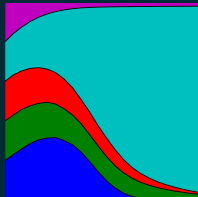


Arcas: Using Python to access open research literature

@NikoletaGlyn



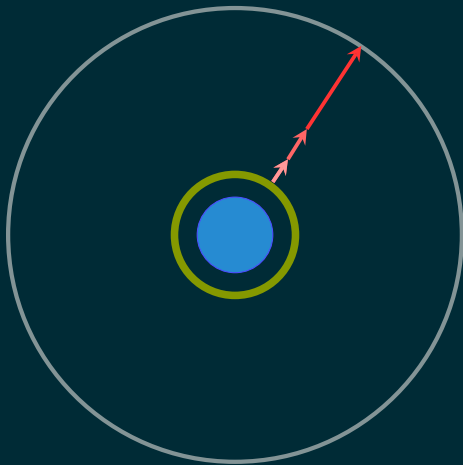


Software
Sustainability
Institute

The illustrated guide to a Ph.D.

Matt Might

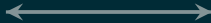
<http://matt.might.net/articles/phd-school-in-pictures/>



ARTICLE



JOURNAL




REVIEW



PUBLISHED

Sustainable Software



Cornell University
Library

We gratefully acknowledge support from
the Simons Foundation
and member institutions

arXiv.org

Login

sustainable softwareAll papers

(Help | Advanced search)

Open access to 1,296,634 e-prints in Physics, Mathematics, Computer Science, Quantitative Biology, Quantitative Finance and Statistics

Subject search and browse:

08 Aug 2017: [A survey for users accessing arXiv programmatically](#)
20 Apr 2017: [Applied Physics subject area added to arXiv](#)
10 Mar 2017: [New members join arXiv Member Advisory Board](#)
06 Mar 2017: [arXiv Scientific Director Search](#)
10 Feb 2017: [Attention Submitters: our TeX processing system has been updated](#)
See cumulative ["What's New"](#) pages. Read [robots beware](#) before attempting any automated download

Physics

102. [arXiv:1309.1828](#) [pdf]

Sustainable Software Development for Next-Gen Sequencing (NGS) Bioinformatics on Emerging Platforms

[Shel Swenson](#), [Yogesh Simmhan](#), [Viktor Prasanna](#), [Manish Parashar](#), [Jason Riedy](#), [David Bader](#), [Richard Vuduc](#)

Comments: 4 pages

Subjects: **Computational Engineering, Finance, and Science (cs.CE)**; Distributed, Parallel, and Cluster Computing (cs.DC)

103. [arXiv:1309.1817](#) [pdf]

Initial Findings from a Study of Best Practices and Models for Cyberinfrastructure Software Sustainability

[Craig A. Stewart](#), [Julie Wernert](#), [Eric A. Wernert](#), [William K. Barnett](#), [Von Welch](#)

Comments: Workshop on Sustainable Software: Practices and Experiences, 4 pages

Subjects: **Software Engineering (cs.SE)**

104. [arXiv:1309.1813](#) [pdf, other]

Reusability in Science: From Initial User Engagement to Dissemination of Results

[Ketan Maheshwari](#), [David Kelly](#), [Scott J. Krieder](#), [Justin M. Wozniak](#), [Daniel S. Katz](#), [Mei Zhi-Gang](#), [Mainak Mookherjee](#)

Comments: 5 pages, WSSSPE 2013 workshop

Subjects: **Software Engineering (cs.SE)**; Distributed, Parallel, and Cluster Computing (cs.DC)

105. [arXiv:1309.1812](#) [pdf, other]

Cactus: Issues for Sustainable Simulation Software

[Frank Löffler](#), [Steven R. Brandt](#), [Gabrielle Allen](#), [Erik Schnetter](#)

Comments: submitted to the Workshop on Sustainable Software for Science: Practice and Experiences 2013

Subjects: **Computational Engineering, Finance, and Science (cs.CE)**; Mathematical Software (cs.MS); Software Engineering (cs.SE)

106. [arXiv:1309.1810](#) [pdf]

Niche Modeling: Ecological Metaphors for Sustainable Software in Science

[Nicholas Weber](#), [Andrea Thomer](#), [Michael Twidale](#)

Comments: Position paper submitted to: Workshop on Sustainable Software for Science: Practice and Experiences (WSSSPE) SC13, Sunday, 17 November 2013, Denver, CO, USA

Subjects: **Software Engineering (cs.SE)**; Computers and Society (cs.CY)

107. [arXiv:1309.1805](#) [pdf]

nanoHUB.org: Experiences and Challenges in Software Sustainability for a Large Scientific Community

[Lynn Zentner](#), [Michael Zentner](#), [Victoria Farnsworth](#), [Michael McLennan](#), [Krishna Madhavan](#), [Gerhard Klimeck](#)

Comments: 4 pages, 1 figure, this version contains minor revisions to correct an acronym, update a quotation, improve grammar, and add a reference

Subjects: **Software Engineering (cs.SE)**; Computational Engineering, Finance, and Science (cs.CE); Digital Libraries (cs.DL)

108. [arXiv:1309.1796](#) [pdf, ps, other]

VisIt: Experiences with Sustainable Software

[Sean Ahern](#), [Eric Brugger](#), [Brad Whittlock](#), [Jeremy S. Meredith](#), [Kathleen Biagas](#), [Mark C. Miller](#), [Hank Childs](#)

Subjects: **Software Engineering (cs.SE)**

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

API

QUERY

```
http://export.arxiv.org/api/query?search_query=ti:  
Sustainable%20Software
```

exportarxiv.org/api/query/search_query=tf%3ASustainable%20Software

```
<?xml version="1.0" encoding="UTF-8"?>
<feed xmlns="http://www.w3.org/2005/Atom">
  <link href="http://arxiv.org/api/query?search_query=tf%3ASustainable%20Software%26id_list%3D%26start%3D0%26max_results%3D10" rel="self" type="a" />
  <title type="html">ArXiv Query: search_query=tf:Sustainable Software&id_list=&start=0&max_results=10</title>
  <id>http://arxiv.org/api/N537kBVULVtW1UduEH05fqj8g5Q</id>
  <updated>2017-08-25T00:00:00-04:00</updated>
  <opensearch:totalResults xmlns:opensearch="http://a9.com/-/spec/opensearch/1.1/">14298</opensearch:totalResults>
  <opensearch:startIndex xmlns:opensearch="http://a9.com/-/spec/opensearch/1.1/">0</opensearch:startIndex>
  <opensearch:itemsPerPage xmlns:opensearch="http://a9.com/-/spec/opensearch/1.1/">10</opensearch:itemsPerPage>
  <entry>
    <id>http://arxiv.org/abs/1309.1796v1</id>
    <updated>2013-09-07T00:16:52Z</updated>
    <published>2013-09-07T00:16:52Z</published>
    <title>VisIt: Experiences with Sustainable Software</title>
    <summary> The success of the VisIt visualization system has been wholly dependent upon the culture and practices of software development that have fostered its welcome by users and embrace by developers and researchers. In the following paper, we, the founding developers and designers of VisIt, summarize some of the major efforts, both successful and unsuccessful, that we have undertaken in the last thirteen years to foster community, encourage research, create a sustainable open-source development model, measure impact, and support production software. We also provide commentary about the career paths that our development work has engendered.
  </summary>
    <author>
      <name>Sean Ahern</name>
    </author>
    <author>
      <name>Eric Brugger</name>
    </author>
    <author>
      <name>Brad Whitlock</name>
    </author>
    <author>
      <name>Jeremy S. Meredith</name>
    </author>
  </entry>
</feed>
```

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

QUERY

```
http://export.arxiv.org/api/query?search_query=ti:  
Sustainable%20Software
```

QUERY

`http://export.arxiv.org/api/query?search_query=ti:
Sustainable%20Software`

`http://api.plos.org/search?q=title:
Sustainable%20Software&rows=100`

QUERY

`http://export.arxiv.org/api/query?search_query=ti:
Sustainable%20Software`

`http://api.plos.org/search?q=title:
Sustainable%20Software&rows=100`

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

...

API₂

Query

XML

API₁

Query

XML

API₃

Query

XML

API₆

Query

XML

API₄

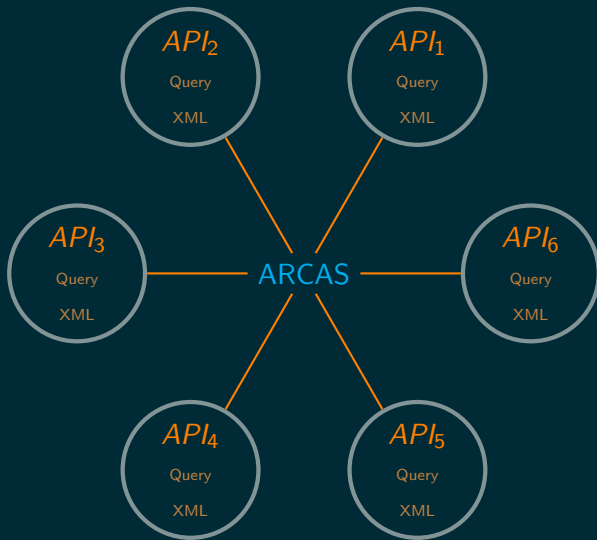
Query

XML

API₅

Query

XML



```
$ pip install arcas
```

```
>>> import arcas

>>> api = arcas.Arxiv()
>>> parameters = api.parameters_fix(
...     title='sustainable software', records=1, start=1)
>>> url = api.create_url_search(parameters)
>>> request = api.make_request(url)
>>> root = api.get_root(request)
>>> raw_article = api.parse(root)

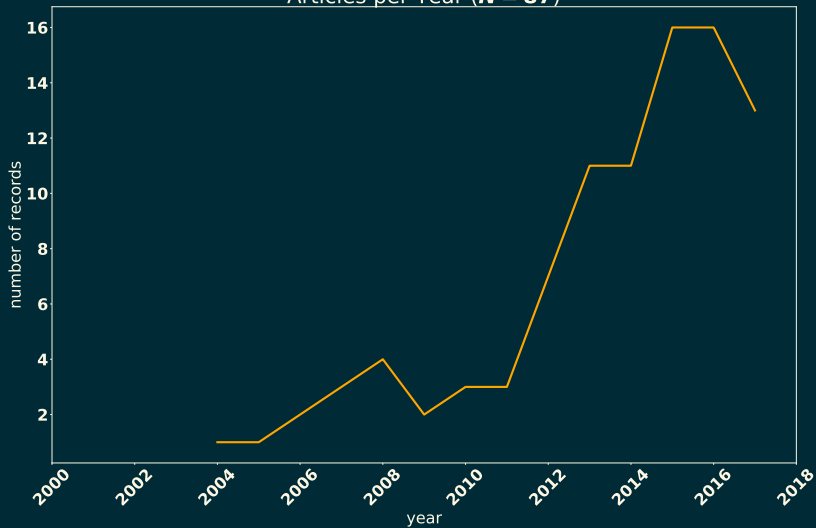
>>> article = api.to_dataframe(raw_article[0])
>>> api.export(article, "result.json")
```

```
{"key":{"0":"Ahern2013"},  
  "unique_key":{"0":"698d27415f69258ef122f46b184a77e0"},  
  "title":{"0":"VisIt: Experiences with Sustainable Software"},  
  "author":{"0":"Sean Ahern","1":"Eric Brugger"},  
  "abstract":{"0":"  The success of the VisIt visualization..."},  
  "date":{"0":2013},  
  "journal":{"0":"arXiv"},  
  "provenance":{"0":"arXiv"}}
```

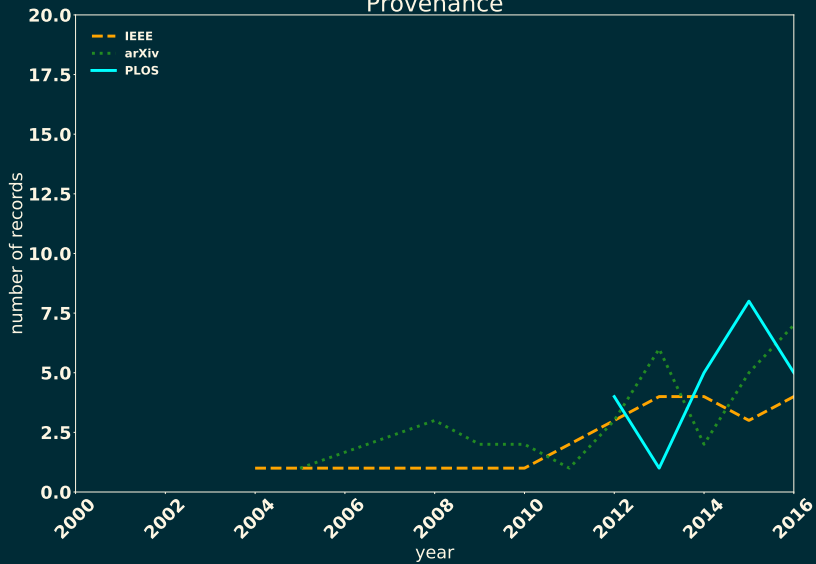
$$15\text{min} + 5\text{min} = 20\text{min}$$

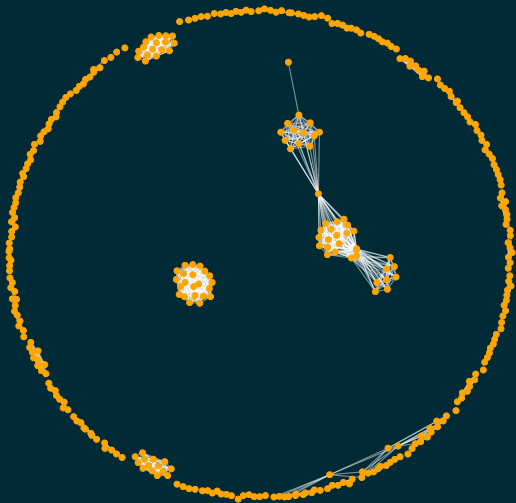
```
>>> for p in [arcas.Arxiv, arcas.Nature, arcas.Ieee, arcas.Plos]:
...     api = p()
...     parameters = api.parameters_fix(
...         title='sustainable software', records=1, start=1)
...     url = api.create_url_search(parameters)
...     request = api.make_request(url)
...     root = api.get_root(request)
...     raw_article = api.parse(root)
...     try:
...         for art in raw_article:
...             article = api.to_dataframe(art)
...             api.export(article, "result_from_{}.json".format(
...                 api.__class__.__name__))
...     except TypeError:
...         pass
```

Articles per Year ($N = 87$)



Provenance





Arcas

tools.py

doc/

arcas.readthedocs.io/

ieee

nature

arxiv

...

test_ieee

test_nature

test_arxiv

...

```
$ arcas_scrape --version  
Arcas 0.0.1
```

```
$ arcas_scrape -p arxiv -t "Sustainable Software" -r 1  
http://export.arxiv.org/api/query?search_query=ti:Sustainable  
Software&max_results=1&start=1
```

@NikoletaGlyn

<https://github.com/Nikoleta-v3/Arcas>

<https://nikoleta-v3.github.io/talks/>