

บทที่ 1

บทนำ

1.1 หลักการและเหตุผล

การใช้โซเชียลมีเดียอย่างแพร่หลายกับสถานการณ์ในปัจจุบันพบว่าการสื่อสารผ่านโซเชียลมีเดียโดยการพิมพ์ผ่านตัวอักษรมากขึ้น ซึ่งเป็นการสื่อสารที่ใช้ภาษาที่หลากหลายจากผู้ใช้งานอินเทอร์เน็ต และขาดความเป็นทางการในข้อความของตัวอักษร ทำให้ข้อความไม่สื่อความหมายใด ๆ หรือสื่อความหมายที่ผิดเพี้ยนไปจากที่ผู้พิมพ์ต้องการจะสื่อสาร ในข้อความที่พิมพ์นั้นจะมีการสะกดคำที่ถูกต้องตามหลักภาษาจะทำให้ข้อความนั้นสื่อความหมายตรงตามผู้พิมพ์ต้องการ โดยมีเทคโนโลยีทางด้านปัญญาประดิษฐ์การรู้จำตัวอักษร ตัวอย่างเช่น โปรแกรมที่สามารถแปลงลายมือเขียนของมนุษย์มาเป็นตัวพิมพ์อัตโนมัติ แต่ผลลัพธ์ที่ได้จากการแปลงลายมือเขียนของมนุษย์ยังมีข้อผิดพลาดอยู่ เช่น ผลลัพธ์ที่ได้จากการแปลงไม่ตรงกับลายมือเขียน ซึ่งอาจจะทำให้เสียเวลาตรวจสอบความถูกต้องในเอกสาร

ดังนั้นโครงการปริญญาโทฉบับนี้ นำเสนอการตรวจสอบความถูกต้องข้อความภาษาไทย (Thai Text Correction) โดยในงานนี้จะอยู่ในลักษณะของการนำข้อมูลที่ได้จากการรู้จำตัวอักษร โดยอาศัยเทคนิคการประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP) นำข้อมูลเข้าไปสู่ขบวนการเรียนรู้ที่เป็นการเรียนรู้เชิงลึก (Deep Learning) โดยมีการปรับปรุงและพัฒนาผลลัพธ์ที่ได้จากโปรแกรมประมวลผลลายมือเขียนเป็นตัวพิมพ์อัตโนมัติ มาทำนายผลข้อความที่มีค่าผิดพลาดมาปรับปรุงและพัฒนาเพื่อเพิ่มประสิทธิภาพความแม่นยำมากขึ้น

1.2 วัตถุประสงค์ของโครงการ

เพื่อเพิ่มประสิทธิภาพความแม่นยำของผลลัพธ์จากโปรแกรมประมวลผลลายมือเขียนเป็นตัวพิมพ์อัตโนมัติให้มีความถูกต้องและเกิดข้อผิดพลาดน้อยที่สุด

1.3 ขอบเขตของโครงการ

1.3.1 ไฟล์ข้อมูลนำเข้ามาเป็นข้อความภาษาไทย หรือ Text file เท่านั้น

1.3.2 ฐานข้อมูลที่ใช้ในการเรียนรู้และการทดสอบ นำมาจากข้อความที่ได้จากการรู้จำลายมืออัตโนมัติ Version 2 เป็นชุดข้อมูลที่รับเข้ามาจะอยู่ใน Text file ที่มีผลเฉลยแล้วมี 1,504,138 อักขระ

จากไฟล์ภาพ 61,184 ข้อความ (บรรทัด/ภาพ) โดย 1 บรรทัด จะประกอบด้วยอักขระตั้งแต่ 3 ตัวอักษร ถึง 104 ตัวอักษร

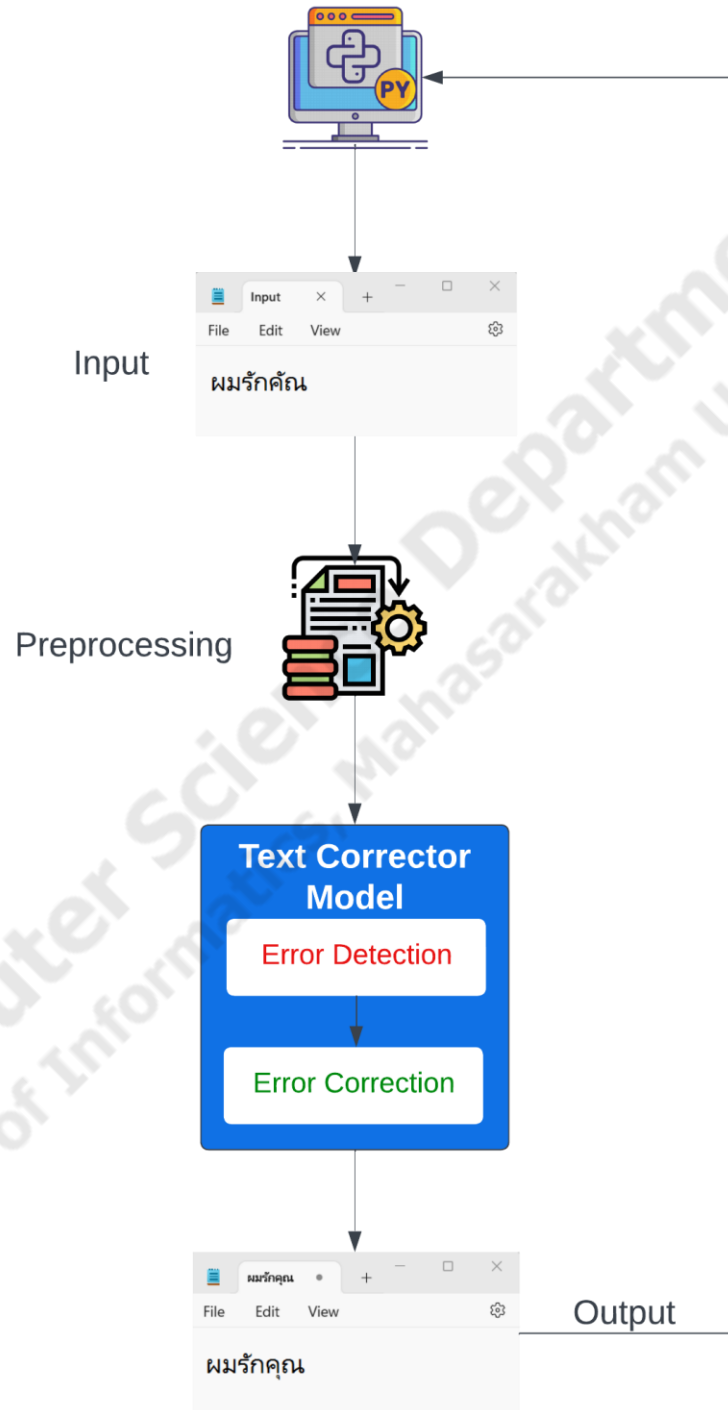
1.3.3 ผู้ใช้สามารถป้อนข้อความทดสอบได้

1.3.4 โปรแกรมสามารถระบุตำแหน่งคำที่ผิดพลาดได้

1.3.5 วิธีการตรวจสอบและแก้ไขข้อความใช้วิธีการแบบ Deep Learning

1.3.6 การประเมินผลมี 2 ขั้นตอน คือ Detection และ Correctness

1.4 ภาพรวมของระบบ



รูปภาพประกอบที่ 1.1 ภาพรวมระบบ

จากรูปภาพประกอบที่ 1.1 คือการทำงานของระบบ โดยที่ผู้ใช้เข้าใช้งานในโปรแกรม ผู้ใช้สามารถป้อนข้อมูลเป็นข้อความภาษาไทย หรือนำเข้าเป็น Text File ที่ต้องการประมวลผล จากนั้นระบบจะทำการ Pre-Processing ในส่วนของการตัดคำ (Word Segmentation) ก่อนนำเข้าสู่ Text Correction Model โดยมี 2 ขั้นตอน คือ Error Detection และ Error Correction โดยใช้ Deep Learning

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1.5.1 ทำให้ผลลัพธ์รู้จำตัวอักษรมีประสิทธิภาพแม่นยำมากขึ้น

1.5.2 ลดระยะเวลาในการตรวจสอบเอกสาร

1.5.3 พัฒนาอัลกอริทึมที่ใช้ในการตรวจสอบความถูกต้องของข้อความที่ได้จากการรู้จำให้มีความแม่นยำยิ่งขึ้น

1.6 อุปกรณ์และเครื่องมือที่ใช้ในการดำเนินงาน

1.6.1 ฮาร์ดแวร์

1.6.1.1 Dell

- Intel(R) Core(TM) i7-4500U CPU @ 1.80GHz 2.40 GHz
- RAM 8.00 GB
- Intel(R) HD Graphics Family
- 64-bit operating system

1.6.1.2 HP

- AMD Ryzen 3 2200U with Radeon Vega Mobile Gfx 2.50 GHz
- RAM 8.00 GB
- AMD Radeon R7 M440
- 64-bit operating system

1.6.2 ซอฟต์แวร์

- ภาษาที่ใช้ในการพัฒนา Python 3.7, JavaScript, Keras+TensorFlow
- โปรแกรม Figma ใช้ออกแบบหน้าโปรแกรม

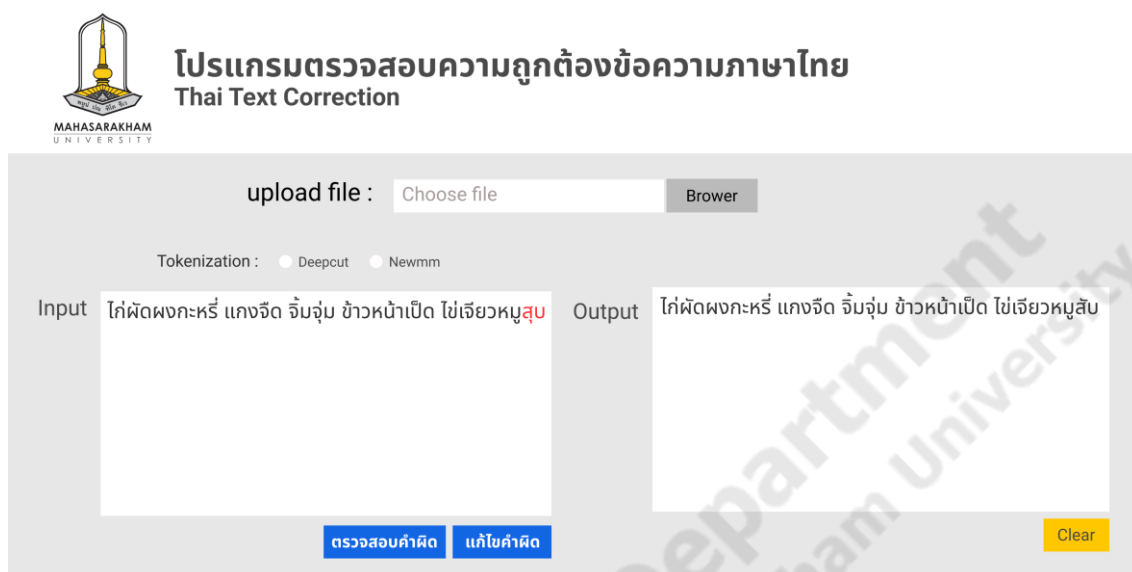
1.7 แผนการดำเนินงาน

โครงการปริญญาโทฉบับนี้ ดำเนินงาน ณ คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม ระหว่างเดือน พฤศจิกายน 2564 ถึง มิถุนายน 2566

ตารางที่ 1.1 แผนการดำเนินงาน

กิจกรรม	เดือน											
	พ.ย.	ธ.ค.	ม.ค.	ก.พ.	มี.ค.	เม.ย.	พ.ค.	มิ.ย.	ก.ค.	ส.ค.	ก.ย.	ต.ค.
1. ศึกษาและรวบรวมข้อมูล	■	■										
2. วิเคราะห์และกำหนดขอบเขต			■	■	■							
3. ออกแบบระบบ						■	■	■				
4. พัฒนาโปรแกรม									■			
5. ทดสอบระบบ										■	■	
6. ทำรายงานสรุป										■	■	
7. นำเสนอโครงการ										■	■	

1.8 ตัวอย่างโปรแกรม



The screenshot shows the interface of the Thai Text Correction application. At the top left is the Mahasarakham University logo. The title is "โปรแกรมตรวจสอบความถูกต้องข้อความภาษาไทย" (Thai Text Correction) and "Thai Text Correction". Below the title, there is an "upload file:" section with a "Choose file" input field and a "Browser" button. Underneath, there are radio buttons for "Tokenization" with options "Deepcut" (selected) and "Newmm". The "Input" field contains the text "ไก่ผัดผงกะหรี่ แองจืด จิ้มจุ่ม ข้าวหน้าเป็ด ไข่เจียวหมูสับ" and the "Output" field contains the corrected text "ไก่ผัดผงกะหรี่ แองจืด จิ้มจุ่ม ข้าวหน้าเป็ด ไข่เจียวหมูสับ". At the bottom, there are three buttons: "ตรวจสอบคำผิด" (Check spelling), "แก้ไขคำผิด" (Correct spelling), and "Clear".

รูปภาพประกอบที่ 1.2 ตัวอย่างโปรแกรม