Project Academic Knowledge

Using the Microsoft Academic API to evaluate institutional repository impact

Joseph Koivisto

https://go.umd.edu/LRIPF -MA

What is Microsoft Academic?

What is Microsoft Academic?

- Commercial discovery interface

- Publicly accessible

- Semantic search vs. keyword search

  - "employs advances in machine learning, semantic inference and knowledge discovery"

https://academic.microsoft.com/faq

What isn't Microsoft Academic?

Why it matters that it isn't Google Scholar...

- An alternative to Google hegemony

- Discovery not limited to publisher outlets and interfaces (includes open access resources*)

- Pathways for systematic access and reuse of data

What is Project Academic Knowledge?

What is Project Academic Knowledge?

"With [the **Project Academic Knowledge**] service, you will be able to interpret user queries for academic intent and retrieve rich information from the **Microsoft Academic Graph (MAG)** . The MAG knowledge base is a web-scale heterogeneous entity graph comprised of entities that model scholarly activities: field of study, author, institution, paper, venue, and event."

What is Project Academic Knowledge?

Four related REST API methods:

- **Interpret:** Interprets a natural language user query string.

- **Evaluate:** Evaluates a query expression and returns Academic Knowledge entity results

- **Calchistogram:** Calculates a histogram of the distribution of attribute values for the academic entities returned by a query expression

- **Similarity:** Calculates the cosine similarity between two strings

What is Project Academic Knowledge?

Four related REST API methods:

- **Interpret:**   Interprets a natural language user query string.

- **Evaluate:** Evaluates a query expression and returns Academic Knowledge entity results

- **Calchistogram:**   Calculates a histogram of the distribution of attribute values for the academic entities returned by a query expression

- **Similarity:**   Calculates the cosine similarity between two strings

Limitations Project Academic Knowledge?

- Not everything is indexed

- Still a commercial scholarly portal

- Requires some technical knowledge (not a lot, but more than zero)

- Other limitations (Hug, Ochsner & Brändle, 2017)
  - Does not provide the publication document type
  - "Fields of study" are dynamic, too specific & field hierarchies incoherent
  - Some publications assigned to incorrect years
  - Metadata of some publications did not include all authors

Project Academic Knowledge - Evaluate

Request Parameters

Name	Value	Required?	Description
expr	Text string	Yes	A query expression that specifies which entities should be returned.
model	Text string	No	Name of the model that you wish to query. Currently, the value defaults to *latest*.
attributes	Text string	No default: Id	A comma-delimited list that specifies the attribute values that are included in the response. Attribute names are case-sensitive.
count	Number	No Default: 10	Number of results to return.
offset	Number	No Default: 0	Index of the first result to return.
orderby	Text string	No Default: by decreasing prob	Name of an attribute that is used for sorting the entities. Optionally, ascending/descending can be specified. The format is: *name:asc* or *name:desc*.

https://docs.microsoft.com/en-us/academic-services/project-academic-knowledge/reference-evaluate-method

Project Academic Knowledge - Evaluate

Paper Entity

04/10/2020 • 4 minutes to read •

> ⓘ **Note**
>
> Below attributes are specific to paper entity. (Ty = '0')

Name	Description	Type	Operations
AA.AfId	Author affiliation ID	Int64	Equals
AA.AfN	Author affiliation name	String	Equals, StartsWith
AA.AuId	Author ID	Int64	Equals
AA.AuN	Normalized author name	String	Equals, StartsWith
AA.DAuN	Original author name	String	None

Use Case Scenario

How to get citation counts for 13,000+ records?

Project Academic Knowledge - Evaluate
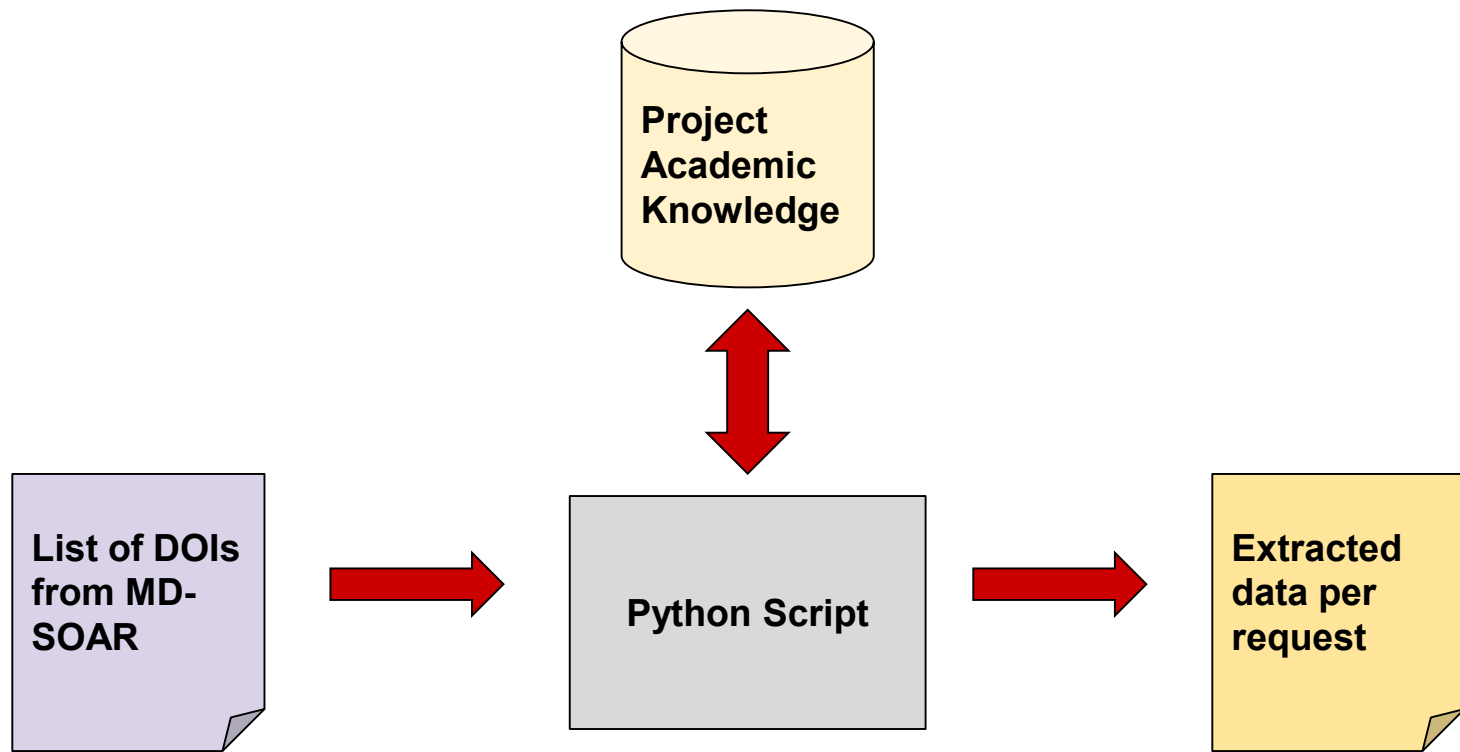
```python
def genTen(start):
    global query
    global place
    doiQuery = 'OR('
    for i in doiList[start:start+30]:
        a = 'DOI=\'' + i.upper() + '\''
        doiQuery = doiQuery + a + ', '
    doiQuery = {'expr':doiQuery[:-2] + ')'}
    query = doiQuery


base_url = 'https://api.labs.cognitive.microsoft.com/academic/v1.0/evaluate'
params = {'expr':'',
          'model':'latest',
          'offset':0,
          'attributes':'Id,DN,DOI,CC,ECC,S'}
headers = {'accept':'application/json',
           "Ocp-Apim-Subscription-Key":    SECRET, SHHH!!!!!!!!    }

with open('edit_doi.csv','r') as f:
    for i in f:
        i = i.strip()
        doiList.append(i)

filename = 'ma_raw_' + str(datetime.date.today()) + '.json'
h = open(filename,'a+')
h.write('[')

#while place < 100:
while place < len(doiList):
    time.sleep(2)
    genTen(place)
    params.update(query)
#    print(params)
    r = requests.get(base_url, params=params, headers=headers)
    j = r.json()
    h.write(json.dumps(j, indent = 3))
    h.write(',')
    place += 30
    print(place)

h.write(']')
h.close()
```

See the code on GitHub at
https://gist.github.com/23koivisto/994063d4c1d8f9d30a49797e58553935

```
"entities": [
    {
        "logprob": -19.015,
        "prob": 5.5193817e-09,
        "Id": 2143899515,
        "CC": 120,
        "ECC": 416,
        "DN": "KQML - A Language and Protocol for Knowledge and Information Exchange",
        "DOI": "10.13016/M24KFA-JUWY",
        "S": [
            {
                "Ty": 3,
                "U": "https://ebiquity.umbc.edu/get/a/publication/741.pdf"
            },
            {
                "Ty": 3,
                "U": "https://www.aaai.org/Papers/Workshops/1994/WS-94-02/WS94-02-007.pdf"
            },
            {
                "Ty": 1,
                "U": "https://ebiquity.umbc.edu/paper/html/id/688/KQML-A-Language-and-Protocol-for-Knowledge-and-Information-Exchange"
            }
        ]
    },
    {
        "logprob": -21.353,
        "prob": 5.327334e-10,
        "Id": 2922266863,
        "CC": 4,
        "ECC": 4,
        "DN": "Selection of Optimal Hyperspectral Wavebands for Detection of Discolored, Diseased Rice Seeds",
        "DOI": "10.13016/M2G5LJ-BVNM",
        "S": [
            {
                "Ty": 1,
                "U": "https://mdsoar.org/handle/11603/13268"
            }
        ]
    },
```

Project Academic Knowledge - Evaluate

	A	B	C	D	E	F	G	H	I	J
1	normalizedTitle	DOI	citationCount	estimatedCitationCount	indexSources					
2	KQML - A Language and Protocol for Knowledge and Information Exchange	10.13016/M24KFA-JUWY	120	416	[{'Ty': 3, 'U': 'https://ebiquity.umbc.edu/get/a/publication/741.pd				
3	Selection of Optimal Hyperspectral Wavebands for Detection of Discolored, Diseased Rice Seeds	10.13016/M2G5LJ-BVNM	4	4	[{'Ty': 1, 'U': 'https://mdsoar.org/handle/11603/13268'}]				
4	Use of Assistive Technology for Cognition Mobile Applications by Breast Cancer Patients Treated with Chemotherapy	10.13016/M2UQUW-AB6L	0	0	[{'Ty': 1, 'U': 'https://mdsoar.org/handle/11603/12746'}]				
5	Ensuring data integrity and user retention within BANDIT	10.13016/M2W6XB-S3L1	0	0	[{'Ty': 1, 'U': 'https://mdsoar.org/handle/11603/13293'}]				
6	Darwin Correspondence Project	10.13016/M2Q921	9	9	[{'Ty': 3, 'U': 'https://works.bepress.com/betty_landesman/5/dow				
7	Disability History Museum	10.13016/M2V059	2	2	[{'Ty': 3, 'U': 'https://works.bepress.com/betty_landesman/6/dow				
8	Altering growth rates and nutritional qualities of microalgal feedstock with symbiotic bacteria	10.13016/M2TW6B	0	0	[{'Ty': 1, 'U': 'https://mdsoar.org/handle/11603/131'}]				
9	An exploration of two online approaches to mathematics teacher education	10.13016/M29W6P	4	4	[{'Ty': 1, 'U': 'http://www.editlib.org/p/27033/'}, {'Ty': 1, 'U': 'https				
10	"Film at 11": True Stories of News Film Collections at the University of Baltimore	10.13016/M26S6Q	0	0	[{'Ty': 1, 'U': 'https://mdsoar.org/handle/11603/188'}]				

2,235 records indexed

5,189 citations recorded

How to get started

How to get started

Visit https://www.microsoft.com/en-us/research/project/academic-knowledge/ and subscribe

How to get started

Microsoft Research APIs

HOME   SUBSCRIBE   KEYS   SIGN IN

Available Projects

Project Academic Knowledge

Tap into the wealth of academic content in the Microsoft Academic Graph. 10,000 transactions per month, 3 per second for interpret, 1 per second for evaluate, 6 per minute for calcHistogram.

How to get started

Create a Microsoft API Management account if you don't already have one

How to get started

You'll get a couple emails, be sure to check your spam folder if you don't see them

How to get started

You'll receive a subscription key

How to get started

Navigate to https://msr-apis.portal.azure-api.net/docs/services/academic-search-api

How to get started

Experiment!

(Suggest starting with **GET Evaluate**)

How to get started

Important note

- Expression (expr) is where you enter your search query
- Attributes is where you will specify what fields you would like returned

How to get started

- Example

expr: DOI='10.1145/2983270'

- Searches for resources identified by this DOI

Attributes: Id, DN, CC, ECC,AA.AuN

- Will return the resource ID, normalized name, citation count, estimated citation count, and normalized author's name

How to get started

Documentation on the evaluate method is available at https://docs.microsoft.com/en-us/academic-services/project-academic-knowledge/reference-evaluate-method

Keep in touch

Let me know what you're working on

Let me know if you need help on a problem

Come to UMD Libraries Coding Workshop

Citations

- How is MA different from other academic search engines? (ND). https://academic.microsoft.com/faq
- Project Academic Knowledge (2020). https://docs.microsoft.com/en-us/academic-services/project-academic-knowledge/introduction
- Hug, S.E., Ochsner, M. & Brändle, M.P. (2017). Citation analysis with microsoft academic. Scientometrics 111. P 371–378. https://doi-org.proxy-um.researchport.umd.edu/10.1007/s11192-017-2247-8
- Evaluate method (2020). https://docs.microsoft.com/en-us/academic-services/project-academic-knowledge/reference-evaluate-method
- Paper entity (2020). https://docs.microsoft.com/en-us/academic-services/project-academic-knowledge/reference-paper-entity-attributes