



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

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

Martín Villalba

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.

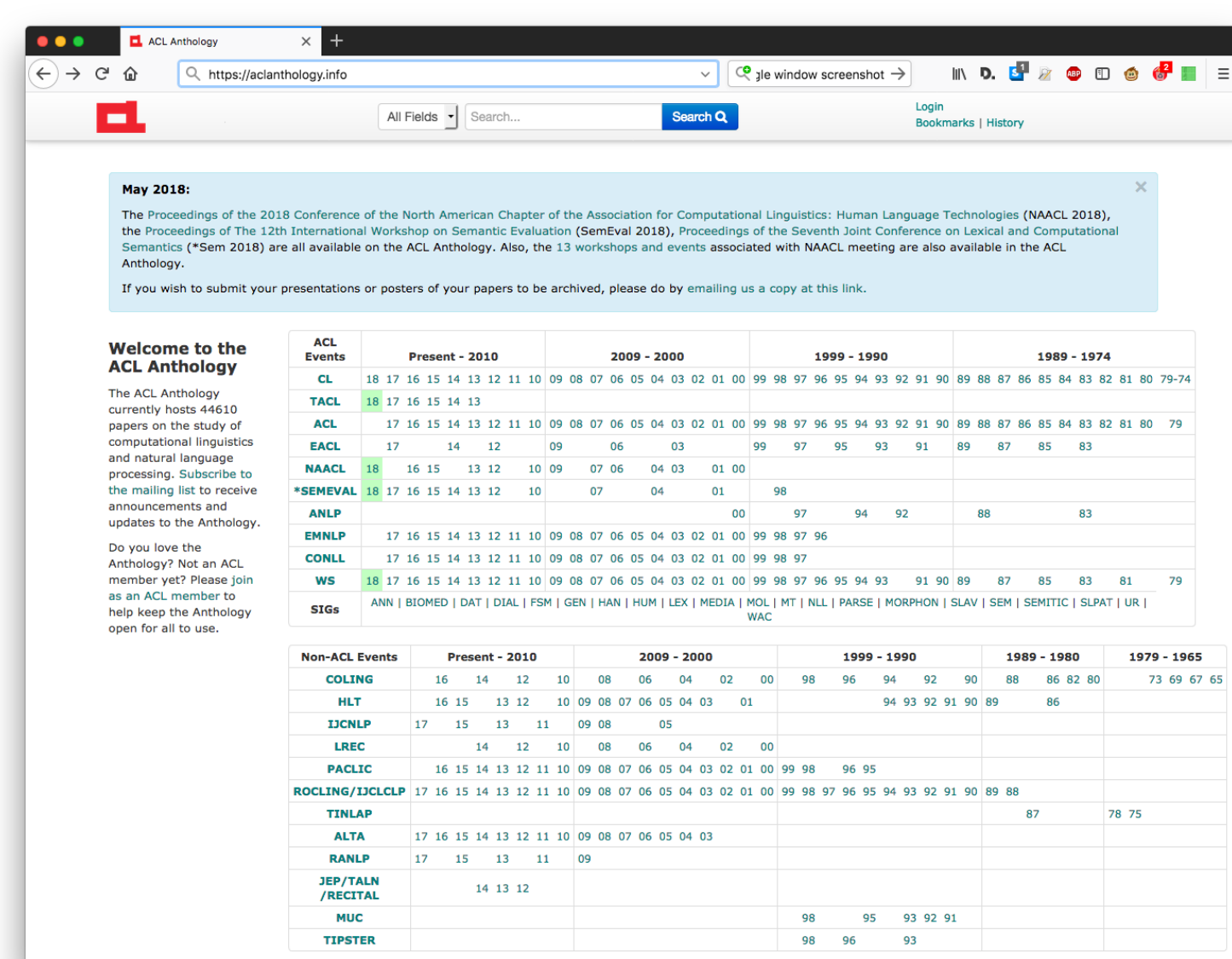
Future proofing the Anthology

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- Update/replace all outdated dependencies.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Upgrade and/or migrate outdated dependencies	Full test coverage and consistency checks	Docker image for releases	Add index support for popular search engines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

Current state of the Anthology



43k
hosted papers

4.5k
daily hits

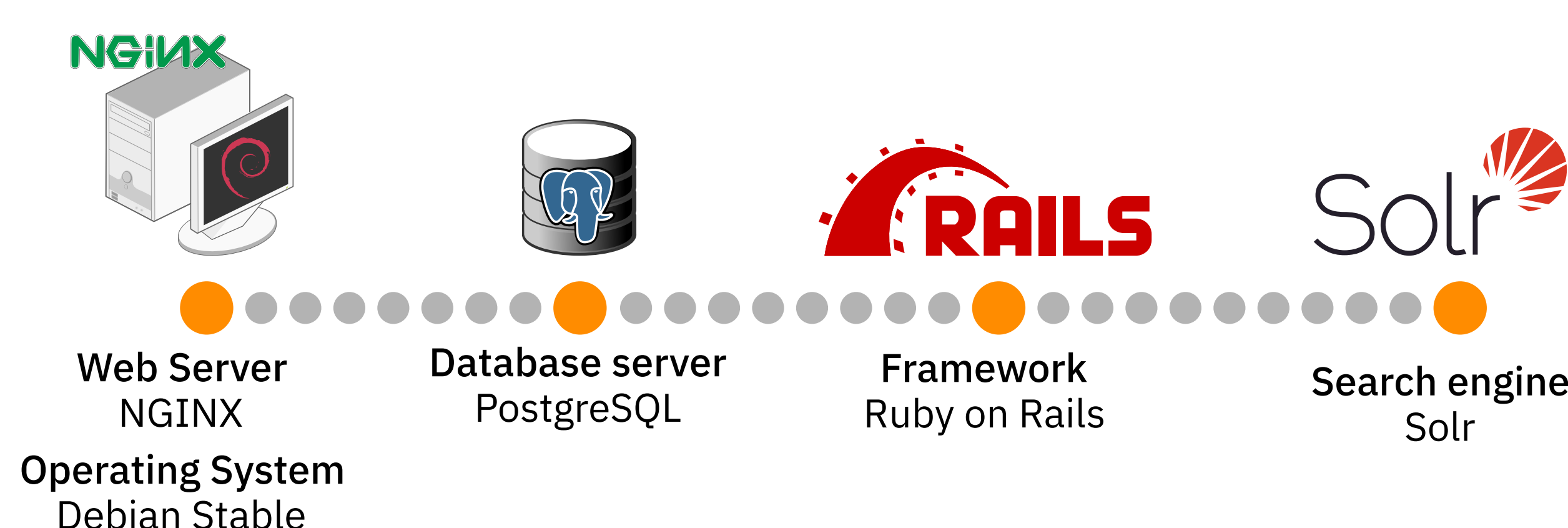
Challenges for the Community

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

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>



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.

