David Wadden

Curriculum Vitae

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

Research Interests

I'm broadly interested in developing systems that can extract, synthesize, and verify information expressed in natural language. I'm motivated in particular by scientific applications, especially in settings that require incorporating and reasoning over background knowledge.

Education

Since 2016 **PhD Computer Science**, *University of Washington*, Seattle.
Advised by Hannaneh Hajishirzi.
Expected graduation: 2022.

2016–2019 **MS Computer Science**, *University of Washington*, Seattle.

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

Research Experience

Since Fall '16 **Graduate Research Assistant**, University of Washington, Seattle, WA.

- **Scientific claim verification**: Formulated task and constructed the SciFact dataset to facilitate the development of scientific claim verification systems. Developed modeling baselines for this new tasks. Currently working on methods to incorporate background knowledge not stated explicitly in evidence documents, but required to verify claims.
- **Information extraction**: Built the DyGIE++ model, a multitask IE system that performs named entity recognition, relation extraction, event extraction, and coreference resolution. State-of-the art performance on datasets in the scientific and newswire domain. Followup work using DyGIE++ to extract mechanisms from COVID-19 literature.
- **NLP for mental health**: Used tools from psycholinguistics to examine the effectiveness of moderation on online mental health discussions.

Fall-Spr '19-20 **Research Intern**, Allen Institute for AI, Seattle, WA.

- **Supervisor**: Kyle Lo
- **Scientific claim verification**: Constructed the SciFact dataset during AI2 internship.

Spr'11-Spr'14 **Associate Computational Biologist**, The Broad Institute of MIT and Harvard, Cambridge, MA.

- **Supervisor**: Aravind Subramanian, PhD
- **Computational analysis of gene expression data**: Built modeling pipeline and performed data analysis for a high-throughput gene expression screen.

Professional Experience

Sum'18 **Software Engineering Intern**, Google, Kirkland, WA.

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

Publications

Conference publications

2021 **David Wadden**, Tal August, Qisheng Li, Tim Althoff. The Effect of Moderation on Online Mental Health Conversations. *ICWSM 2021*.
🏆 **Outstanding study design paper**.

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*.

Preprints

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.

Journal papers

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.

Awards

2010 William Warren Stifler Prize in Physics, Amherst College

2007 Bassett Physics Prize, Amherst College

Mentoring

I've been lucky to work with two mentees on projects resulting in conference publications.

Win-Spr '20 Peter Lin, UW CSE MS student

Win-Spr, '19 Ulme Wennberg, UW CSE MS exchange student

Teaching

Wi'17 UW CSE 446: Machine Learning (Undergraduate)

Fall'16, '17 UW CSE 427: Computational Biology (Undergraduate)

Skills

Programming languages: Python, JavaScript, R, C, Matlab, Scheme.
Deep learning: Tensorflow, PyTorch, AllenNLP, Transformers.
Data science: Pandas, Numpy, Scikit-Learn, Matplotlib.