

ระบบรวบรวมและวิเคราะห์ข้อมูลภาครัฐเพื่อการวิเคราะห์ DMC

Government Data Integration and Analysis System: GDIAS

ผู้จัดทำ นายต้นดิน พูลสวัสดิ์
นายณัฐกิตติ จันทรอารี

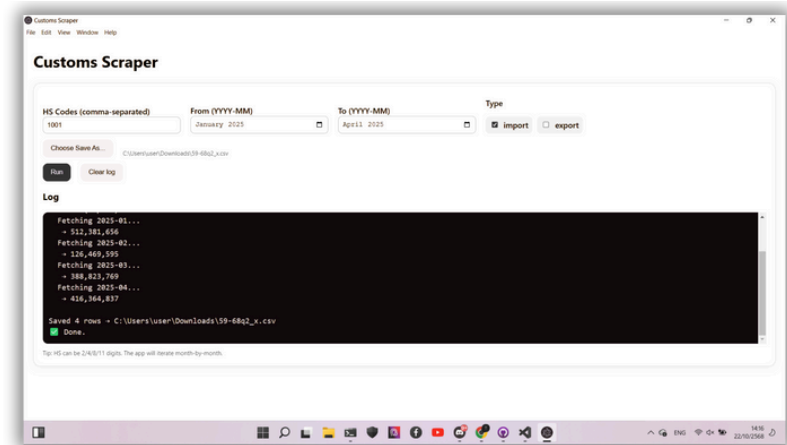
**ผู้ดูแล
ที่ปรึกษา** นายเฉลิมพล ชาญศรีภักย์
นายฤกษ์ณรงค์ พรหมยา

Overview

ระบบรวบรวมและวิเคราะห์ข้อมูลหน่วยงานภาครัฐ เป็นระบบที่ช่วยรวบรวมข้อมูลจากเว็บไซต์ของหน่วยงานภาครัฐ โดยใช้เทคนิค API, HTML Parsing และ Web Scraping ในการดึงข้อมูลจากแหล่งต่าง ๆ จากนั้นโปรแกรมจะนำข้อมูลที่ได้อ่านผ่านกระบวนการวิเคราะห์และทำความสะอาด (Data Cleaning) โดยจะได้เป็นข้อมูลที่เกี่ยวข้องกับการบริโภควัสดุภายในประเทศ (DMC) ในรูปแบบไฟล์ CSV โดยรองรับเว็บไซต์ของกรมต่างๆ ดังนี้

กรมศุลกากร

ดึงข้อมูลจากเว็บไซต์กรมศุลกากรของไทยใช้ Node.js โดยจะอ่านค่ารหัสสินค้า (HS Code) ปี และเดือนจากไฟล์ config.json แล้วส่งคำขอ (POST) ไปยังเว็บศุลกากรเพื่อขอข้อมูลการนำเข้า-ส่งออก จากนั้นใช้ Cheerio ในการอ่าน HTML ที่ได้อ่าน แล้วดึงค่ารวม (Sum) จากแถวสุดท้ายของตารางข้อมูลเมื่อได้ข้อมูลครบทุกเดือนแล้ว โปรแกรมจะบันทึกผลทั้งหมดเป็นไฟล์ CSV ไว้ในตำแหน่งที่กำหนด



START

ดึงและรวมข้อมูล
ทางการเกษตร

ตรวจหาข้อมูลในกลุ่ม
"Production"

END

ทำความสะอาด
ข้อมูล

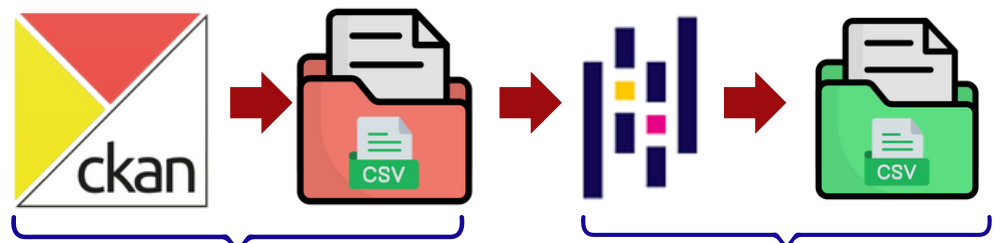
เลือกเฉพาะ
Column ที่อยากได้

สำนักงานเศรษฐกิจการเกษตร

ดำเนินการโดย ดึงและรวมข้อมูลการผลิตทางการเกษตร จากเว็บไซต์กรมส่งเสริมการเกษตร (OAE) ผ่าน API ของ CKAN โปรแกรมจะเริ่มจากเข้าไปดูว่าในกลุ่มข้อมูลชื่อ "production" มีชุดข้อมูล (dataset) อะไรบ้าง จากนั้นมันจะค่อย ๆ ดึงข้อมูลของแต่ละชุดออกมาทีละหน้า แล้วเลือกเฉพาะคอลัมน์ที่สำคัญ เช่น ปี จังหวัด สินค้า ปริมาณ มูลค่า และหน่วย หลังจากนั้นจะทำการ ทำความสะอาดข้อมูล โปรแกรมจะรวมทั้งหมดเป็นไฟล์เดียว

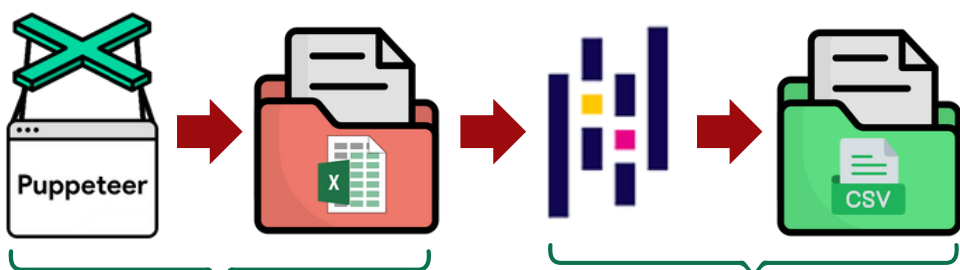
กรมประมง

ดึงข้อมูลสถิติจากระบบ CKAN ของกรมประมงในกลุ่ม import - export โดยจะรวมทุกชุดข้อมูลที่มีคำว่าสถิติไว้ในไฟล์ CSV จากนั้นจะอ่านข้อมูลเพื่อสร้างรายชื่อประเทศทั้งหมด สรุปผลผลิตแยกตามประเทศและชนิดปลา รวมผลผลิตทั้งโลกต่อชนิดปลาและต่อปี แล้วบันทึกไฟล์สรุปทั้งหมด



ขั้นตอนรวบรวมข้อมูล

ขั้นตอนทำความสะอาดข้อมูล



ขั้นตอนรวบรวมข้อมูล

ขั้นตอนทำความสะอาดข้อมูล

กรมพัฒนาพลังงานทดแทนและ อนุรักษ์พลังงาน

ใช้ Node.js (ESM) กับ Puppeteer เพื่อดาวน์โหลดไฟล์ XLSX รายปีจากเว็บไซต์ dede.go.th เฉพาะหน้า annual โดยข้ามไฟล์รายเดือน/ครึ่งปี และเน้นไฟล์ประเภท "Physical" ถ้ามี ไฟล์จะถูกบันทึกตามโครงสร้าง downloads/<product>/<year>/ พร้อมเก็บชื่อไฟล์ต้นฉบับและข้อมูลปีไว้ใน manifest.json เพื่อป้องกันดาวน์โหลดซ้ำ TLS ของ dede.go.th ถูกตั้ง relaxed เฉพาะโดเมนนี้ ส่วนฟังก์ชันช่วยเหลือนจะช่วยแยกปี ตรวจสอบช่วงเวลา และจัดการไฟล์อย่างปลอดภัย

