

การพัฒนาขั้นตอนวิธีและนำเสนอแนวทางการใช้งานการตรวจสอบการเขียนภาษาไทยด้วย
การประมวลผลภาพ สำหรับนักเรียนชั้นประถมศึกษาตอนปลาย
กรณีศึกษาโรงเรียนบ้านสันป่าไร่

Development of an algorithm and guidelines for thai writing verification using
image processing for upper primary school students: a case study of
bansanparai school

ฐิดารรัตน์ สิทธิ^{1*} วุฒิชัย พิสิฏ² ชยันต์ นันทวงศ์³

Thidararat Sitthi^{1*} Wudhijaya Philuek² Chayan Nuntawong³

^{1,2,3} สาขาวิชาคอมพิวเตอร์ศึกษา คณะครุศาสตร์ มหาวิทยาลัยราชภัฏนครสวรรค์

^{1,2,3} Program of Computer Education, The Faculty of Education, Nakhon Sawan Rajabhat University

*Corresponding Author E-mail : Thidararat934@gmail.com

วันที่รับบทความ 24 พฤศจิกายน 2568

วันแก้ไขบทความ 11 มีนาคม 2569

วันที่ตอบรับบทความ 1 เมษายน 2569

บทคัดย่อ(Abstract)

การวิจัยนี้มีวัตถุประสงค์เพื่อ 1) ประยุกต์ใช้กฎของซิปฟ์ (Zipf's Law) ในการวิเคราะห์คำศัพท์ภาษาไทย 2) ประยุกต์ใช้รูปแบบทางภาษาในการจำแนกคำเพื่อตรวจสอบการเขียนภาษาไทย 3) เพื่อพัฒนาตรวจสอบการเขียนประโยคภาษาไทย และ 4) เพื่อนำเสนอแนวทางในการใช้งานการตรวจสอบการเขียนภาษาไทยด้วยการประมวลผลภาพด้วยการวิจัยเชิงปริมาณ ข้อมูลที่ศึกษาเป็นข้อมูลจากการเขียนภาษาไทยของนักเรียนชั้นประถมศึกษาตอนปลาย โรงเรียนบ้านสันป่าไร่ จังหวัดตาก จำนวน 37 คน เครื่องมือวิจัย ได้แก่ แบบทดสอบคำศัพท์ภาษาไทย แบบทดสอบการเขียน ระบบตรวจสอบการเขียนด้วยการประมวลผลภาพและใช้สถิติเชิงพรรณนา การทดสอบประสิทธิภาพโมเดลและการประเมินรับรองด้วยผู้ทรงคุณวุฒิ ผลการวิจัยพบว่า 1) ผลประยุกต์ใช้กฎของซิปฟ์ (Zipf's Law) ในการวิเคราะห์ความรู้คำศัพท์ภาษาไทยสามารถจัดกลุ่มคำได้ 4 กลุ่มตาม ควอไทล์ ได้แก่ คำที่เขียนได้มากที่สุด คำง่าย คำปานกลาง และคำยาก 2) ประยุกต์ใช้โมเดลภาษา Trigram ร่วมกับการรู้จำลายมือและโครงข่ายประสาทแบบคอนโวลูชัน จำแนกคำจากภาพลายมือและจัดลำดับคำได้ตรงกับประโยคต้นฉบับ 100% แม้ข้อมูลนำเข้าจะมีการจัดลำดับไม่แน่นอน 3) ผลการพัฒนาขั้นตอนวิธีตรวจสอบการเขียนประโยคภาษาไทย มีผลลัพธ์ Accuracy, Precision, Recall และ F1-score = 1.0 ทั้งในชุดฝึกและชุดตรวจสอบ รวมถึงไม่มีค่าผิดพลาดใน Confusion Matrix ของทั้งประโยคที่ 1 และ 2 และ 4) แนวทางในการใช้งานการตรวจสอบการเขียนภาษาไทยด้วยการประมวลผลภาพมีความเหมาะสม ค่าเฉลี่ยอยู่ที่ 4.40 ความเหมาะสมอยู่ในระดับ มาก

คำสำคัญ (Keywords) : การตรวจสอบการเขียนภาษาไทย, การนำเสนอแนวทางการใช้งานการตรวจสอบการเขียนภาษาไทย, การประมวลผลภาพ, กฎของซิปฟ์, การประเมินความถูกต้องของการเขียน

Abstract

This research aims to 1) Apply Zipf's Law in the analysis of Thai vocabulary, 2) Apply language patterns in word classification to examine Thai writing, 3) Develop a system for checking Thai sentence writing, and 4) Propose guidelines for using image processing-based Thai writing verification through quantitative research. The studied data consisted of Thai writing samples from 37 upper-elementary students (late primary level) at Ban San Parai School, Tak Province. Research instruments included a Thai vocabulary test, a writing test, an image-processing-based writing verification system, descriptive statistics, model performance testing, and expert validation.

The research findings were as follows: 1) The application of Zipf's Law in analyzing Thai vocabulary knowledge enabled the classification of words into 4 groups based on quartiles: most frequently written words, easy words, medium words, and difficult words. 2) The application of a Trigram language model combined with handwriting recognition and Convolutional Neural Network was able to classify words from handwritten images and correctly reorder them to match the original sentences with 100% accuracy, even when the input data had uncertain ordering. 3) The developed algorithm for checking Thai sentence writing achieved an Accuracy, Precision, Recall, and F1-score = 1.0 on both training and test sets. There were no errors in the Confusion Matrix for both Sentence 1 and Sentence 2. 4) The proposed guidelines for using image-processing-based Thai writing verification were deemed highly appropriate, with an average suitability score of 4.40 (on a 5-point scale), corresponding to a "high" level of appropriateness.

Keywords : Zipf's Law, Thai Handwriting Verification, Guidelines for Thai Writing Verification Application, Image Processing, Writing Accuracy Assessment

บทนำ (Introduction)

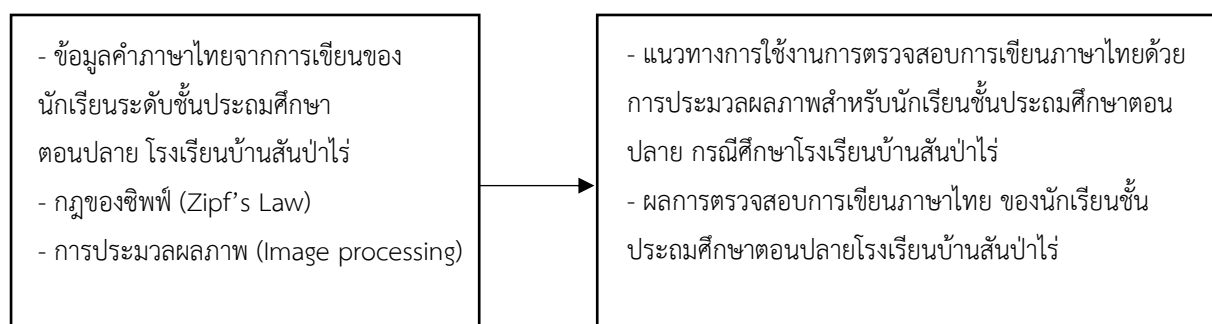
การเขียนภาษาไทยมีความสำคัญอย่างยิ่งต่อการพัฒนาทักษะการอ่านและการเรียนรู้ ซึ่งเป็นพื้นฐานของการศึกษาในทุกวิชา ทั้งยังสะท้อนถึงมรดกทางวัฒนธรรมและช่วยปลูกฝังความภาคภูมิใจในเอกลักษณ์ของชาติ อย่างไรก็ตาม การเขียนด้วยลายมือมักเผชิญปัญหาความแตกต่างเฉพาะบุคคล ความคล้ายคลึงของอักษร และการเขียนที่ไม่ถูกต้อง ซึ่งอาจก่อให้เกิดความสับสน ความล่าช้าในการอ่าน และข้อผิดพลาดในการจดจำ (ศรายุทธ กอนวิรัตน์. 2561: 1) ภาษาไทยในฐานะภาษาหลักมีบทบาทสำคัญต่อการสื่อสารและการเรียนรู้ โดยเฉพาะในระดับประถมศึกษาตอนปลายซึ่งเป็นช่วงพัฒนาทักษะการเขียนที่ซับซ้อนมากขึ้น แต่ในบางพื้นที่ที่มีความหลากหลายทางชาติพันธุ์ เช่น โรงเรียนบ้านสันป่าไร่ นักเรียนจำนวนมากใช้ภาษาถิ่นในชีวิตประจำวัน ทำให้ประสบปัญหาในการเรียนรู้และเขียนภาษาไทยอย่างถูกต้อง ส่งผลกระทบต่อวิชาอื่น ๆ เพื่อแก้ไขปัญหา

งานวิจัย “การพัฒนาขั้นตอนวิธีและแนวทางการใช้งานการตรวจสอบการเขียนภาษาไทยด้วยการประมวลผลภาพ สำหรับนักเรียนชั้นประถมศึกษาตอนปลาย กรณีศึกษาโรงเรียนบ้านสันป่าไร่” มุ่งประยุกต์ใช้เทคโนโลยีการประมวลผลภาพ (Image Processing) เพื่อตรวจสอบการเขียนได้อย่างรวดเร็วและแม่นยำ ลดภาระครู และให้ข้อเสนอแนะกับผู้เรียนทันที โดยอาศัยการวิเคราะห์เชิงภาษาด้วยกฎของซิปฟ์ (Zipf’s Law) เพื่อประเมินการเขียนประโยคภาษาไทยที่มีความซับซ้อน งานวิจัยนี้จึงมีความสำคัญทั้งด้านการพัฒนาเทคโนโลยีการเสริมสร้างทักษะภาษา และการลดความเหลื่อมล้ำทางการศึกษา อีกทั้งการเขียนภาษาไทยช่วยปลูกฝังความภาคภูมิใจในเอกลักษณ์และมรดกทางวัฒนธรรม ทำให้นักเรียนสามารถสื่อสารและแสดงออกทางความคิดสร้างสรรค์ได้อย่างมีประสิทธิภาพ อันเป็นรากฐานสำคัญสำหรับการเรียนรู้และการดำเนินชีวิตในอนาคต (ห้องสมุดวังสนามจันทร์. 2556 : ออนไลน์)

วัตถุประสงค์การวิจัย(Research Objectives)

1. เพื่อประยุกต์ใช้กฎของซิปฟ์ (Zipf’s Law) ในการวิเคราะห์ความรู้คำศัพท์ภาษาไทย สำหรับนักเรียนชั้นประถมศึกษาตอนปลาย กรณีศึกษาโรงเรียนบ้านสันป่าไร่
2. เพื่อประยุกต์ใช้รูปแบบทางภาษาในการจำแนกคำเพื่อตรวจสอบการเขียนภาษาไทย สำหรับนักเรียนชั้นประถมศึกษาตอนปลาย กรณีศึกษาโรงเรียนบ้านสันป่าไร่
3. เพื่อพัฒนาตรวจสอบการเขียนประโยคภาษาไทย สำหรับนักเรียนชั้นประถมศึกษาตอนปลาย กรณีศึกษาโรงเรียนบ้านสันป่าไร่
4. เพื่อนำเสนอแนวทางในการใช้งานการตรวจสอบการเขียนภาษาไทยด้วยการประมวลผลภาพ สำหรับนักเรียนชั้นประถมศึกษาตอนปลาย กรณีศึกษาโรงเรียนบ้านสันป่าไร่

กรอบแนวคิดในการวิจัย (ถ้ามี)



ภาพที่ 1 กรอบแนวคิดของการวิจัย

สมมติฐานการวิจัย (Research Hypothesis) (ถ้ามี)

-

วิธีดำเนินการวิจัย (Research Methods)

การวิจัยครั้งนี้เป็นการวิจัยเชิงปริมาณ โดยใช้วิธีการทดลองและการวิเคราะห์เชิงสถิติ โดยแบ่งกระบวนการออกเป็นสี่ระยะ ได้แก่ ระยะที่ 1 การประยุกต์ใช้กฎของซิปฟ์ (Zipf's Law) ในการวิเคราะห์ความรู้คำศัพท์ภาษาไทย ระยะที่ 2 การประยุกต์ใช้รูปแบบทางภาษาในการจำแนกคำเพื่อตรวจสอบการเขียนภาษาไทย ระยะที่ 3 การพัฒนาตรวจสอบการเขียนประโยคภาษาไทย และระยะที่ 4 การนำเสนอแนวทางในการใช้งานการตรวจสอบการเขียนภาษาไทยด้วยการประมวลผลภาพ ดังนี้

กลุ่มเป้าหมาย

1. การศึกษาครั้งนี้ใช้กลุ่มประชากรเป็นกลุ่มเป้าหมาย โดยเลือกกรณีศึกษาโรงเรียนบ้านสันป่าไร่ อำเภอมะเข่มาต ตำบลพระธาตุ จังหวัดตาก สังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษาตากเขต 2 โดยเป็นนักเรียนระดับชั้นประถมศึกษาปีที่ 4-6 จำนวน 37 คน ได้มาโดยการกำหนดขนาดของกลุ่มตัวอย่างแบบเจาะจง ซึ่งข้อมูลลายมือได้จากครูประจำชั้น ซึ่งได้มีการบันทึกข้อความขอลายมือนักเรียนพร้อมทั้งภาพถ่ายลายมือต่าง ๆ ไม่สามารถระบุตัวตนนักเรียนผู้เขียนได้

2. ผู้ทรงคุณวุฒิด้านการศึกษา ด้านวิทยาการคอมพิวเตอร์ศึกษาและครู รวมจำนวนทั้งหมด 10 ท่าน

เครื่องมือที่ใช้ในการรวบรวมข้อมูล

1. แบบทดสอบคำศัพท์ภาษาไทย จำนวน 100 คำ ซึ่งได้จากแบบประเมินการอ่านคำพื้นฐาน (Highland school and Island school Reading Test: HIS Test) ใช้กำหนดคำศัพท์ภาษาไทย ที่ใช้ในการเก็บข้อมูลความรู้คำศัพท์ภาษาไทยของนักเรียน

2. แบบทดสอบการเขียนประโยคภาษาไทย ซึ่งได้คัดเลือกคำจากแบบทดสอบคำศัพท์ภาษาไทยที่ผ่านการ วิเคราะห์ข้อมูลจัดแบ่งเป็นกลุ่มคำตามกฎของซิปฟ์ (Zipf's Law) นำมาแต่งประโยคที่ประกอบไปด้วยคำในแต่ละกลุ่มคำตามความถี่เพื่อทดสอบการเขียนภาษาไทยของนักเรียน

3. ระบบตรวจสอบการเขียนภาษาไทยด้วยการประมวลผลภาพ โดยใช้เทคนิคการประมวลผลภาพ (Image Processing) ร่วมกับโครงข่ายประสาทเทียมแบบคอนโวลูชัน (CNN) และแบบจำลองภาษา Trigram

4. แบบประเมินแสดงความคิดเห็นแนวทางการใช้งานการตรวจสอบการเขียนภาษาไทยด้วยการประมวลผลภาพ

วิธีรวบรวมข้อมูล และสถิติที่ใช้ในการวิเคราะห์

การเก็บรวบรวมข้อมูลแบ่งเป็น 2 ระยะ ดังนี้

ระยะที่ 1 เก็บข้อมูลคำศัพท์จากนักเรียนชั้นประถมศึกษาตอนปลายจำนวน 37 คน ซึ่งแต่ละคนเขียนคำศัพท์ จำนวน 100 คำ รวมทั้งสิ้น 3,700 คำ โดยใช้แบบทดสอบคำศัพท์ภาษาไทย และนำข้อมูลไปวิเคราะห์ด้วยกฎของซิปฟ์ (Zipf's Law) เพื่อจัดกลุ่มคำตามความถี่ (Quartile Analysis)

ระยะที่ 2 เก็บข้อมูลการเขียนประโยคภาษาไทยด้วยลายมือของนักเรียน เพื่อนำไปผ่านกระบวนการตัดคำ (Text Segmentation) และประมวลผลภาพด้วยโปรแกรม OpenCV แล้วนำเข้าสู่โมเดล CNN และ Trigram Language Model เพื่อจำแนกคำและตรวจสอบการเขียนประโยค

ผลการวิจัย (Research Results)

ระยะที่ 1 การประยุกต์ใช้กฎของซิปฟ์ (Zipf's Law) ในการวิเคราะห์ความรู้คำศัพท์ภาษาไทย สำหรับนักเรียนชั้นประถมศึกษาตอนปลาย กรณีศึกษาโรงเรียนบ้านสันป่าไร่

ผลการวิจัยในระยะที่ 1 การประยุกต์ใช้กฎของซิปฟ์ (Zipf's Law) ในการวิเคราะห์ความรู้คำศัพท์ภาษาไทย สำหรับนักเรียนชั้นประถมศึกษาตอนปลาย กรณีศึกษาโรงเรียนบ้านสันป่าไร่ มีดังต่อไปนี้

ส่วนที่ 1 ผลการเก็บรวบรวมข้อมูลคำศัพท์ภาษาไทย จำนวน 100 คำ กับนักเรียนกลุ่มตัวอย่าง จำนวน 37 คน

ตารางที่ 1 ตัวอย่างความถี่ความรู้คำศัพท์ภาษาไทยของนักเรียน จำนวน 10 คำ จาก 100 คำ

ลำดับ (Rank)	คำภาษาไทย	ความถี่คำ	Log (Rank)	Log (Frequency)
1	กระเจก	36	0.000	1.556
2	ธรรมชาติ	29	1.114	1.462
3	ภูมิใจ	26	1.447	1.415
4	ตกยาก	23	1.531	1.362
5	ขัดสน	22	1.633	1.342
6	กลิ้ง	20	1.708	1.301
7	ทันสมัย	18	1.778	1.255
8	ผ่องใส	13	1.875	1.114
9	ญาติ	11	1.886	1.041
10	พสกนิกร	1	1.964	0.000

จากการวิเคราะห์ข้อมูลการเขียนคำภาษาไทยด้วยวิธีการจัดกลุ่ม (Clustering) พบว่า สามารถแบ่งคำออกเป็น 4 กลุ่มคำตามระดับความยากง่าย คือ กลุ่มคำที่เขียนได้มากที่สุด กลุ่มคำง่าย กลุ่มคำปานกลาง และกลุ่มคำยาก ดังนี้

1. กลุ่มคำที่เขียนได้มากที่สุด กลุ่มนี้มีค่าควอไทล์ที่ 4 (Q4) เท่ากับ 26.25 แสดงถึงคำที่ผู้เรียนเขียนได้ถูกต้องบ่อยที่สุด คำส่วนใหญ่เป็นคำที่พบเห็นในชีวิตประจำวันและคุ้นเคย เช่น กระเจก ดำดิน ชัง กระไร นาค กระปุก ตวง คงคา สันบุญ เขียวขจี จัวเจีย ทนต์ ธรรมชาติ ขอม งูนง ดิฉัน เนื่อ ซุปซิบ ภิกษุ ซูป ผอม จารึก ถอย กล่อม เนื้อหา เป็นต้น

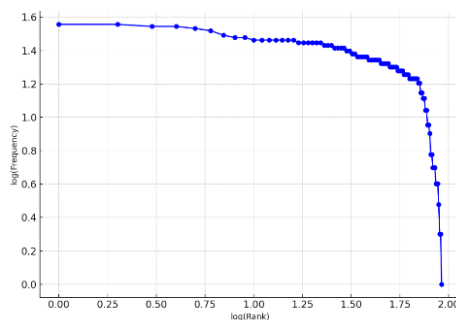
2. กลุ่มคำง่าย กลุ่มนี้มีค่าควอไทล์ที่ 3 (Q3) เท่ากับ 26.25 เช่นเดียวกัน ลักษณะคำอยู่ในระดับง่าย ผู้เรียนส่วนใหญ่สามารถเขียนได้ถูกต้อง คำในกลุ่มนี้ ได้แก่ เนือง ผอง ภูมิใจ จังหะ กระทบ กระทบ กระนั้น บ่ ตกยาก ทุนหัว เขียววัย จีบ สุนัขย สิ้นชีพ ขบ กลุ่ม กริ่ง ขัดสน คุณธรรม ฉิว ฉีก ด้อม ถดถอย กร้ว เป็นต้น

3. กลุ่มคำปานกลาง กลุ่มนี้มีค่าควอไทล์ที่ 2 (Q2) เท่ากับ 20.00 เป็นคำที่ผู้เรียนเขียนได้ในระดับปานกลาง ต้องใช้ความรู้ความเข้าใจมากขึ้น คำตัวอย่าง เช่น ภาควิชา กลิ้ง เทือกเขา โทณ คุณภาพ เปิดเผย ครางทรมาน ชะรอย คด ทนสมัย คลาย ขึ้น เชิญ เปื้อน จักรเย็บผ้า ซอกซ้า บอเลา กุศล ธุระ สิ้นชีวิต คำนวน บากบั่น เขียวขุ่ม กิณร ผ่องใส เป็นต้น

4. กลุ่มคำยาก กลุ่มนี้มีค่าควอไทล์ที่ 1 (Q1) เท่ากับ 12.50 เป็นคำที่นักเรียนส่วนใหญ่เขียนได้ยากที่สุด มีลักษณะเป็นคำที่ซับซ้อนหรือไม่คุ้นเคย เช่น ถ้อยที่ถ้อยอาศัย ภูตผี กิตติศัพท์ เค้างคว้าง ธรณี ธาณี ผ่องแผ้ว พฤษภา นัยนา ทาส ชลเนตร ครุฑา ชะงัก ไคลคลา ราชศัตรู นิวัติ พสกนิกร กรุงรัตนโกสินทร์ จัตุบาท พรหมณ์ รสคนธ์ นิราศ พลการ ภูวนาถ รูปพรรณ เป็นต้น

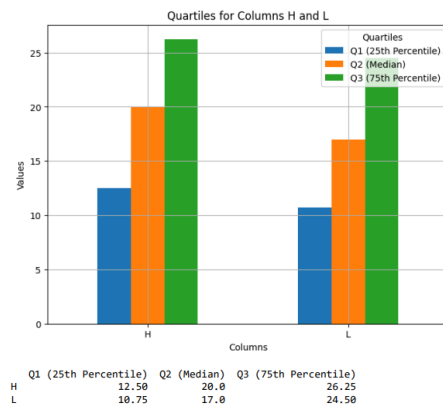
ผลการวิเคราะห์แสดงให้เห็นลำดับความสามารถในการเขียนคำภาษาไทยของนักเรียน เริ่มจากคำที่คุ้นเคยและใช้ในชีวิตประจำวัน (กลุ่มที่เขียนได้มากที่สุดและกลุ่มง่าย) ไปจนถึงคำที่มีความซับซ้อนทั้งด้านรูปคำและการใช้ (กลุ่มปานกลางและกลุ่มยาก) การจัดกลุ่มเช่นนี้ช่วยให้สามารถกำหนดแนวทางการสอนหรือฝึกฝนเพิ่มเติมในคำที่อยู่ในกลุ่มปานกลางและยากได้อย่างตรงจุด

ส่วนที่ 2 กราฟแสดงความสัมพันธ์ระหว่าง $\log(\text{Rank})$ และ $\log(\text{Frequency})$ (ตรวจสอบกฎของชิพฟ์ Zipf's Law) ซึ่งจากกราฟจะเห็นได้ว่าค่าของ $\log(\text{Rank})$ และ $\log(\text{Frequency})$ มีความสัมพันธ์ในลักษณะเส้นตรง แสดงว่าข้อมูลคำชุดนี้มีแนวโน้มเป็นไปตามกฎของชิพฟ์



ภาพที่ 1 กราฟแสดงความสัมพันธ์ระหว่างลอการิทึมของลำดับคำ ($\log(\text{Rank})$) และความถี่คำ ($\log(\text{Frequency})$) ซึ่งสอดคล้องกับกฎของ Zipf's Law

ส่วนที่ 3 การจัดกลุ่มคำศัพท์ภาษาไทยตามความถี่คำที่ได้จากกฎของชิพฟ์ (Zipf's Law) ที่ใช้การวิเคราะห์ด้วยควอไทล์ (Quartile Analysis) ในการจัดกลุ่มคำ โดยใช้ภาษา Python ร่วมกับไลบรารี Pandas เพื่อคำนวณค่าควอไทล์จากข้อมูลของตัวแปร H และ L

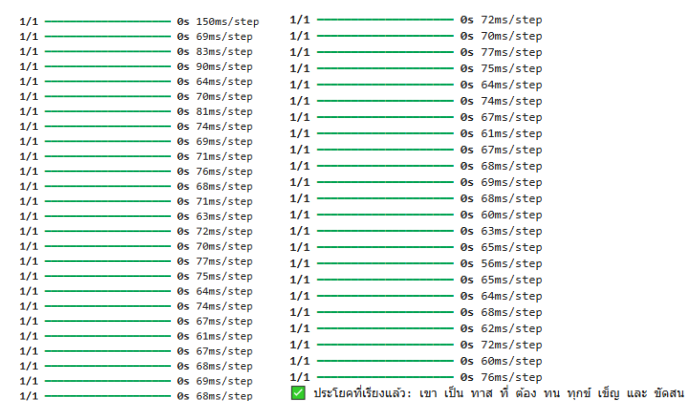


ภาพที่ 2 แผนภาพแท่งเปรียบเทียบค่าควอไทล์ (Quartiles) สำหรับคอลัมน์ที่มีชื่อว่า H และ L

ระยะที่ 2 ประยุกต์ใช้รูปแบบทางภาษาในการจำแนกคำเพื่อตรวจสอบการเขียนภาษาไทย สำหรับนักเรียนชั้นประถมศึกษาตอนปลาย กรณีศึกษาโรงเรียนบ้านสันป่าไร่

ผลการวิจัยในระยะที่ 2 ประยุกต์ใช้รูปแบบทางภาษาในการจำแนกคำเพื่อตรวจสอบการเขียนภาษาไทย สำหรับนักเรียนชั้นประถมศึกษาตอนปลาย กรณีศึกษาโรงเรียนบ้านสันป่าไร่ มีดังต่อไปนี้

1. รวบรวมและคัดเลือกข้อมูลคำศัพท์ภาษาไทย จากประโยคต้นแบบที่กำหนด โดยเลือกคำที่มีความสำคัญในระดับประถมปลาย
2. เตรียมข้อมูลและตัดข้อความ (Text Segmentation) จากภาพประโยคที่นักเรียนเขียนจริง แยกออกมาเป็นคำ ๆ ด้วยเทคนิค image processing โดยใช้ OpenCV
3. สร้างและฝึกฝนแบบจำลองทางภาษา (Language Model Training) โดยใช้รูปแบบ Trigram ในการจับคู่คำสามคำติดต่อกัน เพื่อสร้างโมเดลจำแนกลำดับคำ
4. การประเมินความแม่นยำของแบบจำลอง โดยนำภาพลายมือจริงของนักเรียนมาเป็นข้อมูลทดสอบ ดังรูปภาพนี้



ภาพที่ 3 แสดงผลการประมวลผลของระบบที่ใช้แบบจำลองรู้จำลายมือร่วมกับ Trigram Language Model

จากผลผลลัพธ์การทำนายพบว่า ระบบสามารถประมวลผลภาพลายมือของนักเรียนและจำแนกคำออกจากภาพได้อย่างถูกต้อง เมื่อนำคำที่แยกได้เข้าสู่กระบวนการจำแนกลำดับคำด้วยแบบจำลองภาษา

Trigram Language Model ระบบสามารถจัดเรียงลำดับคำได้ตรงกับประโยคต้นฉบับอย่างถูกต้องครบถ้วนทุกคำ โดยใช้เวลาเฉลี่ยประมาณ 60–80 มิลลิวินาทีต่อคำ ซึ่งสะท้อนถึงความแม่นยำและประสิทธิภาพของการนำโมเดลภาษาแบบ Trigram มาใช้ร่วมกับเทคนิคการรู้จำลายมือเพื่อตรวจสอบการเขียนภาษาไทยของนักเรียนระดับประถมศึกษาได้อย่างมีประสิทธิภาพ แม้ระบบจะได้รับภาพที่มีการจัดลำดับคำไม่แน่นอน แต่สามารถนำแบบจำลองภาษา Trigram มาช่วยจำแนกลำดับคำให้ใกล้เคียงกับประโยคต้นฉบับได้อย่างถูกต้อง ซึ่งแสดงให้เห็นถึงประสิทธิภาพของการประยุกต์ใช้ Trigram ร่วมกับโมเดลรู้จำลายมือในงานตรวจสอบการเขียนภาษาไทยของนักเรียนประถมได้อย่างมีนัยสำคัญ

ระยะที่ 3 พัฒนาระบบขั้นตอนวิธีตรวจสอบการเขียนประโยคภาษาไทย สำหรับนักเรียนชั้นประถมศึกษาตอนปลาย กรณีศึกษาโรงเรียนบ้านสันป่าไร่

ผลการวิจัยในระยะที่ 3 พัฒนาระบบขั้นตอนวิธีตรวจสอบการเขียนประโยคภาษาไทย สำหรับนักเรียนชั้นประถมศึกษาตอนปลาย กรณีศึกษาโรงเรียนบ้านสันป่าไร่ มีดังต่อไปนี้

1. กำหนดเกณฑ์คุณภาพการเขียนประโยคกำหนดเกณฑ์การประเมินคุณภาพการตรวจสอบการเขียน ดังนี้

1.1 ค่าความแม่นยำ (Accuracy) ต้องไม่น้อยกว่า 95%

1.2 ค่าความแม่นยำเชิงบวก (Precision) ค่าการเรียกคืน (Recall) และค่าความถูกต้องเชิงกลมกลืน (F1-score) ของทุกคำต้องมีค่าไม่น้อยกว่า 0.95

1.3 ค่าการสูญเสีย (Loss) อยู่ในระดับต่ำ และมีแนวโน้มลดลงระหว่างการฝึกสอนโมเดล

2. พัฒนาตัวแบบตรวจสอบคุณภาพการเขียน

ในการพัฒนาตัวแบบใช้โครงข่ายประสาทเทียม (Convolutional Neural Network: CNN) สำหรับจำแนกคำจากภาพลายมือ ร่วมกับโมเดล Trigram เพื่อตรวจสอบความถูกต้องของลำดับคำในประโยค โดยใช้ข้อมูลภาพคำที่แยกออกจากประโยคมาฝึกสอนโมเดลเป็นจำนวน 10 รอบการฝึก (Epoch) พร้อมการตรวจสอบด้วยชุดข้อมูล Validation

3. ทดสอบและปรับปรุงตัวแบบ

ตารางที่ 2 การเปรียบเทียบผลการทดสอบระหว่างประโยคที่ 1 และประโยคที่ 2

รายการประเมิน	ประโยคที่ 1	ประโยคที่ 2	สรุป
Accuracy (Train/Val)	1.000	1.000	ทั้งสองประโยคทำได้ แม่นยำเต็ม 100%
Loss (ต่ำกว่า 10^{-4})	ต่ำและลดลงต่อเนื่อง	ต่ำแต่ Val Loss มีแนวโน้มเพิ่ม	ทั้งคู่ Loss ต่ำมาก
Precision/Recall/F1-score	1.0	1.0	แม่นยำสูงเท่ากัน
Confusion Matrix	ถูกต้องทุกคำ	ถูกต้องทุกคำ	ไม่เกิดการทำนายผิด

ตารางที่ 2 (ต่อ)

รายการประเมิน	ประโยคที่ 1				ประโยคที่ 2				สรุป
แนวโน้มความเสถียร ของโมเดล	สูงมาก				สูงมาก				โมเดลทำงานได้ สม่ำเสมอทั้งสอง ประโยค

	precision	recall	f1-score	support
kadson	1.0	1.0	1.0	32.0
ken	1.0	1.0	1.0	36.0
kou	1.0	1.0	1.0	40.0
lae	1.0	1.0	1.0	38.0
pen	1.0	1.0	1.0	37.0
tas	1.0	1.0	1.0	41.0
tee	1.0	1.0	1.0	35.0
ton	1.0	1.0	1.0	40.0
tong	1.0	1.0	1.0	38.0
tuk	1.0	1.0	1.0	39.0
accuracy	1.0	1.0	1.0	1.0
macro avg	1.0	1.0	1.0	376.0
weighted avg	1.0	1.0	1.0	376.0

	precision	recall	f1-score	support
autid	1.0	1.0	1.0	36.0
catha	1.0	1.0	1.0	39.0
hi	1.0	1.0	1.0	37.0
kuson	1.0	1.0	1.0	43.0
laee	1.0	1.0	1.0	37.0
pean	1.0	1.0	1.0	37.0
pong	1.0	1.0	1.0	41.0
putpee	1.0	1.0	1.0	40.0
suad	1.0	1.0	1.0	37.0
tam	1.0	1.0	1.0	58.0
accuracy	1.0	1.0	1.0	1.0
macro avg	1.0	1.0	1.0	405.0
weighted avg	1.0	1.0	1.0	405.0

ภาพที่ 4 ภาพตารางผลการวัดประสิทธิภาพโมเดลตรวจสอบคำลายมือประโยคที่ 1 และประโยคที่ 2

ระยะที่ 4 นำเสนอแนวทางในการใช้งานการตรวจสอบการเขียนภาษาไทยด้วยการประมวลผลภาษาธรรมชาติ สำหรับนักเรียนชั้นประถมศึกษาตอนปลาย กรณีศึกษาโรงเรียนบ้านสันป่าไร่

ผลการวิจัยในระยะที่ 4 นี้ ตามวัตถุประสงค์ของการวิจัยข้อที่ 4 เป็นการนำเสนอแนวทางในการใช้งานการตรวจสอบการเขียนภาษาไทยด้วยการประมวลผลภาษาธรรมชาติ เพื่อรับรองการพัฒนาขั้นตอนวิธีการตรวจสอบการเขียนภาษาไทยด้วยการประมวลผลภาพ สำหรับนักเรียนชั้นประถมศึกษาตอนปลาย กรณีศึกษาโรงเรียนบ้านสันป่าไร่ โดยระยะนี้มีรายละเอียดดังต่อไปนี้

โดยผู้เชี่ยวชาญ ผู้ทรงคุณวุฒิทั้งหมด 10 ท่าน พบว่าแนวทางที่นำเสนอได้รับการประเมินจากผู้ทรงคุณวุฒิในระดับมากถึงมากที่สุด โดยมีค่าเฉลี่ย 4.4 และ S.D. 0.63 โดยหัวข้อที่ได้คะแนนสูงสุดคือความสอดคล้องกับบริบทโรงเรียน และความเหมาะสมโดยรวม ซึ่งบ่งชี้ถึงความเป็นไปได้ในการนำไปใช้ และผู้ทรงคุณวุฒิส่วนใหญ่เห็นพ้องว่าการประยุกต์ใช้กฎของซิปฟ์ (Zipf's Law) ในการวิเคราะห์และจัดกลุ่มคำศัพท์ 4 กลุ่มคำ คือ กลุ่มคำที่เขียนได้มากที่สุด ง่าย ปานกลาง และยาก มีความชัดเจนและสอดคล้องกับหลักภาษาศาสตร์ไทย ช่วยให้ครูเข้าใจระดับความรู้ของนักเรียนและปรับบทเรียนได้ตรงจุด นอกจากนี้ การพัฒนาระบบด้วย Convolutional Neural Network (CNN) ร่วมกับ Trigram Language Model และ OpenCV สำหรับ Text Segmentation ให้ประสิทธิภาพสูง เช่น ความแม่นยำ 100%, Loss ต่ำระดับ 10^{-5} , F1-score 1.0 ทุกคำ และ Confusion Matrix ถูกต้อง 100% ซึ่งแสดงถึงความเสถียรในการจัดการลายมือเด็ก

ระบบยังให้ feedback ทันที ช่วยเสริมแรงจูงใจนักเรียน โดยเฉพาะกลุ่มคำยาก และมีประโยชน์ต่อครูในการตรวจงานเร็ว ลดภาระและส่งเสริมความเท่าเทียมทางการศึกษาในบริบทโรงเรียนชนบทอย่างบ้านสันป่าไร่ ความทันสมัยของเทคโนโลยีที่ผสมผสานกับการสอนเป็นนวัตกรรมที่ตอบโจทย์ปัญหาจริงและช่วยลดช่องว่างดิจิทัล รวมถึงเกณฑ์ประเมินที่ชัดเจน (Accuracy $\geq 95\%$) ทำให้เป็นเกณฑ์อ้างอิงที่ดีสำหรับงานวิจัยต่อยอด การนำเสนอขั้นตอนวิธีการทำมีรายละเอียดครบถ้วนและสามารถใช้ในชั้นเรียนจริงได้ หากมีข้อจำกัดข้อมูลตัวอย่างจากนักเรียนเพียง 37 คนอาจไม่ครอบคลุมความหลากหลายของลายมือ ทำให้ระบบอาจไม่แข็งแกร่งในบริบทจริง นอกจากนี้การนำเสนอเนื้อหาชัดเจนแต่หากปรับเป็นการอธิบายด้วยภาพให้เห็นชัดเจนมากกว่าการบรรยายเป็นตัวหนังสือจะดีมาก เนื่องจากอาจช่วยให้เข้าใจง่ายขึ้น ผู้ทรงคุณวุฒิเสนอแนะให้ต่อยอดระบบโดยเพิ่มข้อมูลความหลากหลายของลายมือเพื่อยกระดับการสอนภาษาไทย และพัฒนาให้มีรายละเอียดครบถ้วนยิ่งขึ้นในอนาคต โดยคงจุดแข็งเดิมไว้เพื่อนำไปต่อยอดได้ต่อไป

อภิปรายผลการวิจัย (Research Discussion)

การวิจัยครั้งนี้เป็นการการพัฒนาขั้นตอนวิธีและนำเสนอแนวทางการใช้งานการตรวจสอบการเขียนภาษาไทยด้วยการประมวลผลภาพ สำหรับนักเรียนชั้นประถมศึกษาตอนปลาย กรณีศึกษาโรงเรียนบ้านสันป่าไร่ ผู้วิจัยมีประเด็นในการสรุปผลการวิจัยตามวัตถุประสงค์งานวิจัย มีรายละเอียดดังนี้

วัตถุประสงค์ที่ 1 การประยุกต์ใช้กฎของซิปฟ์ (Zipf's Law) ในการวิเคราะห์ความรู้คำศัพท์ภาษาไทย สำหรับนักเรียนชั้นประถมศึกษาตอนปลาย กรณีศึกษาโรงเรียนบ้านสันป่าไร่

ผลการวิจัยพบว่า ข้อมูลความรู้คำศัพท์ จำนวน 100 คำ จากนักเรียน 37 คน รวมทั้งสิ้น 3,700 คำ สอดคล้องกับกฎของซิปฟ์ โดยความถี่คำ (Frequency) ลดลงตามลำดับ (Rank) ในลักษณะเส้นตรงเมื่อพล็อตกราฟ $\log(\text{Rank})$ และ $\log(\text{Frequency})$ ซึ่งบ่งชี้ว่าคำศัพท์ที่ใช้บ่อยจะถูกเขียนได้มากกว่า กฎของซิปฟ์ (Zipf's Law) เป็นกฎทางสถิติที่อธิบายการกระจายตัวของความถี่ของคำในภาษาธรรมชาติ โดยระบุว่าความถี่ของคำใด ๆ ในชุดข้อมูลขนาดใหญ่จะแปรผกผันกับอันดับของคำนั้นในตารางความถี่ Zipf. (1949 : Abstract) คำที่ใช้บ่อยที่สุดจะมีความถี่สูงสุด และคำที่มีอันดับต่ำลงไปจะมีความถี่ลดลงตามลำดับในรูปแบบที่คาดเดาได้ การจัดกลุ่มคำศัพท์ภาษาไทยตามความถี่คำที่ได้จากกฎของซิปฟ์ (Zipf's Law) ด้วย Quartile Analysis โดยใช้ Python ร่วมกับไลบรารี Pandas แบ่งได้เป็น 4 กลุ่มคำ คือ กลุ่มคำที่เขียนได้มากที่สุด เช่น กระเจก คำดิน มีความถี่สูงกว่า 26.25 กลุ่มคำง่าย เช่น เนือง ผอง มีความถี่ 26.25 กลุ่มคำปานกลาง เช่น ภาคภูมิ กลิ้ง มีความถี่ 20.0 และกลุ่มคำยาก เช่น ถ้อยที่ถ้อยอาศัย ภูตผี มีความถี่ต่ำกว่า 12.50 ผลนี้สอดคล้องกับงานวิจัยของ Mandelbrot. (1953: 15-27) ที่พบว่ากฎซิปฟ์ใช้ได้กับภาษาธรรมชาติ รวมถึงภาษาไทย ซึ่งช่วยให้ครูสามารถปรับแผนการสอนให้เหมาะกับระดับความรู้ของนักเรียน อย่างไรก็ตามจุดอ่อนคือข้อมูลจากนักเรียนเพียง 37 คน อาจไม่ครอบคลุมประชากรทั้งหมด และคำบางคำ เช่น กรุงรัตน์โกสินทร์ มีความถี่ 0 ซึ่งอาจเกิดจากบริบทวัฒนธรรมที่ไม่คุ้นเคย

วัตถุประสงค์ 2 ประยุกต์ใช้รูปแบบทางภาษาในการจำแนกคำเพื่อตรวจสอบการเขียนภาษาไทย สำหรับนักเรียนชั้นประถมศึกษาตอนปลาย กรณีศึกษาโรงเรียนบ้านสันป่าไร่

ผลการวิจัยแสดงว่าการใช้ Text Segmentation ด้วย OpenCV สามารถแยกคำจากภาพลายมือจริงได้อย่างถูกต้อง และเมื่อนำเข้าสู่ Trigram Language Model สามารถจัดลำดับคำให้ตรงกับประโยคต้นแบบ โดยใช้เวลาเฉลี่ย 60-80 มิลลิวินาทีต่อคำ ซึ่งบ่งชี้ถึงประสิทธิภาพสูง ผลนี้สอดคล้องกับงานวิจัยของ Jurafsky & Martin. (2009: 141-145) ที่ระบุว่า Trigram มีประสิทธิภาพในการจำแนกภาษาธรรมชาติ โดยเฉพาะภาษาที่มีโครงสร้างยืดหยุ่นอย่างภาษาไทย จุดแข็งคือระบบสามารถจัดการกับลายมือที่ไม่แน่นอนได้ดี แต่จุดอ่อนคืออาจมีข้อจำกัดกับลายมือที่อ่านยากมาก ซึ่งอาจต้องปรับปรุงด้วยการเพิ่มข้อมูลฝึกสอนที่สำคัญคือช่วยลดภาระครูในการตรวจงานเขียน และส่งเสริมการเรียนรู้ภาษาไทยให้มีประสิทธิภาพมากขึ้น

วัตถุประสงค์ 3 พัฒนาระบบขั้นตอนวิธีตรวจสอบการเขียนประโยคภาษาไทย สำหรับนักเรียนชั้นประถมศึกษาตอนปลาย กรณีศึกษาโรงเรียนบ้านสันป่าไร่

ผลการวิจัยพบว่าขั้นตอนวิธีที่พัฒนาด้วย CNN ร่วมกับ Trigram ให้ความแม่นยำสูงสุด โดยในประโยคที่ 1 และ 2 มีค่า Accuracy 1.000 (100%) ตลอด 10 Epoch, Precision/Recall/F1-score 1.0 และ Loss ต่ำระดับ 10^{-5} Confusion Matrix แสดงการจำแนกถูกต้องทุกคำโดยไม่มีการผิดพลาด อย่างไรก็ตาม ในประโยคที่ 2 Validation Loss มีแนวโน้มเพิ่มขึ้นเล็กน้อย ซึ่งอาจบ่งชี้ถึง Overfitting เล็กน้อย ผลนี้สอดคล้องกับงานวิจัยของ LeCun et al. (1998: 253-257) ที่พบว่า CNN มีประสิทธิภาพสูงในการประมวลผลภาพ และเมื่อรวมกับ Language Model จะยกระดับการตรวจสอบข้อความ จุดแข็งคือระบบเสถียรและรวดเร็ว แต่จุดอ่อนคืออาจเกิด Overfitting หากข้อมูลฝึกน้อย ซึ่งสามารถแก้ไขด้วย Data Augmentation ที่สำคัญคือช่วยให้นักเรียนได้รับ feedback ทันที ส่งเสริมทักษะการเขียนภาษาไทยให้ดีขึ้น

วัตถุประสงค์ 4 นำเสนอแนวทางในการใช้งานการตรวจสอบการเขียนภาษาไทยด้วยการประมวลผลภาษาธรรมชาติ สำหรับนักเรียนชั้นประถมศึกษาตอนปลาย กรณีศึกษาโรงเรียนบ้านสันป่าไร่

ผลการวิจัยพบว่าแนวทางที่นำเสนอได้รับการประเมินจากผู้ทรงคุณวุฒิในระดับมากถึงมากที่สุด (ค่าเฉลี่ย 4.4, S.D. 0.63) โดยหัวข้อที่ได้คะแนนสูงสุดคือความสอดคล้องกับบริบทโรงเรียน และความเหมาะสมโดยรวม ซึ่งบ่งชี้ถึงความเป็นไปได้ในการนำไปใช้ ผลนี้สอดคล้องกับงานวิจัยของ Prensky. (2001: 15-18) ที่เน้นเทคโนโลยีเพื่อการศึกษาที่เหมาะสมกับบริบทท้องถิ่น จุดแข็งคือแนวทางชัดเจนและทันสมัย แต่จุดอ่อนคืออาจต้องปรับให้เหมาะกับโรงเรียนที่มีทรัพยากรจำกัด ที่สำคัญคือช่วยเชื่อมโยงเทคโนโลยีกับการเรียนการสอนภาษาไทยในโรงเรียนชนบท

ข้อเสนอแนะการวิจัย (Research Suggestions)

1. ข้อเสนอแนะในการนำผลการวิจัยไปใช้

- 1.1 ชุดข้อมูลภาพลายมือมาจากนักเรียนโรงเรียนเดียว อาจไม่ครอบคลุมลายมือทั้งหมด
- 1.2 ระบบยังไม่รองรับภาพข้อมูลลายมือที่เบลอ ตัวเขียนอ่านยาก หรือแสงน้อย ทำให้การทำนายคำเกิด error ได้

1.3 ควรบูรณาการระบบตรวจสอบนี้เข้ากับการเรียนการสอนในรายวิชาภาษาไทย เพื่อให้ นักเรียนได้ฝึกใช้และเห็นผลการตรวจสอบทันทีหลังการเขียน ซึ่งจะช่วยกระตุ้นให้นักเรียนปรับปรุงลายมือและ การสะกดคำอย่างต่อเนื่อง

1.4 ควรจัดทำฐานข้อมูลคำและประโยคให้ครอบคลุมเนื้อหาที่ใช้ในหลักสูตรระดับประถม ปลาย รวมถึงคำที่พบในชีวิตประจำวัน เพื่อให้ระบบตรวจสอบได้อย่างหลากหลายและตรงกับการใช้งานจริง

2. ข้อเสนอแนะเพื่อการทำวิจัยในครั้งต่อไป

2.1 ควรขยายการทดสอบระบบกับประโยคที่มีความซับซ้อนมากขึ้น หรือเป็นประโยค ที่นักเรียนแต่งขึ้นเอง เพื่อประเมินความยืดหยุ่นและความแม่นยำของระบบในสถานการณ์จริง

2.2 ศึกษาความเป็นไปได้ในการนำระบบไปใช้ผ่านแพลตฟอร์มออนไลน์หรือแอปพลิเคชัน เพื่อให้ครูและนักเรียนสามารถใช้งานได้ทุกที่ทุกเวลา

เอกสารอ้างอิง (References)

- ศราวุธ กองวิรัตน์. (2561). การจดจำตัวอักษรลายมือเขียนภาษาไทยโดยใช้โครงข่ายประสาทแบบคอนโวลูชัน (Thai handwriting character recognition using convolutional neural networks). [วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ, จุฬาลงกรณ์มหาวิทยาลัย]. (Chula ETD).
- ห้องสมุดวังสนามจันทร์. (2556, เมษายน). เรื่องของการเขียนภาษาไทยกับลายมือ. Sanamchandra Palace Library's Blog. สืบค้นเมื่อ 14 กันยายน 2568 จาก <https://snc.lib.su.ac.th/snclibblog?p=28243Amrhein>,
- C., & Clematide, S. (2018, July). Supervised OCR Error Detection and Correction Using Statistical and Neural Machine Translation Methods. *Journal for Language Technology and Computational Linguistics (JLCL)*. 33(1); 49-76.
- Jurafsky, D., & Martin, J. H. (2009). *Speech and language processing : An introduction to natural language processing, computational linguistics, and speech recognition* (2nd ed.). Prentice Hall.
- LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324. <https://doi.org/10.1109/5.726791>
- Piantadosi, S. T. (2014). Zipf's word frequency law in natural language : A critical review and future directions. *Psychonomic Bulletin & Review*, 21(5), 1112–1130.
- Prensky, M. (2001). *Fun, play and games : What makes games engaging? Digital game-based learning*. New York : McGraw-Hill.
- Rani, N. S., et al. (2018, July). Deformed Character Recognition Using Convolutional Neural Networks. *International Journal of Engineering & Technology*. 7(3): 1599–1604.

Wulff, P., et al. (2023, January). Enhancing writing analytics in science education research with Machine Learning and Natural Language Processing-Formative assessment of science and non-science preservice teachers' written reflections. *Frontiers in Education*. 7 : 1–18.