My past, present and potential future of releasing software with my publications

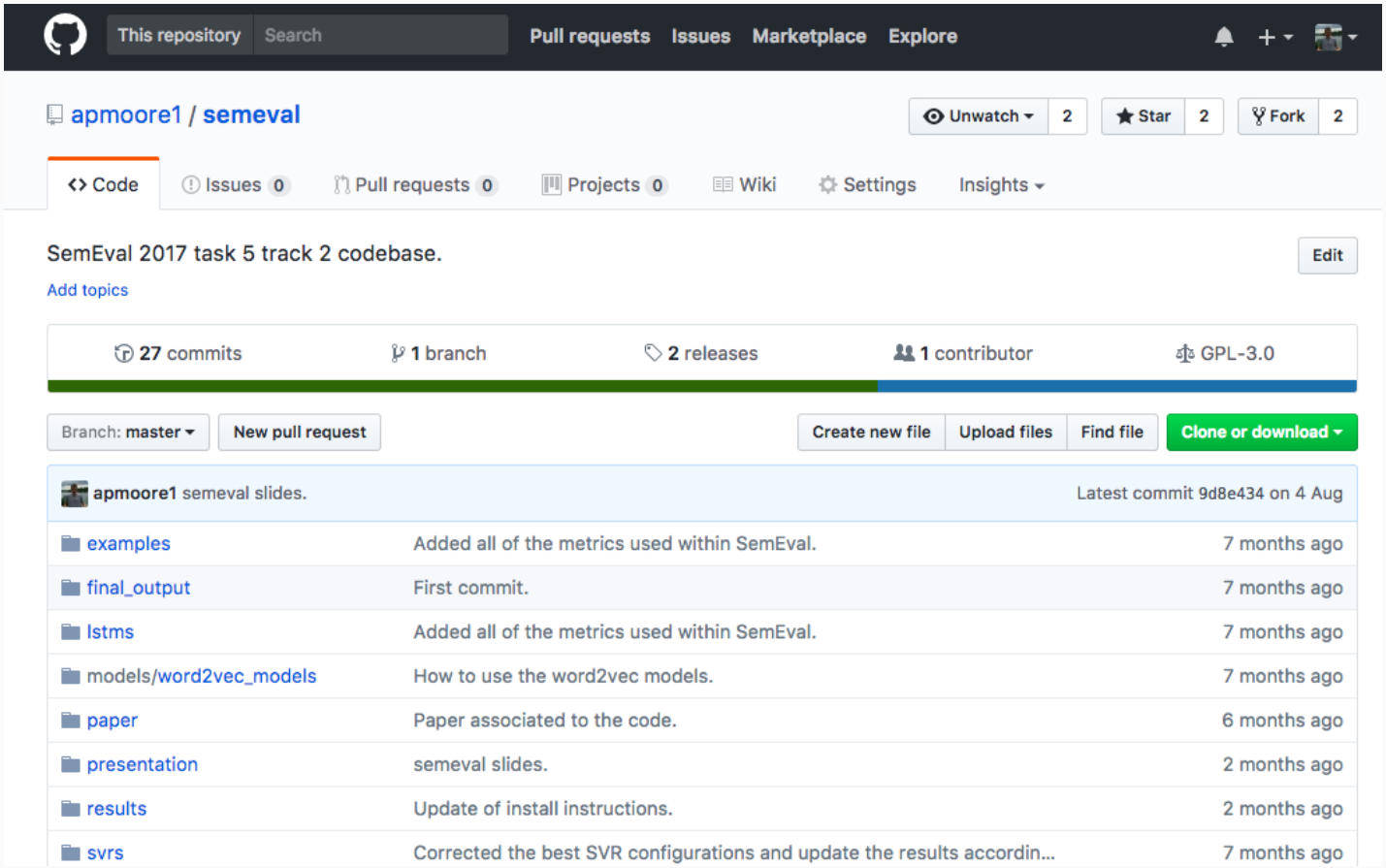
Andrew Moore

February 27, 2018

School of Computing and Communications, Lancaster University.

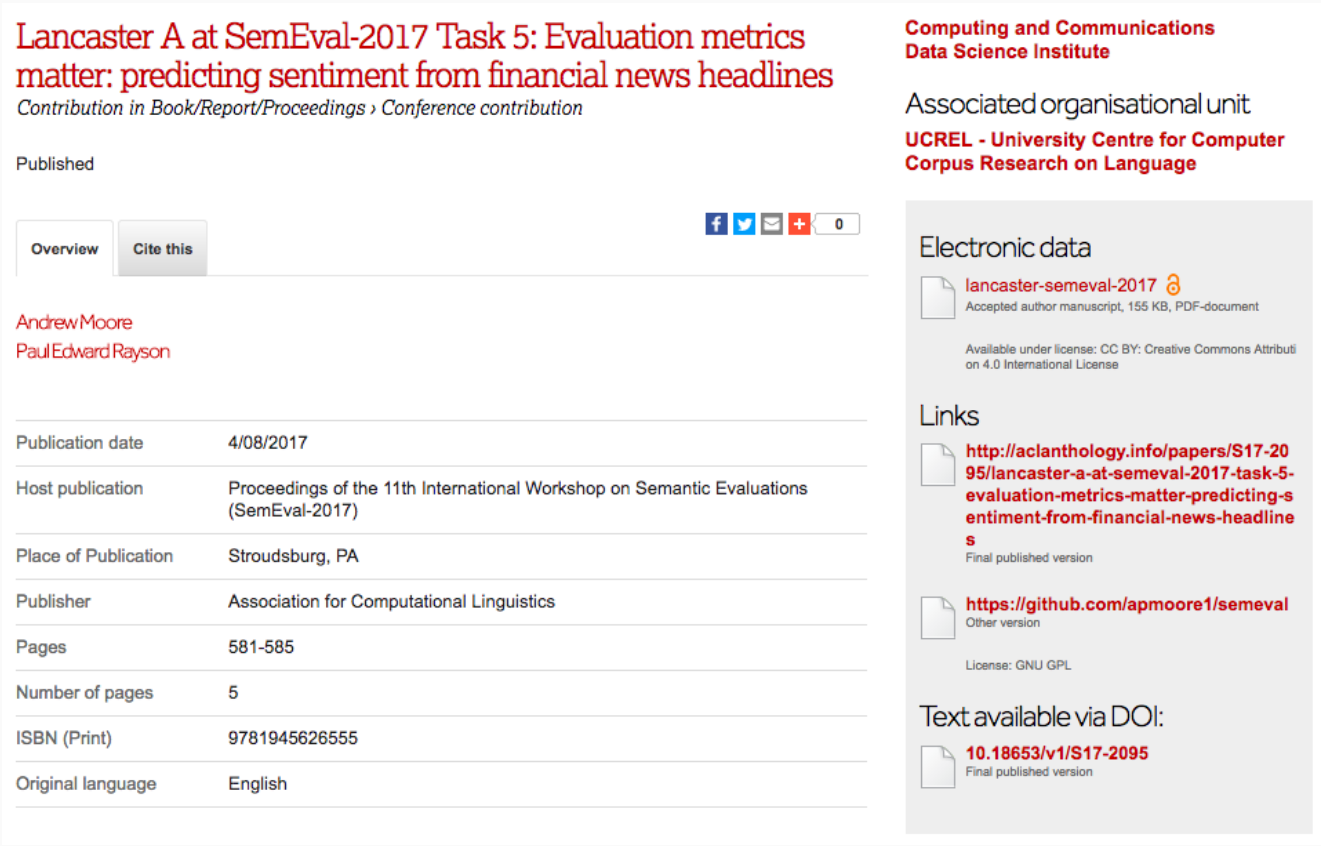
Past

SemEval publication on predicting sentiment in financial news headlines[2]



How people can find the code

1. Link in the paper.
2. Via the research directory



Lancaster A at SemEval-2017 Task 5: Evaluation metrics matter: predicting sentiment from financial news headlines

Contribution in Book/Report/Proceedings › Conference contribution

Published

Overview | Cite this

Andrew Moore
Paul Edward Rayson

Publication date	4/08/2017
Host publication	Proceedings of the 11th International Workshop on Semantic Evaluations (SemEval-2017)
Place of Publication	Stroudsburg, PA
Publisher	Association for Computational Linguistics
Pages	581-585
Number of pages	5
ISBN (Print)	9781945626555
Original language	English

Computing and Communications
Data Science Institute

Associated organisational unit
UCREL - University Centre for Computer Corpus Research on Language

Electronic data
lancaster-semeval-2017
Accepted author manuscript, 155 KB, PDF-document

Available under license: CC BY: Creative Commons Attribution 4.0 International License

Links
http://aclanthology.info/papers/S17-2095/lancaster-a-at-semeval-2017-task-5-evaluation-metrics-matter-predicting-sentiment-from-financial-news-headlines
Final published version

https://github.com/apmoore1/semeval
Other version

License: GNU GPL

Text available via DOI:
10.18653/v1/S17-2095
Final published version

Present

The problems that I found from my past:

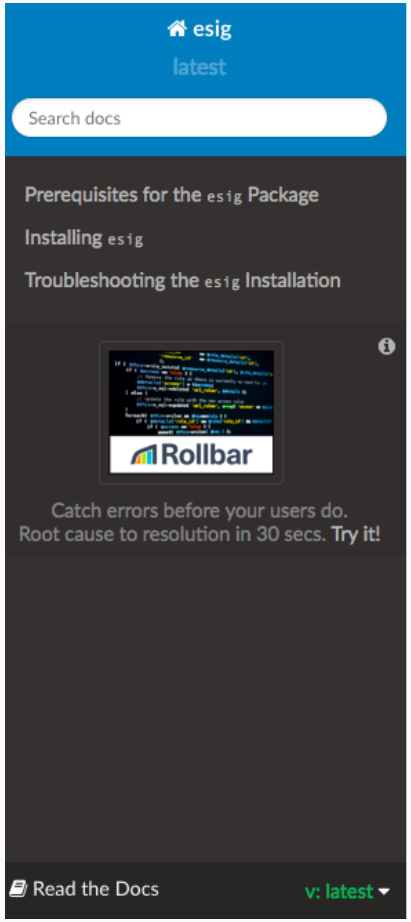
1. The code does not prove what I have done has been implemented correctly.
2. The code lacks detailed documentation.
3. The code could be easier to find.

My solutions to these problems:

1. Create test e.g. unit test.
2. Create detailed documentation like readthedocs.
3. My profile on the research directory to include software tab.

readthedocs example

esig python package[1]



Docs » The esig Python Package — Edit on GitHub

The esig Python Package

This is the online documentation for the Python esig package. Below you'll find the table of contents.

esig: What is it?

[1] http://esig.readthedocs.io/en/latest/

Research directory example

Andrew Moore

Research student

Profile | Publications

Computing and Communications

InfoLab21 Lancaster University Bailrigg Lancaster
LA1 4WA
LA1 4WA
United Kingdom

E-mail: Show email

ORCID: 0000-0002-3395-0841

0

Publications & Outputs

Domain adaptation using stock market prices to refine sentiment dictionaries — Published

Moore, A., Rayson, P. E. & Young, S. E. 23/05/2016 *Proceedings of the 10th edition of Language Resources and Evaluation Conference (LREC2016)*. European Language Resources Association (ELRA), 4 p.

Contribution in Book/Report/Proceedings › Conference contribution

Lancaster A at SemEval-2017 Task 5: Evaluation metrics matter: predicting sentiment from financial news headlines — Published

Moore, A. & Rayson, P. E. 4/08/2017 *Proceedings of the 11th International Workshop on Semantic Evaluations (SemEval-2017)*. Stroudsburg, PA: Association for Computational Linguistics, p. 581-585 5 p.

Contribution in Book/Report/Proceedings › Conference contribution

Future

What I would like to see more of:

1. More researchers releasing their code.

The reasons I hear why researchers don't release their code:

1. I don't have the time.
2. I don't like the way the code is at the moment.
3. It is my code/ I don't want to.

My answers/solutions

1. Releasing the code no matter what it looks like is better than not.
2. When it is released others might help you.
3. Research Software Engineers (RSE)
4. More time/money

Spending more time/money is better

Reasons for RSE and more time on making the code easier to use:

1. Higher impact for your research
2. Allow other researchers to perform their job faster

Example

Stanford CoreNLP[1][2] has:

1. 1711 citations
2. 3771 stars and been forked 1409 times on Github

,

[2] https://stanfordnlp.github.io/CoreNLP/

Questions?

Andrew Moore
@apmoore94
Slides: `https://github.com/apmoore1/software-as-data`

References I

[1] C. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. Bethard, and D. McClosky.
The stanford corenlp natural language processing toolkit.
In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60. Association for Computational Linguistics, 2014.

[2] A. Moore and P. Rayson.
Lancaster a at semeval-2017 task 5: Evaluation metrics matter: predicting sentiment from financial news headlines.
In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 581–585. Association for Computational Linguistics, 2017.