



Project: การดึงข้อมูลผลงานทางวิชาการ  
ของอาจารย์มหาวิทยาลัยธรรมศาสตร์จากเว็บไซต์ Google Scholar  
โดยใช้กระบวนการ web scraping

เสนอ

ผศ. ดร. วสิค ลิมประเสริฐ

จัดทำโดย

นายณัฐชัย นิ่งน้อย เลขทะเบียนนักศึกษา 6109656311

รายงานนี้เป็นส่วนหนึ่งของรายวิชา  
วท. 321 โครงสร้างพื้นฐานคอมพิวเตอร์สำหรับการประมวลข้อมูลขนาดใหญ่  
สาขาวิชาวิทยาศาสตร์และนวัตกรรมข้อมูล  
วิทยาลัยสหวิทยาการ มหาวิทยาลัยธรรมศาสตร์  
ชั้นปีที่ 3 ภาคเรียนที่ 2 ปีการศึกษา 2563

## 1. รายละเอียดข้อมูล

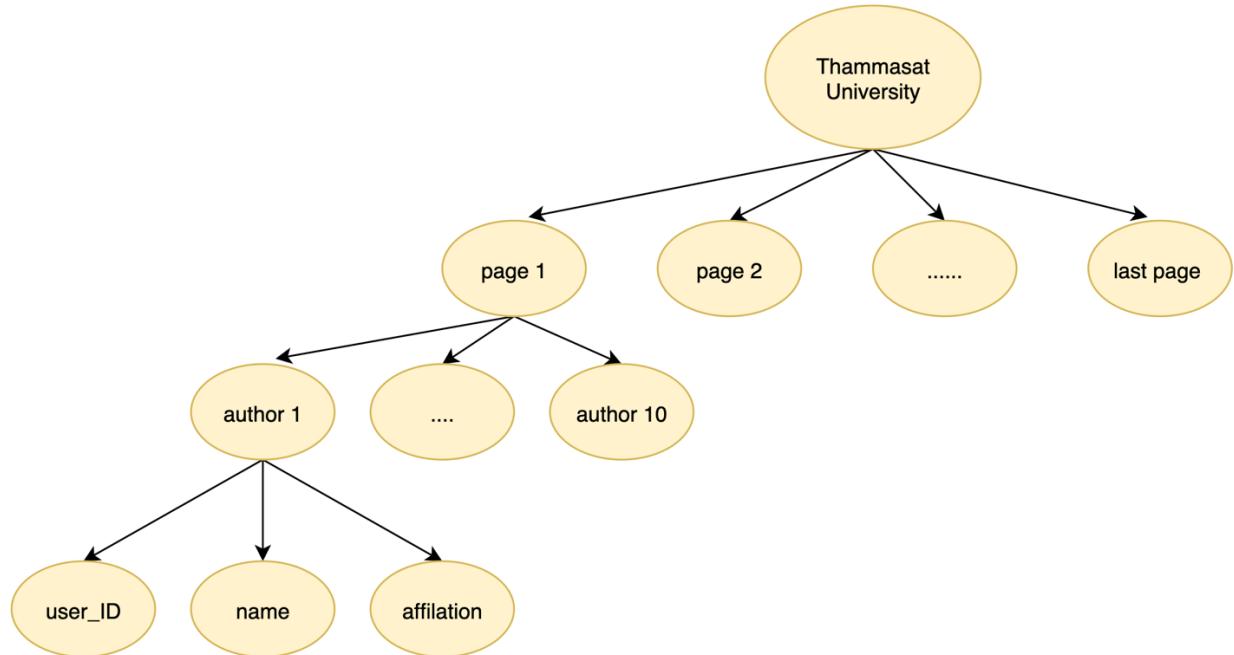
### 1.1. URL ที่ใช้คือในการดึงข้อมูลจาก Google Scholar

[https://scholar.google.com/citations?view\\_op=view\\_org&org=10241031385301082500&hl=en&oi=io](https://scholar.google.com/citations?view_op=view_org&org=10241031385301082500&hl=en&oi=io)

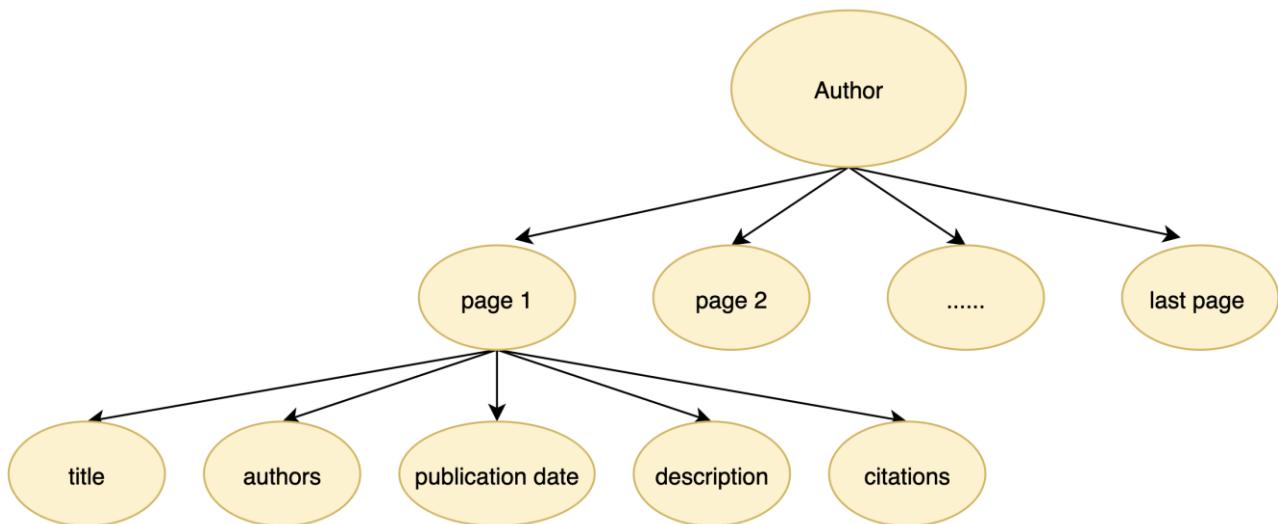
```
driver.get("https://scholar.google.com/citations?view_op=view_org&org=10241031385301082500&hl=en&oi=io")
```

### 1.2. Data structure (tree data structure)

tree data structure ของ author table



tree data structure ของ paper table



โดยข้อมูลที่เราต้องการเก็บมีรูปแบบดังนี้ คือ

Sheet ที่ 1 คือ Author Table โดยมีรูปแบบดังนี้คือ

- user\_ID ใช้ในการเก็บรหัสของ Author
- name ใช้ในการเก็บชื่อของ Author
- affiliation หรือ สังกัด ใช้ในการเก็บข้อมูลเกี่ยวกับคณะหรือสังกัดของ Author

**Author Table**

user_ID	name	affiliation
N5APA98AAAAJ	Wasit Limprasert	Thammasat university
...		

Sheet ที่ 2 คือ Paper Table โดยข้อมูลที่เราต้องการมีรูปแบบดังนี้คือ

- title ใช้ในการเก็บชื่อ paper
- authors ใช้ในการเก็บรายชื่อผู้จัดทำ paper
- publication\_date ใช้ในการเก็บ วันเวลา ที่เผยแพร่ paper นั้น ๆ
- description ใช้ในการเก็บคำอธิบายของ paper นั้น ๆ
- cite\_by ใช้ในการเก็บจำนวนคนที่เข้ามาอ่าน paper

**Paper Table**

title	authors	publication_date	description	cite_by
Real-time people tracking in a camera network	Wasit Limprasert, Andrew Wallace, Greg Michaelson	2013/4/24	We present an approach to track several subjects from video sequences acquired by multiple cameras in real time. We address the key concerns of real time performance and continuity of tracking in overlapping and nonoverlapping fields of view...	19
...				

## ผลลัพธ์ที่ได้

### Sheet ที่ 1 คือ Author Table

Nattachan\_Ningnoi\_6109656311\_Google-Scholar-Assignment

B1	A	B	C	D	E	F
	user_ID	name	affiliation			
1	Re819VUAAAAAJ	sandhya Babel	Professor, Sirindhorn International Institute of Technology, Thammasat University			
2	97Zz_TIAAAAJ	Sombat Muengtaweepongsa	Thammasat University			
3	u-vl_aIAAAAJ	Wutiphol Sintunavarat	Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat ...			
4	9bUei6wAAAAAJ	Bunyari Uyyanonvara	Associated Professor, SIIT, Thammasat University			
5	UOeuXvQAAAAAJ	Chanathip Namprempre	Assistant Professor, Thammasat University			
6	82GgezQAAAAAJ	Prof.Tritos Laosirhongthong	Department of Industrial Engineering, Faculty of Engineering, Thammasat University ...			
7	YMLTjQAAAAAJ	Kriengsak Panuwatwanich	Associate Professor, Sirindhorn International Institute of Technology, Thammasat University			
8	PLCJne4AAAAAJ	Sakun Boon-lt	Professor of Operations Management, Thammasat Business School			
9	yGhz_1oAAAAAJ	Juthathip Jongwanich	Faculty of Economics, Thammasat University			
10	I2IBB5YAAAAAJ	Kritsadarat Wattanasuwan, DPhil (Oxon)	Thammasat Business School, Thammasat University, Bangkok Thailand			
11	debyzkIAAAAJ	Bundit Limmeechokchai	Sirindhorn International Institute of Technology, THAMMASAT University			
12	JGJNbJIAAAAJ	Ruth Banomyong	Thammasat Business School, Thammasat University			
13	UE8D-AAAAAJ	Archanan Kophaloobon	Faculty of Economics, Thammasat University			
14	aVwwa_0AAAAAJ	Nurak Gridanurak	Chemical Engineering Dept., Thammasat University			
15	91yGxx0AAAAAJ	Paiboon Sreearunothai	Sirindhorn International Institute of Technology			
16	HD7ISIUAJJ	Chanatip Samart	Thammasat University			
17	O4WHKcYAAAAAJ	Prof. Dr. Chaoasuan Kanchanomai	Department of Mechanical Engineering, Thammasat University, Thailand			
18	SMJ2150AAAAAJ	Stanislav Makhanov	Professor Sirindhorn International Institute of Technology, Thammasat University			
19	QL5hx5gAAAAAJ	Kongkiat Kulkantarakorn	Thammasat University			
20	I5F9jrQAAAAAJ	Thanaruk Theeramunkong	Sirindhorn International Institute of Technology			
21	debyzkIAAAAJ	Bundit Limmeechokchai	Sirindhorn International Institute of Technology, THAMMASAT University			
22	JGJNbJIAAAAJ	Ruth Banomyong	Thammasat Business School, Thammasat University			
23	UE8D-AAAAAJ	Archanan Kophaloobon	Faculty of Economics, Thammasat University			
24	aVwwa_0AAAAAJ	Nurak Gridanurak	Chemical Engineering Dept., Thammasat University			
25	91yGxx0AAAAAJ	Paiboon Sreearunothai	Sirindhorn International Institute of Technology			
26	HD7ISIUAJJ	Chanatip Samart	Thammasat University			
27	O4WHKcYAAAAAJ	Prof. Dr. Chaoasuan Kanchanomai	Department of Mechanical Engineering, Thammasat University, Thailand			
28	SMJ2150AAAAAJ	Stanislav Makhanov	Professor Sirindhorn International Institute of Technology, Thammasat University			

### Sheet ที่ 2 คือ Paper Table

Nattachan\_Ningnoi\_6109656311\_Google-Scholar-Assignment

B2	A	B	C	D	E	F
	title	authors	publication_date	description	cite_by	
1	Low-cost adsorbents for heavy metals uptake from contaminated water: a review	Sandhya Babel, Tonni Agustiono Kurniawan	2003/2/28	In this article, the technical feasibility of v	4013	
2	Physico-chemical treatment techniques for wastewater laden with heavy metals	Tonni Agustiono Kurniawan, Gilbert YS Chan, Wai-Hung L	2006/5/1	This article reviews the technical applica	1973	
3	Cr (VI) removal from synthetic wastewater using coconut shell charcoal and commercial a	Sandhya Babel, Tonni Agustiono Kurniawan	2004/2/1	In this study, the technical feasibility of cr	948	
4	Comparisons of low-cost adsorbents for treating wastewaters laden with heavy metals	Tonni Agustiono Kurniawan, Gilbert YS Chan, Wai-Hung L	2006/8/1	In this article, the removal performance e	804	
5	Equilibrium and kinetic modelling of biosorption of Remazol Black B by Rhizopus arrhizus	Zümriye Aksu, Sevilay Tezer	2000/12/1	The biosorption of Remazol Black B, a vi	584	
6	Heavy metal removal from contaminated sludge for land application: a review	Sandhya Babel, Dominica del Mundo Dacera	2006/1/1	In recent years, various methods for hea	340	
7	Microfiltration membrane fouling and cake behavior during algal filtration	Sandhya Babel, Satoshi Takizawa	2010/10/15	This study was carried out to investigate	169	
8	A matrix in life cycle perspective for selecting sustainable materials for buildings in Sri Lan	Sandhya Babel, Shabbir Ghee	2009/5/1	This paper presents a matrix to select su	152	
9	Effect of acid speciation on solid waste liquefaction in an anaerobic acid digester	Sandhya Babel, Kensuke Fukushi, Bunpot Sitanrassamee	2004/5/1	Laboratory scale experiments were carri	143	
10	Factors affecting seasonal variation of membrane filtration resistance caused by Chlorella	Sandhya Babel, Satoshi Takizawa, Hiroaki Ozaki	2002/3/1	A seasonal fluctuation pattern was obser	113	
11	Removal of Cr from synthetic wastewater by sorption into volcanic ash soil	Sandhya Babel, Einstine M Opiso	2007/12/1	The possibility of using volcanic ash soli	108	
12	Anaerobic co-digestion of sewage and brewery sludge for biogas production and land app	Sandhya Babel, Jenirja Sae-Tang, Athipol Pecharaply	2009/12/1	In Thailand, sewage sludge production fr	91	
13	Rapid enhanced photocatalytic degradation of dyes using novel N-doped ZrO2	Hangara Sudrajat, Sandhya Babel, Hiroshi Sakai, Satoshi	2016/1/1	A novel N-doped ZrO2 (N-ZrO2) photoc	82	
14	Procion Green H-4G immobilized poly (hydroxyethylmethacrylate/chitosan) composite mei Ô Genç, L Soysal, G Bayramoğlu, MY Arica, S Bektaş		2003/2/28	The effective removal of toxic heavy met	76	
15	Various treatment technologies to remove arsenic and mercury from contaminated ground S Babel, TA Kurniawan		2005/12/1	Contamination of groundwater due to tox	68	
16	Chemical pretreatment for reduction of membrane fouling caused by algae	Sandhya Babel, Satoshi Takizawa	2011/7/1	Batch filtration experiments in dead-end	64	
17	Sustainable utilization of waste palm oil and sulfonated carbon catalyst derived from cocor Indika Thushari, Sandhya Babel		2018/1/1	In this study, an inexpensive, environme	63	
18	A research study on Cr (VI) removal from contaminated wastewater using low-cost adsorb TA Kurniawan, S Babel		2003/2	In this study, the technical feasibility of u	63	
19	Survival of Salmonella spp. in a simulated acid-phase anaerobic digester treating sewage K Fukushi, Sandhya Babel, S Burakrai		2003/1/1	The presence of pathogenic microorgani	62	
20	Removal of heavy metals from contaminated sewage sludge using Aspergillus niger ferme Dominic Del Mundo Dacera, Sandhya Babel		2008/4/1	The environmental benefits derived from	60	
21	Nitrogen, phosphorus and silicon uptake kinetics by marine diatom Chaetoceros calcitrans Chayarat Tantrasarit, Andrew J Englands, Sandhya Babel		2013/8/1	Nitrogen, phosphorus, and silicon uptake	54	
22	Ultrasonic pretreatment of palm oil mill effluent: Impact on biohydrogen production, biolext Emmanuel P Leano, Alfredo J Anceno, Sandhya Babel		2012/9/1	Substrate bioavailability is one of the criti	51	
23	Effects of pretreatment methods on cassava wastewater for biohydrogen production optim EP Leano, Sandhya Babel		2012/3/1	Batch production of biohydrogen from ca	49	
24	Potential for land application of contaminated sewage sludge treated with fermented liquid Dominic Del Mundo Dacera, Sandhya Babel, Preeda Par		2009/8/15	The suitability for land application of ana	46	
25	Environmental, economic and social analysis of materials for doors and windows in Sri Lai UG Yasantha Abeyasundra, Sandhya Babel, Shabbir Ghee		2007/5/1	This paper compares the environmental,	44	
26	Rapid photocatalytic degradation of the recalcitrant dye amaranth by highly active N-WO : Hangara Sudrajat, Sandhya Babel		2016/6	Dye wastewater is a major source of toxi	43	
27	Use of citric acid for heavy metals extraction from contaminated sewage sludge for land a) Dominic Del Mundo Dacera, Sandhya Babel		2006/11	Recent studies revealed that organic acti	43	
28	Treatment of natural rubber processing wastewater by combination of ozonation and activ Sandhya Babel, Nonglak Rungruang		2008	Most of latex concentrate processing fac	39	

### Sheet ที่ 3 คือ Paper Table ในรูปแบบ 1NF

Nattachan\_Ningnoi\_6109656311\_Google-Scholar-Assignment

File Edit View Insert Format Data Tools Add-ons Help Last edit was made 4 minutes ago by NATTACHAN NINGNOI

B158 Effect of drying process on heavy metals fractionation of contaminated sludge for land application

	A	B	C	D	E	F
1	title		authors	publication_date	description	cite_by
2	0 Low-cost adsorbents for heavy metals uptake from contaminated water: a review	Sandhya Babel	2003/2/28	In this article, the technical feasibilit	4013	
3	1 Low-cost adsorbents for heavy metals uptake from contaminated water: a review	Tonni Agustiono Kurniawan	2003/2/28	In this article, the technical feasibilit	4013	
4	2 Physico-chemical treatment techniques for wastewater laden with heavy metals	Tonni Agustiono Kurniawan	2006/5/1	This article reviews the technical ap	1973	
5	3 Physico-chemical treatment techniques for wastewater laden with heavy metals	Gilbert YS Chan	2006/5/1	This article reviews the technical ap	1973	
6	4 Physico-chemical treatment techniques for wastewater laden with heavy metals	Wai-Hung Lo	2006/5/1	This article reviews the technical ap	1973	
7	5 Physico-chemical treatment techniques for wastewater laden with heavy metals	Sandhya Babel	2006/5/1	This article reviews the technical ap	1973	
8	6 Cr (VI) removal from synthetic wastewater using coconut shell charcoal and commercial activated carbon	Sandhya Babel	2004/2/1	In this study, the technical feasibility	948	
9	7 Cr (VI) removal from synthetic wastewater using coconut shell charcoal and commercial activated carbon	Tonni Agustiono Kurniawan	2004/2/1	In this study, the technical feasibility	948	
10	8 Comparisons of low-cost adsorbents for treating wastewaters laden with heavy metals	Tonni Agustiono Kurniawan	2006/8/1	In this article, the removal performa	804	
11	9 Comparisons of low-cost adsorbents for treating wastewaters laden with heavy metals	Gilbert YS Chan	2006/8/1	In this article, the removal performa	804	
12	10 Comparisons of low-cost adsorbents for treating wastewaters laden with heavy metals	Wai-hung Lo	2006/8/1	In this article, the removal performa	804	
13	11 Comparisons of low-cost adsorbents for treating wastewaters laden with heavy metals	Sandhya Babel	2006/8/1	In this article, the removal performa	804	
14	12 Equilibrium and kinetic modelling of biosorption of Remazol Black B by Rhizopus arrhizus in a batch system	Zümrüye Aksu	2000/12/1	The biosorption of Remazol Black E	584	
15	13 Equilibrium and kinetic modelling of biosorption of Remazol Black B by Rhizopus arrhizus in a batch system	Sevilay Tezer	2000/12/1	The biosorption of Remazol Black E	584	
16	14 Heavy metal removal from contaminated sludge for land application: a review	Sandhya Babel	2006/1/1	In recent years, various methods for	340	
17	15 Heavy metal removal from contaminated sludge for land application: a review	Dominica del Mundo Dacera	2006/1/1	In recent years, various methods for	340	
18	16 Microfiltration membrane fouling and cake behavior during algal filtration	Sandhya Babel	2010/10/15	This study was carried out to investi	169	
19	17 Microfiltration membrane fouling and cake behavior during algal filtration	Satoshi Takizawa	2010/10/15	This study was carried out to investi	169	
20	18 A matrix in life cycle perspective for selecting sustainable materials for buildings in Sri Lanka	UG Yasantha Abeywardha	2009/5/1	This paper presents a matrix to sele	152	
21	19 A matrix in life cycle perspective for selecting sustainable materials for buildings in Sri Lanka	Sandhya Babel	2009/5/1	This paper presents a matrix to sele	152	
22	20 A matrix in life cycle perspective for selecting sustainable materials for buildings in Sri Lanka	Shabhir Gheewala	2009/5/1	This paper presents a matrix to sele	152	
23	21 Effect of acid speciation on solid waste liquefaction in an anaerobic acid digester	Sandhya Babel	2004/5/1	Laboratory scale experiments were	143	
24	22 Effect of acid speciation on solid waste liquefaction in an anaerobic acid digester	Kensuke Fukushi	2004/5/1	Laboratory scale experiments were	143	
25	23 Effect of acid speciation on solid waste liquefaction in an anaerobic acid digester	Bunpot Sitranrasamee	2004/5/1	Laboratory scale experiments were	143	
26	24 Factors affecting seasonal variation of membrane filtration resistance caused by Chlorella algae	Sandhya Babel	2002/3/1	A seasonal fluctuation pattern was c	113	
27	25 Factors affecting seasonal variation of membrane filtration resistance caused by Chlorella algae	Satoshi Takizawa	2002/3/1	A seasonal fluctuation pattern was c	113	

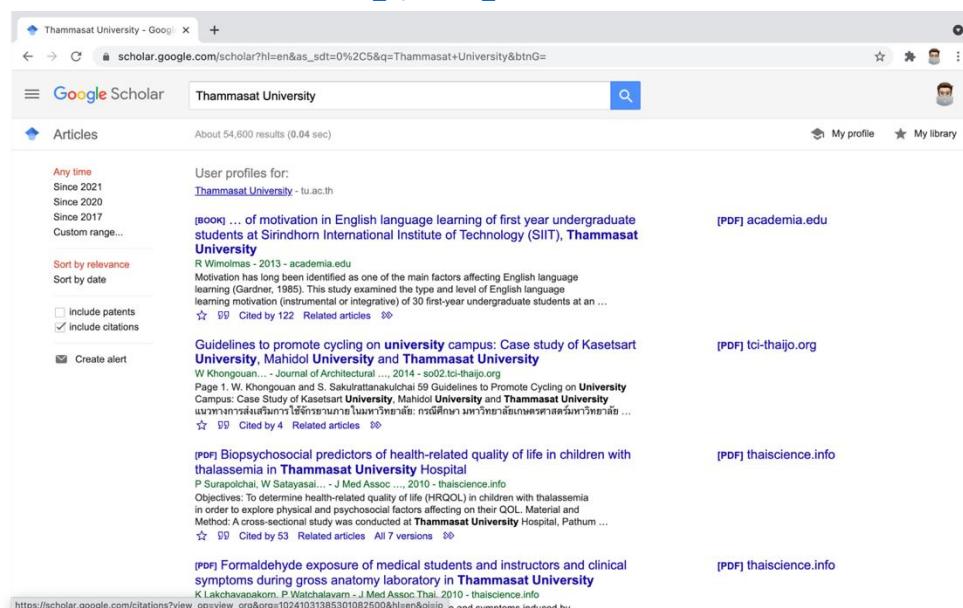
+ author paper paper\_1NF Explore >

## 2. ขั้นตอนการเก็บข้อมูลทำอย่างไร

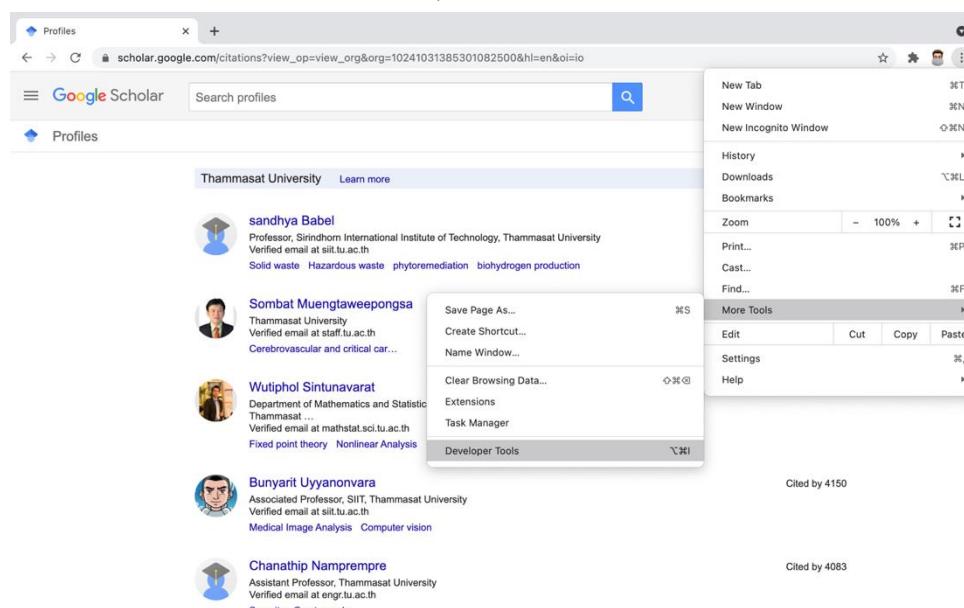
ก่อนที่เราจะเก็บข้อมูลจากส่วนต่าง ๆ ของเว็บไซต์ เราต้องทำการศึกษาโครงสร้างของเว็บไซต์ดังกล่าวก่อน ว่าตัวแปรต่าง ๆ ถูกเก็บไว้ใน class อะไรและในตัวแปรนั้น ๆ จะให้ค่าอะไรออกมา

### 2.1. ขั้นตอนที่ 1 ศึกษารูปแบบเว็บไซต์

- ค้นหา “[Google Scholar](#)” โดยค้นหาคำว่า “Thammasat University” ซึ่งจะได้ผลลัพธ์ดังภาพ ให้คลิกที่ User profiles for: [Thammasat University - tu.ac.th](#) ซึ่ง Google Scholar จะพาเราไปยัง [https://scholar.google.com/citations?view\\_op=view\\_org&org=10241031385301082500&hl=en&oi=io](https://scholar.google.com/citations?view_op=view_org&org=10241031385301082500&hl=en&oi=io)



- เมื่อเราได้เว็บไซต์ที่เราต้องการจะดึงข้อมูลแล้ว เราอาจจะเริ่มศึกษารูปแบบของเว็บไซต์ โดยไปที่ ⋮ แล้วเลือก “More Tools” แล้วเลือก “Developer Tools”



### 3. เมื่อเราคลิกที่ Developer Tools แล้วจะได้หน้าตาดังนี้

The screenshot shows the Google Scholar developer tools interface. The left pane displays a list of profiles from Thammasat University, each with a thumbnail, name, citation count, and a brief description of their research interests. The right pane shows the DOM structure with the following hierarchy:

```
<!DOCTYPE html>
<html class="gs_el_ta gs_el_sm" data-arp-injected="true">
  <head></head>
  <body>...</body>
</html>
```

A large diagram on the right illustrates the HTML structure:

- <!DOCTYPE html>**: The root element.
- <HTML>**: Contains the head and body sections.
- <HEAD>** and **</HEAD>**: The head section.
- <BODY>** and **</BODY>**: The body section, labeled **CONTENT**.
- </HTML>**: Closes the root element.

The bottom of the developer tools window shows the **<!doctype>** tab selected, along with other tabs like Styles, Computed, Layout, etc.

### โดยมีรายละเอียดที่เราควรรู้ดังนี้

- **<!DOCTYPE html>** คือ tag ที่ไว้บอกбраузอร์ว่าเอกสารที่มันกำลังเปิดอยู่ถูกกำหนดมาตรฐานไว้แบบไหน โดยที่ Doctype จะเป็น tag ที่อยู่ก่อน tag html [1] โดย Google scholar ใช้รูปแบบ HTML 5 Doctype [2]
- **<HTML> </HTML>** โดย **<HTML>** เป็นคำสั่งเริ่มต้นในการเขียนโปรแกรม และ **</HTML>** เป็นคำสั่งจุดสิ้นสุด โปรแกรมเหมือนคำสั่ง Beign และ End ใน Pascal [3]
- **<HEAD> </HEAD>** ใช้กำหนดข้อความในส่วนที่เป็นชื่อเรื่อง ภายในคำสั่งนี้จะมีคำสั่งย่ออย อีกหนึ่งคำสั่ง คือ **<TITLE>** [3]
- **<BODY> </BODY>** จะเป็นส่วนเนื้อหาของโปรแกรมจะเริ่มต้นด้วย คำสั่ง **<BODY>** และจบลงด้วย **</BODY>** ภายในคำสั่งนี้ คือ ส่วนที่จะแสดงทางจอภาพ [3]

## 2.2 ศึกษารายละเอียดของสิ่งที่เราจะดึงข้อมูล

Profiles

scholar.google.com/citations?view\_op=view\_org&org=10241031385301082500&hl=en&oi=io

Search profiles

Profiles

Learn more

div.gsc\_1usr 551 x 142

sandhya Babel Cited by 11176

Professor, Sirindhorn International Institute of Technology, Thammasat University  
Verified email at sitt.ac.th  
Solid waste Hazardous waste phytoremediation  
biodegradation production

Sombat Muengtaweepongosa Cited by 6155

Thammasat University  
Verified email at staff.tu.ac.th  
Cerebrovascular and critical care...

Wutiphol Sintunavarat Cited by 4484

Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat ...  
Verified email at mathstat.sci.tu.ac.th  
Fixed point theory Nonlinear Analysis

Bunyarat Uyyanonvara Cited by 4150

Associated Professor, SIT, Thammasat University  
Verified email at sitt.ac.th  
Medical Image Analysis Computer vision

Chanathip Namprempre Cited by 4083

Assistant Professor, Thammasat University  
Verified email at enqr.tu.ac.th

Profiles

scholar.google.com/citations?view\_op=view\_org&org=10241031385301082500&hl=en&oi=io

Search profiles

Profiles

Thammasat University Learn more

div.gsc\_1usr 551 x 110

sandhya Babel Cited by 11176

Professor, Sirindhorn International Institute of Technology, Thammasat University  
Verified email at sitt.ac.th  
Solid waste Hazardous waste phytoremediation

Sombat Muengtaweepongosa Cited by 6155

Thammasat University  
Verified email at staff.tu.ac.th  
Cerebrovascular and critical care...

Wutiphol Sintunavarat Cited by 4484

Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat ...  
Verified email at mathstat.sci.tu.ac.th  
Fixed point theory Nonlinear Analysis

Bunyarat Uyyanonvara Cited by 4150

Associated Professor, SIT, Thammasat University  
Verified email at sitt.ac.th  
Medical Image Analysis Computer vision

Chanathip Namprempre Cited by 4083

Assistant Professor, Thammasat University  
Verified email at enqr.tu.ac.th

- 2.2.1 จากรูปด้านบนเราจะเห็นกล่องสีเขียว ซึ่งกล่องนี้จะเก็บค่าต่าง ๆ ของ user ไว้ ซึ่งอยู่ใน class name ที่มีชื่อว่า 'gsc\_1usr' ซึ่งภายในกล่อง 'gsc\_1usr' ก็จะประกอบไปด้วย ชื่อ, สังกัด, e-mail, Cited by, tag ที่เกี่ยวข้อง



sandhya Babel

Professor, Sirindhorn International Institute of Technology, Thammasat University

Verified email at siit.tu.ac.th

## Solid waste   Hazardous waste   phytoremediation biohydrogen production

Cited by 11217

- 2.2.2 ในส่วนของ user\_ID ใช้ในการเก็บรหัสของ Author เราจะดึงมาจาก class ที่มีชื่อว่า gs\_ai\_ph0 ซึ่ง user\_ID เราสามารถหาได้จาก link คือ href="/citations?hl=en&user=Re819VUAAAAJ" ซึ่งเราสามารถแยกเพื่อนำมาแค่ user\_ID ออกมายได้ โดยการลบ URL ที่อยู่ด้านหน้าออกแล้วจะเหลือแค่สิ่งที่เราต้องการคือ user\_ID คือ Re819VUAAAAJ

Thammasat University Learn more

**sandhya Babel**

Professor, Sirindhorn International Institute of Technology, Thammasat University  
Verified email at siit.tu.ac.th  
Solid waste Hazardous waste phytoremediation  
biohydrogen production

Cited by 11217

```
<div class="gs_ai_gs_scl gs_ai_chpr">
  <a href="/citations?hl=en&user=Re819VUAAAAJ" class="gs_ai_ph0" style="color: #000000; text-decoration: none; font-weight: bold;">sandhya Babel
  <span class="gs_rimg gs_pp_sm">
    
  </span>
</a>
<div class="gs_ai_t">
  <h3 class="gs_ai_name">
    <a href="/citations?hl=en&user=Re819VUAAAAJ" style="color: #000000; text-decoration: none;">sandhya Babel

```

- 2.2.3 ในส่วนของชื่อ (name) จะอยู่ใน class = “gs\_ai\_name”

Profiles scholar.google.com/citations?view\_op=view\_org&org=10241031385301082500&hl=en&oi=io

Search profiles

h3.gs\_ai\_name 323x21 Margin 0px 0px 4px

ACCESSIBILITY

Name sandhya Babel Role heading ore Keyboard-focuable

**sandhya Babel**

Professor, Sirindhorn International Institute of Technology, Thammasat University  
Verified email at siit.tu.ac.th  
Solid waste Hazardous waste phytoremediation  
biohydrogen production

Cited by 11176

**Sombat Muengtaweepongsa**

Thammasat University  
Verified email at staff.tu.ac.th  
Cerebrovascular and critical care

Cited by 6155

**Putiphol Sintunavarat**

Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat ...  
Verified email at mathstat.sci.tu.ac.th  
Fixed point theory Nonlinear Analysis

Cited by 4484

**Bunyarat Uyyanovara**

Associated Professor, SIIT, Thammasat University  
Verified email at siit.tu.ac.th  
Medical Image Analysis Computer vision

Cited by 4150

**Chanathip Namprempre**

Assistant Professor, Thammasat University  
Verified email at engr.tu.ac.th

Cited by 4083

Elements Console Sources Network Performance

```
<div class="gsc_header">...</div>
<div id="gsc_org_help">...</div>
<div class="gsc_lusr">
  <div class="gs_scl gs_ai_chpr">
    <a href="/citations?hl=en&user=Re819VUAAAAJ" class="gs_ai_ph0" style="color: #000000; text-decoration: none; font-weight: bold;">sandhya Babel
    <span class="gs_rimg gs_pp_sm">
      
    </span>
  </a>
  <div class="gs_ai_t">
    <h3 class="gs_ai_name">
      <a href="/citations?hl=en&user=Re819VUAAAAJ" style="color: #000000; text-decoration: none;">sandhya BabelSombat Muengtaweepongsa
    <span class="gs_rimg gs_pp_sm">
      
    </span>
  </a>
  <div class="gs_ai_t">
    <h3 class="gs_ai_name">
      <a href="/citations?hl=en&user=97zz_TIAAAA" style="color: #000000; text-decoration: none;">Sombat Muengtaweepongsa

```

โดยใน class = “gs\_ai\_name” จะมีการเข้ามายิงลิงค์ไปยังหน้าอื่น โดยปกติเว็บไซต์หนึ่ง ๆ จะมีหน้าเพจมากกว่าหนึ่งหน้า อย่างเช่นจากหน้าโฮมเพจก็สามารถคลิกเพื่อเข้าไปดูข้อมูลในหน้าอื่น ๆ ได้ ซึ่งเราเรียกว่าการเข้ามายิงเว็บเพจ หรือ การลิงค์ โดยมีรูปแบบคือ [ไฟล์เว็บเพจ](#)... สิ่งที่จะลิงค์...</a>

```
... <h3 class="gs_ai_name" style="color: #000000; text-decoration: none; font-weight: bold;">sandhya Babel

```

- 2.2.4 ในส่วนของ affiliation หรือ สังกัดจะอยู่ใน class="gs\_ai\_aff"

The screenshot shows a list of profiles from Thammasat University. The 'affiliation' section for each profile is highlighted with a yellow box and its corresponding HTML code is shown below:

```
<div class="gs_ai_aff">
    "Professor, Sirindhorn International Institute of Technology,
    Thammasat University"
</div>
```

- 2.2.5 ในส่วนของ Cited\_by จะอยู่ใน class="gs\_ai\_cby"

The screenshot shows a list of profiles from Thammasat University. The 'Cited\_by' section for each profile is highlighted with a yellow box and its corresponding HTML code is shown below:

```
<div class="gs_ai_cby">Cited by 11176</div>
```

สรุป Sheet ที่ 1 คือ Author Table

- user\_ID จะอยู่ใน class = “gs\_ai\_name”
- ชื่อของ (name) จะอยู่ใน class = “gs\_ai\_name”
- Affiliation จะอยู่ใน class 'gs\_ai\_aff'

## 2.3 ดึงข้อมูลของ Sheet ที่ 2 คือ Paper Table

ศึกษารายละเอียดในแต่ละส่วนของเว็บไซต์ที่เราต้องการจะดึงข้อมูล

**Sombat Muengtaweepongsa**  
Thammasat University  
Verified email at staff.tu.ac.th - Homepage  
Cerebrovascular and critical...

TITLE	CITED BY	YEAR
Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study	3247	2010
Aspirin and extended-release dipyridamole versus clopidogrel for recurrent stroke	1090	2008
Telmisartan to prevent recurrent stroke and cardiovascular events	907	2008
Chinese medicine neuroaid efficacy on stroke recovery: a double-blind, placebo-controlled, randomized study	74	2013
Practice patterns and outcomes after stroke across countries at different economic levels (INTERSTROKE): an international observational study	49	2018
Early outcome after intravenous thrombolysis in patients with acute ischemic stroke	38	2011

**ตัวอย่างหน้าเว็บไซต์ของ "Sombat Muengtaweepongsa"**

**GET MY OWN PROFILE**

**Cited by**

All	Since 2016
Citations 6155	2905
h-index 19	14
I10-index 30	26

**Public access**

0 articles	8 articles
not available	available

Based on funding mandates

**Title** Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study [PDF] from academia.edu

**Authors** Martin J O'Donnell, Denis Xavier, Lisheng Liu, Hongye Zhang, Siu Lim Chin, Purnima Rao-Melacini, Sumathy Rangarajan, Shofiqul Islam, Prem Pais, Matthew J McQueen, Charles Mendo, Alberto Damasceno, Patrício Lopez-Jaramillo, Graeme J Hankey, Antonio L Dans, Khalid Yusuf, Thomas Truelsen, Hans-Christoph Diener, Ralph L Sacco, Danuta Ryglewicz, Anna Czonkowska, Christian Weimar, Xingyu Wang, Salim Yusuf

**Publication date** 2010/7/10

**Description** Background: The contribution of various risk factors to the burden of stroke worldwide is unknown, particularly in countries of low and middle income. We aimed to establish the association of known and emerging risk factors with stroke and its primary subtypes, assess the contribution of these risk factors to the burden of stroke, and explore the differences between risk factors for stroke and myocardial infarction. Methods: We undertook a standardised case-control study in 22 countries worldwide between March 1, 2007, and April 23, 2010. Cases were patients with acute first stroke (within 5 days of symptoms onset and 72 h of hospital admission). Controls had no history of stroke, and were matched with cases for age and sex. All participants completed a structured questionnaire and a physical examination, and most provided blood and urine samples. We calculated odds ratios (ORs) and population-attributable risks ...

**Cited by** Cited by 3247

**ตัวอย่างหัวข้อที่เราสนใจ ซึ่งประกอบด้วย**

- Title
- Authors
- Publication date
- Description
- Cited by

- 2.3.1 จากรูปด้านล่างเราจะเห็นกล่องสีเขียว ซึ่งกล่องนี้จะเก็บค่าต่าง ๆ ของ user ไว้ ซึ่งอยู่ใน class name ที่มีชื่อว่า 'gsc\_a\_at' ซึ่งภายในกล่อง 'gsc\_a\_at' ก็จะประกอบไปด้วยลิงก์ที่จะแสดงข้อมูลของ paper นั้นๆ

Chrome is being controlled by automated test software.

ARTICLES	CITED BY	PUBLIC ACCESS	CITED BY	YEAR
<a href="#">a.gsc_a_at</a> 520.27 x 52				
Low-cost adsorbents for heavy metals uptake from contaminated water: a review S Babel, TA Kurniawan Journal of hazardous materials 97 (1-3), 219-243	4013	2003		
Physico-chemical treatment techniques for wastewater laden with heavy metals TA Kurniawan, GYS Chan, WH Lo, S Babel Chemical engineering journal 118 (1-2), 83-98	1973	2006		
Cr (VI) removal from synthetic wastewater using coconut shell charcoal and commercial activated carbon modified with oxidizing agents and/or chitosan S Babel, TA Kurniawan Chemosphere 54 (7), 951-967	948	2004		
Comparisons of low-cost adsorbents for treating wastewaters laden with heavy metals TA Kurniawan, GYS Chan, W Lo, S Babel Science of the total environment 366 (2-3), 409-426	804	2006		
Equilibrium and kinetic modelling of biosorption of Remazol Black B by Rhizopus arrhizus in a batch system: effect of temperature Z Aksu, S Tezer Process Biochemistry 36 (5), 431-439	584	2000		
Heavy metal removal from contaminated sludge for land application: a review S Babel, D del Mundo Dacera Waste Management 26 (9), 988-1004	340	2006		

[Low-cost adsorbents for heavy metals uptake from contaminated water: a review](javascript:void(0)) == \$0

Chrome is being controlled by automated test software.

ARTICLES	CITED BY	PUBLIC ACCESS	CITED BY	YEAR
<a href="#">a.gsc_a_at</a> 514.05 x 52				
Low-cost adsorbents for heavy metals uptake from contaminated water: a review S Babel, TA Kurniawan	4013	2003		
Physico-chemical treatment techniques for wastewater laden with heavy metals TA Kurniawan, GYS Chan, WH Lo, S Babel Chemical engineering journal 118 (1-2), 83-98	1973	2006		
Cr (VI) removal from synthetic wastewater using coconut shell charcoal and commercial activated carbon modified with oxidizing agents and/or chitosan S Babel, TA Kurniawan Chemosphere 54 (7), 951-967	948	2004		
Comparisons of low-cost adsorbents for treating wastewaters laden with heavy metals TA Kurniawan, GYS Chan, W Lo, S Babel Science of the total environment 366 (2-3), 409-426	804	2006		
Equilibrium and kinetic modelling of biosorption of Remazol Black B by Rhizopus arrhizus in a batch system: effect of temperature Z Aksu, S Tezer Process Biochemistry 36 (5), 431-439	584	2000		
Heavy metal removal from contaminated sludge for land application: a review S Babel, D del Mundo Dacera Waste Management 26 (9), 988-1004	340	2006		

[Low-cost adsorbents for heavy metals uptake from contaminated water: a review](javascript:void(0)) == \$0

### ● 2.3.2 title ใช้ในการเก็บชื่อ paper จะอยู่ใน class 'gsc\_vcd\_title\_link'

The screenshot shows a Google Scholar citation page for a paper titled "Low-cost adsorbents for heavy metals uptake from contaminated water: a review". The title is highlighted in purple, indicating it matches the CSS selector 'a.gsc\_vcd\_title\_link'. The developer tools in Chrome are open, showing the element tree where the title is wrapped in a div with id="gsc\_vcd\_title\_link". The page also displays citation statistics, a bar chart of total citations over time, and a list of related articles.

### ● 2.3.3 authors ใช้ในการเก็บรายชื่อผู้จัดทำ paper จะอยู่ใน class 'gsc\_vcd\_field'

The screenshot shows the same Google Scholar citation page as before, but now the authors' names "Sandhya Babel, Tonni Agustiono Kurniawan" are highlighted in purple, matching the CSS selector 'a.gsc\_vcd\_field'. The developer tools in Chrome are open, showing the element tree where the authors' names are wrapped in a div with id="gsc\_vcd\_field". The page displays the same citation statistics and related articles as the previous screenshot.

- 2.3.4 publication\_date ใช้ในการเก็บ วันเวลา ที่เผยแพร่ paper นั้น ๆ จะอยู่ใน class 'gsc\_vcd\_field'

Chrome is being controlled by automated test software.

The screenshot shows a Google Scholar search result for a paper titled "Low-cost adsorbents for heavy metals uptake from contaminated water: a review". The publication date, "2003/2/28", is highlighted in the DOM inspector under the 'Elements' tab. The page content includes details like authors (Sandhya Babel, Tonni Agustiono Kurniawan), total citations (4013), and a bar chart of citations over time.

```

<div id="gsc_vcd_table">
  <div class="gs_scl">
    <div class="gsc_vcd_field">Authors</div>
    <div class="gsc_vcd_value">Sandhya Babel, Tonni Agustiono Kurniawan</div>
  </div>
  <div class="gs_scl">
    <div class="gsc_vcd_field">Publication date</div> == $0
    <div class="gsc_vcd_value">2003/2/28</div>
  </div>
</div>

```

- 2.3.5 description ใช้ในการเก็บคำอธิบายของ paper นั้น ๆ จะอยู่ใน class 'gsc\_vcd\_field'

Chrome is being controlled by automated test software.

The screenshot shows the same Google Scholar search result as above. The 'Description' section of the paper's summary is highlighted in the DOM inspector under the 'Elements' tab. This section contains a detailed abstract of the research findings.

```

<div class="gsc_vcd_field">Description</div> == $0
  <div class="gsc_vcd_value" id="gsc_vcd_descr">...</div>

```

- 2.3.6 cite\_by ใช้ในการเก็บจำนวนคนที่เข้ามาอ่าน paper จะอยู่ใน class 'gsc\_vcd\_field'

The screenshot shows a Google Scholar citation page for a paper titled "Low-cost adsorbents for heavy metals uptake from contaminated water: a review". The developer tools (Elements tab) highlight the class "gsc\_vcd\_field" which contains author information: "Authors Sandhya Babel, Tonni Agustiono Kurniawan". Below this, there is a bar chart showing citations from 2007 to 2021, and a table of scholar articles.

- 2.3.7 ปุ่ม SHOW MOER เพื่อใช้ในกดคลิกเพื่อเก็บ paper ต่าง ๆ ใน Author ให้ครบ จะอยู่ใน class 'gs\_btnPD'

The screenshot shows a Google Scholar citation page for a paper titled "Chemical pretreatment for reduction of membrane fouling caused by algae". The developer tools (Elements tab) highlight the class "gs\_btnPD" which is part of a button element. Below the button, there is a list of related publications.

## สรุป Sheet ที่ 2 คือ Paper Table

- title ใช้ในการเก็บชื่อ paper จะอยู่ใน class 'gsc\_vcd\_title\_link'
- authors ใช้ในการเก็บรายชื่อผู้จัดทำ paper จะอยู่ใน class 'gsc\_vcd\_field'
- publication\_date ใช้ในการเก็บ วันเวลา ที่เผยแพร่ paper นั้น ๆ จะอยู่ใน class 'gsc\_vcd\_field'
- description ใช้ในการเก็บคำอธิบายของ paper นั้น ๆ จะอยู่ใน class 'gsc\_vcd\_field'
- cite\_by ใช้ในการเก็บจำนวนคนที่เข้ามาอ่าน paper จะอยู่ใน class 'gsc\_vcd\_field'
- ปุ่ม SHOW MOER เพื่อใช้ในการคลิกเพื่อเก็บ paper ต่าง ๆ ใน Author ให้ครบ จะอยู่ใน class 'gs\_btnPD'
- กล่องของ paper นั้น ๆ ซึ่งกล่องนี้จะเก็บค่าต่าง ๆ ของ user ไว้ ซึ่งอยู่ใน class name ที่มีชื่อว่า 'gsc\_a\_at'

## ส่วนของ Code

### Import Necessary Libraries

```
1 from bs4 import BeautifulSoup as bs
2 import pandas as pd
3 import time
4 import requests
5 import random
6 from selenium import webdriver
7 from bs4 import BeautifulSoup
8 from requests import get
9
10 import time
11 from selenium import webdriver
12
13 from random import randint
14 from time import sleep
```

## ส่วนของ Code

Sheet ที่ 1

Author Table

```
1 # เรียกใช้ WebDriver ที่อยู่ใน current directory
2 driver = webdriver.Chrome("./chromedriver")
3 # เรียกใช้เว็บไซต์ที่เราต้องการดึง
4 driver.get("https://scholar.google.com/citations?view_op=view_org&org=10241031385301082500&hl=en&oi=io")
5
6
7
8
9
10 #%%time
11 """
12 %%time คือ magic command ที่ใช้ในการติดตามสิ่งที่เราทำลังบาระมาผลลัพธ์ทั้งหมดภายใน Cell นั้น ๆ โดยจะบอกถึง [100]
13 CPU times:
14   • user คือ เวลาทั้งหมดที่ใช้โดย CPUs ในระหว่างการทำงาน
15   • sys คือ เวลาทั้งหมดที่ใช้โดย CPU ในระหว่างการทำงานที่เกี่ยวข้องกับระบบ เช่น การจัดสรรหน่วยความจำ (memory allocation)
16 Wall time:
17   • คือเวลาทั้งหมดที่มีบังคับเริ่มต้นโปรแกรมถึงสิ้นสุดโปรแกรมหรือความแตกต่างระหว่างเวลาที่งานเสร็จสิ้นและเวลาที่งานเริ่มต้นขึ้น
18 """
19
20 # สร้าง data frame ที่ใช้ในการเก็บข้อมูลของ Author
21 author_df = pd.DataFrame(columns=['user_ID','name','affiliation'])
22 """
23   • 'user_ID' ใช้ในการเก็บรหัสของ Author หรือผู้ใช้งานนั้น ๆ
24   • 'name' ใช้ในการเก็บชื่อของ Author
25   • 'affiliation' หรือ สังกัด ใช้ในการเก็บข้อมูลเกี่ยวกับคณะหรือสังกัดของ Author
26 """
27
28 # กำหนด row = 1
29 """
30 ก้าวหนึ่ง row = 1 เพื่อก้าวเดาแรกของ data frame ที่เราจะเก็บข้อมูลลงไป โดยเริ่มเดาต้นคือ 1 แล้วเพิ่มขึ้นเรื่อย ๆ
31 เมื่อเก็บข้อมูลครบ 1 ครั้ง จำนวนแ嘎ก็จะเพิ่มขึ้น จาก row +=1
32 """
33 row = 1
34
35 while True:
36   """
37   โดยเราจะใช้ While loop ซึ่งเป็นคำสั่งวนซ้ำ คือ เราจะเก็บข้อมูลทั้งหมด
38   จนกว่าเงื่อนไขจะเป็นจริง คือ เก็บข้อมูลครบทั้งหมด
39   """
40
41 author_objects = driver.find_elements_by_class_name('gsc_lusr')
42 """
43 ก้าวหนึ่งที่แปร author_objects ซึ่งเปรียบเสมือนกล่อง ๆ หนึ่ง
44 แล้วในกล่องก็จะมีสิ่งของต่าง ๆ ได้แก่ user_ID , author_name และ affiliation
45 ซึ่งอยู่ใน class ที่มีชื่อว่า 'gsc_lusr'
46 """
47
48 # ใช้ for loop ในการเก็บค่าต่าง ๆ ที่อยู่ในกล่องใบนั้น ได้แก่ user_ID , author_name และ affiliation
49 for item in author_objects:
50   link = item.find_element_by_class_name('gs_ai_phot')
51   """
52   Link โดยเราจะใช้ find_element_by_class_name('class_name') เพื่อใช้ในการหา class name
53   ที่มีชื่อว่า 'gs_ai_phot' สำหรับ class name ที่ตรงกันเราจะได้ตัวที่อยู่ใน class นั้นออกมาก็คือ Link ของ
54   Author แหล่งคน ตัวอย่างเช่น href="/citations?hl=en&user=97Zz_TIAAAAJ"
55   """
56   user_ID = str(link.get_attribute('href')).replace('https://scholar.google.com/citations?hl=en&user=', '')
57   author_name = item.find_element_by_class_name('gs_ai_name')
58   author_name = author_name.text
59   """
60   author_name โดยเราจะใช้ find_element_by_class_name('class_name') เพื่อใช้ในการหา class name ที่มีชื่อว่า 'gs_ai_name'
61   สำหรับ class name ที่ตรงกันเราจะได้ตัวที่อยู่ใน class นั้นออกมาก็คือ ชื่อของ Author แล้วแปลง element ต่าง ๆ ให้อยู่ในรูปแบบของ text
62   """
63
64   affiliation = item.find_element_by_class_name('gs_ai_aff')
65   affiliation = affiliation.text
66   """
67   affiliation จะท้ารูปแบบเดียวกันกับ author_name หรือ class name ที่ตรงกันคือ 'gs_ai_aff'
68   แล้วค่าที่อยู่ใน class นี้ให้อยู่ในรูปแบบของ text
69   """
70
71   # เพิ่มข้อมูลลงในไฟล์
72   author_df.loc[row] =[user_ID,author_name,affiliation]
73   # เพิ่มจำนวน row เพื่อบอก data frame ให้ขึ้น row ใหม่
74   row +=1
75
76   # หา URL ปัจจุบันที่ webdriver กำลังเรียกใช้งานอยู่
77   page = requests.get(driver.current_url)
78   # ส่งคืนค่าเนื้อหาในเว็บกลับมาในรูปแบบของ raw bytes
79   soup = BeautifulSoup(page.content)
80
81   # ใช้ในการคลิกปุ่มเพื่อเบรคหน้า
82   next_buttons = soup.find_all(class_="gs_btnPR")
83   if next_buttons[0].has_attr('onclick'):
84     driver.find_element_by_css_selector(".gs_btnPR").click()
85   else:
86     break
87   # หยุด โดยนั้น 0 ถึง 3 แล้วค่อยเริ่มต้นทำงานใหม่
88   sleep(randint(0,5))
89
90   # save file ในรูปแบบ .CSV
91   author_df.to_csv('author.csv')
92   # ปิดหน้าต่างเว็บ browser ที่ webdriver กำลังเรียกใช้งานอยู่
93   driver.close()
```

CPU times: user 2.98 s, sys: 215 ms, total: 3.2 s  
Wall time: 1min 36s

**บรรทัดที่ 1 : %%time** คือ magic command ที่ใช้ในการติดตามสิ่งที่เรากำลังประมวลผลอยู่ทั้งหมดภายใน Cell นั้น ๆ โดยจะบอกถึง [5]

- CPU times:
  - user คือ เวลาทั้งหมดที่ใช้โดย CPUs ในระหว่างการคำนวณ
  - sys คือ เวลาทั้งหมดที่ใช้โดย CPUs ในระหว่างการทำงานที่เกี่ยวข้องกับระบบ เช่น การจัดสรรหน่วยความจำ (memory allocation)
- Wall time:
  - คือ เวลาทั้งหมดที่นับตั้งแต่เริ่มรันโปรแกรมจนถึงสิ้นสุดโปรแกรมหรือความแตกต่างระหว่างเวลาที่งานเสร็จสิ้นและเวลาที่งานเริ่มต้นขึ้น

**บรรทัดที่ 11 : "author\_df"** สร้าง data frame ที่ใช้ในการเก็บข้อมูลของ Author ได้แก่

- 'user\_ID' ใช้ในการเก็บรหัสของ Author หรือผู้ใช้งานนั้น ๆ
- 'name' ใช้ในการเก็บชื่อของ Author
- 'affiliation' หรือ สังกัด ใช้ในการเก็บข้อมูลเกี่ยวกับคณะหรือสังกัดของ Author

**บรรทัดที่ 23 :** กำหนด row = 1 เพื่อกำหนดแถวของ data frame ที่เราจะเก็บข้อมูลลงไป โดยเริ่มแถวต้นคือ 1 แล้วเพิ่มขึ้นเรื่อย ๆ เมื่อเก็บข้อมูลครบ 1 ครั้ง จำนวนแถวจะเพิ่มขึ้น จาก row +=1

**บรรทัดที่ 22-60 :** โดยเราจะใช้ While loop ซึ่งเป็นคำสั่งวนซ้ำ คือ เราจะเก็บข้อมูลทั้งหมด จนกว่าเงื่อนไขจะเป็นจริง คือ เก็บข้อมูลครบทั้งหมด

**บรรทัดที่ 31 :** โดยเราจะกำหนดตัวแปร author\_objects ซึ่งเปรียบเสมือนกล่อง ๆ หนึ่ง และในกล่องก็จะมีสิ่งของต่าง ๆ ได้แก่ user\_ID , author\_name และ affiliation ซึ่งอยู่ใน class ที่มีชื่อว่า 'gs\_1usr'

**บรรทัดที่ 38 :** ใช้ for loop ในการเก็บค่าต่าง ๆ ที่อยู่ในกล่องใบนั้น ได้แก่ user\_ID , author\_name และ affiliation

**บรรทัดที่ 39 :** Link โดยเราจะใช้ find\_element\_by\_class\_name('class\_name') เพื่อใช้ในการหา class name ที่มีชื่อว่า 'gs\_ai\_phot' ถ้ามี class name ที่ตรงกันเราจะได้ค่าที่อยู่ใน class นั้นออกมาคือ Link ของ Author แต่ละคน ตัวอย่างเช่น href="/citations?hl=en&user=97Zz\_TIAAAJ"

**บรรทัดที่ 42 :** user\_ID โดยเราสามารถหาได้จาก link คือ href="/citations?hl=en&user=u-vl\_aiAAAAJ" ซึ่งเราสามารถแยกเพื่อนำมาแค่ user\_ID ออกมาได้ โดยการลบ URL ที่อยู่ด้านหน้าออกแล้วจะเหลือแค่สิ่งที่เราต้องการคือ user\_ID เช่น u-vl\_aiAAAAJ

**บรรทัดที่ 46-47 :** author\_name โดยเราจะใช้ find\_element\_by\_class\_name('class\_name) เพื่อใช้ในการหา class name ที่มีชื่อว่า 'gs\_ai\_name' ถ้ามี class name ที่ตรงกันเราจะได้ค่าที่อยู่ใน class นั้นออกมาคือ ชื่อของ Author และเปลี่ยน element ต่าง ๆ ให้อยู่ในรูปแบบของ text

**บรรทัดที่ 53-54 :** affiliation จะทำรูปแบบเดียวกันกับ author\_name การหา class name ที่ตรงกันคือ 'gs\_ai\_aff' และค่าที่อยู่ใน class นั้นให้อยู่ในรูปแบบของ text

**บรรทัดที่ 61 :** เพิ่มข้อมูลที่เราได้ลงในตัวแปร author\_df โดยเพิ่มในรูปแบบของ dict โดยใช้ .loc[row] และ row คือจำนวนแถวที่จะเพิ่มข้อมูลลงไป

**บรรทัดที่ 63 :** "row +=1" เพิ่มจำนวน row เพื่อบอก data frame ให้ขึ้น row ใหม่

**บรรทัดที่ 65 :** ใช้ requests.get เพื่อหา URL ปัจจุบันที่ webdriver กำลังเรียกใช้งานอยู่

**บรรทัดที่ 67 :** ส่งคืนค่าเนื้อหาในเว็บกลับมาในรูปแบบของ raw bytes

**บรรทัดที่ 70-75 :** ในส่วนต่อมาใช้ในการเปลี่ยนหน้า ซึ่งเปรียบเสมือนเรากดปุ่ม ไปเรื่อย ๆ จนครบ จำนวนหน้าที่เราต้องการ หรือ ไม่สามารถที่จะคลิกต่อไปได้อีกแล้ว แต่เพียงบอกเราให้เครื่องเป็นคนกดแทนเรา แต่เครื่องไม่ได้ฉลาดเหมือนคนเราจึงต้องบอกเครื่องว่าปุ่มคืออะไร โดย ปุ่มถูกกำหนดโดย .gs\_btnPR โดยใช้ .click() เพื่อใช้จำลองการกดของเรา

**บรรทัดที่ 77 :** เมื่อครบ 1 คน และเราจะทำการหยุด โดยนับ 0 ถึง 5 และค่อยเริ่มต้นทำงานใหม่ เพื่อไม่ให้ถูกกีดกันการเข้าถึงจาก Google เนื่องจากเดี๋ยไม่ต้องการให้ครमาร์ดึงข้อมูลปริมาณมากในเวลาเดียวกัน เพราะทำให้เพิ่มปริมาณการใช้งานไปยัง server จำนวนมากหากโดยไม่จำเป็น เราจึงต้อง act like humans คือ ไม่ดึงข้อมูลเร็วเกินไป ให้เหมือนมนุษย์ เป็นคนเปิดอยู่

**บรรทัดที่ 80 :** แล้วเมื่อคำสั่ง While loop เป็นจริง หรือ หมายถึงเราดึงข้อมูลมาครบทุกคนแล้ว ให้ save file จาก data frame ที่มีชื่อว่า author\_df ให้อยู่ในรูปแบบของ .CSV โดยมีชื่อว่า 'author.csv'

**บรรทัดที่ 82 :** driver.close() ใช้เพื่อปิดหน้าต่างเว็บ browser ที่ webdriver กำลังเรียกใช้งานอยู่ [6]

# Paper Table

```
1 %%time
2
3 # เรียกใช้ webdriver.Chrome()
4 driver = webdriver.Chrome("./chromedriver")
5 # เรียกใช้ไฟล์ author.csv เพื่อตั้งชื่อ默 'user_ID'
6 author = pd.read_csv("author.csv")
7 # สร้าง DataFrame ที่ใช้เก็บ Paper
8 paper_df = pd.DataFrame(columns=['title','authors','publication_date','description','cite_by'])
9
10 # กำหนด row = 1 เพื่อใช้ในการเพื่อกำหนดแถวของ data frame
11 row = 1
12 # ใช้ในการสั่งตัวของ author ที่เรากำลังดึงอยู่
13 user_queue = 0
14
15 # for loop วนซ้ำตามจำนวนของ user_ID
16 for user_ID in author['user_ID']:
17     # ให้ webdriver ใน การ เปิดเว็บ browser ตามตัวย user_ID ที่เราเก็บไว้ในไฟล์ author.csv"
18     driver.get('https://scholar.google.com/citations?hl=en&user=' + user_ID)
19     # คลิกปุ่ม SHOW MORE
20     button_show_more = driver.find_elements_by_class_name('gs_btnPD')
21     # เมื่อคลิกปุ่ม SHOW MORE ให้คลิกปุ่มนั้นจนกว่าปุ่มนั้นจะกดไม่ได้ หรือ เส็ง paper ครบหมดแล้ว
22     while button_show_more[0].is_enabled() == True:
23         # คลิกปุ่ม SHOW MORE
24         driver.find_element_by_class_name('gs_btnPD').click()
25         # เมื่อคลิกแล้วให้รอนับ 0 ถึง 5
26         sleep(randint(0,5))
27
28     # Link ที่บอกรายละเอียดของ paper นั้นๆ
29     link = driver.find_elements_by_class_name('gsc_a_at')
30     # เลือกเฉพาะ link อ่อนน้ำ
31     link = [i.get_attribute('data-href') for i in link]
32
33     # เส็งคิวของ Author ที่เราทำการดึงข้อมูล
34     print("-----")
35     print("User Queue : ", user_queue)
36     print(" ")
37     user_queue += 1
38
39     for url in link:
40         driver.get('https://scholar.google.com/' + url)
41         try :
42             table = driver.find_element_by_id('gsc_vcd_table')
43             body = table.find_elements_by_class_name('gs_scl')
44
45             Title = " "
46             Authors = " "
47             Publication_date = " "
48             Description = " "
49             cited = " "
50
51             for item in body:
52                 Title = driver.find_elements_by_class_name('gsc_vcd_title_link')[0].text
53                 if item.find_elements_by_class_name('gsc_vcd_field')[0].text == 'Authors':
54                     Authors = item.find_elements_by_class_name('gsc_vcd_value')[0].text
55                 elif item.find_elements_by_class_name('gsc_vcd_field')[0].text == 'Publication date':
56                     Publication_date = item.find_elements_by_class_name('gsc_vcd_value')[0].text
57                 elif item.find_elements_by_class_name('gsc_vcd_field')[0].text == 'Description':
58                     Description = item.find_elements_by_class_name('gsc_vcd_value')[0].text
59                 elif item.find_elements_by_class_name('gsc_vcd_field')[0].text == 'Total citations':
56                     cited = item.find_elements_by_css_selector('a')[0].text.replace('Cited by ','')
57                     paper_df.loc[row]=[Title,Authors,Publication_date,Description,cited]
58                     row += 1
59             except :
60                 pass
61             print('Articles number >>> ', len(paper_df))
62             sleep(randint(0,5))
63             paper_df.to_csv('paper.csv')
64             driver.close()
```

ส่วนของ Code

Sheet ที่ 2 Paper Table

## อธิบาย ส่วนของ Code Sheet ที่ 2 Paper Table

บรรทัดที่ 1 : %%time คือ magic command ที่ใช้ในการติดตามสิ่งที่เรากำลังประมวลผลอยู่ทั้งหมดภายใน Cell นั้น ๆ โดยจะบอกถึง [5]

- CPU times:
  - user คือ เวลาทั้งหมดที่ใช้โดย CPUs ในระหว่างการคำนวณ
  - sys คือ เวลาทั้งหมดที่ใช้โดย CPUs ในระหว่างการทำงานที่เกี่ยวข้องกับระบบ เช่น การจัดสรรหน่วยความจำ (memory allocation)
- Wall time:
  - คือ เวลาทั้งหมดที่นับตั้งแต่เริ่มรันโปรแกรมจนถึงสิ้นสุดโปรแกรมหรือความแตกต่างระหว่างเวลาที่งานเสร็จสิ้นและเวลาที่งานเริ่มต้นขึ้น

บรรทัดที่ 4 : เรียกใช้ webdriver.Chrome() โดยกำหนด Path ที่อยู่ของ webdriver คือ "./chromedriver"

บรรทัดที่ 6 : เรียกใช้ไฟล์ author.csv เพื่อดึงข้อมูล 'user\_ID' เพื่อนำ 'user\_ID' ไปใช้ในการตีง paper ในแต่ละ user\_ID นั้น ๆ

บรรทัดที่ 8 : “author\_df” สร้าง data frame ที่ใช้ในการเก็บข้อมูลของ Author ได้แก่

- 'title' ใช้ในการเก็บชื่อ paper
- 'authors' ใช้ในการเก็บรายชื่อผู้จัดทำ paper
- 'publication\_date' ใช้ในการเก็บ วันเวลาที่เผยแพร่ paper
- 'description' ใช้ในการเก็บคำอธิบายของ paper นั้น ๆ
- 'cite\_by' ใช้ในการเก็บจำนวนคนที่เข้ามาอ่าน paper

บรรทัดที่ 11 : กำหนด row = 1 เพื่อกำหนดแถวของ data frame ที่เราจะเก็บข้อมูลลงไป โดยเริ่มแถวต้นคือ 1 แล้วเพิ่มขึ้นเรื่อย ๆ เมื่อเก็บข้อมูลครบ 1 ครั้ง จำนวนแถวจะเพิ่มขึ้น จาก row +=1

บรรทัดที่ 13 : กำหนด user\_queue = 0 ใช้ในการลำดับของ author ที่เราคำนึงดึงอยู่

บรรทัดที่ 14-55 : โดยเราจะใช้ for loop เป็นคำสั่งวนซ้ำที่ใช้ควบคุมการทำงานซ้ำ ๆ ในจำนวนรอบที่แน่นอน โดยจะวนซ้ำตามจำนวนของ user\_ID โดยการทำงานคือจะเก็บ paper ของ ID ที่ 1 จนครั้ง แล้ววนไปเก็บ ID ที่ 2 จนครบแล้วทำซ้ำไปเรื่อย ๆ จนครับ ตามจำนวนของ user\_ID โดยเรามีทั้งหมด 308 user\_ID

บรรทัดที่ 18 : ใช้ webdriver ในการเปิดเว็บ browser โดยมี <https://scholar.google.com/citations?hl=en&user=> แล้วตามด้วย user\_ID ที่เราเก็บไว้ในไฟล์ author.csv" วนไปเรื่อย ๆ จนครบ user\_ID ที่เราเก็บมา

บรรทัดที่ 20 : ใช้ find\_element\_by\_class\_name('class\_name') เพื่อใช้ในการหา class name ที่มีชื่อว่า 'gs\_btnPD' หรือ เป็นปุ่มที่เราจะใช้ในการแสดงผลลัพธ์เพิ่มเติม SHOW MOER

บรรทัดที่ 22-26 : เมื่อเจอปุ่ม SHOW MOER ให้คลิกปุ่มนั้นจนกว่าปุ่มนั้นจะกดไม่ได้ หรือ แสดง paper ครบหมดแล้ว

บรรทัดที่ 24 : คลิกที่ปุ่ม SHOW MOER

บรรทัดที่ 26 : เมื่อคลิกแล้วให้รอนับ 0 ถึง 5

บรรทัดที่ 29 : ค้นหา link ที่บอกรายละเอียดของ paper นั้น ๆ

บรรทัดที่ 31 : เลือกเฉพาะ link อกมา

บรรทัดที่ 36 : แสดงคิวของ Author ที่เราทำการดึงข้อมูล

บรรทัดที่ 37 : นับคิวของ Author เพิ่มขึ้น 1 คิว

บรรทัดที่ 41-64 : นับคิวของ Author เพิ่มขึ้น 1 คิว

บรรทัดที่ 45-49 : ใช้ในการเก็บค่า Title, Authors, Publication\_date, Description, cited

บรรทัดที่ 51-60 : ใช้ในการเก็บค่า Authors, Publication date, Description, Total citations ใน class ที่มีชื่อว่า 'gsc\_vcd\_field' โดยค้นหา เช่น ถ้าคำใน class ('gsc\_vcd\_field')[0].text == 'Authors' ให้เก็บข้อมูลนั้นลงในตัวแปร Authors

บรรทัดที่ 61 : เพิ่มข้อมูลที่เราได้ลงในตัวแปร paper\_df โดยเพิ่มในรูปแบบของแกลฯ โดยใช้ .loc[row] และ row คือจำนวนแกลฯที่จะเพิ่มข้อมูลลงไป

บรรทัดที่ 62 : “row +=1” เพิ่มจำนวน row เพื่อบอก data frame ให้ขึ้น row ใหม่

บรรทัดที่ 65 : แสดงลำดับของ paper ที่ดึงข้อมูลสำเร็จแล้ว

บรรทัดที่ 66 : เมื่อครบ 1 paper และเราจะทำการหยุด โดยนับ 0 ถึง 5 และค่อยเริ่มต้นทำงานใหม่ เพื่อไม่ให้ถูกกีดกัน การเข้าถึงจาก google เนื่องจากเด็กไม่ต้องการให้ครमาร์ดึงข้อมูลบริษัทมากในเวลาเดียวกัน เพราะทำให้เพิ่มปริมาณการใช้งานไปยัง server จำนวนมหาศาลโดยไม่จำเป็น เราจึงต้อง act like humans คือ ไม่ดึงข้อมูลเร็วเกินไป ให้เหมือนมนุษย์ เป็นคนเปิดอยู่

บรรทัดที่ 67 : และเมื่อคำสั่ง for loop เป็นจริง หรือ หมายถึงเราดึงข้อมูลครบทุกคนแล้ว ให้ save file จาก data frame ที่มีชื่อว่า paper\_df ให้อยู่ในรูปแบบของ .CSV โดยมีชื่อว่า 'paper.csv'

บรรทัดที่ 68 : driver.close() ใช้เพื่อปิดหน้าต่างเว็บ browser ที่ webdriver กำลังเรียกใช้งานอยู่ [6]

## スペックคอมพิวเตอร์ที่ใช้

- Processor 2.3 GHz Quad-Core Intel Core i5
- Memory 8 GB 2133 MHz LPDDR3
- Graphics Intel Iris Plus Graphics 655 1536 MB
- OS macOS Big Sur Version 11.2.3 (20D91)

## JupyterLab

- Version 2.2.6

## Chrome

- Version 90.0.4430.93 (Official Build) (x86\_64)

## Chromedriver

- Chrome version 90

## Github

Project: การดึงข้อมูลผลงานทางวิชาการ ของอาจารย์มหาวิทยาลัยธรรมศาสตร์จาก เว็บไซต์ Google Scholar โดยใช้ กระบวนการ web scraping

Readme

No releases published  
Create a new release

No packages published  
Publish your first package

Jupyter Notebook 100.0%

[https://github.com/nattachan-ningnoi/6109656311\\_Google-Scholar-Assignment](https://github.com/nattachan-ningnoi/6109656311_Google-Scholar-Assignment)

## ผลสรุปการดำเนินการ

- เวลาที่ใช้ในการเก็บข้อมูล

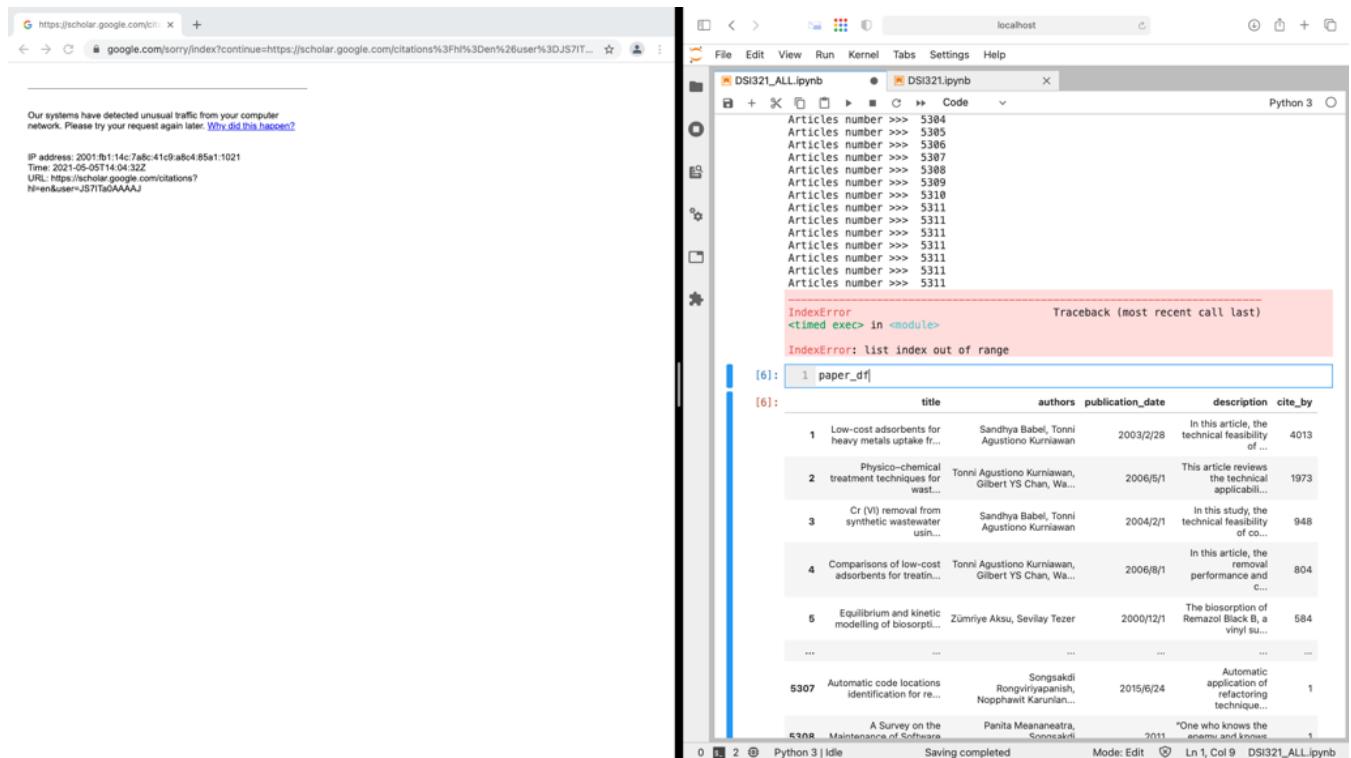
- Sheet ที่ 1 คือ Author Table ใช้เวลา 1 นาที 36 วินาที

**CPU times: user 2.98 s, sys: 215 ms, total: 3.2 s**  
**Wall time: 1min 36s**

- Sheet ที่ 2 คือ Paper Table ใช้เวลา 7 ชั่วโมง 22 นาที 43 วินาที

**CPU times: user 6min 29s, sys: 31.5 s, total: 7min 1s**  
**Wall time: 7h 22min 43s**

## ปัญหาที่พบและแนวทางการแก้ไข



- ดึงข้อมูลเร็วเกินทำให้ถูก Google Block IP

วิธีในการแก้ไขคือ ปรับเวลาในการดึงข้อมูล (time sleep) ใหม่ ให้นานขึ้นกว่าเดิม เพื่อให้การดึงข้อมูลในแต่ละครั้งใช้เวลานานขึ้น เปรียบเสมือนมนุษย์เป็นคนทำ เนื่องจากถ้าเราดึงข้อมูลเร็วจนเกินไป ทำให้ Google คิดว่า เราคือ bot หรือ คนที่ประสงค์ร้าย ต้องการให้ระบบล้ม โดยการเรียกใช้หน้าเว็บไซต์ของเรา โดยเราต้องใช้หลักการคือ act like a human คือ ทำให้เหมือนกับเราเป็นคนปกติเว็บนั้นเอง เนื่องจากถ้าเป็นคนเราต้องมีการลากเมาส์ไปกดแต่ละปุ่ม ซึ่งก็มีเวลาพอดีควรจะใช้เครื่องคอมพิวเตอร์ที่ใช้เวลาเพียงเสี้ยววินาที ทำให้เราถูกบล็อกแล้วไม่สามารถดึงข้อมูลต่อได้

## แปลงข้อมูลให้อยู่ในรูป 1NF

- การอุดแบบวิธีแปลงข้อมูล

- โดย 1NF หรือ First Normal Form มีเงื่อนไขอยู่ว่า ต้องไม่มีคอลัมน์ใดในตารางที่มีค่ามากกว่า 1 ค่า หรือที่เรียกว่า Atomic ซึ่งหมายถึง ข้อมูลที่เก็บไว้ในแต่ละคอลัมน์ต้องมีลักษณะเป็น Single Value ไม่สามารถแยกย่ออย่างได้

A	B	C	D	E	F
1	title	authors	publication_date	description	cite_by
2	1 Low-cost adsorbents for heavy metals uptake from contaminated water: a review	Sandhya Babel, Tonni Agustiono Kurniawan	2003/2/28	In this article, the technical feasibility of v	4013



A	B	C	D	E	F
1	title	authors	publication_date	description	cite_by
2	0 Low-cost adsorbents for heavy metals uptake from contaminated water: a review	Sandhya Babel	2003/2/28	In this article, the technical feasibilit	4013
3	1 Low-cost adsorbents for heavy metals uptake from contaminated water: a review	Tonni Agustiono Kurniawan	2003/2/28	In this article, the technical feasibilit	4013

### Code ที่ใช้

Prepare data into 1NF

```
[13]: 1 import pandas as pd
[14]: 1 # นำเข้าไฟล์ paper.csv
2 paper = pd.read_csv('paper.csv')
[15]: 1 # สร้าง DataFrame ใหม่ที่ใช้ในการเก็บ paper.csv ในรูปแบบ 1NF
2 paper_1NF = pd.DataFrame(columns=['title','authors','publication_date','description','cite_by'])
[16]: 1 paper_1NF
[16]: title authors publication_date description cite_by
[17]: 1 # เริ่มที่ row 0
2 rows = 0
3
4 for index, row in paper.iterrows():
5     # ทำการแยก authors ออกจาก comma (,)
6     for authors in str(row['authors']).split(','):
7         # เพิ่มชื่อผู้เขียนใน row ของ column 'title','publication_date','description','cite_by'
8         paper_1NF.loc[rows]=[row['title'],
9                           authors,row['publication_date'],
10                          row['description'],
11                          row['cite_by']]
12     # เมื่อครบ 1 Author ให้เริ่ม Author ใหม่ ห้าม row ใหม
13     rows += 1
```

- ตัวอย่าง output

paper_1NF						
	title	authors	publication_date	description	cite_by	
0	Low-cost adsorbents for heavy metals uptake fr...	Sandhya Babel	2003/2/28	In this article, the technical feasibility of ...	4013	
1	Low-cost adsorbents for heavy metals uptake fr...	Tonni Agustiono Kurniawan	2003/2/28	In this article, the technical feasibility of ...	4013	
2	Physico-chemical treatment techniques for wast...	Tonni Agustiono Kurniawan	2006/5/1	This article reviews the technical applicabil...	1973	
3	Physico-chemical treatment techniques for wast...	Gilbert YS Chan	2006/5/1	This article reviews the technical applicabil...	1973	
4	Physico-chemical treatment techniques for wast...	Wai-Hung Lo	2006/5/1	This article reviews the technical applicabil...	1973	
...	...	...	...	...	...	...
28258	กลยุทธ์ และ เครื่อง ช่วย การ สืบสาน เพื่อ สร้าง...	อดิ พล เบี้ย ชัย พันธุ์	2018	การ วิจัย เรื่อง "กลยุทธ์ และ เครื่อง ช่วย การ ส...		
28259	กลยุทธ์ และ เครื่อง ช่วย การ สืบสาน เพื่อ สร้าง...	ณัฐ มิชากร กั่ง มาลา	2018	การ วิจัย เรื่อง "กลยุทธ์ และ เครื่อง ช่วย การ ส...		
28260	การ สืบสาน เพื่อ การ พัฒนา เครื่อง ช่วย ตลาด สี...	เมือง ชัย พันธุ์	2015	บทความ ชั้น นี้ จึง มุ่ง นำเสนอ ความ เช้าใจ ...		
28261	การ สืบสาน เพื่อ การ พัฒนา เครื่อง ช่วย ตลาด สี...	อดิ พล	2015	บทความ ชั้น นี้ จึง มุ่ง นำเสนอ ความ เช้าใจ ...		
28262	การ สืบสาน เพื่อ การ พัฒนา เครื่อง ช่วย ตลาด สี...	อดิ พล เบี้ย ชัย พันธุ์	2015	บทความ ชั้น นี้ จึง มุ่ง นำเสนอ ความ เช้าใจ ...		

28263 rows x 5 columns

# Google Sheet

Three Google Sheets are displayed side-by-side:

- Sheet 1: Low-cost adsorbents for heavy metals uptake from contaminated water: a review**
- Sheet 2: Low-cost adsorbents for heavy metals uptake from contaminated water: a review**
- Sheet 3: Low-cost adsorbents for heavy metals uptake from contaminated water: a review**

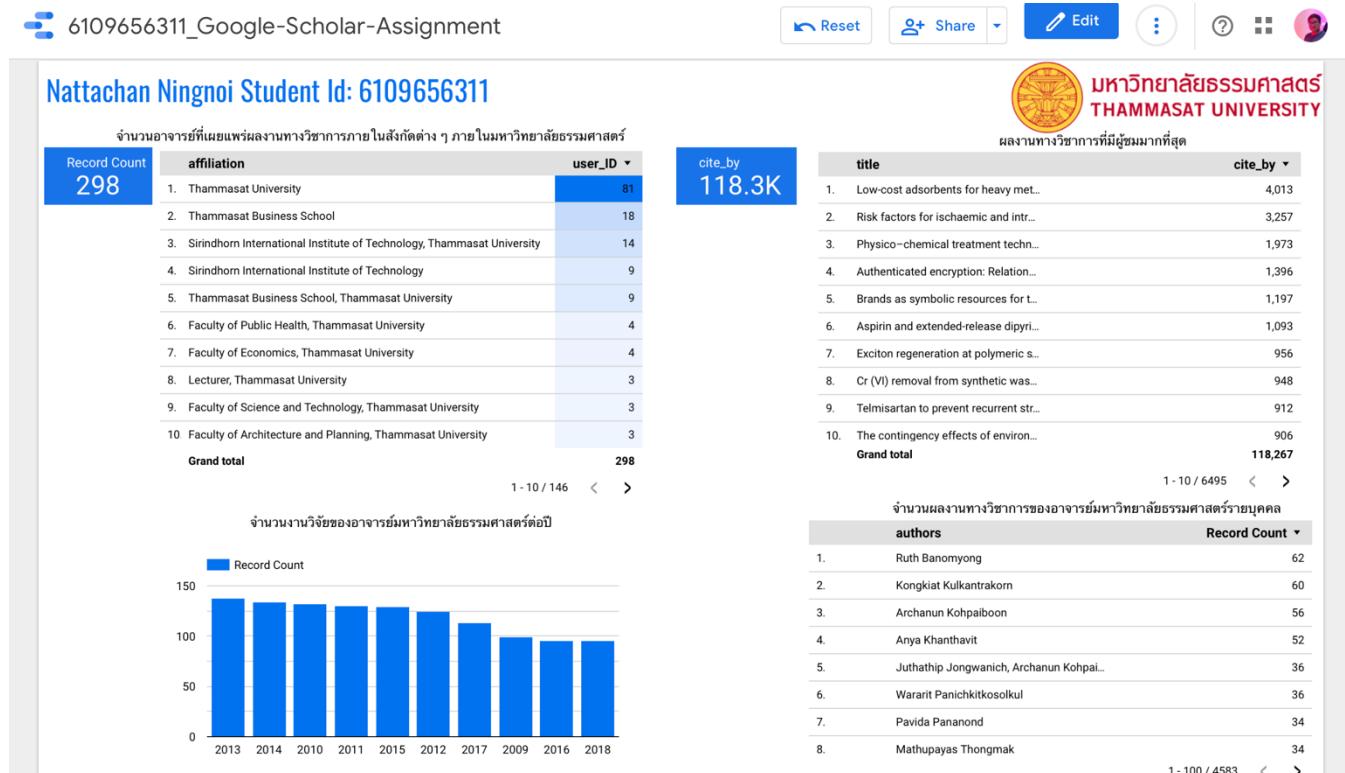
The first two sheets show identical data, while the third sheet shows slightly different data. All three sheets have columns A through F.

A	B	C	D	E	F
1	title	author	publication_date	description	cite_by
2	Low-cost adsorbents for heavy metals uptake from contaminated water: a review	Sandhya Babel, Tonni Agustina Kurniawan	2003/2/28	In this article, the technical feasibility	4013
3	Low-cost adsorbents for heavy metals uptake from contaminated water: a review	Tonni Agustina Kurniawan	2003/2/28	This article reviews the technical ap	4013
4	Physico-chemical treatment techniques for wastewater laden with heavy metals	Tonni Agustina Kurniawan	2005/5/1	This article reviews the technical ap	1973
5	Physico-chemical treatment techniques for wastewater laden with heavy metals	Gilbert YS Chan	2005/5/1	This article reviews the technical ap	1973
6	Physico-chemical treatment techniques for wastewater laden with heavy metals	Wei-Hung Lo	2005/5/1	This article reviews the technical ap	1973
7	Physico-chemical treatment techniques for wastewater laden with heavy metals	Sandhya Babel	2005/5/1	This article reviews the technical ap	1973
8	Cr (VI) removal from synthetic wastewater using coconut shell charcoal and commercial activated carbon	Sandhya Babel	2004/2/1	In this study, the technical feasibility	948
9	Cr (VI) removal from synthetic wastewater using coconut shell charcoal and commercial activated carbon	Tonni Agustina Kurniawan	2004/2/1	In this study, the technical feasibility	948
10	Comparisons of low-cost adsorbents for treating wastewater laden with heavy metals	Sandhya Babel, Tonni Agustina Kurniawan	2005/5/1	The removal performa	804
11	Comparisons of low-cost adsorbents for treating wastewater laden with heavy metals	Gilbert YS Chan	2005/5/1	In this article, the removal performa	804
12	Comparisons of low-cost adsorbents for treating wastewater laden with heavy metals	Wei-hung Lo	2005/5/1	In this article, the removal performa	804
13	Comparisons of low-cost adsorbents for treating wastewater laden with heavy metals	Sandhya Babel	2005/5/1	In this article, the removal performa	804
14	Equilibrium and kinetic modelling of biosorption of Remazol Black B by Rhizopus arrhizus in a batch syst	Zumrinya Aksu	2000/12/1	The biosorption of Remazol Black E	584
15	Equilibrium and kinetic modelling of biosorption of Remazol Black B by Rhizopus arrhizus in a batch syst	Seyvalay Tezer	2000/12/1	The biosorption of Remazol Black E	584
16	Heavy metal removal from contaminated sludge for land application: a review	Sandhya Babel	2006/1/1	In recent years, various methods fo	340
17	Heavy metal removal from contaminated sludge for land application: a review	Dominica Del Mundo Dacera	2006/1/1	In recent years, various methods fo	340
18	Comparison of various adsorbents for removing arsenic from wastewater	Sandhya Babel, Tonni Agustina Kurniawan	2005/5/1	The removal performa	169
19	Mitrofiltration membrane fouling and cake behavior during aqial filtration	Satohi Takizawa	2010/10/15	This study was carried out to investi	169
20	A matrix in life cycle perspective for selecting sustainable materials for buildings in Sri Lanka	UG Yasantha Abeywardena	2009/5/1	This paper presents a matrix to sel	152
21	A matrix in life cycle perspective for selecting sustainable materials for buildings in Sri Lanka	Sandhya Babel	2009/5/1	This paper presents a matrix to sel	152
22	A matrix in life cycle perspective for selecting sustainable materials for buildings in Sri Lanka	Shabbir Gheewala	2009/5/1	This paper presents a matrix to sel	152
23	Effect of acid speciation on solid waste liquification in an anaerobic acid digester	Sandhya Babel	2004/5/1	Laboratory scale experiments were	152
24	Effect of acid speciation on solid waste liquification in an anaerobic acid digester	Kensuke Fukushi	2004/5/1	Laboratory scale experiments were	152
25	Effect of acid speciation on solid waste liquification in an anaerobic acid digester	Bunpol Siranrasamee	2004/5/1	Laboratory scale experiments were	152
26	Factors affecting seasonal variation of membrane filtration resistance caused by Chlorella algae	Sandhya Babel	2002/5/1	A seasonal fluctuation pattern was t	152
27	Factors affecting seasonal variation of membrane filtration resistance caused by Chlorella algae	Satohi Takizawa	2002/5/1	A seasonal fluctuation pattern was t	152



<https://docs.google.com/spreadsheets/d/1Oo47nl0eJzmGOHawJN4XOLWesDJO7EeClxVfq6jig/edit?usp=sharing>

## นำเสนอในรูปแบบ Visualization บน Google data studio



## Data Storytelling

โดยจากการที่เราทำบน Dashboard บน Google data studio โดยต้องการนำเสนอถึง  
จำนวนอาจารย์ที่เผยแพร่ผลงานทางวิชาการภายใต้ชื่อ “มหาวิทยาลัยธรรมศาสตร์”

- ส่วนนี้สามารถบอกได้ว่า สำหรับในปัจจุบัน มีอาจารย์ที่ทำงานผลงานทางวิชาการมากน้อยแค่ไหน โดยมากที่สุดคือ “Thammasat Business School” และครอบคลุมสีน้ำเงินด้านซ้ายมือจะบ่งบอกถึงผลรวมของอาจารย์ในแต่ละสังกัด ภายใต้ชื่อ “มหาวิทยาลัยธรรมศาสตร์” โดย 298 คือ จำนวนอาจารย์ทั้งหมดที่เผยแพร่ผลงานทางวิชาการ

### จำนวนงานวิจัยของอาจารย์มหาวิทยาลัยธรรมศาสตร์ต่อปี

- จะบอกถึงว่าปีไหน มีผลงานทางวิชาการมากหรือน้อย โดยปีที่มากที่สุดคือ ปี 2013

### ผลงานทางวิชาการที่มีผู้อ่านมากที่สุด

- จะบอกถึงผลงานทางวิชาการอันไหนที่เป็นที่นิยม (popular) โดยผลงานที่มีผู้เข้าชมมากที่สุดอันดับ 1 คือ Low-cost adsorbents for heavy metals uptake from contaminated water: a review โดยมีผู้เข้าชมมากถึง 4013 วิว

### จำนวนผลงานทางวิชาการของอาจารย์มหาวิทยาลัยธรรมศาสตร์รายบุคคล

- จะบอกถึงว่าอาจารย์คนไหนที่มีผลงานทางวิชาการมากถึงหรือน้อย โดยคนที่ทำผลงานทางวิชาการมากที่สุดคือ Ruth Banomyong โดยมีผลงานมากที่สุดคือ 62 ผลงาน

## สรุปงานที่ได้รับมอบหมาย

1. รายละเอียดข้อมูล			
1.1	input url	เรียบร้อย	หน้า 2
1.2	Data structure (tree data structure)	เรียบร้อย	หน้า 2
1.3	output data structure (output table; authors, papers)	เรียบร้อย	หน้า 2-5
2. ขั้นตอนการเก็บข้อมูลทำอย่างไร (Study data structure of authors and papers from Google Scholar)			
2.1	การออกแบบวิธีการเก็บ	เรียบร้อย	หน้า 6-16
2.2	สเปคคอมพิวเตอร์	เรียบร้อย	หน้า 25
2.3	Code (Collect data from the website using HTTP request)	เรียบร้อย	หน้า 16-24
2.4	ผลสรุปการดำเนินการ; เวลาที่ใช้ในการเก็บข้อมูล ปัญหาที่พบและแนวทางการแก้ไข	เรียบร้อย	หน้า 26
3. การแปลงข้อมูลให้อยู่ในรูป 1NF (Prepare data into 1NF as shown in expecting output)			
3.1	การออกแบบวิธีแปลงข้อมูล	เรียบร้อย	หน้า 27
3.2	ตัวอย่าง output	เรียบร้อย	หน้า 27
4. แนบไฟล์ CSV ด้วยลิงก์ google drive (Store the output data into Google Sheet)			
4.1	Google Sheet	เรียบร้อย	หน้า 28
5. นำเสนอด้วยรูปแบบ Visualization บน Google data studio			
5.1	Google data studio	เรียบร้อย	หน้า 29

## บรรณานุกรม

1. Doctype สำหรับมาตรฐาน HTML แบบต่างๆ HTML 5, XHTML, 13 ตุลาคม 2563

<https://www.mindphp.com/%E0%B8%9A%E0%B8%97%E0%B9%80%E0%B8%A3%E0%B8%B5%E0%B8%A2%E0%B8%99%E0%B8%AD%E0%B8%AD%E0%B8%99%E0%B9%84%E0%B8%A5%E0%B8%99%E0%B9%8C/%E0%B8%9A%E0%B8%97%E0%B9%80%E0%B8%A3%E0%B8%B5%E0%B8%A2%E0%B8%99-html/2005-doctype-%E0%B8%A1%E0%B8%B2%E0%B8%95%E0%B8%A3%E0%B8%90%E0%B8%B2%E0%B8%99-.html.html>

2. HTML <!DOCTYPE> Declaration, W3Schools, [https://www.w3schools.com/tags/tag\\_doctype.asp](https://www.w3schools.com/tags/tag_doctype.asp)

3. โครงสร้างพื้นฐานของ HTML, [https://sites.google.com/site/hellomysitebylooknamkampong/hnwy-thi-1-khwam-ru-beuxng-tn-keiyw-kab-xinthexrnet/khorngsrang-phun-than-khxng-html](https://sites.google.com/site/hellomysitebylooknamkampong/hnwy-thi-1-khwam-ru-beuxng-tn-keiyw-kab-webphec/baeb-thdsxb-kxn-reiyn-hnwy-thi-1/khwam-ru-beuxng-tn-keiyw-kab-xinthexrnet/khorngsrang-phun-than-khxng-html)

4. jQuery text() Method, W3Schools, [https://www.w3schools.com/jquery/html\\_text.asp](https://www.w3schools.com/jquery/html_text.asp)

5. What do 'real', 'user' and 'sys' mean in the output of time(1)?,

<https://stackoverflow.com/questions/556405/what-do-real-user-and-sys-mean-in-the-output-of-time1>

6. Difference b/w driver.close() and driver.quit(),Kuldeep Rana,January 6, 2020,

<https://artoftesting.com/difference-between-driver-close-and-driver-quit-command-in-selenium-webdriver>