## Accessing open research literature with Python

@NikoletaGlyn







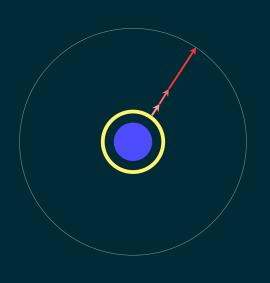
Software Sustainability Institute



# The illustrated guide to a Ph.D.

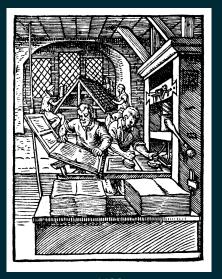
Matt Might

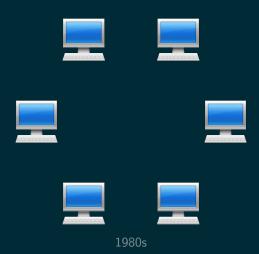
http://matt.might.net/articles/phd-school-in-pictures/



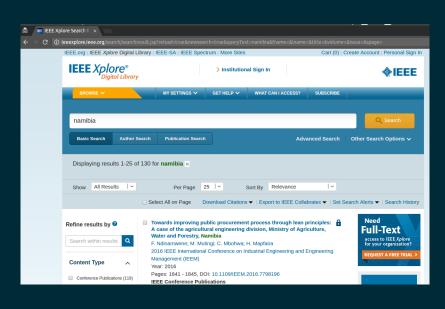
## ARTICLE ← JOURNAL

REVIEW









#### org/search/searchresult.jsp?newsearch=true&queryText=namibia

 Case Study on South Africa and Namibia: A Model for Electronic Evidence for the SADC Region



Amelia Phillips

2011 Sixth IEEE International Workshop on Systematic Approaches to Digital

Forensic Engineering

Year: 2011

Pages: 1 - 6, DOI: 10.1109/SADFE.2011.4

**IEEE Conference Publications** 

▶ Abstract

((html))



**(C)** 

Mapping pegmatite using HyMap data in southern Namibia



Atsushi Momose; Shuichi Miyatake; Yessy Arvelyna; Anna Nguno; Kombada Mhopjeni; Minsozi Sibeso; Aphary Muyongo; Ewereth Muvangua

2011 IEEE International Geoscience and Remote Sensing Symposium

Year: 2011

Pages: 2216 - 2217, DOI: 10.1109/IGARSS.2011.6049608

**IEEE Conference Publications** 

Abstract

((html))





# $0.5\min + 100 \times 1.5\min + 10 \times 0.5\min = 155.5\min \Rightarrow 2h \text{ and } 35.5\min$



# QUERY

http://ieeexplore.ieee.org/gateway/ipsSearch.jsp?ti=

#### **QUERY**

http://ieeexplore.ieee.org/gateway/ipsSearch.jsp?ti= Namibia&hc=100

http://api.plos.org/search?q=title:Namibia&rows=100

#### **QUERY**

http://ieeexplore.ieee.org/gateway/ipsSearch.jsp?ti= Namibia&hc=100

http://api.plos.org/search?q=title:Namibia&rows=100

http://www.nature.com/opensearch/request?queryType=cql&query=dc.title%20adj%20Namibia&maximumRecords=100

. . .

```
ieeexplore.ieee.org/ ×
← → C ① ieeexplore.ieee.org/gateway/ipsSearch.jsp?ti=Namibia&hc=100
                                                                                                                              ① ☆

    ▼<document>

    <rank>6</rank>
   ▼<title>
     <! [CDATAI
        Mapping pegmatite using HyMap data in southern Namibia
      11>
    </title>
   ▼<authors>
     <! [CDATA[
        Atsushi Momose; Shuichi Miyatake; Yessy Arvelyna; Anna Nguno; Kombada Mhopjeni; Minsozi Sibeso; Aphary Muyongo; Ewereth
        Muvangua
      11>
    </authors>
   ▼<affiliations>
     <! [CDATAI
        Japan Oil, Gas and Metals National Corporation, Japan
      11>
    </affiliations>
   ▼<controlledterms>
     ▼<term>
        <![CDATA[ data analysis ]]>
      </term>
     ▼<term>
        <![CDATA[ geophysical image processing 1]>
      </term>
     ▼<term>
        <![CDATA[ geophysical techniques ]]>
      </term>
     ▼<term>
        <![CDATA[ minerals ]]>
      </term>
     ▼<term>
        <![CDATA[ rocks 11>
      </term>
    </controlledterms>
```

# $15\min + 1\min + 50\min = 66\min \Rightarrow 1h \text{ and } 6\min$



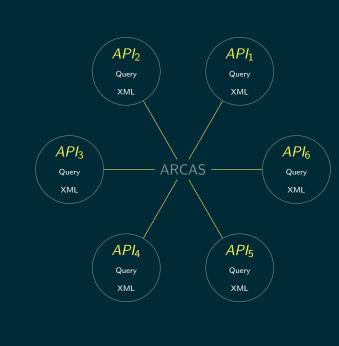






API<sub>4</sub>
Query
XML





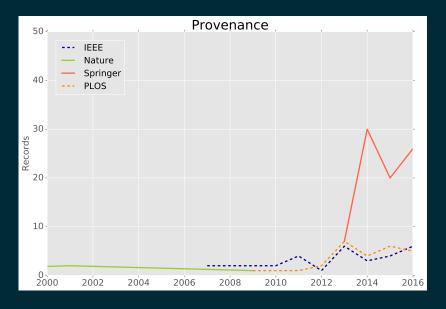
# pip install arcas

```
import arcas
arguments = {'-a': None, '-t': 'Namibia', '-s': None,
api = arcas.leee()
parameters = api.parameters_fix(arguments)
url = api.create_url_search(parameters)
request = api.make_request(url)
response = api.get_root(request)
root = api.get_root(response)
raw_article = api.parse(root)
article = api.to_dataframe(raw_article)
```

```
arguments = {'-a': None, '-t': 'Namibia', '-s': None,
for p in [arcas.Ieee, arcas.Plos, arcas.Arxiv,
         arcas.Nature, arcas.Springer]:
    api = p()
    parameters = api.parameters_fix(arguments)
    url = api.create_url_search(parameters)
    request = api.make_request(url)
    response = api.get_root(request)
    root = api.get_root(response)
    raw_article = api.parse(root)
    for art in raw article:
        article = api.to_dataframe(raw_article)
       api.export(articles, 'results.json')
```

```
{"key":{"0":"Momose2011",
        "1":"Momose2011",
        "2": "Momose2011"}.
"unique key": {"0": "4061b0ca3b823f85a0cb2823a554c524".
              "1": "4061b0ca3b823f85a0cb2823a554c524",
              "2": "4061b0ca3b823f85a0cb2823a554c524"}.
"title": {"0": "Mapping pegmatite using HyMap data in southern Namibia",
         "1": "Mapping pegmatite using HyMap data in southern Namibia",
         "2": "Mapping pegmatite using HyMap data in southern Namibia"},
"author": {"0": "Atsushi Momose".
          "1": "Atsushi Momose",
          "2":"Atsushi Momose"},
"abstract":{"0":"A pegmatite deposit is an ..."},
"date":{"0":2011.
"journal": ["0": "2011 IEEE International Geoscience and Remote Sensing Symposium".
           "1": "2011 IEEE International Geoscience and Remote Sensing Symposium",
           "2": "2011 IEEE International Geoscience and Remote Sensing Symposium"},
"pages":{"0":"2216-2217".
         "2": "2216-2217"},
"key_word":{"0":"data analysis",
           "1": "geophysical image processing",
           "2": "geophysical techniques"},
"provenance": {"0": "IEEE",
              "2":"TEEE"}}
```

# $15\min + 5\min = 20\min$

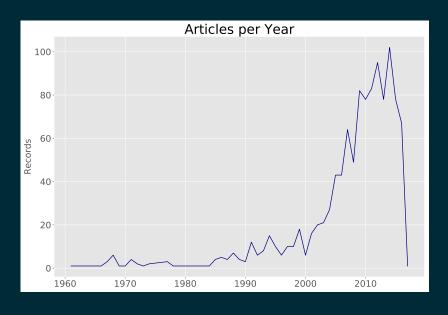


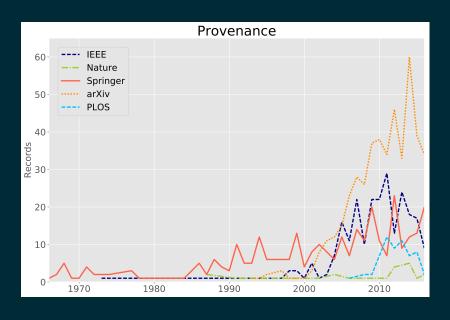
doc/ arcas.readthedocs.io/

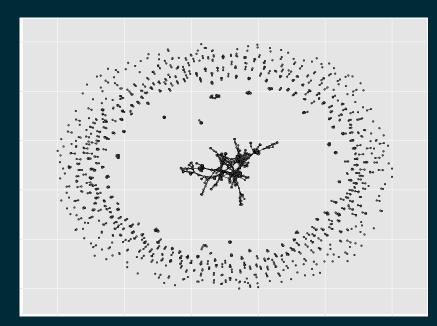
IEEE Nature
PLOS ...

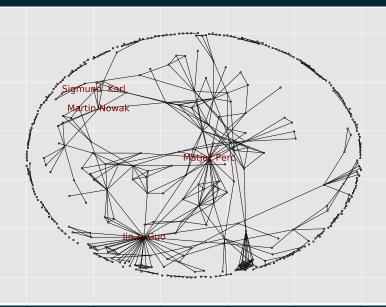
testIEEE testNature

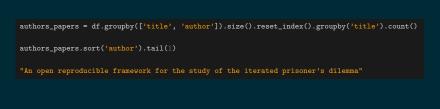
testPLOS ...











```
arcas scrape -h
Arcas. A library to facilitate scraping of APIs for scholarly resources.
Usage:
   arcas_scrape [-h] [-p API] [-a AUTHOR] [-t TITLE] [-b ABSTRACT]
    [-y YEAR] [-r RECORDS] [-s START] [-v VALIDATE] [-f FILENAME]
   arcas_scrape --version
Options:
   -h --help
                          Show this
   --version
                          Show version
   -p API
                          The online API, from a given list, to parse [default: arxiv]
   -a AUTHOR
                          Terms to search for in Author
   -t TITLE
                          Terms to search for in Title
   -b ABSTRACT
                          Terms to search for in the Abstract
   -v YEAR
                          Terms to search for in Year
   -r RECORDS
                          Number of records to fetch
   -s START
                          Sequence number of first record to fetch
   -v VALIDATE
                          Checks if query returned with arguments asked [default: False]
   -f FILENAME
                          Name of json file [default: results.json]
```

## I academic API so you don't have to!

@NikoletaGlyn https://github.com/Nikoleta-v3/Arcas @SoftwateSaved

@PhoenixCUni