AllenNLP

An open-source NLP research library, built on PyTorch

Matt Gardner, Joel Grus, Mark Neumann, Oyvind Tafjord, Pradeep Dasigi, Nelson F. Liu, Matthew Peters, Michael Schmitz, Luke Zettlemoyer

… and the list keeps growing

ALLEN INSTITUTE *for* ARTIFICIAL INTELLIGENCE

AllenNLP

An open-source NLP research library, built on PyTorch

- Made to make NLP research easy
- Abstractions designed for NLP
- Configuration-driven experiments for doing good science
- Reference implementations and demos for a lot of tasks
- An active community

AI2 ALLEN INSTITUTE *for* ARTIFICIAL INTELLIGENCE

What if...

AllenNLP

An open-source NLP research library, built on PyTorch

- Clean implementations of state-of-the-art models for virtually any NLP task
    - Dramatically lowers barrier to entry for doing NLP research

AllenNLP

An open-source NLP research library, built on PyTorch

- Live demos of all of these models that you can play around with and break
    - Mark Johnson used these yesterday to demonstrate a point about linguistics
    - Plenty of usage in twitter conversations about NLP models

ALLEN INSTITUTE *for* ARTIFICIAL INTELLIGENCE

AllenNLP

An open-source NLP research library, built on PyTorch

- Allows for more fundamental, wide-ranging NLP research
    - Test your idea on all NLP tasks, instead of architecture engineering on a single task

ALLEN INSTITUTE
for ARTIFICIAL INTELLIGENCE

- We’re not there yet, but with a little help, we could be
  - We’re a small team, we can’t do everything
  - One possibility: make a model re-implementation a class project in your intro course
  - Issues to solve around control and credit assignment

The ACL Anthology

Current State and Future Directions

Daniel Gildea, **Min-Yen Kan,** Nitin Madnani, Christoph Teichmann, Martin Villalba

What is this presentation about?

The ACL Anthology

Current state and future directions

Daniel Gildea
Department of Computer Science
University of Rochester
gildea@cs.rochester.edu

Min-Yen Kan
School of Computing
National University of Singapore
kanmy@comp.nus.edu.sg

Nitin Madnani
Educational Testing Service
Princeton NJ
nmadnani@ets.org

Christoph Teichmann **Martín Villalba**
Dept. of Language, Science, and Technology
Saarland University
{cteichmann|villalba}@coli.uni-saarland.de

The ACL Anthology is a service offered by the Association for Computational Linguistics (ACL), allowing open access to the proceedings of all ACL-sponsored conferences and journal articles. It offers search of indexed papers, author-specific pages, and services that can be embedded within pages. It is Open Source, and maintained on a volunteer basis.

A brief history of the Anthology

2001	The Anthology is proposed as a project to the ACL Executive by Steven Bird.
2002	First version of the Anthology, with **Steven Bird** taking the position of Editor.
2008	**Min-Yen Kan** takes the role of Editor.
2012	A second version of the Anthology is released. Maintenance continues to this day.
2017	After several years hosted in the **University of Singapore**, the Anthology relocates to **Saarland University**.

Current state of the Anthology

43k
hosted papers

4.5k
daily hits

The Anthology is a community project run by volunteers, coordinated by the Editor. Hosting has been typically provided by Universities free of charge and all code is publicly available on Github:

https://github.com/acl-org/acl-anthology

Web Server
NGINX
Operating System
Debian Stable

Database server
PostgreSQL

Framework
Ruby on Rails

Search engine
Solr

Future proofing the Anthology

With the Anthology in a stable state, it is time to plan ahead. We identified three main areas of work:

- Documenting every aspect of the system, to simplify onboarding of new members when project members change.
- Update/replace all outdated dependencies.
- Reinforce our development process to quickly detect bugs and/or data inconsistencies.

Backlog	In research	In progress	Done
Upgrade and/or migrate outdated dependencies	Full test coverage and consistency checks	Docker image for releases	Add index support for popular search engines
Full text search over uploaded papers		Add a staging server	Document and update the installation process
			Add a test server

Challenges for the Community

We want the Anthology to grow beyond a repository of scientific papers. We invite the community to contribute ideas and implementations on all areas. We suggest two possible first projects:

- Add anonymous pre-print support, helping authors and preserving double-blind review - either as collaboration with pre-print services like ArXiv, or as an extension
- Use the paper database to find suitable reviewers for submissions in future conferences

You can also download our Docker image, including all metadata on ACL papers, and start hacking right away. For references on other cool projects involving the Anthology, make sure to check the **Further Reading** section!

Join the Anthology team!

We are always looking for new volunteers to keep the Anthology running. Programmers (especially Ruby on Rails!), system administrators, software engineers... every little bit helps!

Interested? Know someone? Get in touch with us! Contact any of the authors, or follow us on Github!

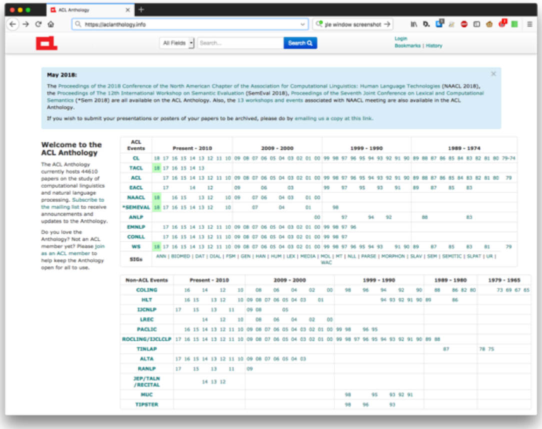
Further reading:

Anderson, McFarland and Jurafsky: *Towards a Computational History of the ACL: 1998-2008*. Proceedings of the ACL Special Workshop 2012 on Rediscovering 50 years of Discoveries.
Bird et. al: *The ACL Anthology Reference Corpus: A Reference Dataset for Bibliographic Research in Computational Linguistics*. Language Resources and Evaluation Conference (LREC) 2008.

UNIVERSITY of ROCHESTER
NUS National University of Singapore
ETS Educational Testing Service
UNIVERSITÄT DES SAARLANDES

- Summarize the history and current state of efforts related to the Anthology
- Illustrate the challenges of maintaining a community Project
- Invite the community to extend the capabilities of the Anthology
- Call you to join the Anthology team

The Anthology in **summary**



May 2018:

The Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL 2018), the Proceedings of The 12th International Workshop on Semantic Evaluation (SemEval 2018), Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics (*Sem 2018) are all available on the ACL Anthology. Also, the 13 workshops and events associated with NAACL meeting are also available in the ACL Anthology.

If you wish to submit your presentations or posters of your papers to be archived, please do by emailing us a copy at this link.

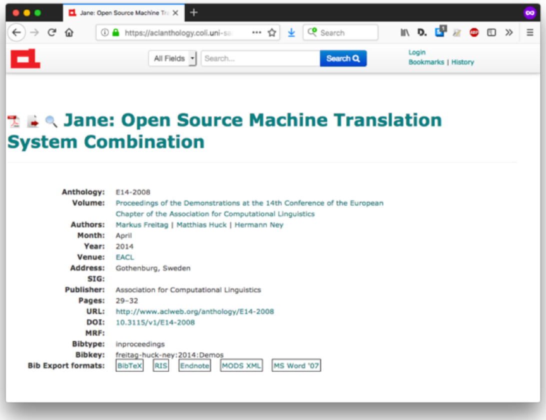
Welcome to the ACL Anthology

The ACL Anthology currently hosts 44610 papers on the study of computational linguistics and natural language processing. Subscribe to the mailing list to receive announcements and updates to the Anthology.

Do you love the Anthology? Not an ACL member yet? Please join as an ACL member to help keep the Anthology open for all to use.

ACL Events	Present - 2010	2009 - 2000	1999 - 1990	1989 - 1974
CL	18 17 16 15 14 13 12 11 10	09 08 07 06 05 04 03 02 01 00	99 98 97 96 95 94 93 92 91 90	89 88 87 86 85 84 83 82 81 80 79-74
TACL	18 17 16 15 14 13			
ACL	17 16 15 14 13 12 11 10	09 08 07 06 05 04 03 02 01 00	99 98 97 96 95 94 93 92 91 90	89 88 87 86 85 84 83 82 81 80 79
EACL	17 14 12	09 06 03	99 97 95 93 91	89 87 85 83
NAACL	18 16 15 13 12 10	09 07 06 04 03 01 00		
*SEMEVAL	18 17 16 15 14 13 12 10	07 04 01	98	
ANLP		00	97 94 92	88 83
EMNLP	17 16 15 14 13 12 11 10	09 08 07 06 05 04 03 02 01 00	99 98 97 96	
CONLL	17 16 15 14 13 12 11 10	09 08 07 06 05 04 03 02 01 00	99 98 97	
WS	18 17 16 15 14 13 12 11 10	09 08 07 06 05 04 03 02 01 00	99 98 97 96 95 94 93 91 90	89 87 85 83 81 79
SIGs	ANN | BIOMED | DAT | DIAL | FSM | GEN | HAN | HUM | LEX | MEDIA | MOL | MT | NLL | PARSE | MORPHON | SLAV | SEM | SEMITIC | SLPAT | UR | WAC			

Non-ACL Events	Present - 2010	2009 - 2000	1999 - 1990	1989 - 1980	1979 - 1965
COLING	16 14 12 10	08 06 04 02 00	98 96 94 92 90	88 86 82 80	73 69 67 65
HLT	16 15 13 12 10	09 08 07 06 05 04 03 01	94 93 92 91 90	89 86	
IJCNLP	17 15 13 11	09 08 05			
LREC	14 12 10	08 06 04 02 00			
PACLIC	16 15 14 13 12 11 10	09 08 07 06 05 04 03 02 01 00	99 98 96 95		
ROCLING/IJCLCLP	17 16 15 14 13 12 11 10	09 08 07 06 05 04 03 02 01 00	99 98 97 96 95 94 93 92 91 90	89 88	
TINLAP				87	78 75
ALTA	17 16 15 14 13 12 11 10	09 08 07 06 05 04 03			
RANLP	17 15 13 11	09			
JEP/TALN /RECITAL	14 13 12				
MUC			98 95 93 92 91		
TIPSTER			98 96 93		

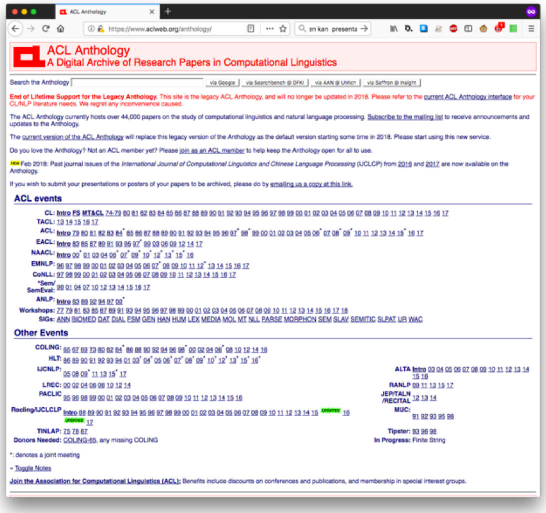


Jane: Open Source Machine Translation System Combination

Anthology:	E14-2008
Volume:	Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics
Authors:	Markus Freitag | Matthias Huck | Hermann Ney
Month:	April
Year:	2014
Venue:	EACL
Address:	Gothenburg, Sweden
SIG:	
Publisher:	Association for Computational Linguistics
Pages:	29–32
URL:	http://www.aclweb.org/anthology/E14-2008
DOI:	10.3115/v1/E14-2008
MRF:	
Bibtype:	inproceedings
Bibkey:	freitag-huck-ney:2014:Demos
Bib Export formats:	BibTeX RIS Endnote MODS XML MS Word '07

- Open access service for all ACL-Sponsored publications
- Also hosts posters and additional data
- Paper search and author pages
- 45K papers and 4.5K daily hits
- Open Source
- Maintained by volunteers
- New papers added in collaboration with proceedings editors

A brief **History** of the Anthology



ACL Anthology
A Digital Archive of Research Papers in Computational Linguistics

Search the Anthology [via Google] [via Searchbench @ DFKI] [via AAN @ UMich] [via Saffron @ Insight]

End of Lifetime Support for the Legacy Anthology. This site is the legacy ACL Anthology, and will no longer be updated in 2018. Please refer to the current ACL Anthology interface for your CL/NLP literature needs. We regret any inconvenience caused.

The ACL Anthology currently hosts over 44,000 papers on the study of computational linguistics and natural language processing. Subscribe to the mailing list to receive announcements and updates to the Anthology.

The current version of the ACL Anthology will replace this legacy version of the Anthology as the default version starting some time in 2018. Please start using this new service.

Do you love the Anthology? Not an ACL member yet? Please join as an ACL member to help keep the Anthology open for all to use.

NEW Feb 2018: Past journal issues of the *International Journal of Computational Linguistics and Chinese Language Processing* (IJCLCP) from 2016 and 2017 are now available on the Anthology.

If you wish to submit your presentations or posters of your papers to be archived, please do by emailing us a copy at this link.

ACL events

CL: Intro FS MT&CL 74-79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17
TACL: 13 14 15 16 17
ACL: Intro 79 80 81 82 83 84* 85 86 87 88 89 90 91 92 93 94 95 96 97* 98* 99 00 01 02 03 04 05 06* 07 08* 09* 10 11 12 13 14 15* 16 17
EACL: Intro 83 85 87 89 91 93 95 97* 99 03 06 09 12 14 17
NAACL: Intro 00* 01 03 04 06* 07* 09* 10* 12* 13* 15* 16
EMNLP: 96 97 98 99 00 01 02 03 04 05 06 07* 08 09 10 11 12* 13 14 15 16 17
CoNLL: 97 98 99 00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17
***Sem/ SemEval:** 98 01 04 07 10 12 13 14 15 16 17
ANLP: Intro 83 88 92 94 97 00*
Workshops: 77 79 81 83 85 87 89 91 93 94 95 96 97 98 99 00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18
SIGs: ANN BIOMED DAT DIAL FSM GEN HAN HUM LEX MEDIA MOL MT NLL PARSE MORPHON SEM SLAV SEMITIC SLPAT UR WAC

Other Events

COLING: 65 67 69 73 80 82 84* 86 88 90 92 94 96 98* 00 02 04 06* 08 10 12 14 16
HLT: 86 89 90 91 92 93 94 01 03* 04* 05 06* 07* 08* 09* 10* 12* 13* 15* 16*
IJCNLP: 05 08 09* 11 13 15* 17
LREC: 00 02 04 06 08 10 12 14
PACLIC 95 96 98 99 00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16
Rocling/IJCLCLP Intro 88 89 90 91 92 93 94 95 96 97 98 99 00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 UPDATED 16 UPDATED 17
TINLAP: 75 78 87
Donors Needed: COLING-65, any missing COLING

ALTA Intro 03 04 05 06 07 08 09 10 11 12 13 14 15 16
RANLP 09 11 13 15 17
JEP/TALN /RECITAL 12 13 14
MUC: 91 92 93 95 98
Tipster: 93 96 98
In Progress: Finite String

*: denotes a joint meeting

- Toggle Notes

Join the Association for Computational Linguistics (ACL): Benefits include discounts on conferences and publications, and membership in special interest groups.

Steven Bird

Min-Yen Kan

- Proposed in 2001 by Steven Bird
- First version online in 2002, with Steven Bird as editor
- Min-Yen Kan becomes the new editor in 2008
- A new version of the Anthology with extra functionality is released in 2012
- Hosting of the Anthology moves from the National University of Singapore to Saarland University

How to **Future-proof** the Anthology

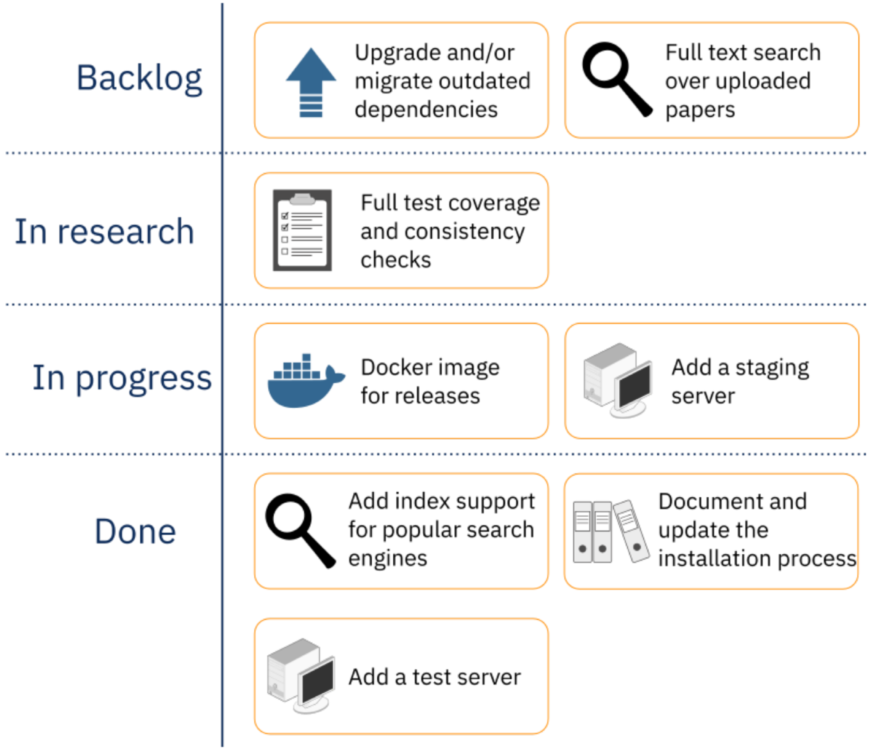
Challenges

- Limited resources for day-to-day code maintenance
- Dependencies become outdated
- Maintainer churn

Solutions

- Docker container for easier set-up and sandboxing
- Collaborative documentation efforts to ease onboarding
- Migration plan on the pipeline, including upgrades and test cases

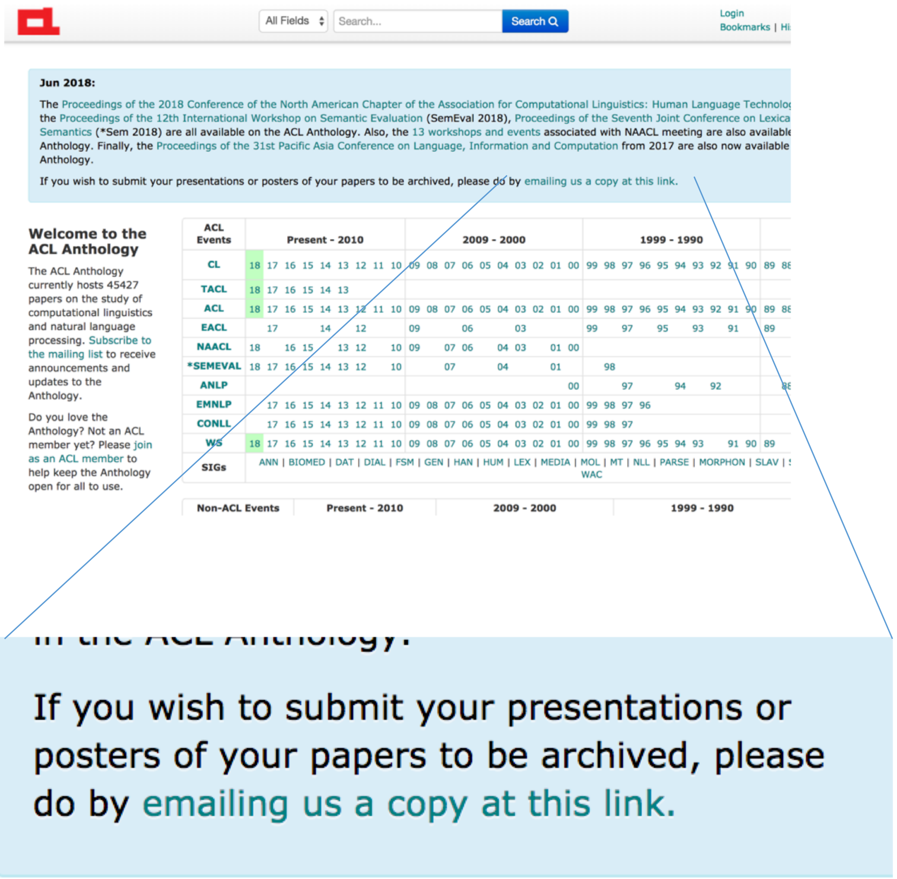
Upcoming major steps



Backlog	Upgrade and/or migrate outdated dependencies	Full text search over uploaded papers
In research	Full test coverage and consistency checks	
In progress	Docker image for releases	Add a staging server
Done	Add index support for popular search engines	Document and update the installation process
	Add a test server	

- Hosting the Anthology within the main ACL website
- Recruit a new Anthology editor
- (possibly) pay for extra support
  for the Anthology

Exercise: Importing of your slides



Jun 2018:

The Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technolog the Proceedings of the 12th International Workshop on Semantic Evaluation (SemEval 2018), Proceedings of the Seventh Joint Conference on Lexica Semantics (*Sem 2018) are all available on the ACL Anthology. Also, the 13 workshops and events associated with NAACL meeting are also available Anthology. Finally, the Proceedings of the 31st Pacific Asia Conference on Language, Information and Computation from 2017 are also now available Anthology.

If you wish to submit your presentations or posters of your papers to be archived, please do by emailing us a copy at this link.

Welcome to the ACL Anthology

The ACL Anthology currently hosts 45427 papers on the study of computational linguistics and natural language processing. Subscribe to the mailing list to receive announcements and updates to the Anthology.

Do you love the Anthology? Not an ACL member yet? Please join as an ACL member to help keep the Anthology open for all to use.

<table>
<tr><th>ACL Events</th><th>Present - 2010</th><th>2009 - 2000</th><th>1999 - 1990</th><th></th></tr>
<tr><td>CL</td><td>18 17 16 15 14 13 12 11 10</td><td>09 08 07 06 05 04 03 02 01 00</td><td>99 98 97 96 95 94 93 92 91 90</td><td>89 88</td></tr>
<tr><td>TACL</td><td>18 17 16 15 14 13</td><td></td><td></td><td></td></tr>
<tr><td>ACL</td><td>18 17 16 15 14 13 12 11 10</td><td>09 08 07 06 05 04 03 02 01 00</td><td>99 98 97 96 95 94 93 92 91 90</td><td>89 88</td></tr>
<tr><td>EACL</td><td>17 14 12</td><td>09 06 03</td><td>99 97 95 93 91</td><td>89</td></tr>
<tr><td>NAACL</td><td>18 16 15 13 12 10</td><td>09 07 06 04 03 01 00</td><td></td><td></td></tr>
<tr><td>*SEMEVAL</td><td>18 17 16 15 14 13 12 10</td><td>07 04 01</td><td>98</td><td></td></tr>
<tr><td>ANLP</td><td></td><td>00</td><td>97 94 92</td><td>88</td></tr>
<tr><td>EMNLP</td><td>17 16 15 14 13 12 11 10</td><td>09 08 07 06 05 04 03 02 01 00</td><td>99 98 97 96</td><td></td></tr>
<tr><td>CONLL</td><td>17 16 15 14 13 12 11 10</td><td>09 08 07 06 05 04 03 02 01 00</td><td>99 98 97</td><td></td></tr>
<tr><td>WS</td><td>18 17 16 15 14 13 12 11 10</td><td>09 08 07 06 05 04 03 02 01 00</td><td>99 98 97 96 95 94 93 91 90</td><td>89</td></tr>
<tr><td>SIGs</td><td colspan="4">ANN | BIOMED | DAT | DIAL | FSM | GEN | HAN | HUM | LEX | MEDIA | MOL | MT | NLL | PARSE | MORPHON | SLAV | S
WAC</td></tr>
</table>

Non-ACL Events	Present - 2010	2009 - 2000	1999 - 1990

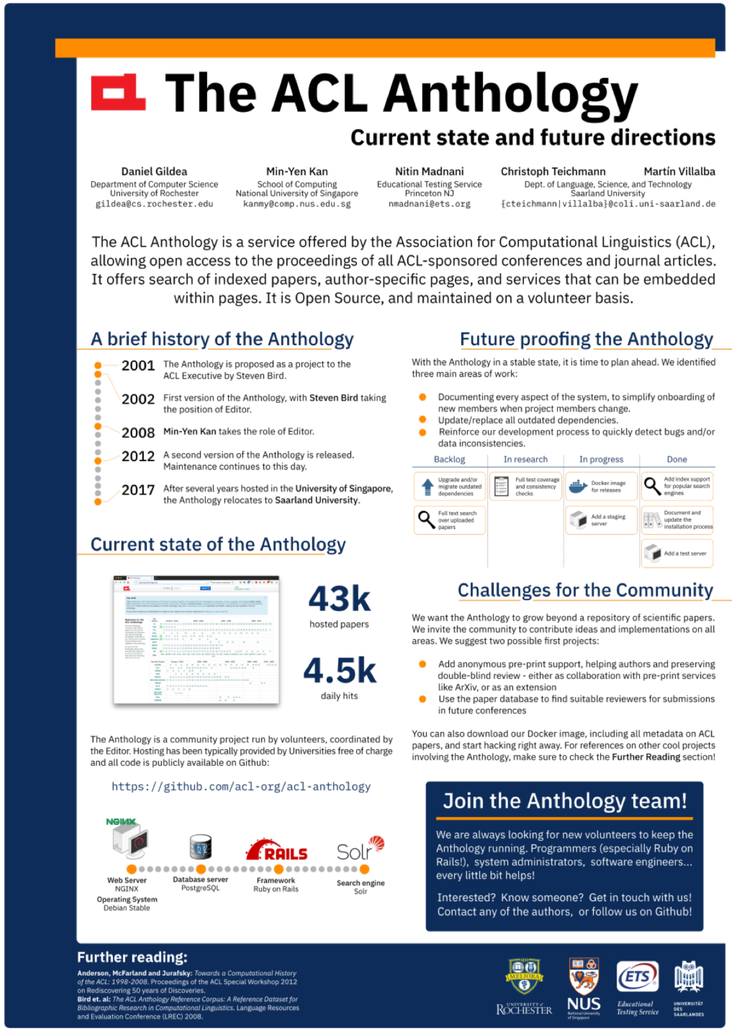
in the ACL Anthology.

If you wish to submit your presentations or posters of your papers to be archived, please do by emailing us a copy at this link.

- We import slides, datasets, videos from your own
- Currently done by email (try it yourself! yes, now)
- Better workflow: pull request against the Anthology XML (à la csrankings.org)

Possible **future directions**

- Contains useful information both *for* CL researchers and *about* CL researchers. Useful for identifying suitable reviewers.
- Move focus from day-to-day operations towards development
- Establish a network of mirrors
- Host anonymized pre-prints



The ACL Anthology

Current state and future directions

Daniel Gildea
Department of Computer Science
University of Rochester
gildea@cs.rochester.edu

Min-Yen Kan
School of Computing
National University of Singapore
kanmy@comp.nus.edu.sg

Nitin Madnani
Educational Testing Service
Princeton NJ
nmadnani@ets.org

Christoph Teichmann **Martín Villalba**
Dept. of Language, Science, and Technology
Saarland University
{cteichmann|villalba}@coli.uni-saarland.de

The ACL Anthology is a service offered by the Association for Computational Linguistics (ACL), allowing open access to the proceedings of all ACL-sponsored conferences and journal articles. It offers search of indexed papers, author-specific pages, and services that can be embedded within pages. It is Open Source, and maintained on a volunteer basis.

A brief history of the Anthology

2001 The Anthology is proposed as a project to the ACL Executive by Steven Bird.

2002 First version of the Anthology, with **Steven Bird** taking the position of Editor.

2008 **Min-Yen Kan** takes the role of Editor.

2012 A second version of the Anthology is released. Maintenance continues to this day.

2017 After several years hosted in the **University of Singapore**, the Anthology relocates to **Saarland University**.

Current state of the Anthology

43k
hosted papers

4.5k
daily hits

The Anthology is a community project run by volunteers, coordinated by the Editor. Hosting has been typically provided by Universities free of charge and all code is publicly available on Github:

https://github.com/acl-org/acl-anthology

Web Server NGINX Operating System Debian Stable	Database server PostgreSQL	Framework Ruby on Rails	Search engine Solr

Future proofing the Anthology

With the Anthology in a stable state, it is time to plan ahead. We identified three main areas of work:

- Documenting every aspect of the system, to simplify onboarding of new members when project members change.
- Update/replace all outdated dependencies.
- Reinforce our development process to quickly detect bugs and/or data inconsistencies.

Backlog	In research	In progress	Done
Upgrade and/or migrate outdated dependencies	Full test coverage and consistency checks	Docker image for releases	Add index support for popular search engines
Full text search over uploaded papers		Add a staging server	Document and update the installation process
			Add a test server

Challenges for the Community

We want the Anthology to grow beyond a repository of scientific papers. We invite the community to contribute ideas and implementations on all areas. We suggest two possible first projects:

- Add anonymous pre-print support, helping authors and preserving double-blind review - either as collaboration with pre-print services like ArXiv, or as an extension
- Use the paper database to find suitable reviewers for submissions in future conferences

You can also download our Docker image, including all metadata on ACL papers, and start hacking right away. For references on other cool projects involving the Anthology, make sure to check the **Further Reading** section!

Join the Anthology team!

We are always looking for new volunteers to keep the Anthology running. Programmers (especially Ruby on Rails!), system administrators, software engineers... every little bit helps!

Interested? Know someone? Get in touch with us! Contact any of the authors, or follow us on Github!

Further reading:

Anderson, McFarland and Jurafsky: *Towards a Computational History of the ACL: 1998-2008*. Proceedings of the ACL Special Workshop 2012 on Rediscovering 50 years of Discoveries.

Bird et. al: *The ACL Anthology Reference Corpus: A Reference Dataset for Bibliographic Research in Computational Linguistics*. Language Resources and Evaluation Conference (LREC) 2008.

UNIVERSITY of ROCHESTER NUS National University of Singapore ETS Educational Testing Service UNIVERSITÄT DES SAARLANDES

- Comments? Questions?
- Ideas for future directions?
- Interested in joining the Anthology team?

Come and visit our **poster**

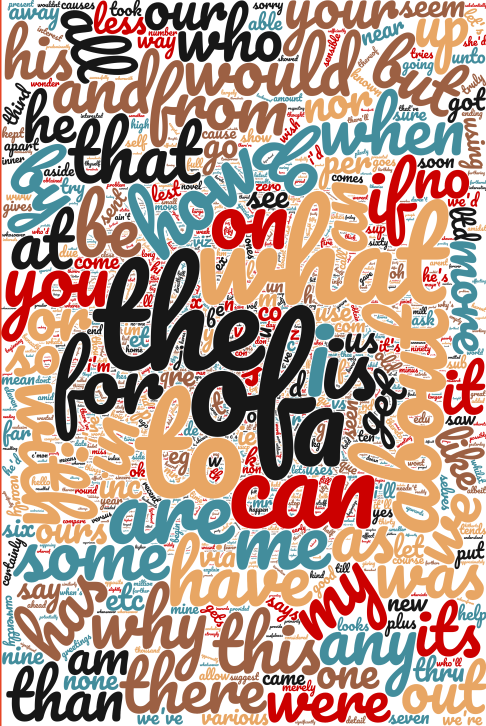
Stop Word Lists in Free Open-source Software Packages

Joel Nothman Hanmin Qin
Roman Yurchak

20 July 2018

scikit learn
machine learning in Python

THE UNIVERSITY OF SYDNEY



In OSS we trust

- Users trust OSS packages to provide good stop word lists
- Maintainers might not have given it much thought
- Lists are adapted from each other
- Lists include surprises and inconsistencies

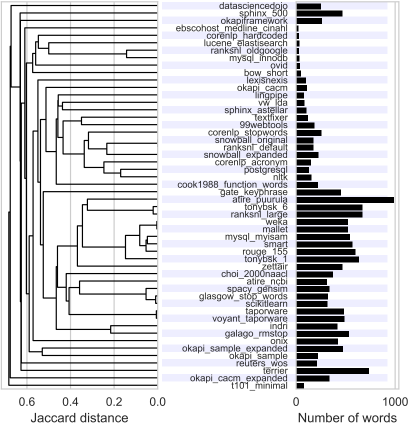
Scikit-learn stop words

- We don't know how our 'english' list was constructed
- but spaCy and Gensim use a similar list

- Has typos: fify corrected to fifty in 2015
- Surprising inclusions: computer (removed 2011); system; cry
- Surprising omissions: seven, does
- Inconsistent with our default tokenizer: ve isn't stopped

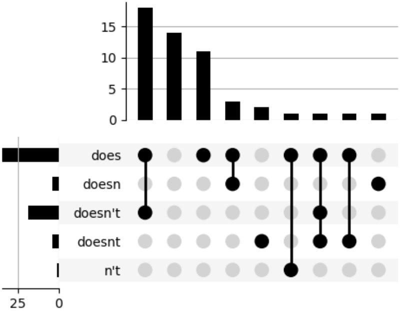
Looking beyond Scikit-learn

- We analyse @igorbrigadir's collection of English stop word lists
- We compare the contents of 52 lists



Looking beyond Scikit-learn

- We analyse @igorbrigadir's collection of English stop word lists
- We compare the contents of 52 lists
- We identify some surprises and inconsistencies



We can improve how we provide stop lists

- Better documentation
- Adapt the list to the NLP pipeline
- Tools for quality control
- Tools for automatic list construction

The risk of sub-optimal use of Open Source NLP Software

UKB is inadvertently state-of-the-art in knowledge-based WSD

Eneko Agirre   Oier López de Lacalle   **Aitor Soroa**

NLP-OSS Workshop, July 2018

IXA NLP group, UPV/EHU

Introduction

- UKB is a collection of programs for WSD
- Graph-based, exploits relations of KB
  - using the Personalized PageRank algorithm
- First released on 2009, attained SOA results
- Free software (GPLv3 license)

Many uses

- Named Entity disambigiation
- Disambiguation of medical entities
- Word similarity
- Create knowledge-based word embeddings

Parameters

- UKB contains many parameters

Parameters

- UKB contains many parameters
    - KB relations
        - Which relations to use
        - Use relation weights

Parameters

- UKB contains many parameters
  - KB relations
    - Which relations to use
    - Use relation weights
  - Dictionary
    - Use sense frequencies

Parameters

- UKB contains many parameters
  - KB relations
    - Which relations to use
    - Use relation weights
  - Dictionary
    - Use sense frequencies
  - Graph algorithms
    - Whole graph: *ppr*, *ppr_w2w*
    - Subgraph: *dfs*, *bfs*
    - Aproximation algorithms: *nibble*
    - Each contains its own hyper-parameters

Parameters

- UKB contains many parameters
    - KB relations
        - Which relations to use
        - Use relation weights
    - Dictionary
        - Use sense frequencies
    - Graph algorithms
        - Whole graph: *ppr*, *ppr_w2w*
        - Subgraph: *dfs*, *bfs*
        - Aproximation algorithms: *nibble*
        - Each contains its own hyper-parameters
    - Input pre-processing
        - Context of at least 20 words

UKB parameters

- Default parameters are sub-optimal
    - they do not obtain best results
- Two main reasons:
    - remain purely unsupervised
    - speed trade-off
- Some authors reported results with the default sub-optimal parameters

	All	S2	S3	S07	S13	S15
UKB (elsewhere)†‡	57.5	60.6	54.1	42.0	59.0	61.2
UKB (this work)	**67.3**	68.8	66.1	53.0	**68.8**	**70.3**

UKB parameters

- Default parameters are sub-optimal
  - they do not obtain best results
- Two main reasons:
  - remain purely unsupervised
  - speed trade-off
- Some authors reported results with the default sub-optimal parameters

	All	S2	S3	S07	S13	S15
UKB (elsewhere)†‡	57.5	60.6	54.1	42.0	59.0	61.2
UKB (this work)	**67.3**	68.8	66.1	53.0	**68.8**	**70.3**
Chaplot and Sakajhutdinov (2018) ‡	66.9	**69.0**	**66.9**	55.6	65.3	69.6
Babelfy (Moro et al., 2014)†	65.5	67.0	63.5	51.6	66.4	70.3
MFS	65.2	66.8	66.2	55.2	63.0	67.8
Basile et al. (2014)†	63.7	63.0	63.7	**56.7**	66.2	64.6
Banerjee and Pedersen (2003)†	48.7	50.6	44.5	32.0	53.6	51.0

Conclusion

- Default parameters are very important
  - extremely important to include precise instructions and optimal default parameters.
- If possible, include end-to-end scripts to automatically reproduce results
- Most recent version (3.0)
  - parameters are now optimal
  - contains scripts for reproducing results on WSD Evaluation Framework (Raganato et al, 2017)
- UKB still SOA among KB methods

Conclusion

Thank you