ARI KOBREN

21 Brookline St. Unit 108 ⋄ Cambridge, MA 02139
(202) · 681 · 2494 ⋄ ari.kobren@gmail.com ⋄ https://akobre01.github.io/

EDUCATION

University of Massachusetts Amherst, Amherst, MA *Sept. 2012 - Present*
M.S./Ph.D. in Computer Science

Tufts University, Medford, MA *Sept. 2006 - May 2010*
Bachelor of Science, Computer Science Engineering
Summa Cum Laude, Dean's List all semesters, GPA: 3.78/4.0

HONORS & AWARDS

UMass Search and Mining Award, for 2 students in data mining, September 2017
NSF Graduate Research Fellowship, for graduate study in STEM fields, March 2012
Paul Utgoff Memorial Scholarship, for 1 student in machine learning, May 2012

EXPERIENCE

University of Massachusetts Amherst Sept. 2012 - Present
Research Assistant, Information Extraction and Synthesis Laboratory *Amherst, MA*

· Develop machine learning methods for clustering, entity resolution and knowledge base population and maintenance (under supervision of Prof. Andrew McCallum)
· Mentor master's and undergraduate students regarding research projects

Google Inc. June 2016 - Sept. 2016
Ph.D. Research Intern *New York City, NY*

· Developed TensorFlow-based probabilistic model for location-attribute prediction
· Improved upon current production model by $> 5\%$ F1 accuracy

Google Inc. June 2014 - Sept. 2014
Ph.D. Research Intern *Mountain View, CA*

· Improved user retention in the Quizz.us crowdsourcing platform by more than 70%
· Developed and trained models of user interaction with the system (contributed production code)

MIT Lincoln Laboratory Sept. 2010 - Sept. 2012
Researcher *Lexington, MA*

· Developed natural language processing tools to support US intelligence analysts
· Obtained United States security clearance (Secret-Level)

FindJoey Sept. 2008 - May 2010
Co-Founder, Developer, CTO *Medford, MA*

· Co-founded company providing service for users to receive departure times for campus shuttle via SMS
· Achieved over 250,000 uses by more than 5000 unique users between 2008-2012

Tufts Computer Science Dept. Feb. 2009 - Sept. 2009
Researcher *Medford, MA*

· Developed modifications of Bloom Filters for Bayesian Classifier (http://osbf-lua.luaforge.net)

PUBLICATIONS

[1] Michael L. Wick, Ari Kobren, and Andrew McCallum. Large-scale author coreference via hierarchical entity representations. In *ICML Workshop: Peer Reviewing and Publishing Models (PEER)*, Atlanta, Georgia, USA, Jun 2013.

[2] Michael L. Wick, Sameer Singh, Ari Kobren, and Andrew McCallum. Assessing confidence of knowledge base content with an experimental study in entity resolution. In *CIKM Workshop on Automated Knowledge Base Construction (AKBC)*, San Francisco, Californa, USA, October 2013. **(Selected for oral presentation)**.

[3] Michael L. Wick, Ari Kobren, and Andrew McCallum. Probabilistic reasoning about human edits in information integration. In *ICML Workshop: Machine Learning Meets Crowdsourcing*, Atlanta, Georgia, USA, Jun 2013.

[4] Sameer Singh, Limin Yao, David Belanger, Ari Kobren, Sam Anzaroot, Michael Wick, Alexandre Passos, Harshal Pandya, Jinho Choi, Brian Martin, and Andrew McCallum. Slot filling by kb construction and universal schema. In *Text Analysis Conference on Knowledge Base Population*, Gaithersburg, Maryland, USA, October 2013.

[5] Ari Kobren, Thomas Logan, Siddarth Sampangi, and Andrew McCallum. Domain specific knowledge base construction via crowdsourcing. In *Advances in Neural Information Processing Systems Workshop: Automated Knowledge Base Construction*, Montreal, Canada, Dec 2014. **(Outstanding Paper Award, Selected for oral presentation)**.

[6] Ari Kobren, Chun How Tan, Panos Ipeirotis, and Evgeniy Gabrilovich. Getting more for less: Optimized crowdsourcing with dynamic tasks and goals. In *Proceedings of the 24th International Conference on the World Wide Web*, Florence, Italy, May 2015.

[7] Ari Kobren*, Nicholas Monath*, Akshay Krishnamurthy, and Andrew McCallum. A hierarchical algorithm for extreme clustering. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 255–264. ACM, 2017 **(Oral Presentation)**.

[8] Ari Kobren, Nicholas Monath, and Andrew McCallum. Entity-centric attribute feedback for interactive knowledge bases. In *NIPS 17' Workshop on Automated Knowledge Base Construction*, Long Beach, California, December 2017.

[9] Nicholas* Monath, A.* Kobren, Akshay Krishnamurthy, and Andrew McCallum. Gradient-based hierarchical clustering. In *NIPS 17' Workshop on Discrete Structures in Machine Learning*, Long Beach, California, December 2017 **(Oral Presentation)**.

[10] Craig Greenberg, Nicholas Monath, Ari Kobren, Patrick Flaherty, Andrew McGregor, and Andrew McCallum. Compact representation of uncertainty in clustering. In *NeurIPS*, Montreal, Canada, 2018.

[11] Ari Kobren, Nicholas Monath, and Andrew McCallum. Integrating user feedback under identity uncertainty in knowledge base construction. In *Automated Knowledge Base Construction*, Amherst, MA, 2019.

[12] Nishant Yadav, Ari Kobren, Nicholas Monath, and Andrew McCallum. Supervised hierarchical clustering with exponential linkage. In *International Conference on Machine Learning*, Long Beach, CA, 2019.

[13] Ari Kobren, Barna Saha, and Andrew McCallum. Paper matching with local fairness constraints. In *International Conference on Knowledge Discovery and Data Mining*, Anchorage, Alaska, 2019 **(Oral Presentation)**.

[14] Nicholas Monath*, Ari Kobren*, Akshay Krishnamurthy, Michael Glass, and Andrew McCallum. Scalable hierarchical clustering via tree grafting. In *International Conference on Knowledge Discovery and Data Mining*, Anchorage, Alaska, 2019 **(Oral Presentation)**.

[15] Ari Kobren, Pablo Bario, Oksana Yakhnenko, Johann Hibschman, and Ian Langmore. Constructing high precision knowledge bases with subjective and factual attributes. In *International Conference on Knowledge Discovery and Data Mining*, Anchorage, Alaska, 2019.

[16] Derek Tam, Nicholas Monath, Ari Kobren, Aaron Traylor, Rajarshi Das, and McCallum Andrew. Optimal transport-based alignment of learned character representations for string similarity. In *Annual Meeting of the Association of Computational Linguistics*, Florence, Italy, 2019.