

CONTACT INFORMATION

8133 Gates-Hillman Complex
Carnegie Mellon University
Pittsburgh, PA 15213 USA

vox: 608-332-3456
email: bsettles@cs.cmu.edu
web: burrsettles.com

EDUCATION AND TRAINING

Carnegie Mellon University , Pittsburgh, PA Postdoctoral Fellow, Machine Learning Department	2009–present
University of Wisconsin , Madison, WI Postdoctoral Fellow, Biostatistics and Medical Informatics	2009
University of Wisconsin , Madison, WI PhD, Computer Sciences (minors: Biology, Linguistics)	2008
DePauw University , Greencastle, IN BA, Computer Science (minor: Studio Art), <i>cum laude</i> with distinction	2000

RESEARCH INTERESTS

I develop machine learning technology that is “active” (creative, curious, or exploratory, if you will) and interacts with humans as part of knowledge acquisition and training. I also strive to put these technologies to use in natural language processing, biological discovery, and social computing applications.

PUBLICATIONS

Books and Book Chapters

- B. Settles. *Active Learning*. Morgan & Claypool, 2012 (in revision).
- B. Settles. Algorithms for active learning. In B. Krishnapuram, S. Yu, and R.B. Rao (Eds.), *Cost-Sensitive Machine Learning*, pages 3–32. Chapman and Hall, 2011.
- B. Settles. A software tool for biomedical information extraction (and beyond). In V. Prince and M. Roche (Eds.), *Information Retrieval in Biomedicine: Natural Language Processing for Knowledge Integration*, pages 326–335. IGI Global Press, 2009.

Journal Articles¹

- B.C. Smith, B. Settles, W.C. Hallows, M.W. Craven, and J.M. Denu. SIRT3 substrate specificity determined by peptide arrays and machine learning. *ACS Chemical Biology*, 6(2):146–157. 2011. (cover article; impact factor: 5.1)
- B. Settles. ABNER: An open-source tool for automatically tagging genes, proteins, and other entity names in text. *Bioinformatics*, 21(14):3191–3192. 2005. (impact factor: 4.3)

Conference Papers

- B. Settles and X. Zhu. Behavioral factors in interactive training of text classifiers. In *Proceedings of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL HLT)*. Short paper, to appear. ACL Press, 2012. (35%, 27% for oral presentation)
- B. Settles. Closing the loop: Fast, interactive semi-supervised annotation with queries on features and instances. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1467–1478. ACL Press, 2011. (23%, 15% for oral presentation)
- M. Burke and B. Settles. Plugged in to the community: Social motivators in online goal-setting groups. *Proceedings of the International Conference on Communities & Technologies (C&T)*, pages 1–10. ACM Press, 2011. (29%)

¹Publications are listed in reverse chronological order. Known journal impact factors and conference acceptance rates are shown in parentheses. Up-to-date citation counts, *h*-index, etc. can be found on my [Google scholar profile](#).

- E. Law, B. Settles and T.M. Mitchell. Learning to tag from open vocabulary labels. In *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD)*, pages 211–226. Springer, 2010. (16%)
- A. Carlson, J. Betteridge, B. Kisiel, B. Settles, E.R. Hruschka Jr. and T.M. Mitchell. Toward an architecture for never-ending language learning. In *Proceedings of the Conference on Artificial Intelligence (AAAI)*, pages 1306–1313. AAAI Press, 2010. (26%)
- G. Druck, B. Settles, and A. McCallum. Active learning by labeling features. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 81–90. ACL Press, 2009. (34%, 20% for oral presentation)
- B. Settles, M. Craven, and S. Ray. Multiple-instance active learning. In *Advances in Neural Information Processing Systems (NIPS)*, volume 20, pages 1289–1296. MIT Press, 2008. (20%)
- B. Settles and M. Craven. An analysis of active learning strategies for sequence labeling tasks. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1069–1078. ACL Press, 2008. (30%, 21% for oral presentation)
- A. Goldberg, D. Andrzejewski, J. van Gael, B. Settles, X. Zhu, and M. Craven. Ranking biomedical passages for relevance and diversity. In *Proceedings of the Text Retrieval Conference (TREC)*. 2007.
- T. Brow, B. Settles, and M. Craven. Classifying biomedical articles by making localized decisions. In *Proceedings of the Text Retrieval Conference (TREC)*. 2006.
- B. Settles and M. Craven. Exploiting zone information, syntactic features, and informative terms in gene ontology annotation from biomedical documents. In *Proceedings of the Text Retrieval Conference (TREC)*. 2005.

Refereed Workshop Papers

- E. Law, B. Settles, A. Snook, H. Surana, L. von Ahn, and T. Mitchell. Human computation for attribute and attribute value acquisition. In *CVPR Workshop on Fine-Grained Visual Categorization*. 2011.
- B. Settles. From theories to queries: Active learning in practice. *Active Learning and Experimental Design*. JMLR W&CP, volume 16, pages 1–18. Microtome Publishing, 2011. (Invited paper)
- B. Settles. Computational creativity tools for songwriters. In *Proceedings of the Workshop on Computational Approaches to Linguistic Creativity (CALC)*, pages 49–57. ACL Press, 2010. (66%)
- B. Settles, M. Craven, and L.A. Friedland. Active learning with real annotation costs. In *Proceedings of the NIPS Workshop on Cost-Sensitive Learning*, pages 1–10. 2008. (35%)
- B. Settles. Biomedical named entity recognition using conditional random fields and rich feature sets. In *Proceedings of the International Joint Workshop on Natural Language Processing in Biomedicine and its Applications (NLPBA)*, pages 104–107. 2004.

Other Publications

- B. Settles. Active learning literature survey. Computer Sciences Technical Report 1648, University of Wisconsin-Madison, 2009. (3,000+ downloads)
- B. Settles. *Curious Machines: Active Learning with Structured Instances*. PhD thesis, University of Wisconsin-Madison, 2008.

HONORS AND AWARDS

- | | |
|--|-----------|
| NIH Genomic Sciences Postdoctoral Fellowship | 2009 |
| NIH Computation and Informatics in Biology and Medicine PhD Fellowship | 2002–2006 |

WORK EXPERIENCE

- | | |
|--|--------------|
| February Album Writing Month, LLC , Pittsburgh, PA. <i>Founder and “Fawmlord”</i> | 2004–present |
| National Library of Medicine , Bethesda, MD. <i>Lister Hill Research Fellow</i> | 2003 |
| SRA International, Inc. , Fairfax, VA. <i>Research Intern</i> | 2001 |
| Lexmark International, Inc. , Lexington, KY. <i>Software Intern</i> | 1997, 1999 |

TEACHING
EXPERIENCE

Carnegie Mellon University, Pittsburgh, PA

- Guest Lecturer, *Machine Learning* (10-701) Spring 2011
- Guest Lecturer, *Social Media Analysis* (10-802) Spring 2011
- Guest Lecturer, *Advanced Statistical Language Processing: Reading the Web* (10-709) Fall 2009

University of Wisconsin, Madison, WI

- Instructor, *Computational Biology & Biostatistics* (summer undergrad seminar) Summer 2008
- Instructor, *Artificial Intelligence* (CS 540) Summer 2003
- Teaching Assistant, *Programming Languages & Compilers* (CS 536) Fall 2001
- Teaching Assistant, *Machine Organization & Architecture* (CS 354) Fall 2000

STUDENT
ADVISING

Thesis Proposal Committee Member

- Mohit Kumar, ABD. *Generalizing Active Learning for Practical Deployments*. PhD thesis proposal, Carnegie Mellon University Language Technologies Institute, 2011.

Project Supervision and Mentoring

- Guo Chen, Yilun Cui, and Liuyi Sun (CMU MS students). Semester project building an automated crowdsourcing-backed evaluation system for information extraction, 2012.
- Malcolm Greaves (CMU undergraduate). Summer independent study on naïve Bayes approaches to Web-scale information extraction, 2011.
- Edith Law (CMU PhD student). Co-advised a Data Analysis Project on learning to tag audio files for classification and retrieval, resulted in ECML'10 publication, 2010.
- Thomas Brow (Stanford University undergraduate). Co-advised a summer internship at UW–Madison on biomedical information retrieval, resulted in TREC'06 publication, 2005.

SELECTED TALKS

Invited Conference and Workshop Talks

- “Combining Learning Strategies to Make the Most of Language Resources”
EMNLP Workshop on Algorithms and Resources for Modeling of Dialects
and Language Varieties, Edinburgh, Scotland July 2011
- “From Theories to Queries: Active Learning in Practice”
AISTATS Active Learning and Experimental Design Workshop, Sardinia, Italy May 2010
- “14,000 Songs in 28 Days: A Case Study”
SXSW Interactive Conference, Austin, TX March 2010
- “Active Learning for NLP: Past, Present, and Future”
NAACL-HLT Active Learning for NLP Workshop, Boulder, CO June 2009

Invited Seminars and Colloquia

- “Machine Learning and Computer-Assisted Creativity”
Carnegie Mellon University, Pittsburgh, PA April 2012
- “Interactive Machine Learning: Combining Learning Strategies with Humans in the Loop”
University of Wisconsin, Madison, WI February 2012
University of Illinois, Chicago, IL February 2012
Johns Hopkins University, Baltimore, MD February 2012
University of Pittsburgh, Pittsburgh, PA February 2012
Carnegie Mellon University, Pittsburgh, PA November 2011
IBM T.J. Watson Research Center, Yorktown Heights, NY September 2011
- “Asking the Right Questions: New Query Types in Active Learning”
Oregon State University, Corvallis, OR November 2010
University of Oregon, Eugene, OR November 2010
University of Texas, Austin, TX March 2010

Tufts University, Medford, MA	March 2010
“Adaptive High-Throughput Peptide Screening and Analysis via Active Learning” Brookhaven National Laboratory, Long Island, NY	May 2009
University of Wisconsin, Madison, WI	February 2009
“Exploiting Document-Passage Relationships in Classification and Learning” University of Cambridge, United Kingdom	March 2007

**SELECTED PRESS
COVERAGE**

Never-Ending Language Learning

- “Aiming to Learn as We Do, a Machine Teaches Itself.” *New York Times*, October 5, 2010.
- “Read the Internet, Speak English.” *Universe/io9* (blog), October 8, 2010.

Computational Creativity Tools

- “CMU Fellow’s Online Tool Prods Songwriters Short on Inspiration.” *Associated Press*, January 2011. (Appeared in dozens of newspapers nationally.)
- “John Uses The Muse to Get Started on a Song.” *Marketplace* (NPR), January 21, 2011.
- “New Tools For February’s Album Writing Month.” *American Songwriter*, January 24, 2011.

February Album Writing Month Project

- “Can Journalism Find Its Own ‘Brooklyn Boheme’ to Inspire Collaboration?” *PBS.org*, March 1, 2012.
- “FAWM Challenges Songwriters to up Their Output.” *Pittsburgh CityPaper*, February 8, 2012.
- “Presidents’ Songs: Of Legends And Mortal Men.” *All Things Considered* (NPR), October 8, 2008.

**PROFESSIONAL
ACTIVITIES AND
SERVICE**

Meeting Organizer

- ICML Workshop on Combining Learning Strategies to Reduce Labeling Cost
Co-organizer with H. Daumé III, P. Rai, and X. Zhu
July 2011
- NAACL-HLT Workshop on Active Learning for NLP (ALNLP)
Co-organizer with K. Small and K. Tomanek
June 2010

Conference and Workshop Program Committees

- IJCAI: International Joint Conference on Artificial Intelligence
Senior PC, 2011
- AACL: Conference on Artificial Intelligence
2011–2012
- ICML: International Conference on Machine Learning
2011–2012
- EMNLP: Conference on Empirical Methods on Natural Language Processing
2009–2012
- NAACL-HLT: North American Chapter of the Association for Computational Linguistics
2012
- ACL: Association for Computational Linguistics
2010–2011
- ALNLP: Workshop on Active Learning in NLP
2009–2010
- BioNLP: Workshop on Biomedical Natural Language Processing
2009
- IJCNLP: International Joint Conference on Natural Language Processing
2008

Journal Reviewing

- *Artificial Intelligence*
- *Machine Learning*
- *Journal of Machine Learning Research*
- *Data Mining and Knowledge Discovery*
- *Information Retrieval*
- *Bioinformatics*
- *IEEE Journal of Selected Topics in Signal Processing*
- *IEEE Transactions on Knowledge and Data Engineering*
- *BMC Bioinformatics*
- *Briefings in Bioinformatics*

Conference Reviewing

- CSCW: ACM Conference on Community-Supported Cooperative Work 2012–2013
- SDM: SIAM International Conference on Data Mining 2010

Other Service

- Manager, [Active Learning \(Machine Learning\)](#) mailing list (300+ members) 2010–present
- Co-organizer (with K. Nigam), CMU ML/Google Distinguished Lecture Series 2010–present
- Organizer, UW–Madison Artificial Intelligence Seminar 2008–2009
- Social Chair, UW–Madison Student ACM Chapter 2000–2001

SOFTWARE PROJECTS

NELL: Never-Ending Language Learner 2009–present

<http://rtw.ml.cmu.edu/rtw/>

Software agent that continuously crawls the Internet 24×7, learning to read (and reading to learn) from English-language Web pages. (80+ citations)

FAWM: February Album Writing Month 2004–present

<http://fawm.org>

Online social network, goal-setting community, and annual challenge for musicians. I developed the concept and all the software, including search, recommender system, and social interaction features. 6,000+ participants from 60+ countries worldwide, collectively authoring 40,000+ new songs to date.

DUALIST: Utility for Active Learning with Instances and Semantic Terms 2011

<http://code.google.com/p/dualist/>

Open source application for quickly and interactively training text-based classifiers for various natural language processing tasks (10+ citations, 350+ downloads).

AMIL: Active Multiple-Instance Library 2007

<http://pages.cs.wisc.edu/~bsettles/amil/>

Open source framework in Java for multiple-instance learning, with particular components for various active learning scenarios (700+ downloads).

ABNER: A Biomedical Named Entity Recognizer 2005

<http://pages.cs.wisc.edu/~bsettles/abner/>

Open source information extraction tool in Java (with API) for mining information from biomedical texts (400+ citations; 10,000+ downloads).

MISCELLANY

Biographical: Married, US citizen.

Languages: English (mother tongue), Français (connaissances de base).

Interests and Hobbies: Songwriting and musical performance (200+ compositions, 2 album releases, and 80+ solo and group tour dates in North America and Europe); travel and backpacking (15+ countries); language, anagrams, and wordplay; bicycling, gardening, and environmental stewardship.

REFERENCES

Available upon request.