

Diane M. Napolitano

28 Fairmont Avenue
Somerville, NJ 08876

(914) 400-8714

✉ dmpolitano@gmail.com

🌐 <http://diane.sdf-us.org>

Education

- 2006 to 2008 **Master of Science, Computer Science**, *State University of New York at Stony Brook*, Stony Brook, NY.
Thesis *Software for Writing Assistance and Improvement for Advanced Learners of English*
- 2002 to 2006 **Bachelor of Science, Computer Science**, *State University of New York at Binghamton*, Binghamton, NY. Minor in History.

Professional Interests and Skills

- Professional Interests Natural Language Processing, Machine Learning, Information Retrieval and Extraction
- Programming Languages Python, Perl, Java, JSON, XML, \LaTeX
- Other *nix (MacOS, Linux), git, Atlassian Suite (JIRA, Bitbucket, Bamboo, Confluence), SQL (PostgreSQL, Oracle, Microsoft), Sun Grid Engine, Stanford CoreNLP, Apache Storm

Experience

Software Engineering

- February 2014 to Present **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.
- Lead Engineer for *TextEvaluator*, ETS' automated NLP-based system for measuring reading complexity (<https://textevaluator.ets.org/>).
 - 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:
 - 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.
 - 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 which serves requests via Apache httpd and facilitates communication between the back-end (*e-rater*, *TextEvaluator*, and a MS SQL Server database) and the front-end (C#). Key contributor to API and front-end discussions 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 procedures between our databases and the internal Data Warehouse (Oracle). Work across teams to secure access to data and mentor Research Assistants.

- May 2011 to February 2014 **Associate Research Engineer**, *Educational Testing Service*, Princeton, NJ.
Promoted to Research Engineer in February 2014.
- March to May 2011 **Quality Assurance Contractor**, *Cablevision Systems Corporation*, Woodbury, NY.
Three-month contract to perform full, daily tests of Cablevision's new iPad app.
- Helped to ensure that the correct programming was being broadcast via the iPad in all of Cablevision's customer regions
 - Performed automated tests via Perl and PHP scripts to verify that customers were receiving the proper services they requested for their accounts
- Feb. 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
- Community Involvement**
- 2018 **Program Committee**, North American Association for Computational Linguistics (NAACL)'s Innovative Use of NLP for Building Educational Applications (BEA) Workshop.
- 2017, 2018 **Reviewer**, Empirical Methods in Natural Language Processing (EMNLP) System Demonstrations Program.
- 2016 **Reviewer**, Association for Computational Linguistics (ACL) System Demonstrations Program.
- 2014, 2016 **Reviewer**, Computer Assisted Language Instruction Consortium (CALICO) Journal.
- 2015 to Present **Member**, Association for Computational Linguistics (ACL).

Selected Publications

- Madnani, N., Burstein, J., Elliot, N., Beigman Klebanov, B., **Napolitano, D.**, Andreyev, S., and Schwartz, M. (2018, August). *Writing Mentor: Self-Regulated Writing Feedback for Struggling Writers*. To appear in the Proceedings of the 27th International Conference on Computational Linguistics (COLING) System Demonstrations Session, Santa Fe, NM.
- **Napolitano, D.**, Sheehan, K.M., and Mundkowsky, R. (2015, June). *Online Readability and Text Complexity Analysis with TextEvaluator*. Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics (NAACL) System Demonstrations Session, Denver, CO.
- Sheehan, K.M., Kostin, I., **Napolitano, D.** and Flor, M. The TextEvaluator Tool: Helping teachers and test developers select texts for use in instruction and assessment. *The Elementary School Journal*. Vol. 115, No. 2 (December 2014), pp. 184-209.
- Cahill, A., Madnani, N., Tetreault, J., and **Napolitano, D.** (2013, June). *Robust Systems for Preposition Error Correction Using Wikipedia Revisions*. Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), Atlanta, GA.
- Sheehan, K.M., Flor, M., and **Napolitano, D.** (2013, June). *A Two-Stage Approach for Generating Unbiased Estimates of Text Complexity*. Proceedings of the 2nd Workshop on Natural Language Processing for Improving Textual Accessibility (NLP4ITA), Atlanta, GA.
- **Napolitano, D.** and Stent, A. "TechWriter: An Evolving System for Writing Assistance for Advanced Learners of English." *CALICO Journal* 26, no. 3, pp. 611-625, May 2009.