EMILY SHENG

ewsheng at gmail • https://ewsheng.github.io
Los Angeles, CA

EDUCATION

PH.D. STUDENT IN COMPUTER SCIENCE Aug 2015 - present
University of Southern California
Advisor: Dr. Prem Natarajan, Mentor: Dr. Nanyun Peng
Research interests: fairness/bias in natural language processing, language generation

M.S. IN COMPUTER SCIENCE Aug 2015 - May 2017
University of Southern California

B.A. IN COMPUTER SCIENCE, COGNITIVE SCIENCE Aug 2010 - May 2014
University of California, Berkeley

RESEARCH EXPERIENCE

USC/ISI: Natural language processing Aug 2015 - present
Research Assistant at University of Southern California/Information Sciences Institute

- Evaluate and mitigate biases in language generation, named entity recognition
- Information extraction, named entity recognition, and faceted search for biomedical literature
- Automatic classification techniques to study the pedagogical "value" of documents

UC Berkeley/ICSI: Resolving prepositional phrase attachment ambiguity Jan 2014 - May 2014
Research project at University of California, Berkeley/International Computer Science Institute
A survey of lexical, semantic, and contextual methods to resolve ambiguity (with Prof. Jerome Feldman)

UC Berkeley/Walker Lab: Sleep study June 2012 - May 2013
Research Assistant at University of California, Berkeley
EEG, MRI, and behavioral tests to study effect of sleep on adolescents

UC Berkeley/Concepts and Cognition Lab: Yahoo Answers study Aug 2011 - May 2013
Research Assistant at University of California, Berkeley
Extracted features of up-voted Yahoo Answers to find those favored in "good" explanations

PUBLICATIONS

Sheng, E., Chang, K.-W., Natarajan, P., Peng, N. (2019). The Woman Worked as a Babysitter: On Biases in Language Generation. *In Proceedings of EMNLP 2019.*

Sheng, E., & Natarajan, P. (2018). A Byte-sized Approach to Named Entity Recognition. *arXiv preprint arXiv:1809.08386.*

Sheng, E., Miller, S., Ambite, J. L., Natarajan, P. (2017). A Neural Named Entity Recognition Approach to Biological Entity Identification. In *Proceedings of the BioCreative VI Workshop.*

Sheng, E., Natarajan, P., Gordon, J., & Burns, G. (2017). An Investigation into the Pedagogical Features of Documents. In *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 109-120).

Gordon, J., Aguilar, S., Sheng, E., & Burns, G. (2017). Structured generation of technical reading lists. In *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 261-270).

Oral presentations

Sheng, E. (2020). Fairness in Natural Language Processing. *Presented at USC/ISI Research Day 2020.*

Sheng, E., Chang, K.-W., Natarajan, P., Peng, N. (2019). The Woman Worked as a Babysitter: On Biases in Language Generation. *Presented at EMNLP 2019.*

Sheng, E., Miller, S., Ambite, J. L., Natarajan, P. (2017). A Neural Named Entity Recognition Approach to Biological Entity Identification. *Presented at the BioCreative VI Workshop.*

Poster presentations

Sheng, E., Natarajan, P., Gordon, J., & Burns, G. (2017). An Investigation into the Pedagogical Features of Documents. *12th Workshop on Innovative Use of NLP for Building Educational Applications.*

Sheng, E., and Natarajan, P. (2016). An Investigation into the Pedagogical "Value" of Documents. *CRA-W Grad Cohort Workshop* and *ISI Graduate Student Symposium*.

PROFESSIONAL ACTIVITY

SEMINAR COORDINATOR Sept 2019 – Present
USC Information Sciences Institute Natural Language Seminar *Marina del Rey, CA*

STUDENT CO-CHAIR Sept 2019
2019 SoCal NLP Symposium *Los Angeles, CA*

PROFESSIONAL EXPERIENCE

RESEARCH INTERN May 2019 – Aug 2019
Google (Research & Machine Intelligence) *Mountain View, CA*

- Evaluate and mitigate biases towards different demographics in poetry generation

SOFTWARE ENGINEERING INTERN May 2018 – Aug 2018
Google (Research & Machine Intelligence) *Mountain View, CA*

- Evaluate semantic textual similarity across text lengths for bag-of-words and convolutional neural network model variants

SOFTWARE ENGINEER July 2014 - July 2015
Expect Labs *San Francisco, CA*

- Prototype classifier for domain-specific named entity recognition to improve a natural language understanding system
- Full-stack development of developer platform tools

SOFTWARE ENGINEERING INTERN May 2013 - Aug 2013
Samsung Telecommunications America *San Jose, CA*

- Build back end of an analytics prototype project, including optimizations and automation
- Create custom ETL process to load data into a column-oriented Vertica database

TEACHING EXPERIENCE

TEACHING ASSISTANT Aug 2015 – May 2016
Introduction to Computing course *University of Southern California*