Curriculum Vitae

Name: Kedar Bellare

Date of Birth: 04 October, 1982

Residential Address:
4355 Renaissance Dr, #104,
San Jose, CA 95134, USA

Office Address:
Yahoo! Research,
4401 Great America Pkwy,
Santa Clara, CA 95054, USA

Phone number:
(413)-265-0343 (cell)
(413)-545-3616 (office)

E-mail:
kedar.bellare@gmail.com
kedar.bellare@yahoo.com
kedarb@cs.umass.edu

Current VISA Status: H-1B visa (April 14, 2011 – April 15, 2014)

Background:

- Education:
  - Ph.D (Doctor of Philosophy) from University of Massachusetts, Amherst (*All qualifications completed on June 2011. Dissertation defense date to be scheduled*).
  - M.S. (Master of Science) from University of Massachusetts, Amherst. (*February 2008*).
  - B.Tech (Bachelor of Technology) from IIT Bombay in Computer Science and Engineering. (*August 2004*).
- Advisor: Prof. Andrew McCallum/Information Extraction and Synthesis Lab (IESL)
  - RA positions: Sept. 2004-June 2011 in IESL lab.
  - TA positions: Computational Social Network Analysis (Taught by David Jensen and Andrew McCallum in Fall '08).

Research:

- Paramveer Dhillon, Sathiya Keerthi, Kedar Bellare, Olivier Chapelle and S. Sundararajan. *Deterministic Annealing for Semi-Supervised Structured Output Learning.* International Conference on Artificial Intelligence and Statistics (AISTATS) 2012,La Palma, Canary Islands (AISTATS '12), 2012.
- Michael Wick, Khashayar Rohanimanesh, Kedar Bellare, Aron Culotta and Andrew McCallum. *SampleRank: training factor graphs with atomic gradients.* International Conference on Machine Learning (ICML) 2011, Bellevue, Washington, USA (ICML '11), 2011.
- Kedar Bellare and Andrew McCallum. *Generalized Expectation Criteria for Bootstrapping Extractors using Record-Text Alignment.* Empirical Methods in NLP (EMNLP) 2009, Singapore (EMNLP'09), 2009.
- Kedar Bellare, Gregory Druck and Andrew McCallum. *Alternating Projections for Learning with Expectation Constraints.* Uncertainty in Artificial Intelligence (UAI) 2009, Montreal, QC, Canada (UAI'09), 2009.
- Kedar Bellare, Koby Crammer and Dayne Freitag. *Loss-Sensitive Discriminative Training of Machine Transliteration Models.* Student Research Workshop at HLT-NAACL 2009, Boulder, CO, USA (HLT-NAACL SRW'09), 2009.
- Kedar Bellare, Partha Pratim Talukdar, Giridhar Kumaran, Fernando Pereira, Mark Liberman, Andrew McCallum and Mark Dredze. *Lightly-Supervised Attribute Extraction for Web Search.* NIPS workshop on Machine Learning for Web Search at NIPS 2007, Vancouver, BC, Canada (NIPS MLWS'07), 2007.
- Kedar Bellare and Andrew McCallum. *Learning Extractors from Unlabeled Text using Relevant Databases.* IIWeb workshop at AAAI 2007, Vancouver, BC, Canada (IIWeb'07), 2007.
- Kuzman Ganchev, Koby Crammer, Fernando Pereira, Gideon Mann, Kedar Bellare, Andrew McCallum, Steve Carroll, Yang Jin, and Pete White. *Penn/Umass/CHOP BiocreativeII Systems.* Proceedings of the Biocreative II Workshop, pp. 119 - 124, 2007.
- Andrew McCallum, Kedar Bellare and Fernando Pereira. *A Conditional Random Field for Discriminatively-trained Finite-state String Edit Distance.* Proceedings of the International Conference on Uncertainty in AI (UAI'05), 2005.
- Kedar Bellare, Anish Das Sarma, Atish Das Sarma, Navneet Loiwal, Vaibhav Mehta, Ganesh Ramakrishnan, Pushpak Bhattacharya. *Generic Text Summarization Using WordNet.* Proceedings of the International Conference on Language Resources and Evaluation (LREC'04), 2004.

Skills:

- **Programming:** Scala, Java, Perl, Python, Matlab, R, JSP with some knowledge of C.
- **Frameworks & Software:** MongoDB, Akka Toolkit, Lucene, Google Web Toolkit, Hadoop, Pig.
- **Open-source Projects:** Contributed code in the following open-source projects:
  - Mallet (http://mallet.cs.umass.edu/): A java-based package for statistical natural language processing, document classification, clustering, topic modeling, information extraction, and other machine learning applications to text. Worked on Conditional Random Fields (CRFs) and their applications.

- Factorie (http://factorie.cs.umass.edu/): A scala-based toolkit for large-scale probabilistic modeling. Worked on scalable learning and inference of large-scale factor graphs.
- Dynprog (https://github.com/kedarbellare/dynprog): A scala-based framework for implementing various dynamic programming learning and inference algorithms based on the abstraction of hypergraphs.
- Entizer (https://github.com/kedarbellare/entizer): A scala-based framework for data integration and information extraction from structured, semi-structured and unstructured text sources. Uses various technologies like MongoDB, Akka actors, Memcache, Lucene to achieve scalability.

Experience:

- **Yahoo! Research**.
  Scientist in the Web Information Management group.
  Managers: Philip Bohannon and Raghu Ramakrishnan.
  June 2011 to Present.
  Work on large-scale information extraction and entity de-duplication problems. Also working on incorporating domain knowledge into models for machine learning.
- **Yahoo! Research**.
  Summer Intern under Sathiya Keerthi and Ashwin Machanavajjhala.
  Summer 2009.
  Worked on a Bayesian network model for performing joint unsupervised information extraction and record linkage.
- **Yahoo! Research**.
  Summer Intern under Sathiya Keerthi and Srujana Merugu.
  Summer 2008.
  Worked on rapid development of classifiers using labeling of instances and features. Also studied active learning using instance/feature labeling.
- **FairIsaac Corporation**.
  Summer Intern under Dayne Freitag.
  Summer 2007.
  Worked on applying on-line machine learning algorithms to sequence/tree edit distance problems.
- **University of Pennsylvania**.
  Fall Intern under Prof. Mark Liberman and Prof. Fernando Pereira.
  Fall 2006.
  Worked on the problem of lightly-supervised attribute extraction.
- **University of Massachusetts, Amherst**.
  Research Assistant under Prof. Andrew McCallum.
  Sept 2004 - June 2011.
  Worked on problems related to information extraction and data integration. Primarily focused on using alignments between knowledge bases and text corpora to automatically induce extractors.

- **University of British Columbia**.
  Summer Intern under Prof. Laks Lakshmanan and Prof. Raymond Ng.
  May - July 2003.
  Worked on algorithms for schema integration of disparate XML databases.

Service:

- Professional:
  - Reviewed conference papers submitted to NIPS'06, HLT-NAACL'07, ACL'07, KDD '07, IIWeb '07, ICML'08, ICML '09, UAI '09, KDD '09, EMNLP '09, ICML '10, ICML '12, EMNLP '12.
  - Reviewed journal papers for the Special Issue of Applied AI on applications of Grammatical Inference, Transactions on Database Systems (TODS), Transactions on Knowledge and Data Engineering (TKDE).
- Laboratory:
  - Helped in the creation of the Rexa search engine, a research paper search engine developed by our lab (IESL).

References:

- **Andrew McCallum**
  Department of Computer Science
  University of Massachusetts Amherst
  Amherst, MA 01003
  E-mail: mccallum@cs.umass.edu
- **Philip Bohannon**
  Yahoo! Research,
  Santa Clara, CA
  E-mail: philbonj@yahoo.com
- **Sathiya Keerthi**
  Yahoo! Research,
  Santa Clara, CA
  E-mail: geethakee@yahoo.com
- **Srujana Merugu**
  IBM Research,
  New Delhi, India
  E-mail: srujanamerugu@in.ibm.com
- **Dayne Freitag**
  SRI International,
  San Diego, CA
  E-mail: freitag@ai.sri.com