Somin Wadhwa

CS Ph.D. Candidate at Northeastern University

Boston, MA | wadhwa.s@northeastern.edu | sominw.com

Research Interests: My work lies broadly in understanding and communicating vast amounts of complex unstructured data encoded in text. I specialize in combining methods from natural language processing (NLP) and *structured* information extraction (IE/IR) to enable effective data analysis.

Education

Northeastern University — Boston, MA

Ph.D. in Computer Science, advised by Byron Wallace and Silvio Amir — *Fall 2021 - Present*

- **Research:** Methods to improve small model (<1B) performance on structured IE tasks through data elicitation from large language models (LLMs).

University of Massachusetts Amherst

Master's in Computer Science — *05/2021*

Indraprastha University

Bachelor's in Computer Science & Engineering — *05/2018*

Work Experience

Amazon — Seattle, WA

Applied Scientist II Intern at AWS — *Fall 2023*

Supervisor: Chris Kong

- Developed methods for generalizable neural entity matching (EM). Work under review.
- Benchmarked existing methods for scalable entity matching. Explored alternatives to *very large* language models to improve cost effectiveness in large scale product EM tasks.
- Performed ablations to assess usefulness of our approach, in both manual (eliciting human annotations) and automated settings (experimental design).

IBM Research — Yorktown Heights, NY

Sr. Researcher Intern — *Summer 2023*

Supervisor: Oktie Hassanzadeh

- Explored probabilistic event model construction through data elicited from LLMs.
- Designed methods to generate high quality event-sequences, filling in gaps in existing knowledge graphs like Wikidata.
- Tested new evaluation methods for event sequences in the absence of high quality reference data due to incompletes of the input knowledge graph.

Amazon — East Palo Alto, CA

Research Intern at AWS AI Labs — *Summer 2020*

Supervisor(s): Ashwin Raut

- Deployed small vision models on proprietary Amazon edge devices to monitor a live video feed for COVID-19 contact tracing.
- Ran extensive set of tests to evaluate vision models, both in terms of performance and specific device latency.

Selected Publications

Note re publication norms: In *computer science*, peer-reviewed publishing largely occurs in conference and workshop proceedings, and rarely in journals. Choice of venue, even within an area, varies greatly with the specific flavor of research (theory, applications etc) and/or geographic constraints (North America, Asia etc). Within the sub-field of natural language processing high impact venues generally include, *but are not limited to*, the *ACL conferences (ACL, EACL, EMNLP etc).

Somin Wadhwa, Silvio Amir, and Byron Wallace. Revising Relation Extraction in the era of Large Language Models. In *61st Annual Meeting of the Association for Computational Linguistics (ACL)*, Canada, 2023.

Somin Wadhwa, Vivek Khetan, Silvio Amir, and Byron Wallace. RedHOT: A Corpus of Annotated Medical Questions, Experiences, and Claims on Social Media. In *Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, Croatia, 2023.

Somin Wadhwa, Jay DeYoung, Benjamin Nye, Silvio Amir, and Byron Wallace. Jointly Extracting Interventions, Outcomes, and Findings from RCT Reports with LLMs. In *Proceedings of the Machine Learning for Healthcare Conference (MLHC)*, New York, NY, USA, 2023.

M. Jasim, Pooya Khaloo, **Somin Wadhwa**, Amy X. Zhang, Ali Sarvghad, and Narges Mahyar. CommunityClick: Capturing and Reporting Community Feedback from Town Halls to Improve Inclusivity. In *ACM Conference on Computer-Supported Cooperative Work (CSCW)*, Minnesota, USA, 2020. ***Best paper award.***

Somin Wadhwa, Kanhua Yin, Kevin S. Hughes, and Byron Wallace. Semi-Automating Knowledge Base Construction for Cancer Genetics. In *Conference on Automated Knowledge Base Construction (AKBC)*, 2020.

Professional Service

Reviewer for ACL Rolling Review, ACL 2024, NAACL 2024, EACL 2024, SemEval Workshop 2023 (co-located with ACL 2023)

CS Ph.D. Admissions Committee for Khoury College (2022)

Member of Association of Computational Anthology (2023-present), Khoury GSA (2023)

Teaching Assistant:

- Probabilistic Graphical Models (CS 688 at UMass Amherst) with Justin Domke in Spring 2021.
- Intro Physics I (Phy. 131 at UMass Amherst) with Paul Bourgeois in Spring 2020.

Graduate Coursework

Advanced Algorithms (CS611), Neural Networks (CS682), Probabilistic Graphical Models (CS688), Systems for Machine Learning (CS692S), Natural Language Processing (CS585), Advanced Information Assurance (CS660), Information Retrieval (CS646)

Technical Skills

Python, PyTorch, HuggingFace, NLTK, SpaCy, Bash, Amazon Mechanical Turk (MTurk).