

28 Fairmont Avenue
Somerville, NJ 08876

☎ (914) 400-8714

✉ dmmnapolitano@gmail.com

🌐 <http://diane.sdf-us.org>

Diane M. Napolitano

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*

Advisor Amanda Stent

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, Java, Perl, JSON, XML, \LaTeX , YAML

Other *nix (MacOS, Linux), git, Atlassian Suite (JIRA, Bitbucket, Bamboo, Confluence), SQL (PostgreSQL, Oracle, Microsoft), flask, Sun Grid Engine, Stanford CoreNLP, Apache Spark, Apache Hadoop, Apache Storm, Apache Hive, vmware

Experience

Software Engineering

January 2019 to Present **Research Programmer**, *Rutgers University*, Piscataway, NJ.

Back-End Software Engineer in the Laboratory for Computer Science Research, supporting the Department of Computer Science

- Contributor to Autograder, the department's custom Java software used to automatically grade introductory-level Java programming assignments.
- Maintain resources and 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/>).
- Create development virtual machines using VMWare for testing Autograder and upgrades to the department's existing Hadoop cluster.

February 2014 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.

- 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.
- o 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.
- o Lead Data Manager responsible for building and maintaining PostgreSQL databases for re-search 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.

- o Helped to ensure that the correct programming was being broadcast via the iPad in all of Cablevision's customer regions
- o 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.

- o 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
- o 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

Sept. 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-enrollment public college. Course taught included introductory-level Java courses and advanced electives in Artificial Intelligence. Courses:

- o CS1500: Introduction to Computer Applications, Summer 2010
- o CS2510: Computer Programming I, Fall 2010
- o CS2511: Computer Programming II, Spring 2011
- o CS3620: Computer Architecture, Fall 2008, Spring 2010, Spring 2011
- o CS3911: C++ in Object-Oriented Design, Fall 2008, Spring 2009, Winter 2011
- o CS4400: Artificial Intelligence, Spring 2010
- o CS5610: Operating Systems, Spring 2009 and Fall 2010
- o CS5720: Advanced Java Programming and Applications, Spring 2011

Community Involvement

- 2019 **Reviewer**, North American Association for Computational Linguistics (NAACL) Social Media Track (Short and Long Papers).
- 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).

Publications

Forsyth, Carolyn M., Stephanie Peters, Jung Aa Moon, and Diane Napolitano. 2019. *Assessing Scientific Inquiry Based on Multiple Sources of Evidence*. Presentation. American Educational Research Association (AERA), Toronto, ON, Canada. Forthcoming.

Burstein, Jill, Norbert Elliot, Beata Beigman Klebanov, Nitin Madnani, Diane Napolitano, Maxwell Schwartz, Patrick Houghton, and Hillary Molloy. 2018. "Writing Mentor: Writing Progress Using Self-Regulated Writing Support". *The Journal of Writing Analytics* 2 (1): 285–313. <https://journals.colostate.edu/analytics/article/view/213>.

Madnani, Nitin, Jill Burstein, Norbert Elliot, Beata Beigman Klebanov, Diane Napolitano, Slava Andreyev, and Maxwell Schwartz. 2018. "Writing Mentor: Self-Regulated Writing Feedback for Struggling Writers". In *Proceedings of the 27th International Conference on Computational Linguistics (COLING): System Demonstrations Session*. Santa Fe, NM. <http://www.aclweb.org/anthology/C18-2025>.

Madnani, Nitin, Aoife Cahill, Daniel Blanchard, Slava Andreyev, Diane Napolitano, Binod Gyawali, Michael Heilman, et al. 2018. *A Robust Microservice Architecture for Scaling Automated Scoring Applications*. ETS Research Report Series. doi:10.1002/ets2.12202. <https://onlinelibrary.wiley.com/doi/abs/10.1002/ets2.12202>.

Malmasi, Shervin, Keelan Evanini, Aoife Cahill, Joel Tetreault, Robert Pugh, Christopher Hamill, Diane Napolitano, and Yao Qian. 2017. "A Report on the 2017 Native Language Identification Shared Task". In *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications (BEA)*. Copenhagen, Denmark: Empirical Methods for Natural Language Processing (EMNLP). <http://www.aclweb.org/anthology/W17-5007>.

Yoon, Su-Youn, Yeonsuk Cho, and Diane Napolitano. 2016. "Spoken Text Difficulty Estimation Using Linguistic Features". In *Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications (BEA)*. San Diego, CA: North American Chapter of the Association for Computational Linguistics (NAACL). <http://m-mitchell.com/NAACL-2016/BEA/pdf/BEA1131.pdf>.

Bhat, Suma, Su-Youn Yoon, and Diane Napolitano. 2015. "Automatic Detection of Grammatical Structures from Non-native Speech". In *Proceedings of the Sixth Workshop on Speech and Language Technology in Education (SLaTE)*. Leipzig, Germany: INTER-SPEECH. <https://www.slate2015.org/files/submissions/Bhat15-AD0.pdf>.

Napolitano, Diane, Kathleen M. Sheehan, and Robert Mundkowsky. 2015. "Online Readability and Text Complexity Analysis with TextEvaluator". In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics (NAACL): System Demonstrations Session*. Denver, CO. <http://www.aclweb.org/anthology/N/N15/N15-3020.pdf>.

Sheehan, Kathleen M., Michael Flor, Diane Napolitano, and Chaitanya Ramineni. 2015a. *Using TextEvaluator to Better Understand the Comprehension Challenges Presented Within Textbooks Targeted at First Grade Readers*. Presentation. American Educational Research Association (AERA), Chicago, IL.

———. 2015b. *Using TextEvaluator to Quantify Sources of Linguistic Complexity in Textbooks Targeted at First-Grade Readers Over the Past Half Century*. ETS Research Report Series. doi:10.1002/ets2.12085. <http://onlinelibrary.wiley.com/doi/10.1002/ets2.12085/full>.

Cho, Yeonsuk, Su-Youn Yoon, Diane Napolitano, and Yuan Wang. 2014. *An Automated Spoken Text Difficulty Evaluation System*. Presentation. Computer Assisted Language Instruction Consortium (CALICO), Athens, OH. <https://calico.org/calico-conference/conferences-from-previous-years/calico-2014-ohio-university/thursday-may-8/>.

Higgins, Derrick, Chris Brew, Michael Heilman, Ramon Ziai, Lei Chen, Aoife Cahill, Michael Flor, et al. 2014. *Is getting the right answer just about choosing the right words? The role of syntactically-informed features in short answer scoring*. arXiv: 1403.0801 [cs.CL]. <http://arxiv.org/abs/1403.0801v2>.

Sheehan, Kathleen M., Irene Kostin, Diane Napolitano, and Michael Flor. 2014. "The TextEvaluator Tool: Helping Teachers and Test Developers Select Texts for Use in Instruction and Assessment". *The Elementary School Journal* 115 (2): 184–209. <http://www.jstor.org/stable/10.1086/678294>.

Sheehan, Kathleen M., and Diane Napolitano. 2014. *Measuring the Difficulty of Inferring Connections Across Sentences*. Presentation. National Council on Measurement in Education (NCME), Philadelphia, PA.

Cahill, Aoife, Joel Madnani Nitin Tetreault, and Diane Napolitano. 2013. "Robust Systems for Preposition Error Correction Using Wikipedia Revisions". In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*. Atlanta, GA. <http://www.aclweb.org/anthology-new/N/N13/N13-1055.pdf>.

Sheehan, Kathleen M., Michael Flor, and Diane Napolitano. 2013. "A Two-Stage Approach for Generating Unbiased Estimates of Text Complexity". In *Proceedings of the 2nd Workshop on Natural Language Processing for Improving Textual Accessibility (NLP4ITA)*. Atlanta, GA: North American Chapter of the Association for Computational Linguistics (NAACL). <http://www.aclweb.org/anthology/W/W13/W13-15.pdf#page=59>.

Sheehan, Kathleen M., Irene Kostin, Diane Napolitano, and Michael Flor. 2013. *Helping teachers and test developers determine the difficulty of text for instruction and assessment*. Presentation. Literacy Research Association, Addressing the Three Legs of the Text Complexity Triangle: Quantitative, Qualitative, and Reader-Task Systems, Dallas, TX.

Sheehan, Kathleen M., Irene Kostin, and Diane Napolitano. 2012a. *SourceRater: An automated approach for generating text complexity classifications aligned with the Common Core Standards*. Presentation. National Council on Measurement in Education (NCME), Vancouver, BC, Canada.

———. 2012b. *SourceRater: Helping Teachers and Test Developers Determine the Difficulty of Text for Instruction and Assessment*. Presentation. National Council on Measurement in Education (NCME), Vancouver, BC, Canada.

Napolitano, Diane, and Amanda Stent. 2009. "TechWriter: An Evolving System for Writing Assistance for Advanced Learners of English". *CALICO Journal* 26 (3): 611–625. <https://www.jstor.org/stable/calicojournal.26.3.611>.

———. 2008. *TechWriter: An individualized approach to writing assistance and improvement*. Poster. Computer Assisted Language Instruction Consortium (CALICO), Workshop on the Automatic Analysis of Learner Language. calico_poster_08.png.

References

Furnished on request.