CiteAs.org: Bridging the Gap in Software Citation

C. Fan Du[1], James Howison[2], Heather Piwowar[3], Jason Priem[4], and Patrice Lopez[5]

Abstract

We present CiteAs.org, a specialized web-crawling search engine that retrieves information for citing software in research publications. CiteAs.org makes citation recommendations to researchers by finding the citation requests made by those developing software, across all languages and platforms. CiteAs.org also helps educate developers of software in ways to make visible citation requests. CiteAs.org bridges the gap between software developers and end-users by providing actionable suggestions. Ultimately, we hope to give those making scientific contributions through software better ability to demonstrate their impact and thus gain the time and money needed for sustained quality software work in science.

Talk Outline

1. Motivation

- **The Undermined Visibility of Software in Science**
  - Software work is relatively invisible in science. Improving the visibility of scientific software work ensures all the software contributors are given due credit, and thus incentivizes better software work in science.

    Great software work → Clear requests for citation → More visibility in publications → More credit → Better Software → Better Research

  - One widely acknowledged way to increase visibility is to make software a citable entity in the record of scientific research. But mentions of software in research publications are often inconsistent, and are frequently informal.

Mention type	Count ($n = 286$)	Proportion (95% CI)
Cite to publication	105	0.37 (0.31–0.43)
Cite to users manual	6	0.02 (0.01–0.05)
Cite to project name or website	15	0.05 (0.03–0.09)
Instrument-like	53	0.19 (0.14–0.24)
URL in text	13	0.05 (0.03–0.08)
In-text name mention only	90	0.31 (0.26–0.37)
Not even name mentioned	4	0.01 (0.00–0.04)

Table 1: Varieties of software mentions in research publications (Howison & Bullard, 2015)

1,2 School of Information, The University of Texas at Austin
3,4 Our Research
5 science-miner

2. The Significance of Software Visibility in Science

- The non-systematic fashion of software citation does not only undermines visibility, but also hinders the general identification of software as an important class of research products within the scholarly ecosystem.
- Software embodies knowledge. Key methodologies and techniques utilized during the research process are often implemented in scientific software.
- Thus, to cite software in research papers can improve the identification, findablity, and accessibility of software. This helps researchers better understand others' research process, facilitates more contribution to the existing software, and improves the reproducibility and thus the credibility of research.

3. How to Cite Software: Best Practice Recommendation (Smith et al., 2016)

- Cite your software within the body text, as well as in the reference list.
- Give credit to all the contributors.
- Adopt a machine actionable unique identifier.
- Include access information.
- Be specific, such as versions numbers or the platform/operation system

4. However, Even a Single Software Citation can be Hard

- **Software end-users**: "Where can I find all the information?"
  - When you are citing a paper, usually you can find all the relevant information in it:

    © 2015 ASIS&T ● Published online 13 May 2015 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/asi.23538

    nication is complementary to recent in
    In sum, then, the relationship of soft

    JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY, 67(9):2137–2155, 2016

  - When you are citing a piece of software, usually the relevant information is fragmentary and everywhere. You cannot afford the time and energy to hunt for everything and then piece them together.
- **Software developers**: "Some of us actually have done a lot of work on this..."

  ### SciPy (the library)

  There is now a journal article available for citing usage of SciPy:

  Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, CJ Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E.A. Quintero, Charles R Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. (2020) **SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python**. *Nature Methods*, in press.

  Here's an example of a BibTeX entry:

  - However, in our 2019 interview study, some researchers told us they had not cited SciPy since it is often a basic building block they just take it for granted and never thought about looking for citation information on the web page of SciPy.

- Platform-specific techniques (such as Python projects duecredit and citepy) are excellent, and CiteAs.org leverages these approaches, drawing them together across languages and platforms.
- To make your software product citable, the best option is to assemble and maintain human- and machine-readable software metadata for citation, including formats like citation.ccf and CodeMeta file.

  Example: R DESCRIPTION file

```
Package: mypackage
Title: What The Package Does (one line, title case required)
Version: 0.1
Authors@R: person("First", "Last", email = "first.last@example.com",
                role = c("aut", "cre"))
Description: What the package does (one paragraph)
Depends: R (>= 3.1.0)
License: What license is it under?
LazyData: true
```

- ***But how to make your citation request more visible?***

5. CiteAs.org links between pieces of software and their requested citations

- CiteAs.org is a specialized search engine that retrieves relevant information and make recommendation for software citation. It also prompts software developers how they can further improve the visibility of their software work.

CiteAs[1]
alpha   About   API

All research products deserve credit.

Get the correct citation for diverse research products, from software
and datasets to preprints and articles.

Paste a URL, DOI, arXiv ID, or any search term (e.g. software name/abbreviation)

Examples:
http://yt-project.org   https://cran.r-project.org/web/packages/stringr   More examples

- **For software end-users**, if you search for "SciPy", you can get**:**

CiteAs[1]
alpha   About   API

Python for Scientific Computing
view website

                                                    American Psychological Association 6th edition ▼

Oliphant, T. E. (2007). Python for Scientific Computing. *Computing in Science & Engineering*, 9(3), 10–20.
http://doi.org/10.1109/mcse.2007.58

COPY   Endnote   Modify   view in API   Results not as expected?

          Refworks

          BibTeX

Citation Pro   Install the "Zotero Connector" extension

- **For software developers,** you can check if you have provided sufficient information for fulfilling a good-quality citation.

- We are working on adding more interactive features to make more intelligent recommendation to software developers

- The ultimate design envision of CiteAs.org is to bridge the gap between end-users and software producers in multiple ways.
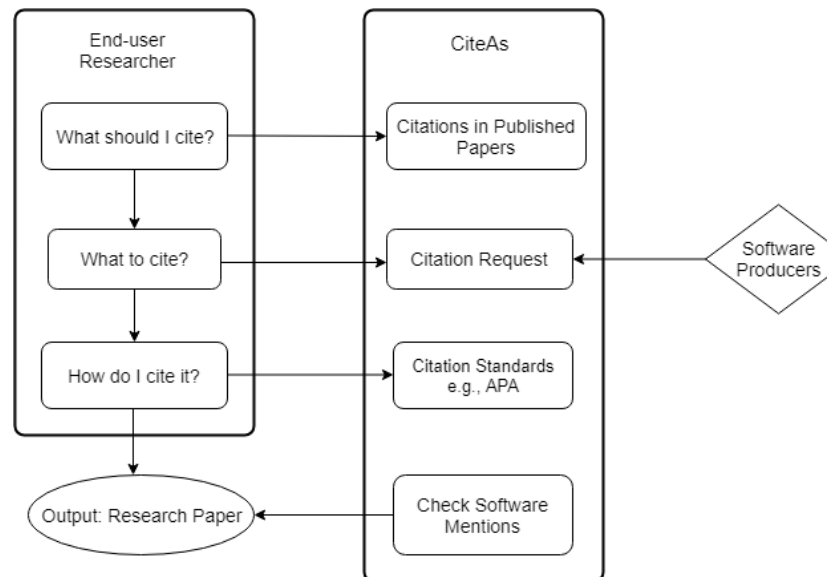
Links

- Link to previous talk @Scientific Software Repository/Registry Collaborative Workshop, The University of Maryland, College Park, November 2019 (funded by Sloan Foundation): https://figshare.com/articles/CiteAs_SSRCW_13Nov2019_pptx/11858412
- Link to previous blog post @URSSI (US Research Software Sustainability Institute, funded by National Science Foundation): CiteAs.org: Discovering and Improving Software Requests for Citation http://urssi.us/blog/2018/10/01/citeas.org-discovering-and-improving-software-requests-for-citation/