

ACL Anthology Status

Matt Post
Anthology Director

ACL 2019
Florence, Italy

28 July 2019

Problems

For your consideration

Videos

- ACL pays Weyond a lot to have videos recorded
 - Weyond is contractually obligated (per Priscilla) to link videos to Anthology IDs, doesn't
 - Contact very unresponsive
 - Has lost entire conferences

Licensing

- Anthology distributed under the following license
 - 2016+: CC-BY-4.0
 - prior to 2016: CC-BY-3.0 (no commercial reuse)
- Currently no way to mark at the volume level
- Needs formal reconsideration

PAST

Static Rewrite

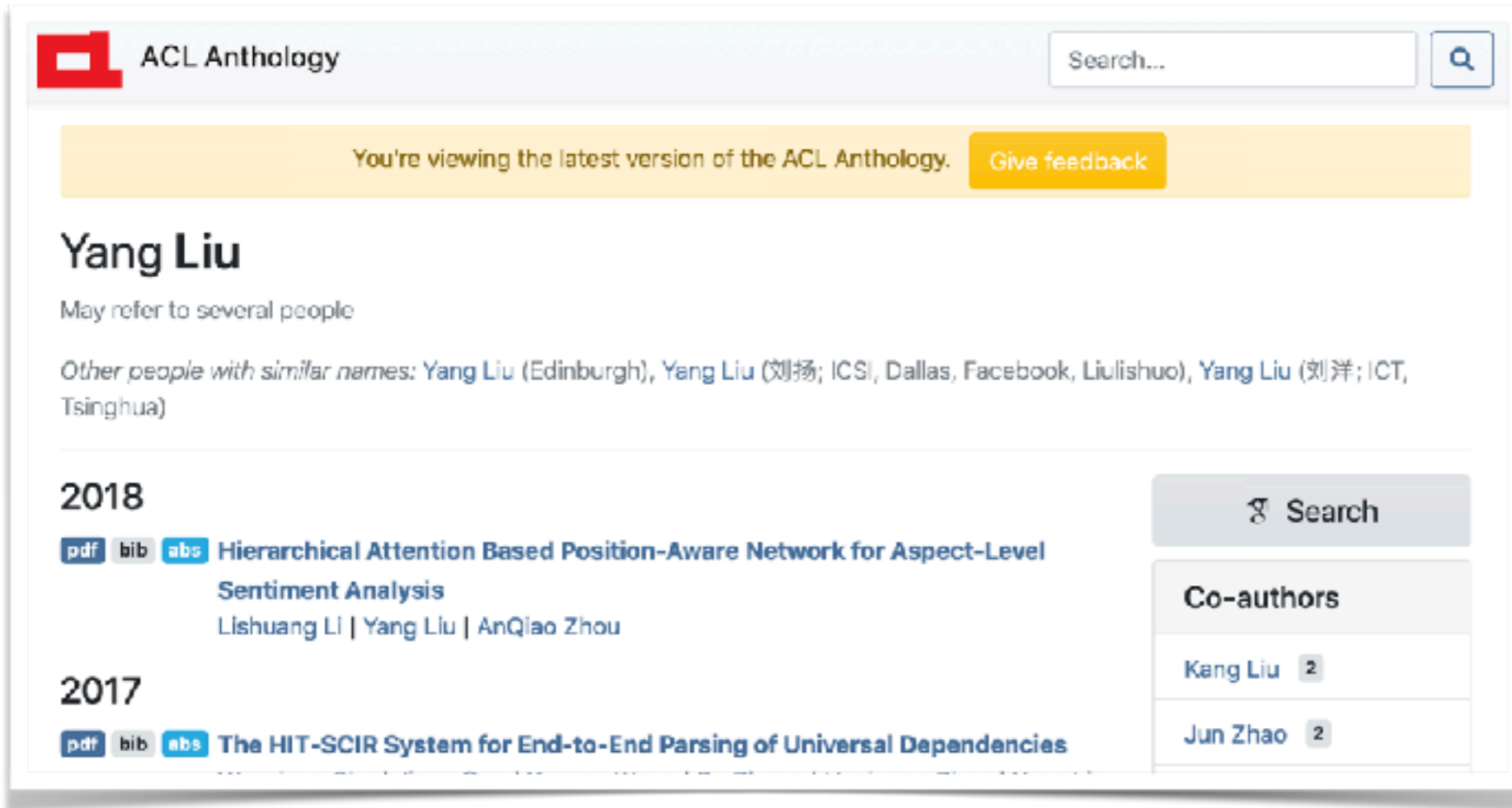
- Moved from dynamic Ruby-on-rails architecture to static site built with Python 3.7 and Hugo
- Much faster, quicker to publish changes
- Also contributed many visual enhancements, fixed Google Scholar issues



Marcel Bollmann
University of Copenhagen

Corrections

- Name consolidation, author disambiguation, restoring diacritics



The screenshot shows the ACL Anthology website interface. At the top, there is a search bar and a notification bar stating "You're viewing the latest version of the ACL Anthology." with a "Give feedback" button. The main content area displays the profile for "Yang Liu", noting that the name "May refer to several people". It lists "Other people with similar names: Yang Liu (Edinburgh), Yang Liu (刘扬; ICSI, Dallas, Facebook, Liulishuo), Yang Liu (刘洋; ICT, Tsinghua)". Below this, there are two sections of publications. The "2018" section shows a paper titled "Hierarchical Attention Based Position-Aware Network for Aspect-Level Sentiment Analysis" by Lishuang Li | Yang Liu | AnQiao Zhou, with links for pdf, bib, and abs. The "2017" section shows a paper titled "The HIT-SCIR System for End-to-End Parsing of Universal Dependencies" with links for pdf, bib, and abs. On the right side, there is a "Search" button and a "Co-authors" section listing "Kang Liu" with a count of 2 and "Jun Zhao" with a count of 2.



David Chiang
Notre Dame

- Also: title case protection, many other data integrity issues

TACL/CL

- Smoothed-out ingestion process
- Got some help from MIT press, their process still a bit cumbersome



Arya McCarthy
Johns Hopkins

Video Linking

- We've been linking videos old and new, thanks to scripts that have been contributed
- Missing: mismatched videos, events not associated with papers (e.g., keynotes)



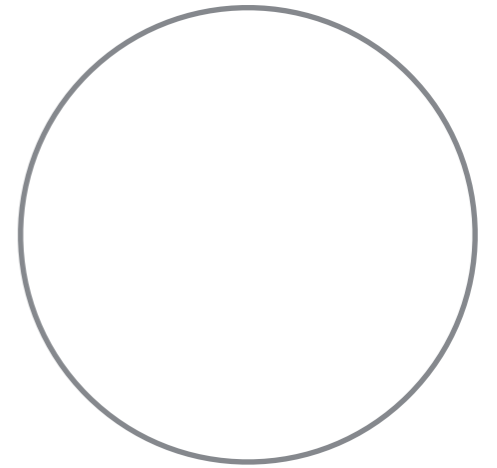
Arne Köhn
Saarland

Present

Ongoing projects

Engineering Village

- Registration with Elsevier is important to many academics in Europe and China
- Currently a time-consuming manual process
- Working on seeing if they'll do the work



Chenliang Li
Wuhan University
www.lichenliang.net/



Aurelie Herbelot
University of Trento

Ingesting old material

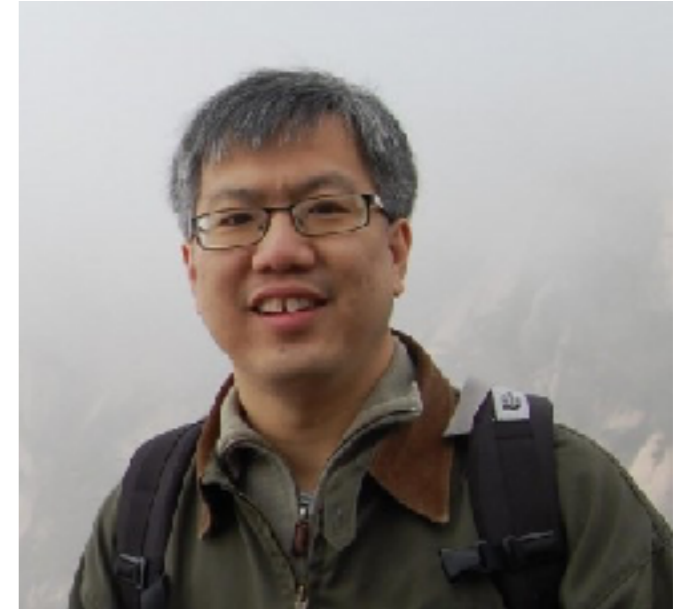
- IWPT: 1989–2011
- Wishlist: mt-archive.info



Kilian Gebhardt
TU Dresden

DOIs

- Have a new ingestion process for DOIs
- Working on adding all ACL papers



Min-Yen Kan
NUS

Portents

Plans for next year

Additional content

- Material with not associated with a paper
 - Keynote talks
 - Conference blog posts
 - NLP Highlights podcast
- Criteria: *citable* content

New ID format

- We're running out of IDs (single letters)
- “W” catchall no longer working
 - supports only 100 papers
 - semantic mismatch
- 2020: new IDs

P18-1001 → acl-2018-1.1

Volume naming

- Want to enforce volume naming format
 - *Proceedings of ACL-08: HLT*
 - *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*
 - *Proceedings of the 57th Conference of the Association for Computational Linguistics*
 - *Proceedings of the ACL 2010 Conference Short Papers*
- Proposed (comment now!)
 - *Proceedings of the 97th Conference of the Association for Computational Linguistics*
 - *Proceedings of the 97th Conference of the Association for Computational Linguistics: Short Papers*

Conflict of Interest

- ACL person information spread out
 - ACL Portal
 - Softconf
 - Anthology
- Could link these together
 - Name matches
 - Asking users for explicit links
 - Other info (ORC ids, START ids)
- Anthology cannot host dynamic content, but author pages are easy to link to