

# Microsoft Academic Graph

i-Planet International Inc.

# Microsoft Academic Graph



- Browse 10,000,000,000,000 web pages
- Read 1,000,000,000 scholarly articles
  - Period: over 1.5 centuries
- Knowledge graph contains
  - Publications (+patents)
  - Citations (+contexts)
  - Fields of Study/ Concepts
  - Authors
  - Organizations
  - Venues

# Microsoft Academic Graph(MAG)

- Microsoft Academic Graph includes Publications, Authors, Topics, Conferences, Journals, and Institutions.
- It's beneficial to releasing publications and cooperation when doing research of Econometric analysis.
- There are 24.7 million papers, including 26.1 million authors, 743,427 thousand topics, 4,523 Conferences, 48,974 journals and 25,811 institutions.



Updated: 2020/12/09

# Homogeneous Product

	Google Scholar	Microsoft Academic	BASE	CORE	Science.gov	Semantic Scholar	Baidu Scholar	Scopus
Range (million)	200	244	166	179	200	184	200	77
Abstract	✓	✓	✓	✓	✓	✓		✓
Related Articles	✓	✓		✓		✓	✓	✓
Reference	✓	✓				✓	✓	✓
Times Cited	✓	✓				✓		✓
Full-Text Link	✓	✓	✓ (Partially)	✓ (Partially)	✓ (Partially)	✓	✓	✓
Output Format	APA, MLA, Chicago, Harvard, Vancouver, RIS, BibTeX	APA, MLA, BibTeX	RIS, BibTeX	BibTeX	APA, MLA, RIS, BibTeX(Partial)	APA, MLA, Chicago, BibTeX	APA, MLA, RIS, BibTeX	

# Coverage Ratio of MAG

## Distribution of Publication Types



- There are about 98,105,723 papers in the MAG.
- It is 38.188% of total released amount.
- Remaining papers can be authorized through web crawler and other cooperative databases.

DisplayName	Publisher	PaperCount	PaperFamilyCount	CitationCount
Nature		256576	256575	16268083
Science		190340	190340	14034113
Proceedings of the National Academy of Sciences		140856	140856	14557170
Journal of Biological Chemistry		193114	193114	12809636
Journal of the American Chemical Society		189854	189854	8984203
Physical Review Letters		137989	137988	6974300
The New England Journal of Medicine		98800	98797	7338498
The Lancet		223897	223897	6140852
Physical Review B		201126	201126	5515215
JAMA		146584	146584	3937112
Applied Physics Letters		125659	125659	4346069
BMJ		268213	268213	3020603
Journal of Applied Physics		141081	141081	3486318
Journal of Chemical Physics		136577	136577	4998331
Cell		21922	21922	5508259
ChemInform		723485	723485	269232
Circulation		109077	109077	4601265
Journal of Geophysical Research		109687	109687	4667831
Cancer Research		136403	136403	3827470
Angewandte Chemie		113819	113819	3802041
Blood		135991	135991	3747198
The Astrophysical Journal		119166	119166	4705642
Journal of Immunology		81654	81654	4006878
Biochimica et Biophysica Acta		112833	112833	3592840
PLOS ONE		226977	226977	4563612
Journal of Clinical Investigation		31245	31245	3321154
Physical Review		46065	46065	1629712
Biochemistry		80341	80341	3023711
Journal of Clinical Oncology		111923	111923	2670887
Lecture Notes in Computer Science		48571	48569	428133
Physical Review D		110275	110275	2526211

Journal Conference List

# Coverage Ratio of MAG

- According to the research of Curtin University, it is revealed the database is average (Figure 1).
- The coverage ratio is higher than WoS and Scopus (Figure 2).

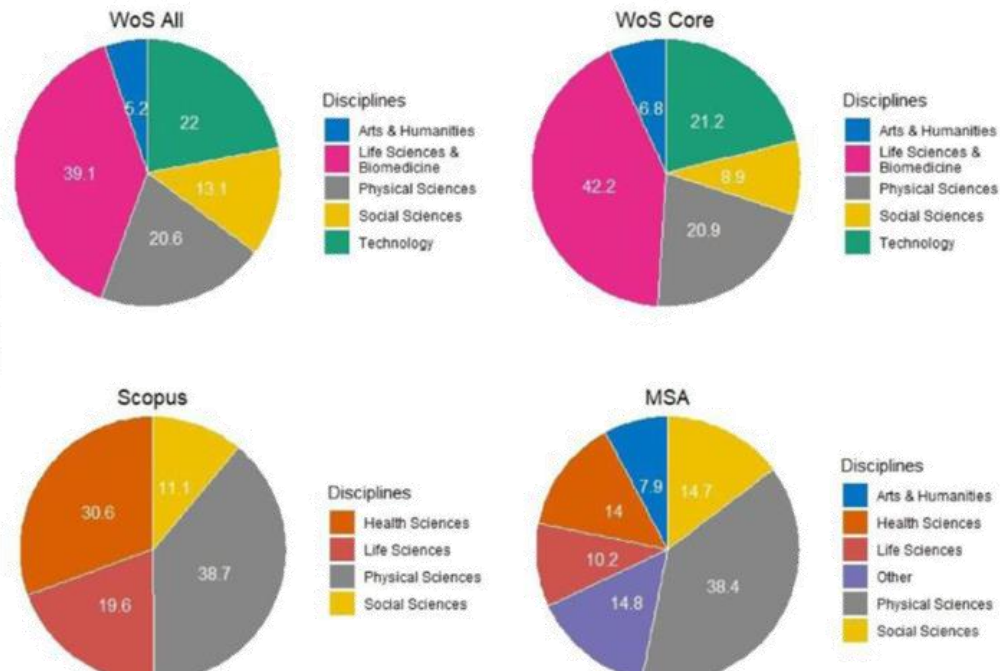


Figure 1



Figure 2

# Application

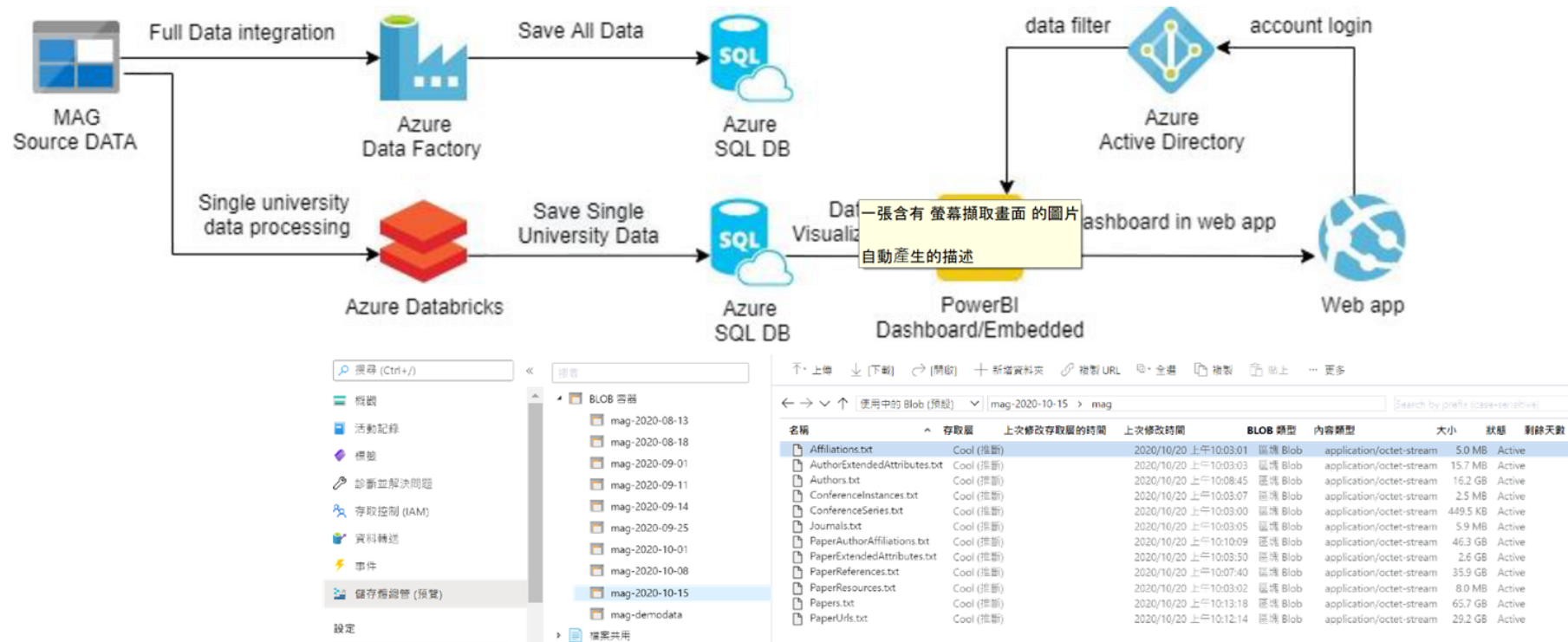
https://academic.microsoft.com/home

The screenshot displays the Microsoft Academic interface for the user 'andrew y ng'. The page is organized into several sections:

- Header:** 'Microsoft Academic' logo and user name 'andrew y ng'. Navigation icons for search, quote, and sign up/sign in are present.
- Filter Section:** 'FILTER BY:' with three main categories highlighted in red boxes:
  - Time (發布年分):** A range selector from 1982 to 2020.
  - Top Topics (主要發表領域):** A horizontal bar chart showing various topics such as 'Artificial intelligence', 'Machine learning', and 'Latent Dirichlet allocation'.
  - Publication Types (發表類型):** A horizontal bar chart showing types like 'Conference publications', 'Journal publications', and 'Patents'.
- Publication List:** A list of papers with the following details:
  - Latent dirichlet allocation:** 2003 JOURNAL OF MACHINE LEARNING RESEARCH. David M. Blei<sup>1</sup>, Andrew Y. Ng<sup>2</sup>, Michael I. Jordan<sup>1</sup>. 31,684 citations\* for all 31,605 citations\* (被引用次數). Topics: Variational message passing, Topic model, Probabilistic latent semantic analysis.
  - Latent Dirichlet Allocation:** 2001 NEURAL INFORMATION PROCESSING SYSTEMS. 79 citations\*. View More.
  - Large Scale Distributed Deep Networks:** 2012 NEURAL INFORMATION PROCESSING SYSTEMS. Jeffrey Dean, Greg Corrado, Rajat Monga, Kai Chen, Matthieu Devin. 2,547 citations\*. Topics: Stochastic gradient descent, Software framework, Multi-core processor.
- Author Profile:** Andrew Y. Ng, STANFORD UNIVERSITY. Bio: 'Andrew Yan-Tak Ng (Chinese: 吳恩達; born 1976) is a Chinese-American businessman, computer scientist, investor, and writer. He is focusing on machine learning and AI. As a businessman and investor, Ng co-founded and led Google Brain and was a former Vice President and Chief Scientist at Baidu, building... MORE'. Stats: PAPERS (250), CITATIONS\* (127,479). 'Authors with similar name' section lists co-authors like Buddha P. Roy, Mykhaylo Andriluka, and Adam Coates.

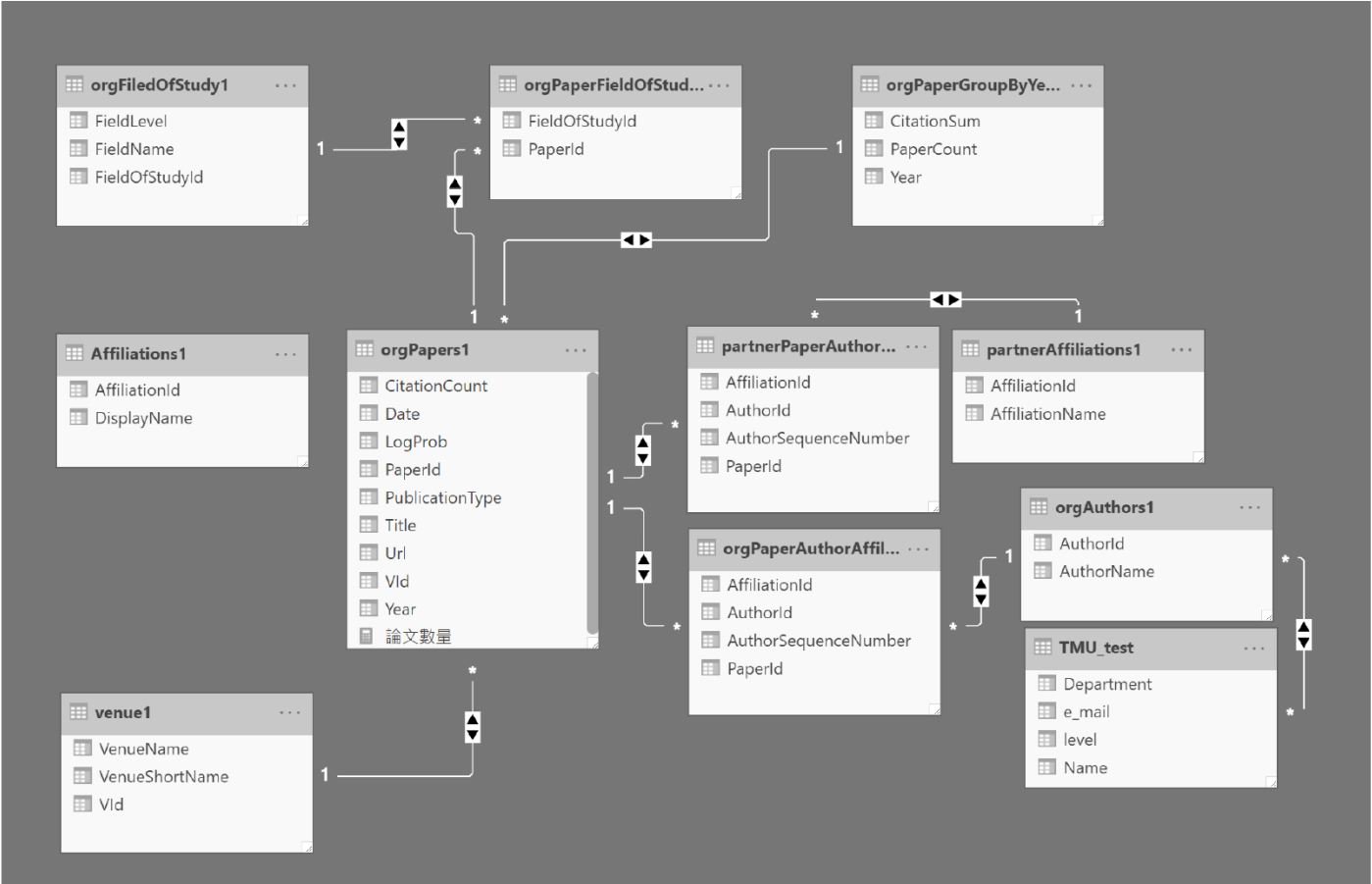
# Insight(1) – Data&Schema

- About 1 million papers are updated every month
- About 500,000 authors are updated per month





# Data & Schema



TableName	BlobDataSize
ConferenceSeries	426.07 KB
ConferenceInstances	2.44 MB
Affiliations	4.82 MB
Journals	5.72 MB
PaperResources	8.26 MB
FieldOfStudyChildren	19.34 MB
FieldOfStudyExtendedAttributes	32.56 MB
FieldsOfStudy	56.21 MB
RelatedFieldOfStudy	289.28 MB
EntityRelatedEntities	463.53 MB
PaperExtendedAttributes	2.37 GB
Authors	14.45 GB
PaperUrls	25.18 GB
PaperReferences	33.78 GB
PaperFieldsOfStudy	40.3 GB
PaperAuthorAffiliations	42.04 GB
PaperCitationContexts	48.72 GB
Papers	61.32 GB
PaperRecommendations	79.7 GB
PaperAbstractsInvertedIndex	196.43 GB

# Insight(2) – Field of Study

## Topic Filter

All > ... > Disease

### Parent Topics

🏠 Diabetes mellitus

### Current Topic

🏠 Disease

### Child Topics

🏠 Acute-phase protein    🏠 Adrenal Cortex Diseases

🏠 Age of onset    🏠 Alzheimer's disease    🏠 Amyloid

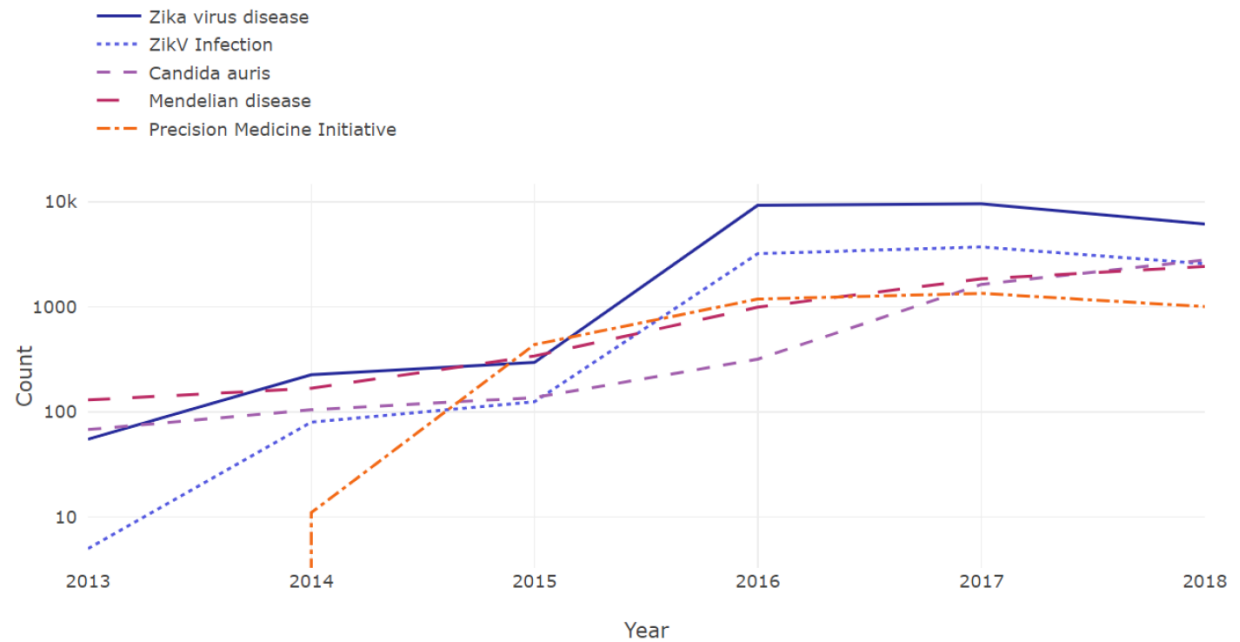
🏠 Amyotrophic lateral sclerosis    🏠 Ankylosing spondylitis

🏠 Apolipoprotein E    🏠 Arterial stiffness    🏠 Arthritis

View More (140+) ▾

## Trending Topics in Disease

Based on citation growth rate in the past 5 years.



# Field of Study Example COVID-19

- Parent Topics


- Child Topics

- Related Topics

Microsoft Academic

Search any topic, author, journal, etc. or any combination of these

Sign up / Sign in



## Coronavirus disease 2019 (COVID-19)

150,135 Publications 532,163 Citations\*

### Definition

COVID-19 is a new illness that can affect your lungs and airways. It's caused by a virus called coronavirus.

### Website Links

webmd.com

### Topics

#### Parent Topics

Infectious disease (medical specialty)

#### Related Topics

Medicine Severe acute respiratory syndrome coronavirus 2 Pandemic 2019-20 coronavirus outbreak Virology

Betacoronavirus MEDLINE Coronavirus Outbreak Coronavirus Infections Internal medicine

Viral Epidemiology Intensive care medicine China Disease Transmission Environmental health

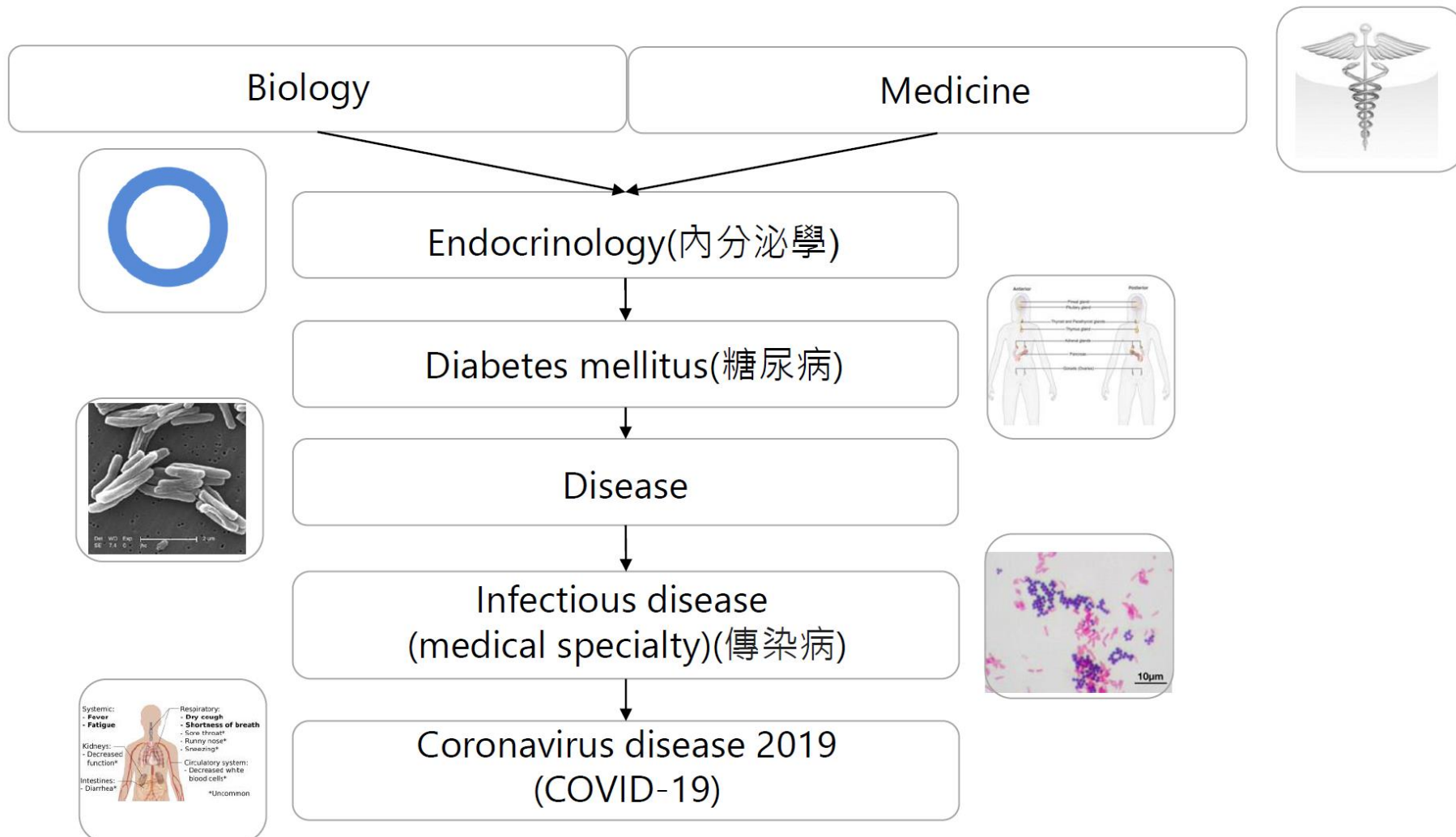
Medical emergency Health care Geography View Less

FOLLOW EXPLORE

### Publications & Citations Over Time

<https://academic.microsoft.com/topic/3008058167>

# Field of Study Example COVID-19



# Insight(3) - Data Redundancy

Different versions of the same paper will be included in the form of articles, and all the parent articles and excerpts will be cited.

The screenshot shows the Microsoft Academic search interface for "Coronavirus disease 2019 (COVID-19)". The search results are filtered by time (2000-2021) and sorted by relevance. The results list three different versions of the same paper, "Clinical Characteristics of Coronavirus Disease 2019 in China", published in 2020. The first version is in THE NEW ENGLAND JOURNAL OF MEDICINE with 9,715 citations\* for all and 9,220 citations\*. The second version is in MEDRXIV with 495 citations\*. The third version is also in THE NEW ENGLAND JOURNAL OF MEDICINE with 7,681 citations\*. Each entry includes the authors' names and affiliations, and a snippet of the abstract. The interface also shows filter options for top topics and publication types.

Microsoft Academic Coronavirus disease 2019 (COVID-19) Sign up / Sign in

PUBLICATIONS AUTHORS CONFERENCES JOURNALS INSTITUTIONS

FILTER BY: 1-10\* of 50,000+ VIEW SORT BY RELEVANCE

**Time**  
2000-2021

**Top Topics**

- Coronavirus disease 2019 (COVI...)
- Medicine
- Severe acute respiratory syndro...
- Pandemic
- 2019-20 coronavirus outbreak
- Virology
- Betacoronavirus
- Coronavirus
- MEDLINE
- Outbreak

**Publication Types**

- Journal publications

**Clinical Characteristics of Coronavirus Disease 2019 in China.** 9,715 citations\* for all  
2020 THE NEW ENGLAND JOURNAL OF MEDICINE 9,220 citations\*  
Wei-jie Guan, Zheng-yi Ni, Yu Hu, Wen-hua Liang, Chun-quan Ou see all 37 authors  
Guangzhou Medical University  
Pandemic Betacoronavirus View More (8+)

Abstract Background Since December 2019, when coronavirus disease 2019 (Covid-19) emerged in Wuhan city and rapidly spread throughout China, data have been needed on the clinical characteristics of...

**Clinical characteristics of 2019 novel coronavirus infection in China** 495 citations\*  
2020 MEDRXIV View Less  
Wei-jie Guan<sup>1</sup>, Zheng-yi Ni<sup>1</sup>, Yu Hu<sup>2</sup>, Wen-hua Liang<sup>1</sup>, Chun-quan Ou<sup>3</sup> see all 37 authors  
<sup>1</sup> Guangzhou Medical University, <sup>2</sup> Huazhong University of Science and Technology, <sup>3</sup> Southern Medical University  
Intensive care unit Hazard ratio View More (8+)

Abstract Background Since December 2019, when coronavirus disease 2019 (Covid-19) emerged in Wuhan city and rapidly spread throughout China, data have been needed on the clinical characteristics of...

**Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia.** 7,681 citations\*  
2020 THE NEW ENGLAND JOURNAL OF MEDICINE  
Qun Li, Xuhua Guan, Peng Wu, Xiaoye Wang, Lei Zhou see all 45 authors

Feedback

# Chinese Name

教授中文名	教授英文名	大學cng	大學	學院	系所
古安	Y.U. Ku	National Chengchi University	政治大學	文學院	中國文學系
古安	Ku, Y.U.	National Chengchi University	政治大學	文學院	中國文學系
古安	Ku, Yu-An	National Chengchi University	政治大學	文學院	中國文學系
古安	Yu-An Ku	National Chengchi University	政治大學	文學院	中國文學系
呂明	C.H. Lu	National Chengchi University	政治大學	文學院	中國文學系
呂明	Lu, C.	National Chengchi University	政治大學	文學院	中國文學系
呂明	Lu, Chao-Ming	National Chengchi University	政治大學	文學院	中國文學系
呂明	Chao Ming Lu	National Chengchi University	政治大學	文學院	中國文學系
宋珊	Y.U. Sung	National Chengchi University	政治大學	文學院	中國文學系
宋珊	Sung, Y.U.	National Chengchi University	政治大學	文學院	中國文學系
宋珊	Sung, Yun-Shan	National Chengchi University	政治大學	文學院	中國文學系
宋珊	Yun-Shan Sung	National Chengchi University	政治大學	文學院	中國文學系
車健	C. Che	National Chengchi University	政治大學	文學院	中國文學系
車健	Che, C.	National Chengchi University	政治大學	文學院	中國文學系
車健	Che, Hsing-Chien	National Chengchi University	政治大學	文學院	中國文學系
車健	Che, Hsing-Chien	National Chengchi University	政治大學	文學院	中國文學系
周煌	C.H. Chou	National Chengchi University	政治大學	文學院	中國文學系
周煌	Chou, C.H.	National Chengchi University	政治大學	文學院	中國文學系
周煌	Chou, Chih-Huang	National Chengchi University	政治大學	文學院	中國文學系

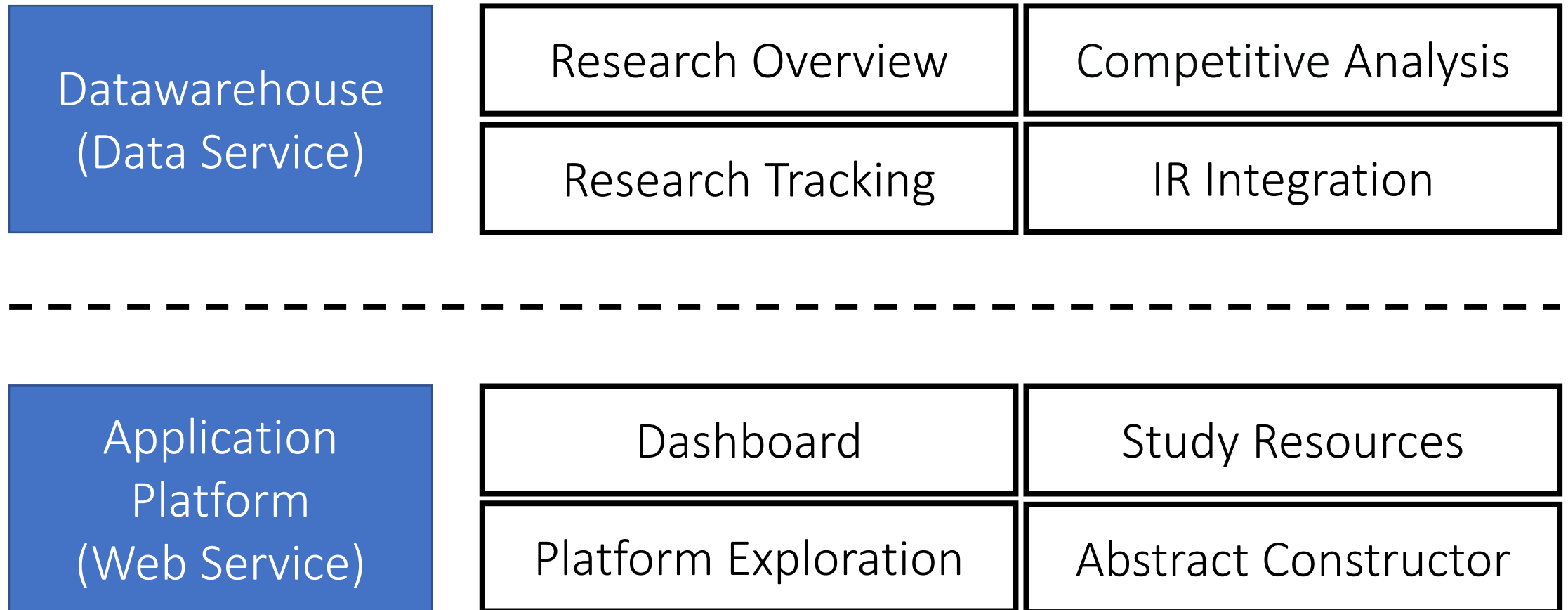
- MAG provides 690,000 Chinese name references for of authors.
- It can be quickly built if Taiwan allows.

AuthorExtendedAttributes - 記事本

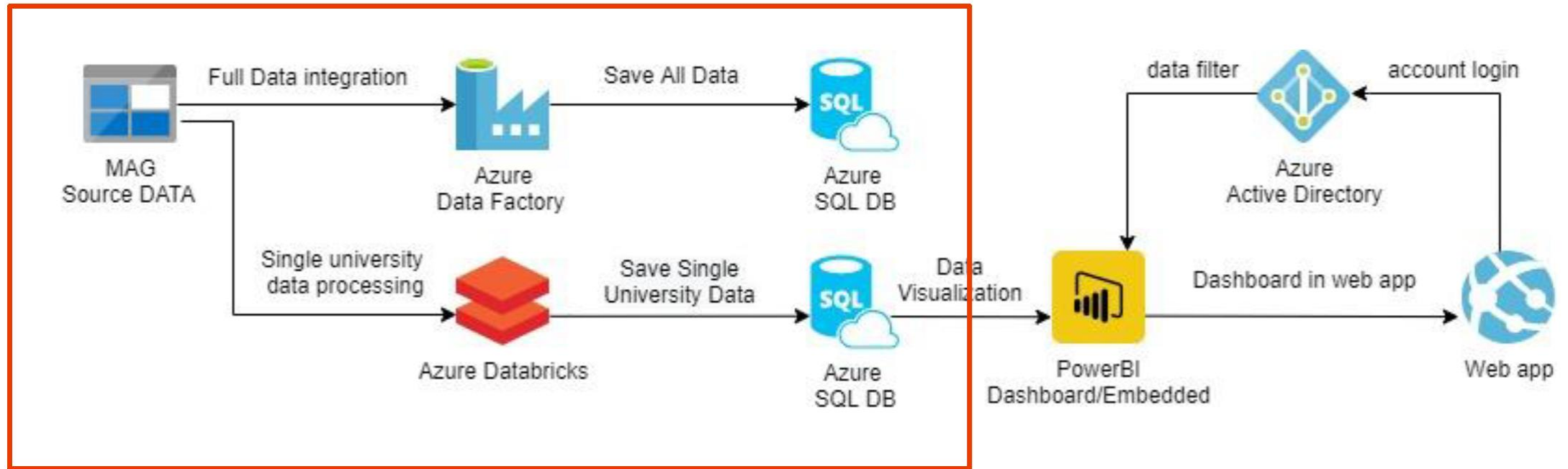
檔案(E) 編輯(E) 格式(O) 檢視(V) 說明

檔案(E)	編輯(E)	格式(O)	檢視(V)	說明
4038052	1			郝均
13772676	1			王璋
19895283	1			王汤
20325981	1			王宇
28950106	1			王陈
35168414	1			王陈
37172629	1			王陈
38640429	1			王陈
45874883	1			王陈
51593008	1			王陈
55853649	1			王陈
60152244	1			王陈
103228277	1			王陈
106220589	1			王陈
107140170	1			王陈
115771842	1			王陈
121149887	1			王陈
124705081	1			王陈
126743660	1			王陈
128231162	1			王陈
130531935	1			王陈
131354520	1			王陈
142676925	1			王陈
146148284	1			王陈
152148550	1			王陈
154491240	1			王陈
159770166	1			王陈
170513195	1			王陈
184488143	1			王陈
186301161	1			王陈
188718954	1			王陈
189800579	1			王陈

# Service - Basic



# Service Infrastructure



- SQL Sources (500MB)
- PBI Report (3 Pages)



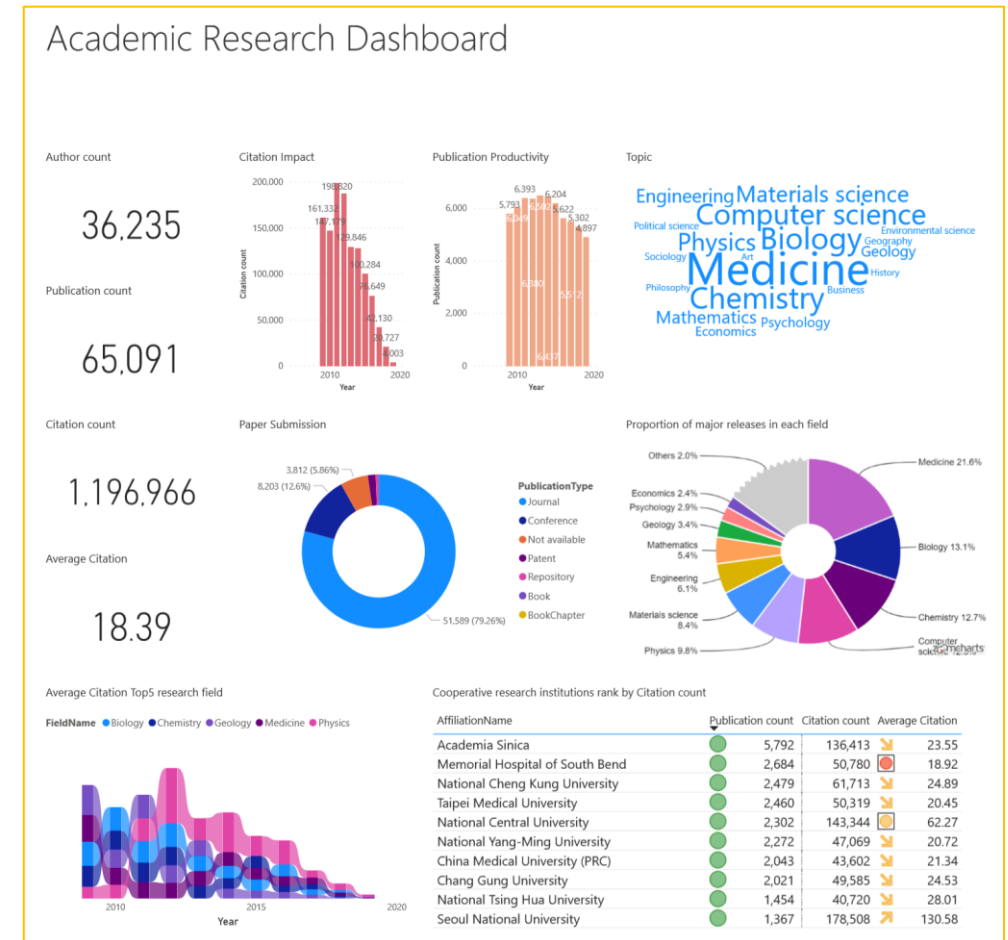
2 hrs PBI Training



# Achievement Dashboard

Track the research situation of researchers, including the following statistical values:

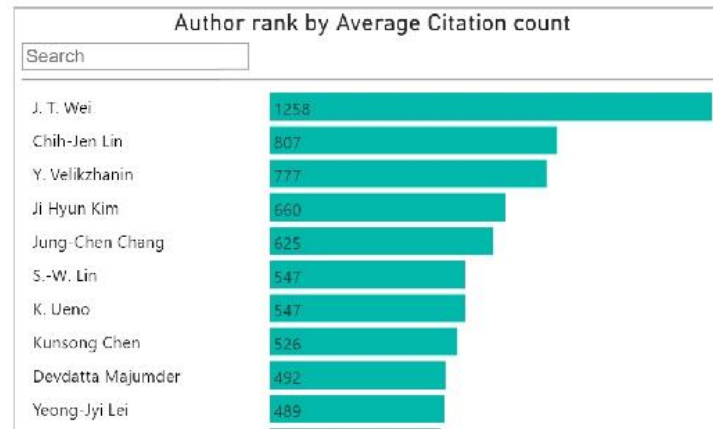
- Total number of researchers
- Number of papers published
- Citations
- Average citations of papers
- Papers published in recent years
- Paper cited in recent years
- Distribution and proportion of main research areas
- Output of cooperative research
- Top5 cited cases



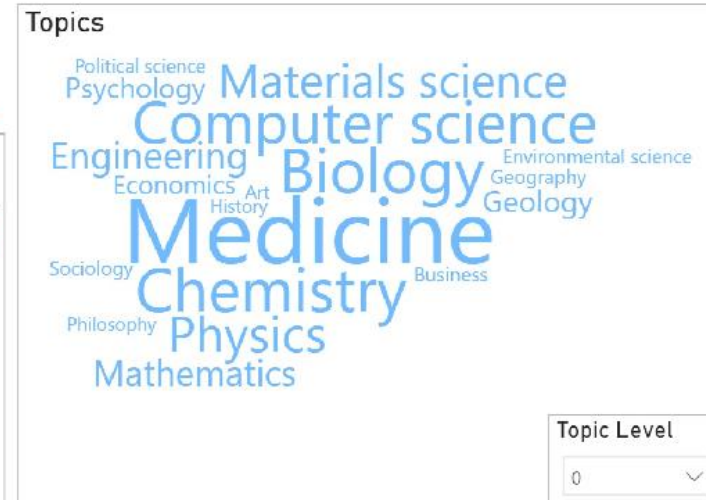
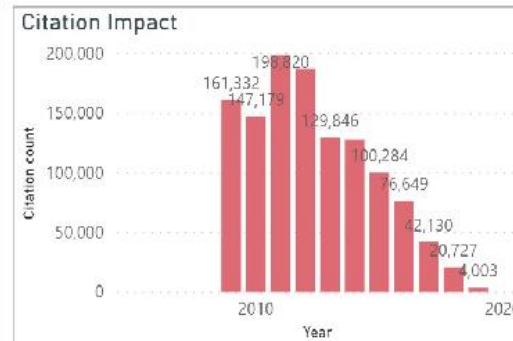
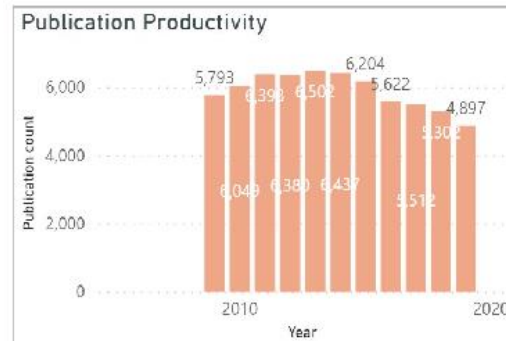
# Achievement Dashboard (1)

## – Research Overview

### Academic Research Overview



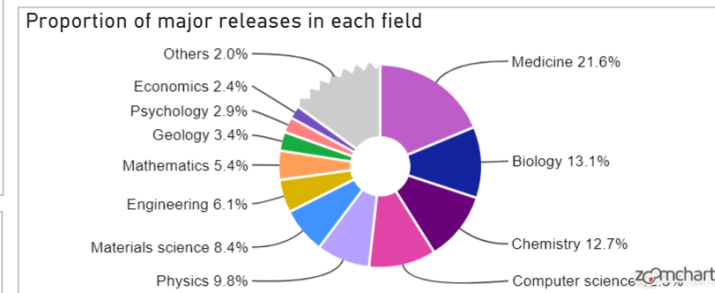
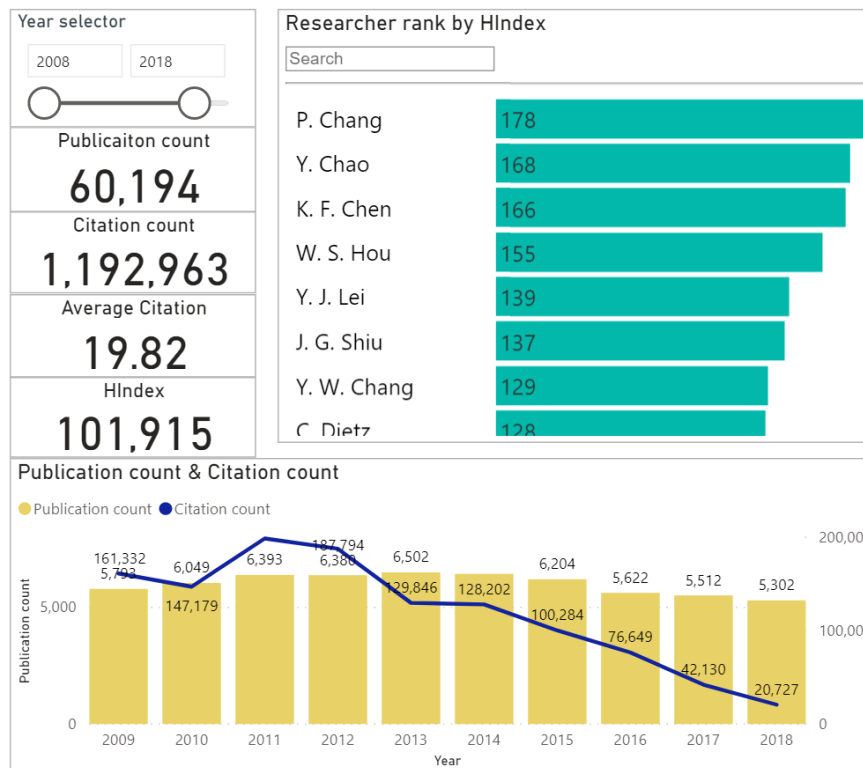
**i** Only when the number of publication is greater than 10 will be included in this ranking.



# Achievement Dashboard (2)

## – Researcher Profile

### Researcher Profile



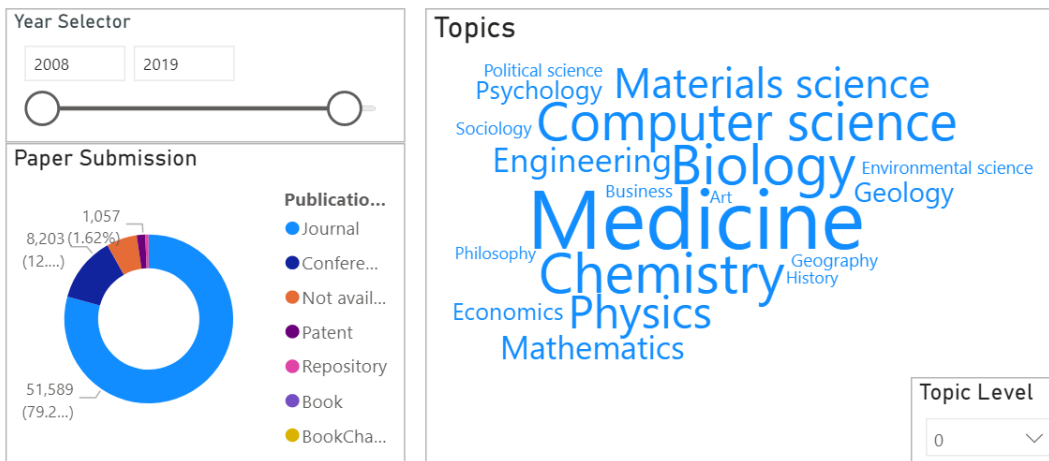
Paper rank by Citation count

Url	Year	Citation count	Title
<a href="#">2011</a>	2011	42,284	libsvm a library for support vector machines
<a href="#">2012</a>	2012	17,659	observation of a new boson at a mass of 125 gev with the cms e
<a href="#">2016</a>	2016	10,957	planck 2015 results xiii cosmological parameters
<a href="#">2014</a>	2014	7,894	global regional and national prevalence of overweight and obes
<a href="#">2009</a>	2009	7,645	gefitinib or carboplatin paclitaxel in pulmonary adenocarcinoma

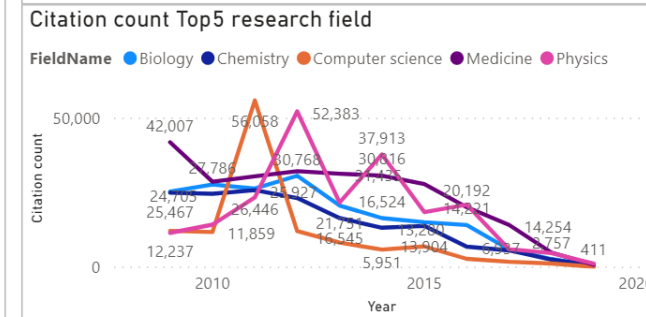
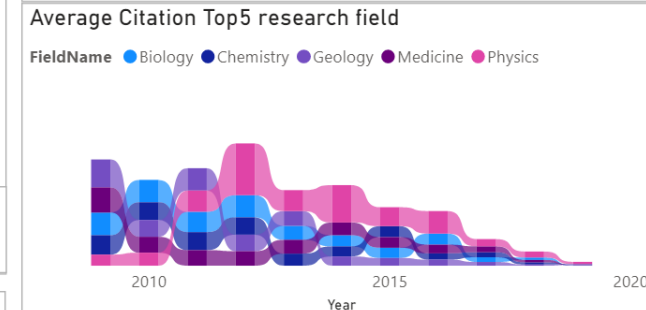
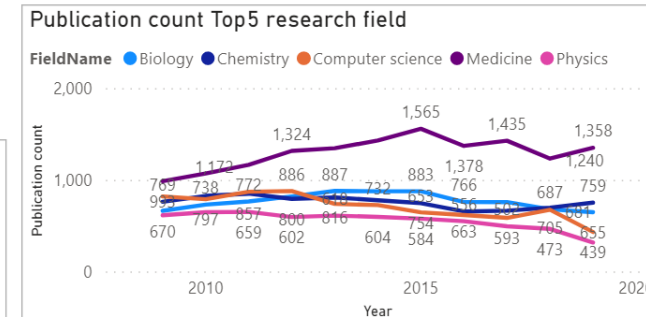
# Achievement Dashboard (3)

## – Paper Submission Main Venue

### Paper Submission main Venue



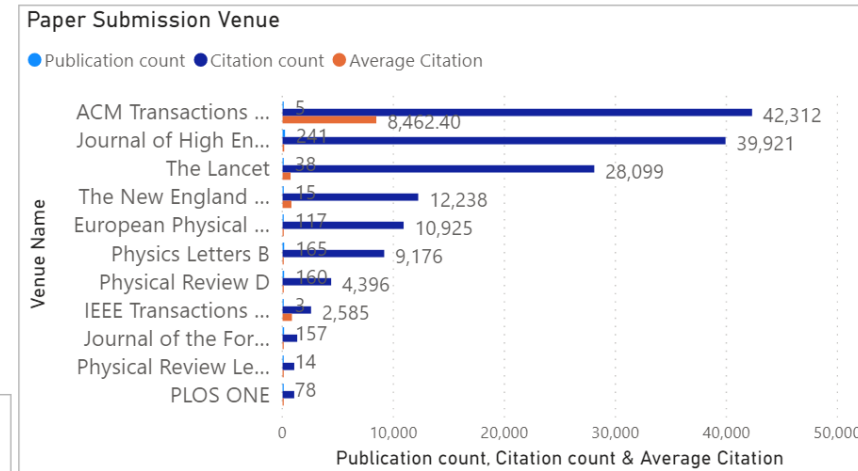
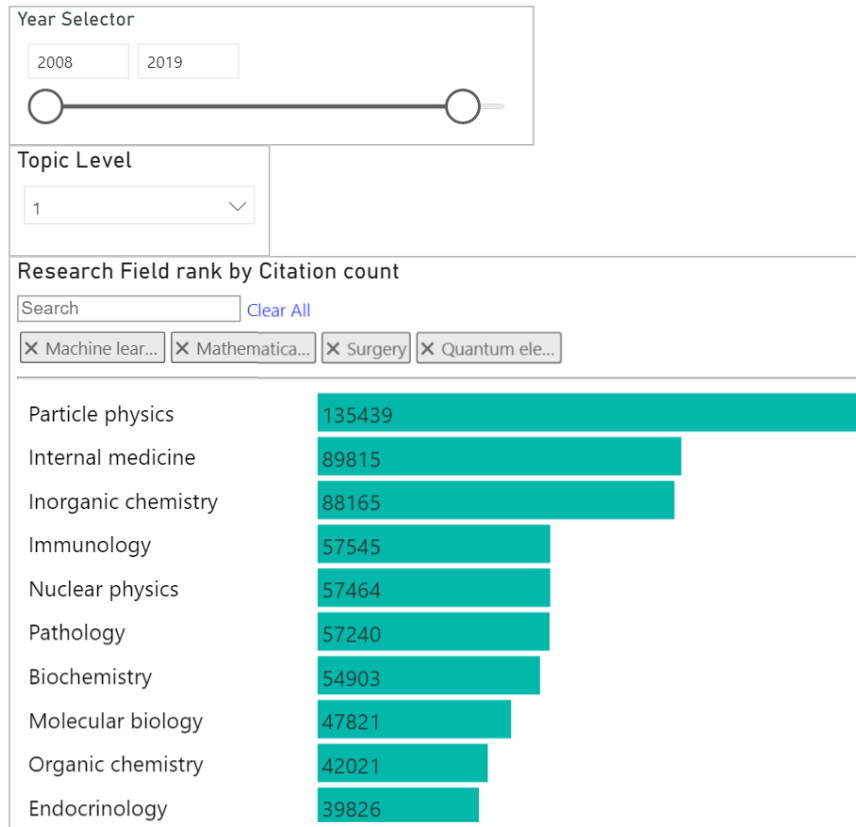
VenueName	Publication count	Citation count	Average Citation
3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video	3	37	12.33
Aaps Journal	3	24	8.00
AASRI Procedia	1	2	2.00
Abacus	2	11	5.50
Abdominal Radiology	1	1	1.00
Abstract and Applied Analysis	1	0	0.00
<b>Total</b>	<b>60,202</b>	<b>1,184,973</b>	<b>19.68</b>



# Achievement Dashboard (4)

## – Research Field Observation

### Research field Observation



Author paper Citation count

Url	AuthorName	Year	CitationCount	Title
<a href="#">🔗</a>	Chih-Chung Chang	2011	42284	libsvm a library for support vector mac
<a href="#">🔗</a>	Chih-Jen Lin	2011	42284	libsvm a library for support vector mac
<a href="#">🔗</a>	Jung-Chen Chang	2014	7894	global regional and national prevalenc adults during 1980 2013 a systematic r
<a href="#">🔗</a>	Chih-Hsin Yang	2009	7645	gefitinib or carboplatin paclitaxel in pu
<a href="#">🔗</a>	J. Alwall	2014	6502	the automated computation of tree lev sections and their matching to parton
<a href="#">🔗</a>	J. G. Shiu	2011	4408	measurement of the inclusive w and z sqrt s 7 tev with the cms experiment
<a href="#">🔗</a>	K. F. Chen	2011	4408	measurement of the inclusive w and z sqrt s 7 tev with the cms experiment

# Achievement Dashboard (5)

## – Cooperative Research Institutions

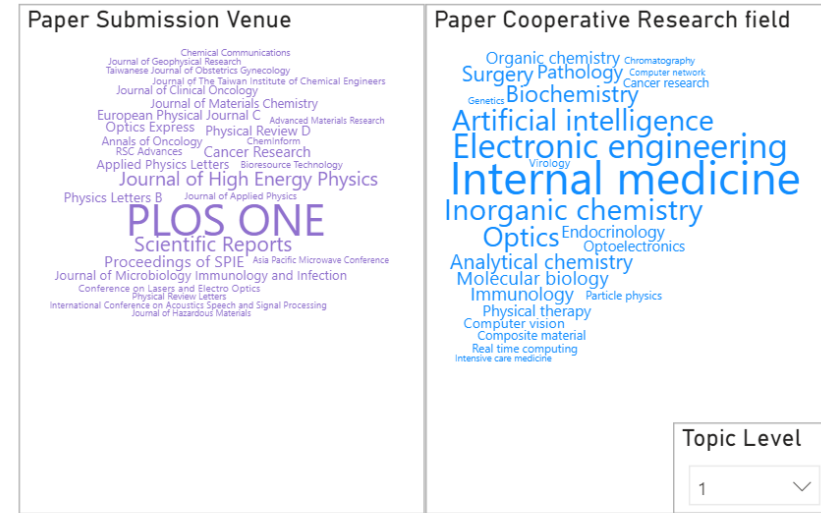
Cooperative research institutions and output situation

**Year Selector**

2008 2019

**Affiliation query**

AffiliationName



Cooperative research institutions rank by Citation count

AffiliationName	Publication count	Citation count	Average Citation
Academia Sinica	5,792	136,413	23.5
Memorial Hospital of South Bend	2,684	50,780	18.9
National Cheng Kung University	2,479	61,713	24.9
Taipei Medical University	2,460	50,319	20.4
National Central University	2,302	143,344	62.3
National Yang-Ming University	2,272	47,069	20.7
China Medical University (PRC)	2,043	43,602	21.3
Chang Gung University	2,021	49,585	24.5
National Tsing Hua University	1,454	40,720	28.0
Seoul National University	1,367	178,508	130.5

Paper Cooperative Research rank by Citation count

Url	Year	CitationCount	Title
<a href="#">libsvm</a>	2011	42284	libsvm a library for support vector machines
<a href="#">boson</a>	2012	17659	observation of a new boson at a mass of 12 experiment at the lhc
<a href="#">planck</a>	2016	10957	planck 2015 results xiii cosmological parameters
<a href="#">prevalence</a>	2014	7894	global regional and national prevalence of celiac disease in children and adults during 1980-2013 a systematic review and the global burden of disease study 2013
<a href="#">gefitinib</a>	2009	7645	gefitinib or carboplatin paclitaxel in pulmonary adenocarcinoma: a phase 3 placebo-controlled study in patients with no prior systemic therapy
<a href="#">tree level</a>	2014	6502	the automated computation of tree level and loop level differential cross sections and their matching to nll order: the case of

# Practical Case– COVID-19

Using abstracts, keywords, and titles to search relevant documents related to COVID-19.

## MAG COVID19 論文收錄

**論文發布日期**

2020/1/1 2020/7/23

**論文發布數量**

**95,515**

**論文篩選方法**

Abstract

FOS

Title

**出版商排序**

Publisher	PaperCount	PaperCount 的 %GT
Cold Spring Harbor Laboratory Press	5217	8.03%
Elsevier	4434	6.83%
Springer Science and Business Media LLC	3499	5.39%
John Wiley & Sons, Ltd	2872	4.42%

**COVID19 論文發布狀態**

CreatedDate 月	PaperCount
2020 Jan...	184
2020 Febr...	602
2020 March	4070
2020 April	11387
2020 May	22134
2020 June	21899
2020 July	18910

**論文發布位置**

Medical Hypotheses, Chemical Engineering News, International Journal of Environmental Research and Public Health, The New England Journal of Medicine, BMC Journal Watch, Authorea Preprints, Current Science, Dermatology Therapy, Journal of Medical Virology, Critical Care, JAMA The Journal of the American Medical Association, Clinical Infectious Diseases, medRxiv, Infection Control and Hospital Epidemiology, Nature, BMJ, Journal of Infection, arXiv Preprint, Population and Evolution, International Journal of Infectious Diseases, arXiv Preprint, Physics and Society, Emerging Infectious Diseases, Science of The Total Environment

**FOS, Title 與 Abstract 論文發布文氏圖**

Legend: FOS (82K, 86%), Title (79K, 83%), Abstract (51K, 53%)

## MAG COVID19 論文收錄

**論文總和數量**

**95,515**

**該項論文總和數量**

**81,972**

**論文關鍵字最高項目論文數量**

**58,865**

**篩選方式**

多重篩取項目

**資料最後更新日期**

2020年7月23日

**COVID19論文關鍵字**

**論文關鍵字出現次數排序**

Description	Count
coronavirus	69,583
Coronavirus disease 2019 (COVID-19)	58,865
covid 19	55,089
covid-19	31,274
Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)	29,913
2019-20 coronavirus outbreak	21,525
sars-cov-2	9,529

**COVID19論文關鍵句**

Severe acute respiratory syndrome coronavirus 2 SARS CoV 2  
 coronavirus  
 2019-20 coronavirus outbreak  
 Coronavirus disease 2019 COVID 19

**論文關鍵字占比**

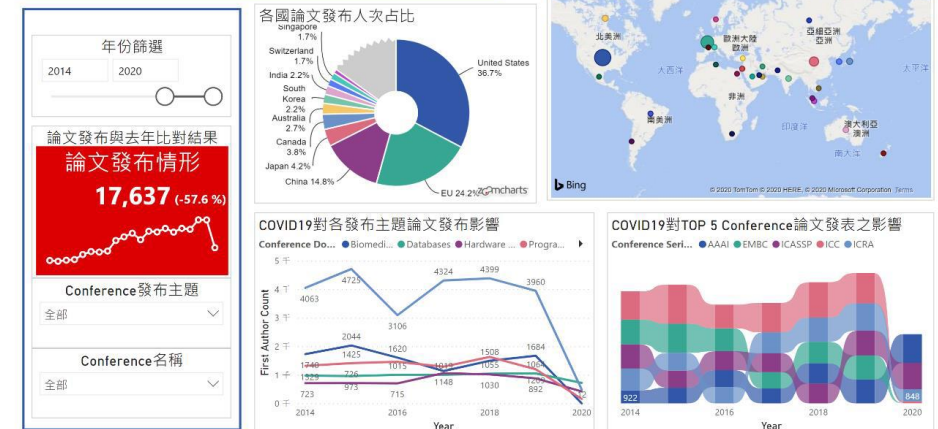


# Practical Case– COVID-19

Take the field of information engineering as an example

- There are 105 most influential CS seminars and conferences. Their data are combined by MAG.
- Those can be used to observe the impact of COVID-19 on paper submissions this year.

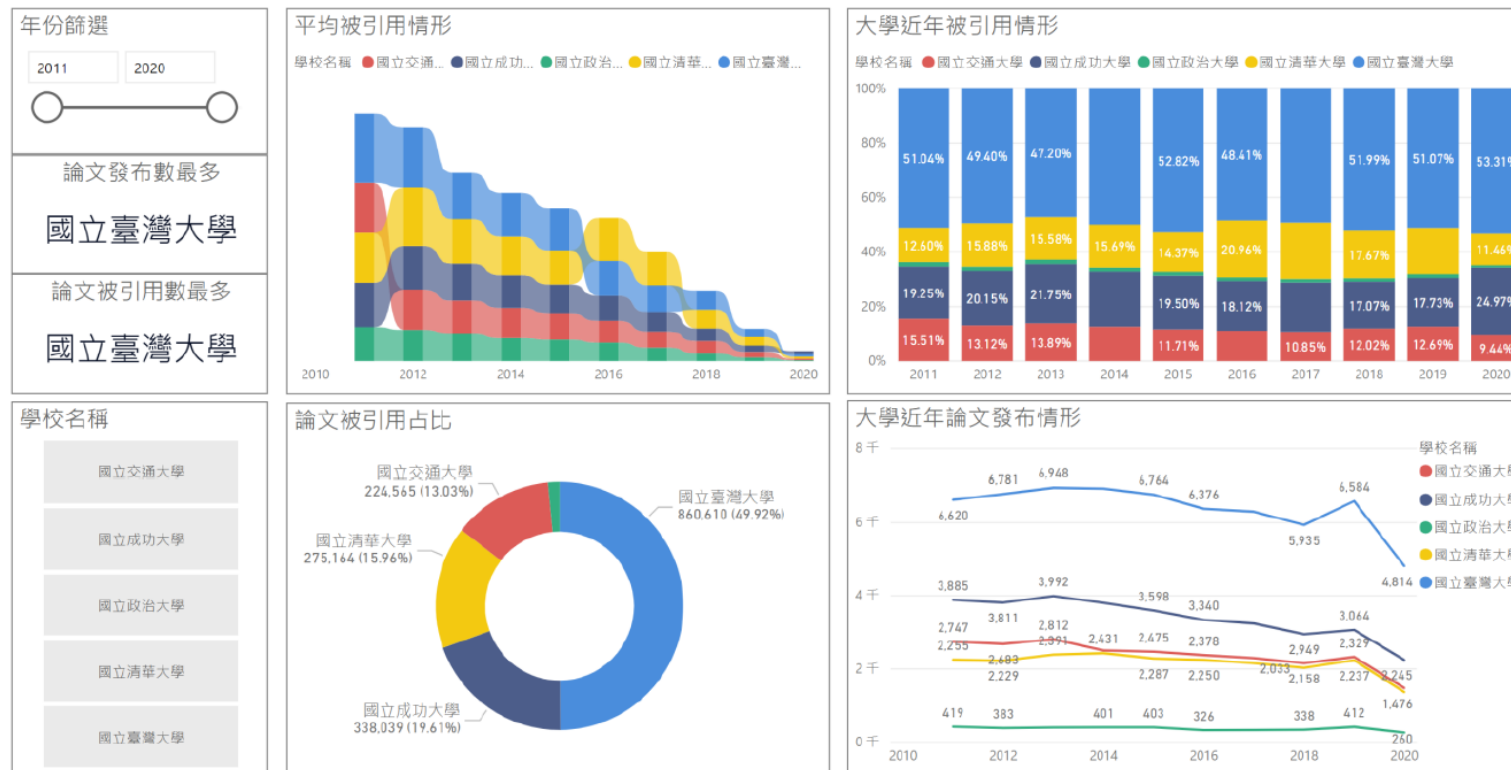
Impact of COVID-19 on the Computer Science Research Community





# Practical Case – Competition Analysis

Find out the data of five top universities in Taiwan from MAG, and compare the output of academic papers.



# Basic Program

This 1 week Azure Microsoft Academic Graph training service helps customers understanding how to use MAG and understanding what are big data processing, visualizing achievement data and producing Power BI dashboard.

- Setting Up Microsoft Azure Cloud Services (Blob /Data factory/Databricks/SQL)
- To sign up for MAG
- End-user personas and their needs
- Analysis of MAG data
- Analysis of planning for data mapping and key performance indicators (KPIs) from MAG data
- Define customer user acceptance testing (UAT) requirements

# Advanced Program

This four-week implementation offering provides a quick on-ramp to design, develop, and deploy Microsoft Power BI dashboards according to Microsoft Academic Graph data and reports for your key decision-makers.

## **Week 1**

I-Planet international ltd. will conduct a three-day workshop that will focus on:

- Setting Up Microsoft Azure Cloud Services
- To sign up for MAG
- End-user personas and their needs
- Analysis of MAG data
- Analysis of planning for data mapping and key performance indicators (KPIs) from MAG data
- Define customer user acceptance testing (UAT) requirements

## **Week 2 – 3**

Weeks 2 and 3 will focus on the following activities:

- ETL jobs for MAG data
- I-Planet international ltd. will provide services to design and develop Microsoft Power BI reports
- User acceptance testing with key stakeholder to ensure data accuracy and integrity in reports

## **Week 4**

Activities for week 4 include:

- Deployment and configuration of Power BI reports and dashboards
- Provide operational guidance so that IT can manage reports within the deployed environment

Thank You