

Migrating ACL Anthology Network / "All About NLP"

Presented by Jay Hou and Clark Xie. Additional group members Dan Keller and Andre Monteiro. Department of Computer Science, Yale School of Engineering and Applied Sciences, New Haven, CT

Introduction

- ACL Anthology Network / "All About NLP"
- Collection of NLP (Natural Language Processing) papers, tutorials, and resources
- More than 21k papers and 17k authors
- Hosted on http://tangra.cs.yale.edu/index.php

Motivation and Goals

- 1. For learning:
- Easily find resources on a given topic
- NLP topics and related topics taxonomized
- Lectures and tutorials in addition to papers
- 2. For research:
- Find related publications and leading papers
- Find authors and colleagues along with the number of papers published, number of times cited, h-indexes, and more
- Visualize credibility of papers and authors

Technical work

- Moving the codebase and database from University of Michigan to Yale on tangra
- Updating from php5 to php7
- Fixing bugs and manual QAing to ensure consistency between sites
- Updating the database by uploading papers from recent conferences through admin site
- Integrating PDF box to perform automatic text extraction and reference building
- Developing a tool for manually adding PDFs and online resources

Figure 1: The index page of the website

ACL Anthology Network (All About NLP) (beta) Home Venues Authors Networks Paper Title Paper Fulltext 🔍 institution 🔍 Author All 💿 Topics (WIP) Updates About Contact Us Welcom ${f W}$ elcome to the University of Michigan CLAIR Group's ACL Anthology Network (All About NLP) interface. Here we have collected information regarding all of the papers included in the many ACL venues. From those papers, we have created several networks, including paper itation, author citation, and author collaboration. Click here to learn more about the project. The network is currently built using ACL papers. This includes all papers up to and including those published in June 2011 which were successfully processed. Feel free to visit The ACL Anthology to browse what is available. Latest Updates 2013 Release: We have added 1,565 new papers to the anthology. The new papers are from the following venues: EMNLP 2013, ACL 2013, NAACL 2013, *SEM 2013, SIGDIAL 2013, CL Volume 39, TACL Volume 1 and all the ACL and NAACL workshops from 201 February 2013 2012 Release: We have added 1,360 new papers to the anthology. The new papers are from the ollowing venues: EMNLP 2012, ACL 2012, NAACL 2012, EACL 2012, SemEval 2012, CL Volume 38, and

Figure 4. Ability to list and sort authors ACL Anthology Network (All About NLP) (beta)

The Association Of Computational Linguistics Anthology Network

0

Venues Authors						Search:		
Networks	Name	♦ #Papers ♦	#Collaborators	♦ #Venues ♦	#Affiliations	#Citations -	h-index	\$
Topics (WIP)	Franz Josef Och	39	58	11	5	3650	20	
Updates	Christopher D. Manning	134	116	30	4	2975	27	
About	Philipp Koehn	77	114	16	4	2930	19	
Contact Us	Dan Klein	97	41	12	3	2659	25	
	Hermann Ney	114	115	21	3	2650	20	\$
	Michael John Collins	48	51	10	4	2255	22	
	Daniel Marcu	57	47	16	5	1959	21	
	Kevin Knight	87	97	21	4	1892	24	•
	Chris Callison-Burch	69	103	19	3	1881	18	
	Vincent J. Della Pietra	14	17	5	3	1852	9	
	Mitchell P. Marcus	38	52	13	5	1699	12	
	Fernando Pereira	45	49	13	9	1671	17	
	Robert L. Mercer	20	29	5	2	1658	13	
	Salim Roukos	35	62	6	3	1607	13	
	Daniel Jurafsky	78	81	18	4	1585	21	

Future Features

- Single document citation-based summary
- Build paper prerequisite chains
- Personalized learning progress tracking and recommendation
- Identify hot, outdated, and developing topics for research

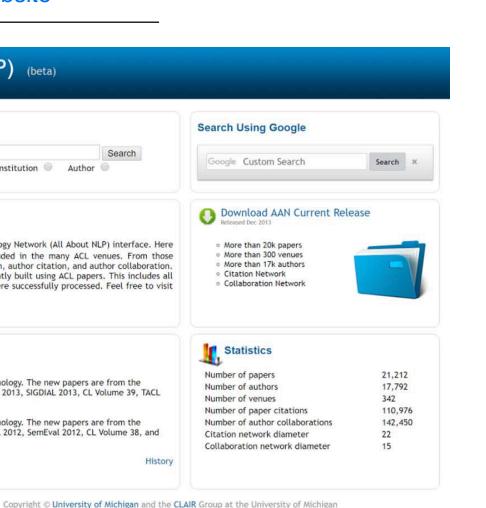


Figure 2. Tools for learning

ubtopics 5										
	Subtopics	AAN Papers	Non-AAN Papers	Tutorials	Lectures	Assignments C	orpora Libr	aries Surve	ys NACLO Pro	olems
Subtopics										
Id #							Name			
41						Que	estion Answerin	g		
42						s	ummarization			
43						Т	ext Generation			
44						Disco	urse and Dialo	jue		
45						Ma	chine Translatio	n		
		Topic 42: Su	Immarization							
		Topic 42: Su #Subtopics	ummarization 6							
			6	Subtopics	AAN Papers	Non-AAN Papers	Tutorials	Lectures	Assignments	Cor
			6	Subtopics	AAN Papers	Non-AAN Papers	s Tutorials	Lectures	Assignments	Cor
		#Subtopics	6	Subtopics	AAN Papers	Non-AAN Papers	Tutorials		Assignments Date	Cor
		#Subtopics	6 Sum	Title		Non-AAN Papers	Lecture Dragomir Ra	r dev		Cor
		#Subtopics	6 Sum	Title marization tion Techniq	ues 1	Non-AAN Papers	Lecture Dragomir Ra Dragomir Ra	r idev idev	Date 2016 2016	Cor
		#Subtopics	6 Sum	Title marization tion Techniq	ues 1 ues 2	Non-AAN Papers	Lecture Dragomir Ra	r Idev Idev Idev	Date 2016	Corp

Return to Parent Topic

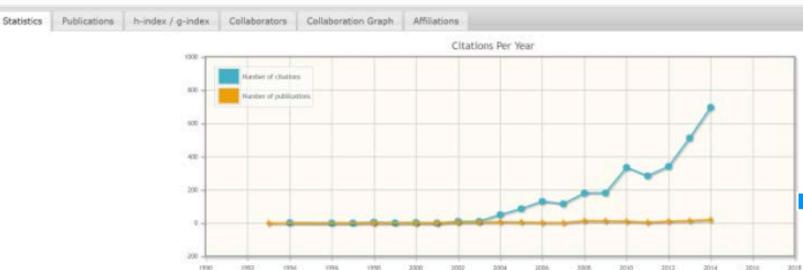
Copyright © University of Michigan and the CLAIR Group at the University of Michig

Dragomir Rade

Figure 3. Tools for researching

Sentence Simplification

with 11 ns 213 ons 110 tions •	6 Co-author(s) from 19 83 Citation(s) 6 Citation(s) Stanford University, Si Stanford University, Si	193 to 2014 tanford CA tanford CA; Palo Al	to Research Center, Pa	alo Alto CA		
Publications	h-index / g-index	Collaborators	Collaboration Graph	Affiliations		
<u>.</u>	Title		¢	Authors	0	#Citation
		ub Categorization	Christopher D.	Manning		54
The Segmentation Problem In Morphology Learning			Christopher D. Manning			2
Enriching The Knowledge Sources Used In A Maximum Entropy Part-Of-Speech Tagger			Kristina Toutan	Kristina Toutanova Christopher D. Manning		
Parsing With Treebank Grammars: Empirical Bounds Theoretical Models And The Structure Of The Penn Treebank			Dan Klein Chr	ristopher D. Manning		7
ons ted with ations itations	131 Paper(s) in 116 Co-author 2183 Citation(116 Citation(s)	1 31 venue(s) (s) from 1993 to s)				
	with 110 ns 211 ons 110 tions • Publications Publications Automatic Ar Dictionary Fr The Segmen Enriching Th Entropy Part Parsing With Theoretical N Treebank	with 116 Co-author(s) from 15 ns 2183 Citation(s) ons 116 Citation(s) itions 116 Citation(s) is Stanford University, Si is Stanford University, Si is University of Sydney, Si Publications h-index / g-index Title Automatic Acquisition Of A Large Si Dictionary From Corpora The Segmentation Problem In Morp Enriching The Knowledge Sources O Entropy Part-Of-Speech Tagger Parsing With Treebank Grammars: Theoretical Models And The Structu Treebank Christopher D. Manning ons 131 Paper(s) in ted with 116 Co-authori ations 2183 Citation(s)	ns 2183 Citation(s) ons 116 Citation(s) tions 116 Citation(s) • Stanford University, Stanford CA • Stanford University, Stanford CA • University of Sydney, Sydney Australia Publications h-index / g-index Collaborators Publications h-index / g-index Collaborators Title Automatic Acquisition Of A Large Sub Categorization Dictionary From Corpore The Segmentation Problem In Morphology Learning Enriching The Knowledge Sources Used In A Maximum Entropy Part-Of-Speech Tagger Parsing With Treebank Grammars: Empirical Bounds Theoretical Models And The Structure Of The Penn Treebank Christopher D. Manning ons 131 Paper(s) in 31 venue(s) ted with 116 Co-author(s) from 1993 to ations 2183 Citation(s) Itations 116 Citation(s)	with 116 Co-author(s) from 1993 to 2014 ns 2183 Citation(s) ons 116 Citation(s) itions - Stanford University, Stanford CA - Stanford University, Stanford CA; Palo Alto Research Center, Ri - University of Sydney, Sydney Australia Publications h-index / g-index Collaborators Collaboration Graph Title Automatic Acquisition Of A Large Sub Categorization Christopher D. Dictionary From Corpora Christopher D. The Segmentation Problem In Morphology Learning Christopher D. Enriching The Knowledge Sources Used In A Maximum Kristina Toutan Entriching The Knowledge Sources Used In A Maximum Kristina Toutan Parsing With Treebank Grammars: Empirical Bounds Dan Klein Christopher D. Parsing With Treebank Grammars: Empirical Bounds Dan Klein Christopher D. Genes 131 Paper(s) in 31 venue(s) Dan Klein Christopher D. ted with 116 Co-author(s) from 1993 to 2014 ations 2183 Citation(s) 116 Citation(s) Entions	with 116 Co-author(s) from 1993 to 2014 ns 2183 Citation(s) ons 116 Citation(s) etions Stanford University, Stanford CA Stanford University, Stanford CA; Palo Alto Research Center, Palo Alto CA • University of Sydney, Sydney Australia Publications h-index / g-index Collaborators Collaboration Graph Affiliations Publications h-index / g-index Collaborators Collaborator Graph Affiliations Publications h-index / g-index Collaborators Collaborator Graph Affiliations Publications h-index / g-index Collaborators Collaborator Graph Authors Automatic Acquisition Of A Large Sub Categorization Christopher D. Manning Christopher D. Manning Enric	with 116 Co-author(s) from 1993 to 2014 Ins 2183 Citation(s) Ins 116 Citation(s) Ins Stanford University, Stanford CA Stanford University, Stanford CA; Palo Alto Research Center, Palo Alto CA University of Sydney, Sydney Australia Publications h-index / g-index Collaborators Collaboration Graph Affiliations Title Automatic Acquisition Of A Large Sub Categorization Dictionary From Corpora Christopher D. Manning The Segmentation Problem In Morphology Learning Christopher D. Manning Enriching The Knowledge Sources Used In A Maximum Entropy Part-Of-Speech Tagger Parsing With Treebank Grammars: Empirical Bounds Theoretical Models And The Structure Of The Penn Treebank Christopher D. Manning Dan Klein Christopher D. Manning Dan Klein Christopher D. Manning The Structure Of The Penn Treebank Christopher D. Manning Interstopher D. Interstopher D. Inter





Libraries Surveys NACLO Problems

p://clair.eecs.umich.edu/aan/595/421.p tp://clair.eecs.umich.edu/aan/595/426.pdf

Find Author On Google Scholar

	Search:				
Venue	٥	Session	0		
ACL		Main Conference			
Workshop On New Meth		2			
2000 Joint SIGDAT Co		2			
ACL		Main Conference			