

Empowering Bibliometric Analysis with Python and API

Henry Guanwen Zhang^a,
Thane Chambers^a,
Joseph Marchand^b

^aUniversity of Alberta Library

^bDept. of Economics, University of Alberta
{guanwen, thane, jmarchan}@ualberta.ca

A
G
E
N
D
A

01

Overview of Bibliometrics

02

API & Python

03

Applications of API & Python to BA

04

Discussion

Bibliometrics

- Methodology
 - Statistical Analysis
 - Quantitative Evaluation
- Scholarly Publications
 - Articles
 - Reviews,
 - Books, etc
- Goal
 - “to measure influence or impact”



Bibliographic Data

- Scopus
- Web of Science
- Dimensions
- OpenAlex
- Google Scholar

Transformation

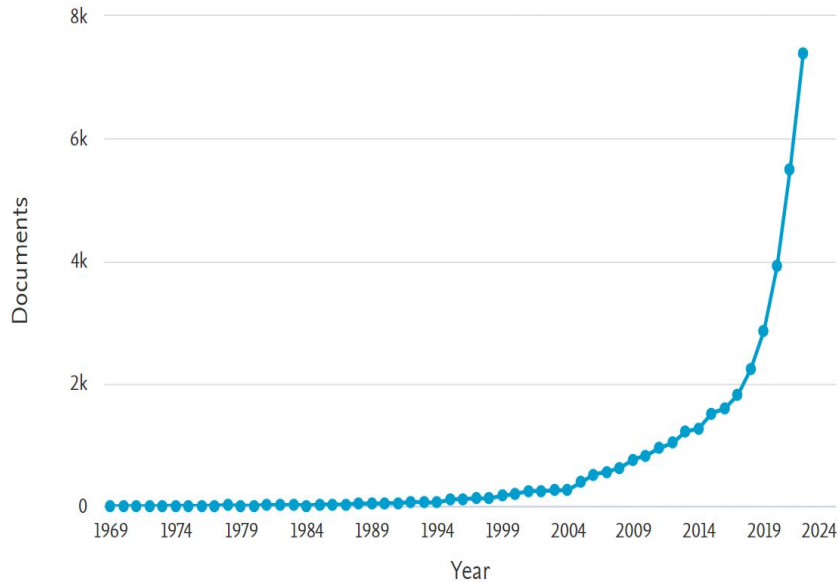
- Count
- Aggregate
- Average
- Mathematical Models

Bibliometrics

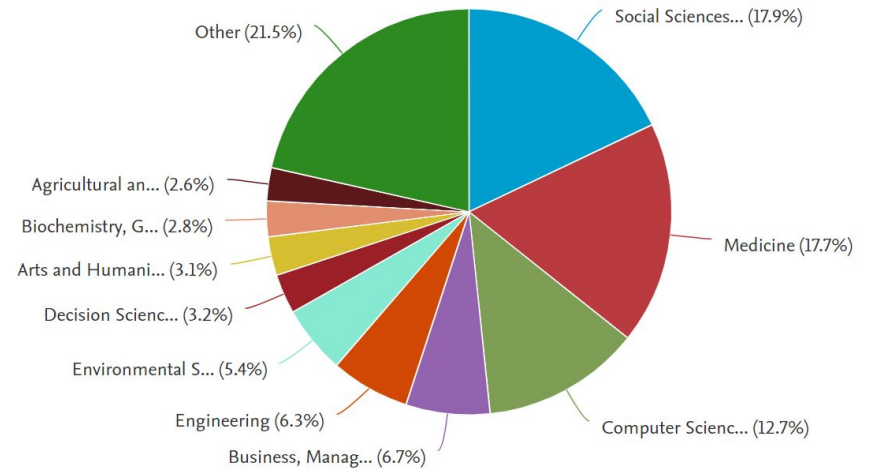
- Productivity
- Impact
- Collaboration
- Open Access
- Interdisciplinarity
- Alternatives

Bibliometrics - Relevance & Significance

Documents by year



Documents by subject area



DB: Scopus; Query: "Bibliometric*" OR "Bibliometric Analysis"

How do we approach
Bibliometric Analysis
?

Bibliometrics - Author Search on Scopus

Start exploring

Discover the most reliable, relevant, up-to-date research. All in one place.

 Documents  **Authors**  Researcher Discovery ^{Pilot}  Affiliations

Search authors using: Author name ORCID Keyword ^{Pilot}

Enter last name *

Marchand

Enter first name

Joseph

Enter affiliation name

University of Alberta

Search 

Bibliometrics - Export Authors' Publications

You have chosen to export 22 documents

Select your method of export

- MENDELEY ExLibris RefWorks SciVal RIS Format
EndNote,
Reference Manager CSV
Excel BibTeX Plain Text
ASCII in HTML

What information do you want to export?

- | | | | | |
|--|---|---|--|--|
| <input type="checkbox"/> Citation information | <input type="checkbox"/> Bibliographical information | <input type="checkbox"/> Abstract & keywords | <input type="checkbox"/> Funding details | <input type="checkbox"/> Other information |
| <input checked="" type="checkbox"/> Author(s) | <input type="checkbox"/> Affiliations | <input type="checkbox"/> Abstract | <input type="checkbox"/> Number | <input type="checkbox"/> Tradenames & manufacturers |
| <input checked="" type="checkbox"/> Author(s) ID | <input type="checkbox"/> Serial identifiers (e.g. ISSN) | <input checked="" type="checkbox"/> Author keywords | <input type="checkbox"/> Acronym | <input type="checkbox"/> Accession numbers & chemicals |
| <input checked="" type="checkbox"/> Document title | <input type="checkbox"/> PubMed ID | <input checked="" type="checkbox"/> Index keywords | <input type="checkbox"/> Sponsor | <input type="checkbox"/> Conference information |
| <input checked="" type="checkbox"/> Year | <input type="checkbox"/> Publisher | | <input type="checkbox"/> Funding text | <input type="checkbox"/> Include references |
| <input checked="" type="checkbox"/> EID | <input type="checkbox"/> Editor(s) | | | |
| <input checked="" type="checkbox"/> Source title | <input type="checkbox"/> Language of original document | | | |
| <input type="checkbox"/> volume, issue, pages | <input type="checkbox"/> Correspondence address | | | |
| <input checked="" type="checkbox"/> Citation count | <input type="checkbox"/> Abbreviated source title | | | |
| <input type="checkbox"/> Source & document type | | | | |
| <input type="checkbox"/> Publication Stage | | | | |
| <input type="checkbox"/> DOI | | | | |
| <input checked="" type="checkbox"/> Open Access | | | | |

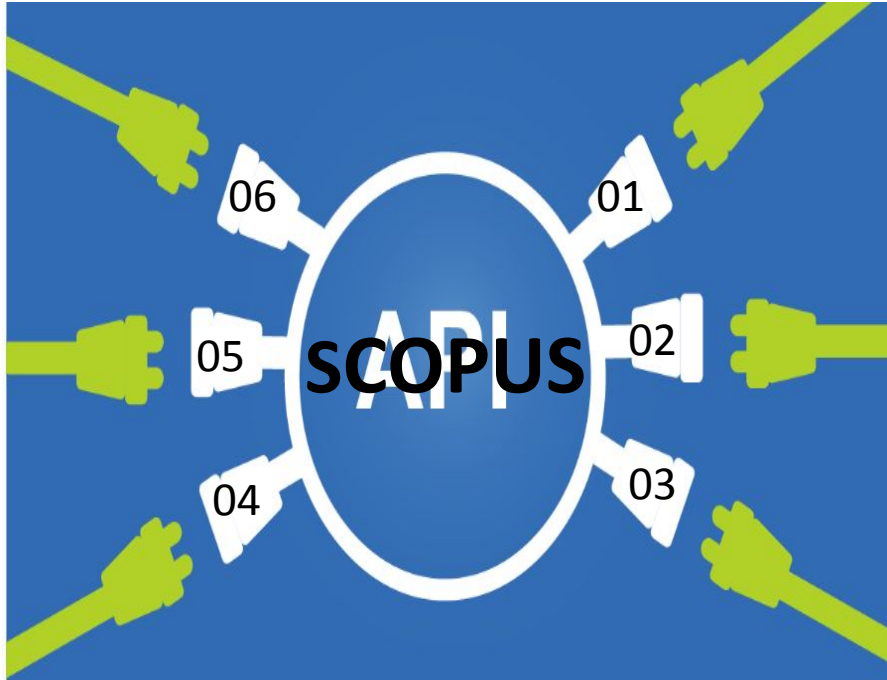
Problems of Traditional Manual Methods

- Each step is manual, which takes time
- For a smaller number of authors, we can do it manually
- What about the number of authors for a department is large?
 - UA, Obstetrics and Gynecology: 64
 - UA, Mechanical Engineering: 55
- What about frequent updates of bibliometrics? For example, the research group of UA “Understanding Labor Markets & Work which consists of professors from different departments to conduct collaborative research work. They require constant updates of their research metrics (weekly, monthly). “
- Worldwide University Network (WUN): 24 universities, 6 continents, frequent updating.

Can anything come
to the rescue
?

Application Programming Interface (API)

Software Product Interfaces



Physical Product Interfaces



Response Content Type

Parameters

Parameter	Value
query	<u>authlast(Marchand) and authfirst(Joseph)</u>
apiKey	974191 *****
httpAccept	
insttoken	7947c *****
access_token	

Try it out!

Hide Response

"document-count": "22",

"subject-area": [

{

"@abbrev": "SOCI",

"@frequency": "12",

"\$": "Social Sciences (all)"

},

{

"@abbrev": "ECON",

"@frequency": "12",

"\$": "Economics, Econometrics and Finance (all)"

},

{

"@abbrev": "BIOC",

"@frequency": "10",

"\$": "Biochemistry, Genetics and Molecular Biology (all)"

}

Python: Scripts

```
22 for au in authors:
23     if len(au)==2:
24         fname=au[0]
25         mname=""
26         lname=au[1]
27     else:
28         fname=au[0]
29         mname=au[1]
30         lname=au[2]
31
32     query=f''AUTHLAST({lname}) and AUTHFIRST({fname}
33           ) AND AFFIL(University of Alberta)''
34     s=AuthorSearch(query)
35     astr=str(s).strip()
36     m=re.search(r"AUTHOR_ID:(\d+) ",astr)
37     if m == None:
38         print(f"{fname},{lname}",file=no_pubs)
39         continue
40
41     ll=len(s.authors)
42     for i in range(0,ll):
43         author=s.authors[i]
44         result=dict()
45         au_id=author.eid.split("-")[-1]
```

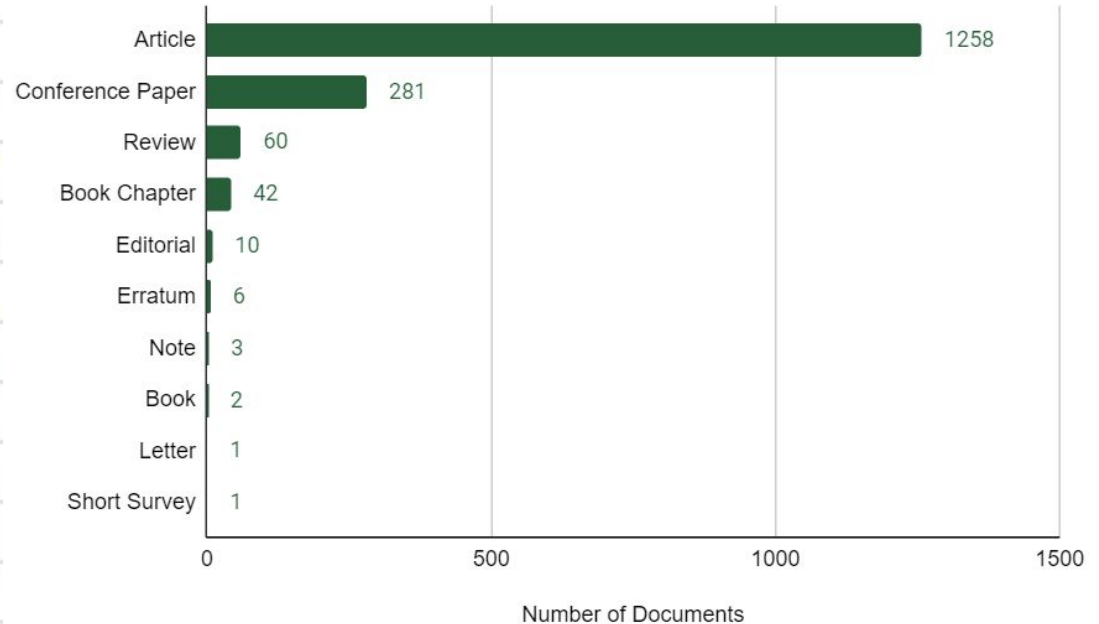
Algorithm:

- For each author, get his/her name and affiliation
- Run a query against SCOPUS
- Process the response
- Generate bibliometric performance data

AU_ID	Full Name	Docs	Citations	H-Index
7003518412	Harvey J. Krahn	79	1930	23
13004895100	Michelle Lee Maroto	34	526	14
57211631360	Karen D. Hughes	30	1271	16
35910081200	Gillian Stevens	30	2013	17
38961871600	Mesbah Fathy Sharaf	29	550	15
16245761500	Joseph Marchand	22	383	11
55440448500	Nicole Denier	19	260	10
35752940600	Haifang Huang	18	1045	12
15026164700	Sara K. Dorow	18	393	8
15768659000	Vera Brenčič	15	132	8
36460153600	Andrew D. McGee	13	139	7
36536915500	Dmytro Hryshko	10	143	6
25650492900	Beyza Ural Marchand	10	179	5
56600353000	Pınar Mine Güneş	10	175	4
57195975288	Jeffrey Penney	9	26	2
57206746899	Xingfei Liu	7	51	3
56323738500	Ashantha Ranasinghe	5	45	4
57742567000	Valentina Kozlova	1	0	0
57226198489	Laurel Wheeler	1	4	1

Applying the Python & API code to Department of Mechanical Engineering

Types of Publication	# of Docs
Article	1258
Conference Paper	281
Review	60
Book Chapter	42
Editorial	10
Erratum	6
Note	3
Book	2
Letter	1
Short Survey	1
Total	1664



Other Performance Metrics by API and Python

- Top 10 journals
- Top 10 most cited papers
- Percentage of Open Access Publications (individual, group, department)
- Co-authorship Analysis
- International Collaboration
- And more

Summary

- Demonstrated the popularity and significance of Bibliometric Analysis
 - Highlighted traditional manual approaches inefficiency in some use cases: such as large number of researchers; and periodical updates of the same group
 - API and Python can obtain the same information that is obtained by manual approach but in seconds after the program is developed
 - API and Python can be used to calculate customized performance metrics
-
- When there are ambiguity in name and institutions, manual process is needed to scrutinize and resolve the ambiguity before feeding to API & Python
 - When the SCOPUS database has erroneous records in it, i.e., the same name, the same affiliation, but more than one author profiles, manual approach is needed
 - The performance metrics must be predetermined and well defined before using API & Python because computer programs need to be implemented

Discussion

- We used SCOPUS as our database to conduct the research because it is more generous in giving back more complete information through its API
- Some other databases also have API, but they are tiered. The more you pay the subscription, the more information you get. That deters API programmers and organizations
- Python is chosen as the programming language because it has modules developed by third-party developers, we don't have to reinvent the wheel.
- If the librarians don't want to do the programming, the freeware Data Fetcher is a ready-to-use API product.
- Lori Ostapowicz-Critz, Worcester Polytechnic Institute and Julie Morris, University of New Brunswick highlighted their applications of APIs through Data Fetcher to analyze media citations with PlumX and collection impact respectively.

Discussion

- This presentation just scratches the surface, and try to gauge the community interest
- Hopefully, we can do a workshop in the future if there are enough people who are interested.



Questions?



bit.ly/APASlides

References

1. Bredhal, L. (2022). The current and evolving landscape of bibliometric tools and technologies, Library Technology Reports, ALA, 58(8).
[DOI:https://doi.org/10.5860/ltr.58n8](https://doi.org/10.5860/ltr.58n8)
2. Durieux V, Gevenois PA. Bibliometric indicators: quality measurements of scientific publication. Radiology. 2010 May;255(2):342-51. doi: 10.1148/radiol.09090626. PMID: 20413749.
3. SCOPUS, <https://www.scopus.com/>
4. Beyond Bibliometrics – Using Scopus to Assess Collections and Support Faculty
<https://www.choice360.org/webinars/beyond-bibliometrics-using-scopus-to-assess-collections-and-support-faculty/>
5. <https://www.planmill.com/wp-content/uploads/2017/09/API-4-1.png>
6. Data Fetcher
https://dev.elsevier.com/data-fetcher-resources/DataFetcherManual_7_4_3.pdf
7. SCOPUS Interactive API, <https://dev.elsevier.com/interactive.html>