Kenton Murray

Kmurray4@nd.edu
www.kentonmurray.com

University of Notre Dame
Department of Computer
Science and Engineering

EDUCATION:

- **University of Notre Dame** South Bend, IN
  Ph.D. in Computer Science August 2014 - Present
  Advisor: David Chiang
  GPA: 4.0

- **Carnegie Mellon University School of Computer Science** Pittsburgh, PA
  M.S. Language Technologies August 2011 – August 2013
  Master's Thesis: "A Semantic Scan Statistic for Novel Disease Outbreak Detection"
  Advisor: Daniel Neill
  * Course Work Includes: Machine Learning, Machine Learning for Policy, Probabilistic Graphical
    Models, Language and Statistics, Social Media Analysis, Software Engineering, Algorithms for
    NLP, Grammars and Lexicons, Seminar on Endangered Languages

- **Princeton University** Princeton, NJ
  Major: B.S.E. in Computer Science September 2005 - June 2009
  Certificates (Minors): Robotics and Intelligent Systems, Finance
  Senior Thesis: **"**Summarization by Latent Dirichlet Allocation"
  Advisor: David Blei
  * Course Work Includes: Algorithms and Data Structures, Advanced Programming Techniques, Logic
    Design, Human Computer Interface, Robotics and Intelligent Systems, Corporate Finance, Micro
    and Macroeconomics, Programming Languages, Calculus through Multivariable, Linear Algebra,
    and Statistics

Awards and Honors:

- Best Poster Presentation – CMU Language Technologies Institute's Annual Student Research Symposium.
  Selected from a committee of Faculty and Students with expertise ranging across Machine Learning, NLP,
  Computer Science, and Language Technologies
- Best performing systems in spoken and written translation in both directions for Arabic/English Language
  Pair at IWSLT 2013.
- Masters Student Representative – Elected to be the sole representative for 50 Masters Students'. Voiced
  student concerns to the faculty. Held regular meetings regarding student's issues with classes.
- Sigma Xi Scientific Research Society Inductee for Senior Thesis on using Statistical NLP to Automatically
  Summarize Text Documents. Required Faculty Nomination. Currently one of the leading algorithms in the
  field.
- Won the Princeton Junior Orator Award for a Speech on Revolutionary America

RESEARCH:

Notre Dame NLP Lab August 2014 – Present
PhD Student P.I. David Chiang
Working on Statistical Machine Translation with a particular emphasis on morphologically rich languages and
languages in a low-resource setting. Focusing on developing more robust and easy to use neural models for research
in Machine Translation and broader NLP. In particular, addressing issues with understandability and ability to use
these methods without exhaustive parameter tuning and grid-search approaches.

Arabic Language Technologies Lab September 2013 – June 2014
Qatar Computing Research Institute P.I. Stephan Vogel

Research Associate
Working on Statistical Machine Translation. In particular, investigating statistical methods for phrase alignment and extraction. Specific focus on morophologically complex languages, such as Arabic. Developing open source phrase alignment toolkit for phrase based MT.

Event and Pattern Detection Lab   August 2011 – August 2013
CMU School of Computer Science   P.I. Daniel Neill
Graduate Research Assistant
Developed unsupervised ML techniques on free text in Hospital Emergency Department Records for Anomalous Pattern Detection. Expanded state-of-the-art pattern detection methods from reliance on expertly labeled data to natural language. Focused on robustness in graphical models and capturing low-frequency terms. Experiments include various Variational Inference Methods, Gibbs Samplers, EM, and corpus preprocessing steps. Received Full Graduate Fellowship for tuition and stipend equivalent to PhD candidates'.

Hierarchical Co-Clustering for Collaborative Filtering   Spring 2013
Class Project for Probabilistic Graphical Models
Investigated extensions of matrix factorization for collaborative filtering through hierarchical clustering. Proposed a new exponential family distribution for use in collaborative filtering tasks. Resulted in a Long Paper at WWW 2014. Joint work with Alex Beutel, Alex Smola, and Christos Faloutsos.

CMU Machine Translation Lab   Spring and Summer 2013
CMU's group focusing on MT. Submitted WMT 2013 paper. Part of reading and discussion group for Machine Translation. Discussed and presented work from recent publications.

Automated Essay Grading   Spring 2012
Class Project for Machine Learning
Used LASSO regularization for grading secondary school student essays after looking at numerous scoring methodologies. Examined numerous lexical and semantic features for scoring essays compared to gold-standard of expert human graders. Achieved performance on par with inter-annotator agreement. Joint work with Naoki Orii.

Crowdsourcing Augmented Language Models for Geotagging Tweets   Spring 2012
Class Project for Machine Learning for Policy
Enhanced data for language models using crowdsourcing. Augmented language models by assigning conditional priors to geographic locations from the use of specific. Inferred additional information by free text actively gathered through Mechanical Turk.

Automatic Text Summarization using Topic Models   September 2008 – May 2009
Independent Research Advised by David Blei
Investigated the use of Probabilistic Topic Models for the use in Summarizing Textual Documents. Investigated a non-standard summarization corpus. Evaluated using Crowdsourcing which had no prior applications in automatic summarization. Work resulted in induction into Sigma Xi.

PUBLICATIONS:

2015:
"Auto-Sizing Neural Networks: With Applications to n-gram Language Models" Kenton Murray and David Chiang. Proceedings of the International Conference on Empirical Methods in Natural Language Processing, September 2015, Lisbon, Portugal.

2014:
"Collaborative Bayesian Filtering" Alex Beutel, Kenton Murray, Alex Smola, and Christos Faloutsos. Proceedings of the International World Wide Web Conference, April 2014, Seoul, Korea.

2013:
"QCRI at IWSLT 2013: Experiments in Arabic-English and English-Arabic Spoken Language Translation". Hassan Sajjad, Fransico Guzman, Preslav Nakov, Ahmed Abdelali, Kenton Murray, Fahad Al Obaidli, Stephan Vogel. Proceedings of IWSLT. December 2013, Heidelberg, Germany. (Best Arabic-English and English-Arabic Systems)

"The CMU machine translation systems at WMT 2013: Syntax, synthetic translation options, and pseudo-references" Waleed Ammar, Victor Chahuneau, Michael Denkowski, Greg Hanneman, Wang Ling, Austin Matthews, Kenton Murray, Nicola Segall, Yulia Tsvetkov, Alon Lavie, Chris Dyer. Proceedings of WMT. August 2013, Sofia, Bulgaria.

"A Semantic Scan Statistic for Novel Disease Outbreak Detection". Kenton Murray and Daniel Neill. Carnegie Mellon University Master's Thesis (available online). August 2013, Pittsburgh, PA.

2009:
"Summarization by Latent Dirichlet Allocation". Kenton Murray and David Blei. Princeton University Senior Thesis (available online and through Princeton University Libraries). May 2009, Princeton, NJ.

PROFESSIONAL EXPERIENCE:

NLP Start-up IP Consultant *Pittsburgh, PA*   Summer 2014
Hired by a lawfirm to compare proprietary and open-source code bases for potential IP overlaps. Evaluated NLP pipelines in Java and Python for potential derivative works. Wrote a report as the technical expert for the due diligence of a start-up acquisition.

Safaba Translation Services *Software Developer (Part-time) | Pittsburgh, PA*   July 2012 – July 2013
Developed Machine Translation Service for Fortune 500 companies using Moses providing translations for dozens of languages. Worked on proprietary MT software that improves state of the art translation for domain specific applications. Developed service monitoring framework.

Lighter Capital *Software Development Engineer Intern | Seattle, WA*   Summer 2011
Implemented Build Automation System, Started Proper QA protocols. Sole developer of a proprietary asset pricing software for mitigating risk in firm's investment portfolio and to determine loan conditions to small businesses.

Microsoft *Software Development Engineer in Test | Bellevue, WA*   Summer Intern: 2008, Full-Time: 2009 - 2010
Dynamics CRM Online. Part of team developing the backend of a cloud service. Sole tester on the Backup Service for all customer data running every 10 minutes to maintain SLA. Authored determining report for the GM that led to expanding product to 41 languages and 40 countries after evaluating feasibility affecting business strategy for the following year. Online Test Expert in Upgrade and Patching Scenarios for the product. In charge of deploying new releases to Asian and European Data Centers. Programmed Automated Test Suites in C# and the .Net used daily on the latest build (BVTs). Focused on preventing security flaws and viruses, consumer ease of use, and stress testing.

LECTURES and TEACHING:

University of Notre Dame Computer Science NLP Class *TA/Guest Lecturer | South Bend, IN*   Spring 2015
Sole TA for an upper-level undergraduate/graduate class on Natural Language Processing. Held weekly office hours, graded and helped write assignments and tests, as well as helped develop with course structure. Wrote required background readings, and gave lectures for the classes focusing on CRFs and Linear Regression.

Carnegie Mellon University Materials Science Department *Guest Lecturer | Pittsburgh, PA*   April 1, 2013
Guest Lecturer on Introduction to Text Mining. Course on basic machine learning techniques on impacts of material for advanced undergraduates (27-566). Covered basic challenges and methods for NLP and the impacts of statistics within corpora.

Carnegie Mellon University Heinz College *Guest Lecturer | Pittsburgh, PA*   December 2011 and 2012
Gave guest lectures to a Master's Level Course for Public Policy Students (Large Scale Data Analysis). Presented current state-of-the-art methods for using ML techniques on Natural Language Data for Public Health Surveillance.

Princeton Computer Science Department *Laboratory TA, Grader | Princeton, NJ*   Spring 2008 - Spring 2009
Held weekly lab hours for students in all 100 and 200 level computer science classes to help with programming assignments. Taught programming techniques, debugging, and general subject knowledge. Graded assignments.

SOFTWARE:

NPLM – Contribute to the NPLM toolkit for continuous space language models. Added functionality for group regularizers for determining hidden-layer sizes.

Bevara – An Android Application for Corpus Collection. Uses camera, microphone, and other hardware on android devices to capture a wide range of data for corpora of interest to computational linguistics. Automatically inserts meta-data into capture process for producing higher quality scientific datasets. Tailored to address low-resource and endangered natural languages. Currently in private Beta. Source available on Github.

MISCELLANEOUS:
- Programming Languages: Java, C#, C, C++, AWK, Python, .NET, PHP, SQL, Haskell
- Princeton Varsity Men's Crew Team (2 Letters)   2005 – 2007
- Princeton Charter Eating Club Officer (Managed $30,000 Social Budget)   2007 - 2009