

CS 2564/CS10

รายงานความก้าวหน้า 2

ระบบการจัดการเอกสารโดยการแบ่งแยกหมวดหมู่อัตโนมัติ

Document Management System with Automatic Classification

โดย

613020224-3 นายปิ่นธร วุฒิอาภรณ์

613020595-8 นายรัชชานนท์ ศิริสาร

อาจารย์ที่ปรึกษา : รศ.ดร.ปัญญาพล หอระตะ

รายงานนี้เป็นส่วนหนึ่งของการศึกษาวิชา SC314775 โครงการวิทยาการคอมพิวเตอร์ 2

ภาคเรียน 2 ปีการศึกษา 2564

สาขาวิชาวิทยาการคอมพิวเตอร์ วิทยาลัยการคอมพิวเตอร์

มหาวิทยาลัยขอนแก่น

(เดือน กุมภาพันธ์ พ.ศ. 2565)

ปิ่นธร วุฒิอาภรณ์, รัชชานนท์ ศิริสาร , 2564,

ระบบการจัดการเอกสารโดยการแบ่งแยกหมวดหมู่อัตโนมัติ.โครงการคอมพิวเตอร์
ปริญญาวิทยาศาสตรบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ วิทยาลัยการ
คอมพิวเตอร์ มหาวิทยาลัยขอนแก่น

อาจารย์ที่ปรึกษา: รศ.ดร.ปัญญาพล หอระตะ

บทคัดย่อ

เนื่องจากปัจจุบันเอกสารราชการมีการจัดเก็บในรูปแบบไฟล์เอกสาร การค้นหาเอกสาร
เหล่านั้นแม้ว่าจะมีการกำหนดหมายเลขเอกสารเพื่อใช้ในการจำแนกหมวดหมู่ได้ง่ายยิ่งขึ้น แต่ใน
กรณีที่มีเอกสารจำนวนมาก การจัดการกับเอกสารให้มีความรวดเร็วและถูกต้องนั้นจึงเป็นเรื่อง
ยากเช่นกัน ดังนั้นโครงการนี้มีแนวคิดที่จะพัฒนาระบบไอซีอาร์เพื่อวิเคราะห์ไฟล์เอกสารโดย
สแกนหาเลขที่และชื่อเรื่องของไฟล์เอกสาร จากนั้นสกัดคำออกมาจากไฟล์เอกสารเพื่อนำมาใช้
ในการค้นหาเอกสาร ซึ่งวิธีการดังกล่าวจะช่วยลดระยะเวลาในการตรวจสอบเอกสารของผู้ใช้

คำสำคัญ: การสกัดคำ, ไฟล์เอกสาร, ไอซีอาร์

Puntorn Wutiarporn and Ruchchanon Sirisan. 2021.

Document management system with automatic classification.

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

Thesis Advisor: Associate Professor Punyaphol Horata

Abstract

In past of decade, a government organization has to get or send the most important government documents, which the documents are in the form of files, and there have many different types. Searching for these documents requires keywords such as document number, title, etc. Due to the number of the documents is increased rapidly which it is hard to keep or search and having a lot of time to input keywords for easier searching. Therefore, this project is to develop a repository query documents system to handle the problem. This system has two main parts. This first part is to extract keywords automatically from input document using the OCR technique and the second is a web application for searching the desired document.

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

ปัทมธร วุฒิอาภรณ์

รัชชานนท์ ศิริสาร

สารบัญ

	หน้า
บทคัดย่อ	ก
ABSTRACT	ข
กิตติกรรมประกาศ	ค
สารบัญ	ง
สารบัญภาพ	จ
บทที่ 1	1
1. ความเป็นมาและความสำคัญของปัญหา	1
2. วัตถุประสงค์	1
3. เป้าหมายและขอบเขต	1
4. ประโยชน์ที่คาดว่าจะได้รับ	2
บทที่ 2	3
1. ทฤษฎีที่เกี่ยวข้อง	3
2. งานวิจัยที่เกี่ยวข้อง	5
บทที่ 3	7
1. ขั้นตอนการทำงาน	7
2. เครื่องมือที่ใช้ในการทำโครงการนี้	7
บทที่ 4	8
1. การวิเคราะห์ระบบ	8
2. การออกแบบระบบ	8
3. การพัฒนาระบบ	12
4. การออกแบบระบบ	17
บทที่ 5	38
1. สรุปผลการดำเนินโครงการ	38
2. ข้อจำกัดของระบบ	38
3. ปัญหาอุปสรรค และ แนวทางแก้ไข	38
4. ข้อเสนอแนะ ในการพัฒนาต่อไป	38
เอกสารอ้างอิง	39
ประวัติผู้วิจัย	40

สารบัญภาพ

หน้า

ภาพที่ 1	ขั้นตอนการทำงานของระบบค้นหาเอกสาร	8
ภาพที่ 2	ขั้นตอนการทำงานของระบบนำเข้าเอกสารโดยรวม	9
ภาพที่ 3	โครงสร้างการทำงานของระบบนำเข้าเอกสาร	10
ภาพที่ 4	สถาปัตยกรรมของระบบ	11
ภาพที่ 5	Use case diagram ของเว็บไซต์	12
ภาพที่ 6	Use case diagram ของระบบนำเข้าเอกสาร	15
ภาพที่ 7	หน้าล็อกอิน	17
ภาพที่ 8	หน้าหลักของหน้าค้นหา	18
ภาพที่ 9	หน้าค้นของหมวดด้านบริการและธุรการ(ทั่วไป)(1)	19
ภาพที่ 10	หน้าแสดงรายละเอียดของเอกสาร (1)	20
ภาพที่ 11	หน้าค้นของหมวดด้านบริการและธุรการ (ทั่วไป) (2)	21
ภาพที่ 12	หน้าแสดงรายละเอียดของเอกสาร (2)	22
ภาพที่ 13	หน้าแสดงข้อมูลของเอกสาร	23
ภาพที่ 14	หน้ารับเข้าเอกสาร	24
ภาพที่ 15	หน้าประเภทส่งออกเอกสาร	25
ภาพที่ 16	หน้าใส่ข้อมูลเกี่ยวกับรายละเอียดเอกสารส่งออก	26
ภาพที่ 17	หน้าเลือกไฟล์เอกสารเข้าสู่ระบบ	27
ภาพที่ 18	แสดงข้อความที่ผ่านระบบออกมาแสดงเพื่อ แก้ไข และยืนยัน	28
ภาพที่ 19	แสดงผลของการนำเข้าเอกสารเข้าสู่ระบบ	29
ภาพที่ 20	ตัวอย่างการ์ดด้านบริหารและบริการทั่วไป	29
ภาพที่ 21	ตัวอย่างการ์ดด้านการเงิน	30
ภาพที่ 22	ตัวอย่างการ์ดด้านพัสดุ	30
ภาพที่ 23	ตัวอย่างการ์ดด้านวิจัย	31
ภาพที่ 24	ตัวอย่างการ์ดด้านแผนยุทธศาสตร์	31
ภาพที่ 25	ตัวอย่างการ์ดด้านทรัพยากรบุคคล	32
ภาพที่ 26	ตัวอย่างการ์ดด้านบริการการศึกษา	32
ภาพที่ 27	ตัวอย่างการ์ดด้านอาคารสถานที่	33

ภาพที่ 28 ตัวอย่างการ์ดด้านสาขาวิชาต่างๆ	33
ภาพที่ 29 ตัวอย่างการ์ดด้านคำสั่ง ประกาศ	34
ภาพที่ 30 ตัวอย่างการ์ดด้านระเบียบข้อบังคับ	34
ภาพที่ 31 ขั้นตอนการทำงานของระบบล็อกอิน	35
ภาพที่ 32 ขั้นตอนการทำงานของระบบสืบค้น	36
ภาพที่ 33 ขั้นตอนการทำงานของระบบจัดเก็บเอกสาร	37

สารบัญตาราง

	หน้า
ตารางที่ 1 รายละเอียดของ Use case login	12
ตารางที่ 2 รายละเอียดของ Use case search doc	13
ตารางที่ 3 รายละเอียดของ Use case upload doc	13
ตารางที่ 4 รายละเอียดของ Use case download doc	13
ตารางที่ 5 รายละเอียดของ Use case manage user	14
ตารางที่ 6 รายละเอียดของ Use case manage doc	14
ตารางที่ 7 รายละเอียดของ Use case manage doc	14
ตารางที่ 8 รายละเอียดของ Use case Input document	15
ตารางที่ 9 รายละเอียดของ Use case Input document	16
ตารางที่ 10 รายละเอียดของ Use case Edit and confirm	16

บทที่ 1

บทนำ

1. ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันเอกสารสำคัญของหน่วยงานราชการนั้นส่วนใหญ่อยู่ในรูปแบบของไฟล์และมีมากมายหลากหลาย การค้นหาเอกสารเหล่านั้นแม้ว่าจะมีการกำหนดหมายเลขเอกสารเพื่อใช้ในการจำแนกหมวดหมู่ได้ง่ายยิ่งขึ้น แต่ในกรณีที่มีเอกสารจำนวนมาก การจัดการกับเอกสารให้มีความรวดเร็วและถูกต้องนั้นจึงเป็นเรื่องยากเช่นกัน ดังนั้นงานวิจัยนี้จึงมีแนวคิดที่จะพัฒนาระบบเพื่อสแกนหาหมายเลขเอกสาร จากนั้นสกัดออกมาจากเอกสาร เพื่อนำมาใช้ในการค้นหาเอกสาร นอกจากนี้ยังมีการพัฒนาระบบสืบค้นตามหมวดหมู่ของระบบสารบรรณหนังสือราชการไทยอีกด้วย

2. วัตถุประสงค์

- 2.1. สร้างระบบจัดกลุ่มคำสำคัญโดยอัตโนมัติจากไฟล์PDF เพื่อการค้นหาเอกสาร
- 2.2. เพื่อพัฒนาระบบรู้จำเลขเอกสารและระบบการค้นหาเอกสารโดยอัตโนมัติ

3. เป้าหมายและขอบเขต

3.1 ส่วนของเว็บไซต์

พัฒนาเว็บไซต์ระบบเอกสารที่ง่ายต่อการใช้งาน พร้อมระบบโอซีอาร์ที่สามารถอ่านไฟล์ภาพหรือไฟล์พีดีเอฟ และแปลงให้อยู่ในรูปแบบของข้อความเพื่อหาคีย์เวิร์ดต่างๆที่ต้องการมาใช้เป็นข้อมูลกำกับของไฟล์นั้นๆ เช่น ชื่อเรื่อง เลขเอกสาร วันเวลา จากนั้นจัดเก็บลงฐานข้อมูล

3.2 ส่วนของผู้ใช้งานเว็บไซต์

ผู้ใช้งานสามารถใช้งานเว็บไซต์ได้อย่างมีประสิทธิภาพ และเพิ่มความสะดวกสบายให้ผู้ใช้งาน เช่น ดึงคีย์เวิร์ดต่างๆออกมาใส่ในช่องอินพุตข้อมูลของไฟล์ให้ผู้ใช้อัตโนมัติ เพื่อให้ผู้ใช้ตรวจสอบความถูกต้องและยืนยันเพื่อบันทึกลงฐานข้อมูล

3.3 ประเภทของเอกสารนำเข้า

3.3.1 เอกสารรับเข้า

3.3.2 เอกสารส่งออก

3.4 หมวดหมู่ของเอกสารรับเข้า

3.4.1 ด้านบริหารและธุรการ (ทั่วไป)

3.4.2 ด้านการเงิน

3.4.3 ด้านพัสดุ

3.4.4 ด้านวิจัย

3.4.5 ด้านแผนยุทธศาสตร์

3.4.6 ด้านทรัพยากรบุคคล

3.4.7 ด้านบริการการศึกษา

3.4.8 ด้านอาคารสถานที่

3.4.9 สาขาวิชาต่างๆ

3.4.10 คำสั่ง ประกาศ

3.4.11 ระเบียบ ข้อบังคับ

3.5 หมวดหมู่ของเอกสารส่งออก

3.5.1 งานบริหารและธุรการ (ทั่วไป)

3.5.2 งานคลังและพัสดุ

3.5.3 งานวิจัย

3.5.4 งานแผนยุทธศาสตร์

3.5.5 งานทรัพยากรบุคคล

3.5.6 งานบริการการศึกษา

3.5.7 สาขาวิชาต่างๆ

4. ประโยชน์ที่คาดว่าจะได้รับ

4.1 เพื่อจัดการเอกสารได้อย่างมีประสิทธิภาพมากยิ่งขึ้น

4.2 เพื่ออำนวยความสะดวกให้ผู้ใช้งาน

4.3 เพื่อช่วยให้การจัดการและจัดเก็บเอกสารมีประสิทธิภาพมากขึ้น

4.4 เพื่อช่วยให้การสืบค้นข้อมูลเป็นไปอย่างรวดเร็ว

บทที่ 2

งานวิจัยและทฤษฎีที่เกี่ยวข้อง

1. ทฤษฎีที่เกี่ยวข้อง

1. การรู้จำอักขระด้วยแสง (Optical Character Recognition)

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

การเรียนรู้เชิงลึกคือการจำลองรูปแบบการประมวลผลของสมองมนุษย์ โดยใช้โครงข่ายคล้ายเซลล์ประสาทในการประมวลผล เมื่อได้รับข้อมูลมาการเรียนรู้เชิงลึกจะทำการแบ่งแยกข้อมูลและรายละเอียดต่างๆ ที่ได้รับมาทั้งหมด แล้วนำมาประมวลผลหาจุดเด่นและจุดแตกต่างของข้อมูลในเชิงลึก คล้ายกับการกรองข้อมูลเป็นชั้นๆ แล้วสรุปผลข้อมูลออกมาเป็นผลลัพธ์และตรวจสอบว่าข้อมูลนั้นส่งผลอย่างไร ผิด หรือถูก

เช่น มีข้อมูลสัตว์ 1 ตัวที่ไม่ทราบว่าจะเป็นอะไรการเรียนรู้เชิงลึกจะทำการตรวจสอบและคาดการณ์ว่า เป็นสัตว์ชนิดนี้ใช่หรือไม่ โดยไม่จำเป็นต้องระบุว่ามีปีกหรือมีหางการเรียนรู้เชิงลึกจะทำการพยากรณ์ผลลัพธ์ออกมา

หากการเรียนรู้เชิงลึกพยากรณ์ไม่ถูกต้อง ตัวมันจะเรียนรู้และปรับเปลี่ยนการประมวลผล เพื่อให้ผลลัพธ์ที่ออกมามีความถูกต้องมากขึ้น และยังการเรียนรู้เชิงลึกได้เรียนรู้มากขึ้นเพียงใด การเรียนรู้เชิงลึกก็จะเข้าใจได้มากขึ้น และลงลึกในรายละเอียดีย่อยได้มากขึ้น จนสามารถสังเกตความแตกต่างของข้อมูลได้แม้เพียงเล็กน้อยก็ตาม โดยที่มนุษย์ไม่จำเป็นต้องแนะนำ

2. รูปแบบการนิยามและระบุข้อมูลชนิดสตริง (Regular Expressions)

รูปแบบการนิยามและระบุข้อมูลชนิดสตริงคือลำดับอักขรรูปแบบที่ใช้ในการค้นหาข้อความในสตริงโดยสามารถใช้ค้นหาอย่างเดียวหรือค้นหาแล้วแทนที่ด้วยข้อความที่กำหนดก็ได้ หรือใช้ตรวจสอบการรับค่าของสตริงว่าเป็นไปตามรูปแบบที่กำหนดหรือไม่ การระบุข้อมูลชนิดสตริงเป็นที่นิยมมาก เนื่องจากแก้ปัญหาที่ทุกคนต้องเจอ ซ้ำ 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. ต้องการข้อมูลจำนวนมากมหาศาล
2. กลไกการทำงานที่ไม่สามารถอธิบายได้

หลักการทำงานของการเรียนรู้เชิงลึก

การเรียนรู้เชิงลึกคือการจำลองรูปแบบการประมวลผลของสมองมนุษย์ โดยใช้โครงข่ายคล้ายเซลล์ประสาทในการประมวลผล เมื่อได้รับข้อมูลมาการเรียนรู้เชิงลึกจะทำการแบ่งแยกข้อมูลและรายละเอียดต่างๆ ที่ได้รับมาทั้งหมด แล้วนำมาประมวลผลหาจุดเด่นและจุดแตกต่างของข้อมูลในเชิงลึก คล้ายกับการกรองข้อมูลเป็นชั้นๆ แล้วสรุปผลข้อมูลออกมาเป็นผลลัพธ์และตรวจสอบว่าข้อมูลนั้นส่งผลอย่างไร ผิด หรือถูก

เช่น มีข้อมูลสัตว์ 1 ตัวที่ไม่ทราบว่าจะเป็นอะไรการเรียนรู้เชิงลึกจะทำการตรวจสอบและคาดการณ์ว่า เป็นสัตว์ชนิดนี้ใช่หรือไม่ โดยไม่จำเป็นต้องระบุว่ามียกหรือมีหางการเรียนรู้เชิงลึกจะทำการพยากรณ์ผลลัพธ์ออกมา

หากการเรียนรู้เชิงลึกพยากรณ์ไม่ถูกต้อง ตัวมันจะเรียนรู้และปรับเปลี่ยนการประมวลผล เพื่อให้ผลลัพธ์ที่ออกมามีความถูกต้องมากขึ้น และยังการเรียนรู้เชิงลึกได้เรียนรู้มากขึ้นเพียงใด การเรียนรู้เชิงลึกก็จะเข้าใจได้มากขึ้น และลงลึกในรายละเอียดีย่อยได้มากขึ้น จนสามารถสังเกตความแตกต่างของข้อมูลได้แม้เพียงเล็กน้อยก็ตาม โดยที่มนุษย์ไม่จำเป็นต้องแนะนำ

2. งานวิจัยที่เกี่ยวข้อง

จากการสรุปผลการค้นคว้า งานวิจัยเรื่อง ระบบสารสนเทศสำหรับจัดการอุตสาหกรรม [1] งานวิจัยนี้นำเสนอวิธีการแปลงข้อมูลการผลิตรายปี โดยแปลงภาพตัวอักษร ให้เป็นข้อมูลตัวอักษร ปรากฏว่าการแปลงข้อมูลจากรูปนั้นให้เป็นตัวอักษรมีทั้งข้อดีและข้อเสียที่เป็นอุปสรรคที่เกิดขึ้นดังนี้

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

แต่อย่างไรก็ตาม งานวิจัยที่มี การใช้การสกัดคำนั้นในปัจจุบันมีประสิทธิภาพในการทำงานสูงและขั้นตอนการทำงานที่รวดเร็ว ทำให้เกิดความสะดวกต่อการทำงานในหลายๆด้าน จากงานวิจัยเรื่อง การเพิ่มประสิทธิภาพการรู้จำอักขระภาษาไทยด้วยแสงโดยใช้เทคนิคเปรียบเทียบสายอักขระโดยประมาณและความแตกต่างของลำดับอักขระ

[2] งานวิจัยนี้นำเสนอ การเพิ่มประสิทธิภาพการทำงานของงานการสกัดคำ สรุปได้ว่า ระบบรู้จำตัวอักขระด้วยแสงที่เหมาะสมกับการทดลองที่สุดคือ การสกัดคำแบบเอพีโอ และพื้นหลังที่ให้ประสิทธิภาพในการทดลองดีที่สุดคือ 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 เพื่อให้ผู้ใช้ สามารถใช้ในการค้นหาเอกสารราชการ

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 Express JS

2.5 Tesseract OCR

2.6 MongoDB

2.7 React Framework

2.8 Regex

2.9 Node JS

บทที่ 4

การวิเคราะห์ระบบ และพัฒนาโปรแกรมวิจัย

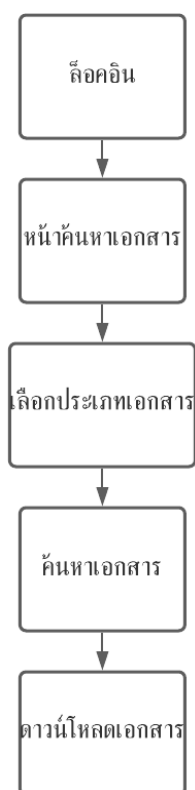
1. การวิเคราะห์ระบบ

ระบบวิเคราะห์ข้อความที่ได้จากเอกสารที่นำเข้าสู่ระบบโดยจะจัดเก็บไว้เพื่อใช้ในการค้นหาเอกสารนั้นๆ ตัวอย่างประเภทของคำค้นหา จะมี ชื่อเรื่อง ประเภท เลขที่คำสั่ง เลขที่หนังสือและเดือนปีที่เขียน เพื่อใช้สำหรับการค้นหาเอกสาร การนำเอกสารเข้าสู่ระบบนั้นจะอัปโหลดข้อมูลขึ้นบนเว็บ แล้วระบบจะทำการโอซีอาร์ข้อความออกมาเป็นข้อความและจะจัดเก็บเข้าระบบ

2. การออกแบบระบบ

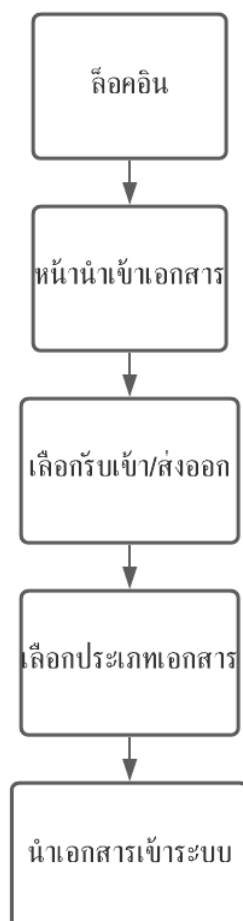
2.1 โครงสร้างของระบบ

จาก ภาพที่ 1 ขั้นตอนการทำงานของระบบค้นหาจะเริ่มจากการเข้าสู่ระบบหลังจากนั้นจะแสดงหน้าค้นหาหลักสามารถเลือกหมวดหมู่ของเอกสารได้ เมื่อเลือกหมวดหมู่ของเอกสารแล้วจะสามารถใส่คำสำคัญเพื่อหาเอกสารนั้นๆได้ เมื่อค้นหาแล้วจะสามารถเข้าไปดูรายละเอียดของเอกสารได้และสามารถดาวน์โหลดเอกสารเหล่านั้น



ภาพที่ 1 ขั้นตอนการทำงานของระบบค้นหาเอกสาร

จาก ภาพที่ 2 ขั้นตอนการทำงานของระบบนำเข้าเอกสารหลังจากเข้าสู่ระบบแล้วจะสามารถเลือกประเภทของเอกสารที่นำเข้าได้ นั่นคือ เอกสารรับเข้าหรือเอกสารส่งออก เมื่อเลือกประเภทแล้วเลือกประเภทของหมวดหมู่เอกสารที่ต้องการจะนำเข้า ต่อมาการเลือกไฟล์เอกสารที่ต้องการนำเข้าสู่ระบบ



ภาพที่ 2 ขั้นตอนการทำงานของระบบนำเข้าเอกสารโดยรวม

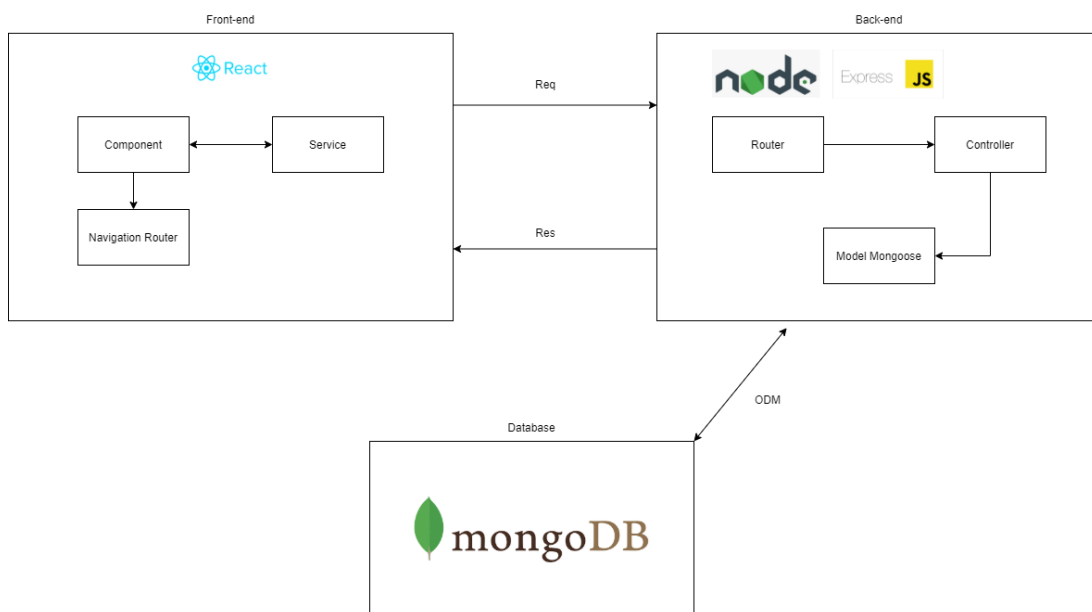
จาก ภาพที่ 3 เป็นโครงสร้างของการทำงานของระบบนำเข้าเอกสาร โดยระบบจะเริ่มจากการที่ผู้นำเอกสารเข้าสู่ระบบผ่านเว็บไซต์ของระบบ หลังจากนั้นระบบจะแปลงไฟล์เอกสารเป็นไฟล์รูปภาพ หลังจากนั้นจะนำไฟล์รูปผ่านแปลงเป็นตัวอักษรโดยกระบวนการโอซีอาร์ หลังจากได้ตัวอักษรมาแล้วจะทำการคัดเฉพาะคำสำคัญออกมาและส่งกลับไปหน้าเว็บไซต์เพื่อให้ผู้ใช้ได้ตรวจสอบและแก้ไขความถูกต้อง เมื่อตรวจสอบและแก้ไขแล้วจะทำการเก็บลงฐานข้อมูล



ภาพที่ 3 โครงสร้างการทำงานของระบบนำเข้าเอกสาร

2.2 สถาปัตยกรรมของระบบ

ในส่วน Frontend จะใช้ React Framework ในการจัดการ ซึ่งในตัว React นั้นจะมี library หลายตัวที่ช่วยในการจัดการกับ component ต่างๆของหน้าเว็บไซต์ส่วนการเรียกใช้ service จากหน้าเว็บไซต์ จะใช้ Axios เพื่อส่ง request จากหน้าเว็บไซต์ไปยัง Backend โดยจุด endpoint จะอยู่ที่ route ของฝั่ง Backend ซึ่งที่ route จะใช้ Express.js เพื่อใช้สำหรับการกำหนดเส้นทางต่างๆของระบบ ซึ่งเมื่อ route ได้รับ request จะทำการเรียกใช้ฟังก์ชันต่างๆจาก Controller และ Middleware ตามที่ Frontend ได้ request มา โดยตัว Middleware จะเกี่ยวข้องกับ Authentication เพื่อตรวจสอบผู้ใช้ที่เข้าสู่ระบบ ส่วน Controller จะประกอบด้วยฟังก์ชันทั่วไปและฟังก์ชันต่างๆที่ใช้ในการจัดการกับข้อมูลต่างๆที่อยู่ใน Mongo Database ผ่านตัวกลางคือ Mongoose Model ซึ่งท้ายสุดจะส่ง response กลับไปยังฝั่ง Frontend โดยทั้งหมดนี้จะทำงานอยู่บน Node.js ซึ่งเป็น Cross Platform Runtime Environment ดังภาพที่ 4

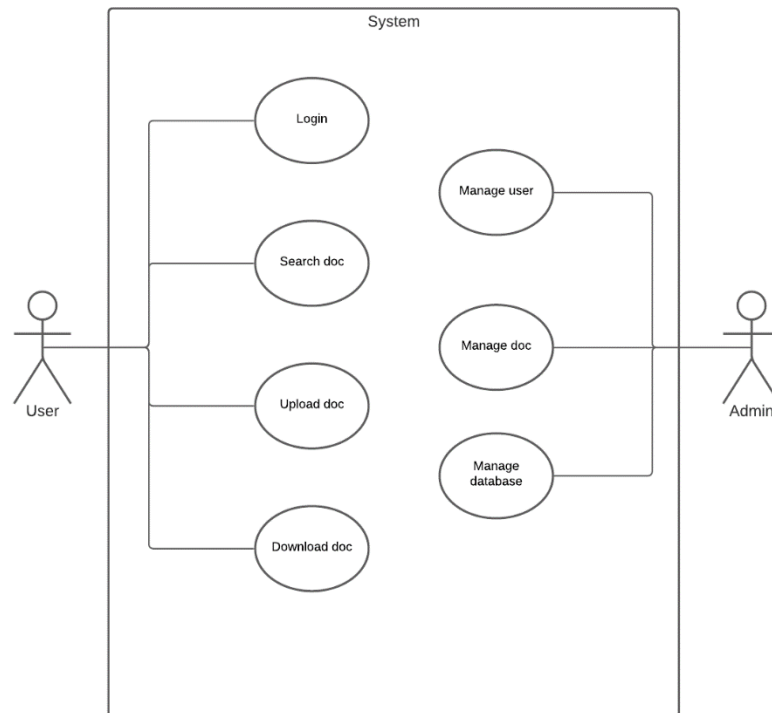


ภาพที่ 4 สถาปัตยกรรมของระบบ

3. การพัฒนาระบบ

3.1 System design / Architecture design

3.1.1 ออกแบบ Use case diagram



ภาพที่ 5 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:	นำเข้าสู่เว็บไซต์	
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

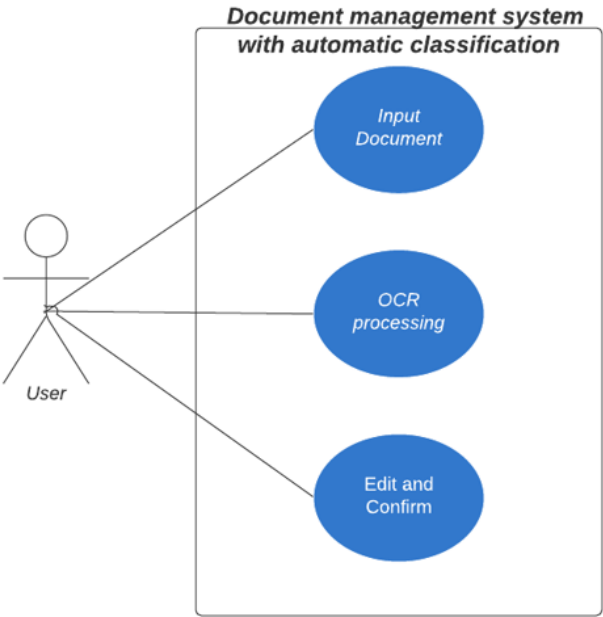
Use case name:	Manage user	
Actor:	ผู้ดูแลระบบ	
Description:	จัดการกับบัญชีผู้ใช้ในระบบ	
Pre-condition:	จัดการกับบัญชีผู้ใช้ในระบบ	
Post-condition:	จัดการกับบัญชีผู้ใช้ในระบบ	
Flow of Events:	Actor Event	System Responses
	ผู้ดูแลระบบจัดการกับ บัญชีผู้ใช้ในระบบ	จัดการกับบัญชีผู้ใช้ในระบบ

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 database

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 6 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

ผู้วิจัยได้ออกแบบเว็บไซต์ที่มีฟังก์ชันการทำงานดังต่อไปนี้

- 4.1 ฟังก์ชันล็อกอินเข้าสู่ระบบ
- 4.2 ฟังก์ชันค้นหาเอกสารแบบละเอียด
- 4.3 ฟังก์ชันการค้นหาตามหมวดหมู่
- 4.4 ฟังก์ชันแสดงผลแบบตาราง
- 4.5 ฟังก์ชันแสดงผลแบบการ์ด
- 4.6 ฟังก์ชันดูรายละเอียดของเอกสาร
- 4.7 ฟังก์ชันดาวน์โหลดเอกสาร
- 4.8 ฟังก์ชันเพิ่มเอกสารเข้าสู่ระบบ
- 4.9 ฟังก์ชันสกัดคำอัตโนมัติ

อธิบายการทำงานของระบบจาก ภาพที่ 7 จะแสดงหน้าการล็อกอินในหน้านี้จะสามารถเลือกได้ว่าจะเข้าสู่ระบบแบบสืบค้นเอกสารหรือนำเข้าเอกสารโดยการกรอกชื่อผู้ใช้และรหัสผ่าน

ภาพที่ 7 หน้าล็อกอิน

จาก ภาพที่ 8 จะแสดงหน้าค้นหาหลักซึ่งสามารถค้นหาได้ทุกหมวดหมู่ และมีหมวดคำค้นหาที่เยอะกว่า

CS
COMPUTER SCIENCE
ภาควิชาวิทยาการคอมพิวเตอร์

ระบบการสืบค้นข้อมูลเอกสารอัตโนมัติ สาขาวิชาวิทยาการคอมพิวเตอร์

โพรวัลย์, ออกจากระบบ

หน้าหลัก

หมวดหมู่

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

มหาวิทยาลัยขอนแก่น

ล้างค่า

หมวดหมู่

ทั้งหมด

เลขที่คำสั่ง/ประกาศ

ระบุเลขที่คำสั่ง/ประกาศ

เลขที่หนังสือ

ระบุเลขที่หนังสือ

ชื่อเรื่อง

ระบุชื่อเรื่อง

เดือน

ปี

จากคณะ/หน่วยงาน

ถึง

ทุกเดือน

2564

ระบุคณะ/หน่วยงาน

ปัจจุบัน

ค้นหา

สาขาวิชาวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

ภาพที่ 8 หน้าหลักของหน้าค้นหา

จาก ภาพที่ 9 เป็นหน้าค้นหของหมวดหมู่ ด้านบริหารและธุรการ (ทั่วไป) ที่แสดงผลออกมาในรูปแบบของการ์ด หน้านี้จะมีการค้นหาโดยใช้ ชื่อเรื่อง เลขที่หนังสือ และวันที่ลงหนังสือ

ระบบการสืบค้นข้อมูลเอกสารอัตโนมัติ สาขาวิชาวิทยาการคอมพิวเตอร์

ไอพรวิสัย, ออกจากระบบ

ค้นหา

ใส่คำค้นหา คำค้นหา เลขที่หนังสือ เลขที่หนังสือ วันที่ 2021-09-16 2021-10-01

หน้าหลัก

หมวดหมู่

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

สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

ด้านบริหารและธุรการ(ทั่วไป)

ระเบียบสำนักนายกรัฐมนตรีว่าด้วย การจัดทำแผนกา...

เงินรายได้ของมหาวิทยาลัย ว.ศ.2540

เงินรายได้ของมหาวิทยาลัย ฉบับที่ 2 ว.ศ.2540

รายได้ของมหาวิทยาลัย ว.ศ.2521

กองทุนวิจัยและศึกษาค้นคว้า ประดิษฐ์ ว.ศ.2529

กองทุนพัฒนาและส่งเสริม คณาจารย์ทางด้านวิชา...

ด้านบริหารและธุรการ(ทั่วไป)

กองทุนส่วนสาธารณะ มหาวิทยาลัยขอนแก่น

การติดตั้ง การย้าย และการใช้โทรศัพท์ ว.ศ.2530

การให้บริการของศูนย์ คอมพิวเตอร์ ว.ศ.2529

การเก็บและการจ่ายเงินตาม ข้อมติของแพทยสภา...

รายได้ของมหาวิทยาลัย ว.ศ.2521

ระเบียบสำนักนายกรัฐมนตรีว่าด้วย การจัดทำแผนกา...

1 2 3 ... กดไป

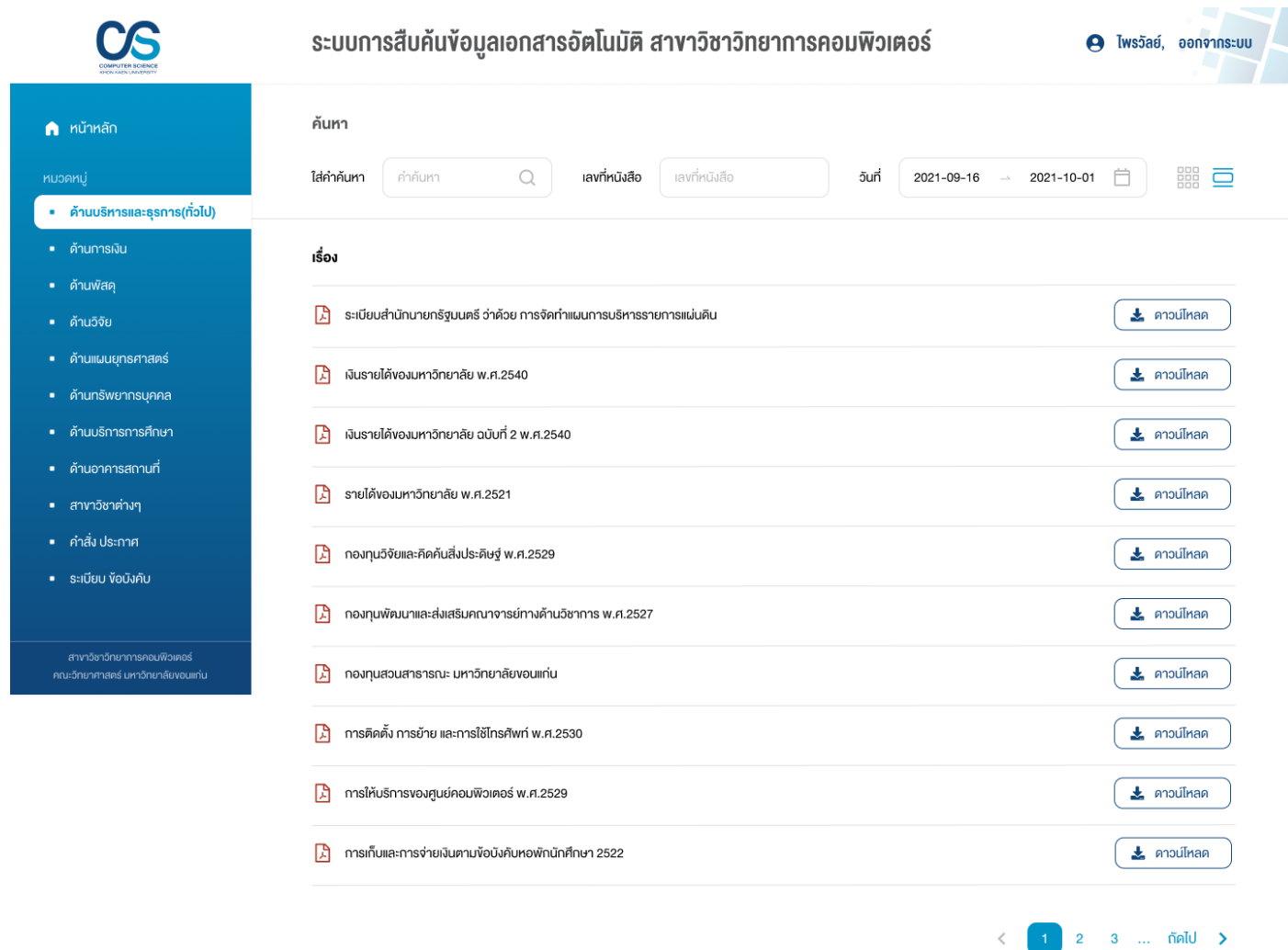
ภาพที่ 9 หน้าค้นหของหมวดด้านบริการและธุรการ(ทั่วไป)(1)

จาก ภาพที่ 10 เมื่อกดเลือกที่การ์ดจะแสดงรายละเอียดของเอกสารออกมา สามารถดาวน์โหลดเอกสารและดูรายละเอียดเอกสารได้



ภาพที่ 10 หน้าแสดงรายละเอียดของเอกสาร (1)

จาก ภาพที่ 11 เป็นหน้าค้นหาของหมวดหมู่ ด้านบริหารและธุรการ (ทั่วไป) ที่แสดงผลออกมาในรูปแบบของตาราง หน้านี้จะมีการค้นหาโดยใช้ ชื่อเรื่อง เลขที่หนังสือ และวันที่ลงหนังสือ



ระบบการสืบค้นข้อมูลเอกสารอัตโนมัติ สาขาวิชาวิทยาการคอมพิวเตอร์

ไอพริตซ์, ออกจากระบบ

ค้นหา

ใส่คำค้นหา คำค้นหา เลขที่หนังสือ เลขที่หนังสือ วันที่ 2021-09-16 2021-10-01

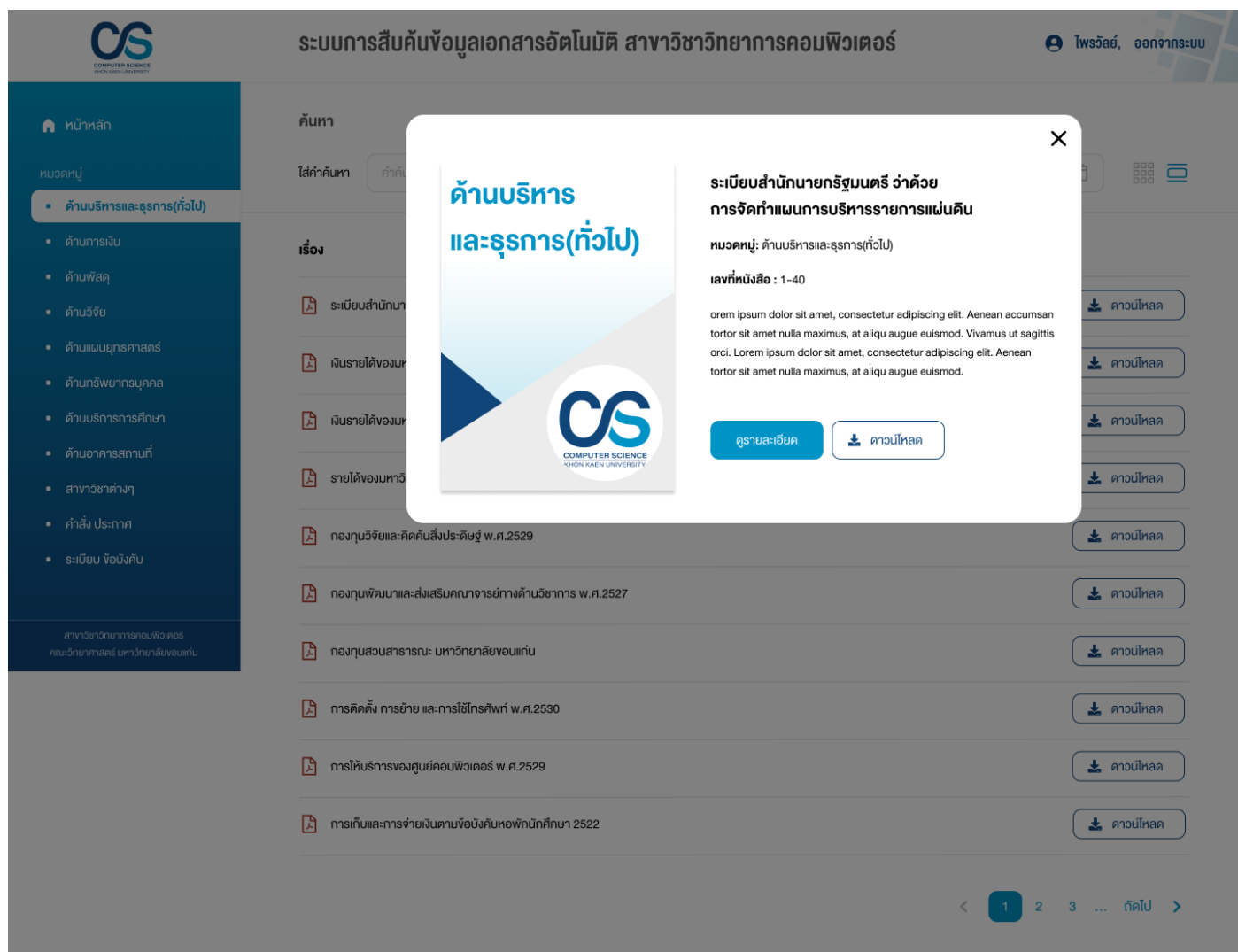
เรื่อง

ระเบียบสำนักนายกรัฐมนตรี ว่าด้วย การจัดทำแผนการบริหารราชการแผ่นดิน	ดาวน์โหลด
เงินรายได้ของมหาวิทยาลัย พ.ศ.2540	ดาวน์โหลด
เงินรายได้ของมหาวิทยาลัย ฉบับที่ 2 พ.ศ.2540	ดาวน์โหลด
รายได้ของมหาวิทยาลัย พ.ศ.2521	ดาวน์โหลด
กองทุนวิจัยและศึกษาค้นคว้า พ.ศ.2529	ดาวน์โหลด
กองทุนพัฒนาและส่งเสริมคณาจารย์ทางด้านวิชาการ พ.ศ.2527	ดาวน์โหลด
กองทุนสวนสาธารณะ มหาวิทยาลัยขอนแก่น	ดาวน์โหลด
การติดตั้ง การย้าย และการใช้โทรศัพท์ พ.ศ.2530	ดาวน์โหลด
การให้บริการของศูนย์คอมพิวเตอร์ พ.ศ.2529	ดาวน์โหลด
การเก็บและการจ่ายเงินค่านักศึกษาหอพักนักศึกษา 2522	ดาวน์โหลด

1 2 3 ... กดไป

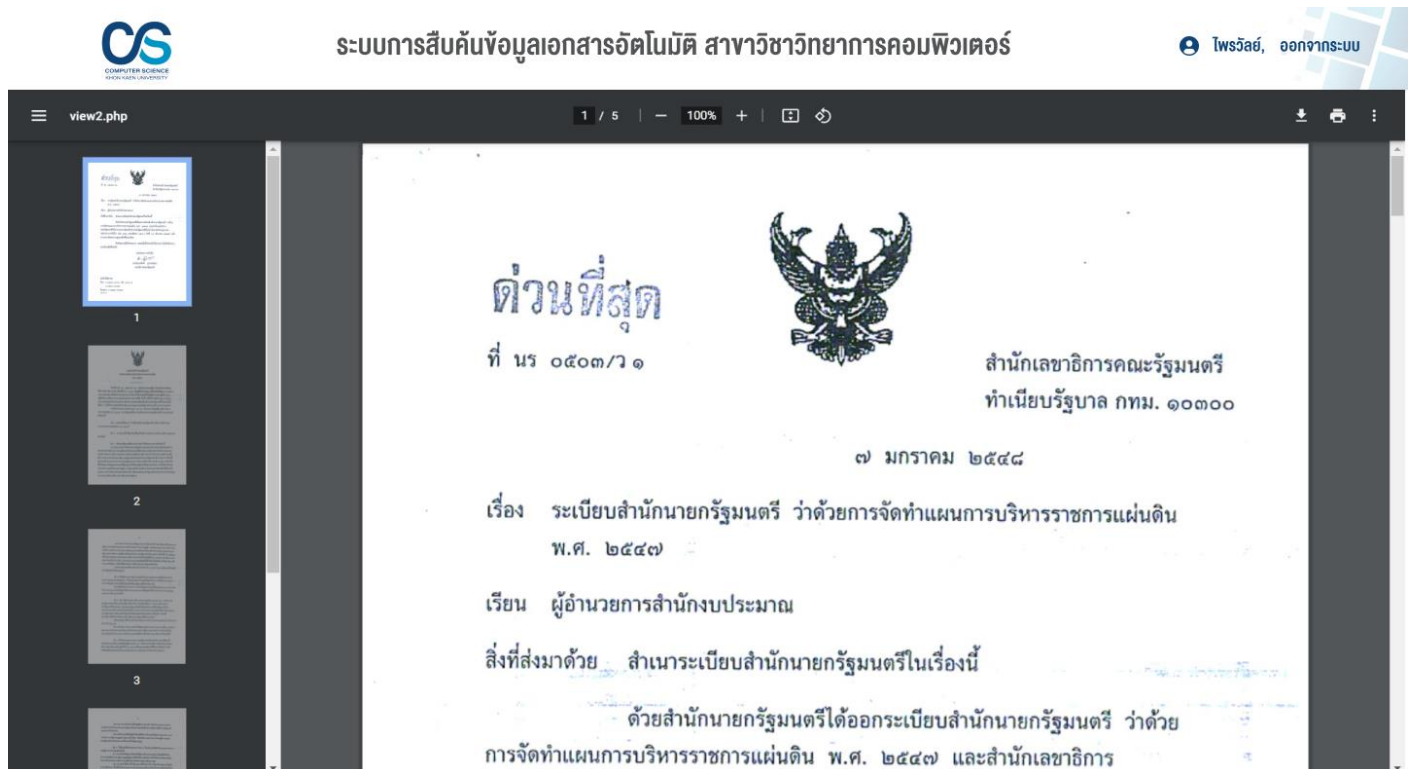
ภาพที่ 11 หน้าค้นหาค้นของหมวดหมู่ด้านบริการและธุรการ (ทั่วไป) (2)

จาก ภาพที่ 12 เมื่อกดเลือกที่ตารางจะแสดงรายละเอียดของเอกสารออกมา สามารถดาวน์โหลดเอกสารและดูแลรายละเอียดเอกสารได้




ภาพที่ 12 หน้าแสดงรายละเอียดของเอกสาร (2)

จาก ภาพที่ 13 เมื่อเลือกกดเข้าไปที่ดูรายละเอียด จะแสดงรายละเอียดของเอกสารออกมา



ภาพที่ 13 หน้าแสดงข้อมูลของเอกสาร

จาก ภาพที่ 14 เป็นหน้าของส่วนในการนำเข้าเอกสาร ในประเภทรับเข้าเอกสารจะแสดงผลออกมาเป็นรูปแบบตาราง



COMPUTER SCIENCE
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี

ระบบรับเข้าข้อมูลเอกสาร สาขาวิชาวิทยาการคอมพิวเตอร์

🔍 โพรวัดย์, ออกจากระบบ

หน้าสือรับเข้า

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

สาขาวิชาวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ มหาวิทยาลัยพระจอมเกล้าธนบุรี

ค้นหา

วันที่ 2021-09-16 → 2021-10-01 ที่ ระบุที่ จาก ระบุจาก รับเข้า ส่งออก

ใส่คำค้นหา


รายการรับเข้าเอกสาร + เพิ่มรับเข้าเอกสาร

วันที่	เลขรับ	ที่	ลงวันที่	จาก	ถึง	เรื่อง	
2021-09-16	0020	ทว5206/217	2020-01-15	สนท.นครจนเท่น	หัวหน้าสาขา	ขอความอนุเคราะห์บุคลากรเป็นวิทยากร...	ดาวน์โหลด
2021-09-16	0020	อว600207.1	2020-03-03	สพพ.	พี่อ้อม พี่ อ.ใจ	แบบสอบถามการดำเนินงานของคณะกรรม...	ดาวน์โหลด
2021-09-16	0020	สอวน 016/2563	2020-03-10	สมาคมวิศวกรรม สิ่งแวดล้อม	ประธานมูลนิธิ ไอส์บปิก	อนุมัติจัดโครงการแลกเปลี่ยนประมาณจัดการ...	ดาวน์โหลด
2021-09-16	0020	ทว0425.06.01	2020-03-14	มูลนิธิไอส์บปิก	หัวหน้าสาขา	รับบริจาคเครื่อง Printer	ดาวน์โหลด
2021-09-16	0020	สอ0850/97	2020-04-16	ร.ร อนุบาลจอนเท่น	หัวหน้าสาขา	เชิญ อ.สุนทภา 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	ประส งานประชุมวิชาการนานาชาติ IAIT ครั้งที่ 11	ดาวน์โหลด
2021-09-16	0020	DM/001/219/63	2020-05-12	บริษัทไอดีไทร์ฟ	อ.มอส	ขอความอนุเคราะห์ อ.นกร เพื่อเป็นที่ปรึกษา...	ดาวน์โหลด
2021-09-16	0020	สอวน.089/2563	2020-06-27	สอวน.	หัวหน้าสาขา	เชิญประชุมปรึกษาการจัดไอส์บปิก	ดาวน์โหลด
2021-09-16	0020	สอวน.074/2563	2020-07-20	สอวน.	หัวหน้าสาขา	แจ้งจำนวนนักเรียน สอวน.ที่จะเข้าร่วมแข่งขัน...	ดาวน์โหลด

แสดงรายการ 10 < >


ภาพที่ 14 หน้ารับเข้าเอกสาร

จาก ภาพที่ 15 เป็นหน้าของส่วนในการนำเข้าเอกสาร ในประเภทส่งออกเอกสารจะแสดงผลออกมาเป็นรูปแบบตาราง



COMPUTER SCIENCE
SODHOMSAK UNIVERSITY

ระบบส่งออกข้อมูลเอกสาร สาขาวิชาวิทยาการคอมพิวเตอร์


 โปรไฟล์, ออกจากระบบ

หนังสือส่งออก

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

สาขาวิชาวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

ค้นหา

วันที่ 2021-09-16 → 2021-10-01 

ใส่คำค้นหา











วันที่

จาก

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	สพ.	พี่อ้อม พี่ อ.ไฉ	แบบสอบถามการดำเนินงานของคณะกรรม...	 ดาวน์โหลด
2021-09-16	0020	สอวน 016/2563	2020-03-10	สมาคมวิศวกรรม สิ่งแวดล้อม	ประธานมูลนิธิ ไอเอ็มป้า	อนุมัติจัดโครงการแลกเปลี่ยนประมา...	 ดาวน์โหลด
2021-09-16	0020	คส04025.06.01	2020-03-14	มูลนิธิไอเอ็มป้า	หัวหน้าสาขา	รับบริจาคเครื่อง Printer	 ดาวน์โหลด
2021-09-16	0020	สอ0850/97	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	สำนักวิชาการสุภาพ จิต จ.นนทบุรี	ประธานป.เอก CS	การยื่นขอผลงานตีพิมพ์ในวารสารวิชาการ...	 ดาวน์โหลด
2021-09-16	0020	อว7601/217	2020-05-11	สำนักงานคณะกรรมการ การส่งเสริมวิทยาศาสตร์	หัวหน้าสาขา	ปอส งานประชุมวิชาการนานาชาติ IAIT ครั้งที่ 11	 ดาวน์โหลด
2021-09-16	0020	DM/001/219/63	2020-05-12	บริษัทไอทีโครฟ	อ.มอส	ขอความอนุเคราะห์ อ.นคร เพื่อเป็นที่ปรึกษา...	 ดาวน์โหลด
2021-09-16	0020	สอวน.089/2563	2020-06-27	สอวน.	หัวหน้าสาขา	เชิญประชุมปรึกษาการจัดไอเอ็มป้า	 ดาวน์โหลด
2021-09-16	0020	สอวน.074/2563	2020-07-20	สอวน.	หัวหน้าสาขา	แจ้งจำนวนนักเรียน สอวน.ที่จะเข้าร่วมแข่งขัน...	 ดาวน์โหลด

แสดงรายการ 10 < >

ภาพที่ 15 หน้าประเภทส่งออกเอกสาร

จาก ภาพที่ 16 หน้านี้จะเป็นการนำไฟล์เอกสารและรายละเอียดของเอกสารเข้าสู่ฐานข้อมูลโดยจะมีรายละเอียดดังนี้ วันที่ เลขรับ ที่ ลงวันที่ จาก ถึง เรื่อง



- ด้านบริหารและธุรการ(ทั่วไป)
- ด้านการเงิน
- ด้านพัสดุ
- ด้านวิจัย
- ด้านแผนยุทธศาสตร์
- ด้านทรัพยากรบุคคล
- ด้านการบริหารการศึกษา
- ด้านอาคารสถานที่
- สารนิเทศต่างๆ
- ฝ่ายสิ่ง ประกาศ
- ระบบคอมพิวเตอร์


ระบบการสืบค้นข้อมูลเอกสารอัตโนมัติ

admin [ออกจากระบบ](#)

Choose File

No file chosen


วันที่

mm/dd/yyyy 

เลขรับ

ที่

ลงวันที่

mm/dd/yyyy 

จาก

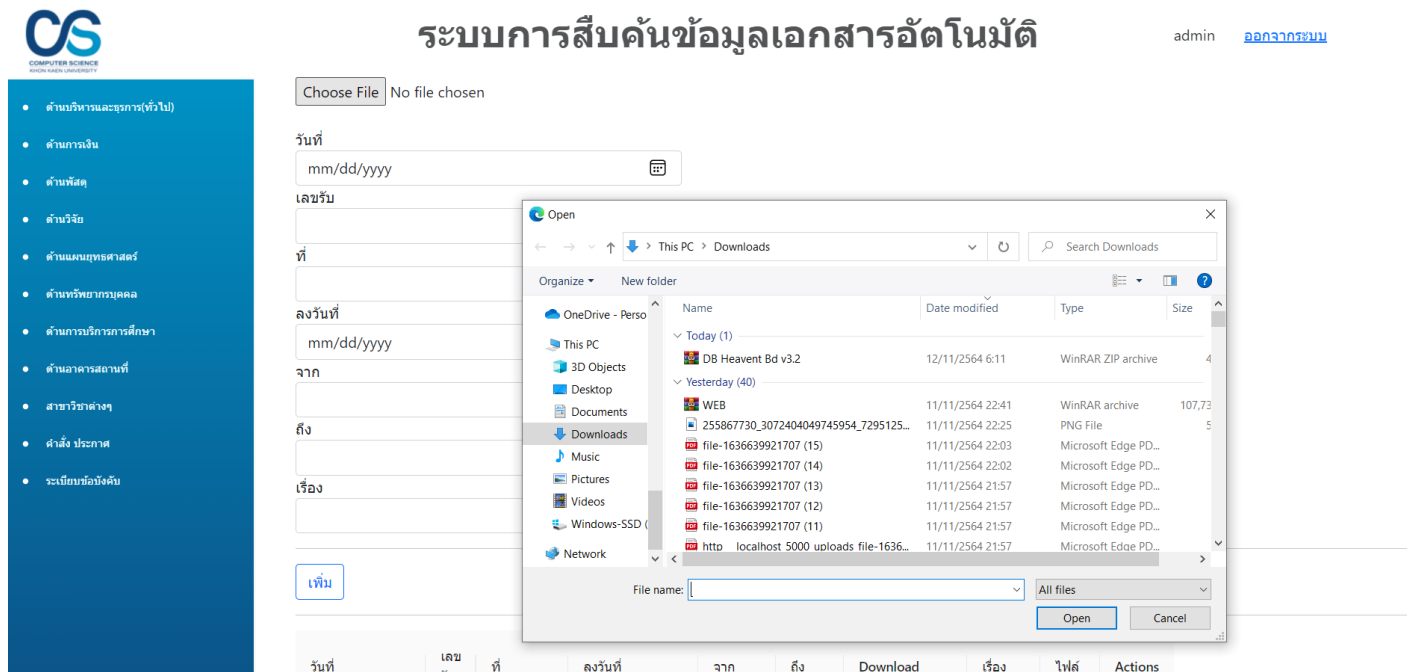
ถึง

เรื่อง

เพิ่ม

ภาพที่ 16 หน้าใส่ข้อมูลเกี่ยวกับรายละเอียดเอกสารส่งออก

จาก ภาพที่ 17 หน้าแสดงการเลือกไฟล์เอกสารที่จะนำเข้าสู่ระบบ



ภาพที่ 17 หน้าเลือกไฟล์เอกสารเข้าสู่ระบบ

จาก ภาพที่ 18 เมื่อไฟล์เอกสารเข้าสู่ระบบแล้วจะแสดงผลพอร์ออกมาในช่องใส่รายละเอียดสามารถตรวจสอบความถูกต้องและแก้ไขเพื่อยืนยันความถูกต้อง



- ด้านบริหารและธุรการ(ทั่วไป)
- ด้านการเงิน
- ด้านพัสดุ
- ด้านวิจัย
- ด้านแผนยุทธศาสตร์
- ด้านทรัพยากรบุคคล
- ด้านการบริการการศึกษา
- ด้านอาคารสถานที่
- สาขาวิชาต่างๆ
- คำสั่ง ประกาศ
- ระเบียบข้อบังคับ

ระบบการสืบค้นข้อมูลเอกสารอัตโนมัติ

admin [ออกจากระบบ](#)

Choose File

เชิญ อ.สมณฑา.pdf

วันที่

11/05/2021

เลขรับ

0020

ที่

สธ0850/97

ลงวันที่

11/20/2021

จาก

ร.ร อนุบาลขอนแก่น

ถึง

หัวหน้าสาขา


เรื่อง

เชิญ อ.สมณฑา เป็นวิทยากร 21 ม.ค.63

เพิ่ม

ภาพที่ 18 แสดงข้อความที่ผ่านระบบออกมาแสดงเพื่อ แก้ไข และยืนยัน

จาก ภาพที่ 19 เมื่อยืนยันเรียบร้อยแล้วระบบจะทำการจัดเก็บลงฐานข้อมูลและแสดงผลลัพธ์ออกมาในรูปแบบของตาราง



- ด้านบริหารและธุรการ(ทั่วไป)
- ด้านการเงิน
- ด้านพัสดุ
- ด้านวิจัย
- ด้านแผนยุทธศาสตร์
- ด้านทรัพยากรบุคคล
- ด้านการบริการการศึกษา
- ด้านอาคารสถานที่
- สาขาริชาต่างๆ
- คำสั่ง ประกาศ
- ระเบียบข้อบังคับ

ระบบการสืบค้นข้อมูลเอกสารอัตโนมัติ

ลงวันที่
11/20/2021

จาก
ร.ร อนุบาลขอนแก่น

ถึง
หัวหน้าสาขา



เรื่อง
เชิญ อ.สมเดช เป็นวิทยากร 21 ม.ค.63

✓

Create เชิญ อ.สมเดช เป็นวิทยากร
 21 ม.ค.63 Success

✕

เพิ่ม

วันที่	เลข รับ	ที่	ลงวันที่	จาก	ถึง	Download	เรื่อง	ไฟล์	Actions
2021-11-05T00:00:00.000Z	20	สธ0850/97	2021-11-20T00:00:00.000Z	ร.ร อนุบาลขอนแก่น	หัวหน้าสาขา	ดาวน์โหลด	เชิญ อ.สมเดช เป็นวิทยากร 21 ม.ค.63	 	

ภาพที่ 19 แสดงผลของการนำเอกสารเข้าสู่ระบบ



ระเบียบสำนักนายกรัฐมนตรี
ว่าด้วย การจัดทำแผนกา...

ภาพที่ 20 ตัวอย่างการ์ดด้านบริหารและบริการทั่วไป



ระเบียบสำนักนายกรัฐมนตรี
ว่าด้วย การจัดทำแผนกา...

ภาพที่ 21 ตัวอย่างการ์ดด้านการเงิน



ระเบียบสำนักนายกรัฐมนตรี
ว่าด้วย การจัดทำแผนกา...

ภาพที่ 22 ตัวอย่างการ์ดด้านพัสดุ



เรียบเรียงสำนักนายกรัฐมนตรี
ว่าด้วย การจัดทำแผนกา...

ภาพที่ 23 ตัวอย่างการ์ดด้านวิจัย



เรียบเรียงสำนักนายกรัฐมนตรี
ว่าด้วย การจัดทำแผนกา...

ภาพที่ 24 ตัวอย่างการ์ดด้านแผนยุทธศาสตร์



ระเบียบสำนักนายกรัฐมนตรี
ว่าด้วย การจัดทำแผนกา...

ภาพที่ 25 ตัวอย่างการ์ดด้านทรัพยากรบุคคล



ระเบียบสำนักนายกรัฐมนตรี
ว่าด้วย การจัดทำแผนกา...

ภาพที่ 26 ตัวอย่างการ์ดด้านบริการการศึกษา



ระเบียบสำนักนายกรัฐมนตรี
ว่าด้วย การจัดทำแผนกา...

ภาพที่ 27 ตัวอย่างการ์ดด้านอาคารสถานที่



ระเบียบสำนักนายกรัฐมนตรี
ว่าด้วย การจัดทำแผนกา...

ภาพที่ 28 ตัวอย่างการ์ดด้านสาขาวิชาต่างๆ



ระเบียบสำนักนายกรัฐมนตรี
ว่าด้วย การจัดทำแผนกา...

ภาพที่ 29 ตัวอย่างการ์ดด้านคำสั่ง ประกาศ



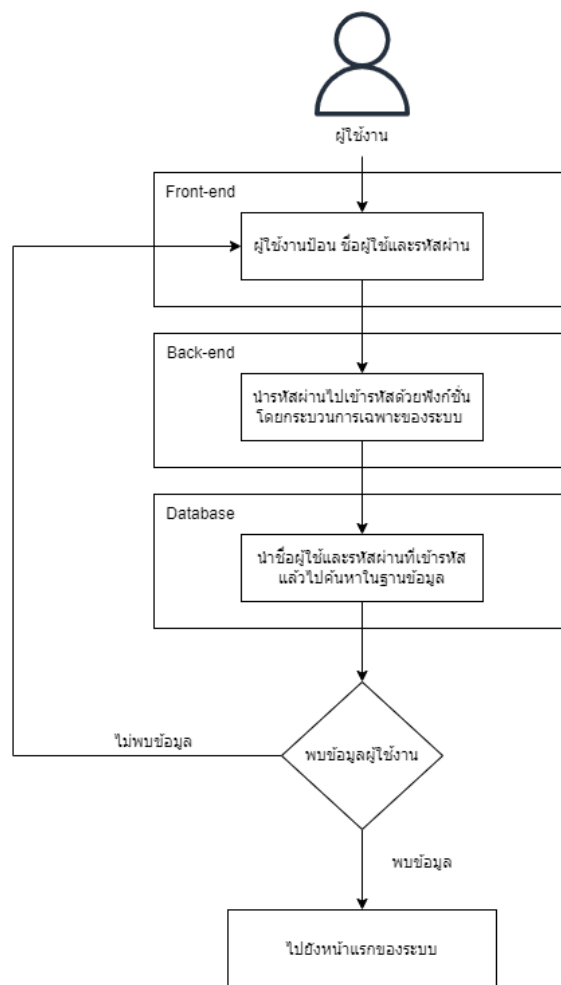
ระเบียบสำนักนายกรัฐมนตรี
ว่าด้วย การจัดทำแผนกา...

ภาพที่ 30 ตัวอย่างการ์ดด้านระเบียบข้อบังคับ

5. ขั้นตอนการทำงานของระบบ

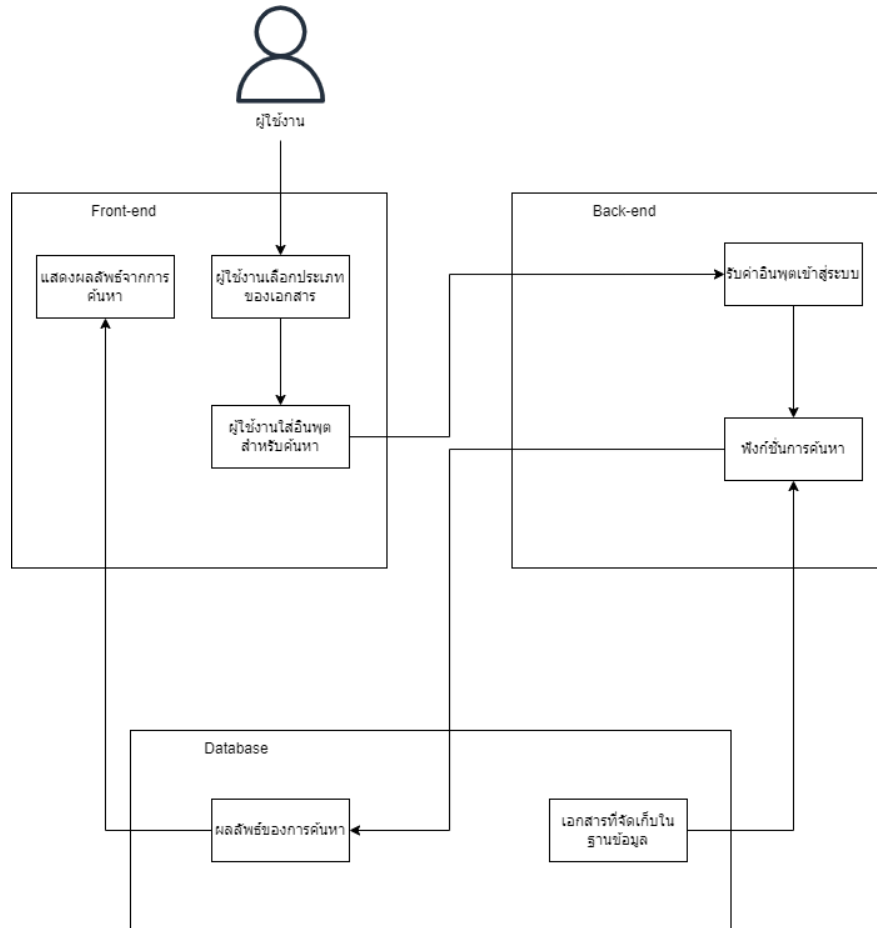
5.1 ขั้นตอนการทำงานของระบบลือคอิน

จาก ภาพที่ 31 จะแสดงถึงการทำงานของระบบลือคอินซึ่งจะเริ่มจากผู้ผู้ใช้ใส่ชื่อผู้ใช้และรหัสผ่านหลังจากนั้นจะถูกส่งไปในระบบ ระบบจะทำการนำรหัสมาถอดรหัสผ่านฟังก์ชันเฉพาะของระบบ หลังจากที่ถูกถอดรหัสแล้วระบบจะนำชื่อผู้ใช้และรหัสที่ถูกถอดรหัสแล้วไปค้นหาในฐานข้อมูล ถ้าพบว่าตรงกับที่บันทึกไว้ในฐานข้อมูลระบบจะนำไปสู่หน้าเริ่มต้นของระบบหากไม่ตรงกับข้อมูลที่บันทึกไว้ในฐานข้อมูลจะไม่สามารถเข้าสู่หน้าเริ่มต้นได้



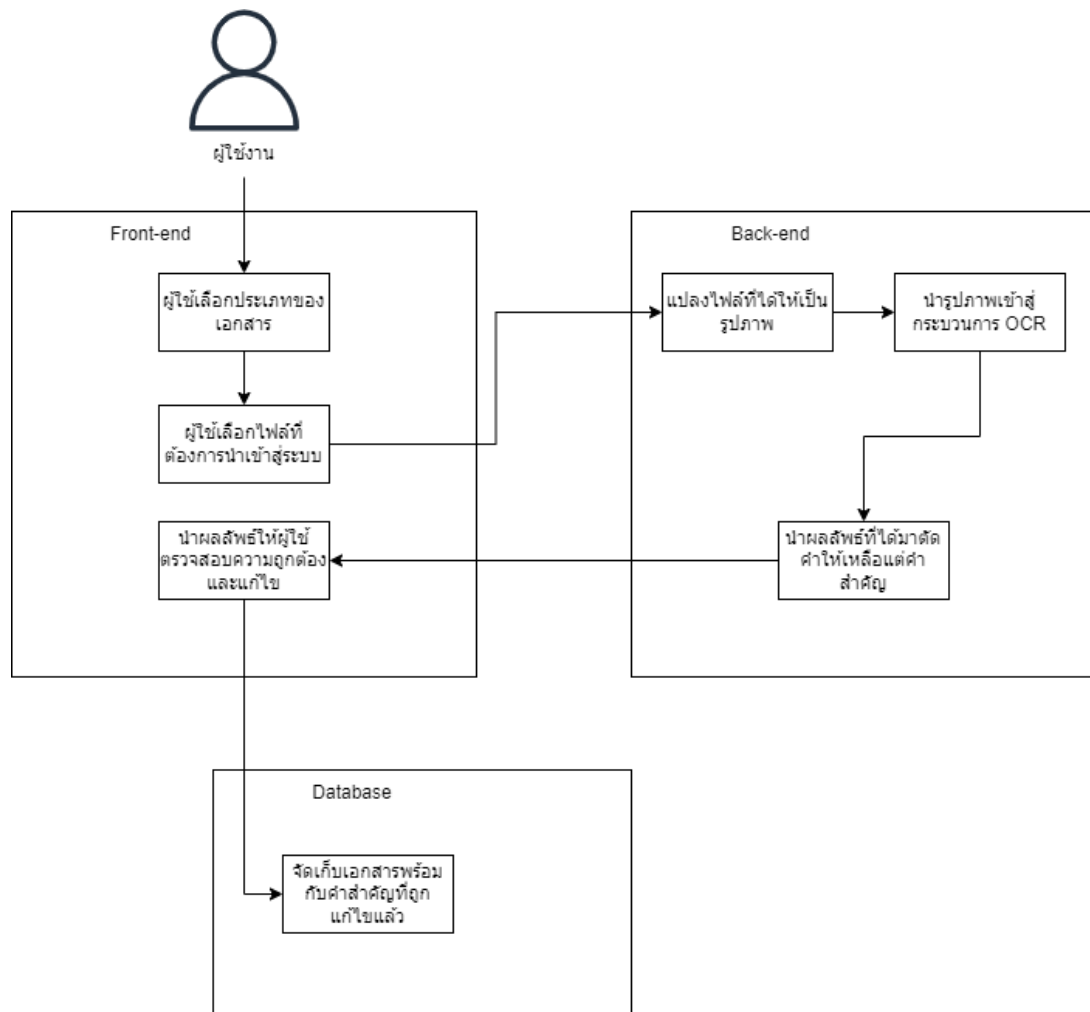
ภาพที่ 31 ขั้นตอนการทำงานของระบบลือคอิน

จาก ภาพที่ 32 จะแสดงถึงการทำงานของระบบสืบค้นโดยผู้ใช้ต้องเลือกประเภทของเอกสารที่ต้องการจะสืบค้นหลังจากนั้นผู้ใช้ใส่คำสำคัญลงในอินพุตของระบบ ระบบจะนำคำสำคัญนั้นไปเปรียบเทียบกับคำสำคัญที่ถูกจัดเก็บไว้ในฐานข้อมูล แล้วจะนำผลลัพธ์ที่ได้ออกมาแสดง



ภาพที่ 32 ขั้นตอนการทำงานของระบบสืบค้น

จาก ภาพที่ 33 จะแสดงถึงการทำงานของระบบนำเข้าเอกสาร โดยใช้การเลือกประเภทของเอกสาร ผู้ใช้จะต้องเลือกไฟล์เอกสารที่ต้องการจัดเก็บระบบจะนำไฟล์ที่ได้นั้นแปลงเป็นรูปภาพเพื่อนำรูปภาพนั้นมาเข้าสู่กระบวนการสกัดคำ เมื่อสกัดคำเสร็จแล้วจะนำผลลัพธ์ที่ได้ไปตัดคำ โดยจะเหลือไว้เพียงคำสำคัญที่ใช้ในการค้นหา โดยจะนำคำสำคัญที่ได้นี้ส่งกลับไปให้ผู้ใช้งานตรวจสอบความถูกต้องและบันทึกลงฐานข้อมูลพร้อมกับไฟล์เอกสาร



ภาพที่ 33 ขั้นตอนการทำงานของระบบจัดเก็บเอกสาร

บทที่ 5

ผลสรุป

1. สรุปผลการดำเนินโครงการ

จากการดำเนินโครงการ สิ่งที่ได้ดำเนินไปแล้ว คือ การสกัดข้อความที่เป็นคำสำคัญจากไฟล์เอกสารสำหรับการใช้ในการค้นหาเอกสารและสามารถเช็คความถูกต้องของข้อความที่ถูกสกัดออกมาแล้วนำไปแก้ไขได้หากผิดพลาดก่อนจะบันทึกข้อความลงระบบ

2. ข้อจำกัดของระบบ

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

3. ปัญหาอุปสรรค และ แนวทางแก้ไข

การจัดเก็บสารต้องใช้เวลา เพราะเอกสารทั้งหมดมีจำนวนมาก

4. ข้อเสนอแนะ ในการพัฒนาต่อไป

จากผลการพัฒนาระบบโอซีอาร์เพื่อนำมาใช้งานกับระบบเอกสาร พบว่าการใช้ระบบโอซีอาร์กับเอกสารที่ใช้ภาษาอังกฤษนั้นได้ผลลัพธ์ที่ค่อนข้างดี แต่พอนำมาประยุกต์ใช้กับเอกสารภาษาไทยนั้นได้ผลลัพธ์ไม่เป็นที่น่าพอใจนัก จึงควรพัฒนาด้วยการนำเทคโนโลยี การรู้จำมาใช้ในการอ่านตัวอักษร มีประสิทธิภาพและความแม่นยำมากยิ่งขึ้น

เอกสารอ้างอิง

- [1] วีรพล มั่นสอานินาท. (2008). ระบบสารสนเทศสำหรับการจัดการข้อมูลอุตสาหกรรม. สาขาวิชาการจัดการงานวิศวกรรม ภาควิชาวิศวกรรมอุตสาหกรรมและการจัดการ บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร
- [2] พรศิริ ภาวุธญาณชัย. (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.
- [7] MERN Stack คืออะไร? ,<https://medium.com/nexthopthai/mern-stack-%E0%B8%84%E0%B8%B7%E0%B8%AD%E0%B8%AD%E0%B8%B0%E0%B9%84%E0%B8%A3-e651592206ce>, 7 August 2019.

ประวัติผู้วิจัย

ชื่อ-นามสกุล ปิ่นธร 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

ลงชื่อผู้ทำโครงการงาน *อัษฎ์*
(นาย)นางสาว นายปิ่นธร วฒิอากรณ
วันที่ 27/1/2565

ลงชื่อผู้ทำโครงการงาน *วิรัชพงศ์*
(นาย)นางสาว นายรัชชานนท์ ศิริสาร
วันที่ 27/1/2565

การตรวจสอบจากอาจารย์ที่ปรึกษาโครงการงาน

.....
.....

(ลงชื่อ)

(.....)

วันที่...../...../.....