Elliot P. Schumacher

elliot@elliotschu.com ⋄ elliotschu.com ⋄ *Google Scholar*

EDUCATION

Johns Hopkins University   *August 2017 - August 2022 (expected)*
Ph.D. in Computer Science
Center for Language and Speech Processing
Advisor: Dr. Mark Dredze, Dr. James Mayfield
Research area: Natural Language Processing and Information Extraction, with a Focus on Linking Tasks in challenging domains

Carnegie Mellon University   *August 2015-August 2017*
Master of Science in Language Technologies
Language Technologies Institute
Advisor: Dr. Maxine Eskenazi, Dr. Kevyn Collins-Thompson
Research area: Natural Language Processing for Educational domain.
Selected Courses: Machine Learning, Algorithms for NLP, Machine Translation, Language and Statistics, Deep Learning, Computational Semantics

Ohio State University   *May 2014*
Bachelors of Science in Computer & Information Science, and Linguistics
with Honors, Cum Laude
Selected Courses: Computational Linguistics, Speech and Language Processing, Advanced Artificial Intelligence, Computer Vision, Machine Learning

RESEARCH EXPERIENCE

Johns Hopkins University   September 2017 - Present
Research Assistant   *Baltimore, MD*

· *Information extraction for medical documents.* Focused on concept linking (linking mentions of medical concepts to an ontology), but also other medical tasks such as phenotyping (classifying whether a patient has a disease given their medical notes).

· *Entity linking in the crosslingual and multilingual settings.* This included building a cross-lingual linker, which linked mentions of entities in non-English languages (e.g. Mandarin Chinese) to an English knowledge bases, that can be trained in a zero-shot manner. Additionally, built a multi-lingual linker, which sought to transfer an English trained entity linker to Chinese mentions and knowledge bases. This work was applied to a search task in SCALE 2021, which was a JHU-based research effort to use external documents to improve search results.

· *Domain transfer for Entity Linking..* Studying how to build entity linkers that can transfer across different domains. This includes work on medical and biomedical text, and on news and forum data. While much recent work within entity linking has focused on wikipedia data, performance often drops on other domains.

Amazon Search   September 2020 - December 2020
Applied Scientist Intern   *Barcelona, Spain*

· Interned with the Subjective NLP group, working on search-related technologies.

Carnegie Mellon University   September 2015 - August 2017
Research Assistant   *Pittsburgh, PA*

- Worked on the DSCoVAR project, a Department of Education grant to build a Vocabulary Tutoring System.
- Developed a method of ranking the reading difficulty of a sentence for first language learners, by running a crowdsourcing task and modeling the data to find important features (see EMNLP 2016 paper).
- Built a pipeline that finds sentences with selected vocabulary words, and annotates them for difficulty and other information.

WORK EXPERIENCE

State Teachers Retirement System of Ohio — January 2013 - August 2015
Student Developer, promoted to Developer — *Columbus, Ohio*

- Developed internal applications in Java and C#.
- Designed and implemented a web application for internal forms.

Ohio State University Wexner Medical Center — May 2011 - January 2013
Student Intern — *Columbus, Ohio*

- Provided technical support within Ohio State's Hospital system.

Ohio State University College of Medicine — Sept 2009 - August 2011
Mobile Services Student — *Columbus, Ohio*

- Provided technical support for students in the College of Medicine

PUBLICATIONS

Cross-Lingual Transfer in Zero-Shot Cross-Language Entity Linking — July 2021
- **E. Schumacher**, M. Dredze, J. Mayfield. *ACL 2021 Findings. (Link)*.

Clinical Concept Linking with Contextualized Neural Representations — July 2020
- **E. Schumacher**, A. Mulyar, M. Dredze. *ACL 2020. (Link)*.

Phenotyping of Clinical Notes with Improved Document Classification Models Using Contextualized Neural Language Models. — Dec 2019
- A. Mulyar, **E. Schumacher**, M. Rouhizadeh, M. Dredze. *NeurIPS Workshop on Machine Learning for Health (ML4H). (Link)*.

Learning Unsupervised Contextual Representations for Medical Synonym Discovery — November 2019
- **E. Schumacher**, M. Dredze. *JAMIA Open. (Link)*.

Discriminative Candidate Generation for Medical Concept Linking. Proceedings of the 2019 AKBC Conference. — May 2019
- **E. Schumacher**, M. Dredze. *AKBC 2019. (Link)*.

Predicting the Relative Difficulty of Single Sentences With and Without Surrounding Context. — November 2016
- **E. Schumacher**, M. Eskenazi, G. Frishkoff, K. Collins-Thompson. *EMNLP 2016. (Link)*.

PREPRINTS

Improving Zero-Shot Multi-Lingual Entity Linking

· **E. Schumacher**, J. Mayfield, M. Dredze. *(Link)*.

TEACHING EXPERIENCE

Johns Hopkins University   Fall 2018
Teaching Assistant for Machine Learning   *Baltimore, MD*

· Responsibilities include holding office hours, writing homework assignments, and teaching recitation.
· Gave guest lectures on Decision Trees and Linear Regression.

TECHNICAL REPORTS AND PRESS COVERAGE

A Readability Analysis of Campaign Speeches from the 2016 US Presidential Campaign
March 2016

E. Schumacher, M. Eskenazi. *(Arxiv Link)*.

· **CMU Press Release**: *Most Presidential Candidates Speak at Grade 6 — 8 Level*
· **Selected Press Coverage**

   Huffington Post: *Trump's Speeches Are At A Middle-School Reading Level, Study Says.*

   Washington Post: *Trump's grammar in speeches 'just below 6th grade level,' study finds.*

   Pittsburgh Tribune-Review: *Speeches dip below 6th-grade level, study says.*

AWARDS AND ACHIEVEMENTS

Graduate Research Fellowship (Carnegie Mellon University)   Recipient

· Selected for a research fellowship, which provides full funding for tuition, stipend, and other fees.
· Scholarship value: USD 75,000 per year, renewable for duration of degree.

Maximus Scholarship Recipient (Ohio State University)   Recipient

· Received a yearly scholarship, which partially covered tuition for 4 years.

TECHNICAL STRENGTHS

Computer Languages	Python, Java, C#.
Toolkits	pytorch, sklearn, nltk, numpy, huggingface transformers
Databases	MySQL, Oracle, Microsoft SQL