Jonathan David Louis May

CONTACT INFORMATION

4676 Admiralty Way, Suite 1001
Marina Del Rey, CA 90292

Phone: (310) 448-9157
Email: jonmay@isi.edu

RESEARCH INTERESTS

Computational Linguistics, Machine Translation,
Machine Learning, Semantics, Automata Theory

WORK EXPERIENCE

USC Information Sciences Institute, Marina Del Rey, California

Research Assistant Professor, Department of Computer Science *July 2016 – present*

Classes:

CSCI 662, Advanced Natural Language Processing *Fall 2019*
CSCI 544, Applied Natural Language Processing *Fall 2017,2018*

Research Lead, Artificial Intelligence Division *May 2018 – present*

Computer Scientist, Intelligent Systems Division *May 2014 – May 2018*

Projects:

DARPA-KAIROS (Co-PI)	2019-2022
DARPA-LWLL	2019-
IARPA-BETTER	2019-
DARPA-ASED	2019-
DARPA-MCS	2019-2023
DARPA-AIDA	2019
IARPA-MATERIAL (Co-PI)	2017-2021
DARPA-LORELEI (Co-PI, PI)	2016-2019
DEEPLANG (PI)	2015
DARPA-BOLT	2014-2015
IARPA-Metaphor	2014-2015

SDL Language Weaver, Los Angeles, California

Research Scientist *Aug. 2010 – May 2014*

Government Projects Technical Lead:

FP7-FAUST (Feedback Analysis for User-Adaptive Statistical Translation)
DARPA-BOLT (Broad Operation Language Translation)
CTTSO/TSWG-MTIL (Machine Translation for Informal Language)

BBN Technologies, Cambridge, Massachusetts

Staff Scientist, Natural Language Processing group *Sep. 2001 – Jun. 2004*

EDUCATION

University of Southern California, Los Angeles, California

Ph.D, Computer Science. *Sep. 2004 – Aug. 2010*

Thesis: "Weighted Tree Automata and Transducers for Syntactic Natural Language Processing" (Committee: Kevin Knight (chair), Daniel Marcu, David Chiang, Sven Koenig, Shri Narayanan, Fernando Pereira)

University of Pennsylvania, Philadelphia, Pennsylvania

M.S.E., Computer and Information Science *Feb. 2000–May 2001*

B.S.E., Computer Science Engineering. *Sep. 1997–May 2001*

Grants Awarded

"CORAL: Combined Representations for Adept Learning." 2019–2022. DARPA.

"LESTAT: Discovering Schemas from Diverse Data." 2019–2023. DARPA.

"Discovering Common Sense from Video, Images, Text and Knowledge Bases." 2019–2022. DARPA.

"CLEAR: Cross-Lingual Event & Argument Retrieval." 2019–2022. IARPA.

"Summarization and domain-Adaptive Retrieval Across Languages." 2017–2021. IARPA.

"DEEPLANG". 2015. MITRE.

Patents

"Efficient online domain adaptation".

F. Hieber and J. May. US 9,213,694. Awarded December 15, 2015.

"Personalized machine translation via online adaptation".

D. Marcu and J. May. US 9,152,622. Awarded October 6, 2015.

"Modification of annotated bilingual segment pairs in syntax-based machine translation".

W. Wang and J. May and K. Knight. US 8,825,466. Awarded September 2, 2014.

"Systems and methods for tuning parameters in statistical machine translation".

M. Hopkins and J. May. US 8,694,303. Awarded April 8, 2014.

Publications

Refereed Conference Papers

James Mullenbach, Jonathan Gordon, Nanyun Peng, and Jonathan May. "Do Nuclear Submarines Have Nuclear Captains? A Challenge Dataset for Commonsense Reasoning over Adjectives and Objects". *Proc. EMNLP*, 2019.

Ananya Subburathinam, Di Lu, Heng Ji, Jonathan May, Shih-Fu Chang, Avirup Sil, and Clare Voss. "Cross-lingual Structure Transfer for Relation and Event Extraction". *Proc. EMNLP*, 2019.

Xiaolei Huang, Jonathan May, and Nanyun Peng. "What Matters for Neural Cross-Lingual Named Entity Recognition: An Empirical Analysis". *Proc. EMNLP*, 2019.

Meryem Mhamdi, Marjorie Freedman, and Jonathan May. "Contextualized Cross-Lingual Event Trigger Extraction with Minimal Resources". *Proc. CoNLL*, 2019.

Elizabeth Boschee, Joel Barry, Jayadev Billa, Marjorie Freedman, Thamme Gowda, Constantine Lignos, Chester Palen-Michel, Michael Pust, Banriskhem Kayang Khonglah, Srikanth Madikeri, Jonathan May and Scott Miller. "SARAL: A Low-Resource Cross-Lingual Domain-Focused Information Retrieval System for Effective Rapid Document Triage". *Proc. ACL Demo Track*, 2019.

Nima Pourdamghani, Nada Aldarrab, Marjan Ghazvininejad, Kevin Knight and Jonathan May. "Translating Translationese: A Two-Step Approach to Unsupervised Machine Translation". *Proc. ACL*, 2019.

Xusen Yin and Jonathan May. "Comprehensible Context-driven Text Game Playing". *Proc. CoG*, 2019.

Ronald Cardenas, Ying Lin, Heng Ji and Jonathan May. "A Grounded Unsupervised Universal Part-of-Speech Tagger for Low-Resource Languages". *Proc. NAACL*, 2019.

Lifu Huang, Heng Ji and Jonathan May. "Cross-lingual Multi-Level Adversarial Transfer to Enhance Low-Resource Name Tagging". *Proc. NAACL*, 2019.

Ulf Hermjakob, Jonathan May, Michael Pust, and Kevin Knight. "Translating a Language You Don't Know In the Chinese Room". *Proc. ACL Demo Track*, 2018.

Ulf Hermjakob, Jonathan May, and Kevin Knight. "Out-of-the-box Universal Romanization Tool *uroman*". *Proc. ACL Demo Track*, 2018. **Best Demo Award.**

Boliang Zhang, Ying Lin, Xiaoman Pan, Di Lu, Jonathan May, Kevin Knight and Heng Ji. "ELISA-EDL: A Cross-lingual Entity Extraction, Linking and Localization System". *Proc. NAACL Demo Track*, 2018.

Yining Chen, Sorcha Gilroy, Andreas Maletti, Jonathan May, and Kevin Knight. "Recurrent Neural Networks as Weighted Language Recognizers". *Proc. NAACL*, 2018. **Outstanding Paper Award.**

Pavlos Papadopoulos, Ruchir Travadi, Colin Vaz, Nikolaos Malandrakis, Ulf Hermjakob, Nima Pourdamghani, Michael Pust, Boliang Zhang, Xiaoman Pan, Di Lu, Ying Lin, Ondřej Glembek, Murali Karthick Baskar, Martin Karafiát, Lukáš Burget, Jonathan May, Heng Ji, Kevin Knight, and Shrikanth Narayanan. "Team ELISA System for DARPA LORELEI Speech Evaluation 2016". *Proc. Interspeech*, 2017.

Xiaoman Pan, Boliang Zhang, Jonathan May, Joel Nothman, Kevin Knight and Heng Ji. "Cross-lingual Name Tagging and Linking for 282 Languages". *Proc. ACL*, 2017.

Barret Zoph, Deniz Yuret, Jonathan May and Kevin Knight. "Transfer Learning for Low-Resource Neural Machine Translation". *Proc. EMNLP*, 2016.

Barret Zoph, Ashish Vaswani, Jonathan May and Kevin Knight. "Simple, Fast Noise-Contrastive Estimation for Large RNN Vocabularies". *Proc. NAACL*, 2016.

Eunsol Choi, Matic Horvat, Jonathan May, Kevin Knight and Daniel Marcu. "Extracting Structured Scholarly Information from the Machine Translation Literature". *Proc. LREC*, 2016.

Michael Pust, Ulf Hermjakob, Kevin Knight, Daniel Marcu, and Jonathan May. "Parsing English into Abstract Meaning Representation Using Syntax-Based Machine Translation". *Proc. EMNLP*, 2015.

Jonathan May, Yassine Benjira, and Abdessamad Echihabi. "An Arabizi-English Social Media Statistical Machine Translation System". *Proc. AMTA*, 2014.

Alberto Barrón-Cedeño, Lluís Màrquez, Carlos A. Henríquez Q., Lluís Formiga, Enrique Romero, and Jonathan May. "Identifying Useful Human Correction Feedback from an On-line Machine Translation Service". *Proc. IJCAI*, 2013.

Mark Hopkins and Jonathan May. "Models of Translation Competitions". *Proc. ACL*, 2013.

Daniel Pighin, Lluís Màrquez, and Jonathan May. "An Analysis (and an Annotated Corpus) of User Responses to Machine Translation Output". *Proc. LREC*, 2012

Mark Hopkins and Jonathan May. "Tuning as Ranking". *Proc. EMNLP*, 2011.

Jonathan May, Kevin Knight, and Heiko Vogler. "Efficient Inference Through Cascades of Weighted Tree Transducers". *Proc. ACL*, 2010.

Jonathan May and Kevin Knight. "Syntactic Re-Alignment Models for Machine Translation". *Proc. EMNLP*, 2007.

Johanna Högberg, Andreas Maletti, and Jonathan May. "Bisimulation Minimisation for Weighted Tree Automata". *Proc. International Conference on Developments in Language Theory (DLT)*, 2007.

Johanna Högberg, Andreas Maletti, and Jonathan May. "Backward and Forward Bisimulation Minimisation of Tree Automata". *Proc. International Conference on Implementation and Application of Automata (CIAA)*, 2007.

Jonathan May and Kevin Knight. "Tiburon: A Weighted Tree Automata Toolkit". *Proc. International Conference on Implementation and Application of Automata (CIAA)*, 2006.

Jonathan May and Kevin Knight. "A Better N-Best List: Practical Determinization of Weighted Finite Tree Automata". *Proc. NAACL-HLT*, 2006.

Refereed Journal Articles

Ulf Hermjakob, Qiang Li, Daniel Marcu, Jonathan May, Sebastian J. Mielke, Nima Pourdamghani, Michael Pust, Xing Shi, Kevin Knight, Tomer Levinboim, Kenton Murray, David Chiang, Boliang Zhang, Xiaoman Pan, Di Lu, Ying Lin, Heng Ji. "Incident-Driven Machine Translation and Name Tagging for Low-resource Languages". *Machine Translation*. . October, 2017.

Lifu Huang, Jonathan May, Xiaoman Pan, Heng Ji, Xiang Ren, Jiawei Han, Lin Zhao, James A. Hendler. "Liberal Entity Extraction: Rapid Construction of Fine-Grained Entity Typing Systems". *Big Data*. 5(1). March, 2017.

Wei Wang, Jonathan May, Kevin Knight, and Daniel Marcu. "Re-Structuring, Re-Labeling, and Re-Aligning for Syntax-Based Machine Translation". *Computational Linguistics*. 36(2). June, 2010.

Johanna Högberg, Andreas Maletti, and Jonathan May. "Backward and Forward Bisimulation Minimisation of Tree Automata". *Theoretical Computer Science*. 410(37). September, 2009.

Jonathan Graehl, Kevin Knight, and Jonathan May. "Training Tree Transducers". *Computational Linguistics*, 34(3). September, 2008.

Jonathan May, Ada Brunstein, Prem Natarajan, and Ralph Weischedel. "Surprise! What's in a Cebuano or Hindi Name?". *ACM Transactions on Asian Language Information Processing*, 2(3). September, 2003.

Refereed Workshop Papers

Xiaoman Pan, Thamme Gowda, Heng Ji, Jonathan May, and Scott Miller. "Cross-lingual Joint Entity and Word Embedding to Improve Entity Linking and Parallel Sentence Mining". *Proc. DeepLo*, 2019.

Jonathan May, Ekaterina Shutova, Auerlie Herbelot, Xiaodan Zhu, Marianna Apidianaki, and Saif M. Mohammad. *Proceedings of the 13th International Workshop on Semantic Evaluation*. 2019.

Nanyun Peng, Marjan Ghazvininejad, Jonathan May, and Kevin Knight. "Towards Controllable Story Generation". Proc. of the 1st Workshop on Storytelling, 2018.

Marianna Apidianaki, Saif M. Mohammad, Jonathan May, Ekaterina Shutova, Steven Bethard, and Marine Carpuat. *Proceedings of the 12th International Workshop on Semantic Evaluation*. 2018.

Jonathan May and Jay Priyadarshi. "SemEval-2017 Task 9: Abstract Meaning Representation Parsing and Generation". *Proc. SemEval*, 2017.

Jonathan May. "SemEval-2016 Task 8: Meaning Representation Parsing". *Proc. SemEval*, 2016.

Jonathan Gordon, Jerry Hobbs, Jonathan May, Michael Mohler, Fabrizio Morbini, Bryan Rink, Marc Tomlinson, and Suzanne Wertheim. "A Corpus of Rich Metaphor Annotation". *Proc. Workshop on Metaphor in NLP*, 2015.

Jonathan Gordon, Jerry Hobbs, Jonathan May, and Fabrizio Morbini. "High-Precision Abductive Mapping of Multilingual Metaphors". *Proc. Workshop on Metaphor in NLP*, 2015.

Matthias Büchse, Jonathan May, and Heiko Vogler. "Determinization of Weighted Tree Automata Using Factorizations". *Proc. ATANLP*, 2009.

Book Chapters

Jonathan May and Joseph Dane. "Evidence and Artificial Intelligence". In Joseph A. Dane et al., *Begging The Question: Chauceriana, Book History, and Humanistic Inquiry (Mythodologies II)*. Los Angeles: Loyola Marymount Univ. Press, 2019.

Kevin Knight and Jonathan May. "Applications of Weighted Automata in Natural Language Processing". In M. Droste, W. Kuich, and H. Vogler, editors, *Handbook of Weighted Automata*. Springer-Verlag, 2009.

Other Publications

Jinxi Xu, Ana Licuanan, Jonathan May, Scott Miller, and Ralph Weischedel. "Answer Selection and Confidence Estimation". *New Directions in Question Answering, Papers from 2003 AAAI Spring Symposium*, Stanford University, Stanford, CA AAAI Press, 2003.

Jinxi Xu, Ana Licuanan, Jonathan May, Scott Miller, and Ralph Weischedel. "TREC 2002 QA at BBN: Answer Selection and Confidence Estimation". *Proc. TREC*, 2002.

Software

Tiburon, a weighted tree automata toolkit that incorporates algorithms from thesis work. Used in several research projects and classes on empirical methods in NLP, and downloaded over 400 times in 31 countries. Downloadable from `http://www.isi.edu/licensed-sw/tiburon`.

Professional Activity

Area chair, Machine Translation and Multilinguality, AACL 2020

Publication chair, *SEM 2020

Area Co-chair, Semantics, ACL 2020

Area Co-chair, Multilinguality, ACL 2019

Symposium Co-Organizer, SoCalNLP 2019

Treasurer, NAACL, 2019–2020

Handbook Chair, NAACL HLT 2018

Symposium Co-Organizer, WeCNLP 2018

Area Co-chair, Semantics, NAACL HLT 2018

Workshop Co-Organizer, SemEval 2018, SemEval 2019

Task Organizer, Task 9 (AMR Parsing and Generation), SemEval 2017

Task Organizer, Task 8 (Meaning Representation Parsing), SemEval 2016

Local Organizer, NACLO 2015–present

Social Media Chair, NAACL HLT 2015

Program Committee, ACL, EMNLP, NAACL, NIPS, AMTA, IJCNLP, CIAA, NACLO

Reviewer, MIT Press, TACL, *Computational Linguistics*, NSF, ANR (France), NWO (Netherlands)

Local Co-Chair for NAACL HLT 2010 (with David Chiang, Jason Riesa, Ed Hovy)

Coordinator, ISI Natural Language Seminar (2006-7)

Invited Talks

"Resource-Constrained Neural Machine Translation".

Arizona State University, Phoenix, AZ *Jan. 2020*

"Machine Translation: 350 years of progress and new challenges in the connectionist age".

Reed College, Portland, OR *Feb. 2019*

AIHAcks High School Hackathon, Playa del Rey, CA *Jun. 2019*

"How I Learned to Stop Worrying and Love Evaluations".

Johns Hopkins University, Baltimore, MD *Nov. 2016*

University of Pennsylvania, Philadelphia, PA *Nov. 2016*

"The Machine Learning of Machine Translation".

University of Southern California, Los Angeles, CA *Aug. 2015*

"Using Syntax-Based Machine Translation to Parse English into Abstract Meaning Representation".

University of Edinburgh, Edinburgh, UK *Mar. 2015*

"Machine Translation: How it Works, Why it's Hard, and How To Make it Better".

Bloomberg, London, UK *Mar. 2015*

University of Edinburgh, Edinburgh, UK *Mar. 2015*

"Toward User-Focused NLP".

USC-ISI, Marina del Rey, CA *Feb. 2014*

"Models of Translation Competitions".

USC-ISI, Marina del Rey, CA *Aug. 2013*

"Tuning As Ranking".

USC-ISI, Marina del Rey, CA *Jul. 2011*

"Natural Language Processing and Weighted Finite-State Machines".

Umeå University, Umeå, Sweden *Oct. 2008*

"Syntactic Re-Alignment Models for Machine Translation".

University of California, Berkeley, CA *Apr. 2008*

"Bisimulation Minimisation for Weighted Tree Automata".

Dresden University of Technology, Dresden, Germany *Jun. 2007*

Honors and Awards

ISI Research Award, "Universal Translators for Asylum Seekers at the Border", 2020

Best Demo Award, ACL, 2018

Outstanding Paper Award, NAACL, 2018
USC School of Engineering Doctoral Fellowship, 2004–2008

University of Pennsylvania, Department of Engineering and Applied Sciences Faculty Appreciation Award, 2001

Rear Admiral Grace Murray Hopper Endowed Scholarship, 1999–2001

Dean's List, 1997–1999

Miscellaneous

Outside Interests: Public Transit, Urban Development, Bicycling, Gardening

Date of Birth: August 7, 1978

Citizenship: United States, Germany