

# Diane M. Napolitano

☎ (914) 400-8714  
✉ [dmnapolitano@gmail.com](mailto:dmnapolitano@gmail.com)  
🌐 <http://diane.sdf-us.org>

---

## Professional Interests and Skills

Professional Interests	Natural Language Processing, Machine Learning, Data Engineering
Programming Languages	Python, Java, Perl
Other	*nix (MacOS, Linux), git, SQL (PostgreSQL, Oracle, Microsoft), flask, Sun Grid Engine, Apache Spark, Apache Hadoop, Apache Storm, PyTorch, Docker, Google Cloud, Tableau
Certifications	AWS Certified Developer Associate (in progress)

---

## Experience

### Projects

- August 2019 to Present **Developer**, *Association for Computational Linguistics*, <https://www.aclweb.org>. Contributor to project aimed at modernizing the organization's website.
- July 2019 to Present **Lead Developer**, *Tate/Lang*, [https://github.com/dmnapolitano/tate\\_lang](https://github.com/dmnapolitano/tate_lang). Attempt to browse the Tate London's collection using information derived using Natural Language Processing techniques, including word2vec-based similarity metrics.
- August 2014 to Present **Lead Developer**, *match*, <https://github.com/EducationalTestingService/match>. Python library to find the offsets of standardized, tokenized substrings within the larger original, untokenized text.

### Software Engineering

- January 2019 to August 2019 **Research Programmer**, *Rutgers University, Piscataway, NJ*  
Application Developer in the Laboratory for Computer Science Research, supporting the Department of Computer Science.
- Contributed new functionality to Autograder, the department's custom Java software used to automatically grade thousands of introductory-level Java programming assignments. Added support for assignments using arrays and matrices.
  - Wrote internal documentation and added internal assignment-archiving functionality to Autolab (<https://github.com/autolab/Autolab>), the Ruby on Rails and Docker-based front-end to Autograder.
  - Created Python-based Hadoop, Spark, and Hive tutorials for students in the Data Science MS degree program (<https://resources.cs.rutgers.edu/docs/data-science-resources/>). Performed tests of the existing infrastructure and subsequent upgrades. Advised on which software packages the department should provide for use by the MS program.
  - Oversaw the development of the MS Computer Science program's website (<https://ms.cs.rutgers.edu>) by reviewing and guiding the work of the student developer.
- May 2011 to December 2018 **Research Engineer**, *Educational Testing Service, Princeton, NJ*  
Back-End Software Engineer in the Natural Language Processing and Speech group in the Research and Development division writing production code.
- Lead Engineer for *TextEvaluator*, ETS' automated NLP-based system for measuring reading complexity and classifying texts into genres (<https://textevaluator.ets.org/>). Rewrote and re-architected the existing Perl script into an object-oriented Python program open to new features and functionality. Modernized sentence- and word-tokenization, part-of-speech tagging, and syntactic parsing. Worked with scientists, clients, and front-end developers to create the website and ensure proper back-end communication.
  - Lead Engineer for The Writing Mentor (<https://mentormywriting.org/>), ETS' Google Docs

Add-on which provides writing feedback for high school and college-level students. With assistance from junior engineers, responsible for maintenance of entire code base (Apache Storm-based architecture written in Python, JSON, Java, Perl), documentation, and collaboration with external front-end developers.

- Key contributor to *e-rater*, ETS' engine for automated essay scoring used to score test-takers' constructed written responses on the GRE, TOEFL, Praxis, and American Institute of Certified Public Accountants (AICPA) exams. Engine is an Apache Storm-based architecture written in Python, Perl, and Java. Responsible for:
  - improving our in-house rule-based sentence tokenizer by fixing bugs and calculating offsets for the standardized, tokenized sentence into the original text,
  - fixing bugs in the existing code base as well as contributing new features and code review,
  - making decisions regarding the future of the architecture and its relationship to other projects,
  - organizing and delivering data used in operational model training and evaluation,
  - on-boarding new Research Engineers, familiarizing them with the code-base and development processes.
- Lead API Engineer for the Assessment DEvelopment Passage Tool (ADEPT), an internal project aimed at assisting content developers in selecting and organizing reading passages for use on exams. Developed a flask-based API controller which serves requests via Apache httpd and facilitates communication between the model (*e-rater*, *TextEvaluator*, and a MS SQL Server database) and the view (C#). Key contributor to API and front-end design and the future of the project and its architecture.
- Lead Data Manager responsible for building and maintaining PostgreSQL databases for research and operational use, as well as performing extract-transform-load (ETL) procedures between our databases and the Data Warehouse (Oracle) intended for use by all of ETS. Worked across teams to secure access to data and mentor Research Assistants in obtaining data and executing ETL code. Databases stored over thirty million test-taker essays, short-answer responses, and spoken responses.
- Promoted from Associate Research Engineer to Research Engineer in February 2014.

February 2009 to  
March 2010

**Natural Language Processing Developer**, *Stroz Friedberg LLC*, Valley Stream, NY.

Responsible for integrating state-of-the-art Natural Language Processing and Machine Learning into existing Information Extraction system, such as a statistical named entity tagger, and a filter for junk documents developed in-house.

- Performed maintenance of existing rule-based system written in Java, using the open-source General Architecture for Text Engineering (GATE), Oracle database, and Oracle Text.
- System used to search for documents, and extract data from documents, which are analyzed by lawyers for potential use as evidence in court cases, otherwise known as Electronic Discovery.

### Teaching

September 2008 to  
May 2011

**Adjunct Instructor**, *Math and Computer Information Sciences Department*, SUNY College at Old Westbury, Old Westbury, NY.

Created lectures, assignments and exams, assigned grades, and designed and planned one or two courses per semester at small (approximately 4,000 students) open-enrolment public college. Taught introductory Java, C++, and Architecture courses along with electives in Artificial Intelligence and Operating Systems.

---

### Education

2006 to 2008

**Master of Science, Computer Science**, *State University of New York at Stony Brook*, Stony Brook, NY.

Thesis  
Advisor

*Software for Writing Assistance and Improvement for Advanced Learners of English*  
Amanda Stent

2002 to 2006

**Bachelor of Science, Computer Science**, *State University of New York at Binghamton*, Binghamton, NY. Minor in History.