

# 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 (History minor)**, *State University of New York at Binghamton*, Binghamton, NY.

---

## Professional Interests and Skills

- Professional Interests Natural Language Processing, Machine Learning, Information Retrieval and Extraction
- Programming Languages Python, Perl, Java,  $\LaTeX$ , C/C++, HTML/CSS/Javascript, PHP
- Operating Systems Linux (Ubuntu/Debian, Slackware, Fedora), Mac OS X, Windows
- Other SQL (PostgreSQL, Oracle, MySQL), Apache httpd/Lucene

---

## Experience

### Software Engineering

- February 2014 to Present **Research Engineer**, *Educational Testing Service*, Princeton, NJ.  
Engineer in the Natural Language Processing and Speech group in the Research and Development division.
- Lead Engineer for *TextEvaluator*, ETS' automated system for measuring and analyzing reading complexity
  - Contributor to code-bases of and research regarding both *e-rater* and *SpeechRater* through numerous sub-projects, utilizing corpus and linguistic analysis, statistics and machine learning, with substantial amount of implementation in Python
  - Contact person for data sets and corpora, both from external sources such as the Linguistic Data Consortium and internal programs such as GRE and TOEFL
- May 2011 to February 2014 **Associate Research Engineer**, *Educational Testing Service*, Princeton, NJ.
- 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

June to Aug. 2004 **Web Development Intern**, *Independent Financial Marketing Group, Inc.*, Purchase, NY.  
Summer internship with third-party financial company.

- Used and gained advanced knowledge of ASP, Visual Basic Scripting Edition, and SQL both with Oracle and Microsoft SQL Server
- Made changes and additions to Infinet21.com—one of their primary products—and developed Genesis, the company's intranet

### Projects

Dec. 2012 to Present **stanford-thrift**, <https://github.com/EducationalTestingService/stanford-thrift>.  
Multithreaded server for the NLP tools produced by Stanford University's Natural Language Processing Group using Apache Thrift.

- Server written in Java but can be communicated with from any programming language via simple API calls in that language, each of which return a string or simple data structure
- Thrift generates the interface between the client, in any language, and the server, in Java, so that to the user, the interactions between client and server are seamless and appear as calls to an API in that language
- By loading the models for all of the various tools (parser, part-of-speech tagger, etc.) once, gain a tremendous boost in runtime performance and memory usage

### Research

May to Aug. 2008 **Affinity-Based and Content-Based Access Control**, SUNY at Stony Brook, Stony Brook, NY.  
Project aims to improve blog security while still encouraging the exchange of ideas.

- Ensured the proper inclusion of Java-based Natural Language Processing and Machine Learning tools (specifically Stanford's NLP tools, OpenNLP) into larger system (Stony Brook's PLOG project)
- Used natural language features (parts of speech, named entities, coreference annotations) to cluster blog entries on topic and sentiment using Latent Semantic Analysis with K-means clustering, written in Java as part of the Weka package

Advisors Amanda Stent and Rob Johnson, Department of Computer Science

May to July 2007 **A Graphical Model for Crawling the Virtual Web**, University of Illinois at Urbana-Champaign, Urbana, IL.  
Research done as part of the Data Sciences Summer Institute, organized by the Multimodal Information and Access Synthesis group in the Department of Computer Science.

- Developed a graphical model for crawling the Virtual Web in Java Swing and powered by Apache's Nutch web crawling engine
- Software allows the user to build complex crawlers by combining a series of crawlers in a manner similar to building a finite state graph, considering crawlers as a set of crawlers and the links between them as vertices and edges in a graph

Advisor Kevin C. Chang, Department of Computer Science

- Jan. 2007 to May 2008 **Writing Assistance Tool for Native and Non-Native Speakers Using Natural Language Processing**, SUNY at Stony Brook, Stony Brook, NY.  
Research done for masters thesis, a requirement for completion of the degree program.
- o Developed TechWriter, a plugin for the jEdit text editor, which uses the OpenNLP Java Natural Language Processing package to help TechWriter detect and correct writing errors in phrasing, word choice, and other mistakes made by English language students
  - o Researched appropriate machine learning and used statistical measures, such as Pearson's r coefficient, to detect and automatically correct mistakes, restructure sentences, and provide assistance to the student to improve their English understanding
- Advisor Amanda Stent, Department of Computer Science

### 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.  
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
- Jan. 2007 to May 2008 **Graduate Teaching Assistant**, *Department of Computer Science*, SUNY at Stony Brook, Stony Brook, NY.  
Responsibilities included: grading of coursework, assisting students with coursework, website maintenance, weekly recitation sections, occasional class lecturing and exam proctoring.  
Courses:
- o CSE 592: Machine Learning, taught by Prof. IV Ramakrishnan
  - o CSE 300: Writing in Computer Science, taught by Prof. Tony Scarlato
  - o CSE 114: Computer Science I, taught by Prof. Michael Tashbook

---

### Publications and Presentations

- o Sheehan, K.M., Kostin, I., Napolitano, D. and Flor, M. (in press). The TextEvaluator Tool: Helping teachers and test developers select texts for use in instruction and assessment. *Elementary School Journal*.
- o Cho, Y., Yoon, S., Napolitano, D. (2014, May). *An Automated Spoken Text Difficulty Evaluation System*. To be presented at the Computer Assisted Language Instruction Consortium (CALICO) Conference, Athens, OH.
- o Sheehan, K.M. and Napolitano, D. (2014, April). *Measuring the Difficulty of Inferring Connections Across Sentences*. Paper presented at the annual meeting of the National Council on Measurement in Education (NCME), Philadelphia, PA.
- o Derrick Higgins, Chris Brew, Michael Heilman, Ramon Ziai, Lei Chen, Aoife Cahill, Michael Flor, Nitin Madnani, Joel Tetreault, Daniel Blanchard, Diane Napolitano, Chong Min Lee, John Blackmore. (2014, March). *Is getting the right answer just about choosing the right words? The role of syntactically-informed features in short answer scoring*. arXiv:1403.0801v2.

- Sheehan, KM, Kostin, I, Napolitano, D, Flor, M. (2013, December). *Helping teachers and test developers determine the difficulty of text for instruction and assessment*. Paper presented at: Addressing the Three Legs of the Text Complexity Triangle: Quantitative, Qualitative, and Reader-Task Systems. Proceedings of the 63rd Annual Conference of the Literacy Research Association, Dallas, TX.
- 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.
- Sheehan, K.M., Kostin, I., and Napolitano, D. (2012, April). *SourceRater: An automated approach for generating text complexity classifications aligned with the Common Core Standards*. Paper presented at the annual meeting of the National Council on Measurement in Education (NCME), Vancouver, BC.
- Sheehan, K.M., Kostin, I., and Napolitano, D. (2012, April). *SourceRater: Helping Teachers and Test Developers Determine the Difficulty of Text for Instruction and Assessment*. Paper presented at the annual meeting of the National Council on Measurement in Education (NCME), Vancouver, BC.
- Diane M. Napolitano and Amanda Stent. "TechWriter: An Evolving System for Writing Assistance for Advanced Learners of English." *CALICO Journal* 26, no. 3, pp. 611-625, May 2009.
- Diane M. Napolitano and Amanda Stent. Poster, *TechWriter: An individualized approach to writing assistance and improvement*. Computer Assisted Language Instruction Consortium (CALICO) workshop, "Automatic Analysis of Learner Language". 2009.

---

## References

Furnished on request.