CiteAs [1] alpha : Bridging the Gaps in Software Citation

Caifan Du   The University of Texas at Austin
November 13, 2019 @ Scientific Software Registry Collaboration Workshop

James Howison   The University of Texas at Austin
Heather Piwowar   Our Research
Jason Priem   Our Research

CiteAs [1]
alpha

- Motivation: Improve the visibility of scientific software work and software citation.

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

- Linking pieces of software to citation requests: We want to discover and honor author's requests.

CiteAs[1] as a Specialized Search Engine

- Input: digital identifiers of a specific piece of software

  e.g., name of a software package; link to a Github repository/project website, DOI

- Output: A recommendation of a formatted software citation based on retrieved citation requests

CiteAs[1] Current Forms of Citation Requests

Machine Readable citation metadata:
- CITATION.cff
- CodeMeta
- R Description file
- DOI-associated metadata

Natural language citation requests:
- Project website
- Documentation
etc.

CiteAs[1]
alpha

Example:

http://citeas.org/

CiteAs [1]
alpha

We are in continuous development:

- We have been conducting stakeholder interviews since 2019
  - Seeking feedbacks for improvement
  - Looking for potential collaborators for sustainability concern

We are in continuous development:

- Towards further motivating software creators to make clear citation requests:
  - Annotating software mentions in scientific publications
  - Developing a machine learning system that automatically identify software mentions in academic texts
  - Expected to be integrated into CiteAs, prompting software creators how their software have been mentioned in research papers; thus motivates the creation of clearer citation requests

CiteAs[1]
alpha

CiteAs [1]
alpha

The column scores (the fraction of entirely correct columns) were reported in addition to Q-scores for BAliBASE 3.0. Wilcoxon signed-ranks tests were performed to calculate statistical significance of comparisons between alignment programs, which include **ProbCons** (**version 1.10**) (23), **MAFFT** (**version 5.667**) (11) with several options, **MUSCLE** (**version 3.52**) (10) and **ClustalW** (**version 1.83**) (7).

PROBCONS

Type: **software**

Raw name: **ProbCons**

Version nb: **version 1.10**

conf: *0.8174*

ProbCons is an open source probabilistic consistency-based multiple alignment of amino acid sequences. It is an efficient protein multiple sequence alignment program, which has demonstrated a statistically significant improvement in accuracy compared to several leading alignment tools.
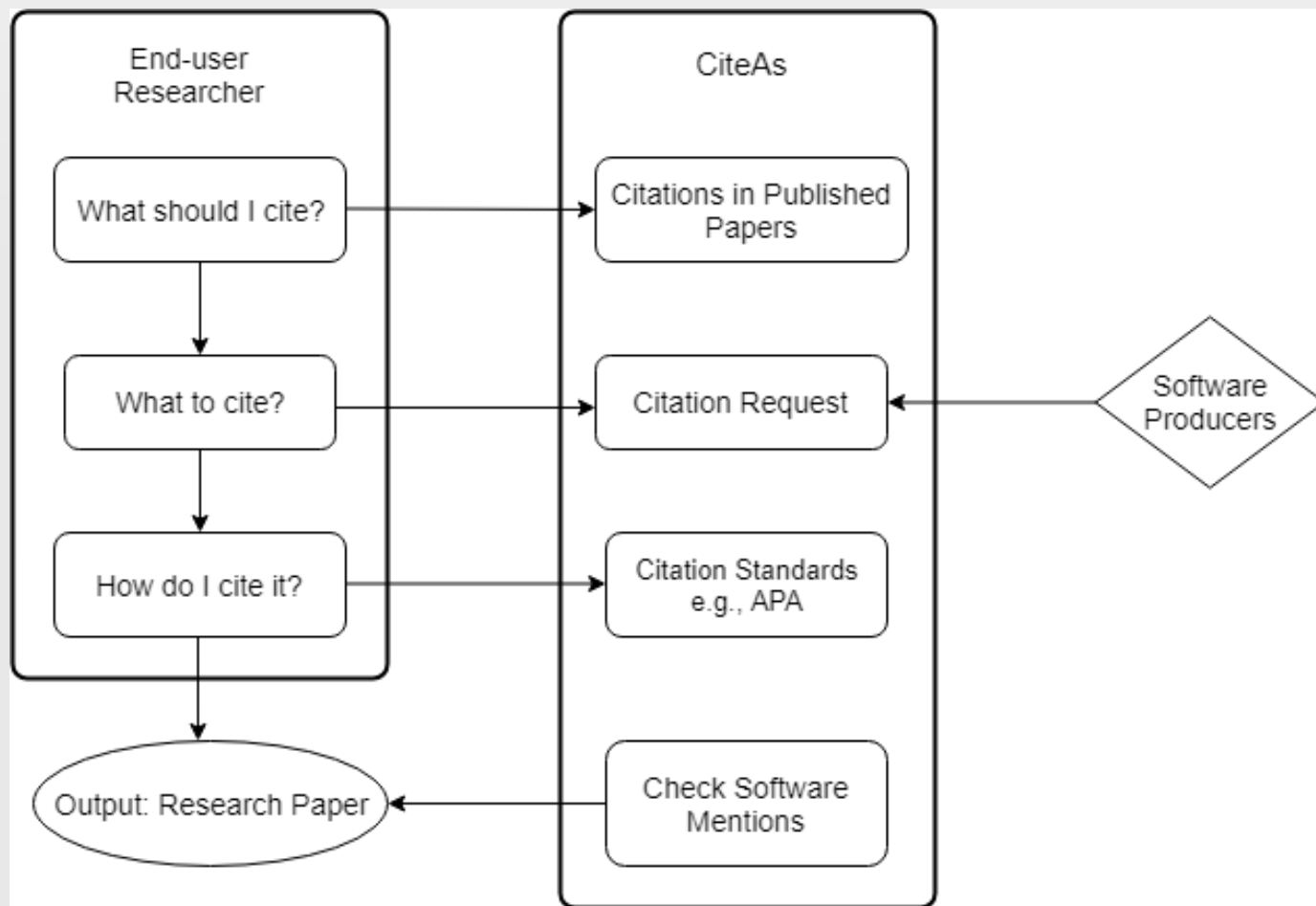
Wikidata statements ⌄

References: W ▌▌▌

Figure 1. CiteAs is designed for bridging the gaps in software citation

CiteAs[1]
alpha

Please try CiteAs and report bugs and request features.

Thanks!