การวิเคราะห์ความรับผิดชอบและข้อกำหนดจากประกาศงานทางด้านเทคโนโลยีสารสนเทศ

นรัชกร ฤกษ์ศิริ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาการสารสนเทศ
คณะวิทยาการสารสนเทศ มหาวิทยาลัยบูรพา
2566
ลิขสิทธิ์เป็นของมหาวิทยาลัยบูรพา

การวิเคราะห์ความรับผิดชอบและข้อกำหนดจากประกาศงานทางด้านเทคโนโลยีสารสนเทศ

นรัชกร ฤกษ์ศิริ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาการสารสนเทศ
คณะวิทยาการสารสนเทศ มหาวิทยาลัยบูรพา
2566
ลิขสิทธิ์เป็นของมหาวิทยาลัยบูรพา

Analyzing responsibilities and requirements from job posts in IT domain

NARATCHAKORN ROEKSIRI

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR MASTER DEGREE OF SCIENCE

IN INFORMATICS

FACULTY OF INFORMATICS

BURAPHA UNIVERSITY

2023

COPYRIGHT OF BURAPHA UNIVERSITY

คณะกรรมการควบคุมวิทยานิพนธ์และคณะกรรมการสอบวิทยานิพนธ์ได้พิจารณาวิทยานิพนธ์ของ นรัชกร ฤกษ์ศิริ ฉบับนี้แล้ว เห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการสารสนเทศ ของมหาวิทยาลัยบูรพาได้

คณะกรรมการควบคุมวิทยานิพนธ์
อาจารย์ที่ปรึกษาหลัก

.....
(ผู้ช่วยศาสตราจารย์ ดร.โกเมศ อัมพวัน)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธาน
(รองศาสตราจารย์ ดร.อนุชิต จิตพัฒนกุล)

..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.อุรีรัฐ สุขสวัสดิ์ชน)

..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.โกเมศ อัมพวัน)

..... คณบดีคณะวิทยาการสารสนเทศ
(ผู้ช่วยศาสตราจารย์ ดร. กฤษณะ ชินสาร)
วันที่.....เดือน.....พ.ศ.....

บัณฑิตวิทยาลัย มหาวิทยาลัยบูรพา อนุมัติให้รับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการสารสนเทศ ของมหาวิทยาลัยบูรพา

..... คณบดีบัณฑิตวิทยาลัย
(รองศาสตราจารย์ ดร.วิทวัส แจ้งเอี่ยม)
วันที่.....เดือน.....พ.ศ.....

61910137: สาขาวิชา: วิทยาการสารสนเทศ; วท.ม. (วิทยาการสารสนเทศ)
คำสำคัญ: ประกาศรับสมัครงานที่เกี่ยวข้องกับด้านเทคโนโลยีสารสนเทศ, การวิเคราะห์ประกาศรับสมัครงาน, การค้นพบข้อมูลที่เกิดขึ้นร่วมกัน, การวิเคราะห์โดยเจาะลึกเฉพาะเจาะจง

นรัชกร ฤกษ์ศิริ : การวิเคราะห์ความรับผิดชอบและข้อกำหนดจากประกาศงานทางด้านเทคโนโลยีสารสนเทศ. (Analyzing responsibilities and requirements from job posts in IT domain) คณะกรรมการควบคุมวิทยานิพนธ์: โกเมศ อัมพวัน ปี พ.ศ. 2566.

งานวิจัยนี้เน้นการวิเคราะห์และสรุปความรับผิดชอบและข้อกำหนดจากประกาศงานทางด้านเทคโนโลยีสารสนเทศ ในการแก้ปัญหานี้จึงนำเสนอวิธีการค้นพบข้อมูลที่เกิดขึ้นร่วมกันในการประกาศรับสมัครงานที่เกี่ยวข้องกับคอมพิวเตอร์ (Co-occurrence Information Discovery in Computer-related Job Advertisements, CID-CJA) วิธีใหม่สำหรับสกัดข้อมูลที่มีประโยชน์จากประกาศรับสมัครงาน โดยดำเนินการศึกษาทดลองข้อมูลประกาศรับสมัครงานในสายงานวิทยาการเทคโนโลยีสารสนเทศที่เป็นภาษาอังกฤษเท่านั้น จากเว็บไซต์รับสมัครงานในประเทศไทยได้แก่ jobsDB, Linkedin และ Indeed รวบรวมเป็นจำนวนทั้งหมด 4,280 ประกาศงาน โดยในการประมวลผลได้จัดกลุ่มตำแหน่งงานได้ทั้งหมด 11 กลุ่มตำแหน่งงานในสายงานที่เกี่ยวข้องกับวิทยาการเทคโนโลยี เพื่อแสดงให้เห็นถึงรูปแบบโครงสร้างข้อมูลของรายละเอียดของการประกาศรับสมัครงานและแสดงการค้นพบข้อมูลเชิงลึกที่เกิดขึ้น ความยืดหยุ่นของวิธีการ CID-CJA ทำให้ผู้ใช้งานสามารถดำเนินการวิเคราะห์ข้อมูลแบบเจาะลึกลงในส่วนของข้อมูลที่เฉพาะเจาะจง สามารถดูผลสรุปของข้อมูลได้ และปรับระดับความละเอียดของข้อมูลตามความต้องการ เพื่อช่วยให้ผู้หางานเข้าใจหน้าที่ที่ต้องรับผิดชอบ ทักษะและคุณสมบัติเฉพาะที่เป็นที่ต้องการ และเพื่อช่วยให้พวกเขาหรือสถานศึกษาสามารถจัดเส้นทางการศึกษาสู่อาชีพได้อย่างมีประสิทธิภาพมากขึ้น นอกจากนี้ยังช่วยให้นายจ้างสามารถสร้างคำอธิบายลักษณะงาน ให้ตรงกับความต้องการขององค์กร วิธีการ CID-CJA ได้ประเมินผลการทดสอบ 3 ด้าน ซึ่งประกอบด้วยความถูกต้องในการจัดกลุ่ม, ความแม่นยำในการสกัดคำสำคัญ, และความถูกต้องในการจัดประเภทและสร้างโครงสร้างข้อมูลที่สำคัญ ศึกษาทดลองโดยใช้ประกาศรับสมัครงานจริง ๆ เพื่อให้ได้มาซึ่งผลลัพธ์ที่สามารถเปิดเผยข้อมูลที่มีค่าสำหรับผู้ที่เกี่ยวข้องกับงานที่เกี่ยวข้องกับศาสตร์ทางด้านเทคโนโลยีสารสนเทศ

61910137: MAJOR: INFORMATICS; M.Sc. (INFORMATICS)
KEYWORDS: computer-related job advertisements, job advertisement analysis, co-occurrence information discovery, drill- down analysis

NARATCHAKORN ROEKSIRI : ANALYZING RESPONSIBILITIES AND REQUIREMENTS FROM JOB POSTS IN IT DOMAIN. ADVISORY COMMITTEE: KOMATE AMPHAWAN, Ph.D. 2023.

This research focuses on analyzing and summarizing responsibilities and requirements from Information Technology job postings. To tackle this, a method known as Co-occurrence Information Discovery in Computer-related Job Advertisements (CID-CJA) is introduced. This novel method is designed to extract valuable information from job postings and has been tested using a dataset of 4,280 English-language job postings from job search websites in Thailand, including jobsDB, LinkedIn, and Indeed.

The process involves categorizing job positions into 11 groups related to technology. This categorization is employed to highlight the structural patterns in job posting details and unveil profound insights into the data. The flexibility of the CID-CJA method allows users to conduct in-depth analyses on specific data elements, access summaries, and adjust the level of detail according to their requirements.

The CID-CJA method has been assessed from three key perspectives: accuracy in grouping, precision in keyword extraction, and correctness in categorizing and structuring crucial data. Real job postings were utilized to validate the method, resulting in insights that can assist job seekers in comprehending responsibilities, necessary skills, and specific job attributes. Furthermore, it can aid employers in crafting job descriptions tailored to their organizational needs. The CID-CJA method represents an effective approach for those engaged in the Information Technology field.

มหาวิทยาลัยบูรพา
BURAPHA UNIVERSITY

กิตติกรรมประกาศ

งานวิจัยนี้สำเร็จลุล่วงด้วยความกรุณาจาก ผู้ช่วยศาสตราจารย์ ดร. โกเมศ อัมพวัน อาจารย์ที่ปรึกษาที่เคารพอย่างยิ่งของผู้วิจัย ที่ได้มอบประสบการณ์ที่ดีในการดำเนินงาน มอบคำแนะนำและแนวทางที่มีคุณค่าอย่างมากสำหรับงานวิจัยของเรา ผู้วิจัยรู้สึกดีใจที่ได้รับความรู้และกระบวนการวิธีคิดจากท่าน ท่านเป็นแรงผลักดันให้ผู้วิจัยต่อสู้ในการทำงานวิจัยนี้ เป็นกำลังใจอันใหญ่ และตลอดระยะเวลาของการดำเนินงานวิจัย ไม่เพียงแต่คอยดูแลและสนับสนุนผู้วิจัยในด้านวิชาการเท่านั้น แต่ยังช่วยเหลือแก้ปัญหาข้อบกพร่องต่าง ๆ ที่เกิดขึ้นในงานวิจัยด้วยความละเอียดถี่ถ้วนและเอาใจใส่ เพื่อให้งานวิจัยนี้มีคุณภาพที่ดีขึ้นอย่างมาก ขอบคุณอาจารย์ที่เป็นแรงบันดาลใจและแนวทางที่แสนมีประสิทธิภาพ

นอกจากนี้ ขอขอบคุณครอบครัว คุณพ่อ คุณแม่ น้องสาว และสัตว์เลี้ยง ที่คอยเป็นที่พึ่งทางจิตใจที่ดีให้กับผู้วิจัย เป็นกำลังสนับสนุนอยู่ตลอดเวลา ขอขอบคุณเป็นอย่างสูง

ขอขอบคุณ พี่ก้อย พี่ต้อม พี่ปิ่น พี่เดียร์ รุ่นพี่ทุก ๆ คนในห้องปฏิบัติการวิจัย Computational Innovation Laboratory (CIL) ที่คอยมอบกำลังใจ ให้คำปรึกษา พาไปกินของอร่อย ให้คำปรึกษาในขั้นตอนของงานวิจัยนี้ พี่ ๆ ทุกคนคือความร่วมมือที่สำคัญที่ช่วยให้ผู้วิจัยสามารถตั้งคำถามและหาคำตอบในงานวิจัยนี้ได้อย่างมีคุณค่า และเป็นมิตรภาพที่ดี เป็นความทรงจำที่ดีสำหรับผู้วิจัยอย่างมาก

ขอขอบคุณน้องวรรณ เพื่อนรุ่นเดียวกันในการศึกษาระดับปริญญาโท ที่เป็นที่พูดคุย พาไปเที่ยวและให้คำปรึกษาเข้าใจปัญหาของกันและกันในการทำวิจัย ตั้งแต่วันแรกที่พบกันคุณเป็นเพื่อนที่มีค่าที่คู่ควรได้รู้จักและมีโอกาสได้เรียนร่วมกัน ตลอดจนขอขอบคุณคณะผู้บริหาร อาจารย์ และบุคลากรในคณะวิทยาการสารสนเทศที่ช่วยเหลือ สนับสนุน แก้ปัญหาให้กับผู้วิจัยเสมอมา

เนื่องจากงานวิจัยฉบับนี้ได้รับการสนับสนุนทุนวิจัยจากคณะวิทยาการสารสนเทศ มหาวิทยาลัยบูรพา จึงขอขอบคุณมา ณ ที่นี้

สุดท้ายนี้ ขอขอบคุณตนเองที่มุ่งมั่น และความพยายามในแต่ละวัน จนสามารถดำเนินงานวิจัยนี้ไปจนสำเร็จลุล่วง และขอบคุณทุกคนที่มีส่วนร่วมในความสำเร็จของงานวิจัยนี้ คุณทุกคนเป็นส่วนสำคัญที่ช่วยให้ผู้วิจัยสามารถสร้างผลงานที่มีคุณค่าและสร้างความเป็นไปได้ในการพัฒนาด้านนี้ได้อย่างดี ผู้วิจัยขอให้ทุกคนมีความสุขและสำเร็จในงานและชีวิตของคุณต่อไป

นรัชกร ฤกษ์ศิริ

มหาวิทยาลัยบูรพา
BURAPHA UNIVERSITY

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ช

สารบัญ ..... ฌ

สารบัญตาราง ..... ฐ

สารบัญภาพ ..... ฒ

บทที่ 1 บทนำ 1

1.1 ที่มาและความสำคัญของงานวิจัย 1

1.2 วัตถุประสงค์ของงานวิจัย 3

1.3 ขอบเขตของงานวิจัย 3

1.4 ขั้นตอนและระยะเวลาในการดำเนินงานวิจัย 5

1.5 ประโยชน์ที่คาดว่าจะได้รับ 7

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 8

2.1 การประมวลผลข้อความเบื้องต้น (Text preprocessing) 8

2.1.1 การแบ่งประโยค/คำ (Sentence/Word tokenization) 8

2.1.2 การตรวจสอบและแก้ไขการสะกดคำ (Word error correction) 9

2.1.3 การแปลงคำย่อ/ชื่อย่อ (Abbreviation resolution) 10

2.1.4 การลบคำหยุด (Stop Word Removal) 10

2.1.5 การแปลงตัวอักษรให้เป็นตัวพิมพ์เล็ก (Lowercase conversion) 10

2.1.6 การลบเครื่องหมายวรรคตอนและอักขระ (Punctuation removal) 11

2.1.7 การลดรูปของคำให้อยู่ในรากศัพท์ (Stemming and lemmatization) 11

2.1.8 การวิเคราะห์ชนิดของคำ (Parts of speech tagging) 11
2.2 คำฝังตัว (Word Embeddings) 12
2.2.1. Word2Vec 13
2.2.2 GloVe (Global Vectors for Word Representation) 13
2.2.3 Cosine Similarity 13
2.3 งานวิจัยที่เกี่ยวข้อง 14
2.3.1 การสรุปใจความสำคัญอัตโนมัติ (automatic text summarization) 14
2.3.1.1 การสรุปใจความสำคัญจากเอกสารหนึ่ง ๆ (single-document summarization) 14
2.3.1.2 การสรุปใจความสำคัญจากหลายเอกสาร (multi-document summarization) 15
2.3.1.3 แบบผสม 16
2.3.1.3.1 การสรุปข่าวสาร 16
2.3.1.3.2 การสรุปอีเมลล์ 16
2.3.1.3.3 การสรุปเว็บไซต์เครือข่ายสังคม Twitter 17
2.3.2 งานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ประกาศงาน 17
บทที่ 3 วิธีการดำเนินการวิจัย 23
3.1 การเตรียมข้อมูลและการรวบรวมข้อมูลประกาศรับสมัครงานเฉพาะศาสตร์ทางด้านเทคโนโลยีสารสนเทศ 24
3.1.1 ศึกษาโครงสร้างข้อมูลประกาศงานของเว็บไซต์ 24
3.1.2 การสกัดข้อมูลจากเว็บไซต์ jobsDB Linkedin และ Indeed 25
3.1.3 การรวบรวมคำย่อของคำศัพท์เฉพาะทางด้านเทคโนโลยีสารสนเทศ 30
3.1.4 การรวบรวมคำศัพท์เฉพาะทางคอมพิวเตอร์เทคโนโลยี 31
3.1.5 การรวบรวมคำกิริยาที่มีความหมายคล้ายคลึงหรือเหมือนกันกับคำกริยาอื่น (Verb synonym) 33

3.1.6 การรวบรวมกฎทางภาษาศาสตร์ 35
3.2 การจัดกลุ่มประกาศงาน 40
3.3 การประมวลผลข้อความเบื้องต้น 41
3.3.1 การตรวจสอบและแก้ไขการสะกดคำ (Word error correction) 44
3.3.2 การแปลงคำย่อ/ชื่อย่อ (Abbreviation resolution) 44
3.3.3 การแปลงตัวอักษรให้เป็นตัวพิมพ์เล็ก (Lowercase conversion) 45
3.3.4 การแบ่งประโยค (Sentence tokenization) 46
3.3.5 ลบเครื่องหมายวรรคตอนและอักขระ (Punctuation removal) 46
3.3.6 การแบ่งคำ (Word tokenization) 46
3.3.7 การลบคำหยุด (Stop word removal) 47
3.3.8 การลดรูปของคำให้อยู่ในรากศัพท์ (Word Stemming and lemmatization) 47
3.3.9 การวิเคราะห์ชนิดของคำ (Parts of speech tagging) 47
3.3.10 การระบุคำศัพท์เฉพาะทางด้านคอมพิวเตอร์เทคโนโลยี(Reserved word identification) 48
3.3.11 การจับคู่คำพ้องความหมายและการเปลี่ยนแปลง (Synonym matching and transformation) 48
3.3 การจัดโครงสร้างและการรวบรวมเนื้อหา 49
3.3.1 การกำหนดโครงสร้างการจัดเก็บคำสำคัญจากข้อมูลประเภทของหน้าที่ความรับผิดชอบ 52
3.3.2 การกำหนดโครงสร้างการจัดเก็บคำสำคัญจากข้อมูลประเภทของอายุ 53
3.3.3 การกำหนดโครงสร้างการจัดเก็บคำสำคัญจากข้อมูลประเภทของเพศ 55
3.3.4 การกำหนดโครงสร้างการจัดเก็บคำสำคัญจากข้อมูลประเภทของสัญชาติ 56
3.3.5 การกำหนดโครงสร้างการจัดเก็บคำสำคัญจากข้อมูลประเภทของการศึกษาที่ต้องการ57
3.3.6 การกำหนดโครงสร้างการจัดเก็บคำสำคัญจากข้อมูลประเภทของประสบการณ์ที่ต้องการ 59

3.3.7 การกำหนดโครงสร้างการจัดเก็บคำสำคัญจากข้อมูลประเภทของทักษะที่จำเป็น 61
3.4 การส่งข้อมูลให้ผู้ใช้ (User) 64
บทที่ 4 ผลการดำเนินงาน 65
4.1 ความถูกต้องในการจัดกลุ่มประกาศงาน 66
4.2 ความแม่นยำในการสกัดคำสำคัญ 67
4.3 ความถูกต้องในการจัดประเภทและสร้างโครงสร้างข้อมูลที่สำคัญ 68
4.4 การนำข้อมูลจากฐานข้อมูลมาแสดงผล 69
4.4.1 เครื่องมือที่ใช้ 69
4.4.2 การแสดงผล 69
บทที่ 5 สรุปและอภิปลายผล 73
5.1 สรุปผลการดำเนินงาน 73
5.2 ข้อเสนอแนะ 74
บรรณานุกรม 75
ประวัติย่อของผู้วิจัย 205

สารบัญตาราง

หน้า

ตารางที่ 1 การจัดอันดับของเว็บไซต์และการสำรวจข้อมูลประกาศรับสมัครงานที่เผยแพร่บนเว็บไซต์ต่าง ๆ เบื้องต้น 4

ตารางที่ 2 ระยะเวลาในการดำเนินงานวิจัย 5

ตารางที่ 3 ชนิดและหน้าที่ของคำ 12

ตารางที่ 4 ตารางงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ประกาศงาน 17

ตารางที่ 5 โครงสร้างข้อมูลของเว็บไซต์ 24

ตารางที่ 6 แสดงคำเป้าหมายและรูปแบบของคำที่ใช้ในกฎการจำแนกประเภทของประโยค 37

ตารางที่ 7 ตัวอย่างของการแปลงคำย่อ/ชื่อย่อให้อยู่ในรูปแบบของคำเต็ม/ชื่อเต็ม 44

ตารางที่ 8 ตัวอย่างแสดงคำศัพท์ที่กำหนดช่วงอายุและการแทนค่าด้วยสัญลักษณ์ 53

ตารางที่ 9 แสดงคำที่บ่งบอกถึงระดับความสามารถของทักษะที่มักปรากฏขึ้นก่อนหน้าคำที่บ่งบอกถึงลักษณะของทักษะ 62

ตารางที่ 10 จำนวนประกาศงานของแต่ละกลุ่มงานจาก 3 เว็บไซต์ 65

ตารางที่ 11 จำนวนประกาศงานที่ถูกต้องและจำนวนประกาศงานที่ถูกจัดกลุ่มได้และค่าความถูกต้อง 66

ตารางที่ 12 จำนวนคำสำคัญที่ถูกต้องและจำนวนสำคัญที่สกัดได้ถูกต้องและค่าความถูกต้องในการสกัดคำสำคัญ 67

ตารางที่ 13 จำนวนประโยคที่จัดประเภทที่ถูกต้องและจำนวนประโยคที่ถูกจัดประเภทได้ถูกต้องและค่าความถูกต้อง 68

ตารางที่ 14 กฎทางภาษาศาสตร์ที่ไม่พิจารณาร่วมกับคำศัพท์เฉพาะและไม่มีคำกริยาอยู่ในประโยคสำหรับวิธีการ CID-CJA 79

ตารางที่ 15 กฎทางภาษาศาสตร์ที่ไม่พิจารณาร่วมกับคำศัพท์เฉพาะและแต่มีคำกริยาอยู่ในประโยคสำหรับวิธีการ CID-CJA 82

ตารางที่ 16 กฎทางภาษาศาสตร์ที่พิจารณาร่วมกับคำศัพท์เฉพาะแต่ไม่มีคำกริยาอยู่ในประโยค สำหรับวิธีการ CID-CJA 90

ตารางที่ 17 กฎทางภาษาศาสตร์ที่พิจารณาร่วมกับคำศัพท์เฉพาะและมีคำกริยาอยู่ในประโยคสำหรับ วิธีการ CID-CJA 110

ตารางที่ 18 โครงสร้างการจัดเก็บคำสำคัญจากข้อมูลประเภทของอายุ 130

ตารางที่ 19 โครงสร้างการจัดเก็บคำสำคัญจากข้อมูลประเภทของเพศ 133

ตารางที่ 20 โครงสร้างการจัดเก็บคำสำคัญจากข้อมูลประเภทของสัญชาติ 134

ตารางที่ 21 กฎทางภาษาศาสตร์สำหรับสาขาวิชาสำหรับประโยคที่ถูกจำแนกประเภทเป็นการศึกษา 134

ตารางที่ 22 โครงสร้างการจัดเก็บคำสำคัญจากข้อมูลประเภทของการศึกษาที่ต้องการ 135

ตารางที่ 23 โครงสร้างการจัดเก็บคำสำคัญจากข้อมูลประเภทของประสบการณ์ที่ต้องการ 179

ตารางที่ 24 โครงสร้างการจัดเก็บคำสำคัญจากข้อมูลประเภทของทักษะที่จำเป็น 187

สารบัญภาพ

หน้า

ภาพที่ 1 เส้นทางการเชื่อมโยงของการลงประกาศรับสมัครงานในตำแหน่งต่าง ๆ ของแต่ละนายจ้างบนเว็บไซต์ต่าง ๆ 1

ภาพที่ 2 ตัวอย่างการแบ่งประโยคและการแบ่งคำจากคำอธิบายของหน้าที่ความรับผิดชอบในตำแหน่งงานหนึ่ง ๆ 9

ภาพที่ 3 ตัวอย่างการการตรวจสอบและแก้ไขการสะกดคำจากคำอธิบายของหน้าที่ความรับผิดชอบในตำแหน่งงานหนึ่ง ๆ 9

ภาพที่ 4 กรอบงานการวิเคราะห์และสรุปความรับผิดชอบและข้อกำหนดจากประกาศรับสมัครงานทางด้านเทคโนโลยีสารสนเทศ 23

ภาพที่ 5 แสดงการนำเสนอโครงสร้างข้อมูลบนเว็บไซต์ https://th.jobsdb.com/ ในส่วนที่ต้องการสกัด 26

ภาพที่ 6 แสดงการนำเสนอโครงสร้างข้อมูลบนเว็บไซต์ https://www.linkedin.com/ ในส่วนที่ต้องการสกัด 27

ภาพที่ 7 แสดงการนำเสนอโครงสร้างข้อมูลบนเว็บไซต์ https://th.indeed.com/ ในส่วนที่ต้องการสกัด 28

ภาพที่ 8 ตัวอย่างข้อมูลประกาศงานที่ถูกจัดเก็บลงในฐานข้อมูลที่จัดทำขึ้นเป็นคลังข้อมูลเพื่อจัดเก็บข้อมูล ประกาศรับสมัครงาน 29

ภาพที่ 9 แสดงการส่วนข้อมูลที่ต้องการสกัดเพื่อรวบรวมรายการคำย่อและคำเต็มจากเว็บไซต์วิกิพีเดีย 31

ภาพที่ 10 ตัวอย่างเลือกรายการคำกริยาที่มีความหมายคล้ายคลึงหรือเหมือนกันของคำกริยาตั้งต้นหนึ่ง ๆ พิจารณาเลือกคำกริยาที่มีค่าความคล้ายคลึงหรือเหมือนกันที่มากกว่า 0.6 34

ภาพที่ 11 ตัวอย่างการจัดกลุ่มประกาศงานจนได้ชื่อกลุ่มตำแหน่งงาน 40

ภาพที่ 12 ตัวอย่างส่วนหน้าที่ความรับผิดชอบ และคุณสมบัติที่ต้องการจากประกาศงาน 42

ภาพที่ 13 ขั้นตอนการประมวลผลข้อความเบื้องต้นจากข้อมูลตัวอย่างส่วนหน้าที่ความรับผิดชอบและคุณสมบัติที่ต้องการจากประกาศงาน 43

ภาพที่ 14 ตัวอย่างการติดป้ายกำกับประเภทที่เกี่ยวข้องให้กับแต่ละประโยค โดยใช้กฎการจำแนกประเภทที่กำหนดไว้ 49

ภาพที่ 15 ตัวอย่างประโยคที่ถูกจำแนกประเภทเป็นประโยคหน้าที่ความรับผิดชอบและถูกจัดเก็บคำสำคัญให้อยู่ในรูปแบบโครงสร้างที่กำหนด 53

ภาพที่ 16 ตัวอย่างผลลัพธ์จากการจัดเก็บข้อมูลจากประโยคที่ถูกจำแนกให้อยู่ในประเภทอายุและถูกจัดเก็บคำสำคัญให้อยู่ในรูปแบบโครงสร้างที่กำหนด 55

ภาพที่ 17 ตัวอย่างผลลัพธ์จากการจัดเก็บข้อมูลจากประโยคที่ถูกจำแนกให้อยู่ในประเภทเพศและถูกจัดเก็บคำสำคัญให้อยู่ในรูปแบบโครงสร้างที่กำหนด 56

ภาพที่ 18 ตัวอย่างผลลัพธ์จากการจัดเก็บข้อมูลจากประโยคที่ถูกจำแนกให้อยู่ในประเภทสัญชาติให้อยู่ในโครงสร้างที่กำหนด 57

ภาพที่ 19 ตัวอย่างผลลัพธ์จากการจัดเก็บข้อมูลจากประโยคที่ถูกจำแนกให้อยู่ในประเภทการศึกษาและถูกจัดเก็บคำสำคัญให้อยู่ในรูปแบบโครงสร้างที่กำหนด 59

ภาพที่ 20 แสดงผลลัพธ์จากการจัดเก็บข้อมูลจากประโยคหนึ่ง ๆ ซึ่งอยู่ในประเภทข้อมูลประสบการณ์ที่ต้องการและถูกจัดเก็บคำสำคัญให้อยู่ในรูปแบบโครงสร้างที่กำหนด 61

ภาพที่ 21 ตัวอย่างผลลัพธ์จากการจัดเก็บข้อมูลจากประโยคที่ถูกจำแนกให้อยู่ในประเภททักษะที่จำเป็นและถูกจัดเก็บคำสำคัญให้อยู่ในรูปแบบโครงสร้างที่กำหนด 63

ภาพที่ 22 หน้ารายงานการแสดงผล 70

ภาพที่ 23 หน้ารายงานการแสดงผลที่ถูกระบุตำแหน่ง “Data Scientist” 70

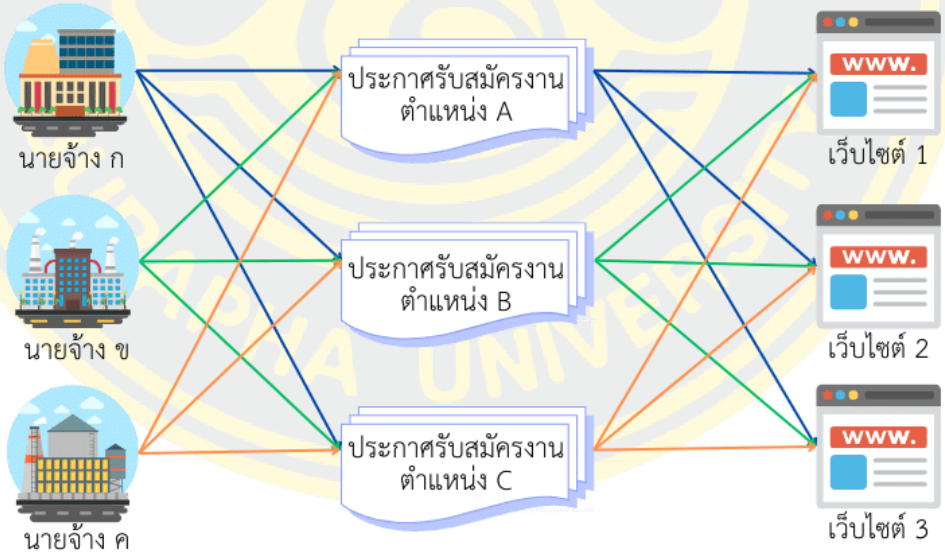
ภาพที่ 24 ข้อมูลในหน้ารายงานผลส่วนตารางแสดงจำนวนของระดับทักษะที่พบจาก 518 ประกาศงาน 71

ภาพที่ 25 ข้อมูลในหน้ารายงานผลส่วนตารางแสดงจำนวนของภาระงานพบจาก 518 ประกาศงาน72

บทที่ 1 บทนำ

1.1 ที่มาและความสำคัญของงานวิจัย

ความก้าวหน้าอย่างรวดเร็วด้านเทคโนโลยีได้ส่งผลกระทบต่อการศึกษา การพัฒนา และการดำเนินงานขององค์กรต่างๆ รวมถึงตลาดงานในสายงานที่เกี่ยวข้องกับเทคโนโลยีเป็นอย่างมาก การเปลี่ยนแปลงที่รวดเร็วในเทคโนโลยีและแบบจำลองธุรกิจที่กำลังเกิดขึ้นได้สร้างความต้องการที่เปลี่ยนแปลงตลอดเวลา ซึ่งในการทราบเกี่ยวกับแนวโน้มและความต้องการของอุตสาหกรรมเป็นเรื่องที่ทุกคนในสายงานนี้ต้องมีการตอบสนอง การติดตามข่าวสารเกี่ยวกับความต้องการทางสายงานอาชีพและแนวโน้มของอุตสาหกรรมที่กำลังเปลี่ยนแปลงอาจเป็นเรื่องที่ยุ่งยาก เนื่องจากข้อมูลมีความหลากหลาย ดังที่แสดงในภาพที่ 1 แสดงให้เห็นถึงเส้นทางการเชื่อมโยงของการลงประกาศรับสมัครงานในตำแหน่งต่าง ๆ ของแต่ละนายจ้างบนเว็บไซต์ต่าง ๆ



ภาพที่ 1 เส้นทางการเชื่อมโยงของการลงประกาศรับสมัครงานในตำแหน่งต่าง ๆ ของแต่ละนายจ้างบนเว็บไซต์ต่าง ๆ

จากภาพที่ 1 จะพบว่านายจ้าง ก,ข และ ค สามารถประกาศรับสมัครงานตำแหน่ง A, B และ C ลงในเว็บไซต์ 1, 2 และ 3 ได้เช่นเดียวกัน และในส่วนประกาศรับสมัครงานตำแหน่งที่ A, B และ C นั้น นายจ้าง ก, ข และ ค สามารถลงประกาศรับสมัครงานในตำแหน่งงานเดียวกันบนเว็บไซต์ที่แตกต่าง

กันได้เช่นกัน จากการที่แต่ละนายจ้างลงประกาศรับสมัครงานแต่ละตำแหน่งงานในเว็บไซต์ต่าง ๆ ซึ่งในแต่ละเว็บไซต์มีโครงสร้างการแสดงข้อมูลการประกาศรับสมัครงานที่แตกต่างกัน และรูปแบบการเขียนรายละเอียดการรับสมัครงานในตำแหน่งเดียวกันของแต่ละนายจ้างก็มีความแตกต่างกันเช่นกัน ซึ่งอาจเป็นสาเหตุที่ทำให้ผู้สมัครงาน, นักศึกษา, สถานศึกษา หรือองค์กรที่กำลังพิจารณาประกาศงานเกิดความสับสนต่อการมุ่งเน้นวัตถุประสงค์ที่สำคัญ อาจทำให้ผู้สมัครงานต้องใช้เวลาในการทำความเข้าใจต่อความต้องการในรายละเอียดงาน ทำให้มีความกังวลว่าไม่ทราบรายละเอียดงานเพียงพอเกี่ยวกับประเด็นต่าง ๆ ส่วนนักศึกษาอาจมีความสับสนในการเลือกสาขาการศึกษาหรือหลักสูตรที่เหมาะสมสำหรับโอกาสในการทำงานในอนาคต ส่วนสถานศึกษาอาจผลิตบุคลากรที่ความสามารถไม่ตอบโจทย์ต่อความต้องการของอุตสาหกรรมอาจต้องปรับหลักสูตรการเรียนการสอนให้สอดคล้องกับความต้องการที่เปลี่ยนแปลงในอุตสาหกรรม ส่วนนายจ้างที่กำลังพิจารณาลงรายละเอียดประกาศการจ้างงานอาจต้องพิจารณาข้อความในประกาศงาน เพื่อตัดสินใจเกี่ยวกับภาระงาน ประสบการณ์ และความสามารถของผู้สมัครให้ตอบโจทย์ต่อความต้องการที่องค์กรต้องการ

จากปัญหาข้างต้นที่กล่าวมา งานวิจัยนี้จึงมุ่งเน้นดำเนินการวิจัยในการวิเคราะห์และสรุปความรับผิดชอบ และข้อกำหนดจากข้อมูลในประกาศรับสมัครงานทางด้านเทคโนโลยีสารสนเทศ นำเสนอวิธีการค้นพบข้อมูลที่เกิดขึ้นร่วมกันในการประกาศรับสมัครงานที่เกี่ยวข้องกับคอมพิวเตอร์ (Co-occurrence Information Discovery in Computer-related Job Advertisements, CID-CJA) เป็นวิธีที่สามารถสกัดข้อมูลที่มีประโยชน์จากเนื้อหาในส่วนหน้าที่ความรับผิดชอบ และส่วนคุณสมบัติที่ต้องการจากประกาศรับสมัครงานให้อยู่ในรูปแบบโครงสร้างข้อมูลที่เหมาะสม โดยมีทั้งหมด 6 รูปแบบโครงสร้าง ประกอบด้วย หน้าที่ความรับผิดชอบ เพศ อายุ การศึกษา ประสบการณ์ ทักษะ ซึ่งวิธีการนี้มีความยืดหยุ่นที่ช่วยให้ผู้ใช้งานสามารถทำการวิเคราะห์ข้อมูลได้อย่างลึกลงในส่วนของข้อมูลที่เฉพาะเจาะจง และสามารถปรับระดับความละเอียดของข้อมูลได้ตามความต้องการ เพื่อช่วยส่งเสริมให้ผู้ที่กำลังมองหางานสามารถเข้าใจได้อย่างชัดเจนถึงหน้าที่ความรับผิดชอบ ทักษะ และคุณสมบัติที่ต้องการ และช่วยให้สถานศึกษาสามารถปรับหลักสูตรการเรียนการสอนให้สอดคล้องกับความต้องการที่เปลี่ยนแปลงไปตามรายละเอียดของอุตสาหกรรม สามารถผลิตบุคลากรได้อย่างมีประสิทธิภาพ และเพื่อช่วยนายจ้างสร้างคำอธิบายงานตรงกับความต้องการของอุตสาหกรรม ให้นายจ้างและผู้สมัครงานทราบถึงข้อมูลที่มีค่าสำหรับช่วยในการตัดสินใจและการวางแผนในอนาคตของพวกเขา

1.2 วัตถุประสงค์ของงานวิจัย

1. เพื่อศึกษาและค้นหาแนวทางวิธีการจัดรูปแบบโครงสร้างให้กับข้อมูลที่ไม่มีโครงสร้างจากประกาศรับสมัครงานในศาสตร์ทางด้านเทคโนโลยีสารสนเทศจากเว็บไซต์รับสมัครงานภายในประเทศไทย
2. เพื่อสร้างวิธีการสำหรับจัดรูปแบบโครงสร้างข้อมูลที่เกี่ยวข้องกับหน้าที่ความรับผิดชอบ และคุณสมบัติที่ต้องการในตำแหน่งงานต่าง ๆ ที่เกี่ยวข้องกับศาสตร์ทางด้านเทคโนโลยีสารสนเทศจากเว็บไซต์รับสมัครงานภายในประเทศไทย
3. เพื่อสร้างข้อสรุปของหน้าที่ความรับผิดชอบ และคุณสมบัติที่ต้องการในตำแหน่งงานต่าง ๆ ที่เกี่ยวข้องกับศาสตร์ทางด้านเทคโนโลยีสารสนเทศจากเว็บไซต์รับสมัครงานภายในประเทศไทย

1.3 ขอบเขตของงานวิจัย

1. ข้อมูลจากประกาศรับสมัครงานที่นำมาวิเคราะห์ในการสร้างวิธีการสำหรับจัดรูปแบบโครงสร้างข้อมูลของหน้าที่ความรับผิดชอบ และคุณสมบัติที่ต้องการในตำแหน่งงานต่าง ๆ เป็นสายงานอาชีพที่เกี่ยวข้องกับศาสตร์ทางด้านเทคโนโลยีสารสนเทศ
2. จากการพิจารณาเพื่อเลือกเว็บไซต์ประกาศรับสมัครงาน เพื่อนำมาเป็นข้อมูลที่ใช้ในการวิเคราะห์ เว็บไซต์ที่เลือกพิจารณาคือ เว็บไซต์ jobsDB, Linkedin และ Indeed ซึ่งอยู่ในอันดับ top 1 2 และ 3 จาก 10 เว็บไซต์ที่ถูกรวบรวมไว้ โดยทั้ง 3 เว็บไซต์เป็นเว็บไซต์ที่เกี่ยวข้องกับการค้นหางานภายในประเทศไทย โดยผู้วิจัยจะพิจารณาเฉพาะข้อมูลประกาศงานที่เป็นภาษาอังกฤษเท่านั้น โดยข้อมูลประกาศงานที่รวบรวมได้ ผู้วิจัยจะพิจารณาเฉพาะข้อมูล i) ชื่อตำแหน่งงาน ii) หน้าที่ความรับผิดชอบ iii) คุณสมบัติที่ต้องการ เท่านั้น เนื่องจากผู้วิจัยได้ดำเนินการสำรวจและพิจารณาโครงสร้างของเว็บไซต์ ซึ่งแสดงรายละเอียดในบทที่ 3 ตารางที่ 1 ทำให้พบว่าทั้ง 3 ส่วนนี้ มีส่วนเนื้อหาอยู่บนหน้าเว็บไซต์ของทั้ง 3 แหล่งเว็บไซต์ และทั้ง 3 ส่วนเป็นเนื้อหาที่สามารถนำมาวิเคราะห์มีความชัดเจน และเป็นส่วนสำคัญของเนื้อหาของเว็บไซต์ ซึ่งผู้วิจัยได้ดำเนินการสำรวจการจัดอันดับเว็บไซต์จาก https://www.alexa.com/siteinfo ซึ่งเป็นเว็บไซต์ที่อยู่ภายใต้การบริหารของเว็บอีคอมเมิร์ซรายใหญ่ Amazon.com โดย Alexa ได้ทำการจัดอันดับและทำข้อมูลสถิติของเว็บไซต์ต่าง ๆ ซึ่งนำข้อมูลมาจากการใช้งานเว็บไซต์ของผู้ใช้อินเทอร์เน็ต ซึ่งแสดงรายละเอียดในตารางที่ 1

ตารางที่ 1 การจัดอันดับของเว็บไซต์และการสำรวจข้อมูลประกาศรับสมัครงานที่เผยแพร่บนเว็บไซต์ต่าง ๆ เบื้องต้น

อันดับของเว็บไซต์	**ชื่อเว็บไซต์**	**ข้อมูลประกาศงาน**	**ประกาศงานในประเทศไทย**
57	linkedin.com	ส่วนใหญ่เป็นภาษาอังกฤษ	✓
105	indeed.com	ส่วนใหญ่เป็นภาษาอังกฤษ บางประกาศงานมีภาษาไทยผสมภาษาอังกฤษ	✓
5,029	jobsdb.com	ส่วนใหญ่เป็นภาษาอังกฤษ	✓
17,099	jobthai.com	ส่วนใหญ่เป็นภาษาไทย บางประกาศงานมีภาษาไทยผสมภาษาอังกฤษ	✓
36,500	jobbkk.com	ส่วนใหญ่เป็นภาษาไทย	✓
64,741	jobth.com	ส่วนใหญ่เป็นภาษาไทย ข้อมูลที่ผู้วิจัยต้องการเก็บไม่ครบ เช่น วันที่ลงประกาศ	✓
70,257	jobtopgun.com	ส่วนใหญ่เป็นภาษาไทย บางประกาศงานมีภาษาไทยผสมภาษาอังกฤษ	✓
177,620	workventure.com	ส่วนใหญ่เป็นภาษาไทย บางประกาศงานมีภาษาไทยผสมภาษาอังกฤษ	✓
4,580,808	thaibestjobs.com	ส่วนใหญ่เป็นภาษาไทย	✓
7,406,118	jobkonthai.com	ส่วนใหญ่เป็นภาษาไทย	✓

1.4 ขั้นตอนและระยะเวลาในการดำเนินงานวิจัย

ตารางที่ 2 ระยะเวลาในการดำเนินงานวิจัย

กิจกรรม	ระยะเวลาตามปีการศึกษาและภาคการศึกษา							
	2/ 62	1/ 63	2/ 63	1/ 64	2/64	1/65	2/65	1/66
1. ศึกษาและสำรวจเพื่อทำความเข้าใจกับข้อมูลประกาศรับสมัครงานในเว็บไซต์ต่าง ๆ	◆	——◆						
2. ศึกษางานวิจัยที่เกี่ยวข้องจากเอกสาร หนังสือและงานวิจัยที่มีการตีพิมพ์เผยแพร่ในแหล่งข้อมูลที่มีความน่าเชื่อถือ	◆	——◆						
3. พิจารณาข้อมูลประกาศรับสมัครงานและกำหนดโครงสร้างในการจัดเก็บข้อมูล	◆	——◆						
4. กำหนดวัตถุประสงค์และขอบเขตงานวิจัยและรวบรวมข้อมูลประกาศรับสมัครตำแหน่งงานในศาสตร์ทางด้านเทคโนโลยีสารสนเทศจากเว็บไซต์และเตรียมข้อมูลเพื่อใช้ในการทดลอง	◆	———	◆					
5. ศึกษาและออกแบบวิธีการในการวิเคราะห์เพื่อให้ได้ข้อสรุปจากเนื้อหาในประกาศรับสมัครงานของตำแหน่งงานหนึ่ง ๆ		◆	——◆					
6. ทดสอบและตรวจสอบผลลัพธ์ที่ได้		◆	——◆					
7. จัดทำเอกสารโครงร่างวิทยานิพนธ์และสอบโครงร่างวิทยานิพนธ์			◆	◆				
8. ดำเนินการเสนอขอพิจารณาจริยธรรมการวิจัย				◆				

กิจกรรม	ระยะเวลาตามปีการศึกษาและภาคการศึกษา							
	2/ 62	1/ 63	2/ 63	1/ 64	2/64	1/65	2/65	1/66
9. ตรวจสอบปัญหาความถูกต้องและแก้ไขปรับปรุงวิธีการในการวิเคราะห์เพื่อให้ได้ข้อสรุปจากเนื้อหาในประกาศรับสมัครงานของตำแหน่งงานหนึ่ง ๆ				◆	—	—	◆	
10. ทดสอบและตรวจสอบผลลัพธ์ที่ได้และประเมินประสิทธิภาพและแก้ไขปรับปรุงเพื่อให้มีประสิทธิภาพดีขึ้น				◆	—	—	—	◆
11. วิเคราะห์และสรุปผลลัพธ์ที่ได้						◆	—	◆
12. จัดทำบทความวิจัยเกี่ยวกับวิธีการในการวิเคราะห์เพื่อหา ข้อสรุปจากเนื้อหาในประกาศรับสมัครงานของตำแหน่งงานหนึ่ง ๆ เพื่อตีพิมพ์ในการประชุมวิชาการระดับนานาชาติ							◆	— ◆
13. เผยแพร่งานวิจัย								◆ — ◆
14.จัดทำเอกสารวิทยานิพนธ์และสอบวิทยานิพนธ์								◆ — ◆

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. ได้ขั้นตอนวิธีการสำหรับจัดรูปแบบโครงสร้างข้อมูลของหน้าที่ความรับผิดชอบ และคุณสมบัติที่ต้องการในตำแหน่งงานต่าง ๆ ที่เกี่ยวข้องกับศาสตร์ทางด้านเทคโนโลยีสารสนเทศจากเว็บไซต์รับสมัครงานภายในประเทศไทย
2. ได้ฐานข้อมูลที่จัดเก็บข้อมูลจากประกาศรับสมัครงานบนเว็บไซต์ต่าง ๆ
3. ได้ข้อสรุปของหน้าที่ความรับผิดชอบ และคุณสมบัติที่ต้องการในแต่ละตำแหน่งงานต่าง ๆ ในศาสตร์ทางด้านเทคโนโลยีสารสนเทศภายในประเทศไทยที่เป็นภาษาอังกฤษ

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

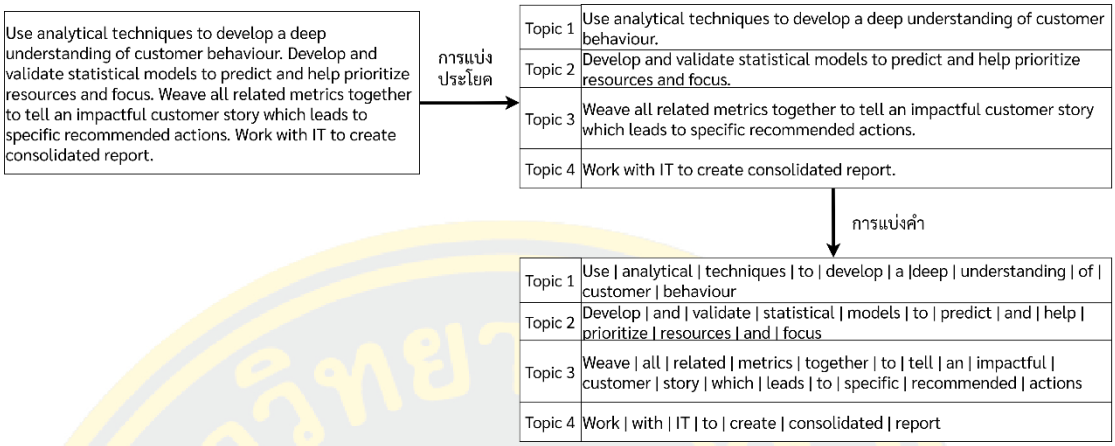
บทนี้จะเป็นการอธิบายถึงทฤษฎีและงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์และหาข้อสรุปรายละเอียดของการประกาศรับสมัครงานจากเว็บไซต์สำหรับประกาศรับสมัครงานต่าง ๆ โดยผู้วิจัยจะอธิบายวิธีการที่ได้นำมาประยุกต์ใช้ดังต่อไปนี้

2.1 การประมวลผลข้อความเบื้องต้น (Text preprocessing)

การประมวลผลข้อความเบื้องต้นเป็นกระบวนการในการจัดการกับข้อความให้อยู่ในรูปแบบที่คอมพิวเตอร์สามารถนำไปวิเคราะห์และประมวลผลเชิงลึกได้ได้อย่างมีประสิทธิภาพ โดยในส่วนของการประมวลผลข้อความเบื้องต้นที่ผู้วิจัยนำมาประยุกต์ใช้ในการดำเนินงานนี้ประกอบไปด้วย การแบ่งประโยค/คำ, การตรวจสอบและแก้ไขการสะกดคำ, การแปลงคำย่อ, การลบคำหยุด, การแปลงตัวอักษรให้เป็นตัวพิมพ์เล็ก, การลบเครื่องหมายวรรคตอนและอักขระ, การลดรูปของคำให้อยู่ในรากศัพท์ และการวิเคราะห์หน้าที่ของคำ

2.1.1 การแบ่งประโยค/คำ (Sentence/Word tokenization)

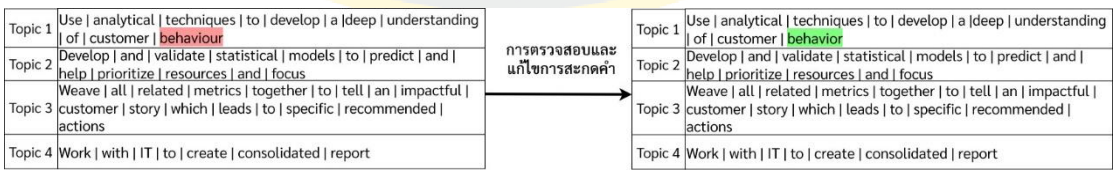
ในส่วนของกระบวนนี้สามารถแบ่งออกได้เป็น 2 ส่วน คือ I.) การแบ่งประโยคคือกระบวนการในการแบ่งประโยคแต่ละประโยคในเอกสารหนึ่ง ๆ ออกจากกัน โดยจะทำการพิจารณาถึงจุด (“.”) หรืออาจจะเป็นอัฒภาค (“;”) ที่อยู่ในส่วนท้ายของแต่ละประโยค และ II.) การแบ่งคำคือกระบวนการในการแบ่งคำที่อยู่ในประโยคออกจากกัน โดยจะพิจารณาถึงช่องว่างระหว่างคำ จากนั้นทำการแบ่งคำแต่ละคำออกจากกัน ซึ่งในงานวิจัยนี้ผู้วิจัยได้นำทั้งส่วนของการแบ่งประโยคมาใช้ในการแบ่งแต่ละประโยคในแต่ละตำแหน่งงานออกจากกัน จากนั้นนำประโยคที่แบ่งแล้วมาทำการแบ่งคำแต่ละคำออกจากกัน โดยมีตัวอย่างของการแบ่งประโยค/คำ ดังตัวอย่างในภาพที่ 3 จะประกอบด้วยข้อความ 1 ย่อหน้าที่มี 4 ประโยค สามารถทำการแบ่งประโยคจากจุด (“.”) ได้เป็นประโยคแต่ละประโยคที่เป็นอิสระต่อกันได้ จากนั้นพิจารณาแต่ละประโยคเช่น ประโยคที่ 1) “Use analytical techniques to develop a deep understanding of customer behavior.” แล้วทำการแยกคำในประโยคได้เป็น “use | analytical | techniques | to | develop | a | deep | understanding | of | customer | behavior”



ภาพที่ 2 ตัวอย่างการแบ่งประโยคและการแบ่งคำจากคำอธิบายของหน้าที่ความรับผิดชอบในตำแหน่งงานหนึ่ง ๆ

2.1.2 การตรวจสอบและแก้ไขการสะกดคำ (Word error correction)

เป็นกระบวนการตรวจสอบและแก้ไขการสะกดคำที่อยู่ในข้อความ จะตรวจสอบความถูกต้องของการสะกดคำโดยการเปรียบเทียบความเหมือนกันของสายอักขระระหว่างคำที่สะกดผิดกับคำศัพท์ที่อยู่ในคลังข้อมูลภาษาอังกฤษที่เป็นทางการ/พจนานุกรม ซึ่งวิธีการตรวจสอบความถูกต้องของการสะกดคำที่ได้รับความนิยม คือ วิธีการตรวจสอบความถูกต้องของการสะกดคำแบบ N-gram เช่น Trigram เป็นต้น และทำการคำนวณหาความน่าจะเป็น (probability) จากการเปรียบเทียบกันระหว่างคำที่มีการสะกดผิดกับคำศัพท์ที่มีอยู่ จากนั้นทำการแทนที่คำที่มีการสะกดผิดด้วยคำศัพท์ที่ถูกต้องจากการพิจารณาคำศัพท์ที่มีค่าความน่าจะเป็นสูงที่สุด โดยมีตัวอย่างของการตรวจสอบและแก้ไขการสะกดคำดังภาพที่ 3 ในประโยคที่ 1 เมื่อทำกระบวนการตรวจสอบและแก้ไขการสะกดคำพบว่ามีคำที่สะกดผิด คือ “behaviour” จะถูกเปลี่ยนเป็น “behavior”



ภาพที่ 3 ตัวอย่างการการตรวจสอบและแก้ไขการสะกดคำจากคำอธิบายของหน้าที่ความรับผิดชอบในตำแหน่งงานหนึ่ง ๆ

2.1.3 การแปลงคำย่อ/ชื่อย่อ (Abbreviation resolution)

เป็นกระบวนในการตรวจสอบและระบุถึงคำย่อ/ชื่อย่อที่ปรากฏขึ้นในประโยค จากนั้นทำการแปลงคำย่อ/ชื่อย่อที่พบให้อยู่ในรูปแบบของคำเต็ม/ชื่อเต็มของคำคำนั้น โดยในส่วนของงานวิจัยนี้ผู้วิจัยจะทำการพิจารณาถึงคำย่อ/ชื่อย่อที่เกี่ยวข้องกับศาสตร์ทางด้านคอมพิวเตอร์เทคโนโลยีสารสนเทศเท่านั้น (list of computing and IT acronyms and abbreviations) ตัวอย่างเช่น คำย่อ “GCP” จะถูกดำเนินการแปลงเปลี่ยนเป็นคำเต็มคือ “Google Cloud Platform”

2.1.4 การลบคำหยุด (Stop Word Removal)

เป็นกระบวนในการลบคำหยุดออกจากประโยคหรือข้อความเพื่อลดความหนาแน่นของข้อความและเพิ่มความชัดเจนของข้อความที่เหลืออยู่ คำหยุดคือคำที่มักใช้บ่อยในภาษาแต่มีความหมายที่โดดเด่นน้อยหรือไม่มีความหมายเฉพาะตัวที่ส่งเสริมความเข้าใจเนื้อหา ตัวอย่างคำหยุดที่มักพบมีดังนี้ “a”, “an”, “the”, “is”, “are” และอื่น ๆ การกำจัดคำหยุดมักทำให้ประโยคหรือข้อความดูสั้นลงและเน้นไปที่คำหลักหรือความสำคัญของเนื้อหาที่เราต้องการสื่อสาร นี่คือตัวอย่างของการกำจัดคำหยุดในประโยคที่ 1 จากภาพที่ 2 คือ “Use analytical techniques to develop a deep understanding of customer behavior.” ซึ่งเป็นประโยคก่อนการกำจัดคำหยุด เมื่อประโยคผ่านกระบวนการหลังการกำจัดคำหยุดแล้วจะได้เป็น “Use analytical techniques develop deep understanding customer behavior.” ในกระบวนการนี้คำหยุดคือ “to” และ “a” จำถูกลบออก และประโยคย่อมเปลี่ยนไปเป็นข้อความที่เข้าใจง่ายและมีการเน้นคำหลักที่สำคัญมากขึ้น การกำจัดคำหยุดเป็นเครื่องมือที่มีประโยชน์ในการเพิ่มความชัดเจนและประสิทธิภาพในการเขียนและการสื่อสารเนื้อหา

2.1.5 การแปลงตัวอักษรให้เป็นตัวพิมพ์เล็ก (Lowercase conversion)

เป็นกระบวนการที่เปลี่ยนอักษรตัวพิมพ์ใหญ่ทั้งหมดในข้อความเป็นอักษรตัวพิมพ์เล็กทั้งหมด การทำนี้เป็นวิธีที่มักใช้ในประมวลผลข้อความเพื่อให้ข้อความดูเหมือนกันทั่วทั้งเอกสารและทำให้การค้นหาข้อมูลที่ตรงกับข้อความเป็นไปอย่างเท่าเทียม ตัวอย่างการแปลงตัวอักษรให้เป็นตัวพิมพ์เล็กในประโยคที่ 1 จากภาพที่ 2 ประโยคก่อนการแปลงตัวอักษรคือ “Use analytical techniques to develop a deep understanding of customer behavior.” ซึ่งเมื่อทำกระบวนการแปลงตัวอักษรเป็นตัวพิมพ์เล็กแล้วจะได้เป็น “use analytical techniques to develop a deep

understanding of customer behavior.” ในกระบวนการนี้ทุกตัวอักษรที่เป็นพิมพ์ใหญ่ในประโยคถูกเปลี่ยนเป็นอักษรตัวพิมพ์เล็ก ทำสิ่งนี้เพื่อให้ข้อความดูเรียบร้อย ทำให้การประมวลผลข้อความและการค้นหาข้อมูลเป็นไปอย่างเท่าเทียมและสะดวกมากขึ้น

2.1.6 การลบเครื่องหมายวรรคตอนและอักขระ (Punctuation removal)

เป็นกระบวนการที่มุ่งหมายที่จะลบอักขระและเครื่องหมายวรรคตอนออกจากข้อความหรือประโยคเพื่อทำให้ข้อความดูเรียบร้อย และเพิ่มความคงเดิมของข้อความ อักขระและเครื่องหมายวรรคตอนไม่มีความหมายทางวิทยาศาสตร์และไม่ส่งเสริมความเข้าใจของเนื้อหา ตัวอย่างของเครื่องหมายวรรคตอนและอักขระมีดังนี้ . , ? ! ;: - ( ) “ “ [ ] เป็นต้น การทำนี้จะทำให้ข้อความดูเรียบร้อยและเหมาะสม เพื่อสื่อสารเนื้อหาอย่างเข้าใจและเป็นระเบียบ กระบวนการลบเครื่องหมายวรรคตอนและอักขระมักใช้ในการประมวลผลข้อมูลที่เน้นความคงเดิมของข้อความและการค้นหาข้อมูลในการค้นหาเว็บไซต์และฐานข้อมูล และอื่น ๆ ที่ต้องการข้อมูลที่ดูเรียบร้อย

2.1.7 การลดรูปของคำให้อยู่ในรากศัพท์ (Stemming and lemmatization)

เป็นกระบวนการลดรูปคำศัพท์ให้อยู่ในรูปเดิมของคำหรือรูปแบบของรากศัพท์ของคำ ๆ นั้น โดยการลดรูปของคำให้อยู่ในรากศัพท์สามารถแบ่งออกได้เป็น 2 ส่วนคือ I.) Stemming เป็นกระบวนการเรียนรู้แบบฮิวริสติก (heuristic process) ซึ่งคำลงท้ายจะถูกตัดออก ตัวอย่างเช่น คำว่า “jumped” และ “jumps” จะถูกลดรูปไปเป็นรากศัพท์ซึ่งได้แก่คำว่า “jump” ในขณะที่คำว่า “jumpiness” อาจถูกแปลงเป็นคำว่า “jumpi” และ II.) Lemmatization เป็นกระบวนการที่ใช้คำศัพท์และการวิเคราะห์ทางสัณฐานวิทยา (morphological analysis) ของคำศัพท์คำนั้น ๆ เพื่อทำการลดรูปของคำศัพท์ให้ได้รูปแบบพื้นฐาน (รูปแบบพจนานุกรม) ตัวอย่างเช่น คำว่า “am”, “are”, “is”, “was” จะถูกลดรูปเป็นคำว่า “be” หรือ “saw”, “seen” จะถูกลดรูปเป็นคำว่า “see”

2.1.8 การวิเคราะห์ชนิดของคำ (Parts of speech tagging)

เป็นการวิเคราะห์คุณลักษณะทางไวยากรณ์ของคำแต่ละคำในประโยคหนึ่ง ๆ เพื่อแสดงหน้าที่ของคำหรือคลาสของคำ (word class) โดยจะบ่งบอกถึงคุณลักษณะและหน้าที่ของแต่ละคำ

ซึ่งเกณฑ์ในการจำแนกชนิดของคำคำนั้นจะแตกต่างกันตามลักษณะโครงสร้างทางไวยากรณ์ในแต่ละประโยค ซึ่งหน้าที่ของคำในภาษาอังกฤษมีทั้งหมด 8 ชนิด ดังตารางที่ 3 แสดงชนิดและหน้าที่ของคำ

ตารางที่ 3 ชนิดและหน้าที่ของคำ

ชนิดของคำ	หน้าที่ของคำ
1) คำนาม (Noun)	คำที่ใช้เรียกแทนชื่อคน, สัตว์, สิ่งของ, สถานที่
2) คำสรรพนาม (Pronoun)	คำที่ใช้แทนคำนาม
3) คำกริยา (Verb)	คำที่แสดงอาการหรือการกระทำในประโยค
4) คำคุณศัพท์ (Adjective)	คำที่ทำหน้าที่ขยายคำนาม โดยตำแหน่งจะอยู่หน้าคำนามเสมอ
5) คำกริยาวิเศษณ์ (Adverb)	คำที่มีหน้าที่ขยายกริยา ขยายคุณศัพท์ และขยายกริยาวิเศษณ์ด้วยกันเอง
6) คำบุพบท (Preposition)	คำที่ใช้บอกตำแหน่ง วันเวลา ทิศทาง สถานที่ หรือบอกความสัมพันธ์ระหว่างนามหรือสรรพนามกับคำอื่น ๆ
7) คำสันธาน (Conjunction)	คำที่ใช้เชื่อมคำกับคำ กลุ่มคำกับกลุ่มคำ หรือประโยคกับประโยค
8) คำอุทาน (Interjection)	คำที่ใช้บ่งบอกอารมณ์ ความรู้สึกที่เกิดขึ้นในตอนนั้น

ยกตัวอย่างเช่น จากรูปภาพที่ 3 ประโยคที่ 1 ที่ถูกแบ่งคำ “use | analytical | techniques | to | develop | a | deep | understanding | of | customer | behavior” เมื่อดำเนินการวิเคราะห์ชนิดคำจะได้เป็น (“use”, “VB”), (“analytical”, “JJ”), (“techniques”, “NNS”), (“to”, “TO”), (“develop”, “VB”), (“a”, “DT”), (“deep”, “JJ”), (“understanding”, “NN”), (“of”, “IN”), (“customer”, “NN”), (“behavior”, “NN”)

2.2 คำฝังตัว (Word Embeddings)

เป็นโมเดลที่มีความสำคัญในการประมวลผลภาษาธรรมชาติและการทำงานกับข้อมูลข้อความและได้รับความนิยมในวงการการประมวลผลภาษาธรรมชาติ (Natural language processing) และการเรียนรู้ของเครื่อง (Machine Learning) ซึ่งคำฝังตัว (Word Embedding) เป็นกระบวนการแปลงคำหรือข้อมูลข้อความให้อยู่ในรูปแบบที่เครื่องคอมพิวเตอร์เข้าใจได้ และนำไปใช้ในการประมวลผลข้อมูลอย่างเป็นประสิทธิภาพ โดยการนำเอาคำหรือข้อมูลข้อความเหล่านี้มาแปลงเป็น

ตัวเลข และเปรียบเทียบความคล้ายคลึงระหว่างคำหรือข้อมูลข้อความเหล่านี้สามารถวัดได้ด้วยการใช้ Cosine Similarity เป็นเทคนิคที่ใช้ในการวัดความคล้ายคลึงระหว่างความหมายของคำ เพื่อให้สามารถนำมาวิเคราะห์ความคล้ายคลึงระหว่างคำต่าง ๆ ในบริบทของข้อความได้อย่างมีประสิทธิภาพ ซึ่งจะช่วยในหลายงานเช่น ค้นหาข้อมูลที่เกี่ยวข้อง, การจัดหมวดหมู่ข้อมูล, และการเข้าใจความหมายของข้อความ โดยงานวิจัยนี้ผู้วิจัยนำเทคนิคการสร้างคำฝัง โดยใช้โมเดล Word2Vec, Glove, และ Cosine Similarity

2.2.1. Word2Vec

Word2Vec ซึ่งพัฒนาโดย Google โดยมี Mikolov เป็นหัวหน้าในการพัฒนา (Mikolov et al, 2013) เป็นโมเดลการเรียนรู้คำและสร้างเวกเตอร์ (vectors) คุณลักษณะสำหรับคำ หรือแปลงคำให้อยู่ในรูปแบบของเวกเตอร์ โดยใช้วิธีการโครงข่ายประสาทเทียม (Neural Networks) 2 ชั้น โดยมีการนำข้อมูลจำนวนมากมาใช้ในการฝึกสอนให้ได้เวกเตอร์คุณลักษณะออกมา โดยถูกปล่อยออกมาให้ได้ใช้งานครั้งแรกในปี 2013 ข้อมูลข้อความใหญ่จะช่วยให้โมเดลมีโอกาสเรียนรู้ความหมายของคำได้อย่างครอบคลุมและถูกฝึกสอนให้เหมาะสมในการสร้างเวกเตอร์คำที่แสดงความคล้ายคลึงทางความหมายของคำอย่างมีประสิทธิภาพ

2.2.2 GloVe (Global Vectors for Word Representation)

จากความสำเร็จในการสร้าง Word2Vec ขึ้น ทางมหาวิทยาลัย Stanford ได้ทำการพัฒนาต่อยอดโดยใช้ชื่อว่า GloVe หรือ Global Vectors (Pennington et al., 2014) เป็นโมเดลการเรียนรู้คำและสร้างเวกเตอร์คุณลักษณะของคำที่สร้างขึ้นโดยอาศัยการเรียนรู้คำจากข้อมูลข้อความขนาดใหญ่ โดยอาศัยความถี่ของคำคู่ และความถี่ของคำเริ่มต้นของคำคู่ในข้อความ โดยใช้หลักการของ Nearest Neighbors ในการนำ Feature ที่ได้ไปใช้ค้นหาคำที่มีความหมายใกล้เคียงกัน และ Linear Substructures เพื่อทำการค้นหาคำที่ใช้ในหมวดหมู่เดียวกัน หรือมีความสัมพันธ์กันอยู่ ทำให้สามารถสร้างเวกเตอร์ที่แสดงความสัมพันธ์ระหว่างคำได้อย่างแม่นยำ

2.2.3 Cosine Similarity

เป็นเทคนิคทางคณิตศาสตร์ที่ใช้ในการวัดความคล้ายคลึงระหว่างเวกเตอร์หรือคำของข้อมูล มันคำนวณค่าความคล้ายคลึงโดยวัดค่าคอสไทน์ของมุมระหว่างเวกเตอร์ค่า cosine similarity มีค่า

ระหว่าง -1 ถึง 1 โดยค่า 1 แสดงถึงความคล้ายคลึงสูงสุด, ค่า 0 แสดงถึงความคล้ายคลึงต่ำที่สุด และค่า -1 แสดงถึงความต่างกันสูงสุด

Cosine similarity มักถูกใช้ในงานค้นหาข้อมูลและการจัดเรียงข้อมูลเพื่อวัดความคล้ายคลึงระหว่างเวกเตอร์หรือเอกสาร

2.3 งานวิจัยที่เกี่ยวข้อง

2.3.1 การสรุปใจความสำคัญอัตโนมัติ (automatic text summarization)

งานวิจัยของการสรุปใจความสำคัญอัตโนมัติ (automatic text summarization) เริ่มต้นโดย (Luhn, 1958) นำเสนอวิธีการสรุปใจความสำคัญโดยมีแนวคิดว่าความถี่ของคำหนึ่ง ๆ ในเอกสารสามารถบ่งบอกถึงใจความสำคัญที่เป็นประโยชน์ Luhn. จึงได้พิจารณาประโยคในเอกสารและทำการรวบรวมคำในเอกสาร โดยแต่ละคำจะนับความถี่โดยเรียงจากคำที่มีความถี่จากน้อยไปมาก จากนั้นจะพิจารณาตัดคำที่ไม่สำคัญออก ดังนั้นจากประโยคทั้งหมดที่มีในเอกสาร ประโยคใดที่มีคำสำคัญที่มีความถี่มากที่สุดจะถูกเลือกเพื่อสร้างบทคัดย่อของเอกสารนั้น ๆ
ก่อให้เกิดความท้าทายในการวิจัยที่เกี่ยวข้องกับการสรุปใจความสำคัญอัตโนมัติเกิดขึ้น โดยการสรุปใจความสำคัญอัตโนมัติสามารถจัดเป็นการสรุปใจความสำคัญจากเอกสารหนึ่ง ๆ (single-document summarization) และ การสรุปใจความสำคัญจากหลายเอกสาร (multi-document summarization)

2.3.1.1 การสรุปใจความสำคัญจากเอกสารหนึ่ง ๆ (single-document summarization)

(Baxendale, 1958) ได้นำเสนอการพิจารณาคุณลักษณะที่เป็นประโยชน์ในการสรุปใจความสำคัญในย่อหน้าหนึ่ง ๆ ของบทความ ซึ่งทำการตรวจสอบข้อความ 200 ย่อหน้า พบว่า 85% ของย่อหน้าทั้งหมดที่นำมาพิจารณา ใจความสำคัญสามารถสรุปได้จากข้อความประโยคแรกของย่อหน้านั้น ๆ และ 7% ของย่อหน้าทั้งหมดที่นำมาพิจารณาใจความสำคัญสามารถสรุปได้จากข้อความประโยคสุดท้ายของย่อหน้านั้น ๆ ดังนั้น Baxendale. จึงได้นำเสนอวิธีการที่เรียบง่ายในการสรุปใจความสำคัญคือการเลือกประโยคแรกหรือประโยคสุดท้ายของข้อความในย่อหน้าหนึ่ง ๆ นอกจากนี้คุณลักษณะนี้ยังสามารถนำไปปรับใช้กับวิธีการทางการเรียนรู้ของเครื่องจักร (Machine learning) อีกด้วย

(Lin & Hovy, 1997) ได้ศึกษาเกี่ยวกับคุณลักษณะตำแหน่งของประโยคที่ปรากฏในย่อหน้าหนึ่ง ๆ ที่อยู่ในเอกสารหนึ่ง ๆ จากที่ (Baxendale., 1958) ได้นำเสนอวิธีการที่เรียบง่ายในการสรุปใจความสำคัญคือการเลือกประโยคแรกหรือประโยคสุดท้ายของข้อความในย่อหน้าหนึ่ง ๆ นั้นไม่เพียงพอ Lin และ Hovy. ได้พิจารณาการให้น้ำหนักของประโยคตามตำแหน่งของประโยคที่ปรากฏในย่อหน้าหนึ่ง ๆ และยังพิจารณาว่าข้อความในประโยคนั้น ๆ มีความสอดคล้องกับชื่อเรื่อง/ชื่อของเอกสารหรือไม่ด้วย ซึ่งเรียกว่า “Position method” จากนั้นจึงรวบรวมข้อความจากคลังข้อมูล TIPSTER ที่เกี่ยวกับคอมพิวเตอร์และฮาร์ดแวร์ ซึ่งแต่ละเอกสารประกอบไปด้วยรายการคำสำคัญและบทความขนาดเล็ก 6 ประโยค จากนั้นประเมินว่าประโยคหนึ่ง ๆ ที่ปรากฏขึ้นในตำแหน่งใด ๆ มีน้ำหนักเกี่ยวเนื่องกับชื่อเรื่อง/ชื่อของเอกสารมากที่สุด และจัดอันดับ ซึ่งการประเมินวิธีการที่นำเสนอวัดได้จากประโยคที่อยู่ในลำดับต้น ๆ ของการจัดอันดับนั้นสามารถครอบคลุมรายการคำสำคัญของเอกสารนั้น ๆ

ซึ่งการสรุปใจความสำคัญที่ถูกสร้างขึ้นจากเอกสารหนึ่ง ๆ สามารถขยายออกไปเพื่อสร้างการสรุปใจความสำคัญจากเอกสารเอกสารหลายชุดได้

2.3.1.2 การสรุปใจความสำคัญจากหลายเอกสาร (multi-document summarization)

(McKeown et al., 1999) ได้นำเสนอวิธีการสรุปใจความสำคัญของหลายเอกสาร โดยพิจารณาการระบุธีม เช่น รายการของคำที่เหมือนกัน เป็นต้น ซึ่งเป็นแนวคิดเดียวกับการจัดกลุ่ม (Clustering) การระบุธีมหนึ่ง ๆ นั้นสามารถทำได้จากการวัดความคล้ายคลึงกันระหว่างคำ/ข้อความ ซึ่งสามารถคำนวณได้จากวิธีการระบุนามวลี (Noun phrases), การระบุสรรพนาม (Proper nouns), การค้นหารายการคำที่มีความหมายเดียวกัน/ใกล้เคียงกัน (Synsets) และการค้นหาความหมายแบบลำดับชั้นของคำกริยา จากนั้นจะทำการเปรียบเทียบคู่ของเอกสาร โดยพิจารณาแต่ละประโยคตามคุณสมบัติข้างต้นที่ได้กล่าวไว้ จากนั้นใช้วิธีกฎการตัดสินใจ (Decision rules) เพื่อนำมาจำแนกเอกสารแต่ละคู่ว่าเหมือนหรือแตกต่างกัน ทำให้สามารถระบุธีมของกลุ่มเอกสารแต่ละกลุ่มได้ จากนั้นจะพิจารณาต่อว่าในกลุ่มของเอกสารกลุ่มหนึ่ง ๆ (ธีมหนึ่ง ๆ)นั้น ประโยคใดที่ปรากฏขึ้นบ่อยที่สุด (มีความถี่มากที่สุด) และใช้ประโยคดังกล่าวเป็นการสรุปใจความสำคัญของกลุ่มของเอกสารกลุ่มหนึ่ง ๆ

(Carbonell & Goldstein, 1998) ได้นำเสนอวิธีการสรุปใจความสำคัญตามหัวข้อที่มีชื่อวิธีการว่า maximal marginal relevance (MMR) โดยมีแนวคิดคือรวมคำค้นที่มีความเกี่ยวเนื่องกันกับความใหม่ของข้อมูล เข้ามาใช้ในการดำเนินการสรุปใจความสำคัญ จากแนวคิดนี้สามารถนำไปประยุกต์ใช้กับงานต่าง ๆ ได้ตั้งแต่การสกัดข้อความไปจนถึงการสรุปใจความสำคัญตามหัวข้อ ซึ่งวิธีการ MMR จะดำเนินการเก็บข้อความที่เกี่ยวเนื่องกันและทำการตัดข้อความที่มีความซ้ำซ้อนกันออกไป จากนั้นทำการเรียงลำดับของข้อความที่มีลำดับสูงสุดขึ้นก่อน และนำเสนอออกมาเป็นผลลัพธ์ ซึ่งจากวิธีการของ Carbonell et al., สามารถสรุปใจความสำคัญได้ตามคำค้นที่ผู้ใช้งานป้อนเข้ามา ซึ่งทำให้ได้ผลลัพธ์ของการสรุปใจความสำคัญขึ้นอยู่กับคำค้นของผู้ใช้งานแต่ละคน

จนถึงขณะนี้มีการพัฒนาเทคนิคและคิดค้นแนวทางจำนวนมากในงานวิจัยที่เกี่ยวข้องกับการสรุปใจความสำคัญอัตโนมัติอย่างต่อเนื่อง เนื่องจากเนื้อหาหรืออินพุตที่นำมาสรุปมีหลายรูปแบบ เช่น บทความข่าว นวนิยาย หนังสือ เอกสารทางกฎหมาย เอกสารชีวการแพทย์ เอกสารทางวิทยาศาสตร์ แหล่งข้อมูลบนอินเทอร์เน็ต(บทวิจารณ์ของผู้ใช้ บทความบนบล็อก เครือข่ายโซเชียลมีเดีย ฯลฯ) เป็นต้น

2.3.1.3 แบบผสม

2.3.1.3.1 การสรุปข่าวสาร

(Sahni & Palwe, 2018) ได้นำเสนอวิธีการสรุปใจความสำคัญสร้างแบบจำลองลูกโซ่คำศัพท์พื้นฐาน โดยใช้อัลกอริธึมการสร้างลูกโซ่คำศัพท์ (lexical chains) มาใช้ในการดำเนินการกำหนดสรรพนามให้สัมพันธ์กับคำนามที่เกิดขึ้นและการให้คะแนนประโยค เพื่อใช้ประโยชน์จากโครงสร้างของบทความข่าวในการสรุปใจความสำคัญออกมา ผลลัพธ์จากสรุปพบว่านักข่าวทำตามรูปแบบที่ตายตัวในการเขียนบทความข่าว โดยในบทความจะเริ่มต้นด้วยสิ่งที่เกิดขึ้นในย่อหน้าแรก ต่อด้วยการอธิบายรายละเอียดว่าทำไมจึงเกิดขึ้นในย่อหน้าต่อไปนี้ และตั้งแต่ประโยคแรกของบทความมักจะมีคำนามที่พูดถึงหัวข้อของบทความ

2.3.1.3.2 การสรุปอีเมลล์

(Muresan et al., 2001) พบว่าข้อความอีเมลล์เป็นข้อความที่ไม่มีโครงสร้าง จึงนำเสนอแนวคิดการสรุปใจความสำคัญที่รวมเทคนิคทางภาษา กับอัลกอริธึมการเรียนรู้ของเครื่อง เพื่อแยกคำ

นามวลี เพื่อสร้างข้อมูลสรุปของข้อความอีเมล ซึ่งสรุปผลลัพธ์ว่าการปรับเปลี่ยนของคำนามวลีมีความสำคัญกับความหมายทางภาษาศาสตร์ในการสรุป

2.3.1.3.3 การสรุปเว็บไซต์เครือข่ายสังคม Twitter

(Alsaedi et al., 2016) ได้นำเสนอการเปรียบเทียบวิธีการสรุปกลุ่มหัวข้อของโพสต์ของทวิต 3 วิธี ได้แก่ I.) วิธี TF-IDF กำหนดให้ค่าน้ำหนักกับคำที่มากที่สุด II.) วิธีการโหวตจากการรีทวิต พิจารณาจำนวนการรีทวิตสูงสุดของโพสต์เพื่อค้นหาโพสต์ที่เป็นตัวแทนมากที่สุด และ III.) วิธี Centroid เลือกโพสต์ที่สอดคล้องกับเซนทรอยด์มากที่สุดและจัดกลุ่ม (cluster) ซึ่งจะพิจารณาโพสต์ทวิตที่อยู่ในกรอบเวลาหนึ่งชั่วโมง จากผลลัพธ์ที่ทำวิเคราะห์ วิธีการ TF-IDF ได้ผลลัพธ์ที่สูงสุดทั้งในส่วนคะแนน ROUGE และคะแนนการประเมินของมนุษย์ และวิธีการ Centroid สะท้อนให้เห็นถึงหัวข้อหรือเหตุการณ์

และนับตั้งแต่เริ่มต้นผู้วิจัยที่เกี่ยวข้องกับการสรุปใจความสำคัญอัตโนมัติ พวกเขายังคงพยายามทำการวิจัยเพื่อปรับปรุงเทคนิคหรือสร้างวิธีการสร้างข้อมูลสรุปเพื่อให้ข้อมูลสรุปที่ถูกสร้างด้วยคอมพิวเตอร์สามารถจับคู่กับบทสรุปที่มนุษย์สร้างขึ้น หรือให้มนุษย์สามารถเข้าใจได้มากยิ่งขึ้น

2.3.2 งานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ประกาศงาน

งานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ประกาศงานมีแนวทางและวิธีการต่าง ๆ ที่ถูกใช้เพื่อวิเคราะห์ข้อมูลที่ปรากฏในประกาศงาน งานวิจัยเหล่านี้ใช้แนวทางและวิธีการที่หลากหลายเพื่อเข้าใจแนวโน้มและความต้องการของตลาดงานในตำแหน่งงานที่แตกต่างกันในสาขาและสถานที่ต่าง ๆ สรุปงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ประกาศงานดังในตารางที่ 4 ตารางงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ประกาศงาน

ตารางที่ 4 ตารางงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ประกาศงาน

ผู้วิจัย	ที่มาข้อมูล	ข้อมูลที่มุ่งเน้น	ภาษา	แนวทาง	วิธีการ
Nelson, และคณะ (2007)	เว็บไซต์ออนไลน์	ทักษะ	อังกฤษ	จุดมุ่งหมายเพื่อกำหนดรูปแบบของทักษะที่จำเป็นสำหรับงานด้าน IT/IS และตรวจสอบความแตกต่างในรูปแบบการจ้างงานระหว่างบริษัทที่มีขนาดแตกต่างกัน	วิเคราะห์โดยใช้การจัดกลุ่มความหมายแฝง (Latent Semantic Clustering)

ผู้วิจัย	ที่มาข้อมูล	ข้อมูลที่มุ่งเน้น	ภาษา	แนวทาง	วิธีการ
Sibarani, และคณะ (2017)	ข้อมูลความต้องการงานที่ได้รับจากความพยายามในการระบุทักษะที่ร้องขอจากตำแหน่งงานว่าง	ทักษะ	อังกฤษ	อธิบายการศึกษาเกี่ยวกับความต้องการของตลาดงานสำหรับทักษะวิทยาศาสตร์ข้อมูลโดยใช้วิธีการดึงข้อมูลที่มีแนวทางแบบออนโทโลยี	ใช้วิธี Ontology-based Information Extraction (OBIE) เพื่อระบุทักษะที่ได้รับการร้องขอจากตำแหน่งงานว่าง
Radovilsky, และคณะ (2018)	เว็บไซต์ที่เกี่ยวข้องกับงานออนไลน์วิเคราะห์ข้อมูลธุรกิจ (BDA) และวิทยาศาสตร์ข้อมูล(DS)	ทักษะ	อังกฤษ	ระบุและเปรียบเทียบขอบเขตความรู้และทักษะที่จำเป็นสำหรับวิชาชีพการวิเคราะห์ข้อมูลธุรกิจ (BDA) และวิทยาศาสตร์ข้อมูล (DS)	ใช้วิธีการขุดข้อความเพื่อวิเคราะห์ข้อมูลการลงประกาศงาน
Parker, และคณะ (2019)	ประกาศรับสมัครงานที่ไม่ซ้ำกัน 196 รายการจากเว็บไซต์พอร์ทัลรับสมัครงาน 5 แห่งในแอฟริกาใต้	ทักษะ	อังกฤษ	ตรวจสอบข้อกำหนดทักษะความรู้ และคุณสมบัติสำหรับผู้เชี่ยวชาญด้านความปลอดภัยทางไซเบอร์ในแอฟริกาใต้	การศึกษานี้ประกอบด้วยวิธีผสมผสานวิธีนิรนัยและอุปนัย
Pejic-Bach, และคณะ (2020)	เว็บไซต์ที่เผยแพร่โฆษณารับสมัครงาน	ทักษะ	อังกฤษ	จุดมุ่งหมายเพื่อพัฒนาโปรไฟล์ของการโฆษณาตำแหน่งงานในอุตสาหกรรม 4.0 โดยใช้การขุดข้อความในโฆษณา	ทำการวิเคราะห์การขุดข้อความ (text mining analysis) เกี่ยวกับโฆษณารับสมัครงาน

ผู้วิจัย	ที่มาข้อมูล	ข้อมูลที่มุ่งเน้น	ภาษา	แนวทาง	วิธีการ
				ตำแหน่งงาน	
Lukauskas, และคณะ (2023)	ไม่ออนไลน์ ประกาศรับสมัครงาน	ทักษะ	อังกฤษ	มุ่งหวังที่จะใช้เทคโนโลยีการประมวลผลภาษาธรรมชาติและวิธีการจัดกลุ่มข้อมูลเพื่อวิเคราะห์ความต้องการทักษะของพนักงานชาวลิทัวเนียและสร้างโปรไฟล์งานอัตโนมัติ	สำรวจการใช้การประมวลผลภาษาธรรมชาติและเทคนิคการจัดกลุ่มเพื่อวิเคราะห์ประกาศรับสมัครงานและสร้างโปรไฟล์งานอัตโนมัติ

งานวิจัยของ (Nelson et al., 2007) มุ่งเน้นการกำหนดรูปแบบของทักษะที่จำเป็นสำหรับงานด้านเทคโนโลยีสารสนเทศและระบบสารสนเทศที่ปรากฏในประกาศงานออนไลน์ และการตรวจสอบความแตกต่างในรูปแบบการจ้างงานระหว่างบริษัทที่มีขนาดแตกต่างกัน วิธีการที่ใช้ในงานวิจัยนี้คือการวิเคราะห์คำอธิบายงานออนไลน์โดยใช้การจัดกลุ่มความหมายแฝง (Latent Semantic Clustering) เพื่อค้นพบแนวโน้มและความสัมพันธ์ในความต้องการที่เกี่ยวข้องกับงานด้านเทคโนโลยีสารสนเทศและระบบสารสนเทศ และสร้างระบบการจัดประเภทในรูปแบบปริมาณ (quantitative classification system) เพื่อทำความเข้าใจความต้องการทักษะในงานเหล่านี้ได้อย่างชัดเจนและเชิงปริมาณ ผลลัพธ์ของงานวิจัยนี้ช่วยในการเข้าใจรูปแบบในข้อกำหนดการจ้างงานและความต้องการของตลาดงานในตำแหน่งงานด้านเทคโนโลยีสารสนเทศและระบบสารสนเทศ โดยเน้นไปที่ความแตกต่างในความต้องการของบริษัทขนาดใหญ่เมื่อเทียบกับบริษัทขนาดเล็กในสาขานี้

งานวิจัยของ (Sibarani et al., 2017) ใช้ข้อมูลเกี่ยวกับความต้องการงานที่ได้รับจากความพยายามในการระบุทักษะที่ร้องขอ จากตำแหน่งงานว่าง และใช้วิธีการดึงข้อมูลตาม Ontology-based Information Extraction (OBIE) เพื่อระบุทักษะที่ได้รับการร้องขอจากตำแหน่งงานว่างในสาขาวิทยาศาสตร์ข้อมูล โดยอ้างอิงคำศัพท์ที่ออกแบบเฉพาะทาง การวิจัยนี้มุ่งเน้นการเข้าใจความต้องการของตลาดงานสำหรับทักษะด้านวิทยาศาสตร์ข้อมูลโดยใช้เทคนิคการวิเคราะห์และการดึงข้อมูลที่อาจช่วยในการสร้างโปรไฟล์งานอัตโนมัติของตำแหน่งงานที่ตรงตามความต้องการของตลาด

ผลการวิจัยเผยให้เห็นทักษะทางเทคนิคที่เป็นที่ต้องการและโครงสร้าง ทำให้เกิดความเชื่อมโยงที่สำคัญสำหรับการวิเคราะห์ขั้นสูงยิ่งขึ้น

งานวิจัยของ (Radovilsky et al., 2018) ได้รวบรวมข้อมูลการประกาศงานวิชาชีพการวิเคราะห์ข้อมูลธุรกิจ (Business Data Analytics) และวิทยาศาสตร์ข้อมูล (Data Scientist) จากเว็บไซต์ที่เกี่ยวข้องกับงานออนไลน์ ข้อมูลที่วิเคราะห์ประกอบด้วยข้อกำหนด ทักษะ และคุณสมบัติของตำแหน่งงาน เป็นข้อมูลลักษณะที่ไม่มีโครงสร้าง เพื่อการวิเคราะห์โดยใช้วิธีการขุดข้อความ (text mining) เพื่อระบุความถี่ของทักษะต่างๆ ที่เกี่ยวข้องกับวิชาชีพวิเคราะห์ข้อมูลธุรกิจ และวิทยาศาสตร์ข้อมูล จัดกลุ่มทักษะเหล่านั้น และเปรียบเทียบ ผลลัพธ์ของการวิเคราะห์การขุดข้อความเผยให้เห็นความเหมือนและความแตกต่างระหว่างขอบเขตความรู้ที่จำเป็นและกลุ่มทักษะสำหรับวิชาชีพวิเคราะห์ข้อมูลธุรกิจ และวิทยาศาสตร์ข้อมูล โดยให้ข้อมูลเชิงลึกสำหรับการออกแบบหลักสูตรและการพัฒนาทักษะทางวิชาชีพ

งานวิจัยของ (Parker & Brown, 2019) จะตรวจสอบทักษะ ความรู้ และข้อกำหนดคุณสมบัติสำหรับผู้เชี่ยวชาญด้านความปลอดภัยทางไซเบอร์ (Cyber Security) ในแอฟริกาใต้ โดยอิงจากการวิเคราะห์เนื้อหาของประกาศรับสมัครงาน 196 รายการ จากเว็บไซต์รับสมัครงาน 5 แห่ง แนวทางการวิจัยที่ใช้ในการศึกษานี้ ประกอบด้วย ทั้งวิธีทางเชิงนิรนัยและวิธีทางเชิงอุปนัยมาใช้ร่วมกันในการศึกษาหรือการวิจัย วิธีการนิรนัยเกี่ยวข้องกับการใช้บทบาทและทักษะด้านความปลอดภัยทางไซเบอร์ที่มีอยู่ซึ่งระบุไว้ในการศึกษาก่อนหน้านี้ เช่น การศึกษาโดย Potter และ Vickers ในออสเตรเลีย วิธีการอุปนัยเกี่ยวข้องกับการวิเคราะห์ประกาศรับสมัครงานเพื่อกำหนดธีมและหมวดหมู่ที่เกี่ยวข้องกับทักษะและความรู้ มีการสร้างความสัมพันธ์ระหว่างหมวดหมู่ ระยะแรกของการศึกษาเกี่ยวข้องกับการค้นหาและทบทวนวรรณกรรมที่มีอยู่เกี่ยวกับข้อกำหนดทักษะสำหรับผู้เชี่ยวชาญด้านความปลอดภัยทางไซเบอร์ ข้อค้นพบจากการทบทวนนี้ใช้เพื่อรวบรวมรายการหมวดหมู่เพื่อวิเคราะห์ประกาศรับสมัครงานในแอฟริกาใต้ ระยะที่สองเกี่ยวข้องกับการรวบรวมประกาศรับสมัครงานในช่วงระยะเวลาสองเดือน และการตรวจสอบรายการที่ซ้ำกันเพื่อให้ได้ตำแหน่งผู้บริหารและตำแหน่งการขาย 196 ตำแหน่งที่ไม่ซ้ำกัน ผลการวิจัยพบว่าผู้เชี่ยวชาญด้านความปลอดภัยทางไซเบอร์ในแอฟริกาใต้จำเป็นต้องมีทักษะด้านเทคนิคและมนุษย์สัมพันธ์ พร้อมด้วยวุฒิ

การศึกษาขั้นต่ำในระดับปริญญาตรีสาขาวิทยาการคอมพิวเตอร์ ระบบสารสนเทศ หรือวิศวกรรมศาสตร์ พร้อมด้วยใบรับรองอุตสาหกรรม การศึกษานี้ทำหน้าที่เป็นการศึกษาเชิงสำรวจเบื้องต้นที่สามารถวางรากฐานสำหรับการวิจัยที่เฉพาะเจาะจงมากขึ้นในอนาคต

งานวิจัยของ (Pejic-Bach et al., 2020) มีจุดมุ่งหมายเพื่อพัฒนาโปรไฟล์ของการประกาศงานตำแหน่งงานในอุตสาหกรรม 4.0 ซึ่งรวบรวมข้อมูลจากโฆษณาตำแหน่งงานเกี่ยวข้องกับอุตสาหกรรม 4.0 ที่เผยแพร่ประกาศงานบนเว็บไซต์ จากการวิเคราะห์ประกาศรับสมัครงาน พบว่าส่วนใหญ่เป็นตำแหน่งงานเต็มเวลา พนักงานร่วม และผู้บริหารระดับกลาง โดยส่วนใหญ่มาจากสหรัฐอเมริกาและเยอรมนี การวิเคราะห์การขุดข้อความ (text mining) ส่งผลให้มีโปรไฟล์งานสองกลุ่ม กลุ่มหนึ่งมุ่งเน้นไปที่ความรู้ที่เกี่ยวข้องกับอุตสาหกรรม 4.0 เช่น ระบบไซเบอร์ฟิสิกส์และอินเทอร์เน็ตของสิ่งต่าง ๆ สำหรับการผลิตโดยใช้หุ่นยนต์ และอีกกลุ่มมุ่งเน้นไปที่ความรู้ทั่วไปที่ปรับให้เข้ากับอุตสาหกรรม 4.0 เช่น การจัดการห่วงโซ่อุปทานและความพึงพอใจของลูกค้า นักวิจัยยังได้ดำเนินการขุดหัวข้อเกี่ยวกับวลีที่แยกออกมาเพื่อสร้างโปรไฟล์งานในสาขาต่างๆ

งานวิจัยของ (Lukauskas et al., 2023) วิเคราะห์ความต้องการด้านทักษะของพนักงานชาวลิทัวเนีย ดำเนินการวิเคราะห์กลุ่ม และสร้างโปรไฟล์งานอัตโนมัติโดยใช้การประมวลผลภาษาธรรมชาติและเทคนิคการจัดกลุ่ม การศึกษานี้ใช้ประกาศรับสมัครงานมากกว่าห้าแสนรายการเพื่อสร้างโปรไฟล์งาน/ตำแหน่ง โดยดึงข้อมูลจากข้อกำหนดของงาน การทำเวกเตอร์ข้อมูลดำเนินการโดยใช้โครงสร้าง BERT และการลดขนาดทำได้โดยใช้เทคนิค UMAP ซึ่งให้ผลลัพธ์ที่ดีที่สุด พบว่าวิธี HDBSCAN มีประสิทธิภาพมากที่สุดสำหรับการจัดกลุ่ม และคำอธิบายโปรไฟล์งานถูกสร้างขึ้นโดยใช้ปัญญาประดิษฐ์เชิงกำเนิดตามทักษะโปรไฟล์งานที่รวบรวมไว้ การประเมินโดยผู้เชี่ยวชาญยืนยันผลลัพธ์ที่ประสบความสำเร็จของอัลกอริทึมการวิเคราะห์ประกาศรับสมัครงานแบบอัตโนมัติ ซึ่งบ่งชี้ถึงการนำไปใช้ได้จริง

แม้ว่าการศึกษาเหล่านี้จะนำเสนอข้อมูลที่มีค่า แต่มักเน้นการระบุชุดทักษะสำคัญที่จำเป็นต้องมีโดยนายจ้างอย่างไรก็ตาม ผู้สมัครงานบ่อยครั้งต้องการเข้าใจประกาศงานโดยรวม รวมถึงรายละเอียดที่ซ่อนอยู่ เช่น ผู้สมัครงานที่วางแผนที่จะเปลี่ยนงานในอนาคตใกล้เข้ามา ต้องการความเข้าใจโดยรวมเกี่ยวกับข้อจำกัดที่ถูกตั้งโดยนายจ้าง เพื่อประเมินความเข้ากันได้กับความคาดหวังใน

อาชีพระยะยาว โดยการพิจารณาบริบทที่กว้างขึ้น ผู้สมัครงานสามารถตัดสินใจโดยมีข้อมูลเพื่อสำรวจหน้าที่ที่ต้องรับผิดชอบโดยเฉพาะ ทักษะที่จำเป็น และข้อจำกัดอื่น ๆ ที่ระบุในประกาศงาน

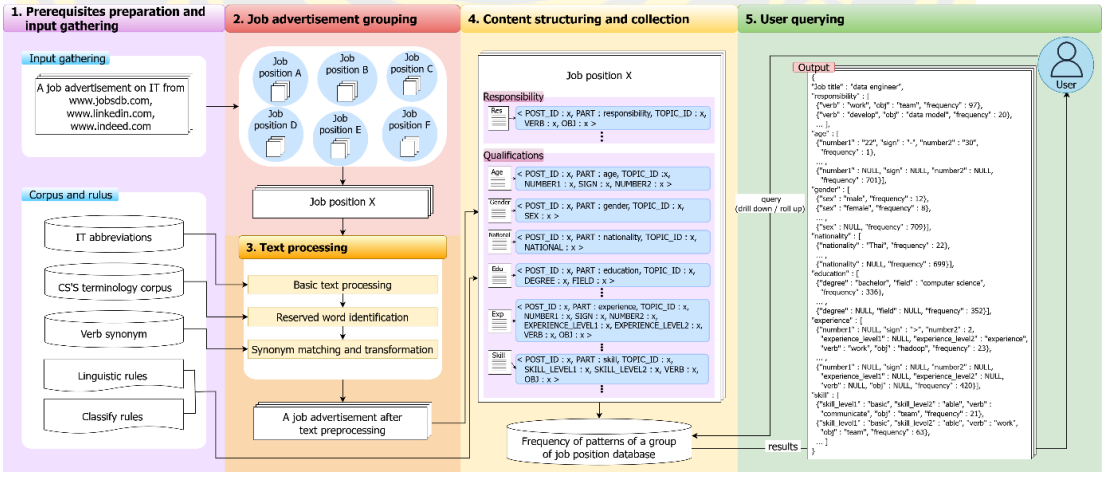
ดังนั้นการรู้และเข้าใจรายละเอียดและบริบทที่ซ่อนอยู่ในประกาศงานช่วยให้ผู้สมัครงานทำการสำรวจตำแหน่งงานอย่างมีประสิทธิภาพและเข้าใจความต้องการของนายจ้างในลักษณะที่ลึกซึ้งมากขึ้น และทำให้พวกเขาสามารถตัดสินใจเกี่ยวกับการสมัครงานได้ดียิ่งขึ้น ในการพิจารณาคาดคะเนตัวเองในตำแหน่งงานนั้นอย่างเหมาะสม

บทที่ 3 วิธีการดำเนินการวิจัย

งานวิจัยนี้มีเป้าหมายในการวิเคราะห์และสรุปความรับผิดชอบ และข้อกำหนดจากข้อมูลในประกาศรับสมัครงานในศาสตร์ทางด้านเทคโนโลยีสารสนเทศจากเว็บไซต์รับสมัครงานภายในประเทศไทย นำเสนอโครงสร้างวิธีการในการค้นพบข้อมูลที่เกิดขึ้นร่วมกันในการประกาศรับสมัครงานที่เกี่ยวข้องกับคอมพิวเตอร์ (Co-occurrence Information Discovery in Computer-related Job Advertisements) หรือเรียกว่า “CID-CJA” โดยดำเนินการประกอบไปด้วย 5 ขั้นตอน ดังต่อไปนี้

1. การเตรียมข้อมูลและการรวบรวมข้อมูลประกาศรับสมัครงานเฉพาะศาสตร์ทางด้านเทคโนโลยีสารสนเทศ
2. การจัดกลุ่มประกาศงาน
3. การประมวลผลข้อความ
4. การจัดโครงสร้างและการรวบรวมเนื้อหา
5. การส่งข้อมูลให้ผู้ใช้

โดยโครงสร้างรวมของ 5 ขั้นตอน แสดงดังภาพที่ 4 เป็นกรอบงานการวิเคราะห์และสรุปความรับผิดชอบและข้อกำหนดจากประกาศรับสมัครงานทางด้านเทคโนโลยีสารสนเทศ



ภาพที่ 4 กรอบงานการวิเคราะห์และสรุปความรับผิดชอบและข้อกำหนดจากประกาศรับสมัครงานทางด้านเทคโนโลยีสารสนเทศ

3.1 การเตรียมข้อมูลและการรวบรวมข้อมูลประกาศรับสมัครงานเฉพาะศาสตร์ทางด้านเทคโนโลยีสารสนเทศ

ในขั้นตอนนี้เน้นการเตรียมความพร้อม รวบรวมข้อมูลที่จำเป็น และรวบรวมข้อมูลนำเข้าโดยเป็นข้อมูลประกาศรับสมัครงานเฉพาะศาสตร์ทางด้านเทคโนโลยีสารสนเทศที่เป็นภาษาอังกฤษ และอยู่ภายในประเทศไทยเท่านั้น ซึ่งจะรวบรวมจากเว็บไซต์ jobsDB, Linkedin และ Indeed

3.1.1 ศึกษาโครงสร้างข้อมูลประกาศงานของเว็บไซต์

ผู้วิจัยทำการสำรวจโครงสร้างข้อมูลประกาศรับสมัครงาน เนื่องจากโครงสร้างรูปแบบหน้าประกาศรับ สมัครงานของแต่ละเว็บไซต์มีการนำเสนอข้อมูลบางส่วนแตกต่างกัน จากสาเหตุนี้ทำให้ต้องมีการศึกษาโครงสร้างของเว็บไซต์และโครงสร้างของข้อมูลประกาศงานของตำแหน่งงานว่ามีความเหมือนหรือแตกต่างกันอย่างไรดังตารางที่ 5 เพื่อนำข้อมูลมาจัดรูปแบบให้อยู่เป็นมาตราฐานเดียวกัน ซึ่งศึกษาโครงสร้างข้อมูลประกาศงานของ 3 เว็บไซต์ดังต่อไปนี้

1. https://th.jobsdb.com
2. https://th.linkedin.com
3. https://th.indeed.com

ตารางที่ 5 โครงสร้างข้อมูลของเว็บไซต์

ชื่อเว็บไซต์	**jobsDB**	**Linkedin**	**Indeed**
โครงสร้างข้อมูลประกาศรับสมัครงาน			
1. ชื่อตำแหน่งงาน	มี	มี	มี
2. ชื่อองค์กร/บริษัท	มี	มี	มี
3. วันที่โพสรับสมัครงาน	มี	มี	มี
4. ที่อยู่องค์กร/บริษัท	มี	มี	มี

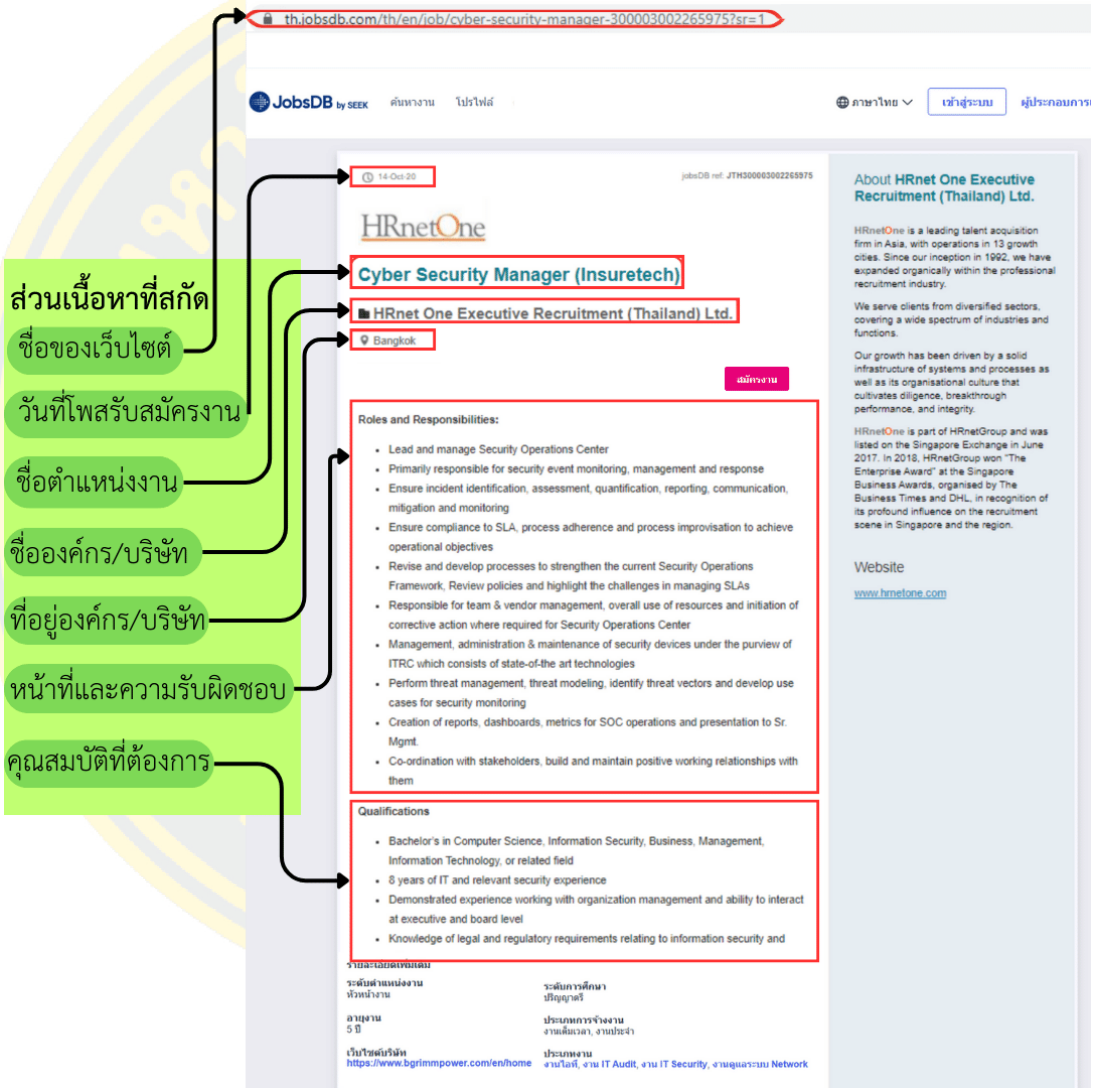
ชื่อเว็บไซต์	jobsDB	Linkedin	Indeed
5. หน้าที่ความรับผิดชอบ	มี	มี	มี
6. คุณสมบัติที่ต้องการ	มี	มี	มี
7. เงินเดือน	มีและไม่มี	ไม่มี	ไม่มี
8. ระดับตำแหน่งงาน	มี	มี	มีและไม่มี
9. ประเภทการจ้างงาน	มี	มี	มีและไม่มี
10. ประเภทงาน	มี	มี	ไม่มี
11. ประเภทอุตสาหกรรมขององค์กร/บริษัท	มี	มี	ไม่มี
12. ที่อยู่ตำแหน่งเว็บไซต์ของประกาศงาน	มี	มี	มี
13. ชื่อของเว็บไซต์	มี	มี	มี

จากตารางที่ 5 จะเห็นว่ามีข้อมูลบางส่วนที่ไม่พบบนหน้าประกาศรับสมัครงานของเว็บไซต์ส่วนมากอยู่ในส่วนของข้อมูล เงินเดือน ระดับตำแหน่งงาน ประเภทการจ้างงาน ประเภทงาน ประเภทอุตสาหกรรมขององค์กร/บริษัท ดังนั้นผู้วิจัยจึงเลือกที่จะรวบรวมข้อมูลในส่วนที่พบในหน้าประกาศรับสมัครงานของทั้ง 3 เว็บไซต์คือ ชื่อตำแหน่งงาน ชื่อองค์กร/บริษัท วันที่โพสรับสมัครงาน ที่อยู่องค์กร/บริษัท หน้าที่ความรับผิดชอบ คุณสมบัติที่ต้องการ ที่อยู่ตำแหน่งเว็บไซต์ของประกาศงาน และชื่อของเว็บไซต์

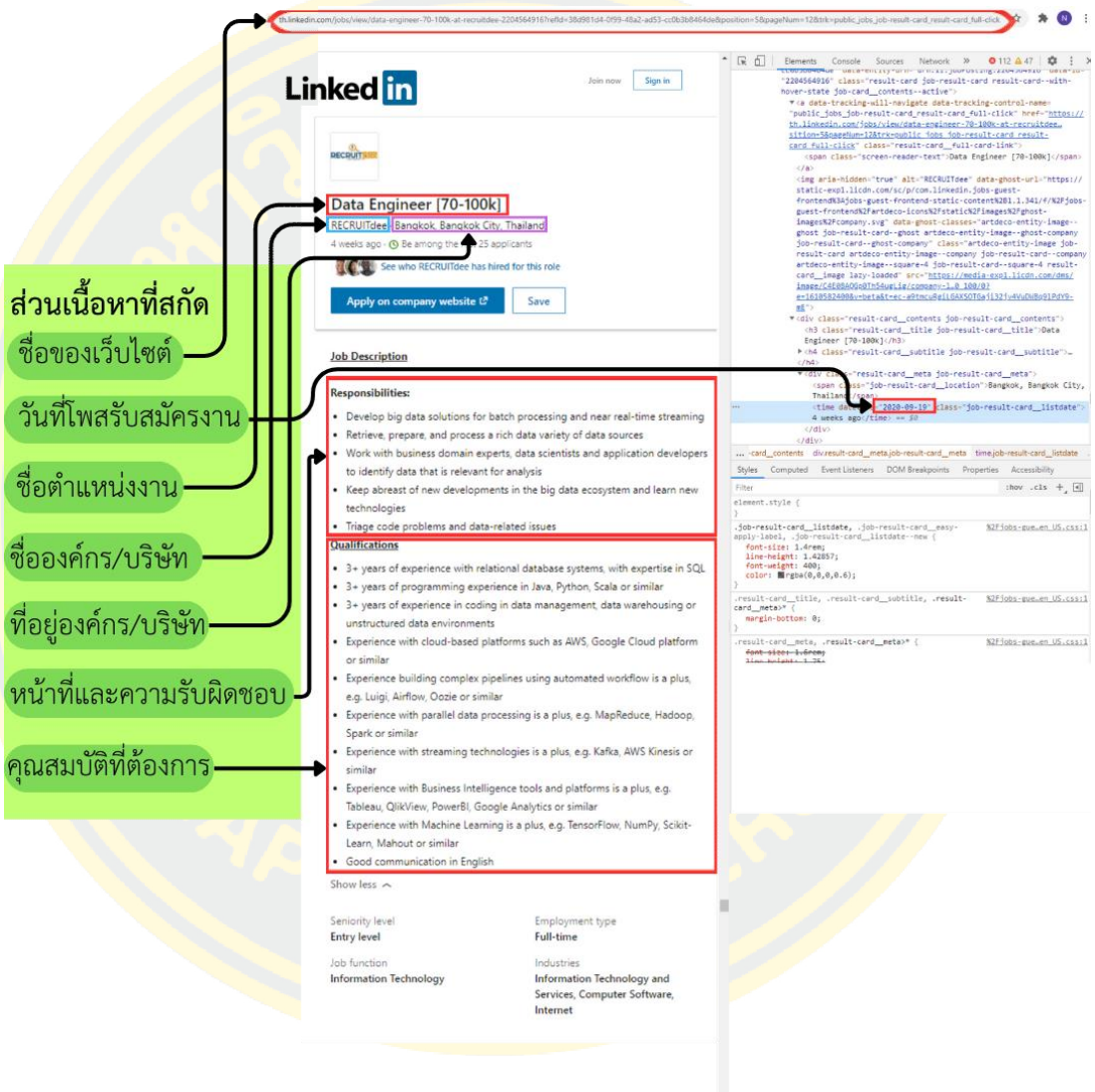
3.1.2 การสกัดข้อมูลจากเว็บไซต์ jobsDB Linkedin และ Indeed

หลังจากศึกษาโครงสร้างข้อมูลประกาศงานของแต่ละเว็บไซต์ พบว่าข้อมูลที่สามารถสกัดได้คือ i) ชื่อตำแหน่งงาน ii) ชื่อองค์กร/บริษัท iii) วันที่โพสรับสมัครงาน iv) ที่อยู่องค์กร/บริษัท v) หน้าที่ความรับผิดชอบ vi) คุณสมบัติที่ต้องการ vii) ที่อยู่ตำแหน่งเว็บไซต์ของประกาศงาน viii) ชื่อของเว็บไซต์จากบนหน้าเว็บไซต์ที่เผยแพร่แล้ว โดยจะใช้วิธีการสกัดข้อมูลจากเว็บไซต์ (web

scraping) ในการรวบรวมข้อมูลในส่วนที่ต้องการบนเว็บไซต์ jobsDB Linkedin และ Indeed แสดงดังภาพที่ 5, 6 และ 7 แสดงการนำเสนอโครงสร้างข้อมูลบนเว็บไซต์ และส่วนเนื้อหาบนหน้าเว็บไซต์ที่ต้องการสกัดข้อมูล



ภาพที่ 5 แสดงการนำเสนอโครงสร้างข้อมูลบนเว็บไซต์ https://th.jobsdb.com/ ในส่วนที่ต้องการสกัด



ภาพที่ 6 แสดงการนำเสนอโครงสร้างข้อมูลบนเว็บไซต์ https://www.linkedin.com/ ในส่วนที่ต้องการสกัด

ภาพที่ 7 แสดงการนำเสนอโครงสร้างข้อมูลบนเว็บไซต์ https://th.indeed.com/ ในส่วนที่ต้องการสกัด

จากภาพที่ 5, 6 และ 7 แสดงโครงสร้างหน้าเว็บไซต์และการนำเสนอข้อมูลประการรับสมัครงาน ซึ่งเลือกดำเนินการสกัดข้อมูลออกมาในส่วน

i) วันที่ลงประกาศงาน
ii) ชื่อตำแหน่งงาน
iii) ชื่อองค์กร/บริษัท
iv) ที่ตั้งขององค์กร/บริษัท
v) หน้าที่ความรับผิดชอบ
vi) คุณสมบัติที่ต้องการ
vii) ที่อยู่ตำแหน่งเว็บไซต์ของประกาศงาน
viii) และชื่อของเว็บไซต์จากบนหน้าเว็บไซต์ที่เผยแพร่แล้ว

แล้วจึงนำข้อมูลประกาศงานที่ถูกสกัด นำไปรวบรวมและจัดเก็บลงในฐานข้อมูลที่จัดทำขึ้นเป็นคลังข้อมูลเพื่อจัดเก็บข้อมูลประกาศรับสมัครงานที่จะนำไปพิจารณาและวิเคราะห์ในกระบวนการวิเคราะห์ต่อไปแสดงตัวอย่างดังภาพที่ 8 ตัวอย่างข้อมูลที่ถูกสกัดจากหน้าเว็บไซต์ที่เผยแพร่นำไปจัดเก็บที่ฐานข้อมูล

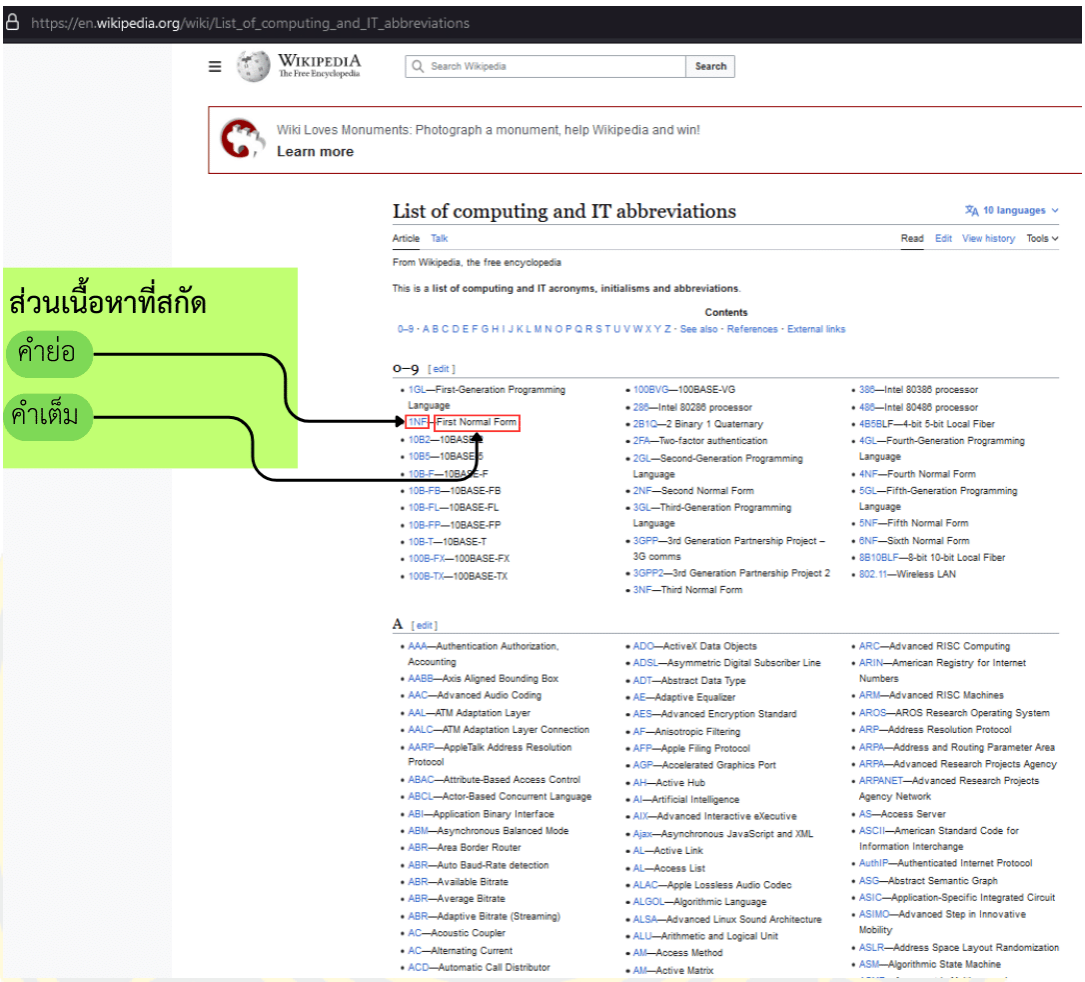
แอททริบิวต์(Attribute) ที่เลือกจัดเก็บ	ข้อมูลประกาศงานที่ถูกจัดเก็บลงในฐานข้อมูล
ID	1
Date	27/02/2020
Job title	Front End Developer
Company name	TELLSCORE CO., LTD
Location	Yan Ta Khao District, Trang, Thailand
Responsibility	Develop object-oriented JavaScript, CSS and HTML for Tellscore website. Design and implement JavaScript objects/modules that are reusable and performant. Introduce innovative processes and techniques to boost front-end performance. Collaborate with project managers, designers, and back-end developers to refine and improve the complete user experience. Connect with the user interface development community to track the latest trends and emerging technologies.
Qualification	Minimum 2 years of experience in web dev. Experienced with HTML5, CSS3, jQuery and Bootstrap. Experienced in Respnsive Web Application. Good communication skills.
Link	https://th.indeed.com/viewjob? jk=16f6307791335c31&tk=1ehp4ll68ohei800&from=serp&vjs=3
Web	indeed

ภาพที่ 8 ตัวอย่างข้อมูลประกาศงานที่ถูกจัดเก็บลงในฐานข้อมูลที่จัดทำขึ้นเป็นคลังข้อมูลเพื่อจัดเก็บข้อมูล ประกาศรับสมัครงาน

3.1.3 การรวบรวมคำย่อของคำศัพท์เฉพาะทางด้านเทคโนโลยีสารสนเทศ

คำย่อของคำศัพท์เฉพาะทางด้านเทคโนโลยีสารสนเทศมีค่อนข้างมาก ซึ่งถ้าไม่เปลี่ยนคำย่อเป็นคำเต็มอาจทำให้ความหมายของคำเปลี่ยนแปลงไปอย่างสิ้นเชิง เช่น คำย่อ “IT” คำเต็มคือ “Information Technology” ซึ่งถ้าไม่แปลงคำย่อเป็นคำเต็มคำย่อ “IT” อาจถูกตีความหมายว่า “มัน” หรือ “สิ่งนี้” โดยมักใช้เพื่ออ้างถึงสิ่งของหรือสิ่งที่ถูกกล่าวถึง ด้วยเหตุนี้จึงพิจารณาเก็บรวบรวมคำย่อและคำเต็มที่เกี่ยวข้องกับคำศัพท์ทางด้านเทคโนโลยีสารสนเทศ เพื่อนำมาแปลงคำย่อให้อยู่ในรูปคำเต็มเพื่อความหมายที่แท้จริงของคำ และเพื่อช่วยในการสกัดคำสำคัญจากประกาศงานให้มีประสิทธิภาพมากขึ้น โดยรวบรวมรายการคำย่อและคำเต็มจากเว็บไซต์วิกิพีเดีย (Wikipedia)[1] แสดงดังภาพที่ 9 ที่แสดงเนื้อหาในส่วนคำย่อและคำเต็มบนหน้าเว็บไซต์ ซึ่งสามารถรวบรวมคำย่อได้ทั้งหมด 1,434 คำ และรวบรวมคำเต็มได้ทั้งหมด 1,434 คำ เช่นกัน โดยรวบรวมและจัดเก็บลงในคลังข้อมูลคำย่อชื่อว่า “IT abbreviations”

[1] https://en.wikipedia.org/wiki/List_of_computing_and_IT_abbreviations



ภาพที่ 9 แสดงการส่วนข้อมูลที่ต้องการสกัดเพื่อรวบรวมรายการคำย่อและคำเต็มจากเว็บไซต์วิกิพีเดีย

3.1.4 การรวบรวมคำศัพท์เฉพาะทางคอมพิวเตอร์เทคโนโลยี

การรวบรวมคลังข้อมูลคำศัพท์เฉพาะทางคอมพิวเตอร์เทคโนโลยีและคำพ้องความหมายที่เกี่ยวข้อง งานวิจัยนี้ได้ทำการประยุกต์ใช้คลังข้อมูลที่รวบรวมจาก (Kamlangpuech & Amphawan, 2020, 2021) สำหรับการจัดทำคลังข้อมูลนี้ ศาสตร์ทางด้านเทคโนโลยีสารสนเทศมีคำศัพท์เฉพาะที่มีความหมายเฉพาะค่อนข้างมาก เช่น “Object Oriented Programming”, “Machine Learning”, “Fuzzy Logic”, “Big Data” เป็นต้น ด้วยเหตุนี้จึงพิจารณาการเก็บรวบรวมคำศัพท์เฉพาะทางด้านเทคโนโลยีสารสนเทศนั้น สามารถนำมาใช้ในการระบุถึงคำสำคัญที่เฉพาะเจาะจง เพื่อช่วยในการสกัดคำสำคัญจากประกาศงานให้มีประสิทธิภาพมากขึ้น โดยรวบรวม

คำศัพท์เฉพาะทางด้านคอมพิวเตอร์เทคโนโลยี จาก 8 คลังคำศัพท์ ได้แก่

1) พจนานุกรมคำศัพท์เฉพาะโดยมหาวิทยาลัยอ๊อกซฟอร์ด[2],

2) รายการคำศัพท์เฉพาะจากเว็บไซต์ Labautopedia[3],

3) คำศัพท์เฉพาะทางด้านคอมพิวเตอร์เทคโนโลยีจากเว็บไซต์ ComputerHope[4]

4) คำศัพท์เฉพาะที่เกี่ยวข้องทางด้านคอมพิวเตอร์จากเว็บไซต์ Glossary of Computer Related Terms[5]

5) คำศัพท์เฉพาะทางด้านคอมพิวเตอร์เทคโนโลยีจากเว็บไซต์ GCSE's Computer Science Glossary[6]

6) คำศัพท์เฉพาะทางด้านวิทยาการคอมพิวเตอร์ที่รวบรวมโดยวิกิพีเดีย (Wikipedia's Glossary of computer science)[7]

7) คำศัพท์เฉพาะที่เกี่ยวข้องกับคอมพิวเตอร์เทคโนโลยีจากเว็บไซต์ PC Glossary [8]และ

8) คำศัพท์เฉพาะที่เกี่ยวข้องกับเทคโนโลยีจากเว็บไซต์ TechTarget [9]

รวมคำศัพท์เฉพาะทั้งสิ้น 28,392 คำ โดยถูกจัดเก็บไว้ในคลังคำศัพท์เฉพาะชื่อว่า "CS's terminology corpus"

---

[2] https://www.oxfordreference.com/view/10.1093/acref/9780199688975.001.0001/acref-9780199688975

[3] http://www.labautopedia.org/mw/List_of_programming_and_computer_science_terms

[4] https://www.computerhope.com/jargon.htm

[5] http://www.math.utah.edu/~wisnia/glossary.html

[6] https://www.computerscience.gcse.guru/glossary

[7] https://en.wikipedia.org/wiki/Glossary_of_computer_science

[8] https://pc.net/glossary/

[9] https://whatis.techtarget.com/definitions/A

3.1.5 การรวบรวมคำกิริยาที่มีความหมายคล้ายคลึงหรือเหมือนกันกับคำกริยาอื่น (Verb synonym)

ในภาษาอังกฤษคำกริยาที่มีความหมายเหมือนกันมักจะใช้ในบริบทที่คล้ายคลึงกัน และสามารถแทนที่กันได้ในประโยค โดยสื่อความหมายเดิมและไม่เปลี่ยนแปลงความหมาย เพื่อช่วยในการสรุปข้อมูลในการวิจัยได้ดียิ่งขึ้น จึงพิจารณาดำเนินการรวบรวมรายการคำกริยาของภาษาอังกฤษจาก 2 คลังคำศัพท์ ได้แก่ 1) รูปแบบกริยาของอังกฤษและอเมริกันใน Oxford Advanced Learner's Dictionary ฉบับที่ 9[10] และ 2) รูปแบบกริยาของอังกฤษและอเมริกันใน Longman Dictionary of Contemporary English ฉบับที่ 6[11] ซึ่งรายการคำกริยาที่เก็บรวบรวมมาจะนำมาเป็นคำกริยาตั้งต้น เพื่อนำไปหารายการคำกริยาที่มีความหมายคล้ายคลึงหรือเหมือนกันของคำกริยาตั้งต้นหนึ่ง ๆ จากเว็บไซต์ google translate[12] ซึ่งบนหน้าเว็บไซต์แสดงแถบความถี่ทั้งหมด 3 แถบ ซึ่งหมายถึงระบุว่าคำแปลปรากฏในเอกสารสาธารณะบ่อยเพียงใด ดังนั้นจึงเลือกพิจารณาคำกริยาที่มีความหมายคล้ายคลึงหรือเหมือนกันของคำกริยาตั้งต้นหนึ่ง ๆ ที่มีแถบความถี่มากกว่า 2 แถบขึ้นไป นำมาเก็บรวบรวมเป็นรายการคำของกริยาตั้งต้นนั้น จากนั้นนำคำไปเปรียบเทียบวิธีการ Word2Vec และจับคู่คำหาค่าเวกเตอร์คำในวิธี GloVe โดยค่าได้จาก "glove.840B.300d" เรียนรู้จากเวกเตอร์คำที่ผ่านการฝึกอบรมเกี่ยวกับสถิติการเกิดขึ้นร่วมกันของคำและคำทั่วโลกรวบรวมจากคลังข้อมูล จะได้ค่าเวกเตอร์ของแต่ละคู่คำออกมาจากนั้นนำค่าเวกเตอร์ของคู่คำไปหาค่าความคล้ายคลึงหรือเหมือนกันจากสมการที่ 1

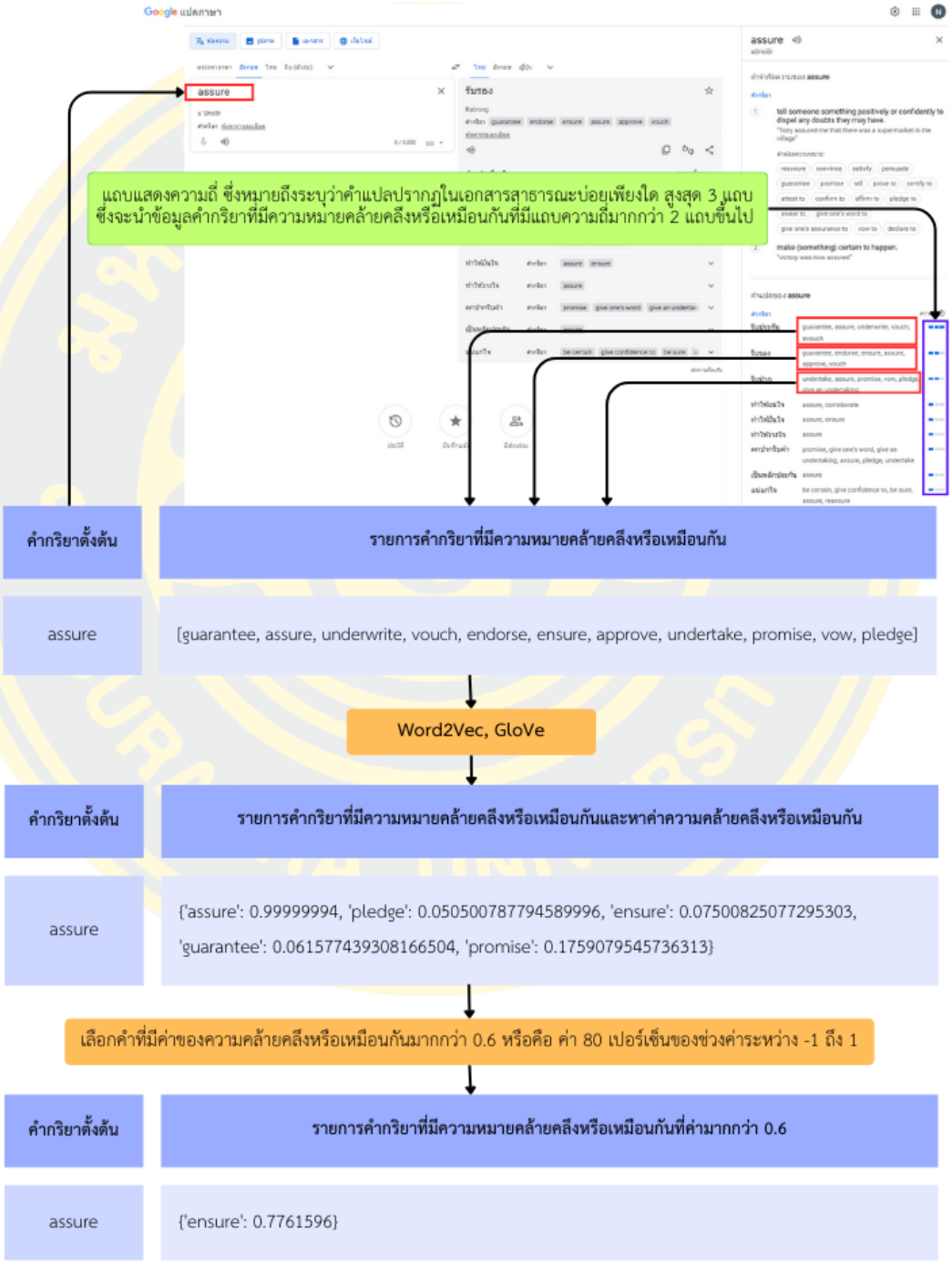
$$Cosine\ Similarity(A,B) = \frac{A \cdot B}{\| A \| \cdot \| B \|} \qquad (1)$$

[10] https://www.oxfordlearnersdictionaries.com/definition/english/

[11] https://www.ldoceonline.com/

[12] https://translate.google.co.th/?hl=th

ซึ่งจะได้ค่าของความคล้ายคลึงหรือเหมือนกันของแต่ละคู่คำออกมา แล้วจึงพิจารณาเลือกคำกริยาที่มีค่าของความคล้ายคลึงหรือเหมือนกันมากกว่า 0.6 หรือคือ ค่า 80 เปอร์เซ็นของช่วงค่าระหว่าง -1 ถึง 1 ตัวอย่างแสดงดังในภาพที่ 10



ภาพที่ 10 ตัวอย่างเลือกรายการคำกริยาที่มีความหมายคล้ายคลึงหรือเหมือนกันของคำกริยาตั้งต้นหนึ่ง ๆ พิจารณาเลือกคำกริยาที่มีค่าความคล้ายคลึงหรือเหมือนกันที่มากกว่า 0.6

จากภาพที่ 10 ตัวอย่างการเก็บรวบรวมรายการคำกริยาที่มีความหมายคล้ายคลึงหรือเหมือนกันของคำกริยาตั้งต้นหนึ่ง ๆ หาค่าเวกเตอร์จาก Word2Vec, glove และนำค่าเวกเตอร์แต่ละคู่คำมาคำนวณความคล้ายแบบคอสไซน์ (cosine similarity) และพิจารณาเลือกคำกริยาที่มีค่าความคล้ายคลึงหรือเหมือนกันกับคำกริยาตั้งต้นที่มากกว่า 0.6 ซึ่งรวบรวมคำกริยาที่ความหมายคล้ายคลึงหรือเหมือนกันได้ทั้งหมด 711 คู่คำ โดยถูกจัดเก็บไว้ในคลังข้อมูลคำย่อชื่อว่า “Verb synonym”

3.1.6 การรวบรวมกฎทางภาษาศาสตร์

งานวิจัยนี้ได้ทำการประยุกต์ใช้คลังข้อมูลที่รวบจาก (Chaisoongnoen & Amphawan, 2020; Chaisoongnoen et al., 2018) สำหรับสร้างและพัฒนากฎทางภาษาศาสตร์นี้ ผู้วิจัยรวบรวมกฎทางภาษาศาสตร์และนำมาแยกคำกริยาออกมา เพื่อนำมาใช้ในกระบวนการสกัดคำสำคัญจากเนื้อหาส่วนหน้าที่ความรับผิดชอบ และส่วนคุณสมบัติที่ต้องการ โดยกฎทางภาษาศาสตร์ถูกแบ่งออกเป็น 4 ประเภทหลัก คือ

1. กฎทางภาษาศาสตร์ที่พิจารณาร่วมกับคำศัพท์เฉพาะแต่ไม่มีคำกริยาอยู่ในประโยค กฎทางภาษาศาสตร์ประเภทนี้จะนำไปใช้ในการสกัดคำสำคัญจากประโยคที่มีการระบุว่ามีคำศัพท์เฉพาะหรือเรียกว่า “Reserved word” หรือแทนว่า “RW” ที่พบร่วมกับคำนาม (Noun) คำบุพบท (Preposition) หรือ คำคุณศัพท์ (Adjective) ตัวอย่างเช่น กฎ: “Noun (NN) + Reserved word (RW)” ⟶ OBJ: (“Noun (NN) + Reserved word (RW)”) จากกฎถ้าในประโยคพบคำที่มีหน้าที่คือ คำนาม (Noun) เกิดขึ้นร่วมกับคำศัพท์เฉพาะ (RW) แบบตามลำดับจากกฎจะดำเนินการสกัดคำสำคัญได้แก่ “คำนาม (Noun) + คำศัพท์เฉพาะ (RW)” ออกมา
2. กฎทางภาษาศาสตร์ที่พิจารณาร่วมกับคำศัพท์เฉพาะและมีคำกริยาอยู่ในประโยค กฎทางภาษาศาสตร์ประเภทนี้จะนำไปใช้ในการสกัดคำสำคัญจากประโยคที่มีการระบุว่ามีคำศัพท์เฉพาะหรือเรียกว่า “Reserved word” หรือแทนว่า “RW” ที่พบร่วมกับคำนาม (Noun) คำบุพบท (Preposition) หรือ คำคุณศัพท์ (Adjective) หรือ คำกริยา (Verb) อยู่ในประโยค ตัวอย่างเช่น กฎ: “Verb (VB) + Noun (NN) + Reserved word (RW)” ⟶ Verb: “Verb (VB) “, OBJ: (“Noun (NN) + Reserved word

(RW)") จากกฎถ้าในประโยคพบคำที่มีหน้าที่คือ คำนาม (Noun) เกิดขึ้นร่วมกับคำศัพท์เฉพาะ (RW) แบบตามลำดับจากกฎจะดำเนินการสกัดคำสำคัญได้แก่ "คำนาม (Noun) + คำศัพท์เฉพาะ (RW)" ออกมา

3. กฎทางภาษาศาสตร์ที่ไม่ได้พิจารณาร่วมกับคำศัพท์เฉพาะและไม่มีคำกริยาอยู่ในประโยค กฎทางภาษาศาสตร์ประเภทนี้จะนำไปใช้ในการสกัดคำสำคัญจากประโยคที่มีการระบุว่ามีคำศัพท์เฉพาะ โดยจะพิจารณาจากคำนาม (Noun) ที่พบร่วมกันกับคำนาม หรือเรียกว่ากลุ่มของคำนาม (Noun phrase), คำคุณศัพท์ (Adjective) ในประโยค แล้วทำการสกัดคำเหล่านี้ออกมา ตัวอย่างเช่น กฎ: "Adjective (JJ) + Noun (NN)" ⟶ OBJ: ("Adjective (JJ) + Noun (NN)") จากกฎถ้าในประโยคพบคำที่มีหน้าที่คือ คำคุณศัพท์ (Adjective) เกิดขึ้นร่วมกับ คำนาม (Noun) แบบตามลำดับจากกฎจะดำเนินการสกัดคำสำคัญได้แก่ "คำคุณศัพท์ (Adjective) + คำนาม (Noun)" ออกมา
4. กฎทางภาษาศาสตร์ที่ไม่ได้พิจารณาร่วมกับคำศัพท์เฉพาะและมีคำกริยาอยู่ในประโยค กฎทางภาษาศาสตร์ประเภทนี้จะนำไปใช้ในการสกัดคำสำคัญจากประโยคที่มีการระบุว่ามีคำศัพท์เฉพาะ โดยจะพิจารณาจากคำนาม (Noun) ที่พบร่วมกันกับคำนาม หรือเรียกว่ากลุ่มของคำนาม (Noun phrase), คำคุณศัพท์ (Adjective) หรือ คำกริยา (Verb) อยู่ในประโยค แล้วทำการสกัดคำเหล่านี้ออกมา ตัวอย่างเช่น กฎ: "Verb (VB) + Adjective (JJ) + Noun (NN)" ⟶ Verb: "Verb (VB) ", OBJ: ("Adjective (JJ) + Noun (NN)") จากกฎถ้าในประโยคพบคำที่มีหน้าที่คือ คำคุณศัพท์ (Adjective) เกิดขึ้นร่วมกับ คำนาม (Noun) แบบตามลำดับจากกฎจะดำเนินการสกัดคำสำคัญได้แก่ "คำคุณศัพท์ (Adjective) + คำนาม (Noun)" ออกมา

ซึ่งแสดงรายละเอียดกฎทางภาษาศาสตร์นี้ที่ภาคผนวก ก ในตารางที่ 14, 15, 16 และ17

3.1.6 การสร้างกฎในการจำแนกประเภทของประโยค

จากการศึกษาและสำรวจข้อมูลในประกาศรับสมัครงานในส่วนหน้าที่ความรับผิดชอบ (Responsibility) พบว่าในแต่ละรูปแบบของประโยคบ่งบอกถึงบทบาทหน้าที่ความรับผิดชอบชัดเจน ไม่พบประโยคลักษณะอื่น และข้อมูลในส่วนของคุณสมบัติที่ต้องการ (Qualification) พบว่าในประโยคมีการปรากฏรูปแบบของคำและคำสำคัญในประโยคที่สามารถบ่งบอกได้ว่าประโยคนั้น ๆ เป็น

ประโยคที่เกี่ยวเนื่องกับข้อมูล อายุ เพศ สัญชาติ การศึกษาที่ต้องการ ประสบการณ์ที่ต้องการ และทักษะที่จำเป็น ดังนั้นผู้วิจัยจึงดำเนินการสร้างกฎเพื่อการจำแนกประเภทของประโยค โดยพิจารณาจากคำเป้าหมาย (Target word) และรูปแบบคำในประโยคที่สามารถบ่งบอกได้ว่าประโยคนั้น ๆ เป็นประโยคประเภทใด เพื่อสามารถใช้วิธีการติดป้ายกำกับ (label) ประเภทที่เกี่ยวข้องให้กับแต่ละประโยค โดยใช้กฎการจำแนกประเภทของประโยคที่กำหนดดังตารางที่ 6 อธิบายถึงคำหลักที่ใช้ในกฎการจำแนกประเภทของประโยคแต่ละประเภท และรูปแบบของคำที่เหมาะสมที่จะเป็นสัญลักษณ์สำหรับแต่ละประเภท

ตารางที่ 6 แสดงคำเป้าหมายและรูปแบบของคำที่ใช้ในกฎการจำแนกประเภทของประโยค

ประเภท	**คำเป้าหมาย**
อายุ (age)	“age”, “year old”, “year of age” ซึ่งปรากฏร่วมกับตัวเลข
เพศ (sex)	“gender”, “all gender”, “male”, “female”
สัญชาติ (nationality)	“nationality”, “nationalities”, “national”, “citizen”
การศึกษา (Education)	“degree”, “bachelor”, “education”, “graduated”, “diploma”, “doctorate”, “associate”, “vocational”
ประสบการณ์ (Experience)	“at least”, “year of experience”, “experience”, “background”, คำว่า “experience” หรือ “background” ซึ่งปรากฏร่วมกับตัวเลข
ทักษะที่จำเป็น (Skills)	“well”, “advance”, “demonstrate strength”, “strong”, “good”, “deep”, “native”, “fluency”, “great”, “advanced”, “excellent”, “knowledge”, “certification”, “demonstrated”, “extensive”, “intermediate”, “superior”, “proficient”, “high”, “basic”, “depth”, “fluent”, “adept”, “expertise”, “successful”, “hands on”, “passion”, “able”, “understand”, “fully”, “awareness”, “interest”, “proficiency”,, “attention”, “passionate”, “capable”, “expert”,

ประเภท	คำเป้าหมาย
	“clear understand”, “familiar”, “familiarity”, “common”, “foundation”, “skill”

ตัวอย่างเช่น

i) ประโยคที่เกี่ยวกับอายุ (age):

ตัวอย่างประโยค: “Applicants should be between 25 and 35 years old.”

กฎการจำแนกประเภท: พบว่าในประโยคปรากฏคำว่า”years old” และมีการใช้ตัวเลขเพื่อบ่งบอกอายุ ประโยคนี้จึงถูกจำแนกประเภทเป็นอายุ “age”

ii) ประโยคที่เกี่ยวกับเพศ (sex):

ตัวอย่างประโยค: “We are an equal opportunity employer and welcome applicants of all genders.”

กฎการจำแนกประเภท: พบว่าในประโยคปรากฏคำ “gender” หรือ “all gender” ประโยคนี้จึงถูกจำแนกประเภทเป็นเพศ “sex”

iii) ประโยคที่เกี่ยวกับสัญชาติ (nationality):

ตัวอย่างประโยค: “Applicants must be a Thai national or have a valid work permit.”

กฎการจำแนกประเภท: พบว่าในประโยคปรากฏคำ “national” ประโยคนี้จึงถูกจำแนกประเภทเป็นสัญชาติ “nationality”

iv) ประโยคที่เกี่ยวกับการศึกษา (Education):

ตัวอย่างประโยค: “Candidates should have a bachelor’s degree in a related field.”

กฎการจำแนกประเภท: พบว่าในประโยคปรากฏคำ “degree” ประโยคนี้จึงถูกจำแนกประเภทเป็นการศึกษา “Education”

v) ประโยคที่เกี่ยวกับประสบการณ์ (Experience):

ตัวอย่างประโยค: “We require at least 3 years of relevant experience in the industry.”

กฎการจำแนกประเภท: พบว่าในประโยคปรากฏคำ “years of experience” หรือ “experience” ร่วมกับตัวเลข ประโยคนี้จึงถูกจำแนกประเภทเป็นประสบการณ์ “Experience”

vi) ประโยคที่เกี่ยวกับทักษะที่จำเป็น (Skills):

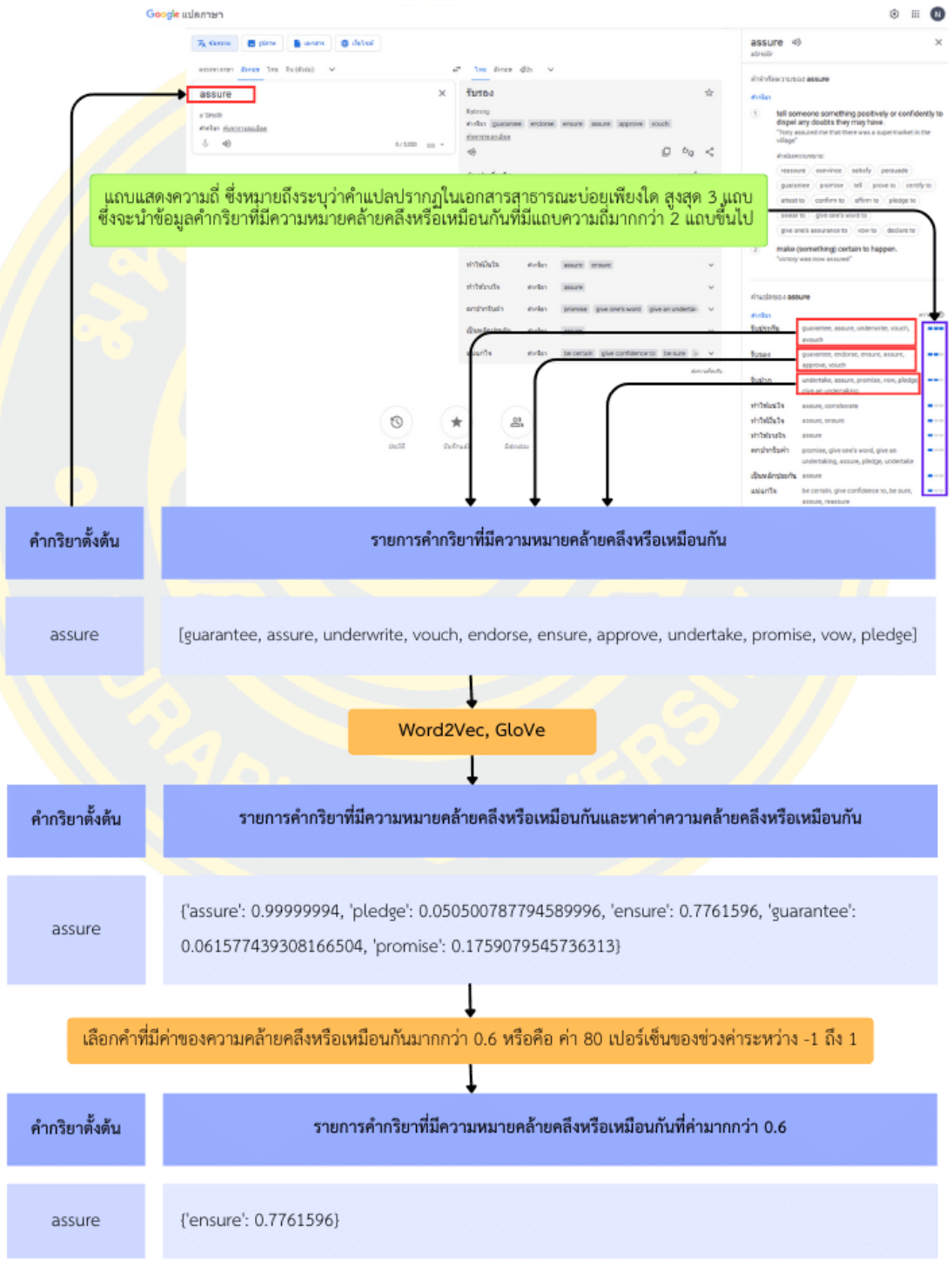
ตัวอย่างประโยค: “We are looking for candidates with advanced proficiency in Microsoft Office.”

กฎการจำแนกประเภท: พบว่าในประโยคปรากฏคำที่บ่งบอกระดับความสามารถหรือคุณสมบัติที่เกี่ยวข้องกับทักษะ ประโยคนี้จึงถูกจำแนกประเภทเป็นทักษะที่จำเป็น “Skills”

โดยการใช้กฎการจำแนกประเภทของประโยคเหล่านี้ สามารถวิเคราะห์เพื่อค้นหาข้อมูลที่เกี่ยวข้องในประกาศรับสมัครงานอย่างมีประสิทธิภาพ

3.2 การจัดกลุ่มประกาศงาน

ในขั้นตอนนี้จะทำการวิเคราะห์และประมวลผลชื่อตำแหน่งงานในแต่ละประกาศงานอย่างละเอียด เพื่อดำเนินการจัดกลุ่มประกาศงาน



ภาพที่ 11 ตัวอย่างการจัดกลุ่มประกาศงานจนได้ชื่อกลุ่มตำแหน่งงาน

จากในภาพที่ 11 เป็นตัวอย่างขั้นตอนการจัดกลุ่ม ดังนี้

ขั้นตอนที่ 1 ใช้เทคนิคการประมวลผลข้อความเพื่อทำความสะอาดและปรับปรุงหรือลบข้อมูลที่ไม่เกี่ยวข้องในชื่อตำแหน่งงานออก

ขั้นตอนที่ 2 นำชื่อตำแหน่งที่ทำความสะอาดแล้ว นำมานับจำนวนของถี่เพื่อจัดอันดับชื่อตำแหน่งงานที่มีจำนวนประกาศงานที่มากที่สุดได้มา 20 อันดับจากชื่อประกาศงานในข้อมูลที่รวบรวมมา

ขั้นตอนที่ 3 ต่อมาดำเนินการนำชื่อตำแหน่งงาน 20 ชื่อนี้มาจับคู่ชื่อตำแหน่งงานให้ครบคู่แล้วทำการหา Word2Vec และความคล้ายแบบคอสไซน์ (cosine similarity) ของแต่ละคู่ ถ้าค่าของชื่อคู่ไหนมีค่ามากกว่า 0.7 จะนำชื่อตำแหน่งงานที่ไปเปรียบเทียบกับชื่อตำแหน่งตั้งต้นแทนในชื่อตำแหน่งตั้งต้น จนได้มาทั้งหมด 11 กลุ่มชื่อตำแหน่งงาน

ขั้นตอนที่ 4 นำทั้ง 11 ชื่อตำแหน่งงานนี้มาจับคู่หาค่า Word2Vec และความคล้ายแบบคอสไซน์ (cosine similarity) อีกครั้งซึ่งไม่มีคู่ที่มีค่ามากกว่า 0.7 ดังนั้นจึงได้ชื่อตำแหน่งงานมาทั้งหมด 11 กลุ่มตำแหน่งงาน ได้แก่

1. Business intelligence
2. Cloud computing
3. Cyber security
4. Data analyst
5. Data engineer
6. Data scientist
7. Database administrator
8. IT support
9. Network engineer
10. Programmer
11. System analyst

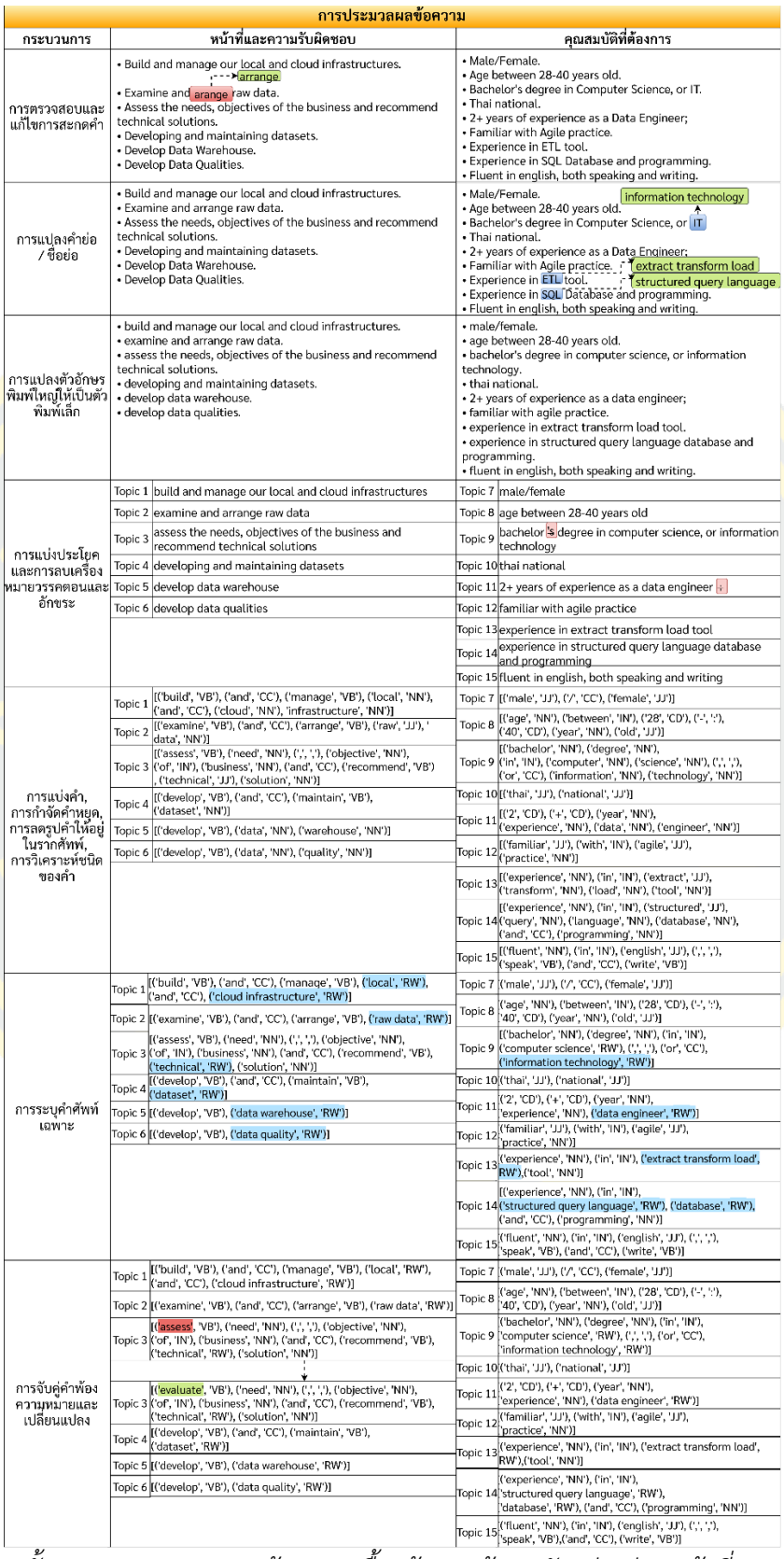
3.3 การประมวลผลข้อความเบื้องต้น

หลังจากสกัดข้อมูลประกาศรับสมัครงานจะได้ข้อมูลอยู่ในไฟล์นามสกุล csv ซึ่งข้อมูลที่ได้มานั้นเป็นข้อมูลที่ยังไม่พร้อมใช้งาน ทำให้ยากต่อการนำไปวิเคราะห์ ดังนั้นในขั้นตอนนี้จึงมีความสำคัญ

ต่อการนำข้อมูลไปวิเคราะห์เพื่อให้มีประสิทธิภาพ โดยจำนะข้อมูลเข้าสู่ขั้นตอนการประมวลผลข้อความเบื้องต้นซึ่งจะดำเนินการกับข้อมูลในส่วนหน้าที่ความรับผิดชอบ และคุณสมบัติที่ต้องการ แสดงตัวอย่างรายละเอียดดังภาพที่ 12 ซึ่งขั้นตอนการประมวลผลข้อความเบื้องต้นเป็นขั้นตอนการดำเนินการที่เชื่อมต่อกันตามลำดับดังภาพที่ 13 แสดงขั้นตอนการประมวลผลข้อความเบื้องต้นจากข้อมูลตัวอย่างส่วนหน้าที่ความรับผิดชอบ และคุณสมบัติที่ต้องการจากประกาศงาน

แอททริบิวต์(Attribute) ที่เลือกจัดเก็บที่เลือกจัดเก็บ	ข้อมูลประกาศงานที่ถูกจัดเก็บลงในฐานข้อมูล
เลขประจำตัวของประกาศงาน	100
วันที่ลงประเทศ	14/02/2023
ชื่อตำแหน่งงาน	Data Engineer
ชื่อองกรค์/บริษัท	AIGEN CO., LTD.
ที่อยู่องค์กร/บริษัท	ห้วยขวาง
หน้าที่และความรับผิดชอบ	• Build and manage our local and cloud infrastructures. • Examine and arange raw data. • Assess the needs, objectives of the business and recommend technical solutions. • Developing and maintaining datasets. • Develop Data Warehouse. • Develop Data Qualities.
คุณสมบัติที่ต้องการ	• Male/Female. • Age between 28-40 years old. • Bachelor's degree in Computer Science, or IT. • Thai national. • 2+ years of experience as a Data Engineer; • Familiar with Agile practice. • Experience in ETL tool. • Experience in SQL Database and programming. • Fluent in english, both speaking and writing.
ลิงก์	https://th.jobsdb.com/th/th/job/data-engineer-%E0%B8%A7-300003002926327?token=0~a6a11339-ab13-47fb-908d-8b846618fd2d§ionRank=12&jobId=jobsdb-th-job-300003002926327
จากเว็บไซต์	jobsdb

ภาพที่ 12 ตัวอย่างส่วนหน้าที่ความรับผิดชอบ และคุณสมบัติที่ต้องการจากประกาศงาน



ภาพที่ 13 ขั้นตอนการประมวลผลข้อความเบื้องต้นจากข้อมูลตัวอย่างส่วนหน้าที่ความรับผิดชอบและคุณสมบัติที่ต้องการจากประกาศงาน

อธิบายแต่ละขั้นตอนดังต่อไปนี้

3.3.1 การตรวจสอบและแก้ไขการสะกดคำ (Word error correction)

ขั้นตอนในการตรวจสอบและแก้ไขการสะกดคำ ใช้วิธีการเปรียบเทียบคำกับคำศัพท์ในพจนานุกรมภาษาอังกฤษ (Dictionary) เพื่อแก้ไขคำที่สะกดผิดให้ถูกต้อง จากภาพที่ 13 จากตัวอย่างในกระบวนการตรวจสอบและแก้ไขการสะกดคำ ประโยคในเนื้อหาส่วนหน้าที่ความรับผิดชอบ “Examine and arange raw data.” ในกรณีนี้คำที่ต้องการแก้คือ “arange” ซึ่งเป็นคำที่สะกดผิดและควรแก้ไขให้ถูกต้อง เมื่อค้นหาในพจนานุกรม คำที่ถูกต้องของคำ “arange” คือ “arrange” จึงดำเนินการแก้ไขคำว่า “arange” เป็น “arrange” เพื่อให้ประโยคเป็นถูกต้อง ดังนั้นประโยคที่ถูกประมวลผลแก้ไขการสะกดคำคือให้ถูกต้องคือ “Examine and arrange raw data.” การแก้ไขคำที่สะกดผิดเป็นการปรับปรุงความถูกต้องและความชัดเจนของข้อความ ซึ่งมีความสำคัญในการสื่อสารและเข้าใจข้อมูลอย่างถูกต้อง

3.3.2 การแปลงคำย่อ/ชื่อย่อ (Abbreviation resolution)

ในส่วนของงานวิจัยนี้ผู้วิจัยดำเนินการพิจารณาถึงคำย่อ/ชื่อย่อที่เกี่ยวข้องกับศาสตร์ทางด้านเทคโนโลยีสารสนเทศเท่านั้น (list of computing and IT acronyms and abbreviations) เพราะคำย่อบางคำสามารถทำให้เกิดความเข้าใจผิดได้ถ้าไม่เป็นเป็นชื่อเต็ม เช่น คำว่า “IT” หรือ “it” เปลี่ยนเป็น “Information Technology” โดยมีตัวอย่างของการแปลงคำย่อ/ชื่อย่อให้อยู่ ในรูปแบบของคำเต็ม/ชื่อเต็มดังตารางที่ 7 แสดงตัวอย่างของการแปลงคำย่อ/ชื่อย่อให้อยู่ในรูปแบบของคำเต็ม/ชื่อเต็ม

ตารางที่ 7 ตัวอย่างของการแปลงคำย่อ/ชื่อย่อให้อยู่ในรูปแบบของคำเต็ม/ชื่อเต็ม

คำย่อ/ชื่อย่อ	**คำเต็ม/ชื่อเต็ม**
AI	Artificial Intelligence
BI	Business Intelligence
CG	Computer Graphics
CX	Customer Experience

คำย่อ/ชื่อย่อ	คำเต็ม/ชื่อเต็ม
DB	Database
EAI	Enterprise Application Integration
GIS	Geographic Information System
HCI	Human Computer Interaction
IT	Information Technology
JDK	Java Development Kit

ตัวอย่างเช่น จากภาพที่ 13 จากตัวอย่างในกระบวนการแปลงคำย่อ/ชื่อย่อ ประโยคในเนื้อหาส่วนคุณสมบัติที่ต้องการ “Bachelor’s degree in Computer Science, or IT.” ในกรณีนี้คำย่อ/ชื่อย่อที่ต้องการแปลงคือ “IT” เมื่อเข้าสู่กระบวนการแปลงคำย่อเป็นคำเต็ม คำย่อ “IT” ถูกแปลงเป็นคำเต็ม “Information Technology” เพื่อทำให้ประโยคถูกต้อง ประโยคที่ถูกประมวลผลแปลงคำย่อ/ชื่อย่อ คือ “Bachelor’s degree in Computer Science, or Information Technology.”

3.3.3 การแปลงตัวอักษรให้เป็นตัวพิมพ์เล็ก (Lowercase conversion)

ในส่วนขั้นตอนกระบวนการเปลี่ยนตัวอักษรทั้งหมดในคำหรือประโยคให้เป็นตัวพิมพ์เล็กทั้งหมด เพื่อทำให้ง่ายต่อการตรวจสอบและการทำงานกับข้อมูลนั้นในขั้นตอนถัดไปของกระบวนการประมวลผลข้อมูล ตัวอย่างเช่น จากภาพที่ 13 จากตัวอย่างในกระบวนการแปลงตัวอักษรพิมพ์ใหญ่ให้เป็นตัวพิมพ์เล็ก ประโยคในเนื้อหาส่วนหน้าที่ความรับผิดชอบ “Develop Data Qualities.” ในกรณีนี้ประโยคที่ต้องการแปลงคือ “Develop Data Qualities.” ดำเนินการเปลี่ยนตัวอักษรทั้งหมดให้เป็นตัวพิมพ์เล็กทั้งหมด ประโยคที่ถูกประมวลผลแปลงตัวอักษรให้เป็นตัวพิมพ์เล็กคือ “develop data qualities.” การแปลงตัวอักษรให้เป็นตัวพิมพ์เล็กช่วยให้ข้อมูลมีความสม่ำเสมอและง่ายต่อการจัดการและประมวลผล และสามารถทำให้ข้อมูลสอดคล้องกับเกณฑ์หรือรูปแบบที่ต้องการในกระบวนการวิเคราะห์และการประมวลผลข้อมูลในภายหลัง

3.3.4 การแบ่งประโยค (Sentence tokenization)

ในส่วนของขั้นตอนการแบ่งประโยค ดำเนินการนำเนื้อหาในส่วนหน้าที่ความรับผิดชอบ และของคุณสมบัติที่ต้องการมาแบ่ง ประโยคออกจากกันในแต่ละประกาศงาน เนื่องจากการสกัดข้อมูลจากเว็บไซต์ทำให้มีการเก็บข้อมูลที่เป็น ภาษา HTML ด้วย จึงพิจารณาการแบ่งประโยคเป็น 2 รูปแบบร่วมกัน คือ I.) พิจารณาโดย “\n” ซึ่งใน ความหมายทางภาษา HTML คือ การขึ้นบรรทัดใหม่ และ II.) พิจารณาจาก จุด (“.”), อัฒภาค (“;”) ที่อยู่ใน ส่วนท้ายของแต่ละประโยค

3.3.5 ลบเครื่องหมายวรรคตอนและอักขระ (Punctuation removal)

เป็นกระบวนการที่ลบเครื่องหมายวรรคตอนและอักขระต่าง ๆ ที่ไม่ใช่ตัวอักษรหรือตัวเลขจากข้อความ เป็นเพื่อลดความแตกต่างที่ไม่จำเป็นในข้อมูล ตัวอย่างของเครื่องหมายวรรคตอนและอักขระที่ถูกลบได้ เช่นเครื่องหมายจุลภาค (,) จุด (.) คอลอน (:) เครื่องหมายคำพูด (“ “) เป็นต้น การลบเครื่องหมายเหล่านี้ช่วยให้ข้อมูลเป็นคำและประโยคที่สะอาดและพร้อมใช้งานในขั้นตอนถัด ๆ ไปของการประมวลผลข้อมูล ตัวอย่างเช่น จากภาพที่ 13 จากตัวอย่างในกระบวนการลบเครื่องหมายวรรคตอนและอักขระ ประโยคในเนื้อหาส่วนคุณสมบัติที่ต้องการ “• 2+ years of experience as a data engineer;” เมื่อเข้าสู่กระบวนการลบเครื่องหมายวรรคตอนและอักขระ ประโยคที่ถูกประมวลผลลบเครื่องหมายวรรคตอนและอักขระคือ “2+ years of experience as a data engineer” ข้อมูลหลังการลบเครื่องหมายวรรคตอนและอักขระจะเหลือเฉพาะตัวอักษรและตัวเลขที่เกี่ยวข้อง ที่ทำให้ข้อมูลเป็นคำและประโยคที่สามารถประมวลผลและวิเคราะห์ให้มีความครบถ้วนในขั้นตอนถัดไปของกระบวนการ

3.3.6 การแบ่งคำ (Word tokenization)

ในส่วนของขั้นตอนนี้ดำเนินการแบ่งรูปของคำในแต่ละประโยค เพื่อลดความแตกต่างกันของข้อมูล ซึ่งสามารถช่วยทำให้ข้อมูลอยู่ในรูปแบบคำ ตัวอย่างเช่น จากภาพที่ 13 จากตัวอย่างในกระบวนการแบ่งคำ ประโยคในเนื้อหาส่วนคุณสมบัติที่ต้องการ “2+ years of experience as a data engineer.” เมื่อเข้าสู่กระบวนการแบ่งคำ ประโยคที่ถูกประมวลผลแบ่งคำคือ (“2”, “+”, “years”, “of”, “experience”, “as”, “a”, “data”, “engineer”) การแบ่งคำจะทำให้คำแต่ละ

คำอยู่ในรูปแบบของคำโดยเอาแต่ละคำที่สำคัญมาตัดออกมาเพื่อนำไปใช้ในการวิเคราะห์และประมวลผลข้อมูลในขั้นตอนถัด ๆ ไป

3.3.7 การลบคำหยุด (Stop word removal)

เป็นกระบวนการลบคำที่ไม่มีความหมายหรือคำหยุดออกจากข้อความ เพื่อทำให้ข้อมูลที่เหลือมีความหมายที่มีประโยชน์มากขึ้นเช่น “a”, “an”, “the” เป็นต้น เป็นคำที่มักจะปรากฏในข้อความแต่ไม่มีความหมายสำคัญ การลบคำเหล่านี้ช่วยทำให้เห็นข้อมูลที่สำคัญชัดเจนขึ้น ตัวอย่างเช่น จากผลลัพธ์ใน 3.3.6 การแบ่งคำ ประโยคที่ถูกประมวลผลแบ่งคำแล้วคือ (“2”, “+”, “years”, “of”, “experience”, “as”, “a”, “data”, “engineer”) เมื่อเข้าสู่กระบวนการลบคำหยุด คำหยุด “of”, “as”, และ “a” ถูกลบออกจากประโยค เพื่อให้เหลือเฉพาะคำที่มีความหมายที่สำคัญเท่านั้น ทำให้ประโยคที่เหลือเป็นดังนี้ (“2”, “+”, “years”, “experience”, “data”, “engineer”) การลบคำที่ไม่จำเป็นช่วยให้การวิเคราะห์และสกัดคำสำคัญจากข้อมูลเป็นไปอย่างมีประสิทธิภาพ

3.3.8 การลดรูปของคำให้อยู่ในรากศัพท์ (Word Stemming and lemmatization)

ในส่วนของขั้นตอนนี้ดำเนินการเพื่อลดความแตกต่างกันของข้อมูล ซึ่งสามารถช่วยทำให้ข้อมูลอยู่ในรูปแบบเดียวกัน ในส่วนขั้นตอนการดำเนินการทำลดรูปของคำให้อยู่ในรากศัพท์ ตัวอย่างเช่น จากผลลัพธ์ใน 3.3.7 การลบคำหยุด ประโยคที่ถูกประมวลผลลบคำหยุดแล้วคือ (“2”, “+”, “years”, “experience”, “data”, “engineer”) ในกรณีนี้เมื่อเข้าสู่กระบวนการลดรูปของคำให้อยู่ในรากศัพท์แล้ว คำว่า “years” จะถูกดำเนินการตัด -s ออก กลายเป็นคำว่า “year” ประโยคที่ผ่านการประมวลผลการลดรูปของคำให้อยู่ในรากศัพท์คือ (“2”, “+”, “year”, “experience”, “data”, “engineer”)

3.3.9 การวิเคราะห์ชนิดของคำ (Parts of speech tagging)

ขั้นตอนการวิเคราะห์หน้าที่ของคำ เป็นกระบวนการที่กำกับหน้าที่หรือประเภทของคำแต่ละคำในประโยคด้วยตัวบ่งชี้ (tag) ที่ระบุลักษณะของคำนั้น ๆ ในประโยค ตัวอย่างเช่น จากผลลัพธ์ที่ได้จาก 3.3.8 การลดรูปของคำให้อยู่ในรากศัพท์ ประโยคที่ผ่านการประมวลผลการลดรูปของคำให้อยู่ในรากศัพท์แล้วคือ (“2”, “+”, “year”, “experience”, “data”, “engineer”) ในกรณีนี้เมื่อเข้าสู่

กระบวนการวิเคราะห์หน้าที่ของคำแล้ว จะถูกระบุหน้าที่ของคำดังนี้ [(“2”, “CD”), (“+”, “CD”), (“year”, “NN”), (“experience”, “NN”), (“data”, “NN”), (“engineer”, “NN”)]

3.3.10 การระบุคำศัพท์เฉพาะทางด้านคอมพิวเตอร์เทคโนโลยี(Reserved word identification)

การค้นหาคำศัพท์เฉพาะในประกาศงาน โดยใช้วิธีการ n-gram ในการเปรียบเทียบแต่ละคำในประโยคกับคำศัพท์เฉพาะทางด้านคอมพิวเตอร์เทคโนโลยีที่ทำการรวบรวมอยู่ในคลังคำศัพท์ในขั้นที่ 3.1.4 ตัวอย่างเช่น จากผลลัพธ์ที่ได้จาก 3.3.9 ประโยคที่ผ่านการประมวลผลการวิเคราะห์ชนิดของคำแล้วคือ “[(“2”, “CD”), (“+”, “CD”), (“year”, “NN”), (“experience”, “NN”), (“data”, “NN”), (“engineer”, “NN”)]” ในกรณีนี้เมื่อเข้าสู่กระบวนการระบุคำศัพท์เฉพาะทางด้านเทคโนโลยีสารสนเทศแล้ว ประโยคที่ถูกประมวลผลได้ “[(“2”, “CD”), (“+”, “CD”), (“year”, “NN”), (“experience”, “NN”), (“data engineer”, “RW”)]” จากประโยคข้างต้นคำว่า “data” และ “engineer” จะถูกระบุเป็นคำศัพท์เฉพาะทางด้านคอมพิวเตอร์เทคโนโลยีเป็นคำว่า “data engineer” เนื่องจากการวิเคราะห์โดยวิธีการ n-gram และนำไปเปรียบเทียบกับคำศัพท์เฉพาะทางด้านคอมพิวเตอร์เทคโนโลยีที่อยู่ในคลังคำศัพท์ผลปรากฏว่า พบคำว่า “data engineer” อยู่ในคลังคำศัพท์เฉพาะทางด้านคอมพิวเตอร์เทคโนโลยี ดังนั้นจึงรวมกลุ่มคำว่า “data” และ “engineer” เป็น “data engineer” และระบุหน้าที่ของคำเป็น “RW” (Reserved word) เป็นต้น

3.3.11 การจับคู่คำพ้องความหมายและการเปลี่ยนแปลง (Synonym matching and transformation)

เป็นกระบวนการที่ใช้ในการจับคู่คำกริยาที่มีความหมายคล้ายกันที่ได้เก็บรวบรวมไว้ในขั้นตอนที่ 3.1.5 นำมาจับคู่คำกริยาที่มีความหมายคล้ายกัน และนำคู่นั้นมาหาความถี่ที่พบในชุดข้อมูล ถ้าคำไหนพบความถี่มากกว่าก็จะดำเนินการใช้กริยานั้นแทนอีกคำ ซึ่งช่วยให้ในการดูผลสรุปของข้อมูลทั้งหมดเห็นคำกริยาที่หมายถึง การกระทำต่อสิ่งหนึ่ง ๆ ที่ความหมายรวมเป็นคำหนึ่ง ๆ ช่วยลดความหลากหลายของคำที่ความหายคล้านคลึงกัน ตัวอย่างการใช้การจับคู่คำพ้องความหมายและการเปลี่ยนแปลง (Synonym matching and transformation) ของคู่คำกริยาพ้องความหมายที่นำไป

หาความถี่ในชุดข้อมูลที่จะนำมาทำการทดลองแล้ว (“buy”: 18 , “purchase”: 32) โดยจะดำเนินการแปลงคำ “buy” เป็น “purchase” เป็นต้น

3.3 การจัดโครงสร้างและการรวบรวมเนื้อหา

ในขั้นตอนนี้จะดำเนินการกับข้อมูลส่วนหน้าที่ความรับผิดชอบ และส่วนคุณสมบัติที่ต้องการที่ประมวลผลข้อมูลในกระบวนการก่อนหน้านี้ นำมาติดป้ายกำกับประเภทที่เกี่ยวข้อง โดยใช้กฎการจำแนกประเภทที่กำหนดไว้ก่อนหน้านี้แสดงดังภาพที่ 14 ตัวอย่างการติดป้ายกำกับประเภทที่เกี่ยวข้องให้กับแต่ละประโยค โดยใช้กฎการจำแนกประเภทที่กำหนดไว้

การประมวลผลข้อความ		กฎเกณฑ์ในการจำแนกประเภทของประโยค
หน้าที่และความรับผิด		ป้ายกำกับประเภท
Topic 1	[('build', 'VB'), ('and', 'CC'), ('manage', 'VB'), ('local', 'RW'), ('and', 'CC'), ('cloud infrastructure', 'RW')]	responsibility
Topic 2	[('examine', 'VB'), ('and', 'CC'), ('arrange', 'VB'), ('raw data', 'RW')]	responsibility
Topic 3	[('evaluate', 'VB'), ('need', 'NN'), (',', ','), ('objective', 'NN'), ('of', 'IN'), ('business', 'NN'), ('and', 'CC'), ('recommend', 'VB'), ('technical', 'RW'), ('solution', 'NN')]	responsibility
Topic 4	[('develop', 'VB'), ('and', 'CC'), ('maintain', 'VB'), ('dataset', 'RW')]	responsibility
Topic 5	[('develop', 'VB'), ('data warehouse', 'RW')]	responsibility
Topic 6	[('develop', 'VB'), ('data quality', 'RW')]	responsibility
คุณสมบัติที่ต้องการ		
Topic 7	[('male', 'JJ'), ('/', 'CC'), ('female', 'JJ')]	sex
Topic 8	[('age', 'NN'), ('between', 'IN'), ('28', 'CD'), ('-', ':'), ('40', 'CD'), ('year', 'NN'), ('old', 'JJ')]	age
Topic 9	[('bachelor', 'NN'), ('degree', 'NN'), ('in', 'IN'), ('computer science', 'RW'), (',', ','), ('or', 'CC'), ('information technology', 'RW')]	education
Topic 10	[('thai', 'JJ'), ('national', 'JJ')]	nationality
Topic 11	[('2', 'CD'), ('+', 'CD'), ('year', 'NN'), ('of', 'IN'), ('experience', 'NN'), ('data engineer', 'RW')]	experience
Topic 12	[('familiar', 'JJ'), ('with', 'IN'), ('agile', 'JJ'), ('practice', 'NN')]	skill
Topic 13	[('experience', 'NN'), ('in', 'IN'), ('extract transform load', 'RW'), ('tool', 'NN')]	experience
Topic 14	[('experience', 'NN'), ('in', 'IN'), ('structured query language', 'RW'), ('database', 'RW'), ('and', 'CC'), ('programming', 'NN')]	experience
Topic 15	[('fluent', 'NN'), ('in', 'IN'), ('english', 'JJ'), (',', ','), ('speak', 'VB'), ('and', 'CC'), ('write', 'VB')]	skill

ภาพที่ 14 ตัวอย่างการติดป้ายกำกับประเภทที่เกี่ยวข้องให้กับแต่ละประโยค โดยใช้กฎการจำแนกประเภทที่กำหนดไว้

จากภาพที่ 14 เมื่อเนื้อหาทั้งส่วนหน้าที่ความรับผิดชอบและส่วนคุณสมบัติที่ต้องการถูกประมวลผลข้อความแล้ว จะถูกนำมาติดป้ายกำกับประเภท ในเนื้อหาส่วนหน้าที่ความรับผิดชอบ กำหนดติดป้ายกำกับให้กับทุกประโยคเป็น “responsibility” ส่วนเนื้อหาในส่วนคุณสมบัติที่ต้องการในประโยคแรกหรือประโยคที่ 7 เมื่อนับต่อจากประโยคในส่วนหน้าที่ความรับผิดชอบ “[(“male”, “JJ”), (“/”, “CC”), (“female”, “JJ”)]” ถูกระบุป้ายกำกับประเภทเป็นเพศคือ “sex” เพราะพบว่ามีคำเป้าหมาย (Target word) จากตารางที่ 6 คำว่า “male” และ “female”

ประโยคที่ 8 “[(“age”, “NN”), (“between”, “IN”), (“28”, “CD”), (“-”, “:”), (“40”, “CD”), (“year”, “NN”), (“old”, “JJ”)]” ถูกระบุป้ายกำกับประเภทเป็นอายุคือ “age” เพราะพบว่ามีคำเป้าหมาย คำว่า “age” และ “year old” ซึ่งปรากฏร่วมกับตัวเลข

ประโยคที่ 9 “[(“bachelor”, “NN”), (“degree”, “NN”), (“in”, “IN”), (“computer science”, “RW”), (“,”, “,”), (“or”, “CC”), (“information technology”, “RW”)]” ถูกระบุป้ายกำกับประเภทเป็นการศึกษาคือ “education” เพราะพบว่ามีคำเป้าหมาย คำว่า “degree”, “bachelor”

ประโยคที่ 10 “[(“thai”, “JJ”), (“national”, “JJ”)]” ถูกระบุป้ายกำกับประเภทเป็นสัญชาติคือ “nationality” เพราะพบว่ามีคำเป้าหมาย คำว่า “national”

ประโยคที่ 11 “[(“2”, “CD”), (“+”, “CD”), (“year”, “NN”), (“of”, “IN”), (“experience”, “NN”), (“data engineer”, “RW”)]” ถูกระบุป้ายกำกับประเภทเป็นประสบการณ์คือ “experience” เพราะพบว่ามีคำเป้าหมาย คำว่า “year of experience”, “experience” ซึ่งปรากฏร่วมกับตัวเลข

ประโยคที่ 12 “[(“familiar”, “JJ”), (“with”, “IN”), (“agile”, “JJ”), (“practice”, “NN”)]” ถูกระบุป้ายกำกับประเภทเป็นทักษะที่จำเป็นคือ “skill” เพราะพบว่ามีคำเป้าหมาย คำว่า “familiar”

ประโยคที่ 13 “[(“experience”, “NN”), (“in”, “IN”), (“extract transform load”, “RW”),(“tool”, “NN”)]” ถูกระบุป้ายกำกับประเภทเป็นประสบการณ์คือ “experience” เพราะพบว่ามีคำเป้าหมาย คำว่า “experience”

ประโยคที่ 14 “[(“experience”, “NN”), (“in”, “IN”), (“structured query language”,

“RW”), (“database”, “RW”), (“and”, “CC”), (“programming”, “NN”)]” ถูกระบุป้ายกำกับประเภทเป็นประสบการณ์คือ “experience” เพราะพบว่ามีคำเป้าหมาย คำว่า “experience”

ประโยคที่ 15 “[(“fluent”, “NN”), (“in”, “IN”), (“english”, “JJ”), (“,”, “,”),(“speak”, “VB”), (“and”, “CC”), (“write”, “VB”)]” ถูกระบุป้ายกำกับประเภทเป็นทักษะที่จำเป็นคือ “skill” เพราะพบคำเป้าหมาย “fluent” ในประโยคนี้โดยอ้างอิงจากรายการคำเป้าหมายในตารางที่ 6 ซึ่งสอดคล้องกับประเภทนี้

ต่อมาประโยคที่มีป้ายกำกับจะถูกจัดระเบียบไปยังโครงสร้างที่กำหนดและรวบรวมไว้ ซึ่งในโครงสร้างของแต่ละประเภทจะแตกต่างกันขึ้นอยู่กับบริบทของประโยคในแต่ละประเภทนั้น ๆ เพื่อสร้างรูปแบบโครงสร้างที่ยังคงความเชื่อมโยงของคำในประโยคไว้ จึงดำเนินการพิจารณาจากรูปแบบประโยคบริบทของคำที่เกิดในประโยคประเภทนั้น ๆ และพิจารณาชนิดหน้าที่ของคำที่เกิดขึ้นร่วมกันร่วมกับกฎทางภาษาศาสตร์ที่รวบรวมไว้ก่อนหน้านี้ ซึ่งในโครงสร้างประกอบด้วย

1. คำเป้าหมาย ซึ่งได้จากการกำหนดกฎการจำแนกประเภทของประโยค
2. คำหรือกลุ่มของคำที่เกิดขึ้นร่วมกันตามกฎทางภาษาศาสตร์ประกอบด้วย คำนาม (Noun) คำศัพท์เฉพาะ (Reserved word) และคำคุณศัพท์ (Adjective) ที่ไม่รวมคำเป้าหมาย และคำกริยา (Verb) โดยเรียกแทนกลุ่มคำนี้ว่า “Object” หรือย่อว่า “Obj”
3. ชนิดของคำที่เป็น คำกริยา (Verb)

โดยโครงสร้างการจัดเก็บคำสำคัญจากข้อมูลมีทั้งหมด 7 รูปแบบโครงสร้าง ได้แก่

1. โครงสร้างการจัดเก็บคำสำคัญจากข้อมูลประเภทของบทบาทหน้าที่ความรับผิดชอบ
2. โครงสร้างการจัดเก็บคำสำคัญจากข้อมูลประเภทของอายุ
3. โครงสร้างการจัดเก็บคำสำคัญจากข้อมูลประเภทของเพศ
4. โครงสร้างการจัดเก็บคำสำคัญจากข้อมูลประเภทของสัญชาติ
5. โครงสร้างการจัดเก็บคำสำคัญจากข้อมูลประเภทของการศึกษาที่ต้องการ
6. โครงสร้างการจัดเก็บคำสำคัญจากข้อมูลประเภทของประสบการณ์ที่ต้องการ
7. โครงสร้างการจัดเก็บคำสำคัญจากข้อมูลประเภทของทักษะที่จำเป็น

ซึ่งจะแสดงโครงสร้างของแต่ละประเภทที่ภาคผนวก ข ในตารางที่ 18, 19, 20, 22, 23 และ 24 โดยจะอธิบายโครงสร้างของแต่ละประเภทที่แตกต่างกันดังต่อไปนี้

3.3.1 การกำหนดโครงสร้างการจัดเก็บคำสำคัญจากข้อมูลประเภทของหน้าที่ความรับผิดชอบ

จากการศึกษาและพิจารณากฎการจำแนกประเภทรูปแบบประโยคในประเภทหน้าที่ความรับผิดชอบในประกาศงาน พบว่ารูปแบบประโยคที่ถูกระบุหน้าที่ของคำแล้ว ประกอบไปด้วยชนิดของคำที่เป็น คำกริยา (Verb) ที่บ่งบอกถึง การกระทำ ตามด้วย คำนาม (Noun) หรือ คำศัพท์เฉพาะ (Reserved word) ที่บ่งบอกถึง สิ่งที่จะกระทำหรือถูกกระทำ และคำคุณศัพท์ (Adjective) ที่อธิบายถึงลักษณะหรือคุณภาพของคำนามและคำศัพท์เฉพาะ ดังนั้นจึงกำหนดโครงสร้างการจัดเก็บข้อมูลบทบาทหน้าที่ความรับผิดชอบให้อยู่ในรูปแบบของ

< “POST_ID”: x_1, “PART”: “responsibility”, “TOPIC_ID”: x_2, “VERB”: x_3, “OBJ”: x_4 >

โดย

i) “POST_ID” คือ ชื่อฟิลด์ลำดับของประกาศงานที่กำลังจัดเก็บข้อมูล โดยจัดเก็บข้อมูลที่ค่า x1 ให้เป็นตัวเลขที่ระบุลำดับของประกาศงานที่ต้องการจัดเก็บข้อมูล

ii) “PART” คือ ชื่อฟิลด์ส่วนของข้อมูลที่กำลังจัดเก็บ ในกรณีนี้ข้อมูลที่ถูกจัดเก็บคือค่า “responsibility” ซึ่งระบุว่าจัดเก็บข้อมูลเกี่ยวกับหน้าที่ความรับผิดชอบ

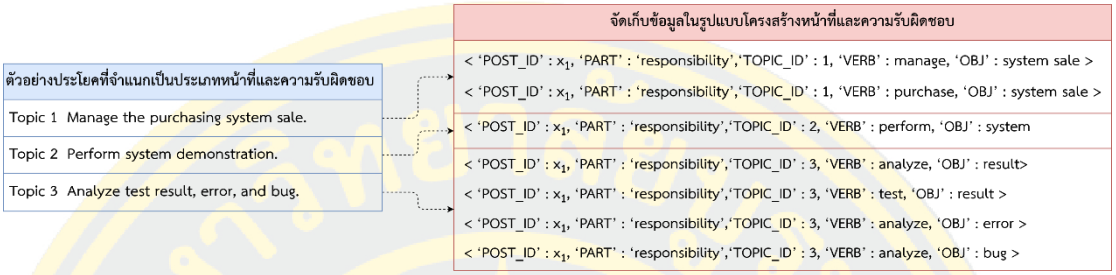
iii) “TOPIC_ID” คือ ชื่อฟิลด์ลำดับของประโยคในประกาศงานที่กำลังจัดเก็บข้อมูล โดยจัดเก็บข้อมูลที่ค่า x_2 ให้เป็นตัวเลขที่ระบุลำดับของประโยคที่กำลังจัดเก็บข้อมูล

iv) “VERB” คือ ชื่อฟิลด์คำกริยา (Verb) โดยจัดเก็บข้อมูลที่ค่า x_3 ให้เป็นคำกริยา

v) “OBJ” คือ ชื่อฟิลด์คำหรือกลุ่มของคำที่เกิดขึ้นร่วมกันตามกฎทางภาษาศาสตร์ โดยจัดเก็บข้อมูลที่ค่า x_4 ให้เป็นคำหรือกลุ่มของคำที่เกิดขึ้นร่วมกันตามกฎทางภาษาศาสตร์

ซึ่งตัวอย่างผลลัพธ์การจัดเก็บคำสำคัญจากประโยคหนึ่ง ๆ ซึ่งอยู่ในประเภทข้อมูลหน้าที่ความรับผิดชอบให้อยู่ในโครงสร้างที่กำหนดแสดงดังภาพที่ 15 ตัวอย่างประโยคที่ถูกระบุว่าเป็นประเภทหน้าที่ความรับผิดชอบ จะถูกจัดเก็บคำสำคัญให้อยู่ในโครงสร้างที่กำหนด เช่นจากประโยคที่ 1 จากภาพที่ 15 “Manage the purchasing system sale.” เมื่อดำเนินการจัดเก็บข้อมูลในรูปแบบโครงสร้างหน้าที่ความรับผิดชอบ จะได้ “< “POST_ID”: x1, “PART”: “responsibility”,

“TOPIC_ID”: “1”, “VERB”: “manage”, “OBJ”: “system sale” >” และ “< “POST_ID”: x1, “PART”: “responsibility”, “TOPIC_ID”: “1”, “VERB”: “purchase”, “OBJ”: “system sale” >”



ภาพที่ 15 ตัวอย่างประโยคที่ถูกจำแนกประเภทเป็นประโยคหน้าที่ความรับผิดชอบและถูกจัดเก็บคำสำคัญให้อยู่ในรูปแบบโครงสร้างที่กำหนด

3.3.2 การกำหนดโครงสร้างการจัดเก็บคำสำคัญจากข้อมูลประเภทของอายุ

จากการศึกษาและพิจารณากฎการจำแนกประเภทรูปแบบประโยคในประเภทอายุ พบว่ารูปแบบประโยคมีคำเป้าหมาย (Target word) คำว่า “age”, “year”, “year old” ที่ปรากฏร่วมกับตัวเลขและจากการศึกษาเพิ่มเติมมีคำศัพท์ที่กำหนดช่วงอายุที่ต้องการจัดเก็บอยู่ในประโยคนั้น ๆ ซึ่งผู้วิจัยได้พิจารณากำหนดสัญลักษณ์ในการแทน ค่าขอบเขตช่วงอายุมี 3 รูปแบบ คือ “>” คือความหมายว่า มากกว่า, “<” คือความหมายว่า น้อยกว่า และ “-” คือความหมายว่า ถึง ตัวอย่างแสดงดังตารางที่ 8

ตารางที่ 8 ตัวอย่างแสดงคำศัพท์ที่กำหนดช่วงอายุและการแทนค่าด้วยสัญลักษณ์

คำศัพท์ที่กำหนดช่วงอายุ	**แทนค่าด้วยสัญลักษณ์**
to	-
not over	<
below	<
more than	>
between ... to ...	-

คำศัพท์ที่กำหนดช่วงอายุ	**แทนค่าด้วยสัญลักษณ์**
up	>
over	>
above	>

ดังนั้นจึงกำหนดโครงสร้างการจัดเก็บข้อมูลช่วงอายุให้อยู่ในรูปแบบของ < “POST_ID”: x_1, “PART”: “age”, “TOPIC_ID”: x_2, “NUMBER1”: x_3, “SIGN”: x_4, “NUMBER2”: x_5 >

โดย

i) “POST_ID” คือ ชื่อฟิลด์ลำดับของประกาศงานที่กำลังจัดเก็บข้อมูล โดยจัดเก็บข้อมูลที่ค่า x_1 เป็นตัวเลขเพื่อระบุลำดับของประกาศงานที่เราต้องการจัดเก็บข้อมูล

ii) “PART” คือ ชื่อฟิลด์ส่วนของข้อมูลที่กำลังจัดเก็บ ในกรณีนี้ข้อมูลที่ถูกจัดเก็บคือค่า “age” ซึ่งระบุว่าจัดเก็บข้อมูลเกี่ยวกับอายุ

iii) “TOPIC_ID” คือ ชื่อฟิลด์ลำดับของประโยคในประกาศงานที่กำลังจัดเก็บ โดยจัดเก็บข้อมูลที่ค่า x_2 เป็นตัวเลขที่ระบุลำดับของประโยคที่กำลังจัดเก็บข้อมูล

iv) “NUMBER1” คือ ชื่อฟิลด์ที่ระบุตัวเลขช่วงอายุต่ำสุด โดยจัดเก็บข้อมูลที่ค่า x_3 เป็นตัวเลขที่ระบุช่วงอายุต่ำสุด

v) “SIGN” คือ ชื่อฟิลด์สัญลักษณ์ที่ใช้แบ่งช่วงอายุ โดยจัดเก็บข้อมูลที่ค่า x_4 เป็นสัญลักษณ์ (เช่น “>“, “<“, “>=” และ “-”)

vi) “NUMBER2” คือ ชื่อฟิลด์ที่ระบุตัวเลขอายุ หรือช่วงอายุสูงสุดที่กำลังจัดเก็บ โดยจัดเก็บข้อมูลที่ค่า x_5 เป็นตัวเลขที่ระบุอายุ หรือช่วงอายุสูงสุด

รายละเอียดแสดงที่ภาคผนวก ข ในตารางที่ 18 และตัวอย่างผลลัพธ์ของการจัดเก็บคำสำคัญจากประโยคหนึ่ง ๆ ซึ่งอยู่ในประเภทข้อมูลอายุ ให้อยู่ในโครงสร้างที่กำหนดแสดงดังภาพที่ 16 ตัวอย่างประโยคที่ถูกระบุว่าเป็นประเภทอายุ จะถูกจัดเก็บคำสำคัญให้อยู่ในโครงสร้างที่

กำหนด เช่นจากประโยคที่ 1 จากภาพที่ 16 “Age 25-40 years.” เมื่อดำเนินการจัดเก็บข้อมูลในรูปแบบโครงสร้างอายุจะได้ “< “POST_ID”: x1, “PART”: “age”, “TOPIC_ID”: “1”, “NUMBER1”: “25”, “SIGN”: “-”, “NUMBER2”: “40” >”

ตัวอย่างประโยคที่จำแนกเป็นประเภทอายุ	จัดเก็บข้อมูลในรูปแบบโครงสร้างอายุ
Topic 1 Age 25-40 years.	< POST_ID : x_1, PART: age, TOPIC_ID : 1, NUMBER1: 25, SIGN : -, NUMBER2 : 40 >
Topic 2 Age between 21 to 24 year old.	< POST_ID : x_1, PART: age, TOPIC_ID : 2, NUMBER1: 21, SIGN : -, NUMBER2 : 24 >
Topic 3 Age not over 35 year old.	< POST_ID : x_1, PART: age, TOPIC_ID : 3, NUMBER1: NULL, SIGN : <, NUMBER2 : 35 >

ภาพที่ 16 ตัวอย่างผลลัพธ์จากการจัดเก็บข้อมูลจากประโยคที่ถูกจำแนกให้อยู่ในประเภทอายุและถูกจัดเก็บคำสำคัญให้อยู่ในรูปแบบโครงสร้างที่กำหนด

3.3.3 การกำหนดโครงสร้างการจัดเก็บคำสำคัญจากข้อมูลประเภทของเพศ

จากการศึกษาและพิจารณากฎการจำแนกประเภทรูปแบบประโยคในประเภทเพศ พบว่ารูปแบบประโยคมีคำเป้าหมาย (Target word) คำว่า “all gender” ,”male”, “female” ซึ่งจากการพิจารณาถ้าพบประโยคว่า “All gender” จะแทนที่ด้วย “male, female and other” ซึ่งเป็นคำสำคัญที่ต้องการจัดเก็บและเป็นคำศัพท์ที่ระบุถึงเพศ ดังนั้นจึงกำหนดโครงสร้างการจัดเก็บข้อมูลเพศให้อยู่ในรูปแบบของ

< “POST_ID”: x_1, “PART”: “sex”, “TOPIC_ID”: x_2, “SEX”: x_3 >

โดย

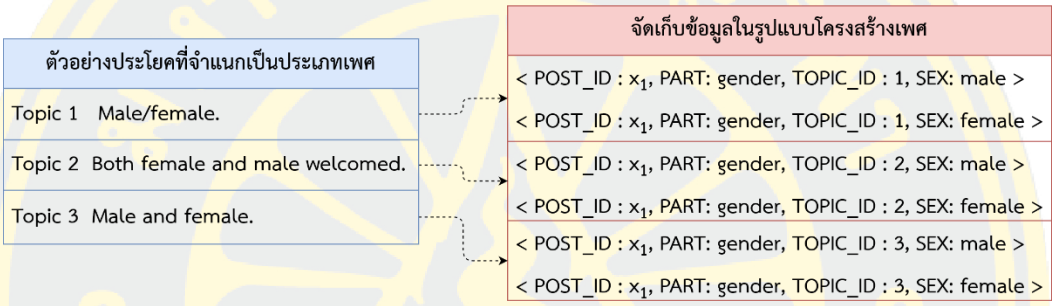
i) “POST_ID” คือ ชื่อฟิลด์ลำดับของประกาศงานที่กำลังจัดเก็บข้อมูล โดยจัดเก็บข้อมูลที่ค่า x_1 เป็นตัวเลขที่ระบุลำดับของประกาศงานที่ต้องการจัดเก็บข้อมูล

ii) “PART” คือ ชื่อฟิลด์ส่วนของข้อมูลที่กำลังจัดเก็บ ในกรณีนี้ข้อมูลที่ถูกจัดเก็บคือค่า “sex” ซึ่งระบุว่าจัดเก็บข้อมูลเกี่ยวกับเพศ

iii) “TOPIC_ID” คือ ชื่อฟิลด์ลำดับของประโยคในประกาศงานที่กำลังจัดเก็บ โดยจัดเก็บข้อมูลที่ค่า x2 ให้เป็นตัวเลขที่ระบุลำดับของประโยคที่กำลังจัดเก็บข้อมูล

iv) “SEX” คือ ชื่อฟิลด์เพศ โดยจัดเก็บข้อมูลที่ค่า x_4 เป็นคำหรือวลีที่ระบุเพศ (เช่น “male”, “female” หรือ “other”)

รายละเอียดที่ภาคผนวก ข ในตารางที่ 19 และตัวอย่างผลลัพธ์ของการจัดเก็บคำสำคัญจากประโยคหนึ่ง ๆ ซึ่งอยู่ในประเภทข้อมูลเพศให้อยู่ในโครงสร้างที่กำหนดแสดงดังภาพที่ 17 ตัวอย่างประโยคที่ถูกระบุว่าเป็นประเภทเพศ จะถูกจัดเก็บคำสำคัญให้อยู่ในโครงสร้างที่กำหนด เช่นจากประโยคที่ 1 จากภาพที่ 17 “Male/female.” เมื่อดำเนินการจัดเก็บข้อมูลในรูปแบบโครงสร้างเพศจะได้ “< “POST_ID”: x1, “PART”: “gender”, “TOPIC_ID”: “1”, “SEX”: “male” >” และ “< “POST_ID”: x1, “PART”: “gender”, “TOPIC_ID”: “1”, “SEX”: “female” >”



ภาพที่ 17 ตัวอย่างผลลัพธ์จากการจัดเก็บข้อมูลจากประโยคที่ถูกจำแนกให้อยู่ในประเภทเพศและถูกจัดเก็บคำสำคัญให้อยู่ในรูปแบบโครงสร้างที่กำหนด

3.3.4 การกำหนดโครงสร้างการจัดเก็บคำสำคัญจากข้อมูลประเภทของสัญชาติ

จากการศึกษาและพิจารณากฎการจำแนกประเภทรูปแบบประโยคในประเภทสัญชาติ พบว่ารูปแบบประโยคมีคำเป้าหมาย (Target word) คำว่า “nationality”, “nationalities”, “national”, “citizen” วลีใดวลีหนึ่ง ซึ่งจะอยู่ด้านหลังของคำสำคัญที่ต้องการจัดเก็บซึ่งเป็นคำศัพท์ที่ระบุถึงสัญชาติที่ต้องการ ดังนั้นจึงกำหนดโครงสร้างการจัดเก็บข้อมูลสัญชาติให้อยู่ในรูปแบบของ

< “POST_ID”: x_1, “PART”: “nationality”, “TOPIC_ID”: “x_2”, “NATIONAL”: “x3” >

โดย

i) “POST_ID” คือ ชื่อฟิลด์ลำดับของประกาศงานที่กำลังจัดเก็บข้อมูล โดยจัดเก็บข้อมูลที่ค่า x_1 เป็นตัวเลขที่ระบุลำดับของประกาศงานที่ต้องการจัดเก็บข้อมูล

ii) “PART” คือ ชื่อฟิลด์ส่วนของข้อมูลที่กำลังจัดเก็บ ในกรณีนี้ข้อมูลที่ถูกจัดเก็บคือค่า “nationality” ซึ่งระบุว่าจัดเก็บข้อมูลเกี่ยวกับสัญชาติ

iii) “TOPIC_ID” คือ ชื่อฟิลด์ลำดับของประโยคในประกาศงานที่กำลังจัดเก็บ โดยจัดเก็บข้อมูลที่ค่า x_2 ให้เป็นตัวเลขที่ระบุลำดับของประโยคที่กำลังจัดเก็บข้อมูล

iv) “NATIONAL” คือ ชื่อฟิลด์สัญชาติ โดยจัดเก็บข้อมูลที่ค่า x_3 เป็นคำหรือวลีที่ระบุสัญชาติ

รายละเอียดที่ภาคผนวก ข ในตารางที่ 20 และตัวอย่างผลลัพธ์ของการจัดเก็บคำสำคัญจากประโยคหนึ่ง ๆ ซึ่งอยู่ในประเภทข้อมูลสัญชาติ ให้อยู่ในโครงสร้างที่กำหนดแสดงดังภาพที่ 19 ตัวอย่างประโยคที่ถูกระบุว่าเป็นประเภทสัญชาติ จะถูกจัดเก็บคำสำคัญให้อยู่ในโครงสร้างที่กำหนด เช่นจากประโยคที่ 1 จากภาพที่ 18 “Thai nationality.” เมื่อดำเนินการจัดเก็บข้อมูลในรูปแบบโครงสร้างสัญชาติจะได้ “< “POST_ID”: “x1”, “PART”: “nationality”, “TOPIC_ID”: “1”, “NATIONAL”: “thai” >”

ตัวอย่างประโยคที่จำแนกเป็นประเภทสัญชาติ	จัดเก็บข้อมูลในรูปแบบโครงสร้างสัญชาติ
Topic 1 Thai nationality.	< POST_ID : x_1, PART: nationality, TOPIC_ID : 1, ‘NATIONAL’: thai >
Topic 2 Thai nationality only.	< POST_ID : x_1, PART: nationality, TOPIC_ID : 2, ‘NATIONAL’: thai >
Topic 3 Thai nationals.	< POST_ID : x_1, PART: nationality, TOPIC_ID : 3, ‘NATIONAL’: thai >

ภาพที่ 18 ตัวอย่างผลลัพธ์จากการจัดเก็บข้อมูลจากประโยคที่ถูกจำแนกให้อยู่ในประเภทสัญชาติให้อยู่ในโครงสร้างที่กำหนด

3.3.5 การกำหนดโครงสร้างการจัดเก็บคำสำคัญจากข้อมูลประเภทของการศึกษาที่ต้องการ

จากการศึกษาและพิจารณากฎการจำแนกประเภทรูปแบบประโยคในประเภทการศึกษาที่ต้องการในประกาศงาน พบว่ารูปแบบประโยคมีคำเป้าหมาย (Target word) คำว่า “bachelor”, “diploma”, “doctorate”, “associate”, “vocational” วลีใดวลีหนึ่งซึ่งเป็นคำสำคัญที่ต้องการจัดเก็บในระดับการศึกษาและมีคำศัพท์ที่ระบุถึงสาขาวิชาที่เกี่ยวข้องในการศึกษาที่ต้องการ ดังนั้นจึงกำหนดโครงสร้างการจัดเก็บข้อมูลการศึกษาที่ต้องการให้อยู่ในรูปแบบของ

< “POST_ID”: x_1, “PART”: “education”, “TOPIC_ID”: x_2, “DEGREE”: x_3, “FIELD”: x_4 >

โดย

i) “POST_ID” คือ ชื่อฟิลด์ลำดับของประกาศงานที่กำลังจัดเก็บข้อมูล โดยจัดเก็บข้อมูลที่ค่า x_1 เป็นตัวเลขที่ระบุลำดับของประกาศงานที่ต้องการจัดเก็บข้อมูล

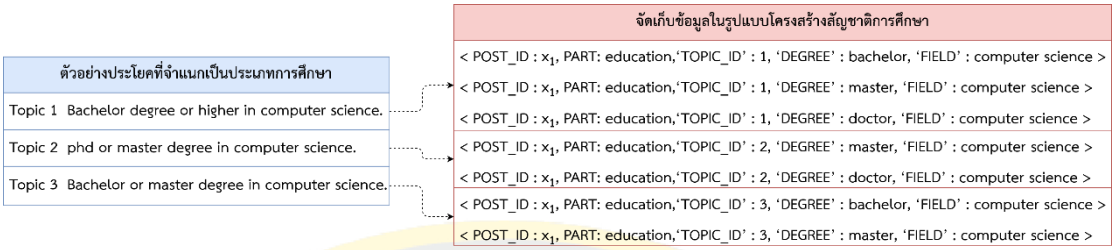
ii) “PART” คือ ชื่อฟิลด์ส่วนของข้อมูลที่กำลังจัดเก็บ ในกรณีนี้ข้อมูลที่ถูกจัดเก็บคือค่า “education” ซึ่งระบุว่าจัดเก็บข้อมูลเกี่ยวกับการศึกษาที่ต้องการ

iii) “TOPIC_ID” คือ ชื่อฟิลด์ลำดับของประโยคในประกาศงานที่กำลังจัดเก็บ โดยจัดเก็บข้อมูลที่ค่า x_2 ให้เป็นตัวเลขที่ระบุลำดับของประโยคที่กำลังจัดเก็บข้อมูล

iv) “DEGREE” คือ ชื่อฟิลด์ระดับการศึกษา โดยจัดเก็บข้อมูลที่ค่า x_3 เป็นคำหรือวลีที่ระบุระดับการศึกษา

v) “FIELD” คือ ชื่อฟิลด์สาขาวิชา โดยจัดเก็บข้อมูลที่ค่า x_4 เป็นคำหรือวลีที่ระบุสาขาวิชา

รายละเอียดที่ภาคผนวก ข ในตารางที่ 22 และตัวอย่างผลลัพธ์ของการจัดเก็บคำสำคัญจากประโยคหนึ่ง ๆ ซึ่งอยู่ในประเภทข้อมูลการศึกษา ให้อยู่ในโครงสร้างที่กำหนดแสดงดังภาพที่ 20 ตัวอย่างประโยคที่ถูกระบุว่าเป็นประเภทการศึกษา จะถูกจัดเก็บคำสำคัญให้อยู่ในโครงสร้างที่กำหนด เช่นจากประโยคที่ 1 จากภาพที่ 19 “Bachelor degree or higher in computer science.” เมื่อดำเนินการจัดเก็บข้อมูลในรูปแบบโครงสร้างการศึกษาจะได้ “< “POST_ID”: x1 , “PART”: “education”, “TOPIC_ID”: “1”, “DEGREE”: “bachelor”, “FIELD”: “computer science” >”, “ < “POST_ID”: x1 , “PART”: “education”, “TOPIC_ID”: “1”, “DEGREE”: “master”, “FIELD”: “computer science” >” และ “< “POST_ID”: x1 , “PART”: “education”, “TOPIC_ID”: “1”, “DEGREE”: “doctor”, “FIELD”: “computer science” >”



ภาพที่ 19 ตัวอย่างผลลัพธ์จากการจัดเก็บข้อมูลจากประโยคที่ถูกจำแนกให้อยู่ในประเภทการศึกษา และถูกจัดเก็บคำสำคัญให้อยู่ในรูปแบบโครงสร้างที่กำหนด

3.3.6 การกำหนดโครงสร้างการจัดเก็บคำสำคัญจากข้อมูลประเภทของประสบการณ์ที่ต้องการ

จากการศึกษาและพิจารณากฎการจำแนกประเภทรูปแบบประโยคในประเภทประสบการณ์ที่ต้องการในประกาศงาน พบว่ารูปแบบประโยคมีคำเป้าหมาย (Target word) คำว่า “at least”, “year of experience”, “experience”, “knowledge”, “background” และกรณีที่คำว่า “experience”, “knowledge” หรือ “background” ปรากฏร่วมกับตัวเลข กรณีใดกรณีหนึ่ง และมีคำสำคัญที่ต้องการจัดเก็บซึ่งเป็นคำศัพท์ที่ระบุถึงประสบการณ์ที่ต้องการ จากการศึกษาเพิ่มเติมพบคำที่บ่งบอกถึงระดับความเชี่ยวชาญคำว่า “basic”, “good”, “ fundamental”, “high”, “deep”, “depth” ปรากฏขึ้นก่อนหน้าคำเป้าหมาย (Target word) คำว่า “experience”, “knowledge”, “background” ดังนั้นจึงกำหนดโครงสร้างการจัดเก็บข้อมูลประสบการณ์ที่ต้องการให้อยู่ในรูปแบบของ

< “POST_ID”: x_1, “PART”: “experience”, “TOPIC_ID”: x_2, “NUMBER1”: x_3, “SIGN”: x_4, “NUMBER2”: x_5, “EXPERIENCE_LEVEL1”: x_6, “EXPERIENCE_LEVEL2”: x_7, “VERB”: x_8, “OBJ”: x_9 >

โดย

i) “POST_ID” คือ ชื่อฟิลด์ลำดับของประกาศงานที่กำลังจัดเก็บข้อมูล โดยจัดเก็บข้อมูลที่ค่า x_1 เป็นตัวเลขที่ระบุลำดับของประกาศงานที่ต้องการจัดเก็บข้อมูล

ii) “PART” คือ ชื่อฟิลด์ส่วนของข้อมูลที่กำลังจัดเก็บ ในกรณีนี้ข้อมูลที่ถูกจัดเก็บคือค่า “experience” ซึ่งระบุว่าจัดเก็บข้อมูลเกี่ยวกับประสบการณ์ที่ต้องการ

iii) “TOPIC_ID” คือ ชื่อฟิลด์ลำดับของประโยคในประกาศงานที่กำลังจัดเก็บ โดยจัดเก็บข้อมูล

ที่ค่า x_2 ให้เป็นตัวเลขที่ระบุลำดับของประโยคที่กำลังจัดเก็บข้อมูล

iv) “NUMBER1” คือ ชื่อฟิลด์ที่ระบุตัวเลขช่วงปีต่ำสุด โดยจัดเก็บข้อมูลที่ค่า x_3 เป็นตัวเลขที่ระบุตัวเลขช่วงปีต่ำสุด

v) “SIGN” คือ ชื่อฟิลด์เครื่องหมายทางคณิตศาสตร์ที่ใช้แบ่งช่วงปี โดยจัดเก็บข้อมูลที่ค่า x_4 เป็นเครื่องหมายทางคณิตศาสตร์ (เช่น <, >, <=, >=)

vi) “NUMBER2” คือ ชื่อฟิลด์ที่ระบุตัวเลขจำนวนปี หรือช่วงปีสูงสุด โดยจัดเก็บข้อมูลที่ค่า x_5 เป็นตัวเลขที่ระบุจำนวนปี หรือช่วงปีสูงสุด

vii) “EXPERIENCE_LEVEL1” คือ ชื่อฟิลด์คำระดับความเชี่ยวชาญ โดยจัดเก็บข้อมูลที่ค่า x_6 เป็นคำที่บ่งบอกถึงระดับความเชี่ยวชาญได้แก่คำว่า “basic”, “good”, “specialized”, “fundamental”, “high”, “deep”, “depth” คำใดคำหนึ่ง

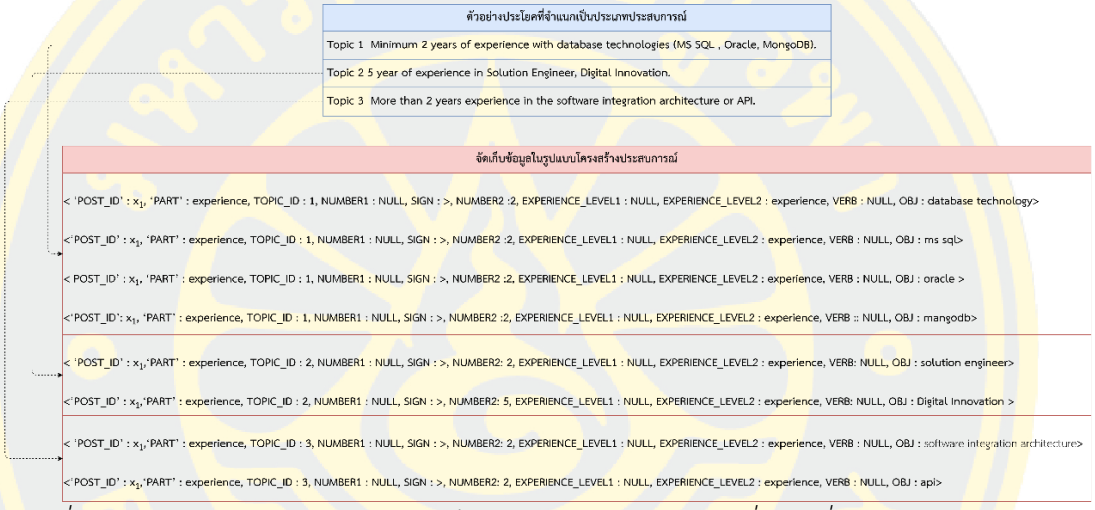
viii) “EXPERIENCE_LEVEL2” คือ ชื่อฟิลด์คำเป้าหมาย โดยจัดเก็บข้อมูลที่ค่า x_7 เป็นคำเป้าหมายคำว่า “experience”, “knowledge”, “background” คำใดคำหนึ่ง

ix) “VERB” คือ ชื่อฟิลด์คำกริยา (Verb) โดยจัดเก็บข้อมูลที่ค่า x_8 เป็นคำหรือวลีที่เป็นคำกริยา

x) “OBJ” คือ ชื่อฟิลด์คำหรือกลุ่มของคำที่เกิดขึ้นร่วมกันตามกฎทางภาษาศาสตร์ โดยจัดเก็บข้อมูลที่ค่า x9 ให้เป็นคำหรือกลุ่มของคำที่เกิดขึ้นร่วมกันตามกฎทางภาษาศาสตร์

รายละเอียดที่ภาคผนวก ข ในตารางที่ 23 และตัวอย่างผลลัพธ์ของการจัดเก็บคำสำคัญจากประโยคหนึ่ง ๆ ซึ่งอยู่ในประเภทข้อมูลประสบการณ์ ให้อยู่ในโครงสร้างที่กำหนดแสดงดังภาพที่ 21 ตัวอย่างประโยคที่ถูกระบุว่าเป็นประเภทประสบการณ์ จะถูกจัดเก็บคำสำคัญให้อยู่ในโครงสร้างที่กำหนด เช่นจากประโยคที่ 1 จากภาพที่ 20 “Minimum 2 years of experience with database technologies (MS SQL , Oracle, MongoDB).” เมื่อดำเนินการจัดเก็บข้อมูลในรูปแบบโครงสร้างประสบการณ์จะได้ “< “POST_ID”: x1, “PART”: “experience”, “TOPIC_ID”: “1” , “NUMBER1”: NULL, “SIGN”: “>”, “NUMBER2” : “2”, “EXPERIENCE_LEVEL1”: NULL, “EXPERIENCE_LEVEL2”: “experience”, “VERB”: NULL, “OBJ”: “database technology”>”, “<“POST_ID”: x1, “PART”: “experience”, “TOPIC_ID”: “1”, “NUMBER1”: NULL, “SIGN”: “>”, “NUMBER2” : “2”, “EXPERIENCE_LEVEL1”: NULL, “EXPERIENCE_LEVEL2”: “experience”, “VERB”: NULL,

“OBJ”: “microsoft sql”>”, “< “POST_ID”: x1 , “PART”: “experience”, “TOPIC_ID”: “1 ”, “NUMBER1 ”: NULL, “SIGN”: “>”, “NUMBER2 ” : “2 ”, “EXPERIENCE_LEVEL1”: NULL, “EXPERIENCE_LEVEL2”: “experience”, “VERB”: NULL, “OBJ”: “oracle” >” และ “ <“POST_ID”: x1 , “PART”: “experience”, “TOPIC_ID”: “1 ”, “NUMBER1 ”: NULL, “SIGN”: “>”, “NUMBER2” : “2”, “EXPERIENCE_LEVEL1”: NULL, “EXPERIENCE_LEVEL2”: “experience”, “VERB” : NULL, “OBJ”: “mangodb”>”



ภาพที่ 20 แสดงผลลัพธ์จากการจัดเก็บข้อมูลจากประโยคหนึ่ง ๆ ซึ่งอยู่ในประเภทข้อมูลประสบการณ์ที่ต้องการและถูกจัดเก็บคำสำคัญให้อยู่ในรูปแบบโครงสร้างที่กำหนด

3.3.7 การกำหนดโครงสร้างการจัดเก็บคำสำคัญจากข้อมูลประเภทของทักษะที่จำเป็น

จากการศึกษาและพิจารณากฎการจำแนกประเภทรูปแบบประโยคในประเภททักษะที่จำเป็นในประกาศงาน พบว่ารูปแบบประโยคมีคำเป้าหมาย (Target word) คำว่า “skill”, “good”, “excellent”, “strong”, “able” และ “basic” เป็นต้น มีคำสำคัญที่ต้องการจัดเก็บซึ่งเป็นคำศัพท์ที่ระบุถึงทักษะจำเป็น จากการศึกษาเพิ่มเติมพบว่ามีคำเป้าหมาย (Target word) ที่สามารถบ่งบอกถึงระดับความสามารถของทักษะที่มักปรากฏขึ้นก่อนหน้าคำเป้าหมาย (Target word) ที่สามารถบ่งบอกถึงลักษณะของทักษะ เช่นคำว่า “fundamental”, “strong”, “good”, “deep”, “fluency”, “great”, “advanced”, “excellent” เป็นต้น ปรากฏขึ้นก่อนหน้าคำเป้าหมาย (Target word) เช่นคำว่า “understand”, “able”, “hand on”, “familiar” เป็นต้นดังตารางที่ 8 แสดงคำที่บ่งบอกถึงระดับความสามารถของทักษะที่มักปรากฏขึ้นก่อนหน้าคำที่บ่งบอกถึง

ลักษณะของทักษะ

ตารางที่ 9 แสดงคำที่บ่งบอกถึงระดับความสามารถของทักษะที่มักปรากฏขึ้นก่อนหน้าคำที่บ่งบอกถึงลักษณะของทักษะ

คำเป้าหมาย (Target word) ที่สามารถบ่งบอกถึงระดับความสามารถของทักษะที่ปรากฏขึ้นก่อนหน้าคำเป้าหมาย (Target word) ที่สามารถบ่งบอกถึงลักษณะของทักษะ	“fundamental”, “strong”, “good”, “deep”, “fluency”, “great”, “advanced”, “excellent”, “superior”, “proficient”, “high”, “basic”, “depth”, “fluent”, “intermediate”, “specialized”, “adept”, “fluency”, “expertise”, “advanced”
คำเป้าหมาย (Target word) ที่สามารถบ่งบอกถึงลักษณะของทักษะ	“understand”, “able”, “hand on”, “familiar”, “passion”, “capability”, “fully”, “awareness”, “interest”, “keen”, “proficiency”, “attention”, “hand on”, “passionate”, “familiar”, “common”

ดังนั้นจึงกำหนดโครงสร้างการจัดเก็บข้อมูลทักษะที่จำเป็นให้อยู่ในรูปแบบของ

< “POST_ID”: x_1, “PART”: “skill”, “TOPIC_ID”: x_2, “SKILL_LEVEL1”: x_3, “SKILL_LEVEL2”: x_4, “VERB”: x_5, “OBJ”: x_6 >

โดย

i) “POST_ID” คือ ชื่อฟิลด์ลำดับของประกาศงานที่กำลังจัดเก็บข้อมูล โดยจัดเก็บข้อมูลที่ค่า x_1 เป็นตัวเลขที่ระบุลำดับของประกาศงานที่ต้องการจัดเก็บข้อมูล

ii) “PART” คือ ชื่อฟิลด์ส่วนของข้อมูลที่กำลังจัดเก็บ ในกรณีนี้ข้อมูลที่ถูกจัดเก็บคือค่า “skill” ซึ่งระบุว่าจัดเก็บข้อมูลเกี่ยวกับทักษะที่จำเป็น

iii) “TOPIC_ID” คือ ชื่อฟิลด์ลำดับของประโยคในประกาศงานที่กำลังจัดเก็บ โดยจัดเก็บข้อมูลที่ค่า x_2 ให้เป็นตัวเลขที่ระบุลำดับของประโยคที่กำลังจัดเก็บข้อมูล

iv) “SKILL_LEVEL1” คือ ชื่อฟิลด์คำระดับความสามารถของทักษะ โดยจัดเก็บข้อมูลที่ค่า x_3 เป็นคำที่บ่งบอกถึงระดับความสามารถของทักษะ

v) “SKILL_LEVEL2” คือ ชื่อฟิลด์คำที่บ่งบอกถึงลักษณะของทักษะ โดยจัดเก็บข้อมูลที่ค่า x_4 เป็นคำที่ที่บ่งบอกถึงลักษณะของทักษะ

vi) “VERB” คือ ชื่อฟิลด์คำกริยา (Verb) โดยจัดเก็บข้อมูลที่ค่า x_5 เป็นคำหรือวลีที่เป็นคำกริยา

vii) “OBJ” คือ ชื่อฟิลด์คำหรือกลุ่มของคำที่เกิดขึ้นร่วมกันตามกฎทางภาษาศาสตร์ โดยจัดเก็บข้อมูลที่ค่า x_6 ให้เป็นคำหรือกลุ่มของคำที่เกิดขึ้นร่วมกันตามกฎทางภาษาศาสตร์

รายละเอียดที่ภาคผนวก ข ในตารางที่ 24 และตัวอย่างผลลัพธ์ของการจัดเก็บคำสำคัญจากประโยคหนึ่ง ๆ ซึ่งอยู่ในประเภทข้อมูลทักษะที่จำเป็น ให้อยู่ในโครงสร้างที่กำหนดแสดงดังภาพที่ 21 ตัวอย่างประโยคที่ถูกระบุว่าเป็นประเภททักษะที่จำเป็น จะถูกจัดเก็บคำสำคัญให้อยู่ในโครงสร้างที่กำหนด เช่นจากประโยคที่ 1 จากภาพที่ 21 “Excellent communicate english.” เมื่อดำเนินการจัดเก็บข้อมูลในรูปแบบโครงสร้างทักษะที่จำเป็นจะได้ “< “POST_ID”: x1, “PART”: “skill”, “TOPIC_ID”: “1 ”, “SKILL_LEVEL1”: “excellent”, “SKILL_LEVEL2”: NULL, “VERB”: “communicate”, “OBJ”: “english” >”

ภาพที่ 21 ตัวอย่างผลลัพธ์จากการจัดเก็บข้อมูลจากประโยคที่ถูกจำแนกให้อยู่ในประเภททักษะที่จำเป็นและถูกจัดเก็บคำสำคัญให้อยู่ในรูปแบบโครงสร้างที่กำหนด

เมื่อข้อมูลทั้งหมดถูกดำเนินการจัดระเบียบโครงสร้างแล้ว จึงดำเนินการรวบรวมข้อมูลทั้งหมดที่อยู่ในรูปแบบโครงสร้างใส่ลงในฐานข้อมูล

3.4 การส่งข้อมูลให้ผู้ใช้ (User)

หลังจากดำเนินการจัดโครงสร้างและรวบรวมเนื้อหาลงในฐานข้อมูล ผู้ใช้ (User) สามารถเข้าถึงข้อมูลในฐานข้อมูล ผู้ใช้สามารถดำเนินการดึงข้อมูลตามความต้องการ เพื่อนำไปวิเคราะห์และค้นหาข้อมูลได้แบบเจาะลึก และสามารถระบุระดับความละเอียดของข้อมูลตามความต้องการจากส่วนประกอบของข้อมูล ทำให้ผู้ใช้ได้รับข้อมูลข้อสรุปที่ต้องการ โดยข้อมูลที่ส่งให้ผู้ใช้ ประกอบด้วย

1. ชื่อตำแหน่งงาน
2. หน้าที่ความรับผิดชอบ
3. อายุ
4. เพศ
5. สัญชาติ
6. ระดับการศึกษาและสาขา
7. ประสบการณ์
8. ทักษะ

ซึ่งสามารถนำข้อมูลในส่วนต่าง ๆ ในฐานข้อมูล นำไปแสดงผลเป็นแผงรายงานข้อมูล (Dashboard) หน้าจอที่สรุปข้อมูลต่าง ๆ ให้อยู่ในหน้าจอเดียว ทำให้ผู้ใช้สามารถเข้าใจข้อมูลได้อย่างรวดเร็ว และสามารถดูข้อมูลรวม ๆ หรือรายละเอียดตามความต้องการ ทั้งนี้ทั้งนั้นข้อมูลที่ถูกแสดงบนแผงรายงานควรสามารถทำให้ผู้ใช้สามารถทำเปรียบเทียบและวิเคราะห์ข้อมูลอย่างมีประสิทธิภาพเพื่อใช้ในการตัดสินใจและวางแผน

บทที่ 4 ผลการดำเนินงาน

ในบทนี้จะอธิบายถึงผลการทดลองจากวิธีการค้นพบข้อมูลที่เกิดขึ้นร่วมกันในการประกาศรับสมัครงานที่เกี่ยวข้องกับคอมพิวเตอร์ (วิธีการ CID-CJA) ที่ได้นำเสนอไปก่อนหน้านี้ โดยในการทดลองผู้วิจัยได้ทำการรวบรวมชุดข้อมูลประกาศรับสมัครงานทางด้านเทคโนโลยีสารสนเทศแบบสุ่มมาจำนวนทั้งหมด 4,280 ประกาศงาน จากเว็บไซต์ค้นหางานที่มีชื่อเสียง 3 เว็บไซต์ ได้แก่

1. www.jobsdb.com เป็นจำนวน 1,905 ประกาศงาน
2. www.linkedin.com เป็นจำนวน 1,504 ประกาศงาน
3. www.indeed.com เป็นจำนวน 871 ประกาศงาน

ซึ่งเก็บรวบรวมชุดข้อมูลนี้ตั้งแต่เดือนสิงหาคม 2020 ถึงเมษายน 2023 โดยเป็นข้อมูลประกาศรับสมัครงานเฉพาะศาสตร์ทางด้านเทคโนโลยีสารสนเทศ ภายในประเทศไทยที่เป็นภาษาอังกฤษเท่านั้น ในวิธีการได้ดำเนินการจัดกลุ่มประกาศงานทั้งหมด แบ่งออกเป็นกลุ่มตามชื่อกลุ่มงานทั้งหมด 11 กลุ่ม โดยจำนวนประกาศงานของแต่ละกลุ่มงานไม่เท่ากัน แสดงรายละเอียดดังตารางที่ 10

ตารางที่ 10 จำนวนประกาศงานของแต่ละกลุ่มงานจาก 3 เว็บไซต์

ลำดับ	ชื่อตำแหน่งงาน	จำนวนประกาศงาน	จำนวนคำสำคัญที่ถูกต้อง
1	Data Scientist	518	26,736
2	Data Engineer	721	29,255
3	Data Analyst	354	18,691
4	Cyber Security	348	11,455
5	Database Administrator	309	10,028
6	Network Engineer	262	10,702
7	Programmer	732	30,218
8	System analyst	253	7,566
9	IT Support	314	6,046
10	Business Intelligence	251	26,736
11	Cloud Engineer	218	11,306

การทดสอบการทำงานของวิธีการ CID-CJA ด้วยการทดสอบ 3 ด้าน ได้แก่ (i) ความถูกต้องในการจัดกลุ่มโฆษณางาน (ii) ความแม่นยำในการสกัดคำสำคัญ และ (iii) ประสิทธิภาพในการแยกและสร้างโครงสร้างข้อมูลที่สำคัญ การทดสอบเหล่านี้เป็นเพื่อประเมินประสิทธิภาพของ CID-CJA

4.1 ความถูกต้องในการจัดกลุ่มประกาศงาน

จากบทก่อนหน้านี้เราทราบได้ว่าการจัดกลุ่มประกาศงานตามชื่อตำแหน่งงาน ช่วยให้ข้อมูลสามารถเปรียบเทียบกับกลุ่มต่าง ๆ ได้อย่างมีประสิทธิภาพ ดังนั้นผู้วิจัยจึงทำการประเมินค่าความถูกต้อง เพื่อวัดประสิทธิภาพวัดประสิทธิภารการจัดกลุ่มประกาศงาน

$$\text{Accuracy} = \frac{\text{จำนวนตำแหน่งที่ถูกจัดกลุ่มได้}}{\text{จำนวนตำแหน่งของกลุ่มที่ถูกต้อง}} \quad (2)$$

ตารางที่ 11 จำนวนประกาศงานที่ถูกต้องและจำนวนประกาศงานที่ถูกจัดกลุ่มได้และค่าความถูกต้อง

ชื่อตำแหน่งงาน	จำนวนประกาศงานที่ถูกต้อง	จำนวนประกาศงานที่ถูกจัดกลุ่มได้	ค่าความถูกต้อง
Business Intelligence	518	518	100
Cloud computing	721	721	100
Cyber Security	354	354	100
Data analyst	348	348	100
Data Engineer	309	309	100
Data Scientist	262	262	100
Database Administrator	732	732	100
IT Support	253	253	100
Network Engineer	314	314	100
Programmer	251	251	100
System analyst	218	218	100
รวมจำนวนประกาศงานทั้งหมด/ค่าเฉลี่ย	4280	4280	100 (ค่าเฉลี่ย)

จากตาราง 11 ค่าความถูกต้อง 100% แสดงให้เห็นว่าการจัดกลุ่มประกาศงานตามชื่อตำแหน่งงานได้ผลลัพธ์ที่ถูกต้อง

4.2 ความแม่นยำในการสกัดคำสำคัญ

การสกัดคำสำคัญให้ได้ผลลัพธ์ที่มีความถูกต้องเป็นแนวทางที่ใช้ในการวัดประสิทธิภาพของกระบวนการสกัดคำสำคัญในงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ข้อมูลหรือข้อความ การวัดความแม่นยำนี้ช่วยให้เราเข้าใจว่ากระบวนการสกัดคำสำคัญที่เราใช้มีประสิทธิภาพแค่ไหนในการคัดเลือกคำสำคัญจากข้อมูล จึงใช้สูตร

$$\text{Accuracy} = \frac{\text{จำนวนคำสำคัญที่สกัดได้ถูกต้อง}}{(\text{จำนวนคำสำคัญที่สกัดได้ถูกต้อง} + \text{จำนวนคำสำคัญที่สกัดได้ไม่ถูกต้อง})} \quad (3)$$

ในตารางที่ 12 การสกัดคำสำคัญสำหรับตำแหน่งงานต่าง ๆ และเปรียบเทียบจำนวนคำสำคัญที่ถูกต้องกับจำนวนคำสำคัญที่สกัดได้ถูกต้อง และค่าความถูกต้องรวมเป็นเปอร์เซ็นต์ตามสูตรที่กำหนดด้านบน

ตารางที่ 12 จำนวนคำสำคัญที่ถูกต้องและจำนวนสำคัญที่สกัดได้ถูกต้องและค่าความถูกต้องในการสกัดคำสำคัญ

ชื่อตำแหน่งงาน	จำนวนคำสำคัญที่ถูกต้อง	จำนวนสำคัญที่สกัดได้ถูกต้อง	ค่าความถูกต้อง
Business Intelligence	10024	9583	95.79
Cloud computing	11306	11054	97.82
Cyber Security	11455	11040	96.5
Data analyst	18691	16934	91.4
Data Engineer	29255	28295	96.82
Data Scientist	26736	25509	95.61
Database Administrator	10028	9699	96.82
IT Support	6046	5873	97.22
Network Engineer	10702	10310	96.47
Programmer	30218	28943	95.95
System analyst	7566	7279	96.34
รวมจำนวนคำสำคัญทั้งหมด/ค่าเฉลี่ย	172,027	164,519	95.96 (ค่าเฉลี่ย)

ค่าความถูกต้องเฉลี่ยของกระบวนการสกัดคำสำคัญทั้งหมดคือ 95.96% ซึ่งแสดงให้เห็นว่ากระบวนการสกัดคำสำคัญมีประสิทธิภาพในการคัดเลือกคำสำคัญจากข้อมูลอย่างมีความแม่นยำสูงในตำแหน่งงานต่าง ๆ ที่ได้ระบุในตารางข้างต้น

4.3 ความถูกต้องในการจัดประเภทและสร้างโครงสร้างข้อมูลที่สำคัญ

ความถูกต้องในการจัดประเภทและสร้างโครงสร้างข้อมูลที่สำคัญเป็นสิ่งสำคัญในการประมวลผลข้อมูลของงานวิจัยนี้เพื่อการวิเคราะห์ข้อมูลหรือข้อความ ดังนั้นผู้วิจัยจึงใช้สูตรดังนี้

$$\text{Accuracy} = \frac{\text{จำนวนประโยคที่ถูกจัดประเภทได้ถูกต้อง}}{(\text{จำนวนประโยคที่ถูกจัดประเภทได้ถูกต้อง} + \text{จำนวนประโยคที่ถูกจัดประเภทได้ไม่ถูกต้อง})} \quad (3)$$

ตารางที่ 13 จำนวนประโยคที่จัดประเภทที่ถูกต้องและจำนวนประโยคที่ถูกจัดประเภทได้ถูกต้องและค่าความถูกต้อง

ประเภท	จำนวนประโยคที่จัดประเภทที่ถูกต้อง	จำนวนประโยคที่ถูกจัดประเภทได้ถูกต้อง	ค่าความถูกต้อง
Responsibilities	29,271	29,271	100
Age	212	211	99.52
Sex	14	14	100
Nationality	34	33	97.05
Education	2411	2411	100
Experience	1892	1892	100
Skills	19186	19185	99.9
รวมจำนวนคำสำคัญทั้งหมด/ค่าเฉลี่ย	53,020	53,017	99.58(ค่าเฉลี่ย)

ค่าความถูกต้องเฉลี่ยของกระบวนการจัดประเภทและสร้างโครงสร้างข้อมูลที่สำคัญคือ 99.58% ซึ่งแสดงให้เห็นว่า CID-CJA มีประสิทธิภาพและแม่นยำในการจัดประเภทและติดป้ายชื่อบริบทให้อย่างถูกต้องในหลายประเภทข้อมูลที่มีความหลากหลายและซับซ้อน นี่เป็นผลมาจาก

ความสามารถของ CID-CJA ในการตรวจจับและระบุคำสำคัญที่เฉพาะเจาะจงสำหรับแต่ละหมวดหมู่ข้อมูล และการใช้กฎการจัดประเภทเพื่อช่วยในการระบุประเภทของบริบท

ตัวอย่างการตรวจจับคำสำคัญที่เฉพาะเจาะจงส่วนหน้าที่ความรับผิดชอบซึ่งสามารถจัดป้ายชื่อบริบททั้งหมดในส่วนนี้เป็น “responsibility” สำหรับ “Sex”, “Education”, และ “Experience” ก็สามารถระบุคำสำคัญที่เฉพาะเจาะจงได้เช่นกัน เช่น “male”, “female”, “bachelor”, “master”, “degree”, “experience” เป็นต้น ซึ่งทำให้ CID-CJA สามารถจัดประเภทและติดป้ายชื่อบริบทในส่วนนี้ได้อย่างแม่นยำ

นอกจากนี้การใช้กฎการจัดประเภทที่กำหนดล่วงหน้าเพื่อช่วยในการระบุประเภทของบริบทเป็นปัจจัยที่สำคัญที่ช่วยให้ CID-CJA ทำงานอย่างแม่นยำและมีประสิทธิภาพในการจัดประเภทและติดป้ายชื่อบริบทในแต่ละหมวดหมู่ข้อมูลที่เกี่ยวข้อง โดยทั่วไป CID-CJA จึงเป็นเครื่องมือที่มีประสิทธิภาพสำหรับการจัดประเภทและการติดป้ายชื่อบริบทในข้อมูลที่มีความซับซ้อนและหลากหลายในหลายประเภทข้อมูลต่าง ๆ อย่างมีความแม่นยำและมั่นใจ

4.4 การนำข้อมูลจากฐานข้อมูลมาแสดงผล

4.4.1 เครื่องมือที่ใช้

Power BI คือแพลตฟอร์มที่พัฒนาโดย Microsoft ซึ่งเป็นเครื่องมือที่ใช้ในการนำข้อมูลมาแสดงผลข้อมูลในรูปแบบกราฟ และการใช้ Power BI จะช่วยให้เห็นภาพรวมของข้อมูลเป็นการรายงานการแสดงผล เพื่อเป็นพื้นฐานในการตัดสินใจ

4.4.2 การแสดงผล

เมื่อจัดโครงสร้างและรวบรวมเนื้อหาลงในฐานข้อมูลแล้ว ผู้วิจัยได้ดำเนินการนำข้อมูลจากฐานข้อมูลมาแสดงผลโดยใช้เครื่องมือ Power BI สร้างหน้าแสดงผลแบบภาพรวมที่รวบรวมจากข้อมูล 4,280 ประกาศงานแสดงดังภาพที่ 22 โดยในหน้าแสดงผลนี้สามารถเลือกระบุข้อมูลตำแหน่งงาน ระดับการศึกษา สาขาวิชา อายุ หรือข้อมูลอย่างใดอย่างหนึ่งก็ได้ ที่ผู้ใช้งานต้องการระบุเพื่อดูผลของข้อมูลในส่วนต่าง ๆ

ภาพที่ 22 หน้ารายงานการแสดงผล

ยกตัวอย่างเช่น ผู้ใช้งานดำเนินการระบุตำแหน่งงาน “Data Scientist” หน้าแสดงผลจะแสดงรายละเอียดดังภาพที่ 23 หน้ารายงานการแสดงผลที่ถูกระบุตำแหน่ง “Data Scientist”

ภาพที่ 23 หน้ารายงานการแสดงผลที่ถูกระบุตำแหน่ง “Data Scientist”

จากภาพที่ 23 หน้ารายงานข้อมูลแสดงข้อมูลของแต่ละส่วน ซึ่งในแต่ละกราฟข้อมูลจะปรับเปลี่ยนเป็นข้อมูลที่เกี่ยวข้องกับตำแหน่ง “Data Scientist” โดยข้อมูลที่ได้จาก 518 ประกาศงานมีดังนี้

i) กราฟอายุ พบว่ามีประกาศงานที่ไม่ได้ระบุข้อมูลอายุ 507 ประกาศงาน ส่วนประกาศงานที่ระบุอายุที่ต้องการคือ 21 – 35 จาก 5 ประกาศงาน, 23 – 28 จาก 3 ประกาศงาน, 28 – 38 จาก 2 ประกาศงาน และ 28 – 35 จาก 1 ประกาศงาน

ii) กราฟสัญชาติ พบว่ามีประกาศระบุต้องการสัญชาติไทย 4 ประกาศงาน

iii) กราฟเพศ พบว่ามีประกาศงานระบุต้องการเพศชายจาก 1 ประกาศ เพศชายหรือหญิงจาก 3 ประกาศงาน ทำให้สรุปได้ว่าจากตำแหน่ง “Data Scientist” ไม่สนใจเพศ

iv) กราฟระดับการศึกษาและสาขา พบว่ามีประกาศงานระบุต้องการระดับศึกษและสาขา 351 ประกาศงาน โดยที่ในระดับปริญญาตรีสาขาที่พบมากที่สุดคือ “computer science” ระดับปริญญาโทสาขาที่พบมากที่สุดคือ “computer science” เช่นกัน และในระดับปริญญาเอกคือ “computer science” และ “statistic”

v) กราฟประสบการณ์ พบว่ามีประกาศงานระบุถึงความต้องการประสบการณ์ 294 ประกาศงาน ซึ่งพบสุดคือ ต้องการประสบการณ์ 3 ปีขึ้นไปในการวิเคราะห์

Skill level Frequency

LEVEL_MAIN	LEVEL_SUB	VERB	OBJ	Count of Count
strong	N/A	solve	problem	56
good	N/A	command	english	49
basic	experience	N/A	python	30
strong	N/A	solve	emphasis	30
basic	experience	N/A	sql	27
basic	experience	N/A	programming	24

ภาพที่ 24 ข้อมูลในหน้ารายงานผลส่วนตารางแสดงจำนวนของระดับทักษะที่พบจาก 518 ประกาศงาน

vi) กราฟตารางระดับทักษะ พบว่ามีประกาศงานที่ระบุถึงทักษะทุกประกาศงาน โดยระดับทักษะที่พบมากที่สุดคือ แข็งแกร่งในการแก้ปัญหา ทำให้สรุปได้ว่าจากตำแหน่ง “Data Scientist” ต้องการบุคคลที่มีทักษะต่าง ๆ ตามลำดับในภาพที่ 24 ข้อมูลในหน้ารายงานผลส่วนตารางแสดงจำนวนของระดับทักษะที่พบจาก 518 ประกาศงาน

Duties Frequency

VERB	OBJ	Count of Count
analyze	data	48
learn	model	47
discover	pattern	41
build	machine	39
use	predictive model	38
develop	model	36
learn	algorithm	35
solve	problem	35
identify	opportunity	33

ภาพที่ 25 ข้อมูลในหน้ารายงานผลส่วนตารางแสดงจำนวนของภาระงานพบจาก 518 ประกาศงาน

vii) กราฟตารางหน้าที่ความรับผิดชอบ พบว่ามีประกาศงานที่ระบุถึงหน้าที่ที่ต้องรับผิดชอบทุกประกาศงาน โดยการกระทำในภาระงานที่พบมากที่สุดคือ การวิเคราะห์ข้อมูล, เรียนรู้โมเดล ตามลำดับในภาพที่ 25 ข้อมูลในหน้ารายงานผลส่วนตารางแสดงจำนวนของภาระงานพบจาก 518 ประกาศงาน

ได้ข้อสรุปว่าจากการวิเคราะห์ข้อมูลของ 4,280 ประกาศงาน โดยระบุความต้องการข้อมูลในตำแหน่ง "Data Scientist" มีจำนวนประกาศงาน 518 ประกาศงานพบว่า องค์กรส่วนใหญ่มีความยืดหยุ่นในเรื่องอายุ สัญชาติ และเพศ ของผู้สมัคร แต่มีข้อกำหนดที่เฉพาะเจาะจงในเรื่องของระดับการศึกษา ประสบการณ์และทักษะที่ต้องการ นั่นคือ สาขา “Computer Science” เป็นสาขาที่ต้องการเป็นส่วนใหญ่ และต้องการสาขา “statistic” ในระดับปริญญาเอกมากเช่นกัน ต้องการประสบการณ์ในการวิเคราะห์ข้อมูล และต้องการบุคคลที่มีทักษะแข็งแกร่งในการแก้ปัญหา ประกอบกับหน้าที่และความรับผิดชอบในการวิเคราะห์ข้อมูลและการเรียนรู้โมเดล เป็นต้น

ดังข้อมูลที่สรุปมานี้อาจช่วยให้ผู้ใช้งานได้เป็นส่วนประกอบในการวางแผนหรือตัดสินใจใน ไม่ว่าจะเป็นในด้านการเลือกตำแหน่งงานให้ตรงกับทักษะและความสนใจ หรือการพัฒนาทักษะเพื่อตอบสนองต่อความต้องการของตลาดงานในตำแหน่งงานนั้น ๆ

บทที่ 5 สรุปและอภิปลายผล

5.1 สรุปผลการดำเนินงาน

งานวิจัยนี้มุ้งเน้นการวิเคราะห์และหาข้อสรุปรายละเอียดของการประกาศรับสมัครงาน นำเสนอวิธีการค้นพบข้อมูลที่เกิดขึ้นร่วมกันในการประกาศรับสมัครงานที่เกี่ยวข้องกับคอมพิวเตอร์ (วิธีการ CID-CJA) โดยนำข้อมูลประกาศงานที่เป็นภาษาอังกฤษ จากเว็บไซต์รับสมัครงานในประเทศไทย ได้แก่ jobsDB Linkedin และ Indeed ซึ่งตำแหน่งงานที่ผู้วิจัยสนใจเป็นตำแหน่งงานศาสตร์ทางด้านเทคโนโลยีสารสนเทศ วิธีการ CID-CJA จะทำให้พบรูปแบบโครงสร้างของข้อมูลสำคัญ โดยดำเนินการสกัดคำสำคัญจากข้อมูลรายละเอียดจากประกาศรับสมัครงาน ซึ่งได้ผลลัพธ์จากการวิเคราะห์และหาข้อสรุปรายละเอียดของการประกาศรับสมัครงานลงฐานข้อมูล ทำให้ข้อมูลอยู่ในรูปแบบโครงสร้าง ซึ่งช่วยให้ 1) ผู้ที่สนใจสมัครงานหรือนักศึกษาสามารถรู้ถึงทักษะที่ตนเองควรจะมีในตำแหน่งงานที่สนใจ 2) สถาบันการศึกษาต่าง ๆ ทำการปรับปรุงและพัฒนาหลักสูตรการสอน ให้มีความสอดคล้องและตรงตามความต้องการกับตำแหน่งงานใน ปัจจุบันมากยิ่งขึ้น 3) องค์กรต่าง ๆ สามารถทำการเปรียบเทียบรายละเอียดต่าง ๆ ในตำแหน่งงานเดียวได้ เพื่อให้ทราบถึงรายละเอียดเพิ่มเติมหรือส่วนที่แตกต่างกับองค์กร ๆ

ในการดำเนินงานของวิธีการ CID-CJA สามารถแบ่งออกได้5 ขั้นตอน 1) การเตรียมข้อมูลและการรวบรวมข้อมูลประกาศรับสมัครงานเฉพาะศาสตร์ทางด้านเทคโนโลยีสารสนเทศ 2) การจัดกลุ่มประกาศงาน 3) การประมวลผลข้อความ 4) การจัดโครงสร้างและการรวบรวมเนื้อหา 5) การส่งข้อมูลให้ผู้ใช้ ซึ่งผลลัพธ์จากวิธีการ CID-CJA จะแสดงให้เห็นถึงข้อมูลในกลุ่มงานที่มีคำสำคัญที่สามารถเจาะลึกลงไปอย่างเฉพาะเจาะจง ให้เห็นถึงข้อมูลที่สำคัญที่แสดงความต้องการในประกาศงาน

วิธีการ CID-CJA ได้ประเมินการทดสอบใน 3 ด้าน ซึ่งประกอบด้วย 1) ความถูกต้องในการจัดกลุ่ม 2) ความแม่นยำในการสกัดคำสำคัญ และ 3) ความถูกต้องในการจัดประเภทและสร้างโครงสร้างข้อมูลที่สำคัญ

การทดสอบความถูกต้องในการจัดกลุ่มแสดงให้เห็นว่า CID-CJA สามารถจัดกลุ่มประกาศรับสมัครงานตามชื่อตำแหน่งงานได้ 100%

การทดสอบความแม่นยำในการสกัดคำสำคัญแสดงให้เห็นว่า CID-CJA มีความแม่นยำในการคัดเลือกคำสำคัญจากข้อมูลในตำแหน่งงานต่าง ๆ ที่ได้ระบุ โดยมีค่าความถูกต้องเฉลี่ยที่ 95.96%

การทดสอบความถูกต้องในการจัดประเภทและสร้างโครงสร้างข้อมูลที่สำคัญแสดงให้เห็นว่า CID-CJA มีความแม่นยำและมีประสิทธิภาพในการจัดประเภทและติดป้ายชื่อบริบท โดยมีค่าความถูกต้องเฉลี่ยที่ 99.58%

สรุปคือ งานวิจัยนี้ได้ทำการพัฒนาและทดสอบวิธีการ CID-CJA ในการวิเคราะห์ประกาศรับสมัครงานทางด้านเทคโนโลยีสารสนเทศในประเทศไทย และพบว่าวิธีการนี้มีประสิทธิภาพในการจัดกลุ่มประกาศรับสมัครงาน สกัดคำสำคัญ และการจัดประเภทข้อมูล ซึ่งสามารถนำไปใช้ในการวิเคราะห์และเข้าใจแนวโน้มของตลาดงานและความต้องการของบุคลากรในอนาคต งานวิจัยนี้ยังมีศักยภาพในการนำไปใช้ในส่วนที่เกี่ยวข้องกับการสรรหาบุคลากรในสาขาด้านเทคโนโลยีสารสนเทศในประเทศไทย เพื่อให้การตัดสินใจเกี่ยวกับการจ้างงานเป็นไปอย่างมีประสิทธิภาพ

5.2 ข้อเสนอแนะ

งานวิจัยนี้สามารถพัฒนาและปรับปรุงวิธีการ CID-CJA ในส่วนของข้อมูลระดับของตำแหน่งงาน หรือระบุตำแหน่งที่ตั้งของบริษัทที่ประกาศรับสมัครงาน หรือพัฒนาเพิ่มประสิทธิภาพและความแม่นยำในการวิเคราะห์และตรวจสอบข้อมูลเกี่ยวกับตำแหน่งงานทางด้านเทคโนโลยีสารสนเทศ

บรรณานุกรม

Alsaedi, N., Burnap, P., & Rana, O. (2016). Automatic summarization of real world events using twitter. Proceedings of the International AAAI Conference on Web and Social Media,

Baxendale, P. B. (1958). Machine-made index for technical literature—an experiment. *IBM Journal of research and development*, *2*(4), 354-361.

Carbonell, J., & Goldstein, J. (1998). The use of MMR, diversity-based reranking for reordering documents and producing summaries. Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval,

Chaisoongnoen, B., & Amphawan, K. (2020). An improvement of supplementary book suggestion system. The 9th International Conference on Smart Media and Applications,

Chaisoongnoen, B., Amphawan, K., & Bunpeng, A. (2018). Supplementary book suggestion for computer science courses. 2018 5th International Conference on Advanced Informatics: Concept Theory and Applications (ICAICTA),

Kamlangpuech, P., & Amphawan, K. (2020). A new system for analyzing contents of computer science courses. 2020 7th International Conference on Advance Informatics: Concepts, Theory and Applications (ICAICTA),

Kamlangpuech, P., & Amphawan, K. (2021). eCSCDA: An efficient system for analyzing contents of Computer Science Courses. 2021 8th International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA),

Lin, C.-Y., & Hovy, E. (1997). Identifying topics by position. Fifth conference on applied natural language processing,

Luhn, H. P. (1958). The automatic creation of literature abstracts. IBM Journal of research and development, 2(2), 159-165.

Lukauskas, M., Šarkauskaitė, V., Pilinkienė, V., Stundžienė, A., Grybauskas, A., & Bruneckienė, J. (2023). Enhancing Skills Demand Understanding through Job Ad Segmentation Using NLP and Clustering Techniques. *Applied Sciences*, *13*(10),

6119.

McKeown, K., Klavans, J. L., Hatzivassiloglou, V., Barzilay, R., & Eskin, E. (1999). Towards multidocument summarization by reformulation: Progress and prospects.

Muresan, S., Tzoukermann, E., & Klavans, J. L. (2001). Combining linguistic and machine learning techniques for email summarization. Proceedings of the ACL 2001 Workshop on Computational Natural Language Learning (ConLL),

Nelson, H. J., Ahmad, A., Martin, N. L., & Litecky, C. R. (2007). A comparative study of IT/IS job skills and job definitions. Proceedings of the 2007 ACM SIGMIS CPR conference on Computer personnel research: The global information technology workforce,

Parker, A., & Brown, I. (2019). Skills requirements for cyber security professionals: a content analysis of job descriptions in South Africa. Information Security: 17th International Conference, ISSA 2018, Pretoria, South Africa, August 15–16, 2018, Revised Selected Papers 17,

Pejic-Bach, M., Bertoncel, T., Meško, M., & Krstić, Ž. (2020). Text mining of industry 4.0 job advertisements. *International journal of information management*, *50*, 416-431.

Radovilsky, Z., Hegde, V., Acharya, A., & Uma, U. (2018). Skills requirements of business data analytics and data science jobs: A comparative analysis. *Journal of Supply Chain and Operations Management*, *16*(1), 82-101.

Sahni, A., & Palwe, S. (2018). Topic Modeling on Online News Extraction. Intelligent Computing and Information and Communication: Proceedings of 2nd International Conference, ICICC 2017,

Sibarani, E. M., Scerri, S., Morales, C., Auer, S., & Collarana, D. (2017). Ontology-guided job market demand analysis: a cross-sectional study for the data science field. Proceedings of the 13th international conference on semantic systems,

ภาคผนวก

ภาคผนวก ก

กฎทางภาษาศาสตร์สำหรับวิธีการ CID-CJA

ตารางที่ 14 กฎทางภาษาศาสตร์ที่ไม่พิจารณาร่วมกับคำศัพท์เฉพาะและไม่มีคำกริยาอยู่ในประโยค สำหรับวิธีการ CID-CJA

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
1	(W_0 = "NN" or "NNS" or "NNP" or "NNPS"), ..., (W_n = "NN" or "NNS" or "NNP" or "NNPS")	• VERB: "-", OBJ: "W_0, ..., W_n"
2	(W_0 = "JJ" or "JJR" or "JJS") + (W_1 = "NN" or "NNS" or "NNP" or "NNPS"), ..., (W_n = "NN" or "NNS" or "NNP" or "NNPS")	• VERB: "-", OBJ: "W_0 + W_1, ..., W_n"
3	(W_0 = "NN" or "NNS" or "NNP" or "NNPS") + (W_1 = "IN") + (W_2 = "NN" or "NNS" or "NNP" or "NNPS") + (W_3 = "IN") + (W_4 = "NN" or "NNS" or "NNP" or "NNPS")	• VERB: "-", OBJ: "W_0 + W_1 + W_2" • VERB: "-", OBJ: "W_2 + W_3 + W_4"
4	(W_0 = "NN" or "NNS" or "NNP" or "NNPS") + (W_1 = "NN" or "NNS" or "NNP" or "NNPS") + (W_2 = "CC") + (W_3 = "NN" or "NNS" or "NNP" or "NNPS")	• VERB: "-", OBJ: "W_0 + W_1" • VERB: "-", OBJ: "W_0 + W_3"
5	(W_0 = "NN" or "NNS" or "NNP" or "NNPS") + (W_1 = "CC") + (W_2 = "NN" or "NNS" or "NNP" or "NNPS") + (W_3 = "NN" or "NNS" or "NNP" or "NNPS")	• VERB: "-", OBJ: "W_0 + W_3" • VERB: "-", OBJ: "W_2 + W_3"
6	(W_0 = "NN" or "NNS" or "NNP" or "NNPS") + (W_1 = "CC") + (W_2 = "NN" or "NNS" or "NNP" or "NNPS")	• VERB: "-", OBJ: "W_0" • VERB: "-", OBJ: "W_2"
7	(W_0 = "JJ" or "JJR" or "JJS") + (W_1 = "NN" or "NNS" or "NNP" or "NNPS") + (W_2 = "CC") + (W_3 = "JJ" or "JJR" or "JJS") + (W_4 = "NN" or "NNS" or "NNP" or "NNPS")	if W_1 = W_4 : • VERB: "-", OBJ: "W_0 + W_1" • VERB: "-", OBJ: "W_3 + W_4"
8	(W_0 = "NN" or "NNS" or "NNP") + (W_1 = "IN") + (W_2 = "NN" or "NNS" or "NNP") + (W_3 = "CC") + (W_4 = "NN" or "NNS" or "NNP")	• VERB: "-", OBJ: "W_0 + W_1 + W_2" • VERB: "-", OBJ: "W_0 + W_1 + W_4"
9	(W_0 = "NN" or "NNS" or "NNP") +	• VERB: "-", OBJ: "W_0 + W_3

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_1 = “CC”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “IN”) + (W_4 = “NN” or “NNS” or “NNP”)	+ W_4” • VERB: “-”, OBJ: “W_2 + W_3 + W_4”
10	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0 + W_1”
11	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “CC”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “IN”) + (W_4 = “JJ” or “JJR” or “JJS”) + (W_5 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0 + W_3 + W_4 + W_5” • VERB: “-”, OBJ: “W_2 + W_3 + W_4 + W_5”
12	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “IN”) + (W_2 = “JJ” or “JJR” or “JJS”) + (W_3 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0 + W_1 + W_2 + W_3”
13	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “CC”) + (W_4 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0 + W_1 + W_2” • VERB: “-”, OBJ: “W_0 + W_1 + W_4”
14	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “CC”) + (W_2 = “JJ” or “JJR” or “JJS”) + (W_3 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0 + W_1” • VERB: “-”, OBJ: “W_2 + W_3”
15	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “CC”) + (W_3 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0 + W_1” • VERB: “-”, OBJ: “W_0 + W_3”
16	(W_0 = “NN” or “NNS” or “NNP”) +	• VERB: “-”, OBJ: “W_0 + W_1”

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “CC”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_3 + W_4”
17	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “JJ” or “JJR” or “JJS”) + (W_3 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0 + W_1” • VERB: “-”, OBJ: “W_2 + W_3”
18	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “IN”) + (W_4 = “JJ” or “JJR” or “JJS”) + (W_5 = “NN” or “NNS” or “NNP”) + (W_6 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0 + W_1 + W_2” • VERB: “-”, OBJ: “W_4 + W_5 + W_6” • VERB: “-”, OBJ: “W_0 + W_1 + W_2 + W_3 + W_4 + W_5 + W_6”
19	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “IN”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0 + W_1 + W_2 + W_3 + W_4”
20	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “CC”) + (W_2 = “JJ” or “JJR” or “JJS”)	• VERB: “-”, OBJ: “W_0” • VERB: “-”, OBJ: “W_2”
21	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0 + W_1 + W_2”
22	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “CC”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0 + W_1” • VERB: “-”, OBJ: “W_0 + W_3 + W_4 “

ตารางที่ 15 กฎทางภาษาศาสตร์ที่ไม่พิจารณาร่วมกับคำศัพท์เฉพาะและแต่มีคำกริยาอยู่ในประโยค สำหรับวิธีการ CID-CJA

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
23	(W_0 = "VB" or "VBP") + (W_1 = "NN" or "NNS" or "NNP" or "NNPS"), ..., (W_n = "NN" or "NNS" or "NNP" or "NNPS")	• VERB: "W_0", OBJ: "W_1, ..., W_n"
24	(W_0 = "VB" or "VBP") + (W_1 = "JJ" or "JJR" or "JJS") + (W_2 = "NN" or "NNS" or "NNP" or "NNPS"), ..., (W_n = "NN" or "NNS" or "NNP" or "NNPS")	• VERB: "W_0", OBJ: "W_1 + W_2, ..., W_n"
25	(W_0 = "VB" or "VBP") + (W_1 = "NN" or "NNS" or "NNP" or "NNPS") + (W_2 = "IN") + (W_3 = "NN" or "NNS" or "NNP" or "NNPS") + (W_4 = "IN") + (W_5 = "NN" or "NNS" or "NNP" or "NNPS")	• VERB: "W_0", OBJ: "W_1 + W_2 + W_3" • VERB: "W_0", OBJ: "W_3 + W_4 + W_5"
*26	(W_0 = "VB" or "VBP") + (W_1 = "NN" or "NNS" or "NNP" or "NNPS") + (W_2 = "NN" or "NNS" or "NNP" or "NNPS") + (W_3 = "CC") + (W_4 = "NN" or "NNS" or "NNP" or "NNPS")	• VERB: "W_0", OBJ: "W_1 + W_2" • VERB: "W_0", OBJ: "W_1 + W_4"
27	(W_0 = "VB" or "VBP") + (W_1 = "NN" or "NNS" or "NNP" or "NNPS") + (W_2 = "CC") + (W_3 = "NN" or "NNS" or "NNP" or "NNPS") + (W_4 = "NN" or "NNS" or "NNP" or "NNPS")	• VERB: "W_0", OBJ: "W_1 + W_3" • VERB: "W_0", OBJ: "W_3 + W_4"
28	(W_0 = "VB" or "VBP") + (W_1 = "NN" or "NNS" or "NNP" or "NNPS") + (W_2 = "CC") + (W_3 = "NN" or "NNS" or "NNP" or "NNPS")	• VERB: "W_0", OBJ: "W_1" • VERB: "W_0", OBJ: "W_3"
29	(W_0 = "VB" or "VBP") + (W_1 = "JJ" or "JJR" or "JJS") + (W_2 = "NN" or "NNS" or "NNP" or "NNPS") + (W_3 = "CC") + (W_4 = "JJ" or "JJR" or "JJS") +	if W_2 = W_5 : • VERB: "W_0", OBJ: "W_1 + W_2" • VERB: "W_0", OBJ: "W_4 + W_5"

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_5 = “NN” or “NNS” or “NNP” or “NNPS”)	
30	(W_0 = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “IN”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “CC”) + (W_5 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1 + W_2 + W_3” • VERB: “W_0”, OBJ: “W_1 + W_2 + W_5”
31	(W_0 = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “CC”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “IN”) + (W_5 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1 + W_4 + W_5” • VERB: “W_0”, OBJ: “W_3 + W_4 + W_5”
32	(W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1 + W_2”
33	(W_0 = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “CC”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “IN”) + (W_5 = “JJ” or “JJR” or “JJS”) + (W_6 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1 + W_4 + W_5 + W_6” • VERB: “W_0”, OBJ: “W_3 + W_4 + W_5 + W_6”
34	(W_0 = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “IN”) + (W_3 = “JJ” or “JJR” or “JJS”) + (W_4 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1 + W_2 + W_3 + W_4”
35	(W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “NN” or “NNS” or “NNP”) +	• VERB: “W_0”, OBJ: “W_1 + W_2 + W_3” • VERB: “W_0”, OBJ: “W_1 +

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “CC”) + (W_5 = “NN” or “NNS” or “NNP”)	$W_2 + W_5$”
36	(W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “CC”) + (W_4 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “$W_1 + W_2$” • VERB: “W_0”, OBJ: “$W_1 + W_4$”
37	(W_0 = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “CC”) + (W_4 = “NN” or “NNS” or “NNP”) + (W_5 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “$W_1 + W_2$” • VERB: “W_0”, OBJ: “$W_4 + W_5$”
38	(W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “JJ” or “JJR” or “JJS”) + (W_4 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “$W_1 + W_2$” • VERB: “W_0”, OBJ: “$W_3 + W_4$”
39	(W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “JJ” or “JJR” or “JJS”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “IN”) + (W_5 = “JJ” or “JJR” or “JJS”) + (W_6 = “NN” or “NNS” or “NNP”) + (W_7 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “$W_1 + W_2 + W_3$” • VERB: “W_0”, OBJ: “$W_5 + W_6 + W_7$” • VERB: “W_0”, OBJ: “$W_1 + W_2 + W_3 + W_4 + W_5 + W_6 + W_7$”
40	(W_0 = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “IN”) + (W_4 = “NN” or “NNS” or “NNP”) + (W_5 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “$W_1 + W_2 + W_3 + W_4 + W_5$”

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
41	(W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “CC”) + (W_3 = “JJ” or “JJR” or “JJS”)	• VERB: “W_0”, OBJ: “W_1” • VERB: “W_0”, OBJ: “W_3”
42	(W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “JJ” or “JJR” or “JJS”) + (W_3 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1 + W_2 + W_3”
43	(W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “CC”) + (W_4 = “NN” or “NNS” or “NNP”) + (W_5 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1 + W_2” • VERB: “W_0”, OBJ: “W_1 + W_4 + W_5 “
44	(W_0 = “VB” or “VBP”) + (W_{n1} = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP” or “NNPS”), ..., (W_{n2} = “NN” or “NNS” or “NNP” or “NNPS”)	• VERB: “W_0, ..., W_{n1}”, OBJ: “W_1, ..., W_{n2}”
45	(W_0 = “VB” or “VBP”) + (W_{n1} = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “NN” or “NNS” or “NNP” or “NNPS”), ..., (W_{n2} = “NN” or “NNS” or “NNP” or “NNPS”)	• VERB: “W_0, ..., W_{n1}”, OBJ: “W_1 + W_2, ..., W_{n2}”
46	(W_0 = “VB” or “VBP”) + (W_n = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP” or “NNPS”) + (W_2 = “IN”) + (W_3 = “NN” or “NNS” or “NNP” or “NNPS”) + (W_4 = “IN”) + (W_5 = “NN” or “NNS” or “NNP” or “NNPS”)	• VERB: “W_0, ..., W_{n1}”, OBJ: “W_1 + W_2 + W_3” • VERB: “W_0, ..., W_{n1}”, OBJ: “W_3 + W_4 + W_5”
47	(W_0 = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP” or “NNPS”) + (W_2 = “NN” or “NNS” or “NNP” or “NNPS”) + (W_3 = “CC”) +	• VERB: “W_0”, OBJ: “W_1 + W_2” • VERB: “W_0”, OBJ: “W_1 + W_4”

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_4 = “NN” or “NNS” or “NNP” or “NNPS”)	
48	(W_0 = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP” or “NNPS”) + (W_2 = “NN” or “NNS” or “NNP” or “NNPS”) + (W_3 = “CC”) + (W_4 = “VB” or “VBP”) + (W_5 = “NN” or “NNS” or “NNP” or “NNPS”)	• VERB: “W_0”, OBJ: “W_1 + W_2” • VERB: “W_4”, OBJ: “W_5”
49	(W_0 = “VB” or “VBP”) + (W_n = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP” or “NNPS”) + (W_2 = “CC”) + (W_3 = “NN” or “NNS” or “NNP” or “NNPS”) + (W_4 = “NN” or “NNS” or “NNP” or “NNPS”)	• VERB: “W_0, …, W_{n1}”, OBJ: “W_1 + W_4” • VERB: “W_0, …, W_{n1}”, OBJ: “W_3 + W_4”“
50	(W_0 = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP” or “NNPS”) + (W_2 = “CC”) + (W_3 = “VB” or “VBP”) + (W_4 = “NN” or “NNS” or “NNP” or “NNPS”) + (W_5 = “NN” or “NNS” or “NNP” or “NNPS”)	• VERB: “W_0”, OBJ: “W_1” • VERB: “W_3”, OBJ: “W_4 + W_5”
51	(W_0 = “VB” or “VBP”) + (W_n = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “CC”) + (W_3 = “NN” or “NNS” or “NNP”)	• VERB: “W_0, …, W_n”, OBJ: “W_1” • VERB: “W_0, …, W_n”, OBJ: “W_3”
52	(W_0 = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “CC”) + (W_3 = “VB” or “VBP”) + (W_4 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1” • VERB: “W_3”, OBJ: “W_4”
53	(W_0 = “VB” or “VBP”) + (W_n = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “IN”) +	• VERB: “W_0, …, W_n”, OBJ: “W_1+ W_2 + W_3” • VERB: “W_0, …, W_n”, OBJ: “W_1 + W_2 + W_5”

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “CC”) + (W_5 = “NN” or “NNS” or “NNP”)	•
54	(W_0 = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “CC”) + (W_3 = “VB” or “VBP”) + (W_4 = “NN” or “NNS” or “NNP”) + (W_5 = “IN”) + (W_6 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1” • VERB: “W_3”, OBJ: “W_4 + W_5 + W_6”
55	(W_0 = “VB” or “VBP”) + (W_n = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “NN” or “NNS” or “NNP”)	• VERB: “W_0, …, W_n”, OBJ: “W_1 + W_2”
56	(W_0 = “VB” or “VBP”) + (W_n = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “CC”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “IN”) + (W_5 = “JJ” or “JJR” or “JJS”) + (W_6 = “NN” or “NNS” or “NNP”)	• VERB: “W_0, …, W_n”, OBJ: “W_1 + W_5 + W_6” • VERB: “W_0, …, W_n”, OBJ: “W_3 + W_5 + W_6”
57	(W_0 = “VB” or “VBP”) + (W_n = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “IN”) + (W_3 = “JJ” or “JJR” or “JJS”) + (W_4 = “NN” or “NNS” or “NNP”)	• VERB: “W_0, …, W_n”, OBJ: “W_1 + W_2 + W_4”

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
58	(W_0 = "VB" or "VBP") + (W_n = "VB" or "VBP") + (W_1 = "JJ" or "JJR" or "JJS") + (W_2 = "NN" or "NNS" or "NNP") + (W_3 = "NN" or "NNS" or "NNP") + (W_4 = "CC") + (W_5 = "NN" or "NNS" or "NNP")	• VERB: "W_0, ..., W_n", OBJ: "W_1 + W_2 + W_3" • VERB: "W_0, ..., W_n", OBJ: "W_1 + W_2 + W_5"
59	(W_0 = "VB" or "VBP") + (W_1 = "JJ" or "JJR" or "JJS") + (W_2 = "CC") + (W_3 = "NN" or "NNS" or "NNP") + (W_4 = "NN" or "NNS" or "NNP")	• VERB: "W_0", OBJ: "W_1 + W_3" • VERB: "W_0", OBJ: "W_1 + W_4"
60	(W_0 = "VB" or "VBP") + (W_n = "VB" or "VBP") + (W_1 = "NN" or "NNS" or "NNP") + (W_2 = "NN" or "NNS" or "NNP") + (W_3 = "CC") + (W_4 = "NN" or "NNS" or "NNP") + (W_5 = "NN" or "NNS" or "NNP")	• VERB: "W_0, ..., W_n", OBJ: "W_1 + W_2" • VERB: "W_0, ..., W_n", OBJ: "W_4 + W_5"
61	(W_0 = "VB" or "VBP") + (W_1 = "NN" or "NNS" or "NNP") + (W_2 = "NN" or "NNS" or "NNP") + (W_3 = "CC") + (W_4 = "VB" or "VBP") + (W_5 = "NN" or "NNS" or "NNP") + (W_6 = "NN" or "NNS" or "NNP")	• VERB: "W_0", OBJ: "W_1 + W_2" • VERB: "W_4", OBJ: "W_5 + W_6"
62	(W_0 = "VB" or "VBP") + (W_n = "VB" or "VBP") + (W_1 = "JJ" or "JJR" or "JJS") + (W_2 = "NN" or "NNS" or "NNP") + (W_3 = "JJ" or "JJR" or "JJS") +	• VERB: "W_0, ..., W_n", OBJ: "W_1 + W_2" • VERB: "W_0, ..., W_n", OBJ: "W_3 + W_4"

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_4 = “NN” or “NNS” or “NNP”)	
63	(W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “VB” or “VBP”) + (W_4 = “JJ” or “JJR” or “JJS”) + (W_5 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1 + W_2” • VERB: “W_3”, OBJ: “W_4 + W_5”
64	(W_0 = “VB” or “VBP”) + (W_n = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “JJ” or “JJR” or “JJS”) + (W_3 = “NN” or “NNS” or “NNP”)	• VERB: “W_0, ..., W_n”, OBJ: “W_1 + W_2 + W_3”
65	(W_0 = “VB” or “VBP”) + (W_n = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “CC”) + (W_4 = “NN” or “NNS” or “NNP”) + (W_5 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1 + W_2” • VERB: “W_0”, OBJ: “W_1 + W_4 + W_5”

ตารางที่ 16 กฎทางภาษาศาสตร์ที่พิจารณาร่วมกับคำศัพท์เฉพาะแต่ไม่มีคำกริยาอยู่ในประโยค สำหรับวิธีการ CID-CJA

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
66	(W_0, ..., W_n = "RW")	• VERB: "-", OBJ: "W_0, ..., W_n"
67	(W_0 = "RW") + (W_1 = "NN" or "NNS" or "NNP")	• VERB: "-", OBJ: "W_0 + W_1"
68	(W_0 = "RW") + (W_1 = "NN" or "NNS" or "NNP"), ..., (W_n = "NN" or "NNS" or "NNP" or "NNPS")	• VERB: "-", OBJ: "W_0 + W_1, ..., W_n"
69	(W_0 = "RW") + (W_1 = "RW") + (W_2 = "NN" or "NNS" or "NNP")	• VERB: "-", OBJ: "W_0 + W_1 + W_2"
70	(W_0 = "RW") + (W_1 = "NN" or "NNS" or "NNP") + (W_2 = "JJ" or "JJR" or "JJS") + (W_3 = "NN" or "NNS" or "NNP")	• VERB: "-", OBJ: "W_0 + W_1" • VERB: "-", OBJ: "W_2 + W_3"
71	(W_0 = "JJ" or "JJR" or "JJS") + (W_1 = "RW") + (W_2 = "RW")	• VERB: "-", OBJ: "W_1 + W_2" • VERB: "-", OBJ: "W_0 + W_1 + W_2"
72	(W_0 = "JJ" or "JJR" or "JJS") + (W_1 = "RW") + (W_2 = "NN" or "NNS" or "NNP")	• VERB: "-", OBJ: "W_1 + W_2" • VERB: "-", OBJ: "W_0 + W_1 + W_2"
73	(W_0 = "JJ" or "JJR" or "JJS") + (W_1 = "JJ" or "JJR" or "JJS") + (W_2 = "RW") + (W_3 = "NN" or "NNS" or "NNP")	• VERB: "-", OBJ: "W_2 + W_3" • VERB: "-", OBJ: "W_0 + W_1 + W_2 + W_3"
74	(W_0 = "NN" or "NNS" or "NNP") + (W_1 = "RW")	• VERB: "-", OBJ: "W_0 + W_1"
75	(W_0 = "JJ" or "JJR" or "JJS") + (W_1 = "RW")	• VERB: "-", OBJ: "W_0 + W_1"
76	(W_0 = "RW") +	• VERB: "-", OBJ: "W_0 +

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_1 = "IN") + (W_2 = "RW")	W_1 + W_2"
77	(W_0 = "JJ" or "JJR" or "JJS") + (W_1 = "NN" or "NNS" or "NNP") + (W_2 = "IN") + (W_3 = "RW")	• VERB: "-", OBJ: "W_3" • VERB: "-", OBJ: "W_0 + W_3" • VERB: "-", OBJ: "W_3 + W_1"
78	(W_0 = "NN" or "NNS" or "NNP") + (W_1 = "IN") + (W_2 = "RW") + (W_3 = "IN") + (W_4 = "RW")	• VERB: "-", OBJ: "W_2 + W_0" • VERB: "-", OBJ: "W_2 + W_0 + W_1 + W_4"
79	(W_0 = "JJ" or "JJR" or "JJS") + (W_1 = "NN" or "NNS" or "NNP") + (W_2 = "IN") + (W_3 = "RW") + (W_4 = "RW")	• VERB: "-", OBJ: "W_3 + W_4" • VERB: "-", OBJ: "W_0 + W_3 + W_4" • VERB: "-", OBJ: "W_3 + W_4 + W_1"
80	(W_0 = "NN" or "NNS" or "NNP") + (W_1 = "NN" or "NNS" or "NNP") + (W_2 = "IN") + (W_3 = "RW") + (W_4 = "RW")	• VERB: "-", OBJ: "W_3 + W_4" • VERB: "-", OBJ: "W_3 + W_4 + W_0" • VERB: "-", OBJ: "W_3 + W_4 + W_1"
81	(W_0 = "NN" or "NNS" or "NNP") + (W_1 = "IN") + (W_2 = "RW") + (W_3 = "RW") + (W_4 = "RW")	• VERB: "-", OBJ: "W_2 + W_3 + W_4" • VERB: "-", OBJ: "W_2 + W_3 + W_4 + W_0"
82	(W_0 = "RW") +	• VERB: "-", OBJ: "W_0 +

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “IN”) + (W_3 = “RW”)	W_1” • VERB: “-”, OBJ: “W_0 + W_1 + W_2 + W_3”
83	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “IN”) + (W_2 = “RW”)	• VERB: “-”, OBJ: “W_2 + W_0”
84	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “IN”) + (W_3 = “RW”) + (W_4 = “RW”) + (W_5 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_3 + W_4 + W_5” • VERB: “-”, OBJ: “W_3 + W_4 + W_5 + W_1” • VERB: “-”, OBJ: “W_0 + W_3 + W_4 + W_5”
85	(W_0 = “RW”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “CC”) + (W_3 = “RW”)	• VERB: “-”, OBJ: “W_0 + W_1” • VERB: “-”, OBJ: “W_3”
86	(W_0 = “RW”) + (W_1 = “RW”) + (W_2 = “CC”) + (W_3 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0 + W_1” • VERB: “-”, OBJ: “W_0 + W_3”
87	(W_0 = “RW”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “CC”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “NN” or “NNS” or “NNP”)	if W_1 = W_4 : • VERB: “-”, OBJ: “W_0 + W_1” • VERB: “-”, OBJ: “W_3 + W_4”
88	(W_0 = “RW”) + (W_1 = “CC”) + (W_2 = “RW”) + (W_3 = “NN” or “NNS” or “NNP”)	if W_0 = W_2 : • VERB: “-”, OBJ: “W_0” • VERB: “-”, OBJ: “W_2 + W_3”
89	(W_0 = “RW”) + (W_1 = “CC”) +	• VERB: “-”, OBJ: “W_0 + W_3”

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_2 = “RW”) + (W_3 = “RW”)	• VERB: “-”, OBJ: “W_2 + W_3”
90	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “CC”) + (W_2 = “RW”)	• VERB: “-”, OBJ: “W_0” • VERB: “-”, OBJ: “W_2”
91	(W_0 = “RW”) + (W_1 = “CC”) + (W_2 = “RW”) + (W_3 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0 + W_3” • VERB: “-”, OBJ: “W_2 + W_3”
92	(W_0 = “RW”) + (W_1 = “CC”) + (W_2 = “RW”)	• VERB: “-”, OBJ: “W_0” • VERB: “-”, OBJ: “W_2”
93	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “RW”) + (W_2 = “CC”) + (W_3 = “RW”)	• VERB: “-”, OBJ: “W_0 + W_1” • VERB: “-”, OBJ: “W_0 + W_3”
94	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “RW”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “CC”) + (W_4 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0 + W_1 + W_2” • VERB: “-”, OBJ: “W_0 + W_1 + W_4”
95	(W_0 = “RW”) + (W_1 = “CC”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0” • VERB: “-”, OBJ: “W_2 + W_3”
96	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “CC”) + (W_2 = “JJ” or “JJR” or “JJS”) + (W_3 = “RW”)	• VERB: “-”, OBJ: “W_0 + W_3” • VERB: “-”, OBJ: “W_2 + W_3”
97	(W_0 = “RW”) + (W_1 = “NN” or “NNS” or “NNP”) +	• VERB: “-”, OBJ: “W_0 + W_1”

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_2 = “CC”) + (W_3 = “RW”) + (W_4 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_3 + W_4”
98	(W_0 = “RW”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “CC”) + (W_3 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0 + W_1” • VERB: “-”, OBJ: “W_0 + W_3”
99	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “IN”) + (W_2 = “RW”) + (W_3 = “CC”) + (W_4 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_2” • VERB: “-”, OBJ: “W_2 + W_0” • VERB: “-”, OBJ: “W_2 + W_4”
100	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “IN”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “CC”) + (W_5 = “JJ” or “JJR” or “JJS”) + (W_6 = “RW”) + (W_7 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_5 + W_6 + W_7” • VERB: “-”, OBJ: “W_2 + W_3” • VERB: “-”, OBJ: “W_5 + W_6 + W_7 + W_0” • VERB: “-”, OBJ: “W_2 + W_3 + W_0”
101	(W_0 = “RW”) + (W_1 = “IN”) + (W_2 = “RW”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “CC”) + (W_5 = “NN” or “NNS” or “NNP”) + (W_6 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_2 + W_3” • VERB: “-”, OBJ: “W_5 + W_6” • VERB: “-”, OBJ: “W_2 + W_3 + W_0” • VERB: “-”, OBJ: “W_5 + W_6 + W_0”
102	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “IN”) +	• VERB: “-”, OBJ: “W_2 + W_0”

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “CC”) + (W_4 = “RW”)	• VERB: “-”, OBJ: “W_4 + W_0”
103	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “CC”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “IN”) + (W_5 = “RW”) + (W_6 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0 + W_1 + W_4 + W_5 + W_6” • VERB: “-”, OBJ: “W_3 + W_1 + W_4 + W_5 + W_6”
104	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “CC”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “IN”) + (W_4 = “RW”)	• VERB: “-”, OBJ: “W_0 + W_3 + W_4” • VERB: “-”, OBJ: “W_2 + W_3 + W_4”
105	(W_0 = “RW”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “CC”) + (W_3 = “JJ” or “JJR” or “JJS”) + (W_4 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0 + W_1” • VERB: “-”, OBJ: “W_3 + W_4”
106	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “CC”) + (W_2 = “RW”) + (W_3 = “RW”)	• VERB: “-”, OBJ: “W_0” • VERB: “-”, OBJ: “W_2 + W_3”
107	(W_0 = “RW”) + (W_1 = “CC”) + (W_2 = “JJ” or “JJR” or “JJS”) + (W_3 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_2 + W_3 + W_0”
108	(W_0 = “RW”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “CC”) +	• VERB: “-”, OBJ: “W_0 + W_1” • VERB: “-”, OBJ: “W_3 +

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_3 = “RW”) + (W_4 = “RW”)	W_4”
109	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “CC”) + (W_2 = “JJ” or “JJR” or “JJS”) + (W_3 = “RW”)	• VERB: “-”, OBJ: “W_0 + W_3” • VERB: “-”, OBJ: “W_2 + W_3”
110	(W_0 = “RW”) + (W_1 = “IN”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “IN”) + (W_5 = “JJ” or “JJR” or “JJS”) + (W_6 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_2 + W_3 + W_0” • VERB: “-”, OBJ: “W_2 + W_3 + W_0 + W_4 + W_5 + W_6”
111	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “CC”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “RW”)	• VERB: “-”, OBJ: “W_0 + W_1” • VERB: “-”, OBJ: “W_3 + W_4”
112	(W_0 = “RW”) + (W_1 = “CC”) + (W_2 = “JJ” or “JJR” or “JJS”) + (W_3 = “RW”) + (W_4 = “RW”)	• VERB: “-”, OBJ: “W_2 + W_3 + W_4” • VERB: “-”, OBJ: “W_0 + W_3 + W_4”
113	(W_0 = “RW”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “IN”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0 + W_1” • W_3 + W_4” • VERB: “-”, OBJ: “W_0 + W_1 + W_2 + W_3 + W_4”
114	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “IN”) +	• VERB: “-”, OBJ: “W_2 + W_3 + W_0”

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_2 = “RW”) + (W_3 = “RW”)	
115	(W_0 = “RW”) + (W_1 = “RW”) + (W_2 = “CC”) + (W_3 = “JJ” or “JJR” or “JJS”) + (W_4 = “RW”) + (W_5 = “RW”)	• VERB: “-”, OBJ: “W_0 + W_1” • VERB: “-”, OBJ: “W_3 + W_4 + W_5”
116	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “RW”) + (W_2 = “CC”) + (W_3 = “JJ” or “JJR” or “JJS”) + (W_4 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0 + W_1” • VERB: “-”, OBJ: “W_3 + W_4”
118	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “RW”) + (W_2 = “RW”) + (W_3 = “CC”) + (W_4 = “RW”) + (W_5 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0 + W_1 + W_2” • VERB: “-”, OBJ: “W_0 + W_4 + W_5”
119	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “IN”) + (W_2 = “RW”) + (W_3 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_2 + W_3 + W_0”
120	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “CC”) + (W_2 = “JJ” or “JJR” or “JJS”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “IN”) + (W_5 = “RW”)	• VERB: “-”, OBJ: “W_0 + W_4 + W_5” • VERB: “-”, OBJ: “W_2 + W_3 + W_4 + W_5”
121	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “IN”) +	• VERB: “-”, OBJ: “W_2 + W_0”

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_2 = “RW”) + (W_3 = “IN”) + (W_4 = “RW”) + (W_5 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_2 + W_0 + W_3 + W_4 + W_5”
122	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “IN”) + (W_2 = “RW”) + (W_3 = “RW”) + (W_4 = “CC”) + (W_5 = “RW”) + (W_6 = “RW”)	• VERB: “-”, OBJ: “W_2 + W_3 + W_0” • VERB: “-”, OBJ: “W_5 + W_6 + W_0”
123	(W_0 = “RW”) + (W_1 = “RW”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “CC”) + (W_4 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0 + W_1 + W_2” • VERB: “-”, OBJ: “W_0 + W_1 + W_4”
124	(W_0 = “RW”) + (W_1 = “RW”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0 + W_1 + W_2 + W_3”
125	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “IN”) + (W_2 = “RW”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_2 + W_3 + W_4 + W_0”
126	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “RW”) + (W_2 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0 + W_1 + W_2”
127	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “IN”) +	• VERB: “-”, OBJ: “W_2 + W_3 + W_0”

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_2 = “JJ” or “JJR” or “JJS”) + (W_3 = “RW”) + (W_4 = “CC”) + (W_5 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_5 + W_0”
128	(W_0 = “RW”) + (W_1 = “CC”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “IN”) + (W_4 = “RW”)	• VERB: “-”, OBJ: “W_0 + W_3 + W_4” • VERB: “-”, OBJ: “W_2 + W_3 + W_4”
129	(W_0 = “RW”) + (W_1 = “IN”) + (W_2 = “RW”) + (W_3 = “RW”)	• VERB: “-”, OBJ: “W_0 + W_1 + W_2 + W_3”
130	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “RW”) + (W_2 = “CC”) + (W_3 = “JJ” or “JJR” or “JJS”) + (W_4 = “RW”) + (W_5 = “IN”) + (W_6 = “RW”)	• VERB: “-”, OBJ: “W_0 + W_1 + W_5 + W_6” • VERB: “-”, OBJ: “W_3 + W_4 + W_5 + W_6”
131	(W_0 = “RW”) + (W_1 = “IN”) + (W_2 = “JJ” or “JJR” or “JJS”) + (W_3 = “RW”)	• VERB: “-”, OBJ: “W_0 + W_1 + W_2 + W_3”
132	(W_0 = “RW”) + (W_1 = “CC”) + (W_2 = “JJ” or “JJR” or “JJS”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_2 + W_3 + W_4” • VERB: “-”, OBJ: “W_0 + W_3 + W_4”
133	(W_0 = “NN” or “NNS” or “NNP”) +	• VERB: “-”, OBJ: “W_2 +

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_1 = “IN”) + (W_2 = “RW”) + (W_3 = “CC”) + (W_4 = “JJ” or “JJR” or “JJS”) + (W_5 = “RW”) + (W_6 = “NN” or “NNS” or “NNP”)	W_0” • VERB: “-”, OBJ: “W_4 + W_5 + W_6 + W_0”
134	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “CC”) + (W_2 = “RW”) + (W_3 = “RW”) + (W_4 = “NN” or “NNS” or “NNP”)	if W_0 = W_2 : • VERB: “-”, OBJ: “W_0” • VERB: “-”, OBJ: “W_2 + W_3 + W_4”
135	(W_0 = “RW”) + (W_1 = “RW”) + (W_2 = “RW”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “CC”) + (W_5 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0 + W_1 + W_2 + W_3” • VERB: “-”, OBJ: “W_0 + W_1 + W_2 + W_5”
136	(W_0 = “RW”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “IN”) + (W_3 = “RW”) + (W_4 = “RW”)	VERB: “-”, OBJ: “W_0 + W_1 + W_2 + W_3 + W_4”
137	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “RW”) + (W_2 = “IN”) + (W_3 = “RW”)	VERB: “-”, OBJ: “W_0 + W_1 + W_2 + W_3”
138	(W_0 = “RW”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “IN”) + (W_3 = “JJ” or “JJR” or “JJS”) +	VERB: “-”, OBJ: “W_0 + W_1 + W_2 + W_3 + W_4”

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_4 = “RW”)	
139	(W_0 = “RW”) + (W_1 = “IN”) + (W_2 = “JJ” or “JJR” or “JJS”) + (W_3 = “JJ” or “JJR” or “JJS”) + (W_4 = “RW”)	VERB: “-”, OBJ: “W_0 + W_1 + W_2 + W_3 + W_4”
140	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “CC”) + (W_3 = “RW”) + (W_4 = “RW”)	• VERB: “-”, OBJ: “W_0 + W_1” • VERB: “-”, OBJ: “W_3 + W_4”
141	(W_0 = “RW”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “CC”) + (W_4 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0 + W_1 + W_2” • VERB: “-”, OBJ: “W_0 + W_1 + W_4”
142	(W_0 = “RW”) + (W_1 = “IN”) + (W_2 = “RW”) + (W_3 = “RW”) + (W_4 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0 + W_1 + W_2 + W_3 + W_4”
143	(W_0 = “RW”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “CC”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “IN”) + (W_5 = “RW”) + (W_6 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0 + W_1 + W_4 + W_5 + W_6” • VERB: “-”, OBJ: “W_3 + W_4 + W_5 + W_6”
144	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “CC”) +	• VERB: “-”, OBJ: “W_0 + W_3”

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “RW”)	• VERB: “-”, OBJ: “W_2 + W_3”
145	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “CC”) + (W_3 = “RW”)	• VERB: “-”, OBJ: “W_0 + W_1” • VERB: “-”, OBJ: “W_0 + W_3”
146	(W_0 = “RW”) + (W_1 = “IN”) + (W_2 = “RW”) + (W_3 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0 + W_1 + W_2 + W_3”
147	(W_0 = “RW”) + (W_1 = “CC”) + (W_2 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0” • VERB: “-”, OBJ: “W_2”
148	(W_0 = “RW”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “RW”) + (W_4 = “IN”) + (W_5 = “JJ” or “JJR” or “JJS”)	• VERB: “-”, OBJ: “W_1 + W_2 + W_3” • VERB: “-”, OBJ: “W_0 + W_1 + W_2 + W_3 + W_4 + W_5”
149	(W_0 = “RW”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “RW”) + (W_4 = “RW”) + (W_5 = “IN”) + (W_6 = “RW”)	• VERB: “-”, OBJ: “W_1 + W_2 + W_3 + W_4” • VERB: “-”, OBJ: “W_0 + W_1 + W_2 + W_3 + W_4 + W_5 + W_6”
150	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “IN”) + (W_2 = “RW”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “CC”) + (W_5 = “RW”) + (W_6 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_2 + W_3 + W_0” • VERB: “-”, OBJ: “W_5 + W_6 + W_0”

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
151	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “RW”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “CC”) + (W_4 = “RW”) + (W_5 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0 + W_1 + W_2” • VERB: “-”, OBJ: “W_0 + W_4 + W_5”
152	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “RW”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0 + W_1 + W_2 + W_3”
153	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “CC”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “IN”) + (W_5 = “RW”)	• VERB: “-”, OBJ: “W_5 + W_3” • VERB: “-”, OBJ: “W_5 + W_1” • VERB: “-”, OBJ: “W_0 + W_5”
154	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “CC”) + (W_2 = “RW”) + (W_3 = “RW”) + (W_4 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0 + W_4” • VERB: “-”, OBJ: “W_2 + W_3 + W_4”
155	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “RW”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = ““CC) + (W_4 = “NN” or “NNS” or “NNP”) + (W_5 = “NN” or “NNS” or “NNP”)	if W_2 = W_5 : • VERB: “-”, OBJ: “W_0 + W_1 + W_2” • VERB: “-”, OBJ: “W_4 + W_5”
156	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “IN”) +	• VERB: “-”, OBJ: “W_0 + W_1” • VERB: “-”, OBJ: “W_3”

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_3 = “RW”)	
157	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “RW”)	• VERB: “-”, OBJ: “W_0 + W_1 + W_2”
158	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “RW”)	• VERB: “-”, OBJ: “W_0 + W_1 + W_2 + W_3”
159	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “RW”) + (W_2 = “CC”) + (W_3 = “RW”) + (W_4 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0 + W_1” • VERB: “-”, OBJ: “W_3 + W_4”
160	(W_0 = “RW”) + (W_1 = “IN”) + (W_2 = “RW”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “CC”) + (W_5 = “RW”) + (W_6 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_2 + W_3” • VERB: “-”, OBJ: “W_5 + W_6” • VERB: “-”, OBJ: “W_2 + W_3 + W_0” • VERB: “-”, OBJ: “W_5 + W_6 + W_0”
161	(W_0 = “RW”) + (W_1 = “RW”) + (W_2 = “CC”) + (W_3 = “RW”)	• VERB: “-”, OBJ: “W_0 + W_1” • VERB: “-”, OBJ: “W_0 + W_3”
162	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “RW”)	• VERB: “-”, OBJ: “W_0 + W_1 + W_2”
163	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “RW”) +	if W_2 = W_4 : • VERB: “-”, OBJ: “W_0 +

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_2 = “NN” or “NNS” or “NNP”) + (W_3 = ““CC) + (W_4 = “NN” or “NNS” or “NNP”)	W_1 + W_2” • VERB: “-”, OBJ: “W_0 + W_1 + W_4”
164	(W_0 = “RW”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “IN”) + (W_3 = “JJ” or “JJR” or “JJS”) + (W_4 = “RW”) + (W_5 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0 + W_1” • VERB: “-”, OBJ: “W_3 + W_4 + W_5”
165	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “IN”) + (W_3 = “JJ” or “JJR” or “JJS”) + (W_4 = “NN” or “NNS” or “NNP”) + (W_5 = “CC”) + (W_6 = “RW”)	• VERB: “-”, OBJ: “W_0 + W_1” • VERB: “-”, OBJ: “W_3 + W_4” • VERB: “-”, OBJ: “W_0 + W_1 + W_6” • VERB: “-”, OBJ: “W_3 + W_4 + W_6”
166	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “IN”) + (W_2 = “JJ” or “JJR” or “JJS”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “CC”) + (W_5 = “RW”)	• VERB: “-”, OBJ: “W_2 + W_3 + W_0” • VERB: “-”, OBJ: “W_4 + W_0”
167	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “CC”) + (W_3 = “JJ” or “JJR” or “JJS”) + (W_4 = “NN” or “NNS” or “NNP”) + (W_5 = “IN”) + (W_6 = “RW”) + (W_7 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0 + W_1 + W_5 + W_6 + W_7” • VERB: “-”, OBJ: “W_3 + W_4 + W_5 + W_6 + W_7”

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
168	(W_0 = “RW”) + (W_1 = “CC”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “RW”)	• VERB: “-”, OBJ: “W_0 + W_4” • VERB: “-”, OBJ: “W_2 + W_3 + W_4”
169	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “CC”) + (W_2 = “JJ” or “JJR” or “JJS”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “RW”)	• VERB: “-”, OBJ: “W_0 + W_3 + W_4” • VERB: “-”, OBJ: “W_2 + W_3 + W_4”
170	(W_0 = “RW”) + (W_1 = “RW”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “CC”) + (W_4 = “RW”) + (W_5 = “NN” or “NNS” or “NNP”) + (W_6 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0 + W_1 + W_2” • VERB: “-”, OBJ: “W_4 + W_5 + W_6”
171	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “CC”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “IN”) + (W_4 = “JJ” or “JJR” or “JJS”) + (W_5 = “RW”)	• VERB: “-”, OBJ: “W_0 + W_3 + W_4 + W_5” • VERB: “-”, OBJ: “W_2 + W_3 + W_4 + W_5”
172	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “RW”)	• VERB: “-”, OBJ: “W_0 + W_1”
173	(W_0 = “RW”) + (W_1 = “RW”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “CC”) + (W_4 = “NN” or “NNS” or “NNP”) +	• VERB: “-”, OBJ: “W_0 + W_1 + W_2” • VERB: “-”, OBJ: “W_4 + W_5”

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_5 = “NN” or “NNS” or “NNP”)	
174	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “RW”) + (W_2 = “RW”) + (W_3 = “CC”) + (W_4 = “RW”)	• VERB: “-”, OBJ: “W_0 + W_1 + W_2” • VERB: “-”, OBJ: “W_0 + W_1 + W_4”
175	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “IN”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “RW”)	• VERB: “-”, OBJ: “W_2 + W_3 + W_0”
176	(W_0 = “RW”) + (W_1 = “RW”) + (W_2 = “CC”) + (W_3 = “RW”) + (W_4 = “RW”)	• VERB: “-”, OBJ: “W_0 + W_1” • VERB: “-”, OBJ: “W_3 + W_4”
177	(W_0 = “RW”) + (W_1 = “RW”) + (W_2 = “RW”)	• VERB: “-”, OBJ: “W_0 + W_1 + W_2”
178	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “IN”) + (W_2 = “RW”) + (W_3 = “RW”)	• VERB: “-”, OBJ: “W_0 + W_2 + W_3”
179	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “IN”) + (W_2 = “JJ” or “JJR” or “JJS”) + (W_3 = “RW”)	• VERB: “-”, OBJ: “W_2 + W_3 + W_0”
180	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “RW”)	• VERB: “-”, OBJ: “W_0 + W_1 + W_2”
181	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “RW”) +	• VERB: “-”, OBJ: “W_0 + W_1 + W_2 + W_3”

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_2 = “RW”) + (W_3 = “NN” or “NNS” or “NNP”)	
182	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “RW”) + (W_3 = “RW”)	• VERB: “-”, OBJ: “W_2 + W_3” • VERB: “-”, OBJ: “W_0 + W_1 + W_2 + W_3”
183	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “IN”) + (W_2 = “RW”) + (W_3 = “RW”) + (W_4 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_2 + W_3 + W_4 + W_0”
184	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “CC”) + (W_2 = “JJ” or “JJR” or “JJS”) + (W_3 = “RW”) + (W_4 = “RW”)	• VERB: “-”, OBJ: “W_0 + W_3 + W_4” • VERB: “-”, OBJ: “W_2 + W_3 + W_4”
185	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “RW”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “RW”)	• VERB: “-”, OBJ: “W_0 + W_1” • VERB: “-”, OBJ: “W_2 + W_3”
186	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “IN”) + (W_2 = “JJ” or “JJR” or “JJS”) + (W_3 = “RW”) + (W_4 = “RW”)	• VERB: “-”, OBJ: “W_2 + W_3 + W_4 + W_0”
187	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “IN”) + (W_2 = “RW”)	• VERB: “-”, OBJ: “W_0 + W_2”
188	(W_0 = “RW”) + (W_1 = “CC”) +	• VERB: “-”, OBJ: “W_0” • VERB: “-”, OBJ: “W_2 +

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_2 = “JJ” or “JJR” or “JJS”) + (W_3 = “RW”)	W_3”
189	(W_0 = “RW”) + (W_1 = “CC”) + (W_2 = “RW”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0 + W_3 + W_4” • VERB: “-”, OBJ: “W_2 + W_3 + W_4”
190	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “IN”) + (W_3 = “RW”) + (W_4 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_3 + W_4 + W_1” • VERB: “-”, OBJ: “W_0 + W_3 + W_4”
191	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “RW”) + (W_3 = “RW”) + (W_4 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_2 + W_3 + W_4” • VERB: “-”, OBJ: “W_0 + W_1 + W_2 + W_3 + W_4”
192	(W_0 = “JJ” or “JJR” or “JJS”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “RW”) + (W_3 = “RW”)	• VERB: “-”, OBJ: “W_2 + W_3” • VERB: “-”, OBJ: “W_0 + W_1 + W_2 + W_3”
193	(W_0 = “RW”) + (W_1 = “RW”) + (W_2 = “RW”) + (W_3 = “CC”) + (W_4 = “RW”) + (W_5 = “RW”)	• VERB: “-”, OBJ: “W_0 + W_1 + W_2” • VERB: “-”, OBJ: “W_4 + W_5”
194	(W_0 = “RW”) + (W_1 = “RW”) + (W_2 = “RW”) +	• VERB: “-”, OBJ: “W_0 + W_1 + W_2” • VERB: “-”, OBJ: “W_0 +

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_3 = “CC”) + (W_4 = “RW”) + (W_5 = “NN” or “NNS” or “NNP”)	W_4 + W_5”
195	(W_0 = “RW”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “CC”) + (W_3 = “JJ” or “JJR” or “JJS”) + (W_4 = “NN” or “NNS” or “NNP”) + (W_5 = “IN”) + (W_6 = “RW”) + (W_7 = “NN” or “NNS” or “NNP”)	• VERB: “-”, OBJ: “W_0 + W_1 + W_5 + W_6 + W_7” • VERB: “-”, OBJ: “W_3 + W_4 + W_5 + W_6 + W_7”

ตารางที่ 17 กฎทางภาษาศาสตร์ที่พิจารณาร่วมกับคำศัพท์เฉพาะและมีคำกริยาอยู่ในประโยคสำหรับวิธีการ CID-CJA

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
196	(W_0 = “VB” or “VBP”) + (W_1, ..., W_n = “RW”)	• VERB: “W_0”, OBJ: “W_1, ..., W_n”
197	(W_0 = “VB” or “VBP”) + (W_n = “VB” or “VBP”) + (W_1, ..., W_n = “RW”)	• VERB: “W_0, ..., W_n”, OBJ: “W_1, ..., W_n”
198	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1 + W_2”
199	(W_0 = “VB” or “VBP”) + (W_n = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “NN” or “NNS” or “NNP”)	• VERB: “W_0, ..., W_n”, OBJ: “W_1 + W_2”
200	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “NN” or “NNS” or “NNP”), ..., (W_n = “NN” or “NNS” or “NNP” or “NNPS”)	• VERB: “W_0”, OBJ: “W_1 + W_2, ..., W_n”
201	(W_0 = “VB” or “VBP”) + (W_n = “VB” or “VBP”) +	• VERB: “W_0, ..., W_n”, OBJ: “W_1 + W_2, ..., W_n”

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_1 = “RW”) + (W_2 = “NN” or “NNS” or “NNP”), ..., (W_n = “NN” or “NNS” or “NNP” or “NNPS”)	
202	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “RW”) + (W_3 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1 + W_2 + W_3”
203	(W_0 = “VB” or “VBP”) + (W_n = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “RW”) + (W_3 = “NN” or “NNS” or “NNP”)	• VERB: “W_0, ..., W_n”, OBJ: “W_1 + W_2 + W_3”
204	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “JJ” or “JJR” or “JJS”) + (W_4 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1 + W_2” • VERB: “W_0”, OBJ: “W_3 + W_4”
205	(W_0 = “VB” or “VBP”) + (W_n = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “JJ” or “JJR” or “JJS”) + (W_4 = “NN” or “NNS” or “NNP”)	• VERB: “W_0, ..., W_n”, OBJ: “W_1 + W_2” • VERB: “W_0, ..., W_n”, OBJ: “W_3 + W_4”
206	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “JJ” or “JJR” or “JJS”) + (W_4 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1 + W_2” • VERB: “W_0”, OBJ: “W_3 + W_4”
207	(W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “RW”) + (W_3 = “RW”)	• VERB: “W_0”, OBJ: “W_2 + W_3” • VERB: “W_0”, OBJ: “W_1 + W_2 + W_3”
208	(W_0 = “VB” or “VBP”) + (W_n = “VB” or “VBP”) +	• VERB: “W_0, ..., W_n”, OBJ: “W_2 + W_3”

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “RW”) + (W_3 = “RW”)	• VERB: “W_0, ..., W_n”, OBJ: “$W_1 + W_2 + W_3$”
209	(W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “RW”) + (W_3 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “$W_2 + W_3$” • VERB: “W_0”, OBJ: “$W_1 + W_2 + W_3$”
210	(W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “JJ” or “JJR” or “JJS”) + (W_3 = “RW”) + (W_4 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “$W_3 + W_4$” • VERB: “W_0”, OBJ: “$W_1 + W_2 + W_3 + W_4$”
211	(W_0 = “VB” or “VBP”) + (W_n = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “RW”)	• VERB: “W_0, ..., W_n”, OBJ: “$W_1 + W_2$”
212	(W_0 = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “RW”)	• VERB: “W_0”, OBJ: “$W_1 + W_2$”
213	(W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “RW”)	• VERB: “W_0”, OBJ: “$W_1 + W_2$”
214	(W_0 = “VB” or “VBP”) + (W_n = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “RW”)	• VERB: “W_0, ..., W_n”, OBJ: “$W_1 + W_2$”
215	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “IN”) + (W_3 = “RW”)	• VERB: “W_0”, OBJ: “$W_2 + W_3 + W_4$”
216	(W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) +	• VERB: “W_0”, OBJ: “W_4” • VERB: “W_0”, OBJ: “W_1 +

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “IN”) + (W_4 = “RW”)	W_4” • VERB: “W_0”, OBJ: “W_4 + W_2”
216	(W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “IN”) + (W_4 = “RW”)	• VERB: “W_0”, OBJ: “W_4” • VERB: “W_0”, OBJ: “W_1 + W_4” • VERB: “W_0”, OBJ: “W_4 + W_2”
217	(W_0 = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “IN”) + (W_3 = “RW”) + (W_4 = “IN”) + (W_5 = “RW”)	• VERB: “W_0”, OBJ: “W_3 + W_1” • VERB: “W_0”, OBJ: “W_3 + W_1 + W_2 + W_5”
218	(W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “IN”) + (W_4 = “RW”) + (W_5 = “RW”)	• VERB: “W_0”, OBJ: “W_4 + W_5” • VERB: “W_0”, OBJ: “W_1 + W_4 + W_5” • VERB: “W_0”, OBJ: “W_4 + W_5 + W_2”
219	(W_0 = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “IN”) + (W_4 = “RW”) + (W_5 = “RW”)	• VERB: “W_0”, OBJ: “W_4 + W_5” • VERB: “W_0”, OBJ: “W_4 + W_5 + W_1” • VERB: “W_0”, OBJ: “W_4 + W_5 + W_2”
220	(W_0 = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “IN”) + (W_3 = “RW”) +	• VERB: “W_0”, OBJ: “W_3 + W_4 + W_5” • VERB: “W_0”, OBJ: “W_3 + W_4 + W_5 + W_1”

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_4 = “RW”) + (W_5 = “RW”)	
221	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “IN”) + (W_4 = “RW”)	• VERB: “W_0”, OBJ: “W_1 + W_2” • VERB: “W_0”, OBJ: “W_1 + W_2 + W_3 + W_4”
222	(W_0 = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “IN”) + (W_3 = “RW”)	• VERB: “W_0”, OBJ: “W_3 + W_1”
223	(W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “IN”) + (W_4 = “RW”) + (W_5 = “RW”) + (W_6 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_4 + W_5 + W_6” • VERB: “W_0”, OBJ: “W_4 + W_5 + W_6 + W_2” • VERB: “W_0”, OBJ: “W_1 + W_4 + W_5 + W_6”
224	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “CC”) + (W_4 = “RW”)	• VERB: “W_0”, OBJ: “W_1 + W_2” • VERB: “W_0”, OBJ: “W_5”
225	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “CC”) + (W_4 = “VB” or “VBP”) + (W_5 = “RW”)	• VERB: “W_0”, OBJ: “W_1 + W_2” • VERB: “W_1”, OBJ: “W_5”
226	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “RW”) + (W_3 = “CC”) + (W_4 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1 + W_2” • VERB: “W_0”, OBJ: “W_1 + W_4”

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
227	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “RW”) + (W_3 = “CC”) + (W_4 = “VB” or “VBP”) + (W_5 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1 + W_2” • VERB: “W_4”, OBJ: “W_5”
228	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “CC”) + (W_4 = “VB” or “VBP”) + (W_5 = “NN” or “NNS” or “NNP”) + (W_6 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1 + W_2” • VERB: “W_4”, OBJ: “W_5 + W_6”
229	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “CC”) + (W_3 = “VB” or “VBP”) + (W_4 = “RW”) + (W_5 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1” • VERB: “W_3”, OBJ: “W_4 + W_5”
230	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “CC”) + (W_3 = “VB” or “VBP”) + (W_4 = “RW”) + (W_5 = “RW”)	• VERB: “W_0”, OBJ: “W_1” • VERB: “W_3”, OBJ: “W_4 + W_5”
231	(W_0 = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “CC”) + (W_3 = “RW”)	• VERB: “W_0”, OBJ: “W_1” • VERB: “W_0”, OBJ: “W_3”
232	(W_0 = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “CC”) + (W_3 = “VB” or “VBP”) + (W_4 = “RW”)	• VERB: “W_0”, OBJ: “W_1” • VERB: “W_3”, OBJ: “W_4”

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
233	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “CC”) + (W_3 = “RW”) + (W_4 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1 + W_4” • VERB: “W_0”, OBJ: “W_3 + W_4”
234	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “CC”) + (W_3 = “RW”)	• VERB: “W_0”, OBJ: “W_1” • VERB: “W_0”, OBJ: “W_3”
235	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “CC”) + (W_3 = “VB” or “VBP”) + (W_4 = “RW”)	• VERB: “W_0”, OBJ: “W_1” • VERB: “W_3”, OBJ: “W_4”
236	W_0 = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “RW”) + (W_3 = “CC”) + (W_4 = “RW”)	• VERB: “W_0”, OBJ: “W_1 + W_2” • VERB: “W_0”, OBJ: “W_1 + W_4”
237	W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “RW”) + (W_3 = “NN” or “NNS” or “NNP”) (W_4 = “CC”) + (W_5 = “VB” or “VBP”) + (W_6 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1 + W_2 + W_3” • VERB: “W_4”, OBJ: “W_6”
238	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “CC”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1” • VERB: “W_0”, OBJ: “W_3 + W_4”
239	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) +	• VERB: “W_0”, OBJ: “W_1”

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_2 = “CC”) + (W_3 = “VB” or “VBP”) + (W_4 = “NN” or “NNS” or “NNP”) + (W_5 = “NN” or “NNS” or “NNP”)	• VERB: “W_3”, OBJ: “W_3 + W_4”
240	(W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “CC”) + (W_3 = “JJ” or “JJR” or “JJS”) + (W_4 = “RW”)	• VERB: “W_0”, OBJ: “W_1 + W_4” • VERB: “W_0”, OBJ: “W_3 + W_4”
241	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “CC”) + (W_4 = “RW”) + (W_5 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1 + W_2” • VERB: “W_0”, OBJ: “W_4 + W_5”
242	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “CC”) + (W_4 = “VB” or “VBP”) + (W_5 = “RW”) + (W_6 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1 + W_2” • VERB: “W_4”, OBJ: “W_5 + W_6”
243	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “CC”) + (W_4 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1 + W_2” • VERB: “W_0”, OBJ: “W_1 + W_4”
244	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “CC”) +	• VERB: “W_0”, OBJ: “W_1 + W_2” • VERB: “W_0”, OBJ: “W_4 + W_5”

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_4 = “JJ” or “JJR” or “JJS”) + (W_5 = “NN” or “NNS” or “NNP”)	
245	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “CC”) + (W_3 = “JJ” or “JJR” or “JJS”) + (W_4 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_3 + W_4 + W_1”
246	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “CC”) + (W_4 = “RW”) + (W_5 = “RW”)	• VERB: “W_0”, OBJ: “W_1 + W_2” • VERB: “W_0”, OBJ: “W_4 + W_5”
247	(W_0 = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “CC”) + (W_3 = “JJ” or “JJR” or “JJS”) + (W_4 = “RW”)	• VERB: “W_0”, OBJ: “W_1 + W_4” • VERB: “W_0”, OBJ: “W_3 + W_4”
248	(W_0 = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “CC”) + (W_3 = “VB” or “VBP”) + (W_4 = “JJ” or “JJR” or “JJS”) + (W_5 = “RW”)	• VERB: “W_0”, OBJ: “W_1 + W_4” • VERB: “W_3”, OBJ: “W_3 + W_4”
249	(W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “CC”) + (W_4 = “NN” or “NNS” or “NNP”) + (W_5 = “RW”)	• VERB: “W_0”, OBJ: “W_1 + W_2” • VERB: “W_0”, OBJ: “W_4 + W_5”
250	(W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “NN” or “NNS” or “NNP”) +	• VERB: “W_0”, OBJ: “W_1 + W_2” • VERB: “W_4”, OBJ: “W_5 +

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_3 = “CC”) + (W_4 = “VB” or “VBP”) + (W_5 = “NN” or “NNS” or “NNP”) + (W_6 = “RW”)	W_6”
251	(W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “RW”) + (W_3 = “CC”) + (W_4 = “JJ” or “JJR” or “JJS”) + (W_5 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1 + W_2” • VERB: “W_0”, OBJ: “W_4 + W_5”
252	(W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “RW”) + (W_3 = “CC”) + (W_4 = “VB” or “VBP”) + (W_5 = “JJ” or “JJR” or “JJS”) + (W_6 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1 + W_2” • VERB: “W_4”, OBJ: “W_5 + W_6”
253	(W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “RW”) + (W_3 = “RW”) + (W_4 = “CC”) + (W_5 = “RW”) + (W_6 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1 + W_2 + W_3” • VERB: “W_0”, OBJ: “W_1 + W_5 + W_6”
254	(W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “RW”) + (W_3 = “RW”) + (W_4 = “CC”) + (W_5 = “VB” or “VBP”) + (W_6 = “RW”) + (W_7 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1 + W_2 + W_3” • VERB: “W_0”, OBJ: “W_1 + W_6 + W_7”
255	(W_0 = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP”) +	• VERB: “W_0”, OBJ: “W_3 + W_4 + W_1”

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_2 = “IN”) + (W_3 = “RW”) + (W_4 = “NN” or “NNS” or “NNP”)	
256	(W_0 = “VB” or “VBP”) + (W_n = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “IN”) + (W_3 = “RW”) + (W_4 = “NN” or “NNS” or “NNP”)	• VERB: “W_0, …, W_n”, OBJ: “W_3 + W_4 + W_1” •
257	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “RW”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “CC”) + (W_5 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1 + W_2 + W_3” • VERB: “W_0”, OBJ: “W_1 + W_2 + W_5”
258	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “RW”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “CC”) + (W_5 = “VB” or “VBP”) + (W_6 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1 + W_2 + W_3” • VERB: “W_5”, OBJ: “W_6”
259	(W_0 = “VB” or “VBP”) + (W_n = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “RW”) + (W_3 = “RW”) +	• VERB: “W_0, …, W_n”, OBJ: “W_1 + W_2 + W_3”
260	(W_0 = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “RW”) + (W_3 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1 + W_2 + W_3”
261	(W_0 = “VB” or “VBP”) + (W_n = “VB” or “VBP”) + (W_1 = “RW”) +	• VERB: “W_0”, OBJ: “W_1 + W_2 + W_3”

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_2 = “IN”) + (W_3 = “RW”) +	
262	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “IN”) + (W_3 = “RW”) + (W_4 = “RW”) +	• VERB: “W_0”, OBJ: “W_1 + W_2 + W_3 + W_4”
263	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “CC”) + (W_3 = “JJ” or “JJR” or “JJS”) + (W_4 = “NN” or “NNS” or “NNP”) + (W_5 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_3 + W_4 + W_5” • VERB: “W_0”, OBJ: “W_1 + W_4 + W_5”
264	(W_0 = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “CC”) + (W_4 = “RW”) + (W_5 = “RW”)	• VERB: “W_0”, OBJ: “W_1 + W_2” • VERB: “W_0”, OBJ: “W_4 + W_5”
265	(W_0 = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “CC”) + (W_4 = “VB” or “VBP”) + (W_5 = “RW”) + (W_6 = “RW”)	• VERB: “W_0”, OBJ: “W_1 + W_2” • VERB: “W_4”, OBJ: “W_5 + W_6”
266	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_4 = “CC”) + (W_5 = “VB” or “VBP”) +	• VERB: “W_0”, OBJ: “W_1 + W_2” • VERB: “W_5”, OBJ: “W_6”

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_6 = “RW”)	
267	(W_0 = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “CC”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “RW”)	• VERB: “W_0”, OBJ: “W_1 + W_2” • VERB: “W_0”, OBJ: “W_4 + W_5”
268	(W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “CC”) + (W_4 = “RW”)	• VERB: “W_0”, OBJ: “W_1 + W_2” • VERB: “W_0”, OBJ: “W_1 + W_4”
269	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “IN”) + (W_3 = “RW”) + (W_4 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1 + W_2 + W_3 + W_4”
270	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “CC”) + (W_3 = “RW”) + (W_4 = “RW”)	• VERB: “W_0”, OBJ: “W_1” • VERB: “W_0”, OBJ: “W_1 + W_4”
271	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “CC”) + (W_3 = “VB” or “VBP”) + (W_4 = “RW”) + (W_5 = “RW”)	• VERB: “W_0”, OBJ: “W_1” • VERB: “W_0”, OBJ: “W_1 + W_4”
272	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “CC”) + (W_3 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1” • VERB: “W_0”, OBJ: “W_3”
273	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “CC”) +	• VERB: “W_0”, OBJ: “W_1” • VERB: “W_0”, OBJ: “W_4”

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_3 = “VB” or “VBP”) + (W_4 = “NN” or “NNS” or “NNP”)	
274	(W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “RW”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “CC”) + (W_5 = “RW”) + (W_6 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1 + W_2 + W_3” • VERB: “W_0”, OBJ: “W_1 + W_5 + W_6”
275	(W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “RW”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “CC”) + (W_5 = “VB” or “VBP”) + (W_6 = “RW”) + (W_7 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1 + W_2 + W_3” • VERB: “W_5”, OBJ: “W_6 + W_7”
276	(W_0 = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “IN”) + (W_4 = “RW”)	• VERB: “W_0”, OBJ: “W_1 + W_2” • VERB: “W_0”, OBJ: “W_4”
277	(W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “RW”)	• VERB: “W_0”, OBJ: “W_1 + W_2 + W_3”
278	(W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “JJ” or “JJR” or “JJS”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “RW”)	• VERB: “W_0”, OBJ: “W_1 + W_2 + W_3”+ W_4”

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
279	(W_0 = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “RW”) + (W_3 = “CC”) + (W_4 = “RW”) + (W_5 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1 + W_2” • VERB: “W_0”, OBJ: “W_4 + W_5”
280	(W_0 = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “RW”) + (W_3 = “CC”) + (W_4 = “VB” or “VBP”) + (W_5 = “RW”) + (W_6 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1 + W_2” • VERB: “W_4”, OBJ: “W_5 + W_6”
281	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “IN”) + (W_3 = “RW”) + (W_4 = “NN” or “NNS” or “NNP”) + (W_5 = “CC”) + (W_6 = “RW”) + (W_7 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_3 + W_4” • VERB: “W_0”, OBJ: “W_6 + W_7” • VERB: “W_0”, OBJ: “W_3 + W_4 + W_1” • VERB: “W_0”, OBJ: “W_6 + W_7 + W_1”
282	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “RW”) + (W_3 = “CC”) + (W_4 = “RW”)	• VERB: “W_0”, OBJ: “W_1 + W_2” • VERB: “W_0”, OBJ: “W_1 + W_4”
283	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “RW”) + (W_3 = “CC”) + (W_4 = “VB” or “VBP”) +	• VERB: “W_0”, OBJ: “W_1 + W_2” • VERB: “W_4”, OBJ: “W_5”

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_5 = “RW”)	
284	(W_0 = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “RW”)	• VERB: “W_0”, OBJ: “W_1 + W_2 + W_3”
285	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “CC”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “NN” or “NNS” or “NNP”) + (W_5 = “RW”)	• VERB: “W_0”, OBJ: “W_1 + W_5” • VERB: “W_0”, OBJ: “W_3 + W_4 + W_5”
286	(W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “CC”) + (W_3 = “JJ” or “JJR” or “JJS”) + (W_4 = “NN” or “NNS” or “NNP”) + (W_5 = “RW”)	• VERB: “W_0”, OBJ: “W_1 + W_4 + W_5” • VERB: “W_0”, OBJ: “W_3 + W_4 + W_5”
287	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “RW”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “CC”) + (W_5 = “RW”) + (W_6 = “NN” or “NNS” or “NNP”) + (W_7 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1 + W_2 + W_3” • VERB: “W_0”, OBJ: “W_5 + W_6 + W_7”
288	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “RW”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “CC”) + (W_5 = “VB” or “VBP”) +	• VERB: “W_0”, OBJ: “W_1 + W_2 + W_3” • VERB: “W_5”, OBJ: “W_6 + W_7 + W_8”

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_6 = “RW”) + (W_7 = “NN” or “NNS” or “NNP”) + (W_8 = “NN” or “NNS” or “NNP”)	
289	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “RW”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “CC”) + (W_5 = “NN” or “NNS” or “NNP”) + (W_6 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1 + W_2 + W_3” • VERB: “W_0”, OBJ: “W_5 + W_6”
290	(W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “RW”) + (W_3 = “RW”) + (W_4 = “CC”) + (W_5 = “RW”)	• VERB: “W_0”, OBJ: “W_1 + W_2 + W_3” • VERB: “W_0”, OBJ: “W_1 + W_2 + W_5”
291	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “RW”) + (W_3 = “CC”) + (W_4 = “RW”) + (W_5 = “RW”)	• VERB: “W_0”, OBJ: “W_1 + W_2” • VERB: “W_0”, OBJ: “W_4 + W_5”
292	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “RW”) + (W_3 = “CC”) + (W_4 = “VB” or “VBP”) + (W_5 = “RW”) + (W_6 = “RW”)	• VERB: “W_0”, OBJ: “W_1 + W_2” • VERB: “W_0”, OBJ: “W_4 + W_5”
293	(W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “RW”) +	• VERB: “W_0”, OBJ: “W_1 + W_2 + W_3 + W_4”

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_3 = “RW”) + (W_4 = “NN” or “NNS” or “NNP”)	
294	(W_0 = “VB” or “VBP”) + (W_1 = “NN” or “NNS” or “NNP”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “RW”) + (W_4 = “RW”)	• VERB: “W_0”, OBJ: “W_3 + W_4” • VERB: “W_0”, OBJ: “W_1 + W_2 + W_3 + W_4”
295	(W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “RW”) + (W_3 = “NN” or “NNS” or “NNP”) + (W_4 = “RW”)	• VERB: “W_0”, OBJ: “W_1 + W_2” • VERB: “W_0”, OBJ: “W_3 + W_4”
296	(W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “RW”) + (W_3 = “VB” or “VBP”) + (W_4 = “NN” or “NNS” or “NNP”) + (W_5 = “RW”)	• VERB: “W_0”, OBJ: “W_1 + W_2” • VERB: “W_0”, OBJ: “W_3 + W_4”
297	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “CC”) + (W_3 = “JJ” or “JJR” or “JJS”) + (W_4 = “RW”)	• VERB: “W_0”, OBJ: “W_1” • VERB: “W_0”, OBJ: “W_3 + W_4”
298	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “CC”) + (W_3 = “RW”) + (W_4 = “NN” or “NNS” or “NNP”) + (W_5 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1 + W_4 + W_5” • VERB: “W_0”, OBJ: “W_3 + W_4 + W_5”
299	(W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “NN” or “NNS” or “NNP”) +	• VERB: “W_0”, OBJ: “W_4 + W_5 + W_2” • VERB: “W_0”, OBJ: “W_1 +

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_3 = “IN”) + (W_4 = “RW”) + (W_5 = “NN” or “NNS” or “NNP”)	W_4 + W_5”
300	(W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “JJ” or “JJR” or “JJS”) + (W_3 = “RW”) + (W_4 = “RW”) + (W_5 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_3 + W_4 + W_5” • VERB: “W_0”, OBJ: “W_1 + W_2 + W_3 + W_4 + W_5”
301	(W_0 = “VB” or “VBP”) + (W_1 = “JJ” or “JJR” or “JJS”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “RW”) + (W_4 = “RW”)	• VERB: “W_0”, OBJ: “W_3 + W_4” • VERB: “W_0”, OBJ: “W_1 + W_2 + W_3 + W_4”
302	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “RW”) + (W_3 = “RW”) + (W_4 = “CC”) + (W_5 = “RW”) + (W_6 = “RW”)	• VERB: “W_0”, OBJ: “W_1 + W_2 + W_3” • VERB: “W_0”, OBJ: “W_5 + W_6”
303	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “RW”) + (W_3 = “RW”) + (W_4 = “CC”) + (W_5 = “VB” or “VBP”) + (W_6 = “RW”) + (W_7 = “RW”)	• VERB: “W_0”, OBJ: “W_1 + W_2 + W_3” • VERB: “W_5”, OBJ: “W_6 + W_7”
304	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “RW”) +	• VERB: “W_0”, OBJ: “W_1 + W_2 + W_3” • VERB: “W_0”, OBJ: “W_1 +

ลำดับ	รูปแบบหัวข้อ	ผลลัพธ์
	(W_3 = “RW”) + (W_4 = “CC”) + (W_5 = “RW”) + (W_6 = “NN” or “NNS” or “NNP”)	W_5 + W_6”
305	(W_0 = “VB” or “VBP”) + (W_1 = “RW”) + (W_2 = “NN” or “NNS” or “NNP”) + (W_3 = “CC”) + (W_4 = “JJ” or “JJR” or “JJS”) + (W_5 = “NN” or “NNS” or “NNP”) + (W_6 = “IN”) + (W_7 = “RW”) + (W_8 = “NN” or “NNS” or “NNP”)	• VERB: “W_0”, OBJ: “W_1 + W_2 + W_6 + W_7 + W_8” • VERB: “W_0”, OBJ: “W_4 + W_5 + W_6 + W_7 + W_8”

ภาคผนวก ข

โครงสร้างการจัดเก็บคำสำคัญจากประโยคแต่ละประเภทสำหรับวิธีการ CID-CJA

ตารางที่ 18 โครงสร้างการจัดเก็บคำสำคัญจากข้อมูลประเภทของอายุ

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
1	age + number_x	● NUMBER1: NULL, SIGN: >=, NUMBER2: number_x
2	age + number_x + year + old	● NUMBER1: NULL, SIGN: >=, NUMBER2: number_x
3	number_x + year + old	● NUMBER1: NULL, SIGN: >=, NUMBER2: number_x

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
4	age + not + over + number_x + year	● NUMBER1: NULL, SIGN: <, NUMBER2: number_x
5	not + over + number_x + year + old	● NUMBER1: NULL, SIGN: <, NUMBER2: number_x
6	age + below + number_x	● NUMBER1: NULL, SIGN: <, NUMBER2: number_x
7	age + not + over + than + number_x + years + old	● NUMBER1: NULL, SIGN: <, NUMBER2: number_x
8	age + over + number_x + year + old	● NUMBER1: NULL, SIGN: >, NUMBER2: number_x
9	over + number_x + years + old	● NUMBER1: NULL, SIGN: >, NUMBER2: number_x
10	age + number_x + year + old + up	● NUMBER1: NULL, SIGN: >, NUMBER2: number_x
11	age + above + number_x + year + and + up	● NUMBER1: NULL, SIGN: >, NUMBER2: number_x
12	age + number_x + up + years + old	● NUMBER1: NULL, SIGN: >, NUMBER2: number_x
13	age + more + than + number_x + years + old	● NUMBER1: NULL, SIGN: >, NUMBER2: number_x
14	age + number_x + years + up	● NUMBER1: NULL, SIGN: >, NUMBER2: number_x
15	up + to + number_x + years + old	● NUMBER1: NULL, SIGN: >, NUMBER2: number_x
16	aged + above + number_x + year	● NUMBER1: NULL, SIGN: >, NUMBER2: number_x

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
17	age + over + than + number_x + years + old	● NUMBER1: NULL, SIGN: >, NUMBER2: number_x
18	age + up + to + number_x + years + old	● NUMBER1: NULL, SIGN: >, NUMBER2: number_x
19	number_x + - + number_y + year + of + age	● NUMBER1: number_x, SIGN: -, NUMBER2: number_y
20	age + between + number_x + - + number_y + year + old	● NUMBER1: number_x, SIGN: -, NUMBER2: number_y
21	age + number_x + to + number_y + years + old	● NUMBER1: number_x, SIGN: -, NUMBER2: number_y
22	age + number_x + - + number_y + year + old	● NUMBER1: number_x, SIGN: -, NUMBER2: number_y
23	age + around + number_x + - + number_y + + year + old	● NUMBER1: NULL, SIGN: >=, NUMBER2: number_x
24	age + over + number_x + but + not + over + number_y + year	● NUMBER1: number_x, SIGN: -, NUMBER2: number_y

ตารางที่ 19 โครงสร้างการจัดเก็บคำสำคัญจากข้อมูลประเภทของเพศ

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
1	for + sex_x	● SEX: sex_x
2	sex_x	● SEX: sex_x
3	sex_x_1 + or + sex_x_2	● SEX: sex_x_1 ● SEX: sex_x_2
4	sex_x_1 + / + sex_x_2	● SEX: sex_x_1 ● SEX: sex_x_2

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
5	all + gender	● SEX: sex_x_1 (“male”) ● SEX: sex_x_2 (“female”)
6	any + genders	● SEX: sex_x_1 (“male”) ● SEX: sex_x_2 (“female”)

ตารางที่ 20 โครงสร้างการจัดเก็บคำสำคัญจากข้อมูลประเภทของสัญชาติ

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
1	national_ x_1 + national	● NATIONAL: national_ x_1
2	national_ x_1 + nationality	● NATIONAL: national_ x_1
3	national_ x_1 + national + only	● NATIONAL: national_ x_1
4	national_ x_1 + nationality + only	● NATIONAL: national_ x_1
5	national_ x_1 + citizen + only	● NATIONAL: national_ x_1
6	national_ x_1 + citizen	● NATIONAL: national_ x_1
7	foreign	● NATIONAL: national_ x_1

ตารางที่ 21 กฎทางภาษาศาสตร์สำหรับสาขาวิชาสำหรับประโยคที่ถูกจำแนกประเภทเป็นการศึกษา

ลำดับ	รูปแบบประโยค	**ระบุเป็น** field_ x ในตารางที่
1	(W_0 = “NN” or “NNS” or “NNP”)	● field_ x
2	(W_0 = “RW”)	● field_ x
3	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “NN” or “NNS” or “NNP”)	● field_ x

ลำดับ	รูปแบบประโยค	ระบุเป็น field_ x ในตารางที่
4	(W_0 = “NN” or “NNS” or “NNP”) + (W_1 = “RW”)	● field_ x
5	(W_0 = “RW”) + (W_1 = “NN” or “NNS” or “NNP”)	● field_ x
6	(W_0 = “RW”) + (W_1 = “RW”)	● field_ x

ตารางที่ 22 โครงสร้างการจัดเก็บคำสำคัญจากข้อมูลประเภทของการศึกษาที่ต้องการ

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
1	bachelor + degree	● DEGREE: bachelor, FIELD: not specified
2	bachelor + also + can + be + considered	● DEGREE:

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
		bachelor, FIELD: not specified
3	diploma or degree	• DEGREE: diploma, FIELD: not specified
4	master + degree + preferred	• DEGREE: master, FIELD: not specified
5	masters + is + a + plus	• DEGREE: master, FIELD: not specified
6	bachelor or higher degrees	• DEGREE: bachelor, FIELD: not specified • DEGREE: master, FIELD: not specified • DEGREE: doctor, FIELD: not specified
7	undergraduate + degree + or + above	• DEGREE: bachelor, FIELD: not specified • DEGREE: master, FIELD: not specified • DEGREE: doctor, FIELD: not

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
		specified
8	welcome new graduated	• DEGREE: bachelor, FIELD: not specified
9	new graduated are also welcome	• DEGREE: bachelor, FIELD: not specified
10	bachelor degree in + field_ x_1	• DEGREE: bachelor, FIELD: field_ x_1
11	degree in + field_ x_1	• DEGREE: bachelor, FIELD: field_ x_1 • DEGREE: master, FIELD: field_ x_1 • DEGREE: doctor, FIELD: field_ x_1
12	graduated in bachelor of field_ x_1	• DEGREE: bachelor, FIELD: field_ x_1
13	bachelor degree or higher in field_ x_1	• DEGREE: bachelor, FIELD: field_ x_1 • DEGREE: master, FIELD: field_ x_1 • DEGREE: doctor,

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
		FIELD: field_ x_1
14	graduated bachelor degree or higher in field_ x_1	• DEGREE: bachelor, FIELD: field_ x_1 • DEGREE: master, FIELD: field_ x_1 • DEGREE: doctor, FIELD: field_ x_1
15	bachelor degree in field_ x_1 + is + preferable	• DEGREE: bachelor, FIELD: field_ x_1
16	educational degree in field_ x_1	• DEGREE: bachelor, FIELD: field_ x_1 • DEGREE: master, FIELD: field_ x_1 • DEGREE: doctor, FIELD: field_ x_1
17	master or phd in + field_ x_1	• DEGREE: master, FIELD: field_ x_1 • DEGREE: doctor, FIELD: field_ x_1
18	master + / + doctoral degree in field_ x_1	• DEGREE: master, FIELD: field_ x_1 • DEGREE: doctor, FIELD: field_ x_1

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
19	bachelor degree in + field_ x_1 + area	• DEGREE: bachelor, FIELD: field_ x_1
20	degree in + field_ x_1 + or + field_x_2	• DEGREE: bachelor, FIELD: field_ x_1 • DEGREE: bachelor, FIELD: field_ x_2 • DEGREE: master, FIELD: field_ x_1 • DEGREE: master, FIELD: field_ x_2 • DEGREE: doctor, FIELD: field_ x_1 • DEGREE: doctor, FIELD: field_ x_2
21	bachelor degree in + field_ x_1 + or + field_ x_2	• DEGREE: bachelor, FIELD: field_ x_1 • DEGREE: bachelor, FIELD: field_ x_2
22	bachelor degree or higher in + field_ x_1 + or + field_ x_2	• DEGREE: bachelor, FIELD: field_ x_1 • DEGREE:

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
		bachelor, FIELD: field_ x_2 • DEGREE: master, FIELD: field_ x_1 • DEGREE: master, FIELD: field_ x_2 • DEGREE: doctor, FIELD: field_ x_1 • DEGREE: doctor, FIELD: field_ x_2
23	bachelor degree or higher degree in + field_ x_1 + or + field_ x_2	• DEGREE: bachelor, FIELD: field_ x_1 • DEGREE: bachelor, FIELD: field_ x_2 • DEGREE: master, FIELD: field_ x_1 • DEGREE: master, FIELD: field_ x_2 • DEGREE: doctor, FIELD: field_ x_1 • DEGREE: doctor, FIELD: field_ x_2
24	master or bachelor degree in field_ x_1 + or + field_x2	• DEGREE: bachelor, FIELD: field_ x_1

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
		• DEGREE: bachelor, FIELD: field_ x_2 • DEGREE: master, FIELD: field_ x_1 • DEGREE: master, FIELD: field_ x_2
25	bachelor degree in the relevant + field_ x_1 + or + field_ x_2 + field	• DEGREE: bachelor, FIELD: field_ x_1 • DEGREE: bachelor, FIELD: field_ x_2
26	bachelor degree or higher in faculty of field_ x_1 + / + field_ x_2	• DEGREE: bachelor, FIELD: field_ x_1 • DEGREE: bachelor, FIELD: field_ x_2 • DEGREE: master, FIELD: field_ x_1 • DEGREE: master, FIELD: field_ x_2 • DEGREE: doctor, FIELD: field_ x_1 • DEGREE: doctor, FIELD: field_ x_2

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
27	bachelor degree or higher in field_ x_1 + and + / + or + field_ x_2	• DEGREE: bachelor, FIELD: field_ x_1 • DEGREE: bachelor, FIELD: field_ x_2 • DEGREE: master, FIELD: field_ x_1 • DEGREE: master, FIELD: field_ x_2 • DEGREE: doctor, FIELD: field_ x_1 • DEGREE: doctor, FIELD: field_ x_2
28	bachelor in field_ x_1 + / + field_ x_2 + or + any + field_x_n	• DEGREE: bachelor, FIELD: field_ x_1 + ... + • DEGREE: bachelor, FIELD: field_x_n

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
29	bachelor or master degree in field_ x_1 + / + field_ x_2 + / + ... + or + field_x_n	• DEGREE: bachelor, FIELD: field_ x_1 + ... + • DEGREE: bachelor, FIELD: field_x_n • DEGREE: master, FIELD: field_ x_1 + ... + • DEGREE: master, FIELD: field_x_n
30	master degree bachelor + field_ x_1 +, + field_ x_2	• DEGREE: bachelor, FIELD: field_ x_1 + ... + • DEGREE: bachelor, FIELD: field_x_n • DEGREE: master, FIELD: field_ x_1 +

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
		... + ● DEGREE: master, FIELD: field_x_n
31	bachelor degree in + field_ x_1 +, + ... + , + field_x_n	● DEGREE: bachelor, FIELD: field_ x_1 + ... + ● DEGREE: bachelor, FIELD: field_x_n
32	bachelor degree or higher in + field_ x_1 + , + field_ x_2 + , + ... + , + field_x_n	● DEGREE: bachelor, FIELD: field_ x_1 + ... + ● DEGREE: bachelor, FIELD: field_x_n ● DEGREE: master, FIELD: field_ x_1 + ... + ● DEGREE: master,

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
		FIELD: field_x_n ● DEGREE: doctor, FIELD: field_ x_1 + ... + ● DEGREE: doctor, FIELD: field_x_n
33	educational degree in + field_ x_1 + , + field_ x_2 + , + ... + , + field_x_n	● DEGREE: bachelor, FIELD: field_ x_1 + ... + ● DEGREE: bachelor, FIELD: field_x_n ● DEGREE: master, FIELD: field_ x_1 + ... + ● DEGREE: master, FIELD: field_x_n ● DEGREE: doctor, FIELD: field_ x_1 + ...

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
		+ • DEGREE: doctor, FIELD: field_x_n
34	bachelor in + field_ x_1 + , + field_ x_2 + , + … + , + field_x_n	• DEGREE: bachelor, FIELD: field_ x_1 + … + • DEGREE: bachelor, FIELD: field_x_n
35	a relevant master degree + (+ field_ x_1 + , + field_ x_2 + , + … + , + field_x_n +)	• DEGREE: master, FIELD: field_ x_1 + … + • DEGREE: master, FIELD: field_x_n
36	education +: + bachelor + degree + or + master + degree + in + field_ x_1 + , + field_ x_2 + , + … + , + field_x_n	• DEGREE: bachelor, FIELD: field_ x_1 + … + • DEGREE: bachelor, FIELD: field_x_n

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
		• DEGREE: master, FIELD: field_ x_1 + ... + • DEGREE: master, FIELD: field_ x_n
37	educated to degree level in + field_ x_1 + , + field_ x_2 + , + ... + , + field_ x_n	• DEGREE: bachelor, FIELD: field_ x_1 + ... + • DEGREE: bachelor, FIELD: field_ x_n • DEGREE: master, FIELD: field_ x_1 + ... + • DEGREE: master, FIELD: field_ x_n • DEGREE: doctor, FIELD: field_ x_1 + ... +

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
		• DEGREE: doctor, FIELD: field_x_n
38	undergraduate or masters degree in + field_ x_1 + , + field_ x_2 + , + ... + and + field_x_n	• DEGREE: bachelor, FIELD: field_ x_1 + ... + • DEGREE: bachelor, FIELD: field_x_n • DEGREE: master, FIELD: field_ x_1 + ... + • DEGREE: master, FIELD: field_x_n
39	educational degree in + field_ x_1 + , + field_ x_2 + , + ... + or + field_x_n	• DEGREE: bachelor, FIELD: field_ x_1 + ... + • DEGREE: bachelor, FIELD: field_x_n
40	degree in + field_ x_1 + , + field_ x_2 + , + ... +	• DEGREE:

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
	or + field_x_n	bachelor, FIELD: field_ x_1 + ... + • DEGREE: bachelor, FIELD: field_x_n • DEGREE: master, FIELD: field_ x_1 + ... + • DEGREE: master, FIELD: field_x_n • DEGREE: doctor, FIELD: field_ x_1 + ... + • DEGREE: doctor, FIELD: field_x_n
41	bachelor degree in + field_ x_1 + , + field_ x_2 + , + ... + or + field_x_n	• DEGREE: bachelor, FIELD: field_ x_1 + ... +

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
		• DEGREE: bachelor, FIELD: field_x_n
42	bachelor degree or higher degree in + field_ x_1 +, + … + , + field_x_n	• DEGREE: bachelor, FIELD: field_ x_1 + … + • DEGREE: bachelor, FIELD: field_x_n • DEGREE: master, FIELD: field_ x_1 + … + • DEGREE: master, FIELD: field_x_n • DEGREE: doctor, FIELD: field_ x_1 + … + • DEGREE: doctor, FIELD: field_x_n
43	bachelor degree or higher in + field_ x_1 + , + field_ x_2 + , + … + or + field_x_n	• DEGREE: bachelor, FIELD:

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
		field_ x_1 + ... + • DEGREE: bachelor, FIELD: field_x_n • DEGREE: master, FIELD: field_ x_1 + ... + • DEGREE: master, FIELD: field_x_n • DEGREE: doctor, FIELD: field_ x_1 + ... + • DEGREE: doctor, FIELD: field_x_n
44	bachelor or master degree in + field_ x_1 + , + field_ x_2 + , + ... + or + field_x_n	• DEGREE: bachelor, FIELD: field_ x_1 + ... + • DEGREE:

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
		bachelor, FIELD: field_x_n ● DEGREE: master, FIELD: field_ x_1 + ... + ● DEGREE: master, FIELD: field_x_n
45	bachelor degree or master degree in + field_ x_1 + , + field_ x_2 + , + ... + or + field_x_n	● DEGREE: bachelor, FIELD: field_ x_1 + ... + ● DEGREE: bachelor, FIELD: field_x_n ● DEGREE: master, FIELD: field_ x_1 + ... + ● DEGREE: master, FIELD: field_x_n
46	degree in + field_ x_1 + , + field_ x_2 + , + ... + or + field_x_n	● DEGREE: bachelor, FIELD: field_ x_1

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
		+ … + • DEGREE: bachelor, FIELD: field_x_n • DEGREE: master, FIELD: field_ x_1 + … + • DEGREE: master, FIELD: field_x_n • DEGREE: doctor, FIELD: field_ x_1 + … + • DEGREE: doctor, FIELD: field_x_n
47	degree of + field_ x_1 + , + field_ x_2 + , + … + or + field_x_n	• DEGREE: bachelor, FIELD: field_ x_1 + … + • DEGREE: bachelor, FIELD:

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
		field_x_n • DEGREE: master, FIELD: field_ x_1 + ... + DEGREE: master, FIELD: field_x_n • DEGREE: doctor, FIELD: field_ x_1 + ... + DEGREE: doctor, FIELD: field_x_n
48	bachelor degree (or above) majoring in + field_ x_1 + , + field_ x_2 + , + ... + or + field_x_n	• DEGREE: bachelor, FIELD: field_ x_1 + ... + DEGREE: bachelor, FIELD: field_x_n • DEGREE: master, FIELD: field_ x_1 + ...

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
		+ • DEGREE: master, FIELD: field_x_n • DEGREE: doctor, FIELD: field_ x_1 + ... + • DEGREE: doctor, FIELD: field_x_n
49	bachelors or higher in + field_ x_1 + , + field_ x_2 + , + ... + or + field_x_n	• DEGREE: bachelor, FIELD: field_ x_1 + ... + • DEGREE: bachelor, FIELD: field_x_n • DEGREE: master, FIELD: field_ x_1 + ... + • DEGREE: master, FIELD: field_x_n • DEGREE: doctor, FIELD: field_ x_1

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
		+ … + • DEGREE: doctor, FIELD: field_x_n
50	at least a bachelor degree in + field_ x_1 + , + field_ x_2 + , + … + or + field_x_n	• DEGREE: bachelor, FIELD: field_ x_1 + … + • DEGREE: bachelor, FIELD: field_x_n • DEGREE: master, FIELD: field_ x_1 + … + • DEGREE: master, FIELD: field_x_n • DEGREE: doctor, FIELD: field_ x_1 + … + • DEGREE: doctor, FIELD: field_x_n

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
51	minimum bachelor degree in + field_ x_1 + , + field_ x_2 + , + ... + or + field_x_n	• DEGREE: bachelor, FIELD: field_ x_1 + ... + • DEGREE: bachelor, FIELD: field_x_n • DEGREE: master, FIELD: field_ x_1 + ... + • DEGREE: master, FIELD: field_x_n • DEGREE: doctor, FIELD: field_ x_1 + ... + • DEGREE: doctor, FIELD: field_x_n
52	minimum of bachelor degree in + field_ x_1 + , + field_ x_2 + , + ... + or + field_x_n	• DEGREE: bachelor, FIELD: field_ x_1 + ...

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
		+ ● DEGREE: bachelor, FIELD: field_x_n ● DEGREE: master, FIELD: field_ x_1 + ... + ● DEGREE: master, FIELD: field_x_n ● DEGREE: doctor, FIELD: field_ x_1 + ... + ● DEGREE: doctor, FIELD: field_x_n
53	bachelor degree in + field_ x_1 + , + field_ x_2 + , + ... + , + or + field_x_n	● DEGREE: bachelor, FIELD: field_ x_1 + ... + ● DEGREE: bachelor, FIELD: field_x_n
54	bachelor or master degree in + field_ x_1 + ,	● DEGREE:

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
	+ field_ x_2 + , + ... + , + or + field_x_n	bachelor, FIELD: field_ x_1 + ... + • DEGREE: bachelor, FIELD: field_x_n • DEGREE: master, FIELD: field_ x_1 + ... + • DEGREE: master, FIELD: field_x_n
55	candidates should possess at least a bachelor's degree in + field_ x_1 + , + field_ x_2 + , + ... + or + field_x_n	• DEGREE: bachelor, FIELD: field_ x_1 + ... + • DEGREE: bachelor, FIELD: field_x_n • DEGREE: master, FIELD: field_ x_1 + ...

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
		+ ● DEGREE: master, FIELD: field_x_n ● DEGREE: doctor, FIELD: field_ x_1 + ... + ● DEGREE: doctor, FIELD: field_x_n
56	bachelor of master degree in + field_ x_1 + , + field_ x_2 + , + ... + or + field_x_n	● DEGREE: bachelor, FIELD: field_ x_1 + ... + ● DEGREE: bachelor, FIELD: field_x_n ● DEGREE: master, FIELD: field_ x_1 + ... + ● DEGREE: master, FIELD: field_x_n
57	master degree in + field_ x_1 + , + field_ x_2 + , + ... + or + field_x_n	● DEGREE: master, FIELD: field_ x_1

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
		+ … + • DEGREE: master, FIELD: field_x_n
58	master or doctorate degree or equivalent experience in + field_ x_1 + , + field_ x_2 + , + … + or + field_x_n	• DEGREE: master, FIELD: field_ x_1 + … + • DEGREE: master, FIELD: field_x_n • DEGREE: doctor, FIELD: field_ x_1 + … + • DEGREE: doctor, FIELD: field_x_n
59	bachelor/master/doctor in + field_ x_1 + , + field_ x_2 + , + … + or + field_x_n	• DEGREE: bachelor, FIELD: field_ x_1 + … + • DEGREE: bachelor, FIELD: field_x_n

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
		• DEGREE: master, FIELD: field_ x_1 + ... + • DEGREE: master, FIELD: field_ x_n • DEGREE: doctor, FIELD: field_ x_1 + ... + • DEGREE: doctor, FIELD: field_ x_n
60	bachelor degree of higher in + field_ x_1 + , + field_ x_2 + , + ... + or + field_ x_n	• DEGREE: bachelor, FIELD: field_ x_1 + ... + • DEGREE: bachelor, FIELD: field_ x_n • DEGREE: master, FIELD: field_ x_1 + ... +

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
		• DEGREE: master, FIELD: field_x_n • DEGREE: doctor, FIELD: field_ x_1 + ... + • DEGREE: doctor, FIELD: field_x_n
61	bachelor degree/master degree in + field_ x_1 + , + field_ x_2 + , + ... + or + field_x_n	• DEGREE: bachelor, FIELD: field_ x_1 + ... + • DEGREE: bachelor, FIELD: field_x_n • DEGREE: master, FIELD: field_ x_1 + ... + • DEGREE: master, FIELD: field_x_n
62	graduated bachelor degree or higher in + field_ x_1 + , + field_ x_2 + , + ... + or + field_x_n	• DEGREE: bachelor, FIELD: field_ x_1

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
		+ … + • DEGREE: bachelor, FIELD: field_x_n • DEGREE: master, FIELD: field_ x_1 + … + • DEGREE: master, FIELD: field_x_n • DEGREE: doctor, FIELD: field_ x_1 + … + • DEGREE: doctor, FIELD: field_x_n
63	fresh graduates / degree in + field_ x_1 + , + field_ x_2 + , + … + or + field_x_n	• DEGREE: bachelor, FIELD: field_ x_1 + … + • DEGREE: bachelor, FIELD:

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
		field_x_n • DEGREE: master, FIELD: field_ x_1 + ... + DEGREE: master, FIELD: field_x_n • DEGREE: doctor, FIELD: field_ x_1 + ... + DEGREE: doctor, FIELD: field_x_n
64	we are also open to candidates with degree in + field_ x_1 + , + field_ x_2 + , + ... + , + or + field_x_n	• DEGREE: bachelor, FIELD: field_ x_1 + ... + DEGREE: bachelor, FIELD: field_x_n • DEGREE: master, FIELD: field_ x_1 + ...

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
		+ ● DEGREE: master, FIELD: field_x_n ● DEGREE: doctor, FIELD: field_ x_1 + ... + ● DEGREE: doctor, FIELD: field_x_n
65	we are also open to candidates with degree + in + field_ x_1 + , + field_ x_2 + , + ... + or + field_x_n	● DEGREE: bachelor, FIELD: field_ x_1 + ... + ● DEGREE: bachelor, FIELD: field_x_n ● DEGREE: master, FIELD: field_ x_1 + ... + ● DEGREE: master, FIELD: field_x_n ● DEGREE: doctor, FIELD: field_ x_1

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
		+ … + ● DEGREE: doctor, FIELD: field_x_n
66	bachelor or master degree, preferably in + field_ x_1 + , + field_ x_2 + , + … + , + or + field_x_n	● DEGREE: bachelor, FIELD: field_ x_1 + … + ● DEGREE: bachelor, FIELD: field_x_n ● DEGREE: master, FIELD: field_ x_1 + … + ● DEGREE: master, FIELD: field_x_n
67	graduated with bachelor degree in + field_ x_1 + , + field_ x_2 + , + … + or + field_x_n	● DEGREE: bachelor, FIELD: field_ x_1 + … + ● DEGREE:

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
		bachelor, FIELD: field_x_n
68	graduated with bachelor degree in + field_ x_1 + , + field_ x_2 + , + ... + or + field_x_n	● DEGREE: bachelor, FIELD: field_ x_1 + ... + ● DEGREE: bachelor, FIELD: field_x_n
69	university degree (master degree is preferable) in + field_ x_1 + , + field_ x_2 + , + ... + or + field_x_n	● DEGREE: bachelor, FIELD: field_ x_1 + ... + ● DEGREE: bachelor, FIELD: field_x_n ● DEGREE: master, FIELD: field_ x_1 + ... + ● DEGREE: master, FIELD: field_x_n
70	master or phd in + field_ x_1 + , + field_ x_2 +	● DEGREE: master,

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
	, + ... + or + field_x_n	FIELD: field_ x_1 + ... + • DEGREE: master, FIELD: field_x_n • DEGREE: doctor, FIELD: field_ x_1 + ... + • DEGREE: doctor, FIELD: field_x_n
71	master + / + doctoral degree in + field_ x_1 + , + field_ x_2 + , + ... + , + field_x_n	• DEGREE: master, FIELD: field_ x_1 + ... + • DEGREE: master, FIELD: field_x_n • DEGREE: doctor, FIELD: field_ x_1 + ... + • DEGREE: doctor, FIELD: field_x_n
72	master + / + doctoral degree in + field_ x_1 +	• DEGREE: master,

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
	, + field_ x_2 + , + ... + or + field_x_n	FIELD: field_ x_1 + ... + • DEGREE: master, FIELD: field_x_n • DEGREE: doctor, FIELD: field_ x_1 + ... + • DEGREE: doctor, FIELD: field_x_n
73	master degree or equivalent in + field_ x_1 + , + field_ x_2 + , + ... + or + field_x_n	• DEGREE: master, FIELD: field_ x_1 + ... + • DEGREE: master, FIELD: field_x_n
74	university degree or comparable higher education in + field_ x_1 + , + field_ x_2 + , + ... + or + field_x_n	• DEGREE: bachelor, FIELD: field_ x_1 + ... + • DEGREE: bachelor, FIELD:

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
		field_x_n • DEGREE: master, FIELD: field_ x_1 + ... + DEGREE: master, FIELD: field_x_n • DEGREE: doctor, FIELD: field_ x_1 + ... + DEGREE: doctor, FIELD: field_x_n
75	graduated bachelor's degree in + field_ x_1 + / + field_ x_2 + / + ... + or + field_x_n	• DEGREE: bachelor, FIELD: field_ x_1 + ... + DEGREE: bachelor, FIELD: field_x_n • DEGREE: master, FIELD: field_ x_1 + ...

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
		+ • DEGREE: master, FIELD: field_x_n
76	diploma or bachelor degree in + field_ x_1 + , + field_ x_2 + , + ... + , + or + field_x_n	• DEGREE: diploma, FIELD: field_ x_1 + ... + • DEGREE: diploma, FIELD: field_x_n • DEGREE: bachelor, FIELD: field_ x_1 + ... + • DEGREE: bachelor, FIELD: field_x_n
77	degree in field_ x_1 + or + field_ x_2 + , + ... + , + or + field_x_n	• DEGREE: bachelor, FIELD: field_ x_1 + ... + • DEGREE:

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
		bachelor, FIELD: field_x_n • DEGREE: master, FIELD: field_ x_1 + ... + • DEGREE: master, FIELD: field_x_n • DEGREE: doctor, FIELD: field_ x_1 + ... + • DEGREE: doctor, FIELD: field_x_n
78	bachelor degree + (+ in + field_ x_1 + , + field_ x_2 + , + ... + , + or + field_x_n +)	• DEGREE: bachelor, FIELD: field_ x_1 + ... + • DEGREE: bachelor, FIELD: field_x_n • DEGREE: master, FIELD: field_ x_1 +

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
		... + ● DEGREE: master, FIELD: field_x_n
79	degree in + field_ x_1 + , + field_ x_2 + , + ... + field_x_n + or + another	● DEGREE: bachelor, FIELD: field_ x_1 + ... + ● DEGREE: bachelor, FIELD: field_x_n ● DEGREE: master, FIELD: field_ x_1 + ... + ● DEGREE: master, FIELD: field_x_n ● DEGREE: doctor, FIELD: field_ x_1 + ... + ● DEGREE: doctor, FIELD: field_x_n
80	bachelor degree or higher degree in + field_	● DEGREE:

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
	x_1 +, + ... + , + field_x_n + fields	bachelor, FIELD: field_ x_1 + ... + • DEGREE: bachelor, FIELD: field_x_n • DEGREE: master, FIELD: field_ x_1 + ... + • DEGREE: master, FIELD: field_x_n • DEGREE: doctor, FIELD: field_ x_1 + ... + • DEGREE: doctor, FIELD: field_x_n
81	graduated with a bachelor degree onward in + the + field + of + field_x1 + , + field_ x_2 + , + ... + , + or + field_x_n	• DEGREE: bachelor, FIELD: field_ x_1 + ... +

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
		• DEGREE: bachelor, FIELD: field_x_n • DEGREE: master, FIELD: field_ x_1 + ... + • DEGREE: master, FIELD: field_x_n • DEGREE: doctor, FIELD: field_ x_1 + ... + • DEGREE: doctor, FIELD: field_x_n
82	bachelor degree in + field_ x_1 + , + field_ x_2 + , + ... + , + or + any + field_x_n	• DEGREE: bachelor, FIELD: field_ x_1 + ... + • DEGREE: bachelor, FIELD: field_x_n
83	bachelor degree in + field_ x_1 + , + field_ x_2 + , + ... + , + or + field_x_n + is + preferable	• DEGREE: bachelor, FIELD:

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
		field_ x_1 + ... + ● DEGREE: bachelor, FIELD: field_x_n
84	bachelor + / + master + degree in + field_ x_1 + , + field_ x_2 + , + ... + , + or + any + field_x_n	● DEGREE: bachelor, FIELD: field_ x_1 + ... + ● DEGREE: bachelor, FIELD: field_x_n ● DEGREE: master, FIELD: field_ x_1 + ... + ● DEGREE: master, FIELD: field_x_n
85	candidate must possess at least diploma/bachelor degree in field_ x_1 + or + field_ x_2 + , + ... + , + or + equivalent	● DEGREE: diploma, FIELD: field_ x_1 + ... + ● DEGREE: diploma, FIELD: field_x_n

ลำดับ	กฎการจำแนกประเภทรูปแบบประโยค	ส่วนโครงสร้างเก็บคำสำคัญจากประโยค
		• DEGREE: bachelor, FIELD: field_ x_1 + ... + • DEGREE: bachelor, FIELD: field_ x_n • DEGREE: master, FIELD: field_ x_1 + ... + • DEGREE: master, FIELD: field_ x_n • DEGREE: doctor, FIELD: field_ x_1 + ... + • DEGREE: doctor, FIELD: field_ x_n
86	graduated with bachelor degree in + field_ x_1 + , + field_ x_2 + , + ... + , + or + any + field_ x_n	• DEGREE: bachelor, FIELD: field_ x_1 + ... + • DEGREE: bachelor, FIELD: field_ x_n

ตารางที่ 23 โครงสร้างการจัดเก็บคำสำคัญจากข้อมูลประเภทของประสบการณ์ที่ต้องการ

ลำดับ	ประโยค	กฎ
1	at least + number_x + year + in + obj	● NUMBER1: NULL, SIGN: >, NUMBER2: number_x, EXPERIENCE_LEVEL1: NULL, EXPERIENCE_LEVEL2: NULL, VERB: NULL,

ลำดับ	ประโยค	กฎ
		OBJ: obj
2	at least + number_x + year + exp_level2 + in + obj	● NUMBER1: NULL, SIGN: >, NUMBER2: number_x, EXPERIENCE_LEVEL1: NULL, EXPERIENCE_LEVEL2: exp_level2, VERB: NULL, OBJ: obj
3	at least + number_x + year + exp_level2 + as + obj	● NUMBER1: NULL, SIGN: >, NUMBER2: number_x, EXPERIENCE_LEVEL1: NULL, EXPERIENCE_LEVEL2: exp_level2, VERB: NULL, OBJ: obj
4	at least + number_x + year + of + exp_level2 + in + obj	● NUMBER1: NULL, SIGN: >, NUMBER2: number_x, EXPERIENCE_LEVEL1: NULL, EXPERIENCE_LEVEL2: exp_level2, VERB: NULL, OBJ: obj
5	at least + number_x + year + of + exp_level1 + exp_level2 + in + obj	● NUMBER1: NULL, SIGN: >, NUMBER2: number_x, EXPERIENCE_LEVEL1:

ลำดับ	ประโยค	กฎ
		exp_level1, EXPERIENCE_LEVEL2: exp_level2, VERB: NULL, OBJ: obj
6	at least + number_x + year + of + relevant + exp_level2 + obj	● NUMBER1: NULL, SIGN: >, NUMBER2: number_x, EXPERIENCE_LEVEL1: NULL, EXPERIENCE_LEVEL2: exp_level2, VERB: NULL, OBJ: obj
7	at least + number_x + - + number_y + year + exp_level2 + in + obj	● NUMBER1: number_x, SIGN: -, NUMBER2: number_y, EXPERIENCE_LEVEL1: NULL, EXPERIENCE_LEVEL2: exp_level2, VERB: NULL, OBJ: obj
8	number_x + - + number_y + year + of + exp_level2 + with + obj	● NUMBER1: number_x, SIGN: -, NUMBER2: number_y, EXPERIENCE_LEVEL1: NULL, EXPERIENCE_LEVEL2: exp_level2, VERB: NULL, OBJ: obj
9	number_x + - + number_y + year + of +	● NUMBER1: number_x,

ลำดับ	ประโยค	กฎ
	exp_level2 + in + obj	SIGN: -, NUMBER2: number_y, EXPERIENCE_LEVEL1: NULL, EXPERIENCE_LEVEL2: exp_level2, VERB: NULL, OBJ: obj
10	number_x + - + number_y + year + of + exp_level2 + as + obj	• NUMBER1: number_x, SIGN: -, NUMBER2: number_y, EXPERIENCE_LEVEL1: NULL, EXPERIENCE_LEVEL2: exp_level2, VERB: NULL, OBJ: obj
11	number_x + - + number_y + + + year + of + exp_level2 + in + obj1 + or + obj2	• NUMBER1: number_x, SIGN: -, NUMBER2: number_y, EXPERIENCE_LEVEL1: NULL, EXPERIENCE_LEVEL2: exp_level2, VERB: NULL, OBJ: obj1 • NUMBER1: number_x, SIGN: -, NUMBER2: number_y, EXPERIENCE_LEVEL1: NULL, EXPERIENCE_LEVEL2:

ลำดับ	ประโยค	กฎ
		exp_level2, VERB: NULL, OBJ: obj2
12	number_x + year + of + exp_level1 + exp_level2 + in + obj	● NUMBER1: NULL, SIGN: >, NUMBER2: number_x, EXPERIENCE_LEVEL1: exp_level1, EXPERIENCE_LEVEL2: exp_level2, VERB: NULL, OBJ: obj
13	over + number_x + year + of + obj	● NUMBER1: NULL, SIGN: >, NUMBER2: number_x, EXPERIENCE_LEVEL1: NULL, EXPERIENCE_LEVEL2: NULL, VERB: NULL, OBJ: obj
14	over + number_x + year + of + exp_level2 + in + obj	● NUMBER1: NULL, SIGN: >, NUMBER2: number_x, EXPERIENCE_LEVEL1: NULL, EXPERIENCE_LEVEL2: exp_level2, VERB: NULL, OBJ: obj
15	over + number_x + year + of + exp_level2 + as + obj1 + or + obj2	● NUMBER1: NULL, SIGN: >, NUMBER2:

ลำดับ	ประโยค	กฎ
		number_x, EXPERIENCE_LEVEL1: NULL, EXPERIENCE_LEVEL2: exp_level2, VERB: NULL, OBJ: obj
16	minimum + number_x + - + number_y + year + of + exp_level1 + exp_level2 + in + obj1 + and + obj2	● NUMBER1: NULL, SIGN: >, NUMBER2: number_x, EXPERIENCE_LEVEL1: exp_level1, EXPERIENCE_LEVEL2: exp_level2, VERB: NULL, OBJ: obj1 NUMBER1: NULL, SIGN: >, NUMBER2: number_x, EXPERIENCE_LEVEL1: exp_level1, EXPERIENCE_LEVEL2: exp_level2, VERB: NULL, OBJ: obj2
17	exp_level2 + more than + number_x + - + number_y + year + of + exp_level2 + in + obj1 + and + obj2 + and + objn	● NUMBER1: NULL, SIGN: >, NUMBER2: number_x, EXPERIENCE_LEVEL1: NULL, EXPERIENCE_LEVEL2: exp_level2, VERB:

ลำดับ	ประโยค	กฎ
		NULL, OBJ: obj1 ● NUMBER1: NULL, SIGN: >, NUMBER2: number_x, EXPERIENCE_LEVEL1: NULL, EXPERIENCE_LEVEL2: exp_level2, VERB: NULL, OBJ: obj2 ● NUMBER1: NULL, SIGN: >, NUMBER2: number_x, EXPERIENCE_LEVEL1: NULL, EXPERIENCE_LEVEL2: exp_level2, VERB: NULL, OBJ: objn
18	at least + number_x + year + of + verb + exp_level2 + in + obj	● NUMBER1: NULL, SIGN: >, NUMBER2: number_x, EXPERIENCE_LEVEL1: NULL, EXPERIENCE_LEVEL2: exp_level2, VERB: verb, OBJ: obj
19	number_x + year + of + exp_level2 + verb + obj	● NUMBER1: NULL, SIGN: >, NUMBER2: number_x, EXPERIENCE_LEVEL1:

ลำดับ	ประโยค	กฎ
		NULL, EXPERIENCE_LEVEL2: exp_level2, VERB: verb, OBJ: obj
20	number_x + + + year + of + verb + exp_level2 + in + obj	• NUMBER1: NULL, SIGN: >, NUMBER2: number_x, EXPERIENCE_LEVEL1: NULL, EXPERIENCE_LEVEL2: exp_level2, VERB: verb, OBJ: obj
21	number_x + year + of + exp_level2 + verb + in + obj1 + and + obj2 + and + objn	• NUMBER1: NULL, SIGN: >, NUMBER2: number_x, EXPERIENCE_LEVEL1: NULL, EXPERIENCE_LEVEL2: exp_level2, VERB: verb, OBJ: obj1 • NUMBER1: NULL, SIGN: >, NUMBER2: number_x, EXPERIENCE_LEVEL1: NULL, EXPERIENCE_LEVEL2: exp_level2, VERB: verb, OBJ: obj2 • NUMBER1: NULL, SIGN:

ลำดับ	ประโยค	กฎ
		>, NUMBER2: number_x, EXPERIENCE_LEVEL1: NULL, EXPERIENCE_LEVEL2: exp_level2, VERB: verb, OBJ: objn
22	number_x + + + year + of + exp_level2 + verb + in + obj + of + exp_level1	● NUMBER1: NULL, SIGN: >, NUMBER2: number_x, EXPERIENCE_LEVEL1: exp_level1, EXPERIENCE_LEVEL2: exp_level2, VERB: NULL, OBJ: obj

ตารางที่ 24 โครงสร้างการจัดเก็บคำสำคัญจากข้อมูลประเภทของทักษะที่จำเป็น

ลำดับ	ประโยค	กฎ
1	verb + obj + skills	● SKILL_LEVEL1: NULL, SKILL_LEVEL2: NULL, VERB: verb, OBJ: obj
2	skill_level$_1$ + with + obj	● SKILL_LEVEL1: skill_level$_1$, SKILL_LEVEL2: NULL,

ลำดับ	ประโยค	กฎ
		VERB: NULL, OBJ: obj
3	$skill_level_1$ + in + obj_1 + and + obj_2	● SKILL_LEVEL1: $skill_level_1$, SKILL_LEVEL2: NULL, VERB: NULL, OBJ: obj_1 SKILL_LEVEL1: $skill_level_1$, SKILL_LEVEL2: NULL, VERB: NULL, OBJ: obj_2
4	$skill_level_1$ + in + obj	● SKILL_LEVEL1: $skill_level_1$, SKILL_LEVEL2: NULL, VERB: NULL, OBJ: obj
5	$skill_level_1$ + obj + skills	● SKILL_LEVEL1: $skill_level_1$, SKILL_LEVEL2: NULL, VERB: NULL, OBJ: obj
6	$skill_level_2$ + of + obj_1 + and + obj_2 + and + ... + obj_n	● SKILL_LEVEL1: NULL, SKILL_LEVEL2: skill_level2, VERB: NULL, OBJ: obj_1 SKILL_LEVEL1: NULL, SKILL_LEVEL2: $skill_level_2$, VERB: NULL, OBJ: obj_2 SKILL_LEVEL1: NULL, SKILL_LEVEL2: $skill_level_2$, VERB: NULL, OBJ: obj_n

ลำดับ	ประโยค	กฎ
7	$skill_level_1$ + $skill_level_2$ + of + obj_1 + and + obj_2	● SKILL_LEVEL1: $skill_level_1$, SKILL_LEVEL2: $skill_level_2$, VERB: NULL, OBJ: obj_1 SKILL_LEVEL1: $skill_level_1$, SKILL_LEVEL2: $skill_level_2$, VERB: NULL, OBJ: obj_2
8	$skill_level_2$ + $verb_1$ + obj_1 + and +$verb_2$ + obj_2	● SKILL_LEVEL1: NULL, SKILL_LEVEL2: $skill_level_2$, VERB: $verb_1$, OBJ: obj_1 SKILL_LEVEL1: NULL, SKILL_LEVEL2: skill_level2, VERB: $verb_2$, OBJ: obj_2
9	$skill_level_1$ + obj_1 + , + obj_2 + and + obj_3 + skills	● SKILL_LEVEL1: $skill_level_1$, SKILL_LEVEL2: NULL, VERB: NULL, OBJ: obj_1 SKILL_LEVEL1: $skill_level_1$, SKILL_LEVEL2: NULL, VERB: NULL, OBJ: obj_2 SKILL_LEVEL1: $skill_level_1$, SKILL_LEVEL2: NULL,

ลำดับ	ประโยค	กฎ
		VERB: NULL, OBJ: obj_3
10	$skill_level_1$ + in + obj_1 + (+ obj_2 + , + obj_3 +)	● SKILL_LEVEL1: $skill_level_1$, SKILL_LEVEL2: NULL, VERB: NULL, OBJ: obj_1 + obj_2 SKILL_LEVEL1: $skill_level_1$, SKILL_LEVEL2: NULL, VERB: NULL, OBJ: obj_1 + obj_3
11	$skill_level_1$ + of + verb + obj	● SKILL_LEVEL1: skill_level1, SKILL_LEVEL2: NULL, VERB: verb, OBJ: obj
12	$skill_level_1$ + in + obj_1 + and + obj_2	● SKILL_LEVEL1: $skill_level_1$, SKILL_LEVEL2: NULL, VERB: NULL, OBJ: obj_1 SKILL_LEVEL1: $skill_level_1$, SKILL_LEVEL2: NULL, VERB: NULL, OBJ: obj_2
13	$skill_level_1$ + $skill_level_2$ + obj_1 + and + obj_2	● SKILL_LEVEL1: $skill_level_1$, SKILL_LEVEL2: $skill_level_2$, VERB: NULL, OBJ: obj_1 SKILL_LEVEL1:

ลำดับ	ประโยค	กฎ
		skill_level$_1$, SKILL_LEVEL2: skill_level$_2$, VERB: NULL, OBJ: obj$_2$
14	skill_level$_1$ + skill_level$_2$ + of + obj$_1$ + and + obj$_2$	● SKILL_LEVEL1: skill_level$_1$, SKILL_LEVEL2: skill_level$_2$, VERB: NULL, OBJ: obj$_1$ SKILL_LEVEL1: skill_level$_1$, SKILL_LEVEL2: skill_level$_2$, VERB: NULL, OBJ: obj$_2$
15	skill_level1 + verb + on + obj	● SKILL_LEVEL1: skill_level$_1$, SKILL_LEVEL2: NULL, VERB: verb, OBJ: obj
16	skill_level$_1$ + skill_level$_2$ + in + obj$_1$ + and + obj$_2$	● SKILL_LEVEL1: skill_level$_1$, SKILL_LEVEL2: skill_level$_2$, VERB: NULL, OBJ: obj$_1$ SKILL_LEVEL1: skill_level$_1$, SKILL_LEVEL2: skill_level$_2$, VERB: NULL, OBJ: obj$_2$
17	skill_level$_2$ + of + verb + obj$_1$ + and + obj$_2$	● SKILL_LEVEL1: NULL,

ลำดับ	ประโยค	กฎ
		SKILL_LEVEL2: skill_level$_2$, VERB: verb, OBJ: obj$_1$ SKILL_LEVEL1: NULL, SKILL_LEVEL2: skill_level$_2$, VERB: verb, OBJ: obj$_2$

ภาคผนวก ค
เอกสารรับรองผลการพิจารณาจริยธรรมการวิจัยในมนุษย์

บันทึกข้อความ

ส่วนงาน สำนักงานอธิการบดี กองบริหารการวิจัยและนวัตกรรม โทร. ๒๖๒๐
ที่ อว ๘๑๐๐/๐๑๕๖๘ **วันที่** ๙ กุมภาพันธ์ พ.ศ. ๒๕๖๔
เรื่อง ขอแจ้งรับรองโครงการวิจัยที่ส่งมาขอรับการพิจารณาจริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยบูรพา

เรียน นางสาวนรัชกร ฤกษ์ศิริ (นิสิตระดับบัณฑิตศึกษา คณะวิทยาการสารสนเทศ)

ตามที่ท่าน ได้ส่งเอกสารโครงการวิจัยเพื่อขอรับการพิจารณาจริยธรรมการวิจัยในมนุษย์ ในหัวข้อโครงการวิจัย เรื่อง **การวิเคราะห์และสรุปความรับผิดชอบและข้อกำหนดจากประกาศงานด้านเทคโนโลยีสารสนเทศ** นั้น

บัดนี้ คณะกรรมการพิจารณาจริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยบูรพา ได้พิจารณาตามวิธีดำเนินการมาตรฐาน (Standard Operating Procedures, SOP) ฉบับที่ ๑.๑ พ.ศ. ๒๕๖๒ ที่ได้ประกาศใช้เมื่อวันที่ ๑ ตุลาคม พ.ศ. ๒๕๖๒ แล้วว่า โครงการวิจัยดังกล่าวไม่ได้ทำการศึกษาวิจัยในมนุษย์ จึงเห็นสมควรให้ดำเนินการวิจัยได้

จึงเรียนมาเพื่อโปรดทราบ

(ผู้ช่วยศาสตราจารย์ แพทย์หญิงรมร แย้มประทุม)
ประธานคณะกรรมการพิจารณาจริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยบูรพา
สำหรับโครงการวิจัย ระดับบัณฑิตศึกษา และระดับปริญญาตรี
ชุดที่ ๓ (กลุ่มคลินิก/ วิทยาศาสตร์สุขภาพ/ วิทยาศาสตร์และเทคโนโลยี)

ภาคผนวก ง
เอกสารเผยแพร่ผลงานวิจัย

ICSEC 2023
THE 27TH INTERNATIONAL COMPUTER SCIENCE AND ENGINEERING CONFERENCE
Koh Samui , Surat Thani, Thailand
13 – 15 September, 2023
The International Computer Science and Engineering Conference (ICSEC) is the premier forum for the presentation of researchers, practitioners, and educators to present and discuss the most recent innovations, research, experience, trends and concerns in the field of Computer Science, Computer Engineering, Software Engineering, Information Technology and Emerging Technology.
ICSEC 2023 will be held over two days with special sessions and presentations delivered by researchers from the international communities, including presentations from well-known keynote speakers. All accepted papers will be submitted for inclusion into IEEE Xplore.
Papers related to this conference theme including theories, methodologies, and emerging applications are encouraged. Contributions covering theoretical developments and practical applications, including but not limited to the following technical areas, are invited:
Artificial Intelligence
Bioinformatics
Cloud Computing
Computer and Internet Application
Computer Architecture
Computer Graphics
Computer Networks
Computer Vision and Image Processing
Cryptography and Security
Data and Knowledge Management
Data Mining
Digital Forensics
Digital Signal Processing
Distributed Computing and Grid Computing
Embedded Systems
Evolutionary Computing
Fuzzy and Neural Network Systems
High Performance Computing
Human Machine Interface
Information Technology Management
Intelligent Systems
Knowledge Discovery
Knowledge Management
Machine Intelligence and Applications
Medical Image Processing
Mobile Computing
Natural Language Processing
Pattern Recognition
Programming Language
Quantum Computing
Robotics and Automation
Software Engineering
Semantic Web Ontology
Simulation and Modeling
Wireless Networks and Communications
Computer in Education
COMMITTEES / ICSEC2023
Steering Committee
General Chair
Technical Program Committee
Publication Chair
General Secretary
PAPER SUBMISSION
Prospective authors are invited to submit original full papers WITHOUT authors' names and affiliations, in English, of 4 – 6 pages in standard IEEE two-column format only, reporting their original work and results, applications, and/or implementation in one or more of the listed areas.
Papers must be submitted online only through the submission system of the conference website.
At least one author of each accepted paper MUST register and present the paper at the conference in order that the paper is to be included in the program. The program will also be submitted for inclusion in the IEEE Xplore. All accepted papers will be submitted for inclusion into IEEE Xplore.
The author should use IEEE manuscript templates. (Microsoft Word: A4)
CONFERENCE SCHEDULE / Important dates
Deadline for Regular Session Submission June 30, 2023
Notification of Acceptance July 31, 2023
Camera Ready Submission August 7, 2023
Regular Registration August 1-7, 2023
Workshop Day September 13,2023
Conference Day September 14-15, 2023
Research Networking Event September 15, 2023 17:00-22:00
CONFERENCE LOCATION
INTERNATIONAL SCHOOL OF TOURISM, SURATTHANI RAJABHAT UNIVERSITY, KOH SAMUI , SURAT THANI, THAILAND
ekkarat@ieee.org

CID-CJA: Co-occurrence Information Discovery in Computer-related Job Advertisements

Naratchakorn Roeksiri and Komate Amphawan
Computational Innovation Laboratory
Faculty of Informatics, Burapha University
Chonburi, Thailand
naratchakorn.r@gmail.com, komate@gmail.com

***Abstract*—The rapid advancements in technology have brought about significant disruptions in the job market for computer-related fields, necessitating a comprehensive understanding of emerging trends and evolving demands. This research focuses on the analysis of computer-related job advertisements to uncover co-occurrence information that employers seek within the context of disruptive technology and evolving business models. To address this need, we propose *CID-CJA* (Co-occurrence Information Discovery in Computer-related Job Advertisements), a novel method for extracting valuable insights from job advertisements. *CID-CJA* empowers job seekers to identify the knowledge and skills that are relevant in the current landscape, while also assisting universities in aligning their curricula with the dynamic needs of the industry. The flexibility of *CID-CJA* enables users to conduct detailed drill-down analyses on specific elements of information. The effectiveness of *CID-CJA* is validated through experimental studies conducted with real job advertisements, which demonstrate its ability to reveal valuable information that is beneficial for computer-related professionals.**

***Index Terms*—computer-related job advertisements, job advertisement analysis, co-occurrence information discovery, drill-down analysis**

I. Introduction

The computer-related field encompasses diverse disciplines such as computer science, information technology, software engineering, cybersecurity, and data science ([1]–[3]). The increasing influence of technology in our lives has created a growing demand for skilled professionals in these areas. However, job seekers in the computer-related field face challenges in keeping up with evolving skill requirements and industry trends. Furthermore, organizations within this field need to adapt to rapid growth and changes to attract and retain top talent.

Analyzing job advertisements in the computer-related field ([4]–[6]) provides valuable insights into industry trends, skill requirements, and employer expectations. Job seekers can benefit from such analyses by understanding the specific skills, qualifications, and competencies in high demand. This knowledge allows them to align their skill set and educational pursuits accordingly, improving their chances of securing desirable positions in this competitive field. Additionally, job seekers can identify emerging trends and adapt their career strategies to stay ahead.

Employers can leverage job advertisement analysis to attract candidates with desired skill sets ([7]). It enables them to create accurate job descriptions that reflect evolving industry needs. By understanding market trends, employers can align their recruitment strategies with current demands, resulting in a more effective and efficient hiring process. Analyzing job advertisements also helps employers benchmark their requirements against industry standards and make informed decisions about salary, benefits, and employee development.

Academic institutes and educational organizations can benefit from job advertisement analysis by tailoring curricula and training programs to meet industry demands. Bridging the gap between academic education and industry needs ensures graduates are well-prepared for the job market. Moreover, analyzing job advertisements provides insights into emerging research areas, allowing academic institutions to align their focus with industry demands.

Policymakers can leverage job advertisement analysis to inform policies related to education, training, and workforce development. Understanding changing skill requirements helps policymakers support the growth of the computer-related field, promote employment opportunities, and foster innovation, contributing to overall economic development and competitiveness.

Several approaches have been developed to analyze job advertisements. For example, Sibarani et al. [8] proposed an Ontology-based Information Extraction (OBIE) method guided by a specialized vocabulary to extract skills requested by job vacancies. Radovilsky et al. [9] focused on characterizing knowledge domains and skills in Business Data Analytics (BDA) and Data Science (DS) professions. Verma et al. [10]) conducted a comprehensive investigation into the skill requirements for positions in artificial intelligence (AI) and machine learning (ML).

While these studies contribute valuable insights, they often focus on identifying essential skill sets required by employers. However, job seekers often desire a comprehensive understanding of job advertisements, including hidden details. For instance, job seekers planning to change jobs in the near future seek an overall understanding of constraints set by employers

979–8–3503–4210–9/23/$31.00 ©2023 IEEE

to evaluate compatibility with long-term career aspirations. By considering the larger context, job seekers can make informed decisions about exploring specific responsibilities, required skill sets, and other constraints outlined in job advertisements.

To address these issues, we present CID-CJA (Co-occurrence Information Discovery in Computer-related Job Advertisements), a novel method for extracting valuable insights from job advertisements. CID-CJA collects computer-related job advertisements, extracts essential information, and organizes it into a flexible structure. This structure allows users to perform deep and comprehensive searches, explore a wide variety of details, and gain in-depth insights into job requirements and conditions. Additionally, CID-CJA enables comparisons between positions and tracks changes in required skills and conditions over time.

To evaluate CID-CJA, we gathered job advertisements for various computer-related positions and demonstrated its effectiveness in discovering co-occurrence information. The insights provided by CID-CJA are valuable for job seekers and universities seeking to validate their curriculum. By utilizing CID-CJA, job seekers can make more informed decisions about job opportunities, and academic institutions can align their programs with industry demands.

II. The Proposed CID-CJA

In this section, we provide a detailed description of the proposed *CID-CJA* (Co-occurrence Information Discovery in Computer-related Job Advertisements) framework. The framework's overall structure, depicted in Figure 1, comprises five distinct steps, which are outlined as follows.

- *Prerequisites Preparation and Input Gathering - Prior to delving into the analysis of job advertisements, we lay a solid foundation by preparing and compiling a corpus of computer terminologies and their corresponding synonyms. We extend our gratitude to Kamlangpuech et al. [11], [12] for their invaluable contribution in organizing this corpus. Moreover, we curate corpora encompassing abbreviations and verb synonyms. Additionally, we establish linguistic rules (credited to Chaisoongnoen et al. [13], [14]) as well as classification rules, enabling us to effectively extract vital information and classify various content types, such as responsibilities, skills, age requirements, educational qualifications, and more. Lastly, we employ the web scraping technique to gather and compile job advertisements from prominent job search websites.*
  As illustrated in Figure 2, a job advertisement sourced from a prominent job search website comprises three primary components: i) the job title, ii) the job responsibilities (spanning multiple lines), and iii) the job qualifications (also spanning multiple lines). As part of the collection process, each job is assigned a distinct identification number, while each line of content is further categorized as either "title," "responsibility," or "qualification" based on its respective nature.
- *Job advertisement grouping – In this task, we embark on a meticulous journey of analyzing and processing the job titles in each advertisement, employing refined text processing techniques to eliminate extraneous information. Our primary objective is to create cohesive groups of jobs with identical titles. To achieve this, we meticulously examine each group of job advertisements, merging those with remarkably similar titles. Leveraging cutting-edge methodologies such as Word2Vec, GloVe (Global Vectors for Word Representation), and cosine similarity, we unlock the power to accurately categorize and label each job advertisement based on the prevailing title that closely aligns with its essence.*
  As depicted in Figure 3, we illustrate the process using the job title "Data Engineer // BTS Asoke // UP TO 60K," which initially contains an abundance of unnecessary information. Through meticulous refinement, we transform it into the succinct and informative title "data engineer" (similarly, "Data Engineer" is adjusted to "data engineer"). Subsequently, we consolidate all job postings bearing the title "data engineer" into a cohesive group. This group is then merged with the one encompassing "data engineer specialist," owing to their significant similarity. Finally, we harmonize the titles of all jobs within this consolidated group, unifying them under the label of "data engineer."
- *Text processing – To extract and analyze vital information from job advertisements, the initial step involves comprehensive text processing. This crucial task encompasses the implementation of various fundamental techniques. Firstly, we employ word error correction and leverage a pre-collected abbreviation corpus to resolve abbreviations. Next, we convert the text to lowercase, perform sentence tokenization, remove punctuation, conduct word tokenization, eliminate stopwords, and apply word stemming and lemmatization techniques. Additionally, part-of-speech tagging is performed on each job advertisement to further enhance the analysis.*
  Subsequently, we shift our focus to each sentence within the job advertisement and input n-grams of words from each sentence into predefined linguistic rules, enabling the identification of "keywords" specific to the context. Upon completing this step, we meticulously examine each word in the job advertisement and replace it with its synonym, selecting the synonym with the highest frequency of occurrence. This process ensures optimal clarity and consistency throughout the analysis.
  Let's delve into the third line of the job responsibilities (as depicted in Figure 2), which states, "Assess the needs, objectives of the business and recommend technical solutions." Through the text processing phase, this context is collected and represented as a list of word and part-of-speech tag pairs: [('assess', 'VB'), ('needs', 'NN'), ('and', 'CC'), ('objectives', 'NN'), ('of', 'IN'), ('business', 'NN'), ('and', 'CC'), ('recommend', 'VB'),

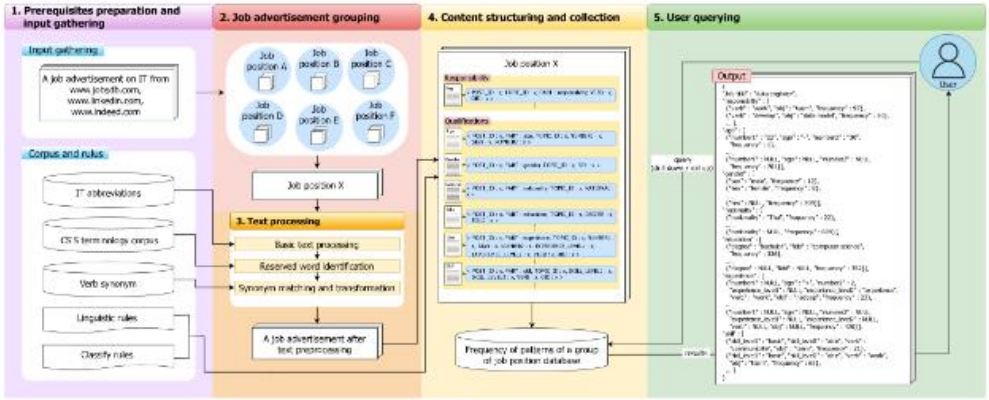


Fig. 1. Framework of the proposed *CID-CJA*

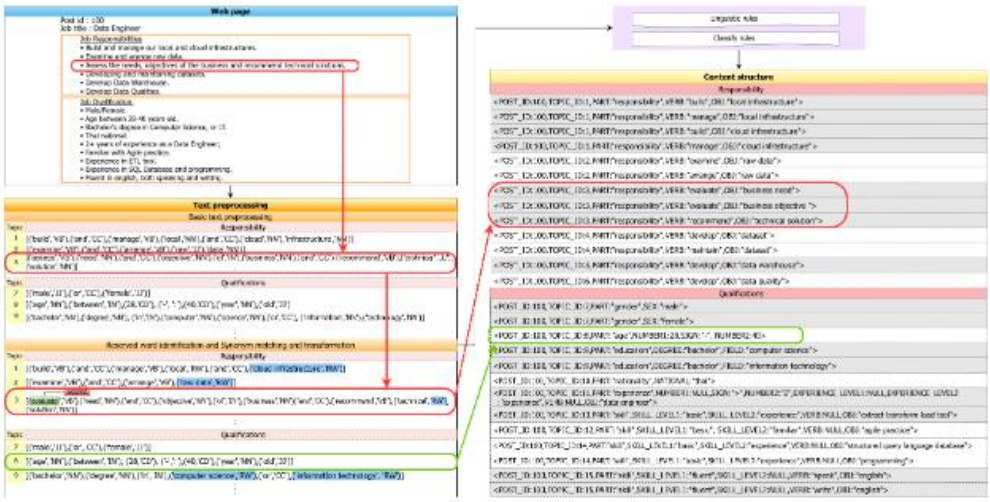


Fig. 2. Example of Job Advertisement and Text Processing

('technical', 'JJ'), ('solutions', 'NN')].
Subsequently, this context undergoes synonym replacement and linguistic rule application. As a result, ('assess', 'VB') is transformed into ('evaluate', 'VB'), while ('technical', 'JJ') is identified as a "reserved word" based on our predefined linguistic rules. Further iterations of the linguistic rules lead to the emergence of three essential contexts:

- *('evaluate', 'VB') + ('business need', KW')*
- *('evaluate', 'VB') + ('business objective', KW')*
- *('recommend', 'VB') + ('technical solution', KW')*

These refined contexts form the basis for subsequent analysis and classification within the framework.

- *Content structuring and collection – Each previously processed context from the preceding task is carefully considered and labeled with its corresponding type, utilizing our predefined classification rules. Subsequently, these labeled contexts are collected and organized according to the predefined structure.*
For instance, let's examine the job advertisement de-

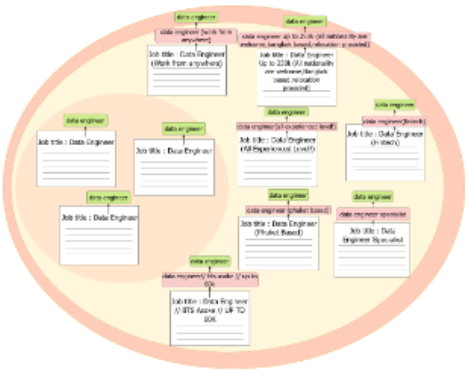


Fig. 3. Example of Job title grouping

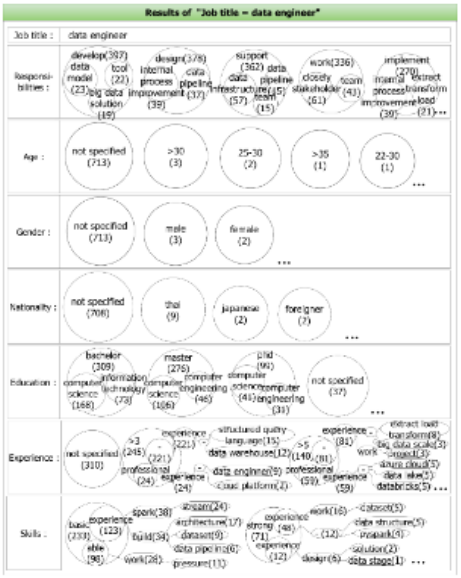


Fig. 4. Example of result under a user-given constraint : "job title = data engineer"

picted in Fig. 2, where there is a mention of the age requirement as "Age between 28 - 40 years old." This context undergoes processing and is transformed into the following list: [('age', 'NN'), ('between', 'IN'), (28, 'CD'), ('-', ':'), (40, 'CD'), ('year', 'NN'), ('old', 'JJ')]. This context is then matched against the rule "RULE → age + between + CD + - + CD + year + old," resulting in the following collected structure:

< "POST ID" : 100, "TOPIC ID" : 8, "PART" : "age", "NUMBER1" : 28, "SIGN" : "-", "NUMBER2" : 40>

With this structured representation, we can infer that in the 100th job advertisement (within our job advertisement portal), on the 8th line, there is a requirement specifying that applicants should be between 28 and 40 years old. These structured representations enable users to easily identify frequently occurring or co-occurring information within job advertisements related to a specific job title, facilitating efficient information retrieval.

- *User querying – This stage, often referred to as "Information Delivery to users," empowers users to perform queries with drill-down and roll-up operations, enabling them to obtain the desired information.*
  For instance, a user might initiate a query to gather an overview of the requirements for the job title "Data Engineer". In response, a comprehensive summary of related information is generated and presented to the user. As exemplified in Fig. 4, the user is provided with an overview of the responsibilities associated with the job title. For instance, it is revealed that there are 23 job advertisements that require applicants to possess the skill of "developing data models", while 39 job advertisements necessitate expertise in "designing internal process improvement", and so on. Furthermore, the user can also observe additional constraints imposed on the age of applicants, such as a single job advertisement specifying a requirement for an experienced applicant over the age of 35. Additionally, gender constraints are present, with three job advertisements exclusively seeking male applicants and two targeting female applicants. Moreover, nationality requirements are highlighted, with nine job advertisements indicating a preference for Thai applicants, two for Japanese applicants, and two for foreigners. Lastly, specific skill requirements are outlined, including the need for "basic experience in building data pipelines" (occurring six times) and a strong proficiency in "pyspark" (occurring four times).
  In addition, users may desire to impose further constraints by generating drill-down queries, such as "job title = data engineer AND age >= 30" (as depicted in Figure 5). This query narrows down the displayed contexts to only show job advertisements related to "data engineer" positions for applicants aged 30 and above. For example, within these filtered contexts, specific re-

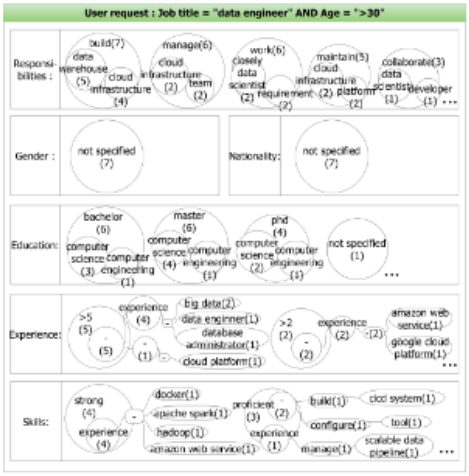


Fig. 5. Example of result under a user-given constraint : "job title = data engineer AND age >= 30"

quirements are highlighted, such as the need to possess skills in "building data warehouses," "managing cloud infrastructure," and "collaborating with data scientists," among others. Additionally, experience of at least 5 years in "big data" and "cloud platforms" is commonly sought after.

On the other hand, users can further focus on essential skills by specifying deeper drill-down queries, such as "job title = data engineer AND age > 35 AND skill = SQL AND skill = ETL". This query narrows down the job advertisements to only consider those labeled as data engineers, with an age requirement of over 35, and where both SQL and ETL skills are explicitly requested. The information from these refined contexts is then summarized and presented to the user.

With the ability to perform drill-down and roll-up operations, users can access a wide variety of information tailored to their specific criteria and interests. This empowers users to efficiently explore job advertisements and extract valuable insights.

III. Experimental & Utilization Studies

In We evaluated the performance of the proposed *CID-CJA* using a dataset of 4,280 randomly collected job advertisements from three popular job search websites: www.jobdb.com (1,905 ads), LinkedIn (1,504 ads), and www.indeed.com (871 ads). The data collection spanned from August 2020 to April 2023.

Three experienced professors in the computer-related field labeled all job advertisements, resulting in 11 distinct groups based on job titles. We conducted three experiments to assess CID-CJA's performance: (i) accuracy of job advertisement grouping, (ii) precision of keyword extraction, and (iii) effectiveness of splitting and structuring essential information. These experiments aimed to evaluate the efficacy of CID-CJA in assisting job seekers and recruiters.

As observed in the table I, job advertisement grouping based on job title (using Word2Vec, GloVe, and cosine similarity) demonstrated accurate results, without any errors. The grouping of similar job titles facilitated efficient comparisons between different groups, enhancing the effectiveness of the approach.

Likewise, the keyword extraction and content structuring also have high accuracy (between 95 - 98%). Thanks to the linguistic rules (provided by [13], [14]) that can help to extract important keywords by considering n-grams of words. Also, we can get benefit from verb synonyms.

Table II illustrates the accuracy of classifying and labeling context using *CID-CJA*. It is evident that *CID-CJA* performs exceptionally well in correctly classifying and labeling context within the "responsibility" part, as this section solely consists of responsibility-related information. Therefore, *CID-CJA* accurately labels all the context within this part as responsibility.

Furthermore, *CID-CJA* demonstrates accurate classification and labeling of context related to "Gender," "Education," and "Experience." This is primarily due to the presence of specific keywords associated with each category. For instance, if a job advertisement includes words such as "male" or "female," *CID-CJA* can directly classify and label it as "Gender." Similarly, terms like"bachelor," "master," "degree," "computer science," or "computer engineering" commonly appear in job advertisements and aid in easy identification as "Education." Moreover, the term"experience" is frequently present in job advertisements, enabling accurate labeling as "Experience." The predefined classification rules contribute to the precise identification of context types.

Lastly, there may be questions regarding whether the information is correctly returned to the user or why the accuracy of user queries was not observed. In response, it is evident that the accuracy is 100% since the frequency of co-occurrence information is determined through exact matching. As users provide conditions that match the pre-structured information, *CID-CJA* can directly return the relevant results, ensuring accurate retrieval of information.

IV. Conclusion

In this paper, we introduce a novel approach called CID-CJA (Co-occurrence Information Discovery in Computer-related Job Advertisements) for uncovering valuable co-occurrence information within computer-related job advertisements. Our proposed method involves extracting crucial details and segmenting the context of each job ad into key components, including: (i) job title, (ii) responsibility, (iii) age, (iv) gender, (v) nationality, (vi) education, (vii) experience, and (viii) skills. This enables users to perform

TABLE I
THE PERFORMANCE ON JOB TITLE GROUPING AND KEYWORD EXTRACTION

Job title	#Job advertisements	Accuracy of grouping(%)	Accuracy of keyword extraction(%)
Data scientist	518	100	95.79
Data engineer	721	100	97.82
Data analyst	354	100	96.50
Cyber security	348	100	91.40
Database administrator	309	100	96.82
Network engineer	262	100	95.61
Programmer	732	100	96.82
IT support	314	100	97.22
Business intelligence	251	100	96.47
Cloud computing	218	100	95.95
System analyst	253	100	96.34
Total(average)	4,280	100 (in average)	95.96 (in average)

TABLE II
THE PERFORMANCE ON CONTEXT CLASSIFICATION AND LABELING

Type	Accuracy (%)
Responsibilities	100
Age	99.53
Gender	100
Nationality	97.05
Education	100
Experience	100
Skills	99.99
average	99.58

drill-down and roll-up operations, revealing a wide range of interesting patterns that occur within the same or across different parts of job advertisements. Through comprehensive examples and experiments, we demonstrate the accurate splitting and structuring of essential job advertisement information achieved by our proposed approach. Furthermore, *CID-CJA* offers flexibility for deep diving and searching for both common and rare co-occurrence information, enhancing the exploration of valuable insights.

REFERENCES

[1] M. F. Ansari, B. Dash, P. Sharma and N. Yathiraju, "The Impact and Limitations of Artificial Intelligence in Cybersecurity: A Literature Review," International Journal of Advanced Research in Computer and Communication Engineering, 2022.

[2] R. Pérez-Castillo, M. A. Serrano and M. Piattini, "Software modernization to embrace quantum technology," Advances in Engineering Software, Vol. 151, 2021.

[3] K. Aggarwal, M. M. Mijwil, Sonia, A. Al-Mistarehi, S. Alomari, M. Gök, A. M. Zein Alaabdin and S. H. Abdulrhman, "Has the Future Started? The Current Growth of Artificial Intelligence, Machine Learning, and Deep Learning," Iraqi Journal For Computer Science and Mathematics, 3(1), 2022, pp. 115–123.

[4] H. J. Nelson, A. Ahmad, N. L. Martin and C. R. Litecky, "A Comparative Study of IT/IS Job Skills and Job Definitions", Proceedings of the 2007 ACM SIGMIS CPR Conference on Computer Personnel Research: The Global Information Technology Workforce, 2007, pp 168 – 170.

[5] M. Lukauskas, V. Šarkauskaitė, V. Pilinkienė, A. Stundžienė, A. Grybauskas and J. Bruneckienė, "Enhancing Skills Demand Understanding through Job Ad Segmentation Using NLP and Clustering Techniques," Applied Sciences, 13(10), 2023.

[6] A. Parker, A. and I. Brown, "Skills Requirements for Cyber Security Professionals: A Content Analysis of Job Descriptions in South Africa", InInformation Security :Communications in Computer and Information Science, vol 973, 2019.

[7] M. Pejic-Bach, T. Bertoncel, M. Meško and Ž. Krstić, "Text mining of industry 4.0 job advertisements", International Journal of Information Management (IJIM), Vol. 50, 2020, pp. 416 – 431.

[8] E. M. Sibarani,S. Scerri, C. Morales, S. Auer, and D. Collarana, "Ontology-guided job market demand analysis: a cross-sectional study for the data science field, " In Proceedings of the 13th International Conference on Semantic Systems, 2017, pp. 25—32.

[9] Z. Radovilsky, V. Hedge, A. Acharya and U. Uma., "Skills Requirements of Business Data Analytics and Data Science Jobs: A Comparative Analysis," Journal of Supply Chain and Operations Management, 16(1), 2018, pp. 82—101.

[10] A. Verma, K. Lamsal and P. Verma, "An investigation of skill requirements in artificial intelligence and machine learning job advertisements," Ind. Higher Educ., 2021.

[11] P. Kamlangpuech and K. Amphawan, "A new system for analyzing contents of Computer Science courses," 2020 7th International Conference on Advance Informatics: Concepts, Theory and Applications (ICAICTA), 2020, pp. 1–6.

[12] P. Kamlangpuech and K. Amphawan, "eCSCDA: An efficient system for analyzing contents of Computer Science Courses," 2021 8th International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA), 2021, pp. 1–6.

[13] B. Chaisoongnoen, K. Amphawan and A. Bunpeng,"Supplementary Book Suggestion for Computer Science Courses," 2018 5th International Conference on Advanced Informatics: Concept Theory and Applications (ICAICTA) 2018, pp. 84–90.

[14] B. Chaisoongnoen and K. Amphawan, "An improvement of supplementary book suggestion system, " In The 9th International Conference on Smart Media and Applications (SMA 2020), 2020, pp. 396—398.

Gmail Naratchakorn eye <naratchakorn.r@gmail.com>

[ICSEC 2023] Decision on your paper 1570930072 (CID-CJA: Co-Occurrence Information Discovery in Computer-Related Job Advertisements)
1 ข้อความ

Ekkarat Boonchieng (ekkarat@boonchieng.net) <ekkarat=boonchieng.net@edas.info> 31 กรกฎาคม 2566 เวลา 13:04
ตอบกลับไปยัง: Ekkarat Boonchieng <ekkarat@boonchieng.net>
ถึง: Komate Amphawan <komate@gmail.com>, Naratchakorn Roeksiri <naratchakorn.r@gmail.com>

Dear Komate Amphawan,

Congratulations! We are delighted to inform you that your manuscript ID: 1570930072 entitled 'CID-CJA: Co-Occurrence Information Discovery in Computer-Related Job Advertisements' has been ACCEPTED for presentation at ICSEC 2023. All papers presented at ICSEC 2023 will be published in IEEE Xplore.

We will send more detail for the instructions later.

Best Regards,

ICSEC 2023 Chair

ICSEC 2023
THE 27TH INTERNATIONAL COMPUTER SCIENCE AND ENGINEERING CONFERENCE

Certificate of Presentation

Naratchakorn Roeksiri

for presenting paper entitled

CID-CJA: Co-Occurrence Information Discovery in Computer-Related Job Advertisements

The 27th International Computer Science and Engineering Conference (ICSEC 2023)
held between 13-15 September 2023 Koh Samui , Surat thani , Thailand

Prabhas Chongstitvatana
Chulalongkorn University

Apichat Sopadang
Dean of Graduate School, Chiang Mai University

Ekkarat Boonchieng
ICSEC 2023 General Chair

ประวัติย่อของผู้วิจัย

ชื่อ-สกุล	นรัชกร ฤกษ์ศิริ
วัน เดือน ปี เกิด	09/05/2537
สถานที่เกิด	จังหวัด ชลบุรี
สถานที่อยู่ปัจจุบัน	บ้านเลขที่ 326/132 หมู่ 8 ตำบล สุรศักดิ์ อำเภอ ศรีราชา จังหวัด ชลบุรี 20110
ตำแหน่งและประวัติการทำงาน	วิทยาศาสตรบัญฑิต สาขาวิทยาการคอมพิวเตอร์ คณะวิทยาการสารสนเทศ มหาวิทยาลัยบูรพา
ประวัติการศึกษา	ประถมศึกษา - มัธยมศึกษาตอนต้น โรงเรียนเซนต์ปอลคอนแวนต์ มัธยมศึกษาตอนปลาย โรงเรียนอัสสัมชัญศรีราชา ปริญญาตรี วิทยาศาสตรบัญฑิต สาขาวิทยาการคอมพิวเตอร์ คณะวิทยาการสารสนเทศ มหาวิทยาลัยบูรพา

BURAPHA UNIVERSITY