

David A. Jurgens

- CONTACT INFORMATION NLP and SNAP Groups *E-mail:* jurgens@cs.stanford.edu
Gates, Rm 406 *WWW:* <http://cs.stanford.edu/~jurgens>
Stanford CA 94305-9035
- EDUCATION **University of California, Los Angeles**, Ph.D. in Computer Science, June 2014
Washington University in St. Louis, M.S. in Computer Science, December 2004
and B.A. in Philosophy and Political Science, May 2004
- RESEARCH EXPERIENCE **Stanford University , NLP and SNAP Groups**
Postdoctoral Scholar, August 2015 – *present*
McGill University, Network Dynamics Lab
Postdoctoral Scholar, June 2014 – July 2015
Sapienza University of Rome, Linguistics Computing Laboratory
Research Scientist, Jan. 2013 – Feb. 2014
HRL Laboratories, Information and Systems Science Lab (ISSL)
Visiting Researcher, Sept. 2010 – Dec. 2012
- FUNDING **Volkswagen Foundation** Co-Principal Investigator, with Fabian Flöck, Przemyslaw Grabowicz, and Scott Hale. “Current Affairs 2.0: Agenda Setting in the European Union.” Call for Computational Social Science. 842,900 € over four years.
- PUBLICATIONS **Refereed**

David Jurgens, Srijan Kumar, Raine Hoover, Dan McFarland, and Dan Jurafsky. “Citation Classification for Behavioral Analysis of a Scientific Field.” *In submission*.

Rob Voigt, Nicholas P. Camp, Vinodkumar Prabhakaran, William L. Hamilton, David Jurgens, Camilla M. Griffiths, Dan Jurafsky, and Jennifer L. Eberhardt. “Racial disparities in police language in traffic stops.” *In submission*.

David Jurgens and Yulia Tsvetkov. “Language Identification.” *In submission*.

Edward Newell*, David Jurgens*, Hardik Vala, Jad Sassine, Caitrin Armstrong, Derek Ruths and Haji Mohammad Saleem. 2016. Social Migration in the Digital Age: A Case Study on Reddit During A Period of Community Unrest. Proceedings of the 10th International AAAI Conference on Web and Social Media (ICWSM).

Osman Başkaya and David Jurgens. 2016. Word Sense Induction and Disambiguation Rivaling Supervised Methods. *Journal of Artificial Intelligence Research (JAIR)*. 55(1) pp 1025-1058.

Hardik Vala, Stefan Dimitrov, David Jurgens, Andrew Piper and Derek Ruths. 2016. Annotating Characters in Literary Corpora: A Scheme, the CHARLES Tool, and an Annotated Novel. Proceedings of the 10th edition of the Language Resources and Evaluation Conference (LREC).

David Jurgens and Mohammad Taher Pilehvar. 2016. SemEval-2016 Task 14: Semantic Taxonomy Enrichment. In Proceedings of the 10th Workshop on Semantic Evaluation.

Hardik Vala, David Jurgens, Andrew Piper, and Derek Ruths. 2015. Mr. Bennet, his coachman, and the Archbishop walk into a bar but only one of them gets recognized: On The

Difficulty of Detecting Characters in Literary Texts. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). ACL.

Jussi Karlgren, Jimmy Callin, Kevyn Collins-Thompson, Amaru Cuba Gyllensten, Ariel Ekgren, David Jurgens, Anna Korhonen, Fredrik Olsson, Magnus Sahlgren, and Hinrich Schütze. 2015. Evaluating learning language representations. Proceedings of Conference and Labs of Evaluation Forum (CLEF).

David Jurgens, Mohammad Taher Pilehvar, and Roberto Navigli. Cross Level Semantic Similarity: An Evaluation Framework for Universal Measures of Similarity. *Journal of Language Resources and Evaluation*. 2015

Silvia Neculescu, Sara Mendes, David Jurgens, Núria Bel, and Roberto Navigli. 2015. Reading Between the Lines: Overcoming Data Sparsity for Accurate Classification of Lexical Relationships. Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics (*SEM). ACL.

David Jurgens and Mohammad Taher Pilehvar. 2015. Reserating the awesometastic: An automatic extension of the WordNet taxonomy for novel terms. In Proceeding of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics – Human Language Technologies (NAACL-HLT). ACL.

David Jurgens, Tyler Finethy, Caitrin Armstrong, and Derek Ruths. 2015a. Everyone’s Invited: A New Paradigm For Evaluation on Non-transferable Datasets. In Proceedings of ICWSM Workshop on Standards and Practices in Large-Scale Social Media Research. AAAI.

David Jurgens, Tyler Finethy, James McCorriston, Yi Tian Xu, and Derek Ruths. 2015b. Geolocation Prediction in Twitter Using Social Networks: A Critical Analysis and Review of Current Practice. In Proceedings of the 9th International Conference on Weblogs and Social Media (ICWSM). AAAI.

David Jurgens, James McCorriston, and Derek Ruths. 2015c. An Analysis of Exercising Behavior in Online Populations. In Proceedings of the 9th International Conference on Weblogs and Social Media (ICWSM). AAAI.

James McCorriston, David Jurgens, and Derek Ruths. 2015. Organizations are Users Too: Characterizing and Detecting the Presence of Organizations on Twitter. In Proceedings of the 9th International Conference on Weblogs and Social Media (ICWSM). AAAI.

Ryan Compton, David Jurgens, and David Allan. 2014. Geotagging One Hundred Million Twitter Accounts with Total Variation Minimization. In Proceedings of the IEEE International Conference on Big Data. IEEE.

David Jurgens and Roberto Navigli. 2014. It’s All Fun and Games until Someone Annotates: Video Games with a Purpose for Linguistic Annotation. *Transactions of the Association for Computational Linguistics* (TACL). ACL.

David Jurgens, Stefan Dimitrov, and Derek Ruths. 2014. Twitter users #CodeSwitch hashtags! #MoltoImportante #wow #헐. Proceedings of the EMNLP Workshop on Code Switching. ACL.

David Jurgens, Mohammad T. Pilehvar, and Roberto Navigli. 2014. SemEval-2014 Task 3: Cross-Level Semantic Similarity. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). ACL.

Daniele Vannella, David Jurgens, Daniele Scarfini, Domenico Toscani, and Roberto Navigli. 2014. Validating and Extending Semantic Knowledge Bases using Video Games with a Purpose. In Proceedings of the Annual Meeting for the Association for Computational Linguistics

(ACL). ACL.

David Jurgens. 2014. An analysis of ambiguity in word sense annotations. David Jurgens. Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC).

Mohammad T. Pilehvar, David Jurgens, and Roberto Navigli. 2013. Align, Disambiguate and Walk: A Unified Approach for Measuring Semantic Similarity. In Proceedings of the Annual Meeting for the Association for Computational Linguistics (ACL). ACL.

David Jurgens. 2013b. That's what friends are for: Inferring location in online communities based on social relationships. In Proceedings of the 7th International Conference on Weblogs and Social Media (ICWSM). AAAI.

Roberto Navigli, David Jurgens and Daniele Vannella. 2013. SemEval-2013 Task 12: Multilingual Word Sense Disambiguation. In Proceedings of the 7th International Workshop on Semantic Evaluation (SemEval 2013). ACL.

David Jurgens and Ioannis Klapaftis. 2013. SemEval-2013 Task 13: Word Sense Induction for Graded and Non-Graded Senses. In Proceedings of the 7th International Workshop on Semantic Evaluation (SemEval 2013). ACL.

David Jurgens. 2013a. Embracing Ambiguity: A Comparison of Annotation Methodologies for Crowdsourcing Word Sense Labels. In Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL). ACL.

Veronika Strnadova, David Jurgens, Tsai-Ching Lu. 2013. Characterizing Online Discussions in Microblogs Using Network Analysis. In Proceedings of the AAAI Spring Symposium on Analyzing Microtext. AAAI.

David Jurgens. 2012. An Evaluation of Graded Sense Disambiguation using Word Sense Induction. In Proceedings of *SEM, the First Joint Conference on Lexical and Computational Semantics. ACL.

David Jurgens, Saif M. Mohammad, Peter D. Turney, and Keith J. Holyoak. 2012. SemEval-2012 Task 2: Measuring Degrees of Relational Similarity. In Proceedings of the 6th International Workshop on Semantic Evaluation (SemEval 2012). ACL.

David Jurgens and Tsai-Ching Lu. 2012. Temporal Motifs Reveal the Dynamics of Editor Interactions in Wikipedia. In Proceedings of the 6th International Conference on Weblogs and Social Media (ICWSM). AAAI.

David Jurgens and Tsai-Ching Lu. 2012. Friends, Enemies, and Lovers: Detecting Communities in Networks Where Relationships Matter. In Proceedings of Web Science. ACM.

David Jurgens. 2011. Word sense induction by community detection. In Proceedings of Sixth ACL Workshop on Graph-based Methods for Natural Language Processing (TextGraphs-6). ACL.

David Jurgens and Keith Stevens. 2011. Measuring the impact of sense similarity on word sense induction. In Proceedings of the First Workshop on Unsupervised Learning in NLP (UNSUP-2011), pages 113–123. ACL.

David Jurgens and Keith Stevens. 2010. Capturing Nonlinear Structure in Word Spaces Through Dimensionality Reduction. Proceedings of the ACL 2010 Workshop on GEometrical Models of Natural Language Semantics (GEMS). ACL.

David Jurgens and Keith Stevens. 2010. HERMIT: Flexible Clustering for the SemEval-2 WSI Task. In Proceedings of the 5th International Workshop on Semantic Evaluation (SemEval

2010). ACL.

David Jurgens and Keith Stevens. 2010. The S-Space Package: An Open-Source Framework for Word Space Algorithms. Proceedings of the ACL System Demonstrations. ACL.

David Jurgens and Keith Stevens. 2009. Event Detection in Blogs using Temporal Random Indexing. In Proceedings of the International Workshop on Events on Emerging Text Types (eETTs), pages 21-28.

Robert Pless and David Jurgens. Road extraction from motion cues in aerial video. In Proceedings of the ACM Conference on Geographic Information Systems, pages 31--38, 2004. ACM.

Manuscripts

David Jurgens. "Making the Most of It: Word Sense Annotation and Disambiguation in the Face of Data Sparsity and Ambiguity." Doctoral Dissertation, Computer Science Department, University of California, Los Angeles, 2014.

David Jurgens and Keith Stevens. "Word Ordering with Reduced Dimensionality for Sense Induction." Technical Report 090020, University of California Los Angeles, 2010.

David Jurgens. "Road Extraction From Aerial Video Using Active Contours and Motion Cues." Masters Thesis, Computer Science Department, Washington University in St. Louis, 2005.

TECHNICAL TUTORIALS

Roberto Navigli and David Jurgens. Multilingual Semantic Processing. Presented at the International Conference on Language Resources and Evaluation (LREC), 2014. (33 registered participants)

David Jurgens and Mohammad Taher Pilehvar. Semantic Similarity Frontiers: From Concepts to Documents. Presented at the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2015. (Over 155 registered participants)

INVITED TALKS

University of California, Berkeley. 14 November 2016.

The workshop on Challenges for Data-driven Natural Language Analysis beyond Standard Data at the Institut für Maschinelle Sprachverarbeitung. 14 October 2016.

Stanford Data Science Initiative. 11 October 2016.

Information Sciences Institute at the University of Southern California. 4 March 2016.

University of Stuttgart. 3 July 2014.

Washington University in St. Louis. 23 November 2015.

AWARDS AND FELLOWSHIPS

Best paper nomination at ACL for Pilehvar, Jurgens and Navigli (2013).

Travel Grant for ICWSM 2013

UCLA Computer Science Department Fellowship, September 2005 – June 2006

First place in both 2010 and 2011 at \$CONTEST, a research presentation competition for the UCLA Computer Science Department.

TEACHING
EXPERIENCE

Sapienza University of Rome, Teaching Assistant

- Natural Language Processing

University of California, Los Angeles, Teaching Assistant

- Computer Science 111 - Operating Systems Principles
- Computer Science 136 - Computer Security Assistant
- Computer Science 132 - Compiler Construction
- Computer Science 161 - Artificial Intelligence
- Engineering 111 - Introduction to Finance and Accounting for Engineers
- Engineering 110 - Introduction to Technology Management and Economics for Engineers
- Engineering 183 - Engineering and Society Assistant

Washington University in St. Louis, Teaching Assistant

- Computer Science 102 - Introduction to Computer Science II
- Computer Science 519 - Computer Vision

Guest Lectures

Stanford University

- Linguistics 1

McGill University

- Computer Science 189: Computers and Society
- Computer Science 250: Introduction to Algorithms and Data Structures

INDUSTRIAL
EXPERIENCE

Microsoft, Bing, Research Software Development Intern (RSDE), Summer 2010

Sun Microsystems Laboratories, Research Intern, Summers 2006, 2007, and 2008

ADVISING AND
MENTORING

Stanford University – Jade Huang (MS)

McGill University – James McCorrison (BS), Tyler Finnethy (BS), Caitrin Armstrong (BS), Yi-Tian Xu (BS), Hardik Vala (MS), Koustuv Sinha (MS)

Sapienza University – Mohammad Taher Pilehvar (PhD; Co-advised with Roberto Navigli), Moreno Di Vincenzi (MS; Co-advised with Roberto Navigli)

UCLA – Sky Lin, Grace Park, Alex Nau, Moon Bae, Sean August Scot Moon, Alexander Honda, Vasishta Jayanti and David Cohen (all BS)

SERVICE

Organizing Committees

Conferences

- Data Chair for ICWSM 2017.

Workshops and Shared Tasks

- Workshop on NLP and Computational Social Science (2017) at ACL
- Workshop on NLP and Computational Social Science (2016) at EMNLP

- Workshop on NLP and Computational Social Science (2016) at ACM WebScience
- Workshop on Semantic Evaluation (SemEval) 2016
- Workshop on Semantic Evaluation (SemEval) 2015
- SemEval-2016 Task 14: Semantic Taxonomy Enrichment
- SemEval-2014 Task 3: Cross-Level Semantic Similarity
- SemEval-2013 Task 12: Multilingual Word Sense Disambiguation
- SemEval-2013 Task 13: Word Sense Induction for Graded and Non-Graded Senses
- SemEval-2012 Task 2: Measuring Degrees of Relational Similarity

Reviewing

Journals: Artificial Intelligence, Behavior Research Methods, International Journal of Corpus Linguistics, Language Resources and Evaluation, Natural Language Engineering, PLOS ONE, Transactions on Knowledge Discovery from Data (TKDD), Computer Speech & Language

Senior Program Committee ICWSM (2016, 2017)

Program Committee: ACL (2013, 2016, 2017), NAACL (2013, 2015, 2016), EMNLP (2015,2016), ICWSM (2013, 2014, 2015), *SEM (2012, 2013, 2016), SemEval (2013, 2014, 2015, 2016), EACL (2014), LREC (2014, 2016), WWW (2014, 2015, 2016, 2017), IJCAI (2016), AAAI (2016, 2017) Workshop on Multilingual and Crosslingual Methods in NLP (2016), Workshop on Evaluating Vector-Space Representations for NLP (2016), Joint Workshop on Statistical Parsing and Semantic Processing of Morphologically Rich Languages (2016), Workshop on Ethics in NLP (2017)

External Reviewer CHI (2016, 2017), CSCW (2017)

PATENTS

“Automated collaborative behavior analysis using temporal motifs” with Tsai-Ching Lu. USPTO 9,465,504, 2016.

“Method for stage-based cost analysis for task scheduling,” with Seth T. Proctor and James Megquier. USPTO Patent 8,250,579, 2012.

“Methods and apparatus for window-based fair priority scheduling,” with Seth T. Proctor and David R. Chase. USPTO Patent 8,286,173, 2012.