William Hogan

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

SUMMARY

Computer Science Ph.D. student specializing in machine learning and natural language processing. Published novel information extraction methodologies. Interested in unsupervised methods within natural language processing. 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 — Summer 2022
Dell Technologies — Round Rock, TX

- Applied state-of-the-art NLP and computer vision methods to identify fraudulent purchase orders
- Designed novel algorithm that prevented $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 deep learning to extract microbiome knowledge from scientific literature
- Apply natural language processing and machine learning models on large amounts of raw medical texts
- Developed high-performing models for relationship extraction, acronym resolution, and bacteria normalization
- Co-created and maintained web-based annotator tool to test 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

Digital Media Specialist — 2011 – 2014
DataCenter — Oakland, CA

- Collaboratively designed and developed infographics of research findings
- Developed an interactive and educational online game about how to conduct a community-lead research project

PUBLICATIONS AND AWARDS

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

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

"Creating a Knowledge Base of Microbes and Diseases Associated with COVID-19" † — TBD
Hogan, Huang, Katsis, Baldwin, Kim, Baeza, Bartko, Hsu

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

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 – June 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, GPA: 3.81 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.

SKILLS

Machine Learning

- Natural Language Processing: information extraction, named entity recognition, named entity normalization, knowledge graph completion, knowledge graph augmented learning, text summarization, text classification
- Computer Vision: semantic segmentation, image style transfer, caption generation
- Architectures: transformers, deep neural networks, recurrent neural networks (LSTM, GRU, and vanilla RNNs) convolutional neural networks, generative adversarial networks, ensemble models, hidden markov models
- Foundation Models: experience with BERT, BioBERT, SciBERT, RoBERTa, GPT2, T5, ResNet, VGG
- Methods: supervised learning, distant supervision, unsupervised learning, transfer learning, reinforcement learning, random forests, support vector machines, logistic and linear regression
- Software: PyTorch, Tensorflow, NumPy, SciPy, pandas, Scikit-learn, spaCy, Sci-spaCy, NLTK, Git, Docker, Kubernetes, Unix, Linux, LaTeX, MATLAB, SQL, MongoDB, PBS and Slurm job management systems
- Strong background in linear algebra, statistics, probability, and optimization

Software Development

- Languages: Python, JavaScript, SQL, MongoDB, CSS, PHP, Ruby, C++, Java, Angular
- Strong background in object-oriented programming, test-driven development, agile workflows, Don't Repeat Yourself (DRY) coding, modular design, pair programming, and version control

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
Participated as a program committee member for the 21st BioNLP workshop, co-located with ACL, 2022.

GradPal Mentor, UCSD 2021
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.