

การพัฒนาระบบตรวจสอบข้อสอบอัตนัยภาษาไทยบนพื้นฐานการวัดค่าความคล้ายคลึง Development of Thai Subjective Scoring System Based on Cosine-Similarity Measure

รักษาติภูมิ วงศ์สาระ¹ ขนิษฐา หอมจันทร์² และ นงนุช เกตุย^{3*}

Rukchartpoom Wongsara¹, Kanitha Homjun² and Nongnuch Ketui^{3*}

หลักสูตรวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยีการเกษตร มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา น่าน^{1,2,3}

Computer Science Program, Faculty of Science and Agricultural Technology,

Rajamangala University of Technology Lanna Nan^{1,2,3}

E-Mail: pueert1324@gmail.com¹, kanithahomjun@rmutl.ac.th² and nongnuchketui@rmutl.ac.th³

บทคัดย่อ

การวิจัยเรื่องการพัฒนาการตรวจสอบข้อสอบอัตนัยภาษาไทยบนพื้นฐานการวัดค่าความคล้ายคลึงมีวัตถุประสงค์เพื่อ 1) พัฒนาระบบตรวจสอบข้อสอบอัตนัยภาษาไทยบนพื้นฐานการวัดค่าความคล้ายคลึง และ 2) ประเมินผลการใช้ระบบตรวจสอบข้อสอบอัตนัยภาษาไทยบนพื้นฐานการวัดค่าความคล้ายคลึง สถิติที่ใช้ในการวิเคราะห์ข้อมูลคือ ค่าเฉลี่ยเลขคณิต ($\bar{X}$) และส่วนเบี่ยงเบนมาตรฐาน (S.D.) กลุ่มเป้าหมาย คือ อาจารย์ และนักศึกษา หลักสูตรวิทยาการคอมพิวเตอร์ จำนวน 20 คน และแบบทดสอบที่มีคำตอบ จำนวน 170 คำตอบ

ผลการวิจัยพบว่า 1) ระบบตรวจสอบข้อสอบอัตนัยภาษาไทยบนพื้นฐานการวัดค่าความคล้ายคลึงได้แบ่งผู้ใช้งาน 2 กลุ่ม ได้แก่ ผู้ออกข้อสอบ และผู้ทำข้อสอบ ซึ่งมีขั้นตอนคือสร้างคำถาม คำตอบ บันทึกหลักฐานข้อมูล แล้วประยุกต์คำตอบเข้าสู่กระบวนการทางเทคโนโลยีประมวลผลภาษาธรรมชาติ เริ่มต้นจากกระบวนการเตรียมข้อมูล แบ่งคำและตัดคำไม่สำคัญ คำนวณค่าความถี่ของคำและค่าน้ำหนักของคำที่เกิดขึ้นในคำตอบ และคำนวณค่าความเหมือนความคล้าย ระหว่างคำตอบผู้ออกข้อสอบและคำตอบของผู้ทำข้อสอบ ผ่านรูปแบบเว็บแอปพลิเคชันที่พัฒนาด้วยภาษาไพทอน พบว่า ระบบสามารถใช้งานได้ และผู้เชี่ยวชาญประเมินอยู่ในระดับมาก และ 2) ผลการประเมินความพึงพอใจต่อการใช้งานระบบมีผลประเมินโดยรวมทุกด้าน อยู่ในระดับพึงพอใจมาก ($\bar{X}$ =4.05, S.D.=0.81)

คำสำคัญ: ระบบตรวจสอบข้อสอบอัตนัย, ค่าความคล้ายคลึง, ประมวลผลภาษาธรรมชาติ

ABSTRACT

The purposes of the research were to develop a Thai subjective scoring system based on cosine-similarity measure and to evaluate a satisfaction assessment of development of Thai subjective scoring system based on cosine-similarity measure. The statistic used to analyze the data is the mean ($\bar{X}$) and standard deviation (S.D.). The target group of computer science lecturer and students is the sample and 170 answers are used to evaluate.

The research findings showed that the development of Thai subjective scoring system based on cosine-similarity measure divided users into 2 groups who are the question makers and answering-students. The processes are starting from creating questions with their answers which are saved into database, applying the answer following as the natural language processing technique (NLP). This NLP technique arranges to preprocessing data, word weight calculation, and cosine-similarity calculation between answer-lecturer and answer-student through a web application as python web framework. The development system can be used with high quality, and the result of user satisfaction achievement is at high level ($\bar{X}$ =4.05, S.D. = 0.81).

Keywords: Thai subjective scoring, Cosine similarity, Natural language processing

บทนำ

ปัจจุบันการใช้เทคโนโลยีปัญญาประดิษฐ์มาประยุกต์ในงานต่าง ๆ มีแพร่หลายมากขึ้น โดยเทคโนโลยีปัญญาประดิษฐ์ (Artificial Intelligence หรือ AI) ที่ทำให้คอมพิวเตอร์เข้าใจภาษามนุษย์เพื่อสื่อสารและวิเคราะห์ข้อมูล เลียนแบบสมองของมนุษย์ได้ เทคโนโลยีการประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP) เป็นวิทยาการแขนงหนึ่งในเทคโนโลยีปัญญาประดิษฐ์ ที่ศึกษา เลียนแบบพฤติกรรม หรือคุณลักษณะของภาษาที่ใช้ในสื่อสารของมนุษย์มาประมวลผลด้วยเครื่องคอมพิวเตอร์ ยกตัวอย่างภาษาพูดทั่วไปที่ใช้พูดคุยกันตามปกติในชีวิตประจำวัน การเขียนข้อความความคิดเห็นในรูปแบบต่าง ๆ หลักการทำงานของ การประมวลผลภาษาธรรมชาติจะมีลักษณะคล้ายกันกับภาษาของมนุษย์โดยมีเทคนิคบางอย่างที่ทำให้เครื่องคอมพิวเตอร์สามารถทำความเข้าใจและประมวลผลภาษาเหล่านั้นได้ ซึ่งทางสถาบัน SAS [1] ได้กล่าวไว้ และจะช่วยให้ประสิทธิภาพการทำงานของมนุษย์ดีขึ้น

การประยุกต์ใช้งานเทคโนโลยีประมวลผลภาษาธรรมชาติสามารถนำไปใช้ในงานต่าง ๆ มากมาย การศึกษาก็เช่นกัน [2] ส่งผลทำให้การเรียนการสอนมีประสิทธิภาพมากขึ้น โดยเฉพาะงานการตรวจความถูกต้องของข้อมูล ซึ่งถือเป็นแนวทางการวัดและประเมินผลหนึ่งของการศึกษา การตรวจข้อสอบอัตโนมัติก่อนเข้ามีความท้าทาย เนื่องจากต้องใช้วิจารณ์ของครูอาจารย์ และความถูกต้องของคำตอบที่ค่อนข้างแม่นยำ และใช้เวลานานในการตรวจสอบ [3] โดยเฉพาะภาษาไทย จะพบปัญหาเรื่องการแบ่งคำ แบ่งประโยคของข้อความ ซึ่งถ้ามองในด้านเทคนิคการประมวลผลภาษาธรรมชาติแล้ว ค่อนข้างยากในการพัฒนาระบบ เพื่อให้เกิดความสะดวกสบายและการตรวจคำตอบที่ง่ายด้วยโปรแกรมที่ถูกพัฒนาขึ้น การตรวจข้อสอบอัตโนมัติออนไลน์จึงเป็นที่น่าสนใจ ที่จะพัฒนาเป็นนวัตกรรมด้านซอฟต์แวร์ และเผยแพร่ผลงานให้กว้างขวางมากยิ่งขึ้น

ดังนั้นทางผู้วิจัยเล็งเห็นว่าการพัฒนาระบบตรวจข้อสอบอัตโนมัติภาษาไทยบนพื้นฐานการวัดค่าความคล้ายคลึง เริ่มต้นจากการสร้างคำถาม คำตอบ แล้วบันทึกลงฐานข้อมูล แล้วนำคำตอบที่ได้ ผ่านกระบวนการทางเทคโนโลยีประมวลผลภาษาธรรมชาติ ซึ่งประกอบไปด้วยการเตรียมข้อมูล การตัดคำไม่สำคัญ [4] การคำนวณค่าความถี่ของคำและค่าน้ำหนักของคำที่เกิดขึ้นในคำตอบ [5] และการคำนวณค่าความคล้ายคลึง หรือความเหมือนความคล้าย [6] ระหว่างคำตอบของผู้ออกข้อสอบและคำตอบของผู้ทำข้อสอบในโครงสร้างทางคณิตศาสตร์หรือที่เรียกว่าเวกเตอร์ และนำแนวทางที่ได้ไปใช้งาน ทั้งนี้ได้เลือกใช้ภาษาไพทอน รองรับการทำงานผ่านเว็บไซต์แอปพลิเคชันที่เขียนขึ้นสำหรับภาษาไพทอนที่ทำงานร่วมกับเว็บเซิร์ฟเวอร์ได้ [7]

1. วัตถุประสงค์การวิจัย

- 1.1 เพื่อพัฒนาระบบตรวจข้อสอบอัตโนมัติภาษาไทยบนพื้นฐานการวัดค่าความคล้ายคลึง
- 1.2 เพื่อประเมินผลการใช้ระบบตรวจข้อสอบอัตโนมัติภาษาไทยบนพื้นฐานการวัดค่าความคล้ายคลึง

2. เอกสารและงานวิจัยที่เกี่ยวข้อง

มาสวี่ร์ มาศศิธรโชติ [8] ได้ให้ความหมายของ การประมวลผลภาษาธรรมชาติเป็นวิธีการที่จะช่วยทำให้คอมพิวเตอร์เข้าใจภาษาธรรมชาติของมนุษย์ที่ไม่มีโครงสร้างที่แน่นอน โดยการแปลให้อยู่ในรูปแบบที่เป็นโครงสร้าง การประมวลผล ซึ่งการวิเคราะห์ภาษาธรรมชาติในภาษาไทยส่วนใหญ่เป็นการวิเคราะห์องค์ประกอบและการวิเคราะห์ระดับโครงสร้างทางภาษาหรือวากยสัมพันธ์เท่านั้น

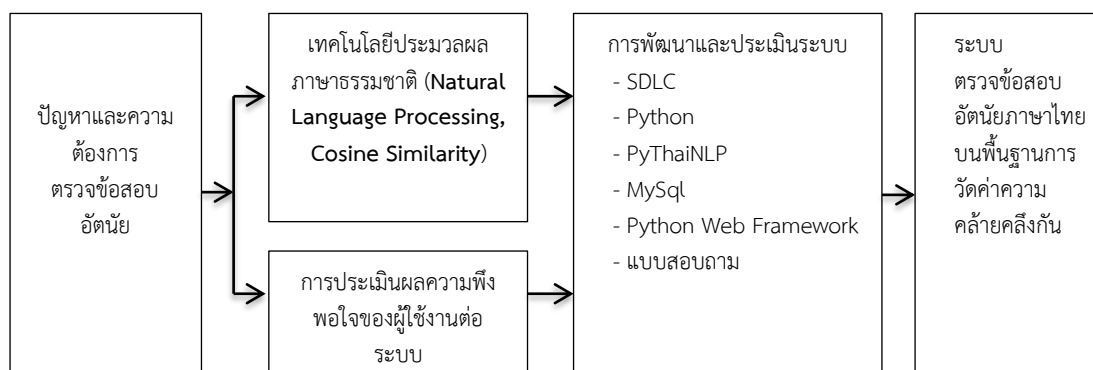
นราธร มั่นสุวรรณ [9] ได้ทำวิทยานิพนธ์เกี่ยวกับอิทธิพลของภาษาต่อประสิทธิภาพการสื่อสารของธนาคารแห่งประเทศไทยโดยนำโมดูลหรือไลบรารี (PyThaiNLP) ที่พัฒนาโดยภาษาไพทอนมาใช้สำหรับตัดคำและแบ่งคำรวมหลักการ PYICU เพื่อให้ได้คำที่เหมาะสม เช่น เงินเพื่อ ตัวโมดูล เป็นต้น เนื่องจาก เงินเพื่อ ถือเป็นคำประสมที่มีความหมาย หากทำการแยกคำความหมายของข้อความก็จะเปลี่ยนไป

ทัศนีย์ อุทัยสุริ [10] ได้ศึกษาวิจัยเกี่ยวกับการสกัดคำสำคัญจากบทคัดย่อภาษาอังกฤษ โดยได้ให้ความหมายของการให้น้ำหนักคำ จากวิธีการถ่วงน้ำหนักคำ (Term Frequency – Inverse Document Frequency: TF-IDF) ซึ่งเป็นวิธีที่นิยมใช้กันมากด้านการสืบค้นสารสนเทศและเหมืองข้อความ ใช้หลักการทางด้านสถิติ เพื่อประเมินความสำคัญของคำต่อเอกสาร ความสำคัญของคำเป็นสัดส่วนโดยตรงกับจำนวนครั้งที่คำคำนั้นปรากฏในเอกสารนั้น แต่ถ่วงลดความสำคัญโดยความถี่ของคำนั้นในกลุ่มเอกสารทั้งหมด

สายเพชร พงศกร และพฤษดี ศิริแสงตระกูล [11] ได้ให้คำอธิบายเกี่ยวกับเวกเตอร์ (Vector) คือ ปริมาณที่บอกทั้งขนาดและทิศทาง หรือในทางคอมพิวเตอร์เป็นการแปลงข้อมูลเหล่านั้น เป็นข้อมูลแบบ 2 มิติ ในการวิจัยนี้ ในการแปลงข้อมูลระหว่างคำตอบของอาจารย์และนักเรียนกลายเป็นเวกเตอร์เพื่อใช้การคำนวณหาความเหมือนความคล้าย (Cosine Similarity)

ไกรศักดิ์ เกษร [12] ได้ทำวิจัยเกี่ยวกับเสริชเอ็นจิน เรื่องการแสดงผลการค้นหาข้อมูล: จุดเล็ก ๆ ที่สำคัญและไม่ควรมองข้ามสำหรับเสริชเอ็นจิน โดยได้อธิบายค่าความเหมือนความคล้าย เป็นการวัดความคล้ายความเหมือนระหว่างเอกสารและคิวรี หรือการเปรียบเอกสารและคิวรีเป็นเอกสารข้อมูลที่ทำให้การค้นหาและเอกสารคิวรีจากตัวฐานข้อมูล โดยการคำนวณความคล้ายเหมือนความคล้ายของทั้ง 2 เอกสารออกมาเป็นตัวเลข ซึ่งถ้าตัวเลขมีค่ามากแสดงว่าเอกสารทั้ง 2 เอกสารนั้นมีความเหมือนความคล้ายมากกว่าเอกสารอื่นที่มีค่าน้อยกว่า

การวิจัยในครั้งนี้จึงมีแนวคิดนำเทคโนโลยีประมวลผลภาษาธรรมชาติซึ่งเป็นแขนงหนึ่งของเทคโนโลยีปัญญาประดิษฐ์มาใช้แก้ปัญหาด้านการศึกษา โดยพัฒนาเป็นระบบตรวจสอบอัตโนมัติภาษาไทยบนพื้นฐานการวัดค่าความคล้ายคลึงกันภายใต้กรอบแนวคิดดังภาพที่ 1



ภาพที่ 1 กรอบแนวคิดในการวิจัย

วิธีดำเนินการวิจัย

1. เครื่องมือการวิจัย

ผู้วิจัยใช้เครื่องมือในการดำเนินงานวิจัยดังนี้

1.1 ด้านการพัฒนาระบบตรวจสอบอัตโนมัติภาษาไทยบนพื้นฐานการวัดค่าความคล้ายคลึง โดยเครื่องมือการพัฒนาดังนี้ (1) การสัมภาษณ์ศึกษาความต้องการระบบ (2) การวิเคราะห์และออกแบบระบบด้วยแผนภาพความสัมพันธ์ข้อมูล (Entity Relationship Diagram) และแผนภาพที่อธิบายความสัมพันธ์ข้อมูลในภาพรวมของระบบ (Unified Modeling Language) (3) การพัฒนาฐานข้อมูลด้วยมายเอสคิวแอล (Mysql) และ (4) การพัฒนาระบบด้วยภาษาไพทอนรองรับการทำงานผ่านเว็บไซต์แอปพลิเคชัน (Python Web Framework) [7] และใช้ไลบรารีจัดการคำและข้อความภาษาไทยด้วยไพไทยเอ็นแอลพี (PyThaiNLP) [4]

1.2 ด้านการประเมินผลการใช้งานระบบ ได้ใช้แบบประเมินความพึงพอใจต่อการใช้งานระบบ ที่มีแบบสอบถามมาตราส่วน 5 ระดับ จากงานวิจัยของ Likert [13] โดยแยกการประเมินออกเป็น 5 ด้าน

ได้แก่ (1) การออกแบบหน้าจอภาพที่เหมาะสม (2) การออกแบบใช้งานง่าย (3) ความสะดวกในการใช้งาน (4) ความเร็วในการแสดงผล และ (5) ประโยชน์และสิ่งที่ได้รับ

2. กลุ่มเป้าหมาย

ผู้ใช้งาน ประกอบด้วยอาจารย์ และนักศึกษา ในหลักสูตรวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยีการเกษตร จำนวน 20 คน

3. ขั้นตอนการดำเนินการวิจัย

การพัฒนาระบบตรวจสอบข้อสอบอัตนัยภาษาไทยบนพื้นฐานการวัดค่าความคล้ายคลึงในครั้งนี ผู้วิจัยได้พัฒนาระบบตามขั้นตอนวงจรการพัฒนาระบบ (System Development Life Cycle: SDLC) ตามแนวทางการวิจัยของ Loo และคณะ [14] ซึ่งมีรายละเอียดดังนี้

3.1 ขั้นตอนการศึกษาความเป็นไปได้

3.1.1 ศึกษาความเป็นไปได้ จากเอกสาร สอบถาม และสัมภาษณ์ อาจารย์ถึงแนวการสร้างข้อสอบและการตรวจข้อสอบอัตนัย และนักศึกษาที่ทำข้อสอบ รวมถึงศึกษาระบบงานเดิมที่มีอยู่

3.1.2 เก็บรวบรวมข้อมูลที่ได้จากเอกสาร สอบถาม และสัมภาษณ์ รวมถึงงานวิจัยที่มีการพัฒนาระบบในลักษณะนี้มาก่อนหน้า

3.1.3 สังเคราะห์รายละเอียด เนื้อหาของความต้องการระบบ งานวิจัยที่เกี่ยวข้อง มาเป็นแนวคิดของระบบใหม่

3.2 ขั้นตอนการวิเคราะห์ระบบ

3.2.1 รวบรวมข้อมูลและความต้องการระบบ มาวิเคราะห์โครงสร้างของระบบและความสัมพันธ์ในการออกแบบตามขั้นตอนต่าง ๆ เพื่อให้สอดคล้องกับระบบงานเดิม และตรงกับความต้องการของระบบ

3.2.2 วิเคราะห์ความต้องการของระบบใหม่จากข้อมูลที่มีการรวบรวมและสังเคราะห์ให้อยู่ในรูปแบบแผนภาพความสัมพันธ์ข้อมูล (Entity Relationship Diagram) และแผนภาพที่อธิบายความสัมพันธ์ข้อมูลในภาพรวมของระบบ (Unified Modeling Language)

3.3 ขั้นตอนการออกแบบระบบ

3.3.1 เตรียมข้อมูล แบ่งคำ และตัดคำไม่สำคัญ โดยใช้ไลบรารี (PyThaiNLP) [8]

3.3.2 คำนวณค่าความถี่ของคำและค่าน้ำหนักของคำที่เกิดขึ้นในคำตอบ (TFIDF) [5]

3.3.3 คำนวณค่าความคล้ายคลึง หรือความเหมือนความคล้าย ระหว่างคำตอบของผู้ออกข้อสอบและคำตอบของผู้ทำข้อสอบในโครงสร้างทางคณิตศาสตร์หรือเวกเตอร์ [6] เป็นการนำค่า Vector ที่ได้จากการแปลงค่า TF-IDF ระหว่างชุดข้อมูลคำตอบของอาจารย์และนักเรียนมาเข้าสมการ ดังภาพที่ 2

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

ภาพที่ 2 สมการความเหมือนความคล้าย

โดย A คือ ค่า TF-IDF คำตอบของอาจารย์ และ B คือ ค่า TF-IDF คำตอบของนักเรียน

3.3.4 กำหนดเกณฑ์ค่าความคล้ายคลึงของคำตอบ พิจารณาจากค่าความถี่ของคำที่ปรากฏในการทดลองเทียบกับระดับคะแนนที่ใช้ประเมินใน มคอ.3 ของรายวิชาในหลักสูตร

3.4 ขั้นตอนการพัฒนาระบบ

พัฒนาฐานข้อมูลด้วยมายเอสคิวแอล (Mysql) และพัฒนาระบบด้วยภาษาไพทอนที่รองรับการทำงานผ่านเว็บไซต์แอปพลิเคชัน (Python Web Framework) ตามรูปแบบโครงสร้างฐานข้อมูลที่ออกแบบไว้ในแผนภาพความสัมพันธ์ของข้อมูล และโครงสร้างภาษาที่ออกแบบไว้ตามลำดับ

3.5 ขั้นตอนการทดสอบและประเมินผลระบบ

ติดตั้งระบบบนเซิร์ฟเวอร์ แล้วทดสอบกับผู้ใช้งานที่เป็นกลุ่มเป้าหมาย โดยกำหนดให้อาจารย์ผู้สอนออกข้อสอบอัตโนมัติ และให้นักศึกษาตอบคำถามดังกล่าว จากนั้นจึงทำการประเมินผลความพึงพอใจในรูปแบบแบบสอบถามออนไลน์ผ่านกูเกิลฟอร์ม (Google Form) โดยแบ่งเป็น 2 ส่วน ได้แก่ ส่วนข้อมูลส่วนตัว และข้อมูลความพึงพอใจพร้อมข้อเสนอแนะ

3.6 ขั้นตอนการปรับปรุงดูแลระบบ

นำข้อเสนอแนะและข้อคิดเห็นของแบบประเมินความพึงพอใจที่ได้มาปรับปรุงระบบ เพื่อให้ใช้งานได้ง่ายและมีประสิทธิภาพมากขึ้น

4. สถิติที่ใช้ในการวิจัย ได้แก่ ค่าเฉลี่ย สำหรับหาค่าเฉลี่ยของระดับประเมินความพึงพอใจต่อการใช้งานระบบ และส่วนเบี่ยงเบนมาตรฐาน ใช้สำหรับวัดค่าการกระจายของข้อมูลที่กลุ่มตัวอย่างทำการประเมินความพึงพอใจ โดยนำผลการประเมินที่ได้เทียบกับเกณฑ์การประเมินของ Likert [12] ดังนี้

ค่าเฉลี่ยเท่ากับ 4.51 – 5.00 หมายความว่า ระดับมากที่สุด

ค่าเฉลี่ยเท่ากับ 3.51 – 4.50 หมายความว่า ระดับมาก

ค่าเฉลี่ยเท่ากับ 2.51 – 3.50 หมายความว่า ระดับปานกลาง

ค่าเฉลี่ยเท่ากับ 1.51 – 2.50 หมายความว่า ระดับน้อย

ค่าเฉลี่ยเท่ากับ 1.01 – 1.50 หมายความว่า ระดับน้อยที่สุด

ผลการวิจัย

1. ผลการพัฒนาระบบตรวจสอบอัตนัยภาษาไทยบนพื้นฐานการวัดค่าความคล้ายคลึง

ผู้วิจัยได้ดำเนินการพัฒนาระบบตรวจสอบอัตนัยภาษาไทยบนพื้นฐานการวัดค่าความคล้ายคลึง

1.1 ผลการวิเคราะห์ข้อมูล

1.1.1 การแสดงขอบเขตการทำงานของผู้ใช้งานแต่ละระดับ โดย

- ผู้ใช้งานประเภทอาจารย์ สามารถสมัครสมาชิก เข้าสู่ระบบ สร้างข้อสอบ และคำตอบ ปรับปรุงแก้ไขคำถามคำตอบ และเรียกดูผลคะแนนได้

- ผู้ใช้งานประเภทนักศึกษา สามารถสมัครสมาชิก เข้าสู่ระบบ ทำข้อสอบ และเรียกดูผลคะแนนได้

- ระบบตรวจสอบข้อสอบอัตโนมัติ สามารถคำนวณความเหมือนความคล้ายระหว่างคำตอบของอาจารย์และนักศึกษา แล้วนำมาคำนวณตามเกณฑ์คะแนนที่กำหนดไว้ และบันทึกลงในฐานข้อมูลได้

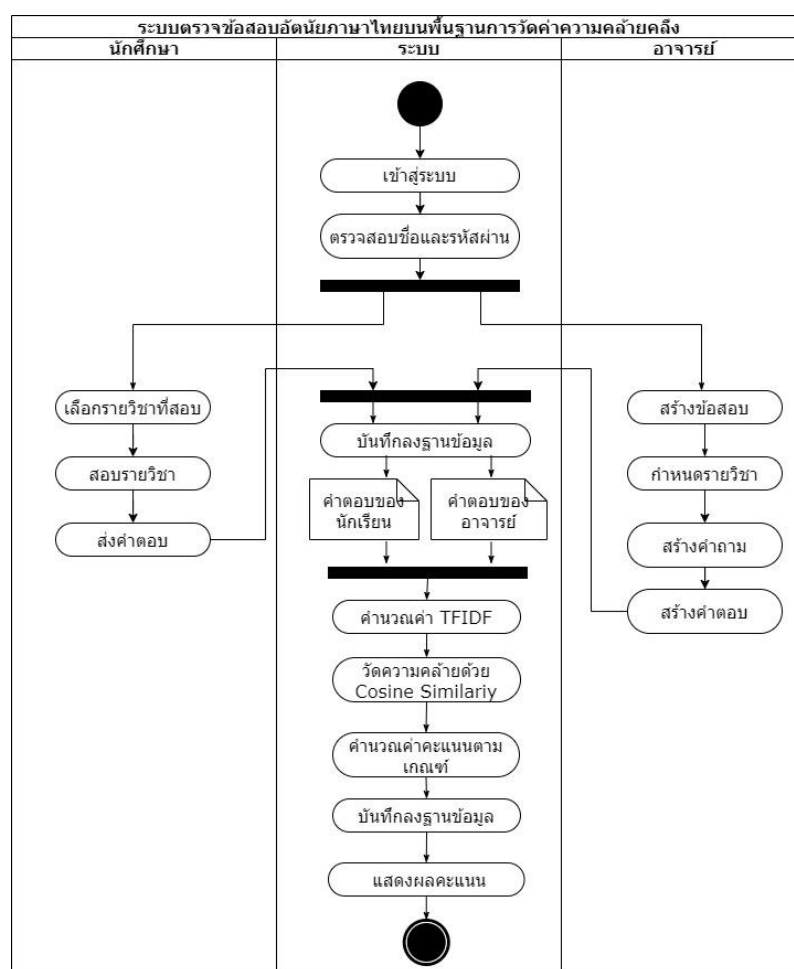
1.1.2 การวิเคราะห์แผนภาพกิจกรรมที่แสดงถึงการทำงานของระบบตามขอบเขตของผู้ใช้งาน สามารถอธิบายถึงขั้นตอนการทำงานของระบบได้ ดังภาพที่ 3

1.1.3 การกำหนดเกณฑ์การให้คะแนนพิจารณาจากความคล้ายคลึง พิจารณาความเหมือนความคล้าย และแบ่งเป็นช่วงคะแนน ตามความถี่ของค่าที่เกิดขึ้นในการทดลองแล้วเทียบเคียงเป็นระดับคะแนนที่ใช้ประเมินใน มคอ.3 ของรายวิชาในหลักสูตร แล้วจึงความเหมือนความคล้ายวัดจากความเหมือนของจำนวนค่าระหว่างคำตอบของอาจารย์และนักศึกษาในรูปแบบของเวกเตอร์ (Vector) และมีช่วงคะแนนความเหมือนความคล้ายจากน้อยที่สุดไปมากที่สุด คือ คะแนน 0 ถึง 5 (คะแนนเต็ม 5) ดังนี้

ค่าความเหมือนความคล้ายเท่ากับ 0	ได้คะแนน 0
ค่าความเหมือนความคล้ายเท่ากับ 1-20	ได้คะแนน 1
ค่าความเหมือนความคล้ายเท่ากับ 21-40	ได้คะแนน 2
ค่าความเหมือนความคล้ายเท่ากับ 41-60	ได้คะแนน 3
ค่าความเหมือนความคล้ายเท่ากับ 61-80	ได้คะแนน 4
ค่าความเหมือนความคล้ายเท่ากับ 81-100	ได้คะแนน 5

1.2 ผลการพัฒนาระบบ

จากผลการวิเคราะห์ข้อมูลนั้น การพัฒนาระบบจะเริ่มต้นจากเข้าสู่ระบบ สมัครงานศึกษา เพื่อบันทึกข้อมูลส่วนตัวของอาจารย์และนักศึกษา จากนั้นอาจารย์สามารถสร้างข้อสอบและคำตอบในรูปแบบอัตโนมัติ ดังภาพที่ 4 ทั้งนี้อาจารย์สามารถแก้ไขคำถามและคำตอบของตนเองได้



ภาพที่ 3 แผนภาพกิจกรรมแสดงการทำงานของระบบ

หลังจากอาจารย์ได้สร้างข้อสอบและคำตอบเรียบร้อยแล้ว นักศึกษาสามารถเข้าสู่ระบบเพื่อทำข้อสอบ ดังภาพที่ 5 ซึ่งสามารถทำข้อสอบตั้งแต่ข้อแรกถึงข้อสุดท้าย ภายใต้หัวข้อของวิชานั้น ๆ เมื่อสิ้นสุดการทำข้อสอบ ระบบทำการวิเคราะห์และคำนวณค่าความคล้ายคลึงระหว่างคำตอบของอาจารย์และนักศึกษา เพื่อให้ นักศึกษาสามารถเรียกดูผลคะแนนของแต่ละข้อ ในขณะที่อาจารย์สามารถเรียกดูคะแนนของ นักศึกษาที่สอบในหัวข้อของตนเองได้ทั้งหมด ดังภาพที่ 6 และ 7 ตามลำดับ จากคะแนนเต็ม 5 ซึ่งกำหนดเกณฑ์การให้คะแนนพิจารณาจากความคล้ายคลึงจากหัวข้อก่อนหน้านี้

สร้างข้อสอบ

ความรู้เบื้องต้นการวิเคราะห์และออกแบบระบบ (UML Dia)

คำถาม

การวิเคราะห์และออกแบบระบบคืออะไร

คำตอบ

การศึกษา รวมรวม ระบบการทำงานปัจจุบัน เพื่อให้ทราบและเข้าใจปัญหา และความต้องการ Requirement ที่มาจากระบบปัจจุบัน แล้วนำมากำหนดแผนหรือวิธีการสร้างระบบขึ้นมาใหม่ โดยผลลัพธ์การทำงานของระบบใหม่ั้น จะต้องสอดคล้องกับความต้องการที่ได้จากการวิเคราะห์ระบบและแก้ปัญหานั้นจากระบบปัจจุบัน

สร้างคำถาม

กลับไปหน้าหลัก

ภาพที่ 4 หน้าจอสร้างคำถามและคำถามของอาจารย์

ข้อ	คำถามครู	คำตอบนักเรียน	ความถี่ (ข้อ 5)	ความถี่ (ข้อ 5)
1	การศึกษา รวมรวม ระบบการทำงานปัจจุบัน เพื่อให้ทราบและเข้าใจปัญหา และความต้องการ Requirement ที่มาจากระบบปัจจุบัน แล้วนำมากำหนดแผนหรือวิธีการสร้างระบบขึ้นมาใหม่ โดยผลลัพธ์การทำงานของระบบใหม่ั้น จะต้องสอดคล้องกับความต้องการที่ได้จากการวิเคราะห์ระบบและแก้ปัญหานั้นจากระบบปัจจุบัน	ข้อใดที่เป็นการศึกษาและออกแบบระบบขึ้นมาใหม่เพื่อให้ระบบมีประสิทธิภาพที่ดีขึ้น	50	3
2	7 ขั้นตอน การวิเคราะห์และออกแบบระบบ 1. ศึกษาและวิเคราะห์ปัญหา 2. กำหนดความต้องการ 3. วิเคราะห์ปัญหา 4. ออกแบบระบบ 5. สร้างแบบจำลอง 6. ตรวจสอบและปรับปรุง 7. นำไปใช้	7 ขั้นตอน 1. ศึกษาและวิเคราะห์ปัญหา 2. กำหนดความต้องการ 3. วิเคราะห์ปัญหา 4. ออกแบบระบบ 5. สร้างแบบจำลอง 6. ตรวจสอบและปรับปรุง 7. นำไปใช้	30	2

ภาพที่ 6 หน้าจอเรียกดูผลคะแนนของนักศึกษา

สอบ

คำถามข้อที่ 1 การวิเคราะห์และออกแบบระบบคืออะไร

การศึกษา รวมรวม ระบบการทำงานปัจจุบัน เพื่อให้ทราบและเข้าใจปัญหา และความต้องการ Requirement ที่มาจากระบบปัจจุบัน แล้วนำมากำหนดแผนหรือวิธีการสร้างระบบขึ้นมาใหม่

ต่อไป

ภาพที่ 5 หน้าจอตอบคำถามของนักศึกษา

ข้อ	คำถามครู	คำตอบนักเรียน	ความถี่ (ข้อ 5)	ความถี่ (ข้อ 5)
1	การศึกษา รวมรวม ระบบการทำงานปัจจุบัน เพื่อให้ทราบและเข้าใจปัญหา และความต้องการ Requirement ที่มาจากระบบปัจจุบัน แล้วนำมากำหนดแผนหรือวิธีการสร้างระบบขึ้นมาใหม่ โดยผลลัพธ์การทำงานของระบบใหม่ั้น จะต้องสอดคล้องกับความต้องการที่ได้จากการวิเคราะห์ระบบและแก้ปัญหานั้นจากระบบปัจจุบัน	ข้อใดที่เป็นการศึกษาและออกแบบระบบขึ้นมาใหม่เพื่อให้ระบบมีประสิทธิภาพที่ดีขึ้น	50	3
2	7 ขั้นตอน การวิเคราะห์และออกแบบระบบ 1. ศึกษาและวิเคราะห์ปัญหา 2. กำหนดความต้องการ 3. วิเคราะห์ปัญหา 4. ออกแบบระบบ 5. สร้างแบบจำลอง 6. ตรวจสอบและปรับปรุง 7. นำไปใช้	7 ขั้นตอน 1. ศึกษาและวิเคราะห์ปัญหา 2. กำหนดความต้องการ 3. วิเคราะห์ปัญหา 4. ออกแบบระบบ 5. สร้างแบบจำลอง 6. ตรวจสอบและปรับปรุง 7. นำไปใช้	45	3

ภาพที่ 7 หน้าจอเรียกดูผลคะแนนของอาจารย์

2. ผลการทดลองใช้ระบบตรวจสอบอัตนัยภาษาไทยบนพื้นฐานการวัดค่าความคล้ายคลึง

ผู้วิจัยได้นำระบบตรวจสอบอัตนัยภาษาไทยบนพื้นฐานการวัดค่าความคล้ายคลึงที่พัฒนาขึ้นให้ผู้เชี่ยวชาญ จำนวน 3 คน ประเมินสอบถามความคิดเห็นที่มีต่อระบบ ซึ่งแสดงดังตารางที่ 1 และหลังจากประเมินจากผู้เชี่ยวชาญจึงนำระบบไปทดลองใช้ กับอาจารย์และนักศึกษาระดับปริญญาตรี หลักสูตรวิทยาการคอมพิวเตอร์ จำนวน 20 คน เพื่อสอบถามความพึงพอใจของอาจารย์และนักศึกษาที่มีต่อการใช้งานระบบ จากนั้นนำผลการทดลองมาวิเคราะห์ด้วยค่าสถิติพื้นฐานเทียบกับเกณฑ์และสรุปผล แสดงดังตารางที่ 2 จากผลการประเมินสอบถามความคิดเห็นของผู้เชี่ยวชาญที่มีต่อระบบ พบว่า ด้านความง่ายต่อการใช้งานระบบ ($\bar{X}$ = 3.84 และ SD. = 0.75) และด้านการรักษาความปลอดภัยของข้อมูล ($\bar{X}$ = 3.96 และ SD. = 0.81) มีค่า ค่าเฉลี่ยน้อยกว่าด้านประสิทธิภาพของการทำงานของระบบ ($\bar{X}$ = 4.30 และ SD. = 0.75) และการทำงานเป็นไปตามฟังก์ชันงาน ($\bar{X}$ = 4.28 และ SD. = 0.83) และโดยรวมของระบบมีค่าเฉลี่ย 4.09 และค่าเบี่ยงเบนมาตรฐานเท่ากับ 0.78 ซึ่งอยู่ในระดับมาก

ตารางที่ 1 ผลการประเมินความคิดเห็นของผู้เชี่ยวชาญที่มีต่อระบบ

รายการ	$\bar{X}$	SD.	ระดับความคิดเห็น
1. ด้านประสิทธิภาพของการทำงานของระบบ			

1.1. ความเร็วในการติดต่อฐานข้อมูล	4.22	0.73	มาก
1.2 ความเร็วในการนำเสนอข้อมูล	4.32	0.81	มาก
1.3 ความเร็วในการทำงานของระบบในภาพรวม	4.35	0.70	มาก
รวม	4.30	0.75	มาก
2. ด้านการทำงานเป็นไปตามฟังก์ชันงาน			
2.1 การเพิ่ม ปรับปรุง และลบข้อมูลได้ถูกต้อง	4.35	0.82	มาก
2.2 การทำงานตามฟังก์ชันได้ถูกต้องในภาพรวม	4.20	0.83	มาก
รวม	4.28	0.83	มาก
3. ด้านความง่ายต่อการใช้งานระบบ			
3.1 ออกแบบหน้าจอ ข้อความ ที่เหมาะสม	3.70	0.72	มาก
3.2 ความสะดวกในการใช้งาน	3.95	0.81	มาก
3.3 ความน่าใช้งานของระบบในภาพรวม	3.86	0.73	มาก
รวม	3.84	0.75	มาก
4. ด้านการรักษาความปลอดภัยของข้อมูลในระบบ			
4.1 การกำหนดสิทธิเข้าถึงระบบของผู้ใช้งาน	3.98	0.83	มาก
4.2 ความปลอดภัยของการเข้าถึงข้อมูล	3.95	0.76	มาก
4.3 การรักษาความปลอดภัยของข้อมูลในภาพรวม	3.95	0.83	มาก
รวม	3.96	0.81	มาก
โดยรวม	4.09	0.78	มาก

ตารางที่ 2 ผลการประเมินความพึงพอใจต่อการใช้งานระบบของผู้ใช้งาน

รายการ	$\bar{X}$	SD.	ระดับความคิดเห็น
1. ด้านความง่ายต่อการใช้งานระบบ			มาก
1.1 การออกแบบหน้าจอ ข้อความ ที่เหมาะสม	3.65	0.86	มาก
1.2 การออกแบบใช้งานง่าย และมีความสะดวก	4.06	0.75	มาก
1.3 ระบบมีความสอดคล้องกับความต้องการของผู้ใช้	4.26	0.83	มาก
1.4 ความถูกต้องของข้อมูลผลลัพธ์และรูปแบบข้อมูล	4.25	0.83	มาก
รวม	4.06	0.82	มาก
2. ด้านประโยชน์ต่อการใช้งานระบบ			
2.1 อาจารย์สามารถจัดการข้อมูล และตรวจสอบข้อมูลได้	3.98	0.82	มาก
2.2 นักศึกษาสามารถจัดการข้อมูล และตรวจสอบข้อมูลได้	4.12	0.76	มาก
2.3 ระบบสามารถนำไปใช้ประโยชน์ต่อการเรียนการสอนได้	4.06	0.83	มาก
รวม	4.05	0.80	มาก
โดยรวม	4.05	0.81	มาก

จากตารางที่ 2 ผลการประเมินความพึงพอใจต่อการใช้งานระบบตรวจสอบอัตลักษณ์ภาษาไทยบนพื้นฐานการวัดค่าความคล้ายคลึง พบว่า มีค่าความพึงพอใจอยู่ในระดับมาก และมีความแตกต่างของระดับความพึงพอใจมีแตกต่างกันน้อยหรือมีความใกล้เคียงกัน โดยค่าเฉลี่ยรวมเท่ากับ 4.05 และค่าเบี่ยงเบนมาตรฐานเท่ากับ 0.81

อภิปรายผลการวิจัย

ผลการดำเนินการวิจัย เรื่องระบบตรวจข้อสอบอัตโนมัติภาษาไทยบนพื้นฐานการวัดค่าความคล้ายคลึง มีวัตถุประสงค์เพื่อต้องการใช้เทคโนโลยีประมวลผลภาษาธรรมชาติมาเพิ่มประสิทธิภาพด้านการศึกษา ลดเวลาการตรวจข้อสอบอัตโนมัติ หรือเป็นแนวทางสนับสนุนให้การพิจารณาคำตอบของนักศึกษาในรูปแบบอัตโนมัติได้อย่างรวดเร็วมากขึ้น การประเมินระบบได้ให้ผู้เชี่ยวชาญเสนอแนะ โดยภาพรวมความง่ายต่อการใช้งาน การออกแบบหน้าจอหรือเว็บไซต์ยังไม่น่าสนใจ อาจเนื่องจากระบบเน้นกระบวนการการทำงานของฟังก์ชันให้ถูกต้องและประสิทธิภาพของฐานข้อมูลเป็นหลัก รวมถึงต้องอาศัยความถูกต้องตามแนวคิดด้านปัญญาประดิษฐ์มาใช้ สอดคล้องกับการรักษาความปลอดภัยของข้อมูล เช่น การออกจากระบบ ที่ต้องปรับปรุงแก้ไข เพื่อรักษาความเป็นส่วนตัวของอาจารย์และนักศึกษาโดยเฉพาะข้อสอบ เป็นต้น นอกจากนี้การวิจัยยังได้รับการประเมินผลความพึงพอใจต่อการใช้งานระบบ โดยใช้อาจารย์และนักศึกษา จำนวน 20 คน ประกอบด้วย ด้านความง่ายต่อการใช้งานระบบและด้านประโยชน์ต่อการใช้งานระบบ มีค่าเฉลี่ยอยู่ในระดับมาก แต่การออกแบบหน้าจอ ข้อความที่เหมาะสมนั้น จะได้ค่าเฉลี่ยน้อยกว่าข้ออื่น เนื่องจากการใช้เทคโนโลยีเว็บที่พัฒนาโดยภาษาไพทอน ต้องใช้เครื่องมืออื่นมาร่วมพัฒนาต่อยอดด้วย จึงจะทำให้หน้าจอกภาพมีความสวยงามยิ่งขึ้น การพัฒนาระบบนี้เป็นไปตามขั้นตอนวงจรการพัฒนากระบวนการ มีการเก็บรวบรวมข้อมูลสอบถามความคิดเห็นของอาจารย์และนักศึกษา โดยมุ่งเน้นการตรวจข้อสอบอัตโนมัติเป็นหลัก การออกแบบจึงเน้นการบันทึกข้อมูลแบบง่าย เริ่มจากสมัครสมาชิกทั้งอาจารย์และนักศึกษา สามารถเข้าถึงการบันทึกข้อสอบอัตโนมัติตามหมวดวิชาที่ได้กำหนดไว้เบื้องต้น อาจารย์สามารถกำหนดข้อสอบและคำตอบ ซึ่งคำตอบต้องมีความยาวไม่เกิน 500 อักขระ (ถ้าผลลัพธ์มีจำนวนคำมาก และมีหลายคำตอบจากนักศึกษา จะค่อนข้างใช้เวลานานในการประมวลผล เพราะต้องมีการแบ่งคำ ตัดคำสำคัญ คำนวณความถี่ และหาค่าความคล้ายคลึง) สำหรับมุมมองเรื่องเกณฑ์การให้คะแนนพิจารณาจากความคล้ายคลึง พบว่า นักศึกษามีผลคะแนนอยู่ในช่วงระดับ 0-3 เนื่องจากนักศึกษาต้องเขียนคำตอบให้ตรงกับคำตอบของอาจารย์มากที่สุด เพราะวิเคราะห์คำที่ตรงกัน ยังไม่ได้วิเคราะห์จากความหมาย (Semantic) จึงเป็นข้อเสนอแนะที่สามารถนำไปพัฒนาต่อยอดได้

การพัฒนาระบบตรวจข้อสอบอัตโนมัติภาษาไทยบนพื้นฐานการวัดค่าความคล้ายคลึง ซึ่งเป็นระบบเว็บไซต์สำหรับช่วยการตรวจข้อสอบอัตโนมัติภาษาไทยออนไลน์ จึงทำให้เพิ่มประสิทธิภาพการตรวจข้อสอบอัตโนมัติ และช่วยการตัดสินใจให้คะแนนของอาจารย์แก่นักศึกษา ซึ่งสอดคล้องนิยามและความหมายของ มาสวี่ มาสดีศรีโชติ [8] ที่การประมวลผลภาษาธรรมชาติเป็นวิธีการที่จะช่วยทำให้คอมพิวเตอร์เข้าใจภาษาธรรมชาติของมนุษย์ที่ไม่มีโครงสร้างที่แน่นอน เนื่องจากภาษาไทยไม่มีเว้นวรรคเมื่อสุดสิ้นประโยค ทำให้การแบ่งคำมักเกิดปัญหาคำกำกวมแบ่งไม่ถูกต้อง เช่น กลุ่มคำประสม เป็นต้น ในขณะที่เดียวกันผลการพัฒนาระบบนี้เมื่อใช้ไลบรารี PyThaiNLP สามารถช่วยให้การตัดคำและแบ่งคำเป็นคำประสมที่มีความหมาย ได้ถูกต้องมากขึ้น เช่นเดียวกับงานวิจัยของ นราธร มั่นสุวรรณ [9] หลังจากได้คำที่ถูกแบ่งเรียบร้อยแล้วจึงมาคำนวณให้คำน้ำหนักคำจากวิธีการถ่วงน้ำหนักคำ (Term Frequency – Inverse Document Frequency: TF-IDF) เช่นเดียวกับงานวิจัยของ ทศนีย์ อุทัยสุริ [10] ที่สามารถใช้หลักการทางด้านสถิติ เพื่อประเมินความสำคัญของคำต่อเอกสาร ความสำคัญของคำเป็นส่วนโดยตรงกับจำนวนครั้งที่คำนั้นปรากฏในเอกสารนั้น แต่ถูกลดความสำคัญโดยความถี่ของคำนั้นในกลุ่มเอกสารทั้งหมด ซึ่งหมายความว่าคำตอบของนักศึกษาจะเน้นคำสำคัญที่เกิดขึ้นในประโยค โดยตัดคำที่เกิดขึ้นปรากฏบ่อย ๆ เช่น กลุ่มคำสันธาน คำบุพบท เป็นต้น เพราะปกติการให้พิจารณาคะแนนของข้อสอบอัตโนมัติจะเน้นคำสำคัญที่เกิดขึ้นในคำตอบของอาจารย์และนักศึกษาเป็นหลัก ทั้งนี้คำน้ำหนักที่ได้ นำมาแปลงรูปทางคณิตศาสตร์หรือเวกเตอร์ เพื่อพิจารณาความเหมือนความคล้ายคลึง สอดคล้องกับงานวิจัยของสายเพชร พงศกร และพฤษดี ศิริแสงตระกูล [11] และไกรศักดิ์ เกษร [12] ที่สามารถวิเคราะห์ทางความเหมือนความคล้ายของเอกสารข้อมูลระหว่างสองเอกสารได้

ข้อเสนอแนะ

1. การพัฒนาเพื่อเพิ่มประสิทธิภาพของระบบ ควรนำเทคนิคการวิเคราะห์เชิงความหมาย (Semantic Analysis) มาประยุกต์ใช้ร่วมกับกระบวนการตรวจข้อสอบ เนื่องจากภาษาไทยมีหลาย ๆ คำ สามารถบ่งบอกความหมายเดียวกัน เช่น กิน รับประทาน เป็นต้น

2. การปรับปรุงให้ระบบดังกล่าวนี้ ให้ทำงานได้จริงบนสภาพแวดล้อมอื่นได้ นอกเหนือการเรียกใช้งานบนตัวเชื่อมโยงของภาษาไพทอน (Python API) เพียงอย่างเดียว เพื่อให้ทำงานได้สะดวกและใช้งานง่ายมากขึ้น

เอกสารอ้างอิง

- [1] SAS Institute Inc. (2562). *การประมวลผลภาษาธรรมชาติ (Natural Language Processing)*. เรียกใช้เมื่อ 31 ตุลาคม 2563 จาก https://www.sas.com/th_th/insights/analytics/what-is-natural-language-processing-nlp.html
- [2] Ferreira-Mello, R., André, M., Pinheiro, A., Costa, E., & Romero, C. (2019). *Text mining in education*. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9(6), e1332.
- [3] เศรษฐชัย ใจอีก และสุศักดิ์ มั่งสิงห์. (2563). ระบบตรวจข้อสอบอัตโนมัติภาษาไทยอัตโนมัติด้วยการสืบค้นเชิงความหมาย. *วารสารเทคโนโลยีสารสนเทศ*, 16(1), 15-23.
- [4] Korakot Chaovavanich, Charin Polpanumas, Arthit Suriyawongkul, Lalita Lowphansirikul, & Pattarawat Chormai Wannaphong Phatthiyaphaibun. (27 June 2016). *PyThaiNLP: Thai Natural Language Processing in Python*. เรียกใช้เมื่อ 1 November 2019 จาก <https://pythainlp.github.io/docs/2.2/>
- [5] Webb G.I. (eds) Sammut C. (2011). TF-IDF. (Boston, MA. Springer, บ.ก.) *Encyclopedia of Machine Learning*.
- [6] Han L. Li B. (2013). Distance Weighted Cosine Similarity Measure for Text Classification. ใน Yin H. et al. (eds), และ Yin H. et al. (eds) (บ.ก.), *Intelligent Data Engineering and Automated Learning – IDEAL 2013, Lecture Notes in Computer Science* (เล่มที่ 8206). Berlin, Heidelberg: Springer.
- [7] Armin Ronacher. (2015). *Flask web development one drop at a time*. เรียกใช้เมื่อ 31 ตุลาคม 2563 จาก <https://flask.palletsprojects.com/>
- [8] มาสวีร์ มาศดิศโรติ. (มกราคม-ธันวาคม 2557). การทำเหมืองความคิดเห็นภาษาไทย. *ศรีปทุมปริทัศน์ ฉบับวิทยาศาสตร์และเทคโนโลยี*, 6(1), 120-128.
- [9] นราธร มั่นสุวรรณ. (2559). *อิทธิพลของภาษาต่อประสิทธิภาพการสื่อสารของธนาคารแห่งประเทศไทย*. กรุงเทพมหานคร: จุฬาลงกรณ์มหาวิทยาลัย.
- [10] ทศนีย์ อุทัยสุริ. (2556). *การสกัดคำสำคัญจากบทคัดย่อภาษาอังกฤษ*. กรุงเทพมหานคร: มหาวิทยาลัยศิลปากร.
- [11] P. Saipech and P. Seresangtakul. (2018). Automatic Thai Subjective Examination using Cosine Similarity. *Proceeding of the 5th International Conference on Advanced Informatics: Concept Theory and Applications (ICAICTA)* (หน้า 214-218). IEEE.
- [12] ไกรศักดิ์ เกษร. (มกราคม-มิถุนายน 2555). การแสดงผลลัพธ์การค้นหาข้อมูล : จุดเล็ก ๆ ที่สำคัญและไม่ควรมองข้ามสำหรับเสิร์ชเอนจิน. *วารสารเทคโนโลยีสารสนเทศ*, 8(1), 62-68.
- [13] Y. B., Loo, W. K., Tham, W. Y., & Tan, S. F. Leau. (2012). Software Development Life Cycle AGILE vs Traditional Approaches. In *Proceeding of International Conference on Information and Network Technology (ICINT 2012)*, 37.
- [14] Rensis Likert. (1932). A Technique for the Measurement of Attitudes. *Archives of Psychology*, 140, 1-55.