

Sujay Kumar Jauhar - Curriculum Vitae

Office Address

Office 3971, Microsoft Building 99
Redmond, WA 98052
USA

Contact Information

sjauhar@microsoft.com
(904) 4JA-UHAR
www.microsoft.com/en-us/research/people/sjauhar

Professional Experience

Microsoft Research AI Applied Scientist (2017 – present)
Redmond, WA, USA

- Group: Knowledge Technologies Group
- Mission: Infuse AI into Microsoft Office products

Education

PhD in Language Technology (2012 – 2017)
Carnegie Mellon University, Pittsburgh, PA, USA

- Thesis: A Relation-Centric View of Semantic Representation Learning
- Advisor: Eduard Hovy

Masters in Language Technology (2012 – 2014)
Carnegie Mellon University, Pittsburgh, PA, USA

- Advisor: Eduard Hovy
- GPA: 4.0

Masters in Natural Language Processing and Human Language Technology (2010 – 2012)
Jointly from University of Wolverhampton, UK & Université de Franche-Comté, France

- Thesis: A New Task for English Lexical Simplification
- Advisor: Lucia Specia

Bachelors Degree (2007 – 2009)
Sri Aurobindo International Centre of Education (SAICE), India

- Academic Foci: Mathematics, Computer Science, English Writing and Literature

Internships

Allen Institute for AI Research Intern Summer 2015
Seattle, WA, USA

- Mentor: Peter Turney
- Project: Multiple-choice question answering with tables

Google Research Intern Summer 2014
Mountain View, CA, USA

- Mentor: Marta Recasens
- Project: Coreference resolution of discourse deictic pronouns

Microsoft Research Research Intern Summer 2011
Bangalore, India

- Mentor: A. Kumaran
- Project: Gaming paradigm to generate language data

Profession Activities

Co-organizer, English Lexical Simplification shared task at SemEval-2012
Reviewer, ACL, NAACL, EMNLP, COLING, *SEM

Honors and Prizes

- Young Researcher attendee (among 200 selected worldwide) of the 5th Heidelberg Laureate Forum (2017)
- AI3 award at Allen Institute for AI (2015)
- Best paper award at the NAACL-2015 main conference (2015)
- Graduated with distinction from Erasmus Mundus program (2012)
- Lexical Simplification system ranked 1st overall on SemEval-2012 shared task (2012)
- Erasmus Mundus Scholarship (2011 - 2012)
- Prize for the Best Student SAICE (2009)
- Prize for Academic Excellence SAICE (2007 - 2009)

Computer Skills

Languages: Python, Scala, Java, C/C++, Octave/Matlab
Software: Experience working with most popular open-source NLP tools
Platforms: Windows, Mac OS and Linux

Languages

Fluent: English, French, Hindi, Oriya
Moderate: Bengali
Beginner: Spanish, Tamil

Teaching Experience

- TA for 11-721 Grammars and Lexicons in Fall 2014 at CMU
- TA for 10-601 Machine Learning in Spring 2014 at CMU
- English to intermediate 2nd language learner student (from France) at EMANCI Language Institute, Pondicherry
- English Writing, Literature, and Mathematics to High School Students at SAICE

Invited Talks

- “Knowledge Representation with Structured Semantic Feature Spaces” – at Allen Institute for AI, Amazon Alexa, Bloomberg and Google, May–Jun. 2017.
- “Vectors and Ontologies: A Love Story in Two Acts” – at AI2’s AI Academy Talk series, Seattle, Aug. 2015.
- “Leveraging Semantic Relations for Better Distributional Representation Learning” – at Google NLP Reading Group, Mountain View, Aug. 2014.
- “An Introduction to Natural Language Processing” – at SAICE, Pondicherry, India, Aug. 2013.
- “Integral Education” - at The Mother’s International School, New Delhi, India, Aug. 2010.

Publications

Theses

- **Sujay K. Jauhar.** (2017). A Relation-Centric View of Semantic Representation Learning. PhD Thesis: Carnegie Mellon University, Pittsburgh, PA, USA.
- **Sujay K. Jauhar.** (2012). A New Task for Lexical Simplification. Masters Thesis: University of Wolverhampton, UK & Université de Franche-Comté, France

Peer-reviewed Conference & Workshop Proceedings

2017

- **Sujay K. Jauhar** and Eduard Hovy. (2017). Embedded Semantic Lexicon Induction with Joint Global and Local Optimization. Proceedings of the 6th Joint Conference on Lexical and Computational Semantics, *SEM-2017.

2016

- **Sujay K. Jauhar**, Peter Turney and Eduard Hovy. (2016). Tables as Semi-structured Knowledge for Question Answering. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL-2016.

2015

- **Sujay K. Jauhar**, Raul D. Guerra, Edgar González and Marta Recasens. (2015). Resolving Discourse-Deictic Pronouns: A Two-stage Approach to Do It. In Proceedings of the 4th Joint Conference on Lexical and Computational Semantics, *SEM-2015.
- **Sujay K. Jauhar**, Chris Dyer and Eduard Hovy. (2015). Ontologically Grounded Multi-sense Representation Learning for Semantic Vector Space Models. In Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-2015.
- Manaal Faruqui, Jesse Dodge, **Sujay K. Jauhar**, Chris Dyer, Eduard Hovy and Noah A. Smith. (2015). Retrofitting Word Vectors to Semantic Lexicons. In Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-2015. **Best Student Paper Award**

2014

- Manaal Faruqui, Jesse Dodge, **Sujay K. Jauhar**, Chris Dyer, Eduard Hovy and Noah A. Smith. (2014). Retrofitting Word Vectors to Semantic Lexicons. In the Deep Learning and Representation Learning Workshop, NIPS-2014.
- **Sujay K. Jauhar** and Eduard Hovy. (2014). Inducing Latent Semantic Relations for Structured Distributional Semantics. In Proceedings of the 25th International Conference on Computational Linguistics, COLING-2014.

2013

- **Sujay K. Jauhar**, Yun-Nung Chen and Florian Metze. (2013). Prosody Based Speech Summarization with Two-layer Mutually Reinforced Random Walk. In Proceedings of the 6th International Joint Conference on Natural Language Processing, IJCNLP-2013.
- Kartik Goyal, **Sujay K. Jauhar**¹, Huiying Li, Mrinmaya Sachan, Shashank Srivastava and Eduard Hovy. (2013). A Structured Distributional Semantic Model: Integrating Structure with Semantics. In Proceedings of the Workshop on Continuous Vector Space Models and their Compositionality, ACL-2013.
- Kartik Goyal, **Sujay K. Jauhar**¹, Huiying Li, Mrinmaya Sachan, Shashank Srivastava and Eduard Hovy. (2013). A Structured Distributional Semantic Model for Event Co-reference. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, ACL-2013.
- Dirk Hovy, Shashank Srivastava, **Sujay K. Jauhar**, Mrinmaya Sachan, Kartik Goyal, Huiying Li, Whitney Sanders and Eduard Hovy. (2013) Identifying

¹Equally contributing first author

Metaphorical Word Use with Tree Kernels. In Proceedings of the Meta4NLP Workshop, NAACL-2013.

2012

- A Kumaran, **Sujay K. Jauhar** and Sumit Basu (2012) Doodling: A Gaming Paradigm for Generating Language Data. In Proceedings of the 4th Human Computation Workshop, HComp-2012.
- **Sujay K. Jauhar** and Lucia Specia. (2012). UOW-SHEF: SimpLex Lexical Simplicity Ranking based on Contextual and Psycholinguistic Features. In Proceedings of the 6th International Workshop on Semantic Evaluation, SemEval-2012.
- Lucia Specia, **Sujay K. Jauhar**, and Rada Mihalcea. (2012). SemEval-2012 Task 1: English Lexical Simplification. In Proceedings of the 6th International Workshop on Semantic Evaluation, SemEval-2012.

Journal Articles

- **Sujay K. Jauhar**. (2012) A New Task for Lexical Simplification. In BULAG Journal, 37, pp. 43 – 64.
- Chamanlal Gupta and **Sujay K. Jauhar**. (2005) Application Engineering Data for Solar Energy Utilization: Solar Tables for Water Heating at Pondicherry. In SESI Journal, 15, pp. 13 - 20.

Others

- **Sujay K. Jauhar**, Peter Turney and Eduard Hovy. (2016). TabMCQ: A Dataset of General Knowledge Tables and Multiple-choice Questions. arXiv preprint arXiv:1602.03960v1 [cs.CL].
- **Sujay K. Jauhar** and Lucia Specia. (2012). Re-assessing Lexical Simplification Quality from a Practically Motivated Standpoint. Extended abstract at Language Technologies Institute Student Research Symposium 2012, Carnegie Mellon University.
- **Sujay K. Jauhar** and Lucia Specia. (2012). SimpLex V0.1 - Scripts for Streamlining Lexical Simplification Learning. Extended abstract at Language Technologies Institute Student Research Symposium 2012, Carnegie Mellon University.

Patents

- “Game paradigm for language learning and linguistic data generation”. Patent US20130084976 A1. Jointly with A. Kumaran & Sumit Basu (Microsoft Research)