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 TeichmannMartín VillalbaDept. of Language, Science, and Technology
Saarland UniversitySaarland 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



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



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



8 Min-Yen Kan takes the role of Editor.



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



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

Current state of the Anthology

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- Update/replace all outdated dependencies.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.



ଟ 🕜 🗘 🤇 https://aclan	Q https://aclanthology.info								~	ار جاe window screenshot $ ightarrow$								l	N I	D.	5	1	ABP	1	۲	2				
d		All F	ields	·	earch	1					Sea	arch (۹.							Log Boo		rks	Hist	ory						
May 2018:																												×		
The Proceedings of the 201																														
the Proceedings of The 12t Semantics (*Sem 2018) ar																											onal			
Anthology.																														
If you wish to submit your	presentations	s or poste	ers of y	our p	aper	s to b	e ar	chived	, plea	ise do	o by e	emai	ing u	is a (сору	at th	is lin	k.												
Welcome to the ACL Anthology	ACL Events		Present - 2010					2009 - 2000								1999 - 1990					1989 - 1			- 19	974					
The ACL Anthology	CL	18 17	16 15 :	14 13	3 12	11 10	09	08 07	06	05 04	1 03	02 0	1 00	99	98 9	97 96	95 9	94 9	3 92	91	90	89 8	8 87	86	85 8	4 83	82	81 80	79-74	
currently hosts 44610	TACL	18 17	16 15 :	14 13	3																									
papers on the study of	ACL	17	16 15 :	14 13	3 12	11 10	09	08 07	06	05 04	1 03	02 0	1 00	99	98 9	97 96	95 9	94 9	3 92	91	90	89 8	8 87	86	85 8	4 83	82	81 80	79	
computational linguistics and natural language	EACL	17			12		09		06		03			99	9	97	95	9	3	91		89	87	7	85	83				
processing. Subscribe to	NAACL	18	16 15	13	3 12	10	09	07	06	04	1 03	0	1 00																	
the mailing list to receive	*SEMEVAL	18 17	16 15 :	14 13	3 12	10)	07	,	04	1	0	1		98															
announcements and updates to the Anthology.	ANLP												00		9	97	9	94	92			8	8			83				
Do you love the	EMNLP	17	16 15 :	14 13	3 12	11 10	09	08 07	06	05 04	1 03	02 0	1 00	99	98 9	97 96														
Anthology? Not an ACL	CONLL	17	16 15 :	14 13	3 12	11 10	09	08 07	06	05 04	1 03	02 0	1 00	99	98 9	97														
member yet? Please join	ws	18 17	16 15 :	14 13	3 12	11 10	09	08 07	06	05 04	1 03	02 0	1 00	99	98 9	97 96	95 9	94 9	3	91	90	89	87	7	85	83	1	81	79	
as an ACL member to help keep the Anthology open for all to use.	SIGs	ANN E	BIOMED	DA	r DI	AL F	SM	GEN	HAN I	ним	LEX	MEI		MOL WAC		NLL	PAR	SE	MOR	PHO	N S	LAV	SEN	1 SE	MITIC	C SLF	PAT	UR		
	Non-ACL	Present - 2010					2009 - 200					D			1999 - 199				0			1989 - 1980				197	9 - 19	55		
	COLI	NG	16		14	12	1	0	08	06	04	ļ.	02	00)	98	96		94	92	2	90	1	88	86	82 80	D	7	3 69	67 6
	HLT	r	16	15	13	3 12	1	0 09	08 07	06	05 04	03	0	1					94 9	3 92	2 91	90	89		86					
	IJCN	LP	17	15	13	3	11	09	08	(05																			
	LRE	с			14	12	1	0	08	06	04	ļ.	02	00)															
	PACL	IC	16	15	14 13	12	11 1	0 09	08 07	06	05 04	03	02 0	1 00	99	98	96	95												
	ROCLING/	IJCLCLP	17 16	15	14 13	12	11 1	0 09	08 07	06	05 04	03	02 0	1 00	99	98 9	7 96	95	94 9	3 92	2 91	90	89	88						
	TINL	AP																						8	7		78	75		
	ALT	A	17 16	15	14 13	3 12	11 1	0 09	08 07	06	05 04	03																		
	RANI	LP	17	15	13	3	11	09																						
	JEP/T/				14 13	12																								
	/RECI															98		95		3 92	0.01									
	MUG	-														90		32	9	3 92	- 91									



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

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.

