บทความวิจัย

Computer Science Department
Faculty of Informatics, Mahasarakham University

การตรวจสอบความถูกต้องข้อความภาษาไทย

Thai Text Correction

ตีรณา สุระนารถ, อารีนันท์ ชัยณรงค์, ผู้ช่วยศาสตราจารย์ ดร.รพีพร ช่ำชอง
สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม
tirana.fai@gmail.com, areenunf.f@gmail.com, rapeeporn.c@msu.ac.th

Computer Science Department
Faculty of Informatics, Mahasarakham University

บทนำ

เทคโนโลยีสำหรับภาษาไทยได้มีหลากหลายมีการพัฒนาอยู่ตลอดทำให้เทคโนโลยีเหล่านั้นต่างมีประสิทธิภาพต่างกัน แต่เนื่องจากอักษรภาษาไทยมีความซับซ้อนมาก เช่น การสื่อสารผ่านโซเชียลโดยการพิมพ์ผ่านตัวอักษร ทำให้ข้อความสื่อความหมายที่ผิดเพี้ยนไปจากที่ผู้พิมพ์ต้องการจะสื่อสาร โครงงานปริญญานิพนธ์นี้ นำเสนอการตรวจสอบความถูกต้องข้อความภาษาไทย (Thai Text Correction) โดยในงานนี้จะอยู่ในลักษณะของการนำข้อมูลที่ได้จากการรู้จำตัวอักษร โดยอาศัย เทคนิคการประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP) นำข้อมูลเข้าไปสู่ ขบวนการเรียนรู้ที่เป็นการเรียนรู้เชิงลึก (Deep Learning) โดยมีการปรับปรุงและพัฒนาผลลัพธ์ที่ได้ จากโปรแกรมประมวลผลลายมือเขียนเป็นตัวพิมพ์อัตโนมัติ มาทำนายผลข้อความที่มีค่าผิดพลาดมา ปรับปรุงและพัฒนาเพื่อเพิ่มประสิทธิภาพความแม่นยำมากขึ้น

1.บทนำ

การใช้โซเชียลมีเดียอย่างแพร่หลายกับสถานการณ์ในปัจจุบันพบว่ามีการสื่อสารผ่านโซเชียล โดยการพิมพ์ผ่านตัวอักษรมากขึ้น ซึ่งเป็นการสื่อสารที่ใช้ภาษาที่หลากหลายจากผู้ใช้บนอินเทอร์เน็ต และขาดความเป็นทางการในข้อความของตัวอักษร ทำให้ข้อความไม่สื่อความหมายใด ๆ หรือสื่อความหมายที่ผิดเพี้ยนไปจากที่ผู้พิมพ์ต้องการจะสื่อสาร ในข้อความที่พิมพ์นั้นจะมีการสะกดค่าที่ถูกต้องตามหลักภาษาจะทำให้ข้อความนั้นสื่อความหมายตรงตามที่ผู้พิมพ์ต้องการ โดยมีเทคโนโลยีทางด้านปัญญาประดิษฐ์การรู้จำตัวอักษร ตัวอย่างเช่น โปรแกรมที่สามารถแปลงลายมือเขียนของมนุษย์มาเป็นตัวพิมพ์อัตโนมัติ แต่ผลลัพธ์ที่ได้จากการแปลงลายมือเขียนของมนุษย์ยังมีข้อผิดพลาดอยู่ เช่น ผลลัพธ์ที่ได้จากการแปลงไม่ตรงกับลายมือเขียน ซึ่งอาจจะทำให้เสียเวลาตรวจสอบความถูกต้องในเอกสาร

2.ทฤษฎีและระบบงานที่เกี่ยวข้อง

2.1 หลักการสะกดคำภาษาไทย การสะกดค่า [1] คือ การอ่านโดยน่าพยัญชนะต้น สระ วรรณยุกต์ และตัวสะกดมาประสมเป็นคำอ่าน ถือเป็นเครื่องมือการอ่านคำใหม่ โดยการสะกดคำนั้นมีหลากหลายวิธี ได้แก่สะกดตามรูปคำ สะกดโดยสะกดแม่ ก กา ก่อน แล้วจึงสะกดมาตราตัวสะกด ค่าที่มีสระอยู่หน้าพยัญชนะ ให้สะกดพยัญชนะก่อนสระเสมอ ค่าที่เป็นสระลดรูปหรือสระเปลี่ยนรูปค่าอักษรควบ สะกดเรียง

ตัวอักษร มุ่งเพื่อการเขียนให้ถูกต้องค่าอักษรน่าสะกดเรียงตัวอักษร มุ่งเพื่อการเขียนให้ถูกต้องค่าที่ตัวสะกดไม่ตรงตามมาตราตัวสะกด ค่าที่มีตัวการันต์ให้ใช้หลักสังเกตรูปค่า รู้ความหมายของคำและจำคำให้ได้โดยอ่านและเขียน

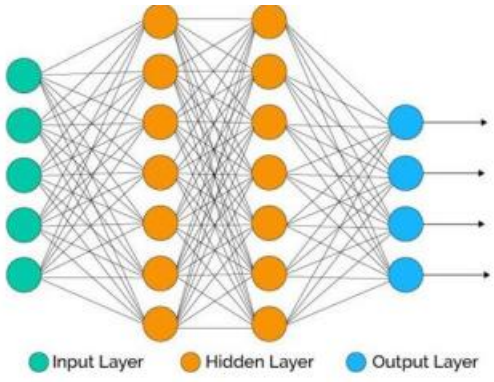
2.2 การตรวจสอบการสะกดผิด (Error Detection) การตรวจสอบการสะกดผิด [2] เป็นการสกัดเอาค่าจากข้อความที่ป้อนเข้ามาไปเทียบกับคำศัพท์ที่มีค่าที่สะกดถูกต้อง เช่น คำศัพท์จากพจนานุกรม ว่าตรงกันหรือไม่ ถ้าตรวจสอบเจอคำที่สะกดไม่ถูกต้องจะทำการเปลี่ยนสีของคำที่สะกดผิด โดยวิธีนี้สามารถตรวจสอบได้เฉพาะค่าที่เป็นการสะกดผิดแบบไม่เป็นคำเท่านั้น

2.3 การแก้ไขการสะกดผิด (Error Correction) การแก้ไขการสะกดผิด [2] สามารถแบ่งย่อยออกเป็น ขั้นตอนการค้นหาค่าที่น่าจะเป็นไปได้และขั้นตอนการคัดเลือกค่าที่ถูกต้องเหมาะสมมากที่สุด ในขั้นตอนของการค้นหาค่าที่น่าจะเป็นไปได้ โดยส่วนใหญ่แล้วค่าที่น่าจะเป็นมักจะได้มาจากการปรับแก้ไขค่าที่สะกดผิดนั้นให้ถูกต้องโดยมีจำนวนของการแก้ไขน้อยครั้งที่สุด และน่าไปใช้แก้ไขคำที่สะกดผิด แต่ถ้าเจอคำที่น่าจะเป็นไปได้มากกว่าหนึ่งคำจะต้องเข้าขั้นตอนสุดท้าย คือ การคัดเลือกคำที่ถูกต้องเหมาะสมมากที่สุดจากรายการคำที่น่าจะเป็น ซึ่งวิธีการคัดเลือกคำที่ เหมาะสมที่สุดนั้นจะขึ้นอยู่กับผู้วิจัยของแต่ละคน คือ ถ้าได้คำที่น่าจะเป็นคำที่สะกดถูกก็จะแนะนำคำเหล่านั้นให้ผู้ใช้ได้เลือกคำที่เหมาะสม

2.4 การประมวลผลภาษาธรรมชาติ การประมวลผลภาษาธรรมชาติ หรือ Natural Language Processing (NLP) [3] เทคนิคที่ใช้ในการประมวลผลภาษาธรรมชาติที่ผู้วิจัยได้นำมาใช้ในงานวิจัยมีดังนี้

การตัดคำ (word segmentation) [4] คือ การแยกข้อความในเอกสารออกเป็นคำ โดยใช้วิธีการต่าง ๆ ได้แก่ การใช้กฎการตัดคำ การใช้พจนานุกรมาโดยการใช้พจนานุกรมเป็นการตัดค่าที่ใช้สายอักขระมาเทียบกับคำที่มีอยู่ในพจนานุกรม ซึ่งวิธีนี้จะต้องจัดเก็บคำไว้ในพจนานุกรม และวิธีนี้ทำให้ได้ความถูกต้องในการตัดคำสูงมากกว่าการใช้กฎแต่จะใช้เวลานานมากกว่า การใช้คลังข้อความซึ่งการตัดคำโดยใช้คลังข้อมูลแบ่งออกเป็น 2 วิธีคือการตัดคำโดยอาศัยความน่าจะเป็น(Probabilistic Word Segmentation) และวิธีการตัดคำโดยอาศัยคุณลักษณะของคำ (Feature-based word segmentation)

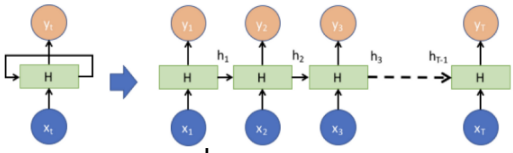
2.4 Neural Networks (NN) [6] หรือโครงข่ายประสาทเทียมนั้นมาจากเลียนแบบการทำงานของระบบประสาทของมนุษย์ จะจำลองการทำงานเหมือนกลุ่มเซลล์ประสาทที่เชื่อมโยงกัน สามารถรับรู้หลาย ๆ ได้ในเวลาเดียวกัน



รูปภาพประกอบที่ 2.1 โครงสร้าง Neural Networks

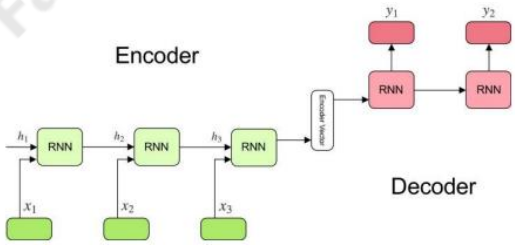
Computer Science Department
Faculty of Informatics, Mahasarakham University

2 . 5 Recurrent Neural Network (RNN) [7] เป็นลักษณะทั่วไปของโครงข่ายประสาทเทียมแบบ Feed Forward ที่มีหน่วยความจำภายใน RNN จะเกิดขึ้นซ้ำอีกเนื่องจากทำหน้าที่เดียวกันสำหรับทุกอินพุตของข้อมูลในขณะที่เอาต์พุตของอินพุตปัจจุบันขึ้นอยู่กับการคำนวณหนึ่งครั้งที่ผ่านมา หลังจากสร้างเอาต์พุตแล้วจะถูกคัดลอกและส่งกลับไปยังเครือข่ายที่เกิดซ้ำ สำหรับการตัดสินใจจะพิจารณาอินพุตปัจจุบันและเอาต์พุตที่ได้เรียนรู้จากอินพุตก่อนหน้า



รูปภาพประกอบที่ 2.2 การทำงานของ RNN

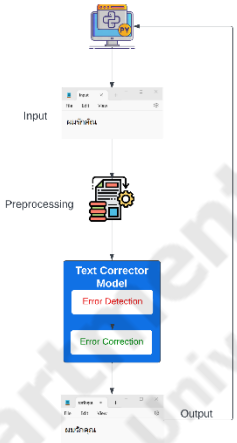
2.6 Sequence-to-Sequence [8] หรือเรียกอีกชื่อว่า RNN Encoder-Decoder ซึ่งโมเดลนี้จะแบ่งเป็นสองส่วน เรียกว่า Encoder กับ Decoder โดยส่วน Encoder จะรับ Input เข้ามาทีละหน่วยผ่านทาง RNN และเก็บสะสม Information ที่จำเป็นไว้ จากนั้นจะผ่าน Information นี้ไปยังส่วน Decoder ซึ่งก็จะเป็น RNN อีกตัวหนึ่งที่ให้ Output ออกมาทีละหน่วย โดยดูจาก Information ที่ได้รับมา และ Output ตัวก่อนหน้า ซึ่งแผนภาพของ Seq2Seq



รูปภาพประกอบที่ 2.3 การทำงานของ Seq2Seq

3. วิธีดำเนินงานวิจัย

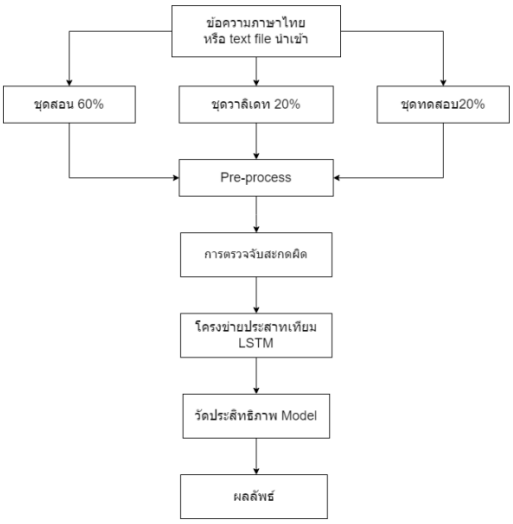
3.1 กรอบการดำเนินงาน



รูปภาพประกอบที่ 3.1 ภาพรวมระบบ

ผู้ใช้เข้าใช้งานบน Website โดยผู้ใช้สามารถป้อนข้อมูลเป็นข้อความภาษาไทย 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

3.2 การทำงานของการสร้าง Model



รูปภาพประกอบที่ 3.2 ตัวอย่างการทำงานการสร้าง Model

Computer Science Department
Faculty of Informatics, Mahasarakham University

3.2.1 การรวบรวมข้อมูล ในส่วนนี้จะอธิบายถึงการรวบรวมข้อมูลและลักษณะข้อมูลที่ใช้ในการทดสอบการตรวจสอบความถูกต้องข้อความภาษาไทย ข้อมูลเอกสารที่ใช้ในโครงงานฉบับนี้ ได้นำข้อมูลมาจากโครงงานการประมวลผลลายมือเขียนเป็นตัวพิมพ์อัตโนมัติ จำนวน 1,504,138 อักขระ จากไฟล์ภาพ 61,184 ข้อความ โดยเอกสารจัดเก็บในรูปแบบไฟล์ นามสกุล .txt

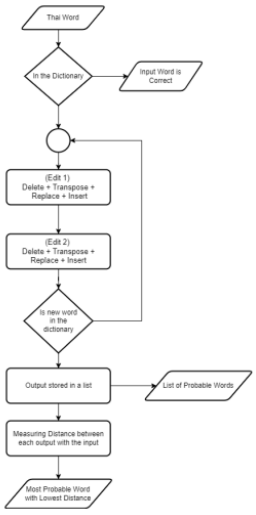
3.2.2 ขั้นตอนการทำงานของ Pre Processing 1.การตัดคำด้วยอัลกอริทึม Newmm [15] เป็นการตัดคำโดยการหา Maximum Matching จากใน Dictionary (TCC) ซึ่งวิธีการตัดคำแบบ Maximal Matching (MM) นี้ จะน่าเอาเฉพาะประโยคภาษาไทย โดยจะน่าประโยคนั้นไปเปรียบเทียบกับคำศัพท์ที่อยู่ในพจนานุกรม หากเปรียบเทียบแล้วพบว่ามีในพจนานุกรมแสดงว่าเป็นคำที่มีความหมาย แต่หากไม่พบในพจนานุกรมจะทำการตัดอักขระโดยตัดอักขระทีละตัวจากข้างหลัง และนำไปเปรียบเทียบอีกครั้ง และท่าซ้ำไปเรื่อย ๆ จนหมดประโยค และจะทำไปเรื่อย ๆ จนหมดบทความนั้น ๆ โดยข้อความที่ใช้ตัดอาจเป็นประโยคที่มีบรรทัดเดียวหรือหลายบรรทัดก็ได้ โดยมีอัลกอริทึมทำงาน Maximal Matching โดยเริ่มจากการหาทางเลือกของรูปแบบการตัดค่าทั้งหมดที่เป็นไปได้ก่อนโดยทำการย้อนกลับ (Backtracking) ทีละคำหลังจากได้คำตอบจากวิธีการเทียบค่าที่ยาวที่สุดแล้ว แล้วจึงเลือกทางเลือกที่มีจำนวนคำน้อยที่สุด

2. การตัดคำด้วย Deep Cut [17] จะใช้ Deep Learning Model CNN หาความน่าจะเป็นของอักษรแต่ละตัวว่าเป็นตัวเริ่มต้นของคำหรือไม่ โดยการทำนายแบบ Binary Classification ซึ่งคำที่ได้จะออกมาเป็น 0 หรือ 1 เท่านั้นโดยความหมายของ 0 คือไม่เป็นตัวเริ่มต้นของค่า และ 1 เป็นตัวเริ่มต้นของคำ โดยการสร้าง Model Deep Cut นี้จะน่าข้อมูลในการเรียนรู้ Model มาจาก BEST I Corpus ซึ่งเป็นข้อความภาษาไทยที่มีการรวบรวมไว้เพื่อพัฒนามาตรฐานการประมวลภาษาไทย ในการเรียนรู้ของ Model Deep Cut นั้นจะใช้ข้อความที่เป็นประโยคเป็นข้อมูลนำเข้าและมีผลเฉลยเป็นประโยคที่ถูกตัดคำแล้วโดยจะเป็นค่าตัวเลขแทนตัวอักษรแต่ละตัวโดยมีค่า 0 และ 1 ซึ่งมี 0 นั้นมีความหมายว่าไม่เป็นอักษรเริ่มต้นของคำ ส่วน 1 นั้นมีความหมายว่าเป็นอักษรเริ่มต้นของคำ

3.2.3 หลักการทำงานของการแก้ไขข้อความ 1.การตรวจสอบคำผิด (Error Detection) การตรวจจับการสะกดผิด เป็นขั้นตอนแรกของการตรวจแก้การสะกดค่ามีหน้าที่ตรวจหาคำที่สะกดผิดในข้อความ การแก้การสะกดผิดเป็นงานที่มีนักวิจัยเสนอวิธีการต่าง ๆ แต่ยังไม่มีการระบุที่ชัดเจนว่าใช้วิธีใด โดยส่วนใหญ่วิธีการที่นำมาประยุกต์ใช้นั้นมีความแตกต่างกันตามวัตถุประสงค์ ในส่วนนี้จะกล่าวถึงวิธีการที่น่ามาใช้จัดการตรวจแก้ปัญหาการสะกดผิดโดยใช้ Spellchecker ใน Pythainpl ใช้อัลกอริทึม Norvig Spell Checker algorithm [14] คือ อัลกอริทึมสำหรับ Word based Spell Corrector ที่

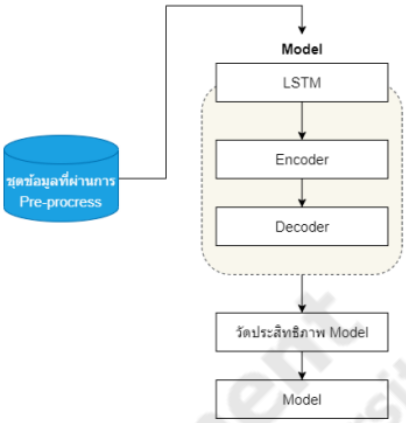
Computer Science Department
Faculty of Informatics, Mahasarakham University

สร้างขึ้นโดย Peter Norvig ในการตรวจจับคำผิด ซึ่งหลักการท่างานของอัลกอริทึม จะอาศัยคลังศัพท์พจนานุกรมจาก Thai National Corpus (TNC) [16]



รูปภาพประกอบที่ 3.10 ผังงานของวิธีการแก้ไขคำภาษาไทย

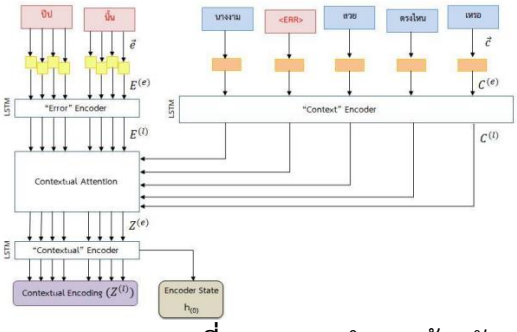
2. การแก้ไขการสะกดผิด (Error Correction) การแก้ไขการสะกดผิดนั้น จะทำการแก้ไขได้นั้นต้องผ่านขั้นตอนแรกคือการตรวจจับการสะกดคำผิดมาก่อน ซึ่งในการแก้ไขการสะกดคำผิดนั้น จะท่าการแก้ไขได้จำเป็นต้องมีข้อมูลของบริบท ดังนั้นบริบทที่เกิดขึ้นมีหลายรูปแบบ ซึ่ง Model ที่ใช้และสามารถตอบสนองได้ดีในตอนนี้คือ Recurrent Neural Network ที่เป็น Long Short-Term Memory (LSTM) ดังนั้นในการแก้ไขคำผิดนั้น เราจะใช้หลักการ Long Short-Term Memory โดยจะทำการ Input ข้อมูลเข้าไปเพื่อฝึกฝนโดยขั้นตอนการฝึกฝน



รูปภาพประกอบที่ 3.15 การทำงานของ Model LSTM

ขั้นตอนในส่วนตัวแก้ไขข้อความ Sequence-to-Sequence Model (Seq2Seq) เป็นโมเดลสร้างเวกเตอร์ของค่า โดยน่าเข้าข้อมูลเป็น Sequence ของหน่วยภาษา และผลลัพธ์เป็น Sequence ของหน่วยภาษา ภายในโมเดลประกอบด้วย Encoder ท่าหน้าที่รับข้อความที่อาจจะมีข้อผิดมาแล้วแปลงให้อยู่ในรูปของเวกเตอร์โดยเทคนิค “ord Em edding” และ Decoder ทำหน้าที่รับเวกเตอร์ โมเดลแบบ LSTM เช่นกัน ทำหน้าที่รับเวกเตอร์ไปสร้างเป็นข้อความที่ไม่มีข้อผิดที่ต้องการ

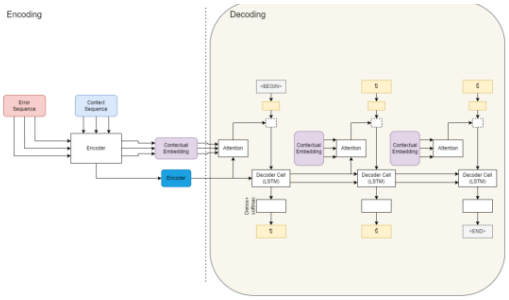
ขั้นตอนการทำงานในส่วนของ Encoder



รูปภาพประกอบที่ 3.19 การทำงานเข้ารหัส Encoder

Computer Science Department Faculty of Informatics, Mahasarakham University

ขั้นตอนการทำงานในส่วนของ Decoder



รูปภาพประกอบที่ 3.20 การทำงานการถอดรหัส Decoder

3.3 การประเมินประสิทธิภาพ

3.3.1 การประเมินประสิทธิภาพในการตัดคำ จะใช้ในการตรวจสอบการตัดคำที่ถูกต้องที่สุดซึ่งจะเปรียบเทียบประสิทธิภาพการตัดค่าของ 2 อัลกอริทึมด้วยว่าเหมาะสมที่จะนำมาใช้หรือไม่ โดยการนำผลลัพธ์การตัดคำของ Deep Cut และNewmm มาเปรียบเทียบกับผลเฉลยที่ได้จากการให้คนตรวจสอบมาทำการประเมินผลประเมินประสิทธิภาพในการตัดคำโดยใช้ Benchmarks ใน Pythainlp [15] โดยเลือกใช้วิธีเทียบจาก Word Level โดย Benchmarks วัดประสิทธิภาพจาก Precision, Recall และf1 **3.3.2 ประสิทธิภาพในการตรวจสอบคำผิด**การการประเมินประสิทธิภาพในการตรวจสอบค่าผิด จะใช้ประโยคที่มีคำผิดปะปนอยู่ และนำผลลัพธ์ของการตัดคำ มาเปรียบเทียบกับผลเฉลยที่ได้จากการให้คนตรวจสอบ จากนั้นจะประเมินประสิทธิภาพ [18] ตามเงื่อนไขต่อไปนี้

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

รูปภาพประกอบที่ 3.22 Confusion Matrix

3.3.3 ประสิทธิภาพในการแก้ไขคำผิด [19] โดยผู้จัดทำได้ประเมินประสิทธิภาพของผลลัพธ์ที่ได้จากการตรวจสอบความถูกต้องขอข้อความด้วย Word Error Rate (WER) [27] คือ อัตราความผิดพลาดของคำจากการทดสอบ การคำนวณ WER ต้องมีข้อความที่ถูกต้อง และผลลัพธ์จากกระบวนการทำนาย ซึ่งจะเปรียบเทียบระหว่างสองตัวอักษรหรือคำแต่ละตัว จำนวนคำที่ถูกต้องและคำที่ไม่ถูกต้องจะนับได้จากจำนวนของการแก้ไข การลบ และการเพิ่ม ที่เกิดขึ้น การคำนวณ WER จะใช้สูตรดังนี้

$$WER = (S + D + I) / N \qquad (15)$$

4.ผลการทดลอง

โครงงานปริญญานิพนธ์นี้นำเสนอการตรวจสอบความถูกต้องข้อความภาษาไทย ที่เกิดจากการป้อนข้อมูลที่อาจจะมีความผิดพลาดของคำ การสะกดผิด ซึ่งในบทนี้จะเป็นผลการทดลองในการประเมินประสิทธิภาพของผลลัพธ์ที่ได้จากโปรแกรมการตรวจสอบความถูกต้องข้อความภาษาไทย โดยแบ่งการทดสอบออกเป็น 3 ส่วน คือ 1.ผลการทดสอบการตัดคำ 2.ผลการตรวจสอบคำผิด 3.ผลการทดสอบแก้ไขคำผิด

Computer Science Department Faculty of Informatics, Mahasarakham University

5.สรุปผลและข้อเสนอแนะ

5.1 สรุปผลและอภิปราย โครงงานปริญญานิพนธ์นี้จะช่วยเพิ่มประสิทธิภาพของผลลัพธ์จากการป้อนข้อมูลข้อความที่ผิดพลาดในแบบภาษาไทย โดยอาศัยเทคนิคด้าน NLP มาใช้งาน เพื่อวิเคราะห์ผลลัพธ์ที่มีการตรวจสอบคำของข้อความที่สะกดผิดให้ถูกต้อง (Text Correction) ซึ่งวิธีการนี้จะช่วยให้แก้ไขคำในภาษาไทยมีความถูกต้องแม่นยำมากขึ้น จากการทดลองวัดประสิทธิภาพในการตรวจสอบคำผิดโดยใช้ไลบรารี่ Spell Checker ของ Pythainlp ที่มีพจนานุกรม เพื่อตรวจสอบคำผิดที่เกิดขึ้นในประโยค และ ใช้โมเดล LSTM (Long Short-Term Memory)เป็นแบบจำลองเครือข่ายประสาทเทียม (Neural Network) ทำในส่วนของการแก้ไขคำผิด จากประโยคทดสอบทั้งหมด 5,375 ประโยค วัดประสิทธิภาพจากค่าอัตราความผิดพลาดของคำคิดเป็น 0.201 เปอร์เซ็นต์ของประโยคทดสอบ

5.2 ปัญหาและอุปสรรคในการทำงาน

ขั้นตอนการทำงานเพื่อให้ได้ผลของการตรวจสอบคำของข้อความที่สะกดผิดและแก้ไขข้อความที่ผิดพลาดให้มีความถูกต้อง ได้พบปัญหาดังนี้ 1.เวลาในการเทรนชุดข้อมูล แต่ละครั้งใช้เวลานาน

5.3 ข้อเสนอแนะ

1.ควรใช้ข้อมูลในการเทรนที่หลากหลายและจำนวนมาก เพื่อให้เกิดความแม่นยำถูกต้อง
2.สามารถนำไปศึกษาและพัฒนาในการแก้ไขคำผิดต่อกับโมเดลที่เกิดขึ้นใหม่ ๆ

เอกสารอ้างอิง

[1]สถาบันภาษาไทย. (2559). หลักการสะกดคำภาษาไทย. สำนักวิชาการและมาตรฐานการศึกษา สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ.

[2]พลวัฒน์ ไหลมนู. (2559). การตรวจจับการสะกดผิดและการแก้ไขการสะกดผิด (วิทยานิพนธ์ ปริญญาบัณฑิต). กรุงเทพมหานคร: จุฬาลงกรณ์มหาวิทยาลัย.

[3]Keng Surapong. (2561). การประมวลผลภาษาธรรมชาติNaturalLanguageProcessing

[4]ปรางค์ปราณี ล้ำเลิศพระคุณ. (2561). การตัดคำ (วิทยานิพนธ์ ปริญญาบัณฑิต). กรุงเทพมหานคร: มหาวิทยาลัยธุรกิจบัณฑิตย์.

[5]ผศ. ดร. กานดารุณนะพงศาและนางสาวปโยธรอุราธรรมกุล. (2549). การตัดคาภาษาไทยโดยการปรับปรุงกฎและพจนานุกรมแบบใหม่. ขอนแก่น: มหาวิทยาลัยขอนแก่น.

[6]Artificial Neural Networks. (2561).

[7]Pisit Bee. (2563). Recurrent Neural Network(RNN).

[8]Ppp Mena. (2561). sequence-to-sequence(seq2seq).

[9]Long Short-Term Memory (LSTM). (2558)

[10]อนุรุทธ์ เลิศปิยะ, ตะวันรัตน์ ชโลธร และเอกพล ช่วงสุวนิช. (2563). Thai Spelling Correction and Word Normalization on Social Text Using a Two-Stage Pipeline With Neural Contextual Attention. IEEE.

Computer Science Department
Faculty of Informatics, Mahasarakham University

[11]จตุพร อินพานิชและวัชรพล นิลทัย. (2561). การแก้ไขข้อความที่ผิดพลาดโดยการประมวลผล ภาษาธรรมชาติ (วิทยานิพนธ์ปริญญาบัณฑิต). มหาสารคาม: มหาวิทยาลัยมหาสารคาม.

[12]อรทัย คงธรรม, ณัฐโชติ พรหมฤทธิ์ และสัจจะพร ไวจรรยา. (2562). Text-based LSTM Networks for Automatic Thai Love Quotes Generation on Twitter. Information Technology Journal.

[13]Wu, Hui & Huang, Yong & Li, Cancheng & Zhuang, Haoyu & Xu, Xinyu. (2021). Thai Word Segmentation Based on Sequence-to-Sequence Model. Journal of Physics.

[14]Anan Methasate. (2562). Thai Spell Correction. ม.ป.พ.

[15]Wannaphong Phatthiyaphaibun. (2559). PyThaiNLP:ThaiNatural Language Processing in Python. ม.ป.พ.

[16]คณาจารย์จากจุฬาลงกรณ์มหาวิทยาลัย. (2552). Thai National Corpus. กรุงเทพมหานคร:จุฬาลงกรณ์มหาวิทยาลัย.

[17]ตัดคำภาษาไทยโดยใช้ Deep learning (AI). (2560). ม.ป.พ.

[18]Pagon Gatchalee. (2562). Confusion Matrix เครื่องมือสำคัญในการประเมินผลลัพธ์ของการทำนาย ในMachine learning. ม.ป.พ.

[19]นางสาวจตุพร อินพานิชและนายวัชรพล นิลทัย. (2561). การแก้ไขข้อความที่ผิดพลาดโดยการประมวลผลภาษาธรรมชาติ. มหาสารคาม:มหาวิทยาลัยมหาสารคาม.

[20]นายทวีศักดิ์ ไชยวังราช. (2549). โปรแกรมวิเคราะห์ประโยคบอกเล่าของภาษาไทยผ่านเว็บเทคโนโลยี. พิษณุโลก:มหาวิทยาลัยนเรศวร.

[21]PradyaSin.(2562).Convolution Neural Network. ม.ป.พ.

[22]Rakpong Kittinaradorn, Titipat Achakulvisut, Korakot Chaovavanich, Kittinan Srithaworn, Pattarawat Chormai, Chanwit Kaewkasi, Tulakan Ruangrong, Krichkorn Oparad.(2019, September 23). DeepCut: A Thai word tokenization library using Deep Neural Network. Zenodo.

[23]A. Lertpiya, T. Chalothorn, and E. Chuangsuwanich,"ThaiSpellingCorrection and Word Normalization on Social Text Using a Two-Stage Pipeline With Neural Contextual Attention," in IEEE Access, vol. 8, pp. 133403-133419, 2020, DOI: 10.1109/ACCESS.2020.3010828.

[24]S. Islam, M. F. Sarkar, T. Hussain, M. M. Hasan, D. M. Farid, and S. Shatabda, "Bangla Sentence Correction Using Deep Neural Network Based Sequence to Sequence Learning," 2018 21st

Computer Science Department
Faculty of Informatics, Mahasarakham University

International Conference of Computer andInformationTechnology(ICCIT),2018,pp.1-6,DOI: 10.1109/ICCITECHN.2018.8631974.

[26]สิรินาถ ตั้งร่วมทรัพย์. (2017). Long Short-Term Memory (LSTM). ม.ป.พ.

[27]Word Error Rate in Python. (2566). ม.ป.พ.

Computer Science Department
Faculty of Informatics, Mahasarakham University