Paul G. Allen Center for Computer Science & Engineering
University of Washington
dwadden@cs.washington.edu
dwadden.github.io
Google Scholar
Semantic Scholar

David Wadden

Curriculum Vitae

Research Interests

My research focuses on developing systems to extract, synthesize, and verify natural language information. I'm motivated by scientific literature understanding and its associated NLP research goals, including: learning from weak or distant supervision, modeling long-range semantic dependencies, and developing new tasks which challenge the capabilities of existing models.

Education

Since 2016 **PhD Computer Science**, *University of Washington*, Seattle
Advisor: Hannaneh Hajishirzi

2006–2010 **BA Physics**, *Amherst College*, Amherst
Phi Beta Kappa, With Distinction

Publications

Conference

2022 David Wadden, Nikita Gupta, Kenton Lee, Kristina Toutanova. **Entity-centric query refinement.** *AKBC 2022*.
Best Paper Honorable Mention.

2022 David Wadden, Kyle Lo, Bailey Kuehl, Arman Cohan, Iz Beltagy, Lucy Lu Wang, Hannaneh Hajishirzi. **SciFact-Open: Towards open-domain scientific claim verification.** *EMNLP Findings 2022*.

2022 David Wadden, Kyle Lo, Lucy Lu Wang, Arman Cohan, Iz Beltagy, Hannaneh Hajishirzi. **MultiVerS: Improving scientific claim verification with weak supervision and full-document context.** *NAACL Findings 2022*.

2022 Dustin Wright, David Wadden, Kyle Lo, Bailey Kuehl, Arman Cohan, Isabelle Augenstein, Lucy Lu Wang. **Generating Scientific Claims for Zero-Shot Scientific Fact Checking.** *ACL 2022*.

2021 David Wadden, Tal August, Qisheng Li, Tim Althoff. **The Effect of Moderation on Online Mental Health Conversations.** *ICWSM 2021*.
Best Paper for Outstanding Study Design.

2021 Rahul Nadkarni, David Wadden, Iz Beltagy, Noah A. Smith, Hannaneh Hajishirzi, Tom Hope. **Scientific Language Models for Biomedical Knowledge Base Completion: An Empirical Study.** *AKBC 2021*.

2021 Aida Amini, Tom Hope, David Wadden, Madeleine van Zuylen, Eric Horvitz, Roy Schwartz, Hannaneh Hajishirzi. **Extracting a Knowledge Base of Mechanisms from COVID-19 Papers.** *NAACL 2021*.

2020 David Wadden, Shanchuan Lin, Kyle Lo, Lucy Lu Wang, Madeleine van Zuylen, Arman Cohan, Hannaneh Hajishirzi. **Fact or Fiction: Verifying Scientific Claims.** *EMNLP 2020*.

2019 David Wadden, Ulme Wennberg, Yi Luan, Hannaneh Hajishirzi. **Entity, Relation, and Event Extraction with Contextualized Span Representations.** *EMNLP 2019*.

2019 Yi Luan, David Wadden, Luheng He, Hannaneh Hajishirzi, Mari Ostendorf. **A General Framework for Information Extraction using Dynamic Span Graphs.** *NAACL 2019*.

Workshop

2021 David Wadden, Kyle Lo. **Overview and Insights from the SciVer Shared Task on Scientific Claim Verification.** *Scholarly Document Processing Workshop @ NAACL 2021*.

Journal

2019 The ALS Stratification Consortium. **Stratification of amyotrophic lateral sclerosis patients: a crowdsourcing approach**. *Scientific Reports* 9, Article number: 690 (2019).

2018 Oana M Enache, David L Lahr, Ted E Natoli, Lev Litichevskiy, David Wadden, Corey Flynn, Joshua Z Gould, Jacob K Asiedu, Rajiv Narayan, Aravind Subramanian. **The GCTx format and cmap{Py, R, M, J} packages: resources for the optimized storage and integrated traversal of dense matrices of annotated dense matrices**. *Bioinformatics*, bty784.

2017 Aravind Subramanian et al. (David Wadden author #17 of 52). **A Next Generation Connectivity Map: L1000 Platform And The First 1,000,000 Profiles**. *Cell* 171(6):1437-1452.e17, 2017.

2017 Ian Smith, Peyton Greenside, Ted Natoli, David L. Lahr, David Wadden, Itay Tirosh, Rajiv Narayan, David E. Root, Todd R. Golub, Aravind Subramanian, John G. Doench. **Evaluation of RNAi and CRISPR technologies by large-scale gene expression profiling in the Connectivity Map**. *PLoS Biol* 15(11):e2003213.

Research and Professional Experience

Since 2016 **Graduate Research Assistant**, University of Washington, Seattle, WA

- **Scientific claim verification**: Proposed the task of *scientific claim verification*, which requires NLP systems to fact-check scientific claims against a corpus of research literature. Collected SciFact, the first dataset for this task. Developed the MultiVerS system, which leverages weak supervision to achieve state-of-the-art performance on three scientific claim verification datasets.
- **Information extraction and knowledge graph construction**: Developed the DyGIE++ system, which achieved state-of-the-art performance on 3 information extraction tasks across four datasets spanning the newswire and scientific domains. Used DyGIE++ to construct a searchable knowledge graph of mechanistic relations reported in COVID-19 research literature.
- **NLP for mental health**: Identified a natural experiment to analyze the effects of moderator presence and involvement on online mental health conversations. Used tools from psycholinguistics to demonstrate that moderation leads to greater civility, engagement, and self-disclosure.

Summer 2021 **Research Intern**, Google, Seattle, WA

- **Supervisors**: Nikita Gupta, Kenton Lee, and Kristina Toutanova
- **Entity-Centric Query Refinement**: Proposed the task of *entity-centric query refinement*, which aims to assist users in efficient domain exploration and entity discovery by offering informative refinements in response to open-ended or ambiguous input queries. Developed approach for weakly-supervised dataset creation, and trained encoder-decoder model to generate novel refinements.

Fall 2019 - Spring 2020 **Research Intern**, Allen Institute for AI, Seattle, WA

- **Supervisors**: Kyle Lo and Lucy Lu Wang
- **Scientific claim verification**: Proposed the scientific claim verification task and collected the SciFact dataset.

Summer 2018 **Software Engineering Intern**, Google, Kirkland, WA

- **Supervisor**: Evgeny Skvortsov
- **Affinity reach**: Designed and implemented a machine learning approach to forecast and report ad reach into affinity (i.e. interest) categories.

2011-2014 **Associate Computational Biologist**, The Broad Institute, Cambridge, MA
- **Supervisor**: Aravind Subramanian
- **Computational analysis of gene expression data**: Contributed to data pipeline, software tools, analysis and visualization for the Connectivity Map, a large-scale gene expression dataset created to discover novel connections between genes, drugs, and disease states.

Honors and Awards

2022 Best Paper Honorable Mention, AKBC 2022. For *Entity-Centric Query Refinement*.

2021 Best Paper Award for Outstanding Study Design, ICWSM 2021. For *The Effect of Moderation on Online Mental Health Conversations*.

2010 William Warren Stifler Prize in Physics, Amherst College.

2007 Bassett Physics Prize, Amherst College.

Open-source software

DyGIE++: Multitask system for entity, relation, and event extraction (450+ GitHub ★).

SciFact: Dataset and modeling baselines for the SciFact scientific claim verification dataset (140+ GitHub ★).

MultiVerS: State-of-the-art scientific claim verification system.

Service

Professional service

2021 Organizer, SciVer shared task. Hosted at SDP workshop @ NAACL 2021.

Reviewer: EMNLP, ACL, NAACL, ACL Rolling Review, Journal of Biomedical Informatics.

Departmental service

2019, 2020 PhD admissions committee member.

2020 Prospective PhD student committee member.

Mentoring

Win-Spr '20 Shanchuan Peter Lin, UW CSE MS student → Software engineer, Microsoft.

Win-Spr, '19 Ulme Wennberg, UW CSE MS exchange student → PhD student, KTH Royal Institute of Technology.

Technical skills

Programming languages: Python, JavaScript, R, C, Matlab.

Deep learning frameworks: Tensorflow, Jax, PyTorch, PyTorch Lightning, AllenNLP, Hugging Face Transformers.

Data science packages: Pandas, Numpy, Scikit-Learn, Matplotlib.

Community engagement

2019-Present **Mentor**, Minds Matter Seattle
Weekly tutoring and college advising for high school students from underrepresented backgrounds.