ระบบการแปลงไฟล์ภาพเอกสารให้เป็นข้อความอิเล็กทรอนิกส์

OPTICAL CHARACTER RECOGNITION (OCR) SYSTEM

ภีรดา ชีพสมทรง

สุชาครีย์ ชาติพจน์

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2562

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ระบบการแปลงไฟล์ภาพเอกสารให้เป็นข้อความอิเล็กทรอนิกส์

นางสาวภีรดา ชีพสมทรง 59011068

นายสุชาครีย์ ชาติพจน์ 59011427

ผศ.ดร.ชุติเมษฏ์ ศรีนิลทา อาจารย์ที่ปรึกษา

ปีการศึกษา 2562

บทคัดย่อ

เนื่องจากพบปัญหาในบริษัทรับทำบัญชี ที่เมื่อมีบริษัทอื่นมาจ้างในจัดการด้านบัญชี ก็จะให้มาแต่เอกสารในรูปแบบเอกสารสิ่งพิมพ์ที่มีแบบฟอร์มที่เหมือน ๆ กันจำนวนมาก แต่ปัญหาก็คือไม่มีไฟล์ต้นฉบับแนบมาด้วย ดังนั้นถ้าหากต้องการจัดการกับข้อมูล จึงจำเป็นที่จะต้องใช้ปริมาณคนและเวลาที่มาก อีกทั้งอาจจะมีปัญหาเอกสารสูญหาย ด้วยเหตุนี้ทางบริษัททำบัญชีต่างๆ จึงให้ความสนใจกับเทคโนโลยีที่จะสามารถมาประยุกต์ใช้ในการจัดการกับเอกสาร ซึ่งก็คือ โปรแกรมที่สามารถแปลงไฟล์ภาพเอกสารให้เป็นข้อความอิเล็กทรอนิกส์แทนการพิมพ์เอกสารขึ้นมาใหม่ หรือที่รู้จักกันในชื่อ “โอซีอาร์” โดยมีเปิดให้ใช้บริการทั้งแบบไม่คิดค่าใช้จ่าย และแบบคิดค่าใช้จ่าย แต่โอซีอาร์ที่เปิดให้ใช้งานส่วนใหญ่นั้น จะให้ผลลัพธ์เป็นข้อความอิเล็กทรอนิกส์ หรือไม่ก็ข้อความอิเล็กทรอนิกส์กับตำแหน่งของข้อความอิเล็กทรอนิกส์นั้น หรือก็คือไม่มีระบบฐานข้อมูลและแม่แบบ จึงทำให้ต้องทำการเก็บข้อมูลหลังจากการใช้บริการโอซีอาร์อีก ซึ่งการจัดเก็บข้อมูลของข้อความอิเล็กทรอนิกส์ที่ผ่านการแปลงโดยโอซีอาร์ให้เป็นหมวดหมู่ หรือหัวข้อเดียวกันให้อยู่ด้วยกันก็จะต้องให้ผู้ใช้เป็นผู้ตัดสินใจและจัดเก็บตามดุลพินิจของผู้ใช้ที่ต้องผ่านจากการอ่านข้อความอิเล็กทรอนิกส์ที่ผ่านการแปลงโดยโอซีอาร์อีกครั้ง

โดยปริญญานิพนธ์นี้ได้จะนำการเรียนรู้เชิงลึกมาประยุกต์ใช้ในการทำระบบการแปลงไฟล์ภาพเอกสารแบบฟอร์มในรูปแบบตัวพิมพ์สำหรับภาษาไทยหรือภาษาอังกฤษให้เป็นข้อความอิเล็กทรอนิกส์ โดยจะมีเว็บแอปพลิเคชันที่เป็นเครื่องมือที่ช่วยในการจัดการแม่แบบก่อนแปลงผลผ่านโอซีอาร์ และจัดการในการเก็บข้อมูลหลังแปลงผลจากโอซีอาร์ เพื่อให้สามารถนำข้อมูลไปใช้ต่อได้สะดวกยิ่งขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Optical Character Recognition (OCR) System

Ms. Peerada Cheepsomsong 59011068

Mr. Suchakree Chatpoch 59011427

Asst.Prof.Dr. Chutimet Srinilta Advisor

Academic Year 2019

ABSTRACT

A pre-printed form contains a number of fields to be filled out. Each field has certain meaning. It holds data. Most of time, documents given to accounting firms are paper-based pre-printed forms that are filled with printed text. Before these documents can be processed further, data must be extracted from them. Such data must be added to the database according to their labels in the form.

Optical Character Recognition(OCR) converts image documents to digital text. Objective of our project is to do the OCR for printed text on pre-printed form and insert the extracted text into the database.

We have developed a web application that helps us manage document templates and data collection easily by OCR in order to make the information ready to be processed further.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

ปริญญานิพนธ์ฉบับนี้ สามารถสำเร็จลุล่วงได้ด้วยความกรุณาจาก ผศ.ดร.ชุติเมษฏ์ ศรีนิลทา อาจารย์ที่ปรึกษาปริญญานิพนธ์ที่ได้ให้คำแนะนำ คำปรึกษา ตลอดจนแก้ไขข้อบกพร่องต่าง ๆ มาโดยตลอด จนปริญญานิพนธ์นี้เสร็จสมบูรณ์ คณะผู้จัดทำปริญญานิพนธ์จึงขอขอบพระคุณอาจารย์เป็นอย่างสูง

ขอขอบพระคุณคณะอาจารย์ทุกท่านที่สั่งสอนให้ความรู้แก่คณะผู้จัดทำปริญญานิพนธ์ตั้งแต่อดีตจนถึงปัจจุบัน จนทำให้มีความสามารถที่ทำปริญญานิพนธ์นี้จนเสร็จสมบูรณ์ได้ ซึ่งคณะผู้จัดทำปริญญานิพนธ์ได้ซาบซึ้งพระคุณคณะอาจารย์ทุกท่านเป็นอย่างสูง

ขอขอบพระคุณภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ที่เอื้อเฟื้อสถานที่ในการทำงาน หาความรู้ และอำนวยความสะดวกในการทำปริญญานิพนธ์นี้

และสุดท้ายขอขอบพระคุณครอบครัว เพื่อนและบุคคลทุกท่านที่เป็นกำลังใจ และสนับสนุนทุกสิ่งที่ส่งผลให้ปริญญานิพนธ์นี้สำเร็จลุล่วงได้ด้วยดี

ภีรดา ชีพสมทรง

สุชาครีย์ ชาติพจน์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

ระบบการแปลงไฟล์ภาพเอกสารให้เป็นข้อความอิเล็กทรอนิกส์ I

Optical Character Recognition (OCR) System II

กิตติกรรมประกาศ III

สารบัญ IV

สารบัญตาราง VI

สารบัญรูป VII

บทที่ 1 1

1.1 ความเป็นมาของปัญหา 1

1.2 วัตถุประสงค์ของปริญญานิพนธ์ 2

1.3 ประโยชน์ที่คาดว่าจะได้รับ 2

1.4 ขอบเขตของปริญญานิพนธ์ 2

บทที่ 2 3

2.1 ทฤษฎีที่เกี่ยวข้อง 3

2.2 งานวิจัยที่เกี่ยวข้อง 15

บทที่ 3 17

3.1 ภาพรวมของระบบ 17

3.2 แผนภาพการทำงานของผู้ใช้ 19

3.3 การทำงานของระบบโอซีอาร์ 20

3.4 แผนภาพแสดงความสัมพันธ์ของข้อมูล 21

บทที่ 4 23

4.1 การทดลองส่วนโมเดลทำนายข้อความอิเล็กทรอนิกส์จากไฟล์ภาพเอกสาร 23

4.2 การใช้งานส่วนเว็บแอปพลิเคชัน 29

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ(ต่อ)

หน้า

บทที่ 5 41

5.1 สรุปผลที่ได้จากปริญญานิพนธ์ 41

5.2 ปัญหาและอุปสรรค 42

5.3 แนวทางการพัฒนาต่อ 42

บรรณานุกรม 43

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตาราง **หน้า**

ตาราง 4.1 ตัวอย่างผลการทดลองจากชุดข้อมูลสำหรับทดสอบจากการสุ่มตัดภาพเอกสารตัวพิมพ์ที่ผ่านการแสกนจากอินเทอร์เน็ต 150 รูป.....26
ตาราง 4.2 ตัวอย่างผลการทดลองจากชุดข้อมูลสำหรับทดสอบจากการ ICDAR2003 1157 รูป และ ICDAR2019 848 รูป.....28

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูป **หน้า**

รูป 2.1 เครื่องมือที่ใช้ในเก็บรูปภาพสำหรับนำไปใช้ในกระบวนการ 3
รูป 2.2 แสดงตัวอย่างการทำงานของ Character Region Awareness For Text Detection (CRAFT)5
รูป 2.3 ตัวกรอง 3x3 สำหรับหาเส้นตรงทะแยงสีขาว 6
รูป 2.4 การคูณเวกเตอร์เชิงสเกลลาร์ระหว่างพื้นที่ของภาพที่ถูกตัวกรองทาบกับ ตัวกรอง แล้วทำการเก็บผลลัพธ์ลงในผังคุณลักษณะ 6
รูป 2.5 การทำ Max Pooling 7
รูป 2.6 โครงสร้างทั่วไปของ Bidirectional Recurrent Neural Networks 7
รูป 2.7 ภาพรวมของโครงข่ายประสาทสำหรับการรู้จำลายมือ 8
รูป 2.8 โลโก้ภาษา Python 9
รูป 2.9 โลโก้ภาษา Typescript 9
รูป 2.10 โลโก้โปรแกรม PyCharm 10
รูป 2.11 โลโก้โปรแกรม Visual Studio Code 10
รูป 2.12 โลโก้ไลบรารี PyTorch 11
รูป 2.13 โลโก้ไลบรารี OpenCV 11
รูป 2.14 โลโก้ฐานข้อมูล MongoDB 12
รูป 2.15 โลโก้ Node.js 12
รูป 2.16 โลโก้ React 13
รูป 2.17 โลโก้ Redux 13
รูป 2.18 โลโก้ Docker 14
รูป 2.19 โลโก้ Jenkins 14
รูป 2.20 สถาปัตยกรรมเครือข่าย 15
รูป 2.21 การแสดงภาพของบริบทและเฟรมสัญลักษณ์สำหรับการระบุสัญลักษณ์ที่เกี่ยวกับตัวอักษรโดยรอบ แต่ละเฟรมเป็นอินพุตโอซีอาร์สำหรับหนึ่งขั้นตอน 16
รูป 3.1 ภาพรวมของระบบ 17
รูป 3.2 แผนภาพการทำงานของผู้ใช้ 19
รูป 3.3 การทำงานของระบบโอซีอาร์ 20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป(ต่อ)

รูป **หน้า**

รูป 3.4 แผนภาพแสดงความสัมพันธ์ของข้อมูล.....21
รูป 4.1 แผนภาพขั้นตอนการทดลอง.....25
รูป 4.2 หน้าแสดงผลของการยืนยันตัวตน.....29
รูป 4.3 หน้าแสดงผลของการสมัครสมาชิก.....31
รูป 4.4 หน้าแสดงผลของโปรเจค.....32
รูป 4.5 หน้าแสดงผลของแม่แบบในโปรเจค.....34
รูป 4.6 หน้าแสดงผลของการสร้างแม่แบบในโปรเจค.....37
รูป 4.7 หน้าแสดงผลของตารางแสดงผลของโปรเจค.....39

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ความเป็นมาของปัญหา

ในปัจจุบัน มีไฟล์เอกสารที่ทำการพิมพ์ลงบนกระดาษแล้วเมื่อเวลาผ่านไปไฟล์เอกสารนั้น ๆ ก็อาจสูญหาย หรือทำการลบออกไปเพื่อให้พื้นที่ในการเก็บกับไฟล์เอกสารใหม่ ดังนั้นเอกสารต่าง ๆ นั้นถูกจัดเก็บในรูปแบบกระดาษ ทำให้อาจมีปัญหาที่เกิดจากการจัดเก็บเอกสารจำนวนมากรวมกันก็จะส่งผลให้การกลับมาศึกษาและค้นหาเอกสารเพื่อนำข้อมูลกลับมาใช้ใหม่นั้นเป็นไปได้อย่างยากลำบาก เนื่องจากเอกสารมีจำนวนที่มากเกินกว่าจะสามารถศึกษาและค้นหาได้ในช่วงเวลาสั้น ๆ แล้วอาจมีกรณีที่มีเอกสารบางส่วนสูญหายจากการจัดเก็บต่างที่กัน หรือสูญหายจากการนำออกมาใช้แล้วนำกลับไปส่งคืนที่จัดเก็บไม่ครบตามจำนวนที่นำออกมา หรืออาจมีปัญหาอื่น ๆ อีกมากมาย ซึ่งจะส่งผลทำให้การนำข้อมูลกลับมาใช้ใหม่นั้นอาจมีข้อผิดพลาดได้ หรือไม่พบข้อมูลที่ต้องการนำมาใช้

ในขณะเดียวกันก็มีการใช้งานคอมพิวเตอร์อย่างแพร่หลาย โดยมีผู้ใช้จำนวนมากที่ต้องการโปรแกรมที่สามารถแปลงไฟล์ภาพเอกสารให้เป็นข้อความอิเล็กทรอนิกส์แทนการพิมพ์เอกสารขึ้นมาใหม่ ซึ่งเป็นที่รู้จักกันในชื่อ “โอซีอาร์” โดยทั่วไปแล้วโอซีอาร์ตัวพิมพ์มีให้เลือกใช้บริการจำนวนมากพอสมควร ซึ่งมีเปิดให้ใช้บริการทั้งแบบไม่คิดค่าใช้จ่าย และแบบคิดค่าใช้จ่าย โดยโอซีอาร์ที่เปิดให้ใช้งานนั้นก็จะให้ผลลัพธ์จากการแปลงโดยโอซีอาร์เป็นข้อความอิเล็กทรอนิกส์หรือไม่ก็ข้อความอิเล็กทรอนิกส์กับตำแหน่งของข้อความอิเล็กทรอนิกส์นั้น หรือก็คือไม่มีระบบฐานข้อมูลและแม่แบบ หากต้องการจะจัดเก็บข้อความอิเล็กทรอนิกส์ที่ผ่านการแปลงโดยโอซีอาร์ให้เป็นหมวดหมู่ หรือหัวข้อเดียวกันให้อยู่ด้วยกันก็จะต้องให้ผู้ใช้เป็นผู้ตัดสินใจและจัดเก็บตามดุลพินิจของผู้ใช้

ตัวอย่างที่เห็นได้ชัดที่ได้เจอกับปัญหาที่ได้กล่าวไว้ก่อนหน้า ก็คงจะเป็นปัญหาของบริษัทรับทำบัญชี เมื่อเวลามีบริษัทอื่นมาจ้างให้จัดการด้านบัญชี บริษัทที่มาทำการจ้างนั้นก็จะให้เอกสารมาในรูปแบบเอกสารสิ่งพิมพ์ที่มีจำนวนมาก และไม่มีไฟล์ต้นฉบับแนบมาให้ด้วย ดังนั้นถ้าหากต้องการจะจัดการข้อมูล จึงต้องใช้จำนวนคนและเวลาที่มาก อีกทั้งยังอาจมีปัญหาเอกสารสูญหายด้วยเหตุนี้ทางบริษัททำบัญชีต่าง ๆ จึงให้ความสนใจกับเทคโนโลยีที่จะสามารถมาประยุกต์ใช้ในการจัดการกับเอกสารเหล่านี้ได้เป็นอย่างมาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ทำให้คณะผู้จัดทำปริญญานิพนธ์มีความสนใจที่จะทำการศึกษาและจัดทำการแปลงไฟล์ภาพเอกสารแบบฟอร์มในรูปแบบตัวพิมพ์สำหรับภาษาไทยหรือภาษาอังกฤษให้เป็นข้อความอิเล็กทรอนิกส์ โดยระบบจะเป็นเครื่องมือที่ช่วยการจัดการแม่แบบก่อนแปลงผลผ่านโอซีอาร์และจัดการในการเก็บข้อมูลหลังแปลงผลผ่านโอซีอาร์ ทำให้สามารถนำข้อมูลไปใช้ต่อได้สะดวกสบายยิ่งขึ้น

1.2 วัตถุประสงค์ของปริญญานิพนธ์

1) เพื่อศึกษาการเรียนรู้ข้อมูลโดยใช้วิธีการเรียนรู้เชิงลึก (Deep Learning)
2) เพื่อศึกษาการทำเว็บแอปพลิเคชัน (Web Application)
3) เพื่อประยุกต์ใช้การเรียนรู้เชิงลึก ในการสร้างโมเดลทำนายข้อความอิเล็กทรอนิกส์จากไฟล์ภาพเอกสาร
4) สร้างและพัฒนาเว็บแอปพลิเคชัน เพื่อเป็นระบบที่ใช้ในการสร้างแม่แบบ โดยการตีกรอบและติดป้ายส่วนที่ต้องการแปลงเป็นข้อความอิเล็กทรอนิกส์ และแสดงผลลัพธ์ที่ได้จากการแปลงไฟล์ภาพเอกสารให้เป็นข้อความอิเล็กทรอนิกส์

1.3 ประโยชน์ที่คาดว่าจะได้รับ

1) ได้รับความรู้และความเข้าใจเกี่ยวกับการเรียนรู้เชิงลึก (Deep Learning)
2) ได้รับความรู้และความเข้าใจเกี่ยวการทำเว็บแอปพลิเคชัน (Web Application) อย่างละเอียด
3) ได้ระบบที่ช่วยในการจัดการการแปลงไฟล์ภาพเอกสารให้เป็นข้อความอิเล็กทรอนิกส์ให้เก็บตามหมวดหมู่ หรือหัวข้อที่ต้องการ และสะดวกสบายต่อผู้ใช้งานในการจัดเก็บและค้นหาข้อมูล

1.4 ขอบเขตของปริญญานิพนธ์

1) การศึกษาเกี่ยวกับการเรียนรู้เชิงลึก ในการประยุกต์ใช้ในการสร้างโมเดลทำนายข้อความอิเล็กทรอนิกส์จากไฟล์ภาพเอกสารที่ได้จากการพิมพ์เท่านั้น
2) การแปลงไฟล์ภาพเอกสารให้เป็นข้อความอิเล็กทรอนิกส์ จะทำการแปลงเฉพาะส่วนที่ทำการตีกรอบและติดป้ายในแม่แบบเท่านั้น
3) ส่วนของเว็บแอปพลิเคชัน ไว้ใช้ในการตีกรอบและติดป้ายในแม่แบบ และแสดงผลลัพธ์ที่ได้จากการแปลงไฟล์ภาพเอกสารให้เป็นข้อความอิเล็กทรอนิกส์เท่านั้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 การรู้จำอักขระด้วยแสง

ชื่อในภาษาอังกฤษก็คือ Optical Character Recognition ซึ่งเป็นที่รู้จักกันใน “โอซีอาร์ (OCR)” โดยโอซีอาร์เป็นกระบวนการทางกลไกหรืออิเล็กทรอนิกส์ของการแปลสื่อสิ่งพิมพ์หรือรูปภาพของข้อความจากการเขียน หรือการพิมพ์ให้กลายเป็นข้อความที่สามารถแก้ไขได้โดยเครื่องคอมพิวเตอร์

2.1.1.1 เครื่องมือที่ใช้ในเก็บรูปภาพสำหรับนำไปใช้ในกระบวนการ

1) เครื่องสแกนเนอร์
2) กล้องถ่ายรูประบบดิจิทัล

ก) ข)

รูป 2.1 เครื่องมือที่ใช้ในเก็บรูปภาพสำหรับนำไปใช้ในกระบวนการ

ก) เครื่องสแกนเนอร์

ข) กล้องถ่ายรูประบบดิจิทัล

(ที่มา:
https://mediaserver.goepson.com/ImConvServlet/imconv/d5f1e98f412f91278735f36201b5fc923f84f44d/300Wx300H?use=productpictures&assetDescr=Support-Scanner-Type-Photo-Scanners-V39_RT_ANG_OPEN,
https://www.jib.co.th/img_master/product/icon/20180828110640_30072_24_1.png)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.1.2 สาขาวิจัยของโอซีอาร์

2.1.1.2.1 การรู้จำแบบ

หรือที่รู้จักกันในชื่อ pattern recognition ซึ่งเป็นสาขาย่อยหนึ่งของวิทยาการคอมพิวเตอร์ เป็นศาสตร์ที่วัตถุประสงค์ในการจำแนกวัตถุ(object) ออกเป็นประเภท (classes) ตามรูปแบบของวัตถุ

2.1.1.2.2 ปัญญาประดิษฐ์

หรือที่รู้จักกันในชื่อ artificial intelligence หรือ “เอไอ(AI)” ซึ่งเป็นสาขาหนึ่งในวิทยาการคอมพิวเตอร์และวิศวกรรมเป็นหลัก อีกทั้งยังมีศาสตร์ในด้านอื่น ๆ โดยเป็นศาสตร์ที่เรียนรู้เกี่ยวกับกระบวนการคิด การกระทำ การให้เหตุผล การปรับตัว หรือการอนุมาน และการทำงานของสมอง

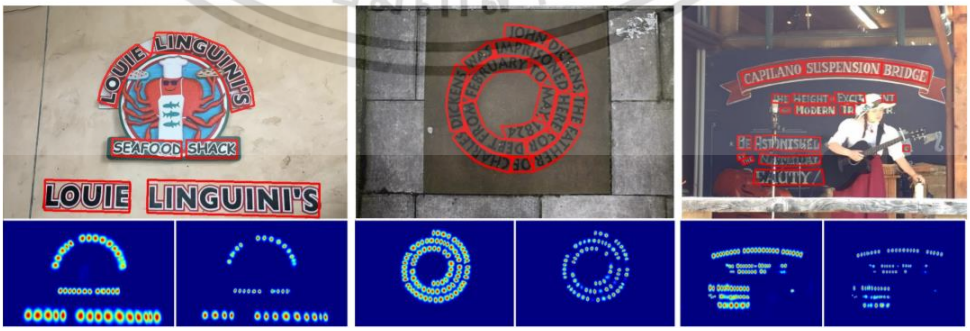
2.1.1.2.3 คอมพิวเตอร์วิทัศน์

หรือที่รู้จักกันในชื่อ computer vision เป็นสาขาหนึ่งของวิทยาการคอมพิวเตอร์ เป็นศาสตร์ที่เกี่ยวกับการดึงสารสนเทศจากรูปภาพ หรือวิดิทัศน์

2.1.2 เทคนิคและเทคโนโลยีที่ใช้

2.1.2.1 Character Region Awareness For Text Detection (CRAFT)

วิธีการตรวจหาพื้นที่ข้อความที่มีประสิทธภาพ โดยการสำรวจแต่ละตัวอักษรและความสัมพันธ์ระหว่างตัวอักษร เพื่อที่จะลดข้อบกพร่องของการอธิบายประกอบในระดับตัวอักษรแต่ละตัว โดยวิธีนี้มีความยืดหยุ่นสูงในการตรวจหาภาพข้อความในฉากที่ซับซ้อน อย่างเช่น ข้อความมีการเขียนเป็นเส้นโค้ง หรือข้อความผิดรูป



ก)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ข)

รูป 2.2 แสดงตัวอย่างการทำงานของ Character Region Awareness For Text Detection (CRAFT)

ก) แต่ละคอลัมน์จะแสดงภาพ ด้านบนเป็นภาพที่นำเข้ามา, ด้านล่างซ้ายเป็นแผนที่คะแนนบริเวณที่เกี่ยวข้อง และด้านล่างขวาเป็นแผนที่ความสัมพันธ์

ข) แต่ละคอลัมน์จะแสดงภาพ ด้านซ้ายเป็นภาพที่นำเข้ามา และด้านขวาเป็นแผนที่คะแนนบริเวณที่เกี่ยวข้อง

2.1.2.2 Convolutional Neural Network (CNN)

คือ โครงข่ายประสาทแบบคอนโวลูชัน(Convolution) เป็นโครงข่ายประสาทเทียมหนึ่งในกลุ่ม bio-inspired โดยจะทำการจำลองการมองเห็นของมนุษย์ที่มองพื้นที่เป็นที่ย่อย ๆ จะเสมือนเป็นการแยกคุณลักษณะ(feature)ของพื้นที่ย่อยนั้น อย่างเช่น ลายเส้น และการตัดกันของสี และนำกลุ่มของพื้นที่ย่อย ๆ มาผสานกัน เพื่อดูว่าสิ่งที่เห็นว่าคืออะไร

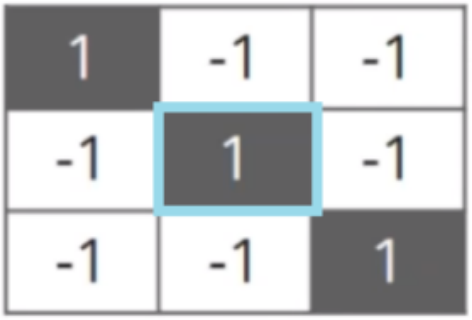
2.1.2.2.1 Feature Extraction

คือ การสกัดคุณลักษณะ โดยแนวคิดของ ซีเอ็นเอ็น(CNN) จะเริ่มจากการกำหนดค่าใน ตัวกรอง(filter) หรือ เคอร์เนล(kernel) ที่ช่วยดึงคุณลักษณะที่ใช้ในการรู้จำวัตถุออก โดยปกติหนึ่งตัวกรองจะดึงคุณลักษณะที่สนใจออกมาได้หนึ่งอย่าง จึงทำให้จำเป็นต้องตัวกรองหลายตัวกรอง เพื่อหาคุณลักษณะทางพื้นที่หลายอย่างประกอบกัน

2.1.2.2.1.1 ลักษณะของตัวกรอง

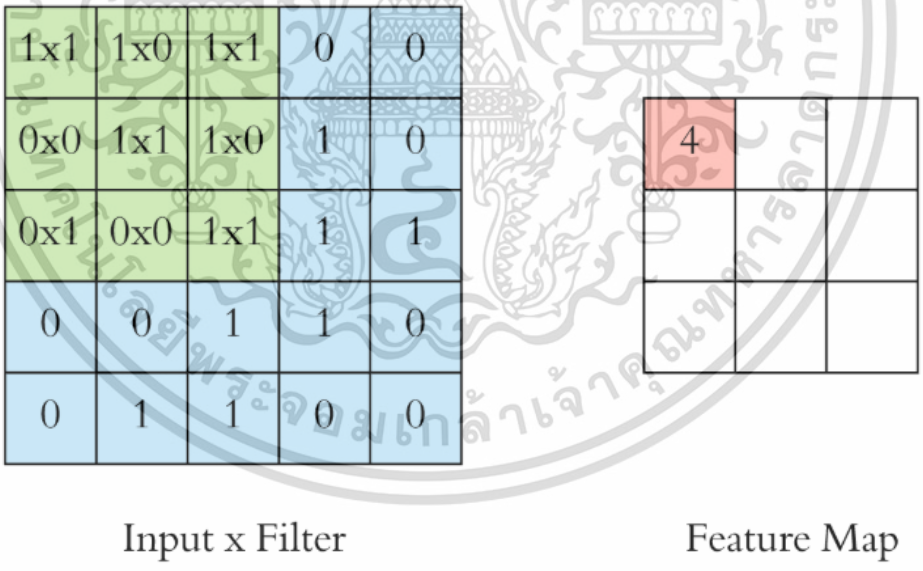
สำหรับลักษณะของตัวกรองของภาพดิจิทัล โดยปกติจะเป็นตารางสองมิติที่มีขนาดตามพื้นที่ย่อย ๆ ที่สนใจจะพิจารณา อย่างเช่น ต้องการหาเส้นทแยงสีขาว ตัวกรองก็อาจจะอยู่ในลักษณะดังรูปต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูป 2.3 ตัวกรอง 3x3 สำหรับหาเส้นตรงทะแยงสีขาว

ตำแหน่งตรงกลางที่มีกรอบสีฟ้าคือ Anchor ที่เอาไว้ทาบบนพิกเซลของภาพข้อมูลเข้า ตัวกรองจะถูกทาบลงในพิกเซลแรกของภาพข้อมูลเข้า จากนั้นจะถูกเลื่อนไปทาบบนพิกเซลอื่นในภาพทีละพิกเซลจนครบทุกพิกเซลในภาพ และอาจจะไม่ทาบตัวกรองบนพิกเซลที่อยู่ใกล้กรอบภาพ เพราะตัวกรองจะล้นออกไปนอกภาพ โดยขณะทำการทาบจะทำการคูณเวกเตอร์เชิงสเกลลาร์(dot product)ระหว่างพื้นที่ของภาพที่ถูกตัวกรองทาบกับตัวกรอง แล้วทำการเก็บผลลัพธ์ลงในผังคุณลักษณะ(feature map)



รูป 2.4 การคูณเวกเตอร์เชิงสเกลลาร์ระหว่างพื้นที่ของภาพที่ถูกตัวกรองทาบกับ ตัวกรอง แล้วทำการเก็บผลลัพธ์ลงในผังคุณลักษณะ

2.1.2.2.1.2 Stride และ Padding

Stride เป็นตัวกำหนดว่าจะเลื่อนตัวกรอง ไปด้วยจังหวะ (step)เท่าไร และถ้า Stride ยิ่งมีค่ามากก็จะยิ่งทำให้ผังคุณลักษณะมีขนาดที่เล็กลง

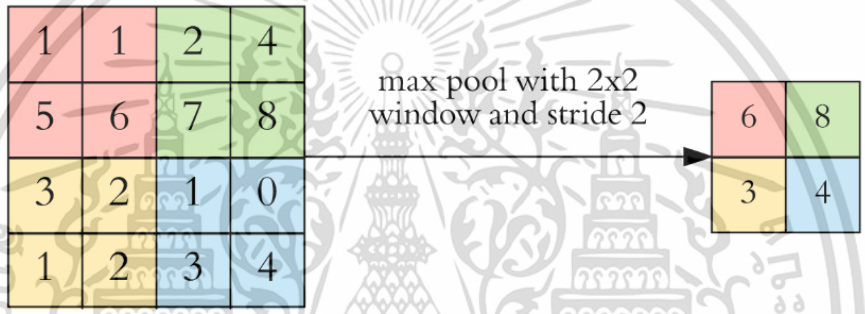
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Padding เป็นการขยายกรอบให้กับภาพ เพื่อให้เมื่อผ่านกระบวนการสกัดคุณลักษณะแล้วผังคุณลักษณะยังคงมีขนาดเท่ากับภาพเริ่มแรก

2.1.2.2.1.3 Max Pooling

Pooling คือความสามารถในการย่อรูปแบบหนึ่งที่ยังคงคุณลักษณะเด่นที่ตัวกรองทำไว้ ซึ่งมีสองประเภทหลักที่นิยมกันคือ max pooling และ mean pooling โดยสาเหตุที่ต้องใช้เนื่องจากอาจจะมีปัญหาที่ข้อมูลอาจะมีหลายขนาด

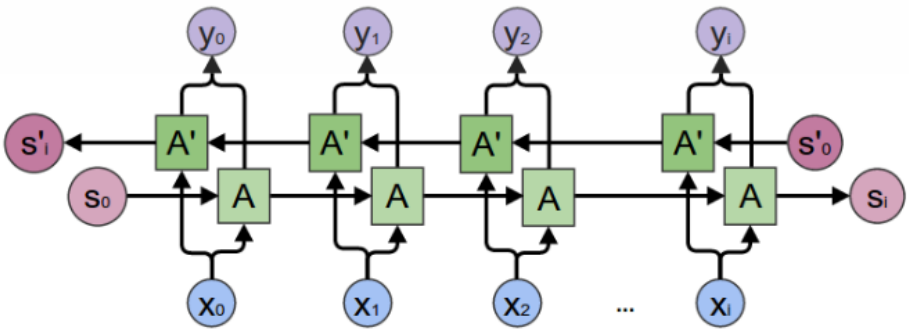
Max Pooling เป็นตัวกรองแบบหนึ่งที่หาค่าสูงสุดในบริเวณที่ตัวกรองทาบอยู่มาเป็นผลลัพธ์ โดยจะนิยมเรียกว่า “pool size”



รูป 2.5 การทำ Max Pooling

2.1.2.3 Bidirectional Gated Recurrent Unit (GRU)

เป็นประเภทของโครงข่ายประสาท(Neural Network) ที่เกิดซ้ำแบบสองทิศทาง จะช่วยให้การใช้ข้อมูลจากทั้งขั้นตอนก่อนหน้าและขั้นตอนถัดไปการคาดการณ์เกี่ยวกับสถานการณ์ปัจจุบัน

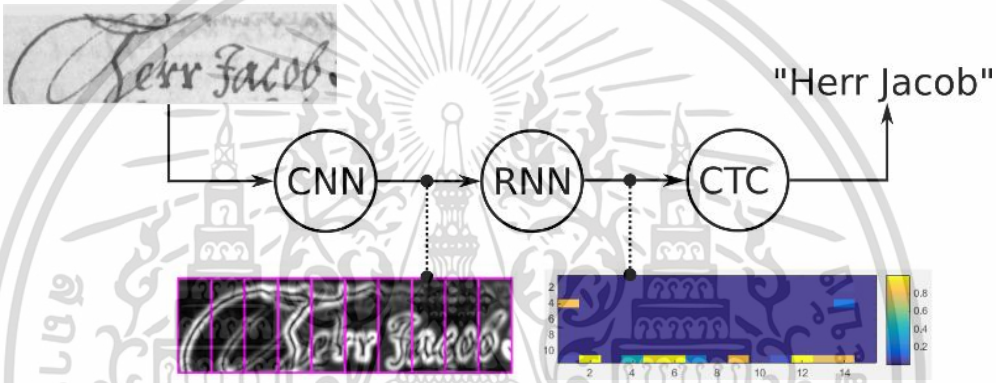


รูป 2.6 โครงสร้างทั่วไปของ Bidirectional Recurrent Neural Networks

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.2.4 Connectionist Temporal Classification(CTC) Loss

เป็นประเภทของเอาท์พุตโครงข่ายประสาท และฟังก์ชันการให้คะแนนที่เกี่ยวข้องสำหรับการฝึกอบรมเครือข่ายประสาทเป็นตัวเลือกที่ดี เนื่องจากมีประสิทธิภาพเหนือกว่าวิธีอื่น ในกรณีที่ทำการรู้จำลายข้อความ ส่วนมากจะประกอบไปด้วย 3 ขั้นได้แก่ Convolutional Layers (CNN) เพื่อแยกลำดับของคุณสมบัติ, Recurrent Layers (RNN) เพื่อเผยแพร่ข้อมูลผ่านลำดับนี้ โดยจะได้คะแนนของตัวอักษรสำหรับแต่ละลำดับองค์ประกอบอยู่ในรูปเมทริกซ์ และทำการการคำนวณค่าการสูญเสียในการฝึกอบรมเครือข่ายประสาท ตามด้วยการสรุปถอดรหัสเมทริกซ์ที่จะได้รับข้อความที่มีอยู่ในภาพที่นำเข้า โดยขั้นสุดท้ายเป็นการดำเนินการของ CTC



รูป 2.7 ภาพรวมของโครงข่ายประสาทสำหรับการรู้จำลายมือ

2.1.3 ภาษาและเครื่องมือที่ใช้ในการพัฒนา

2.1.3.1 ภาษา Python

คือ ภาษาโปรแกรมคอมพิวเตอร์ระดับสูง ซึ่งถูกพัฒนาขึ้นมาโดยไม่ยึดติดกับแพลตฟอร์ม หรือก็คือสามารถรันภาษา Python ได้ทั้งบนระบบ Unix, Linux , Windows NT และอื่น ๆ อีกมากมาย และภาษานี้ยังถูกพัฒนาออกมาในรูปแบบของโอเพนซอร์ส(OpenSource) ทำให้ทุกคนสามารถที่จะนำ Python มาพัฒนาโปรแกรมของได้ฟรี อีกทั้งยังทำให้มีคนเข้ามาช่วยกันพัฒนาให้ Python มีความสามารถสูงขึ้น และใช้งานได้ครบคุมกับทุกลักษณะงาน

ไวยากรณ์ของภาษาถูกพัฒนาขึ้นมาโดยมีความตั้งใจว่าจะให้เป็นภาษาที่อ่านง่าย และมีโครงสร้างที่มองเห็นได้โดยไม่ซับซ้อน ในส่วนของการแปลงชุดคำสั่งที่เขียนให้เป็นภาษาเครื่อง Python มีการทำงานแบบ Interpreter คือเป็นการแปลชุดคำสั่งทีละบรรทัด เพื่อป้อนเข้าสู่หน่วยประมวลผลให้คอมพิวเตอร์ทำงานตามที่ต้องการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

อีกทั้งมีไลบรารีทางด้านการคำนวณ, Machine learning, Deep learning เป็นจำนวณมาก

รูป 2.8 โลโก้ภาษา Python

(ที่มา: https://saixiii.com/python-programming/)

2.1.3.2 ภาษา Typescript

เป็นภาษาโปรแกรมที่รวมความสามารถที่ ES2015(ES6) มีอยู่ และมีคุณสมบัติที่เพิ่มเติมมาก็คือ Enum และความสามารถที่เพิ่มขึ้นของการโปรแกรมเชิงวัตถุ และภาษานี้เป็นจะมีตัวแปลภาษาของ TypeScript ให้เป็น JavaScript ก่อน เพื่อให้เบราว์เซอร์ทำงานได้ และที่สำคัญภาษานี้ยังถูกพัฒนาออกมาในรูปแบบของโอเพนซอร์ส ทำให้ทำงานได้กับระบบปฏิบัติการหลัก ๆ

โดยภาษานี้มีข้อดี ได้แก่

1) สามารถใช้งาน JavaScript ล่าสุดได้ อีกทั้งยังมีความสามารถของ ES2015 และอื่น ๆ ได้
2) เนื่องจากตัวแปรที่ประกาศไปแล้วจะไม่สามารถเปลี่ยนชนิดข้อมูลได้ ทำให้ข้อผิดพลาดในโปรแกรมน้อยลง
3) ภาษานี้มีการตรวจสอบโค้ดในช่วงเวลา compile ทำให้สามารถดักจับข้อผิดพลาดได้แต่ต้นไม่ให้ข้อผิดพลาดหลุดไปตอนทำงานจริง

รูป 2.9 โลโก้ภาษา Typescript

(ที่มา: https://devblogs.microsoft.com/typescript/)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.3.3 PyCharm

เป็นโปรแกรม IDE (Integrated Development Editor) สำหรับพัฒนาโปรแกรมภาษา Python ที่มีการใช้งานอย่างแพร่หลายในการพัฒนาโปรแกรมคอมพิวเตอร์ รองรับระบบปฏิบัติการ Windows, Linux, และMacOS ถูกพัฒนาขึ้นโดย JetBrains มีความสามารถในการวิเคราะห์โค้ด, ดีบัก (Debug), ทดสอบ (Test) ได้อย่างมีประสิทธิภาพ

รูป 2.10 โลโก้โปรแกรม PyCharm

(ที่มา: https://www.jetbrains.com/pycharm/)

2.1.3.4 Visual Studio Code (VSCode)

เป็นโปรแกรม Code Editor ที่ใช้ในการแก้ไขและปรับแต่งโค้ด จากค่ายไมโครซอฟท์ มีการพัฒนาออกมาในรูปแบบของโอเพนซอร์ส จึงสามารถนำมาใช้งานได้แบบฟรี ที่ต้องการความเป็นมืออาชีพ เหมาะสำหรับนักพัฒนาโปรแกรมที่ต้องการใช้งานข้ามแพลตฟอร์ม รองรับการใช้งานทั้งบน Windows, macOS และ Linux สนับสนุนทั้งภาษา JavaScript, TypeScript และ Node.js สามารถเชื่อมต่อกับ Git ได้ ทำให้ใช้งานได้ง่ายไม่ซับซ้อน อีกทั้งยังมีเครื่องมือส่วนขยายต่าง ๆ ให้เลือกใช้อย่างมาก

รูป 2.11 โลโก้โปรแกรม Visual Studio Code

(ที่มา: https://code.visualstudio.com/)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.3.5 PyTorch

เป็นไลบรารีซอฟต์แวร์โอเพนซอร์สของการเรียนรู้ของเครื่อง(Machine Learning) บนพื้นฐานของไลบรารีของTorch ที่ใช้สำหรับแอปพลิเคชัน อย่างเช่น คอมพิวเตอร์วิทัศน์(computer vision) และการประมวลผลภาษาธรรมชาติหรือภาษามนุษย์ (Natural Language Processing)

รูป 2.12 โลโก้ไลบรารี PyTorch

(ที่มา: https://pytorch.org/)

2.1.3.6 Open source Computer Vision (OpenCV)

เป็นไลบรารีฟังก์ชันการเขียนโปรแกรม (Library of Programming Functions) โดยส่วนใหญ่จะมุ่งเป้าไปที่การแสดงผลด้วยคอมพิวเตอร์แบบเรียลไทม์ (Real-Time Computer Vision) อีกทั้งยังสนับสนุนเฟรมเวิร์กการเรียนรู้เชิงลึก (Deep Learning Frameworks) ได้แก่ TensorFlow, Torch/PyTorch และ Caffe

รูป 2.13 โลโก้ไลบรารี OpenCV

(ที่มา: https://opencv.org/)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.3.7 MongoDB

เป็นฐานข้อมูลที่โอเพนซอร์ส โดยเป็นฐานข้อมูลแบบ NoSQL(Not Only SQL) คือไม่มีความสัมพันธ์(relation) ของตารางแบบ SQL(Structured Query Language) ทั่วไป แต่จะเก็บข้อมูลเป็นแบบ JSON (JavaScript Object Notation) แทนการบันทึกข้อมูลทุก ๆ บันทึก(record) ใน MongoDB เราจะเรียกมันว่า “Document” ซึ่งจะเก็บค่าเป็น key และ value และการเก็บข้อมูล Document ใน MongoDB จะถูกเก็บไว้ใน “Collections” (เปรียบเทียบได้กับตาราง(Table) ใน Relational Database ทั่วไป) แต่แตกต่างกันที่ Collections ไม่จำเป็นที่จะต้องมี Schema เหมือนกันก็สามารถบันทึกข้อมูลได้

รูป 2.14 โลโก้ฐานข้อมูล MongoDB

(ที่มา: https://www.mongodb.com/brand-resources)

2.1.3.8 Node.js

คือ Cross Platform Runtime Environment ถูกพัฒนาโดย Ryan Dahl ในปี 2009 เป็นโอเพนซอร์ส สำหรับฝั่งเซิร์ฟเวอร์ ซึ่งเขียนด้วยภาษา JavaScript สร้างบน Chrome’s V8 JavaScript engine โดยที่ Node.js ใช้ event-driven และnon-blocking I/O model อีกทั้งยังมีไลบรารีที่เป็นส่วนเสริมจำนวนมากแยกเป็น JavaScript โมดูล(Module) ต่าง ๆ สำหรับทำงานเฉพาะในแต่ละเรื่อง โดยสามารถติดตั้งเพิ่มเติมได้อย่างง่ายผ่าน npm (Node Package Manager) เป็นตัวจัดการ Package ของ JavaScript และสามารถหา Package ที่เหมาะสมกับงานแต่ละอย่างได้

รูป 2.15 โลโก้ Node.js

(ที่มา: https://nodejs.org/en/about/resources/)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.3.9 React

เป็น JavaScript Library ที่ถูกสร้างโดย Facebook โดยเป็น User Interface (UI) ที่สร้างมาจากพื้นฐานแนวความคิดแบบ MVC(Model View Controller) ซึ่งก็คือมีหน้าที่ในการจัดการกับ Model หรือ View แต่ส่วนใหญ่จะเป็น View

รูป 2.16 โลโก้ React

(ที่มา: https://reactjs.org/)

2.1.3.10 Redux

เป็นแนวคิดหนึ่งสำหรับจัดการให้กับจุดอ่อนของ React ในส่วนของการไหลของ Data ในรูปแบบ Unique Directional Data Flow Data ไหลทางเดียว ไหลจากแม่ไปหาลูก โดยจะเข้าไปจัดการกับ state data ให้กับแอปพลิเคชัน ซึ่งจะเน้นในเรื่องของการจัดการ State หรือสถานะต่าง ๆ ของแอปพลิเคชัน เนื่องจากในระบบงานนั้น จะมี State จำนวนมากยิ่งระบบงานมีขนาดใหญ่ขึ้น ยิ่งมากขึ้นไปตามนั้น รวมทั้ง State ส่วนใหญ่มักจะถูกเปลี่ยนแปลงได้ตลอดเวลา (Mutable) จะส่งผลให้มีความซับซ้อนมากยิ่งขึ้นอย่างมาก ทำให้จัดการได้ยากมาก

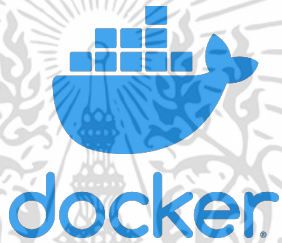
รูป 2.17 โลโก้ Redux

(ที่มา: https://github.com/reduxjs/redux/tree/master/logo)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.3.11 Docker

คือ engine ตัวหนึ่งที่มีการทำงานในลักษณะจำลองสภาพแวดล้อมขึ้นมาบนเครื่อง เซิร์ฟเวอร์ เพื่อใช้ในการ run service ที่ต้องการ ซึ่งมีการทำงานคล้ายคลึงกับ Virtual Machine เช่น VMWare, VirtualBox, XEN, KVM แต่ข้อแตกต่างที่ชัดเจนคือ Virtual Machine ที่รู้จักกันก่อนหน้านี้นั้น เป็นการจำลองทั้งระบบปฏิบัติการเพื่อใช้งานและหากต้องการใช้งาน service ใด ๆ จึงทำการติดตั้งเพิ่มเติมบนระบบปฏิบัติการนั้น แต่สำหรับ Docker แล้วจะใช้คอนเทนเนอร์(container) ในการจำลองสภาพแวดล้อมขึ้นมา เพื่อใช้งานสำหรับ 1 service ที่ต้องการใช้งานเท่านั้น โดยไม่ต้องมีส่วนของระบบปฏิบัติการเข้าไปเกี่ยวข้องเหมือน Virtual Machines อื่น ๆ จึงสามารถใช้งานได้บนหลายแพลตฟอร์ม ทั้งบน Linux, Mac และ Windows



รูป 2.18 โลโก้ Docker

(ที่มา: https://www.docker.com/company/newsroom/media-resources)

2.1.3.12 Jenkins

เป็นเครื่องมือ(Tools) ที่โอเพนซอร์ส ที่ถูกเขียนด้วยภาษา Java โดยถูกสร้างขึ้นมาเพื่อตอบสนอง 2 แนวคิดหลัก ๆ คือ Continuous integration และ Continuous Delivery โดยเมื่อมีการเปลี่ยนแปลงของโค้ดใน Version control software Jenkins จะนำโค้ดมา Compile ลองทดสอบซอฟต์แวร์ และทำการ Deploy ซอฟต์แวร์

รูป 2.19 โลโก้ Jenkins

(ที่มา: https://jenkins.io/)

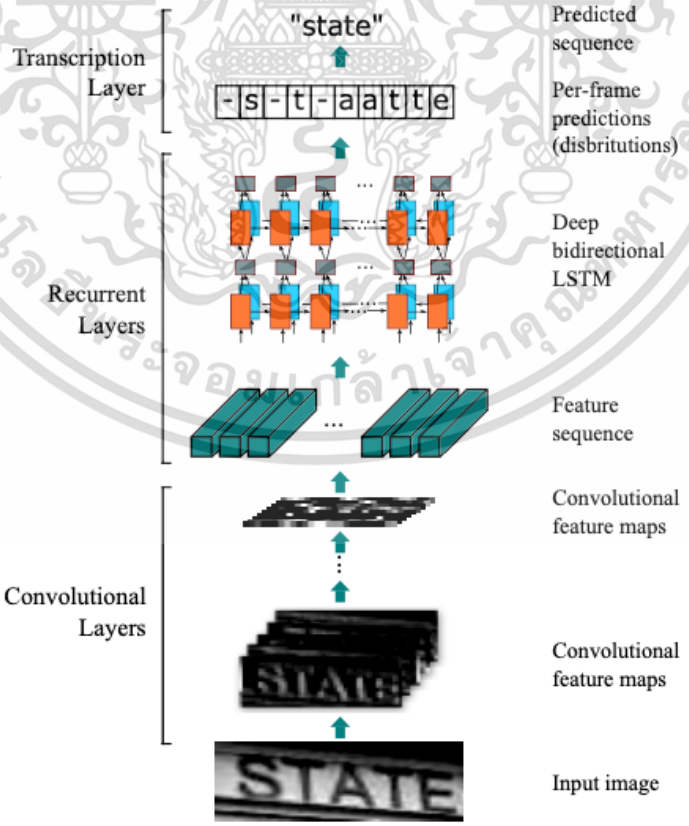
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2 งานวิจัยที่เกี่ยวข้อง

2.2.1 An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition [3]

งานวิจัยชิ้นนี้มีการนำหลักการวิเคราะห์และจำแนกตัวอักษรจากรูปมาใช้ โดยในชีวิตจริงการที่จะจำแนกตัวอักษรจากรูปนั้นจะอยู่ในรูปแบบลำดับ(sequence) ไม่ได้แยกออกจากกันเป็นตัว ๆ ไม่เหมือนการทำการรู้จำวัตถุ(Object recognition) ปกติ โมเดล Deep convolution neural network(DCNN) ที่เป็นที่นิยมนั้นไม่สามารถใช้กับข้อมูลที่เป็นลำดับได้ เนื่องจากจำเป็นต้องมีอินพุต(input) แบบกำหนดขนาด(Dimensions) ที่คงที่ โดยมีความพยายามที่จะแก้ไขปัญหานี้ด้วยการทำการตรวจจับ(detect) ตัวอักษรแต่ละตัวแล้วใช้ DCNN ในการทำนาย(predict) แต่วิธีนี้ต้องมีการตรวจจับที่ดีมาก ทำให้วิธีนี้ใช้จริงได้ไม่ดีเท่าที่ควร

Recurrent neural networks (RNN) เป็นโครงข่ายประสาทอีกประเภทหนึ่ง ที่ออกแบบมาเพื่อข้อมูลที่เป็นลำดับ จึงทำให้เลือกใช้ Convolutional Recurrent Neural Network (CRNN) ซึ่งเป็นการนำข้อดีจากโครงข่ายประสาททั้งสองแบบมารวมกัน



รูป 2.20 สถาปัตยกรรมเครือข่าย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

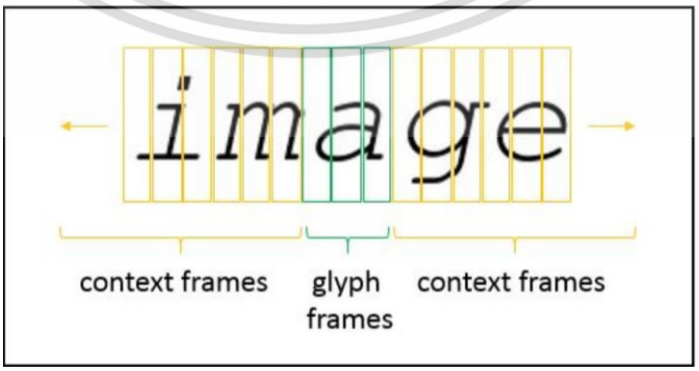
จากรูป 2.20 สถาปัตยกรรมเครือข่าย จะประกอบด้วย 3 ส่วน ได้แก่ 1) Convolutional Layers ทำหน้าที่แยกลำดับคุณลักษณะจากภาพอินพุต 2) Recurrent Layers ทำหน้าที่ทำนายการกระจายป้ายสำหรับแต่ละเฟรม และ 3) Transcription Layer ทำหน้าแปลการทำนายต่อเฟรมเป็นลำดับป้ายสุดท้าย

2.2.2 High-Performance OCR for Printed English and Fraktur using LSTM Networks [2]

งานวิจัยชิ้นนี้ได้ทำการทดลองโดยนำ Long-Short-Term Memory (LSTM) ซึ่งเป็นโครงข่ายประสาทแบบหนึ่งที่ถูกออกแบบมาสำหรับการประมวลผลลำดับ (sequence) โดยจะให้ผลที่ยอดเยี่ยมกับงานการรู้จำลายมือเขียน (handwriting recognition) มาประยุกต์ใช้ ซึ่งก่อนที่จะนำไฟล์ภาพเข้า Long-Short-Term Memory Networks จะต้องผ่านการทำ Text Line Normalization ก่อน โดยงานวิจัยนี้ได้ทดลองทั้ง 1D และ 2D LSTM จะพบว่าการใช้ 1D LSTM ได้ผลที่ดีกว่าสำหรับการใช้งานกับภาพที่ถูกพิมพ์ออกมา

2.2.3 Implicit Language Model in LSTM for OCR [1]

งานวิจัยชิ้นนี้ใช้ Convolutional Neural Network(CNN) และ Bidirectional Long-Short-Term Memory(LSTM) กับ Connectionist Temporal Classification(CTC) Loss เพื่อทำโมเดลโอซีอาร์ สำหรับ Generated font image และพบว่า LSTM พยายามจะทำความเข้าใจภาพในเฟรมก่อนและหลัง มาประกอบการตัดสินใจตัวอักษรในเฟรมนั้นๆด้วย โดยทำการทดลองกับภาพที่เป็นประโยคทั่วไปเทียบกับการสุ่มตัวอักษรที่ไม่ได้มีใจความ พบว่าโมเดลจะทำงานได้ดีกับคำที่เป็นประโยคทั่วไปได้ดีกว่า ทำให้เห็นว่า LSTM พยายามทำความเข้าใจความหมายในตัวประโยคโดยรวมมาประกอบการตัดสินใจในการทำนายตัวอักษรด้วย



รูป 2.21 การแสดงภาพของบริบทและเฟรมสัญลักษณ์สำหรับการระบุสัญลักษณ์ที่เกี่ยวกับตัวอักษรโดยรอบ แต่ละเฟรมเป็นอินพุตโอซีอาร์สำหรับหนึ่งขั้นตอน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3
การออกแบบและพัฒนา

การออกแบบและการพัฒนาแบ่งออกเป็น 4 ส่วน ได้แก่ ภาพรวมของระบบ(Conceptual Design), แผนภาพการทำงานของผู้ใช้(Use Case Diagram), การทำงานของระบบโอซีอาร์ และแสดงความสัมพันธ์ของข้อมูล (Logical Data Model)

3.1 ภาพรวมของระบบ

ภาพรวมของระบบได้รับการออกแบบดังรูปต่อไปนี้

รูป 3.1 ภาพรวมของระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ระบบการแปลงไฟล์ภาพเอกสารให้เป็นข้อความอิเล็กทรอนิกส์ ประกอบไปด้วย 3 ส่วน ได้แก่

3.1.1 ส่วนเว็บแอปพลิเคชัน

เป็นส่วนของเว็บที่ให้บริการกับผู้ใช้งานสามารถแบ่งงานออกเป็นแต่ละโปรเจค โดยแต่ละโปรเจคจะมีส่วนให้จัดการแม่แบบได้หลายแม่แบบ, จัดการอัปโหลดไฟล์ภาพเอกสารแบบฟอร์มในรูปแบบตัวพิมพ์ที่ต้องการทำการแปลงผลผ่านโอซีอาร์ และดูข้อมูลที่แสดงผลที่ได้จากการแปลงผลผ่านโอซีอาร์

ในส่วนจัดการแม่แบบ โดยแม่แบบนี้จะสร้างขึ้น เพื่อใช้ในการระบุตำแหน่งบนไฟล์ภาพเอกสารในตำแหน่งที่ต้องการทำการแปลงผลผ่านระบบโอซีอาร์ และจะทำการติดป้ายให้แต่ละตำแหน่ง เพื่อใช้ในการเก็บลงฐานข้อมูลว่าผลที่ได้จากการแปลงผ่านโอซีอาร์ตรงนี้คือหัวข้อ หรือป้ายอะไร ซึ่งในส่วนนี้จะสามารถสร้าง, แก้ไข, ดู และติดป้ายแม่แบบในโปรเจคได้

ส่วนในการจัดการอัปโหลดไฟล์ภาพเอกสารแบบฟอร์ม จะสามารถอัปโหลดไฟล์ภาพเอกสารแบบฟอร์มในรูปแบบตัวพิมพ์ที่ได้ทำการสร้างแม่แบบไว้เรียบร้อยในโปรเจคนี้แล้วต้องการนำมาทำการแปลงผลผ่านโอซีอาร์ได้

และในส่วนดูข้อมูลที่แสดงผลที่ได้จากการแปลงผลผ่านโอซีอาร์ของในโปรเจคนี้ จะแสดงออกมาในรูปแบบตารางที่สามารถค้นหา และนำออกมาเป็นไฟล์ CSV หรือ EXCEL ได้ โดยข้อมูลต่าง ๆ ที่ได้จากการแปลงผลผ่านโอซีอาร์จะถูกส่งไปยังเซิร์ฟเวอร์ แล้วนำข้อมูลไปเก็บในฐานข้อมูลก่อนจะดึงข้อมูลทั้งหมดออกมาแสดง

3.1.2 ส่วนเซิร์ฟเวอร์และฐานข้อมูล

เป็นส่วนที่ให้บริการเว็บแอปพลิเคชัน และรวบรวมข้อมูลทั้งหมดของระบบ

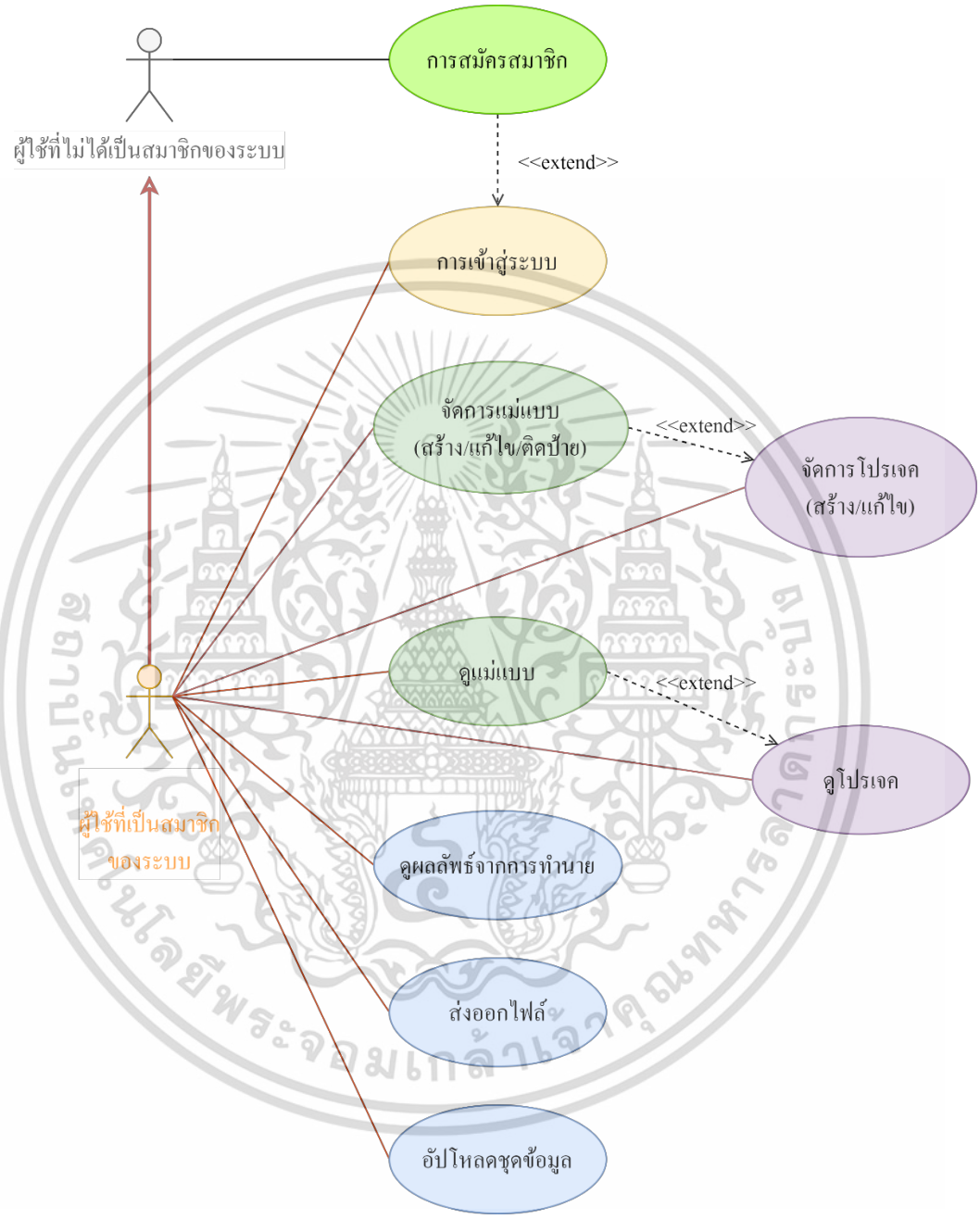
3.1.3 ส่วนบริการโอซีอาร์

เป็นส่วนของบริการในแปลงไฟล์ภาพเอกสารแบบฟอร์มในรูปแบบตัวพิมพ์สำหรับภาษาไทยหรือภาษาอังกฤษให้เป็นข้อความอิเล็กทรอนิกส์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2 แผนภาพการทำงานของผู้ใช้

แผนภาพการทำงานของผู้ใช้ได้รับการออกแบบดังรูปต่อไปนี้



รูป 3.2 แผนภาพการทำงานของผู้ใช้

จากแผนภาพการทำงานของผู้ใช้แบ่งได้ 2 ประเภทและสามารถอธิบายได้ดังนี้

3.2.1 ผู้ใช้ที่ไม่ได้เป็นสมาชิกของระบบ

1) ผู้ใช้ที่ไม่ได้เป็นสมาชิกของระบบสามารถสมัครสมาชิกเพื่อเข้าใช้งานระบบได้

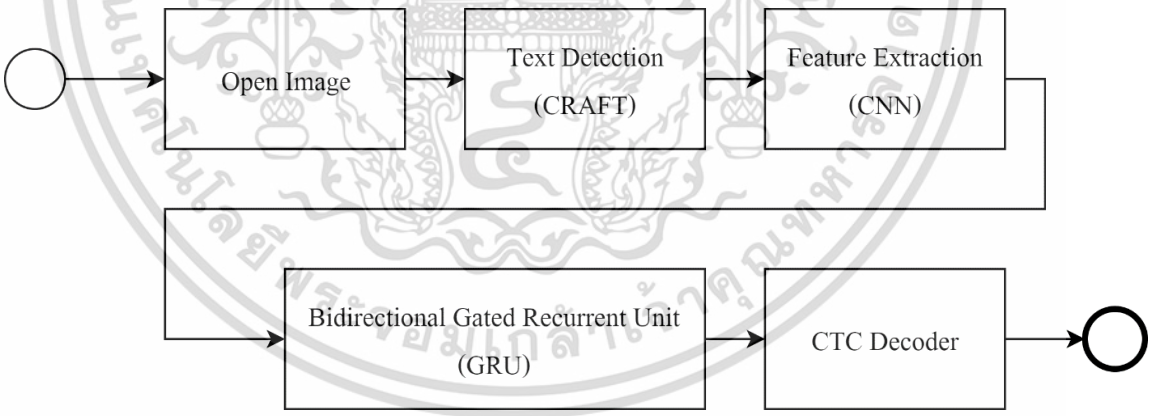
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.2 ผู้ใช้ที่เป็นสมาชิกของระบบ

1) ผู้ใช้ที่เป็นสมาชิกของระบบสามารถเข้าสู่ระบบได้
2) ผู้ใช้ที่เป็นสมาชิกของระบบสามารถสร้าง และ แก้ไขโปรเจคได้
3) ผู้ใช้ที่เป็นสมาชิกของระบบสามารถดูข้อมูลของโปรเจคได้
4) ผู้ใช้ที่เป็นสมาชิกของระบบสามารถสร้าง, แก้ไข และติดป้ายในแม่แบบของโปรเจคได้
5) ผู้ใช้ที่เป็นสมาชิกของระบบสามารถดูข้อมูลของแม่แบบในโปรเจคได้
6) ผู้ใช้ที่เป็นสมาชิกของระบบสามารถดู และค้นหาผลลัพธ์ของข้อมูลที่ผ่านการแปลงผลจากโอซีอาร์ในตารางของแม่แบบในโปรเจคได้
7) ผู้ใช้ที่เป็นสมาชิกของระบบสามารถส่งออกไฟล์ผลลัพธ์ของข้อมูลที่ผ่านการแปลงผลจากโอซีอาร์ได้
8) ผู้ใช้ที่เป็นสมาชิกของระบบสามารถอัปโหลดชุดข้อมูลที่ต้องการจะนำมาแปลงผลผ่านโอซีอาร์ได้

3.3 การทำงานของระบบโอซีอาร์

การทำงานของระบบโอซีอาร์ได้รับการออกแบบดังรูปต่อไปนี้



รูป 3.3 การทำงานของระบบโอซีอาร์

ขั้นตอนการทำงานมีดังนี้

1) เปิดไฟล์ภาพเอกสารในรูปแบบตัวพิมพ์
2) หาตำแหน่งของข้อความตัวพิมพ์ด้วยวิธี Character Region Awareness For Text Detection (CRAFT) และตัดเฉพาะส่วนของข้อความ
3) สกัดคุณลักษณะด้วย Convolutional Neural Network (CNN)

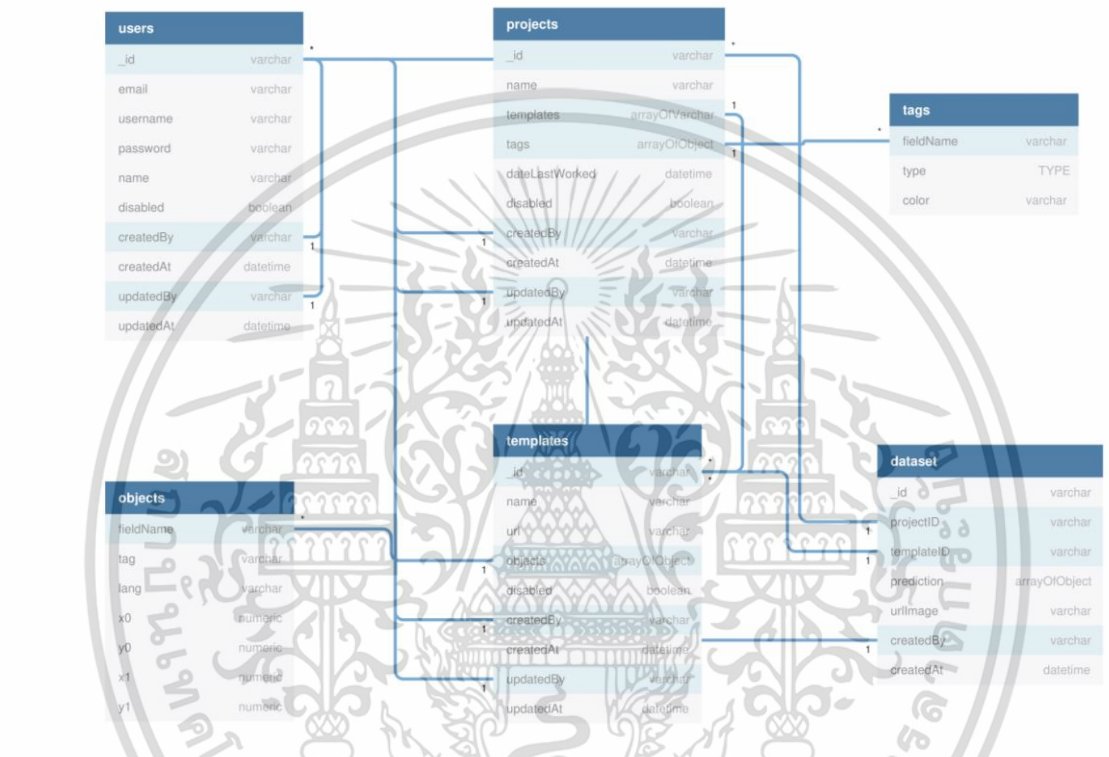
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4) นำผังคุณลักษณะ(feature map) เข้า Bidirectional Gated Recurrent Unit (GRU)

5) Connectionist Temporal Classification(CTC) Loss

6) ได้ประโยคที่เสร็จสมบูรณ์

3.4 แผนภาพแสดงความสัมพันธ์ของข้อมูล



รูป 3.4 แผนภาพแสดงความสัมพันธ์ของข้อมูล

การออกแบบฐานข้อมูลประกอบด้วยตารางจำนวน 6 ตาราง แต่ละตารางมีรายละเอียดดังนี้

3.4.1 ตารางผู้ใช้งาน (user)

เป็นตารางที่แสดงข้อมูลของผู้ใช้งานทั้งหมด เก็บข้อมูลของผู้ใช้งาน ได้แก่ อีเมล, ชื่อผู้ใช้งาน, รหัสผ่าน, ชื่อ, การเปิดใช้งาน, วันและเวลาที่สร้าง, แก้ไขโดย และวันและเวลาที่แก้ไข

3.4.2 ตารางโปรเจค (project)

เป็นตารางที่แสดงข้อมูลของโปรเจคทั้งหมด เก็บข้อมูลของโปรเจค ได้แก่ ชื่อโปรเจค, รายการไอดีแม่แบบ, รายการไอดีป้าย, วันที่ทำงานล่าสุด, การเปิดใช้งาน, สร้างโดย, วันและเวลาที่สร้าง, แก้ไขโดย และวันและเวลาที่แก้ไข

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.4.3 ตารางป้าย (tag)

เป็นตารางที่แสดงข้อมูลของป้ายทั้งหมด เก็บข้อมูลของป้าย ได้แก่ ชื่อป้าย, ประเภท และสี

3.4.4 ตารางแม่แบบ (template)

เป็นตารางที่แสดงข้อมูลของแม่แบบทั้งหมด เก็บข้อมูลของแม่แบบ ได้แก่ ชื่อแม่แบบ, ยูอาร์แอล(URL) ของไฟล์ภาพเอกสารแบบฟอร์มแม่แบบ, รายการไอดีวัตถุ, การเปิดใช้งาน, สร้างโดย, วันและเวลาที่สร้าง, แก้ไขโดย และวันและเวลาที่แก้ไข

3.4.5 ตารางวัตถุ (object)

เป็นตารางที่แสดงข้อมูลของวัตถุทั้งหมด เก็บข้อมูลของวัตถุ ได้แก่ ชื่อวัตถุ, ป้าย, ภาษา, ตำแหน่ง x0, ตำแหน่ง y0, ตำแหน่ง x1 และ ตำแหน่ง y1

3.4.6 ตารางชุดข้อมูล (dataset)

เป็นตารางที่แสดงข้อมูลของชุดข้อมูลทั้งหมด เก็บข้อมูลแต่ละชุดข้อมูลที่ผู้ใช้ต้องการแปลงผ่านโอซีอาร์ ได้แก่ ไอดีโปรเจค, ไอดีแม่แบบ, ผลการแปลงผ่านโอซีอาร์, ยูอาร์แอล(URL) ของไฟล์ภาพเอกสารแบบฟอร์ม, วันที่ทำงานล่าสุด, การเปิดใช้งาน, สร้างโดย, วันและเวลาที่สร้าง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4
การทดลองและการใช้งาน

4.1 การทดลองส่วนโมเดลทำนายข้อความอิเล็กทรอนิกส์จากไฟล์ภาพเอกสาร

เนื่องจากได้ทำการทดลองตามงานวิจัย CRNN แล้วพบว่าใช้เวลาในการฝึกอบรม และทรัพยากรปริมาณมาก คณะผู้จัดทำปริญญานิพนธ์จึงได้ทำการทดลองแค่สถาปัตยกรรม Resnet-50 เท่านั้น ส่วนค่าที่ใช้ในการวัดประสิทธิภาพของโมเดลนั้น จะใช้ค่า CER(Character Error Rate) โดยมีสูตรในการคำนวณดังนี้

$$CER = \frac{(i+s+d)}{min(n_ground_truth, n_predicted)} \quad (4.1)$$

เมื่อ *i* คือ จำนวนที่น้อยที่สุดของ character insertions เพื่อแปลงคำที่ทำนายมาเป็น ground truth

s คือ จำนวนที่น้อยที่สุดของ character substitutions เพื่อแปลงคำที่ทำนายมาเป็น ground truth

d คือ จำนวนที่น้อยที่สุดของ character deletions เพื่อแปลงคำที่ทำนายมาเป็น ground truth

n คือ จำนวน character ทั้งหมด

4.1.1 รายละเอียดในการทดลอง

1) ภาพความละเอียดขนาด 64x1024 พิกเซล โหมดRGB
2) จำนวนฟอนต์ 97 ฟอนต์
3) กระบวนการทำ Augmentation
   - Translate
   - Rotate
   - Elastic transform
   - Blur
4) คลังข้อความ(Text corpus)

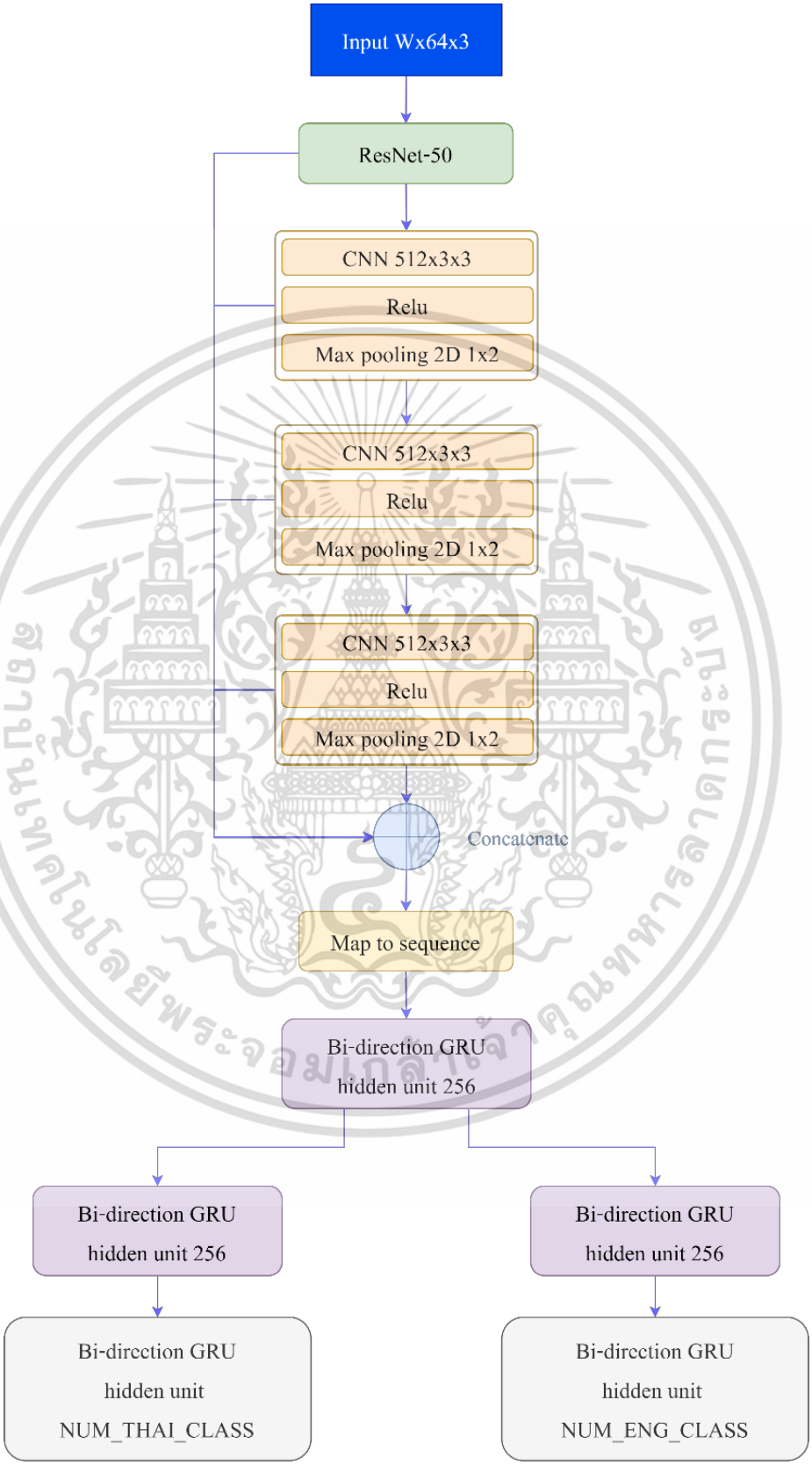
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- ภาษาไทย
  - prachathai-67k
  - wongnai-corpus
- ภาษาอังกฤษ
  - Book Review

5) ชุดข้อมูลสำหรับทดสอบ(Test set)
   - ภาษาไทย
     - สุ่มตัดภาพเอกสารตัวพิมพ์ที่ผ่านการแสกนจากอินเทอร์เน็ต 150 รูป
   - ภาษาอังกฤษ
     - ICDAR2003 1157 รูป
     - ICDAR2019 848 รูป
6) รายละเอียดฮาร์ดแวร์คอมพิวเตอร์ที่ใช้ในการฝึกอบรมแบบจำลอง
   - CPU : AMD Threadripper 2990WX
   - GPU : NVIDIA RTX 2080Ti
   - Memory : 128 GB
7) การตั้งค่าตัวแปรที่ใช้ในการฝึกอบรมแบบจำลอง
   - Learning rate : 0.001
   - Optimize : Adam
   - Loss function : CTC Loss
   - Batch size : 32

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.2 ขั้นตอนการทดลอง



รูป 4.1 แผนภาพขั้นตอนการทดลอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

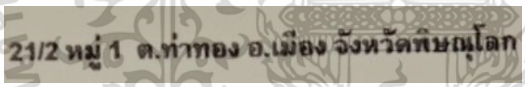
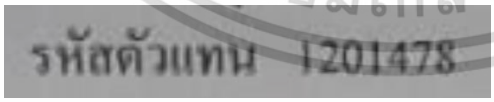
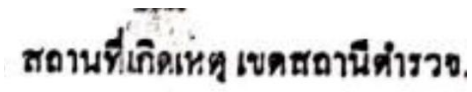
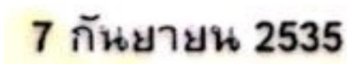
4.1.3 ผลการทดลอง

4.1.3.1 การออกแบบการทดลองของการทำนายข้อความอิเล็กทรอนิกส์จากไฟล์ภาพเอกสารภาษาไทยและตัวเลข

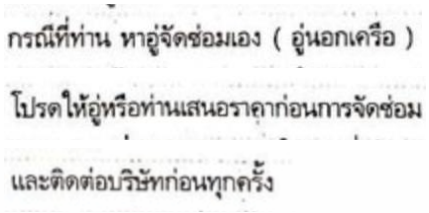
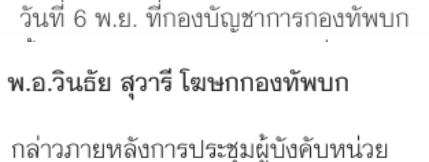
จากการนำชุดข้อมูลสำหรับทดสอบไปทดสอบกับโมเดล ได้ผลลัพธ์ดังต่อไปนี้

- ชุดข้อมูลสำหรับทดสอบจากการสุ่มตัดภาพเอกสารตัวพิมพ์ที่ผ่านการแสกนจากอินเทอร์เน็ต 150 รูป
  - CER Our : 7.19
  - CER Tesseract 4.1 : 25.27

ตาราง 4.1 ตัวอย่างผลการทดลองจากชุดข้อมูลสำหรับทดสอบจากการสุ่มตัดภาพเอกสารตัวพิมพ์ที่ผ่านการแสกนจากอินเทอร์เน็ต 150 รูป

ลำดับที่	ภาพเอกสารตัวพิมพ์	ผลการทำนาย	
		Our	Tesseract
1		400.000.00 บาท	
2		212หมู่1 ต.ทำทอง อ.เมือง จังหวัดพิษณุโลกุ	21)2 หมู่ 1 ต.ท่าทอง อ.เมือง จังหวัดพิษณุโลก
3		99/105 หมู่ 6 ต.ทับมา อ.เมือง จังหวัดระยอง	99!105 หมู่ 6 ต.ทับมา อ.เมือง จังหวัดระยอง
4		รหัดตัวแทน 1201478	รหัสตัวแทน 1!1201478
5		ฉบับที่3ชบ. 00204/62	ฉบับที่3ชบ. 00204/62
6		สถานที่เกิดหนุ เขตสถานีตำรวจ่	สถานที่เกิดเหตุ เขตสถานีตำรวย.
7		7กันยายน 2535	7 กันยายน 2535

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลำดับที่	ภาพเอกสารตัวพิมพ์	ผลการทำนาย	
		Our	Tesseract
8		กรณีที่ทาน หา ู่จัดซ่อมเอง / อูนอกเครือ 1 โปรดให้อู่หรือท่านเสนอราคาก่อนการจัดซ่อมและติดต่อบริษัทก่อนทุกครั้ง	กรณีที่ท่าน หาอู่จัดซ่อมเอง (อุ่นอกเครือ) โปรดให้อู่หรือท่านเสนอราคาก่อนการจัดซ่อมและติดต่อบริษัทก่อนทุกครั้ง
9		การทดลองของไมโครซอฟท์ญี่ปุ่นในเดือนสิงหาคมที่ผ่านมา โดยให้พนักงานทำงาน 4 วันต่อสัปดาห์ พบประสิทธิภาพการทำงานดีขึ้นถึง 4096	ก็ถ่จไจใจ1เล็ดจเลีญี่ย: เงจโรไว วัจเต่: ค็ริฑธี. ค๊กจาจเสีซี้จเถีน การทดลองของไมโครซอฟท์ญี่ปุ่นในเดือนสิงหาคมที่ผ่านมา โดยให้พนักงานทำงาน 4 วันต่อสัปดาห์ พบประสิทธิภาพการทำงานดีขึ้นถึง 40%
10		วันที่ 6 พ.ย. ที่กองบัญชาการกองทัพบก พ.อ. วินธัย สุวารี โฆษกกองทัพบก กล่าวภายหลังการประชุมผู้บังคับหน่วย	วันที่ 6 พ.ย. ที่กองบัญชาการกองทัพบก พ.อ. วินธัย สุวารี โฆษกกองทัพบก กล่าวภายหลังการประชุมผู้บังคับหน่วย

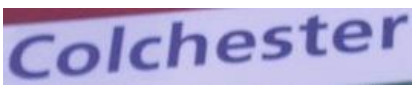
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.3.2 การออกแบบการทดลองของการทำนายข้อความอิเล็กทรอนิกส์จากไฟล์ภาพเอกสารภาษาอังกฤษและตัวเลข

จากการนำชุดข้อมูลสำหรับทดสอบไปทดสอบกับโมเดล ได้ผลลัพธ์ดังต่อไปนี้

- ชุดข้อมูลสำหรับทดสอบจากการ ICDAR2003 1157 รูป
  - CER Our : 20.51
  - CER Tesseract 4.1 : 74.88
- ชุดข้อมูลสำหรับทดสอบจากการ ICDAR2019 848 รูป
  - CER Our : 12.83
  - CER Tesseract 4.1 : 67.63

ตาราง 4.2 ตัวอย่างผลการทดลองจากชุดข้อมูลสำหรับทดสอบจากการ ICDAR2003 1157 รูป และ ICDAR2019 848 รูป

ลำดับที่	ภาพเอกสารตัวพิมพ์	ผลการทำนาย	
		Our	**Tesseract**
1		the	the
2		WILLCOCKS	
3		Sofetvare	
4		MIDDLEBOROUGH	MIDDLEBOROUGH
5		INTERNATIONAL	
6		Associates	NOS O CTA TiS
7		A134	
8		Colchester	Colchester
9		WILEY	
10		order	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.4 สรุปผลการทดลอง

เนื่องจากโมเดลใช้เวลาในการฝึกอบรมนาน คณะผู้จัดทำปริญญานิพนธ์จึงสามารถทดสอบได้เพียงสถาปัตยกรรมเดียว และไม่มีชุดข้อมูลเอกสารตัวพิมพ์ของภาษาไทยและภาษาอังกฤษจริง จึงต้องอาศัยเพียงชุดข้อมูลจากฟอนต์ที่จำลองขึ้นมา ทำให้ CER(Character Error Rate) ของภาษาไทยมีค่าที่ 7.19% และของภาษาอังกฤษมีค่าน้อยสุดที่ 12.83% โดยสรุปโมเดลนี้ไม่สามารถใช้ได้จริง จึงควรหาชุดข้อมูลจากเอกสารจริงมาเพิ่ม เพื่อช่วยให้โมเดลทำงานได้ดียิ่งขึ้น

4.2 การใช้งานส่วนเว็บแอปพลิเคชัน

แอปพลิเคชันของระบบประกอบไปด้วย 6 ส่วนหลัก ได้แก่

4.2.1 ส่วนเข้าส่วนระบบเพื่อใช้งาน

ก)

ข)

รูป 4.2 หน้าแสดงผลของการยืนยันตัวตน

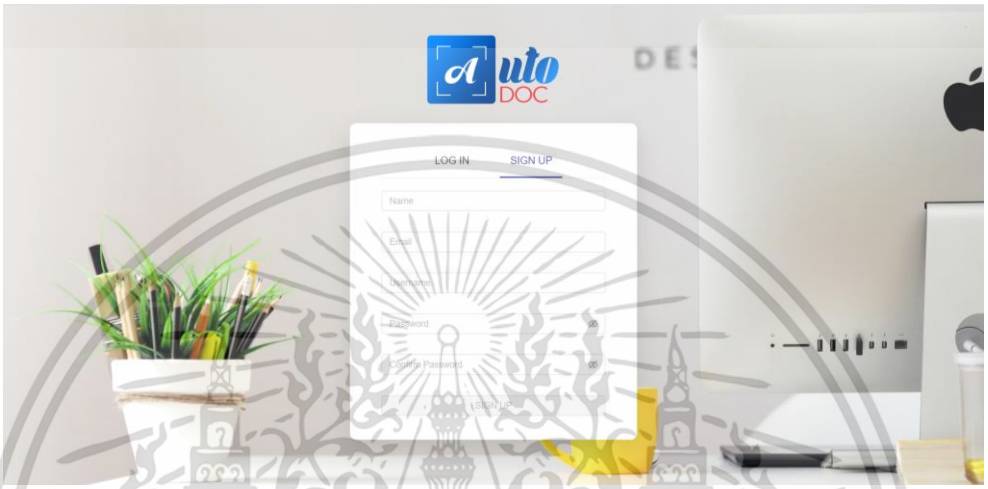
ก) เมื่อไม่ได้กรอกข้อมูล

ข) เมื่อกรอกข้อมูลแล้ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ส่วนของการยืนยันเข้าสู่ระบบ ผู้ใช้ที่เป็นสมาชิกของระบบจะต้องใส่ชื่อผู้ใช้และรหัสผ่านที่ลงทะเบียนสมัครไว้กับระบบในช่องดังรูป 4.2 ก) และเมื่อทำการกรอกข้อมูลครบแล้วก็จะสามารถกดปุ่ม “LOGIN” เพื่อเข้าสู่ระบบได้ดังรูป 4.2 ข)

4.2.2 ส่วนสมัครสมาชิกเพื่อเข้าใช้งาน



ก)

ข)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค)

รูป 4.3 หน้าแสดงผลของการสมัครสมาชิก

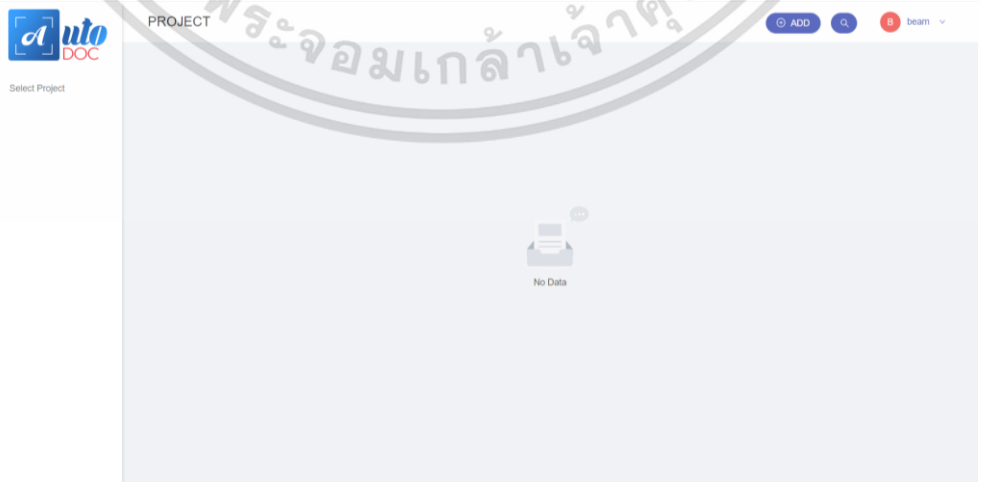
ก) เมื่อไม่ได้กรอกข้อมูล

ข) เมื่อกรอกข้อมูลไม่ถูกต้อง

ค) เมื่อกรอกข้อมูลถูกต้องแล้ว

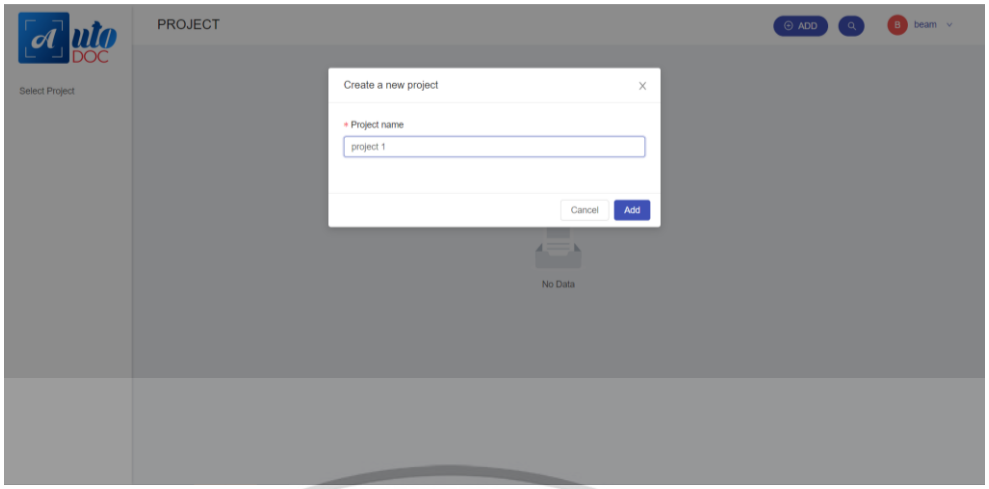
ส่วนของการสมัครสมาชิก ผู้ใช้ที่ไม่ได้เป็นสมาชิกของระบบจะต้องกรอกชื่อ, อีเมล, ชื่อผู้ใช้, รหัสผ่าน และยืนยันรหัสผ่านในช่องดังรูป 4.3 ก) ถ้าหากกรอกข้อมูลไม่ตรงกับเงื่อนไขที่กำหนดไว้ก็จะแจ้งเตือนดังรูป 4.3 ข) แล้วเมื่อการกรอกข้อมูลได้ถูกต้องตรงตามเงื่อนไขก็จะสามารถทำการกดปุ่ม “SIGN UP” เพื่อทำการสมัครสมาชิกได้ดังรูป 4.3 ค)

4.2.3 ส่วนโปรเจค



ก)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ข)

ค)

รูป 4.4 หน้าแสดงผลของโปรเจค

ก) เมื่อไม่มีโปรเจค

ข) เมื่อทำการเพิ่มโปรเจค

ค) เมื่อมีโปรเจค

ส่วนของโปรเจคจะแสดงโปรเจคที่ของผู้ใช้ที่เป็นสมาชิกของระบบสร้างไว้ดังรูป 4.4 ก) เมื่อไม่มีโปรเจค และดังรูป 4.4 ค) เมื่อมีแม่แบบ โดยแต่ละโปรเจคจะจำนวนแม่แบบในโปรเจค

ส่วนการเพิ่มโปรเจคสามารถได้โดยการกดปุ่ม “ADD” ทางบนขวา แล้วเมื่อกดแล้วก็จะเป็นดังรูป 4.4 ข) โดยจะให้กรอกชื่อโปรเจค แล้วทำการกดปุ่ม “Add” เพื่อเพิ่มโปรเจค หรือกดปุ่ม “Cancel” เพื่อยกเลิกการเพิ่มโปรเจค

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

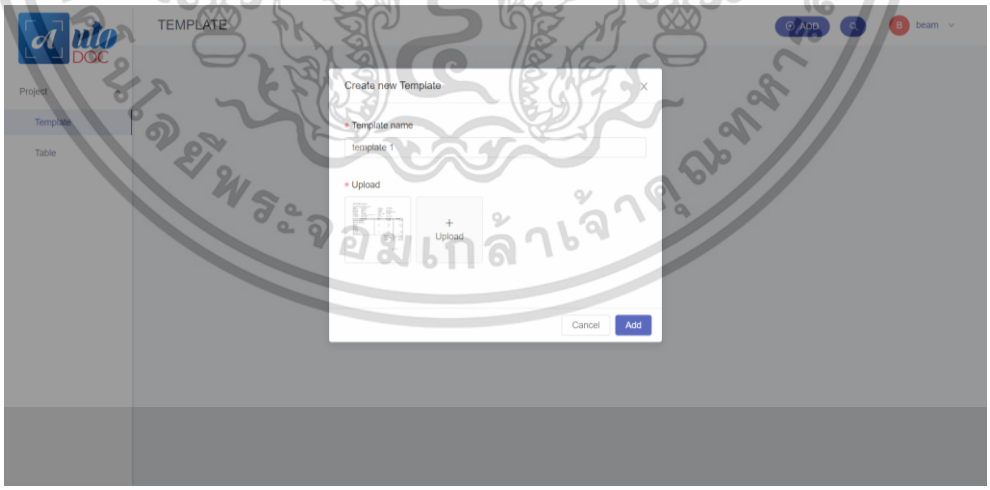
ส่วนการค้นหาชื่อของโปรเจคสามารถทำได้โดยกดที่ปุ่มรูปแว่นขยาย จะสามารถใส่ชื่อและกดค้นหา

ส่วนการแก้ไขโปรเจคสามารถทำได้โดยกดปุ่มจดสามจุดหลังชื่อโปรเจคในกรอบจะมีตัวเลือกให้จัดการ

ส่วนการเลือกที่จะเข้าไปดูโปรเจคไหนให้ทำการคลิกเลือกโปรเจคที่ต้องการได้เลย

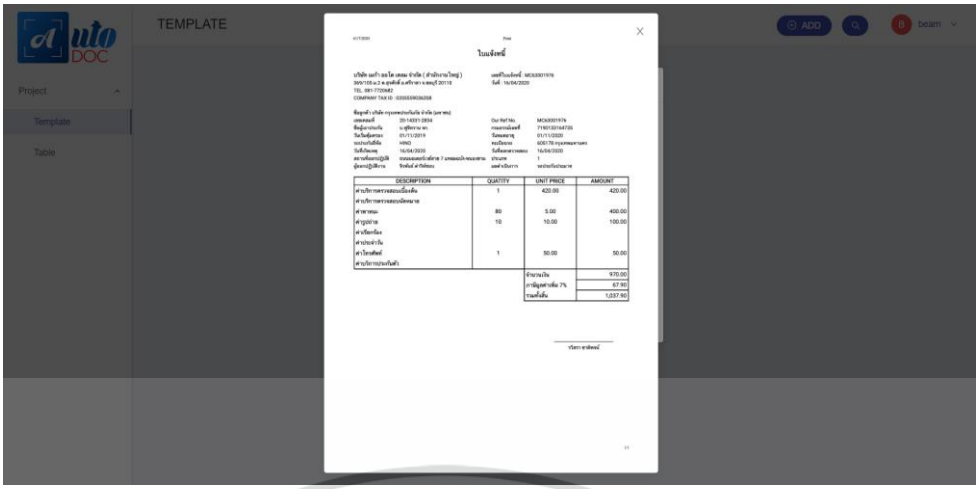
4.2.4 ส่วนแม่แบบของโปรเจค

ก)



ข)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ค)

ง)

รูป 4.5 หน้าแสดงผลของแม่แบบในโปรเจค

ก) เมื่อไม่มีแม่แบบ

ข) เมื่อทำการเพิ่มแม่แบบ

ค) เมื่อดูตัวอย่างไฟล์ภาพเอกสารที่อัปโหลด

ง) เมื่อมีแม่แบบ

ส่วนของแม่แบบจะแสดงแม่แบบในโปรเจคที่ผู้ใช้ที่เป็นสมาชิกของระบบสร้างไว้ดังรูป 4.5 ก) เมื่อไม่มีแม่แบบ และดังรูป 4.5 ง) เมื่อมีแม่แบบ โดยแต่ละแม่แบบจะแสดงเวลาที่เข้าดูล่าสุด และรูปของไฟล์ภาพเอกสารที่เป็นแม่แบบ

ส่วนการเพิ่มแม่แบบสามารถทำได้โดยการกดปุ่ม “ADD” ทางบนขวา แล้วเมื่อกดแล้วก็จะเป็นดังรูป 4.5 ข) โดยจะให้กรอกชื่อแม่แบบ และอัปโหลดไฟล์ภาพเอกสาร ซึ่งสามารถดูตัวอย่าง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

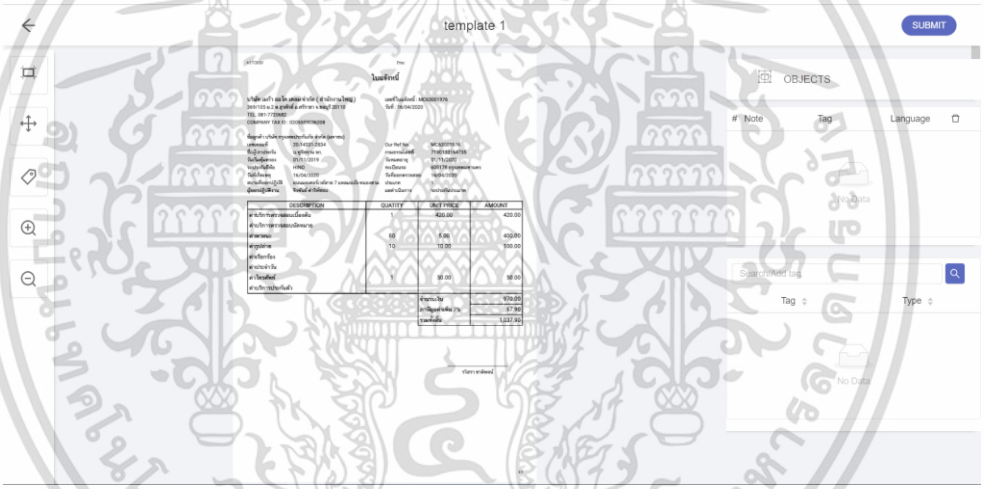
ไฟล์ภาพเอกสารได้จะเป็นดังรูป 4.5 ค) แล้วทำการกดปุ่ม “Add” เพื่อเพิ่มแม่แบบ หรือกดปุ่ม “Cancel” เพื่อยกเลิกการเพิ่มแม่แบบ

ส่วนการค้นหาชื่อของแม่แบบสามารถทำได้โดยกดที่ปุ่มรูปแว่นขยาย จะสามารถใส่ชื่อและกดค้นหา

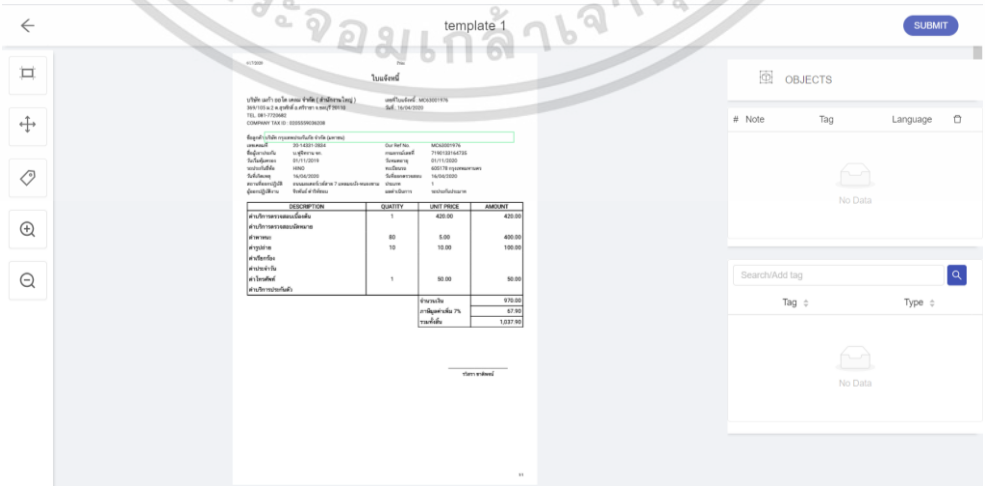
ส่วนการแก้ไขแม่แบบสามารถทำได้โดยกดปุ่มจุดสามจุดหลังชื่อแม่แบบในกรอบจะมีตัวเลือกให้จัดการ

ส่วนการแก้ไขการทำแม่แบบสามารถทำได้โดยทำการคลิกเลือกแม่แบบที่ต้องการ เพื่อเข้าไปแก้การทำแม่แบบที่ทำไว้ก่อนหน้าได้

4.2.5 ส่วนทำแบบแม่

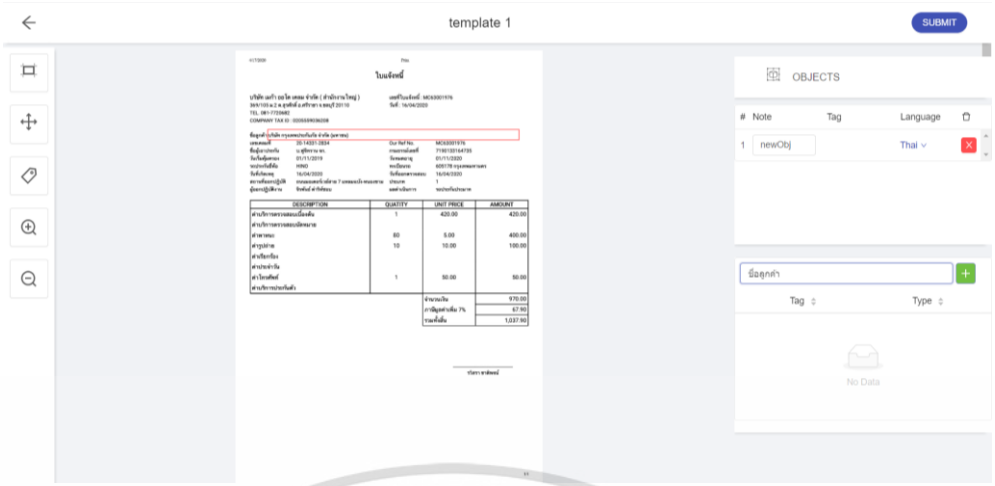


ก)

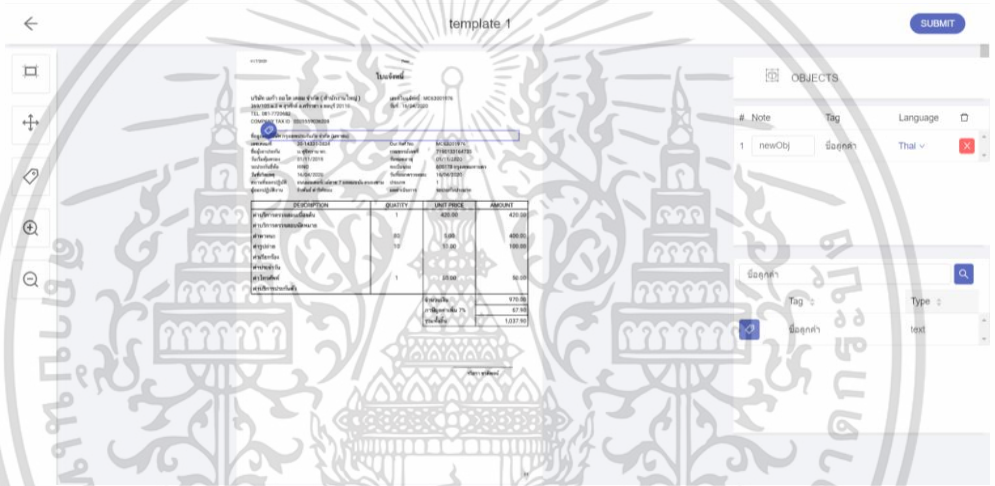


ข)

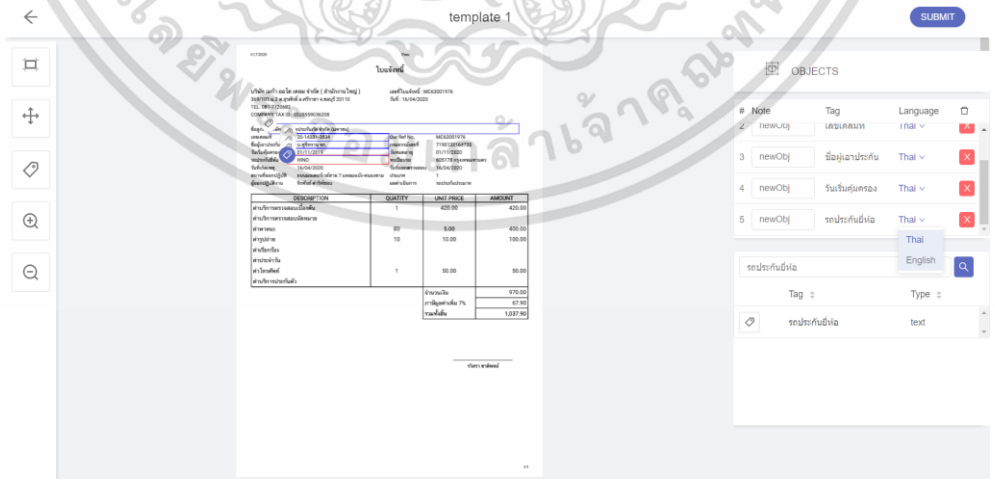
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ค)

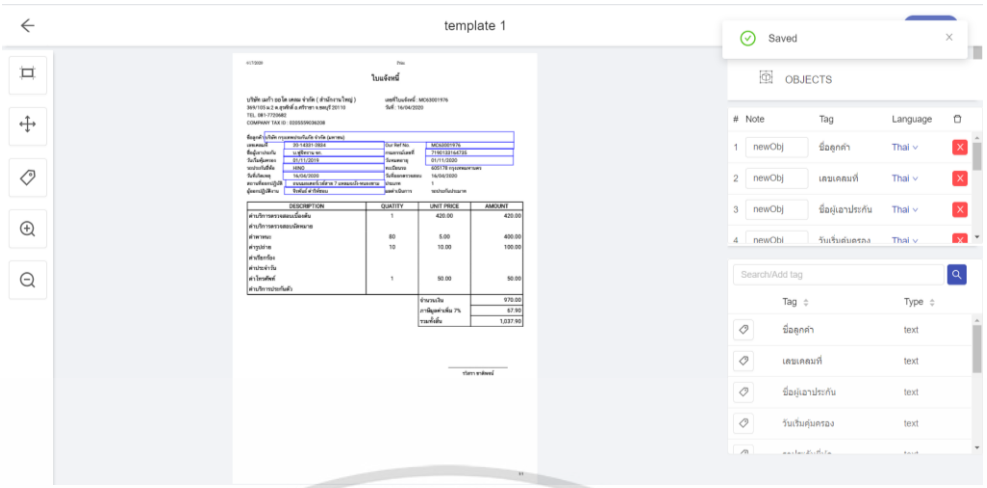


ง)



จ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ฉ)

รูป 4.6 หน้าแสดงผลของการสร้างแม่แบบในโปรเจค

ก) เมื่อเริ่มต้นการทำแม่แบบ

ข) เมื่อทำการตีกรอบ

ค) เมื่อทำการเพิ่มป้าย

ง) เมื่อทำการติดป้าย

จ) เมื่อทำการเลือกภาษาให้วัตถุที่ตีกรอบ

ฉ) เมื่อทำแม่แบบเสร็จ แล้วกด “SUBMIT”

ส่วนของการทำแม่แบบ จะแสดงไฟล์ภาพเอกสารแบบฟอร์มที่ต้องการมาทำแปลงผลผ่านโอซีอาร์ที่แสดงอยู่ตรงกลาง, เครื่องมือทางที่แสดงอยู่ทางด้านซ้าย โดยมีเครื่องมือ ได้แก่ ตีกรอบ, ย้ายกรอบ, ติดป้าย, ขยายรูป และลดรูป และส่วนแสดงวัตถุทั้งหมดในแม่แบบนี้ กับส่วนแสดงป้ายที่ทำการเพิ่มทั้งหมดในโปรเจคนี้ดังรูป 4.6 ก)

ส่วนการตีกรอบ จะทำโดยเลือกเครื่องมือตีกรอบมาตีส่วนที่ต้องการนำไปแปลงผลผ่านโอซีอาร์ในรูปของเอกสารฟอร์มดังรูป 4.6 ข)

ส่วนค้นหา/เพิ่มป้าย จะทำโดยการค้นหาชื่อป้ายที่ต้องการ ซึ่งจะกรอกชื่อป้ายลงในช่องค้นหา/เพิ่มป้าย และกดปุ่มรูปแว่นขยาย เพื่อค้นหาในส่วนรายการป้ายทั้งหมดในโปรเจค ถ้าหากไม่พบชื่อป้ายที่ค้นหาปุ่มรูปแว่นขยายจะเปลี่ยนเป็นปุ่มบวกสีเขียวแทน เมื่อทำการกดที่ปุ่มบวกสีเขียวก็จะเป็นการเพิ่มป้ายดังรูป 4.6 ค)

ส่วนติดป้าย จะทำโดยเลือกเครื่องมือติดป้าย ก็ทำการเลือกกดรูปป้ายบนวัตถุที่ตีกรอบไว้แล้วที่ต้องการ หลังจากนั้นก็จะทำการเลือกป้ายที่ต้องการจากส่วนรายการป้ายทั้งหมดในโปรเจค

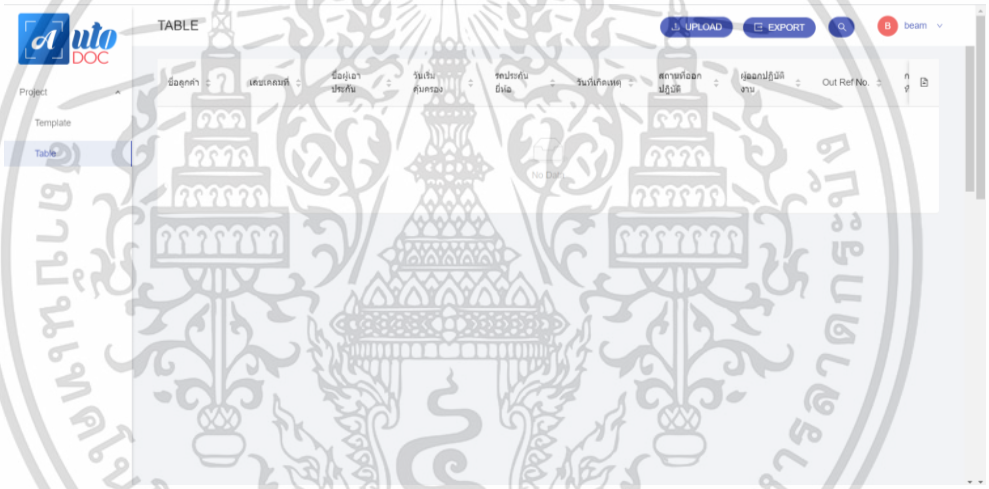
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยทำการเลือกกดรูปป้ายหน้าชื่อป้ายที่ต้องการ เพื่อเป็นการติดป้าย ซึ่งเมื่อทำการติดป้ายเสร็จก็จะไปแสดงบนรายการวัตถุในแม่แบบว่าวัตถุนี้ติดป้ายอะไรดังรูป 4.6 ง)

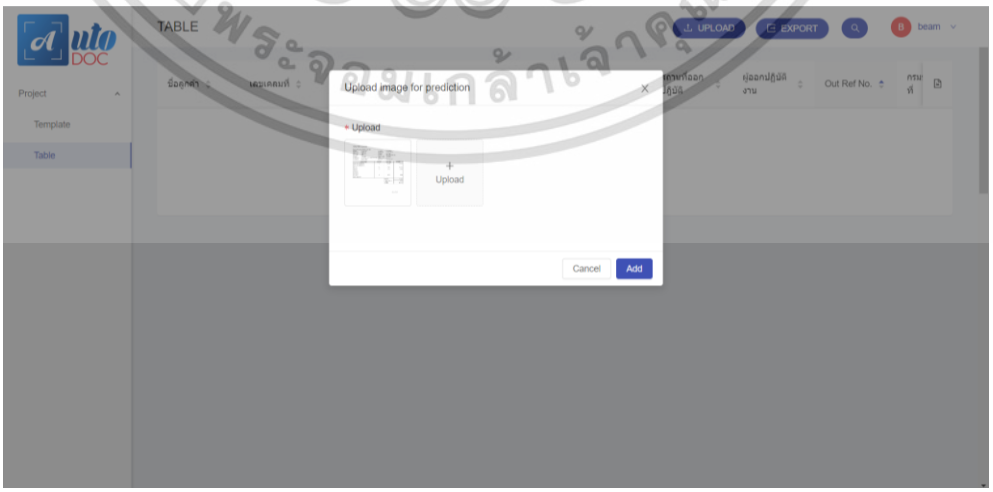
ส่วนแสดงวัตถุในแม่แบบ จะสามารถดูลำดับวัตถุ, ดูและแก้ไขการจดโน้ตให้แต่ละวัตถุ, ดูป้ายของวัตถุที่ติดไว้, ดูภาษาที่ใช้ในการแปลงแปลงผลผ่านโอซีอาร์ แล้วก็สามารถเปลี่ยนภาษาได้ โดยมีให้เลือกระหว่างภาษาไทย กับภาษาอังกฤษดังรูป 4.7 จ) และลบวัตถุ หรือก็คือรูปกากบาทสีแดง

สุดท้ายเมื่อทำการตีกรอบและติดป้ายเรียบร้อยแล้วก็เลือกกดที่ปุ่ม “SUBMIT” แล้วก็จะเป็นดังรูป 4.6 ฉ) แต่ถ้าต้องการยกเลิกการแก้ไขการทำแม่แบบนี้ให้เลือกกดที่ปุ่มรูปลูกศรย้อนกลับ

4.2.6 ส่วนตารางแสดงผลจากการแปลงผ่านโอซีอาร์ของแม่แบบในโปรเจค

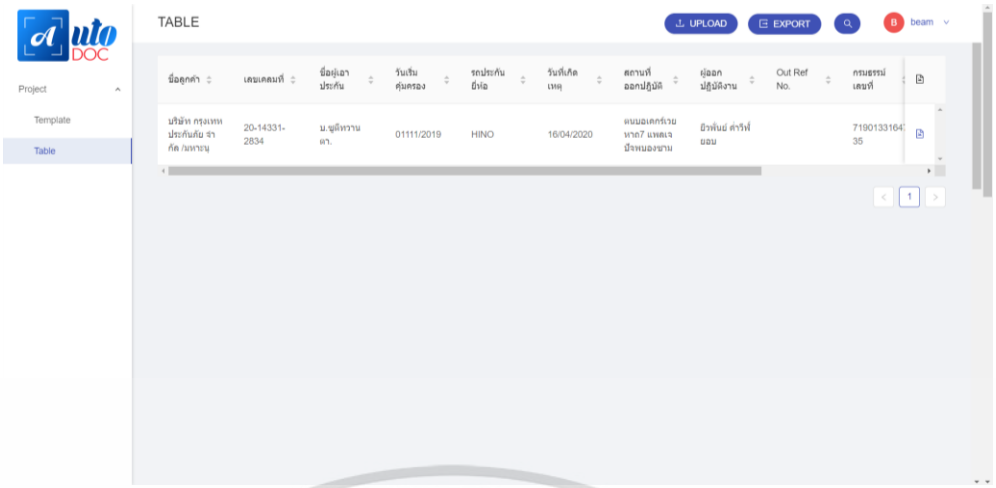


ก)



ข)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ค)

ง)

รูป 4.7 หน้าแสดงผลของตารางแสดงผลของโปรเจค

ก) เมื่อไม่มีข้อมูลการแปลงผลผ่านโอซีอาร์ในฐานข้อมูล

ข) เมื่อทำการเพิ่มไฟล์ภาพเอกสารแบบฟอร์ม

ค) เมื่อมีข้อมูลการแปลงผลผ่านโอซีอาร์ในฐานข้อมูล

ง) เมื่อทำการเลือกการส่งออกข้อมูลในตารางเป็นไฟล์

ส่วนของตารางแสดงผลของแม่แบบ จะแสดงผลจากการแปลงผ่านโอซีอาร์ของทั้งหมดในโปรเจคดังรูป 4.7 ก) เมื่อไม่มีข้อมูลการแปลงผลผ่านโอซีอาร์ในฐานข้อมูล และเป็นดังรูป 4.7 ค) เมื่อมีข้อมูลการแปลงผลผ่านโอซีอาร์ในฐานข้อมูล โดยหัวข้อจะได้มาจากป้ายในแม่แบบที่เลือก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ส่วนการอัปโหลดไฟล์ภาพเอกสารแบบฟอร์มที่มีรูปแบบเหมือนกับแม่แบบในโปรเจคนี้เพื่อทำการแปลงผลผ่านโอซีอาร์ จะสามารถเพิ่มได้โดยการกดปุ่ม “UPLOAD” ทางบนขวา แล้วเมื่อกดแล้วก็จะเป็นดังรูป 4.7 ข) โดยจะมีปุ่มให้อัปโหลดไฟล์ภาพเอกสารแบบฟอร์ม

ส่วนการส่งออกข้อมูลในตารางเป็นไฟล์ได้โดยการกดปุ่ม “EXPORT” ทางบนขวา แล้วเมื่อกดแล้วก็จะเป็นดังรูป 4.7 ง) โดยจะมีตัวเลือกให้เลือกรูปแบบการส่งออกเป็นไฟล์ CSV และ EXCEL

ส่วนการค้นหาข้อมูลสามารถทำได้โดยกดที่ปุ่มรูปแว่นขยาย จะสามารถใส่ข้อความที่ต้องการค้นหาและกดค้นหาได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
บทสรุปและข้อเสนอแนะ

5.1 สรุปผลที่ได้จากปริญญานิพนธ์

5.1.1 ส่วนโมเดลทำนายข้อความอิเล็กทรอนิกส์จากไฟล์ภาพเอกสาร

เนื่องจากการจัดหาชุดข้อมูลเอกสารภาษาไทยของจริงทำได้ยากปัญหาและอุปสรรค ทางคณะผู้จัดทำปริญญานิพนธ์จึงทำการสร้างชุดข้อมูลมาสำหรับการฝึกอบรมขึ้นมาทั้งภาษาไทย และภาษาอังกฤษ โดยทำการใช้หลายฟอนต์ในการสร้างชุดข้อมูล เพื่อให้มีความหลากหลายของข้อมูล ประกอบด้วยการใช้เทคนิค Image Augmentation เข้ามาช่วย หลังจากที่ทำการเตรียมข้อมูลเสร็จเรียบร้อย ทางคณะผู้จัดทำปริญญานิพนธ์ได้ทำการฝึกอบรมกับสถาปัตยกรรมตามงานวิจัย CRNN ซึ่งตัวสถาปัตยกรรมนี้รองรับเพียงภาษาเดียว คณะผู้จัดทำปริญญานิพนธ์จึงนำเทคนิค Multitask Learning มาช่วย เพื่อแบ่งปัน Bi-Directional GRU ของทุกภาษา และแต่ละภาษาจะมี Bi-Directional GRU ของแต่ละภาษาอีกทีในขั้นตอนต่อไป ทำให้โมเดลนี้สามารถเลือกได้ว่าจะให้ทำนายภาษาใด และเนื่องจากการฝึกอบรมโมเดลแต่ละครั้งใช้เวลานานมาก คณะผู้จัดทำปริญญานิพนธ์จึงทดลองได้เพียงแค่สถาปัตยกรรม Resnet-50 หลังจากการฝึกอบรมเสร็จแล้วพบว่าโมเดลสามารถรู้จำอักษรภาษาไทยมี Character error rate อยู่ที่ 7.19 % ซึ่งยังไม่สามารถใช้ในชีวิตจริงได้ ทำให้ต้องทดลองกับสถาปัตยกรรมอื่นและหาชุดข้อมูลของจริงมาเพิ่ม เพื่อให้โมเดลทำงานได้ดีกว่านี้ในอนาคต

5.1.2 ส่วนเว็บแอปพลิเคชัน

เนื่องจากคณะผู้จัดทำปริญญานิพนธ์ได้เล็งเห็นปัญหาที่ระบบโอซีอาร์ที่เปิดให้ใช้งานทั่วไปนั้นจะให้ผลลัพธ์จากการแปลงโดยโอซีอาร์เป็นข้อความอิเล็กทรอนิกส์ หรือไม่ก็ข้อความอิเล็กทรอนิกส์กับตำแหน่งของข้อความอิเล็กทรอนิกส์นั้น หรือก็คือไม่มีระบบฐานข้อมูลและแม่แบบ จึงได้ทำเว็บแอปพลิเคชันสำหรับเป็นเครื่องมือที่ช่วยการจัดการแม่แบบก่อนแปลงผลผ่านโอซีอาร์และจัดการในการเก็บข้อมูลหลังแปลงผลผ่านโอซีอาร์ โดยเว็บแอปพลิเคชันจะประกอบไปด้วย ระบบโปรเจคที่ให้บริการกับผู้ใช้งานในการจัดการโปรเจค โดยแต่ละโปรเจคจะมีระบบแม่แบบที่ใช้ในการจัดการแม่แบบ, ระบบอัปโหลดที่ใช้ในการจัดการอัปโหลดไฟล์ภาพเอกสารแบบฟอร์มในรูปแบบตัวพิมพ์ที่ต้องการทำการแปลงผลผ่านโอซีอาร์ โดยจะนำผลที่ได้จากโอซีอาร์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

มาเก็บไว้ในฐานข้อมูล และระบบแสดงผลที่ใช้ในการดูข้อมูลผลที่ได้จากการแปลงผ่านโอซีอาร์ที่ถูกเก็บไว้ในฐานข้อมูล

5.2 ปัญหาและอุปสรรค

5.2.1 ส่วนโมเดลทำนายข้อความอิเล็กทรอนิกส์จากไฟล์ภาพเอกสาร

1) การจัดหาชุดข้อมูลภาษาไทยได้ยาก
2) เนื่องจากโมเดลใช้ RNN ทำให้การฝึกอบรมไม่สามารถทำแบบขนานได้มากนักจึงใช้เวลาฝึกอบรมนาน
3) ความแม่นยำที่ได้จากโมเดลไม่แม่นยำเท่าที่ควร

5.2.2 ส่วนเว็บแอปพลิเคชัน

1) เนื่องจากใช้ภาษาในการเขียนเว็บแอปพลิเคชัน เป็นภาษา TypeScript จึงต้องมีความระมัดระวังในการกำหนดและเรียกใช้ตัวแปรเป็นอย่างมาก

5.3 แนวทางการพัฒนาต่อ

5.3.1 ส่วนโมเดลทำนายข้อความอิเล็กทรอนิกส์จากไฟล์ภาพเอกสาร

1) ทำการทดลองกับสถาปัตยกรรมอื่นเพิ่มเติม
2) ทำการหาชุดข้อมูลภาษาไทยจากไฟล์ภาพเอกสารในรูปแบบตัวพิมพ์จริงมาใช้
3) ทำการพัฒนาโมเดลให้สามารถจำแนกภาษาได้อัตโนมัติ

5.3.2 ส่วนเว็บแอปพลิเคชัน

1) สร้างระบบสมาชิกให้โปรเจค เพื่อทำให้แต่ละผู้ใช้แบ่งปันการจัดการได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

[1] Ekraam Sabir, Stephen Rawls, and Prem Natarajan.2018. “Implicit Language Model in LSTM for OCR” [Online]. Available : https://arxiv.org/pdf/1805.09441.pdf

[2] Thomas M. Breuel, Adnan Ul-Hasan, Mayce Al Azawi and Faisal Shafait.2013. “High-Performance OCR for Printed English and Fraktur using LSTM Networks” [Online]. Available : https://www.researchgate.net/publication/260341302_High-Performance_OCR_for_Printed_English_and_Fraktur_using_LSTM_Networks

[3] Youngmin Baek, Bado Lee, Dongyoon Han, Sangdoo Yun, and Hwalsuk Lee.2019. “Character Region Awareness for Text Detection.” [Online]. Available : https://arxiv.org/abs/1904.01941

Arden Dertat.2560. **Applied Deep Learning - Part 4: Convolutional Neural Networks.** [Online]. Available : https://towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2

devahoy. **React คืออะไร ? + เริ่มต้นเขียน React.** [Online]. Available : https://devahoy.com/blog/2015/11/getting-started-with-reactjs/

Felix Silwimba.2561. **Bidirectional GRU for Text classification by relevance to SDG#3 indicators.** [Online]. Available : https://medium.com/@felixs_76053/bidirectional-gru-for-text-classification-by-relevance-to-sdg-3-indicators-2e5fd99cc341

Harald Scheidl.2561. **An Intuitive Explanation of Connectionist Temporal Classification.**[Online]. Available : https://towardsdatascience.com/intuitively-understanding-connectionist-temporal-classification-3797e43a86c

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

mindphp. **ขั้นตอนการติดตั้ง Pycharm (ไพชาร์ม) โปรแกรมที่ใช้เขียนภาษา Python (ไพทอน).** [Online]. Available : https://www.mindphp.com/บทเรียนออนไลน์/83-python/4881-installpycharm.html

mindphp.2560. **รู้จักกับ Visual Studio Code (วิชวล สตูดิโอ โค้ด) โปรแกรมฟรีจากค่ายไมโครซอฟท์.** [Online]. Available : https://mindphp.com/บทความ/microsoft/4829-visual-studio-code.html

N2N Solution Provider. **OCR คืออะไร.** [Online]. Available : http://n2n.co.th/?page_id=113

Natthawat Phongchit.2561. **Convolutional Neural Network (CNN) คืออะไร.** [Online]. Available : https://blog.datawow.io/มาลองดูวิธีการคิดของ-cnn-กัน-e3f5d73eebaa

Nuttakan Chuntra.2561. **OpenCV คืออะไร?.** [Online]. Available : https://medium.com/@nut.ch40/opencv-%E0%B8%84%E0%B8%B7%E0%B8%AD%E0%B8%AD%E0%B8%B0%E0%B9%84%E0%B8%A3-8771e2a4c414

Nuttavut Thongjor.2559. **[TypeScript#1] TypeScript คืออะไร? เรียนรู้ชนิดข้อมูลพื้นฐานของ TypeScript.** [Online]. Available : https://www.babelcoder.com/blog/posts/typescript-data-types

Sarayut Nonsiri, PhD. **ภาษาโปรแกรม Python คืออะไร ?.** [Online]. Available : https://www.9experttraining.com/articles/python-คืออะไร

Wikipedia, the free encyclopedia.2562**. Connectionist temporal classification.** [Online]. Available : https://en.wikipedia.org/wiki/Connectionist_temporal_classification

Wikipedia, the free encyclopedia.2562. **Python (programming language).** [Online]. Available : https://en.wikipedia.org/wiki/Python_(programming_language)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Wikipedia, the free encyclopedia.2562. **PyTorch.** [Online]. Available : https://en.wikipedia.org/wiki/Python_(programming_language)

บริษัทเอโอซอฟต์จำกัด.2561. **Python คืออะไร - ภาษา python ใช้ทำอะไร.** [Online]. Available : https://www.aosoft.co.th/article/322/Python-คืออะไร-ภาษา-python-ใช้ทำอะไร.html

วิกิพีเดีย สารานุกรมเสรี.2557. **การรู้จำอักขระด้วยแสง.** [Online]. Available : https://th.wikipedia.org/wiki/การรู้จำอักขระด้วยแสง

วิกิพีเดีย สารานุกรมเสรี.2560. **การรู้จำแบบ.** [Online]. Available : https://th.wikipedia.org/wiki/การรู้จำแบบ

วิกิพีเดีย สารานุกรมเสรี.2561. **คอมพิวเตอร์วิทัศน์.** [Online]. Available : https://th.wikipedia.org/wiki/คอมพิวเตอร์วิทัศน์

วิกิพีเดีย สารานุกรมเสรี.2562. **ปัญญาประดิษฐ์.** [Online]. Available : https://th.wikipedia.org/wiki/ปัญญาประดิษฐ์

วิทยาลัยเซาธ์อีสท์บางกอก. **พื้นฐาน ภาษา Typescript.** [Online]. Available : http://www.theerapone.com/sbc/courses/angular/TypeScript1.1.pdf

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้