



ระบบตรวจข้อสอบอัตนัยภาษาไทยอัตโนมัติด้วยการสืบค้นเชิงความหมาย

Scoring Thai Language Subjective Answer Automatic System by Sematic

เศรษฐชัย ใจฮึก (Seatachai Jaihuek)* และ สุรศักดิ์ มั่งสิงห์ (Surasak Mungsing)*

Received : May 19, 2018
Revised : August 31, 2018
Accepted : September 21, 2018

บทคัดย่อ

ข้อสอบอัตนัย คือ เป็นข้อสอบที่ไม่มีตัวเลือกคำตอบ แต่ใช้วิธีตอบด้วยการเขียนบรรยาย สำหรับการตรวจให้คะแนน ผู้ตรวจอาจใช้เวลานานเพื่อพิจารณาคะแนนให้เกิดความเหมาะสม มหาวิทยาลัยราชภัฏเชียงราย ได้พัฒนาระบบการตรวจข้อสอบอัตนัยแบบออนไลน์ สำหรับทดลองใช้กับวิชา “GEN1102 ระบบสารสนเทศ ในชีวิตประจำวัน” โดยใช้วิธีการค้นหาและเปรียบเทียบคำศัพท์ อย่างไรก็ตามนักวิจัยพบรูปแบบปัญหา ที่เกิดจากคำศัพท์ที่เขียนต่างกัน แต่มีความหมายเหมือนกัน ผู้วิจัยจึงได้นำแนวคิดเกี่ยวกับออนโทโลยี มาใช้แก้ปัญหาด้วยการสร้างความสัมพันธ์เชิงความหมายแล้วส่งต่อไปยังการประมวลผลคะแนนการทดสอบประสิทธิภาพ พบว่า ค่าเฉลี่ยความถูกต้อง (Precision) ที่ร้อยละ 94.42 ค่าความครบถ้วน (Recall) ที่ร้อยละ 59.92 และ ค่าประสิทธิภาพโดยรวม (F-measure) ที่ร้อยละ 72.52 ส่วนความแตกต่างระหว่างการให้คะแนนจากผู้เชี่ยวชาญกับการให้คะแนนจากระบบที่ถูกพัฒนาขึ้นพบว่า ไม่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ผลการสำรวจความพึงพอใจจากการใช้งาน พบว่า ผู้เชี่ยวชาญมีความพึงพอใจในระดับมาก ($\bar{x} = 4.49$, S.D. = 0.52) และผู้ใช้งานมีความพึงพอใจในระดับมาก ($\bar{x} = 4.21$, S.D. = 0.70) เช่นกัน

คำสำคัญ: ออนโทโลยี เว็บแอปพลิเคชัน ข้อสอบอัตนัย ตรวจข้อสอบ อัตโนมัติ

Abstract

A subjective test is the test that does not have options to choose to answer the question but narrative writing is used to answer questions, which requires an amount of time for the examiner to score and takes a long time to determine the accurate score. An online Thai-language examination system was developed and implemented at Chiang Rai Rajabhat University for “GEN1102 The Information System in Daily Life” subject, by using keywords searching and comparison. However, the researcher found problems of different vocabularies with the same meaning. In this research, the ontology concept was introduced to solve the semantic relations problems before processing score evaluation. As the results, the system performance test showed that the new developed system had an average value of the precision at 94.42%, the recall at 59.92 % and the F-measure at 72.52%. The result of t-test also showed that the similarity between expert and system developed was no statistically significant difference at the 0.05 level. The satisfaction survey of the system found that the satisfaction of the experts was at good level ($\bar{x} = 4.49$, S.D. = 0.52) and the satisfaction of users was also at good level ($\bar{x} = 4.21$, S.D. = 0.70).

Keywords: Ontology, Web Application, Subjective, Examination, Automatic.

* สาขาวิชาเทคโนโลยีสารสนเทศ มหาวิทยาลัยศรีปทุม

* Programs in Information Technology, Sripatum University.

1. บทนำ

ข้อสอบอัตนัย (Subjective Test) เป็นลักษณะกระบวนการตอบคำถามโดยการใช้ความสามารถ ด้านการรู้จำ และการคิดวิเคราะห์ เพื่อเขียนบรรยาย เป็นคำตอบ [1] จึงทำให้วิธีการตรวจให้คะแนนไม่เหมือนการตรวจของข้อสอบปรนัย (Objective Test) ที่จัดเตรียมตัวเลือกคำตอบไว้สำหรับผู้สอบเลือกคำตอบ สำหรับให้ระบบรวบรวมคำตอบถูกเป็นผลคะแนน

ปี พ.ศ. 2557 สำนักวิชาคอมพิวเตอร์และเทคโนโลยีสารสนเทศได้ทำการทดลองสร้างต้นแบบระบบตรวจสอบอัตนัยชนิดภาษาไทยผ่านระบบออนไลน์ สำหรับใช้ทำกิจกรรมตอบคำถามในชั้นเรียน โดยเลือกใช้โมเดลเทคนิคการตรวจข้อสอบอัตนัยแบบพื้นฐาน [2] ที่มีลักษณะการเปรียบเทียบคำสำคัญ (คำศัพท์) ระหว่างคำตอบของผู้สอบและคำตอบเฉลย ทั้งนี้ได้จำกัด การกรอกข้อมูลเป็นชนิดภาษาไทยเท่านั้น การทดลอง ใช้งานพบปัญหาการเขียนภาษาไทยที่สะกดผิด ส่งผลให้การตรวจคะแนนเกิดความผิดพลาด ต่อมาปี พ.ศ. 2558 ทดลองใช้กับแบบเรียนรายวิชา GEN1102 สารสนเทศในชีวิตประจำวัน โดยเพิ่มเติมเทคนิคการสะกดคำศัพท์ก่อนส่งคำตอบจริงเพื่อคัดกรองคำศัพท์ที่เขียนผิด ด้วยการตรวจหาความคล้ายของคำศัพท์ ที่เทียบไม่พบในพจนานุกรม และต่อมาปี 2559 ปรับปรุงรูปแบบการประเมินคะแนนร้อยละให้คะแนนเป็น 3 กลุ่ม เพื่อลดความผิดพลาดในการให้คะแนน ตามเทคนิค Point-score Method [1] ดังนี้ 0 คะแนน หมายถึง ไม่พบความถูกต้องของคำตอบ (จับคู่คำสำคัญได้น้อยกว่า ร้อยละ 20) 1 คะแนน หมายถึง มีคำตอบที่ถูกต้อง ในระดับที่ยอมรับได้ (จับคู่คำสำคัญได้ตั้งแต่ร้อยละ 20 ถึง 69) และ 2 คะแนน หมายถึง มีคำตอบที่ถูกต้องใกล้เคียงกับเฉลย (จับคู่คำสำคัญได้มากกว่าร้อยละ 65) ทั้งนี้กระบวนการดังกล่าวสามารถดำเนินการได้ดีในระดับหนึ่ง ต่อมาพบว่าผู้สอบใช้เขียนคำศัพท์ที่มีความหมายเหมือนกันแต่เขียนต่างกันทำให้การจับคู่คำสำคัญ เกิดการผิดพลาด

จากปัญหาดังกล่าว ผู้วิจัยจึงมุ่งสนใจไปในการเพิ่มประสิทธิภาพการสืบค้นคำศัพท์ในส่วนของข้อความ ซึ่งได้นำเอาแนวคิดของออนโทโลยี (Ontology) มาใช้ เป็นโครงสร้างภายใต้ขอบเขตขององค์ความรู้ ของแต่ละบทเรียนที่ทำการสอบ เพื่อจัดความสัมพันธ์เชิงความหมายก่อนส่งต่อ

ผลการทำงานไปยัง ระบบประเมินผลความคล้ายกันระหว่างเฉลยและคำตอบของผู้สอบต่อไป

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ระบบการตรวจข้อสอบอัตนัยชนิดภาษาไทย เป็น การบูรณาการศาสตร์ความรู้ร่วมกันระหว่างการศึกษาและเทคโนโลยีคอมพิวเตอร์ เพื่ออำนวยความสะดวกในการวัดผลและประเมินผลทางการศึกษา

2.1 การวัดผลและประเมินผลทางการศึกษา

Guilford [3] นิยามการวัดผล (Measurement) เป็นการพิจารณา หรือตีค่าข้อมูลในรูปตัวเลข นอกจากนั้น Glonlund [4] นิยามการประเมินผล (Evaluation) คือ กระบวนการที่เป็นระบบในการตัดสินใจภายใต้ขอบเขตของวัตถุประสงค์ในการสอนให้เกิดความสัมฤทธิ์ผลของนักเรียน ส่วน Wichian [5] ให้ความหมายของแบบทดสอบ (Testing) คือ ข้อสอบ ที่วัดสมรรถภาพสมองด้านต่างๆ ที่นักเรียนได้รับการเรียนรู้ผ่านมา แบ่งการทดสอบได้ 2 ประเภท คือ แบบทดสอบอัตนัย (Subjective) และข้อสอบปรนัย (Objective Test) นอกจากนั้น Boonsri [6] ให้คำนิยามสอดคล้องกับ Wichian ว่าแบบทดสอบเป็นกระบวนการวัดผลใช้เปรียบเทียบพฤติกรรมของบุคคลตั้งแต่สองคนขึ้นไป หรือเปรียบเทียบมาตรฐานที่กำหนดไว้ โดยมีเครื่องมือเป็นสิ่งเร้าให้ผู้ตอบแสดงพฤติกรรมที่สังเกต และสามารถวัดผลออกมาในรูปแบบหน่วยวัดได้ ต่อมา Uthumporn [1] นำเสนอข้อดีของข้อสอบอัตนัยในด้านความเหมาะสมที่จะใช้วัดการสังเคราะห์ และการประเมินค่า ความเข้าใจ การวิเคราะห์ การสังเคราะห์ การประเมินค่า กระตุ้นให้ผู้เรียนจัดระบบผสมผสานและแสดงออกซึ่งความคิด แต่ข้อสอบอัตนัยมีข้อจำกัด คือ การสุ่มเนื้อหาทำได้น้อย ไม่ครอบคลุมกับเนื้อหา การเรียนทั้งหมด การตรวจด้วยมนุษย์อาจมีความลำเอียง อารมณ์ของผู้ตรวจคำตอบที่อ่านลายมือได้ยาก

สำหรับรูปแบบเพื่อลดปัญหายุ่งยากและข้อบกพร่องการตรวจข้อสอบอัตนัยจาก Gronlund และ Norman [4] มี 2 รูปแบบซึ่งยังใช้ในปัจจุบัน คือ holistic method หรือ Rating Method โดยผู้ตรวจอ่านคำตอบและประเมินค่าคุณภาพคำตอบจำแนกออกเป็น 5 ถึง 10 กลุ่ม ซึ่งแต่ละกลุ่มจะมีช่วงคะแนนกำหนดไว้ ซึ่งเหมาะกับแบบทดสอบอัตนัยชนิดไม่จำกัดคำตอบ และ แบบ Analytical Method หรือ

Point-score Method ให้ค่าความเชื่อมั่นมากกว่าแบบแรก โดยเทียบคู่มือเฉลยแยกคำตอบเป็นส่วนๆ วิธีดังกล่าวเหมาะกับแบบทดสอบอัตนัยชนิดจำกัดคำตอบ

2.2 ทฤษฎีการสกัดคำภาษาไทยออกจากรูปแบบประโยค

นักวิจัยด้านคอมพิวเตอร์พิจารณาลักษณะภาษาไทยที่เขียนรูปประโยคโดยใช้คำศัพท์เชื่อมติดกัน ไม่เว้นช่องไฟ การสกัดคำที่ใช้ทั่วไปจะมี 2 รูปแบบ คือ รูปแบบที่ 1. การใช้กฎการสร้างพยางค์ภาษาไทย (Rule Base Approach) ใช้พิจารณาการผสมอักขระ พยัญชนะ สระ วรรณยุกต์ ตัวสะกด ตัวการันต์ อย่างถูกหลักไวยากรณ์ วิธีดังกล่าวตัดคำภาษาไทยได้ง่ายและเร็ว แต่ไม่เหมาะกับคำกำกวม รูปแบบที่ 2. การใช้พจนานุกรมคำศัพท์ (Dictionary Approach) หรือรูปแบบของคลังข้อมูล (Corpus Based Approach) โดยลำดับอักขระเทียบเคียงกับคำศัพท์ที่อยู่ในพจนานุกรมคำศัพท์เป็นวิธีที่ไม่เร็วหากมีปริมาณคำจำนวนมาก และไม่เหมาะสมกับกลุ่มคำขาด คำเกิน หรือคำกำกวม ตัวอย่างรูปแบบที่ใช้การตัดคำลักษณะดังกล่าว ได้แก่

- 1) N-GRAM คือ คำนวณค่าความน่าจะเป็นของชุดอักขระที่รวมเป็นคำศัพท์ และอ้างอิงจากสถิติจาก การเลือกใช้งาน
- 2) Longest Word Pattern Matching คือ การเทียบอักขระซ้ำสุดที่ละตัวของคำศัพท์ที่ยาวที่สุดกับพจนานุกรม
- 3) Shortest Word Pattern Matching มีลักษณะคล้าย Longest Word Pattern Matching แต่จะพิจารณาคำศัพท์ที่สั้นที่สุดก่อน
- 4) Word usage Frequency คือ การคำนวณเชิงสถิติความถี่ของคำศัพท์ที่เกิดขึ้นบ่อยครั้ง
- 5) Back Tracking เมื่อคำศัพท์ลำดับเดียวกันมีมากกว่า 1 คำ โปรแกรมจะเลือกคำที่ยาวที่สุด แต่หากคำศัพท์นั้นถูกต้อง จะย้อนกลับไปเลือกคำศัพท์อื่นๆ ที่สั้นกว่า

2.3 ทฤษฎีออนโทโลยี

Gruber [7] และ Chandrasekaran [8] นิยามออนโทโลยีสอดคล้องกัน คือ การให้รายละเอียดเชิงมโนภาพที่ต้องมีการระบุแนวคิด ประกอบไปด้วย Classes, Relation และ Function ซึ่งครอบคลุมส่วนประกอบ ที่สัมพันธ์กัน ดังนั้นออนโทโลยี จึงหมายถึง การบรรยายแนวคิดหรือแนวความรู้ในรูปแบบโครงสร้างลำดับชั้น (Hierarchical Model) ภายใต้กรอบโดเมนที่สนใจ ซึ่งจะประกอบไปด้วย คลาส (Class),

อินสแตนซ์ (Instance) ความสัมพันธ์ (Relationship) คุณสมบัติ (Property) และ กฎ (Rule) จึงทำข้อมูลมีลักษณะคล้ายกราฟหรือทรี ซึ่งจะทำให้คอมพิวเตอร์สามารถเข้าใจความหมายของคำ ส่งผลให้การสืบค้นข้อมูลครอบคลุมในเรื่องที่ค้นหาออนโทโลยีสามารถสร้างขึ้นจากซอฟต์แวร์ เช่น โพรทีเจ (Protégé) หรือ โฮโซ (Hozo) มีภาษาใช้งาน ได้แก่ OWL (Web Ontology Working Group) ที่ถูกเพิ่มส่วนขยายต่อจากภาษา RDF (Resource Description Framework) Mcguinness & Harmelen [9] ที่เก็บข้อมูลแบบทริเปิล (Triple) ทำให้โครงสร้างของการเก็บข้อมูลที่มีความยืดหยุ่น เพราะเก็บข้อมูลแบบ Subject-predicate-object (ประธาน-กริยา-กรรม) โดยมีรูปแบบการเก็บข้อมูลได้ 2 แบบ คือ เก็บตามโครงสร้างเอกสาร XML หรือ ไฟล์ข้อความ (Text Document)

SPARQL (Query Language) คือ ภาษาสอบถาม เป็นภาษาที่ใช้ดึงข้อมูลจากไฟล์ RDF หรือ OWL คล้ายภาษา SQL ที่มีการทำงาน 2 ส่วนคือ Select เป็นส่วนที่เก็บค่าตัวแปรผลลัพธ์ และ Where เป็นเงื่อนไขสำหรับการสืบค้น

2.4 งานวิจัยที่เกี่ยวข้อง

Kolas [10] ใช้ภาษา SPARQL เพื่อสืบค้นข้อมูลทางภูมิศาสตร์ผ่านคลังข้อมูลออนไลน์

Quiliz [11] ได้พัฒนา DARQ ที่เป็นชุดคำสั่ง และข้อกำหนดสำหรับสืบค้นข้อมูลบนเว็บที่หลากหลาย

Harting [12] อธิบายถึงแนวคิดและสำหรับการพัฒนาระบบค้นหาด้วยภาษา SPARQL

Elbassuoni [13] พัฒนาประสิทธิภาพของการสืบค้นข้อมูลจากเว็บได้มากขึ้นด้วยการลดข้อจำกัดของไฟล์ RDF

Seatachai [14] พัฒนาระบบสืบค้นข้อมูลการประกันคุณภาพทางการศึกษาด้วยเทคโนโลยีออนโทโลยี โดยใช้ภาษา SPARQL ขยายการสืบค้นคำสำคัญให้ครอบคลุมกับประเด็นที่สนใจ ผ่านโครงสร้างข้อมูลของ OWL ของสำนักวิชาคอมพิวเตอร์และเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏเชียงราย

3. วิธีดำเนินการวิจัย

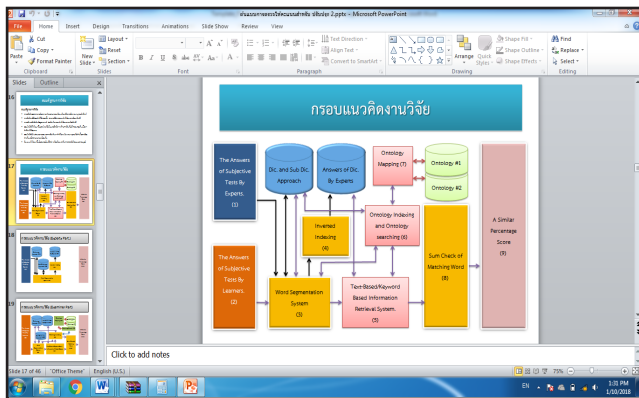
3.1 แผนการดำเนินการ

ใช้วิธีดำเนินการแบบ SDLC ซึ่งมีขั้นตอน ดังนี้

- 1) วิเคราะห์ความต้องการ (Requirement) เมื่อผู้สอบเขียนตอบด้วยคำศัพท์ที่มีความหมายเหมือนกัน จะเกิดปัญหาได้

2 กรณี คือ 1) คำศัพท์นั้นไม่มีในพจนานุกรม ทำให้การรวมคะแนนคำตอบน้อยกว่าคะแนนการตรวจประเมินจากผู้เชี่ยวชาญ และ 2) คำศัพท์ที่เขียนต่างกันแต่มีความหมายเหมือนกัน ทำให้การรวมคะแนนคำตอบมากกว่าคะแนนการตรวจจากผู้เชี่ยวชาญ เมื่อวิเคราะห์ความต้องการของระบบสำหรับการแก้ปัญหาดังกล่าว ด้วยวิธีการสืบค้นคำสำคัญจากฐานข้อมูลเฉลยประสานออนโทโลยี เพื่อกำหนดโครงสร้างและความสัมพันธ์ด้านความหมาย

2) การวิเคราะห์และออกแบบระบบ (Analysis and Design) ปรับปรุงต้นแบบระบบตรวจสอบข้อสอบอัตโนมัติภาษาไทยผ่านระบบออนไลน์ด้วยการสืบค้นเชิงความหมาย มีวิธีการลำดับดังภาพที่ 1



ภาพที่ 1 แผนภาพการทำงานของระบบ

การออกแบบและจัดสร้างซอฟต์แวร์ตามแผนภาพรูปที่ 1 ประกอบด้วย

- (1) The answer of subjective tests by expert ทำหน้าที่รับคำตอบเฉลยจากผู้เชี่ยวชาญแต่ละคน
- (2) The answer of subjective tests by learners ทำหน้าที่รับคำตอบแต่ละข้อของผู้เรียนแต่ละคน และปรับคำศัพท์ให้ถูกต้องก่อนส่งไปสู่กระบวนการถัดไป
- (3) Word segmentation system ทำหน้าที่แปลงข้อความ (Un-structure Data) เป็นการแยกคำสำคัญให้อยู่ในรูปแบบโครงสร้างข้อมูลที่คอมพิวเตอร์สามารถนำมาวิเคราะห์ได้ (Un-structure Data) โดยเทียบเคียงคำศัพท์และแก้ไขคำศัพท์ที่เข้ามาเปรียบเทียบกับ ให้ใกล้เคียงถูกต้องตามพจนานุกรมอิเล็กทรอนิกส์
- (4) Inverted Indexing ทำหน้าที่เก็บคำสำคัญ ลงในฐานข้อมูลคำตอบเฉลยจากผู้เชี่ยวชาญ
- (5) Text-based/keyword based information retrieval

system ทำหน้าที่เทียบคำสำคัญจากคำตอบผู้เชี่ยวชาญกับคำตอบของผู้สอบ

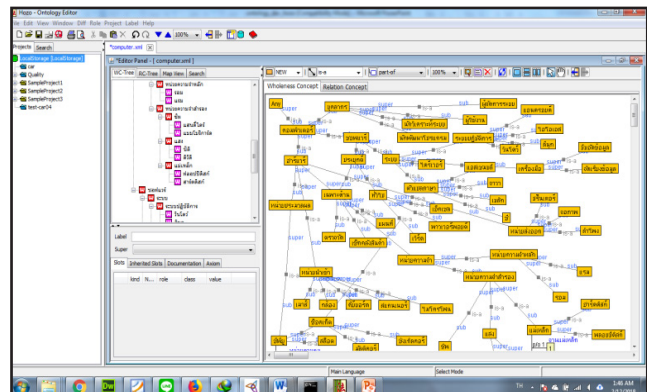
(6) Ontology indexing and ontology searching กรณีที่มีคำสำคัญที่ไม่สามารถเทียบในขั้นตอน Text-based/Keyword based information retrieval system คำสำคัญนั้นจะถูกสืบค้นในโครงสร้างของออนโทโลยี ที่พัฒนาขึ้น เพื่อใช้พิจารณาหาคำสำคัญที่เขียนต่างกันแต่มีความหมายเหมือนกัน

(7) Ontology mapping ทำหน้าที่ประสาน หากใช้ออนโทโลยีมากกว่า 1 โครงสร้าง

(8) Sum check of matching word ทำหน้าที่รวบรวมจำนวนของคำสำคัญที่ค้นพบ และเทียบเป็นค่าร้อยละของการค้นพบคำสำคัญ

(9) A similar percentage score ทำหน้าที่เทียบเคียงอัตราร้อยละให้อยู่ในรูปแบบ Point-score method ตามที่ระบบต้นแบบกำหนดไว้

3) การพัฒนาระบบ (Development/Coding) ด้วยเว็บแอปพลิเคชันที่พัฒนาด้วยภาษา PHP JavaScript และ CSS เชื่อมต่อระบบฐานข้อมูล MySQL และให้บริการจาก Apache web server เพื่อระบบจะนำคำสำคัญ (Key Word) ไปสืบค้นและขยายขอบเขตการค้นหาเชิงความหมาย



ภาพที่ 2 ออนโทโลยีที่ถูกสร้างขึ้นโดยโปรแกรม HOZO

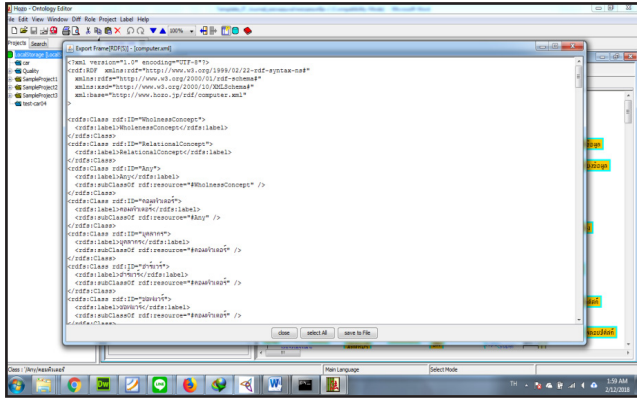
ได้ดำเนินการเก็บข้อมูลเนื้อหาของรายวิชาบทที่ 1 ความรู้เบื้องต้น เพื่อพัฒนาออนโทโลยีโดยในแต่ละ Tag นั้นจะมีค่า Property ที่แสดงถึงการเก็บข้อมูล สามารถอธิบายในแต่ละ Tag ดังนี้

- 1 lesson: docOfName, หัวข้อบทเรียน
- 1 topic: docOfTopic, หัวข้อย่อย
- 1 type: docOfWord, กลุ่มประเภท
- 1 item: docOfItem, ส่วนประกอบ



- 1 word: docOfWord, คำศัพท์
- 1 relation: docOfRelation, ความสัมพันธ์

การออกแบบในส่วนติดต่อกับผู้ใช้งานผ่านเว็บไซต์โดยใช้ภาษา PHP, JavaScript และ CSS เชื่อมต่อฐานข้อมูลด้วย MySQL และใช้ RDF API for PHP V0.9.1 สำหรับสืบค้นจากโครงสร้างไฟล์ computer3.owl



ภาพที่ 3 ภาพไฟล์ computer3.owl



ภาพที่ 4 การรันผล computer3.owl ผ่านทาง RDF API for PHP V0.9.1

4) การทดสอบและบูรณาการระบบ (Testing/System integration) ดำเนินการจัดสอบในปีการศึกษา 3/2559 รายวิชา GEN1102 สารสนเทศในชีวิตประจำวัน นักศึกษาจำนวน 169 คน เลือกข้อสอบจำนวน 5 ข้อ ของบทที่ 1 ความรู้เบื้องต้น ดังตารางที่ 1

การตอบคำถาม ผู้สอบนำเข้าคำตอบด้วยวิธีการพิมพ์ประโยคภาษาไทย จำกัดข้อละไม่เกิน 100 อักขระ และส่งคำตอบที่ละเอียดเพื่อทำการจัดเก็บคำตอบในฐานะข้อมูล

5) การดำเนินงานและการบำรุงรักษา (Operation and Maintenance) การทำงานแบ่งเป็น 3 ส่วน คือ (1) การวิเคราะห์หาประสิทธิภาพโดยรวม (2) การพิจารณาผลคะแนนจาก

ตารางที่ 1 ตัวอย่างคำถาม บทที่ 1 ความรู้เบื้องต้น

ข้อที่	คำถาม
1	จงยกตัวอย่างอุปกรณ์นำเข้า (อย่างน้อย 4 อย่าง)
2	องค์ประกอบของระบบคอมพิวเตอร์มีอะไรบ้าง
3	หน่วยประมวลผลกลางคืออะไร
4	ประเภทของซอฟต์แวร์มีอะไรบ้าง
5	ระบบปฏิบัติการคืออะไร

ตารางที่ 2 ตัวอย่างคำสำคัญของคำตอบเฉลย

ข้อที่	คำสำคัญจากเฉลย
1	เมาส์ คีย์บอร์ด สแกนเนอร์ ไมโครโฟน เครื่องอ่านบาร์โค้ด
2	บุคลากร ขั้นตอน ฮาร์ดแวร์ ซอฟต์แวร์ ข้อมูล
3	ซีพียู สมอของคอมพิวเตอร์
4	ซอฟต์แวร์ระบบ ซอฟต์แวร์ประยุกต์
5	โปรแกรมควบคุม จัดการ บริหาร ตรวจสอบ ฮาร์ดแวร์

การตรวจสอบด้วยระบบและผู้เชี่ยวชาญ และ (3) ประเมินผลความพึงพอใจจากผู้เชี่ยวชาญ และผู้ใช้งานระบบ

6) การวิเคราะห์และการประเมินผล ใช้สถิติที่เกี่ยวข้องดังนี้

(1) การวัดผลแบบ Analytical Method ดังสมการที่ 1

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

if (pr < 20) {score = 0}

if (pr ≥ 20) and (pr < 65) {score = 1}

if (pr ≥ 65) {score = 2}

โดย pr คือ ร้อยละของคำศัพท์ที่พบในคำตอบ

SA คือ จำนวนคำสำคัญที่ค้นพบของผู้สอบ

ST คือ จำนวนคำสำคัญของเฉลย

score คือ ระดับคะแนนประเมิน

(2) วัดประสิทธิภาพของระบบ ดังสมการที่ 2

$$f\text{-measure} = \frac{2RP}{R+P} \quad (2)$$

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

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



โดย P คือ ค่าความถูกต้อง

R คือ ค่าความครบถ้วน

A คือ จำนวนคำศัพท์ที่สามารถเลือกได้ถูกต้อง

B คือ จำนวนคำศัพท์ที่เลือกมาไม่ถูกต้อง

C คือ จำนวนคำศัพท์ที่ถูกต้องแต่ไม่ถูกเลือก

(3) ค่าเฉลี่ยเลขคณิต (Arithmetic Mean) ดังสมการที่ 3

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

โดย $\bar{x}$ คือ ค่าเฉลี่ย

$\sum x$ คือ ผลรวมทั้งหมด

n คือ จำนวนประชากร

(4) ค่าเบี่ยงเบนมาตรฐาน (S.D.) ดังสมการที่ 4

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

โดย $S.D.$ คือ ค่าเบี่ยงเบนมาตรฐาน

$\sum x$ คือ ผลบวกของค่าแต่ละตัว

n คือ จำนวนประชากร

(5) การวิเคราะห์ข้อมูลเชิงอนุมานด้วยค่าสถิติ t-test

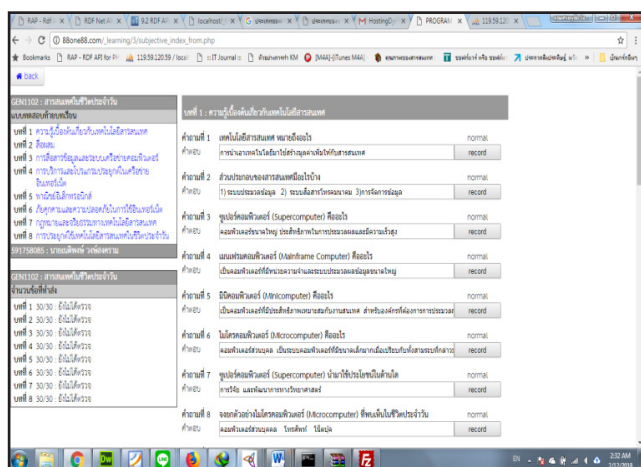
ทดสอบค่าคล้ายกันของค่าเฉลี่ยคะแนนจากผู้เชี่ยวชาญ

จำนวน 5 คน และระบบที่พัฒนาขึ้น

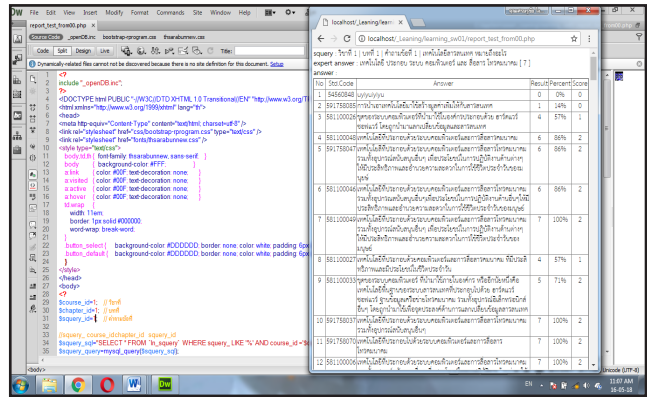
4. ผลการดำเนินงาน

4.1 ผลการพัฒนาระบบ

ผู้สอบเข้าสู่ระบบด้วยชื่อผู้ใช้และรหัสผ่าน โดยผู้สอบจะเข้าสอบข้อสอบอัตโนมัติแบบออนไลน์ของตนเอง ตัวอย่างภาพที่ 5



ภาพที่ 5 หน้าจอของการตอบคำถามข้อสอบอัตโนมัติ



ภาพที่ 6 การประมวลคะแนนผลการสอบ

ข้อมูลการสอบส่งไปประมวลผลคะแนนด้วยเกณฑ์ 3 ระดับ (Point-score Method) ตามสมการที่ 1 เพื่อเปรียบเทียบผลคะแนนจากผู้เชี่ยวชาญ

4.2 ผลการดำเนินงาน

เมื่อระบบได้ทำการตรวจสอบเสร็จสิ้น ได้ผลการสืบค้นคำสำคัญของข้อสอบแต่ละฉบับ ดังตารางที่ 3

ตารางที่ 3 ผลเปรียบเทียบระหว่างระบบงานต้นแบบ (Original System) และระบบที่พัฒนาร่วมกับการสืบค้นด้วยออนโทโลยี (Modify System)

ข้อที่ คนที่	Original System					Modify System				
	1	2	3	4	5	1	2	3	4	5
1	3	4	1	2	4	4	5	1	2	5
2	4	5	1	1	3	5	3	1	1	5
3	4	3	0	2	3	4	5	0	2	3
4	4	3	1	2	3	4	4	2	2	3
5	4	4	0	1	4	4	4	1	1	5
...										
169	3	3	0	1	2	4	4	0	2	4

เมื่อพิจารณาผลการสืบค้นคำสำคัญของข้อสอบแต่ละข้อจากผู้สอบ 169 คน พบว่า ค่าเฉลี่ยของความถูกต้องที่ร้อยละ 94.40, ค่าความครบถ้วนที่ร้อยละ 59.90 และค่าวัดประสิทธิภาพโดยรวมที่ร้อยละ 72.50 ดังตารางที่ 4

สำหรับการเปรียบเทียบระดับการให้คะแนน ด้วยเกณฑ์ประเมิน 3 ระดับ จากสมการที่ 1 ผู้วิจัยได้สุ่มเลือกการตอบคำถามของผู้เข้าสอบจำนวน 25 คน เพื่อให้ผู้เชี่ยวชาญจำนวน 5 คนตรวจและประเมินผลคะแนน ผลการดำเนินการแสดงในตารางที่ 5



ตารางที่ 4 ผลเปรียบเทียบระหว่างระบบงานต้นแบบ (Original System) และระบบที่พัฒนาร่วมกับการสืบค้นด้วยออนโทโลยี (Modify System)

ข้อสอบที่	Precision	Recall	F-measure
1	95.30	54.70	69.50
2	96.40	58.30	72.70
3	98.20	53.00	68.80
4	85.80	80.20	82.90
5	96.40	53.40	68.70
ค่าเฉลี่ย	94.42	59.92	72.52

ตารางที่ 5 การประเมินให้คะแนนระหว่างคะแนนเฉลี่ยจากผู้เชี่ยวชาญ 5 คน และคะแนนที่ได้จากระบบ

คนที่	ค่าเฉลี่ยคะแนนแต่ละข้อจากผู้เชี่ยวชาญ					คะแนนแต่ละข้อที่ระบบประเมินผล				
	1	2	3	4	5	1	2	3	4	5
1	0.00	2.00	2.00	1.60	2.00	0	2	2	2	2
2	2.00	2.00	0.00	1.20	0.20	2	2	0	1	0
3	0.80	0.60	0.20	0.00	2.00	1	1	0	0	1
4	1.40	0.00	2.00	0.00	0.00	1	0	2	0	2
5	1.00	0.60	0.80	2.00	2.00	1	1	0	2	2
...										
25	2.00	2.00	0.80	1.60	0.00	2	2	1	2	0

การเปรียบเทียบค่าเฉลี่ยการให้คะแนนของกลุ่มตัวอย่าง 2 กลุ่ม ด้วย T-test ซึ่งให้ผลตามตารางที่ 6

ตารางที่ 6 แยกผลการเปรียบเทียบ T-test การให้คะแนนข้อสอบ 5 ข้อ ข้อละ 25 ฉบับ โดยเปรียบเทียบทีละข้อ

ข้อที่	t Stat	t Critical one-tail	P(T<=t) one-tail	Pearson Correlation
1	1.41	1.71	0.09	0.54
2	0.90	1.71	0.19	0.54
3	0.70	1.71	0.25	0.55
4	1.07	1.71	0.15	0.40
5	1.44	1.71	0.08	0.74

สมมติฐานหลัก (H0): คะแนนจากระบบคล้ายกับคะแนนจากผู้เชี่ยวชาญ เมื่อเปรียบเทียบค่า T-test ของผลการตรวจ

คะแนนทั้ง 5 ข้อ พบว่า มีค่าแปรผลไม่แตกต่างกัน ค่า t stat น้อยกว่าหรือเท่ากับ t critical one-tail อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05

ตารางที่ 7 ผลการวิเคราะห์ผลรวมการตรวจให้คะแนนของทั้งสองกลุ่มด้วย T-test จากข้อสอบ 25 ฉบับ

รายการ	คะแนน	
	ระบบ	ผู้เชี่ยวชาญ
mean	1.31	1.17
variance	0.10	0.14
observations	25.00	25.00
pearson correlation	0.55	
hypothesized mean diff.	0.00	
df	24.00	
t Stat	1.11	
P(T<=t) one-tail	0.15	
t critical one-tail	1.71	
P(T<=t) two-tail	0.30	
t critical two-tail	2.06	

ผลรวมค่าเฉลี่ยคะแนนของผู้เชี่ยวชาญเปรียบเทียบผลรวมค่าเฉลี่ยคะแนนจากระบบ สมมติฐานหลัก (H0): ค่าเฉลี่ยคะแนนของผู้เชี่ยวชาญคล้ายกับค่าเฉลี่ยคะแนนจากระบบ พบว่าค่าแปรผลไม่แตกต่างกัน โดย t stat น้อยกว่าหรือเท่ากับ t critical one-tail อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05

4.2 ผลการประเมินความพึงพอใจจากผู้เชี่ยวชาญ

การประเมินความพึงพอใจจากผู้เชี่ยวชาญจำนวน 5 คน แสดงผลดังตารางที่ 8

ตารางที่ 8 ผลการวิเคราะห์หาประสิทธิภาพของระบบ

ข้อความถามของแบบประเมินคุณภาพ	$\bar{x}$	S.D.
1. การทำงานของระบบ	4.30	0.56
2. การออกแบบหน้าจอ	4.35	0.60
3. การทดสอบใช้งาน	4.70	0.45
4. ด้านความปลอดภัย	4.60	0.45
ค่าเฉลี่ย	4.49	0.52

จากตารางที่ 8 พบว่า ประสิทธิภาพโดยรวมของระบบ มีค่าเฉลี่ยเท่ากับ 4.49 ซึ่งอยู่ในระดับดีมาก

4.3 ผลการประเมินความพึงพอใจจากกลุ่มผู้ใช้

การประเมินความพึงพอใจจากกลุ่มผู้ใช้ ดังตารางที่ 9

ตารางที่ 9 ผลการประเมินความพึงพอใจของผู้ใช้

ข้อคำถามของแบบประเมินคุณภาพ	$\bar{x}$	S.D.
1. เชื่อถือได้ของระบบ (Reliable)	4.10	0.82
2. ระบบเข้าใจง่าย (Simple)	4.10	0.70
3. รายงานผลทันต่อเวลา (Timely)	4.36	0.55
4. คุ่มราคาต่อการใช้งาน (Economical)	4.17	0.62
5. ผลรายงานตรวจสอบได้ (Verifiable)	4.35	0.48
6. ระบบมีความยืดหยุ่นสามารถใช้ได้กับอุปกรณ์ที่หลากหลาย (Flexible)	4.21	0.61
7. สอดคล้องกับความต้องการขององค์กร (Relevant)	4.23	0.58
8. ระบบมีสะดวกในการเข้าถึง และใช้งาน (Accessible)	4.15	0.55
9. ระบบมีความปลอดภัย (Secure)	4.19	0.38
ค่าเฉลี่ย	4.21	0.70

จากตารางที่ 9 พบว่า ผู้ใช้มีความพึงพอใจระบบตรวจสอบข้อสอบอัตโนมัติภาษาไทยอัตโนมัติ อยู่ในระดับ พึงพอใจมาก โดยมีค่า $\bar{x}$ เท่ากับ 4.21

5. สรุป

ระบบตรวจสอบข้อสอบอัตโนมัติภาษาไทยอัตโนมัติของมหาวิทยาลัยราชภัฏเชียงราย ได้ทดลองใช้งานระบบตรวจสอบข้อสอบในรายวิชา GEN1102 สารสนเทศ ในชีวิตประจำวัน โดยแต่เดิมใช้วิธีการตรวจด้วยวิธีการเทียบคำสำคัญ และได้พัฒนาปรับปรุงประสิทธิภาพ ด้วยการสืบค้นเชิงความหมายจัดการกับปัญหาของคำศัพท์ที่เขียนต่างกันแต่มีความหมายเหมือนกัน เพื่อให้ระบบประเมินคะแนนแบบเกณฑ์ 3 ระดับ มีความถูกต้องหรือใกล้เคียงกับการตรวจคำตอบจากผู้เชี่ยวชาญ ผลการวิจัยพบว่า มีค่าเฉลี่ยความถูกต้องที่ร้อยละ 94.42 ความครบถ้วนที่ร้อยละ 59.92 และค่าประสิทธิภาพโดยรวมที่ร้อยละ 72.52 ซึ่งเมื่อพิจารณาการเชื่อมต่อแนวคิดทางออนโทโลยี ทำให้สืบค้นคำสำคัญมีปริมาณของตัวเลือกภายในโดเมนที่เพิ่มขึ้น แต่ทั้งนี้ระบบยังไม่สามารถจัดการประโยคคำตอบที่สลับตำแหน่งคำศัพท์จนเกิดความหมายรูปประโยคที่ผิดพลาดได้ ซึ่งผู้เชี่ยวชาญพิจารณาว่า อาจเกิดจากเจตนาของผู้สอบที่จะต้องการ

ทดสอบระบบ หรือเกิดความเข้าใจผิด ในเนื้อหาสาระการสอน นอกจากนั้นรูปแบบข้อสอบ ที่ตั้งคำถามให้กรอกคำตอบมากกว่า 1 ตัวอย่าง หากมี คำสำคัญของเฉลยมาก จะส่งผลกับระดับคะแนน ดังนั้น ผู้บันทึกข้อสอบต้องใช้ วิจารณ์งานในการออกข้อสอบให้เหมาะสมกับคำสำคัญของเฉลย และควรมีระบบที่สามารถปรับแต่งข้อกำหนดของการให้คะแนน เพื่อสร้างความยืดหยุ่นเหมือนการตรวจโดยใช้มนุษย์ ในส่วนความคล้ายกันระหว่างการตรวจเฉลยด้วยผู้เชี่ยวชาญและระบบที่พัฒนาขึ้นด้วย t-test พบว่า มีค่าแปรผลไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 และในส่วนค่าเฉลี่ยด้านความพึงพอใจ ต่อระบบจากผู้เชี่ยวชาญอยู่ในระดับมาก ($\bar{x} = 4.49$, S.D. = 0.52) และผู้ใช้งานมีความพึงพอใจอยู่ในระดับมาก ($\bar{x} = 4.21$, S.D. = 0.70) เช่นกัน

การทำงานของระบบตรวจสอบข้อสอบอัตโนมัติภาษาไทยอัตโนมัตินั้นยังมีข้อจำกัดของระบบ ซึ่งต้องอาศัยการพัฒนาของเทคโนโลยี และผู้เชี่ยวชาญด้านภาษาไทย ในการช่วยพัฒนาระบบฐานข้อมูลคำศัพท์ และเชื่อมต่อกับออนโทโลยีให้เกิดประสิทธิภาพครอบคลุมใน กลุ่มโดเมนหลักของเนื้อหา รายวิชา ทั้งนี้ผู้วิจัย ได้มีแนวทางที่จะพัฒนาเพิ่มเติมในอนาคต ให้ระบบมีประสิทธิภาพใกล้เคียงกับผู้เชี่ยวชาญ แต่ละเนื้อหาวิชาให้เหมาะสมกับการตรวจสอบข้อสอบอัตโนมัติ และสามารถตรวจสอบข้อสอบอัตโนมัติหลากหลายรูปแบบมากขึ้น

6. เอกสารอ้างอิง

- [1] U. Charmonman. *Measurement and Evaluation of Higher Education Bangkok: Funy publishing Book*, 1987.
- [2] J. Jerrams-Smith, V. Soh, and D. Callar. "Bridging gaps in computerized assessment of texts." *Proceedings of the International Conference on Advanced Learning Technologies, IEEE*, pp. 139-140, 2001.
- [3] J.P. Guilford. *The Nature of Human Intelligence*. New York: McGraw-Hill Book, 1973.
- [4] N. E. Gronlund. *Society in the Classroom*. Harper New York: Macmillan, 1976.
- [5] W. Ketsingh. *Principles of creating and analyzing*. Bangkok: Odeon Publishing, 1972.
- [6] B. Sisaart, Nipa Sripairoj And Nuch Wattana. *Measurement and evaluation of education. Maha Sarakham: Preeda*



- Printing, 1985.
- [7] T.R. Gruber. "A Translation Approach to Portable Ontology Specification." *Knowledge Acquisition*, Vol. 5, No. 2, pp. 199-220, 1993.
- [8] B. Chandrasekaran, J.R. Josephson, and V.R. Benjamins. "What are ontologies And Why Do We Need Them." *IEEE Intelligent System*, Vol. 14, No. 1, 20-26, 1999.
- [9] D.L. McGuinness and F.V. Harmelen. *Owl Web Ontology Language Overview, World Wide Web Consortium (W3C) Recommendation*, Available Online at <http://www.w3.org/TR/owl-features>, accessed on 20 September 2012.
- [10] D. Kolas. *Supporting Spatial Semantics with SPARQL, Terra Cognita Work-shop*, 2008.
- [11] B. Quilitz and U. Leser. "Querying distributed RDF data sources with SPARQL." *Proceedings of the 5th European Semantic*, 2008.
- [12] O. Hartig, C. Bizer and J. Freytag. Executing SPARQL Queries over the Web of Linked Data, In *International Semantic Web Conference*, Vol. 5823, pp. 293-309, 2009.
- [13] S. Elbassuoni, M. Ramanath, R. Schenkel, and G. Weikum. *Searching RDF Graphs with SPARQL and Keywords*, IEEE, 2010.
- [14] S. Jaihuek And Gp.Capt. S. Mungsing D.Eng. "Development of Insurance Information System Quality." *Journal National level 2015*, Technology for development National College of Technology, 2015.
-