

รายงานการฝึกงาน

KKU smart city(scope)

โดย

นายวรปรัชญ์ โลมาอินทร์ รหัส 623040486-2

รายงานนี้เป็นส่วนหนึ่งของการฝึกงาน สาขาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ มหาวิทยาลัยขอนแก่น

วันที่ 20 มิถุนายน 2565

สารบัญ

หัวข้อ	หน้า
ลักษณะงานองค์กร	1
งานที่ทำหรือได้รับมอบหมายในระหว่างฝึกงาน	1
ปัญหาระหว่างฝึก	7
สรุป	8

1. ลักษณะงานองค์กร

สถานที่ฝึกงาน: ศูนย์ scopc อาคาร north eastern science park

สถานที่ตั้ง: เลขที่ 123 หมู่ที่ 16 ถนนกัลปพฤกษ์ ตำบลในเมือง อำเภอเมือง จังหวัดขอนแก่น 40002

ลักษณะงาน: ศูนย์ scopc เป็นศูนย์ข้อมูล เพื่อบริการให้หน่วยงานต่างๆ และจัดทำกราฟวิเคราะห์ข้อมูลจากหน่วยงานต่างๆ ทั้งเอกชน หรือ ภาครัฐ อีกทั้งยังพัฒนาระบบ สิ่งอำนวยความสะดวกต่างๆ ให้กับมหาวิทยาลัยขอนแก่น

2. งานที่ทำหรือได้รับมอบหมายในระหว่างฝึกงาน

งานที่ได้รับมอบหมายในการฝึกงานภาคการศึกษาพิเศษ 2565 ทาง ศูนย์ scopc ได้มอบหมายงานที่เกี่ยวข้องกับ data analysis โดยโจทย์ของงาน คือ เนื่องจากมหาวิทยาลัยขอนแก่นมีบุคลากรที่ผลงานตีพิมพ์ทางวิชาการมากมาย และ กระจายอยู่เว็บไซต์ต่างๆ google scholar เป็นหนึ่งในเว็บไซต์ที่เก็บข้อมูลเหล่านี้ จึงได้มอบหมายในการดึงข้อมูลต่างๆ ที่เกี่ยวข้องบน google scholar ออกมาเก็บไว้ใน database และนำมาวิเคราะห์

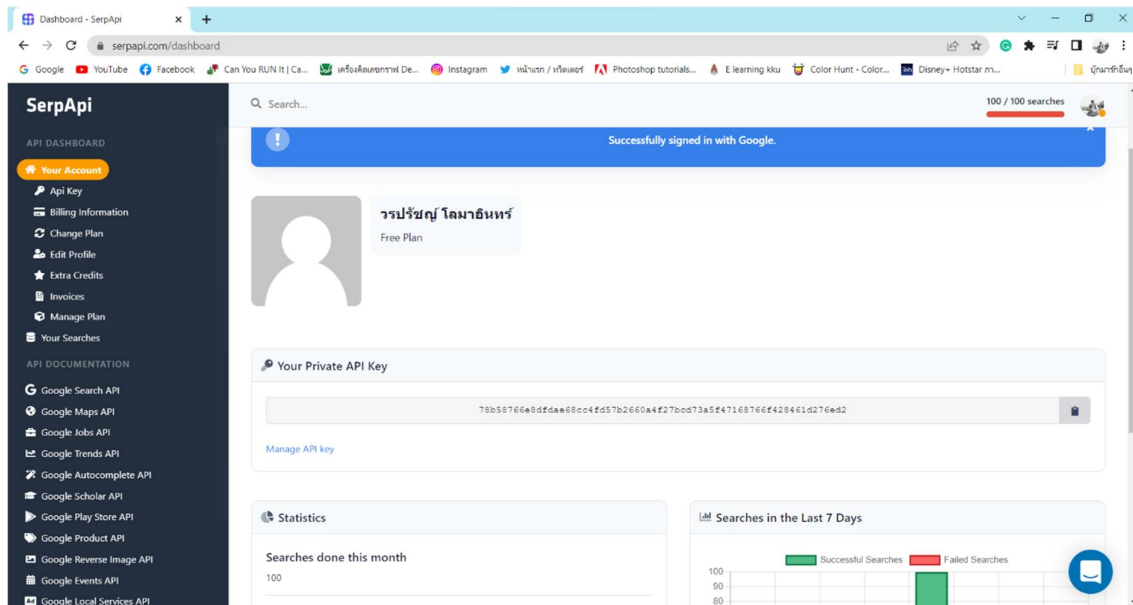
โดยมีเครื่องมือที่ใช้ดังนี้

- Microsoft visual studio
 - Python
 - SQL
- Microsoft power bi
- PostgreSQL
- SerbAPI

2.1 ขั้นตอนที่ 1 - การใช้งาน API

เครื่องมือที่ใช้: SerbAPI

รายละเอียด: การใช้งาน API ของ google scholar นั้น ทาง google เองไม่ได้แจก API มาให้ทางผู้พัฒนามาใช้งานจึงมีความจำเป็นที่จะต้องไปใช้เว็บไซต์ SerbAPI สร้าง api key ของ google scholar ขึ้นมาแทน เพื่อนำมาใช้ในการดึงข้อมูลต่อไป

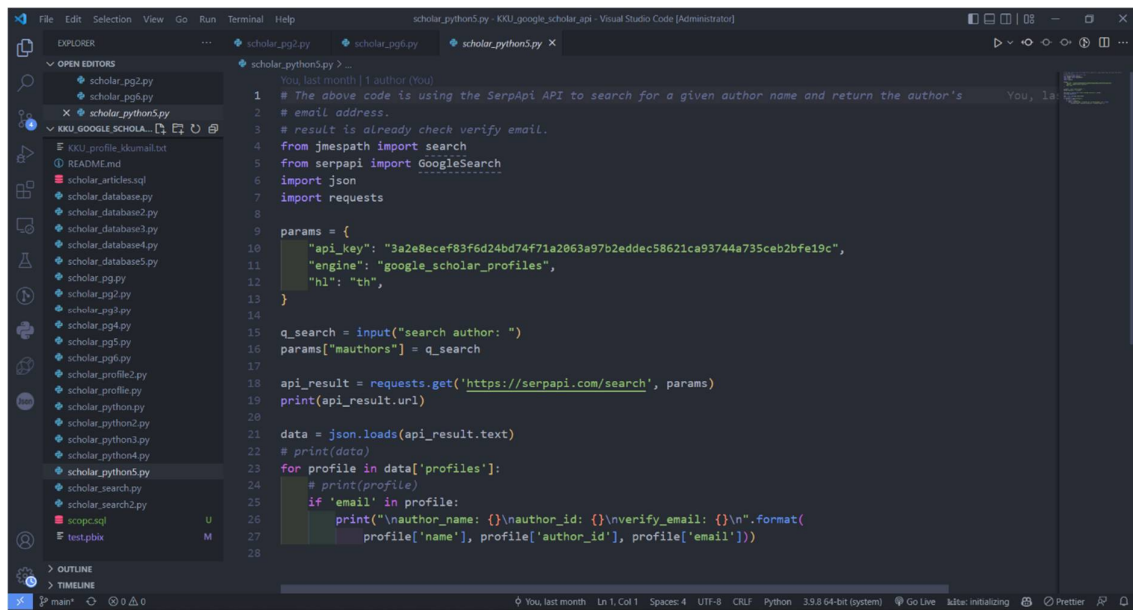


รูปที่ 1 หน้าเว็บไซต์ SerbAPI

2.2 ขั้นที่ 2 - เขียนโปรแกรมเพื่อดึงข้อมูล

เครื่องมือที่ใช้: Microsoft visual studio, python language

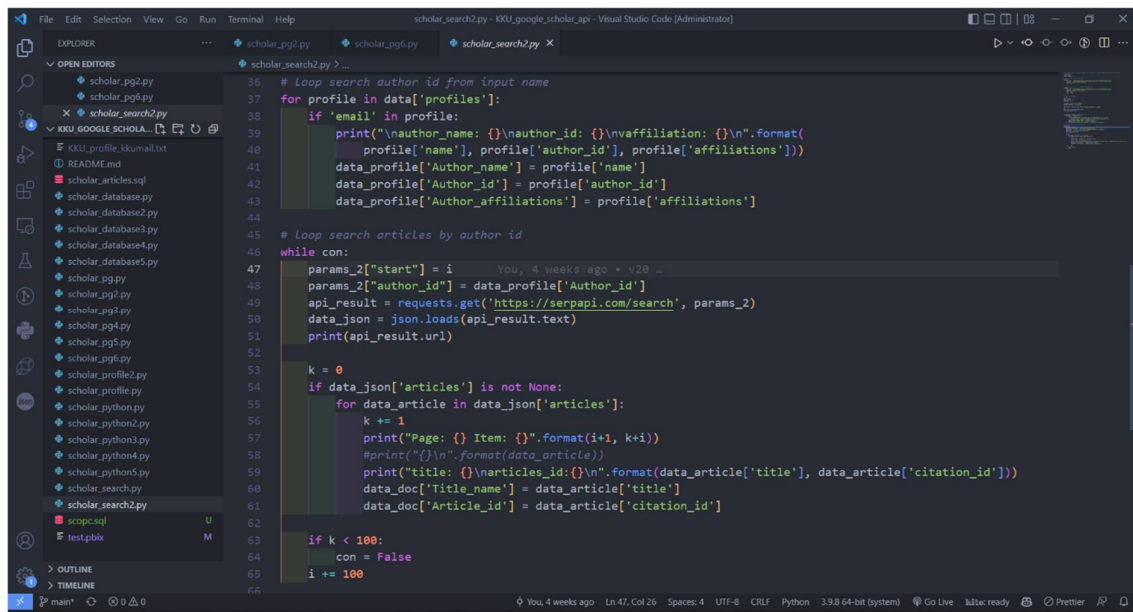
รายละเอียด: เมื่อมี api key ของ google scholar ก็สามารถดึงข้อมูลมาใช้งานได้ โดยการเขียนโปรแกรมครั้งนี้ใช้ภาษา python และเขียนบน Microsoft visual studio ด้วยการดึงข้อมูลไม่ได้เป็นการดึงข้อมูลจากฐานข้อมูลโดยตรง แต่ดึงจากหลายๆหน้า จึงเริ่มทดสอบการดึงข้อมูล แล้วจึงดึงข้อมูลเพิ่มมากขึ้น ในโปรแกรมนี้เริ่มจากการดึง author_id จากการรับรายชื่อจากโปรแกรม และส่งค่ากลับมาเป็น author_id



รูปที่ 2 ตัวอย่างการดึงข้อมูล author_id

และเมื่อมี author_id แล้วก็จะสามารถนำไปค้นหาโปรไฟล์ใน google scholar ได้เพื่อที่จะดึงรายชื่อผลงานตีพิมพ์พร้อมรายละเอียดอื่นๆ อีกที

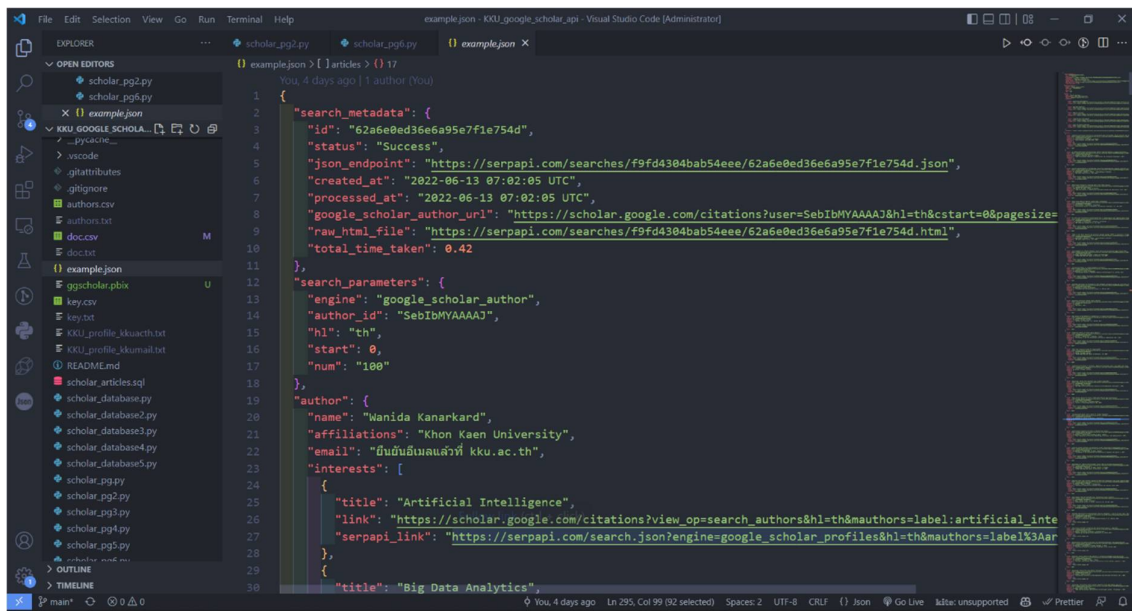
หลังจากสามารถดึงข้อมูลได้แล้วจึงทำการเขียนโปรแกรมที่ดึงข้อมูล author_id และนำไปค้นหาโปรไฟล์ ของบุคคลนั้นๆ แล้วจึงดึงข้อมูลในส่วนของ ผลงานตีพิมพ์พร้อมรายละเอียดต่างๆออกมา



```
36 # Loop search author id from input name
37 for profile in data['profiles']:
38     if 'email' in profile:
39         print("\nauthor_name: {}\nauthor_id: {}\nvaaffiliation: {}".format(
40             profile['name'], profile['author_id'], profile['affiliations']))
41         data_profile['Author_name'] = profile['name']
42         data_profile['Author_id'] = profile['author_id']
43         data_profile['Author_affiliations'] = profile['affiliations']
44
45 # Loop search articles by author id
46 while con:
47     params_2["start"] = i
48     params_2["author_id"] = data_profile['Author_id']
49     api_result = requests.get('https://serpapi.com/search', params_2)
50     data_json = json.loads(api_result.text)
51     print(api_result.url)
52
53     k = 0
54     if data_json['articles'] is not None:
55         for data_article in data_json['articles']:
56             k += 1
57             print("Page: {} Item: {}".format(i+1, k+i))
58             #print("{}\n".format(data_article))
59             print("title: {}\narticles_id: {}".format(data_article['title'], data_article['citation_id']))
60             data_doc['Title_name'] = data_article['title']
61             data_doc['Article_id'] = data_article['citation_id']
62
63     if k < 100:
64         con = False
65     i += 100
66
```

รูปที่ 3 โปรแกรมการดึงข้อมูลชื่อผลงานและอื่นๆ จาก author_id

ข้อมูลที่ได้มาจะมาเป็นรูปแบบของไฟล์ JSON จึงต้องนำมารองเอาเฉพาะข้อมูลที่ต้องการและเก็บใน directory ใหม่และแสดงผลออกมาเพื่อเช็คความถูกต้องข้อมูลก่อนจะนำไปบันทึกลงฐานข้อมูลต่อไป



รูปที่ 4 ไฟล์ข้อมูลต้นทางที่เป็นประเภท JSON File

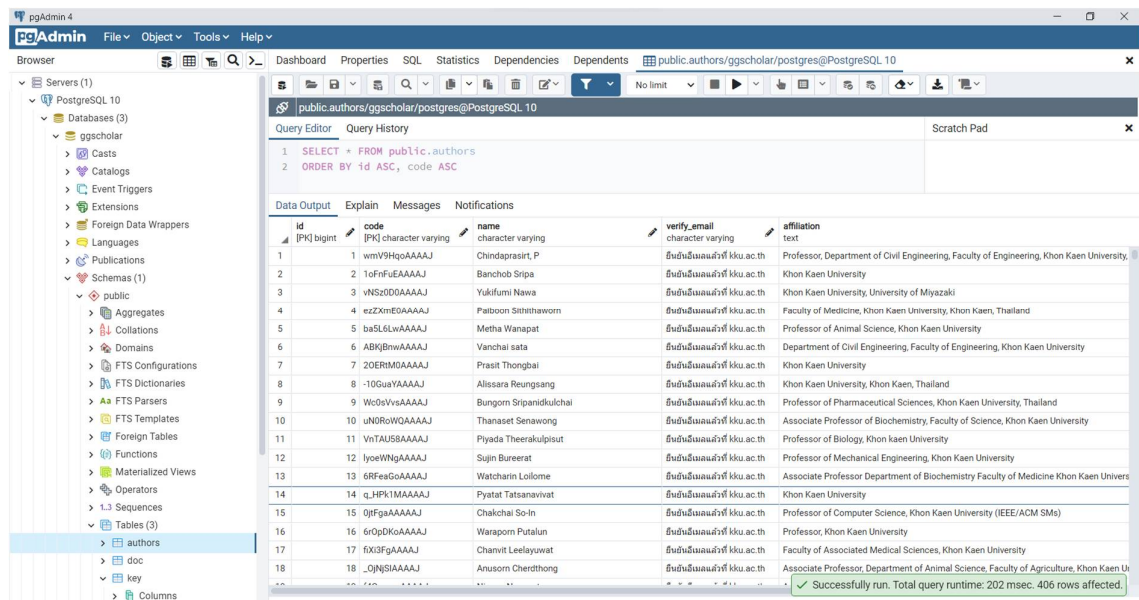
2.3 ขั้นตอนที่ 3 – การทำฐานข้อมูล

เครื่องมือที่ใช้: PostgreSQL, SQL language

รายละเอียด: หลังจากทำทดสอบโปรแกรมว่าทำงานถูกต้องเรียบร้อยแล้ว ต่อไปจะนำข้อมูล

นั้นมาเขียนลงบนฐานข้อมูล โดย เครื่องที่ใช้เป็น PostgreSQL เป็นตัวช่วยในการทำ

ฐานข้อมูล



รูปที่ 5 GUI ของโปรแกรม PostgreSQL

เมื่อทำการสร้างฐานข้อมูลเรียบร้อยแล้วก็ทำการสร้างตารางเพื่อจัดเก็บข้อมูล ซึ่งในส่วน
ของตารางนั้นจะแบ่งเป็น 3 ตาราง ได้แก่ authors, doc และ key เพื่อที่จะทำให้การจัดการ
ฐานข้อมูลทำได้ง่าย หลังจากนั้นได้เขียน โปรแกรมเพื่อที่จะนำข้อมูลจากการค้นหา มาเขียน
บันทึกลงฐานข้อมูล โดยใช้ภาษา python และ SQL การร่วมกันทำงาน

```

File Edit Selection View Go Run Terminal Help
scholar_pg6.py - KKU_google_scholar_api - Visual Studio Code [Administrator]

EXPLORER
scholar_pg2.py
scholar_pg6.py x example.json

OPEN EDITORS
scholar_pg6.py
scholar_pg2.py
X scholar_pg6.py
example.json
KKU GOOGLE SCHOLAR
doc.csv
doc.txt
example.json
ggscholar.pbix
key.csv
key.txt
KKU_profile_kkuacth.txt
KKU_profile_kkumail.txt
README.md
scholar_articles.sql
scholar_database.py
scholar_database2.py
scholar_database3.py
scholar_database4.py
scholar_database5.py
scholar_pg.py
scholar_pg2.py
scholar_pg3.py
scholar_pg4.py
scholar_pg5.py
scholar_pg6.py
scholar_profile2.py
scholar_profile.py
scholar_python.py
scholar_python2.py
scholar_python3.py
OUTLINE
TIMELINE

params_2['author_id'] = row[1]
api_result = requests.get('https://serpapi.com/search', params_2)
data_json = json.loads(api_result.text)

k = 0
if data_json['articles'] is not None:
    for articles in data_json['articles']:
        k += 1
        print("Page: {} Item: {}".format(i+1, k+i))

if 'publication' in articles:
    data_article['Publication'] = articles['publication'].replace("'", " ")
else:
    data_article['Publication'] = "None"
data_article['Link'] = articles['link']
data_article['Value'] = str(articles['cited_by']['value'])
data_article['Citation_id'] = articles['citation_id']

script_insert = "UPDATE doc SET publication = \'\" + \
    data_article['Publication'] + \"\' WHERE code = \'\" + \
    data_article['Citation_id'] + \"\'"

cur2.execute(script_insert)
conn2.commit()

script_insert2 = "UPDATE doc SET cited_by_value = \'\" + \
    data_article['Value'] + \"\' WHERE code = \'\" + \
    data_article['Citation_id'] + \"\'"

cur2.execute(script_insert2)
conn2.commit()

```

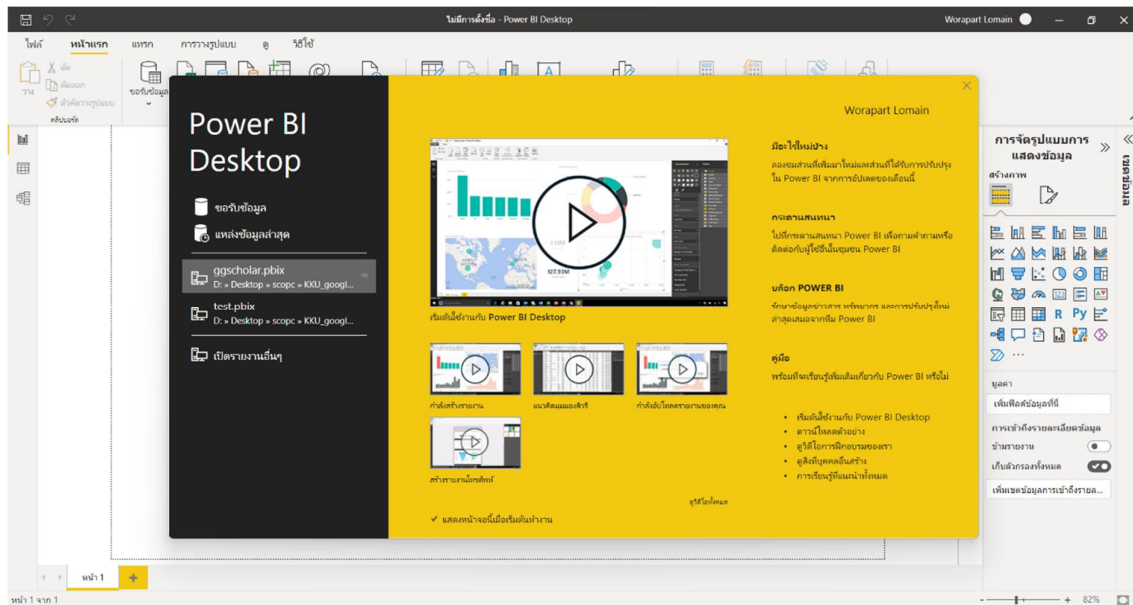
รูปที่ 6 ตัวอย่างการนำข้อมูลบันทึกคลังฐานข้อมูล

ข้อมูลที่ได้ทั้งหมดจะไปเก็บตาม column ต่างๆที่ได้ระบุไว้ในตารางฐานข้อมูล

2.4 ขั้นตอนที่ 4 - ตีออกข้อมูลมาวิเคราะห์

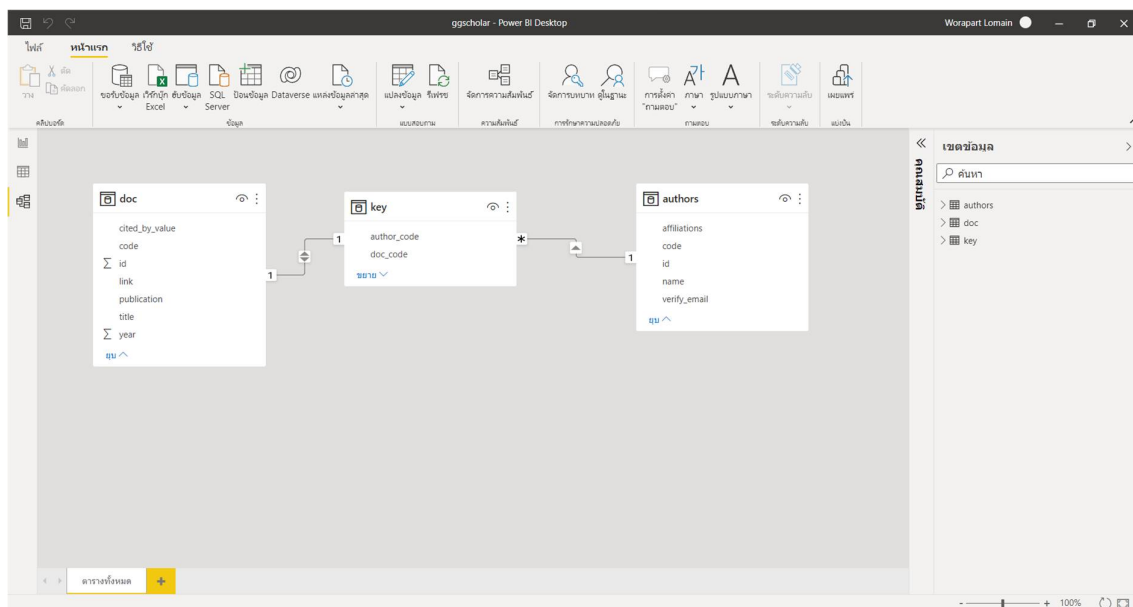
เครื่องมือที่ใช้: Microsoft power bi

รายละเอียด: เมื่อมีข้อมูลพร้อมในฐานข้อมูลแล้ว จึงสามารถส่งออกข้อมูล โดยแปลงเป็นไฟล์นามสกุล csv แล้วนำไปเข้าโปรแกรม Microsoft power bi เพื่อสร้าง dashboard สำหรับการแสดงผลและวิเคราะห์ที่ได้ทำการค้นหามา



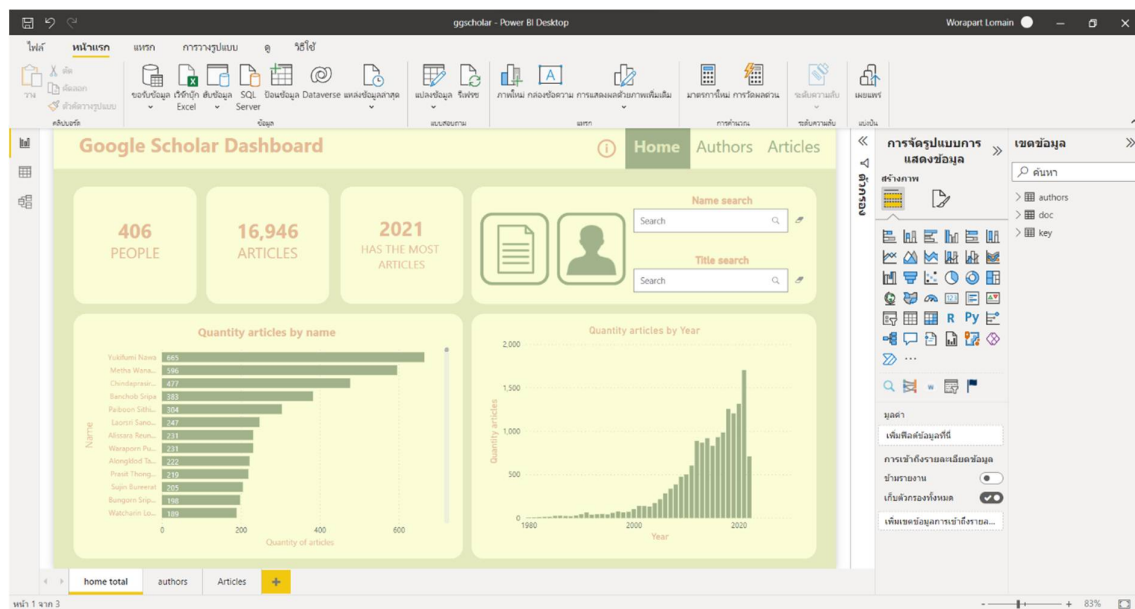
รูปที่ 7 โปรแกรม power bi

เมื่อนำข้อมูลเข้ามาแล้วจึงทำการเชื่อมข้อมูล และจัดการข้อมูลที่ไม่เรียบร้อย แล้วทำการสร้างกราฟจากข้อมูลที่ได้นำเข้า



รูปที่ 8 หน้าต่างการเชื่อมต่อข้อมูล

แล้วนำมาจัดหน้า dashboard เพื่อนำเสนอข้อมูล ตัว power bi สามารถสร้างการแสดงผลได้มากกว่ากราฟ และผลลัพธ์จะได้ประมานนี้



รูปที่ 9 หน้า dashboard นำเสนอข้อมูล

3. ปัญหาด้านเครื่องมือที่ใช้งาน

3.1 ปัญหาด้านเครื่องมือที่ใช้งาน

- SerbAPI มีข้อจำกัดในการใช้งาน และต้องจ่ายเงินเพิ่มเติมทำให้ต้องสมัครหลายอีเมลมาทดแทน
- การดึงข้อมูลของไฟล์ JSON เนื่องจากเป็นครั้งแรกในการทำงานร่วมกับไฟล์ JSON จึงต้องศึกษาเพิ่มเติม ในเรื่องของโครงสร้างข้อมูล
- library python ที่ใช้งานร่วมกับโปรเจกต์นี้ มีการลงที่เกิดการผิดพลาด จึงใช้เวลานานในการแก้ไข
- ภาษา SQL เนื่องจากทำงานใช้ภาษานี้เป็นครั้งแรกจึงต้องใช้เวลาในการศึกษา จึงจะสามารถนำมาทำงานจึงได้
- การใช้งาน PostgreSQL เป็นครั้งแรกที่ต้องงานกับฐานข้อมูลจึงต้องเรียนรู้เพิ่มเติมและใช้เวลาในการทำความเข้าใจ

3.2 ปัญหาด้านข้อมูล

- ข้อมูลบางส่วนไม่ถูกระบุมา (ไม่เป็น null และไม่มีข้อมูลนั้นเลย) จึงทำให้โปรแกรมเกิดความผิดพลาดบ่อยครั้ง
- ข้อมูลที่เป็น null เป็นข้อมูลที่ไม่สามารถนำมาวิเคราะห์ได้ซึ่งมีเยอะพอสมควร
- ข้อมูลที่มีหลายภาษา ทำให้การส่งออกข้อมูลมีปัญหาในการเข้ารหัสไฟล์ ทำให้แสดงผลไม่ถูกต้อง
- เครื่องหมายที่ทับซ้อน ข้อมูลบางตัวเมื่อนำส่งต่อไปยังฐานข้อมูล จะทำให้ติดภาษา SQL และเครื่องหมายที่เป็นปัญหาคือ ' (single quote)

4. สรุป

จากการฝึกงานที่ศูนย์ scope ที่ตั้งอยู่ภายในตึกของ north eastern sci park ตั้งแต่วันที่ 2 พฤษภาคม 2565 จนถึง วันที่ 15 มิถุนายน 2565 ที่ศูนย์มีการอบรมให้ความรู้มากมายและมีผลงานที่เกี่ยวข้องกับ มหาวิทยาลัยขอนแก่น และ เมืองขอนแก่น ซึ่งเป็นการสร้างแรงบันดาลใจในการเข้าร่วมฝึกงานเป็นอย่างมาก จนกระทั่งได้รับมอบหมายงาน ที่เป็นงานเกี่ยวข้องกับ data analysis ของ google scholar โดยดึงข้อมูลที่เกี่ยวข้องกับบุคลากรในมหาวิทยาลัยขอนแก่น มาเก็บไว้ใน database และวิเคราะห์ต่ด้วย Microsoft power bi ที่สามารถนำไปต่อยอดได้

นอกจากนั้นยังระหว่างฝึกงานยังได้ความรู้ทางด้านอื่นๆเพิ่มเติม เช่น การถ่ายภาพ 3 มิติ การทำ metaverse และความรู้เรื่อง ai เป็นต้น และการลงมือปฏิบัติงานจริงมีพี่เลี้ยงคอยช่วยเหลือตลอดทั้งให้ความรู้ คอยแนะนำการใช้เครื่องมือ รวมถึงการ coding