Maria Antoniak

✉ maa343@cornell.edu
http://maria-antoniak.github.io

Areas of interest

Broad: Natural language processing, machine learning

Narrow: Unsupervised methods (word embeddings, topic modeling), narrative structure, sentiment analysis, paraphrase detection and synonym extraction, computational social science

Education

2016-Present **Ph.D. in Information Science**, Cornell University
- Adviser: David Mimno
- Committee Members: Lillian Lee and Mor Naaman
- Minor: Computer Science

2013-2014 **M.S. in Computational Linguistics**, University of Washington
- Adviser: Fei Xia
- Thesis: *Extracting Topically Related Synonyms from Twitter*

2007-2011 **B.A. in Program of Liberal Studies**, University of Notre Dame
- Glynn Family Honors Program

Industry experience

2018 Summer Research Intern, **Microsoft**
- Mentor: Ranjani Ramamurthy
- Worked on Project EmpowerMD (production of an automated medical scribe) and focused on unsupervised extractive summarization of a small dataset of clinical dialogues.

2017 Summer Software Engineering Intern, **Rakuten**
- Developed training data and a classifier for product gender labels for the Rakuten Catalog while restricting the classifier's use based on ethical considerations for each category of products. Also developed an anomaly detection system to discover mislabeled training data.

2014-2016 Data Scientist, **Maana**
- Led the data science design and implementation of successful proof-of-concept projects for Fortune 100 companies (e.g. Shell, GE, Chevron) using various NLP and ML techniques.
- Created a custom named entity recognition and normalization system for messy, technical text using bootstrapping and word embeddings.
- Constructed a query expansion and completion search system for the core product.
- Implemented the NLP sections of a Scala machine learning pipeline for the core product.

2014 Summer — National Security Intern, **Pacific Northwest National Laboratory**
- Mentor: Eric Bell
- Researched automatic synonym extraction for Twitter to boost performance for paraphrase detection, and implemented two baseline systems for paraphrase detection using SVM models trained using machine translation metrics, matrix factorization, and a reweighting scheme.
- Supported other NLP software development projects, including sentiment analysis, topic modeling, and data extraction.

2013 Summer — Mobile Developer Intern, **Pacific Northwest National Laboratory**
- Mentor: Courtney Corley
- Designed, created, and presented *FoodFeed*, a foodborne illness tracking application, using scraped public health data from the FDA and CDC, which won PNNL's biosurveillance mobile app competition.

Teaching experience

2011-2012 — Teaching Assistant, **CS 3300: Data-Driven Web Applications**, Cornell University

2011-2012 — English Teacher, **Ukrainian Catholic University**, Lviv, Ukraine
- Taught English to university students and seminarians and led the university's English club and organized activities for 100+ students.

Summer 2010 & 2011 — Teaching Assistant, **Humanities Spring**, Assisi, Italy
- Taught a poetry and creative writing course for high school students and assisted courses in Latin, Greek, art history, opera, and classic literature.

2010-2011 — Teaching Assistant, **First Year Writing Tutorial**, University of Notre Dame
- Taught university-level writing skills to a group of struggling freshmen (ESL students and high profile athletes), and conducted both individual sessions (6-8 per week) and either assisted or led weekly group sessions.

2008-2011 — Writing Tutor, **University Writing Center**, University of Notre Dame
- Directed 70+ discussion-based consultations per semester with individual students, and completed a one semester training course, with continued monthly training led by experts in various writing styles and disciplines (e.g. chemistry, philosophy, ESL).

Publications

2018 — **Evaluating the Stability of Embedding-based Worth Similarities**. Maria Antoniak and David Mimno. *Transactions of the Association for Computational Linguistics (TACL).*

2016 — **Natural Language Processing Techniques on Oil and Gas Drilling Data**. Maria Antoniak, Jeff Dagliesh, and Justin Lo. *SPE Intelligent Energy Conference and Exhibition.*

2015 — **Leveraging Paraphrase Labels to Extract Synonyms from Twitter**. Maria Antoniak, Eric Bell, and Fei Xia. *The Twenty-Eighth International Florida Artificial Intelligence Research Society Conference (FLAIRS).*

Posters & presentations

2015 **Extracting Topically Related Synonyms from Twitter**. Maria Antoniak, Eric Bell, and Fei Xia. Poster presented at *Workshop for Women in Machine Learning (WiML)*, collocated with *Neural Information Processing Systems (NIPS) 2015*.

Honors, awards, & scholarships

2018 Information Science Outstanding Service Award, *Cornell University*
2016-2017 Information Science Fellowship, *Cornell University*
2013 1st Place, *PNNL Biosurveillance Mobile App Development Competition*
2007-2011 Glynn Family Honors Program, *University of Notre Dame*
2007 National Merit Scholar, *National Merit Scholarship Program*

Service & leadership

2018-2019 Social Chair, *Cornell Graduate Women in Science*
2018-2019 Treasurer, *Information Science Graduate Student Association*
2017-2018 Vice President & Treasurer, *Information Science Graduate Student Association*
2010-2011 Editorial Board Member, *Notre Dame Journal of Undergraduate Research*

Non-degree studies

2016 Summer Lisbon Machine Learning Summer School (LxMLS)
2011 Summer Ukrainian Language Summer School, Lviv, Ukraine
2010 Spring French Language & Culture, Universite Catholique de l'Ouest, Angers, France

Skills

Computing

Python Stack: spaCy, Gensim, scikit-learn, pandas, SciPy, NumPy, PySpark, matplotlib
Other: Mallet, Stanford CoreNLP, NLTK, openNLP, Solr/Lucene

Languages

English: native speaker
French: intermediate level
Ukrainian & Italian: elementary level

Last updated: October 14, 2018 •