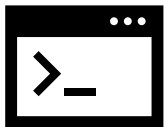


Improving the visibility of scientific software work

James Howison Lab, University of Texas at Austin

Collegeville
Workshop
2020



Software contributions to science and their developers need to be credited.

But software work is relatively invisible to scientific reputation system based on publication citations.



To increase the visibility of scientific software work:

1. Make software citation normative academic practice
2. Mine software informally mentioned in research literature.

Moving forward Software Citation

- Follow [*Software Citation Principle*](#)
- Make human- and machine-readable citation request for your software work
- Try citeas.org (and help us improve it!)

```
codemeta.json
```

```
CITATION.cff
```

BIBTEX
citation()



Automatic Extraction of Software Mentions

by James Howison Lab
& Patrice Lopez



page 1/5

Nucleic Acids Research, 2007, Vol. 35, Web Server issue W325–W329
doi:10.1093/nar/gkm303

taveRNA: a web suite for RNA algorithms and applications

Cagri Aksay¹, Raheleh Salari¹, Emre Karakoc¹, Can Alkan² and S. Cenk Sahinalp^{1,*}

¹Lab for Computational Biology, SFU, Canada and ²Department of Genome Sciences, University of Washington

Received January 31, 2007; Revised April 3, 2007; Accepted April 14, 2007

ABSTRACT

We present **taveRNA**, a web server package that hosts three RNA web services: **alterRNA**, **interRNA** and **pRuNA**. **alterRNA** is a new alternative for RNA secondary structure prediction. It is based on a dynamic programming solution that minimizes the sum of energy density and free energy of an RNA structure. **interRNA** is the first RNA-RNA interaction structure prediction web service. It also employs a dynamic programming algorithm to minimize the free energy of the resulting joint structure of the two interacting RNAs. Lastly, **pRuNA** is an efficient database pruning service; which given a query RNA, eliminates a significant portion of an ncRNA database and returns only a few ncRNAs as potential regulators. **taveRNA** is available at <http://compbio.cs.sfu.ca/taverna>.

INTRODUCTION

Until recently RNA was thought to have only two functions: (i) primarily as an information transmitter between DNA and proteins in the form of a messenger

Regulatory ncRNAs that are generally responsible for regulating gene expression exhibit an exact or partial complementarity to their target mRNAs. Their interaction forms a complex that consists of several non-contiguous helical segments which prevent ribosomal access to the target mRNA. Generally, regulatory ncRNAs contain one or more stem loop structures that are (almost) complementary to specific sequences in the target mRNAs. Interaction with a target RNA is either initiated at such a loop structure of the antisense RNA and a loop structure from the target (forming kissing loop pairs) or between a loop structure and a single-stranded segment of the complementary RNA.

As the number of ncRNAs and in particular regulatory RNAs increase it has become of crucial importance to establish software tools that can help identify their functionality. For this purpose we introduce **taveRNA**, a web-based computational tool set that can help identify structure and functionality of ncRNA molecules. **taveRNA** involves tools whose algorithmic foundations were developed by Simon Fraser University's Lab for Computational Biology over the past few years. The tools aim to solve the following key problems:

1. RNA secondary structure prediction problem, which

TAVERNA

Type: software

Raw name: taveRNA

Creator: the Lab for Computational Biology at Simon Fraser University

conf: 0.748

taveRNA is a software suite for RNA/DNA secondary structure. It is developed in the laboratories for computational biology of the School of Computing Science at the Simon Fraser University. The suite is composed by **alterRNA**, for RNA density fold computing, **interRNA**, for RNA-RNA interaction prediction, **pRuNA**, for predicting the joint partition function, equilibrium concentration, ensemble energy, and **melting temperature** for two RNA sequences, **pRuNA**, a sequence based **pruning** RNA interaction search engine, and **smyRNA**, a platform independent C program novel ab initio ncRNA finder.

Wikidata statements

official website	http://compbio.cs.sfu.ca/taverna
use	Science
use	Bioinformatics
instance of	Software

References:  

Gold-standard Dataset of Software Mentions

Existing efforts are constrained by the lack of gold-standard labeled data. Our research team have manually annotated more than 4,971 full-text academic articles over 2 years , generated 8,335 labels of software mentions and their details.

We are going to release our softcite-dataset as an open dataset.



```
<text xml:lang="en">
  <body>
    <p>All the analysis was performed in the <rs cert="1.0" resp="#annotator12"
      subtype="used" type="software" xml:id="f33d05cff5-software-0">MATLAB</rs> environment
      (<rs corresp="#f33d05cff5-software-0" resp="#curator" type="publisher">The MathWorks</
      rs>, Natick, MA) using <rs corresp="#f33d05cff5-software-1" resp="#curator"
      type="publisher">OMLAB</rs> software (<rs cert="1.0" resp="#curator" subtype="used"
      type="software" xml:id="f33d05cff5-software-1">OMtools</rs>, downloadable from <rs
      corresp="#f33d05cff5-software-1" resp="#curator" type="url">http://www.omlab.org</rs>).
      Eye position was sampled directly; it was prefiltered using a low-pass filter with a
```

With Softcite Dataset...

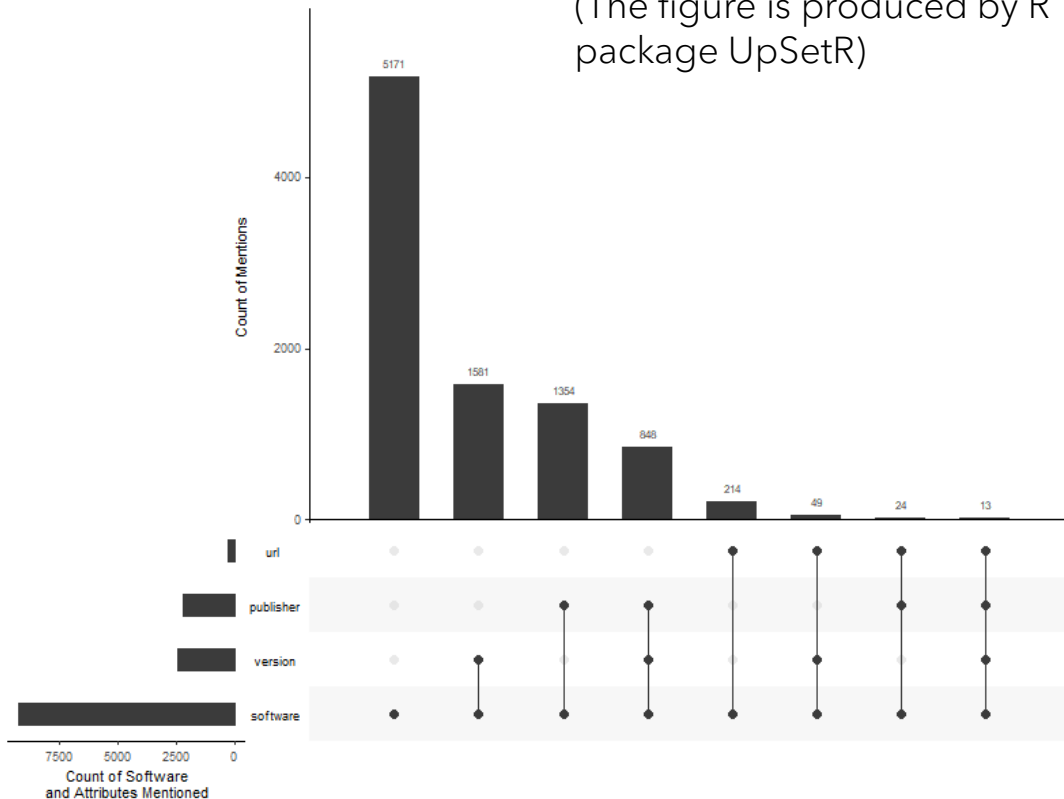
- We are scaling up software entity recognition in tens of millions of open access academic publications
- We plan to link the recognized software in research literature to software dependency data to analyze the dependency risk of domain scientific software
- We intend to further expand the gold-standard labeled dataset to advance the generalizability, scalability, and performance of software entity recognition

In Softcite Dataset

We found software is rarely mentioned with its access information.

How do you request citation for your software work?

(The figure is produced by R package UpSetR)



Contact Info

C. Fan Du | cfdu@utexas.edu

Howison Lab, University of Texas at Austin



**ALFRED P. SLOAN
FOUNDATION**

This work is supported by Alfred P. Sloan Foundation Digital Science Program.

Collegeville
Workshop
2020