

Tae Yano	Graduate Research Assistant Language Technologies Institute School of Computer Science Carnegie Mellon University, Pittsburgh, PA	e-mail: taey [at] cs.cmu.edu http://www.cs.cmu.edu/~taey
Research Interests	Natural language processing, statistical text analysis, probabilistic topic modeling, and computational social science.	
Education	Carnegie Mellon University Language Technologies Institute, School of Computer Science Ph.D candidate in Language and Information Technologies Concentration: natural language processing	Pittsburgh, Pennsylvania Sept 2007 - Present
	Columbia University The Fu Foundation School of Engineering and Applied Science M.S. in Computer Science Concentration: natural language processing Honors: Master of Science Teaching Assistance Fellowship	New York, New York Sept 2005 - May 2007
	Hunter College of the City University of New York B.A. in Computer Science with honor Honors: <i>magna cum laude</i> , Dean's honoree	New York, New York Sept 1998 - Jan 2001
	Lewis and Clark College B.A. in International Affairs Honors: Dean's honoree, international student grant recipient	Portland, Oregon Sept 1994 - Aug 1997
Professional Experience	Research Intern Natural Language Processing Group Microsoft Research Redmond Supervisors: Michael Gamon, Patrick Pantel	Redmond, Washington May 2012 - Aug 2012
	Research Programmer Center for Computational Learning Systems Columbia University Python Programming for Electric Power Grid Con Edison Challenge Project	New York, New York Jun 2007 - Aug 2007
	System Engineer Software Engineering Group Ricoh Corporation Main duties included: - Design and development of large-scale embedded application suits in C/C++ for UNIX based document management devices. - Development and support of proprietary API for the device development. - Education and training of 3rd party software developers	West Caldwell, New Jersey Mar 2001- Sept 2005
	System Administrator Department of Computer Science Hunter College	New York, New York Oct 2000 - Mar 2001

Professional Experience (continued)	Oracle Application Programming Intern Instructional Computing and Information Technology (ICIT) Hunter College	New York, New York Sept 2000 - Dec 2000
	UNIX System Administrator/Programmer Instructional Computing and Information Technology (ICIT) Hunter College	New York, New York Feb 1999 - Jan 2001
Research and Teaching Experience	Graduate Research Assistant Language Technologies Institute School of Computer Science, Carnegie Mellon University	Pittsburgh, Pennsylvania Sept 2007 - Present
	Advisors: Noah A. Smith, William Cohen	
	Project: - Statistical NLP and text analysis in political blogs and tweets - Congressional Roll Call voting prediction from texts - Modeling and tracing of dynamic voting process	
	Teaching Assistant Language Technologies Institute School of Computer Science, Carnegie Mellon University	Pittsburgh, Pennsylvania Sept 2009 - May 2010
	Courses taught: - Text Driven Forecasting A graduate level course investigating the statistical techniques in text analysis and their use in predicting real world event. - Language and Statistics II A graduate level statistical NLP course. - Advanced NLP Seminar	
	Supervised Individual Research Project Department Research Assistant (DRA) Computer Science Department School of Engineering and Applied Science, Columbia University	New York, New York Sept 2006 - May 2007 May 2006 - Aug 2006
	Advisor: Rebecca J. Passonneau	
	Projects: - CLIMB 2 - Computational Linguistics for Metadata Building. (http://www.umiacs.umd.edu/~climb/) - AntArt (ANaTomy of ARt Text) - Interface for annotating texts about art historical images with functional semantic categories)	
	Master of Science Teaching Assistant Fellow Computer Science Department School of Engineering and Applied Science, Columbia University	New York, New York Jan 2005 - May 2007
	The Master of Science Teaching Assistant (MSTA) Fellowship is a department funded fellowship awarded to distinguished masters students. Recipients are compensated by stipend and tuition exemption.	
	Courses Taught: - Advanced Programming (3 semesters) - Data Structure in C/C++ (2 semesters) - Programming Languages and Translator (2 semesters) - Discrete Mathematics (1 semester)	

Publications

Tae Yano, John Wilkerson and Noah A. Smith

Textual Predictors of Bill Survival in Congressional Committees

In Proceedings of the North American Association for Computational Linguistics Human Language Technologies Conference (NAACL). Montreal, Quebec. June 2012

Jacob Eisenstein, Tae Yano, William W. Cohen, Noah A. Smith, and Eric Xing

Bayesian Nonparametrics for Information Extraction

In Proceedings of the EMNLP Workshop on Unsupervised Learning in NLP. Edinburgh, UK, July 2011

Justin Cranshaw and Tae Yano.

Seeing a Home away from the Home: Distilling proto-Neighborhood from Incidental data with Topic Modeling

In Proceedings of the Workshop on Computational Social Science and the Wisdom of Crowds, Annual Conference on Neural Information Processing System (NIPS), Vancouver, B.C., Canada. Dec 2010

Tae Yano, Philip Resnik, and Noah A. Smith.

Shedding (a Thousand Points of) Light on Biased Language

In Proceedings of the NAACL-HLT Workshop on Creating Speech and Language Data With Mechanical Turk. Los Angeles, CA. June 2010

Tae Yano and Noah A. Smith.

What's Worthy of Comment? Content and Comment Volume in Political Blogs with Topic Models

In Proceedings of the International AAAI Conference on Weblogs and Social Media (ICWSM). Washington, D.C., May 2010

Tae Yano, William W. Cohen, and Noah A. Smith.

Predicting Response to Political Blog Posts with Topic Models

In Proceedings of the North American Association for Computational Linguistics Human Language Technologies Conference (NAACL). Boulder, CO. May/June 2009

Rebecca Passonneau, Tom Lippincott, Tae Yano, and Judith Klavans.

Relation between Agreement Measures on Human Labeling and Machine Learning Performance: Results from an Art History Domain

In Proceedings of the International Conference on Language Resources and Evaluation (LREC). Marrakesh, Morocco. May/Jun 2008

Judith Klavans, Carolyn Sheffield, Eileen Abels, Joan Beaudoin, Laura Jenemann, Jimmy Lin, Tom Lippincott, Rebecca Passonneau, Tandeep Sidhu, Dagobert Soergel, and Tae Yano.

Computational Linguistics for Metadata Building: Aggregating Text Processing Technologies for Enhanced Image Access

In proceedings of the LREC Workshop on Language Resources for Content-Based Image Retrieval (OntoImage 2008). Marrakesh, Morocco. May/Jun 2008

Rebecca Passonneau, Tae Yano, and Judith Klavans

Functional Semantic Categories for Art History Text: Human Labeling and Preliminary Machine Learning

In Proceedings of the Workshop on Metadata Mining for Image Understanding, International Conference on Computer Vision Theory and Applications (VISAPP). Funchal, Portugal. Jan 2008

Academic Projects	<i>Polling Made Easy: Aggregating voter sentiment in reader comments with topic modeling</i> Term project for Information Retrieval Lab Language Technologies Institute, Carnegie Mellon University	Pittsburgh, Pennsylvania Spring 2008
	<i>Taking Advantage of Wikipedia: A literature review of the NLP research on Wikipedia</i> Term project for Language and Statistics II Language Technologies Institute, Carnegie Mellon University	Pittsburgh, Pennsylvania Fall 2008
	<i>Experiments on Non-Topical Paragraph Classification of Art History Textbooks</i> Term project for Search Engine Technology Columbia University	New York, New York Fall 2006
	<i>KP: A knitting language</i> Term project for Programming Languages and Translator Columbia University	New York, New York Fall 2004
Presentations and other participations	CAOSS 2012: Workshop on Computational and Online Social Science Participant Columbia University	New York, New York Oct 12, 2012
	New Directions in Analyzing Text as Data (2012) Presentation: Textual Predictors of Bill Survival in Congressional Committees Harvard University, Institute for Quantitative Social Science	Cambridge, Massachusetts Oct 05 - 06, 2012
	Social Computational Systems (SoCS) Ph.D Symposium Presentation: Predicting Voter Behavior from the Observables University of Minnesota, Twin Cities	Minneapolis, Minnesota Jun10 - 12, 2011
	Text-as-Data Conference 2011 Presentation: Observing Motivated Reasoners and Affective Tipping Point in Political Decision Making Northwestern University, Kellogg School of Management	Evanston, Illinois Mar 10 - 12, 2011
	Text-as-Data Conference 2010 Participant Northwestern University, Kellogg School of Management	Evanston, Illinois Mar 08 - 10, 2010
	2010 Google Grad CS Forum Participant Google. Inc	Mountain View, California Jan 21 - 23, 2010
	IBM Statistical Machine Learning and Its Application (SMiLe) Presentation: <i>Modeling Political Blogs with Response</i> IBM Thomas J. Watson Research Center	Yorktown Heights, New York Oct 08 - 09, 2009
	LTI Student Research Symposium Poster presentation <i>Modeling Political Blogs with Response</i> Language Technologies Institute, Carnegie Mellon University	Pittsburgh, Pennsylvania Sept 24, 2009

Recent Extracurricular Activities	PhD Pipeline Panel Discussion Invited participant Computer Science Department Retreat Hunter College	New York, New York Dec 2007
--	--	--------------------------------

Fashion Project Magazine Contributing Editor The current issue can be found at http://fashionprojects.org/	New York, New York Sept 2004 - Aug 2007
---	--

Computer Skills

- C and Python: extensive, hands on experience with system level application programming.
- Solid understanding of OOP concepts with working knowledge in C++ and Java.
- Other programming experience: LISP, PL/SQL, Matlab, R, Perl, various shell scripting, assembly (8086).

Languages Native Japanese, Fluent English, College Level Spanish and French