William Hogan

510.362.1060 | whogan@ucsd.edu | linkedin.com/in/wphogan | San Diego, CA

SUMMARY

Computer Science Ph.D. candidate specializing in machine learning and natural language processing. Published novel information extraction methodologies. Interested in the intersection of natural language processing, unsupervised methods, and biomedical applications. Self-starter and problem-solver with excellent communication skills. Passionate about honing expertise in the service of solving real-world problems with cutting-edge technologies.

EXPERIENCE

Research Data Scientist Graduate Intern — Summers of 2022 & 2023
Dell Technologies — Round Rock, TX

- Built a text-to-SQL language model to access data within a database
- Applied state-of-the-art NLP and computer vision methods to identify fraudulent purchase orders
- Designed novel algorithm that prevented up to $2.1M in company losses

Graduate Student Researcher — Sept. 2019 – June 2022
Center for Microbiome Innovation, UCSD — San Diego, CA

- Researcher within the UCSD-IBM Artificial Intelligence for Healthy Living program using NLP to extract microbiome knowledge from scientific literature
- Co-developed end-to-end NLP pipeline to extract information from large amounts of raw biomedical texts
- Developed high-performing models for relationship extraction, acronym resolution, and bacteria normalization
- Co-created and maintained web-based annotator tool to develop NLP models

Co-owner, Full-stack Developer — 2015 – 2019
Design Action Collective — Oakland, CA

- Lead developer on over 30 websites and apps while also co-managing a web development company
- Improved department-wide workflow to create cleaner, more efficient code
- Improved internal standards for code commenting, git usage, pair programming, and website accessibility

PUBLICATIONS AND AWARDS

"Open-world Semi-supervised Generalized Relation Discovery Aligned in a Real-world Setting" — EMNLP, 2023
Hogan, Li, Shang — See Publication

"Fine-grained Contrastive Learning for Relation Extraction" — EMNLP, 2022
Hogan, Li, Shang — See Publication

"An Overview of Distant Supervision for RE with a Focus on Denoising and Pre-training Methods" — arXiv
Hogan — See Publication

"Abstractified Multi-instance Learning (AMIL) for Biomedical Relation Extraction" — AKBC, 2021
Hogan, Huang, Katsis, Baldwin, Kim, Baeza, Bartko, Hsu — See Publication

"BLAR: Biomedical Local Acronym Resolver" — ACL, 2021
Hogan, Baeza , Katsis, Baldwin, Kim, Hsu — See Publication

"Normalization of Predominant and Long-tail Bacterial Entities with a Hybrid CNN-LSTM" — AKBC, 2020
Hogan, Mehta, Baeza, Katsis, Kim, Bartko, Hsu — See Publication

"MDAD: An Annotated Corpus for Disease-Bacterium Association" † — TBD
Huang, Hogan, Katsis, Baldwin, Kim, Baeza, Bartko, Hsu

First place in UCSD NLP Text-mining Kaggle Competition — 2020
First place in American Society of Civil Engineers National Student Robotics Competition — 2008
Chancellor's Award for Outstanding Achievement — 2008
Dean's Award for Outstanding Achievement — 2008

†: work in progress

EDUCATION

University of California, San Diego — 2021 – Dec 2024 (expected)
Doctor of Philosophy in Computer Science, GPA: 4.0 — San Diego, CA

- Specialization in Natural Language Processing

University of California, San Diego — 2019 – 2021
Master of Science in Computer Science — San Diego, CA

- Specialization in Machine Learning and Natural Language Processing

University of California, Santa Cruz — 2003 – 2008
Bachelor of Science in Electrical Engineering, graduated with honors — Santa Cruz, CA

SELECTED PROJECTS

"Generating Position-specific Scoring Matrices for Protein Secondary Structure Prediction" — Dec. 2020
Designed and built a transformer to generate position-specific scoring matrices for protein sequences. See report and repository.

"Expanding News Timeline Summarization" — Dec. 2020
Improved on existing state-of-the-art date-wise and clustering news timeline summarization (TLS) approaches, introduced more representative evaluation metrics, and expanded the available datasets to train news TLS models. See report and repository.

"8-state Protein Secondary Structure Prediction" — June 2020
Built a convolutional, residual, and recurrent neural network (CRRNN) that uses protein sequences and corresponding position-specific scoring matrices to predict protein secondary structures. See report and repository.

"Deep Photo Style Transfer" — Mar. 2020
Reproduced results from recent works in image style transfer using convolutional neural networks. See report and repository.

VOLUNTEERING

Program Committee Member, EMNLP — 2022
Participated as a program committee member for the Unsupervised and Weakly-Supervised Methods in NLP workshop.

Program Committee Member, BioNLP — 2022 – 2023
Participated as a program committee member for the 21st BioNLP workshop, co-located with ACL, 2022.

GradPal Mentor, UCSD — 2021 – 2023
Welcomed incoming students to campus and the Computer Science and Engineering program.

Developer Mentor, Design Action Collective — 2016 – 2019
Mentored junior web developers on coding best practices. Conducted code reviews and developed curricula to address gaps in understanding.