Chandler May

Education

2013–2022 **Ph.D., Computer Science**.
The Johns Hopkins University, Baltimore, MD

2013–2017 **M.S.E., Computer Science**.
The Johns Hopkins University, Baltimore, MD

2007–2011 **B.S., Mathematics**, with Distinction.
Harvey Mudd College, Claremont, CA
Honors:
- Courtney S. Coleman Prize (2009)
- Meritorious Designation, Interdisciplinary Contest in Modeling (2009)

Professional Interests

- Software engineering (Full stack)
- Crowd annotation
- Human-computer interaction
- Visualization & statistics
- Topic modeling

Work Experience

2022–present **Research Associate**, *Human Language Technology Center of Excellence*, Baltimore, MD.
Providing software engineering support for NLP research projects.
Supervisor: Benjamin Van Durme

Baltimore, MD
cmay14@jh.edu • *www.ccmaymay.net* • *chandler-may*
ccmaymay • *she/her*

2013–2022 **Graduate Research Assistant**, *The Johns Hopkins University*, Baltimore, MD.

Supported completion of deliverables for NLP research grants and proposals.
Supervisor: Benjamin Van Durme
Detailed achievements:

- Developed template information extraction web interface using Docker Compose, Python, Redis, Node.js, Vue.js, and Bootstrap (in progress);
- Developed template information extraction prototype in Google Sheets using Python;
- Integrated NLP systems developed by other researchers to implement third party REST API using Docker Compose, Python, Redis, and S3.
  - Developed mock API implementation and instituted continuous integration tests to ensure compliance with evolving specifications.
- Developed web interface for coreference annotation using Backbone.js and Bootstrap.
  - Interface has been used directly in three projects and shared as starter code for others.
- Launched and managed crowd annotation tasks using Amazon Mechanical Turk and local annotation system; analyzed results using R and tidyverse.
- Maintained Python interface to Thrift NLP communication protocol ("Concrete").
  - Contributed two bug fixes to Thrift Python library.
  - Implemented compact serialization protocol in C and GLib, contributed to Thrift.
  - Integrated Thrift library with C implementation, reducing serialization time by 90%.
  - Implemented custom UUID scheme, reducing compressed file sizes by 30%.

2017 **Research Intern**, *Microsoft Research*, New York, NY.

Investigated a hypothesized empirical trade-off between fairness and interpretability in practice using Python with scikit-learn, Weka, and R.
Supervisor: Hanna Wallach

2014 **Graduate Research Assistant**, *Summer Camp for Applied Language Exploration (SCALE)*, Baltimore, MD.

Implemented stochastic variational inference for nested hierarchical Dirichlet process topic model in Python and Cython and mentored high school student to develop D3.js web visualization of model.
Supervisor: Benjamin Van Durme

2011–2013 **Post Bachelors Research Associate**, *Pacific Northwest National Laboratory*, Baltimore, MD.

Implemented machine learning software solutions for various projects.
Supervisor: Andrew Stevens
Manager: Andrew Cowell
Detailed achievements:

- Developed text analytics for social media and cyber data in Python and Java.
- Implemented parallelized framework for task-driven dictionary learning in C++.
- Implemented information retrieval component of NLP pipeline in Java.
- Proposed and developed internal desktop grid computing prototype using BOINC.

2010–2011 **System Administrator**, *Harvey Mudd College*, Claremont, CA.

Developed dynamic web visualization of dorm energy usage.
Supervisors: Mike Erlinger, Richard Haskell

2009 **Student Researcher**, *North Carolina State University*, Raleigh, NC.
Participated in Modeling and Industrial Applied Mathematics REU: Extracted NFL game data from the web and analyzed relative performance of sports ranking models on NFL games using MATLAB.
Advisor: Carl Meyer

2007–2009 **CS Department Consultant**, *Harvey Mudd College*, Claremont, CA.
Migrated internal knowledge base to wiki; migrated web content to new site.
Advisor: Mike Erlinger

2008 **Student Researcher**, *Harvey Mudd College*, Claremont, CA.
Participated in summer research on GPS-ACS Network Management Project: Designed computer network to simulate GPS ground control network; configured, tested, and documented network monitoring device and server.
Advisor: Mike Erlinger

Teaching Experience

2017 **Teaching Assistant**, *The Johns Hopkins University*, Baltimore, MD.
Assisted with instruction and grading of Knowledge Discovery from Text class: Provided office hours; delivered two lectures; co-developed homework assignments, exams, and projects; led grading of all student work.
Supervisor: Benjamin Van Durme

2009–2011 **Academic Excellence Facilitator**, *Harvey Mudd College*, Claremont, CA.
Facilitated tutoring sessions for core undergraduate math classes.
Supervisor: Wendy Menefee-Libey

2008–2009 **Math Department Grader**, *Harvey Mudd College*, Claremont, CA.
Graded Discrete Mathematics and Operations Research assignments.
Supervisors: Kimberly Kindred, Susan Martonosi

Ph.D. Thesis

Title *Topic Modeling in Theory and Practice*

Advisor Benjamin Van Durme

Description While topic modeling research rarely involves the formulation and application of explicit theories, I show that it does develop and utilize implicit theories. Through four case studies, I demonstrate that these theories do not always hold in practice and argue that such knowledge gaps may exist in scientific research more broadly.

Publications

Michelle Yuan, Patrick Xia, Chandler May, Benjamin Van Durme, and Jordan Boyd-Graber. Adapting coreference resolution models through active learning. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7533–7549, Dublin, Ireland, May 2022. Association for Computational Linguistics.

Os Keyes, Chandler May, and Annabelle Carrell. You keep using that word: Ways of thinking about gender in computing research. *Proc. ACM Hum.-Comput.*

Interact., 5(CSCW1), April 2021.

Patrick Xia, Guanghui Qin, Siddharth Vashishtha, Yunmo Chen, Tongfei Chen, Chandler May, Craig Harman, Kyle Rawlins, Aaron Steven White, and Benjamin Van Durme. LOME: Large ontology multilingual extraction. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, pages 149–159, Online, April 2021. Association for Computational Linguistics.

Chandler May, Alex Wang, Shikha Bordia, Samuel R. Bowman, and Rachel Rudinger. On measuring social biases in sentence encoders. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 622–628, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

Benjamin Van Durme, Tom Lippincott, Kevin Duh, Deana Burchfield, Adam Poliak, Cash Costello, Tim Finin, Scott Miller, James Mayfield, Philipp Koehn, Craig Harman, Dawn Lawrie, Chandler May, Max Thomas, Annabelle Carrell, Julianne Chaloux, Tongfei Chen, Alex Comerford, Mark Dredze, Benjamin Glass, Shudong Hao, Patrick Martin, Pushpendre Rastogi, Rashmi Sankepally, Travis Wolfe, Ying-Ying Tran, and Ted Zhang. CADET: Computer assisted discovery extraction and translation. In *Proceedings of the IJCNLP 2017, System Demonstrations*, pages 5–8, Tapei, Taiwan, November 2017. Association for Computational Linguistics.

Rachel Rudinger, Chandler May, and Benjamin Van Durme. Social bias in elicited natural language inferences. In *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*, pages 74–79, Valencia, Spain, April 2017. Association for Computational Linguistics.

Chandler May, Kevin Duh, Benjamin Van Durme, and Ashwin Lall. Streaming word embeddings with the space-saving algorithm. *Arxiv preprint*, abs/1704.07463, 2017.

Chandler May, Ryan Cotterell, and Benjamin Van Durme. Analysis of morphology in topic modeling. *Arxiv preprint*, abs/1608.03995, 2016.

Chandler May, Francis Ferraro, Alan McCree, Jonathan Wintrode, Daniel Garcia-Romero, and Benjamin Van Durme. Topic identification and discovery on text and speech. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2377–2387, Lisbon, Portugal, September 2015. Association for Computational Linguistics.

Chandler May, Alex Clemmer, and Benjamin Van Durme. Particle filter rejuvenation and Latent Dirichlet Allocation. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages

446–451, Baltimore, Maryland, June 2014. Association for Computational Linguistics.

Chandler May, Michael Henry, Liam McGrath, Eric Bell, Eric Marshall, and Michelle Gregory. Semantic features for classifying referring search terms. In *The Pacific Northwest Regional NLP Workshop*, 2012.

Chandler May. Verification of solutions to the sensor location problem. Bachelor's thesis, Harvey Mudd College, Claremont, CA, 2011.

Talks

Chandler May. Deconstructing Gender Prediction in NLP. Queer in AI Workshop, Vancouver, BC, December 9, 2019.

Chandler May. On Measuring Social Biases in Sentence Encoders. Conference of the North American Chapter of the Association for Computational Linguistics, Minneapolis, MN, June 3, 2019.

Chandler May. Ethics in artificial intelligence (in context). CLSP Student Seminar, Baltimore, MD, September 5, 2017.

Chandler May. Topic Identification and Discovery on Text and Speech. Conference on Empirical Methods on Natural Language Processing, Lisbon, Portugal, September 21, 2015.

Chandler May. Topic Identification and Discovery on Text and Speech. CLSP Student Seminar, Baltimore, MD, September 8, 2015.

Chandler May. Compressive Sensing: Introduction, Implementation, and Applications. PNNL Candidate Presentation, Richland, WA, May 2, 2012.

Chandler May. Verification of Solutions to the Sensor Location Problem. Thesis Presentation in HMC Presentation Days, Claremont, CA, May 2, 2011.

Chandler May. Verification of Solutions to the Sensor Location Problem. Pacific Coast Undergraduate Mathematics Conference, Los Angeles, CA, March 12, 2011.

Service

2018–2019	Graduate representative, Diversity and inclusion committee, JHU CS
2017–2018	Reviewer, EMNLP Workshop on Noisy User-generated Text
2018	Reviewer, EMNLP; Best Reviewer Award (top 7%)
2018	Reviewer (secondary), NAACL-HLT
2017	Reviewer (secondary), EACL Workshop on Ethics in NLP
2017	Organizer, JHU CLSP Wiki Day
2014, 2017	Reviewer (secondary), TACL

2013–2017 Student volunteer, Student seminar committee, JHU CLSP
2016 Student volunteer, WiML
2013–2016 Coffee czar, JHU CS
2015 Reviewer (secondary), WWW
2014 Reviewer (secondary), EACL
2014 Student volunteer, ACL
2011–2013 Mentor, Delta High School Programming Club
2011–2013 Tutorial mentor, Delta High School
2012 Presenter, Delta High School STEMcon
2012 Judge, Delta High School Science and Engineering Fair
2008–2009 President, Harvey Mudd College Math Club
2007–2008 Webmaster, Harvey Mudd College Math Club