

การพัฒนาโปรแกรมตรวจให้คะแนนข้อสอบอัตนัย แบบอัตโนมัติด้วยวิธีการเทียบคำ

Development of automatic subjective score test with word matching

นายเศรษฐชัย ใจอีก¹, ผู้ช่วยศาสตราจารย์.ดร.สุศักดิ์ มั่งสิงห์²
Seatachai Jaihuek¹, Asst.Prof. Surasak Mungsing, Ph.D.²



บทคัดย่อ

ข้อสอบอัตนัยเป็นวิธีวัดผลทางการศึกษาที่มีประสิทธิภาพ เพราะผู้สอบต้องเขียนบรรยายคำตอบ ผู้สอบต้องอาศัยทักษะทางด้านความรู้ ความจำ และการวิเคราะห์ แต่ปัจจุบันโปรแกรมที่ใช้สำหรับอำนวยความสะดวกมีจำนวนจำกัด และไม่ตรงต่อความต้องการของผู้ใช้งาน ผู้วิจัยจึงมีแนวทางในการพัฒนาโปรแกรมตรวจให้คะแนนข้อสอบอัตนัยแบบอัตโนมัติด้วยวิธีการเทียบคำ ซึ่งจะจัดทำเป็นต้นแบบของหน่วยงาน เพื่อใช้พัฒนาศักยภาพการเรียนการสอนโดยเลือกปรับแต่งจากเทคนิค e-rater ที่ใช้งานกับภาษาอังกฤษให้สามารถใช้งานร่วมกับภาษาไทยที่ใช้การเปรียบเทียบคำสำคัญจากพจนานุกรม ทั้งนี้ผู้วิจัยได้พัฒนาระบบการให้คะแนนที่สามารถคาดคะเนความถูกต้องของคำศัพท์ และจำกัดการสืบค้นคำสำคัญเพียงคำถาม/คำตอบเฉลยเพื่อลดเวลาการค้นหา นอกจากนั้นในส่วนการประเมินผลคะแนนได้ใช้เทคนิค rubric assessment เพื่อลดข้อผิดพลาดในการให้คะแนน ผลการทดสอบพบว่า ค่าเฉลี่ยของค่าความถูกต้อง คำร้อยละ 74.40, ค่าความครบถ้วนมีค่าร้อยละ 37.90 และวัดประสิทธิภาพมีค่าร้อยละ 54.50 ในส่วนความพึงพอใจจากผู้เชี่ยวชาญพบว่า ระดับพึงพอใจมาก ($\bar{X} = 4.11$, S.D. = 0.48)

คำสำคัญ: ตัดคำภาษาไทย, การแบ่งแยกคำ, ข้อสอบอัตนัย, ตรวจข้อสอบ

¹ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยศรีปทุม

² คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยศรีปทุม

* Corresponding author. Tel. 08-9755-5257 E-mail: seatachai@gmail.com



Abstract

A subjective test is a powerful tool for measuring educational performance. Students need skills in knowledge, memory and analysis. At present, the facilities are limited. And do not meet the needs of users.

The researcher developed the concept of automatic assessment of subjective test items by using vocabulary matching method for develop the potential of teaching. The original system was developed by E-rater. It is used in English only, to be compatible with Thai language. The researcher has developed a system to predict the accuracy of vocabulary and Limit search queries to just questions / answers. Solution to reduce search time. In addition, in scoring rubric assessments technical were used to reduce errors in scoring. The results showed that average precision = 74.40%, average recall = 37.90 and average F-Measure = 54.50%. In terms of professional satisfaction, the level of satisfaction was very high. ($\bar{X} = 4.11$, S.D. = 0.48)

Keywords: Thai words, word discrimination, subjective exams, exams



ความสำคัญของปัญหา

การจัดการเรียนการสอนผ่านระบบออนไลน์จะช่วยให้นักเรียนสามารถเข้าถึงเนื้อหาการเรียนได้ตลอดเวลา และรองรับจำนวนผู้เรียนจำนวนมากได้ในเวลาเดียวกัน นอกจากนั้นการเรียนการสอนผ่านโปรแกรมออนไลน์ยังสามารถประยุกต์ร่วมกับการวัดผลทางการศึกษา โดยการจัดทำโปรแกรมข้อสอบออนไลน์ในรูปแบบข้อสอบอัตนัยและข้อสอบปรนัยได้ ผลการสอบจะอยู่ในรูปแบบข้อมูลดิจิทัลที่คอมพิวเตอร์สามารถนำไปใช้ประมวลผลคะแนนหรือผลิตข้อมูลสารสนเทศต่อไปได้

การสอบซึ่งจัดเป็นวิธีการวัดผลสัมฤทธิ์ทางการศึกษา โดยทั่วไปโปรแกรมการสอบออนไลน์จะจัดทำรูปแบบการสอบออกเป็น 2 รูปแบบคือ การสอบด้วยข้อสอบปรนัย (objective test) โดยโปรแกรมจัดเตรียมตัวเลือกคำตอบที่ถูกต้องและตัวเลือกลวงไว้ให้เลือก เมื่อทำการประเมินผลคะแนน ระบบจะรวบรวมจำนวนคำตอบถูกต้องเพื่อแสดงผลคะแนน การสอบอีกรูปแบบหนึ่งคือ ข้อสอบอัตนัย (subjective test) เป็นการตอบคำถามแบบเขียนพรรณนา ไม่มีตัวเลือกคำตอบให้ (Uthumporn Charmonman, 1987) ในส่วนการตรวจให้คะแนน ผู้ตรวจต้องใช้การอ่านและพิจารณาคำตอบ เพื่อประเมินคะแนนที่เหมาะสมในแต่ละข้อ ความคลาดเคลื่อนของการประเมินผลคะแนนข้อสอบอัตนัยเกิดจากเจตคติ อารมณ์และความเชี่ยวชาญของผู้ตรวจ จึงมีนักวิจัยทางคอมพิวเตอร์พยายามที่จะสร้างโปรแกรมตรวจข้อสอบอัตนัยแบบอัตโนมัติ แต่เนื่องด้วยกระบวนการภาษาที่ซับซ้อน และการตอบในรูปแบบประโยคพรรณนาจัดเป็นข้อมูลรูปแบบไม่เป็นโครงสร้าง (unstructured data) จึงทำให้โปรแกรมตรวจข้อสอบอัตนัยมีความยากต่อการพัฒนาเป็นโปรแกรมการตรวจแบบอัตโนมัติ

เทคนิคการตรวจข้อสอบอัตนัยแบบอัตโนมัติโดยทั่วไปจะใช้โปรแกรมตัดคำศัพท์ออกจากประโยค ปัจจุบันมีงานวิจัยการตัดคำศัพท์ออกจากประโยคข้อความ 2 กลุ่มคือ การใช้กฎ (rule base approach) หรือ อัลกอริทึม (algorithm approach) และการใช้พจนานุกรม (dictionary base approach) หรือ คลังข้อความ (corpus base approach) เมื่อนำประโยคข้อความมาผ่านเข้าสู่โปรแกรมการตัดคำ และจัดเก็บจะทำให้รูปแบบข้อมูลเปลี่ยนไปเป็นข้อมูลแบบมีโครงสร้าง (structured data) ซึ่งเหมาะสมต่อการประยุกต์ใช้งานกับการประมวล

ภาษาธรรมชาติ (natural language processing) อื่นๆ เช่น การสรุปใจความสำคัญ (text summarization) การแปลภาษา (machine translation) การค้นคืนสารสนเทศ (information retrieval) เป็นต้น

งานวิจัยนี้เป็นส่วนหนึ่งของโปรแกรมการตรวจข้อสอบอัตนัยภาษาไทยแบบอัตโนมัติด้วยออนไลน์ ผู้วิจัยได้เลือกใช้โปรแกรมข้อสอบอัตนัยแบบออนไลน์ของรายวิชา GEN1102 สารสนเทศในชีวิตประจำวัน ร่วมกับเทคนิคการตัดคำศัพท์ที่สำคัญจากประโยคคำตอบ เนื่องจากเป็นวิชาที่มีความหลากหลายของภาษาไทย ทั้งเขียนทับศัพท์และมีการเขียนผิด/ถูกจำนวนมาก โดยปรับแต่งจากเทคนิค e-rater ที่ค้นหาเปรียบเทียบ คำศัพท์ทั้งฐานข้อมูลแบบเป็นการค้นหาเฉพาะคำศัพท์จากคำตอบในแต่ละข้อเพื่อลดเวลาการประเมินผล และเพิ่มเติมเทคนิคการสะกดคำศัพท์ด้วยการตรวจหาความคล้ายของคำศัพท์จากอักขระที่ประกอบคำศัพท์ โดยพัฒนาโปรแกรมทั้งหมดด้วยภาษา PHP และใช้ฐานข้อมูล MySQL



วัตถุประสงค์การวิจัย

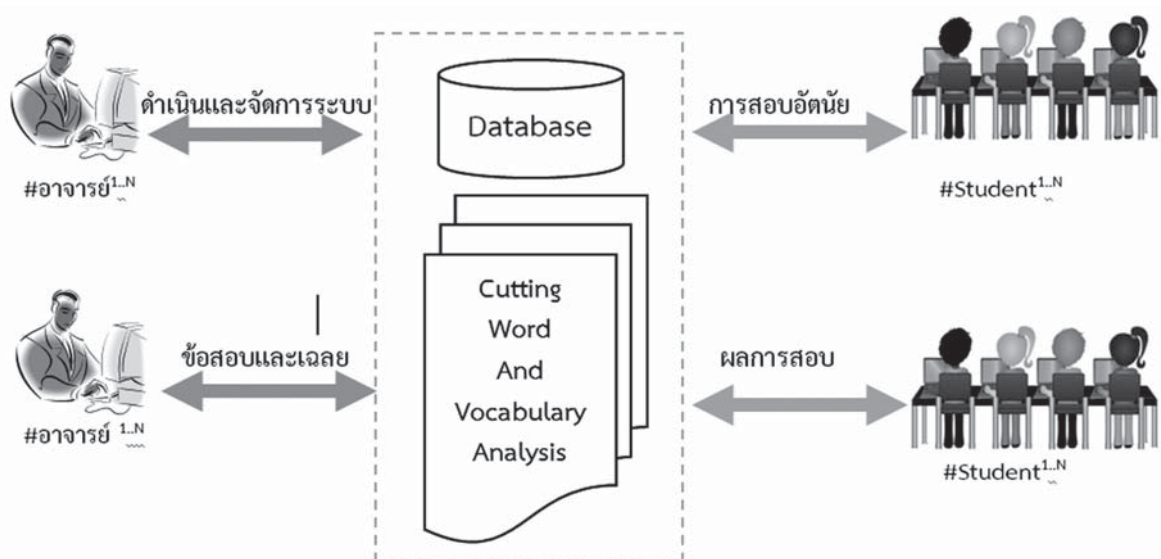
1. พัฒนาโปรแกรมให้คะแนนข้อสอบอัตนัยแบบออนไลน์ด้วยเทคนิคการตัดคำและสะกดคำ
2. ออกแบบและพัฒนาขั้นตอนวิธีสำหรับการประเมินคำตอบของข้อสอบอัตนัยภาษาไทย



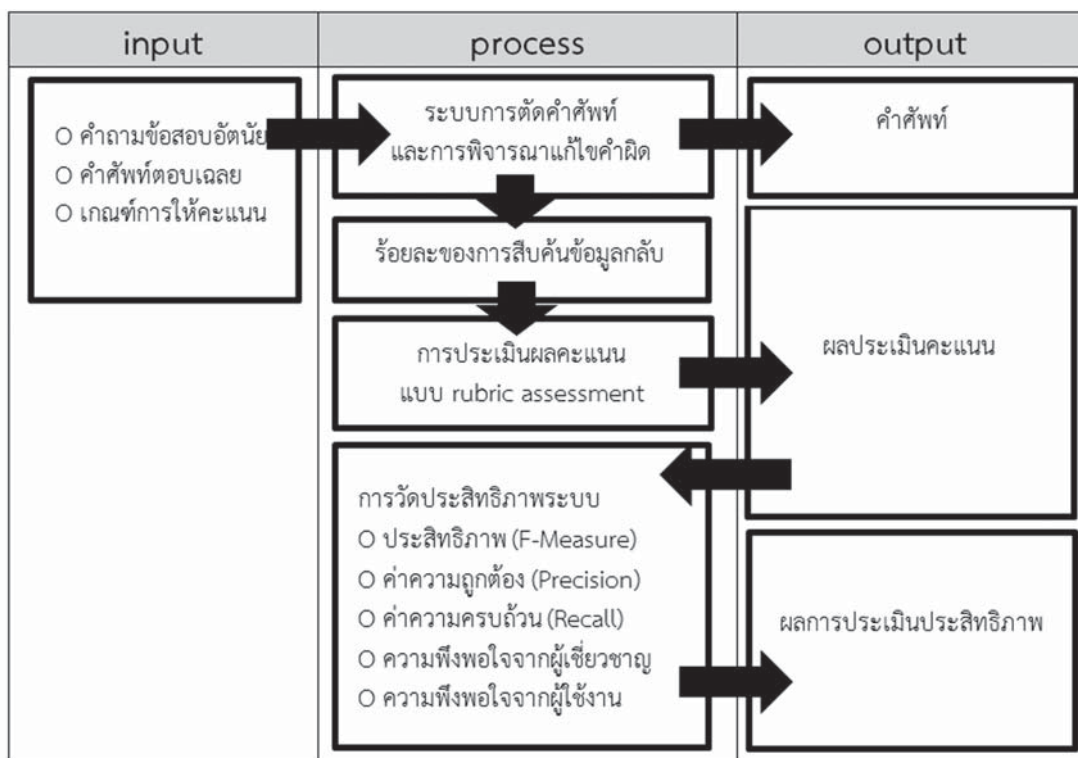
วิธีดำเนินการวิจัย

การดำเนินการประกอบด้วยการพัฒนาระบบการสอบอัตนัยและตรวจข้อสอบด้วยโปรแกรมคอมพิวเตอร์บนเครือข่ายอินเทอร์เน็ต ดังต่อไปนี้

ขั้นตอนที่ 1 เป็นขั้นตอนของการวางแผนการทำงาน เป็นส่วนในการวิเคราะห์ภาพรวมการทำงานทุกภาคส่วน โดยจะปรากฏเป็น 3 ด้านคือ 1. ผู้ทำข้อสอบ 2. ผู้ตรวจข้อสอบ 3. กระบวนการตรวจ นอกจากนั้นต้องพิจารณาถึงเทคโนโลยีและข้อจำกัดของทรัพยากรที่มีอยู่ดังรูปประกอบที่ 1



รูปที่ 1 กรอบแนวคิดกระบวนการพัฒนาระบบตรวจข้อสอบอัตนัยแบบออนไลน์



รูปที่ 2 กรอบแนวคิดกระบวนการพัฒนาโปรแกรมตรวจข้อสอบอัตนัยแบบเทียบคำศัพท์

ขั้นตอนที่ 2 เป็นขั้นตอนของการออกแบบโปรแกรม เป็นการออกแบบการทำงานประกอบด้วยแบ่งการออกแบบ 3 งาน คือ การออกแบบฐานข้อมูลสำหรับคำตอบ, การออกแบบโปรแกรมการตัดคำและพยากรณ์คำ และการออกแบบการติดต่อประสานหน้าจอ

การออกแบบ 1. การออกแบบฐานข้อมูลสำหรับคำตอบ

ผู้วิจัยได้พัฒนาฐานข้อมูลต้นแบบสำหรับพัฒนาโปรแกรมการตรวจข้อสอบออนไลน์แบบเทียบคำศัพท์ซึ่งประกอบไปด้วย 3 ฐานข้อมูลดังนี้

ตารางที่ 1 ฐานข้อมูลโครงสร้างตารางคำถามอัตนัยและคำตอบ

ลำดับ	รายการ	คำอธิบาย	ชนิด
1	question_id	เลขอ้างอิงคำถาม	int
2	question_detail	รายละเอียดคำถาม	character
3	question_answer	รายละเอียดคำตอบ	character
4	question_status	สถานะคำถาม	int

ตารางที่ 2 ฐานข้อมูลโครงสร้างตารางพจนานุกรมคำศัพท์เฉลย

ลำดับ	รายการ	คำอธิบาย	ชนิด
1	dictionary_id	เลขอ้างอิงคำตอบ	int
2	question_id	เลขอ้างอิงคำถาม	int
3	dictionary_keyword	คำศัพท์	character
4	dictionary_rate	จำนวนความถี่ของคำตอบ	int

ตารางที่ 3 ฐานข้อมูลโครงสร้างตารางผลการสอบ

ลำดับ	รายการ	คำอธิบาย	ชนิด
1	score_id	ลำดับอ้างอิงการตรวจ	int
2	student_code	เลขที่ผู้สอบ	int
3	score_keyword	จำนวนคำที่ค้นพบ	int
4	score_rubric	ค่าประเมิน rubric assessment	int

การออกแบบ 2. การออกแบบโปรแกรมการตัดคำและพยากรณ์คำ

กระบวนการแบ่งออกได้เป็น 2 ขั้นตอนคือ กระบวนการตัดคำศัพท์ออกมาจากคำตอบของผู้สอบ และกระบวนการพิจารณาคำที่เขียนผิด ทั้งนี้กระบวนการตัดคำและเทียบคำศัพท์จะทำการตัดคำศัพท์ที่ซ้ำและไม่ถูกนำมาคิดในการประเมินค้นหาคำ โดยนำเสนอในรูปแบบชุดโค้ดดังนี้

ตารางที่ 4 ชุดโค้ดการตัดคำศัพท์ออกจากประโยคคำตอบ และการแก้ไขคำศัพท์ที่เขียนผิด

ลำดับ	ชนิด
1	Variable Answer of Student="", Confirm="", Point=0, Long=0, Half=0, Confirm=0
2	Variable Array Answer[0...n]="", Array Box[0...n]= ""
3	Input Answer of Student (limit 200 character)
4	Compare Array Answer[0...n] = Answer of Student by character;
5	While Record set <= EOF In Database
6	If Answer[0...n] = Answer by Expert in record set Then
7	Array Box[Point] = Answer by Expert in record set
8	Delete value Answer[0...n] amount Answer by Expert in record set
9	Point=Point+1
10	End if
11	End while
	Algorithm : Check and Repair word
12	If Array Answer[0...n] != null Than

ลำดับ	ชนิด
13	While data set <= EOF In Database
14	Compute Long =data set of Length
15	Compute Half = Long/2
16	Compare Array Data[0...n] = data set by character;
17	Compute Key1 = character of Data[1] + character of second in Data[2]
18	Compute Key2 = character of Data[Half] + character of Data[Half+1]
19	Compute Key3 = character of Data[Long-1] + character of Data[Long]
20	If Array Answer[0...n]=Key1 And Array Answer[0...n]=Key2 And point between Key1 to Key 2 <= Half +1 Then
21	Confirm= Confirm+1
22	End if
23	If Array Answer[0...n]=Key2 And Array Answer[0...n]=Key3 And point between Key2 to Key 3 <= Half +1 Then
24	Confirm= Confirm+1
25	End if
26	If Array Answer[0...n]=Key1 And Array Answer[0...n]=Key3 And point between Key1 to Key 3 <= Long +1 Then
27	Confirm= Confirm+1
28	End if
29	FOR i=1 to Point
30	If Confirm>=2 And Array Box[Point] <> data set Then
31	Box[Point+1]= data set
32	Array Answer[Key1 To Key2] =""
33	Array Answer[Key2 To Key3] =""
34	Array Answer[Key1 To Key3] =""
35	End if
36	Next
37	Confirm=0
38	End if
39	Compute sent value Array Box[0...n] to Next Process

การออกแบบ 3. การออกแบบการติดต่อประสานหน้าจอ

การออกแบบและจัดทำหน้าเว็บเพื่อเป็นเครื่องมือสำหรับใช้ในการทดสอบหลักการและเหตุผลของงานวิจัยประกอบไปด้วยหน้าตัวอย่างดังต่อไปนี้

BACK | Admin : 000000001 อาจารย์เศรษฐชัย ใจฮึก : สำนักวิชาคอมพิวเตอร์และเทคโนโลยีสารสนเทศ

ID	รหัสวิชา	ชื่อวิชา	ดำเนินการ		
คำถาม			search	insert	clean
1	GEN1102	สารสนเทศในชีวิตประจำวัน	edit	subject	object
2	SN1002	สังคมวิวัฒน์	edit	subject	object
3	GD2211	ธุรกิจเพื่องานกราฟิกดีไซน์	edit	subject	object
4	GD2308	ความคิดสร้างสรรค์	edit	subject	object

รูปที่ 3 หน้าจอบริหารจัดการข้อสอบตามรายวิชา

BACK | การกรอกข้อสอบ #1 : อาจารย์เศรษฐชัย ใจฮึก : สำนักวิชาคอมพิวเตอร์และเทคโนโลยีสารสนเทศ

ID	คำถาม	คำสำคัญ	กำหนดร้อยละระดับต่ำ			ดำเนินการ		
			ผิด	กลาง	ถูก			
คำถาม			0	33	66	search	insert	clean
1	จงยกตัวอย่างอุปกรณ์นำเข้า (อย่างน้อย 4 อุปกรณ์)	อุปกรณ์นำเข้า	0	33	60	edit	answer	delete
2	องค์ประกอบของระบบคอมพิวเตอร์มีอะไรบ้าง	ระบบคอมพิวเตอร์	0	33	66	edit	answer	delete
3	หน่วยประมวลผลกลางคืออะไร	หน่วยประมวลผลกลาง	0	33	66	edit	answer	delete
4	ประเภทของซอฟต์แวร์มีอะไรบ้าง	ซอฟต์แวร์	0	33	66	edit	answer	delete
5	ระบบปฏิบัติการคืออะไร	ระบบปฏิบัติการ	0	33	59	edit	answer	delete
11	จงยกตัวอย่างซอฟต์แวร์ระบบ (จำนวน 3 ระบบปฏิบัติการ)	ซอฟต์แวร์ระบบ	0	33	60	edit	answer	delete
10	xxxx	xxxx	0	33	66	edit	answer	delete

รูปที่ 4 หน้าจอสำหรับสร้างคำถามอัตนัยและคำตอบ

BACK | #1 รหัสนักศึกษา 357010059 นายเศรษฐชัย ใจฮึก : สำนักวิชาคอมพิวเตอร์และเทคโนโลยีสารสนเทศ

รายวิชา	รวมคะแนน 9 / 21 คะแนน	
GD2211 : ธุรกิจเพื่องานกราฟิกดีไซน์	ข้อที่ 1	จงยกตัวอย่างอุปกรณ์นำเข้า (อย่างน้อย 4 อุปกรณ์)
GD2308 : ความคิดสร้างสรรค์		เมาส์ คีย์บอร์ด โมด็ม 60 ส่งคำตอบ [2] ถูกต้อง
GEN1102 : สารสนเทศในชีวิตประจำวัน	ข้อที่ 2	องค์ประกอบของระบบคอมพิวเตอร์มีอะไรบ้าง
SN1002 : สังคมวิวัฒน์		เป็นส่วนหนึ่งของบุคลากร ฮาร์ดแวร์ ซอฟต์แวร์ ข้อมูล เจ้าหน้าที่ วิธีการ บุคลากร 100 ส่งคำตอบ [2] ถูกต้อง
	ข้อที่ 3	หน่วยประมวลผลกลางคืออะไร
		คอนเตนเซอร์ 0 ส่งคำตอบ [0] ผิด
	ข้อที่ 4	ประเภทของซอฟต์แวร์มีอะไรบ้าง
		ซอฟต์แวร์ระบบและซอฟต์แวร์ประยุกต์ 150 ส่งคำตอบ [2] ถูกต้อง
	ข้อที่ 5	ระบบปฏิบัติการคืออะไร
		ซอฟต์แวร์ควบคุมฮาร์ดแวร์ 44 ส่งคำตอบ [1] ปานกลาง
	ข้อที่ 6	จงยกตัวอย่างซอฟต์แวร์ระบบ (จำนวน 3 ระบบปฏิบัติการ)
		แอนดรอยด์ วินโดวส์ ไอโอเอส 60 ส่งคำตอบ [2] ถูกต้อง
	ข้อที่ 7	xxxx
		ส่งคำตอบ ส่งคำตอบ [0] ผิด

รูปที่ 5 การสอบและผลคะแนนของนักศึกษา

การออกแบบที่ 4. การวิเคราะห์ผล

งานวิจัยนี้ได้กำหนดกลุ่มประชากรคือ นักศึกษาหรือผู้ที่สนใจที่ลงทะเบียนเพื่อทดสอบความรู้เรื่อง GEN1102 ระบบสารสนเทศในชีวิตประจำวัน ของสำนักวิชาคอมพิวเตอร์และเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏเชียงราย จำนวน 169 คน

การวิเคราะห์ข้อมูลแบ่งเป็น 2 ประเด็นคือ การประเมินประสิทธิภาพการเปรียบเทียบประสิทธิภาพของการตรวจให้คะแนน และความพึงพอใจของผู้เชี่ยวชาญในผลของการตรวจจากโปรแกรม

การประเมินประสิทธิภาพการเปรียบเทียบประสิทธิภาพของการตรวจให้คะแนน

1. การพิจารณาค่า Point-Score Method เกณฑ์คะแนน 3 กลุ่ม (กำหนดเปลี่ยนแปลงจากหน้าเว็บไซต์ได้) พิจารณาได้จากสมการที่ 1 ดังต่อไปนี้

$$\left[\left(\frac{\sum_{k=0}^n SA}{\sum_{k=0}^m ST} \right) * 100 \right] = pr \quad (1)$$

$$\begin{aligned} & \text{if}(pr < 20)\{ score = 0\} \\ & \text{if}(pr \geq 20)\text{and}(pr < 65)\{ score = 1\} \\ & \text{if}(pr \geq 65)\{ score = 2\} \end{aligned}$$

โดย pr คือ ร้อยละของคำศัพท์ที่พบในคำตอบ
SA คือ จำนวนคำสำคัญที่ค้นพบของผู้สอบ
ST คือ จำนวนคำสำคัญของเฉลย
score คือ ระดับคะแนนประเมิน

2. ผลการทดสอบวัดประสิทธิภาพของโปรแกรมด้วยค่าเอฟเมเชอร์คำนวณได้จากสมการดังนี้

$$F - Measure = \frac{2RP}{R+P} \quad (2)$$

$$P = \frac{A}{A+B}$$

$$R = \frac{A}{A+C}$$

โดย P คือ ค่าความถูกต้อง (Precision)
R คือ ค่าความครบถ้วน (Recall)
A คือ จำนวนคำศัพท์ที่สามารถเลือกได้ถูกต้อง
B คือ จำนวนคำศัพท์ที่เลือกมาไม่ถูกต้อง
C คือ จำนวนคำศัพท์ที่ถูกต้องแต่ไม่ถูกเลือก

3. ค่าเฉลี่ยเลขคณิต (Arithmetic Mean)

$$\bar{x} = \frac{\sum x}{n} \quad (3)$$

โดย $\bar{x}$ คือ ค่าเฉลี่ย
 $\sum x$ คือ ผลรวมทั้งหมด
 n คือ จำนวนประชากร

4. ค่าเบี่ยงเบนมาตรฐาน (Standard Deviation)

$$S.D. = \sqrt{\frac{n \sum x^2 - (\sum x)^2}{n(n-1)}} \quad (4)$$

โดย S.D. คือ ค่าเบี่ยงเบนมาตรฐาน
 $\sum x$ คือ ผลบวกของค่าแต่ละตัว
 n คือ จำนวนประชากร

5. การประเมินผลความพึงพอใจจากผู้เชี่ยวชาญและผู้ใช้งาน (ประเมินตามรูปแบบสารสนเทศ) โดยให้คะแนน 5 ระดับ เป็นแบบสอบถามเกี่ยวกับการประเมินผลความพึงพอใจ มีลักษณะเป็นแบบมาตราส่วนประมาณค่าระบบเฉลี่ย 5 ระดับ (Rating Scale) วิเคราะห์ข้อมูลโดยใช้วิธีหาค่าเฉลี่ย (Average) และค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) โดยกำหนดเกณฑ์การแปลความหมายเพื่อจัดระดับค่าเฉลี่ยออกเป็นช่วงดังต่อไปนี้

ค่าเฉลี่ย 4.50-5.00 หมายความว่า พึงพอใจมากที่สุด
 ค่าเฉลี่ย 3.50-4.49 หมายความว่า พึงพอใจมาก
 ค่าเฉลี่ย 2.50-3.49 หมายความว่า พึงพอใจปานกลาง
 ค่าเฉลี่ย 1.50-2.49 หมายความว่า พึงพอใจน้อย
 ค่าเฉลี่ย 1.00-1.49 หมายความว่า พึงพอใจน้อยที่สุด



ผลการวิจัย

ตารางที่ 5 ตัวอย่างคำถาม และคำสำคัญของคำตอบเฉลยที่ได้กำหนดไว้

คำถาม	คำสำคัญ
1. จงยกตัวอย่างอุปกรณ์นำเข้า 4 อย่าง	เมาส์, คีย์บอร์ด, สแกนเนอร์, ไมโครโฟน, เครื่องอ่านบาร์โค้ด
2. องค์ประกอบของระบบคอมพิวเตอร์มีอะไรบ้าง	บุคลากร, ขั้นตอน, ฮาร์ดแวร์, ซอฟต์แวร์, ข้อมูล
3. หน่วยประมวลผลกลางคืออะไร	ซีพียู, สมอของคอมพิวเตอร์
4. ประเภทของซอฟต์แวร์มีอะไรบ้าง	ซอฟต์แวร์ระบบ, ซอฟต์แวร์ประยุกต์
5. ระบบปฏิบัติการคืออะไร	โปรแกรม, ควบคุม, จัดการ, บริหาร, ตรวจสอบ, ฮาร์ดแวร์

write_detail : ซีฟียู ดูแลดี้

คำตอบจากผู้สอบ : ซีฟียู ดูแลดี้

จำนวนอักขระ : 14

คำตอบเฉลย : ซีฟียู | (จำนวน 1 keyword)

0. พบข้อความ ซีฟียู ในข้อความ ซีฟียู ดูแลดี้ (+1) ตำแหน่งที่ 1 จำนวน 18 อักขระ แล้วแทนด้วยค่าว่าง x ไม่พบข้อความ สมองของคอมพิวเตอร์ ในข้อความ ดูแลดี้ (0)

จำนวนที่เหลือ : 8 อักขระ(บวกค่าว่าง) : ดูแลดี้(6)

คำตอบเฉลย :

- ซีฟียู [ซี - ย - ยู] = 0 + 0 + 0

- สมองของคอมพิวเตอร์ [สม - มพ - ร] = 0 + 0 + 0

การตัดคำออกจากประโยค : ซีฟียู

รูปที่ 5 รูปแบบการประมวลผลการตัดคำ การพิจารณาคำผิด และการเชื่อมต่อกตามโครงสร้าง

การประมวลผลการตัดคำศัพท์ถูกพิจารณาคำสำคัญในฐานข้อมูล จนเหลือคำที่เกินในประโยค และส่งต่อไปยังโปรแกรมสะกดคำเพื่อพิจารณาคำที่เขียนผิด จนได้คำศัพท์ที่เหมาะสมส่งไปยังการเทียบคำตามตามโครงสร้างที่พัฒนาขึ้น

ตารางที่ 6 ตัวอย่างผลการตรวจคำตอบ

เลขที่ผู้สอบ	คำถามที่	คำตอบ	จำนวนคำศัพท์เฉลย	จำนวนคำตอบที่พบ	ค่าสำคัญ
60001	1	เป็นส่วนหนึ่งของบุคลากร ขั้นตอนฮาร์ดแวร์ ขั้นตอน	5	3	60.00
60001	2	เมาส์ คีย์บอร์ด ไมค์	3	2	60.70
60001	3	ซีฟียู ดูแลดี้	2	1	50.00
60001	4	ซอฟต์แวร์ระบบและซอฟต์แวร์ประยุกต์	3	2	60.70
60001	5	แอนดรอยด์ วินโดวส์ ไอโอเอส	3	2	60.70

ตารางที่ 7 แสดงค่าเฉลี่ยของการวัดประสิทธิภาพจากการตรวจคำตอบ

ลำดับข้อสอบ	ค่าความถูกต้อง (ร้อยละ) (Precision)	ค่าความครบถ้วน (ร้อยละ) (Recall)	ประสิทธิภาพ (ร้อยละ) (F-Measure)
1	75.30	32.70	51.50
2	76.40	36.30	54.70
3	78.20	31.00	50.80
4	65.80	58.20	64.90
5	76.40	31.40	50.70
ค่าเฉลี่ย	74.40	37.90	54.50

การจากทดลองและเก็บผลพบว่า ค่าเฉลี่ยของค่าความถูกต้อง (Precision) มีค่าร้อยละ 74.40, ค่าความครบถ้วน (Recall) มีค่าร้อยละ 37.90 และวัดประสิทธิภาพ (F-Measure) มีค่าร้อยละ 54.50

ตารางที่ 8 การประเมินให้คะแนนด้วยเกณฑ์ 3 ระดับ Point-Score Method

คนที่	ค่าเฉลี่ยคะแนนแต่ละข้อจากผู้เชี่ยวชาญ					คะแนนแต่ละข้อที่โปรแกรมฯ				
	ข้อที่ 1	ข้อที่ 2	ข้อที่ 3	ข้อที่ 4	ข้อที่ 5	ข้อที่ 1	ข้อที่ 2	ข้อที่ 3	ข้อที่ 4	ข้อที่ 5
1	0.00	2.00	2.00	1.60	2.00	0.00	2.00	2.00	2.00	2.00
2	2.00	2.00	0.00	1.20	0.20	2.00	2.00	0.00	1.00	0.00
3	0.80	0.60	0.20	0.00	2.00	1.00	1.00	0.00	0.00	1.00
4	1.40	0.00	2.00	0.00	0.00	1.00	0.00	2.00	0.00	2.00
...										
25	2.00	2.00	0.80	1.60	0.00	2.00	2.00	1.00	2.00	0.00

จากการทดลองได้ทำการสุ่มข้อมูลผู้เข้าสอบจำนวน 25 คน เพื่อทำการตรวจข้อสอบ และทำการเปลี่ยนแปลงค่าให้อยู่ในรูปแบบเกณฑ์ Rubric Assessment เพื่อใช้ในเกณฑ์การประเมิน 3 ระดับ โดยเปรียบเทียบระหว่างค่าเฉลี่ยคะแนนแต่ละข้อจากผู้เชี่ยวชาญกับคะแนนแต่ละข้อจากโปรแกรมฯ

ผลการประเมินประสิทธิภาพโดยผู้เชี่ยวชาญ

การประเมินประสิทธิภาพโดยผู้เชี่ยวชาญจากผู้เชี่ยวชาญ 5 ท่านในการใช้โปรแกรมฯ และประเมินผลการทดลองเมื่อผ่านการสอบจากนักศึกษา

ตารางที่ 9 ผลการวิเคราะห์หาความพึงพอใจของโปรแกรม

ข้อความถามของแบบประเมินคุณภาพ	$\bar{X}$	S.D.
1. ด้านกระบวนการ/ขั้นตอนการใช้งานโปรแกรม		
1.1 ทำงานตามอัลกอริทึมอย่างถูกต้อง	4.30	0.49
1.2 มีลำดับทำงานแต่ละโปรแกรมครบถ้วน	4.30	0.63
ผลรวม	4.30	0.56
2. ด้านประสิทธิภาพของโปรแกรม		
2.1 ความถูกต้องแม่นยำของโปรแกรม	3.50	0.65
2.2 ตรงตามวัตถุประสงค์ที่ต้องการ	3.20	0.59
2.3 รายงานข้อมูลเป็นปัจจุบัน	4.00	0.75
2.4 โปรแกรมมีความน่าเชื่อถือ	3.80	0.46
ค่าเฉลี่ยผลรวม	3.62	0.61

ข้อความถามของแบบประเมินคุณภาพ		S.D.
3. ด้านความสะดวก สบายงาม		
3.1 ทำความเข้าใจกับโปรแกรมได้ง่าย	4.00	0.30
3.2 โปรแกรมมีความเหมาะสมน่าใช้	4.60	0.32
ค่าเฉลี่ยผลรวม	4.30	0.31
4. ด้านคุณภาพของโปรแกรม		
4.1 รายงานมีความถูกต้อง	4.30	0.45
4.2 ความพึงพอใจในผลรายงาน	4.10	0.40
ค่าเฉลี่ยผลรวม	4.20	0.43
ค่าเฉลี่ยรวมทั้งหมด	4.11	0.48

จากตาราง ความพึงพอใจโดยรวมของโปรแกรมมีค่า $\bar{X}$ เท่ากับ 4.11 ซึ่งอยู่ในระดับพึงพอใจมาก



สรุป อภิปรายผล

โปรแกรมตรวจสอบข้อสอบอัตนัยภาษาไทยอัตโนมัติที่จัดขึ้นเป็นต้นแบบของหน่วยงาน ได้ทดลองใช้งาน โปรแกรมตรวจสอบข้อสอบในรายวิชา GEN1102 สารสนเทศในชีวิตประจำวัน ซึ่งถูกพัฒนาด้วยวิธีการเทียบคำสำคัญ และได้พัฒนาปรับปรุงประสิทธิภาพการเพิ่มโปรแกรมวิเคราะห์คำผิดและปรับลดวิธีการสืบค้นจากโปรแกรมเดิม เป็นการหาคำสำคัญเป็นการสืบค้นเฉพาะกลุ่มคำตอบจากคำถามในแต่ละข้อเท่านั้น

นอกจากนั้นโปรแกรมประเมินคะแนนแบบเกณฑ์ 3 ระดับ มีความถูกต้องหรือใกล้เคียงกับการตรวจคำตอบจากผู้เชี่ยวชาญ ผลการวิจัยพบว่า ค่าเฉลี่ยของค่าความถูกต้อง มีค่าร้อยละ 74.40, ค่าความครบถ้วน มีค่าร้อยละ 37.90 และวัดประสิทธิภาพ มีค่าร้อยละ 54.50 ในส่วนความพึงพอใจจากผู้เชี่ยวชาญพบว่า ระดับพึงพอใจมาก ($\bar{X} = 4.11$, S.D. = 0.48)



ข้อเสนอแนะ

งานวิจัยดังกล่าวเป็นเพียงส่วนหนึ่งของงานวิจัยเรื่องการตรวจสอบข้อสอบอัตนัยภาษาไทยแบบอัตโนมัติด้วยออนโทโลยีที่จะต้องพัฒนาต่อยอด ซึ่งงานวิจัยนี้ยังไม่ได้แก้ปัญหาคำศัพท์ที่เขียนต่างกัน แต่มีความหมายเหมือนกันในรูปแบบของซีแมนติก ดังนั้นผู้วิจัยจึงมีแนวทางประยุกต์แนวคิดออนโทโลยีมาใช้เพื่อให้เกิดความเหมาะสมและยืดหยุ่นกับระบบโครงสร้างการตรวจสอบข้อสอบซึ่งจะสามารถเทียบเคียงคำศัพท์ที่ผู้สอบเขียนต่างจากคำศัพท์เดิม



เอกสารอ้างอิง

- Boonsri Sisaart, Nipa Sripairoj And Nuch Wattana. (1985). *Measurement and evaluation of education*. Maha Sarakham: Preeda Printing.
- Gronlund, Norman E. (1976). *Society in the Classroom*. Harper New York: Macmillan.
- Guilford, J.P. (1973). *The Nature of Human Intelligence*. New York: McGraw-Hill Book.
- Jerrams-Smith, J., Soh, V., & Callear D. (2001). *Bridging gaps in computerized assessment of texts*. Proceedings of the International Conference on Advanced Learning Technologies, 139-140, IEEE.
- Seatachai Jaihuek And Gp.Capt. Asst.Prof. Surasak Mungsing D.Eng. (2015). *Development of Insurance Information System Quality*. Journal National level 2015 “Technology for development National College of Technology”.
- Uthumporn Charmonman. (1987). *Measurement and Evaluation of Higher Education*. Bangkok: Funy publicing Book.
- Wichian Ketsingh. (1972). *Principles of creating and analyzing*. Bangkok: Odeon Publishing.