

CS 2564/10

รายงานความก้าวหน้า

ระบบการจัดการเอกสารโดยการแบ่งแยกหมวดหมู่อัตโนมัติ

Document management system with automatic classification

โดย

613020224-3 นายปิ่นภัทร วุฒิอาภรณ์

613020595-8 นายรัชชานนท์ ศิริสาร

อาจารย์ที่ปรึกษา : รศ.ดร.ปัญญาพล หอระตะ

รายงานนี้เป็นส่วนหนึ่งของการศึกษาวิชา SC314774 COMPUTER SCIENCE PROJECT I

ภาคเรียน 1 ปีการศึกษา 2564

ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์

มหาวิทยาลัยขอนแก่น

(เดือน พฤศจิกายน พ.ศ. 2564)

ปิ่นธร วุฒิอาภรณ์, รัชชานนท์ ศิริสาร , 2564,

ระบบการจัดการเอกสารโดยการแบ่งแยกหมวดหมู่อัตโนมัติ.โครงการคอมพิวเตอร์

ปริญญาวิทยาศาสตรบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ภาควิชาวิทยาการ

คอมพิวเตอร์คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

อาจารย์ที่ปรึกษา: รศ.ดร.ปัญญาพล หอระตะ

บทคัดย่อ

ในปัจจุบันเอกสารสำคัญของหน่วยงานราชการนั้นส่วนใหญ่อยู่ในรูปแบบของไฟล์และมีมากมายหลากหลาย ซึ่งการค้นหาเอกสารเหล่านี้จำเป็นต้องมีคำสำคัญเพื่อใช้ในการค้นหา เช่น เลขที่เอกสาร ชื่อเรื่อง ฯลฯ ดังนั้นงานวิจัยนี้จึงเสนอวิธีการสกัดเลขที่เอกสารและชื่อเรื่องออกมาจากไฟล์เอกสารเพื่อนำไปเป็นคำสำคัญในการค้นหาโดยอัตโนมัติ เพื่อเพิ่มความสะดวกรวดเร็วในการค้นหาเอกสาร

คำสำคัญ: การสกัดคำออกมาไฟล์เอกสาร, วิเคราะห์พีดีเอฟ, ไอซีอาร์

Puntorn Wutiarporn and Ruchchanon Sirisan. 2021.

Document management system with automatic classification.

Computer project Bachelor of Science degree, Computer Science,
Science, Khon Kaen University.

Thesis Advisor: Associate Professor Punyaphol Horata

Abstract

Nowadays, the most important government documents are in the form of files and there are many different types. Searching for these documents requires keywords such as document number, title, etc. Therefore, this research proposes a method for extracting document numbers and titles from document files to be automatically used as search keywords to increase the convenience of searching for documents.

keyword: Word extraction from document files, Analyzing PDF, OCR

กิตติกรรมประกาศ

ในการดำเนินโครงการครั้งนี้ ผู้จัดทำโครงการได้รับความอนุเคราะห์ และความช่วยเหลือจากบุคคลหลายท่านด้วยกันจึงขอขอบคุณไว้ ณ โอกาสนี้

ขอขอบพระคุณอาจารย์ทุกท่านที่ให้ความรู้และอบรมสั่งสอน

ขอขอบพระคุณอาจารย์ รศ.ดร.ปัญญาพล หอระตะ ที่เป็นที่ปรึกษาโครงการและได้ให้คำชี้แนะแนวทาง ใน การคิดวิเคราะห์ และการพัฒนาระบบของโครงการนี้

ขอขอบคุณ คุณพ่อ คุณแม่ ที่เป็นกำลังใจในการเรียนและการทำโครงการมาโดยตลอด

ขอบคุณเพื่อนๆที่ให้ คำปรึกษาและช่วยเหลือในด้านการวิเคราะห์ระบบ

ผู้จัดทำ

ปณิธร วุฒิอาภรณ์

รัชชานนท์ ศิริสาร

สารบัญ

	หน้า
บทคัดย่อ	ข
ABSTRACT	ค
กิตติกรรมประกาศ	ง
สารบัญ	จ
สารบัญภาพ	ฉ
บทที่ 1	2
1. ความเป็นมาและความสำคัญของปัญหา	2
2. วัตถุประสงค์	2
3. เป้าหมายและขอบเขต	2
4. ประโยชน์ที่คาดว่าจะได้รับ	3
บทที่ 2	4
1. ทฤษฎีที่เกี่ยวข้อง	4
2. งานวิจัยที่เกี่ยวข้อง	7
บทที่ 3	9
1. ขั้นตอนการทำงาน	9
2. เครื่องมือที่ใช้ในการทำโครงงานนี้	9
บทที่ 4	10
1. การวิเคราะห์ระบบ	10
2. การออกแบบระบบ	10
3. การพัฒนาระบบ	13
4. การออกแบบระบบ	18
บทที่ 5	30
1. สรุปผลการดำเนินโครงงาน	30
2. ข้อจำกัดของระบบ	30
3. ปัญหาอุปสรรค และ แนวทางแก้ไข	30
4. ข้อเสนอแนะ ในการพัฒนาต่อไป	30
เอกสารอ้างอิง	31
ประวัติผู้วิจัย	32

สารบัญภาพ

	หน้า
ภาพที่ 1 ขั้นตอนการทำงานของระบบค้นหาเอกสาร	10
ภาพที่ 2 ขั้นตอนการทำงานของระบบนำเข้าเอกสารโดยรวม	11
ภาพที่ 3 โครงสร้างการทำงานของระบบนำเข้าเอกสาร	12
ภาพที่ 4 Use case diagram ของเว็บไซต์	13
ภาพที่ 5 Use case diagram ของระบบนำเข้าเอกสาร	16
ภาพที่ 6 ตัวอย่างเอกสาร	18
ภาพที่ 7 หน้าลือคอิน	19
ภาพที่ 8 หน้าหลักของหน้าค้นหา	20
ภาพที่ 9 หน้าค้นของหมวดด้านบริการและธุรการ(ทั่วไป)(1)	21
ภาพที่ 10 หน้าแสดงรายละเอียดของเอกสาร(1)	22
ภาพที่ 11 หน้าค้นของหมวดด้านบริการและธุรการ(ทั่วไป)(2)	23
ภาพที่ 12 หน้าแสดงรายละเอียดของเอกสาร(2)	24
ภาพที่ 13 หน้าแสดงข้อมูลของเอกสาร	25
ภาพที่ 14 หน้ารับเข้าเอกสาร	26
ภาพที่ 15 หน้าใส่ข้อมูลเกี่ยวกับเอกสารรับเข้า	27
ภาพที่ 16 หน้าส่งออกเอกสาร	28
ภาพที่ 17 หน้าใส่ข้อมูลเกี่ยวกับเอกสารส่งออก	29

สารบัญตาราง

	หน้า
ตารางที่ 1 รายละเอียดของ Use case login	13
ตารางที่ 2 รายละเอียดของ Use case search doc	14
ตารางที่ 3 รายละเอียดของ Use case upload doc	14
ตารางที่ 4 รายละเอียดของ Use case download doc	14
ตารางที่ 5 รายละเอียดของ Use case manage user	15
ตารางที่ 6 รายละเอียดของ Use case manage doc	15
ตารางที่ 7 รายละเอียดของ Use case manage doc	15
ตารางที่ 8 รายละเอียดของ Use case Input document	16
ตารางที่ 9 รายละเอียดของ Use case Input document	17
ตารางที่ 10 รายละเอียดของ Use case Edit and confirm	17

บทที่ 1

บทนำ

1. ความเป็นมาและความสำคัญของปัญหา

เนื่องจากปัจจุบันเอกสารราชการมีการจัดเก็บในรูปแบบไฟล์เอกสารจำนวนมาก การค้นหาเอกสารเหล่านั้นแม้ว่าจะมีการกำหนดหมายเลขเอกสารซึ่งมีการจำแนกเป็นหมวดหมู่อยู่แล้ว แต่อย่างไรก็ตามในกรณีที่เอกสารจำนวนมาก ก็ยังต้องใช้แรงงานมนุษย์เป็นจำนวนมากและยังต้องใช้เวลามากด้วยเช่นกัน ดังนั้นโครงการนี้มีแนวคิดที่จะพัฒนาระบบเพื่อสแกนหาหมายเลขเอกสาร จากนั้นสกัดออกมาจากเอกสาร เพื่อนำมาใช้ในการค้นหาเอกสาร นอกจากนั้นยังมีการพัฒนาระบบสืบค้นตามหมวดหมู่ของระบบสารบรรณหนังสือราชการไทย

2. วัตถุประสงค์

- 2.1. สร้างระบบจัดกลุ่มคำสำคัญโดยอัตโนมัติจากไฟล์PDF เพื่อการค้นหาเอกสาร
- 2.2. เพื่อพัฒนาระบบรู้จำเลขเอกสารและระบบการค้นหาเอกสารโดยอัตโนมัติ
- 2.3 เพื่อช่วยให้การจัดการและจัดเก็บเอกสารมีประสิทธิภาพมากขึ้น
- 2.4 เพื่อช่วยให้การสืบค้นข้อมูลเป็นไปอย่างรวดเร็ว

3. เป้าหมายและขอบเขต

3.1 ส่วนของเว็บไซต์

พัฒนาเว็บไซต์ระบบเอกสารที่ง่ายต่อการใช้งาน พร้อมระบบโอซีอาร์ที่สามารถอ่านไฟล์ภาพหรือไฟล์พีดีเอฟ และแปลงให้อยู่ในรูปแบบของข้อความเพื่อหาคีย์เวิร์ดต่างๆที่ต้องการมาใช้เป็นข้อมูลกำกับของไฟล์นั้นๆ เช่น ชื่อเรื่อง เลขเอกสาร วันเวลา จากนั้นจัดเก็บลงฐานข้อมูล

3.2 ส่วนของผู้ใช้งานเว็บไซต์

ผู้ใช้งานสามารถใช้งานเว็บไซต์ได้อย่างมีประสิทธิภาพ และเพิ่มความสะดวกสบายให้ผู้ใช้งาน เช่น ดึงคีย์เวิร์ดต่างๆออกมาใส่ในช่องอินพุตข้อมูลของไฟล์ให้ผู้ใช้อัตโนมัติ เพื่อให้ผู้ใช้ตรวจสอบความถูกต้องและยืนยันเพื่อบันทึกลงฐานข้อมูล

3.3 หมวดหมู่ของเอกสาร

- 3.3.1 ด้านบริหารและธุรการ(ทั่วไป)
- 3.3.2 ด้านการเงิน
- 3.3.3 ด้านพัสดุ
- 3.3.4 ด้านวิจัย

3.3.5 ด้านแผนยุทธศาสตร์

3.3.6 ด้านทรัพยากรบุคคล

3.3.7 ด้านบริการการศึกษา

3.3.8 ด้านอาคารสถานที่

3.3.9 สาขาวิชาต่างๆ

3.3.10 คำสั่ง ประกาศ

3.3.11 ระเบียบ ข้อบังคับ

3.4 ประเภทของเอกสารนำเข้า

3.4.1 เอกสารรับเข้า

3.4.2 เอกสารส่งออก

4. ประโยชน์ที่คาดว่าจะได้รับ

4.1 เพื่อจัดการเอกสารได้อย่างมีประสิทธิภาพมากยิ่งขึ้น

4.2 เพื่ออำนวยความสะดวกให้ผู้ใช้งาน

บทที่ 2

งานวิจัยและทฤษฎีที่เกี่ยวข้อง

1. ทฤษฎีที่เกี่ยวข้อง

1. OCR (โอซีอาร์)

โอซีอาร์ คือการแปลงไฟล์ภาพเอกสาร ให้เป็นไฟล์ข้อความโดยอัตโนมัติ ทำให้ไม่ต้องเสียเวลาพิมพ์งาน

โครงสร้างทั่วไปของระบบโอซีอาร์

1. ขบวนการประมวลผลขั้นต้น (Pre-Processing)

- 1.1 การกรองข้อมูลแทรกซ้อน (Noise Filtering)
- 1.2 การปรับแต่งข้อมูล (Normalization)
- 1.3 การตัดแบ่งพื้นที่ใช้งาน (Cropping)
- 1.4 การสกัดลักษณะสำคัญ (Feature Extraction)

2. การรู้จำ (Recognition)

- 2.1 วิธีทางการเข้าคู่รูปแบบ (Template Matching)
- 2.2 วิธีทางสถิติ (Statistical Approach)
- 2.3 วิธีการวิเคราะห์ทางโครงสร้าง (Structural Analysis)
- 2.4 วิธีทางโครงข่ายประสาทเทียม (Neural Network)

3. ขบวนการประมวลผลขั้นปลาย (Post-Processing)

ประโยชน์ที่ได้รับจากการใช้โปรแกรม OCR

1. ประหยัดพื้นที่จัดเก็บข้อมูล เนื่องจากไฟล์ข้อความมีขนาดเล็กกว่าไฟล์ภาพมาก
2. สะดวกในการปรับแต่งและแก้ไขเอกสาร เนื่องจากไฟล์ข้อความสามารถปรับแต่งและแก้ไขได้ง่ายกว่าไฟล์ภาพ

วิธีการทำงานของโอซีอาร์

ในส่วนนี้จะกล่าวถึงเฉพาะในส่วนของการประมวลผลก่อนการรู้จำตัวอักษรเท่านั้น โดยเริ่มจากการนำข้อมูลเข้าด้วยการสแกนภาพเอกสารที่เป็นตัวอักษร โดยเรากำหนดค่าเริ่มต้นของการสแกนรูปให้เป็นขาวดำ และทำการเก็บภาพที่ได้จากการสแกนจากนั้นจึงนำเอกสารที่ได้มาทำการรู้จำตัวอักษรต่อไป ซึ่งในกระบวนการนี้จะประกอบด้วย การประมวลผลภาพขั้นต้น และการประมวลผลขั้นสุดท้าย จากนั้นเราจะได้ข้อความที่ผ่านระบบรู้จำตัวอักษรซึ่งเป็นข้อความที่เราสามารถนำไปใช้จริง

โดยทั่วไปกระบวนการรู้จำตัวอักษรที่มีรูปแบบคือ การนำเข้าของข้อมูล การประมวลผลภาพ เบื้องต้นจะมีการแยกลักษณะเฉพาะของตัวอักษรและการประมวลผลภาพ ขั้นสุดท้ายจะนำตัวอักษรที่แยกได้นั้นมาทำการรู้จำตัวอักษรเพื่อให้ได้ผลลัพธ์ออกมา

2. รูปแบบการนิยามและระบุข้อมูลชนิดสตริง (Regular Expressions)

รูปแบบการนิยามและระบุข้อมูลชนิดสตริงคือลำดับอักขรรูปแบบที่ใช้ในการค้นหาข้อความในสตริงโดยสามารถใช้ค้นหาอย่างเดียวก่อนหรือค้นหาแล้วแทนที่ด้วยข้อความที่กำหนดก็ได้ หรือใช้ตรวจสอบการรับค่าของสตริงว่าเป็นไปตามรูปแบบที่กำหนดหรือไม่ การระบุข้อมูลชนิดสตริงเป็นที่นิยมมาก เนื่องจากแก้ปัญหาที่ทุกคนต้องเจอ ช้า ๆ อย่างมีประสิทธิภาพ เราจะพบการระบุข้อมูลชนิดสตริง ได้ในทุกภาษาโปรแกรมมิ่งชั้นนำ แต่รูปแบบของการระบุข้อมูลชนิดสตริง ที่ใช้อาจจะต่างกันเล็กน้อย

3. การเรียนรู้เชิงลึก (Deep Learning)

การเรียนรู้เชิงลึกคือวิธีการเรียนรู้แบบอัตโนมัติด้วยการ เลียนแบบการทำงานของโครงข่ายประสาทของมนุษย์โดยนำระบบโครงข่ายประสาทมาซ้อนกัน หลายชั้นและทำการเรียนรู้ข้อมูลตัวอย่าง ซึ่งข้อมูล ดังกล่าวจะถูกนำไปใช้ในการตรวจจับรูปแบบหรือจัดหมวดหมู่ข้อมูล

ข้อได้เปรียบสำคัญของการเรียนรู้เชิงลึก เมื่อเปรียบเทียบกับ การเรียนรู้ของเครื่องจักร (Machine Learning) รูปแบบอื่นๆ คือ

1. ไม่จำเป็นต้องจัดโครงสร้างข้อมูล
2. ไม่จำเป็นต้องจัดหมวดหมู่ข้อมูล
3. ไม่จำเป็นต้องกำหนดการจับคู่ข้อมูลล่วงหน้า

ข้อเสียของการเรียนรู้เชิงลึก เมื่อเปรียบเทียบกับ การเรียนรู้ของเครื่องจักรรูปแบบอื่นๆ คือ

1. ต้องการข้อมูลจำนวนมาก
2. กลไกการทำงานที่ไม่สามารถอธิบายได้

หลักการทำงานของการเรียนรู้เชิงลึก

การเรียนรู้เชิงลึกคือการจำลองรูปแบบการประมวลผลของสมองมนุษย์ โดยใช้โครงข่ายคล้ายเซลล์ประสาทในการประมวลผล เมื่อได้รับข้อมูลมาการเรียนรู้เชิงลึกจะทำการแบ่งแยกข้อมูลและรายละเอียดต่างๆ ที่ได้รับมาทั้งหมด แล้วนำมาประมวลผลหาจุดเด่นและจุดแตกต่างของข้อมูลในเชิงลึก คล้ายกับการกรองข้อมูลเป็นชั้นๆ แล้วสรุปผลข้อมูลออกมาเป็นผลลัพธ์และตรวจสอบว่าข้อมูลนั้นส่งผลอย่างไร ผิด หรือถูก

เช่น มีข้อมูลสัตว์ 1 ตัวที่ไม่ทราบว่าจะเป็นอะไรการเรียนรู้เชิงลึกจะทำการตรวจสอบและคาดการณ์ว่า เป็นสัตว์ชนิดนี้ใช่หรือไม่ โดยไม่จำเป็นต้องระบุว่ามียีคหรือมีหางการเรียนรู้เชิงลึกจะทำการพยากรณ์ผลลัพธ์ออกมา

หากการเรียนรู้เชิงลึกพยากรณ์ไม่ถูกต้อง ตัวมันจะเรียนรู้และปรับเปลี่ยนการประมวลผล เพื่อให้ผลลัพธ์ที่ออกมามีความถูกต้องมากขึ้น และยังการเรียนรู้เชิงลึกได้เรียนรู้มากขึ้นเพียงใด การเรียนรู้เชิงลึกก็จะเข้าใจได้มากขึ้น และลงลึกในรายละเอียดีย่อยได้มากขึ้น จนสามารถสังเกตความแตกต่างของข้อมูลได้แม้เพียงเล็กน้อยก็ตาม โดยที่มนุษย์ไม่จำเป็นต้องแนะนำ

4. MSER (Maximally Stable Extremal Regions)

คืออัลกอริทึมเอนเอสอีอาร์ถูกใช้ในการตรวจจับข้อความโดยการรวมเอนเอสอีอาร์เข้ากับอัลกอริทึมสำหรับตรวจจับมุมของภาพ (Canny Edge Detection) ใช้เพื่อช่วยรับมือกับจุดอ่อนของเอนเอสอีอาร์ ที่จะเบลอเอนเอสอีอาร์ถูกนำไปใช้กับรูปภาพที่เป็นปัญหาก่อนเพื่อกำหนดขอบเขตของอักขระ เพื่อปรับปรุงขอบเขตเอนเอสอีอาร์พิกเซลใด ๆ ที่อยู่นอกขอบเขตที่เกิดจากขอบอัลกอริทึมสำหรับตรวจจับมุมของภาพจะถูกลบออก การแยกขอบที่จัดเตรียมไว้ในภายหลังช่วยเพิ่มความสามารถในการใช้งานเอนเอสอีอาร์ ในการแยกข้อความเบลอ การใช้เอนเอสอีอาร์ ทางเลือกในการตรวจจับข้อความคือการทำงานของ Shi โดยใช้แบบจำลองกราฟ วิธีนี้ใช้เอนเอสอีอาร์ กับอิมเมจอีกครั้งเพื่อสร้างขอบเขตเบื้องต้น จากนั้นจะใช้เพื่อสร้างแบบจำลองกราฟตามระยะตำแหน่งและระยะสีระหว่าง เอนเอสอีอาร์ แต่ละตัวซึ่งถือว่าเป็นโหนด ถัดไปโหนดจะถูกแยกออกเป็นส่วนหน้าและพื้นหลังโดยใช้ฟังก์ชันต้นทุน ฟังก์ชันต้นทุนอย่างหนึ่งคือการกำหนดระยะห่างจากโหนดไปยังพื้นหน้าและพื้นหลัง อีกโหนดที่แตกต่างจากเพื่อนบ้านอย่างมีนัยสำคัญ เมื่อย่อขนาดแล้วกราฟจะถูกตัดเพื่อแยกโหนดข้อความออกจากโหนดที่ไม่ใช่ข้อความ เพื่อเปิดใช้งานการตรวจจับข้อความในฉากทั่วไปซึ่งใช้อัลกอริทึมเอนเอสอีอาร์ในการคาดการณ์ที่หลากหลาย นอกเหนือจากการฉายภาพความเข้มระดับสีเทาแล้วเขายังใช้ช่องสีแดงสีน้ำเงินและสีเขียวเพื่อตรวจจับบริเวณข้อความที่มีสีแตกต่างกัน แต่ไม่จำเป็นต้องแตกต่างกันในความเข้มของระดับสีเทา วิธีนี้ช่วยให้ตรวจจับข้อความได้มากกว่าการใช้ฟังก์ชัน เอนเอสอีอาร์พลัสและเอนเอสอีอาร์ ที่กล่าวถึงข้างต้นเพียงอย่างเดียว

หลักการทำงานของ MSER

อัลกอริทึมดั้งเดิมของเอนเอสอีอาร์คือ $O(n \log(\log(n)))$ ในจำนวน n พิกเซล ดำเนินการโดยการจัดเรียงพิกเซลตามความเข้มก่อน สิ่งนี้จะใช้เวลา $O(n)$ เวลาโดยใช้อัลกอริทึมการเรียงลำดับที่ทำงานโดยการกระจายองค์ประกอบของอาร์เรย์ (binsort) หลังจากการเรียงลำดับพิกเซลจะถูกทำเครื่องหมายในภาพและรายการของส่วนประกอบที่เชื่อมต่อที่เพิ่มขึ้น และการรวมเข้าด้วยกันและพื้นที่จะถูกดูแลโดยใช้ หาสภาพ อัลกอริทึม สิ่งนี้จะใช้เวลา $O(n \log(\log(n)))$ เวลา. ในทางปฏิบัติขั้นตอนเหล่านี้เร็วมาก ในระหว่างกระบวนการนี้พื้นที่ของแต่ละองค์ประกอบที่เชื่อมต่อเป็นฟังก์ชันของความเข้มจะถูกจัดเก็บไว้เพื่อสร้างโครงสร้างข้อมูลการรวมกันของสององค์ประกอบถูกมองว่าเป็นการยุติการมีอยู่ของส่วนประกอบที่เล็กกว่าและการแทรกพิกเซลทั้งหมดของส่วนประกอบที่เล็กกว่าลงในส่วนที่ใหญ่กว่า ในบริเวณสุดขีดค่าที่ 'เสถียรสูงสุด' คือค่าที่สอดคล้องกับกรณีประตู่ที่พื้นที่สัมพัทธ์เปลี่ยนไปเนื่องจากฟังก์ชันของการเปลี่ยนแปลงของเกณฑ์สัมพัทธ์เป็นค่าต่ำสุดในพื้นที่กล่าวคือเอนเอสอีอาร์ซึ่งเป็นส่วนของภาพการทำภาพขาวดำสองระดับ (Binarization) ในพื้นที่ที่มีความเสถียรมากกว่า เกณฑ์ที่หลากหลาย

2. งานวิจัยที่เกี่ยวข้อง

จากการสรุปผลการค้นคว้า งานวิจัยเรื่อง ระบบสารสนเทศสำหรับการจัดการอุตสาหกรรม [1] งานวิจัยนี้นำเสนอวิธีการแปลงข้อมูลการผลิตรายปี โดยแปลงภาพตัวอักษร ให้เป็นข้อมูลตัวอักษร ปรากฏว่าการแปลงข้อมูลจากรูปนั้นให้เป็นตัวอักษรมีทั้งข้อดีและข้อเสียที่เป็นอุปสรรคที่เกิดขึ้นดังนี้

1. กรณีต้นฉบับเป็นลายมือเขียน ยังไม่สามารถแปลงเอกสารเป็นตัวหนังสือได้
2. ต้นฉบับมีสิ่งรบกวน ที่ทำให้โฟล์มีความไม่สมบูรณ์ ก็อาจจะทำให้เกิดข้อผิดพลาดได้
3. เอกสารที่ผ่านการสแกนที่เอียงนั้นทำให้เกิดข้อผิดพลาดในการสกัดคำได้
4. ต้นฉบับตัวอักษรติดกันเกินไปอาจทำให้การสกัดคำผิดพลาดได้

แต่อย่างไรก็ตาม งานวิจัยที่มี การใช้การสกัดคำนั้นในปัจจุบันมีประสิทธิภาพในการทำงานสูงและขั้นตอนการทำงานที่รวดเร็ว ทำให้เกิดความสะดวต่อการทำงานในหลายๆด้าน จากงานวิจัยเรื่อง การเพิ่มประสิทธิภาพการรู้จำอักขระภาษาไทยด้วยแสงโดยใช้เทคนิคเปรียบเทียบสายอักขระโดยประมาณและความแตกต่างของลำดับอักขระ

[2] งานวิจัยนี้นำเสนอ การเพิ่มประสิทธิภาพการทำงานของงานของการสกัดคำ สรุปได้ว่า ระบบรู้จำตัวอักขระด้วยแสงที่เหมาะสมกับการทดลองที่สุดคือ การสกัดคำแบบเอพีไอ และพื้นที่หลังที่ให้ประสิทธิภาพในการทดลองดีที่สุดคือ สีขาวและสีของตัวอักษรที่ได้ผลที่สุดคือสีดำ ความละเอียดของภาพขั้นต่ำคือ 13 พิกเซล และมุมกล้องที่ดีที่สุดคือ 90 องศา เมื่อนำระบบที่พัฒนาขึ้นมาเปรียบเทียบกับลายมือเขียน โดยใช้ประชากรกลุ่มตัวอย่างเป็นเพศหญิงมากกว่าเพศชาย ส่วนมากอยู่ในระดับการศึกษา มัธยมศึกษา การทดลองเปรียบเทียบระบบที่ผู้วิจัยได้ทำการพัฒนาโดยการนำเอาเทคนิค เอเอสเอ็ม และ ดีซีโอ มาประยุกต์ใช้ร่วมกับระบบรู้จำตัวอักษรด้วยแสง ผลการทดลองคือสามารถเพิ่มประสิทธิภาพให้กับผลลัพธ์มากสำหรับการใช้ระบบรู้จำด้วยแสงเพียงอย่างเดียว จากงานวิจัยเรื่อง การพัฒนาวิธีการฮิวริสติกเพื่อเพิ่มประสิทธิภาพการรู้จำลายมือชื่อ [3] งานวิจัยนี้นำเสนอ แนวทางในการเพิ่มประสิทธิภาพระบบไบโอเมตริกซ์สำหรับระบุตัวตนบุคคลด้วยลายมือ ผลสรุปของงานวิจัยนี้การใช้ภาพลายมือที่ยังไม่ผ่านกระบวนการใช้ฮิวริสติก พบว่าอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนที่ใช้ฟังก์ชันพหุนามเป็นเคอร์เนล ให้ค่าแม่นยำถึง 95.33% และนอกจากนั้นยังสังเกตเห็นได้ว่าอัลกอริทึมเคเนียร์เรสเนเบอร์ในการทดลองนี้ให้ค่าความแม่นยำที่มีความสัมพันธ์กับค่า K อย่างมีนัยสำคัญโดยความสัมพันธ์ที่ได้มีลักษณะการแปรผันแบบผกผัน นั่นคือ K ที่มีค่าความแม่นยำในการรู้จำลายมือชื่อสูงกว่า K ที่มีค่ามาก

สรุปผลและแนวทางการนำไปใช้

สรุปผลการศึกษาค้นคว้างานวิจัยพบว่า ในงานวิจัยเรื่อง ระบบสารสนเทศสำหรับจัดการอุตสาหกรรม [1] การทำการรู้จำตัวยังมีความผิดพลาดได้ขึ้นอยู่กับปัจจัยต่างที่ทำให้การสกัดคำมีความผิดพลาดได้ ดังนั้นในงานงานวิจัยเรื่อง การเพิ่มประสิทธิภาพการรู้จำอักขระภาษาไทยด้วยแสงโดยใช้เทคนิคเปรียบเทียบสายอักขระโดยประมาณและความแตกต่างของลำดับอักขระ [2] นั้นจะสามารถนำมาช่วยเพิ่มประสิทธิภาพในการรู้จำได้โดยการทำงานต้องมีข้อจำกัดที่มากขึ้น และงานวิจัยเรื่อง การพัฒนาวิธีการฮิวริสติกเพื่อเพิ่มประสิทธิภาพการรู้จำลายมือชื่อ [3] นั้นจะสามารถนำไปพัฒนาการรู้จำลายมือให้มีประสิทธิภาพมากขึ้นเพื่อทำให้การรู้จำเกิดความถูกต้องสูงสุด

บทที่ 3

วิธีการดำเนินงาน

1. ขั้นตอนการทำงาน

1.1 ค้นคว้าศึกษาเอกสารราชการ

ศึกษารูปแบบและความหมายของเอกสารราชการในแบบต่างๆ เพื่อให้เข้าใจ และสามารถแยกหมวดหมู่ได้

1.2 การเก็บรวบรวมข้อมูล

รวบรวมข้อมูลเอกสารราชการที่เกี่ยวข้องทั้งหมด

1.3 สร้างโครงสร้างของเอกสารราชการ

สร้างในส่วนของโครงสร้าง และที่จัดเก็บเอกสารต่างๆ

1.4 สกัดคำจากเอกสารราชการ

สกัดคำออกจากเอกสารราชการซึ่งเอกสารราชการนั้นอยู่ในรูปแบบของไฟล์ PDF

1.5 สร้างเว็บไซต์

สร้างในส่วนของเว็บไซต์และวางองค์ประกอบต่างๆโดยรวม เพื่อให้ผู้ใช้สามารถใช้ในการค้นหาเอกสารราชการ

1.6 พัฒนาเว็บไซต์

เชื่อมในส่วนของเว็บไซต์กับโครงสร้างเอกสาร และปรับปรุงจนกว่าจะสามารถนำคีย์เวิร์ดที่ผู้ใช้ป้อน ไปดึงเอกสารมาแสดงได้อย่างถูกต้อง จากนั้นตกแต่งเว็บไซต์ให้มีความสวยงามน่าใช้งาน

2. เครื่องมือที่ใช้ในการทำโครงงานนี้

- 2.1 Visual Studio Code
- 2.2 AWS Amazon Web Services
- 2.3 AWS S3
- 2.4 AWS Lambda
- 2.5 Tesseract
- 2.6 MongoDB
- 2.7 React Framework
- 2.8 Regex

บทที่ 4

การวิเคราะห์ระบบ และพัฒนาโปรแกรมวิจัย

1. การวิเคราะห์ระบบ

ระบบวิเคราะห์ข้อความที่ได้จากเอกสารที่นำเข้าสู่ระบบโดยจะจัดเก็บไว้เพื่อใช้ในการค้นหาเอกสารนั้นๆ ตัวอย่างประเภทของคำค้นหา จะมี ชื่อเรื่อง ประเภท เลขที่คำสั่ง เลขที่หนังสือและเดือนปีที่เขียน เพื่อใช้ในการค้นหาเอกสาร การนำเอกสารเข้าสู่ระบบนั้นจะอัปโหลดข้อมูลขึ้นบนเว็บ แล้วระบบจะทำการโอซีอาร์ข้อความออกมาเป็นข้อความและจะจัดเก็บเข้าระบบ

2. การออกแบบระบบ

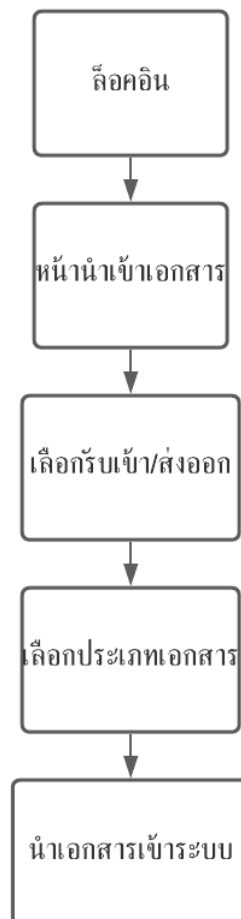
2.1 โครงสร้างของระบบ



ภาพที่ 1 ขั้นตอนการทำงานของระบบค้นหาเอกสาร

จาก ภาพที่ 1 ขั้นตอนการทำงานของระบบค้นหาจะเริ่มจากการเข้าสู่ระบบหลังจากนั้นจะแสดงหน้าค้นหาหลักสามารถเลือกหมวดหมู่ของเอกสารได้ เมื่อเลือกหมวดหมู่ของเอกสารแล้วจะ

สามารถใส่คำสำคัญเพื่อหาเอกสารนั้นๆได้ เมื่อค้นหาแล้วจะสามารถเข้าไปดูรายละเอียดของเอกสารได้และสามารถดาวน์โหลดเอกสารเหล่านั้น



ภาพที่ 2 ขั้นตอนการทำงานของระบบนำเข้าเอกสารโดยรวม

จาก ภาพที่ 2 ขั้นตอนการทำงานของระบบนำเข้าเอกสารหลังจากเข้าสู่ระบบแล้วจะสามารถเลือกประเภทของเอกสารที่นำเข้าได้ นั่นคือ เอกสารรับเข้าหรือเอกสารส่งออก เมื่อเลือกประเภทแล้วเลือกประเภทของหมวดหมู่เอกสารที่ต้องการจะนำเข้า ต่อมาการเลือกไฟล์เอกสารที่ต้องการนำเข้าสู่ระบบ



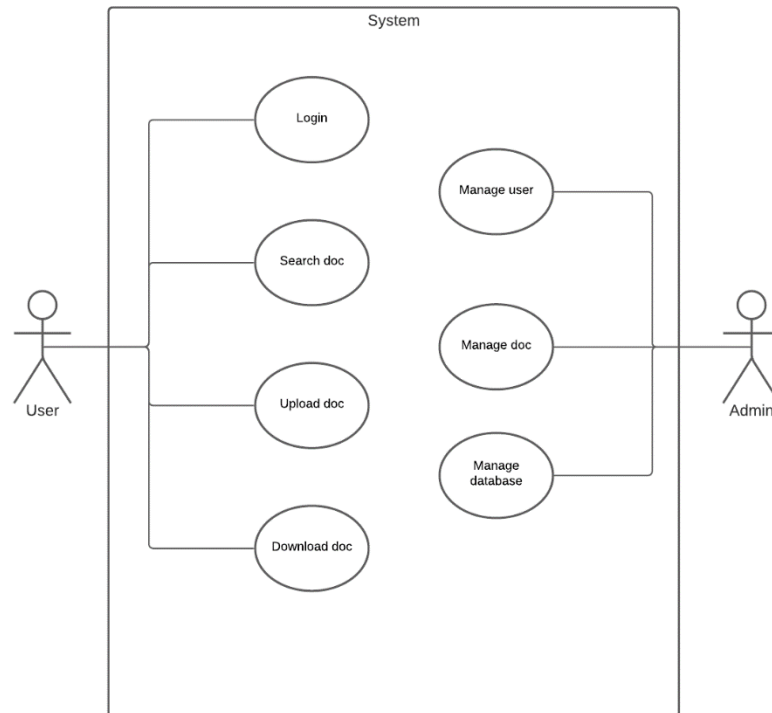
ภาพที่ 3 โครงสร้างการทำงานของระบบนำเข้าเอกสาร

จาก ภาพที่ 3 เป็นโครงสร้างของการทำงานของระบบนำเข้าเอกสาร โดยระบบจะเริ่มจากการที่ผู้ใช้งานเอกสารเข้าสู่ระบบผ่านเว็บไซต์ของระบบ หลังจากนั้นระบบจะแปลงไฟล์เอกสารเป็นไฟล์รูปภาพ หลังจากนั้นจะนำไฟล์รูปผ่านแปลงเป็นไฟล์ เบสหกสิบสี่ เพื่อที่จะได้สามารถอัปโหลดไฟล์รูปเข้าสู่ เอสสาม ได้ หลังจากนั้นระบบจะนำไฟล์รูปที่อัปโหลดไปทำการสกัดคำจากไฟล์รูปภาพ หลังจากได้ข้อความทั้งหมดแล้วระบบจะทำการ ใช้ฟังก์ชัน รูปแบบการนิยามและระบุข้อมูลชนิดสตริง เพื่อนำเฉพาะข้อความที่ต้องการมาส่งกลับไปหน้าเว็บ หลังจากนั้นผู้จะสามารถตรวจสอบความถูกต้องของข้อความที่สกัดออกมาและแก้ไขให้ถูกต้องได้ หลังจากแก้ไขแล้วกดบันทึกเข้าสู่ระบบ ระบบจะทำการผูกข้อความเหล่านั้นไว้กับไฟล์ที่นำเข้าสู่ระบบเพื่อที่จะได้ใช้เป็นคำสำคัญสำหรับค้นหาไฟล์นั้น

3. การพัฒนาระบบ

3.1 System design / Architecture design

3.1.1 ออกแบบ Use case diagram



ภาพที่ 4 Use case diagram ของเว็บไซต์

รายละเอียดของ Use case diagram

ตารางที่ 1 รายละเอียดของ Use case login

Use case name:	Login	
Actor:	ผู้ใช้งาน	
Description:	เข้าสู่ระบบ	
Pre-condition:	ต้องมีUsername และ Password	
Post-condition:	นำเข้าสู่หน้าเว็บไซต์	
Flow of Events:	Actor Event	System Responses
	ผู้ใช้ใส่ Username และ Password	แสดงหน้าเว็บหลัก

ตารางที่ 2 รายละเอียดของ Use case search doc

Use case name:	Search doc	
Actor:	ผู้ใช้งาน	
Description:	ค้นหาเอกสาร	
Pre-condition:	มีคำสำคัญเพื่อค้นหาเอกสาร	
Post-condition:	แสดงเอกสารที่ค้นหา	
Flow of Events:	Actor Event	System Responses
	ผู้ใช้คำสำคัญเพื่อค้นหาเอกสาร	ค้นหาเอกสารในระบบ

ตารางที่ 3 รายละเอียดของ Use case upload doc

Use case name:	Upload doc	
Actor:	ผู้ใช้งาน	
Description:	นำเอกสารเข้าสู่ระบบ	
Pre-condition:	เอกสารสำหรับนำเข้าสู่ระบบ	
Post-condition:	นำเอกสารเข้าสู่ระบบ	
Flow of Events:	Actor Event	System Responses
	ผู้ใช้นำเอกสารเข้าสู่ระบบ	เอกสารเข้าสู่ระบบ

ตารางที่ 4 รายละเอียดของ Use case download doc

Use case name:	Download doc	
Actor:	ผู้ใช้งาน	
Description:	ดาวน์โหลดเอกสาร	
Pre-condition:	เลือกเอกสารสำหรับดาวน์โหลด	
Post-condition:	ดาวน์โหลดเอกสาร	
Flow of Events:	Actor Event	System Responses
	ผู้ใช้นำดาวน์โหลดเอกสารจากระบบ	เอกสารถูกดาวน์โหลด

ตารางที่ 5 รายละเอียดของ Use case manage user

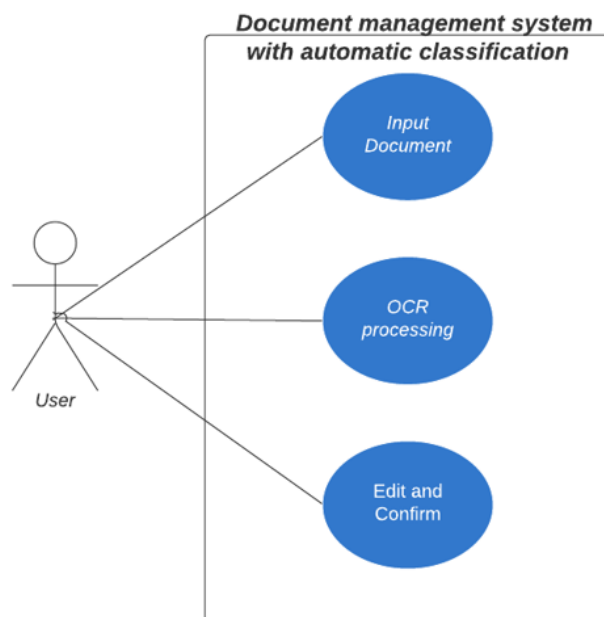
Use case name:	Manage user	
Actor:	ผู้ดูแลระบบ	
Description:	จัดการกับบัญชีผู้ใช้ในระบบ	
Pre-condition:	จัดการกับบัญชีผู้ใช้ในระบบ	
Post-condition:	จัดการกับบัญชีผู้ใช้ในระบบ	
Flow of Events:	Actor Event	System Responses
	ผู้ดูแลระบบจัดการกับ บัญชีผู้ใช้ในระบบ	จัดการกับบัญชีผู้ใช้ในระบบ

ตารางที่ 6 รายละเอียดของ Use case manage doc

Use case name:	Manage doc	
Actor:	ผู้ดูแลระบบ	
Description:	จัดการกับเอกสารในระบบ	
Pre-condition:	จัดการกับเอกสารในระบบ	
Post-condition:	จัดการกับเอกสารในระบบ	
Flow of Events:	Actor Event	System Responses
	ผู้ดูแลระบบจัดการกับ เอกสารในระบบ	จัดการกับเอกสารในระบบ

ตารางที่ 7 รายละเอียดของ Use case manage doc

Use case name:	Manage database	
Actor:	ผู้ดูแลระบบ	
Description:	จัดการกับฐานข้อมูลในระบบ	
Pre-condition:	จัดการกับฐานข้อมูลในระบบ	
Post-condition:	จัดการกับฐานข้อมูลในระบบ	
Flow of Events:	Actor Event	System Responses
	ผู้ดูแลระบบจัดการกับ ฐานข้อมูลในระบบ	จัดการกับฐานข้อมูลในระบบ



ภาพที่ 5 Use case diagram ของระบบนำเข้าเอกสาร

รายละเอียดของ Use case diagram

ตารางที่ 8 รายละเอียดของ Use case Input document

Use case name:	Input document	
Actor:	ผู้ใช้งาน	
Description:	นำเข้าเอกสาร	
Pre-condition:	ต้องมีเอกสาร	
Post-condition:	ดึงเอกสารเข้าสู่ระบบ	
Flow of Events:	Actor Event	System Responses
	ผู้ใช้เลือกเอกสาร	แสดงช่องอัปโหลดเอกสาร

ตารางที่ 9 รายละเอียดของ Use case Input document

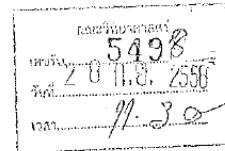
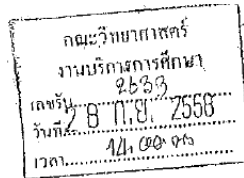
Use case name:	OCR Processing	
Actor:	ระบบ	
Description:	สกัดข้อความออกมาจากเอกสาร	
Pre-condition:	เอกสารต้องมีข้อความ	
Post-condition:	ระบบสกัดข้อความออกมาจากเอกสาร	
Flow of Events:	Actor Event	System Responses
	ผู้ใช้ใส่เอกสาร	ทำการสกัดข้อความออกจากเอกสาร

ตารางที่ 10 รายละเอียดของ Use case Edit and confirm

Use case name:	Edit and confirm	
Actor:	ระบบ	
Description:	นำข้อความที่สกัดได้ตามข้อกำหนดส่งคืนไปหน้าเว็บ	
Pre-condition:	สามารถแก้ไขข้อความที่ถูกส่งมาได้	
Post-condition:	แก้ไขข้อความที่ผิดและยืนยันความถูกต้อง	
Flow of Events:	Actor Event	System Responses
	ตรวจสอบความถูกต้องและแก้ไข	ข้อความที่ถูกยืนยันเข้าสู่ระบบ

4. การออกแบบระบบ

อธิบายการทำงานของระบบ



บันทึกข้อความ

ส่วนราชการ บัณฑิตวิทยาลัย

โทร. 42419

1. ที่ ศธ 0514.10/3534

3. เรื่อง การเสนอขอผู้รับทุนวิจัยสำหรับคณาจารย์บัณฑิตศึกษา ประจำปีการศึกษา 2558 แบบร่วมทุน (joint funding)

เรียน คณะบดีคณะวิทยาศาสตร์

ตามบันทึกที่ ศร 0514.2.1/4583 ลงวันที่ 18 กันยายน 2558 เรื่อง ขอเสนอชื่อผู้รับทุนวิจัยสำหรับ คณาจารย์บัณฑิตศึกษา ประจำปีการศึกษา 2558 แบบร่วมทุน (joint funding) ซึ่งตามประกาศบัณฑิต จิทยาลัย มหาวิทยาลัยขอนแก่น (ฉบับที่ 66/2558) เรื่อง การรับสมัครทุนวิจัยสำหรับคณาจารย์ฯ ประจำปี การศึกษา 2558 แบบร่วมทุน (joint funding) นั้น ผศ.ดร.พชรดี ศิริแสงตระกูล เป็นผู้ได้รับอนุมัติทุนระดับ ปริญญาโท และได้เสนอ นางสาวปวีณา อุ่นลิ รหัศ 585020118-3 เพื่อรับทุนดังกล่าว จากการตรวจสอบ ข้อมูลนักศึกษสำเร็จการศึกษาในระดับปริญญาตรี สาขาวิชาวิทยาการคอมพิวเตอร์ มีผลการเรียน 3.15

บัณฑิตวิทยาลัยโดยคณะกรรมการพิจารณาฯฯ ได้พิจารณาแล้วไม่สามารถอนุมัติทุนวิจัยสำหรับ คณาจารย์ฯฯ ดังกล่าวได้ เนื่องจากคุณสมบัติของนักศึกษาที่ได้รับทุนในระดับปริญญาโท ไม่เป็นไปตามเกณฑ์ ดังที่ระบุไว้ในประกาศ คือ เป็นนักศึกษามหาวิทยาลัยขอนแก่นที่กำลังศึกษาอยู่ในภาคการศึกษาสุดท้ายของ หลักสูตรระดับปริญญาตรี หรือสำเร็จการศึกษาไปแล้ว ที่มีผลการเรียนเฉลี่ยไม่ต่ำกว่า 3.25 หรืออยู่ในกลุ่ม 5% แรกของผู้สำเร็จการศึกษาในสาขาวิชานั้น ทั้งนี้ ผศ.ดร.พชชติ ศิริแสงตระกูล สามารถแจ้งรายชื่อ

ภาพที่ 6 ตัวอย่างเอกสาร

จาก ภาพที่ 6

- (1) เลขที่หนังสือของเอกสาร
- (2) วันที่เขียนเอกสาร
- (3) ชื่อเรื่องของเอกสาร

ระบบสืบค้นเอกสารอัตโนมัติ สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์

สำหรับสืบค้นข้อมูลเอกสาร

สำหรับรับเข้า-ส่งออกข้อมูลเอกสาร

🔒 เข้าสู่ระบบ

ชื่อผู้ใช้


รหัสผ่าน

🗑️

เข้าสู่ระบบ

ภาพที่ 7 หน้าล็อกอิน

จาก **ภาพที่ 7** จะแสดงหน้าการล็อกอินในหน้านี้จะสามารถเลือกได้ว่าจะเข้าสู่ระบบแบบสืบค้นเอกสารหรือนำเข้าเอกสารโดยการกรอกชื่อผู้ใช้และรหัสผ่าน



มหาวิทยาลัยขอนแก่น
ศูนย์วิทยบริการ

ระบบการสืบค้นข้อมูลเอกสารอัตโนมัติ สาขาวิชาวิทยาการคอมพิวเตอร์

ไอพริต, ออกจากระบบ

หน้าหลัก

หมวดหมู่

- ด้านบริหารและธุรการ(ทั่วไป)
- ด้านการเงิน
- ด้านพัสดุ
- ด้านวิจัย
- ด้านแผนยุทธศาสตร์
- ด้านทรัพยากรบุคคล
- ด้านบริการการศึกษา
- ด้านอาคารสถานที่
- สาขาวิชาต่างๆ
- คำสั่ง ประกาศ
- ระเบียบ ข้อบังคับ

สาขาวิชาวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

กรุณารอกข้อมูลที่ต้องการสืบค้น

มหาวิทยาลัยขอนแก่น

▼

ล้างค่า

หมวดหมู่

ทั้งหมด

เลขที่คำสั่ง/ประกาศ

ระบุเลขที่คำสั่ง/ประกาศ

เลขที่หนังสือ

ระบุเลขที่หนังสือ

ชื่อเรื่อง

ระบุชื่อเรื่อง

เดือน

ทุกเดือน

ปี

2564

จากคณะ/หน่วยงาน


ระบุคณะ/หน่วยงาน

ถึง

โปรดระบุ

ค้นหา

ภาพที่ 8 หน้าหลักของหน้าค้นหา



COMPUTER SCIENCE
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี

ระบบการสืบค้นข้อมูลเอกสารอัตโนมัติ สาขาวิชาวิทยาการคอมพิวเตอร์

โพรวัดย์, ออกจากระบบ

หน้าหลัก

หมวดหมู่

- **ด้านบริหารและธุรการ(ทั่วไป)**
- ด้านการเงิน
- ด้านพัสดุ
- ด้านวิจัย
- ด้านแผนยุทธศาสตร์
- ด้านทรัพยากรบุคคล
- ด้านบริการการศึกษา
- ด้านอาคารสถานที่
- สาขาวิชาต่างๆ
- คำสั่ง ประกาศ
- ระเบียบ ข้อบังคับ

สาขาวิชาวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี


ค้นหา

เลขที่หนังสือ


วันที่

→


☐
☐




ระเบียบสำนักนายกรัฐมนตรีว่าด้วย การจัดทำแผนกา...




เงินรายได้ของมหาวิทยาลัย พ.ศ.2540




เงินรายได้ของมหาวิทยาลัย ฉบับที่ 2 พ.ศ.2540




รายได้ของมหาวิทยาลัย พ.ศ.2521




กองทุนวิจัยและคิดค้นสิ่งประดิษฐ์ พ.ศ.2529




กองทุนพัฒนาและส่งเสริมคณาจารย์ทางด้านวิชา...




กองทุนสวนสาธารณะมหาวิทยาลัยขอนแก่น




การติดตั้ง การย้าย และการใช้โทรศัพท์ พ.ศ.2530




การให้บริการของศูนย์คอมพิวเตอร์ พ.ศ.2529



การเก็บและการจ่ายเงินตามจ็อบกับหอพักนักศึกษา...



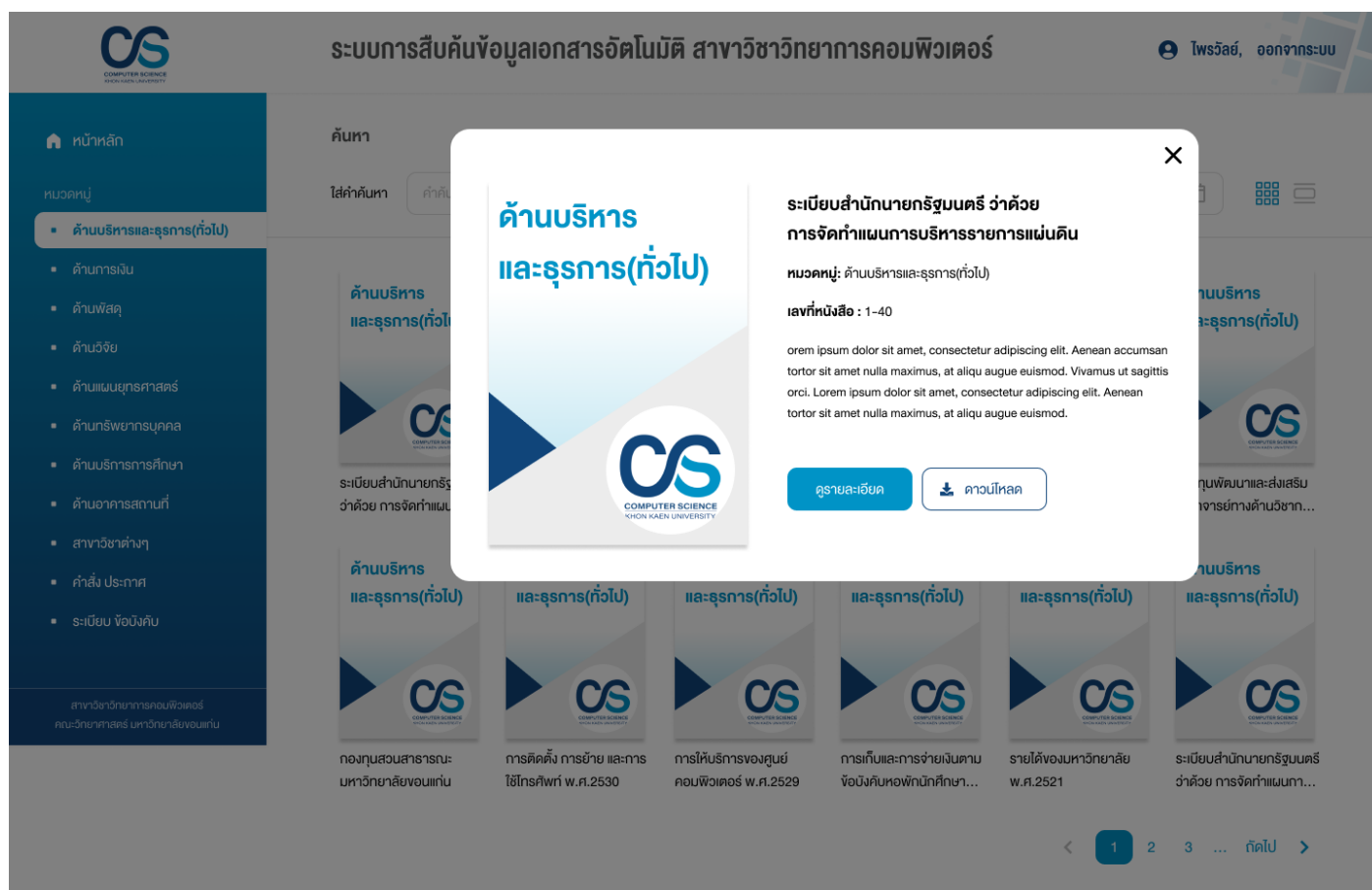
รายได้ของมหาวิทยาลัย พ.ศ.2521




ระเบียบสำนักนายกรัฐมนตรีว่าด้วย การจัดทำแผนกา...

< 1 2 3 ... [ถัดไป](#) >


ภาพที่ 9 หน้าค้นของหมวดด้านบริการและธุรการ(ทั่วไป)(1)



ภาพที่ 10 หน้าแสดงรายละเอียดของเอกสาร(1)



ระบบการสืบค้นข้อมูลเอกสารอัตโนมัติ สาขาวิชาวิทยาการคอมพิวเตอร์


โปรดล็อกเอาท์จากระบบ

หน้าหลัก

หมวดหมู่

- **ด้านบริหารและธุรการ(ทั่วไป)**
- ด้านการเงิน
- ด้านพัสดุ
- ด้านวิจัย
- ด้านแผนยุทธศาสตร์
- ด้านทรัพยากรบุคคล
- ด้านบริการการศึกษา
- ด้านอาคารสถานที่
- สาขาวิชาต่างๆ
- คำสั่ง ประกาศ
- ระเบียบ ข้อบังคับ

สาขาวิชาวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

ค้นหา

ใส่คำค้นหา

🔍

เลขที่หนังสือ

📄

เลขที่หนังสือ



วันที่

2021-09-16











→

2021-10-01

📅

เรื่อง

 ระเบียบสำนักนายกรัฐมนตรี ว่าด้วย การจัดทำแผนการบริหารราชการแผ่นดิน <div style="float: right; border: 1px solid #ccc; padding: 2px 5px;">ดาวน์โหลด</div>
 เงินรายได้ของมหาวิทยาลัย พ.ศ.2540 <div style="float: right; border: 1px solid #ccc; padding: 2px 5px;">ดาวน์โหลด</div>
 เงินรายได้ของมหาวิทยาลัย ฉบับที่ 2 พ.ศ.2540 <div style="float: right; border: 1px solid #ccc; padding: 2px 5px;">ดาวน์โหลด</div>
 รายได้ของมหาวิทยาลัย พ.ศ.2521 <div style="float: right; border: 1px solid #ccc; padding: 2px 5px;">ดาวน์โหลด</div>
 กองทุนวิจัยและศึกษาค้นคว้าวิจัย พ.ศ.2529 <div style="float: right; border: 1px solid #ccc; padding: 2px 5px;">ดาวน์โหลด</div>
 กองทุนพัฒนาและส่งเสริมคณาจารย์ทางด้านวิชาการ พ.ศ.2527 <div style="float: right; border: 1px solid #ccc; padding: 2px 5px;">ดาวน์โหลด</div>
 กองทุนสวนสาธารณะ มหาวิทยาลัยขอนแก่น <div style="float: right; border: 1px solid #ccc; padding: 2px 5px;">ดาวน์โหลด</div>
 การติดตั้ง การย้าย และการใช้โทรศัพท์ พ.ศ.2530 <div style="float: right; border: 1px solid #ccc; padding: 2px 5px;">ดาวน์โหลด</div>
 การให้บริการกองทุนคอมพิวเตอร์ พ.ศ.2529 <div style="float: right; border: 1px solid #ccc; padding: 2px 5px;">ดาวน์โหลด</div>
 การเก็บและการจ่ายเงินค่านักเรียนหอพักนักศึกษา 2522 <div style="float: right; border: 1px solid #ccc; padding: 2px 5px;">ดาวน์โหลด</div>

<

1

2
3
...
ถัดไป
>

ภาพที่ 11 หน้าค้นของหมวดด้านบริการและธุรการ(ทั่วไป)(2)

ระบบการสืบค้นข้อมูลเอกสารอัตโนมัติ สาขาวิชาวิทยาการคอมพิวเตอร์

โพรวัดย์, ออกจากระบบ

หน้าหลัก

หมวดหมู่

- ด้านบริหารและธุรการ(ทั่วไป)
- ด้านการเงิน
- ด้านพัสดุ
- ด้านวิจัย
- ด้านแผนยุทธศาสตร์
- ด้านทรัพยากรบุคคล
- ด้านบริการการศึกษา
- ด้านอาคารสถานที่
- สาขาวิชาต่างๆ
- คำสั่ง ประกาศ
- ระเบียบ ข้อบังคับ

สาขาวิชาวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

ค้นหา

ใส่คำค้นหา

เรื่อง

ระเบียบสำนักนายกรัฐมนตรี ว่าด้วย การจัดทำแผนการบริหารราชการแผ่นดิน

หมวดหมู่: ด้านบริหารและธุรการ(ทั่วไป)

เลขที่หนังสือ : 1-40

ดูรายละเอียด

ดาวน์โหลด

กองกฏวินัยและคดีกันสิ่งประดิษฐ์ พ.ศ.2529

กองกฏพัฒนาและส่งเสริมคนงานจ้างทางด้านวิชาการ พ.ศ.2527

กองกฏสวนสาธารณะ มหาวิทยาลัยขอนแก่น

การติดตั้ง การย้าย และการใช้โทรศัพท์ พ.ศ.2530

การให้บริการของศูนย์คอมพิวเตอร์ พ.ศ.2529

การเก็บและการจ่ายเงินค่าน้ำค้ำประกันนักศึกษา 2522


ดาวน์โหลด

1 2 3 ... กดไป


ภาพที่ 12 หน้าแสดงรายละเอียดของเอกสาร(2)

view2.php


1 / 5 | 100%




1



2



3



ด่วนที่สุด
 ที่ นร ๐๕๐๓/ว ๑
 สำนักเลขาธิการคณะรัฐมนตรี
 ทำเนียบรัฐบาล กทม. ๑๐๓๐๐
 ๗ มกราคม ๒๕๕๘
 เรื่อง ระเบียบสำนักนายกรัฐมนตรี ว่าด้วยการจัดทำแผนการบริหารราชการแผ่นดิน
 พ.ศ. ๒๕๕๗
 เรียน ผู้อำนวยการสำนักงานงบประมาณ
 สิ่งที่ส่งมาด้วย สำเนาระเบียบสำนักนายกรัฐมนตรีในเรื่องนี้
 ด้วยสำนักนายกรัฐมนตรีได้ออกระเบียบสำนักนายกรัฐมนตรี ว่าด้วย
 การจัดทำแผนการบริหารราชการแผ่นดิน พ.ศ. ๒๕๕๗ และสำนักเลขาธิการ

ภาพที่ 13 หน้าแสดงข้อมูลของเอกสาร



ระบบรับเข้าข้อมูลเอกสาร สาขาวิชาวิทยาการคอมพิวเตอร์

โพรวัดย์, ออกจากระบบ

หนังสือรับเข้า

- ด้านบริหารและธุรการ(ทั่วไป)
- ด้านการเงิน
- ด้านพัสดุ
- ด้านวิจัย
- ด้านแผนยุทธศาสตร์
- ด้านทรัพยากรบุคคล
- ด้านบริการการศึกษา
- ด้านอาคารสถานที่
- สาขาวิชาต่างๆ
- คำสั่ง ประกาศ
- ระเบียบ ข้อบังคับ

สาขาวิชาวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่

ค้นหา

ใส่คำค้นหา

คำค้นหา



วันที่

2021-09-16 → 2021-10-01



ที่

ระบุที่

จาก

ระบุจาก

รับเข้า

ส่งออก

รายการรับเข้าเอกสาร

+ เพิ่มรับเข้าเอกสาร

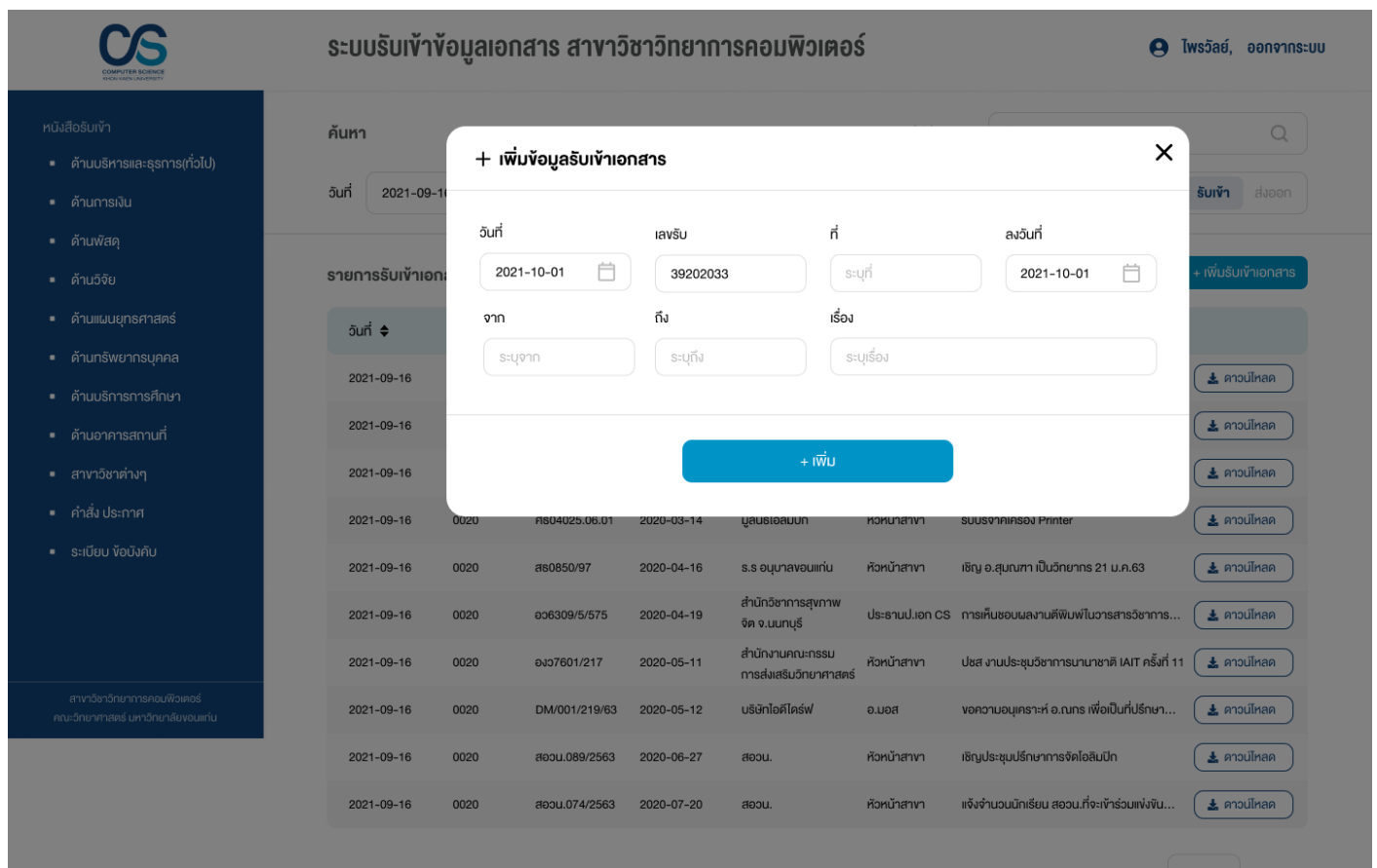
วันที่	เลขรับ	ที่	ลงวันที่	จาก	ถึง	เรื่อง	
2021-09-16	0020	ทว5206/217	2020-01-15	สนท.นครขอนแก่น	หัวหน้าสาขา	ขอความอนุเคราะห์บุคลากรเป็นวิทยากร...	ดาวน์โหลด
2021-09-16	0020	อว600207.1	2020-03-03	สพพ.	พี่อ้อม พี่ธี อ.ใจ	แบบสอบถามการดำเนินงานของคณะกรรม...	ดาวน์โหลด
2021-09-16	0020	สอวน 016/2563	2020-03-10	สมาคมวิศวกรรม สิ่งแวดล้อม	ประธานมูลนิธิ ไอเอ็มบี	อนุมัติจัดโครงการและงบประมาณจัดการ...	ดาวน์โหลด
2021-09-16	0020	ทว4025.06.01	2020-03-14	มูลนิธิไอเอ็มบี	หัวหน้าสาขา	รับรืจากเครื่อง Printer	ดาวน์โหลด
2021-09-16	0020	สอ850/97	2020-04-16	ร.ร อนุบาลขอนแก่น	หัวหน้าสาขา	เชิญ อ.สุนทรา เป็นวิทยากร 21 ม.ค.63	ดาวน์โหลด
2021-09-16	0020	อว6309/5/575	2020-04-19	สำนักวิชาการสุขภาพ จิต จ.นนทบุรี	ประธานป.เอก CS	การเห็นชอบผลงานตีพิมพ์ในวารสารวิชาการ...	ดาวน์โหลด
2021-09-16	0020	อว7601/217	2020-05-11	สำนักงานคณะกรรมการ การส่งเสริมวิทยาศาสตร์	หัวหน้าสาขา	ปะส งานประชุมวิชาการนานาชาติ IAIT ครั้งที่ 11	ดาวน์โหลด
2021-09-16	0020	DM/001/219/63	2020-05-12	บริษัทไอดีไคร์ฟ	อ.มอส	ขอความอนุเคราะห์ อ.นกร เพื่อเป็นที่ปรึกษา...	ดาวน์โหลด
2021-09-16	0020	สอวน.089/2563	2020-06-27	สอวน.	หัวหน้าสาขา	เชิญประชุมปรึกษาการจัดไอเอ็มบี	ดาวน์โหลด
2021-09-16	0020	สอวน.074/2563	2020-07-20	สอวน.	หัวหน้าสาขา	แจ้งจำนวนนักเรียน สอวน.ที่จะเข้าร่วมแข่งขัน...	ดาวน์โหลด

แสดงรายการ

10



ภาพที่ 14 หน้ารับเข้าเอกสาร



ภาพที่ 15 หน้าใส่ข้อมูลเกี่ยวกับเอกสารรับเข้า



ระบบส่งออกข้อมูลเอกสาร สาขาวิชาวิทยาการคอมพิวเตอร์

โพรวัดย์, ออกจากระบบ

หนังสือส่งออก

- งานบริหารและธุรการ(ทั่วไป)
- งานคลังและพัสดุ
- งานวิจัย
- งานแผนยุทธศาสตร์
- งานทรัพยากรบุคคล
- งานบริการการศึกษา
- สาขาวิชาต่างๆ

สาขาวิชาวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

ค้นหา

ใส่คำค้นหา

คำค้นหา



วันที่ 2021-09-16 → 2021-10-01

ที่ ระบุที่

จาก ระบุจาก

รับเข้า

ส่งออก

รายการส่งออกเอกสาร

+ เพิ่มส่งออกเอกสาร

วันที่	เลขรับ	ที่	ลงวันที่	จาก	ถึง	เรื่อง	
2021-09-16	0020	ท5206/217	2020-01-15	สนท.นครขอนแก่น	หัวหน้าสาขา	ขอความอนุเคราะห์บุคลากรเป็นวิทยากร...	ดาวน์โหลด
2021-09-16	0020	อว600207.1	2020-03-03	สพพ.	พี่อ้อม พี่ธี อ.ใจ	แบบสอบถามการดำเนินงานของคณะกรรม...	ดาวน์โหลด
2021-09-16	0020	สอวน 016/2563	2020-03-10	สมาคมวิศวกรรม สิ่งแวดล้อม	ประธานมูลนิธิ ไอเอ็มบี	อนุมัติจัดโครงการและงบประมาณจัดการ...	ดาวน์โหลด
2021-09-16	0020	คร804025.06.01	2020-03-14	มูลนิธิไอเอ็มบี	หัวหน้าสาขา	รับบริจาคเครื่อง Printer	ดาวน์โหลด
2021-09-16	0020	สอ850/97	2020-04-16	ร.ร.อนุบาลขอนแก่น	หัวหน้าสาขา	เชิญ อ.สุนทรา เป็นวิทยากร 21 ม.ค.63	ดาวน์โหลด
2021-09-16	0020	อว6309/5/575	2020-04-19	สำนักวิชาการสุขภาพ จิต จ.นนทบุรี	ประธานป.เอก CS	การเห็นชอบผลงานตีพิมพ์ในวารสารวิชาการ...	ดาวน์โหลด
2021-09-16	0020	อว7601/217	2020-05-11	สำนักงานคณะกรรมการ การส่งเสริมวิทยาศาสตร์	หัวหน้าสาขา	ประชุม ประชุมวิชาการนานาชาติ IAIT ครั้งที่ 11	ดาวน์โหลด
2021-09-16	0020	DM/001/219/63	2020-05-12	บริษัทไอทีโครฟ	อ.มอส	ขอความอนุเคราะห์ อ.นกร เพื่อเป็นที่ปรึกษา...	ดาวน์โหลด
2021-09-16	0020	สอวน.089/2563	2020-06-27	สอวน.	หัวหน้าสาขา	เชิญประชุมปรึกษาการจัดไอเอ็มบี	ดาวน์โหลด
2021-09-16	0020	สอวน.074/2563	2020-07-20	สอวน.	หัวหน้าสาขา	แจ้งจำนวนนักเรียน สอวน.ที่จะเข้าร่วมแข่งขัน...	ดาวน์โหลด

แสดงรายการ

10



ภาพที่ 16 หน้าส่งออกเอกสาร

ภาพที่ 17 หน้าใส่ข้อมูลเกี่ยวกับเอกสารส่งออก

บทที่ 5

ผลสรุป

1. สรุปผลการดำเนินโครงการ

จากการดำเนินโครงการ สิ่งที่ได้ดำเนินไปแล้ว คือ การสกัดข้อความที่เป็นคำสำคัญจากไฟล์เอกสารสำหรับใช้ในการค้นหาเอกสารและสามารถเช็คความถูกต้องของข้อความที่ถูกสกัดออกมาแล้วนำไปแก้ไขได้หากผิดพลาดก่อนจะบันทึกข้อความลงระบบ

2. ข้อจำกัดของระบบ

- 2.1 ไม่สามารถสกัดคำที่เขียนเป็นลายมือได้
- 2.2 หากตัวอักษรมีการขาดหรือหายทำให้การสกัดคำมีปัญหาได้
- 2.3 ในการสกัดคำของ Tesseract ในภาษาไทยนั้น มีความแม่นยำไม่มาก

3. ปัญหาอุปสรรค และ แนวทางแก้ไข

- 3.1 ไม่สามารถอ่านข้อความที่เป็นลายมือได้
- 3.2 การจัดเก็บสารต้องใช้เวลา เพราะเอกสารทั้งหมดมีจำนวนมาก

4. ข้อเสนอแนะ ในการพัฒนาต่อไป

จากผลการพัฒนาระบบโอซีอาร์เพื่อนำมาใช้งานกับระบบเอกสาร พบว่าการใช้ระบบโอซีอาร์กับเอกสารที่ใช้ภาษาอังกฤษนั้นได้ผลลัพธ์ที่ค่อนข้างดี แต่พอนำมาประยุกต์ใช้กับเอกสารภาษาไทยนั้นได้ผลลัพธ์ไม่เป็นที่น่าพอใจนัก จึงควรพัฒนาด้วยการนำเทคโนโลยี ดีพี เกล็นิง มาใช้เพื่อให้การอ่านตัวอักษร มีประสิทธิภาพและความแม่นยำมากยิ่งขึ้น

เอกสารอ้างอิง

- [1] วีรพล มั่นสอริณาท. (2008). ระบบสารสนเทศสำหรับการจัดการข้อมูลอุตสาหกรรม. สาขาวิชาการจัดการงานวิศวกรรม ภาควิชาวิศวกรรมอุตสาหการและการจัดการ บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร
- [2] พรศิริ ภาณุตญาณชัย. (2015). การเพิ่มประสิทธิภาพการรู้จำอักขระภาษาไทยด้วยแสงโดยใช้เทคนิคเปรียบเทียบสายอักขระโดยประมาณและความแตกต่างของลำดับอักขระ. หลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ มหาวิทยาลัยนเรศวร
- [3] รศ. นิตยา เกิดประสพม, รศ. กิตติศักดิ์ เกิดประสพ (2018). การพัฒนาวิธีการฮิวริสติกเพื่อเพิ่มประสิทธิภาพการรู้จำลายมือชื่อ. มหาวิทยาลัยเทคโนโลยีสุรนารี
- [4] Regular Expression (Regex), <https://www.bualabs.com/archives/3070/what-is-regular-expression-regex-regexp-teach-how-to-regex-python-nlp-ep-7/?fbclid=IwAR1UCuWfwClnRufYKad4XI95zPudEuNv5RbQrlWj7TJS1rClyj2ACqvhZP0>, December 03, 2019.
- [5] THAI OCR, <http://thaiocr.phaisarn.com> , JULY 24, 2012.
- [6] Deep learning, <https://www.thaiprogrammer.org/2018/12/deep-learning-%E0%B8%84%E0%B8%B7%E0%B8%AD%E0%B8%AD%E0%B8%B0%E0%B9%84%E0%B8%A3/> , December 16, 2018.

ประวัติผู้วิจัย

ชื่อ-นามสกุล ปันตร์ธร วุฒิอาภรณ์
 วันเกิด วันที่ 10 เดือนกุมภาพันธ์ 2543
 ภูมิลำเนาบ้านเลขที่ 199/256 ม.5 ต.ปรุใหญ่ อ.เมือง จ.นครราชสีมา 30000
 ที่อยู่ปัจจุบัน บ้านเลขที่ 199/256 ม.5 ต.ปรุใหญ่ อ.เมือง จ.นครราชสีมา 30000

ประวัติการศึกษา

พ.ศ.2560 วิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

ชื่อ-นามสกุล รัชชานนท์ ศิริสาร
 วันเกิด วันที่ 6 เดือนตุลาคม 2541
 ภูมิลำเนา 150/148 หมู่บ้านเพชรมุกดาวิลล่า ถ.สำราญชายโขงใต้ ต.ศรีบุญเรือง อ.เมือง จ.
 มุกดาหาร 49000
 ที่อยู่ปัจจุบัน 150/148 หมู่บ้านเพชรมุกดาวิลล่า ถ.สำราญชายโขงใต้ ต.ศรีบุญเรือง อ.
 เมือง จ.มุกดาหาร 49000

ประวัติการศึกษา

พ.ศ.2560 วิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

ลงชื่อผู้ทำโครงการ *ปิ่นเพชร*
(นาย/นางสาว ปิ่นเพชร วดีอักษร.)
วันที่ 04/11/2564

ลงชื่อผู้ทำโครงการ *วิรัชชนก*
(นาย/นางสาว วิรัชชนนธ์ ธีรวิสาร)
วันที่ 04/11/2564

การตรวจสอบจากอาจารย์ที่ปรึกษาโครงการ

.....
.....

(ลงชื่อ)

(.....)

วันที่...../...../.....