาษาที่ต้องการสนับสนุนในการถอดข้อความ

9.2 การวิเคราะห์ความต้องการ สอบถามความต้องการข้อมูลที่ได้จากการถอดข้อความจากทรานสคริปต์

เช่น เอกสารหรือรูปภาพ ทรานสคริปต์เมื่อถอดข้อความแล้วจำเป็นต้องได้ข้อมูล เช่น ชื่อ-นามสกุล รหัสวิชา ผลการเรียนรู้ ระดับคะแนนเฉลี่ย เป็นต้น เพื่อนำข้อมูลที่ได้ไปต่อยอดในงานอื่นต่อไป

9.3 การออกแบบระบบ

9.3.1 สถาปัตยกรรมระบบแบบ Web-based Application

9.3.2 โมดูลหลักได้แก่ Image Preprocessing, OCR Processing, Postprocessing, และ File Handling

9.3.3 การออกแบบ UI/UX ให้ใช้งานง่าย เช่น การอัปโหลดไฟล์และการแสดงผลลัพธ์ที่ชัดเจน

9.3.4 การจัดการประสิทธิภาพด้วย Multithreading, Caching, และ Load Balancing

9.4 การสร้างและพัฒนาระบบ

สร้างและพัฒนาตามขั้นตอนการทำงานของระบบระบบถอดข้อความจากทรานสคริปต์ ตามแบบ Block Diagram

9.5 การทดสอบระบบ

9.5.1 Unit Testing ทดสอบการทำงานของโมดูลแต่ละตัว เช่น การทดสอบฟังก์ชันการแปลงภาพเป็นข้อความว่าทำงานถูกต้องหรือไม่

9.5.2 Integration Testing ทดสอบการทำงานร่วมกันของโมดูล เช่น การทำงานของ OCR ร่วมกับการอัปโหลดไฟล์และการแสดงผลลัพธ์, การทำงานร่วมกันของภาษาไทย ภาษาอังกฤษ ตัวเลข เป็นต้น

9.5.3 System Testing ทดสอบระบบทั้งหมดเพื่อดูว่าทำงานได้ถูกต้องตามความต้องการ เช่น ระบบสามารถถอดข้อความจากไฟล์ทรานสคริปต์ในรูปแบบต่าง ๆ ได้อย่างถูกต้องและครบถ้วน

9.5.4 Performance Testing ทดสอบประสิทธิภาพของระบบ เช่น ความเร็วในการประมวลผลทรานสคริปต์ขนาดใหญ่ หรือระบบสามารถประมวลผลไฟล์หลาย ๆ ไฟล์พร้อมกันได้อย่างเสถียร

9.5.5 User Acceptance Testing (UAT) ทดสอบระบบกับผู้ใช้จริง เพื่อให้มั่นใจว่าระบบตอบสนองความต้องการของผู้ใช้ได้ครบถ้วน

9.6 การติดตั้งและการใช้งาน

นำ OCR Transcript ที่เราสร้างและพัฒนาไปรันบน Web Application เพื่อให้ผู้ใช้สามารถอัปโหลดรูปภาพใบทรานทรานสคริปต์และใช้งาน ตัว OCR ได้

9.7 การบำรุงรักษาระบบ

9.7.1 การแก้ไขข้อผิดพลาด (Bug Fixing) เมื่อระบบถูกใช้งานในระยะยาว อาจพบข้อผิดพลาดบางประการที่ต้องได้รับการแก้ไข

9.7.2 การอัปเดตและปรับปรุงระบบ หากต้องมีการเพิ่มฟีเจอร์ใหม่ ๆ เช่น การรองรับภาษาหรือ
 ฟอนต์ใหม่ ระบบจะต้องถูกปรับปรุงอย่างต่อเนื่อง

9.7.3 การตรวจสอบประสิทธิภาพ ตรวจสอบให้แน่ใจว่าระบบยังคงทำงานได้อย่างเสถียรและประสิทธิภาพไม่ลดลงเมื่อปริมาณงานเพิ่มขึ้น

9.8 จัดทำปริญญานิพนธ์

ตารางแผนการดำเนินโครงการ

[illegible]

10. ประโยชน์ที่คาดว่าจะได้รับ

- 10.1 ได้รับความรู้เกี่ยวกับปัญญาประดิษฐ์ และ Computer Vision
- 10.2 ได้ระบบถอดข้อความจากทรานสคริปต์
- 10.3 ได้รับเครื่องมือที่ช่วยประหยัดเวลาในงานเอกสารเกี่ยวกับทรานสคริปต์
- 10.4 ได้ชุดข้อมูลตัวอักษรไทย อังกฤษ และตัวเลข ที่ผู้อื่นสามารถนำไปใช้งานหรือต่อยอดได้

11. งบประมาณการดำเนินโครงการ

ลำดับ	รายการ	จำนวนเงิน (บาท)
1	ค่าเช่าเซิร์ฟเวอร์ Cloud รายปี	5,000
2	ค่าคอร์สออนไลน์เสริมทักษะความรู้	1,000
3	ค่าทำรูปเล่มรายงาน	500

รวมทั้งสิ้น 6,500 บาท

12. เอกสารอ้างอิง

- [1] ทศนัย ชุ่มวัฒนะ, วรัมพร รัตนอำนวยชัย, & ภัทรานุช ช่วยชู. (2565). การสกัดข้อมูลผู้ป่วยด้วยเทคนิคตัวช่วยแปลงไฟล์เอกสาร. วารสารเทคโนโลยีสารสนเทศทางการแพทย์. สืบค้นจาก <https://he03.tci-thaijo.org/index.php/jtmi/article/download/198/91/416>
- [2] วิรุฬห์ ศรีบริรักษ์. (2562). เทคโนโลยีเฝ้าชันสารสนเทศแสดงข้อมูลหลากหลายเอกสารกำกับยาแบบอัตโนมัติสำหรับบริหารจัดการคลังยาปฏิชีวนะเพื่อความปลอดภัย. บัณฑิตวิทยาลัย, มหาวิทยาลัยบูรพา. สืบค้นจาก https://buuir.buu.ac.th/bitstream/1234567890/3738/3/2563_231.pdf
- [3] สรวิทย์ โพธิ์พยอม. (2562). การตรวจจับและรู้จำข้อความเพื่อการทำงานแบบอัตโนมัติในภาคอุตสาหกรรม. วิทยานิพนธ์ปริญญาโท, มหาวิทยาลัยนเรศวร. สืบค้นจาก <https://nuir.lib.nu.ac.th/dspace/bitstream/123456789/1542/3/61062373.pdf>
- [4] นริศรา เหม้ง, & ปณชัย เทพรักษ์. (2564). ได้ศึกษาการตรวจสอบผลิตภัณฑ์นมโดยใช้เทคนิค Deep Learning. สถาบันดิจิทัล, มหาวิทยาลัยมหาสารคาม. สืบค้นจาก <http://digital.csmsu.net:8080/library/handle/123456789/133>

- [5] จิรโรจน์ โสอุดร. (2562). ระบบส่งเสริมการท่องเที่ยวจากข้อมูลการเข้าพักของนักท่องเที่ยว. จุฬาลงกรณ์มหาวิทยาลัย. สืบค้นจาก <https://digital.car.chula.ac.th/chulaetd/9199/>
- [6] วิสรุต เหล่าดารา. (2565). ระบบเซ็นเซอร์ที่ออกแบบคัดลอกจากเอกสารสแกนคำพิพากษาด้วยปัญญาประดิษฐ์. มหาวิทยาลัยธุรกิจบัณฑิต. สืบค้นจาก <https://libdoc.dpu.ac.th/thesis/Wisarut.Kang.pdf>

13. คณะกรรมการ

- [illegible]