

# Sandeep Soni

 sandeepsoni |  sandeepsoni.github.io |  soni.sandeepb@gmail.com |  +1.678.702.3780

## RESEARCH INTERESTS

---

Natural language processing, Machine learning, Computational social science, Cultural analytics.

## WORK EXPERIENCE

---

Postdoctoral Scholar, UC Berkeley	Sep 2021 - present
Software Engineer, Yahoo!	Jun 2014 - Jun 2015
Research Associate, Oneirix Labs	Aug 2010 - Jul 2012

## INTERNSHIPS

---

Information Sciences Institute, USC	May 2018 - Aug 2018
Max Plank Institute for Software Systems, Germany	May 2017 - Aug 2017
Fuse Labs, Microsoft Research	May 2016 - Aug 2016
Yahoo!	May 2013 - Aug 2013

## EDUCATION

---

PhD (Computer Science), Georgia Institute of Technology	Aug 2015-Jun 2021
M.S (Computer Science), Georgia Institute of Technology	Aug 2012-May 2014
B.E (Computer Engg.), Pune Institute of Computer Technology	Aug 2006-Jun 2010

## SELECTED PUBLICATIONS

---

For a full list of publications, see my [Google Scholar](#) page.

- [1] Kent K. Chang, Mackenzie Cramer, Sandeep Soni, and David Bamman. “Speak, Memory: An Archaeology of Books Known to ChatGPT/GPT-4”. In: *arXiv preprint* (2023). URL: <https://arxiv.org/abs/2305.00118>.
- [2] Sandeep Soni, Amanpreet Sihra, Elizabeth Evans, Matthew Wilkens, and David Bamman. “Grounding Characters and Locations in Narrative Texts”. In: *forthcoming Association for Computational Linguistics: ACL* (2023).
- [3] Sandeep Soni, David Bamman, and Jacob Eisenstein. “Predicting Long-Term Citations from Short-Term Linguistic Influence”. In: *Findings of the Association for Computational Linguistics: EMNLP* (2022).
- [4] Sandeep Soni, Lauren F Klein, and Jacob Eisenstein. “Abolitionist Networks: Modeling Language Change in Nineteenth-Century Activist Newspapers”. In: *Journal of Cultural Analytics* 6.1 (2021).
- [5] Sandeep Soni, Kristina Lerman, and Jacob Eisenstein. “Follow the Leader: Documents on the Leading Edge of Semantic Change Get More Citations”. In: *Journal of the Association for Information Science and Technology* 72.4 (2021), pp. 478–492.

## PRESENTATIONS, TUTORIALS & INVITED TALKS

---

- “Grounding Characters and Locations in Narrative Texts” at the *Literary Lab Seminar, Stanford University* (April, 2023)
- “Measuring Influence using Computational Models of Language” at the *Natural Language Processing Seminar, University of California, Berkeley* (March, 2023)
- “Using word embeddings to measure group and temporal differences in meaning” at *ICWSM* (January, 2022)
- “From Diachronic Text to Semantic Leadership Network”, *Guest Lecture, Quantitative Analysis, Emory Uni-*

versity (October, 2021)

- “Comparing Word Embedding Models”, at the *NLP+CSS 201 Online Tutorial Series* (September, 2021)
- “Computational Models of Language Change from Diachronic Text”, at the *Natural Language Seminar, University of Southern California, Information Sciences Institute* (January, 2021)
- “Follow the Leader: Documents on the Leading Edge of Semantic Change Get More Citations”, at the *3rd Behavioral Insights from Text Conference, Philadelphia, University of Pennsylvania* (2020)
- “Measuring and Modeling Language Change”, at the *5th International Conference on Computational Social Science, Amsterdam, Netherlands* (July, 2019)
- “Multivariate Hawkes Process to Model Spread of Lexical Innovations in a Network”, at the *3rd Workshop on Computational Social Science, Atlanta, GA* (2015)
- “Informational Determinants of Shaming Intensity: A Text Analytic Study of Human Rights Reporting”, at the *1st Workshop on Computational Social Science, Atlanta, GA* (2013)

## SKILLS

---

Languages (Python, Java, C, SQL) | Libraries (Pandas, Scikit-Learn, PyTorch, Transformers)