

Sam Anzaroot

CONTACT INFORMATION	2084 Ocean Parkway. Brooklyn, NY 11223 USA <i>Mobile:</i> +1-718-207-2887 <i>E-mail:</i> samanzaroot@gmail.com	<i>Github:</i> github.com/samanz <i>Web:</i> http://www.samanz.com/about
PROFESSIONAL INTERESTS	Machine Learning, Natural Language Processing, Information Extraction, Graphical Models, Scalability, Data Science	
EDUCATION	University of Massachusetts Amherst , Amherst, MA M.S. Computer Science, February 2014 <ul style="list-style-type: none">• Advisor: Professor Andrew McCallum• Area of Study: Machine Learning, NLP, Information Extraction Queens College - City University of New York , Flushing, NY B.S., Computer Science, June 2011 <ul style="list-style-type: none">• <i>Magna cum Laude</i>, with Honors• Minor in Mathematics• Advisor: Professor Heng Ji	
EXPERIENCE	Research Intern at PLRG lab at Oracle Labs East , 2/2014-6/2014 <ul style="list-style-type: none">• Investigated methods of highly parallel probabilistic inference on conditional random fields (CRFs) using Nvidia GPUs.• Created a GPU version of the belief propagation algorithm written in CUDA. I've optimized this implementation to allow for 200x speedup in inference and 100x speedup in training of CRFs over sequential implementation. Graduate Research Assistant at IESL Lab at the University of Massachusetts Amherst with Dr. Andrew McCallum , 9/2011-2/2014 <ul style="list-style-type: none">• Conducted research in the area of machine learning, using graphical models with applications in natural language processing and information extraction.• Built state-of-the-art named entity recognition model.• Organized the creation of a new dataset for segmenting and labeling parts of citations in research papers for automatic extraction.• Investigated methods of adding global information for inference in conditional random fields using extensions to Lagrange relaxation methods with applications in citation extraction, retrieving new state-of-the-art results on the task.• Contributed to Factorie, an open-source machine learning and NLP toolkit. Undergrad Research Assistant at BLENDER Lab at City University of New York with Dr. Heng Ji , 8/2010-5/2011 <ul style="list-style-type: none">• Improved relation extraction by jointly performing inference over a large amount of automatically collected relations. This project worked by detecting unlikely collections of relations in a network and formulating these collections as constraints. The project then proceeded to find the best configuration of relations that satisfied these learned constraints using integer linear programming (ILP) methods.• Created interactive web-based demo allowing the browsing of relational information networks.	

Undergrad Research Assistant at Speech Lab at City University of New York with Dr. Andrew Rosenberg, 1/2011-9/2011

- Worked on improving HTK (speech recognition software) hidden Markov model (HMM) training in order to provide transcription services over the web.
- Created a system in Adobe Flex for the capture and storage of speech data over the web.

SimStudent REU (Research Experience for Undergraduates) Researcher at Carnegie Mellon University Human Computer Interaction Lab (CMU HCII) with Dr. Noboru Matsuda, 6/2010-8/2010

- Worked on SimStudent project which uses a first order inductive logic machine learning algorithm (FOIL) in a learning-by-demonstration setting to simulate methods in which students learn rules, specifically in the context of learning algebra.
- Evaluated the system to determine which background knowledge and order of learning operators causes the system to most closely match student errors. Furthermore, I modified the machine learning algorithm to allow for selective generalization in rules generated for the selection of working memory elements to examine if such changes allowed for better error matching.

MetroBotics REU Researcher at City University of New York with Dr. Elizabeth Sklar and Dr. Simon Parsons, 9/2009-6/2010

- Created a player/stage driver for the Scribbler + Fluke robotic system.
- Implemented object classification vision algorithms using openCV and machine learning classifiers for use with low-bandwidth camera data connections.

REU Program at Texas A&M with Dr. Yoonsuck Choe and Dr. John Keyser, 6/2009-8/2009

- Created a web-based map data viewer similar to Google Maps for deployment over the web for viewing the volumetric tissue data.
- Implemented a new method of annotating three-dimensional structures within web-based interface.

PUBLICATIONS

- [1] **Sam Anzaroot**, Alexandre Passos, David Belanger, Andrew McCallum. Learning Soft Linear Constraints with Application to Citation Field Extraction. In: *Proc. the 52nd Annual Meeting of the Association for Computational Linguistics (ACL2014)*. 2014
- [2] **Sam Anzaroot**, Andrew McCallum. A New Dataset for Fine-Grained Citation Field Extraction. In: *ICML Workshop on Peer Reviewing and Publishing Models (PEER)* 2013
- [3] Qi Li, **Sam Anzaroot**, Wen-Pin Lin, Xiang Li and Heng Ji. Joint Inference for Crossdocument Information Extraction. In: *Proc. 20th ACM Conference on Information and Knowledge Management (CIKM2011)*. 2011
- [4] Matthew Snover, Xiang Li, Wen-Pin Lin, Zheng Chen, Suzanne Tamang, Mingmin Ge, Adam Lee, Qi Li, Hao Li, **Sam Anzaroot**, Heng Ji. Cross-lingual Slot Filling from Comparable Corpora. In: *Proc. ACL2011 Workshop on Building and Using Comparable Corpora*. 2011

- [5] Elizabeth Sklar, Simon Parsons, Susan Epstein, Arif T. Ozgelen, George Rabanca, **Sam Anzaroot**, Joel Gonzalez, Jesse Lopez, Mitch Lustig, Linda Ma, Mark Manashirov, J. Pablo Munoz, S. Bruno Salazar, Miriam Schwartz. Developing a Framework for Team-based Robotics Research. In: *AAAI 2010 Robotics Exhibition and Workshop*. 2011
- PRESENTATIONS [6] **Sam Anzaroot**, Javier Artilles, Hao Li, Qi Li, Zheng Chen, Suzanne Tamang, Heng Ji, Hongbo Deng, Jiawei Han. Search, Mining and Browsing Self-Boosting Multi-Dimensional Text-Rich Information Networks In *The Network Science Collaborative Technology Alliance Annual Meeting*. 2011
- PROGRAMMING LANGUAGES **Expert:**
- Scala, Java, JavaScript, MATLAB, Python, PHP
- Proficient:**
- C++, C, Ruby
- Some experience:**
- Perl, C#, LISP, Haskell
- TECHNOLOGIES
- Git, Mercurial, Maven, SBT, MongoDB, SQL, CUDA, Linux, Apple OS X, HTML, CSS, JSON, XML, L^AT_EX
- LANGUAGES
- English, Hebrew (intermediate)
- GRADUATE LEVEL COURSEWORK **UMass Amherst:**
- Machine Learning
 - Statistical Inference I
 - Automated Knowledge Based Construction
 - Graphical Models
 - Research Methods
 - Algorithms
 - Advanced Databases
 - Distributed Operating Systems
 - Optimization
 - Artificial Intelligence
- Queens College:**
- Natural Language Processing
 - Bioinformatics
 - Machine Learning
 - Human Computer Interaction