การพัฒนาแพลตฟอร์มข้อมูลชุมชนเพื่อการวิเคราะห์และวางแผนจัดการ

สุนิสา จุลรัตน์[1*] จิรศักดิ์ นพรัตน์[2] มงคล มโนพิรุฬห์พร[3] วรรณรัช สันติอมรทัต[4] สกุณา เจริญปัญญาศักดิ์[5]

[1*,2,3,4,5]*ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยสงขลานครินทร์, สงขลา, ประเทศไทย*

*ผู้ประพันธ์บรรณกิจ อีเมล : 6210120042@email.psu.ac.th

รับต้นฉบับ : 4 มีนาคม 2567; รับบทความฉบับแก้ไข : 23 เมษายน 2567; ตอบรับบทความ : 24 เมษายน 2567
เผยแพร่ออนไลน์ : 26 ธันวาคม 2567

บทคัดย่อ

ข้อมูลแต่ละชุมชนประกอบไปด้วยข้อมูลหลากหลายแขนง ซึ่งแต่ละชุมชนมีข้อมูลที่อาจเหมือนหรือแตกต่างกัน ขึ้นอยู่กับหลายปัจจัย เช่น สภาพพื้นที่ ภูมิอากาศ วิถีชีวิต วัฒนธรรม อาหารท้องถิ่น แหล่งพืชสมุนไพร ประเภทธุรกิจในชุมชน เป็นต้น ข้อมูลโดยส่วนใหญ่ ยังไม่ได้ถูกจัดเก็บเป็นข้อมูลดิจิทัล ทำให้การวิเคราะห์ข้อมูลแต่ละชุมชน ตำบล จังหวัด หรือระดับประเทศ ยังทำได้ล่าช้าและไม่ครอบคลุม ข้อมูลบางชุมชนซึ่งเคยถูกจัดเก็บลงระบบดิจิทัล ยังคงมีการแยกระบบจัดเก็บกันอย่างกระจัดกระจายและยังไม่เชื่อมโยงกัน ส่งผลให้ไม่สามารถบริหารจัดการข้อมูลร่วมกันได้ งานวิจัยนี้ จึงนำเสนอการออกแบบและพัฒนาแพลตฟอร์มข้อมูลชุมชนเพื่อการวิเคราะห์และวางแผนจัดการ ซึ่งรองรับการจัดเก็บข้อมูลชนิดข้อความ ตัวเลข ตำแหน่งที่อยู่ วันเวลา รูปภาพ เอกสาร เป็นต้น โดยสามารถเชื่อมโยงข้อมูลจากระบบต่าง ๆ ที่มีอยู่แล้วผ่าน Application Programming Interface (API) ที่กำหนดได้ โดยระบบดำเนินการจัดเก็บข้อมูลดิบบนระบบ Google Cloud Firestore จัดเก็บไฟล์บนระบบ Google Cloud Storage และจัดเก็บข้อมูลสำหรับการสืบค้นอย่างรวดเร็วบนระบบ Elasticsearch นอกจากนั้นได้ออกแบบและพัฒนาระบบสำหรับบันทึกข้อมูลความหลากหลายของแต่ละชุมชน ซึ่งรองรับการสร้างโครงสร้างข้อมูลของแต่ละหมวดหมู่โดยอัตโนมัติ เพื่ออำนวยความสะดวกให้ผู้สำรวจข้อมูลชุมชนสามารถเก็บบันทึกข้อมูลได้อย่างสะดวกยิ่งขึ้น การออกแบบและพัฒนาระบบจัดการข้อมูลหลากหลายนี้ จึงช่วยรวบรวมกลุ่มข้อมูลประเภทต่าง ๆ ของแต่ละชุมชนได้อย่างยืดหยุ่น ช่วยให้ผู้ใช้งานข้อมูลสามารถนำไปประยุกต์ใช้ในการแก้ปัญหาและพัฒนาเมืองได้ตามต้องการ

คำสำคัญ : การจัดการข้อมูลขนาดใหญ่ การวิเคราะห์ข้อมูล การเก็บข้อมูล ระบบการจัดการข้อมูล ข้อมูลหลากหลายชนิด

Development of a Community Data Platform for Analytics and Management Planning

Sunisa Junrat[1*] Jirasak Nopparat[2] Mongkon Manopiroonporn[3] Wannarat Suntiamorntut[4]
Sakuna Charoenpanyasak[5]

[1*,2,3,4,5]*Department of Computer Engineering, Faculty of Engineering, Prince of Songkla University, Songkhla, Thailand*

*Corresponding Author. E-mail address: 6210120042@email.psu.ac.th

Received: 4 March 2024; Revised: 23 April 2024; Accepted: 24 April 2024
Published online: 26 December 2024

Abstract

Data from each community comprises various categories, influenced by factors such as community condition, climate, lifestyle, culture, local cuisine, plant types, and businesses. Most of this data has not been stored digitally. Analyzing data from individual communities, sub-districts, provinces, and on a national scale can be delayed and incomplete. Some of the data that was previously stored digitally are in separate, scattered, and unconnected storage systems. As a result, managing this data collectively is challenging. This study proposes the design and development of a community data platform for analytics and management planning. The platform supports storing diverse data types such as text, numbers, locations, dates, times, images, and documents. It can link data from other existing systems through defined APIs and stores raw data on Google Firebase Firestore, files on Google Cloud Storage, and data for quick retrieval on Elasticsearch. Furthermore, the platform includes the design and development of applications to record various data categories for each community. It supports the automatic creation of data structures for each data category to make data collection by community members more convenient. The design and development of this multi-data management system allow for flexible gathering of different types of data from each community. As a result, it can be used to address problems and develop cities as desired.

***Keywords*:** Big data management, Data analytics, Data collection, Data platform, Diversity data

1) บทนำ

ข้อมูลแต่ละชุมชนประกอบไปด้วยข้อมูลหลากหลายหมวดหมู่ ไม่ว่าจะเป็นข้อมูลพืชท้องถิ่น อาหารประจำถิ่น ภูมิปัญญาท้องถิ่น แหล่งที่พัก ความเชี่ยวชาญของบุคคลในพื้นที่ สภาพภูมิอากาศและแหล่งน้ำ ข้อมูลเหล่านี้มีจำนวนมากและขนาดใหญ่ขึ้นเรื่อย ๆ อีกทั้งข้อมูลส่วนใหญ่ยังไม่ได้รับการสำรวจและบันทึกลงบนระบบดิจิทัลที่มีการจัดเก็บข้อมูลหลากหลายชนิดอย่างเหมาะสม เพื่อการสืบค้นตามช่วงเวลาที่สนใจ นอกจากนั้นยังมีข้อมูลชุมชนที่ถูกจัดเก็บกระจัดกระจายอยู่บนระบบดิจิทัลอื่น ๆ ซึ่งยังไม่มีการเชื่อมโยงข้อมูลเหล่านั้นเข้าด้วยกัน ทำให้ยังเป็นอุปสรรคต่อการบริหารจัดการและวิเคราะห์ข้อมูลหลากหลายร่วมกัน

งานวิจัยนี้ นำเสนอการออกแบบและพัฒนาระบบจัดเก็บข้อมูลความหลากหลายสำหรับแต่ละชุมชน โดยรองรับชนิดข้อมูลรูปแบบข้อความ ตัวเลข ตำแหน่งที่อยู่ วันเวลา รูปภาพ เอกสาร เป็นต้น และสามารถเชื่อมโยงข้อมูลจากระบบต่าง ๆ ที่เคยบันทึกข้อมูลในรูปแบบดิจิทัลแล้วผ่าน APIs [1] ที่กำหนดโดยดำเนินการจัดเก็บข้อมูลดิบบนระบบ Google Cloud Firestore [2] จัดเก็บไฟล์บนระบบ Google Cloud Storage [3] และข้อมูลสำหรับการสืบค้นอย่างรวดเร็วบนระบบ Elasticsearch [4] นอกจากนั้นยังได้ออกแบบและพัฒนาระบบสำหรับบันทึกข้อมูลหลากหลายหมวดหมู่ตามความต้องการของแต่ละชุมชน ซึ่งรองรับการสร้างโครงสร้างการจัดเก็บข้อมูลของแต่ละหมวดหมู่โดยอัตโนมัติ เพื่ออำนวยความสะดวกให้กับผู้ลงสำรวจเก็บข้อมูลในการรวบรวมข้อมูลที่มีลักษณะเฉพาะถิ่น เพื่อนำมาวิเคราะห์และวางแผนบริหารจัดการเมืองต่อไป ข้อมูลที่ผ่านการจัดเก็บลงสู่ระบบจัดเก็บข้อมูลความหลากหลายนี้ ผู้ใช้งานข้อมูล เช่น นักวิทยาศาสตร์ข้อมูล วิศวกรข้อมูล ผู้จัดการข้อมูลของแต่ละเมือง สามารถเลือกชุดข้อมูลที่สนใจ เพื่อนำมาวิเคราะห์ร่วมกับข้อมูลจากหลากหลายแหล่งได้ เพื่อนำมาประยุกต์แก้ไขปัญหาและพัฒนาเมืองได้อย่างเหมาะสม ดังเช่นตัวอย่าง การนำข้อมูลมาบริหารจัดการความเสี่ยงและภัยพิบัติ [5] การรวบรวมและสรุปความต้องการและความคาดหวังของประชาชนในเมืองลอนดอน [6] เป็นต้น

การเก็บรวบรวมข้อมูลที่มีปริมาณมากขึ้นเรื่อย ๆ จากหลากหลายแหล่ง มีความจำเป็นอย่างยิ่งและส่งผลกระทบต่อการวิเคราะห์ชุมชน เมือง และประเทศ นักวิจัยจึงพยายามรวบรวมข้อมูลเพื่อวิเคราะห์ปัญหาสำหรับการบริหารจัดการเมือง โดยพบว่าข้อมูลส่วนใหญ่ยังไม่ถูกแปลงสู่ระบบดิจิทัล และมีการเก็บข้อมูลกระจัดกระจายแยกออกจากกัน แต่ละเมืองจะมีวิธีการจัดเก็บข้อมูลเป็นของตนเอง ทำให้ไม่สามารถเชื่อมโยงและแลกเปลี่ยนข้อมูลระหว่างกันได้ หลายภาคส่วนมีการรวบรวมข้อมูลเดียวกันแต่ใช้วิธีการที่ต่างกัน โดยวิธีการจัดเก็บข้อมูลแตกต่างกันออกไปตั้งแต่การจัดเก็บในตาราง เอกสาร คอมพิวเตอร์ ปัญหาที่พบคือข้อมูลมีความซ้ำซ้อนและไม่สอดคล้องกัน ส่งผลให้ยังไม่สามารถเชื่อมโยงข้อมูลเพื่อสืบค้นและนำมาใช้ในการวางแผนบริหารจัดการเมืองหรือชุมชนได้ จึงยังมีความต้องการเครื่องมือสำหรับเก็บข้อมูลความหลากหลายของแต่ละชุมชน เพื่อสืบค้นร่วมกันและนำมาประยุกต์สำหรับการพัฒนาเมืองได้อย่างมีประสิทธิภาพ [7] โดยการเก็บรวบรวมข้อมูลในหลากหลายด้าน เช่น ด้านการเกษตร มีผู้วิจัยออกแบบและพัฒนาระบบเก็บข้อมูลด้านการเกษตรขนาดใหญ่ โดยรวบรวมข้อมูลจากหลายแหล่ง เช่น เซนเซอร์ชนิดต่าง ๆ อุปกรณ์ระบุตำแหน่ง เครื่องมือทางการเกษตร ชาวนา และธุรกิจด้านการเกษตร เป็นต้น ซึ่งข้อมูลที่ได้มีหลากหลายชนิดข้อมูล มีขนาดใหญ่ และมีความซับซ้อน มีทั้งข้อมูลที่มีโครงสร้างและไร้โครงสร้าง จึงได้เลือกฐานข้อมูลแบบ NoSQL เป็นเครื่องมือในการจัดการคลังข้อมูล (data warehouse) เนื่องจากสามารถเก็บข้อมูลได้ยืดหยุ่น และอ่านเขียนข้อมูลได้อย่างรวดเร็ว [8], [9] ช่วยให้สืบค้นข้อมูลและนำมาช่วยบริหารจัดการคลังข้อมูลได้อย่างมีประสิทธิภาพ โดยยังใช้การออกแบบโครงสร้างข้อมูลและการเชื่อมโยงความสัมพันธ์ระหว่างกันด้วยผู้เชี่ยวชาญเฉพาะ ยังไม่มีระบบช่วยอำนวยความสะดวกในการจัดทำโครงสร้างการจัดเก็บลงฐานข้อมูล และการสืบค้นข้อความจำเป็นต้องป้อนคำหรือข้อความที่มีความหมายจะทำให้สามารถสืบค้นข้อมูลออกมาได้ [10] นอกจากนั้นการเก็บข้อมูลสำหรับด้านการแพทย์ได้มีผู้ออกแบบและพัฒนาระบบจัดการข้อมูลเกี่ยวกับโรคมะเร็ง ซึ่งมีข้อมูลหลากหลายชนิดทั้งที่มีโครงสร้างและไม่มีโครงสร้าง โดยได้นำข้อมูลจำนวนมหาศาลจากหลากหลายแหล่งมาจัดเก็บรวมกัน เพื่อนำมาประมวลผลและแสดงรายงานให้บุคลากรผู้เกี่ยวข้องสามารถติดตามและวิเคราะห์ร่วมกันได้ โดยพบปัญหาข้อมูลบางส่วนยังมีความผิดพลาด เนื่องจากยังต้องป้อนเข้ามาด้วยมนุษย์และยังคงใช้เจ้าหน้าที่ในการตรวจสอบความถูกต้องด้วยตนเอง ทำให้เกิดความไม่สะดวกในประการนี้ จึงส่งผลต่อการสืบค้นเพื่อนำข้อมูลไปใช้ประโยชน์ได้ไม่ครอบคลุม [11]

ด้านการเลือกใช้ระบบเก็บข้อมูลหรือเทคนิคที่เหมาะสม ได้มีผู้ออกแบบและพัฒนาระบบสำหรับวิเคราะห์การบริโภคภายในมหาวิทยาลัย เจอปัญหาความเร็วในการวิเคราะห์ข้อมูลขนาดใหญ่บนฐานข้อมูลเชิงสัมพันธ์ (relational database) ซึ่งใช้เวลานานในการประมวลผลและการค้นหาข้อมูล จึงได้นำฐานข้อมูลชนิด NoSQL [9] มาจัดเก็บข้อมูลการบริโภคแทน พบว่า สามารถประมวลผลการค้นหาข้อมูลและตอบสนองได้ไวกว่าฐานข้อมูลเชิงสัมพันธ์เป็นอย่างมาก และรองรับการขยายตัวของข้อมูลเป็นอย่างดี [12] ส่วนเทคนิคการสืบค้นให้ได้ผลลัพธ์ที่ถูกต้องสอดคล้องนั้น มีผู้วิจัยออกแบบระบบสำหรับการค้นหาผลลัพธ์ด้วยภาษาพูดหรือภาษาที่ไม่เป็นทางการ โดยใช้การเก็บข้อมูลคำหรือข้อความ ซึ่งมีทั้งที่สะกดถูกต้องและสะกดผิด เพื่อทดสอบว่าหากใช้ภาษาพูดในการสืบค้นข้อมูลจะได้ผลลัพธ์ที่สอดคล้องตรงกันกับคำค้นหามากน้อยเพียงใด โดยผลลัพธ์ที่ได้จากการทดสอบยังไม่ค่อยถูกต้องและไม่สอดคล้องกับข้อมูลที่ต้องการค้นหา และมีหลายครั้งที่ทำการสืบค้นแล้วไม่ได้ผลลัพธ์ตอบสนองกลับมา [13]

2) วัตถุประสงค์ของงานวิจัย

1. การออกแบบและพัฒนาสถาปัตยกรรมของระบบจัดเก็บข้อมูลความหลากหลายของแต่ละชุมชน ซึ่งรองรับการค้นหาข้อมูลขนาดใหญ่ได้อย่างรวดเร็ว

2. ผู้ใช้งานข้อมูลสามารถสืบค้นชุดข้อมูลที่สนใจ เพื่อนำข้อมูลความหลากหลายไปวิเคราะห์และวางแผนการบริหารจัดการชุมชนได้ตามวัตถุประสงค์

3) วิธีดำเนินการวิจัย

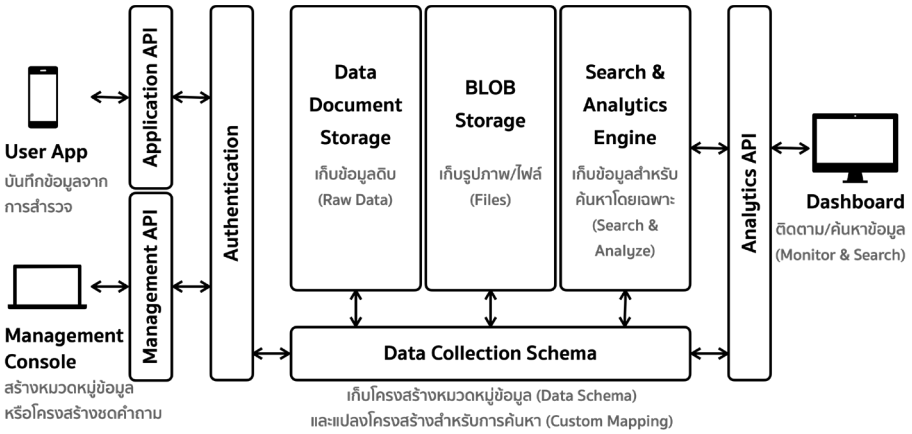
ในหัวข้อนี้จะอธิบายถึงการออกแบบและวิธีการทดสอบระบบที่เกี่ยวข้องกับงานวิจัยนี้ ซึ่งประกอบไปด้วย สถาปัตยกรรมของระบบจัดการข้อมูลความหลากหลาย โครงสร้างการเก็บข้อมูลรูปแบบโครงสร้างการเก็บข้อมูล ความแตกต่างของโครงสร้างข้อมูลแต่ละรูปแบบ และวิธีการทดสอบประสิทธิภาพ

3.1) สถาปัตยกรรมของระบบจัดการข้อมูลความหลากหลาย

การจัดการข้อมูลหลากหลาย ซึ่งมีปริมาณมากขึ้นอยู่ตลอดเวลา จำเป็นต้องใช้แพลตฟอร์มสำหรับบริหารจัดการข้อมูลชุมชนขนาดใหญ่ ซึ่งได้รับการออกแบบและพัฒนาขึ้นมาโดยบุคลากรผู้เชี่ยวชาญเฉพาะด้าน เช่น วิศวกรข้อมูล

การจัดเก็บข้อมูลอย่างมีประสิทธิภาพ จำเป็นต้องออกแบบตั้งแต่โครงสร้างการเก็บข้อมูลให้มีความเหมาะสมสอดคล้องกับข้อมูลหลากหลาย รวมถึงการตรวจสอบความถูกต้องของข้อมูล เพื่อให้ข้อมูลที่ถูกจัดเก็บสามารถนำไปใช้ประโยชน์ได้ทุกส่วน จึงได้ออกแบบโครงสร้างของระบบจัดเก็บข้อมูลความหลากหลายขึ้นมา ดังรูปที่ 1

จากรูปที่ 1 แสดงโครงสร้างของระบบจัดการข้อมูลความหลากหลาย ซึ่งใช้สำหรับการจัดการข้อมูลแต่ละหมวดหมู่ที่ต้องการจัดเก็บ โดยสามารถสร้างและจัดการโครงสร้างข้อมูลตามหมวดหมู่หรือโครงสร้างของชุดคำถามจากส่วน Management Console ซึ่งแลกเปลี่ยนข้อมูลผ่าน Management API และยืนยันสิทธิ์การเข้าถึงข้อมูลผ่าน Authentication เพื่อเชื่อมต่อกับส่วนของ Data Collection Schema สำหรับบันทึกโครงสร้างข้อมูล ซึ่งส่วนนี้จะเชื่อมโยงกับ Data Document Storage สำหรับเก็บข้อมูลดิบ และ BLOB Storage สำหรับเก็บข้อมูลไฟล์ เช่น รูปภาพ เป็นต้น



รูปที่ 1 : โครงสร้างระบบจัดการข้อมูลความหลากหลาย

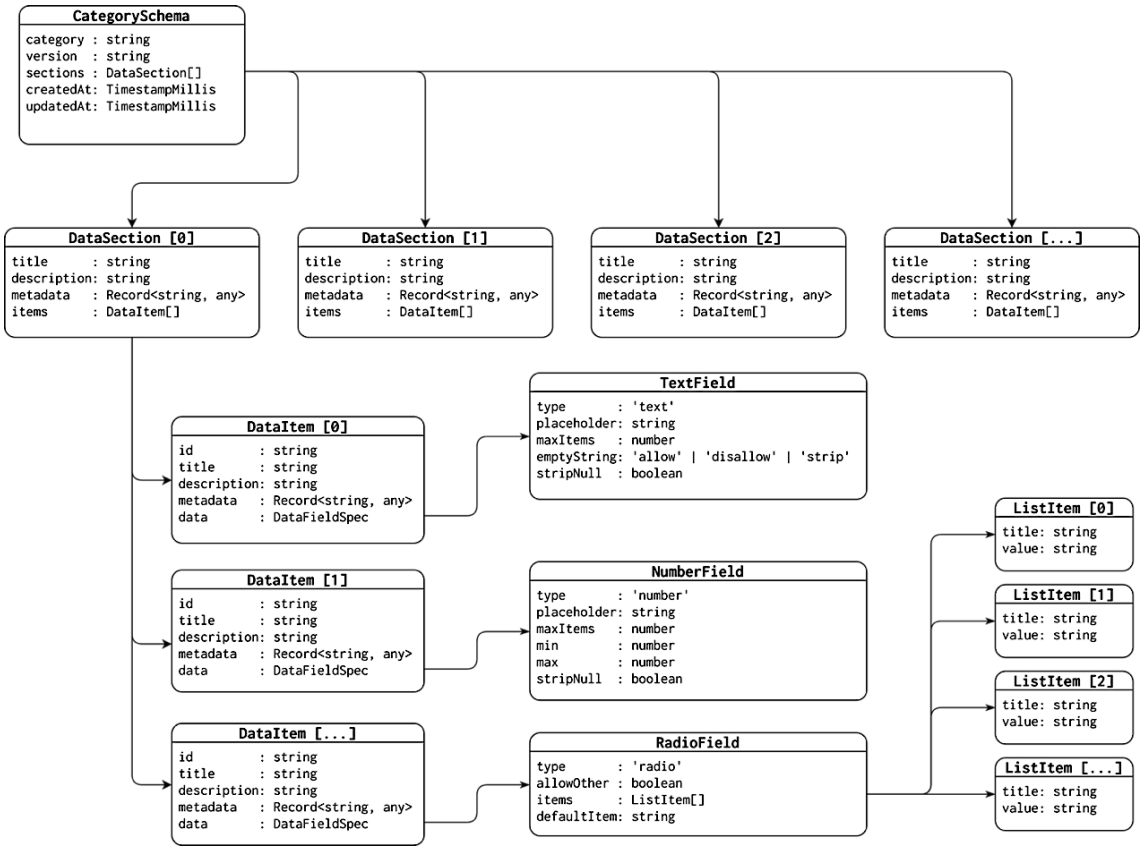
เมื่อสร้างหมวดหมู่ข้อมูลเรียบร้อยแล้ว ผู้สำรวจและบันทึกข้อมูลในแต่ละชุมชน สามารถบันทึกและจัดการข้อมูลได้ผ่าน User App ซึ่งแลกเปลี่ยนข้อมูลผ่าน Application API และยืนยันสิทธิ์การเข้าถึงข้อมูลผ่าน Authentication เพื่อเชื่อมต่อกับ Data Collection Schema เพื่อนำโครงสร้างข้อมูลของหมวดหมู่ที่มีการบันทึกมาแสดงผล ซึ่งส่วนนี้เชื่อมต่อกับ Data Document Storage และ BLOB Storage เพื่อนำข้อมูลดิบและไฟล์ที่สัมพันธ์กันนำกลับไปแสดงผลบน User App

ส่วนของ Data Collection Schema เป็นส่วนที่ทำหน้าที่เก็บโครงสร้างแต่ละหมวดหมู่ข้อมูล (Data Schema) และเชื่อมโยงการเข้าถึงข้อมูลดิบซึ่งจัดเก็บใน Data Document Storage และเชื่อมโยงการเข้าถึงไฟล์ข้อมูล ซึ่งจัดเก็บใน BLOB Storage นอกจากนั้น ยังทำหน้าที่แปลงโครงสร้างแต่ละหมวดหมู่ข้อมูล เป็นโครงสร้างสำหรับการค้นหา (Custom Mapping) ซึ่งใช้งานในส่วนของ Search & Analytics Engine

เมื่อมีการบันทึกข้อมูลซึ่งได้จากการสำรวจแล้ว สามารถติดตามหรือค้นหาข้อมูลได้ผ่าน Dashboard ซึ่งแลกเปลี่ยนข้อมูลผ่าน Analytics API เพื่อเข้าถึงฐานข้อมูลสำหรับการค้นหาหรือส่วนของ Search & Analytics Engine ซึ่งจะทำหน้าที่ประมวลผลตามข้อกำหนดของ Custom Mapping และนำผลลัพธ์ที่ได้กลับไปแสดงผลยัง Dashboard

3.2) รูปแบบโครงสร้างการเก็บข้อมูล (JSON Schema Definition)

รูปแบบโครงสร้างการเก็บข้อมูลที่สอดคล้องกันและมีความยืดหยุ่นในการปรับเพิ่มหรือลดข้อมูลเป็นสิ่งสำคัญอย่างยิ่งในการรองรับกับข้อมูลที่มีความหลากหลายหรือหมวดหมู่ข้อมูลใหม่ ๆ ซึ่งจะถูกเพิ่มเข้ามาในอนาคตอีกมากมาย เพื่อให้สะดวกต่อการจัดการข้อมูลหลากหลายสาขานี้ จึงได้ออกแบบโครงสร้างแต่ละหมวดหมู่ข้อมูล ซึ่งอ้างอิงรูปแบบโครงสร้างตาม JSON Schema Definition ที่ได้ออกแบบดังรูปที่ 2 โดยโครงสร้างแต่ละหมวดหมู่ข้อมูล จะถูกบันทึกโครงสร้างไว้ใน Data Collection Schema ซึ่งเป็นส่วนประกอบที่ระบุไว้ในรูปที่ 1



รูปที่ 2 : JSON Data Definition

จากรูปที่ 2 แสดงโครงสร้างข้อมูลในรูปแบบ JSON (JSON Schema Definition) ซึ่งใช้ประกอบการสร้างหมวดหมู่สำหรับการเก็บข้อมูล (CategorySchema) ซึ่งประกอบด้วย 5 หัวข้อ (Fields) คือ

1. category ระบุรหัสของหมวดหมู่ข้อมูล
2. version ระบุเวอร์ชันของหมวดหมู่ข้อมูล
3. createdAt ระบุเวลาที่หมวดหมู่ข้อมูลถูกสร้างขึ้น
4. updatedAt ระบุเวลาที่หมวดหมู่ข้อมูลถูกแก้ไขล่าสุด
5. sections เป็นรายการของ DataSection ซึ่งจะอธิบายในส่วนถัดไป

DataSection ประกอบไปด้วย 4 Fields ดังนี้

1. title ระบุชื่อของ section
2. description ระบุคำอธิบายของ section
3. metadata สำหรับระบุข้อมูลเพิ่มเติมตามต้องการ
4. items เป็นรายการของ DataItem ซึ่งจะอธิบายในส่วนถัดไป

DataItem ประกอบไปด้วย 5 Fields ดังนี้

1. id ระบุรหัสของ DataItem
2. title ระบุชื่อของ DataItem
3. description ระบุคำอธิบายของ DataItem
4. metadata สำหรับระบุข้อมูลเพิ่มเติมตามต้องการ
5. data เป็นรายการของ DataFieldSpec ซึ่งจะอธิบายในส่วนถัดไป

DataFieldSpec คือ ชนิดของข้อมูลที่จัดเก็บ ซึ่งนอกจากจะใช้ในการระบุชนิดของข้อมูลแล้ว ยังใช้ในการตรวจสอบความถูกต้องของข้อมูลด้วย (validation) โดยในปัจจุบันได้ออกแบบให้รองรับข้อมูล 11 ชนิด ชนิดของข้อมูลมีดังนี้

1. Text Field เก็บข้อมูลที่เป็นข้อความ
2. Number Field เก็บข้อมูลเชิงตัวเลข
3. Date Field เก็บข้อมูลวันที่
4. Time Field เก็บข้อมูลเวลา
5. Date and Time Field เก็บข้อมูลวันที่และเวลา
6. Geo Point Field เก็บข้อมูลพิกัด (ละติจูด ลองจิจูด)
7. Address Field เก็บข้อมูลที่อยู่
8. Drop Down Field เก็บข้อมูล 1 รายการ จากตัวเลือกที่กำหนดไว้
9. Radio Field เก็บข้อมูล 1 รายการ จากตัวเลือกที่กำหนดไว้ หรือระบุข้อมูลอื่น ๆ เพิ่มเติม
10. Check Boxes Field เก็บข้อมูลได้หลายตัวเลือกจากที่กำหนดไว้ หรือระบุข้อมูลอื่น ๆ เพิ่มเติม
11. File Field เก็บข้อมูลไฟล์ เช่น รูปภาพ ไฟล์เอกสาร

3.3) ความแตกต่างของโครงสร้างข้อมูลแต่ละรูปแบบ

โครงสร้างข้อมูลแต่ละรูปแบบ ซึ่งมีการเลือกใช้เครื่องมือหรือเทคนิคที่ไม่เหมือนกัน ย่อมส่งผลต่อประสิทธิภาพในการจัดเก็บและค้นหาข้อมูล จึงขอนำเสนอความแตกต่างระหว่างโครงสร้างข้อมูลแต่ละรูปแบบ ดังต่อไปนี้

3.3.1) โครงสร้างข้อมูลซึ่งจัดเก็บลงใน Google Firestore [2] ซึ่งอยู่ในรูปแบบของ JSON นั้นมีความยืดหยุ่นในการบันทึกหมวดหมู่ข้อมูลที่มีจำนวนข้อมูลไม่เท่ากันได้ โดยหน้าตาข้อมูลแต่ละหมวดหมู่มีการจัดแบ่งโครงสร้างเป็นสองกลุ่ม ประกอบไปด้วยโครงสร้างข้อมูลสำหรับจัดเก็บคำถาม ซึ่งระบบมีการจัดเก็บข้อมูลตามรูปแบบของ JSON Schema Definition ดังรูปที่ 2 และโครงสร้างข้อมูลสำหรับจัดเก็บคำตอบ ของหมวดหมู่นั้น ๆ โดยการจัดเก็บลงใน Google Firestore [2] ซึ่งเป็นฐานข้อมูลชนิด NoSQL [9] สามารถช่วยในการเก็บข้อมูลได้อย่างยืดหยุ่นสำหรับข้อมูลหลากหลายสาขา

3.3.2) โครงสร้างข้อมูลรูปแบบ Dynamic Mapping เป็นโครงสร้างข้อมูลที่ Elasticsearch [4] จัดเตรียมและแปลงข้อมูลที่ต้องการจัดเก็บให้แบบอัตโนมัติ ซึ่งผู้จัดการฐานข้อมูลไม่จำเป็นต้องกำหนด Mappings [14] ขณะเริ่มต้นสร้าง Index [15] สำหรับเก็บข้อมูลหมวดหมู่ต่าง ๆ แต่อย่างใด เนื่อง Elasticsearch [4] จัดการประมวลผลให้โดยอัตโนมัติ

3.3.3) โครงสร้างข้อมูลซึ่งใช้งาน Thai Tokenizer [16] เป็นโครงสร้างข้อมูลที่กำหนดวิธีการตัดคำไทยรูปแบบหนึ่งของฐานข้อมูล Elasticsearch [4] ซึ่งสามารถตั้งค่าการใช้งานได้ผ่านการกำหนด Mappings [14] ส่วนของ analyzer เป็น thai เพื่อเปิดใช้คุณสมบัติการตัดคำสำหรับภาษาไทย โดยส่วนที่ไม่ได้ตั้งค่าหรือกำหนดคุณสมบัติ (properties) ไว้ Elasticsearch [4] จะจัดการให้โดยอัตโนมัติ

3.3.4) โครงสร้างข้อมูลรูปแบบ Custom Mapping เป็นโครงสร้างข้อมูลซึ่งได้ออกแบบและพัฒนา สำหรับจัดเก็บลงฐานข้อมูล Elasticsearch [4] ซึ่งคำนึงถึงประสิทธิภาพให้การสืบค้นให้รวดเร็วและถูกต้องมากยิ่งขึ้น โดยข้อมูลที่ได้ผ่านการแปลงจากส่วน Data Collection Schema ซึ่งมีการกำหนด

วิธีการตัดคำภาษาไทยโดยใช้ N-gram Tokenizer [17] รูปแบบ N-gram2, 3 เพิ่มเติม

3.4) วิธีการทดสอบประสิทธิภาพ

งานวิจัยนี้ มุ่งเน้นไปที่หัวข้อการสืบค้นด้วยคำหรือข้อความ จึงได้ออกแบบการทดลองสำหรับเปรียบเทียบประสิทธิภาพการค้นหาของแต่ละรูปแบบโครงสร้างข้อมูล จึงแบ่งหัวข้อการทดสอบออกเป็นสามการทดลอง ดังต่อไปนี้

3.4.1) การทดลองที่ 1 ออกแบบมาเพื่อเปรียบเทียบความถูกต้องสอดคล้องกันระหว่างโครงสร้างข้อมูลแบบ Dynamic Mapping การใช้งาน Thai Tokenizer [16] และโครงสร้างข้อมูลรูปแบบ Custom Mapping โดยพิจารณาจากผลลัพธ์จากการค้นหาด้วยคำหรือข้อความใน 1 ลำดับ 5 ลำดับ และ 10 ลำดับแรก โดยข้อมูลที่นำมาใช้ทดลองมาจากหมวดหมู่พืชในท้องถิ่น 1,952,571 รายการ ซึ่งเป็นหมวดหมู่ข้อมูลที่ได้มาจากการสำรวจ 7,431 ตำบล ในประเทศไทย โดยคำค้นหาที่จะใช้ทดสอบประกอบไปด้วย สัก ต้นสัก สักทอง ฟ้าทะลายโจร ฟ้าทะลาย ทะลายโจร ถั่วฝักยาว ถั่วฝัก และฝักยาว

3.4.2) การทดลองที่ 2 ออกแบบมาเพื่อเปรียบเทียบความเร็วในการค้นหาข้อมูลระหว่างโครงสร้างข้อมูลแบบ Dynamic Mapping การใช้งาน Thai Tokenizer [16] และโครงสร้างข้อมูลรูปแบบ Custom Mapping โดยข้อมูลที่นำมาใช้ทดลองมาจากหมวดหมู่พืชในท้องถิ่น 1,952,571 รายการ ซึ่งเป็นชุดข้อมูลเดียวกันกับการทดลองที่ 1 โดยคำค้นหาที่จะใช้ทดสอบประกอบไปด้วย สับปะรดภูแล กะเพรา สะตอ ผักเหลียง เห็ดเผาะ บอระเพ็ด ว่านหางจระเข้ ถั่วแระ มะเขือเทศสีดา ข้าวหอมมะลิ

3.4.3) การทดลองที่ 3 ออกแบบมาเพื่อทดสอบประสิทธิภาพการใช้งาน Autocomplete [18] ซึ่งเป็นการระบุคำที่ต้องการค้นหาทีละตัวอักษร โดยต้องการเปรียบเทียบผลลัพธ์ระหว่างโครงสร้างข้อมูลแบบ Dynamic Mapping การใช้งาน Thai Tokenizer [16] และโครงสร้างข้อมูลรูปแบบ Custom Mapping โดยข้อมูลที่นำมาใช้ทดลองมาจากหมวดหมู่พืชในท้องถิ่น 1,952,571 รายการ ซึ่งเป็นชุดข้อมูลเดียวกันกับการทดลองที่ 1 และการทดลองที่ 2 โดยคำค้นหาที่จะใช้ทดสอบประกอบไปด้วย ไม ไมย ไมยร ไมยรา ไมยราบ มะ มะเ มะเข มะเขื เมเขือ กร กระ กระถ กระถิ กระถิน

4) ผลการทดลอง

ผลการทดลองและผลการวิเคราะห์ข้อมูล รวมถึงสถิติการใช้งานระบบเก็บข้อมูลความหลากหลาย จำนวนผู้ใช้งานและการรองรับของระบบ มีรายละเอียดดังต่อไปนี้

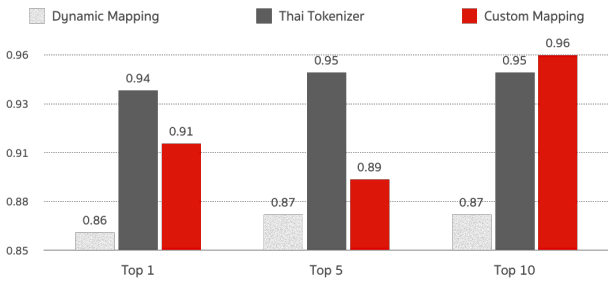
4.1) ผลการวิเคราะห์จากการทดลอง

4.1.1) ผลการวิเคราะห์จากการทดลองที่ 1 ซึ่งเป็นการทดสอบความถูกต้องสอดคล้องกันหรือความแม่นยำของผลลัพธ์ที่ได้จากการค้นหาคำระหว่างโครงสร้างข้อมูลแบบ Dynamic Mapping การใช้งาน Thai Tokenizer [16] และโครงสร้างข้อมูลรูปแบบ Custom Mapping โดยผลการทดลองแสดงดังตารางที่ 1 ตารางที่ 2 และรูปที่ 3

จากผลการทดลองในตารางที่ 1 ตารางที่ 2 และกราฟเปรียบเทียบดังรูปที่ 3 สรุปได้ว่าโครงสร้างข้อมูลแบบ Dynamic Mapping มีความแม่นยำต่ำที่สุด ส่วนโครงสร้างที่ใช้งาน Thai Tokenizer [16] และโครงสร้างแบบ Custom Mapping มีความแม่นยำใกล้เคียงกัน แต่เมื่อพิจารณาร่วมกันจำนวนผลลัพธ์ที่ค้นหาได้แล้ว โครงสร้างแบบ Custom Mapping สามารถสืบค้นข้อมูลได้ครอบคลุมมากกว่าอีกสองรูปแบบอย่างเห็นได้ชัดเจน

ตารางที่ 1 : ความแม่นยำในการค้นหาคำจากการทดลองที่ 1

รูปแบบโครงสร้างข้อมูล	ค่าความแม่นยำในการค้นหาคำเฉลี่ย		
	Top 1	Top 5	Top 10
Dynamic Mapping	0.86	0.87	0.87
Thai Tokenizer	0.94	0.95	0.95
Custom Mapping	0.91	0.89	0.96



รูปที่ 3 : กราฟเปรียบเทียบความแม่นยำในการค้นหาคำจากการทดลองที่ 1

ตารางที่ 2 : จำนวนผลลัพธ์ที่ค้นหาได้จากการทดลองที่ 1

คำค้นหา	จำนวนผลลัพธ์ที่ค้นหาได้ (รายการ)		
	Dynamic Mapping	Thai Tokenizer	Custom Mapping
สัก	969	8,620	192,700
ต้นสัก	4,025	977,624	1,162,212
สักทอง	408	29,162	292,896
ฟ้าทะลายโจร	5,256	18,965	343,422
ฟ้าทะลาย	11	18,919	328,679
ทะลายโจร	0	9,601	147,294
ถั่วฝักยาว	4,569	6,920	453,400
ถั่วฝัก	2	9,734	267,326
ฝักยาว	2	270	372,256

4.1.2) ผลการวิเคราะห์จากการทดลองที่ 2 ซึ่งเป็นการทดสอบความเร็วในการค้นหาด้วยคำระหว่างโครงสร้างข้อมูลแบบ Dynamic Mapping การใช้งาน Thai Tokenizer [16] และโครงสร้างข้อมูลรูปแบบ Custom Mapping โดยได้พิจารณาผลลัพธ์จากการค้นหาเป็น 1 ลำดับแรก (Top 1) 5 ลำดับแรก (Top 5) และ 10 ลำดับแรก (Top 10) ซึ่งค่าความแม่นยำจะอยู่ระหว่าง 0 ถึง 1 สามารถดูผลการทดลองแสดงดังตารางที่ 3 ถึงตารางที่ 5 โดยเกณฑ์การคำนวณคะแนนความแม่นยำ กำหนดตามรายละเอียดดังนี้

- กรณีผลลัพธ์ตรงกันกับคำค้นหาทุกตัวอักษร จะได้คะแนนเท่ากับ 1.00
- กรณีพบคำซ้ำหรือคำเกินมาจากคำค้นหา และสะกดคำถูกต้อง โดยสามารถสื่อสารถึงชื่อชนิดพืชได้ตามคำค้นหา จะได้คะแนนเท่ากับ 0.90
- กรณีพบคำซ้ำหรือคำเกินมาจากคำค้นหา แต่สะกดคำผิด ซึ่งส่งผลให้สื่อสารถึงชื่อชนิดพืชผิดเพี้ยน จะได้คะแนนเท่ากับ 0.50
- กรณีผลลัพธ์ไม่สอดคล้องหรือไม่พบผลลัพธ์ จะได้คะแนนเท่ากับ 0

จากตารางที่ 3 ถึงตารางที่ 5 เมื่อเปรียบเทียบเวลาที่ใช้ในการค้นหาคำและจำนวนผลลัพธ์ที่ได้จากโครงสร้างข้อมูลทั้งสามรูปแบบแล้ว สรุปได้ว่าโครงสร้างข้อมูลรูปแบบ Custom Mapping สามารถค้นหาคำได้รวดเร็วที่สุดและครอบคลุมมากที่สุดโดยใช้เวลาในการค้นหาเฉลี่ยต่อรายการเป็น 0.12 µs ลำดับถัดมาคือโครงสร้างแบบ Thai Tokenizer [16] โดยใช้เวลาในการค้นหาเฉลี่ยต่อรายการเป็น 0.51 µs และโครงสร้างที่ค้นหาได้ช้ามากที่สุดคือโครงสร้างแบบ Dynamic Mapping โดยใช้เวลาในการค้นหาเฉลี่ยต่อรายการเป็น 14.22 µs

ตารางที่ 3 : ความเร็วในการค้นหาด้วยคำ (ms) จากการทดลองที่ 2

คำค้นหา	เวลาที่ใช้ในการค้นหา (ms)		
	Dynamic Mapping	Thai Tokenizer	Custom Mapping
สับปะรดภูแล	1	3	20
กะเพรา	3	2	31
สะตอ	1	1	9
ผักเหลียง	2	5	30
เห็ดเผาะ	11	3	11
บอระเพ็ด	2	1	22
ว่านหางจระเข้	2	5	93
ถั่วแระ	1	1	12
มะเขือเทศสีดา	2	7	55
ข้าวหอมมะลิ	1	5	36

ตารางที่ 4 : จำนวนผลลัพธ์ที่ค้นหาได้จากการทดลองที่ 2

คำค้นหา	จำนวนผลลัพธ์ที่ค้นหาได้ (รายการ)		
	Dynamic Mapping	Thai Tokenizer	Custom Mapping
สับปะรดภูแล	14	5,323	85,733
กะเพรา	3,525	9,263	279,099
สะตอ	883	2,239	32,571
ผักเหลียง	831	62,312	365,397
เห็ดเผาะ	232	5,889	93,345
บอระเพ็ด	890	1,643	305,789
ว่านหางจระเข้	5,707	29,751	616,189
ถั่วแระ	218	429	265,675
มะเขือเทศสีดา	181	52,527	558,283
ข้าวหอมมะลิ	1,366	46,679	668,452

ตารางที่ 5 : เวลาที่ใช้ในการค้นหาต่อรายการ (µs) จากการทดลองที่ 2

คำค้นหา	เวลาที่ใช้ในการค้นหา / 1 รายการ (µs)		
	Dynamic Mapping	Thai Tokenizer	Custom Mapping
สับปะรดภูแล	71.43	0.56	0.23
กะเพรา	0.85	0.22	0.11
สะตอ	1.13	0.45	0.28
ผักเหลียง	2.41	0.08	0.08
เห็ดเผาะ	47.41	0.51	0.12
บอระเพ็ด	2.25	0.61	0.07

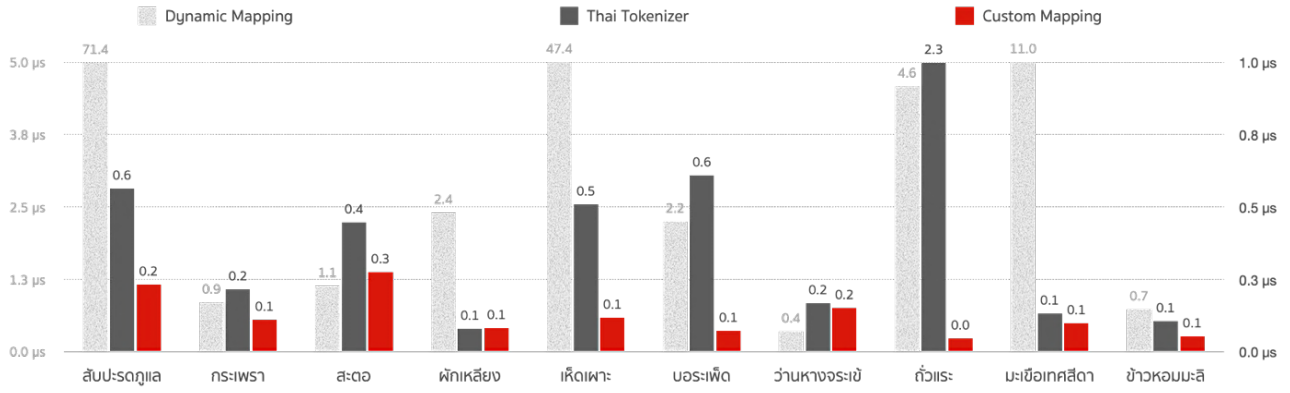
ตารางที่ 5 : เวลาที่ใช้ในการค้นหาต่อรายการ (µs) จากการทดลองที่ 2 (ต่อ)

คำค้นหา	เวลาที่ใช้ในการค้นหา / 1 รายการ (µs)		
	Dynamic Mapping	Thai Tokenizer	Custom Mapping
ว่านหางจระเข้	0.35	0.17	0.15
ถั่วแระ	4.59	2.33	0.05
มะเขือเทศสีดา	11.05	0.13	0.10
ข้าวหอมมะลิ	0.73	0.11	0.05
เฉลี่ย	**14.22**	**0.51**	**0.12**

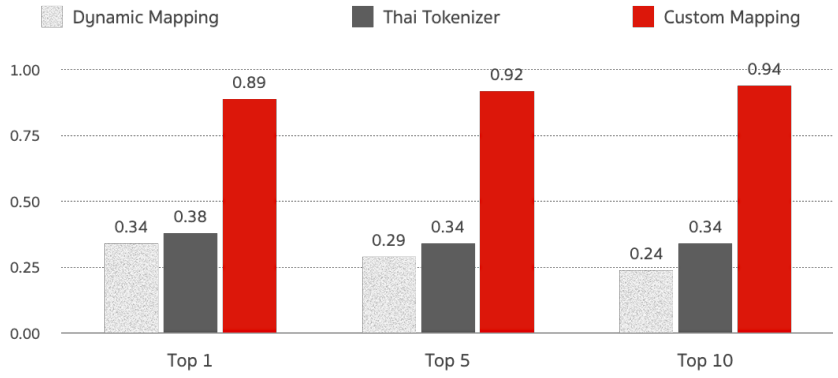
4.1.3) ผลการวิเคราะห์จากการทดลองที่ 3 ซึ่งเป็นการทดสอบการค้นหาผ่านการใช้งาน autocomplete [18] ระหว่างโครงสร้างข้อมูลแบบ Dynamic Mapping การใช้งาน Thai Tokenizer [16] และโครงสร้างข้อมูลรูปแบบ Custom Mapping โดยได้พิจารณาผลลัพธ์จากการค้นหาเป็น 1 ลำดับแรก (Top 1) 5 ลำดับแรก (Top 5) และ 10 ลำดับแรก (Top 10) ซึ่งค่าความแม่นยำจะอยู่ระหว่าง 0 ถึง 1 สามารถดูผลการทดลองดังรูปที่ 4 รูปที่ 5

ตารางที่ 7 และตารางที่ 8 โดยเกณฑ์การคำนวณคะแนนความแม่นยำ กำหนดตามรายละเอียดดังนี้

- กรณีผลลัพธ์ตรงกันกับคำค้นหาทุกตัวอักษร จะได้คะแนนเท่ากับ 1.00
- กรณีพบคำซ้ำหรือคำเกินมาจากคำค้นหา และสะกดคำถูกต้อง โดยสามารถสื่อสารถึงชื่อชนิดพืชได้ตามคำค้นหา จะได้คะแนนเท่ากับ 0.90
- กรณีพบคำซ้ำหรือคำเกินมาจากคำค้นหา แต่สะกดคำผิด ซึ่งส่งผลให้สื่อสารถึงชื่อชนิดพืชผิดเพี้ยน จะได้คะแนนเท่ากับ 0.50
- กรณีผลลัพธ์ไม่สอดคล้องหรือไม่พบผลลัพธ์ จะได้คะแนนเท่ากับ 0



รูปที่ 4 : กราฟเปรียบเทียบเวลาที่ใช้ในการค้นหาต่อรายการ (µs) จากการทดลองที่ 2



รูปที่ 5 : กราฟเปรียบเทียบความแม่นยำในการค้นหาคำจากการทดลองที่

ตารางที่ 6 : ความแม่นยำในการค้นหาคำจากการทดลองที่ 3

รูปแบบโครงสร้างข้อมูล	ค่าความแม่นยำในการค้นหาคำเฉลี่ย		
	Top 1	Top 5	Top 10
Dynamic Mapping	0.34	0.29	0.24
Thai Tokenizer	0.38	0.34	0.34
Custom Mapping	0.89	0.92	0.94

ตารางที่ 7 : จำนวนผลลัพธ์ที่ค้นหาได้จากการทดลองที่ 3

คำค้นหา	จำนวนผลลัพธ์ที่ค้นหาได้ (รายการ)		
	Dynamic Mapping	Thai Tokenizer	Custom Mapping
ไม	1	1	33,188
ไมย	0	0	34,472
ไมยร	0	0	37,982
ไมยรา	0	0	110,914
ไมยราบ	731	2,451	130,904
มะ	5	11,478	319,361
มะเ	0	0	394,875
มะเข	0	11,605	425,610
มะเขื	0	0	425,639
มะเขือ	4,045	25,038	486,504
กร	0	179	148,417
กระ	2	16,676	200,307
กระถ	0	16,852	201,343
กระถิ	0	0	203,042
กระถิน	4,335	13,513	243,198

จากรูปที่ 5 ตารางที่ 6 และตารางที่ 7 สรุปได้ว่าโครงสร้างข้อมูลแบบ Custom Mapping สามารถค้นหาในลักษณะการใช้งาน autocomplete [18] ซึ่งมีการค้นหาโดยเพิ่มไปทีละตัวอักษรนั้น ได้อย่างถูกต้องสอดคล้องและครอบคลุมมากที่สุดมากที่สุด ซึ่งมีข้อแตกต่างจากโครงสร้างรูปแบบ Dynamic Mapping และรูปแบบการใช้งาน Thai Tokenizer [16] ที่พบว่ายังไม่เหมาะสมสำหรับใช้การค้นหาด้วย autocomplete [18] เนื่องจากไม่สามารถหาผลลัพธ์ได้จากการพิมพ์ทีละตัวอักษรซึ่งยังไม่เป็นคำ จึงไม่เหมาะสมสำหรับการใช้งานแนะนำคำที่สอดคล้อง

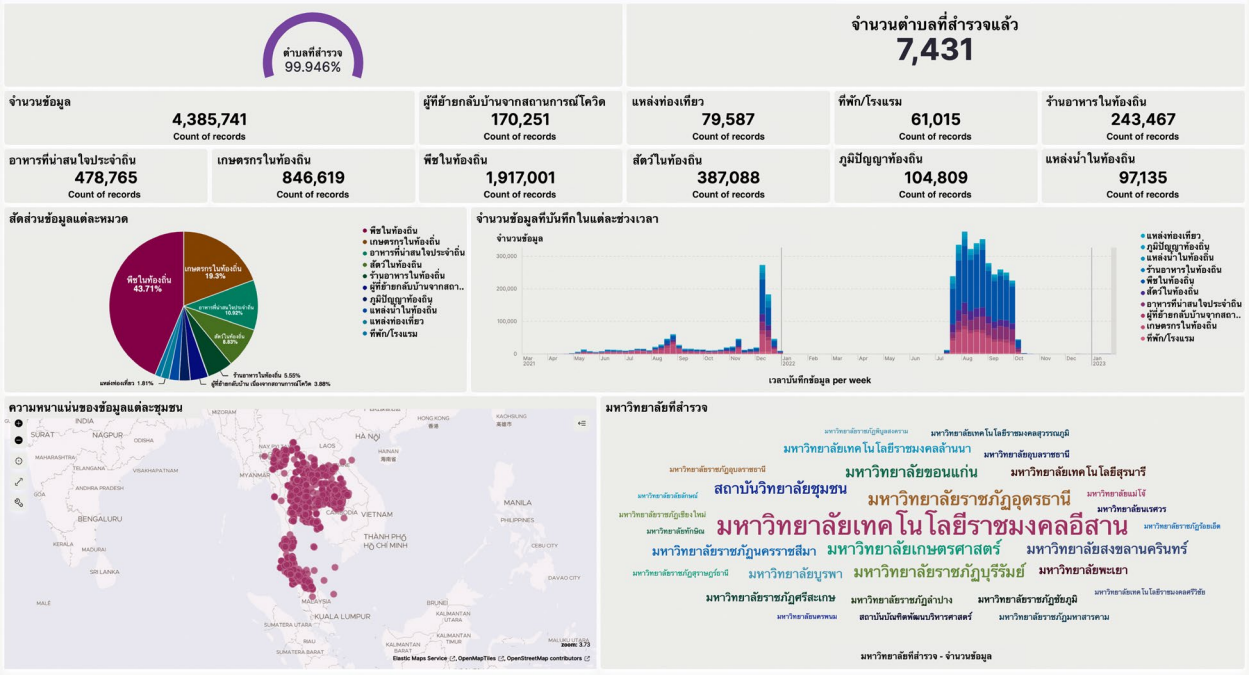
4.2) ผลการวิเคราะห์สถิติการใช้งานระบบ

ข้อมูลที่ถูกจัดเก็บในระบบจัดการข้อมูลหลากหลายนี้สามารถนำมาวิเคราะห์ต่อด้วยเครื่องมือ (data visualization tools) ตามความถนัดของผู้วิเคราะห์ เช่น Kibana [19], Power BI [20] เป็นต้น บทความวิจัยฉบับนี้ได้นำข้อมูลไปวิเคราะห์และจัดเรียงเป็นกระดานสรุปผล

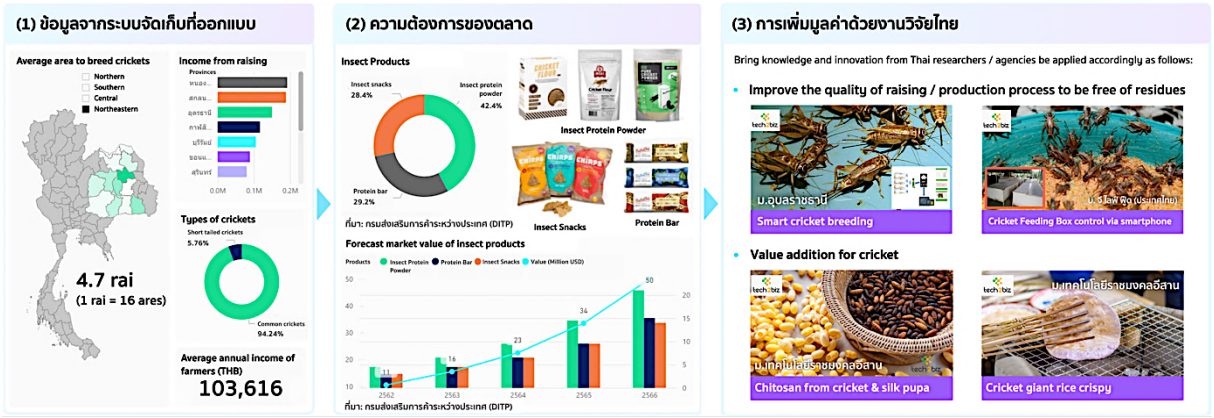
การติดตามข้อมูลสถิติหรือข้อมูลสรุปสำหรับผู้บริหารจัดการพื้นที่หรือชุมชนระดับต่าง ๆ ได้ในรูปแบบหลากหลาย เนื่องด้วยระบบจัดการข้อมูลความหลากหลายซึ่งได้ออกแบบมีการจัดเก็บข้อมูลสำหรับการสืบค้นลงใน Elasticsearch [4] ซึ่งสามารถเชื่อมโยงกับ Kibana [19] เพื่อแสดงผลข้อมูลสถิติในหลากหลายมุมมอง สามารถเลือกกรองเฉพาะชุดข้อมูลที่สนใจ และระบุวันเวลาที่ต้องการค้นหาข้อมูลได้ จึงได้จัดทำ Dashboard ด้วย Kibana เพื่อช่วยในการตัดสินใจวางแผนบริหารจัดการพัฒนาชุมชนต่อไปดังรูปที่ 6

จากรูปที่ 6 แสดงข้อมูลสถิติภาพรวมซึ่งได้จากระบบจัดการข้อมูลความหลากหลายที่ออกแบบตั้งแต่ มีนาคม 2564 ถึง สิงหาคม 2565 ซึ่งจำนวนข้อมูลทั้งหมดที่ได้จากการสำรวจและบันทึกเข้าสู่ระบบ คือ 4,385,741 รายการ จาก 7,431 ตำบลในประเทศไทย ซึ่งคิดเป็น 99.95% ของตำบลทั้งหมดในประเทศไทยซึ่งมีจำนวน 7,435 ตำบล โดยมีการเก็บข้อมูลในหลากหลายด้านจำนวน 10 หมวดหมู่ข้อมูล โดยหมวดหมู่ที่มีการบันทึกข้อมูลมากที่สุด 3 ลำดับแรก ประกอบด้วย พืชในท้องถิ่น 43.71% (1,917,001 รายการ) เกษตรกรในท้องถิ่น 19.3% (846,619 รายการ) และอาหารที่น่าสนใจประจำถิ่น 10.92% (478,765 รายการ)

ข้อมูลซึ่งถูกวิเคราะห์และแสดงผลในหลากหลายรูปแบบนี้สามารถนำไปประกอบการตัดสินใจและใช้สำหรับการวางแผนบริหารจัดการชุมชนของผู้บริหารในระดับประเทศ โดยสามารถนำข้อมูลเหล่านี้ไปต่อยอดเพื่อใช้งานในหลากหลายด้าน เช่น การพัฒนาโปรตีนทางเลือก การพัฒนาองค์ความรู้การเลี้ยงโคนม การวิเคราะห์สัดส่วนการผลิตเส้นใยธรรมชาติภายในประเทศเพื่อการต่อเป็นผลิตภัณฑ์ซึ่งช่วยเพิ่มมูลค่า เป็นต้น โดยตัวอย่าง ซึ่งได้จัดทำเป็นข้อมูลสรุป เพื่อการวิเคราะห์ข้อมูลได้สะดวกมากยิ่งขึ้น ดังแสดงรูปที่ 7



รูปที่ 6 : การนำข้อมูลจากระบบเก็บข้อมูลความหลากหลายที่ออกแบบและพัฒนามาวิเคราะห์ร่วมกับแหล่งข้อมูลอื่น ๆ



รูปที่ 7 : Dashboard แสดงข้อมูลสถิติภาพรวมทั้งหมด

จากรูปที่ 7 แสดงผลการวิเคราะห์ข้อมูลด้วย Power BI [20] ซึ่งสามารถนำเข้าชุดข้อมูลที่สนใจเข้ามาแสดงผลร่วมกับรูปภาพหรือสร้างรูปแบบเฉพาะสำหรับการแสดงผล โดยได้นำข้อมูลที่เกี่ยวกับจิ้งหรีดในประเทศไทย แนวโน้มความต้องการของตลาดจากสหรัฐอเมริกา และข้อมูลงานวิจัยของไทยมาวิเคราะห์ร่วมกัน โดยสรุปข้อมูลเป็น 3 ส่วนด้วยกันประกอบไปด้วย

1. ข้อมูลการเลี้ยงจิ้งหรีดในประเทศไทย ซึ่งมาจากการสำรวจและเก็บบันทึกลงในระบบจัดการข้อมูลความหลากหลายซึ่งได้ออกแบบนี้ พบว่ามีการเลี้ยงจิ้งหรีดเป็นจำนวนมากบริเวณภาคตะวันออกเฉียงเหนือ ซึ่งมีพื้นที่การเลี้ยงประมาณ 4.7 ไร่ โดยจังหวัดที่มีการเพาะเลี้ยงจิ้งหรีดมากที่สุดสามลำดับแรก คือ หนองคาย สกลนคร และอุดรธานี ซึ่งรายได้เฉลี่ยของเกษตรที่เพาะเลี้ยงจิ้งหรีดต่อปีอยู่ที่ 103,616 บาท

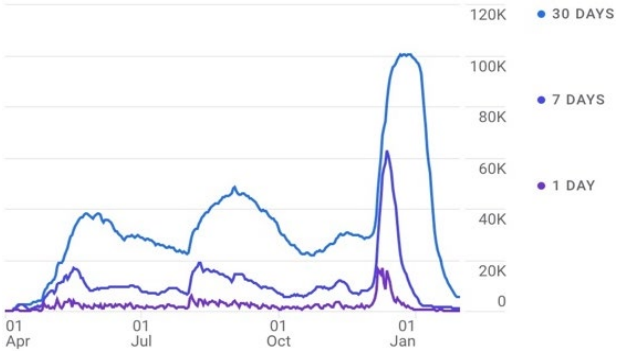
2. ข้อมูลความต้องการของตลาด ซึ่งนำมาจากกรมส่งเสริมการค้าระหว่างประเทศ โดยได้นำข้อมูลแนวโน้มความต้องการสินค้าเกี่ยวกับแมลงจากสหรัฐอเมริกา ตั้งแต่ปี พ.ศ. 2562 ถึง 2566 มาวิเคราะห์ร่วมด้วย พบว่าแนวโน้มสัดส่วนประเภทของผลิตภัณฑ์แปรรูปจากแมลงสามลำดับประกอบด้วย การนำมาทำขนมขบเคี้ยว (snacks) ประมาณร้อยละ 28.40 นำมาทำโปรตีนบาร์ (protein bar) ร้อยละ 29.20 และนำมาทำแป้งจากแมลง (insect protein powder) ประมาณร้อยละ 42.40 โดยมูลค่าทางการตลาดมีแนวโน้มสูงขึ้นตามลำดับ คาดว่าในปี พ.ศ. 2565

มูลค่าผลิตภัณฑ์จากแมลงจะสูงถึง 34 ล้านเหรียญสหรัฐ และในปี พ.ศ. 2566 จะสูงขึ้นไปอีกอยู่ที่ประมาณ 50 ล้านเหรียญสหรัฐ จะเห็นได้ว่าข้อมูลความต้องการของตลาดสามารถช่วยในการตัดสินใจวางแผนการสนับสนุนอุตสาหกรรมการเพิ่มมูลค่าและการแปรรูปแมลงของประเทศไทยเพื่อไปสู่ตลาดโลก

3. การเพิ่มมูลค่าด้วยงานวิจัยไทย จากข้อมูลเกี่ยวกับจิ้งหรีดในประเทศไทย ข้อมูลความต้องการของตลาดของสหรัฐอเมริกาช่วยให้เห็นแนวโน้มความต้องการต่ออุตสาหกรรมการบริโภคแมลง จึงได้นำข้อมูลเกี่ยวกับงานวิจัยของประเทศไทยจาก tech2biz [21] มาวิเคราะห์ร่วมกัน เพื่อนำองค์ความรู้หรือนวัตกรรมจากงานวิจัยทั้งจากหน่วยงานรัฐและเอกชนมาประยุกต์ เพื่อส่งเสริมการทำฟาร์มจิ้งหรีดให้ทันสมัยและได้มาตรฐาน รวมถึงสนับสนุนการแปรรูปเป็นผลิตภัณฑ์ที่หลากหลาย เพื่อเข้าถึงผู้บริโภคได้อย่างกว้างขวาง ช่วยสร้างรายได้ให้กับเกษตรกร ชุมชน ซึ่งส่งผลต่อการพัฒนาเศรษฐกิจของประเทศ

4.3) จำนวนผู้ใช้งานและการรองรับของระบบ

เนื่องจากมีการบันทึกข้อมูลจากการสำรวจชุมชมทั้งประเทศ ซึ่งในบางช่วงเวลาจะพบการเข้าใช้งานระบบบันทึกข้อมูลพร้อมกันดังรูปที่ 8



รูปที่ 8 : จำนวนผู้ใช้บริการระบบจัดเก็บข้อมูลความหลากหลาย

จากรูปที่ 8 แสดงจำนวนผู้ใช้งานที่บันทึกข้อมูลจากการสำรวจของแต่ละพื้นที่ในประเทศไทย ช่วงเวลาประมาณ เมษายน 2564 ถึง มกราคม 2565 พบว่ามีผู้ใช้งานสูงสุดมากกว่า 100,000 คนต่อเดือน โดยระบบจัดการข้อมูลความหลากหลายสามารถให้บริการการจัดเก็บข้อมูล การแสดงผลข้อมูล การสืบค้นข้อมูลได้ตามปกติแม้จะเผชิญกับผู้ใช้ปริมาณมากในเวลาเดียวกัน

5) สรุปผล

การรวมกลุ่มข้อมูลซึ่งมีความหลากหลายของแต่ละชุมชนและจัดเก็บลงระบบดิจิทัลหรือระบบจัดการข้อมูลหลากหลาย ซึ่งรองรับการขยายตัวของข้อมูลซึ่งมีปริมาณเพิ่มขึ้นเรื่อย ๆ และสามารถช่วยประมวลผล ตรวจสอบความถูกต้องของข้อมูล และจัดเก็บข้อมูลอย่างถูกต้อง ส่งผลให้สามารถสืบค้นและนำข้อมูลกลับมาใช้งานได้อย่างครอบคลุม ซึ่งบทความวิจัยฉบับนี้ได้ออกแบบและพัฒนาระบบจัดการข้อมูลหลากหลาย เพื่อจัดเก็บข้อมูลการลงสำรวจชุมชนต่าง ๆ ในประเทศไทย โดยได้บริหารจัดการการจัดเก็บข้อมูลดิบ ข้อมูลชนิดไฟล์ ข้อมูลสำหรับการสืบค้นอย่างรวดเร็ว และได้พัฒนาระบบสำหรับอำนวยความสะดวกในการสร้างโครงสร้างการจัดเก็บข้อมูลตามหมวดหมู่ ระบบสำหรับอำนวยความสะดวกในการบันทึกข้อมูลจากการลงพื้นที่สำรวจ รวมถึงนำเครื่องมือวิเคราะห์ข้อมูลมาช่วยในการติดตามข้อมูลสถิติหรือข้อมูลสรุปสำหรับผู้บริหารจัดการพื้นที่หรือชุมชนในระดับต่าง ๆ ในรูปแบบหลากหลาย

ระบบจัดการข้อมูลความหลากหลายนี้ จึงสามารถช่วยให้ผู้ใช้งานข้อมูล เช่น นักวิจัย นักพัฒนาโปรแกรม นักวิทยาศาสตร์ข้อมูล วิศวกรข้อมูล ผู้บริหารจัดการเมือง สามารถเรียกใช้กลุ่มข้อมูลเหล่านี้ เพื่อนำไปใช้ประโยชน์และช่วยในการตัดสินใจในการแก้ไขปัญหาของเมืองหรือวางแผนพัฒนาทั้งในระดับชุมชน ตำบล อำเภอ จังหวัด และระดับประเทศ อย่างไรก็ตามระบบจัดการข้อมูลความหลากหลายขนาดใหญ่ ถือเป็นรากฐานสำคัญในการบริหารจัดการประเทศในยุคดิจิทัล จึงมีความสำคัญอย่างยิ่งในการพัฒนาต่อยอดให้ระบบจัดการข้อมูลนี้ มีความยืดหยุ่นในการจัดเก็บข้อมูล สามารถขยายฐานการเก็บข้อมูลได้อย่างไม่จำกัด โดยสามารถสืบค้นกลุ่มข้อมูลที่สนใจได้อย่างถูกต้องครอบคลุมและมีความรวดเร็วในการประมวลผลมากยิ่งขึ้น

กิตติกรรมประกาศ

งานวิจัยนี้ ได้รับทุนอุดหนุนจากหน่วยงานบริหารและจัดการทุนด้านการเพิ่มความสามารถในการแข่งขันของประเทศ (บพข.) ภายใต้โครงการ “Quality data platform and Tambol-ERP” คณะวิจัยขอขอบคุณการสนับสนุนจาก บพข. เป็นอย่างยิ่ง

REFERENCES

[1] D. Jacobson, G. Brail, and D. Woods, *APIs: A Strategy Guide*. Sebastopol, CA, USA: O'Reilly Media, 2012.

[2] *Cloud Firestore*. (2023). Firebase. Accessed: Dec. 19, 2023. [Online]. Available: https://firebase.google.com/docs/firestore

[3] *Cloud Storage documentation*. (2023). Google Cloud. Accessed: Dec. 19, 2023. [Online]. Available: https://cloud.google.com/storage/docs

[4] *Elasticsearch Platform — Find real-time answers at scale | Elastic*. (2023). Elasticsearch. Accessed: Dec. 21, 2023. [Online]. Available: https://www.elastic.co

[5] H. S. Munawar, S. Qayyum, F. Ullah, and S. Sepasgozar, "Big data and its applications in smart real estate and the disaster management life cycle: A systematic analysis," *Big Data Cogn. Comput.,* vol. 4, no. 2, 2020, Art. no. 4, doi: 10.3390/bdcc4020004.

[6] J. Vodák, D. Šulyová, and M. Kubina, "Advanced technologies and their use in smart city management," *Sustainability*, vol. 13, no. 10, Jan. 2021, Art. no. 10, doi: 10.3390/su13105746.

[7] M. Ku and J. R. Gil-Garcia, "Ready for data analytics?: Data collection and creation in local governments," in *Proc. 19th Annu. Int. Conf. Digit. Government Res.: Governance in Data Age,* Delft, Netherlands, May 2018, pp. 1–10.

[8] U. Aftab and G. F. Siddiqui, "Big data augmentation with data warehouse: A survey," in *Proc. 2018 IEEE Int. Conf. Big Data,* Seattle, WA, USA, Dec. 2018, pp. 2775–2784.

[9] A. Meier and M. Kaufmann, "NoSQL Databases," in *SQL & NoSQL Databases: Models, Languages, Consistency Options and Architectures for Big Data Management*. Wiesbaden, Germany: Springer Vieweg, 2019, pp. 201–218.

[10] V. M. Ngo, N.-A. Le-Khac, and M-T. Kechadi, "Designing and implementing data warehouse for agricultural big data," in *Proc. 8th Int. Congr., Held as Part of the Services Conf. Federation,* San Diego, CA, USA, Jun. 2019, pp. 1–17, doi: 10.1007/978-3-030-23551-2_1.

[11] H. S. Cha *et al.,* "The Korea cancer big data platform (K-CBP) for cancer research," *Int. J. Environ. Res. Public. Health,* vol. 16, no. 13, Jun. 2019, Art. no. 2290, doi: 10.3390/ijerph16132290.

[12] H. Zheng and X. Jiang, "Design and implementation of college consumption analysis system based on NoSQL database," in *Proc. 13th Int. Conf. Comput. Sci. Educ. (ICCSE),* Colombo, Sri Lanka, Aug. 2018, pp. 1–5.

[13] H. Ganu and P. Viswa Datha, "Fast query expansion on an accounting corpus using sub-word embeddings," in *Proc. 2nd Workshop Subword/Character LEvel Models,* New Orleans, LA, USA, Jun. 2018, pp. 61–65.

[14] Elasticsearch. *Mapping | Elasticsearch Guide [8.12] | Elastic.* (2024). Accessed: Feb. 26, 2024. [Online]. Available: https://www.elastic.co/guide/en/elasticsearch/reference/current/mapping.html

[15] D. Brimley. "What is an Elasticsearch index?." ELASTIC.co. https://www.elastic.co/blog/what-is-an-elasticsearch-index (accessed Feb. 26, 2024).

[16] Elasticsearch. *Thai tokenizer | Elasticsearch Guide [8.11] | Elastic*. (2024). Accessed: Jan. 5, 2024. [Online]. Available: https://www.elastic.co/guide/en/elasticsearch/reference/current/analysis-thai-tokenizer.html

[17] Elasticsearch. *N-gram tokenizer | Elasticsearch Guide [8.11] | Elastic.* (2024). Accessed: Jan. 5, 2024. [Online]. Available: https://www.elastic.co/guide/en/elasticsearch/reference/current/analysis-ngram-tokenizer.html

[18] Elasticsearch. *Autocomplete | Documentation*. (2024). Accessed: Feb. 27, 2024. [Online]. Available: https://docs.elastic.co/search-ui/solutions/ecommerce/autocomplete

[19] *Kibana: Explore, Visualize, Discover Data.* (2020). Elasticsearch. Accessed: Dec. 13, 2020. [Online]. Available: https://www.elastic.co/kibana

[20] *Power BI - Data Visualization | Microsoft Power Platform*. (2024). Microsoft. Accessed: Jan. 5, 2024. [Online]. Available: https://www.microsoft.com/en-us/power-platform/products/power-bi

[21] Tech2Biz. "TechPropose." TECH2BIZ.net. https://www.tech2biz.net/content/inventor (accessed Feb. 27, 2024).