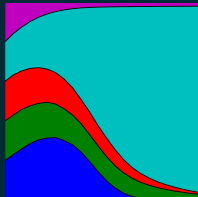


Arcas: Using Python to access open research literature

@NikoletaGlyn



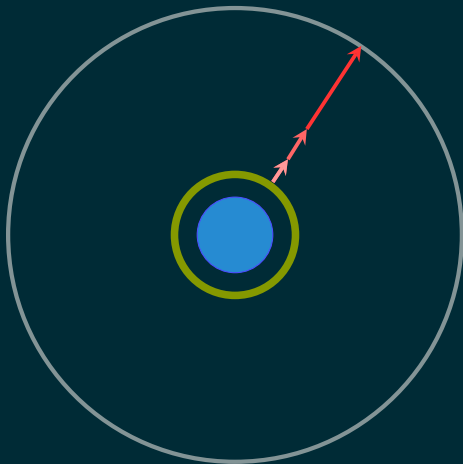


Software
Sustainability
Institute

The illustrated guide to a Ph.D.

Matt Might

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



ARTICLE



JOURNAL





REVIEW



PUBLISHED

Sustainable Software





Open access to 1,298,024 e-prints in Physics, Mathematics, Computer Science, Quantitative Biology, Quantitative Finance and Statistics

Subject search and browse: [Physics](#) [Mathematics](#) [Computer Science](#) [Quantitative Biology](#) [Quantitative Finance and Statistics](#)

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
 12 Feb 2017: Abstract Submitters, our DOI processing system has been updated
 See cumulative "What's New" pages. Read robots beware before attempting any automated download

Physics

- [Astrophysics](#) ([astro-ph](#) new, recent, find)
 includes: Astrophysics of Galaxies; Cosmology and Non-Galactic Astrophysics; Earth and Planetary Astrophysics; High Energy Astrophysical Phenomena; Instrumentation and Methods for Astrophysics; Solar and Stellar Astrophysics
- [Condensed Matter](#) ([cond-mat](#) new, recent, find)
 includes: Disordered Systems and Neural Networks; Materials Science: Mesoscale and Nanoscale Physics; Other Condensed Matter; Quantum Gases; Soft Condensed Matter; Statistical Mechanics; Strongly Correlated Electrons; Superconductivity
- [General Relativity and Quantum Cosmology](#) ([gr-qc](#) new, recent, find)
- [High Energy Physics - Experiment](#) ([hep-ex](#) new, recent, find)
- [High Energy Physics - Lattice](#) ([hep-lat](#) new, recent, find)
- [High Energy Physics - Phenomenology](#) ([hep-ph](#) new, recent, find)
- [High Energy Physics - Theory](#) ([hep-th](#) new, recent, find)
- [Mathematical Physics](#) ([math-ph](#) new, recent, find)
- [Nonlinear Sciences](#) ([nlin](#) new, recent, find)
 includes: Advection and Self-Organizing Systems; Cellular Automata and Lattice Gases; Chaotic Dynamics; Exactly Solvable and Integrable Systems; Pattern Formation and Solitons
- [Nuclear Experiment](#) ([nucl-ex](#) new, recent, find)
- [Nuclear Theory](#) ([nucl-th](#) new, recent, find)
- [Physics](#) ([physics](#) new, recent, find)
 includes: Accelerator Physics; Applied Physics; Atmospheric and Oceanic Physics; Atomic Physics; Atomic and Molecular Clusters; Biological Physics; Chemical Physics; Classical Physics; Computational Physics; Data Analysis, Statistics and Probability; Fluid Dynamics; General Physics; Geophysics; History and Philosophy of Physics; Instrumentation and Detectors; Medical Physics; Optics; Physics Education; Physics and Society; Plasma Physics; Popular Physics; Special Physics
- [Quantum Physics](#) ([quant-ph](#) new, recent, find)

Mathematics

- [Mathematics](#) ([math](#) new, recent, find)
 includes (see detailed description): Algebraic Geometry; Algebraic Topology; Analysis of PDEs; Category Theory; Classical Analysis and ODEs; Combinatorics; Commutative Algebra; Complex Variables; Differential Geometry; Dynamical Systems; Functional Analysis; General Mathematics; General Topology; Geometric Topology; Group Theory; History and Overview; Information Theory; K-Theory and Homology; Logic; Mathematical Physics; Metric Geometry; Number Theory; Numerical Analysis; Operator Algebras; Optimization and Control; Probability; Quantum Algebra; Representation Theory; Rings and Algebras; Spectral Theory; Statistics Theory; Symplectic Geometry

Computer Science

- [Computing Research Repository](#) ([CoRR](#) new, recent, find)
 includes (see detailed description): Artificial Intelligence; Computation and Language; Computational Complexity; Computational Engineering, Finance, and Science; Computational Geometry; Computer Science and Game Theory; Computer Vision and Pattern Recognition; Computers and Society; Cryptography and Security; Data Structures and Algorithms; Database; Digital Libraries; Discrete Mathematics; Distributed, Parallel, and Cluster Computing; Emerging Technologies; Formal Languages and Automata Theory; General Literature; Graphics; Hardware Architecture; Human-Computer Interaction; Information Retrieval; Information Theory; Learning Logic in Computer Science; Mathematical Software; Multiagent Systems; Multimedia; Networking and Internet Architecture; Neural and Evolutionary Computing; Numerical Analysis; Operating Systems; Other Computer Science; Performance; Programming Languages; Robotics; Social and Information Networks; Software Engineering; Sound; Symbolic Computation; Systems and Control

Quantitative Biology

Secure | <https://arxiv.org/abs/1507.05414v4> - arXiv:1507.05414v4 [cs.LG] - 100 papers, 347586941 to 7768804

Software Engineering (cs.SE)

104. [arXiv:1306.1812 \[pdf, other\]](#)
Reusability in Science: From Initial User Engagement to Dissemination of Results
 Kostas Malyshchikov, David Kelly, Scott J. Kiecker, Justin M. Weisman, Daniel S. Katz, Mai Zhu-Gang, Mahesh Mookherjee
 Comments: 5 pages, WISSE 2012 workshop
 Subjects: Software Engineering (cs.SE), Distributed, Parallel, and Cluster Computing (cs.DC)

105. [arXiv:1306.1812 \[pdf, other\]](#)
Cactus: Issues for Sustainable Simulation Software
 Frank Löffler, Steven R. Brandt, Gabriele Allan, Erik Schreiber
 Comments: submitted to the Workshop on Sustainable Software for Science Practice and Experiences 2013
 Subjects: Computational Engineering, Finance, and Science (cs.CR), Mathematical Software (cs.MS), Software Engineering (cs.SE)

106. [arXiv:1306.1810 \[pdf\]](#)
Niche Modeling: Ecological Metaphors for Sustainable Software in Science
 Nicholas Weber, Andrea Thimm, Michael Twidale
 Comments: Position paper submitted to: Workshop on Sustainable Software for Science Practice and Experiences (WISSEPE) SC13, Sunday, 17 November 2013, Denver, CO, USA
 Subjects: Software Engineering (cs.SE), Computer and Society (cs.CS)

107. [arXiv:1306.1809 \[pdf\]](#)
nanoHUB.org: Experiences and Challenges In Software Sustainability for a Large Scientific Community
 Lynn Zentgraf, Michael Zentgraf, Victoria Farnsworth, Michael McLennan, Kristina Mathewson, Gerhard Klenck
 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.CS), Digital Libraries (cs.DL)

108. [arXiv:1306.1786 \[pdf, ps, other\]](#)
Visit: Experiences with Sustainable Software
 Saeed Akbari, Eric Bruggier, Brad Whitlock, Jeremy S. Mendeth, Kathleen Biagus, Mark C. Miller, Hank Chads
 Subjects: Software Engineering (cs.SE)

109. [arXiv:1306.1782 \[pdf, ps, other\]](#)
DUNE as an Example of Sustainable Open Source Scientific Software Development
 Markus Blatt
 Subjects: Mathematical Software (cs.MS), Software Engineering (cs.SE)

110. [arXiv:1306.1677 \[pdf\]](#)
Toward a Research Software Security Maturity Model
 Randy Heister, Betty Thomas, Von Welch, Craig Jackson
 Comments: Submission to Workshop on Sustainable Software for Science Practice and Experiences. [This help URL](#)
 Subjects: Cryptography and Security (cs.CR), Software Engineering (cs.SE)

111. [arXiv:1306.1640 \[pdf\]](#)
Towards a Software Product Sustainability Model
 Coral Calvert, M Angeles Moraga, Manuel F. Berton
 Comments: 4 pages, 1 figure. Paper sent to Workshop on Sustainable Software for Science Practice and Experiences 2013. [This help URL](#)
 Subjects: Software Engineering (cs.SE)

112. [arXiv:1306.1630 \[pdf, other\]](#)
SimGrid: a Sustained Effort for the Versatile Simulation of Large Scale Distributed Systems
 Yann Grossebois, Arnaud Garrecq, Arnaud Legendre, Martin Quinson, Frédéric Suter
 Comments: 4 pages, submission to WISSEPE13
 Subjects: Distributed, Parallel, and Cluster Computing (cs.DC)

113. [arXiv:1306.1533 \[pdf, ps, other\]](#)

API

QUERY

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

[illegible]

QUERY

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

QUERY

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

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

QUERY

`http://export.arxiv.org/api/query?search_query=
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

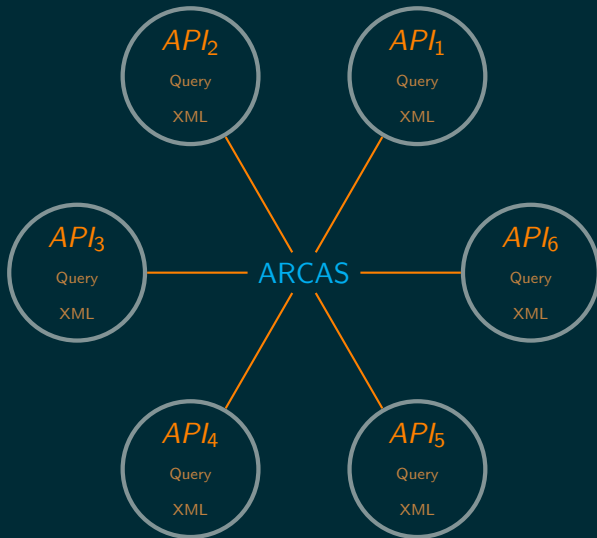
Query

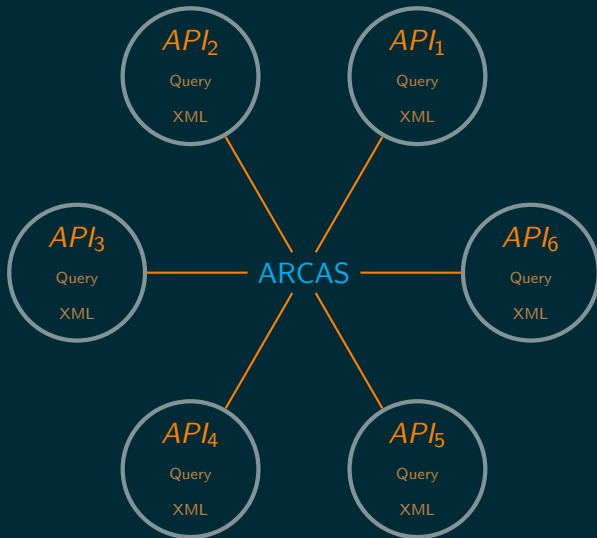
XML

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"}}
```

```
>>> 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
```

Arcas

tools.py

doc/

arcas.readthedocs.io/

ieee

nature

arxiv

...

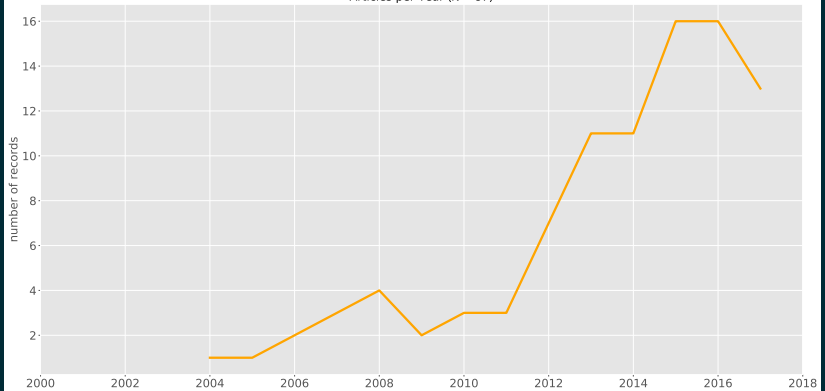
test_ieee

test_nature

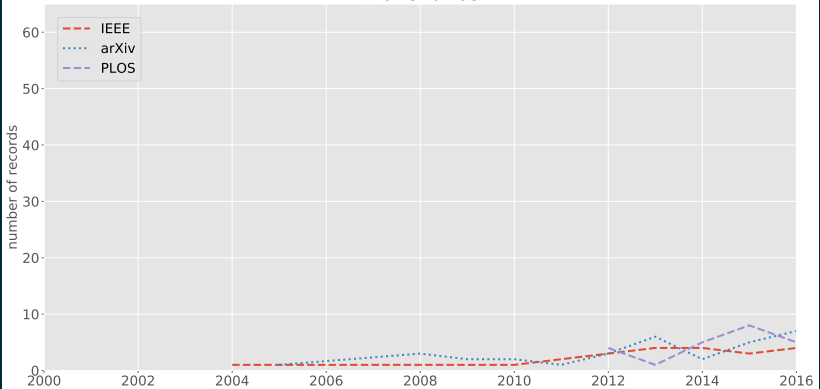
test_arxiv

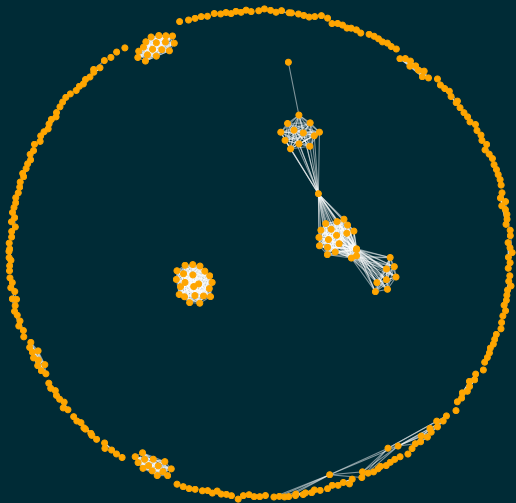
...

Articles per Year ($N = 87$)



Provenance






```
$ 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/>