

โครงการพัฒนาระบบจัดเก็บข้อมูลขนาดใหญ่ (Big Data) ระยะที่ 4

กรมศิลปากร



เอกสารรายงานฉบับสมบูรณ์

โครงการพัฒนาระบบจัดเก็บข้อมูลขนาดใหญ่ (Big Data) ระยะที่ 4

มหาวิทยาลัยธรรมศาสตร์

โดยสำนักงานศูนย์วิจัยและให้คำปรึกษาแห่งมหาวิทยาลัยธรรมศาสตร์

สัญญาเลขที่ 103/2565 ลงวันที่ 7 มิถุนายน 2565

โครงการพัฒนาระบบจัดเก็บข้อมูลขนาดใหญ่ (Big Data) ระยะที่ 4

กรมศิลปากร



เอกสารรายงานฉบับสมบูรณ์

โครงการพัฒนาระบบจัดเก็บข้อมูลขนาดใหญ่ (Big Data) ระยะที่ 4

มหาวิทยาลัยธรรมศาสตร์

โดยสำนักงานศูนย์วิจัยและให้คำปรึกษาแห่งมหาวิทยาลัยธรรมศาสตร์

สัญญาเลขที่ 103/2565 ลงวันที่ 7 มิถุนายน 2565

สารบัญ

1.	ภาพรวมของระบบ	1
2.	ระบบสืบค้นข้อมูลด้วยคำสำคัญ	3
2.1.	ภาพรวมและแนวทางในการพัฒนาโมเดลสืบค้นข้อมูลด้วยคำสำคัญ	3
2.1.1.	Keyword และ Relate Word Mapping	7
2.1.2.	Keyword Indexing	8
2.1.3.	Relate Word Indexing	9
2.2.	การพัฒนาโมเดลสืบค้นข้อมูลด้วยคำสำคัญ	10
2.2.1.	การเตรียมข้อมูล	10
2.2.2.	การฝึกสอนข้อมูล	13
2.2.3.	การแปลง Tag สำหรับการฝึกสอน	14
2.2.4.	ขั้นตอนและการดำเนินการพัฒนาโมเดลสืบค้นข้อมูลด้วยคำสำคัญ	19
2.2.5.	การปรับปรุงโมเดลในการสร้างแท็กสำหรับคีย์เวิร์ดจากประโยคในฐานข้อมูล	21
2.2.6.	การสร้างโมเดลในการสกัดคีย์เวิร์ดด้วย Named Entity Recognition แบบหลายแท็กด้วย เทคนิค Ensemble classification	31
2.2.7.	ผลการทดลองจากการฝึกสอนโมเดล	33
2.2.8.	ตัวอย่างแสดงข้อมูล Keyword ที่ได้จากการจำแนก Tag	38
2.2.9.	ปัญหาที่เกิดขึ้นและการแก้ไขปัญหา	39
3.	ระบบสืบค้นข้อมูลโบราณสถานด้วยภาพถ่าย	40
3.1.	การออกแบบระบบ	40
3.2.	การพัฒนาโมดูล	41
3.2.1.	โมดูลการเข้ารหัสภาพ (Image Encoding module)	41
3.2.2.	โมดูลการค้นหาข้อมูลภาพ	47
3.2.3.	โมดูลจัดอันดับ (Ranking module)	54
4.	การเพิ่มชุดข้อมูลใหม่ในระบบและการจัดการฐานข้อมูล	64
4.1.	การเพิ่มฐานข้อมูลงานอนุรักษ์	64
4.2.	การเพิ่มการค้นหารูปภาพโบราณสถาน	69

4.3.	การปรับปรุงฐานข้อมูลการค้นหาทั้งหมด.....	76
5.	ระบบวิเคราะห์พฤติกรรมผู้ใช้งาน	83
5.1.	การออกแบบระบบวิเคราะห์พฤติกรรมผู้ใช้งาน.....	83
5.2.	การพัฒนาระบบวิเคราะห์พฤติกรรมผู้ใช้งาน	84
6.	ระบบรายงานข้อมูลสถิติช่อง YouTube ของกรมศิลปากร	87
6.1.	การออกแบบระบบรายงานข้อมูลสถิติช่อง YouTube ของกรมศิลปากร.....	87
6.1.1.	การออกแบบสถาปัตยกรรมของระบบ	87
1.1.1.	การพัฒนาฐานข้อมูลของระบบ	88
6.1.2.	การพัฒนาหน้ารายงาน	89
6.2.	การพัฒนาระบบรายงานข้อมูลสถิติช่อง YouTube ของกรมศิลปากร.....	92
6.2.1.	หน้าหลักของระบบจัดการรายงานสถิติช่อง YouTube.....	92
6.2.2.	หน้าจัดการช่องของระบบจัดการรายงานสถิติช่อง YouTube.....	93
6.2.3.	หน้าจัดการสถิติช่องของระบบจัดการรายงานสถิติช่อง YouTube.....	93
6.2.4.	หน้าจัดการวิดีโอของระบบจัดการรายงานสถิติช่อง YouTube	94
6.2.5.	หน้าจัดการสถิติวิดีโอของระบบจัดการรายงานสถิติช่อง YouTube	94
7.	การอบรมถ่ายทอดองค์ความรู้ วิธีการใช้งาน และการดูแลระบบ	96

สารบัญรูปภาพ

รูปที่ 1 ภาพรวมการทำงานของระบบการค้นหาข้อมูลกรมศิลปากร.....	1
รูปที่ 2 ตัวอย่าง input ของการสร้างตาราง TextLabel.....	6
รูปที่ 3 ผลลัพธ์ related word ที่ผ่านการทำ Relate Word Training Model	7
รูปที่ 4 Keyword และ Relate Word mapping	8
รูปที่ 5 ตัวอย่าง Keyword Indexing	8
รูปที่ 6 กระบวนการสร้าง Relate Word Indexing	9
รูปที่ 7 ตัวอย่าง Relate Word Indexing	9
รูปที่ 8 ตัวอย่างคำศัพท์ที่เพิ่ม.....	12
รูปที่ 9 ตัวอย่างชุดคำสั่งที่ใช้ในการเพิ่มคำศัพท์.....	12
รูปที่ 10 รูปแบบของข้อมูล	13
รูปที่ 11 การแปลงข้อมูลเพื่อให้เหมาะสมในการฝึกสอน	14
รูปที่ 12 การสร้างโมเดลในการสกัดชื่อเวิรด์ด้วย Named Entity Recognition แบบหลายแท็กด้วยเทคนิค Ensemble classification.....	31
รูปที่ 13 ตัวอย่างของ Keyword ในแต่ละ Tag.....	38
รูปที่ 14 ตัวอย่างของ Keyword ในแต่ละ Tag.....	39
รูปที่ 15 ระบบสืบค้นข้อมูลโบราณสถานด้วยภาพถ่าย	40
รูปที่ 16 แสดงตัวอย่างของภาพถ่ายโบราณสถานที่มีลักษณะเฉพาะของข้อมูล	42
รูปที่ 17 โครงข่ายประสาทเทียม CNN สำหรับการเรียนรู้เพื่อสกัดลักษณะเด่น	44
รูปที่ 18 แสดงสถาปัตยกรรมโครงข่ายประสาทเทียมแบบ Multi-task CNN	45
รูปที่ 19 สถาปัตยกรรมโครงข่ายประสาทเทียม ResNet50	47
รูปที่ 20 การสร้างดัชนีเพื่อสืบค้นข้อมูลด้วยต้นไม้ KD tree	48
รูปที่ 21 จำนวนภาพของโบราณสถาน	54
รูปที่ 22 โครงข่ายประสาทเทียม DELF.....	55
รูปที่ 23 แสดงผลลัพธ์การจับคู่ keypoints ของภาพที่มาจากโบราณสถานแห่งเดียวกัน	56
รูปที่ 24 แสดงผลลัพธ์การจับคู่ keypoints ของภาพที่มาจากโบราณสถานที่แตกต่างกัน	56
รูปที่ 25 ผลลัพธ์การสืบค้นด้วยภาพควิรี “เสาชิงช้า”.....	57

รูปที่ 26 ผลลัพธ์การสืบค้นด้วยภาพควีรี “สะพานม้ฆวาวฝรั่งเศส”	58
รูปที่ 27 ผลลัพธ์การสืบค้นด้วยภาพทดสอบ “บ้านอับดุลลาฮิม”	59
รูปที่ 28 ผลลัพธ์การสืบค้นด้วยภาพทดสอบ “วัดเขมาภิถาราม”	59
รูปที่ 29 แสดงผลลัพธ์ของการทดสอบระบบโดยชุดข้อมูลทดสอบ 219 ภาพ บนค่าเฉลี่ยของ Precision@K.....	60
รูปที่ 30 แสดงการเปรียบเทียบผลลัพธ์ระหว่างการใช้ single attention, adaptive attention และ no attention ในโมเดล PAN.....	62
รูปที่ 31 แสดงตัวอย่างผลลัพธ์ 7 ลำดับแรกที่โมเดลสามารถจับคู่ได้ โดยภาพคอลัมน์ซ้ายมือสุดเป็นภาพอินพุต ขณะที่ภาพในกรอบสีเขียวเป็นภาพของสถานที่เดียวกันกับภาพอินพุต ขณะที่ภาพกรอบสีแดงแสดงผลลัพธ์ที่คาดเคลื่อน ในชุดข้อมูล TAH-10.....	62
รูปที่ 32 แสดงตัวอย่างผลลัพธ์ 7 ลำดับแรกที่โมเดลสามารถจับคู่ได้ โดยภาพคอลัมน์ซ้ายมือสุดเป็นภาพอินพุต ขณะที่ภาพในกรอบสีเขียวเป็นภาพของสถานที่เดียวกันกับภาพอินพุต ขณะที่ภาพกรอบสีแดงแสดงผลลัพธ์ที่คาดเคลื่อน ในชุดข้อมูล Oxford5k และ Paris6k ตามลำดับ	63
รูปที่ 33 แสดงข้อมูลรูปแบบโครงสร้าง web site ในระบบเดิม	64
รูปที่ 34 แสดงข้อมูลรูปแบบโครงสร้าง web site ในระบบใหม่ได้เพิ่มฐานข้อมูลงานอนุรักษ์	66
รูปที่ 35 แสดงการค้นหารูปภาพระบบใหม่	71
รูปที่ 36 เริ่มต้นค้นหารูปภาพ	74
รูปที่ 37 เลือกรูปภาพ	74
รูปที่ 38 แสดงผลการค้นหารูปภาพโบราณวัตถุ	75
รูปที่ 39 แสดงผลการค้นหารูปภาพโบราณสถาน	75
รูปที่ 40 แสดงข้อมูลรูปแบบโครงสร้าง web site ในระบบเดิม	76
รูปที่ 41 แสดงข้อมูลรูปแบบโครงสร้าง web site ในระบบใหม่.....	77
รูปที่ 42 แสดงข้อมูลรูปแบบโครงสร้าง web site ในระบบใหม่.....	78
รูปที่ 43 แสดงผลลัพธ์จากการค้นหาตารางข้อมูล keyword_search.....	79
รูปที่ 44 แสดงขั้นตอนการทำงานของการค้นหา keyword_search.....	79
รูปที่ 45 แสดงขั้นตอนการสร้างตาราง related_word	80
รูปที่ 46 ตัวอย่างข้อมูลของตารางข้อมูล new_relative_word	81
รูปที่ 47 แสดงขั้นตอนการสร้างตาราง new_relative_word	82

รูปที่ 48 สถาปัตยกรรมของระบบวิเคราะห์พฤติกรรมผู้ใช้งาน	83
รูปที่ 49 การเข้าคิวของ Message ในการวิเคราะห์พฤติกรรมของผู้ใช้งาน	85
รูปที่ 50 การเก็บข้อมูลภายใน Message	86
รูปที่ 51 การเก็บข้อมูลพฤติกรรมผู้ใช้งานลงในฐานข้อมูล	86
รูปที่ 52 สถาปัตยกรรมของระบบรายงานสถิติช่อง YouTube.....	87
รูปที่ 53 การออกแบบฐานข้อมูลของระบบรายงานสถิติช่อง YouTube.....	88
รูปที่ 54 หน้ารายงานสถิติช่อง YouTube ในภาพรวม.....	89
รูปที่ 55 หน้ารายงานสถิติช่อง YouTube.....	90
รูปที่ 56 หน้ารายงานสถิติของวิดีโอ	91
รูปที่ 57 หน้าหลักของระบบจัดการรายงานสถิติช่อง YouTube.....	92
รูปที่ 58 หน้าจัดการช่องของระบบจัดการรายงานสถิติช่อง YouTube.....	93
รูปที่ 59 หน้าจัดการสถิติช่องของระบบจัดการรายงานสถิติช่อง YouTube.....	94
รูปที่ 60 หน้าจัดการวิดีโอของระบบจัดการรายงานสถิติช่อง YouTube	94
รูปที่ 61 หน้าจัดการสถิติวิดีโอของระบบจัดการรายงานสถิติช่อง YouTube	95
รูปที่ 62 การจัดการอบรม.....	97

สารบัญตาราง

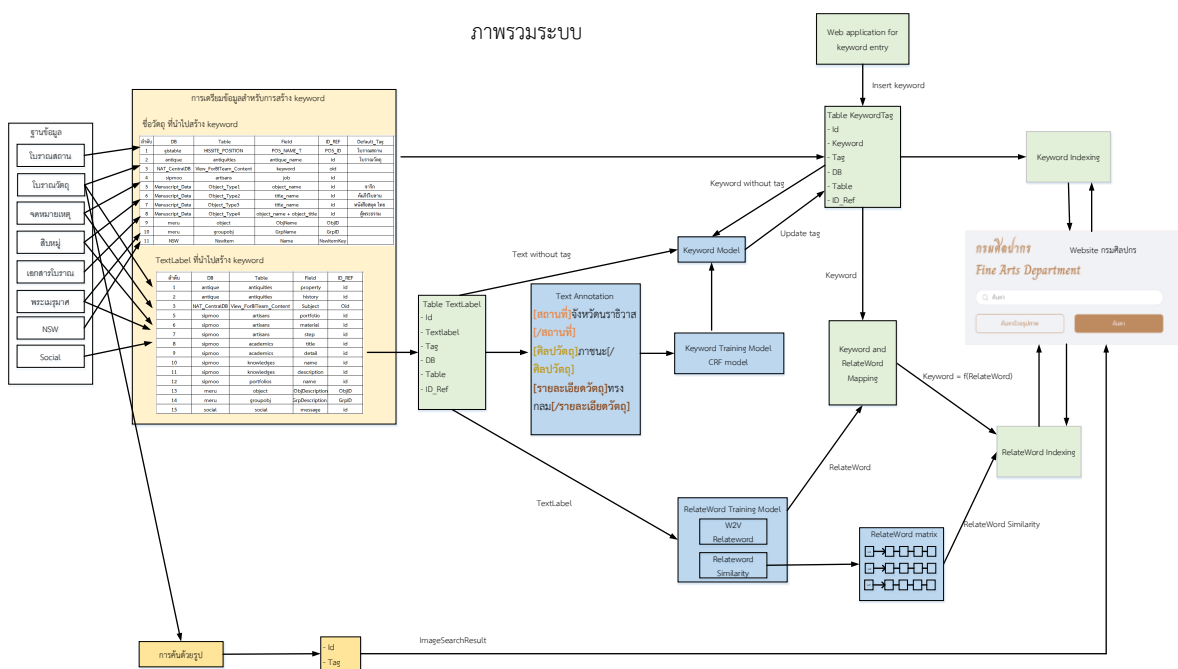
ตารางที่ 1 ชื่อวัตถุที่นำไปสร้าง keyword ที่มาจากฐานข้อมูล ตารางและฟิลด์ต่าง ๆ ของกรมฯ	3
ตารางที่ 2 Text Label ที่นำไปสร้าง keyword ที่มาจากฐานข้อมูล ตารางและฟิลด์ ต่าง ๆ ของกรมฯ	4
ตารางที่ 3 ตัวอย่างการสร้าง keyword_tag แบบ Default	4
ตารางที่ 4 การสร้าง Label หรือ Tag สำหรับคำสำคัญ	10
ตารางที่ 5 ข้อมูลที่ใช้ในการฝึกสอน	13
ตารางที่ 6 การแปลงข้อความให้อยู่ในรูป CoNLL2002	14
ตารางที่ 7 ตัวอย่างตาราง keyword	32
ตารางที่ 8 ตารางสรุปจำนวนคำที่ได้ในแต่ละ Tag	35
ตารางที่ 9 รายการโบราณสถานที่ถูกเลือกมาทำรายการดัชนี	49
ตารางที่ 10 แสดงการเปรียบเทียบประสิทธิภาพของโมเดลค้นคืนภาพที่นำเสนอในงานวิจัยและโมเดลต่างๆ ที่มีใช้ ในปัจจุบัน โดยเปรียบเทียบกับชุดข้อมูล benchmark ได้แก่ Oxford5k และ Paris6k ขณะที่ชุดข้อมูล โบราณสถานของกรมศิลปากร ได้แก่ TAH-100 จะมีการรวบรวมจาก 100 แหล่งโบราณสถาน	61

1. ภาพรวมของระบบ

ข้อมูลกรมศิลปากรที่ประกอบไปด้วย

- ฐานข้อมูลโบราณสถาน
- ฐานข้อมูลโบราณวัตถุ
- ฐานข้อมูลจดหมายเหตุ
- ฐานข้อมูลช่างสิบหมู่
- ฐานข้อมูลเอกสารโบราณ
- ฐานข้อมูลพระเมรุมาศ
- ฐานข้อมูล NSW
- ฐานข้อมูล Social

ภาพรวมการทำงานของระบบการค้นหาค้นหาข้อมูลกรมศิลปากรแสดงดังภาพที่ 1



รูปที่ 1 ภาพรวมการทำงานของระบบการค้นหาค้นหาข้อมูลกรมศิลปากร

จากรูปที่ 1 จากฐานข้อมูลที่ได้ ฐานข้อมูลโบราณสถาน, ฐานข้อมูลโบราณวัตถุ, ฐานข้อมูลจดหมายเหตุ, ฐานข้อมูลช่างสิบหมู่, ฐานข้อมูลเอกสารโบราณ, ฐานข้อมูลพระเมรุมาศ และ ฐานข้อมูล NSW จะถูกนำมาสร้าง Keyword โดยขั้นตอนการสร้าง Keyword จะถูกสร้างได้ 3 แบบ คือ 1) การกำหนด Keyword จากชื่อวัตถุโดยตรง 2) การกำหนด Keyword จากโมเดล CBF ที่สกัด keyword จากข้อความหรือเอกสาร และ 3) การกำหนด Keyword จากผู้ใช้ นอกจากนี้ระบบยังเพิ่มฟังก์ชันการทำงานในการหาคำที่มีความสัมพันธ์กัน (related word) ที่เอาไว้ช่วยแสดงผลลัพธ์ที่เกี่ยวข้องของคำค้นที่ผู้ใช้ค้นหา โดยวิธี word2vec จะถูกนำมาประยุกต์ใช้ในการหา related word เมื่อระบบได้ keyword และ related word แล้วทางผู้พัฒนาจะต้องไปทำการ matching ความสัมพันธ์ระหว่าง keyword กับ related word แสดงดังกระบวนการ Keyword and related word mapping และเพื่อช่วยให้การค้นมีความเร็วขึ้น ตาราง keyword และตาราง related word จะถูกไปทำการ Indexing แสดงดังกระบวนการ Keyword indexing และ related word indexing ตามลำดับ สุดท้ายระบบยังเพิ่มฟังก์ชันการค้นหาดูรูปภาพด้วย AI โดยรูปภาพที่ใช้จะมาจากฐานข้อมูล โบราณสถานและโบราณวัตถุ นอกจากนี้ยังมีฐานข้อมูล social ที่ได้มาจาก social media ของกรมศิลปากร รายละเอียดกระบวนการทำงานแสดงในหัวข้อถัดไป

2. ระบบสืบค้นข้อมูลด้วยคำสำคัญ

2.1. ภาพรวมและแนวทางในการพัฒนาโมเดลสืบค้นข้อมูลด้วยคำสำคัญ

การเตรียมข้อมูลสำหรับการสร้าง keyword

การเตรียมข้อมูลสำหรับการสร้าง Keyword แบ่งออกเป็น 2 ส่วน คือ ส่วน ชื่อวัตถุที่นำไปสร้าง keyword ที่ได้จากฐานข้อมูล ฐานข้อมูลโบราณสถาน (gistable), ฐานข้อมูลโบราณวัตถุ (antique), ฐานข้อมูลจดหมายเหตุ (NAT_CentralDB), ฐานข้อมูลช่างสิบหมู่ (sipmoo), ฐานข้อมูลเอกสารโบราณ (manuscript_Data), ฐานข้อมูลพระเมรุมาศ (meru) และ ฐานข้อมูล NSW และ ส่วน Text Label ที่นำไปสร้างวัตถุ จะมาจาก ฐานข้อมูลโบราณวัตถุ (antique) โดยใช้ตาราง antiquities และ ฟิลด์ property และ history, ฐานข้อมูลจดหมายเหตุ (NAT_CentralDB) โดยใช้ตาราง View_ForBITeam_Content ฟิลด์ subject, ฐานข้อมูลช่างสิบหมู่ (sipmoo) โดยใช้ตาราง artisans ฟิลด์ portforio, material, step , ตาราง academics ฟิลด์ title, detail, ตาราง knowledges ฟิลด์ name, decription และตาราง portforios ฟิลด์ name, ฐานข้อมูลพระเมรุมาศ (meru) ตาราง object ฟิลด์ ObjDecription และตาราง groupObj ฟิลด์ GrpDecription นอกจากนั้นยังมี ฐานข้อมูล social ตาราง social ฟิลด์ message โดยรายละเอียดของข้อมูลนำเข้าแสดงตารางที่ 1 และตารางที่

2

ลำดับ	DB	Table	Field	ID_REF	Default_Tag
1	gistable	HISSITE_POSITION	POS_NAME_T	POS_ID	โบราณสถาน
2	antique	antiquities	antique_name	id	โบราณวัตถุ
3	NAT_CentralDB	View_ForBITeam_Content	keyword	oid	
4	sipmoo	artisans	job	id	
5	Manuscript_Data	Object_Type1	object_name	id	จารึก
6	Manuscript_Data	Object_Type2	title_name	id	คัมภีร์โบราณ
7	Manuscript_Data	Object_Type3	title_name	id	หนังสือสมุด ไทย
8	Manuscript_Data	Object_Type4	object_name + object_title	id	ตู้พระธรรม
9	meru	object	ObjName	ObjID	
10	meru	groupobj	GrpName	GrpID	
11	NSW	NswItem	Name	NswItemKey	

ตารางที่ 1 ชื่อวัตถุที่นำไปสร้าง keyword ที่มาจากฐานข้อมูล ตารางและฟิลด์ต่าง ๆ ของกรมฯ

ลำดับ	DB	Table	Field	ID_REF
1	antique	antiquities	property	id
2	antique	antiquities	history	id
3	NAT_CentralDB	View_ForBITeam_Content	Subject	Oid
5	sipmoo	artisans	portfolio	id
6	sipmoo	artisans	material	id
7	sipmoo	artisans	step	id
8	sipmoo	academics	title	id
9	sipmoo	academics	detail	id
10	sipmoo	knowledges	name	id
11	sipmoo	knowledges	description	id
12	sipmoo	portfolios	name	id
13	meru	object	ObjDescription	ObjID
14	meru	groupobj	GrpDescription	GrpID
15	social	social	message	id

ตารางที่ 2 Text Label ที่นำไปสร้าง keyword ที่มาจากฐานข้อมูล ตารางและฟิลด์ ต่าง ๆ ของกรมฯ

กระบวนการทำ keyword_tag แบ่งออกเป็น 3 แบบ คือ แบบ Default keyword_tag แบบ Text without tag หรือแบบ keyword without tag และแบบ user keyword_tag โดยผู้ใช้เป็นคนกำหนด

- **แบบ Default keyword_tag**

เป็นการสร้าง keyword และ กำหนด tag ตามฐานข้อมูลที่ได้ เช่น tag โบราณสถาน โบราณวัตถุ จารึก เป็นต้น ตัวอย่างการสร้าง keyword_tag แบบ Default แสดงดังตาราง

ID	Keyword	Tag	DB	Table	ID_Ref
1	วัดพระเงิน	โบราณสถาน	gistable	HISSITE_POSITION	0000222
2	วัดเจดีย์ทอง	โบราณสถาน	gistable	HISSITE_POSITION	0003935
3	เทวสตรี	โบราณวัตถุ	antique	antiquities	1
4	ประติมากรรมรูปช้าง	โบราณวัตถุ	antique	antiquities	3

ตารางที่ 3 ตัวอย่างการสร้าง keyword_tag แบบ Default

- **แบบ Text without tag หรือแบบ keyword without tag**
เป็นการสร้าง keyword และ กำหนด tag โดยผ่าน keyword model ที่ได้จากระบบการเรียนรู้ keyword (keyword training model) ด้วยวิธี CFR
- **แบบ user keyword_tag**
เป็นการสร้าง keyword และ กำหนด tag โดยผ่าน Web application โดยผู้ใช้งานจำเป็นต้องใส่ข้อมูลให้ครบตามฟิลด์ของ Table KeywordTag

Table KeywordTag

Table KeywordTag จะมี Schema ดังนี้

NO	NAME	DESCRIPTION
1	id	Auto id
2	Keyword	keyword
3	Tag	แท็ก
4	DB	ฐานข้อมูลที่ได้ keyword เช่น ฐานข้อมูลโบราณสถาน โบราณวัตถุ เป็นต้น
5	Table	ตารางในฐานข้อมูลที่ได้ keyword
6	ID_Ref	รหัสอ้างอิงที่ใช้ในการอ้างอิงกับไปยังฐานข้อมูลของกรมศิลปากร

Table TextLabel

Table TextTag จะมี Schema ดังนี้

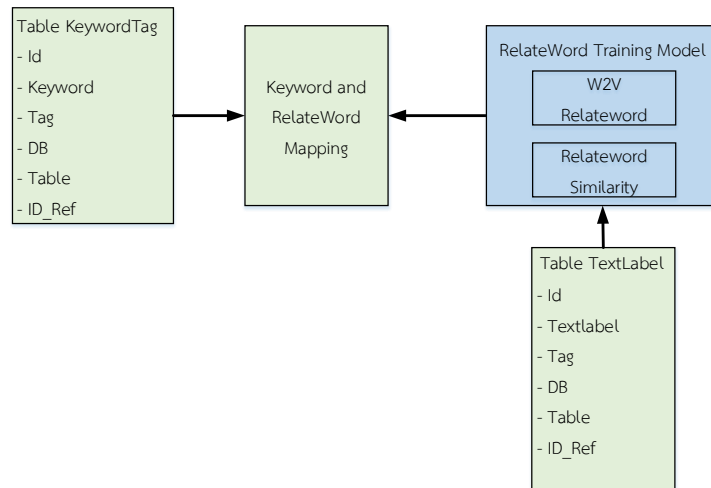
NO	NAME	DESCRIPTION
1	id	Auto id
2	TextLabel	ข้อความ ข้อมูล คำอธิบายคุณลักษณะของวัตถุ
3	Tag	แท็ก
4	DB	ฐานข้อมูลที่ได้ TextLabel เช่น ฐานข้อมูลโบราณสถาน โบราณวัตถุ เป็นต้น
5	Table	ตารางในฐานข้อมูลที่ได้ TextLabel

	abc keyword	abc related_word	myindex
1	วัดพระเงิน	วัดพระนอน,วัดพระเชตุพน,วัดพระเจ้าล้านทอง,วัดพระงาม,วัดพระราม,วัดพริก,วัดพระปรมาศ,วัดพระประโทน,วัดพระ	1
2	วัดพระเจดีย์ทอง	วัดพระนอน,วัดพระปฐมเจดีย์,วัดพระปรมาศ,วัดพริก,วัดพระธาตุ,วัดพระเจ้าล้านทอง,วัดพระประโทน,วัดพระงาม,วัดพ	2
3	วัดพระเจ้าคอกหัก	วัดพระเจ้าล้านทอง,วัดพระนอน,สมเด็จพระเจ้าภคินี,วัดพระเชตุพน,วัดพระราม,สมเด็จพระเจ้าธรรม,สมเด็จพระเจ้า,วัด	3
4	วัดพระเจ้าองค์ดำ	วัดพระเจ้าล้านทอง,สมเด็จพระเจ้าภคินี,สมเด็จพระเจ้าธรรม,พระเจ้าอุทอง,วัดพระนอน,วัดพระเชตุพน,วัดพระราม,สม	4
5	วัดพระเจ้าตนหลวง	วัดพระเจ้าล้านทอง,สมเด็จพระเจ้าธรรม,สมเด็จพระเจ้าภคินี,สมเด็จพระเจ้า,สมเด็จพระเจ้าพระบรมราชินี,สมเด็จพระ	5
6	วัดพระเจ้าทันใจ	วัดพระเจ้าล้านทอง,วัดพระนอน,วัดพระเชตุพน,สมเด็จพระเจ้าภคินี,วัดพระราม,สมเด็จพระเจ้าธรรม,สมเด็จพระเจ้า,วัด	6
7	วัดพระเจ้าล้านทอง	วัดพระนอน,สมเด็จพระเจ้าภคินี,สมเด็จพระเจ้าธรรม,สมเด็จพระเจ้า,วัดพระเชตุพน,สมเด็จพระเจ้าอุกยาเธอ,วัดพริก,	7
8	วัดพระเจ้าลิ้นม่าน (ร้าง)	วัดพระเจ้าล้านทอง,วัดพระนอน,สมเด็จพระเจ้าภคินี,สมเด็จพระเจ้าธรรม,วัดพระเชตุพน,สมเด็จพระเจ้า,วัดพระราม,ส	8
9	วัดพระเจ้าเหลื่อม	วัดพระเจ้าล้านทอง,วัดพระนอน,สมเด็จพระเจ้าธรรม,สมเด็จพระเจ้าภคินี,วัดพระเชตุพน,วัดพระราม,สมเด็จพระเจ้า,วัด	9
10	วัดใต้เห็ง	วัดญาณเสน,วัด,วัดส้ม,วัดแค,วัดบัว,วัดขุนแผน,วัดตะกุก,วัดอโยธยา,วัดอุโบสถ,วัดสมุน,	10
11	วัดพระจำอมเมียง(ร้าง)	วัดพระเจ้าล้านทอง,วัดพระนอน,วัดพระราม,พระเจ้าอุทอง,วัดพระเชตุพน,สมเด็จพระเจ้าธรรม,สมเด็จพระเจ้าภคินี,วัด	11
12	วัดพระเชตุพนวิมลมังคลาราม	วัดพระเชตุพน,พระเชตุพน,วัดอรุณราชวราราม,วัดอรุณ,วัดพระราม,วัดพระนอน,วัดสวนโมกขพลาราม,พระที่นั่งธรรม	12
13	วัดพระไชยเชษฐาธิราช	วัดพระนอน,วัดราชบูรณะ,วัดพระเชตุพน,วัดพระแก้ว,วัดพระสิงห์,สมเด็จพระบรมราชินี,วัดพระราม,สมเด็จพระ	13
14	วัดปากแขง	วัดญาณเสน,วัด,วัดป่าแสง,วัดบัว,วัดแค,วัดส้ม,วัดป่าบาง,วัดลาว,วัดอโยธยา,วัดสมุน,	14

รูปที่ 3 ผลลัพธ์ related word ที่ผ่านการทำ Relate Word Training Model

2.1.1. Keyword และ Relate Word Mapping

ตาราง KeywordTag จะประกอบไปด้วยเซตของ keyword (Keyword Tag) จำนวน 293,708 คำ และขั้นตอนการทำ Relate Word Training Model ก็จะประกอบไปด้วยเซตของ keyword (Relate Word) จำนวน 539,744 คำ และเนื่องจาก keyword (Relate Word) เป็นคีย์ที่ได้จากความสัมพันธ์เชิงสถิติขอประโยคในข้อความทำให้มีจำนวนมากกว่า (Keyword Tag) ซึ่ง keyword (Relate Word) จะ mapping keyword (Keyword Tag) ผ่านค่า DB, Table, ID_ref จากทั้ง Table KeywordTag และ Table TextLabel โดยการ mapping จะเป็นการบอกว่า keyword (KeywordTag) มีความสัมพันธ์กับ keyword (Relate Word)



รูปที่ 4 Keyword และ Relate Word mapping

2.1.2. Keyword Indexing

Keyword Indexing เป็นกระบวนการสร้าง indexing ที่ช่วยในการค้นหาข้อมูลให้มีความรวดเร็วขึ้น โดย Keyword Indexing จะถูกสร้างมาจาก Table keywordTag โดยที่ keyword 1 keyword สามารถเกิดขึ้นได้หลายๆ ฐานข้อมูล และในแต่ละฐานข้อมูลก็ได้หลายวัตถุ และแต่ละวัตถุก็จะเก็บ tag เอาไว้ด้วย ตัวอย่าง Keyword Indexing

```

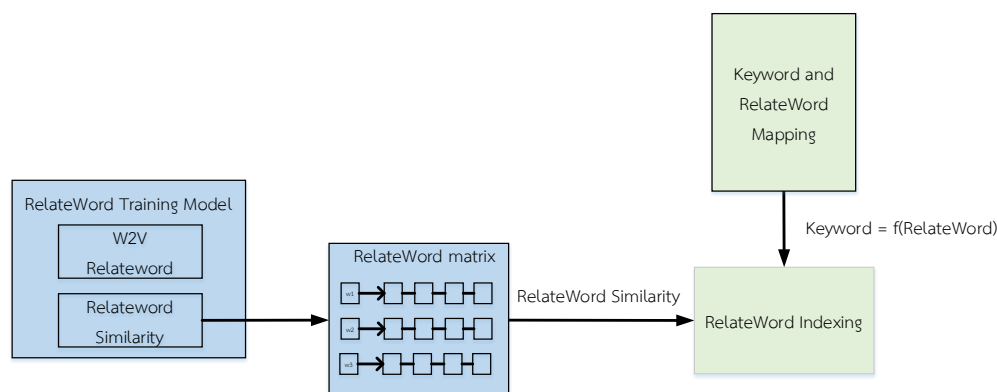
_id: ObjectId("614e9cba585f89c993594fa8")
> __meru: Array
> __antique: Array
> __NSW: Array
> __archives: Array
> __sipmoo: Array
  keyword: "วัดพระเงิน"
> __hissite: Array
  > 0: Object
  > 1: Object
> __ancient_document: Array

_id: ObjectId("614e9cc4585f89c993594faa")
> __meru: Array
> __antique: Array
> __NSW: Array
> __archives: Array
> __sipmoo: Array
  keyword: "วัดพระเจดีย์ทอง"
> __hissite: Array
> __ancient_document: Array
  
```

รูปที่ 5 ตัวอย่าง Keyword Indexing

2.1.3. Relate Word Indexing

RelateWord Indexing เป็นกระบวนการสร้าง indexing ที่ช่วยในการค้นหาข้อมูลให้มีความรวดเร็วขึ้น โดย RelateWord Indexing จะถูกสร้างมาจาก RelateWord Similarity ที่บอกว่า keyword (RelateWord) มีความสัมพันธ์เชิงสถิติกัน โดยนำ keyword ที่มีความสัมพันธ์กันถูกนำมาจัดเก็บที่ RelateWord matrix โดยภาพการทำงานแสดงดังรูปที่ 5 จากนั้น ตัวอย่าง RelateWord Indexing แสดงดังรูปที่ 6



รูปที่ 6 กระบวนการสร้าง Relate Word Indexing

```
_id: ObjectId("61852319dbea692e2430a276")
keyword: "สัตว์ป่าน้ำ"
  list: Array
    0: Object
      ID_labels: "153"
      relative_keyword: "ยี่นตรีกังค"
      ID_Ref: "157"
      Table: "antiquities"
      ID_keyword: "12428"
      _id: ObjectId("614c17f8585f8963bafa3b6d")
      Tag: "โบราณวัตถุ"
      detail: Object
        DB: "antique"
    1: Object
    2: Object
    3: Object
    4: Object
    5: Object
    6: Object
    7: Object
```

รูปที่ 7 ตัวอย่าง Relate Word Indexing

2.2. การพัฒนาโมเดลสืบค้นข้อมูลด้วยคำสำคัญ

2.2.1. การเตรียมข้อมูล

ในการปรับปรุงและพัฒนาโมเดลสืบค้นข้อมูลด้วยคำสำคัญในรายงานฉบับนี้ ได้มีการเพิ่มจำนวน Tag สำหรับคำสำคัญจำนวนทั้งสิ้น 85 Tag ดังรายละเอียดต่อไปนี้

ตารางที่ 4 การสร้าง Label หรือ Tag สำหรับคำสำคัญ

Tag สำหรับคำสำคัญ					
1	สถานที่	31	จดหมายเหตุ	61	เลขไปรษณีย์
2	วันเวลา	32	ฟิล์มกระจก	62	การอบรม/สัมมนา
3	ชื่อโครงการ	33	เอกสารลายลักษณ์	63	สกุลเงิน/เงินตรา
4	กิจกรรม	34	บันทึกเหตุการณ์	64	วารสาร/นิตยสาร
5	สัตว์	35	ยุคสมัย	65	หนังสือ
6	พืช	36	ก่อนประวัติศาสตร์	66	ชนเผ่าชาติพันธุ์
7	กลุ่มบุคคล/องค์กร/บริษัท	37	ทวารวดี	67	ดวงตรา/สัญลักษณ์
8	บุคคล	38	สุโขทัย	68	ราชวงศ์
9	ประติมากร	39	อยุธยา	69	พระราชทาน
10	ภัณฑารักษ์	40	รัตนโกสินทร์	70	พระเครื่อง
11	ศิลปวัตถุ	41	ธนบุรี	71	อวัยวะ
12	ปฐพีมากรรม	42	ลพบุรี	72	ข้อเสนอแนะ/ความคิดเห็น
13	ประติมากรรม	43	เขมรในประเทศไทย	73	วิชา/หลักสูตร
14	จิตรกรรม	44	ศรีวิชัย	74	อาหาร
15	นาฏศิลป์-ดนตรี	45	ล้านนา	75	แบบแปลน/แผนผัง
16	โบราณสถาน	46	ล้านช้าง	76	มาตราส่วน
17	พิพิธภัณฑ์	47	ช่วยเหลือ	77	ภัยพิบัติ
18	อนุสาวรีย์	48	โรค	78	กีฬา
19	อุทยานประวัติศาสตร์	49	กฎหมาย	79	ภาพยนตร์
20	โบราณวัตถุ	50	พระราชพิธี	80	เครื่องราช
21	รายละเอียดวัตถุ	51	ประเพณี	81	รางวัล

22	พระพุทธรูป	52	พิธีกรรม	82	เทพ/เทพเจ้า/พระเจ้า
23	วรรณกรรม	53	ยานพาหนะ	83	ซื้อ/ขาย
24	ตัวละครหรือสัตว์ใน วรรณกรรม	54	วัตถุ	84	สถาปัตยกรรม
25	เอกสารหนังสือ	55	ยารักษา	85	ข่าวสาร
26	จารึก	56	แผนที่, พิกัด		
27	สมุดไทย	57	วัน		
28	คัมภีร์ไบเบิล	58	เดือน		
29	โบราณคดี	59	ปี		
30	ประวัติศาสตร์	60	เครื่องแต่งกาย		

การเตรียมข้อมูลโดยเพิ่มคำศัพท์ที่เกี่ยวข้อง

เพิ่มคำศัพท์ที่เกี่ยวข้องเพื่อใช้ในการตัดประโยคที่ยาวๆ ให้มีความแม่นยำมากขึ้น โดยอัลกอริทึมในการตัดคำที่ใช้ คือ newmm โดยตัวอย่างคำศัพท์ที่เพิ่มแสดงดังภาพที่ 1

addkeywords - Notepad
File Edit Format View Help

นางมณีเมขลาถือลูกแก้ว
นางมณีเมขลา
นางมณีเมขลา
นางมณีวรรณ
นางมณีวรรณกมลพัฒนะ
นางมน
นางมนตรา นิลจันทร์พันธุศรี
นางมนทนี ปุษาปาคม
นางมนัส ดิถีเฮอร์ (ยะฮัมพันธ)
นางมนัสนันท์ ชันธตันธ
นางมโน จันทรอินทร์
นางมยุรี จันยานนท
นางมยุรี ปิ่นแก้ว
นางมลชา
นางมลธิพา ภูมิเพ็ง
นางมลวรรณ ไมตรีสวัสดิ์
นางมลวรรณ รองพิจิตรสรรพการ
นางม่วง
นางม้วน
นางม้วน มัธยมจันทร์

รูปที่ 8 ตัวอย่างคำศัพท์ที่เพิ่ม

ตัวอย่างชุดคำสั่งที่ใช้ในการเพิ่มคำศัพท์แสดงดังภาพที่ 2 ดังนี้

```
from pythainlp import Tokenizer
# from pythainlp.tokenize import dict_trie
from pythainlp.util import dict_trie
from pythainlp.corpus.common import thai_words

words = ["ไทยบางแก้ว", "บางแก้ว"]
custom_words_list = set(thai_words ())
## add multiple words
custom_words_list.update(words)
trie = dict_trie(dict_source=custom_words_list)

custom_tokenizer = Tokenizer (custom_dict=trie, engine='newmm')
```

รูปที่ 9 ตัวอย่างชุดคำสั่งที่ใช้ในการเพิ่มคำศัพท์

2.2.2. การฝึกสอนข้อมูล

ข้อมูลที่ใช้ในการฝึกสอนประกอบด้วยข้อความ คีย์เวิร์ดและ Tag จากผู้เชี่ยวชาญตามจำนวนต่อไปนี้

ลำดับ ที่	ชื่อข้อมูล	ชื่อไฟล์	จำนวนข้อความ
1.	ข้อความ คีย์เวิร์ดและ Tag จากผู้เชี่ยวชาญประจำเดือนมิถุนายน 2565	Keyword_bigdata_062565.xlsx	34304
2.	ข้อความ คีย์เวิร์ดและ Tag จากผู้เชี่ยวชาญประจำเดือนกรกฎาคม 2565	Keyword_bigdata_0765.xlsx	11567
3.	ข้อความ คีย์เวิร์ดและ Tag จากผู้เชี่ยวชาญประจำเดือนกรกฎาคม 2565	Keyword_bigdata_072565.xlsx	24633
4.	ข้อความ คีย์เวิร์ดและ Tag จากผู้เชี่ยวชาญประจำเดือนสิงหาคม 2565	Keyword_bigdata_0865.xlsx	27952
4.	ข้อความ คีย์เวิร์ดและ Tag จากผู้เชี่ยวชาญประจำเดือนกันยายน 2565	Keyword_bigdata_0965.xlsx	27315

ตารางที่ 5 ข้อมูลที่ใช้ในการฝึกสอน

โดยรูปแบบของข้อมูลมีลักษณะดังนี้

	A	B	C	D	E	F
1	ลำดับ	ข้อความ	keyword1	tag1	keyword2	tag2
2	1	การชุดแต่งและบุรณะวัดมหาธาตุ จ. อุทธรยา (พ.ศ.2512)	วัดมหาธาตุ	โบราณสถาน	จ. อุทธรยา	สถานที่
3	2	พบที่วัดพระประโทณเจดีย์วัดนครปฐม	วัดพระประโทณ	โบราณสถาน	จังหวัดนครปฐม	สถานที่
4	3	พระพุทธรูปปางมารวิชัยประทับภายในกรอบสามเหลี่ยมแถบขอกประดับลวดลายไข่มุก	พระพุทธรูป	ปฎิมากรรม		
5	4	พระพิมพ์ มีรูปพระพุทธรูปปางนาคปรกประทับยืนบนชศนาค ด้านหลังเขียน เนื้อโลหะสีดำ	พระพิมพ์	มีรูปพระพุทธรูปปางนาคปรก	พระพุทธรูป	เนื้อโลหะสีดำ
6	5	งานบูรณะโบราณสถานปราสาทหินพนมรุ้ง อำเภอหนอง จังหวัดบุรีรัมย์	โบราณสถานปราสาทหินพนมรุ้ง	โบราณสถาน	อำเภอหนอง จังหวัดบุรีรัมย์	สถานที่
7						
8						

รูปที่ 10 รูปแบบของข้อมูล

เมื่อได้ข้อความจะทำการแปลงข้อมูลเพื่อให้เหมาะสมในการฝึกสอนโดยในชุดข้อมูลแต่ละชุดจะมีการ
ปรับปรุงแบบของข้อมูลดังนี้

	B	C	D	E	F	G
1	ข้อความ	ข้อความ2	keyword1	tag1		keyword2
2	นางมณีเมขลาถือลูกแก้ว		นางมณีเมขลาถือลูกแก้ว	วรรณกรรม	[วรรณกรรม]นางมณีเมขลาถือลูกแก้ว[/วรรณกรรม]	นางมณีเมขลา
3	นางมณีเมขลาฆ่าสีก		นางมณีเมขลา	ตัวละครหรือสัตว์ในวรรณกรรม	[ตัวละครหรือสัตว์ในวรรณกรรม]นางมณีเมขลา/[ตัว	
4	นางมณีเมขลา		นางมณีเมขลา	ตัวละครหรือสัตว์ในวรรณกรรม	ละครหรือสัตว์ในวรรณกรรม]นางมณีเมขลา/[ตัว	
5	นางมณีวรรณ		นางมณีวรรณ	บุคคล	[บุคคล]นางมณีวรรณ[/บุคคล]	
6	นางมณีวรรณกมลพัฒนะ		นางมณีวรรณกมลพัฒนะ	บุคคล	[บุคคล]นางมณีวรรณกมลพัฒนะ[/บุคคล]	
7	นางมน		นางมน	บุคคล	[บุคคล]นางมน[/บุคคล]	
8	นางมณฑรา นิลจันทร์พินัยศรี		นางมณฑรา นิลจันทร์พินัยศรี	บุคคล	[บุคคล]นางมณฑรา นิลจันทร์พินัยศรี[/บุคคล]	
9	นางนันทน์ ปุษาปาคม		นางนันทน์ ปุษาปาคม	บุคคล	[บุคคล]นางนันทน์ ปุษาปาคม[/บุคคล]	
10	นางนัสนี สัตติเอร์ (ประจักษ์พันธ์)		นางนัสนี สัตติเอร์ (ประจักษ์พันธ์)	บุคคล	[บุคคล]นางนัสนี สัตติเอร์ (ประจักษ์พันธ์)/[บุคคล]	
11	นางนัสนีสินธุ์สินธุ์		นางนัสนีสินธุ์ สินธุ์สินธุ์	บุคคล	[บุคคล]นางนัสนีสินธุ์ สินธุ์สินธุ์[/บุคคล]	
12	นางนโม จันทน์อินทร์		นางนโม จันทน์อินทร์	บุคคล	[บุคคล]นางนโม จันทน์อินทร์[/บุคคล]	
13	นางนฤ จันทน์อินทร์		นางนฤ จันทน์อินทร์	บุคคล	[บุคคล]นางนฤ จันทน์อินทร์[/บุคคล]	
14	นางนฤ ปิ่นแก้ว		นางนฤ ปิ่นแก้ว	บุคคล	[บุคคล]นางนฤ ปิ่นแก้ว[/บุคคล]	
15	นางนลชา		นางนลชา	บุคคล	[บุคคล]นางนลชา[/บุคคล]	
16	นางนลชา ภูมิเพ็ง		นางนลชา ภูมิเพ็ง	บุคคล	[บุคคล]นางนลชา ภูมิเพ็ง[/บุคคล]	
17	นางนลวรรณ โมศรีสวัสดิ์		นางนลวรรณ โมศรีสวัสดิ์	บุคคล	[บุคคล]นางนลวรรณ โมศรีสวัสดิ์[/บุคคล]	
18	นางนลวรรณ รองพิจิตรสรพการ		นางนลวรรณ รองพิจิตรสรพการ	บุคคล	[บุคคล]นางนลวรรณ รองพิจิตรสรพการ[/บุคคล]	
19	นางนวง		นางนวง	บุคคล	[บุคคล]นางนวง[/บุคคล]	
20	นางนวัน		นางนวัน	บุคคล	[บุคคล]นางนวัน[/บุคคล]	
21	นางนวัน มัยยมจันทร์		นางนวัน มัยยมจันทร์	บุคคล	[บุคคล]นางนวัน มัยยมจันทร์[/บุคคล]	
22	นางนวัน วิเศษศักดิ์		นางนวัน วิเศษศักดิ์	บุคคล	[บุคคล]นางนวัน วิเศษศักดิ์[/บุคคล]	
23	นางนวย		นางนวย	บุคคล	[บุคคล]นางนวย[/บุคคล]	
24	นางนวย แซ่สี		นางนวย แซ่สี	บุคคล	[บุคคล]นางนวย แซ่สี[/บุคคล]	
25	นางนวย พงศศิริ		นางนวย พงศศิริ	บุคคล	[บุคคล]นางนวย พงศศิริ[/บุคคล]	
	นางนวย ยืนสกุล มีซากศาสนสถาน สมัยอยุธยา		นางนวย ยืนสกุล	บุคคล	[บุคคล]นางนวย ยืนสกุล[/บุคคล]	ซากศาสนสถาน

รูปที่ 11 การแปลงข้อมูลเพื่อให้เหมาะสมในการฝึกสอน

2.2.3. การแปลง Tag สำหรับการฝึกสอน

ในการแปลง tag เพื่อให้สามารถฝึกสอนในโมเดลได้จะทำการแปลงดังนี้

ทำการแปลงข้อความให้อยู่ในรูป CoNLL2002 โดยเพิ่ม tag [word] เข้าไปใส่ทุกคำที่ไม่อยู่ภายใต้ tag ที่เป็นประเภท
นิพจน์ แล้วนำเอาทุก tag ไปตัดคำที่ละ tag ภายใน แล้วรวมเป็น list เดียวกันที่พร้อมแปลงเป็น CoNLL2002 ภายหลัง

ตารางที่ 6 การแปลงข้อความให้อยู่ในรูป CoNLL2002

ลำดับ	ชื่อ Tag ภาษาไทย	ชื่อ Tag ภาษาอังกฤษ	แปลง tag ในรูป CoNLL2002	แปลง tag ในรูป CoNLL2002	แปลง tag ในรูป CoNLL2002	แปลง tag ในรูป CoNLL2002
1	สถานที่	place	'[สถานที่]'	'[/สถานที่]'	'[place]'	'[/place]'
2	วันเวลา	datetime	'[วันเวลา]'	'[/วันเวลา]'	'[datetime]'	'[/datetime]'
3	ชื่อโครงการ	project	'[ชื่อโครงการ]'	'[/ชื่อโครงการ]'	'[project]'	'[/project]'
4	กิจกรรม	activity	'[กิจกรรม]'	'[/กิจกรรม]'	'[activity]'	'[/activity]'
5	สัตว์	animal	'[สัตว์]'	'[/สัตว์]'	'[animal]'	'[/animal]'
6	พืช	plant	'[พืช]'	'[/พืช]'	'[plant]'	'[/plant]'
7	กลุ่มบุคคล/บุคคล/	organization	'[กลุ่มบุคคล/องค์กร/บริษัท]'	'[/กลุ่มบุคคล/องค์กร/บริษัท]'	'[organization]'	'[/organization]'

	องค์กร/ บริษัท					
8	บุคคล	personal	'[บุคคล]'	'[/บุคคล]'	'[personal]'	'[/personal]'
9	ประติมากร	sculptor	'[ประติมากร]'	'[/ประติมากร]'	'[sculptor]'	'[/sculptor]'
10	ภัณฑารักษ์	storekeeper	'[ภัณฑารักษ์]'	'[/ภัณฑารักษ์]'	'[storekeeper]'	'[/storekeeper]'
11	ศิลปวัตถุ	antique	'[ศิลปวัตถุ]'	'[/ศิลปวัตถุ]'	'[antique]'	'[/antique]'
12	ปฏิมากรรม	sculpture	'[ปฏิมากรรม]'	'[/ปฏิมากรรม]'	'[sculpture]'	'[/sculpture]'
13	ประติมากรรม	objsculture	'[ประติมากรรม]'	'[/ประติมากรรม]'	'[objsculture]'	'[/objsculture]'
14	จิตรกรรม	painting	'[จิตรกรรม]'	'[/จิตรกรรม]'	'[painting]'	'[/painting]'
15	นาฏศิลป์-ดนตรี	Dance-Music	'[นาฏศิลป์-ดนตรี]'	'[/นาฏศิลป์-ดนตรี]'	'[Dance-Music]'	'[/Dance-Music]'
16	โบราณสถาน	ancient	'[โบราณสถาน]'	'[/โบราณสถาน]'	'[ancient]'	'[/ancient]'
17	พิพิธภัณฑ	museum	'[พิพิธภัณฑ]'	'[/พิพิธภัณฑ]'	'[museum]'	'[/museum]'
18	อนุสาวรีย์	monument	'[อนุสาวรีย์]'	'[/อนุสาวรีย์]'	'[monument]'	'[/monument]'
19	อุทยานประวัติศาสตร์	historicalpark	'[อุทยานประวัติศาสตร์]'	'[/อุทยานประวัติศาสตร์]'	'[historicalpark]'	'[/historicalpark]'
20	โบราณวัตถุ	antiquities	'[โบราณวัตถุ]'	'[/โบราณวัตถุ]'	'[antiquities]'	'[/antiquities]'
21	รายละเอียดวัตถุ	objectdetail	'[รายละเอียดวัตถุ]'	'[/รายละเอียดวัตถุ]'	'[objectdetail]'	'[/objectdetail]'
22	พระพุทธรูป	buddhastatue	'[พระพุทธรูป]'	'[/พระพุทธรูป]'	'[buddhastatue]'	'[/buddhastatue]'
23	วรรณกรรม	literature	'[วรรณกรรม]'	'[/วรรณกรรม]'	'[literature]'	'[/literature]'
24	ตัวละครหรือสัตว์ในวรรณกรรม	novelanimal	'[ตัวละครหรือสัตว์ในวรรณกรรม]'	'[/ตัวละครหรือสัตว์ในวรรณกรรม]'	'[novelanimal]'	'[/novelanimal]'

	วรรณกรรม					
25	เอกสารหนังสือ	bookdocument	'[เอกสารหนังสือ]'	'[/เอกสารหนังสือ]'	'[bookdocument]'	'[/bookdocument]'
26	จารึก	inscription	'[จารึก]'	'[/จารึก]'	'[inscription]'	'[/inscription]'
27	สมุดไทย	Thaibook	'[สมุดไทย]'	'[/สมุดไทย]'	'[Thaibook]'	'[/Thaibook]'
28	คัมภีร์ใบลาน	palmleafscripture	'[คัมภีร์ใบลาน]'	'[/คัมภีร์ใบลาน]'	'[palmleafscripture]'	'[/palmleafscripture]'
29	โบราณคดี	archeology	'[โบราณคดี]'	'[/โบราณคดี]'	'[archeology]'	'[/archeology]'
30	ประวัติศาสตร์	history	'[ประวัติศาสตร์]'	'[/ประวัติศาสตร์]'	'[history]'	'[/history]'
31	จดหมายเหตุ	Archives	'[จดหมายเหตุ]'	'[/จดหมายเหตุ]'	'[Archives]'	'[/Archives]'
32	ฟิล์มกระจก	glassfilm	'[ฟิล์มกระจก]'	'[/ฟิล์มกระจก]'	'[glassfilm]'	'[/glassfilm]'
33	เอกสารลายลักษณ์	writtendocuments	'[เอกสารลายลักษณ์]'	'[/เอกสารลายลักษณ์]'	'[writtendocuments]'	'[/writtendocuments]'
34	บันทึกเหตุการณ์	eventlog	'[บันทึกเหตุการณ์]'	'[/บันทึกเหตุการณ์]'	'[eventlog]'	'[/eventlog]'
35	ยุคสมัย	era	'[ยุคสมัย]'	'[/ยุคสมัย]'	'[era]'	'[/era]'
36	ก่อนประวัติศาสตร์	prehistoric	'[ก่อนประวัติศาสตร์]'	'[/ก่อนประวัติศาสตร์]'	'[prehistoric]'	'[/prehistoric]'
37	ทวารวดี	Dvaravati	'[ทวารวดี]'	'[/ทวารวดี]'	'[Dvaravati]'	'[/Dvaravati]'
38	สุโขทัย	Sukhothai	'[สุโขทัย]'	'[/สุโขทัย]'	'[Sukhothai]'	'[/Sukhothai]'
39	อยุธยา	Ayutthaya	'[อยุธยา]'	'[/อยุธยา]'	'[Ayutthaya]'	'[/Ayutthaya]'
40	รัตนโกสินทร์	Rattanakosin	'[รัตนโกสินทร์]'	'[/รัตนโกสินทร์]'	'[Rattanakosin]'	'[/Rattanakosin]'
41	ธนบุรี	Thonburi	'[ธนบุรี]'	'[/ธนบุรี]'	'[Thonburi]'	'[/Thonburi]'
42	ลพบุรี	LopBuri	'[ลพบุรี]'	'[/ลพบุรี]'	'[LopBuri]'	'[/LopBuri]'

43	เขมรใน ประเทศ ไทย	KhmerinThailand	'[เขมรในประเทศไทย]'	'[/เขมรในประเทศไทย]'	'[KhmerinThailand]'	'[/KhmerinThailand]'
44	ศรีวิชัย	Sriwichai	'[ศรีวิชัย]'	'[/ศรีวิชัย]'	'[Sriwichai]'	'[/Sriwichai]'
45	ล้านนา	Lanna	'[ล้านนา]'	'[/ล้านนา]'	'[Lanna]'	'[/Lanna]'
46	ล้านช้าง	LanXang	'[ล้านช้าง]'	'[/ล้านช้าง]'	'[LanXang]'	'[/LanXang]'
47	ช่วยเหลือ	help	'[ช่วยเหลือ]'	'[/ช่วยเหลือ]'	'[help]'	'[/help]'
48	โรค	disease	'[โรค]'	'[/โรค]'	'[disease]'	'[/disease]'
49	กฎหมาย	law	'[กฎหมาย]'	'[/กฎหมาย]'	'[law]'	'[/law]'
50	พระราชพิธี	ceremonial	'[พระราชพิธี]'	'[/พระราชพิธี]'	'[ceremonial]'	'[/ceremonial]'
51	ประเพณี	tradition	'[ประเพณี]'	'[/ประเพณี]'	'[tradition]'	'[/tradition]'
52	พิธีกรรม	rite	'[พิธีกรรม]'	'[/พิธีกรรม]'	'[rite]'	'[/rite]'
53	ยานพาหนะ	vehicle	'[ยานพาหนะ]'	'[/ยานพาหนะ]'	'[vehicle]'	'[/vehicle]'
54	วัตถุ	object	'[วัตถุ]'	'[/วัตถุ]'	'[object]'	'[/object]'
55	ยารักษา	medicine	'[ยารักษา]'	'[/ยารักษา]'	'[medicine]'	'[/medicine]'
56	แผนที่, พิกัด	map	'[แผนที่, พิกัด]'	'[/แผนที่, พิกัด]'	'[map]'	'[/map]'
57	วัน	day	'[วัน]'	'[/วัน]'	'[day]'	'[/day]'
58	เดือน	month	'[เดือน]'	'[/เดือน]'	'[month]'	'[/month]'
59	ปี	year	'[ปี]'	'[/ปี]'	'[year]'	'[/year]'
60	เครื่องแต่งกาย	costume	'[เครื่องแต่งกาย]'	'[/เครื่องแต่งกาย]'	'[costume]'	'[/costume]'
61	เลขไปรษณีย์	postalnumber	'[เลขไปรษณีย์]'	'[/เลขไปรษณีย์]'	'[postalnumber]'	'[/postalnumber]'
62	การอบรม/สัมมนา	Training	'[การอบรม/สัมมนา]'	'[/การอบรม/สัมมนา]'	'[Training]'	'[/Training]'
63	สกุลเงิน/เงินตรา	Currency	'[สกุลเงิน/เงินตรา]'	'[/สกุลเงิน/เงินตรา]'	'[Currency]'	'[/Currency]'
64	วารสาร/นิตยสาร	Journal	'[วารสาร/นิตยสาร]'	'[/วารสาร/นิตยสาร]'	'[Journal]'	'[/Journal]'

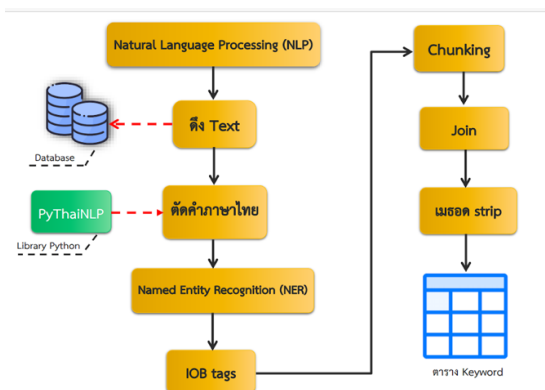
65	หนังสือ	book	'[หนังสือ]'	'[/หนังสือ]'	'[book]'	'[/book]'
66	ชนเผ่าชาติพันธุ์	ethnictribe	'[ชนเผ่าชาติพันธุ์]'	'[/ชนเผ่าชาติพันธุ์]'	'[ethnictribe]'	'[/ethnictribe]'
67	ดวงตรา/สัญลักษณ์	badge	'[ดวงตรา/สัญลักษณ์]'	'[/ดวงตรา/สัญลักษณ์]'	'[badge]'	'[/badge]'
68	ราชวงศ์	dynasty	'[ราชวงศ์]'	'[/ราชวงศ์]'	'[dynasty]'	'[/dynasty]'
69	พระราชทาน	bestowed	'[พระราชทาน]'	'[/พระราชทาน]'	'[bestowed]'	'[/bestowed]'
70	พระเครื่อง	amulet	'[พระเครื่อง]'	'[/พระเครื่อง]'	'[amulet]'	'[/amulet]'
71	อวัยวะ	organ	'[อวัยวะ]'	'[/อวัยวะ]'	'[organ]'	'[/organ]'
72	ข้อเสนอแนะ/ความคิดเห็น	Suggestions	'[ข้อเสนอแนะ/ความคิดเห็น]'	'[/ข้อเสนอแนะ/ความคิดเห็น]'	'[Suggestions]'	'[/Suggestions]'
73	วิชา/หลักสูตร	subject	'[วิชา/หลักสูตร]'	'[/วิชา/หลักสูตร]'	'[subject]'	'[/subject]'
74	อาหาร	food	'[อาหาร]'	'[/อาหาร]'	'[food]'	'[/food]'
75	แบบแปลน/แผนผัง	Schematic	'[แบบแปลน/แผนผัง]'	'[/แบบแปลน/แผนผัง]'	'[Schematic]'	'[/Schematic]'
76	มาตราส่วน	scale	'[มาตราส่วน]'	'[/มาตราส่วน]'	'[scale]'	'[/scale]'
77	ภัยพิบัติ	disaster	'[ภัยพิบัติ]'	'[/ภัยพิบัติ]'	'[disaster]'	'[/disaster]'
78	กีฬา	sport	'[กีฬา]'	'[/กีฬา]'	'[sport]'	'[/sport]'
79	ภาพยนตร์	movies	'[ภาพยนตร์]'	'[/ภาพยนตร์]'	'[movies]'	'[/movies]'
80	เครื่องราช	regalia	'[เครื่องราช]'	'[/เครื่องราช]'	'[regalia]'	'[/regalia]'
81	รางวัล	reward	'[รางวัล]'	'[/รางวัล]'	'[reward]'	'[/reward]'
82	เทพ/เทพเจ้า/พระเจ้า	God	'[เทพ/เทพเจ้า/พระเจ้า]'	'[/เทพ/เทพเจ้า/พระเจ้า]'	'[God]'	'[/God]'
83	ซื้อ/ขาย	Buy	'[ซื้อ/ขาย]'	'[/ซื้อ/ขาย]'	'[Buy]'	'[/Buy]'
84	สถาปัตยกรรม	architecture	'[สถาปัตยกรรม]'	'[/สถาปัตยกรรม]'	'[architecture]'	'[/architecture]'

85	ข่าวสาร	message	'[ข่าวสาร]'	'[/ข่าวสาร]'	'[message]'	'[/message]'
----	---------	---------	-------------	--------------	-------------	--------------

2.2.4. ขั้นตอนและการดำเนินการพัฒนาโมเดลสืบค้นข้อมูลด้วยคำสำคัญ

กระบวนการตัดคำสำหรับพัฒนาโมเดลสืบค้นข้อมูลด้วยคำสำคัญมีขั้นตอนดังนี้

1. เริ่มจากการดึงข้อความที่ต้องการตัดคำจากฐานข้อมูล
2. นำข้อความผ่านกระบวนการตัดคำ โดยใช้ Library PyThaiNLP
3. ทำการสกัดนิพจน์ หรือ ชื่อเฉพาะในข้อความ เช่น ชื่อคน ชื่อสถานที่ โดยใช้ Named-Entity Recognition
4. นำข้อความที่ได้จากขั้นตอนที่ 3 มาใช้เทคนิค Inside-outside-beginning (tagging) เพื่อระบุว่าข้อความนั้นเป็น beginning และเป็น Inside
5. ตัดแท็กที่ไม่ต้องการออก หรือเลือกแท็กที่ต้องการใช้งานเก็บลงในลิสต์
6. นำข้อความที่อยู่ในลิสต์เดียวกัน มารวมกันเป็นคำเดียว
7. ทำการลบ whitespace จากข้างหน้าและท้ายข้อความออก โดยใช้เมธอด strip()
8. นำคำสำคัญที่ตัดได้ไปใส่ลงตาราง keyword



ภาพที่ 5 ตัวอย่างกระบวนการตัดคำ

การคัดกรองคีย์เวิร์ดจะทำการคัดเลือกคำที่มีความหมายหรือคำเฉพาะ เช่น ชื่อบุคคล ชื่อ สถานที่ วัน เดือน ปี เป็นต้น และนำคีย์เวิร์ดดังกล่าวมาเชื่อมตารางฐานข้อมูล

ตัวอย่างที่ 1 กระบวนการตัดคำ

1. ทำการดึง text

2. คำสั่ง, Named Entity Recognition (NER), IOB tags ใน `ThaiNameTagger()`

3. chunking

4. join

5. Methods strip()

6. ตาราง keyword

ตัวอย่างที่ 2 กระบวนการตัดคำ

1. ทำการดึง text

น.ส.พ.ฝรั่งขอถ่ายภาพ ที่บ้านเกศโกศล (12 ก.พ.2509)

2. ตัดคำ, Named Entity Recognition (NER), IOB tags ใน ThaiNameTagger()

[('น.ส.พ.', 'O'), ('ฝรั่ง', 'O'), ('ขอ', 'O'), ('ถ่ายภาพ', 'O'), (' ', 'O'), ('ที่', 'O'), ('บ้าน', 'B-LOCATION'), ('เกศ', 'I-LOCATION'), ('โกศล', 'I-LOCATION'), (' ', 'O'), ('(', 'O'), ('12', 'B-DATE'), (' ', 'I-DATE'),

3. chunking

[('น.ส.พ.', 'O'), ('ฝรั่ง', 'O'), ('ขอ', 'O'), ('ถ่ายภาพ', 'O'), (' ', 'O'), ('ที่', 'O'), ('บ้าน', 'B-LOCATION'), ('เกศ', 'I-LOCATION'), ('โกศล', 'I-LOCATION'), (' ', 'O'), ('(', 'O'), ('12', 'B-DATE'), (' ', 'I-DATE'), ('ก.พ.', 'I-DATE'), ('2509', 'I-DATE'), (')', 'O'), ('\n', 'O')]

4. join

['บ้านเกศโกศล', '12 ก.พ.2509']

5. Methods strip()

['บ้านเกศโกศล', '12 ก.พ.2509']

6. ตาราง keyword

Keyword_ID	Description	Keyword_ID	Description
55946	บ้านเกศโกศล	578	12 ก.พ. 2509

2.2.5. การปรับปรุงโมเดลในการสร้างแท็กสำหรับคีย์เวิร์ดจากประโยคในฐานข้อมูล

ในกระบวนการปรับปรุงโมเดลและพัฒนาโมเดลสืบค้นข้อมูลด้วยคำสำคัญจากฐานข้อมูลจะนำข้อมูลมาจากฐานข้อมูลดังต่อไปนี้

1. ฐานข้อมูลด้านโบราณสถาน
2. ฐานข้อมูลด้านโบราณวัตถุ
3. ฐานข้อมูลหอจดหมายเหตุแห่งชาติ
4. ฐานข้อมูล NSW
5. ฐานข้อมูลหอสมุดแห่งชาติ
6. ฐานข้อมูลพระเมรุมาศและเครื่องประกอบพระเมรุมาศ
7. ฐานข้อมูลระบบศูนย์ข้อมูลงานศิลปกรรม สำนักช่างสิบหมู่

การสร้างโมเดลสำหรับสกัดคำสำคัญจากเอกสารการสกัดคีย์เวิร์ดจากฐานข้อมูล

เมื่อได้ข้อมูลจากฐานข้อมูลจะนำข้อมูลดังกล่าวมาแปลงให้ข้อมูลเหมาะสมในการฝึกสอนโมเดล ซึ่งใน 1 คีย์เวิร์ดจะประกอบไปด้วยแท็กมากกว่า 1 แท็ก โดยจะอยู่ในรูปแบบดังต่อไปนี้

ถ้วยเขียนสี [กิจกรรม]ลายเทพพนมรสิงห์[กิจกรรม] พร้อมฝา [รายละเอียดวัตถุ]ปากปิ่น[รายละเอียดวัตถุ]
 [กิจกรรม]การประชุมสัมมนาทางโบราณคดี[กิจกรรม] ประจำปี[วันเวลา]ปี 2503[วันเวลา]
 อำนาจความสะดวกให้แก่ [บุคคล]MONSIEUR LEMONNIER[บุคคล] และ [บุคคล]MONSIEUR HEURTEMATTE/[บุคคล] นายทหารฝรั่งเศสชั้นผู้ใหญ่ 2 นาย
 [บุคคล]สมเด็จพระเจ้าลูก เจ้าฟ้าจุฬารามวลัยลักษณ์ อัครราชกุมารี[บุคคล] เสด็จพร้อมด้วย[บุคคล]พระเจ้าหลานเธอ พระเจ้าองค์เจ้าสิริภาจุฑาภรณ์[บุคคล] และ[บุคคล]พระเจ้าหลานเธอพระองค์เจ้าอทิตยาทรกิติคุณ[บุคคล] เพื่อ ทรงตก บาตร ณ [สถานที่]ค่ายจุฬารามณ์ กรมนาวิกโยธิน [สถานที่][สถานที่]บ้านทอน[สถานที่] [สถานที่]อำเภอเมือง/[สถานที่] [สถานที่]จังหวัดนครราชสีมา[สถานที่]
 [ศิลปวัตถุ]ภาชนะ[ศิลปวัตถุ][รายละเอียดวัตถุ]ทรงกลม[รายละเอียดวัตถุ] [รายละเอียดวัตถุ]ก้นโค้งมน[รายละเอียดวัตถุ] [รายละเอียดวัตถุ]ตัวภาชนะป่องกลม[รายละเอียดวัตถุ] [รายละเอียดวัตถุ]ปากผายเล็กน้อย[รายละเอียดวัตถุ] [รายละเอียดวัตถุ]เนื้อดินสีน้ำตาล[รายละเอียดวัตถุ]
 [ศิลปวัตถุ]ปางมารวิชัย ขัดสมาธิราบ ประทับบนฐานเขียง ครองจักรท่มเฉียง เปิดพระอังสาขวา[ศิลปวัตถุ]
 [ประติมากรรม]พระพิฆเนศยืน[ประติมากรรม]
 [บุคคล]นายถวิล ณ ตะกั่วทุ่ง[บุคคล] มอบให้เมื่อ[วันเวลา]วันที่ 29 ตุลาคมพ.ศ. 2517[วันเวลา]
 [ปฏิมากรรม]พระพุทธรูปปางสมาธิ [/ปฏิมากรรม] เป็น[พระพุทธรูป]พระพุทธรูปด้วยเงิน [/พระพุทธรูป] โดยมี [รายละเอียดวัตถุ]แกนในเป็นไม้ พระพักตร์กลม พระกรรณเป็นแผ่น พระรัศมีทรงกรวยแหลมสูง ครองจักรเปิดพระอังสา ขวา ประทับนั่งขัดสมาธิราบบนฐานหน้ากระดาน ประดับลายกลีบบัว[รายละเอียดวัตถุ]
 [พระพุทธรูป]พระพุทธรูปปางสมาธิอย่างแผงมี 21 องค์[พระพุทธรูป]

ท่อน[พระพุทธรูป]องค์พระพุทธรูป[พระพุทธรูป] [รายละเอียดวัตถุ]แกะสลักชายสังฆาฏิเป็นริ้ว[รายละเอียดวัตถุ]
ได้จาก[โบราณสถาน]กรุพระปรางค์วัดราชบูรณะ[โบราณสถาน] เมื่อ [วันเวลา]พ.ศ.2500[วันเวลา]
[รายละเอียดวัตถุ]พระพุทธรูปปางสมาธิอย่างแผงมี 21 องค์[รายละเอียดวัตถุ]
[รายละเอียดวัตถุ]ฐานโปร่ง ปากหยัก[รายละเอียดวัตถุ]
แหล่งโบราณคดีออบหลวง [สถานที่]อำเภอฮอด จังหวัดเชียงใหม่[สถานที่]
เขียนำนาญ[บุคคล]ร.ต.ฉัตรรุ่งเรืองโพธิ์[บุคคล]
[บุคคล]คุณสรวง เสหาวิช[บุคคล] มอบให้[พิพิธภัณฑ์]พิพิธภัณฑสถานแห่งชาติ นครศรีธรรมราช[พิพิธภัณฑ์] เมื่อ
[วันเวลา]วันที่ 29 ตุลาคม พ.ศ. 2517[วันเวลา]

การสร้างโมเดลในการสกัดคีย์เวิร์ดด้วย Named Entity Recognition

Named Entity Recognition หรือ NER คือ การสกัดนิพจน์เฉพาะหรือชื่อเฉพาะในประโยค
สมมติเรามีประโยค "ภาษาชนทรงกลมกันไค้มนตัวภาษาชนปะองกลมปากผายเล็กน้อยเนื้อดินสีน้ำตาล"

จากประโยคข้างบน เราจะเห็นนิพจน์เฉพาะ คือ

[ศิลปวัตถุ]ภาษาชน[/ศิลปวัตถุ][รายละเอียดวัตถุ]ทรงกลม[/รายละเอียดวัตถุ] [รายละเอียดวัตถุ]กันไค้มน
[/รายละเอียดวัตถุ] [รายละเอียดวัตถุ]ตัวภาษาชนปะองกลม[/รายละเอียดวัตถุ] [รายละเอียดวัตถุ]ปากผายเล็กน้อย/
รายละเอียดวัตถุ [รายละเอียดวัตถุ]เนื้อดินสีน้ำตาล[/รายละเอียดวัตถุ]

ดังนั้น NER ควรดักจับนิพจน์นี้ได้

ขั้นตอนการทำ Named Entity Recognition ภาษาไทย

ในส่วนนี้เราสามารถทำ Named Entity Recognition ใหม่ขึ้นมาได้ โดยใช้ sklearn-crfsuite ซึ่งเป็น
โมดูลสำหรับทำ CRF model โดยเชื่อมกับ CRFsuite อีกที ในการประมวลผลสามารถทำได้โดยนำคลังข้อมูล
CoNLL2002 ไปทำการฝึกสอน(train) กับโมเดล Conditional Random Fields (CRFs) ซึ่งเป็นโมเดลสำหรับ
การสร้าง Named Entity Recognition

รูปแบบ CoNLL2002 ประกอบไปด้วย

1. คำ
2. postag
3. tag

โดยรูปแบบข้อมูลที่ sklearn-crfsuite ต้องการเพื่อนำไปใช้ในการฝึกสอนคือ

[[[คำ,postag,tag),...],...]

โดย 1 List ภายใน List คือ 1 ประโยค

การสร้างคลังข้อมูลสำหรับทำ Named Entity Recognition เพื่อให้รองรับภาษาไทย

ในการพัฒนา Named Entity Recognition ภาษาไทย จะต้องมีการแปลงคลังข้อมูลให้อยู่ในรูปแบบ CoNLL2002 โดยในส่วนของการตัดคำจะใช้การตัดคำภาษาไทยโดยใช้ thaicut="newmm" โดยขั้นตอนในการสอนโมเดลให้รู้จักคำใหม่ๆ มีขั้นตอนในการทำงานดังนี้

1. ส่งข้อความในรูปแบบต่อไปนี้เพื่อนำไปฝึกสอน

[ศิลปวัตถุ]ภาษาชะ/ศิลปวัตถุ[รายละเอียดวัตถุ]ทรงกลม[รายละเอียดวัตถุ]	[รายละเอียดวัตถุ]กันโค้งมน/[
รายละเอียดวัตถุ]	[รายละเอียดวัตถุ]ตัวภาษาชะป่องกลม[รายละเอียดวัตถุ]
[รายละเอียดวัตถุ]ปากผายเล็กน้อย[รายละเอียดวัตถุ]	[รายละเอียดวัตถุ]เนื้อดินสีน้ำตาล[รายละเอียดวัตถุ]

2. ทำการแปลงข้อความให้อยู่ในรูปแบบ CoNLL2002 โดยเพิ่ม tag [word] เข้าไปใส่ทุกคำที่ไม่อยู่ภายใต้ tag ที่เป็นประเพณีพจน์ แล้วนำเอาทุก tag ไปตัดคำทีละ tag ภายใน แล้วรวมเป็น list เดียวกันที่พร้อมแปลงเป็น CoNLL2002 ภายหลัง

3. ทำการออกแบบคลังข้อความว่าควรมี tag ประเพณีพจน์อะไรบ้าง โดยในส่วนนี้จะมี tag ที่จะทำการฝึกสอนประกอบด้วย

['[กลุ่มบุคคล/องค์กร/บริษัท]',
'[ศิลปวัตถุ]',
'[โบราณสถาน]',
'[วันเวลา]',
'[ปฏิมากรรม]',
'[รายละเอียดวัตถุ]',
'[บุคคล]',
'[อุทยานประวัติศาสตร์]',
'[สถานที่]',
'[เอกสารหนังสือ]',

'กิจกรรม',
 'ฟิล์มกระจก',
 'จิตรกรรม',
 'คัมภีร์ไบเบิล',
 'ชื่อโครงการ',
 'สุโขทัย',
 'อนุสาวรีย์',
 'พิพิธภัณฑ์',
 'โบราณวัตถุ',
 'พระพุทธรูป',
 'รัตนโกสินทร์',
 'อยุธยา',
 'เอกสารลายลักษณ์',
 'โบราณคดี',
 'ประติมากรรม']

4. นำข้อความที่ได้ผ่านตัวตัดคำและให้ตัวที่อยู่ใน tag อื่นที่ไม่ใช่ [word] แปลงให้เป็น B-tag และ I-tag ส่วน [word] ให้เป็น O แล้วบรรจุลง List

5. นำข้อความที่ได้ไปผ่าน postag โดยจะได้รายละเอียดดังนี้

```

[(['ถ้วย', 'NN', 'O'),
 ('เขียน', 'VV', 'O'),
 ('สี', 'NN', 'O'),
 (' ', 'PU', 'O'),
 ('ลาย', 'NN', 'B-painting'),
 ('เทพ', 'NN', 'I-painting'),
 ('พนม', 'NN', 'I-painting'),
 ('นรสิงห์', 'NN', 'I-painting'),
 (' ', 'PU', 'O'),
 ('พร้อม', 'W', 'O'),
 ('ฝา', 'NN', 'O'),

```

(' ', 'PU', 'O'),
 ('ปาก', 'NN', 'B-objectdetail'),
 ('ปิ่น', 'NN', 'I-objectdetail'],
 [('การประชุม', 'NN', 'B-activity'),
 ('สัมมนา', 'NN', 'I-activity'),
 ('ทาง', 'NN', 'I-activity'),
 ('โบราณคดี', 'NN', 'I-activity'),
 (' ', 'PU', 'O'),
 ('ประจำ', 'AJ', 'O'),
 ('ปี', 'NN', 'B-datetime'),
 (' ', 'PU', 'I-datetime'),
 ('2503', 'NU', 'I-datetime']],
 [('อำนวยความสะดวก', 'VV', 'O'),
 ('ให้', 'AV', 'O'),
 ('แก่', 'PS', 'O'),
 (' ', 'PU', 'O'),
 ('MONSIEUR', 'NN', 'B-personal'),
 (' ', 'PU', 'I-personal'),
 ('LEMONNIER', 'NN', 'I-personal'),
 (' ', 'PU', 'O'),
 ('และ', 'CC', 'O'),
 (' ', 'PU', 'O'),
 ('MONSIEUR', 'NN', 'B-personal'),
 (' ', 'PU', 'I-personal'),
 ('HEURTEMATTE', 'NN', 'I-personal'),
 (' ', 'PU', 'O'),
 ('นายทหาร', 'NN', 'O'),
 ('ฝรั่งเศส', 'NN', 'O'),
 ('ชั้นผู้ใหญ่', 'NN', 'O'),
 (' ', 'PU', 'O'),
 ('2', 'NU', 'O'),
 (' ', 'PU', 'O'),

```
('นาย', 'NN', 'O')]]
```

6.จะได้ข้อมูลที่พร้อมเอาไปฝึกสอน กับ sklearn-crfsuite โดยแปลงคลังให้อยู่ในรูปแบบที่ sklearn-crfsuite รองรับ โดยการติดตั้ง sklearn-crfsuite กับ pythainlp โดยชุดคำสั่งของการฝึกสอนประกอบไปด้วยชุดคำสั่งดังนี้

```
trainer = pycrfsuite.Trainer(verbose=False)
i=0
for xseq, yseq in zip(X_data, y_data):
    try:
        trainer.append(xseq, yseq)
        i+=1
    except:
        pass
trainer.set_params({
    'c1': 0.1, # coefficient for L1 penalty
    'c2': 0.1, # coefficient for L2 penalty
    'max_iterations': 500, # stop earlier
    'feature.possible_transitions': True
})
trainer.params()
trainer.train('thai-ner-1-5-newmm-lst20.crfsuite')
```

หรือใช้ชุดคำสั่งดังนี้

```
crf = sklearn_crfsuite.CRF(
    algorithm='lbfgs',
    c1=0.1,
    c2=0.1,
    max_iterations=500,
    all_possible_transitions=True,
    model_filename="thai-ner-1-5-newmm-lst20.crfsuite"
)
```

7. โดยเมื่อทำการฝึกสอนเสร็จ และต้องการทำนาย tag ของข้อความจะใช้ชุดคำสั่งดังนี้

```
labels = list(crf.classes_)
labels.remove('O')
y_pred = crf.predict(X_data)
```

8. โดยผลลัพธ์ที่ได้จากการทำนายประกอบไปด้วย

```
['O',
 'O',
 'O',
 'O',
 'B-painting',
 'I-painting',
 'I-painting',
 'I-painting',
 'O',
 'O',
 'O',
 'O',
 'B-objectdetail',
 'I-objectdetail'],
['B-personal',
 'I-personal',
 'I-personal',
 'I-personal',
 'O',
 'O',
 'B-datetime',
 'I-datetime',
 'I-datetime'],
['O',
 'O',
 'O',
```

```
'O',  
'B-personal',  
'I-personal',  
'I-personal',  
'I-personal',  
'I-personal',  
'I-personal',  
'I-personal',  
'I-personal',  
'I-personal',  
'I-personal',  
'I-personal',  
'I-personal',  
'O',  
'O',  
'O',  
'O']]
```

9. เมื่อต้องการทำนาย tag ของข้อความใหม่ที่เข้ามาจะสามารถส่งข้อความและนำข้อความดังกล่าวมาทำการทำนาย tag ได้ เช่น

ข้อความต้นฉบับ

เมื่อวันที่ 15 พฤศจิกายน 2519 เวลา 13.00 น. สมเด็จพระศรีนครินทราบรมราชชนนี เสด็จฯ พร้อมด้วยสมเด็จพระเจ้าพี่นางเธอ เจ้าฟ้ากัลยาณิวัฒนาโดยเฮลิคอปเตอร์พระที่นั่ง จากอุทยานแห่งชาติภูกระดึง ไปทรงเยี่ยมราษฎรบ้านเมี่ยงตำบลท่าลี่ อำเภอท่าลี่ จังหวัดเลย และพระราชทานธงประจำรุ่นลูกเสือชาวบ้าน จำนวน 11 รุ่น พระราชทานเหรียญที่ระลึกแก่ลูกเสือชาวบ้าน จำนวน 1,130 คน พระราชทานเครื่องแบบนักเรียน อุปกรณ์การเรียนแก่นักเรียนชายหญิง จำนวน 292 คน พระราชทานเครื่องนุ่งห่ม ยาชุดรักษาโรค ประจำบ้านแก่คนชรา และของเล่นแก่เด็ก ในโอกาสนี้โปรดเกล้าฯ ให้หน่วยแพทย์เคลื่อนที่ พอ.สว. ส่วนกลาง ตั้งทำการตรวจรักษาพยาบาลและแจกจ่ายยาแก่ผู้เจ็บป่วยที่มาขอรับการรักษาโดยทั่วกัน

เมื่อผ่านการประมวลผลจะได้คีย์เวิร์ดและ tag ดังนี้

Tag

['[วันเวลา]', '[บุคคล]', '[สถานที่]', '[กลุ่มบุคคล/องค์กร/บริษัท]', '[สถานที่]', '[กลุ่มบุคคล/องค์กร/บริษัท]', '[กลุ่มบุคคล/องค์กร/บริษัท]']

คีย์เวิร์ด

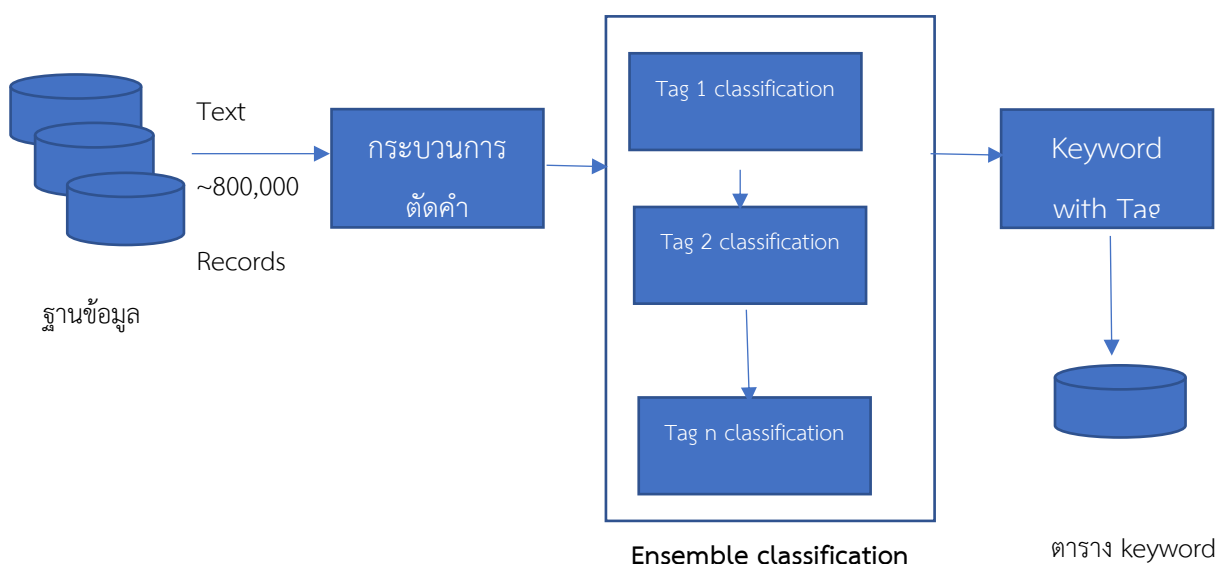
['[วันที่ 15 พฤศจิกายน 2519 เวลา 13.00 น. ', 'datetime')', ('สมเด็จพระเจ้าพี่นางเธอ เจ้าฟ้ากัลยาณิวัฒนา', 'personal'), ('อุทยานแห่งชาติภูกระดึง', 'place'), ('ราษฎรบ้านเมี่ยงตำบลท่าลี่', 'organization'), ('อำเภอท่าลี่ จังหวัดเลย', 'place'), ('นักเรียนชายหญิง', 'organization'), ('หน่วยแพทย์เคลื่อนที่ พอ.สว.', 'organization')]

10. เมื่อได้คีย์เวิร์ดดังกล่าวจะนำไปเก็บในฐานข้อมูลโดย คีย์เวิร์ด 1 คีย์เวิร์ดจะสามารถมี Tag สำหรับคำสำคัญได้มากกว่า 1 Tag

2.2.6. การสร้างโมเดลในการสกัดคีย์เวิร์ดด้วย Named Entity Recognition แบบหลายแท็กด้วยเทคนิค

Ensemble classification

ในการฝึกสอน กับ sklearn-crfsuite โดยแปลงข้อความให้อยู่ในรูปแบบที่ sklearn-crfsuite รองรับ โดยการติดตั้ง sklearn-crfsuite กับ pythainlp จะพบว่าในการฝึกสอนแต่ละครั้งตัวอัลกอริทึมจะสามารถเรียนรู้ Tag ได้เพียง 1 Tag ใน 1 คีย์เวิร์ด ด้วยเหตุนี้เพื่อให้ตัวคีย์เวิร์ดสามารถที่จะเรียนรู้ Tag ได้มากกว่า 1 จึงจำเป็นต้องใช้วิธีการ Ensemble classification ซึ่งเทคนิคนี้จะทำการเรียนรู้ แท็กของคีย์เวิร์ดทีละแท็กและส่งต่อไปยังโมเดลสำหรับการเรียนรู้ในแท็กถัดไป ในการทำวิธีนี้จะเป็นการรับประกันได้ว่าใน 1 คีย์เวิร์ดสามารถมีแท็กได้มากกว่า 1 แท็กแต่จะใช้เวลาในการเรียนรู้นานเพิ่มมากขึ้น โดยกระบวนการในการเรียนรู้มีดังนี้



รูปที่ 12 การสร้างโมเดลในการสกัดคีย์เวิร์ดด้วย Named Entity Recognition แบบหลายแท็กด้วยเทคนิค

Ensemble classification

ตัวอย่างการสร้างโมเดลในการสกัดคีย์เวิร์ดด้วย Named Entity Recognition แบบหลายแท็กด้วยเทคนิค Ensemble classification มีดังนี้

ข้อความต้นฉบับ

เมื่อวันที่ 15 พฤศจิกายน 2519 เวลา 13.00 น. สมเด็จพระศรีนครินทราบรมราชชนนี เสด็จฯ พร้อมด้วยสมเด็จพระเจ้าพี่นางเธอ เจ้าฟ้ากัลยาณิวัฒนาโดยเฮลิคอปเตอร์พระที่นั่ง จากอุทยานแห่งชาติภูกระดึง ไปทรงเยี่ยมราษฎรบ้านเมี่ยงตำบลท่าลี่ อำเภอท่าลี่ จังหวัดเลย และพระราชทานธงประจำรุ่นลูกเสือชาวบ้าน จำนวน 11 รุ่น พระราชทานเหรียญที่ระลึกแก่ลูกเสือชาวบ้าน จำนวน 1,130 คน พระราชทานเครื่องแบบนักเรียน อุปกรณ์การเรียนแก่นักเรียนชายหญิง จำนวน 292 คน พระราชทานเครื่องนุ่งห่ม ยาชุดรักษาโรค ประจำบ้านแก่คนชรา และของเล่นแก่เด็ก ในโอกาสนี้โปรดเกล้าฯ ให้หน่วยแพทย์เคลื่อนที่ พอ.สว. ส่วนกลาง ตั้งทำการตรวจรักษาพยาบาลและแจกจ่ายยาแก่ผู้เจ็บป่วยที่มาขอรับการรักษาโดยทั่วกัน

เมื่อผ่านการประมวลผลจะได้คีย์เวิร์ดและ tag ดังนี้

Tag

['[วันเวลา]', '[บุคคล]', '[สถานที่]', '[กลุ่มบุคคล/องค์กร/บริษัท]', '[สถานที่]', '[กลุ่มบุคคล/องค์กร/บริษัท]', '[กลุ่มบุคคล/องค์กร/บริษัท]']

คีย์เวิร์ด

['[วันที่ 15 พฤศจิกายน 2519 เวลา 13.00 น. ', 'datetime)', ('สมเด็จพระเจ้าพี่นางเธอ เจ้าฟ้ากัลยาณิวัฒนา', 'personal)', ('อุทยานแห่งชาติภูกระดึง', 'place)', ('ราษฎรบ้านเมี่ยงตำบลท่าลี่', 'organization)', ('อำเภอท่าลี่ จังหวัดเลย', 'place)', ('นักเรียนชายหญิง', 'organization)', ('หน่วยแพทย์เคลื่อนที่ พอ.สว.', 'organization)']

ตารางที่ 7 ตัวอย่างตาราง keyword

Keyword_ID	Keyword	Tag
0001	วันที่ 15 พฤศจิกายน 2519 เวลา 13.00 น.	วันเวลา
0002	สมเด็จพระเจ้าพี่นางเธอ เจ้าฟ้ากัลยาณิวัฒนา	บุคคล
0003	อุทยานแห่งชาติภูกระดึง	สถานที่

0004	ราษฎรบ้านเมี่ยงตำบล ท่าลี่	กลุ่มบุคคล/องค์กร/ บริษัท
0005	อำเภอท่าลี่ จังหวัดเลย	สถานที่
0006	นักเรียนชายหญิง	กลุ่มบุคคล/องค์กร/ บริษัท
0007	หน่วยแพทย์เคลื่อนที่ พอ.สว.	กลุ่มบุคคล/องค์กร/ บริษัท

2.2.7. ผลการทดลองจากการฝึกสอนโมเดล

ตารางแสดง Precision recall จากการฝึกสอนโมเดล

	precision	recall	f1-score	support
B-	0.90	0.53	0.67	88
I-	0.97	0.71	0.82	238
B-/dynasty	0.00	0.00	0.00	2
I-/dynasty	0.00	0.00	0.00	4
B-Archives	1.00	1.00	1.00	3
I-Archives	1.00	1.00	1.00	12
B-Ayutthaya	1.00	1.00	1.00	1
I-Ayutthaya	1.00	1.00	1.00	1
B-Currency	0.50	1.00	0.67	1
I-Currency	0.60	1.00	0.75	3
B-LopBuri	0.00	0.00	0.00	1
B-Training	1.00	1.00	1.00	2
I-Training	1.00	1.00	1.00	23
B-activity	0.80	0.80	0.80	10
I-activity	0.97	0.94	0.95	32
B-ancient	0.67	0.50	0.57	4
I-ancient	1.00	1.00	1.00	4
B-annimal	1.00	0.33	0.50	6
I-annimal	1.00	1.00	1.00	1
B-antiquities	1.00	0.89	0.94	19
I-antiquities	1.00	1.00	1.00	45
B-bestowed	1.00	1.00	1.00	1
I-bestowed	1.00	1.00	1.00	1
B-book	1.00	0.86	0.92	7
I-book	1.00	1.00	1.00	23

B-bookdocument	1.00	0.81	0.90	16
I-bookdocument	1.00	1.00	1.00	30
B-ceremonial	1.00	1.00	1.00	3
I-ceremonial	1.00	1.00	1.00	16
B-costume	1.00	0.37	0.54	19
I-costume	1.00	0.97	0.98	29
B-datetime	1.00	1.00	1.00	1
I-datetime	1.00	1.00	1.00	4
B-day	0.67	0.80	0.73	5
B-disaster	0.00	0.00	0.00	1
B-disease	1.00	0.67	0.80	6
I-disease	1.00	1.00	1.00	6
B-dynasty	0.54	0.70	0.61	10
I-dynasty	0.74	0.84	0.79	31
B-ethnictribe	1.00	0.50	0.67	4
I-ethnictribe	1.00	0.75	0.86	4
B-law	1.00	1.00	1.00	2
I-law	1.00	1.00	1.00	3
B-literature	1.00	1.00	1.00	2
I-literature	1.00	1.00	1.00	7
B-month	1.00	0.11	0.20	9
I-month	1.00	1.00	1.00	1
B-museum	0.00	0.00	0.00	1
B-novelandimal	1.00	0.20	0.33	5
I-novelandimal	1.00	1.00	1.00	2
B-object	0.81	0.82	0.82	114
I-object	0.96	0.92	0.94	155
B-objectdetail	0.96	0.97	0.96	2401
I-objectdetail	0.98	0.99	0.99	6978
B-organ	0.67	0.22	0.33	9
I-organ	0.75	0.60	0.67	5
B-organization	0.90	0.71	0.79	127
I-organization	0.96	0.95	0.95	339
B-painting	0.85	0.39	0.53	44
I-painting	0.93	0.51	0.66	80
B-palmleafscripture	1.00	0.86	0.92	28
I-palmleafscripture	1.00	0.99	0.99	82
B-personal	0.96	0.99	0.98	4217
I-personal	0.99	1.00	1.00	13043
B-place	0.93	0.94	0.93	458
I-place	0.97	0.98	0.98	773
B-plant	1.00	0.38	0.56	13
I-plant	1.00	0.60	0.75	10
B-project	1.00	1.00	1.00	4

I-project	1.00	1.00	1.00	15
B-regalia	1.00	1.00	1.00	1
I-regalia	1.00	1.00	1.00	8
B-reward	1.00	1.00	1.00	1
I-reward	1.00	1.00	1.00	6
B-sculpture	0.00	0.00	0.00	1
I-sculpture	0.00	0.00	0.00	2
B-subject	0.82	0.75	0.78	12
I-subject	0.91	0.98	0.94	51
B-writtendocuments	1.00	0.82	0.90	11
I-writtendocuments	1.00	0.95	0.97	41
B-year	1.00	1.00	1.00	6
I-year	1.00	1.00	1.00	1

ตารางที่ 8 ตารางสรุปจำนวนคำที่ได้ในแต่ละ Tag

count(DISTINCT keyword)	count(keyword)	tag
164098	535675	รายละเอียดวัตถุ
24878	239415	โบราณวัตถุ
13899	153863	วันเวลา
32289	123319	บุคคล
11523	114260	สถานที่
12788	51037	โบราณสถาน
9464	37633	คัมภีร์โบราณ
467	35169	พิพิธภัณฑ์
2317	23216	วัตถุ
1357	20187	กลุ่มบุคคล/องค์กร/บริษัท
134	18061	ปฏิมากรรม
14164	17991	เอกสารลายลักษณ์
11381	17746	
669	10311	พระเครื่อง
847	8729	พระพุทธรูป

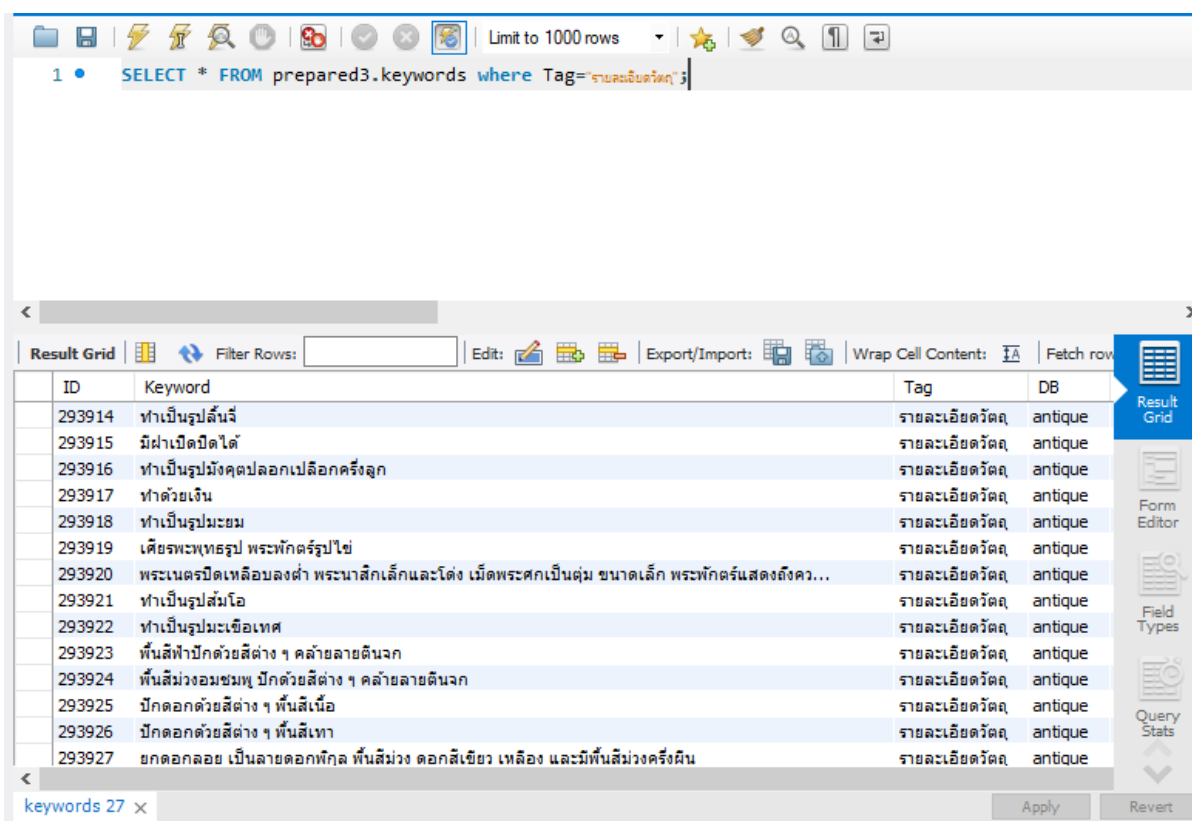
180	7830	โบราณคดี
3156	5320	หนังสือสมุด ไทย
1705	4243	กิจกรรม
2341	4036	เอกสารหนังสือ
971	3740	ปี
339	3564	อวัยวะ
1067	2852	ประติมากรรม
1662	2737	ศิลปวัตถุ
1691	2258	จดหมายเหตุ
1241	1616	บันทึกเหตุการณ์
1	1410	สุโขทัย
995	1351	ชื่อโครงการ
70	1087	สัตว์
547	1003	จิตรกรรม
540	956	เครื่องแต่งกาย
332	849	สกุลเงิน/เงินตรา
343	848	ราชวงศ์
130	632	เดือน
57	486	วัน
2	485	ทวารวดี
1	394	ฟิล์มกระจก
310	355	หนังสือ
260	317	จารึก
143	246	ดวงตรา/สัญลักษณ์
112	229	มาตราส่วน
198	225	พระราชพิธี
28	205	พืช

12	185	ตู้พระธรรม
7	170	ยานพาหนะ
53	165	เทพ/เทพเจ้า/พระเจ้า
150	161	วรรณกรรม
76	137	พระราชทาน
3	127	อยุธยา
2	88	อุทยานประวัติศาสตร์
1	65	ประวัติศาสตร์
58	65	พิธีกรรม
55	60	วารสาร/นิตยสาร
5	43	สถาปัตยกรรม
27	38	ข่าวสาร
6	36	อาหาร
13	31	โรค
23	23	วิชา/หลักสูตร
1	22	สมุดไทย
12	16	แบบแปลน/แผนผัง
2	13	ล้านนา
3	13	ภัยพิบัติ
1	8	รัตนโกสินทร์
6	7	เครื่องราช
6	6	อนุสาวรีย์
2	6	ตัวละครหรือสัตว์ใน วรรณกรรม
1	5	ข้อเสนอแนะ/ความคิดเห็น
2	3	แผนที่, พิกัด
2	2	กฎหมาย

1	2	ชื่อ/ชาย
2	2	การอบรม/สัมมนา
1	1	รางวัล

2.2.8. ตัวอย่างแสดงข้อมูล Keyword ที่ได้จากการจำแนก Tag

เมื่อทำการจำแนกแท็กจากประโยคจำนวน 739168 ประโยคจากฐานข้อมูลจะพบว่าอัลกอริทึมสามารถแยก Keyword จากประโยคดังกล่าวได้จำนวน 1587873 Keywords โดยมีตัวอย่างของ Keyword ในแต่ละ Tag ดังนี้



ID	Keyword	Tag	DB
293914	ทำเป็นรูปปั้น	รายการอิเล็กทรอนิกส์	antique
293915	มีฝาเปิดปิดได้	รายการอิเล็กทรอนิกส์	antique
293916	ทำเป็นรูปมั่งคุดปลอกเปลือกครึ่งลูก	รายการอิเล็กทรอนิกส์	antique
293917	ทำด้วยเงิน	รายการอิเล็กทรอนิกส์	antique
293918	ทำเป็นรูปมะยม	รายการอิเล็กทรอนิกส์	antique
293919	เศียรพระพุทธรูป พระพักตร์รูปไข่	รายการอิเล็กทรอนิกส์	antique
293920	พระเนตรปิดเหลือบลงต่ำ พระนาสิกเล็กและโตง เม็ดพระศกเป็นตุ้ม ขนาดเล็ก พระพักตร์แสดงถึงคว...	รายการอิเล็กทรอนิกส์	antique
293921	ทำเป็นรูปส้มโอ	รายการอิเล็กทรอนิกส์	antique
293922	ทำเป็นรูปมะเขือเทศ	รายการอิเล็กทรอนิกส์	antique
293923	พื้นสีฟ้าปักด้วยสีต่าง ๆ คล้ายลายดินง	รายการอิเล็กทรอนิกส์	antique
293924	พื้นสีม่วงอมชมพู ปักด้วยสีต่าง ๆ คล้ายลายดินง	รายการอิเล็กทรอนิกส์	antique
293925	ปักดอกด้วยสีต่าง ๆ พื้นสีเนื้อ	รายการอิเล็กทรอนิกส์	antique
293926	ปักดอกด้วยสีต่าง ๆ พื้นสีเทา	รายการอิเล็กทรอนิกส์	antique
293927	ยกดอกลอย เป็นลายดอกพิกุล พื้นสีม่วง ดอกสีเขียว เหลือง และมีพื้นสีม่วงครึ่งผืน	รายการอิเล็กทรอนิกส์	antique

รูปที่ 13 ตัวอย่างของ Keyword ในแต่ละ Tag

1 • `SELECT * FROM prepared3.keywords where Tag="โบราณสถาน";`

ID	Keyword	Tag	DB	Table
1	วัดพระเงิน	โบราณสถาน	gistable	HISST
2	วัดพระเจดีย์ทอง	โบราณสถาน	gistable	HISST
3	วัดพระเจ้าคอกหัก	โบราณสถาน	gistable	HISST
4	วัดพระเจ้าองค์ดำ	โบราณสถาน	gistable	HISST
5	วัดพระเจ้านาคหลวง	โบราณสถาน	gistable	HISST
6	วัดพระเจ้าทันใจ	โบราณสถาน	gistable	HISST
7	วัดพระเจ้าล้านทอง	โบราณสถาน	gistable	HISST
8	วัดพระเจ้าล้านช้าง (ร้าง)	โบราณสถาน	gistable	HISST
9	วัดพระเจ้าเหลื่อม	โบราณสถาน	gistable	HISST
10	วัดไผ่เหือง	โบราณสถาน	gistable	HISST
11	วัดพระเจ้านวมเมือง(ร้าง)	โบราณสถาน	gistable	HISST
12	วัดพระเชตุพนวิมลมังคลาราม	โบราณสถาน	gistable	HISST
13	วัดพระไชยเชษฐาธิราช	โบราณสถาน	gistable	HISST
14	วัดปากแซง	โบราณสถาน	gistable	HISST

รูปที่ 14 ตัวอย่างของ Keyword ในแต่ละ Tag

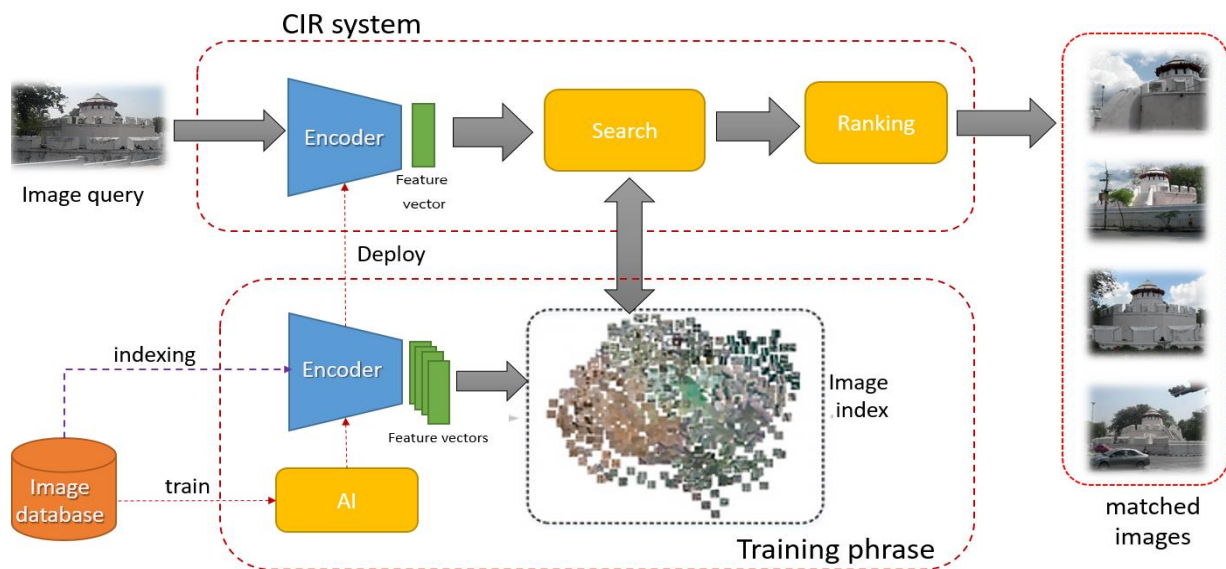
2.2.9. ปัญหาที่เกิดขึ้นและการแก้ไข้ปัญหา

- ปรับรูปแบบการเทรนใหม่ โดยการนำบริบทของคำมาช่วยในการเรียนรู้ Tag
- เพิ่ม dictionary สำหรับการตัดคำ
- ผลลัพธ์ของการทำนาย Tag ผลลัพธ์ดีขึ้นกว่าเดิม คำที่มีความยาวมากมีจำนวนน้อยลงแต่ยังมีปรากฏในบาง Tag ที่มีตัวอย่างการเรียนรู้ของคำจำนวนน้อย
- คำที่มีความยาวมากจะเกิดใน tag ที่มีคำที่ใช้ในการเรียนรู้จำคำไม่มาก เช่น จดหมายเหตุ

3. ระบบสืบค้นข้อมูลโบราณสถานด้วยภาพถ่าย

3.1. การออกแบบระบบ

จากรูปที่ 15 จะแสดงภาพรวมของระบบสืบค้นข้อมูลโบราณสถานของกรมศิลปากร โดยเริ่มจากผู้ใช้อัปโหลดภาพถ่าย (image query) ผ่านเว็บไซต์กรมศิลปากร จากนั้นระบบก็จะเริ่มทำการวิเคราะห์ภาพถ่ายเพื่อสืบค้นภาพโบราณสถานที่มีความคล้ายคลึง (similar image) เพื่อช่วยให้ผู้ใช้สามารถสืบค้นข้อมูลโบราณสถานได้สะดวกและรวดเร็วขึ้น จากภาพที่ 1 เมื่อผู้ใช้ทำการเข้ารหัส (encoding) ภาพถ่ายของผู้ใช้ให้อยู่ในรูปแบบที่เหมาะสมสำหรับการประมวลผลข้อมูล ซึ่งได้แก่ เวกเตอร์คุณลักษณะเด่น (feature vector) หลังจากทีภาพถ่ายถูกนำไปเข้ารหัสให้อยู่ในรูปแบบของเวกเตอร์คุณลักษณะเด่นเรียบร้อยแล้ว ขั้นตอนถัดไปก็คือการนำเวกเตอร์ดังกล่าวไปทำการเปรียบเทียบกับคลังของเวกเตอร์คุณลักษณะเด่นที่ได้จากชุดภาพโบราณสถานของกรมศิลปากรที่มีการจัดทำไว้ล่วงหน้า ผ่านโมดูล Search อย่างไรก็ตามเนื่องจากจำนวนภาพโบราณสถานอาจมีมากและเพิ่มเติมได้ในภายหลัง ทีมที่ปรึกษาจึงต้องออกแบบโมดูล Search เพื่อรองรับปัญหาดังกล่าวที่อาจเกิดขึ้นได้ในอนาคต ผลลัพธ์ของโมดูล Search ก็จะเป็นรายการเวกเตอร์ที่มีความคล้ายคลึงหรือใกล้เคียงกับเวกเตอร์ที่ได้จากการเข้ารหัสภาพถ่ายของผู้ใช้ โดยระบบจะต้องนำผลลัพธ์ดังกล่าวมาประมวลผลเพื่อเรียงลำดับและคัดกรองเฉพาะภาพที่มีความใกล้เคียงกับภาพโบราณสถาน que ผู้ใช้ต้องการสืบค้นต่อไป



รูปที่ 15 ระบบสืบค้นข้อมูลโบราณสถานด้วยภาพถ่าย

จากรูปที่ 15 ระบบสืบค้นข้อมูลโบราณสถานประกอบไปด้วย 3 โมดูล (module) หลัก ได้แก่

1. โมดูลการเข้ารหัสภาพ (Encoding module)
2. โมดูลการค้นหาภาพ (Searching module)
3. โมดูลการจัดลำดับ (Ranking module)

3.2. การพัฒนาโมดูล

3.2.1. โมดูลการเข้ารหัสภาพ (Image Encoding module)

ระบบสืบค้นข้อมูลด้วยภาพ (Image retrieval system) ทัวไปจะมีขั้นตอนที่สำคัญได้แก่ ขั้นตอนการสกัดลักษณะเด่น (feature extraction) เพื่อใช้อธิบายภาพ และแทนภาพ (image representation) ให้อยู่ในรูปแบบที่คอมพิวเตอร์สามารถนำไปวิเคราะห์และประมวลผลได้ง่ายขึ้น หรือที่รู้จักกันว่าเป็นขั้นตอนการเข้ารหัสภาพ (image encoding) ปัจจุบันมีเทคนิคการสกัดลักษณะเด่นของภาพมากมายถูกนำเสนอและสามารถแบ่งออกเป็น 2 ประเภทหลัก ได้แก่

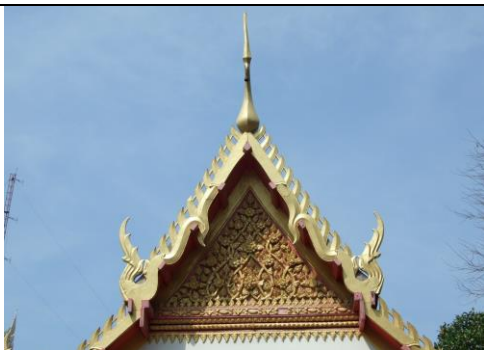
- การสกัดลักษณะเด่นแบบองค์รวม (Global feature extraction) ซึ่งมีเป้าหมายในการสกัดลักษณะเด่นเพื่อใช้อธิบายองค์ประกอบรวมของภาพถ่าย เทคนิคที่สำคัญในกลุ่มนี้ ได้แก่ Local Binary Patterns (LBP), Edge detection, Color histogram และ Convolutional Neural network (CNN) เป็นต้น
- การสกัดลักษณะเด่นแบบเฉพาะบริเวณ (Local feature extraction) ซึ่งมุ่งเน้นที่จะสกัดลักษณะเด่นในบริเวณพิกเซลที่น่าสนใจของภาพผ่านการตรวจจับ keypoints เทคนิคที่สำคัญในกลุ่มนี้ ได้แก่ Scale Invariance Feature Transformation (SIFT), VLAD และ DELF เป็นต้น

ถึงแม้ว่าจะมีเทคนิคการเข้ารหัสภาพถูกนำเสนออย่างมากมาย อย่างไรก็ตามภาพโบราณสถานมักมีลักษณะเฉพาะที่ค่อนข้างแตกต่างจากภาพถ่ายทั่วไป ดังต่อไปนี้

1. ความหลากหลายของประเภทโบราณสถาน
เช่น อาคาร สิ่งปลูกสร้าง เสา สะพาน รูปปั้น อนุสาวรีย์ และซากปรักหักพัง เป็นต้น ซึ่งอาจไม่สามารถจำแนกประเภทได้อย่างชัดเจน
2. ความคล้ายคลึงของแหล่งโบราณสถาน
โบราณสถานบางแหล่ง เช่น โบสถ์ อาคาร และซากปรักหักพัง เป็นต้น มักมีลักษณะคล้ายคลึงกัน ส่งผลให้เทคนิคที่มีอยู่ในปัจจุบันอาจไม่สามารถแยกแยะความแตกต่างของโบราณสถานเหล่านี้ได้อย่างมีประสิทธิภาพ
3. ความแตกต่างของภาพในโบราณสถาน

ถึงแม้ว่าจะเป็นโบราณสถานแหล่งเดียวกันก็อาจมีภาพที่แตกต่างกันอย่างมาก เนื่องจากมุมของการถ่ายภาพ บริเวณที่ถ่ายภาพ เช่น ถ่ายจากภายในหรือภายนอก รวมไปถึงการเปลี่ยนแปลงที่เกิดจากความเสียหายหรือการบูรณะซ่อมแซม เป็นต้น ซึ่งจะส่งผลให้ประสิทธิภาพของเทคนิคเฉพาะบริเวณอาจจับคู่ภาพได้ไม่แม่นยำ

รูปที่ 16 แสดงตัวอย่างของภาพถ่ายโบราณสถานที่มีลักษณะเฉพาะของข้อมูล

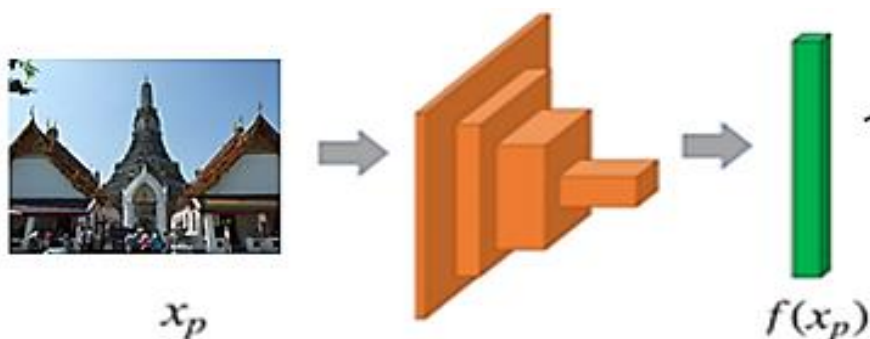
ปัญหา	ตัวอย่างภาพในฐานข้อมูล	
ความหลากหลายของประเภทโบราณสถาน		
		
ความคล้ายคลึงของโบราณสถาน		

ปัญหา	ตัวอย่างภาพในฐานะข้อมูล	
		
ความแตกต่าง ที่มาจาก โบราณสถาน แหล่งเดียวกัน		

จากปัญหาดังกล่าว ผู้วิจัยจึงมีแนวคิดที่จะพัฒนาโมดูลเข้ารหัสภาพโบราณสถานด้วยการสกัดลักษณะเด่นแบบผสมผสาน (hybrid feature extraction) และนำเสนอวิธีการใช้ลักษณะเด่นของภาพเหล่านี้ในการสืบค้นข้อมูลภาพขนาดใหญ่ ข้อดีของการใช้ลักษณะเด่นแบบองค์รวมในการสืบค้นข้อมูลด้วยภาพโบราณสถาน ได้แก่ ความสามารถในการอธิบายโครงสร้างโบราณสถานโดยรวมและความรวดเร็วในการทำงาน จึงเหมาะสมอย่างยิ่งที่จะ

ถูกนำมาใช้ในการคัดกรองภาพที่ไม่เกี่ยวข้องส่วนใหญ่ในฐานข้อมูลภาพ อย่างไรก็ตามเทคนิคการสกัดลักษณะเด่นแบบองค์รวม เช่น การตรวจจับพื้นผิว (texture) สี (color) และขอบภาพ (edge) อาจไม่เหมาะสมสำหรับการสกัดลักษณะเด่นของภาพโบราณวัตถุที่มีความแตกต่างจากภาพทั่วไป

ไม่นานมานี้งานวิจัยด้านคอมพิวเตอร์วิชัน (Computer vision) นำเสนอการใช้โครงข่ายประสาทเทียมคอนโวลูชัน (Convolution Neural Network: CNN) ซึ่งประกอบด้วยชั้นคอนโวลูชัน (Convolution layer) สลับกับชั้นพูลลิง (pooling layer) เพื่อสกัดลักษณะเด่นของภาพสำหรับการทำนายลาเบล (label) หลังจากโครงข่ายประสาทเทียมเรียนรู้สำเร็จแต่ละชั้นของโครงข่ายประสาทเทียมสามารถถูกพิจารณาว่าเป็นลักษณะเด่นที่สามารถในการอธิบายเนื้อหาของภาพแตกต่างกัน โดยลักษณะเด่นที่สกัดได้ในแต่ละชั้นก็สามารถที่จะนำไปใช้ในการจับคู่ภาพ (image matching) เปรียบเทียบกับวิธีการสกัดลักษณะเด่นแบบดั้งเดิม วิธีการสกัดลักษณะเด่นโดยการใช้โครงข่ายประสาทเทียมมีข้อดีในแง่ของความสามารถในการสกัดลักษณะเด่นที่เหมาะสมสำหรับชุดข้อมูลภาพผ่านกระบวนการฝึกฝน (training) ภาพที่ 4 แสดงการสกัดลักษณะเด่นของภาพด้วยวิธีการโครงข่ายประสาทเทียม CNN



รูปที่ 17 โครงข่ายประสาทเทียม CNN สำหรับการเรียนรู้เพื่อสกัดลักษณะเด่น

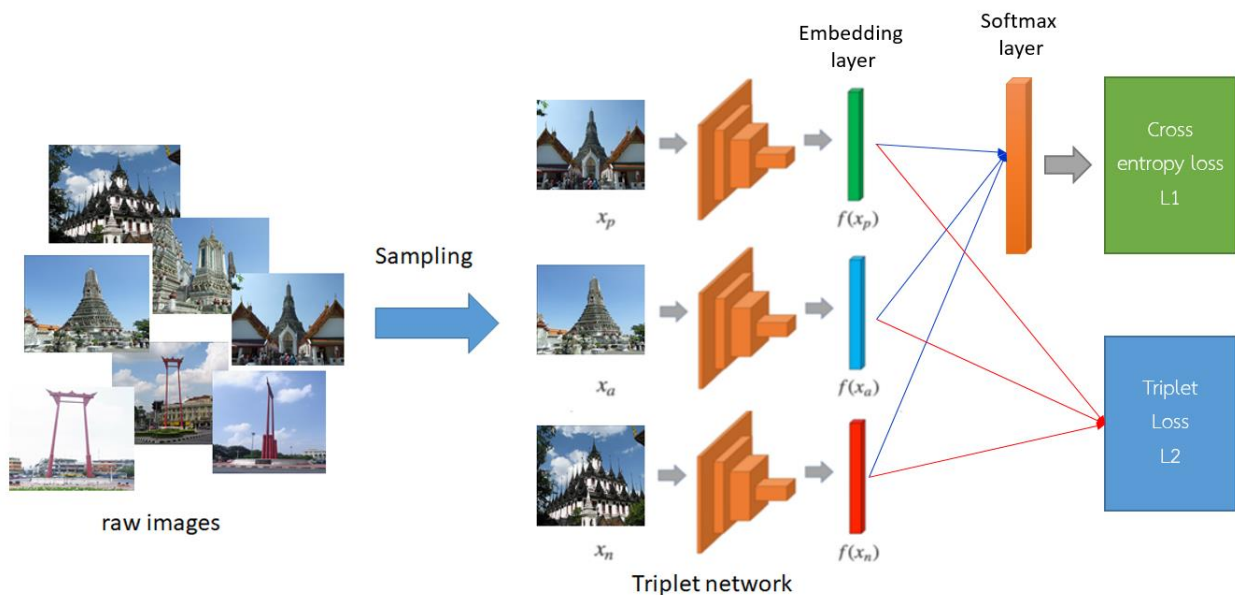
จากรูปที่ 17 ข้อมูลภาพโบราณสถาน x_p จะถูกส่งเข้าไปยังชั้นข้อมูลต่างๆ ของโครงข่ายประสาทเทียม CNN เพื่อสกัดเวกเตอร์ลักษณะเด่น $f(x_p)$ สำหรับทำนายลาเบล (label) ผ่านการออปติไมซ์ (optimization) ฟังก์ชันสูญเสียแบบ Categorical Cross-entropy

$$L_{CE}(\hat{y}_i, y_i) = - \sum_c (y_i^c \log \hat{y}_i^c + (1 - y_i^c) \log(1 - \hat{y}_i^c))$$

โดยที่ y_i^c หมายถึงความน่าจะเป็นของโครงข่ายประสาทเทียมในการทำนายลาเบล c

ถึงแม้ว่าวิธีการโครงข่ายประสาทเทียมแบบ CNN จะสามารถสกัดลักษณะเด่นที่มีความแม่นยำในการทำนายลาเบลของภาพได้ดี แต่ก็ยังมีข้อจำกัดที่สำคัญ ได้แก่ 1) ต้องการข้อมูลภาพในแต่ละลาเบลจำนวนมากและเท่าๆ กัน ซึ่งเป็นไปได้ยาก โดยเฉพาะภาพโบราณสถาน และ 2) ลักษณะเด่นที่สกัดได้อาจไม่สามารถนำไปใช้กับข้อมูลที่ไม่เคยเห็น (unseen data) เช่น โบราณสถานที่ค้นพบใหม่ เนื่องจากโครงข่ายประสาทเทียมจะมุ่งเน้นที่จะทำนายเฉพาะลาเบลที่อยู่ในชุดข้อมูลฝึกฝนเท่านั้น เรียกปัญหานี้ว่า Overfitting

เพื่อป้องกันปัญหาดังกล่าว งานวิจัยนี้นำเสนอวิธีการเพิ่มประสิทธิภาพในการสกัดลักษณะเด่นโดยใช้ข้อมูลไม่มากและลด overfitting โดยทำการปรับปรุงสถาปัตยกรรมโครงข่ายประสาทเทียมให้สามารถสกัดลักษณะเด่นของภาพที่สามารถทำนายลาเบลข้อมูลพร้อมทั้งเพิ่มความสามารถในการเปรียบเทียบความคล้ายคลึงและความแตกต่างระหว่างภาพที่มาจากลาเบลข้อมูลในเวลาเดียวกัน เรียกวิธีการนี้ว่า Multi-task CNN แสดงดังรูปที่ 18



รูปที่ 18 แสดงสถาปัตยกรรมโครงข่ายประสาทเทียมแบบ Multi-task CNN

จากรูปที่ 18 โครงสร้างของโครงข่ายประสาทเทียม Multi-task CNN จะประกอบไปด้วยโครงข่ายประสาทเทียม CNN ที่มีโครงสร้างภายในเหมือนกันทั้ง 3 ชุด ที่มีการแชร์พารามิเตอร์ร่วมกัน (parameter sharing) โดยโครงข่ายประสาทเทียมแต่ละชุดจะรับอินพุตภาพที่มาจากลาเบลประเภทเดียวกัน 2 ภาพ เรียกว่า

ภาพ anchor (X_a) และภาพ positive (X_p) และภาพ negative (X_n) ตามลำดับ โดยโครงข่ายประสาทเทียมแต่ละตัวจะทำการเรียนรู้เพื่อค้นหาความคล้ายคลึงระหว่าง anchor และ positive ขณะเดียวกันก็จะพยายามหาความแตกต่างระหว่าง anchor และ negative โดยจะเปรียบเทียบเวกเตอร์ลักษณะเด่นที่สกัดได้จากโครงข่ายประสาทเทียมแต่ละชุด ขณะเดียวกันเวกเตอร์ลักษณะเด่นที่สกัดได้แต่ละชุดก็จะถูกนำไปผ่านชั้น Softmax เพื่อทำนายลาเบลของแต่ละภาพ

ผลลัพธ์จากการทำนายลาเบลภาพและการเปรียบเทียบความคล้ายคลึงและความแตกต่างระหว่าง anchor, positive และ negative จะถูกนำไปคำนวณค่า loss เพื่อใช้ในการปรับแต่งพารามิเตอร์ของโครงข่ายประสาทเทียม CNN ผ่านการออปติไมซ์ฟังก์ชัน loss รวม

$$L = \gamma \cdot L_{CE}(\hat{Y}_t, Y_t) + \beta \cdot L_t(a, p, n)$$

โดยที่ L_{CE} หมายถึง Cross Entropy loss $\hat{Y}_t$ และ Y_t แทนเวกเตอร์ลาเบลของ triplet

$$t = \langle \hat{y}_a, \hat{y}_p, \hat{y}_n \rangle \text{ และเวกเตอร์ actual label ของ triplet } t$$

ขณะที่ L_t แทน triplet loss function ซึ่งจะเปรียบเทียบระยะทางยูคลีเดียน (Euclidean distance) ระหว่างภาพ anchor, ภาพ positive และภาพ negative ตามลำดับ

$$L_t(a, p, n) = \max\{\|f(a) - f(p)\|_2^2 - \|f(a) - f(n)\|_2^2 + \alpha, 0\}$$

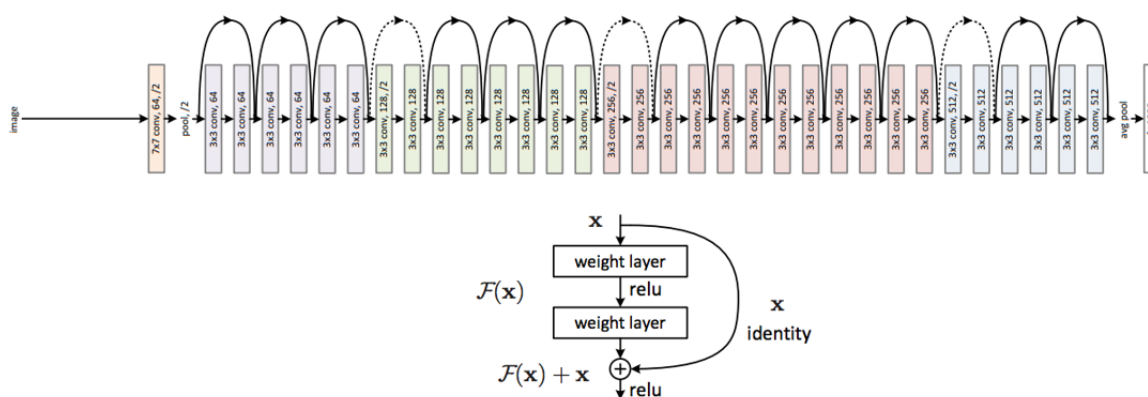
โดยที่ $f(a)$, $f(p)$ และ $f(n)$ หมายถึงลักษณะเด่นที่สกัดได้จากภาพ anchor, positive และ negative

ตามลำดับ

α หมายถึงระยะมาร์จิ้น margin ระหว่าง anchor กับ negative ตามลำดับ

สำหรับงานวิจัยนี้ คณะผู้วิจัยเลือกใช้โครงข่ายประสาทเทียม Residual Network ขนาด 50 ชั้น หรือ ResNet50 ซึ่งเป็นโครงข่ายประสาทเทียม Convolution ที่ถูกออกแบบมาเพื่อทำนายลาเบลภาพขนาด 1,000 object ผ่านการฝึกฝนด้วยชุดข้อมูลภาพ ImageNet ขนาด 1.2 ล้านภาพ โครงข่ายประสาทเทียม ResNet50 จะมีการนำชั้น Convolution มาซ้อนกันและมี skip connection เพื่อแก้ปัญหาของ gradient vanish เรียกการเชื่อมต่อแบบนี้ว่า Residual block

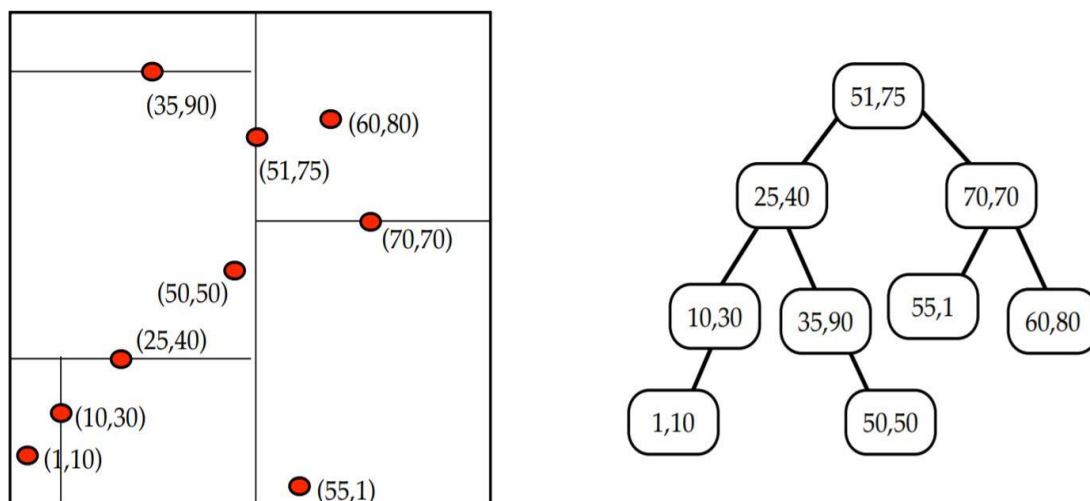
Residual Networks (ResNet50)



รูปที่ 19 สถาปัตยกรรมโครงข่ายประสาทเทียม ResNet50

3.2.2. โมดูลการค้นหาข้อมูลภาพ

หลังจากการฝึกฝนโครงข่ายประสาทเทียม Multi-task ResNet50 เพื่อสกัดเวกเตอร์ลักษณะเด่นของภาพเสร็จเรียบร้อยแล้ว ภาพโบราณสถานในฐานะข้อมูลกรมศิลปากรจะถูกนำมาสกัดเวกเตอร์ลักษณะเด่นโดยนำไปผ่านโครงข่ายประสาทเทียม anchor อย่างไรก็ตามจำนวนภาพโบราณสถานอาจมีมากส่งผลให้ใช้เวลาในการสืบค้นเพื่อจับคู่ภาพที่มีลักษณะใกล้เคียงกันกับภาพควรี (Query image) เพื่อที่จะลดเวลาในการเปรียบเทียบ งานวิจัยนี้จะมีการทำดัชนี (index) ของเวกเตอร์ภาพโดยใช้โครงสร้างข้อมูล KD tree ซึ่งเป็นโครงสร้างข้อมูลประเภทต้นไม้ไบนารี (binary tree) ที่ถูกนำมาใช้ในการค้นหาข้อมูลที่มีความคล้ายคลึงมากที่สุด (nearest neighbors) อย่างมีประสิทธิภาพ แสดงได้ดังรูปที่ 20



รูปที่ 20 การสร้างดัชนีเพื่อสืบค้นข้อมูลด้วยต้นไม้ KD tree

จากรูปที่ 20 การสร้างต้นไม้ KD tree จะเริ่มจากการแบ่งแยกข้อมูลออกเป็นสองฝั่ง ฝั่งที่ถูกแบ่งนั้นจะมีจุดครึ่งหนึ่งจากทั้งหมดของตอนแรกที่ยังไม่ถูกแบ่ง จะแก้ปัญหาผ่านการแหว่ผ่านต้นไม้เริ่มตั้งแต่รากถึงใบโดยการแก้จุดที่กำหนด ในทุกๆที่ที่ถูกแบ่ง วิธีการนี้ขึ้นอยู่กับระยะทางที่กล่าวถึงในคำถาม กิ่งเพื่อนบ้านที่มีตัวที่กำลังค้นหาอยู่นั้นอาจจะต้องถูกนำมาวิเคราะห์ด้วย โดยใช้เวลาในการหาคำตอบเท่ากับ $O(\log N)$ ในกรณีที่เลวร้ายที่สุดคือมีจุดกระจายกันอย่างไม่เป็นระเบียบ ในทางปฏิบัติ เรานั้นสนใจแค่การหาค่าสับเซตอันใดก็ได้ของกลุ่มของจุดทั้งหมดที่ปรากฏให้เห็นในระยะทางที่สั้นที่สุดจากจุดที่กำหนดไว้ก่อนหน้านี้ จุดประสงค์ของแต่ละกิ่งของต้นไม้คือการเดาว่าจุดที่ใกล้ที่สุดจากในกลุ่มที่ตั้งอยู่ในครึ่งหนึ่งของปริภูมิที่ประกอบไปด้วยจุดที่ต้องการรู้ วิธีนี้อาจไม่ได้ผลเสมอไปแต่มันเป็นการเริ่มต้นในการแก้ไขปัญหาของตัวเองที่ดี หลังจากที่ใช้วิธีนี้วนซ้ำครบทุกกิ่งแล้ว เราจะสามารถเปรียบเทียบว่าระยะทางที่ได้จากผลลัพธ์กับระยะทางที่สั้นที่สุดจากจุดที่ต้องการถึงระนาบที่แบ่งไว้ ว่ากิ่งไหนนั้นใช้ระยะทางที่สั้นที่สุด โดยระยะทางที่สั้นที่สุดอาจจะเกิดขึ้นอีกจากฐานข้อมูลอีกอันที่ถูกแบ่งก็ได้ เพราะฉะนั้นบางทีคุณอาจจะต้องพิจารณาทั้งสองฝั่งที่ถูกแบ่ง แต่ถึงอย่างไรก็ตามถ้าระยะทางไปกลับจากจุดที่ต้องการกับระนาบที่แบ่งไว้ นั้นยาวกว่าระยะทางที่ถูกค้นหาก่อนหน้านี้ เราก็ไม่จำเป็นที่จะต้องวิเคราะห์อีกฝั่งหนึ่งแล้ว วิธีนี้ใช้เวลาสั้นกว่าวิธีการค้นหาแบบเส้นตรงเพราะว่าระยะทางระหว่าง จุดที่ต้องการทราบกับกลุ่มของจุดที่ใกล้ที่สุดนั้นเกือบจะเทียบเท่ากับศูนย์ วิธีนี้จะใช้ได้ผลก็ต่อเมื่อมีการค้นหาโดยใช้จุดที่ต้องการทราบ

สำหรับการสร้างดัชนีภาพโบราณสถาน ปัจจุบันมีการจัดทำดัชนีภาพโบราณสถานจำนวน 110 แห่ง จำนวน 3,561 ภาพ แสดงดังตารางที่ 9 รายการโบราณสถานที่ถูกเลือกมาทำรายการดัชนี

ตารางที่ 9 รายการโบราณสถานที่ถูกเลือกมาทำรายการดัชนี

รหัสโบราณสถาน	ชื่อโบราณสถาน
0000002	ป้อมมหากาฬพร้อมด้วยปราการ
0000003	เสาชิงช้า
0000005	วัดสุทัศน์เทพวราราม
0000008	วัดปรินายก
0000009	วัดสระเกศ
0000010	วัดชนะสงคราม
0000011	สะพานม้ฆวาทรงสรรค์
0000013	สะพานดำรงสถิตย
0000015	พระราชวังบวรสถานมงคล และกำแพงพระราชวังบวรสถานมงคล(วังหน้า)
0000019	วัดเทพธิดาราม
0000021	วัดบูรณศิริมตยาราม
0000024	กำแพงเมืองหน้าโรงเรียนวัดบวรนิเวศ
0000025	ซุ้มประตูแพร่งสรรพศาสตร์
0000030	สะพานช้างโรงสี
0000031	สะพานอุบลรัตน์
0000032	สะพานผ่านฟ้าลีลาศ
0000033	อนุสาวรีย์รัชกาลที่ 1
0000036	สะพานเจริญรัช 31
0000037	สะพานเจริญศรี
0000039	วัดราชนััดดาราม

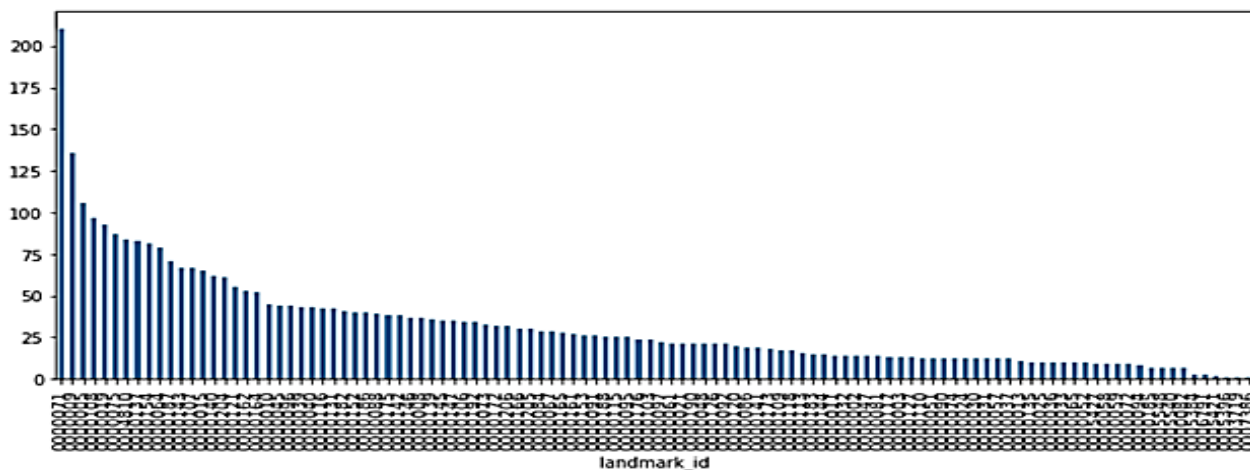
รหัสโบราณสถาน	ชื่อโบราณสถาน
0000040	วัดมหรณพาราม
0000045	เรือนจำพิเศษกรุงเทพมหานคร
0000047	วัดตรีทศเทพวรวิหาร
0000048	วังกรมพระราชวังบวรวิชัยชาญ (วังใหม่)
0000051	สะพานสมมตอมรรค
0000052	อาคารเลขที่ 208 (บ้านตัดทรง)
0000057	สะพานภาณุพันธุ์
0000058	สะพานมอญ
0000059	สะพานผ่านพิภพลีลา
0000061	อาคารโรงพิมพ์คุรุสภา (โรงเรียนช่างพิมพ์วัดสังเวช)
0000064	วัดมกุฏกษัตริยาราม
0000065	ป้อมพระสุเมรุ
0000067	ท้องพระโรงตำหนักกลางวังท่าพระ และตำหนักพรรณราย
0000071	วัดพระเชตุพนวิมลมังคลาราม
0000072	พระราชวังสราญรมย์ (กระทรวงการต่างประเทศเดิม)
0000073	คลังราชการ
0000074	สะพานปีกุน
0000075	วัดราชาธิวาส
0000081	สะพานเสาวนี
0000082	สะพานขมิ้มมรุเชฐ
0000084	พระอุโบสถวัดโบสถ์สามเสน
0000086	ศาลเจ้าแม่ทับทิม
0000088	วัดใหม่ทองเสน

รหัสโบราณสถาน	ชื่อโบราณสถาน
0000090	วัดธรรมาภิรตาราม
0000091	วัดจอมสุตาราม(วัดไผ่รงาม)
0000092	พระตำหนักเพชรรัตน์
0000094	สะพานวรเสรษฐ
0000095	พระตำหนักสวนจิตรลดา วังปารุสกวัน (ด้านเหนือ)
0000096	วัดเบญจมบพิตรดุสิตวนาราม
0000097	กระทรวงศึกษาธิการ(วังจันทรวงษ์)
0000103	บ้านอับดุลราฮิม
0000106	อาคารศุลกสถาน (สถานีตำรวจดับเพลิงบางรัก)
0000107	วัดพุทธวนาราม ราชวรวิหาร (วัดประทุมวนาราม)
0000108	วัดบรมนิวาส
0000109	สถานีวิทยุศาลาแดง
0000110	สะพานมหาตมาทิลยอุทิศ
0000113	ตึก S.A.B.
0000115	วัดสิตาราม
0000116	สำนักงานประปาสาขาแม่น้ำศรี
0000117	ศาลากลางจังหวัดมณฑลบุรี(เดิม) (สำนักงานเขตมณฑลบุรี)
0000118	สถานีตำรวจนครบาลมณฑลบุรี
0000120	สะพานพิทยเสถียร
0000121	วัดจักรวรรดิราชาวาส (สามปลื้ม)
0000126	ศาลเจ้ากวางตุ้ง
0000127	ธนาคารกรุงเทพ สาขาสำเพ็ง
0000131	วัดราชคฤห์

รหัสโบราณสถาน	ชื่อโบราณสถาน
0000135	วัดบุคคโล
0000137	วัดหงส์รัตนาราม
0000139	พระราชวังเดิม
0000140	วัดอรุณราชวราราม
0000141	วัดท่าพระ
0000142	วัดสังข์กระจาย
0000143	วัดโมลีโลกยาราม
0000144	โรงเรียนฤทธิธรรงค์รอน
0000145	วัดอนงคาราม
0000146	วัดทองนพคุณ
0000147	วัดทองธรรมชาติ
0000152	วัดรัชฎาธิษฐาน
0000153	วัดกาญจนสิงหาสน์
0000154	วัดสุวรรณาราม
0000161	สำนักงานเกษตรกรุงเทพมหานคร(บ้านเจ้าพระยารัตนาธิเบศร์ ศาล รัฐธรรมนูญ)
0000162	วัดดุสิตาราม (วัดเสาประโคน)
0000164	กรมอุทกหารเรือ (ปล่อยคว้นหมอน้ำ)
0000165	วัดอ่างแก้ว
0000166	วัดอัปสรสวรรค์
0000172	วัดประเสริฐสุทธาวาส
0000173	โรงเรียนวัดเปาโรหิตย์
0000175	วัดคฤหบดี

รหัสโบราณสถาน	ชื่อโบราณสถาน
0000176	วัดใหม่เทพนิมิตร
0000182	บ้านเลขที่ 19 (บ้าน พลตรี หม่อมราชวงศ์คึกฤทธิ์ ปราโมช)
0000183	หอการค้าไทย-จีน (เดิม)
0000185	กรมสรรพาวุธทหารบก
0000186	วัดทองสุทธาราม
0000187	วัดราชโอรส
0000188	วัดแก้วไพฑูรย์
0000190	วัดไทร
0000192	วัดหนัง
0000204	วัดเขมาภิรตาราม
0000205	พระอุโบสถวัดปราสาท
0000206	พระอุโบสถและพระวิหารวัดชมพูเวก
0003719	วัดเวียงเดิม(วัดโบราณ)
0004810	วัดมหาธาตุวรวิหาร
0005384	หน่วยบัญชาการกำลังสำรอง
0005396	วัดแก้วฟ้าจุฬามณี
0005421	วัดจำปา
0005427	วัดดาวคนอง
0005540	วัดโสมนัสวิหารราชวรวิหาร
0005548	สะพานระพีพัฒนภาค
0006791	วัดสรรพยาวัฒนาราม
0007386	เจดีย์วัดสระเกศ

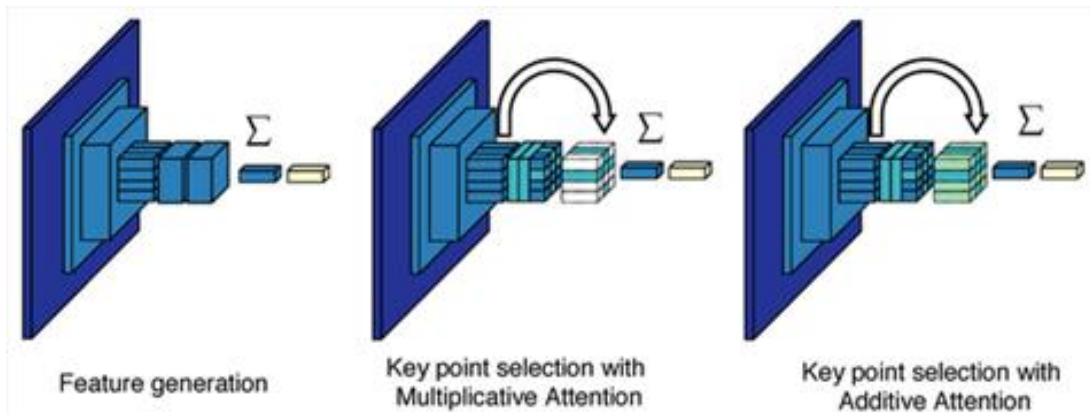
รูปที่ 21 จำนวนภาพของโบราณสถาน แสดงจำนวนภาพซึ่งสัมพันธ์กับโบราณสถานแต่ละแห่ง เรียงลำดับจากมากที่สุดไปน้อยที่สุด จากรูปที่ 21 จำนวนภาพของโบราณสถาน จะเห็นได้ว่าโบราณสถานที่มีจำนวนภาพมากที่สุด ได้แก่ “วัดพระเชตุพนวิมลมังคลาราม” ซึ่งมีจำนวนถึง 210 ภาพ ขณะที่โบราณสถานที่มีจำนวนภาพน้อยที่สุด ได้แก่ “เจดีย์วัดสระเกศ” “วัดเวียงเดิม(วัดโบราณ)” และ “วัดแก้วฟ้าจุฬามณี” ซึ่งมีจำนวนเพียง 1 ภาพเท่านั้น



รูปที่ 21 จำนวนภาพของโบราณสถาน

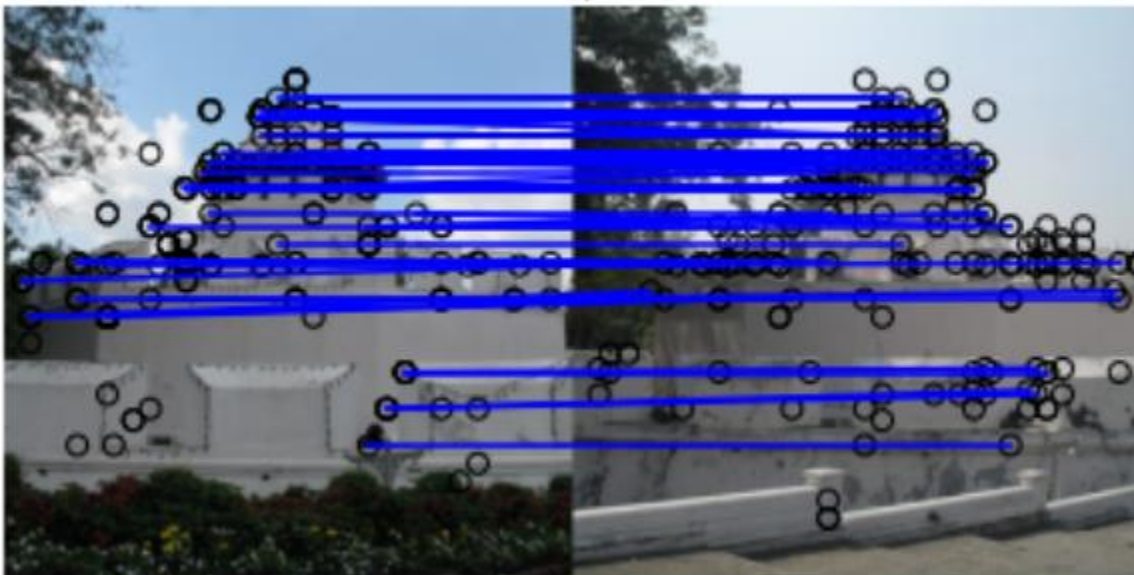
3.2.3. โมดูลจัดอันดับ (Ranking module)

หลังจากการจับคู่ภาพระหว่างภาพคิวรีและภาพจากฐานข้อมูลโดยเปรียบเทียบเวกเตอร์ลักษณะเด่นของภาพที่สกัดได้จากโครงข่ายประสาทเทียม ผลลัพธ์ในการจับคู่ภาพอาจได้รับที่มีลักษณะคล้ายคลึงกับภาพคิวรีจำนวนมาก ในทางปฏิบัติจำเป็นอย่างยิ่งที่จะต้องมีการจัดเรียงภาพตามลำดับความคล้ายคลึงจากมากที่สุดไปหาน้อยที่สุด เนื่องจากข้อจำกัดของการวิเคราะห์ภาพแบบองค์รวม ส่งผลให้ระบบอาจดึงผลลัพธ์ที่ไม่เกี่ยวข้องขึ้นมาแสดงแก่ผู้ใช้ เพื่อที่จะลดปัญหาดังกล่าวการวิเคราะห์ภาพแบบเฉพาะบริเวณจะช่วยจัดลำดับและกรองผลลัพธ์ที่ไม่เกี่ยวข้องออกจากรายการผลลัพธ์ ในงานวิจัยนี้นำเสนอการประยุกต์ใช้โครงข่ายประสาทเทียมเพื่อสกัดลักษณะเด่นเฉพาะบริเวณ โดยใช้โครงข่ายประสาทเทียม DELF (Deep Local Feature) ซึ่งมีความสามารถในการทำนายและเลือก keypoints ที่น่าสนใจในการอธิบายเนื้อหาของภาพ แสดงดังภาพที่ 9

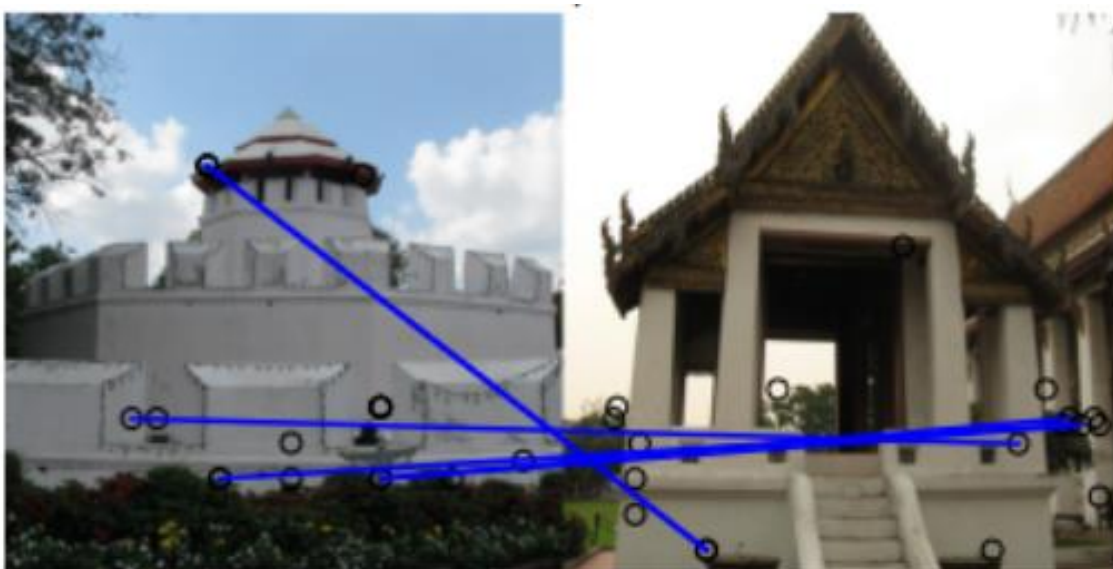


รูปที่ 22 โครงข่ายประสาทเทียม DELF

จากรูปที่ 22 โครงข่ายประสาทเทียม DELF โครงข่ายประสาทเทียม DELF จะมีการปรับแต่งโครงข่ายประสาทเทียม ResNet50 เพื่อให้สกัดลักษณะเด่นของภาพและใช้เวกเตอร์ลักษณะเด่นดังกล่าวในการทำนาย keypoints ที่น่าสนใจในภาพผ่านชั้น attention ปัจจุบันโครงข่ายประสาทเทียม DELF มีโมเดลพร้อมใช้งานทันที โดยมีการฝึกฝนด้วยชุดข้อมูลภาพขนาดใหญ่ Google Landmark Dataset ในงานวิจัยนี้ประยุกต์ใช้ DELF เพื่อจัดลำดับและกรองผลลัพธ์ที่ไม่เกี่ยวข้องกับภาพควีรี โดยการคำนวณค่าคะแนนที่ได้จากการจับคู่ระหว่าง keypoints ของภาพควีรีและภาพผลลัพธ์ที่ถูกเลือกจากโครงข่ายประสาทเทียม Multi-task CNN แสดงดังรูปที่ 23 และรูปที่ 24



รูปที่ 23 แสดงผลลัพธ์การจับคู่ keypoints ของภาพที่มาจากโบราณสถานแห่งเดียวกัน



รูปที่ 24 แสดงผลลัพธ์การจับคู่ keypoints ของภาพที่มาจากโบราณสถานที่แตกต่างกัน

เวกเตอร์ของภาพคิวรีและเวกเตอร์ของภาพในฐานข้อมูลมาพิจารณาพร้อมกับผลลัพธ์ของการจับคู่ keypoints ระหว่างภาพทั้งสอง และสร้างเป็นฟังก์ชัน score ดังนี้

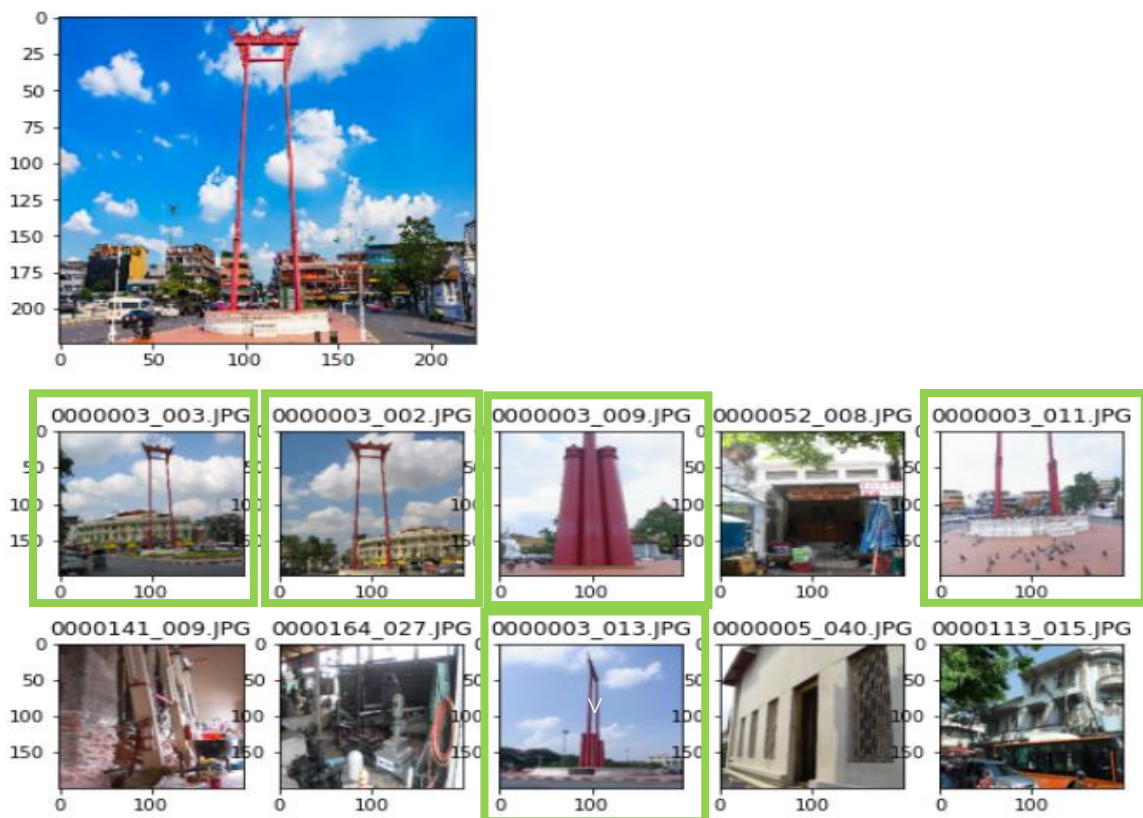
$$R(I_q, I_k) = (1 - \cosine(f(q), f(k))) * \log(\text{RANSAC}(I_q, I_k))$$

โดยที่ $R(I_q, I_k)$ แทนค่า score ระหว่างภาพควีรี I_q และภาพ candidate I_k

$\cosine(f(q), f(k))$ แทนระยะทางโคไซน์บนเวกเตอร์ภาพ I_q และ I_k

$RANSAC(I_q, I_k)$ แทนจำนวน keypoint ที่ถูกจับคู่ระหว่างภาพได้

ค่าที่ได้จากฟังก์ชัน R จะถูกใช้เป็นตัวคะแนนสำหรับการจัดลำดับรายการภาพในฐานข้อมูลโบราณสถานที่ถูกจับคู่ได้โดยจะเรียงลำดับผลลัพธ์จากภาพที่มีคะแนนมากที่สุดไปยังภาพที่มีคะแนนน้อยที่สุดเพื่อนำเสนอผู้ใช้ ภาพที่ 10 แสดงผลลัพธ์การสืบค้นข้อมูลด้วยภาพควีรีทดสอบ (test query)



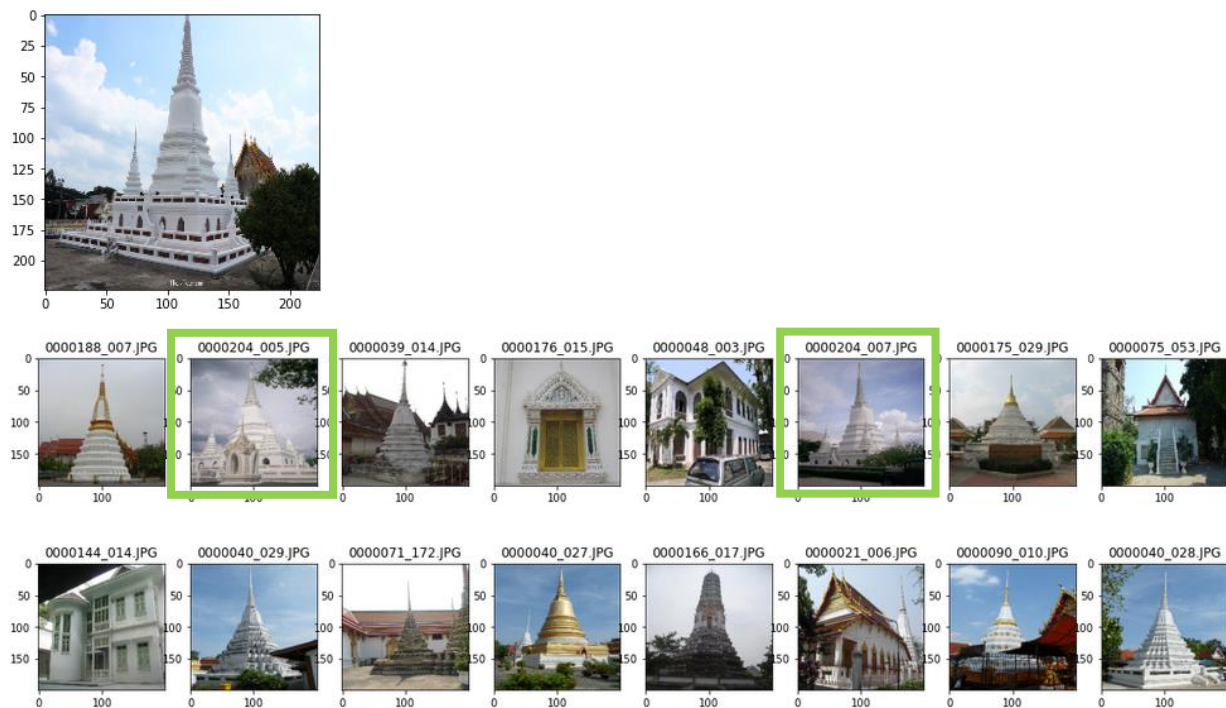
รูปที่ 25 ผลลัพธ์การสืบค้นด้วยภาพควีรี “เสาชิงช้า”



รูปที่ 26 ผลลัพธ์การสืบค้นด้วยภาพควีรี “สะพานไม้หว้าฝรั่งเศส”



รูปที่ 27 ผลลัพธ์การสืบค้นด้วยภาพทดสอบ “บ้านอับดุลลาฮิม”



รูปที่ 28 ผลลัพธ์การสืบค้นด้วยภาพทดสอบ “วัดเขมาภิไธยาราม”

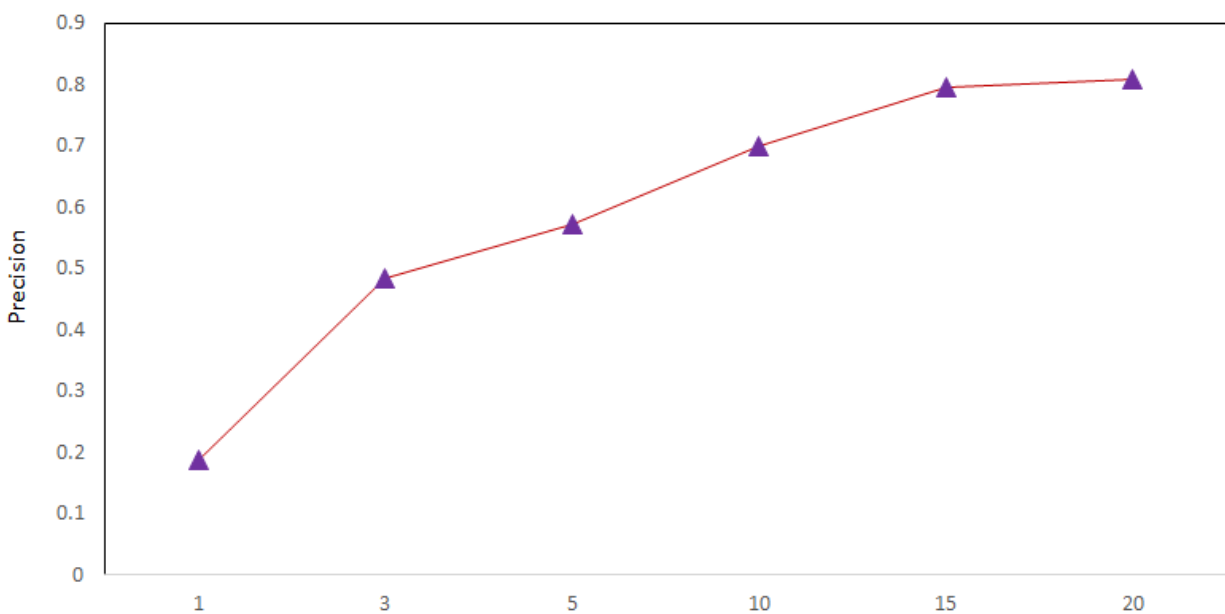
ในการทดสอบประสิทธิภาพของระบบสืบค้นทั่วไป จะมีการแบ่งชุดข้อมูลภาพในฐานข้อมูลโบราณสถาน ออกเป็นชุดข้อมูลฝึกฝน (training set) และชุดข้อมูลทดสอบ (test set) โดยจะสุ่มหยิบภาพในโบราณสถานที่มี ภาพมากกว่า 5 ภาพขึ้นไป อย่างละ 2 ภาพ ซึ่งจะทำให้มีภาพทดสอบทั้งสิ้น 219 ภาพ จากนั้นทำการวัดความ แม่นยำจากรายการผลลัพธ์ที่ระบบสามารถสืบค้นได้ K ลำดับแรกว่ามีภาพโบราณสถานที่ตรงกันกับภาพทดสอบ มากน้อยเพียงใด หรือ Precision@K ดังนี้

$$\text{Precision@K (q, R)} = \frac{|\{j \in R_K | C(j) \cap C(q)\}|}{|R_K|}$$

โดยที่ q หมายถึงภาพทดสอบ

R_K แทนเซตของผลลัพธ์ที่ระบบสืบค้นได้ K ลำดับแรก และ

$C(j)$ หมายถึงโบราณสถานของภาพ j



รูปที่ 29 แสดงผลลัพธ์ของการทดสอบระบบโดยชุดข้อมูลทดสอบ 219 ภาพ บนค่าเฉลี่ยของ Precision@K

ภาพที่ 15 ผลลัพธ์การทดสอบบนค่าเฉลี่ย Precision@K

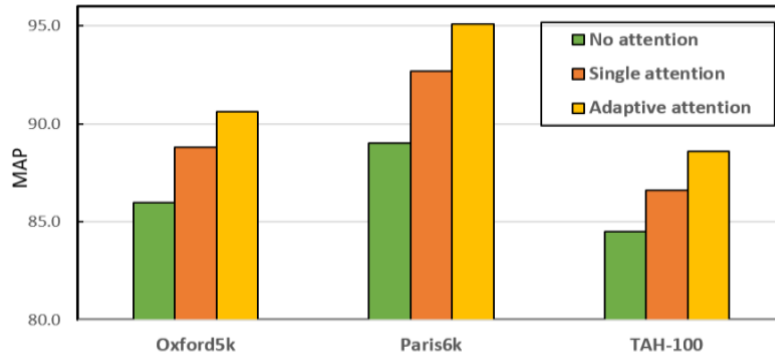
Method	Dim.	Datasets		
		Oxford5k	Paris6k	TAH-100
SPoC	1024	70.8	78.3	67.2
R-MAC	256	87.8	93.5	84.8
GeM (<i>power p = 3</i>)	512	85.1	90.6	82.6
*PAN	512	90.6	95.4	89.4

ตารางที่ 10 แสดงการเปรียบเทียบประสิทธิภาพของโมเดลค้นคืนภาพที่นำเสนอในงานวิจัยและโมเดลต่างๆ ที่มีใช้ในปัจจุบัน โดยเปรียบเทียบกับชุดข้อมูล benchmark ได้แก่ Oxford5k และ Paris6k ขณะที่ชุดข้อมูลโบราณสถานของกรมศิลปากร ได้แก่ TAH-100 จะมีการรวบรวมจาก 100 แหล่งโบราณสถาน

จากตารางที่ 10 แสดงการเปรียบเทียบประสิทธิภาพของโมเดลค้นคืนภาพที่นำเสนอในงานวิจัยและโมเดลต่างๆ ที่มีใช้ในปัจจุบัน โดยเปรียบเทียบกับชุดข้อมูล benchmark ได้แก่ Oxford5k และ Paris6k ขณะที่ชุดข้อมูลโบราณสถานของกรมศิลปากร ได้แก่ TAH-100 จะมีการรวบรวมจาก 100 แหล่งโบราณสถาน พบว่าโมเดลที่นำเสนอ PAN มีความแม่นยำในการค้นคืนภาพโบราณสถานที่มีความคล้ายคลึงกัน เมื่อเปรียบเทียบกับวิธีการอื่นๆ เช่น SPoC R-MAC และ GeM ในการเปรียบเทียบประสิทธิภาพของ SPoC พบว่าวิธีการที่นำเสนอมีความแม่นยำเพิ่มขึ้น +27.96% สำหรับชุดข้อมูล Oxford5k และ +21.83% สำหรับชุดข้อมูล Paris6k ตามลำดับ ยิ่งไปกว่านั้นในการเปรียบเทียบกับ GeM โดยกำหนดพารามิเตอร์ $p = 3$ พบว่า

PAN ยังให้ประสิทธิภาพเพิ่มขึ้น +6.56% สำหรับ Oxford5k และ +5.29% สำหรับ Paris6k ตามลำดับ ในชุดข้อมูลกรมศิลปากร TAH-100 ซึ่งเป็นชุดข้อมูลที่มีความยากกว่าชุดข้อมูล Oxford5k และ Paris6k ยังพบว่าวิธีการที่นำเสนอยังมีประสิทธิภาพมากกว่าวิธีการปัจจุบันอีกด้วย

สำหรับการเปรียบเทียบระหว่างโมเดล R-MAC และโมเดล PAN ซึ่งเป็นโมเดลที่ถูกนำเสนอ ยังพบว่าโมเดล PAN มีความแม่นยำสูงกว่าในชุดข้อมูลทั้ง 3 ถึงแม้ว่า R-MAC จะมีวิธีการสกัด feature ในแต่ละพื้นที่ (region) ของภาพ ซึ่งมีลักษณะใกล้เคียงกับวิธีการ PAN ที่นำเสนอ อย่างไรก็ตาม เนื่องจากโมเดล PAN ใช้วิธีการ attention เพื่อที่จะเพิ่มประสิทธิภาพของ features ที่สกัดได้จากภาพ จึงส่งผลให้โมเดล PAN ได้รับ feature ที่มีความแม่นยำมากกว่าในการค้นคืนภาพโบราณสถาน



รูปที่ 30 แสดงการเปรียบเทียบผลลัพธ์ระหว่างการใช้ single attention, adaptive attention และ no attention ในโมเดล PAN

จากผลลัพธ์ในรูปที่ 30 การใช้ adaptive attention เพื่อปรับปรุงคุณภาพของ feature สำหรับค้นคืนภาพโบราณสถานจะให้ผลลัพธ์ที่แม่นยำกว่าวิธีการ single attention ที่ใช้งานกันอยู่ปัจจุบัน ขณะที่การไม่ใช้ feature attention ก็จะทำให้ประสิทธิภาพของโมเดลลดลง แสดงให้เห็นว่าการปรับปรุงคุณภาพของ feature เป็นเรื่องสำคัญในการค้นคืนภาพ เนื่องจากคุณภาพของภาพอินพุตมีความหลากหลายและแตกต่างกัน



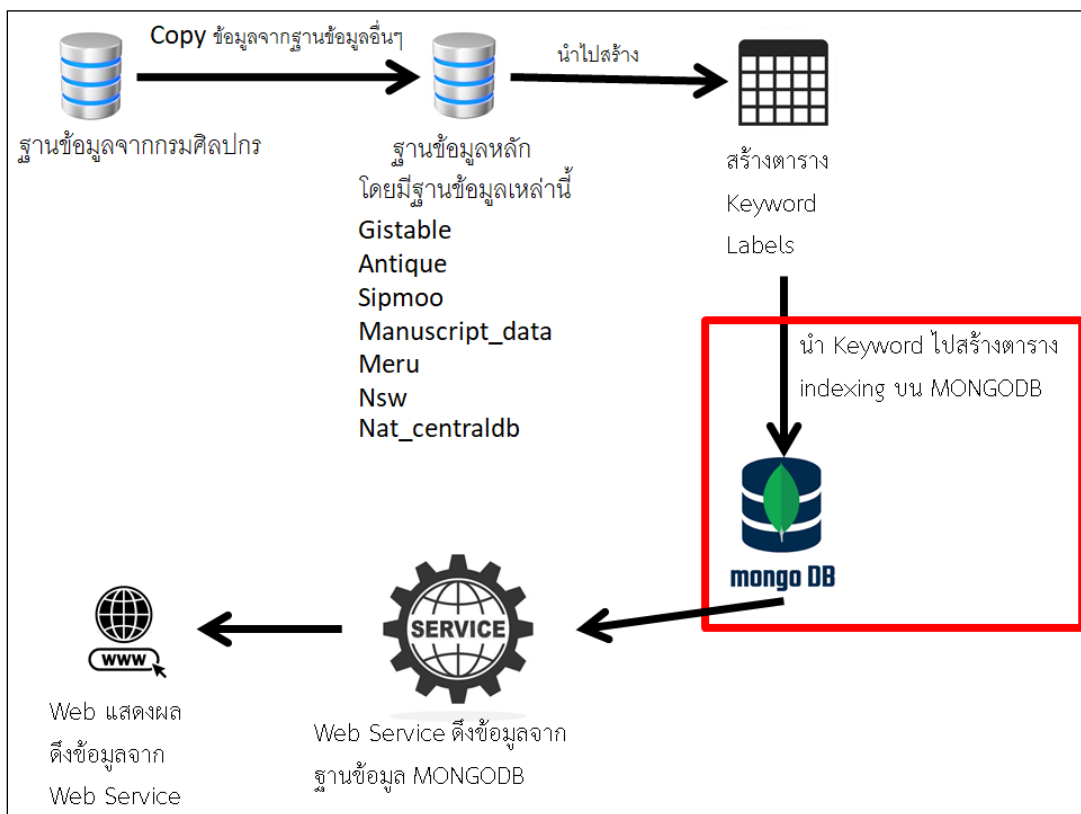
รูปที่ 31 แสดงตัวอย่างผลลัพธ์ 7 ลำดับแรกที่โมเดลสามารถจับคู่ได้ โดยภาพคอลัมน์ซ้ายมือสุดเป็นภาพอินพุต ขณะที่ภาพในกรอบสีเขียวเป็นภาพของสถานที่เดียวกันกับภาพอินพุต ขณะที่ภาพกรอบสีแดงแสดงผลลัพธ์ที่คัดเคลื่อน ในชุดข้อมูล TAH-10



รูปที่ 32 แสดงตัวอย่างผลลัพธ์ 7 ลำดับแรกที่โมเดลสามารถจับคู่ได้ โดยภาพคอลัมน์ซ้ายมือสุดเป็นภาพอินพุต ขณะที่ภาพในกรอบสีเขียวเป็นภาพของสถานที่เดียวกันกับภาพอินพุต ขณะที่ภาพกรอบสีแดงแสดงผลลัพธ์ที่คาดเคลื่อน ในชุดข้อมูล Oxford5k และ Paris6k ตามลำดับ

4. การเพิ่มชุดข้อมูลใหม่ในระบบและการจัดการฐานข้อมูล

4.1. การเพิ่มฐานข้อมูลงานอนุรักษ์



รูปที่ 33 แสดงข้อมูลรูปแบบโครงสร้าง web site ในระบบเดิม

จากรูปที่ 33 โดยระบบเริ่มต้นจากการดึงข้อมูลจาก ฐานข้อมูลกรมศิลปากร มาเก็บไว้ในฐานข้อมูลหลัก ซึ่งในขั้นตอนนี้จะดำเนินการเพียงครั้งเดียว ทำให้เมื่อกรมศิลปากร ดำเนินการปรับปรุงข้อมูลที่ฐานข้อมูลของตนเอง ข้อมูลเหล่านั้นจะไม่ได้รับการปรับตาม เมื่อข้อมูลถูกดึงมาเก็บไว้ที่ฐานข้อมูลหลัก โดยฐานข้อมูลที่ดึงมาเพื่อใช้ในการสืบค้นคือฐานข้อมูลเหล่านี้

- 1) โบราณวัตถุ
- 2) โบราณสถาน
- 3) การนำเข้าส่งออก
- 4) จดหมายเหตุ

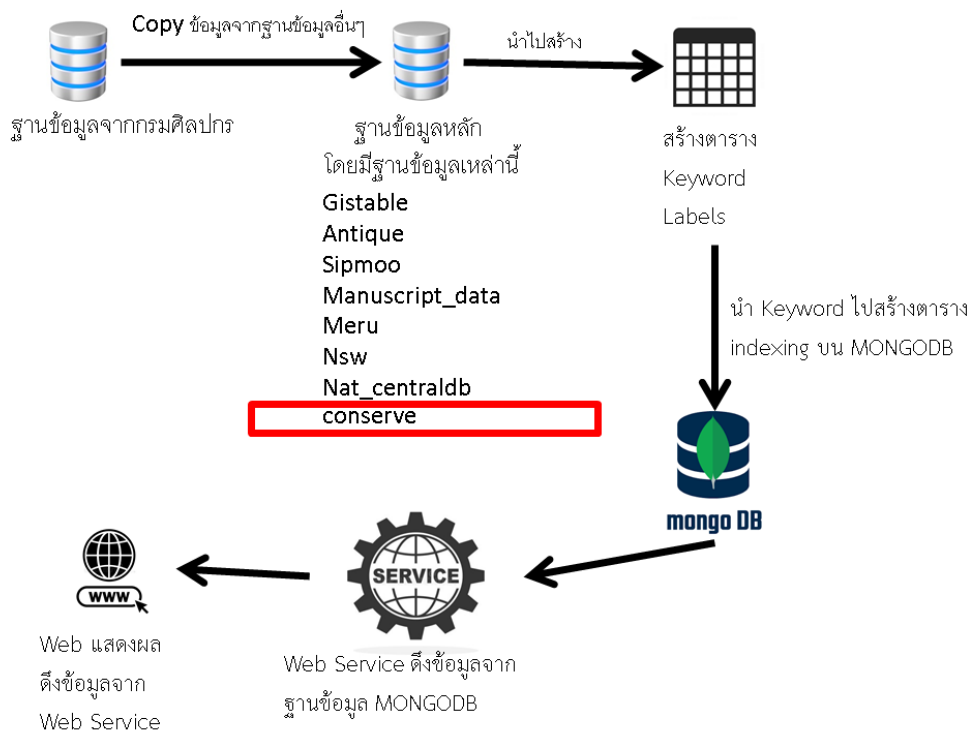
5) เอกสารโบราณ

6) พระเมรุ

7) สิบหมื่น

ฐานข้อมูลทั้งหมดจะถูกนำไปสร้างตาราง Labels เก็บไว้ในฐานข้อมูล MYSQL หลังจากสร้างตาราง Labels ในฐานข้อมูล MYSQL ซึ่งฐานข้อมูล MYSQL เป็นฐานข้อมูล relation ทำให้การดึงข้อมูลมาแสดงผลค่อนข้างช้า จึงให้โปรแกรม ดึงข้อมูลจาก ฐานข้อมูล MYSQL มาเก็บไว้ในฐานข้อมูล MONGODB สำหรับการสร้างฐานข้อมูล MONGODB ต้องดึงข้อมูลจาก ฐานข้อมูล MYSQL เนื่องจาก MYSQL เป็นฐานข้อมูล ในรูปแบบฐานข้อมูลเชิงสัมพันธ์ ในการดึงข้อมูลมีความจำเป็นต้องดึงข้อมูลหลายตาราง ทำให้การค้นหาข้อมูลใช้เวลานาน จึงทำให้แสดงผลของ web site ช้า โดยการจัดเรียงข้อมูลใหม่ในรูปแบบของฐานข้อมูลดัชนี โดยข้อมูลนั้นได้มีการจัดเรียงไว้เรียบร้อยแล้ว หลังจากนั้น WEB SERVICE สามารถดึงข้อมูลไปแสดงผลได้โดยตรง ทำให้การค้นหาข้อมูลรวดเร็วกว่าเดิมมาก ทำให้ web site มีความเร็วมากขึ้น

จากปัญหาระบบไม่ได้ดึงข้อมูลอัตโนมัติมาแก้ไข นอกจากการปัญหาจากการค้นหาที่ในบางครั้งพบว่าข้อความที่ค้นหามีช่องว่างเพียงเล็กน้อยทำให้ไม่พบความสัมพันธ์ระหว่างคำที่ค้นหา หรือสร้างความสัมพันธ์ไม่ถูกต้อง นอกจากนี้ถ้าหากทางกรมศิลปากรปรับปรุงข้อมูล ข้อมูลที่ปรับปรุงจะไม่สามารถปรับปรุงตามได้ จึงได้มีการพัฒนาโปรแกรมให้สามารถปรับปรุงข้อมูล



รูปที่ 34 แสดงข้อมูลรูปแบบโครงสร้าง web site ในระบบใหม่ได้เพิ่มฐานข้อมูลงานอนุรักษ์

ได้เพิ่มฐานข้อมูลงานอนุรักษ์เข้าไปในฐานข้อมูล MONGODB โดยได้เชื่อมข้อมูลระหว่างโบราณสถานกับฐานข้อมูลอนุรักษ์เข้าด้วยกัน โดยความสัมพันธ์ของฐานข้อมูลทั้ง 2 มีได้มีการเชื่อมต่อ POS_ID ของฐานข้อมูลโบราณสถานกับ POS_ID ของฐานข้อมูลอนุรักษ์ และดึงข้อมูลของฐานข้อมูลอนุรักษ์ต่างๆ ดังนี้

```
{
  "year": "2550",
  "project": "งานอนุรักษ์จิตรกรรมฝาผนังพระอุโบสถวัดสุทัศน์เทพวราราม",
  "id": "1940",
  "operator": "สำนักโบราณคดี",
  "budget": "4000000"
}
```

ได้ดึงข้อมูล year คือ ปีที่ทำการอนุรักษ์

ได้ดึงข้อมูล project คือ ชื่องานที่ทำการอนุรักษ์

ได้ดึงข้อมูล id คือ key ในการอ้างอิงข้อมูลของการอนุรักษ์

ได้ดึงข้อมูล operator คือ หน่วยงานที่ดำเนินการอนุรักษ์

ได้ดึงข้อมูล budget คือ งบประมาณของการอนุรักษ์

```
{  
  "path": "http://bigdata.finearts.go.th/....."  
}
```

นอกจากนี้ได้ดึงรูปภาพจากการอนุรักษ์และนำมาแสดง โดยเก็บข้อมูลในรูปแบบ path การอ้างอิง เพื่อจะ
ทำให้สามารถนำรูปภาพนี้ไปแสดงบนหน้าเว็บได้

โดยข้อมูลต่างๆเหล่านี้จะถูกเพิ่มเข้าไปในฐานข้อมูล MONGODB โดยแทรกเพิ่มเข้าไปในก้อนข้อมูล
โบราณสถานตัวอย่างข้างล่างนี้ ตัวอย่างนี้ข้อมูลได้เพิ่ม ก้อน object conserve และ object picture_conserve
เข้าไปใน object __hissite

```
"__hissite": [  
  {  
    "image": "http://58.181.189.35/DATAFILE/PICTURE/HISSITE/PICTURE/0000005_003.JPG",  
    "year": 15,  
    "REG_HISTORY": "วัดสุทัศน์.....",  
    "type": "โบราณสถาน",  
    "conserve": [  
      {  
        "year": "2550",  
        "project": "งานอนุรักษ์จิตรกรรมฝาผนังพระอุโบสถวัดสุทัศน์เทพวราราม",  
        "id": "1940",  
        "operator": "สำนักโบราณคดี",  
        "budget": "4000000"
```

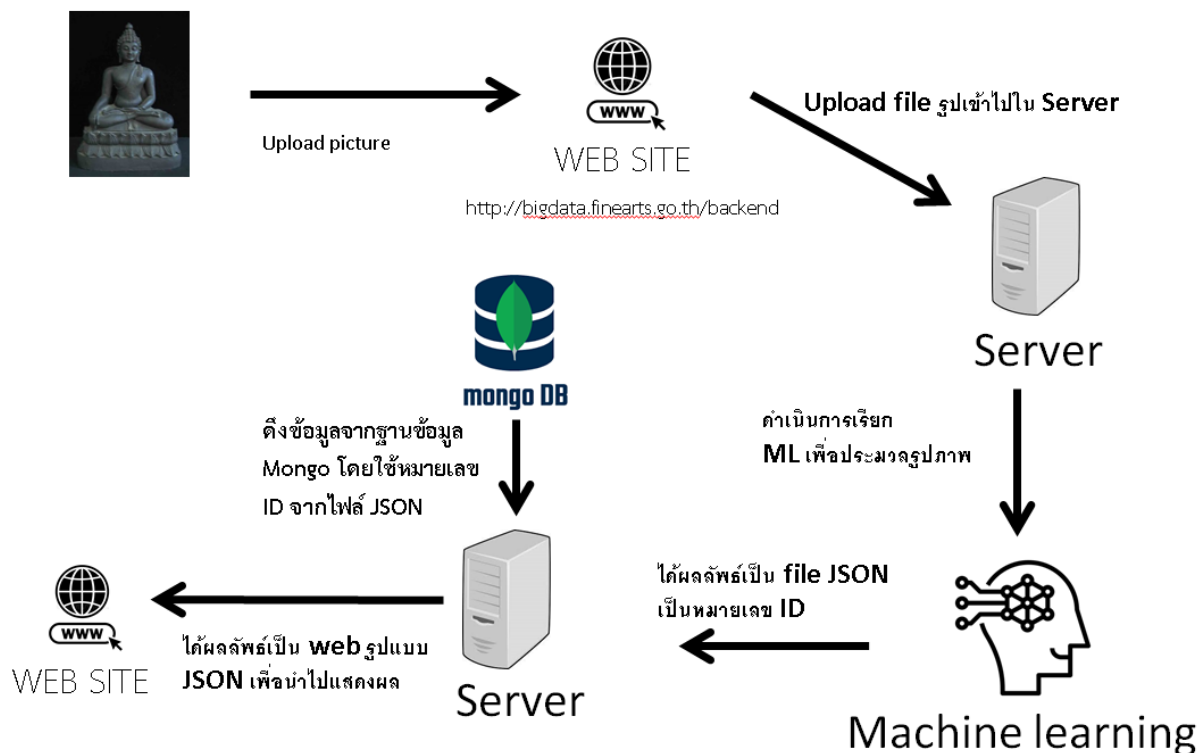
```

    }
  ],
  "province": "",
  "material": "-",
  "picture_conserve": [
    {
      "path":
"http://bigdata.finearts.go.th/backend/finheart_conserve/\\selpicture\\กรุงเทพมหานคร\\0000005\\
00000050000_1_1.jpg"
    },.....
  ],
  "name": "วัดสุทัศน์เทพวราราม",
  "id": "0000005",
  "tag": "โบราณสถาน",
  "age": "สมัยรัตนโกสินทร์",
  "register": "Y"
},
{
  "total": 1
}
],

```

4.2. การเพิ่มการค้นหารูปภาพโบราณสถาน

การค้นหารูปภาพในระบบเดิมจะค้นหาได้เฉพาะภาพเป็นโบราณวัตถุเท่านั้น ไม่สามารถค้นหาโบราณสถานได้ โดยรูปแบบการทำงานของระบบเดิม



ภาพที่ 3 แสดงการค้นหารูปภาพระบบเดิม

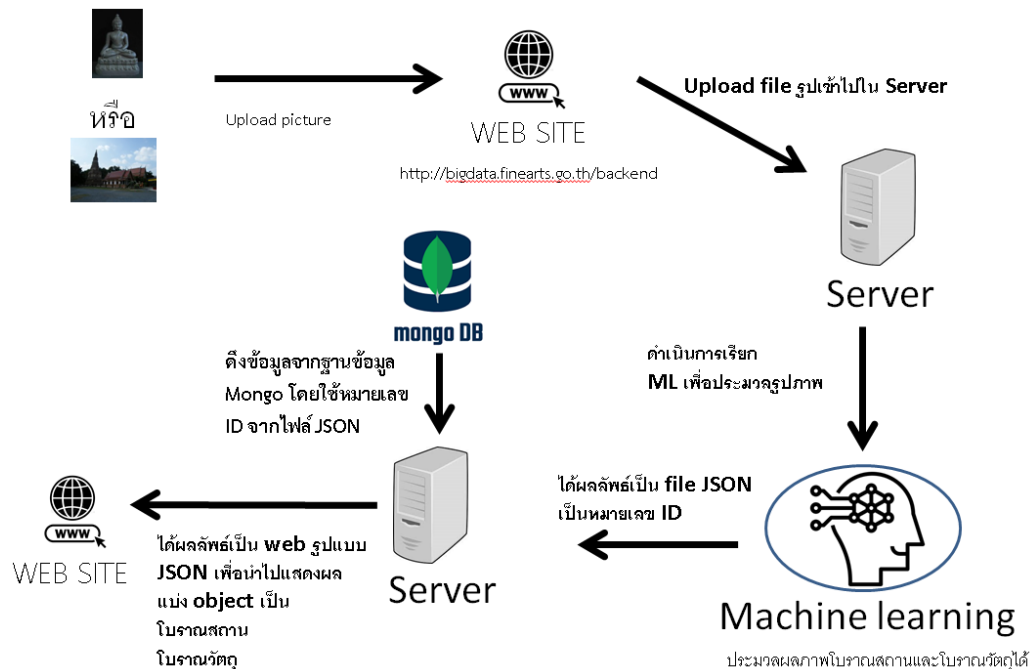
ขั้นตอนการดำเนินการมีรายละเอียดดังนี้

1. จากระบบเดิมเริ่มต้นจากการ upload รูปภาพที่ต้องการค้นหา เข้าไปที่ <http://bigdata.finearts.go.th>
2. จะนำรูปที่ upload เข้าไปเก็บใน server
3. ดำเนินการ run โปรแกรม machine learning บน server ให้ประมวลรูปภาพที่ upload.
4. ผลจากการประมวลผลจะได้ไฟล์ JSON ซึ่งภายในจะมี ID ที่ใช้อ้างอิงข้อมูลที่เก็บใน MONGO
5. ดึงข้อมูลจากฐานข้อมูล MONGO โดยดึงเฉพาะโบราณวัตถุเท่านั้น

6. แสดงผลออกมาในรูปแบบ JSON โดยแสดงรายละเอียดทั้งหมดของโบราณวัตถุนั้น โดยรูปแบบ JSON ที่แสดงมีรูปแบบดังนี้

```
{  
    "image":  
    "https://antique.finearts.go.th/storage/antique/2015/6/55726339e0a3f.jpg",  
    "material": "โลหะ",  
    "culture": "",  
    "name": "พระพุทธรูปปางมารวิชัย",  
    "property": "พระพุทธรูปปางมารวิชัย.....",  
    "id": "268387",  
    "history": "ได้จากการขุดค้น....",  
    "tag": "โบราณวัตถุ",  
    "type": "โบราณวัตถุ"  
}
```

สำหรับในส่วนนี้ทางผู้พัฒนาได้เพิ่มการค้นหาโบราณสถานเข้าไปโดยรูปแบบการทำงานของระบบใหม่มีดังนี้



รูปที่ 35 แสดงการค้นหารูปภาพพระบัพใหม่

ขั้นตอนการดำเนินการมีรายละเอียดดังนี้

1. จากระบบเดิมเริ่มต้นจากการ upload รูปภาพที่ต้องการค้นหา ซึ่งเป็นโบราณสถานหรือโบราณวัตถุก็ได้ เข้าไปที่ <http://bigdata.finearts.go.th>
2. จะนำรูปที่ upload เข้าไปเก็บใน server
3. ดำเนินการ run โปรแกรม machine learning บน server ให้ประมวลรูปภาพที่ upload โดยสามารถประมวลผลภาพโบราณสถานและโบราณวัตถุพร้อมกันได้.
4. ผลจากการประมวลผลจะได้ไฟล์ JSON ซึ่งภายในจะมี ID ที่ใช้อ้างอิงข้อมูลที่เก็บใน MONGO ซึ่ง ID จะมีทั้งโบราณสถานและโบราณวัตถุ
5. ดึงข้อมูลจากฐานข้อมูล MONGO โดยดึงโบราณสถานและโบราณวัตถุ
6. แสดงผลออกมาในรูปแบบ JSON โดยแสดงรายละเอียดทั้งหมดของโบราณสถานและโบราณวัตถุนั้น โดยรูปแบบ json ที่แสดงมีรูปแบบดังนี้ โดยมีโบราณสถานและโบราณวัตถุ โดยแบ่งเป็น 2 object

```

{
  "hissite": [
    {
      "image": "-",
      "province": "",
      "material": "-",
      "year": "-",
      "name": "วัดพระแก้ว",
      "id": "0009037",
      "tag": "โบราณสถาน",
      "REG_HISTORY": "วัดพระแก้ว เป็นวัดที่อยู่ภายในเมืองโบราณ ซึ่งมีคูน้ำคันดินล้อมรอบ เดิมมีประวัติเล่าสืบต่อกันมาเคยเป็นที่ประดิษฐานของพระแก้วมรกต เจดีย์เดิมเป็นเจดีย์ร้างเมื่อประมาณห้าสิบล้านปีมาแล้วพังทลายลง ครุฑทองคำและคณะได้ดำเนินการปฏิสังขรณ์โดยสร้างเจดีย์ใหม่ครอบองค์เดิมไว้ \ววิหารเป็นวิหารทรงล้านนา ไม่ทราบประวัติการสร้างแน่ชัด เดิมมุงด้วยแป้นเกล็ด แต่เปลี่ยนมาเป็นหลังคามุงกระเบื้องเมื่อประมาณสามสิบล้านปีมาแล้ว",
      "type": "โบราณสถาน",
      "age": "-",
      "register": "Y"
    }
  ],
  "antique": [
    {
      "image": "https://antique.finearts.go.th/storage/antique/2015/6/55726339e0a3f.jpg",
      "material": "โลหะ",
      "culture": "",
      "name": "พระพุทธรูปปางมารวิชัย",
    }
  ]
}

```

```

    "property": "พระพุทธรูปปางมารวิชัย ประทับนั่งขัดสมาธิราบบนฐานบัวหงายซ้อนบนฐานเขียงหกเหลี่ยม  

    รัศมีเปลวเพลิง เม็ดพระศกเล็ก พระพักตร์ค่อนข้างเหลี่ยม มีเส้นรอบพระศอสามเส้น พระกรรมมีขนาดใหญ่  

    ด้านบนปลายแหลม ครองจีวรห่มเฉียง ชายสังฆาฏิยาวจรดพระนาภี ปลายตัดตรง ปลายพระหัตถ์ขวายาวจรดเกสร  

    บัว",
    "id": "268387",
    "history": "ได้จากการขุดค้น-ขุดแต่งทางโบราณคดี วัดพระธาตุแสงจันทร์(ร้าง) ตำบลสุเทพ อำเภอเมือง  

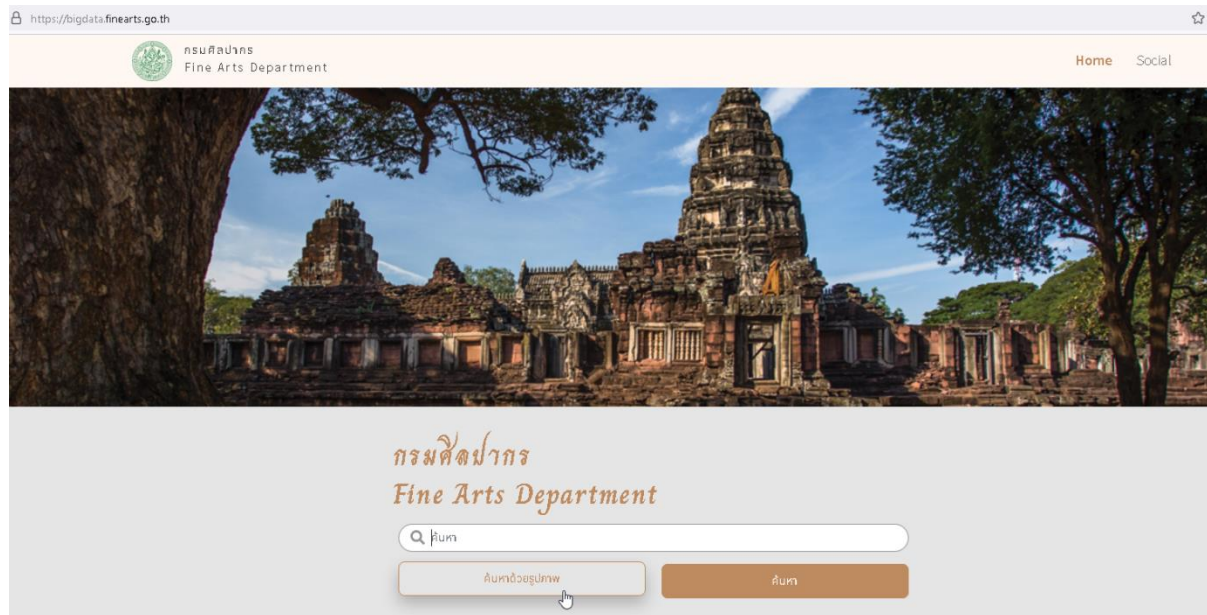
    จังหวัดเชียงใหม่ สำนักศิลปากรที่ 8 เชียงใหม่ ส่งมอบให้พิพิธภัณฑสถานแห่งชาติ เชียงใหม่ เมื่อวันที่ 22 กันยายน  

    2553",
    "tag": "โบราณวัตถุ",
    "type": "โบราณวัตถุ"
  }
]
}

```

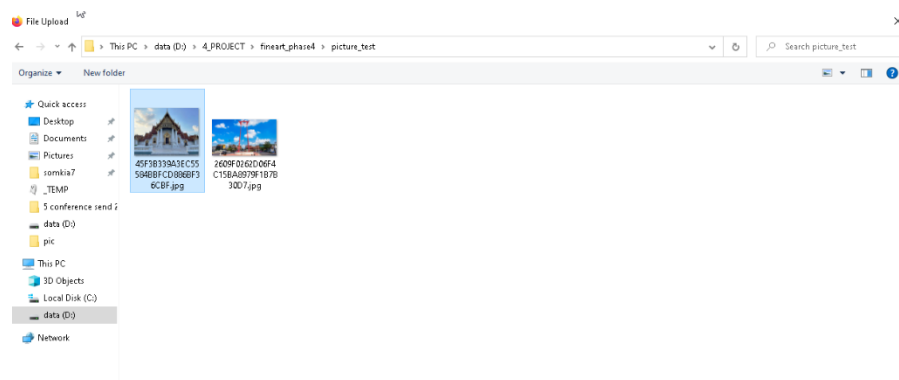
จากส่วนนี้ทำให้การค้นหาได้รับผลลัพธ์ออกมาเป็น 2 ส่วน คือส่วนของโบราณวัตถุและส่วนของ
โบราณสถานดังตัวอย่างต่อไปนี้

แสดงตัวอย่างการค้นหาด้วยภาพ โดยกดปุ่มค้นหาภาพ

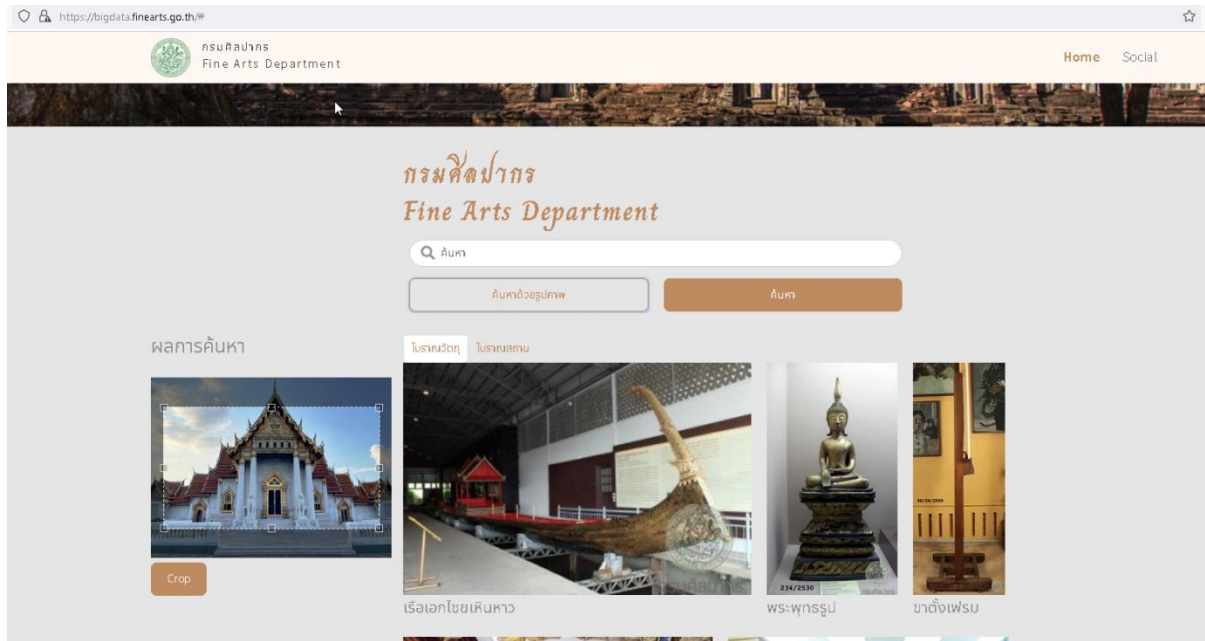


รูปที่ 36 เริ่มต้นค้นหารูปภาพ

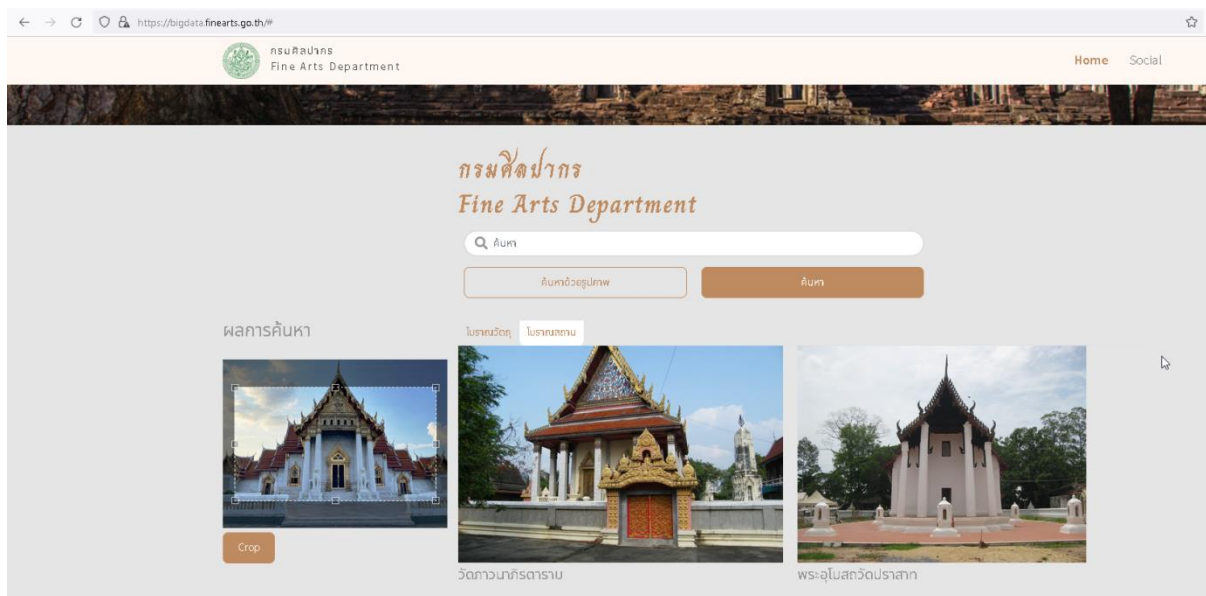
เลือกภาพที่ต้องการค้นหา และรอให้เว็บไซต์ดำเนินการประมวลผลภาพให้เสร็จสมบูรณ์



รูปที่ 37 เลือกรูปภาพ



รูปที่ 38 แสดงผลการค้นหารูปภาพโบราณวัตถุ



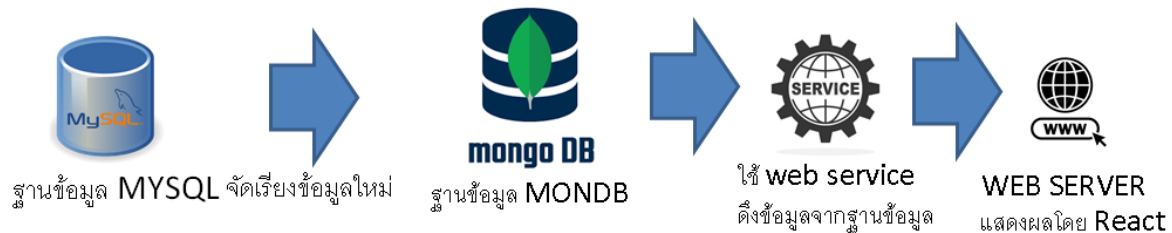
รูปที่ 39 แสดงผลการค้นหารูปภาพโบราณสถาน

รูปที่ 38 แสดงผลลัพธ์การค้นหาจะได้ภาพที่ใกล้เคียงกับภาพที่ค้นหาคือโบราณสถานและโบราณวัตถุ โดยแสดงเป็น 2 tab โดย tab แรกคือ โบราณวัตถุ ดังรูปที่ 38 และรูปที่ 39 แสดงผลลัพธ์การค้นหาโบราณสถาน ใน

ระบบเดิมจะแสดงผลเพียงโบราณวัตถุเท่านั้น แต่ในระบบใหม่ได้เพิ่มโบราณสถาน และยังคงค้นหาโบราณวัตถุได้ด้วยเช่นกัน

4.3. การปรับปรุงฐานข้อมูลการค้นหาทั้งหมด

รูปแบบเดิม



รายชื่อฐานข้อมูลใหม่

- 1) โบราณวัตถุ
- 2) โบราณสถาน
- 3) การนำเข้าส่งออก
- 4) จดหมายเหตุ
- 5) เอกสารโบราณ
- 6) พระเมรุ
- 7) สิบหมื่น

รูปที่ 40 แสดงข้อมูลรูปแบบโครงสร้าง web site ในระบบเดิม

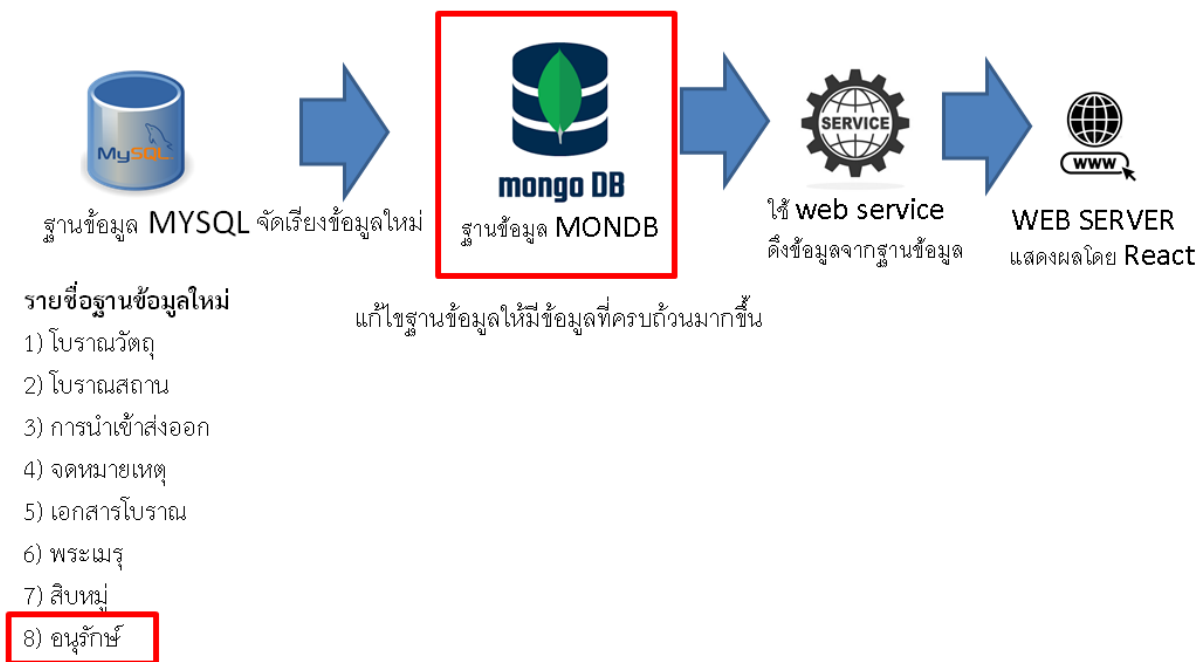
จากรูปที่ 40 จากระบบเดิมการดึงข้อมูลต้องดึงจาก ฐานข้อมูล MySQL แล้ว WEB SERVER ที่ถูกพัฒนาด้วย REACT จะทำการดึงข้อมูลจากฐานข้อมูล MONGODB โดยทำหน้าที่ดึงข้อมูลจาก WEB SERVICE เนื่องจาก MONGODB เป็นฐานข้อมูล ในระบบเก่ามีปัญหาการแสดงผลข้อมูลที่มีจะไม่ถูกต้อง เพราะข้อมูลมีความสัมพันธ์ไม่ครบถ้วน สำหรับระบบเดิมมี ฐานข้อมูลทั้งหมดดังนี้

- 1) โบราณวัตถุ
- 2) โบราณสถาน
- 3) การนำเข้าส่งออก
- 4) โบราณวัตถุ
- 5) จดหมายเหตุ

6) บุรณะ

7) สิบหมู่

รูปแบบใหม่

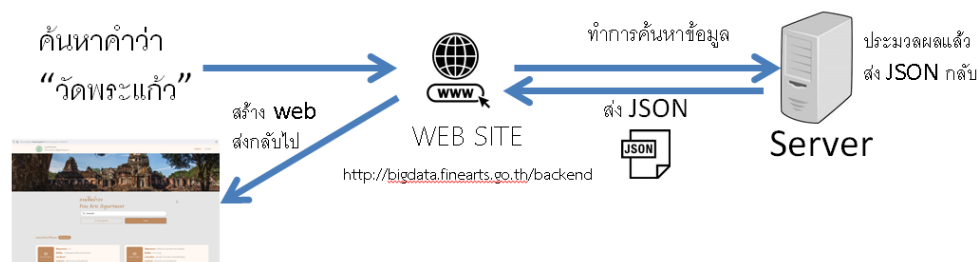


รูปที่ 41 แสดงข้อมูลรูปแบบโครงสร้าง web site ในระบบใหม่

จากรูปที่ 41 ระบบ web site ได้เพิ่มส่วนฐานข้อมูลดัชนี คือ MONGODB ช่วยในการเก็บข้อมูล เป็นข้อมูลที่มีการจัดเรียงไว้เรียบร้อยแล้ว โดย WEB SERVICE สามารถดึงข้อมูลไปแสดงผลได้โดยตรง ทำให้การค้นหาข้อมูลรวดเร็ว ทำให้ web site มีความเร็วมากขึ้น แต่ข้อมูลยังแสดงผลความสัมพันธ์ไม่ครบถ้วน จึงได้ทำการสร้างฐานข้อมูลใหม่ให้มีความสัมพันธ์ที่ดียิ่งขึ้น นอกจากนี้ ได้เพิ่มฐานข้อมูล อนุรักษ์ เข้าไปในระบบฐานข้อมูลใหม่ ดังที่ได้กล่าวไปก่อนหน้านี้แล้ว โดยตารางใน MONGODB ได้แบ่งการทำงานทั้งหมด ออกเป็นดังนี้

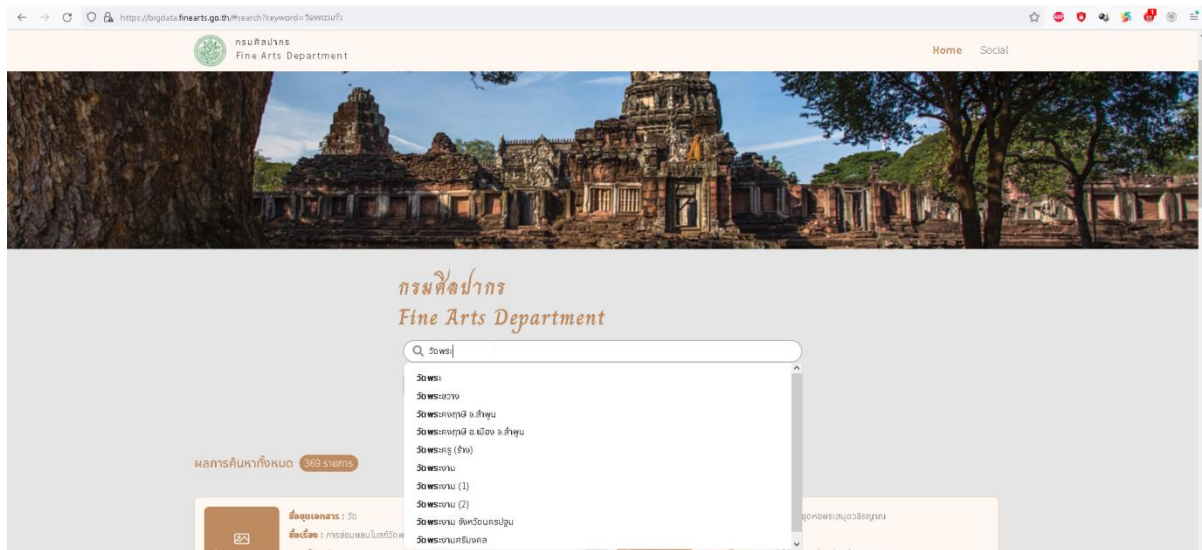
1) ตารางข้อมูล keyword ทำหน้าที่เก็บข้อมูลคำค้นที่ใช้ในการค้นหา จำนวนคำทั้งหมดที่ใช้ในตารางข้อมูลนี้ คือ 307894 คำ ซึ่งผู้ใช้งานจะดำเนินการค้นหาคำต่างๆ เช่น ต้องการค้นหาคำว่า วัดพระแก้ว เมื่อพิมพ์คำว่าวัดพระแก้วให้บน web site หลังจากนั้น web site จะดำเนินการดึงข้อมูลจากตารางข้อมูลนี้ ผลลัพธ์จะได้ก่อนข้อมูลรูปแบบไฟล์ JSON ซึ่งสามารถนำไปแสดงบน web site ดังภาพที่ 10 ในส่วนนี้เนื่องจาก

ดำเนินการแก้ไขข้อมูลตาราง KEYWORD บนตาราง keyword ในฐานข้อมูล MYSQL เพราะได้เพิ่มข้อมูลส่วนของอนุรักษ์ และได้แก้ไขคำค้นทั้งหมด จึงทำให้ ตารางข้อมูล keyword บน MONGODB ต้องถูกทำให้ใหม่ทั้งหมด

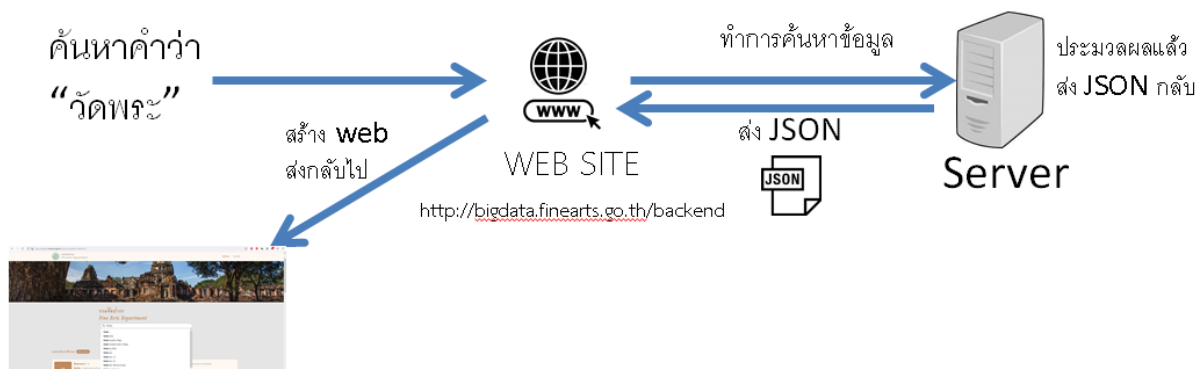


รูปที่ 42 แสดงข้อมูลรูปแบบโครงสร้าง web site ในระบบใหม่

2) ตารางข้อมูล keyword_search ทำหน้าที่เก็บข้อมูลคำที่ใช้แนะนำคำค้นหาคือผู้ใช้แค่พิมพ์คำค้นหาที่ยังไม่สมบูรณ์ คำค้นหาเหล่านั้นจะได้ดึงข้อมูลที่ตารางข้อมูล keyword_search แล้วนำคำที่ใกล้เคียงมาแสดงผลจำนวน 10 คำ โดยผู้ใช้ได้พิมพ์คำว่า “วัดพระ” ผลการค้นหาแสดงคำว่า “วัดพระครู” “วัดพระงาม” เป็นต้น ซึ่งเป็นคำใกล้เคียงคำที่ค้นหา เพื่อเป็นการแนะนำการค้นหาหรือเป็นคำให้เลือกดังรูปภาพที่ 11 และสำหรับภาพที่ 12 แสดงการทำงานของการทำงานการดึงข้อมูลบนตารางข้อมูล keyword_search คือเมื่อผู้ใช้พิมพ์คำบนช่องค้นหาเกิน 3 คำ web site จะดำเนินการดึงข้อมูลจากตารางข้อมูล keyword_search มาจำนวนหนึ่งเพื่อนำมาแสดงให้ผู้ใช้ชมเพื่อที่จะตัดสินใจเลือกใช้คำค้นหาที่ต้องการ โดยการดึงข้อมูลจะส่งคำสั่งไปดึงข้อมูลบนตาราง keyword_search เมื่อดึงได้ 10 ตัวตามต้องการแล้วจะดำเนินการแสดงผลออกมา ถ้ามีการพิมพ์คำเพิ่มเข้าไปอีกเช่น คำว่า วัดพระแกก็จะดึงข้อมูลจากฐานข้อมูลใหม่ ตามการเปลี่ยนแปลงของคำ โดยจำนวนคำที่เก็บในตารางข้อมูล keyword_search คือ 95,570 คำ



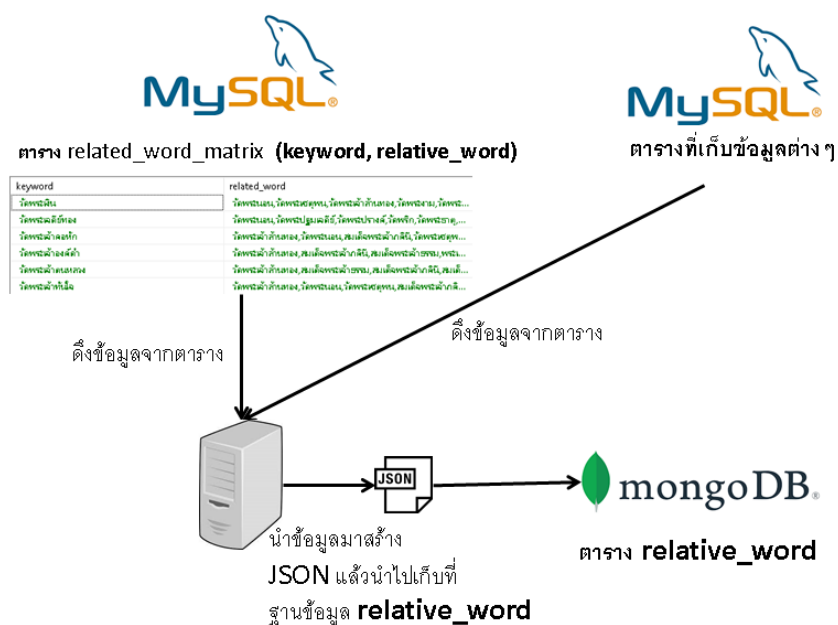
รูปที่ 43 แสดงผลลัพธ์จากการค้นหาตารางข้อมูล keyword_search



รูปที่ 44 แสดงขั้นตอนการทำงานของการทำงานการค้นหา keyword_search

3) ตารางข้อมูล relative_word ทำหน้าที่เก็บข้อมูลเพื่อนำไปคำค้นที่ใกล้เคียง เพื่อสำหรับสร้างตารางค้นหาคำค้นที่ใกล้เคียง เช่น ต้องการหาพระแก้วมรกต จะแสดงข้อมูลที่เกี่ยวข้องกับพระแก้วมรกต เช่น วัดพระแก้ว เจริญพระแก้วมรกต ตำนานพระแก้วมรกต เป็นต้น สำหรับการสร้างตารางนี้ เป็นดังภาพที่ 13 โดยเริ่มต้นจากการดึงข้อมูลจากตาราง related_word_matrix ที่เป็นตารางบนฐานข้อมูลรูปแบบ mysql โดยภายในตาราง related_word_matrix จะมีข้อมูลของ keyword หรือคำค้น และ related_word หรือคำค้นที่ใกล้เคียง ดังนั้นจึงนำ keyword ที่ได้ และ related_word ที่ได้ ไป สร้างไฟล์ json ซึ่งการสร้างไฟล์ json ต้องดึงข้อมูลต่างๆ ที่เป็นรายละเอียดเกี่ยวกับคำค้นนั้นจากตารางบนฐานข้อมูล mysql ได้แก่ ฐานข้อมูล โบราณวัตถุ โบราณสถาน

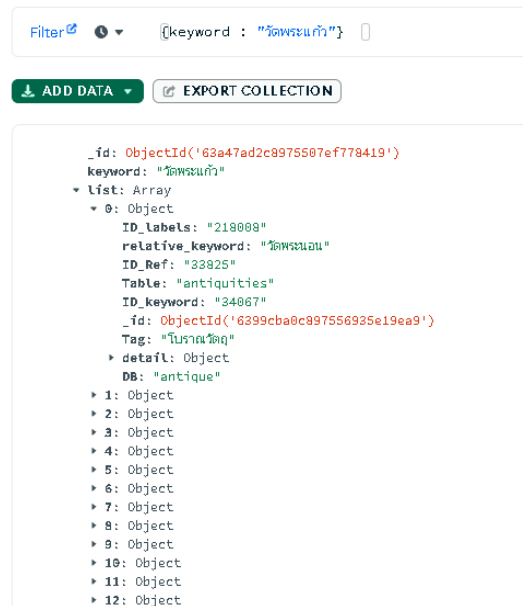
การนำเข้าส่งออก จดหมายเหตุ เอกสารโบราณ พระเมรุ สิบหมู่ อนุรักษสำหรับไฟล์ JSON นั้นจะมีเพียงคำที่เป็นที่ เกี่ยวข้อง และภายในคำที่เกี่ยวข้องจะมีข้อมูลเหมือนกับตาราง keyword แต่ข้อมูลบางอย่างอาจจะไม่ได้แสดงผล ออกมาทั้งหมด เป็นข้อมูลที่น่าไว้แสดงบนหน้า web เท่านั้น จึงมีข้อมูลเพียงบางส่วนเท่านั้น ไม่ได้มีข้อมูล เหมือนกับตาราง keyword สำหรับตารางนี้ ยังไม่สามารถแสดงผล ได้เพราะเป็นข้อมูลเล็กๆ แต่ได้มีคำที่เกี่ยวข้อง ติดอยู่ในข้อมูลเหล่านั้น จึงต้องมีขั้นตอนต่อไปในการนำข้อมูลเหล่านี้มาประกอบกันเป็นข้อมูลขนาดใหญ่ ที่ไว้ สำหรับแสดงผลในขั้นตอนต่อไป โดยจำนวนคำที่เก็บในตารางข้อมูล relative_word คือ 85,557,441 คำ



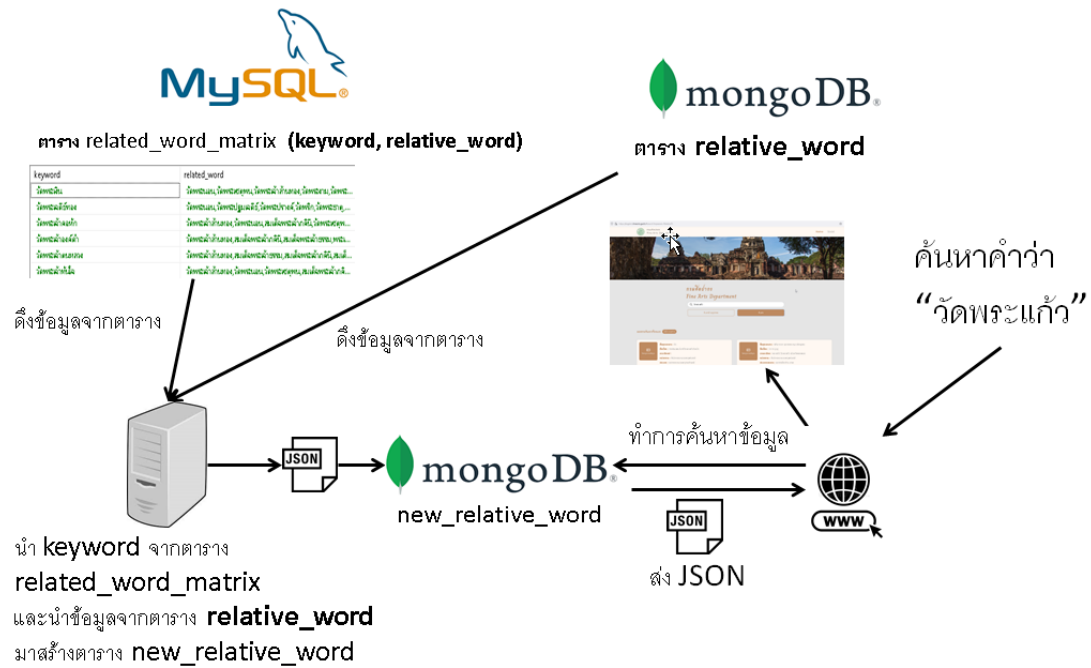
รูปที่ 45 แสดงขั้นตอนการสร้างตาราง related_word

4) ตารางข้อมูล new_relative_word ทำหน้าที่เก็บข้อมูลค้นหาคำค้นที่ใกล้เคียง เช่น ต้องการหาพระ แก้วมรกต จะแสดงข้อมูลที่เกี่ยวข้องกับพระแก้วมรกต เช่น วัดพระแก้ว เจริญพระแก้วมรกต ดำนานพระแก้ว มรกต เป็นต้น สำหรับการสร้างตารางนี้ เริ่มต้นจากการดึงข้อมูลจากตาราง relative_word นำข้อมูลมาประกอบ กันเป็น Array หรือข้อมูลขนาดใหญ่ ดังภาพที่ 14 มีจำนวนไม่เกิน 500 ชุด โดยแต่ละชุด จะประกอบได้ด้วยข้อมูล จาก relative_word สำหรับหนึ่งชุด คือ ข้อมูลจากตาราง relative_word หนึ่งก้อน ดังตัวอย่างถ้าค้นหาคำว่า “วัดพระแก้ว” จะมีความสัมพันธ์กับวัดพระนอน โดยรายละเอียดของวัดพระนอนจะถูกเก็บไว้ในก้อนข้อมูลชุดนี้ ขั้นตอนการสร้างตารางข้อมูล new_relative_word จะเริ่มต้นจากการดึงคำต่างๆจากตาราง

related_word_matrix ที่เป็นตารางบนฐานข้อมูลรูปแบบ mysql นำค่าเหล่านี้มาดึงก้อน JSON ในตาราง relative_word มาใส่ใน Array ที่ละตัวโดยไม่เกิน 500 ตัว นำมาสร้างเป็นก้อนข้อมูล 1 ก้อน โดยก้อนข้อมูลเหล่านี้ เมื่อดำเนินการค้นหาจะดึงก้อนข้อมูลเหล่านี้ไป เพื่อนำไปแสดงผลดังภาพที่ 15 โดยจำนวนคำที่เก็บในตารางข้อมูล relative_word คือ 58,714 คำ



รูปที่ 46 ตัวอย่างข้อมูลของตารางข้อมูล new_relative_word

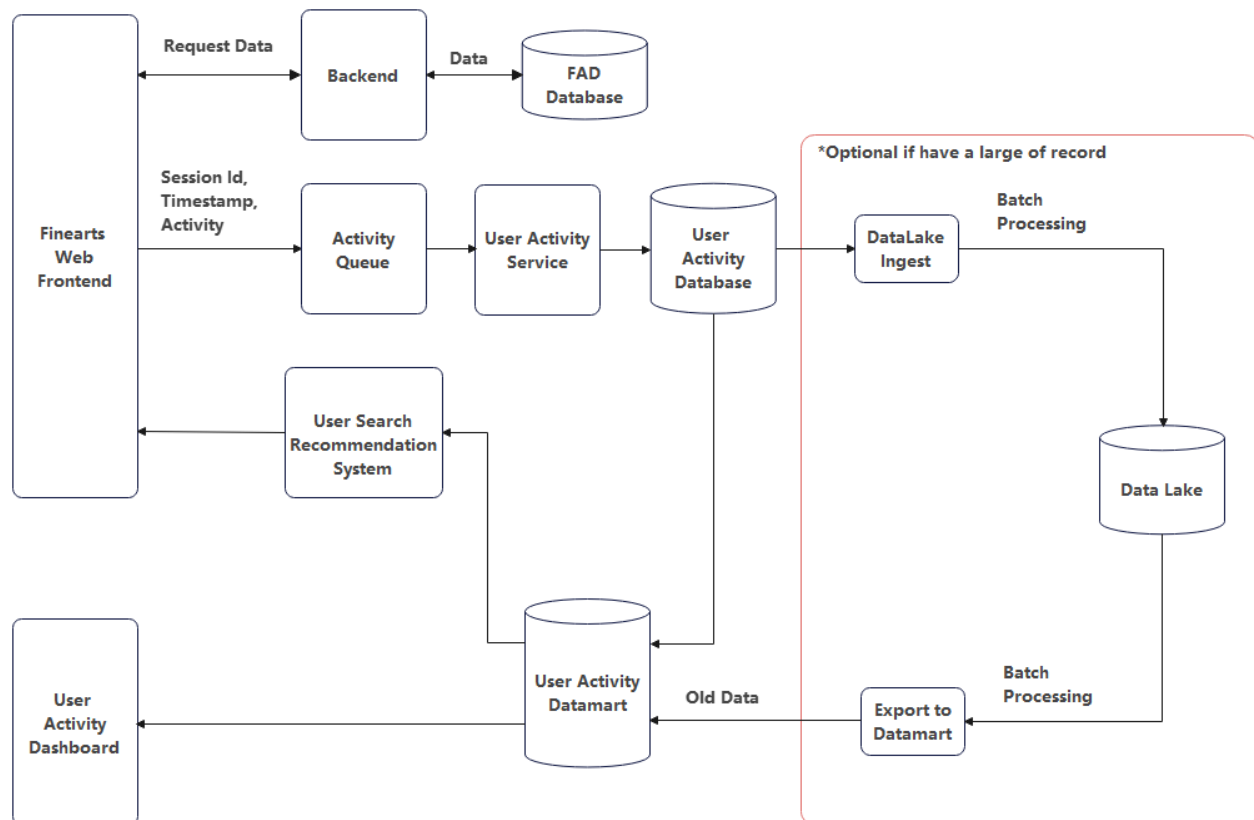


รูปที่ 47 แสดงขั้นตอนการสร้างตาราง new_relative_word

5. ระบบวิเคราะห์พฤติกรรมผู้ใช้งาน

5.1. การออกแบบระบบวิเคราะห์พฤติกรรมผู้ใช้งาน

จากความต้องการที่ต้องการทำการเก็บสถิติการเข้าใช้งานของผู้ใช้งานเพื่อนำมาวิเคราะห์เพื่อใช้ประโยชน์ในการเพิ่มประสิทธิภาพที่ดีของผู้ใช้งานจึงจำเป็นต้องมีการพัฒนาระบบเพิ่มเติมจากระบบเดิมของโครงการ เพื่อให้รองรับความต้องการดังกล่าวผู้พัฒนาจึงได้ออกแบบสถาปัตยกรรมของระบบเพื่อให้รองรับการเก็บข้อมูลที่มีขนาดใหญ่และเป็นข้อมูลที่มีความเคลื่อนไหวตลอดเวลา ดังนี้



รูปที่ 48 สถาปัตยกรรมของระบบวิเคราะห์พฤติกรรมผู้ใช้งาน

จากรูปที่ 48 จะเห็นว่าระบบส่วนที่มีการออกแบบเพิ่มเติมคือในส่วนของ User Activity Service โดยระบบดังกล่าวจะทำงานร่วมกับเว็บแอปพลิเคชันเดิมของโครงการเพื่อให้ทำการรวบรวมข้อมูลต่าง ๆ และนำกลับมาวิเคราะห์ต่อไปในอนาคตโดยส่วนประกอบหลักของระบบจะประกอบไปด้วย Message Queue เพื่อให้การรับข้อมูลในลักษณะที่เป็น Transaction ไม่กระทบกับการทำงานเดิมข้อมูลต่าง ๆ จำเป็นจะต้องทำงานในลักษณะ

ที่ไม่มีการรอดูคำสั่งเหล่านี้การทำงานจึงต้องทำงานในลักษณะของ Asynchronous ระบบจึงต้องมี Message Queue ในการรับก่อนจะทำการประมวลผลต่อไป ส่วนต่อมาคือการบันทึกข้อมูลที่ถูกส่งมาจะถูกจัดเก็บในฐานข้อมูลเพื่อพักไว้รอจังหวะในการสร้างชุดข้อมูล Data Mart ที่พร้อมใช้งานต่อไป

ในกรณีที่ข้อมูลมีขนาดใหญ่ขึ้นซึ่งอาจจะเกิดขึ้นได้ในอนาคตทางผู้พัฒนาได้ทำการออกแบบเพิ่มเติมคือการส่งข้อมูลไปเก็บในลักษณะของ Unstructured Data บน Data Lake ไว้เพิ่มเติมก่อนที่จะข้อมูลจะถูกนำมารวมกับฐานพักข้อมูลเพื่อสร้างชุดข้อมูลที่จำเป็นในการใช้งานต่อไป แต่เนื่องจากการนำข้อมูลเข้าถึง Data Lake มีส่วนที่จะเพิ่มเวลาในการทำงานจากการทำงานปกติซึ่งจะไม่คุ้มค่าในการใช้งานถ้าหากข้อมูลยังไม่มากพอ ทางผู้พัฒนาจึงคิดว่าในส่วนที่เป็นกรอบสีแดงจะยังไม่ถูกนำมาใช้งานในโครงการปัจจุบัน

5.2. การพัฒนาระบบวิเคราะห์พฤติกรรมผู้ใช้งาน

จากการออกแบบสถาปัตยกรรมในหัวข้อ 5.1 ทางผู้พัฒนาได้ทำการพัฒนาในส่วนของการเก็บข้อมูลพฤติกรรมการใช้งานของผู้ใช้งานเป็นที่เรียบร้อยแล้วโดยมีรายละเอียดการพัฒนาต่อไปนี้

ทีมพัฒนาได้เลือกใช้ Rabbit MQ ที่เป็นซอฟต์แวร์ Open Source สำหรับการจัดการ Message ที่ถูกส่งเข้ามาในระบบโดยทางทีมพัฒนาได้ทำการเพิ่ม `fineart_analytic_events.UpdateUserTracking` และ `fineart_analytic_events.UpdateUserTrackingStatus` สำหรับเก็บข้อมูลที่ถูกส่งมาจากเว็บแอปพลิเคชัน

Exchanges

► All exchanges (9)

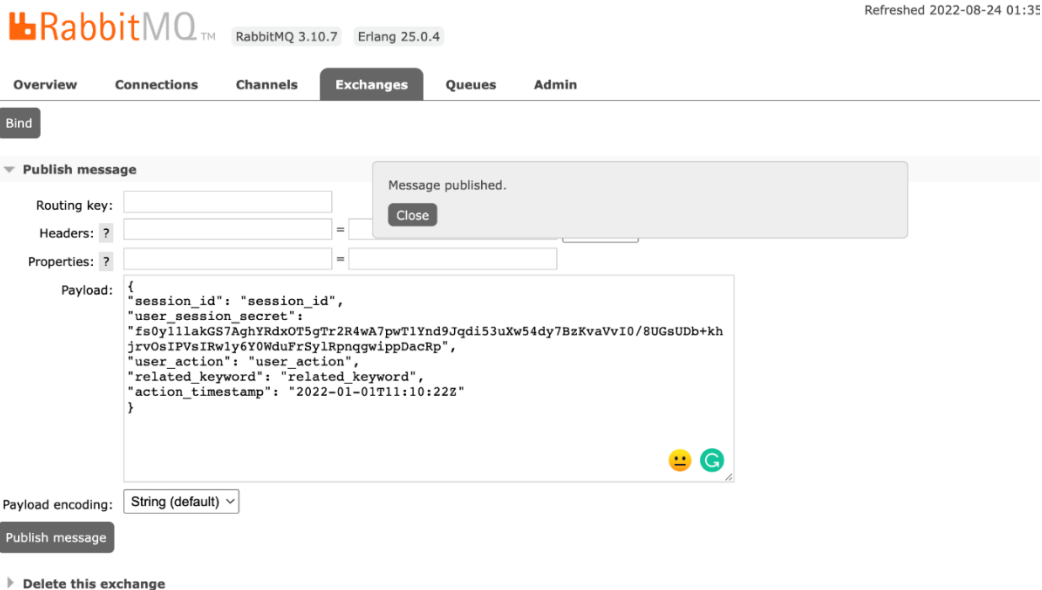
Name	Type	Features	Message rate in	Message rate out	+/-
(AMQP default)	direct	D			
amq.direct	direct	D			
amq.fanout	fanout	D			
amq.headers	headers	D			
amq.match	headers	D			
amq.rabbitmq.trace	topic	D I			
amq.topic	topic	D			
fineart_analytic_events.UpdateUserTracking	fanout	D	0.00/s	0.00/s	
fineart_analytic_events.UpdateUserTrackingStatus	fanout	D	0.00/s		

► Add a new exchange

รูปที่ 49 การเข้าคิวของ Message ในการวิเคราะห์พฤติกรรมของผู้ใช้งาน

โดย Message ที่ถูกส่งเข้ามาจะถูกส่งมาโดยมีข้อมูลต่อไปนี้

1. session_id สำหรับการจัดเก็บ session ของผู้ใช้งานทำให้ระบบสามารถติดตามผู้ใช้งานได้
2. user_session_secret สำหรับการป้องกันการส่งข้อมูลปลอมโดย Secret จาก Client ต้องตรงกับ Backend ถึงจะถูกรับว่าเป็นข้อมูลจริง
3. user_action สำหรับการจัดเก็บการใช้งานของผู้ใช้งาน
4. related_keyword สำหรับการจัดเก็บ Keyword ในการใช้งาน
5. action_timestamp สำหรับจัดเก็บเวลาของการกระทำ
6. parameter สำหรับการจัดเก็บพารามิเตอร์เพิ่มเติมสำหรับการทำ Action ของผู้ใช้งาน



รูปที่ 50 การเก็บข้อมูลภายใน Message

หลังจากนั้นระบบจะเริ่มทำการถอดข้อมูลจาก Message ที่เข้ามาใน Message Queue เพื่อทำการบันทึกข้อมูลลงฐานข้อมูล Buffer โดยการเก็บในฐานข้อมูล RDBMS (MySQL) และเพิ่มในส่วนของ Id และ Created_at และ Updated_at ลงในตารางเพิ่มเติมจาก Message ที่ได้รับเพื่อให้ง่ายต่อการใช้งานตามการออกแบบต่อไป

noc session_id	noc user_session_id	noc user_action	noc related_keyword	action_timestamp	created_at	updated_at
ea0a6a62-7caf-4e69-b877-ef2c34b86210	FA_US-9c6107ab-0d9a-4a5c-9627-dac815f46d02	select_tab_ancient	พระที่นั่งจก	2023-01-03 06:04:30	2023-01-03 06:04:32	2023-01-03 06:04:32
ea0a6a62-7caf-4e69-b877-ef2c34b86210	FA_US-9c6107ab-0d9a-4a5c-9627-dac815f46d02	select_tab_antique	พระที่นั่งจก	2023-01-03 06:04:32	2023-01-03 06:04:34	2023-01-03 06:04:34
ea0a6a62-7caf-4e69-b877-ef2c34b86210	FA_US-9c6107ab-0d9a-4a5c-9627-dac815f46d02	select_image_310806	พระที่นั่งจก	2023-01-03 06:04:36	2023-01-03 06:04:38	2023-01-03 06:04:38
ea0a6a62-7caf-4e69-b877-ef2c34b86210	FA_US-9c6107ab-0d9a-4a5c-9627-dac815f46d02	search_image	วัดพระอิน	2023-01-03 06:04:42	2023-01-03 06:04:44	2023-01-03 06:04:44
ea0a6a62-7caf-4e69-b877-ef2c34b86210	FA_US-9c6107ab-0d9a-4a5c-9627-dac815f46d02	search_keyword	พระที่นั่งจก	2023-01-03 06:04:44	2023-01-03 06:04:46	2023-01-03 06:04:46
ea0a6a62-7caf-4e69-b877-ef2c34b86210	FA_US-9c6107ab-0d9a-4a5c-9627-dac815f46d02	select_item_310806	พระที่นั่งจก	2023-01-03 06:04:51	2023-01-03 06:04:53	2023-01-03 06:04:53
ea0a6a62-7caf-4e69-b877-ef2c34b86210	FA_US-9c6107ab-0d9a-4a5c-9627-dac815f46d02	select_tab_antique	พระที่นั่งจก	2023-01-03 06:06:15	2023-01-03 06:06:17	2023-01-03 06:06:17
ea0a6a62-7caf-4e69-b877-ef2c34b86210	FA_US-9c6107ab-0d9a-4a5c-9627-dac815f46d02	select_item_2736	พระที่นั่งจก	2023-01-03 06:06:22	2023-01-03 06:06:24	2023-01-03 06:06:24
ea0a6a62-7caf-4e69-b877-ef2c34b86210	FA_US-9c6107ab-0d9a-4a5c-9627-dac815f46d02	search_keyword	วัดพระอิน	2023-01-03 06:06:24	2023-01-03 06:06:26	2023-01-03 06:06:26
ea0a6a62-7caf-4e69-b877-ef2c34b86210	FA_US-9c6107ab-0d9a-4a5c-9627-dac815f46d02	search_keyword	วัดพระอิน	2023-01-03 06:06:38	2023-01-03 06:06:40	2023-01-03 06:06:40
ea0a6a62-7caf-4e69-b877-ef2c34b86210	FA_US-9c6107ab-0d9a-4a5c-9627-dac815f46d02	search_keyword	วัดพระอิน	2023-01-03 06:06:40	2023-01-03 06:06:42	2023-01-03 06:06:42
ea0a6a62-7caf-4e69-b877-ef2c34b86210	FA_US-9c6107ab-0d9a-4a5c-9627-dac815f46d02	select_item_0003752	วัดพระอิน	2023-01-03 06:06:46	2023-01-03 06:06:48	2023-01-03 06:06:48
ea0a6a62-7caf-4e69-b877-ef2c34b86210	FA_US-9c6107ab-0d9a-4a5c-9627-dac815f46d02	select_tab_ancient	วัดพระอิน	2023-01-03 06:06:48	2023-01-03 06:06:50	2023-01-03 06:06:50
ea0a6a62-7caf-4e69-b877-ef2c34b86210	FA_US-9c6107ab-0d9a-4a5c-9627-dac815f46d02	select_tab_archives	วัดพระอิน	2023-01-03 06:06:49	2023-01-03 06:06:51	2023-01-03 06:06:51
ea0a6a62-7caf-4e69-b877-ef2c34b86210	FA_US-9c6107ab-0d9a-4a5c-9627-dac815f46d02	select_tab_antique	วัดพระอิน	2023-01-03 06:06:50	2023-01-03 06:06:52	2023-01-03 06:06:52
ea0a6a62-7caf-4e69-b877-ef2c34b86210	FA_US-9c6107ab-0d9a-4a5c-9627-dac815f46d02	select_tab_antique	วัดพระอิน	2023-01-03 06:06:53	2023-01-03 06:06:55	2023-01-03 06:06:55
ea0a6a62-7caf-4e69-b877-ef2c34b86210	FA_US-9c6107ab-0d9a-4a5c-9627-dac815f46d02	select_tab_hissite	วัดพระอิน	2023-01-03 06:06:53	2023-01-03 06:06:55	2023-01-03 06:06:55
ea0a6a62-7caf-4e69-b877-ef2c34b86210	FA_US-9c6107ab-0d9a-4a5c-9627-dac815f46d02	search_keyword	พระที่นั่งจก	2023-01-03 06:07:10	2023-01-03 06:07:12	2023-01-03 06:07:12
ea0a6a62-7caf-4e69-b877-ef2c34b86210	FA_US-9c6107ab-0d9a-4a5c-9627-dac815f46d02	search_keyword	พระที่นั่งจก	2023-01-03 06:07:20	2023-01-03 06:07:22	2023-01-03 06:07:22
ea0a6a62-7caf-4e69-b877-ef2c34b86210	FA_US-9c6107ab-0d9a-4a5c-9627-dac815f46d02	select_item_317181	พระที่นั่งจก	2023-01-03 06:07:21	2023-01-03 06:07:23	2023-01-03 06:07:23
ea0a6a62-7caf-4e69-b877-ef2c34b86210	FA_US-9c6107ab-0d9a-4a5c-9627-dac815f46d02	search_keyword	วัดพระอิน	2023-01-03 06:07:52	2023-01-03 06:07:54	2023-01-03 06:07:54
ea0a6a62-7caf-4e69-b877-ef2c34b86210	FA_US-9c6107ab-0d9a-4a5c-9627-dac815f46d02	search_keyword	วัดพระอิน	2023-01-03 06:07:52	2023-01-03 06:07:55	2023-01-03 06:07:55
ea0a6a62-7caf-4e69-b877-ef2c34b86210	FA_US-9c6107ab-0d9a-4a5c-9627-dac815f46d02	search_keyword	วัดพระอิน	2023-01-03 06:08:44	2023-01-03 06:08:46	2023-01-03 06:08:46
ea0a6a62-7caf-4e69-b877-ef2c34b86210	FA_US-9c6107ab-0d9a-4a5c-9627-dac815f46d02	select_item_0005063	วัดพระอิน	2023-01-03 06:08:54	2023-01-03 06:08:56	2023-01-03 06:08:56
ea0a6a62-7caf-4e69-b877-ef2c34b86210	FA_US-9c6107ab-0d9a-4a5c-9627-dac815f46d02	select_tab_hissite	วัดพระอิน	2023-01-03 06:08:55	2023-01-03 06:08:57	2023-01-03 06:08:57

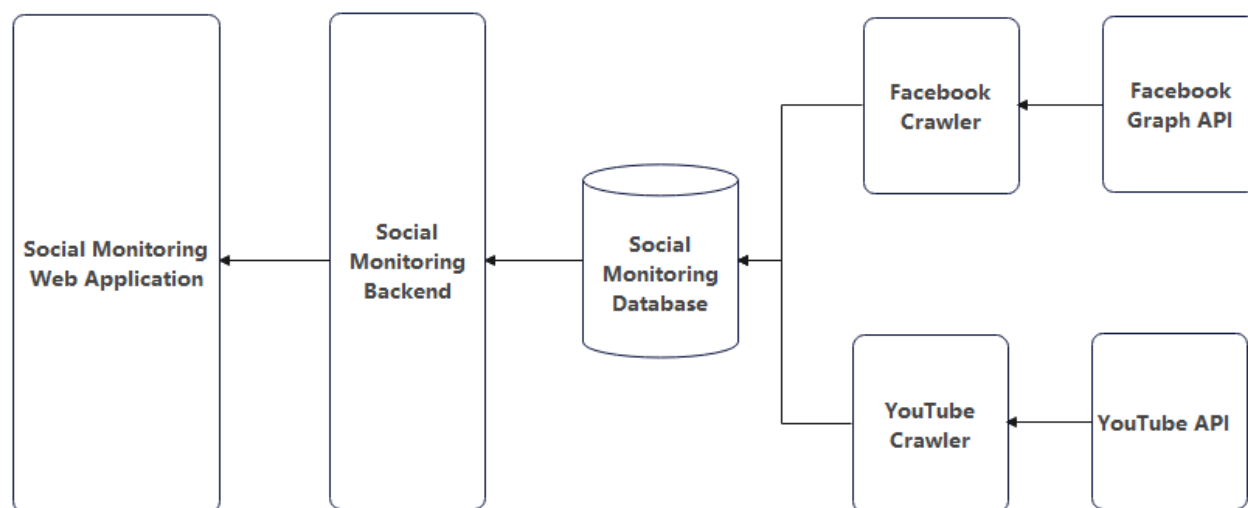
รูปที่ 51 การเก็บข้อมูลพฤติกรรมผู้ใช้งานลงในฐานข้อมูล

6. ระบบรายงานข้อมูลสถิติช่อง YouTube ของกรมศิลปากร

6.1. การออกแบบระบบรายงานข้อมูลสถิติช่อง YouTube ของกรมศิลปากร

6.1.1. การออกแบบสถาปัตยกรรมของระบบ

จากความต้องการของผู้ใช้งานที่ต้องการการรายงานสถิติจากสื่อสังคมออนไลน์ YouTube เพิ่มเติมทางทีมผู้พัฒนาได้ทำการออกแบบสถาปัตยกรรมระบบเพิ่มเติมจากระบบรายงานสถิติสื่อสังคมออนไลน์เดิมจากช่องทางของ Facebook เพื่อให้ง่ายต่อการใช้งานและการจัดการเนื้อหาและการตั้งค่าต่าง ๆ ดังนี้



รูปที่ 52 สถาปัตยกรรมของระบบรายงานสถิติช่อง YouTube

จากรูปที่ 52 ส่วนที่จะทำการพัฒนาเพิ่มเติมจากระบบเดิมคือส่วนของการรวบรวมข้อมูลจาก YouTube และ ส่วนของผู้ดูแลระบบได้แก่

1. YouTube Crawler สำหรับการรวบรวมข้อมูลผ่าน YouTube Data API และ YouTube Analytic API
2. Social Monitoring Backend สำหรับการจัดการเนื้อหาและการตั้งค่าต่าง ๆ ของผู้ดูแลระบบ

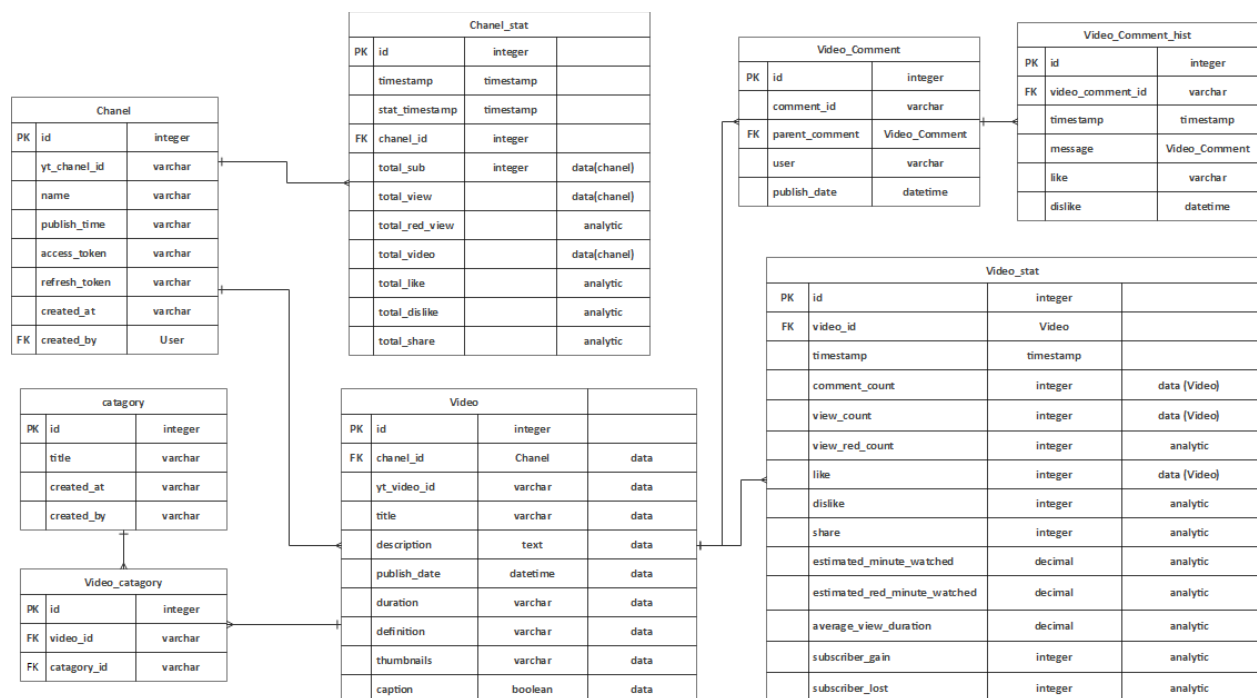
โดยทั้ง 2 ระบบจะทำงานร่วมกันผ่านตัวกลางคือ Social Monitoring Database ที่เป็นส่วนประกอบเดิมของระบบเพื่อให้ระบบที่จะทำการพัฒนาเพิ่มเติมสามารถทำงานร่วมกับระบบเดิมได้

1.1.1. การพัฒนาฐานข้อมูลของระบบ

จากการศึกษาการให้บริการข้อมูลของ YouTube Data API และ YouTube Analytic API ทางทีมพัฒนาจึงได้ออกแบบฐานข้อมูลในการเก็บสถิติจาก API ทั้งสองโดยประกอบไปด้วย 8 ตารางดังนี้

1. Channal สำหรับการจัดเก็บข้อมูลของช่อง YouTube ที่ระบบจัดการ
2. Channal_stat สำหรับการจัดเก็บข้อมูลสถิติของช่อง YouTube
3. Category สำหรับการจัดเก็บประเภทของวิดีโอ
4. Video สำหรับการจัดเก็บข้อมูลวิดีโอ
5. Video_stat สำหรับจัดเก็บข้อมูลสถิติสำหรับวิดีโอ
6. Video_catagory สำหรับจัดเก็บประเภทของวิดีโอ
7. Video_comment สำหรับการจัดเก็บข้อคิดเห็นของวิดีโอ
8. Video_comment_hist สำหรับจัดเก็บประวัติการแก้ไขข้อคิดเห็น

โดยทั้ง 8 ตารางมีความสัมพันธ์และการจัดเก็บข้อมูลตามรูปที่ 53



รูปที่ 53 การออกแบบฐานข้อมูลของระบบรายงานสถิติช่อง YouTube

6.1.2. การพัฒนาหน้ารายงาน

จากการออกแบบหน้ารายงานทางทีมผู้พัฒนาได้อิงจากข้อมูลที่ทางผู้พัฒนาสามารถดึงข้อมูลมาใช้งานได้ ทำให้ทางทีมผู้พัฒนาสามารถแบ่งการรายงานผลออกเป็น 3 หน้าหลักได้แก่

3. หน้ารายงานสถิติช่อง YouTube ในภาพรวม
4. หน้ารายงานสถิติช่อง YouTube
5. หน้ารายงานสถิติของวิดีโอ

6.1.2.1. หน้ารายงานสถิติช่อง YouTube ในภาพรวม

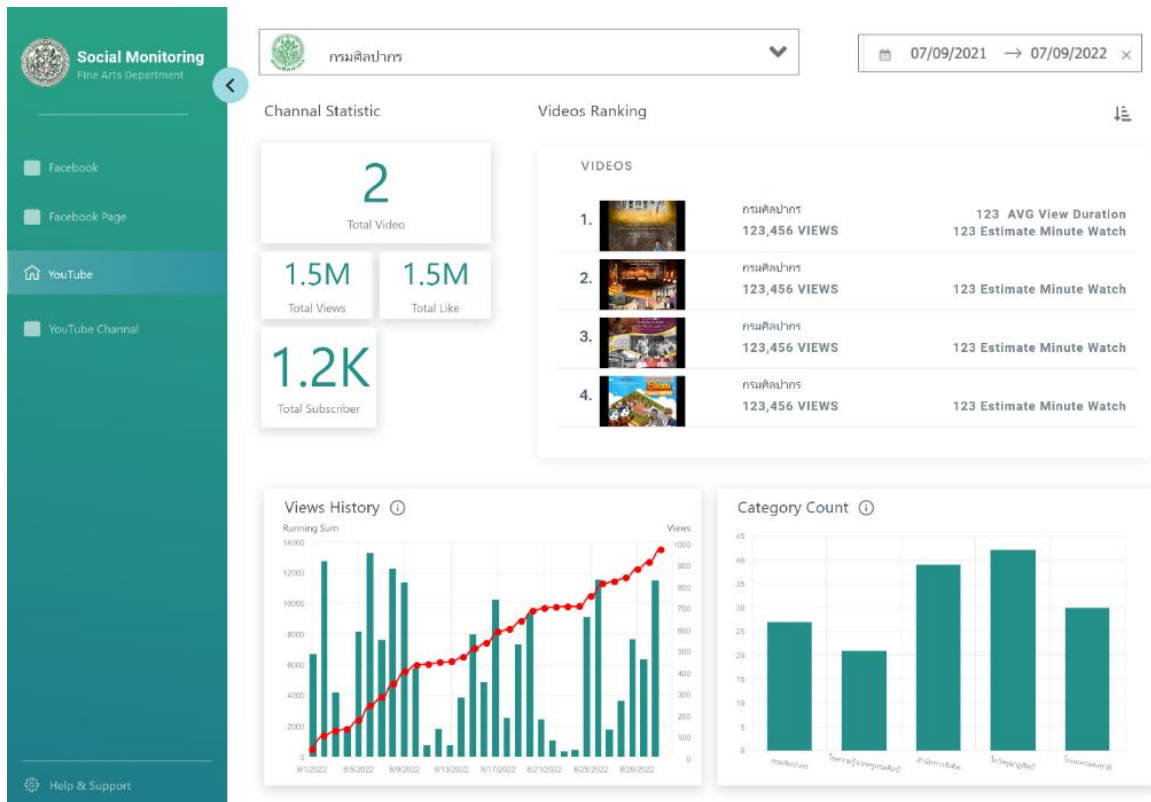
หน้ารายงานสถิติช่อง YouTube ในภาพรวมถูกออกแบบสำหรับการรายงานสถิติในทุกช่องและทุกวิดีโอที่ถูกตั้งค่าโดยผู้ดูแลให้ทำการดึงข้อมูลผ่าน YouTube API โดยการรวบรวมข้อมูลจากช่องทางต่าง ๆ และทำการสรุปออกมาเป็นภาพรวมของค่าสถิติทั้งหมดในชุดข้อมูล



รูปที่ 54 หน้ารายงานสถิติช่อง YouTube ในภาพรวม

6.1.2.2. หน้ารายงานสถิติช่อง YouTube

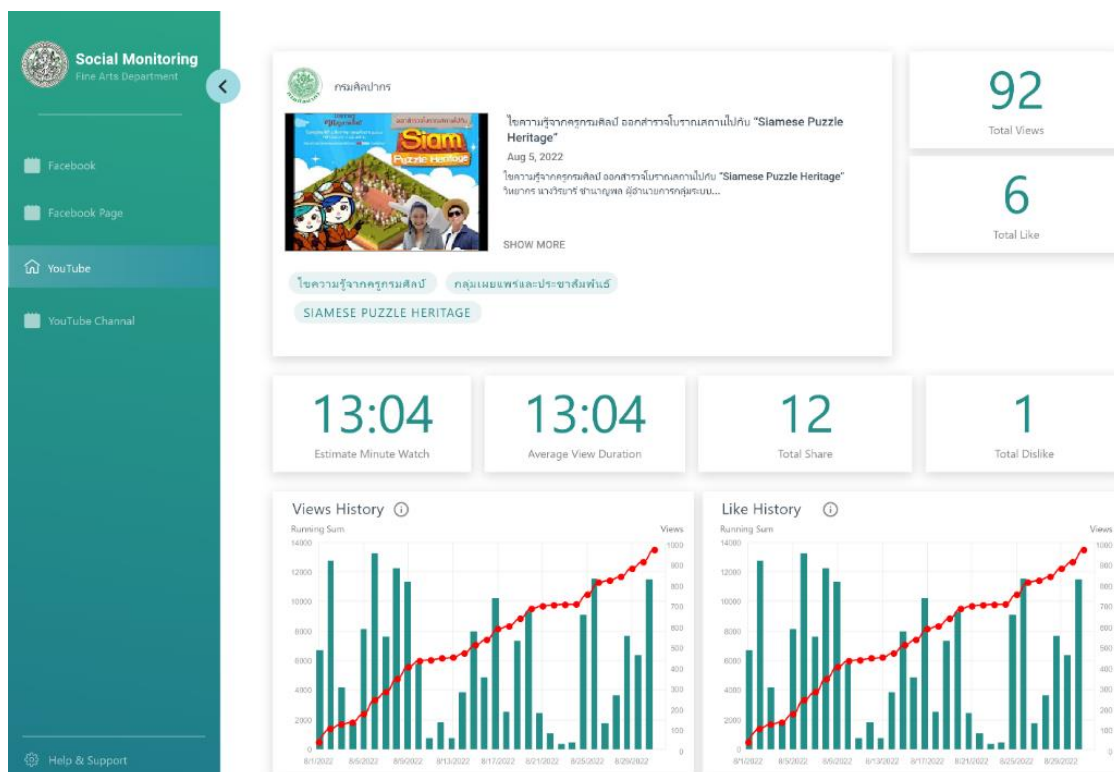
หน้ารายงานสถิติช่อง YouTube ถูกออกแบบสำหรับการแสดงผลข้อมูลและรายงานสถิติของช่อง YouTube เพียงช่องเดียวที่ผู้ใช้งานได้ทำการเลือก โดยจะรวบรวมข้อมูลจากทุกวิดีโอในช่องดังกล่าวมาสรุปเป็นภาพรวมของช่อง ๆ นั้น



รูปที่ 55 หน้ารายงานสถิติช่อง YouTube

6.1.2.3. หน้ารายงานสถิติของวิดีโอ

หน้ารายงานสถิติของวิดีโอถูกออกแบบสำหรับการแสดงผลข้อมูลและรายงานสถิติของวิดีโอเพียงหนึ่งวิดีโอที่ผู้ใช้งานเลือก โดยจะแตกต่างจากสองหน้าที่กล่าวมาข้างต้นเพราะการแสดงผลข้อมูลของหน้านี้เป็นเพียงการดึงข้อมูลต่าง ๆ ขึ้นมาแสดงโดยไม่ได้ทำการสรุปผลข้อมูลแต่อย่างใดเนื่องจากข้อมูลของวิดีโอถือเป็นหน่วยที่เล็กที่สุดของข้อมูล



รูปที่ 56 หน้ารายงานสถิติของวิดีโอ

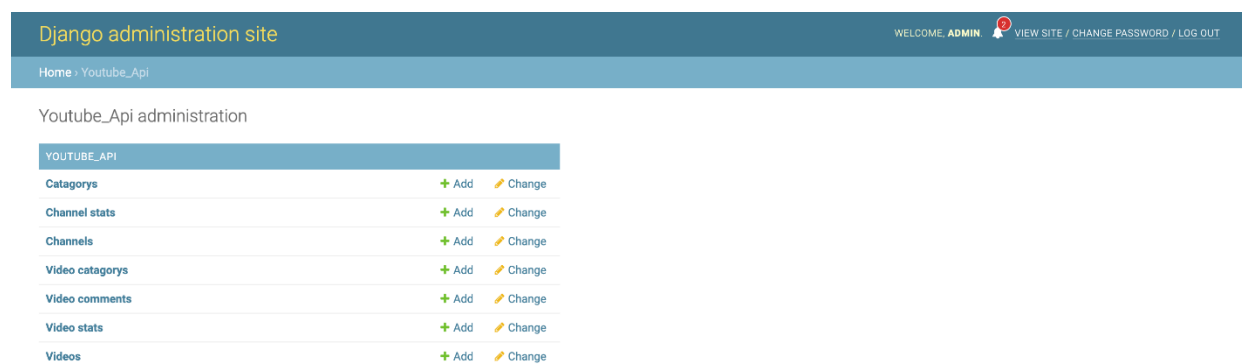
6.2. การพัฒนาระบบรายงานข้อมูลสถิติช่อง YouTube ของกรมศิลปากร

ในการพัฒนาระบบรายงานข้อมูลสถิติช่อง YouTube ของกรมศิลปากรทางทีมพัฒนาได้ดำเนินการในส่วนของการรวบรวมข้อมูลจาก YouTube API และหน้าต่างการจัดการตั้งค่าระบบของผู้ดูแลจนแล้วเสร็จ โดยในหน้าต่างของผู้ใช้งานจะใช้หน้าต่างผู้ใช้งานเดียวกับระบบ Social Monitoring เดิมเพื่อให้ง่ายต่อการดูแลของผู้ดูแลระบบโดยในหน้าต่างการจัดการตั้งค่าและจัดการเนื้อหาประกอบไปด้วยหน้าต่างผู้ใช้งานทั้งหมดจำนวน 5 หน้าต่าง ผู้ใช้งานประกอบไปด้วย

1. หน้าหลักของระบบจัดการรายงานสถิติช่อง YouTube
2. หน้าจัดการช่องของระบบจัดการรายงานสถิติช่อง YouTube
3. หน้าจัดการสถิติช่องของระบบจัดการรายงานสถิติช่อง YouTube
4. หน้าจัดการวิดีโอของระบบจัดการรายงานสถิติช่อง YouTube
5. หน้าจัดการสถิติวิดีโอของระบบจัดการรายงานสถิติช่อง YouTube

6.2.1. หน้าหลักของระบบจัดการรายงานสถิติช่อง YouTube

หน้าหลักของระบบจัดการรายงานสถิติช่อง YouTube เป็นหน้าต่างหลักเมื่อผู้ดูแลระบบเลือกเข้ามาใช้งานในระบบจัดการของ YouTube โดยระบบจะปรากฏเมนูย่อยต่าง ๆ ที่ผู้ดูแลสามารถเข้าไปจัดการเนื้อหาหรือตั้งค่าของระบบได้



รูปที่ 57 หน้าหลักของระบบจัดการรายงานสถิติช่อง YouTube

6.2.2. หน้าจัดการช่องของระบบจัดการรายงานสถิติช่อง YouTube

หน้าจัดการช่องของระบบจัดการรายงานสถิติช่อง YouTube เป็นหน้าต่างสำหรับการจัดการช่องของ YouTube ที่ระบบทำการดูแลโดยช่องที่ปรากฏในหน้าต่างนี้ระบบจะทำการรวบรวมข้อมูลผ่าน YouTube API ทั้งสอง แต่การรวบรวมข้อมูลผ่าน YouTube Analytic API ผู้ดูแลระบบจะต้องทำการตั้งค่าด้วยการเข้าบัญชีของเจ้าของช่องนั้น ๆ ด้วย

Django administration site

WELCOME, ADMIN VIEW SITE / CHANGE PASSWORD / LOG OUT

Home > Youtube_Api > Channels

Select channel to change

ADD CHANNEL +

FILTER

By Status Token

- All
- Re Token
- Active
- Inactive

YT CHANNEL ID	STATUS COLORED	NAME	DESCRIPTION	PUBLISH DATETIME	ACCESS TOKEN
☐ UCZFw0B1s60kTV1LOExkmUw	ACTIVE				ya29.a0AVA9y1v8zzdGymv6X11avL1RnDmf
☐ UCgMRAVlQsjW902GYP4MAAcQ	INIT	AxD Harumi-FC	All talesrunner video play by keyboard, Thanks for watch and enjoy my video XD ** Status : Quit playing talesrunner ** -- Because i'm studying very hard in university and boring talesrunner's new patch :(, haha thank you everybody for watching my talesrunner video :) --	2007-12-23T11:26:30Z	-
☐ UCpzacK4trFrHBizIMmEZHeQ	EXPIRED	MookKa 030	สวัสดีค่ะ ทุกคน ที่มาดูคลิปของเรานะคะ หวังว่าทุกคนจะชอบนะคะ อย่าลืมกดซับ กดLike กดแชร์ด้วยนะคะ ขอขอบคุณค่ะ นี่คือเพจของเราในเฟสบุ๊ค : https://www.facebook.com/MookKa030 ติ๊กต็อกของเรา : https://www.instagram.com/mookka_030/?hl=th แล้วเรามีช่องใน tiktok : https://vt.tiktok.com/D71qbj/ ฝากติดตามด้วยนะคะ	2014-03-26T15:02:52Z	ya29.A0AVA9y1vNCCE5IQ736Uc6gmLQN0N

3 channels

รูปที่ 58 หน้าจัดการช่องของระบบจัดการรายงานสถิติช่อง YouTube

6.2.3. หน้าจัดการสถิติช่องของระบบจัดการรายงานสถิติช่อง YouTube

หน้าจัดการสถิติช่องของระบบจัดการรายงานสถิติช่อง YouTube เป็นหน้าสำหรับการจัดการเนื้อหาของข้อมูลสถิติที่ได้รวบรวมมาได้จาก YouTube API

Select channel stat to change

ADD CHANNEL STAT +

Action: Go 0 of 38 selected

☐	CHANNEL ID	TIMESTAMP	STAT TIMESTAMP	TOTAL SUB	TOTAL VIEW	TOTAL RED VIEW	TOTAL VIDEO	TOTAL LIKE	TOTAL DISLIKE	TOTAL SHARE
☐	UCgMRAVlQsjW9O2GYP4MAAcQ	Aug. 23, 2022, 1:26 a.m.	-	111	3120	-	9	-	-	-
☐	UCpzacK4trFrHBlzlMmEZHeQ	Aug. 23, 2022, 1:26 a.m.	Aug. 23, 2022, 1:26 a.m.	3	381	14	5	-	0	6
☐	UCgMRAVlQsjW9O2GYP4MAAcQ	Aug. 23, 2022, 1:24 a.m.	-	111	3120	-	9	-	-	-
☐	UCpzacK4trFrHBlzlMmEZHeQ	Aug. 23, 2022, 1:24 a.m.	Aug. 23, 2022, 1:24 a.m.	3	381	14	5	-	0	6
☐	UCgMRAVlQsjW9O2GYP4MAAcQ	Aug. 23, 2022, 1:23 a.m.	-	111	3120	-	9	-	-	-
☐	UCpzacK4trFrHBlzlMmEZHeQ	Aug. 23, 2022, 1:23 a.m.	Aug. 23, 2022, 1:23 a.m.	3	381	14	5	-	0	6
☐	UCpzacK4trFrHBlzlMmEZHeQ	Aug. 23, 2022, 1:21 a.m.	Aug. 23, 2022, 1:21 a.m.	3	381	14	5	-	0	6

รูปที่ 59 หน้าจัดการสถิติช่องของระบบจัดการรายงานสถิติช่อง YouTube

6.2.4. หน้าจัดการวิดีโอของระบบจัดการรายงานสถิติช่อง YouTube

หน้าจัดการวิดีโอของระบบจัดการรายงานสถิติช่อง YouTube เป็นหน้าต่างสำหรับการจัดเก็บเนื้อหาของวิดีโอที่ระบบได้รวบรวมมา

Select video to change

ADD VIDEO +

Action: Go 0 of 6 selected

☐	CHANNEL ID	YT VIDEO ID	TITLE	DESCRIPTION	PUBLISH DATETIME	DURATION	DEFINITION	THUMBNAILS	CAPTION
☐	UCgMRAVlQsjW9O2GYP4MAAcQ	1CaD2bK7qr0	Sun & Moon 5 5 1'21'60 By JAMES2BOW	[description]	2013-11-27T11:18:32Z	PT1M48S	hd	https://i.ytimg.com/vi/1CaD2bK7qr0/default.jpg	
☐	UCpzacK4trFrHBlzlMmEZHeQ	TEGDfs9crZc	รีวิว AIBOT CAFFINE โดคริมที่ให้นอนหลับ	[description]	2020-06-03T12:46:54Z	PT59S	hd	https://i.ytimg.com/vi/TEGDfs9crZc/default.jpg	
☐	UCpzacK4trFrHBlzlMmEZHeQ	aoKQs_1yb5M	The sims 4 เล่นไปเรื่อย : Ep.1 การเริ่มต้นของโมด ในเมือง	[description]	2020-06-14T17:09:00Z	PT24M45S	hd	https://i.ytimg.com/vi/aoKQs_1yb5M/default.jpg	
☐	UCpzacK4trFrHBlzlMmEZHeQ	Z1ybNLdAB_J	Line Play: มาเล่นเทพ Avengers กันละ	[description]	2020-06-23T08:54:40Z	PT8M5S	hd	https://i.ytimg.com/vi/Z1ybNLdAB_J/default.jpg	
☐	UCpzacK4trFrHBlzlMmEZHeQ	Wozgu_5A8-4	ุBehind The Scenes In TikTok : าวครูปน้องแมว	[description]	2020-07-15T03:55:45Z	PT1M4S	hd	https://i.ytimg.com/vi/Wozgu_5A8-4/default.jpg	
☐	UCpzacK4trFrHBlzlMmEZHeQ	DXTJ52YVgr4	[LINE PLAY TH] เนเธ่านาม Line Play	[description]	2020-07-17T11:53:46Z	PT5M54S	hd	https://i.ytimg.com/vi/DXTJ52YVgr4/default.jpg	

6 videos

รูปที่ 60 หน้าจัดการวิดีโอของระบบจัดการรายงานสถิติช่อง YouTube

6.2.5. หน้าจัดการสถิติวิดีโอของระบบจัดการรายงานสถิติช่อง YouTube

หน้าจัดการสถิติวิดีโอของระบบจัดการรายงานสถิติช่อง YouTube เป็นหน้าต่างสำหรับการจัดการเนื้อหาทางสถิติของวิดีโอที่ได้รวบรวมมาจาก YouTube API

Select video stat to change

ADD VIDEO STAT +

Action: Go 0 of 30 selected

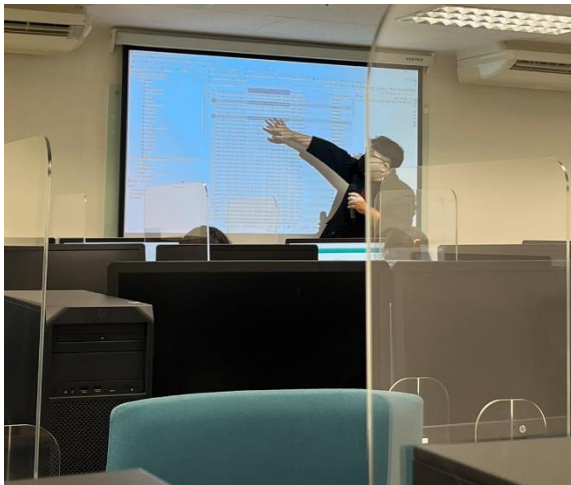
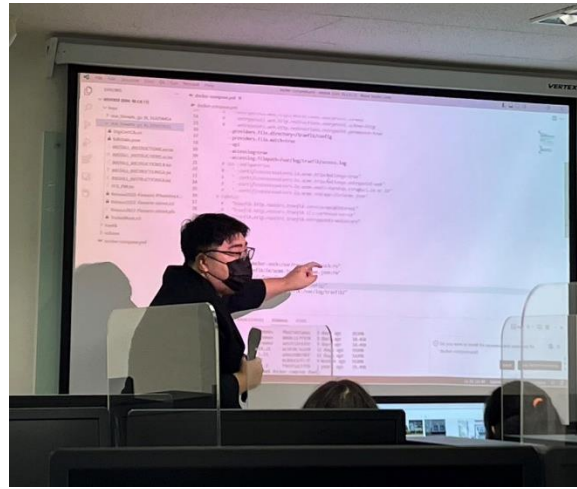
☐	VIDEO ID	TIMESTAMP	STAT TIMESTAMP	COMMENT COUNT	VIEW COUNT	VIEW RED COUNT	LIKE	DISLIKE	SHARE	ESTIMATED MINUTE WATCHED	ESTIMATED RED MINUTE WATCHED	AVERAGE V
☐	TEGDfs9crZc	Aug. 23, 2022, 1:26 a.m.	Aug. 23, 2022, 1:26 a.m.	0	27	0	1	0	1	8.0	0.0	18.0
☐	aoKQs_1yb5M	Aug. 23, 2022, 1:26 a.m.	Aug. 23, 2022, 1:26 a.m.	0	30	1	1	0	3	81.0	0.0	163.0
☐	Z1ybNLdAB_J	Aug. 23, 2022, 1:26 a.m.	Aug. 23, 2022, 1:26 a.m.	0	10	0	1	0	1	17.0	0.0	102.0
☐	Wozgu_5A8-4	Aug. 23, 2022, 1:26 a.m.	Aug. 23, 2022, 1:26 a.m.	0	15	0	0	0	0	4.0	0.0	26.0
☐	DXTJ52YVgr4	Aug. 23, 2022, 1:26 a.m.	Aug. 23, 2022, 1:26 a.m.	0	299	7	7	0	0	112.0	2.0	22.0
☐	TEGDfs9crZc	Aug. 23, 2022, 1:24 a.m.	Aug. 23, 2022, 1:24 a.m.	0	27	0	1	0	1	8.0	0.0	18.0
☐	aoKQs_1yb5M	Aug. 23, 2022, 1:24 a.m.	Aug. 23, 2022, 1:24 a.m.	0	30	1	1	0	3	81.0	0.0	163.0
☐	Z1ybNLdAB_J	Aug. 23, 2022, 1:24 a.m.	Aug. 23, 2022, 1:24 a.m.	0	10	0	1	0	1	17.0	0.0	102.0
☐	Wozgu_5A8-4	Aug. 23, 2022, 1:24 a.m.	Aug. 23, 2022, 1:24 a.m.	0	15	0	0	0	0	4.0	0.0	26.0
☐	DXTJ52YVgr4	Aug. 23, 2022, 1:24 a.m.	Aug. 23, 2022, 1:24 a.m.	0	299	7	7	0	0	112.0	2.0	22.0
☐	TEGDfs9crZc	Aug. 23, 2022, 1:23 a.m.	Aug. 23, 2022, 1:23 a.m.	0	27	0	1	0	1	8.0	0.0	18.0
☐	aoKQs_1yb5M	Aug. 23, 2022, 1:23 a.m.	Aug. 23, 2022, 1:23 a.m.	0	30	1	1	0	3	81.0	0.0	163.0
☐	Z1ybNLdAB_J	Aug. 23, 2022, 1:23 a.m.	Aug. 23, 2022, 1:23 a.m.	0	10	0	1	0	1	17.0	0.0	102.0
☐	Wozgu_5A8-4	Aug. 23, 2022, 1:23 a.m.	Aug. 23, 2022, 1:23 a.m.	0	15	0	0	0	0	4.0	0.0	26.0
☐	DXTJ52YVgr4	Aug. 23, 2022, 1:23 a.m.	Aug. 23, 2022, 1:23 a.m.	0	299	7	7	0	0	112.0	2.0	22.0

รูปที่ 61 หน้าจัดการสถิติวิดีโอของระบบจัดการรายงานสถิติช่อง YouTube

7. การอบรมถ่ายทอดองค์ความรู้ วิธีการใช้งาน และการดูแลระบบ

ทางทีมผู้พัฒนาได้ทำการจัดการอบรมประกอบไปด้วยการอบรมทั้งหมด 6 หลักสูตรในวันที่ 23 มกราคม พ.ศ. 2566 ถึงวันที่ 25 มกราคม พ.ศ. 2566 โดยกำหนดหัวข้อในการอบรมเป็นไปตามสิ่งที่ได้ทำการพัฒนาโครงการจนแล้วเสร็จ รวมถึงการใช้งานและการดูแลระบบต่อไปในอนาคตตามตารางต่อไปนี้

ลำดับ	หลักสูตร	วันที่ฝึกอบรม	เวลา	รูปแบบการอบรม
1	Introduction to Name Entity Recognition (NER) with PythaiNLP	23 มกราคม 2566	09.00 – 12.00	บรรยายและสาธิตการปฏิบัติ
2	Introduction to Image Retrieval with a Case Study of Historical Site Retrieval	23 มกราคม 2566	13.00 – 16.00	บรรยายและสาธิตการปฏิบัติ
3	System Deployment and Management	24 มกราคม 2566	09.00 – 12.00	บรรยายและสาธิตการปฏิบัติ
4	Social Monitoring (YouTube)	24 มกราคม 2566	13.00 – 16.00	บรรยายและสาธิตการปฏิบัติ
5	Introduction to SQL and NoSQL with MySQL and MongoDB	25 มกราคม 2566	09.00 – 12.00	บรรยายและสาธิตการปฏิบัติ
6	Project Summary	25 มกราคม 2566	13.00 – 16.00	บรรยายและสาธิตการปฏิบัติ



รูปที่ 62 การจัดการอบรม