

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, paraphrase detection and synonym extraction, computational social science |

Education

DEGREE SEEKING PROGRAMS

- | | |
|--------------|--|
| 2016-Present | Ph.D. in Information Science , Cornell University <ul style="list-style-type: none">• Adviser: David Mimno• Committee: Lillian Lee• Minor: Computer Science |
| 2013-2014 | M.S. in Computational Linguistics , University of Washington <ul style="list-style-type: none">• Adviser: Fei Xia• Thesis: <i>Extracting Topically Related Synonyms from Twitter</i> |
| 2007-2011 | B.A. in Program of Liberal Studies , University of Notre Dame <ul style="list-style-type: none">• Glynn Family Honors Program• Adviser: Bernd Goehring |

NON-DEGREE STUDIES

- | | |
|-------------|---|
| 2016 Summer | Lisbon Machine Learning Summer School (LxMLS) |
|-------------|---|

Industry experience

- | | |
|-------------|---|
| 2018 Summer | Research Intern, MICROSOFT , Redmond, WA <ul style="list-style-type: none">• 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 , Seattle, WA <ul style="list-style-type: none">• 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**, Bellevue, WA
- 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 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.

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 Word 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 NIPS 2015.

Honors, awards, & scholarships

- 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

- 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*

Skills

COMPUTING

Python Stack: spaCy, Gensim, scikit-learn, pandas, SciPy, NumPy, PySpark, seaborn, matplotlib

Other: Mallet, Stanford CoreNLP, NLTK, openNLP, Solr/Lucene

LANGUAGES

English: native speaker

French: intermediate level

Ukrainian and Italian: elementary level

Last updated: September 3, 2018 •