Aron Culotta

University of Massachusetts
Department of Computer Science
140 Governor's Drive
Amherst, MA 01003

(413) 222-0775 (cell)
(413) 545-3616 (office)
culotta@cs.umass.edu
http://www.cs.umass.edu/~culotta

Research Interests

Designing statistical machine learning algorithms to discover knowledge from text.

[natural language processing, information extraction, coreference resolution, data mining, graphical models, conditional random fields, weighted logic]

Education

University of Massachusetts at Amherst Amherst, MA
Ph.D., Computer Science, 2008 (expected)

University of Massachusetts at Amherst Amherst, MA
M.S., Computer Science, 2004

Tulane University New Orleans, LA
B.S., Computer Science (Math minor), *summa cum laude*, 2002

Work History

Research Assistant September 2002 - present
University of Massachusetts Amherst, MA

Research in machine learning, information extraction, and knowledge discovery in the *Information Extraction and Synthesis Lab* with Professor Andrew McCallum.

Research Intern June 2007 - August 2007
Microsoft Research Redmond, WA

Designed statistical machine learning algorithm to extract and synthesize information from search results.

Research Intern September 2005 - December 2005
Google, Inc. New York, NY

Designed machine learning algorithm to combine relation extraction and knowledge discovery from Wikipedia documents.

Research Intern June 2003 - September 2003
International Business Machines Yorktown Heights, NY

Developed novel support vector machine algorithm to extract relations between people and organizations in newswire text.

Research Assistant June 2001 - September 2001
University of Alabama Huntsville, AL

Optimized memory allocation algorithms for Java's Virtual Machine. Program sponsored by National Science Foundation.

Technology Consultant January 2007 - present
IT.com Washington, DC

Designed and implemented large-scale statistical topic models for knowledge discovery from email data.

Co-owner November 1999 - September 2002
Home Cyber Guide New Orleans, LA

Provided computer training, trouble-shooting, and consulting to clients in greater New Orleans. In addition to technical work, also contributed to employee recruitment, advertising, and lesson development.

Instructor September 1999 - May 2002
Princeton Review New Orleans, LA

Taught test preparation classes for the SAT, ACT, and GRE.

Honors

- Microsoft Live Labs Fellow, 2006-2008
- Passed Ph.D. Qualifier with Distinction, 2005
- "Best Paper Honorable Mention," Proceedings of AAAI, 2004
- Dean's Honor Scholarship, Tulane University, 1998-2002

Teaching Experience

- Guest lecturer and teaching assistant for Introduction to Natural Language Processing with Prof. Andrew McCallum, Fall and Spring semesters 2006, University of Massachusetts.

Skills

Languages (in order of decreasing experience): Java, Perl, C++, C, HTML, Python, R, Matlab, C#, MySQL, XQuery, Flash, Prolog, Occam, Visual Basic

Operating Systems: Mac OS X, Linux, Windows, Solaris

Course work:

Machine Learning	Advanced Machine Learning
Statistical Information Extraction	Information Retrieval
Computational Social Network Analysis	Reasoning and Acting under Uncertainty
Bioinformatics	Multivariate Data Analysis
Advanced Algorithms	Computational Complexity
Database Design and Implementation	Research Methods for Empirical Computer Science

Service Activities

Professional Service

- **Program Committee:** NAACL-HLT, 2007; Workshop on Learning Structured Information in Natural Language Applications, EACL 2006; Northeast Student Colloquium on Artificial Intelligence, 2007; AAAI 2008 Workshop on Enhanced Messaging; AAAI 2008
- **Reviewer:** AAAI, ACL, CoNLL, HLT, ICML, IEEE Trans. on Knowledge Engineering, IJCAI, JMLR, KDD, NIPS, UAI.
- **Webmaster:** NIPS Online proceedings, 2004-2007.

Departmental Service

- **Graduate student representative, 2006**: Participate in faculty meetings, interview faculty candidates, mediate student-department issues.
- **Organizer, Machine Learning Seminar, 2004-2006**: Helped schedule and organize a continuing departmental lunch seminar series on machine learning topics.

Publications

Journal Publications

Aron Culota, Trausti Kristjansson, Andrew McCallum, and Paul Viola. Corrective feedback and persistent learning for information extraction. *Artificial Intelligence*, 170:1101–1122, 2006.

Conference Publications

Aron Culotta, Michael Wick, Robert Hall, Matthew Marzilli, and Andrew McCallum. Canonicalization of database records using adaptive similarity measures. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, San Jose, CA, 2007. (19% accepted).

Aron Culotta, Michael Wick, Robert Hall, and Andrew McCallum. First-order probabilistic models for coreference resolution. In *Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics (HLT/NAACL)*, pages 81–88, 2007. (24% accepted).

Michael Wick, Aron Culotta, and Andrew McCallum. Learning field compatibilities to extract database records from unstructured text. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 603–611, Sydney, Australia, 2006. (18% accepted; 6 citations in Google Scholar).

Aron Culotta, Andrew McCallum, and Jonathan Betz. Integrating probabilistic extraction models and data mining to discover relations and patterns in text. In *Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics (HLT/NAACL)*, pages 296–303, New York, NY, June 2006. (25% accepted; 17 citations in Google Scholar).

Aron Culotta and Andrew McCallum. Joint deduplication of multiple record types in relational data. In *2005 ACM CIKM International Conference on Information and Knowledge Management*, pages 257–258, 2005. (25% accepted; 6 citations in Google Scholar).

Aron Culotta and Andrew McCallum. Reducing labeling effort for stuctured prediction tasks. In *The Twentieth National Conference on Artificial Intelligence (AAAI)*, pages 746–751, Pittsburgh, PA, 2005. (27% accepted; 6 citations in Google Scholar).

Aron Culotta and Jeffery Sorensen. Dependency tree kernels for relation extraction. In *42nd Annual Meeting of the Association for Computational Linguistics (ACL)*, Barcelona, Spain, 2004. (25% accepted; 86 citations in Google Scholar).

Trausti Kristjannson, Aron Culotta, Paul Viola, and Andrew McCallum. Interactive information extraction with constrained conditional random fields. In *Nineteenth National Conference on Artificial Intelligence (AAAI)*, San Jose, CA, 2004. (26% accepted; Best Paper Honorable Mention; 27 citations in Google Scholar).

Aron Culotta and Andrew McCallum. Confidence estimation for information extraction. In *Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT/NAACL)*, Boston, MA, 2004. (26% accepted; 31 citations in Google Scholar).

Aron Culotta, Ron Bekkerman, and Andrew McCallum. Extracting social networks and contact information from email and the web. In *First Conference on Email and Anti-Spam (CEAS)*, Mountain View, CA, 2004. (35% accepted; 47 citations in Google Scholar).

Refereed Workshop Publications

Aron Culotta, Pallika Kanani, Robert Hall, Michael Wick, and Andrew McCallum. Author disambiguation using error-driven machine learning with a ranking loss function. In *Sixth International Workshop on Information Integration on the Web (IIWeb-07)*, Vancouver, Canada, 2007.

Aron Culotta and Andrew McCallum. Tractable learning and inference with high-order representations. In *International Conference on Machine Learning Workshop on Open Problems in Statistical Relational Learning*, Pittsburgh, PA, 2006.

Aron Culotta and Andrew McCallum. Practical markov logic containing first-order quantifiers with application to identity uncertainty. In *Human Language Technology Workshop on Computationally Hard Problems and Joint Inference in Speech and Language Processing (HLT/NAACL)*, June 2006.

Aron Culotta and Andrew McCallum. Learning clusterwise similarity with first-order features. In *Neural Information Processing Systems (NIPS) Workshop on the Theoretical Foundations of Clustering*, Whistler, B.C., December 2005.

Unrefereed Workshop Publications

Aron Culotta, Andrew McCallum, Bart Selman, and Ashish Sabharwal. Sparse message passing algorithms for weighted maximum satisfiability. In *New England Student Colloquium on Artificial Intelligence (NESCAI)*, Ithaca, NY, 2007.

Technical Reports

Aron Culotta and Andrew McCallum. A conditional model of deduplication for multi-type relational data. Technical Report IR-443, University of Massachusetts, September 2005.

Aron Culotta, David Kulp, and Andrew McCallum. Gene prediction with conditional random fields. Technical Report UM-CS-2005-028, University of Massachusetts, Amherst, April 2005. (10 citations in Google Scholar).

Aron Culotta. Maximizing cascades in social networks. Technical report, University of Massachusetts, 2003.

References

Available upon request.