

บทที่ 4

ผลการทดลอง

โครงการปริญญาโทนำเสนองานตรวจสอบความถูกต้องข้อความภาษาไทย ที่เกิดจากการป้อนข้อมูลที่อาจมีความผิดพลาดของคำ การสะกดผิด ซึ่งในบทนี้จะเป็นผลการทดลองในการประเมินประสิทธิภาพของผลลัพธ์ที่ได้จากโปรแกรมการตรวจสอบความถูกต้องข้อความภาษาไทย โดยแบ่งการทดสอบออกเป็น 3 ส่วน คือ

- 1.ผลการทดสอบการตัดคำ
- 2.ผลการตรวจสอบคำผิด
- 3.ผลการทดสอบแก้ไขคำผิด

4.1 ผลการทดลอง

4.1.1 ผลการทดสอบการตัดคำ

ได้ทดสอบประสิทธิภาพของการตัดคำ โดยการนำผลลัพธ์การตัดคำของ Deep Cut และ Newmm มาเปรียบเทียบกับผลเฉลยที่ได้จากการให้คนตรวจสอบ จากนั้นได้ทำการประเมินประสิทธิภาพในการตัดคำโดยเลือกใช้วิธีเทียบจาก Word Level โดย Benchmarks จากการวัดประสิทธิภาพของการตัดคำจากชุดทดสอบทั้งหมดที่มี 5,377 ประโยค 38,579 คำ โดยหาจากค่า Precision, Recall, F1 ได้ผลลัพธ์ดังตารางต่อไปนี้

ตารางที่ 4.1 ผลการคำนวณโดยหาค่า Accuracy จากชุดทดสอบทั้งหมด

ตัวอย่างประโยค	จำนวนคำที่โปรแกรมตัด	TP	TN	FP	FN	ความถูกต้อง
Newmm	36,269	23,358	-63,66	9,691	11,896	$23,358+(-63,66)/(23,358+(-63,66)+11,896)\times 100 = 44\%$
Deepcut	38,237	30,750	-788	4,113	4,504	$30,750+(-788)/(30,750+(-788)+45,04)\times 100 = 77.7\%$
ผลเฉลยจากที่ให้คนตรวจสอบ	38,579	38,579	0	0	0	$38,579+0/(38,579+0+0)\times 100 = 100\%$

ตารางที่ 4.2 ผลลัพธ์จากค่า Precision, Recall, F1 จากชุดข้อมูลทั้งหมด

Precision	Recall	F1 Score
0.707	0.663	0.684
0.882	0.872	0.877
1	1	1

ซึ่งจากการทดสอบข้างต้น พบว่าการตัดคำโดย Deep Cut ได้มีความใกล้เคียงกับผลเฉลยที่ได้จากการให้คนตรวจสอบมากที่สุด

4.1.2 ผลการตรวจสอบคำผิด

จากการวัดประสิทธิภาพของการตรวจสอบคำผิด โดยนำประโยคที่อาจจะมีคำผิดปะปนอยู่ และนำผลลัพธ์ของการตัดคำมาเปรียบเทียบกับผลเฉลย โดยทดสอบข้อมูลจากการตัดคำของ Deep Cut โดยใช้ Library ของ Pythainlp Spell Checker มาทำการตรวจสอบคำผิด

ตารางที่ 4.3 ตัวอย่างการคิดประสิทธิภาพของการตรวจสอบคำ

ประโยคที่มีคำผิด	ตำแหน่งคำที่ผิด	จำนวนคำ	จำนวนคำที่ผิด
ไข่เจียวหมูสุบ	ไข่เจียวหมู สุ บ	4	1
ควมมีโสดไม่ใช่เนื้อคู่	ควม มีโสดไม่ใช่เนื้อคู่	6	1
ทกหมื่นแปดพันสองร้อยยี่สิบ บาทเจ็ดสิบห้าสตางค์	ทกหมื่นแปดพันสองร้อยยี่สิบ บาทเจ็ดสิบห้าสตางค์	13	0
เครื่องสแกนนิ้ว	เครื่องสแกนนิ้ว	2	0
เพราะลายมือเราเปลี่ยนตาม อารมาคะ	เพราะลายมือเราเปลี่ยนตาม อารมา คะ	8	1
ความทรงจำที่ไม่มีวันจางหาย	ความทรงจำที่ไม่มีวัน จาง หาย	6	1
จำนวนทั้งหมด		39	4

จากประโยคข้างต้นมีจำนวนคำทั้งหมด 4 คำ จากคำทั้งหมด 39 คำ โดยแบ่งการทดสอบคำผิดตามวิธีการตัดคำ โดยใช้ Library ของ Pythainlp Spell Checker โดยการวัดประสิทธิภาพของการตรวจสอบคำผิดโดยใช้วิธี Accuracy จะได้ว่า

$$\text{ความถูกต้อง} = \frac{\text{True Negative} + \text{True Positive}}{\text{True Negative} + \text{True Positive} + \text{False Negative} + \text{False Positive}} \quad (16)$$

ตารางที่ 4.4 ผลการทดสอบ

ผลการทดลอง				
TP	TN	FP	FN	ความถูกต้อง
3	34	1	1	$\frac{3+34}{3+34+1+1} \times 100 = 94.87\%$

โดยประสิทธิภาพของการตรวจสอบคำผิด ซึ่งจากการทดสอบข้างต้น พบว่าการตรวจสอบคำผิดใช้ Library ของ Pythainlp Spell Checker ให้ความถูกต้องของการตรวจสอบคำผิดได้ดีตรงกับผลเฉลยเช่นกัน

สรุปการหาประสิทธิภาพตรวจสอบคำผิดจากชุดทดสอบทั้งหมด 5,377 ข้อความ โดยมีข้อความที่ผิดพลาด 1,385 ข้อความ เป็นจำนวนคำ 7,689 คำ โดยที่ Library ของ Pythainlp Spell Checker ตรวจสอบคำผิดได้ 604 คำ จากจำนวนคำที่ผิดพลาดทั้งหมด 1,962 คำ คิดเป็นอัตราความผิดพลาด 0.30 หรือ 30%

4.1.3 ผลการแก้ไขคำผิด

ผู้จัดทำได้ทดสอบประสิทธิภาพของการแก้ไขคำผิด โดยการนำผลลัพธ์ที่ทำนายผลมาจากการทดสอบโดยนำมาหาค่าอัตราความผิดพลาดของคำ โดยเปรียบเทียบกับชุดข้อมูลที่เป็นผลเฉลย จากนั้นหาค่าเฉลี่ยของอัตราความผิดพลาดของคำทั้งหมด

ตารางที่ 4.5 ตัวอย่างผลการทดลองการหาค่าอัตราความผิดพลาดของคำ

ประโยคตัดคำที่เป็นผลเฉลย	ประโยคตัดคำที่เกิดจากการทำนาย	ค่าอัตราความผิดพลาดของคำ
ไข่ เจียว หมู สับ	ไข่ เจียว หมู สุก	$(1+0+0)/4 = 0.25$
ความ โสด ไม่ ไข่ เนื้อคู่	ความ โสด ไม่ ไข่ เนื้อคู่	$(0+0+0)/5 = 0$
หก หมิ่น แปด พัน สอง ร้อย ยี่สิบ บาท เจ็ด สิบ ห้า สตางค์	หก หมิ่น แปด พัน สอง ร้อย ยี่สิบ บาท เจ็ด สิบ ห้า สตางค์	$(0+0+0)/13 = 0$
เครื่อง สแกนนิ้ว	เครื่อง สแกนนิ้ว	$(0+0+0)/2 = 0$
เพราะ ลาย มือ เรา เปลี่ยน ตาม อารมณ์ ค่ะ	เพราะ ลาย มือ เรา เปลี่ยน ตาม อารมณ์ ค่ะ	$(0+0+0)/8 = 0$
ความทรงจำ ที่ ไม่มี วัน จางหาย	ความทรงจำ ที่ ไม่มี วัน จางหาย	$(0+1+0)/6=0.167$

ตารางที่ 4.5 ตัวอย่างผลการทดลองการหาค่าอัตราความผิดพลาดของคำ (ต่อ)

ประโยคตัดคำที่เป็นผลเฉลย	ประโยคตัดคำที่เกิดจากการทำนาย	ค่าอัตราความผิดพลาดของคำ
		$(0.25+0+0+0+0+0.16)/6$ = 0.065 หรือ 6.5%

สรุปการหาประสิทธิภาพแก้ไขคำผิดจากชุดทดสอบทั้งหมด 5,377 ข้อความ โดยมีจำนวน 38,579 คำพบคำผิดพลาด 1,083 คำ จากคำทั้งหมด คิดเป็นอัตราความผิดพลาด 0.028 หรือ 2.8%