

# Maria Antoniak

✉ maa343@cornell.edu   🏠 maria-antoniak.github.io   💻 www.linkedin.com/in/mariaantoniak

## Education

### Cornell University

Ithaca, NY

PHD IN INFORMATION SCIENCE

August 2016 - Present

- Focus on Natural Language Processing and Machine Learning
- Advisor: David Mimno

### University of Washington

Seattle, WA

MS IN COMPUTATIONAL LINGUISTICS

June 2013 - December 2014

- Advisor: Fei Xia
- Thesis: *Extracting Topically Related Synonyms from Twitter*
- GPA: 3.95

### University of Notre Dame

Notre Dame, IN

BA IN PROGRAM OF LIBERAL STUDIES

August 2007 - May 2011

- Advisor: Bernd Goehring
- Glynn Family Honors Program
- GPA: 3.59

## Research Interests

- Natural language processing, machine learning, unsupervised and semi-supervised learning, word embeddings, topic modeling, synonym extraction, paraphrase detection, digital humanities, computational social science.

## Experience

### Rakuten

Seattle, WA

SOFTWARE ENGINEERING INTERN

May 2017 - August 2017

- Investigated implications of imbalance of gender labels for products in the Rakuten Catalog.
- Developed training data and a classifier for product gender labels while restricting the classifier's use based on ethical considerations for each category of products.
- Developed an anomaly detection system to discover mislabeled training data.

### MAANA

Seattle, WA

DATA SCIENTIST

December 2014 - May 2016

- 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.

### Pacific Northwest National Laboratory

Seattle, WA

NATIONAL SECURITY INTERN

July - December 2014

- Researched automatic synonym extraction for Twitter to boost performance for paraphrase detection.
- Implemented two baseline systems for paraphrase detection using SVM models trained using machine translation metrics, a re-weighting scheme (TF-KLD), and non-negative matrix factorization (NMF).
- Supported other NLP software development projects, including sentiment analysis, topic modeling, and data extraction.

### Pacific Northwest National Laboratory

Richland, WA

MOBILE APP DEVELOPER INTERN

June - August 2013

- Designed, created, and presented FoodFeed, a foodborne illness tracking application that won PNNL's biosurveillance mobile app competition.
- Developed the front-end of the Android application for Nexus 4 and 7.
- Scraped and indexed public health data from FDA and CDC using Solr/Lucene.

## Teaching Experience

---

### Ukrainian Catholic University

ENGLISH TEACHER

Lviv, Ukraine

August 2011 - July 2012

- Taught English to university students and seminarians.
- Led the university's English club and organized activities for 100+ students.

### Humanities Spring

TEACHING ASSISTANT

Assisi, Italy

Summers 2010 & 2011

- Taught a poetry and creative writing course for high school students.
- Assisted courses in Latin, Greek, art history, opera, and classic literature.
- Led student excursions to historical sites in Umbria and Tuscany.

### First Year Writing Tutorial, University of Notre Dame

TEACHING ASSISTANT

Notre Dame, IN

September 2010 - May 2011

- Taught university-level writing skills to a core group of struggling freshmen.
- Conducted both individual sessions (6-8 per week) and either assisted or led weekly group sessions.
- Worked one-on-one with high profile athletes (ND Football) and ESL students.

### Writing Center, University of Notre Dame

WRITING TUTOR

Notre Dame, IN

September 2008 - May 2011

- Directed 70+ discussion-based consultations per semester with individual students.
- Led specialized workshops for grant applicants.
- Completed one semester training course, with continued monthly training led by experts in various writing styles and disciplines (e.g. chemistry, philosophy, ESL).

## Skills

---

**Computer Languages** Python (primary), Java (secondary), C++ (some experience), Scala (limited experience)

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

**Other Toolkits** Mallet, Stanford CoreNLP, NLTK, openNLP, Solr/Lucene

**Human Languages** English (native), French (intermediate), Ukrainian (beginning), Italian (beginning), Russian (beginning)

## Publications & Posters

---

### PUBLICATIONS

- Antoniak, Maria and David Mimno. "Evaluating the Stability of Embedding-based Word Similarities" *TACL*. 2017.
- Antoniak, Maria, Jeff Dugliesh, and Justin Lo. "Natural Language Processing Techniques on Oil and Gas Drilling Data." *SPE Intelligent Energy Conference and Exhibition*. 2016.
- Antoniak, Maria, Eric Bell, and Fei Xia. "Leveraging Paraphrase Labels to Extract Synonyms from Twitter." *The Twenty-Eighth International Florida Artificial Intelligence Research Society Conference*. 2015.

### POSTERS

- Antoniak, Maria, Eric Bell, and Fei Xia. "Extracting Topically Related Synonyms from Twitter." In *Workshop for Women in Machine Learning (WiML)*, collocated with NIPS 2015.

## Honors & Awards

---

- 2016 **Information Science Fellowship**, Cornell University
- 2013 **1st Place**, PNNL Biosurveillance Mobile App Development Competition
- 2010 **Editorial Board Member**, Notre Dame Journal of Undergraduate Research
- 2007 **Glynn Family Honors Program**, University of Notre Dame
- 2007 **National Merit Scholar**, National Merit Scholarship Program

## Service

---

2017-2018 **Vice President**, Cornell Information Science Graduate Student Association (ISGSA)

## Other Academic Experience

---

2016 **Lisbon Machine Learning Summer School (LxMLS)**, Instituto Superior Tecnico (Lisboa)