William Yang Wang

Carnegie Mellon University
Language Technologies Institute
School of Computer Science
5000 Forbes Ave.
Pittsburgh, PA 15213

Phone: (412) 370-8296
Fax: (412) 268-6298
Email: yww@cs.cmu.edu
Homepage: http://www.cs.cmu.edu/~yww/

Education

- **Ph.D., Computer Science - Language Technologies**
  Carnegie Mellon University, 2011 - 2016
  Advisor: William W. Cohen (Machine Learning Department).
  Additional thesis committee members: Tom Mitchell, Christos Faloutsos, Eric Horvitz.

- **M.S., Computer Science - Natural Language Processing**
  Columbia University, 2009 - 2011
  Advisors: Kathleen McKeown and Julia Hirschberg.

- **B.Eng. Summa Cum Laude in Computer Science**
  School of Information Engineering, Shenzhen University, 2005 - 2009.

Research Interests

- Machine Learning - Statistical Relational Learning:
  [IJCAI 2016] [MLJ 2015] [IJCAI 2015] [ACL-IJCNLP 2015a] [CIKM 2014] [StarAI 2014] [CIKM 2013]

- Computational Social Science:
  [EMNLP 2015] [NAACL 2015a] [ACL 2014] [EMNLP 2013a] [CSL 2013] [ACL 2012]
  [SIGDIAL 2012] [INTERSPEECH 2011] [SIGDIAL 2011] [COLING 2010]

- Natural Language Processing:
  [NAACL 2016] [EMNLP 2014] [EMNLP 2013b] [SLT 2012] [IJCNLP 2011]

- Speech & Spoken Language Understanding:
  [ACL-IJCNLP 2015b] [NAACL 2015b] [INTERSPEECH 2015] [SLT 2014] [ASRU 2013]
  [ICASSP 2013] [ASRU 2011] [FLAIRS 2011]

Research Experience

- **Carnegie Mellon University**, Pittsburgh, PA 01/2013 – 05/2016.
  Research Fellow, Advisor: William W. Cohen
  I investigated scalable learning and inference algorithms [CIKM 2013] for large knowledge graphs [MLJ 2015] [StarAI 2014], as well as structure learning [CIKM 2014] [IJCAI 2016], a soft version of predicate invention [IJCAI 2015] for statistical relational learning, natural language parsing [EMNLP 2014], and joint information extraction and reasoning [ACL 2015a].

- **Yahoo! Labs**, New York, NY 05/2015 – 08/2015.
  Research Scientist Intern, Mentors: Amanda Stent, Drago Radev, and Yashar Mehdad.
  I worked with the NLP group on low-rank approximation techniques for scalable NLP [NAACL HLT 2016].

- **Microsoft Research**, Redmond, WA 05/2012 – 08/2012 and 10/2012 – 01/2013.
  Research Intern, Mentors: Dan Bohus, Ece Kamar, and Eric Horvitz.
  I worked on an open-world learning problem with the adaptive systems team, as well as a crowd-sourcing approach to natural language elicitation [SLT 2012].

- **Carnegie Mellon University**, Pittsburgh, PA 09/2011 – 05/2012.
  R. K. Mellon Presidential Fellow, Advisors: Justine Cassell and Alan W Black.
  I worked on the problem of rapport building and automatic prediction of impoliteness and positivity in peer tutoring dialogues [SIGDIAL 2012].

- **Columbia University**, New York, NY 09/2009 – 08/2011.

  - *Research Assistant*, Natural Language Processing Group, Advisor: Kathleen McKeown. 09/2010 – 05/2011 I worked with Kathy on question answering [IJCNLP 2011], DARPA GALE evaluation, and automatic vandalism detection [COLING 2010] on Wikipedia.

  - *Research Assistant*, Spoken Language Processing Group, Advisor: Julia Hirschberg. 09/2009 – 05/2011 I worked with Julia on speaker state detection [CSL 2012] [SIGDIAL 2011] [INTERSPEECH 2011], and automatic dialect identification.

  - *Visiting Research Assistant*, Biomedical Informatics, Host: Noemie Elhadad. 05/2011 – 07/2011 We studied sentiment classification techniques on 15 fine-grained emotion types in real suicide notes.

  - *Visiting Research Assistant*, School of International and Public Affairs, Host: Suresh Naidu. 07/2011 – 08/2011
    We studied Bayesian generative models to analyze the legal opinions related to slaves from 18-19 century U.S. State Supreme Courts [ACL 2012].

- Institute for Creative Technology, **University of Southern California**, Playa Vista, CA 05/2010 – 08/2010.
  Visiting Research Assistant, Hosts: Ron Artstein, Kallirroi Georgila, Anton Leuski, and David Traum.
  We explored the use of phonetic mixture models [Invited Book Chapter 2011] [FLAIRS 2011] for natural language understanding, and the problem of automatic detection of unnatural segments in unit-speech synthesis [ASRU 2011].

- **Chinese Academy of Sciences/Chinese University of Hong Kong**, China 11/2008 – 08/2009.
  Visiting Research Student, Host: Lan Wang.
  We investigated several data-driven and knowledge-based approaches to letter-to-sound rules generation [CISE 2010].

Best Paper Awards or Nominations

- **Notable Data Set Award, EMNLP 2015** (awarded to 2 out of 1300 submissions)

- **Best Paper Honorable Mention Award, CIKM 2013** (awarded to 3 out of 848 submissions)

- **Best Student Paper Award, ASRU 2013** (awarded to 1 out of 172 submissions)

- **Best Paper Award finalist, FLAIRS 2011** (awarded to 6 out of 179 submissions)

Grants and Fellowships

- Facebook Fellowship Finalist, 2014-2015 and 2015-2016
  (awarded to top ~20 out of hundreds of applicants globally)

- The CIKM Student Travel Grant from SIGIR, SIGWEB, and Facebook, 2013

- The Student Volunteer Award from EMNLP, 2013

- The Donald and Betty Walker Student Scholarship from ACL (1 of 5 awardees worldwide), 2012

- Richard King Mellon Presidential Fellowship, Carnegie Mellon University, 2011 - 2012

- Research Fellowship, Language Technologies Institute, Carnegie Mellon University, 2011

- Computer Science Chair's Distinguished Fellowship, Columbia University (declined), 2011

- International Speech Communication Association (ISCA) Grant, 2011

- Graduate Research Assistantship with full tuition waiver and monthly stipend, Columbia University, 2010-2011

- Second Prize, the First Annual Spoken Dialog Systems Demo Competition, Columbia University, 2010

- Champion, Microsoft Elites Competition, Microsoft Shenzhen Technology Center, 2008

- University Scholarship, Top Class Honor, 2006-2007, 2007-2008

- Namtai Fellowship (Top 5%), 2006-2007, 2007-2008

Honors

- Best Reviewer Award, NAACL 2015 (awarded to a small fraction of reviewers)

- Best Presentation Runner-Up, LTI Student Research Symposium, Carnegie Mellon University, 2011

- Outlook Award, China Central Television "Star of Outlook" National Final, Beijing (Rank No. 1 in Guangdong Province), 2007

- Shenzhen University Innovative Credits Award (6+3 Credits), 2006, 2007

Peer-Reviewed Conference, Journal Papers, and Book Chapters

1. **William Yang Wang**, and William W. Cohen, "Learning First-Order Logic Embeddings via Graph Rewriting", to appear in Proceedin gs of the 25th International Joint Conference on Artificial Intelligence (**IJCAI 2016**), full paper for oral presentation, New York, NY, U.S.A., July 9-15, IJCAI.

2. **William Yang Wang**, Yashar Mehdad, Dragomir Radev, and Amanda Stent, "A Low-Rank Approximation Approach to Learning Joint Embeddings of News Stories and Images for Timeline Summarization", to appear in Proceedings of the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (**NAACL HLT 2016**), long paper, San Diego, CA, U.S.A., June 13-15, ACL.

3. **William Yang Wang** and Diyi Yang, "That's So Annoying!!!: A Lexical and Frame-Semantic Embedding Based Data Augmentation Approach to Automatic Categorization of Annoying Behaviors using #petpeeve Tweets", to appear in Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (**EMNLP 2015**), short paper, Lisbon, Portugal, Sept 19-21, ACL. (EMNLP 2015 Inaugural Notable Data Set Award, 2/1300=0.2%)

4. **William Yang Wang**, Kathryn Mazaitis, and William W. Cohen, "A Soft Version of Predicate Invention Based on Structured Sparsity", to appear in Proceedings of the 24th International Joint Conference on Artificial Intelligence (**IJCAI 2015**), full paper for oral presentation, Buenos Aires, Argentina, July 25-31, IJCAI.

5. **William Yang Wang** and William W. Cohen, "Joint Information Extraction and Reasoning: A Scalable Statistical Relational Learning Approach", to appear in Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and The 7th International Joint Conference of the Asian Federation of Natural Language Processing (**ACL-IJCNLP 2015**), long paper for oral presentation, Beijing, China, July 26-31, ACL.

6. Yun-Nung Chen, **William Yang Wang**, Anatole Gershman and Alexander Rudnicky, "Matrix Factorization with Knowledge Graph Propagation for Unsupervised Spoken Language Understanding", to appear in Proceedings of the 53rd Annual Meeting of th e Association for Computational Linguistics and The 7th International Joint Conference of the Asian Federation of Natural Language Processing (**ACL-IJCNLP 2015**), long paper for oral presentation, Beijing, China, July 26-31, ACL.

7. **William Yang Wang**, Kathryn Mazaitis, Ni Lao, William W. Cohen, "Efficient Inference and Learning in a Large Knowledge Base: Reasoning with Extracted Information using a Locally Groundable First-Order Probabilistic Logic", in Machine Learning Journal (**MLJ**), 2015, Springer.

8. **William Yang Wang**, Miaomiao Wen, "I Can Has Cheezburger? A Nonparanormal Approach to Combining Textual and Visual Information for Predicting and Generating Popular Meme Descriptions", in Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (**NAACL HLT 2015**), long paper, Denver, CO, May 31 - June 5, ACL. Covered by U.S. and Spanish media.

9. Yun-Nung Chen, **William Yang Wang**, and Alex Rudnicky, "Jointly Modeling Inter-Slot Relations by Random Walk on Knowledge Graphs for Unsupervised Spoken Language Understanding", in Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (**NAACL HLT 2015**), long paper, Denver, CO, May 31 - June 5, ACL.

10. Yun-Nung Chen, **William Yang Wang**, and Alex Rudnicky, "Learning Semantic Hierarchy with Distributed Representations for Unsupervised Spoken Language Understanding", to appear in Proceedings of the 16th Annual Conference of the International Speech Communication Association (**INTERSPEECH 2015**), Dresden, Germany, Sept. 6-10, ISCA.

11. **William Yang Wang**, Kathryn Mazaitis, William W. Cohen, "Structure Learning via Parameter Learning", in Proceedings of the 23rd ACM International Conference on Information and Knowledge Management (**CIKM 2014**), full oral paper, Shanghai, China, Nov. 3-7th, ACM.

12. **William Yang Wang**, Lingpeng Kong, Kathryn Mazaitis, William W. Cohen, "Dependency Parsing for Weibo: An Efficient Probabilistic Logic Programming Approach", in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (**EMNLP 2014**), short paper, Doha, Qatar, Oct. 25-29, ACL.

13. **William Yang Wang**, and Zhenhao Hua, "A Semiparametric Gaussian Copula Regression Model for Predicting Financial Risks from Earnings Calls", in Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (**ACL 2014**), long paper, Baltimore, MD, June 22-27, ACL.

14. **William Yang Wang**, Kathryn Mazaitis, William W. Cohen, "ProPPR: Efficient First-Order Probabilistic Logic Programming for Structure Discovery, Parameter Learning, and Scalable Inference", in Proceedings of the AAAI 2014 International Workshop on Statistical Relational AI (**StarAI 2014**), position paper, Québec City, Québec, Canada, July 27-31, AAAI.

15. Yun-Nung Chen, **William Yang Wang**, and Alex Rudnicky, "Leveraging Frame Semantics and Distributional Semantics for Unsupervised Semantic Slot Induction for Spoken Dialogue Systems", in Proceedings of IEEE **SLT 2014**, Lake Tahoe, California and Nevada, Dec 7-10, IEEE.

16. **William Yang Wang**, Kathryn Mazaitis, William W. Cohen, "Programming with Personalized PageRank: A Locally Groundable First-Order Probabilistic Logic", in Proceedings of the 22nd ACM International Conference on Information and Knowledge Management (**CIKM 2013**), full oral paper, Burlingame, CA, USA, Oct. 27th - Nov. 1st, ACM. (Best Paper Honorable Mention Award)

17. **William Yang Wang**, Edward Lin, John Kominek, "This Text has the Scent of Starbucks: A Laplacian Structured Sparsity Model for Computational Branding Analytics", in Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (**EMNLP 2013**), full paper, Seattle, WA, USA, Oct. 18-21, ACL.

18. Di Wang, Chenyan Xiong, **William Yang Wang**, "Automatic Domain Partitioning for Multi-Domain Learning", in Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (**EMNLP 2013**), short paper, Seattle, WA, USA, Oct. 18-21, ACL.

19. Yun-Nung Chen, **William Yang Wang**, and Alex Rudnicky, "Unsupervised Induction and Filling of Semantic Slots for Spoken Dialogue Systems using Frame-Semantic Parsing", in Proceedings of the IEEE **ASRU 2013**, Olomouc, Czech Republic, Dec. 8-12, 2013, IEEE. (Best Student Paper Award)

20. Yun-Nung Chen, **William Yang Wang**, and Alex Rudnicky, "An Empirical Investigation of Sparse Log-Linear Models for Improved Dialogue Act Classification", in Proceedings of 38th International Conference on Acoustics, Speech, and Signal Processing (**ICASSP 2013**), Vancouver, May 26-31, IEEE.

21. **William Yang Wang**, Dan Bohus, Ece Kamar, and Eric Horvitz, "Crowdsourcing the Aquisition of Natural Language Corpora: Methods and Observations", in Proceedings of IEEE **SLT 2012**, Miami, Florida, Dec 2-5, IEEE.

22. **William Yang Wang**, Elijah Mayfield, Suresh Naidu, and Jeremiah Dittmar, "Historical Analysis of Legal Opinions with a Sparse Mixed-Effects Latent Variable Model", in Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (**ACL 2012**), full oral paper, Jeju Island, Korea, July 8-14, ACL. (Oral Presentation Long Paper Acceptance Rate: 14%)

23. **William Yang Wang**, Samantha Finkelstein, Amy Ogan, Alan W Black and Justine Cassell, "'Love ya, jerkface': using Sparse Log-Linear Models to Build Positive (and Impolite) Relationships with Teens", in Proceedings of the 13th annual SIGdial Meeting on Discourse and Dialogue (**SIGDIAL 2012**), full oral paper, Seoul, Korea, July 5-6, ACL.

24. **William Yang Wang**, Fadi Biadsy, Andrew Rosenberg, Julia Hirschberg, "Automatic Detection of Speaker State: Lexical, Prosodic, and Phonetic Approaches to Level-of-Interest and Intoxication Classification", in the **Computer Speech and Language** journal, 2012.

25. **William Yang Wang** and Kallirroi Georgila, "Automatic Detection of Unnatural Word-Level Segments in Unit-Selection Speech Synthesis", in Proceedings of IEEE **ASRU 2011**, Hawaii, USA, Dec. 11-15, IEEE.

26. **William Yang Wang**, Kapil Thadani and Kathleen McKeown, "Identifying Event Descriptions using Co-training with Online News Summaries", in Proceedings of the 5th International Joint Conference

on Natural Language Processing (**IJCNLP 2011**), full oral paper, Chiang Mai, Thailand, Nov. 8-13, ACL-AFNLP.

27. **William Yang Wang**, Ron Artstein, Anton Leuski, and David Traum, "Improving Spoken Dialogue Understanding Using Phonetic Mixture Models", in Cross-Disciplinary Advances in Applied Natural Language Processing: Issues and Approaches, IGI Global. (This is an invited book chapter based on the FLAIRS 2011 paper.)

28. Fadi Biadsy, **William Yang Wang**, Andrew Rosenberg, Julia Hirschberg, "Intoxication Detection using Phonetic, Phonotactic and Prosodic Cues", in Proceedings of the 12th Annual Conference of the International Speech Communication Association (**INTERSPEECH 2011**), oral paper, Florence, Italy 28-31 Aug., 2011.

29. **William Yang Wang** and Julia Hirschberg, "Detecting Levels of Interest from Spoken Dialog with Multistream Prediction Feedback and Similarity Based Fusion Learning", in Proceedings of the 12th Annual SIGdial Meeting on Discourse and Dialogue (**SIGDIAL 2011**), oral paper, Portland, Oregon, USA 17-18 June 2011.

30. **William Yang Wang**, Ron Artstein, Anton Leuski, and David Traum, "Improving spoken dialogue understanding using phonetic mixture models", in Proceedings of the 24th International Florida Artificial Intelligence Research Society Conference (**FLAIRS-24**), oral paper, Palm Beach, Florida, USA 18-20 May 2011 (Best Paper Honorable Mention Award)

31. **William Yang Wang** and Kathleen McKeown, "Got You!: Automatic Vandalism Detection in Wikipedia with Web-based Syntactic-Semantic Modeling", in Proceedings of The 23rd International Conference on Computational Linguistics (**COLING 2010**) (Oral Presentation Paper Acceptance Rate: 19%)

32. **William Yang Wang**, Lan Wang and Chongguo Li, "Using Extended Letter-to-sound Rules to Detect Pronunciation Errors Made by Chinese Learner of English", in Proceedings of The 2010 International Conference on Computational Intelligence and Software Engineering (**CiSE 2010**) (Note: This is an extended version of 2009 CAS Technical Report.)

Workshop Papers

1. **William Yang Wang**, "Position Paper", in Proceedings of the 8th Young Researchers' Roundtable on Spoken Dialog Systems (**YRRSDS 2012**), July 3-4, Seoul, South Korea.

2. **William Yang Wang**, "Position Paper", in Proceedings of the 7th Young Researchers' Roundtable on Spoken Dialog Systems (**YRRSDS 2011**), June 15-16, Portland, Oregon, USA.

3. **William Yang Wang**, "Automatic Intoxication Detection using Accent Identification Approaches"', in Proceedings of the 2011 CMU LTI Student Research Symposium (**SRS 2011**) and the first Mid-Atlantic Student Colloquium on Speech, Language and Learning (**MASC**), oral presentations. (This is based on the joint work with F. Biadsy, A. Rosenberg, and J. Hirschberg at Interspeech 2011.)

Professional Activities

- Session chair for two text mining sessions, *the 23rd ACM International Conference on Information and Knowledge Management* (CIKM) 2014

- Session chair for machine learning and data mining, *the 22nd ACM International Conference on Information and Knowledge Management* (CIKM) 2013

- PC co-chair, *the 11th annual Language Technologies Institute Student Research Symposium* (SRS) 2013

- Program committee member, *the 24th International Joint Conference on Artificial Intelligence* (IJCAI) 2015

- Program committee member, *the 2015 Conference on Empirical Methods in Natural Language Processing* (EMNLP) 2015, 2016

- Technical program committee member, *the 16th Annual Conference of the International Speech Communication Association* (INTERSPEECH) 2015, 2016

- Program committee member, *the North American Chapter of the Association for Computational Linguistics - Human Language Technologies* (NAACL-HLT) 2015,2016

- Program committee member, *the Annual Meeting of the Association for Computational Linguistics Student Research Workshop* (ACL SRW) 2016

- Program committee member, *the 5th CCF Conference on Natural Language Processing & Chinese Computing* (NLPCC) 2016

- Program committee member, *the ACM International Conference on Information and Knowledge Management* (CIKM) 2014,2015,2016

- Program committee member, *the 2nd Mid-Atlantic Student Colloquium on Speech, Language, and Learning* (MASC-SLL) 2012

- Journal reviewer, *IEEE Transactions on Affective Computing (IEEE TAC)* 2010

- Journal reviewer, *IEEE Signal Processing Letters (IEEE SPL)* 2011

- Journal reviewer, *IEEE Communications Letters (IEEE CL)* 2010, 2011

- Journal reviewer, *Journal of the American Society for Information Science and Technology (JASIST)* 2012, 2014 x 2

- Journal reviewer, *Educational Research Review* (ERR) 2012

- Journal reviewer, *IEEE Transactions on Audio, Speech, and Language Processing (IEEE TASL)* 2013 x 3, 2014

- Journal reviewer, *Bioinformatics* 2013

- Journal reviewer, *Journal of Medical Internet Research (JMIR)* 2013, 2014

- Journal reviewer, *Artificial Intelligence (AIJ)* 2014

- Journal reviewer, *Natural Language Engineering (NLE)* 2015,2016

- Reviewer, *the IEEE International Conference on Acoustics, Speech and Signal Processing* (ICASSP) 2015,2016

- Reviewer, *the 18th ACM Conference on Computer Supported Cooperative Work and Social Computing* (CSCW) 2015,2016

- Reviewer, *the ACM CHI Conference on Human Factors in Computing Systems* (CHI) 2015

- Reviewer, *the 20th ACM Conference on Intelligent User Interfaces* (IUI) 2015

- Reviewer, *the 17th ACM Conference on Computer Supported Cooperative Work and Social Computing* (CSCW) 2014

- Organizing committee, *the 7th Young Researchers' Roundtable on Spoken Dialog Systems (YRRSDS)* 2011

- Organizing committee, *the 10th and 11th annual LTI Student Research Symposium (SRS)* 2012, 2013

- Volunteer, *the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP)* 2013

- Volunteer, *the 12th and 13th annual SIGdial Meeting on Discourse and Dialogue (SIGDIAL)* 2011, 2012

- Student member, *Association for Computational Linguistics (ACL)*, 2011-

- Student member, *Association for Computing Machinery (ACM)*, 2013-

- Graduate student member, *IEEE* 2011-

- Student member, *IEEE Signal Processing Society (IEEE SPS)*, 2011-

- Affiliate member, *Speech and Language Processing Technical Committee, IEEE Signal Processing Society*, 2011-

- Member, *SIGdial/SIGSEM/SIGNLL/SIGDAT*, 2011-

Recent Presentations and Invited Talks

1. "Scalable Learning and Reasoning for Large Knowledge Graphs", at **University of Virginia**, Charlottesville, VA, Apr. 25, 2016.

2. "Scalable Learning and Reasoning for Large Knowledge Graphs", at **Toyota Technological Institute at Chicago**, Chicago, IL, Apr. 13, 2016.

3. "Scalable Learning and Reasoning for Large Knowledge Graphs", at **Northeastern University**, Boston, MA, Apr. 8, 2016.

4. "Scalable Learning and Reasoning for Large Knowledge Graphs", at **University of Rochester**, Rochester, NY, Mar. 31, 2016.

5. "Scalable Learning and Reasoning for Large Knowledge Graphs", at **Rutgers University**, New Brunswick, NJ, Mar. 29, 2016.

6. "Scalable Learning and Reasoning for Large Knowledge Graphs", at **University of Texas, Dallas**, Richardson, TX, Mar. 21, 2016.

7. "Scalable Learning and Reasoning for Large Knowledge Graphs", at **University of California, Santa Barbara**, Santa Barbara, CA, Mar. 16, 2016.

8. "Scalable Learning and Reasoning for Large Knowledge Graphs", at **Pennsylvania State University**, State College, PA, Mar. 2, 2016.

9. "Scalable Learning and Reasoning for Large Knowledge Graphs", at **New York University**, New York, NY, Nov. 11, 2015.

10. "Scalable Learning and Reasoning for Large Knowledge Graphs", at **Columbia University**, New York, NY, Nov. 9, 2015.

11. "Interdisciplinary Approaches to Computational Social Science", at **Academia Sinica**, Taipei, Taiwan (R.O.C.), Oct 29th,2015.

12. "Scalable Learning and Reasoning for Large Knowledge Graphs", at **Academia Sinica**, Taipei, Taiwan (R.O.C.), Oct 28th., 2015.

13. "Joint Information Extraction and Reasoning: A Scalable Statistical Relational Learning Approach", at **Microsoft Research Asia**, Beijing, China, July 31st, 2015.

14. "Joint Information Extraction and Reasoning: A Scalable Statistical Relational Learning Approach", at **ACL-IJCNLP 2015**, Beijing, China, July 27th, 2015.

15. "A Low-Rank Approximation Approach to Learning Joint Embeddings of News Stories and Images for Timeline Summarization", at **Yahoo! Summer Science Week**, Sunnyvale, CA, July 21th, 2015.

16. "Joint Information Extraction and Reasoning: A Scalable Statistical Relational Learning Approach", at **Yahoo! NLP group**, New York, NY, July 16th, 2015.

17. "I Can Has Cheezburger? A Nonparanormal Approach to Combining Textual and Visual Information for Predicting and Generating Popular Meme Descriptions", at **Yahoo! Labs NYC** lunch talk, New York, NY, June 25th, 2015.

18. "Structure Learning via Parameter Learning", at the 23rd ACM International Conference on Information and Knowledge Management (**CIKM 2014**), Shanghai, China, Nov. 6th, 2014.

19. "Programming with Personalized PageRank: a Locally Groundable First-Order Probabilistic Logic", at the 22nd ACM International Conference on Information and Knowledge Management (**CIKM 2013**), Burlingame, CA, USA, Oct. 30th, 2013.

20. "Programming with Personalized PageRank: a Locally Groundable First-Order Probabilistic Logic", at the 2013 ICML workshop on Inferning: Interactions between Inference and Learning. (**ICML 2013 workshop**), Atlanta, GA, Jun. 20th, 2013.

21. "Crowdsourcing Natural Language Corpora: Methods and Observations", at the 2013 Microsoft TechFest (**TechFest 2013**), Redmond, WA, Mar. 6-7th, 2013.

22. "Crowdsourcing Natural Language Corpora: Methods and Observations", at the 2013 CMU LTI Student Open House poster session, Pittsburgh, PA, Mar. 1st, 2013.

23. "Learning to Recognize Things Unknown: Studies of Out-of-Domain Detection", **Microsoft Research**, Redmond, Jan. 11th, 2013.

24. "Crowdsourcing Natural Language Corpora: Methods and Observations", **Microsoft Online Services Division**, Bellevue, WA, Dec. 11th, 2012.

25. "Structured Natural Language Elicitation via Crowdsourcing: Methods and Observations", at the second Mid-Atlantic Student Colloquium on Speech, Language and Learning (**MASC 2012**), University of Maryland, College Park, Oct. 19th.

26. "Historical Analysis of Legal Opinions with a Sparse Mixed-Effects Latent Variable Model", at the 50th Annual Meeting of the Association for Computational Linguistics (**ACL 2012**), Jeju Island, Korea, July 10th, ACL.

27. "Love ya, jerkface: using Sparse Log-Linear Models to Build Positive (and Impolite) Relationships with Teens", at the 13th annual SIGdial Meeting on Discourse and Dialogue (**SIGDIAL 2012**), Seoul National University, Seoul, Korea, July 5, 2012.

28. "Crowd-sourcing Natural Language Corpora for Time-Efficient Construction of Natural Language Processing Systems", **Microsoft Research**, Redmond, WA. June 29, 2012.

29. "Automatic Intoxication Detection using Accent Identification Approaches", at the 2011 Mid-Atlantic Student Colloquium on Speech, Language and Learning (**MASC**), Baltimore, MD, Sept. 23th, 2011.

30. "Automatic Intoxication Detection using Accent Identification Approaches", at the 2011 CMU LTI Student Research Symposium (**SRS 2011**), Pittsburgh, PA, Sept. 9th, 2011.

31. "Detecting Levels of Interest from Spoken Dialog with Multistream Prediction Feedback and Similarity Based Hierarchical Fusion Learning", at **SIGDIAL 2011**, Portland, OR., June 18th, 2011.

32. "Improving Spoken Language Understanding with Phonetic Mixture Models", Spoken Language Processing Group of **Columbia University**, NY, USA, Feb. 14th, 2011.

33. "Automatic Vandalism Detection in Wikipedia", invited talk at the Spoken Language Processing Group of **Columbia University**, NY, USA, Oct. 26th, 2010.

34. "Got You!: Automatic Vandalism Detection in Wikipedia with Web-based Syntactic-Semantic Modeling", presented at the 23rd International Conference on Computational Linguistics (**COLING 2010**), Beijing, China, Aug. 27th, 2010.

35. "Automatic Vandalism Detection in Wikipedia", invited talk at Information Science Institute (**ISI**), CA, USA, July 30th, 2010.

36. "Improving Spoken Language Understanding with Phonetic Mixture Models", invited talk at Institute for Creative Technology, University of Southern California (**USC ICT**), CA, USA, June 29th, 2010.

Student Mentoring

- Tristan Hull (Cornell, 2014)
- Chao-Yuan Wu (MLD, CMU, 2014)
- Sneha Chaudhari (LTI, CMU, 2014-2015)
- Adele Chase (Columbia University, 2010)

Natural Languages

- Mandarin Chinese (Native), English (Fluent), Cantonese (Fluent)

Programming Languages and Tools

- Languages: Matlab, Java, C#, Shell, Perl, C, C++, LISP, JSP, JavaScript, ASP, PHP, ColdFusion, Python
- Tools: HTK, Praat, Festival, SRILM, NLTK, Weka, LibSVM, LibLinear, & many optimization toolkits

Teaching Experience

- The 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, USA Summer 2016
  Tutorial Speaker, NAACL HLT 2016, *Scalable Statistical Relational Learning for NLP*.

- International Joint Conference on Artificial Intelligence, New York, USA Summer 2016
  Tutorial Speaker, IJCAI 2016, *Scalable Probabilistic Logics*.

- Peking University, China, Summer 2015
  Instructor, CIPS Summer School, *Information Extraction*, 250 students.

- Carnegie Mellon University, USA, Spring 2014
  Head Teaching Assistant, *Machine Learning for Large Datasets*, 150 students.

- Carnegie Mellon University, USA, Fall 2013
  Head Teaching Assistant, *Machine Learning*, 60 students.

- Shenzhen University, China, Spring 2008
  Teaching Assistant, *Programming Language in Visual Basics*, 150 students.

Media Coverage and Invited Stories

- Machine Learning: The Present and Future, Yehua Zhu, **Science & Technology Review**, 34(7), Apr. 15th, 2016, China Association for Science and Technology.

- AlphaGo Beats Lee Sedol, Scientists: Not Surprising, Shiqi Zhu, **IFENG.com**, Mar. 9th, 2016.

- 'Software' con sentido del humor, o cómo crear memes automóticos (y graciosos), José Manuel Blanco, **El Confidencial**, Oct. 20, 2015.

- Can a Computer Make a Meme?, Gretchen McCulloch, **Mental Floss**, Sept. 22, 2015.

- The Theory of "Technological Singularity": A Scientific Theory of Artificial Intelligence? Or Junk Science?, William Wang, **Communications of the China Computer Federation**, June Issue, 2015.

- Proof: the Science of Booze, Adams Roger, **Mariner Books**, June, 2014. Best Science Book of 2014 by Amazon, Wired, the Guardian, and NBC, Winner of the 2014 Gourmand Award for Best Spirits Book in the United States.

- Big Data and Machine Learning: a Retrospective of ICML 2013, **Programmer Magazine**, Sept. issue, 2013.

- Acquiring Corpora using Crowdsourcing, Matt Marge, **IEEE SLTC Newsletter**, May, 2013.

- Detecting Intoxication in Speech, Matt Marge, **IEEE SLTC Newsletter**, Oct., 2011.

Public Outreach

NLP/ML/speech microblog on Weibo: 45,000+ followers, and 4,000,000+ monthly views (as of Mar. 2016).

References

- **William W. Cohen**, *Carnegie Mellon University*
  Professor of Machine Learning.
  wcohen@cs.cmu.edu

- **Tom M. Mitchell**, *Carnegie Mellon University*
  E. Fredkin University Professor, and Chair of Machine Learning Department.
  tom.mitchell@cmu.edu

- **Eric J. Horvitz**, *Microsoft Research*
  Technical Fellow and Managing Director.
  horvitz@microsoft.com

- **Kathleen R. McKeown**, *Columbia University*
  Rothschild Professor of Computer Science, and Director of the Data Science Institute.
  kathy@cs.columbia.edu

- **Amanda Stent**, *Yahoo! Labs*
  Director of Research.
  stent@yahoo-inc.com

- **Dragomir Radev**, *University of Michigan*
  Professor of EECS.
  radev@umich.edu

- **Alan W. Black**, *Carnegie Mellon University*
  Professor of Language Technologies.
  awb@cs.cmu.edu

- **Julia B. Hirschberg**, *Columbia University*
  Percy K. and Vida L. W. Hudson Professor and Head of Computer Science Department.
  julia@cs.columbia.edu

- **David R. Traum**, *University of Southern California*
  Principle Scientist, Institute for Creative Technologies.
  traum@ict.usc.edu

Last updated: May 15, 2016