Maria Antoniak

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

Research interests

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 researched unsupervised extractive summarization 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).
- 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 Twitter synonym extraction to improve paraphrase detection performance.
- Implemented baseline systems for paraphrase detection using SVM models trained using machine translation metrics, matrix factorization, and a re-weighting 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.

Publications

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

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

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)*, 2016.

Workshops

Diverging Paths in Birth Stories: A Medical Dataset for Narrative Analysis. Maria Antoniak and David Mimno. Poster presented at *Workshop for Narrative Understanding*, *NAACL*, 2019.

Comparative Analysis of Cultural Biases with Word Embeddings. Maria Antoniak and David Mimno. Poster presented at *Text as Data (TADA)*, 2018.

Evaluating the Stability of Embedding-based Word Similarities. Maria Antoniak and David Mimno. Poster presented at *Amazon Graduate Research Symposium*, 2017.

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 *NeurIPS*, 2015.

Teaching experience

Spring 2019   Teaching Assistant, **CS 3300: Data-Driven Web Applications**, Cornell University
Spring 2017   Teaching Assistant, **CS 3300: Data-Driven Web Applications**, Cornell University
2011-2012   English Teacher, **Ukrainian Catholic University**, Lviv, Ukraine
2010 & 2011   Teaching Assistant, **Humanities Spring**, Assisi, Italy
2010-2011   Teaching Assistant, **First Year Writing Tutorial**, University of Notre Dame
2008-2011   Writing Tutor, **University Writing Center**, University of Notre Dame

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
Limited/Older Experience: Java, C++, C, Scala, Javascript

LANGUAGES

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