

BURR SETTLES

CONTACT INFORMATION

6775 Medical Sciences Center
1300 University Avenue
University of Wisconsin
Madison, WI 53706 USA

Vox: 608-265-6868
Fax: 608-265-7916
Email: bsettles@cs.wisc.edu
Web: pages.cs.wisc.edu/~bsettles/

RESEARCH INTERESTS

Machine learning, natural language processing, and bioinformatics. In particular, efficient active learning and multiple-instance learning approaches to facilitate real-world applications in information extraction, information retrieval, and data mining from social and biological networks.

ACADEMIC EDUCATION

University of Wisconsin, Madison, WI USA

- PhD, Computer Sciences (December 2008)
 - Thesis: *Curious Machines: Active Learning with Structured Instances*
 - Advisor: Mark Craven
 - Minors: Linguistics and Biology
- MS, Computer Sciences (May 2002)

DePauw University, Greencastle, IN USA

- BA, Computer Science (May 2000)
 - Graduated cum laude with distinction
 - Minor: Studio Art

HONORS AND AWARDS

NIH Genomic Sciences Postdoctoral Fellow (2009)
CIBM Research Fellow, University of Wisconsin (2002–2006)
Lister Hill Research Fellow, National Library of Medicine (2003)
Management Fellow, DePauw University (1996–2000)
Faculty Development Council Research Grant, DePauw University (1999)

RESEARCH EXPERIENCE

University of Wisconsin, Madison, WI USA

- *Postdoctoral Fellow* (January 2009–present)
Member of the machine learning and biomedical computing research groups. Developing software systems for biomedical text processing and functional annotation of biological experiments. Also designing algorithms for machine-guided biological experimentation and data mining.
- *Research Assistant* (August 2002–December 2008)
Developed algorithms and conducted experiments with machine learning approaches to information extraction and retrieval from text and image sources. Thesis work on active learning, multiple-instance learning, and human-computer interaction.

National Library of Medicine, Bethesda, MD USA

- *Lister Hill Research Fellow* (September 2003–December 2003)
Investigated hybrid logical/statistical machine learning methods to generate novel predicate rules in the SemRep information extraction system for clinical records.

SRA International, Inc., Fairfax, VA USA

- *Research Intern* (May 2001–September 2001)
Studied various approaches to text categorization. Implemented machine learning software for large-scale text processing, including feature selection and support vector machine modules.

TEACHING
EXPERIENCE

University of Wisconsin, Madison, WI USA

- *Instructor* (September 2001–June 2008)
Designed syllabi and taught courses for: Artificial Intelligence (CS540, Summer 2003), Introduction to Bioinformatics (IBS/CBB undergraduate research program, Summer 2008), and Graphics & Web Design (CS136, Spring 2001).
- *Teaching Assistant* (September 2000–May 2001)
Held office hours, administered lab groups, and assisted in grading for: Programming Languages and Compilers (CS536), Machine Organization and Architecture (CS354), and Object-Oriented Programming (CS302).

OTHER WORK
EXPERIENCE

FAWM.ORG, LLC, Madison, WI USA

- *Founder and Owner* (February 2004–present)
Established an international online social network and annual creative challenge for musicians. Developing and implementing database-driven community software for the FAWM.ORG website, including user management, information retrieval, and collaborative filtering recommender systems. Overseeing a team of volunteers for various related projects.

Lexmark International, Inc., Lexington, KY USA

- *Research and Development Intern* (May 1999 – August 1999)
Designed an internal network distribution system for third party printer standards. Developed intranet applications to facilitate code management in the business printer division.
- *Research and Development Intern* (May 1997–August 1997)
Developed and debugged firmware for the Optra S laser printer. Responsible for re-factoring code in a hardware simulation project.

Freelance Consulting

- *Web Designer and Application Developer* (January 1997–present)
Several large and small scale website design and development projects, including custom-built content management systems and interactive content websites.

COMPUTER
SKILLS

Languages

Java, Perl, HTML/CSS, Javascript, PHP, Python, C++, Prolog, R

Frameworks

MALLET, Weka (machine learning); jQuery, Django (web development)

Databases

MySQL, SQLite, PostgreSQL

Operating Systems

Mac OS X, Linux, Windows, Solaris

PUBLICATIONS

Journal Articles and Book Chapters

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.

B. Settles. ABNER: An Open Source Tool for Automatically Tagging Genes, Proteins, and Other Entity Names in Text. *Bioinformatics*, 21(14):3191–3192. 2005.

Conference and Workshop Articles

B. Settles and M. Craven. An Analysis of Active Learning Strategies for Sequence Labeling Tasks. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1069–1078. ACL Press, 2008.

B. Settles, M. Craven, and L. Friedland. Active learning with real annotation costs. *Proceedings of the NIPS Workshop on Cost-Sensitive Learning*, pages 1–10. 2008.

B. Settles, M. Craven, and S. Ray. Multiple-Instance Active Learning. *Advances in Neural Information Processing Systems (NIPS)*, volume 20, pages 1289–1296. MIT Press, 2008.

A. Goldberg, D. Andrzejewski, J.V. Gael, B. Settles, X. Zhu, and M. Craven. Ranking Biomedical Passages for Relevance and Diversity. *Proceedings of the Text Retrieval Conference (TREC)*. 2007.

T. Brow, B. Settles, and M. Craven. Classifying Biomedical Articles by Making Localized Decisions. *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. *Proceedings of the Text Retrieval Conference (TREC)*. 2005.

B. Settles. Biomedical Named Entity Recognition Using Conditional Random Fields and Rich Feature Sets. *Proceedings of the International Joint Workshop on Natural Language Processing in Biomedicine and its Applications (NLPBA)*, 104–107. 2004.

Technical Reports and Other Publications

B. Settles. Active Learning Literature Survey. Computer Sciences Technical Report 1648, University of Wisconsin-Madison. 2009.

B. Settles. *Curious Machines: Active Learning with Structured Instances*. PhD thesis, University of Wisconsin-Madison. 2008.

INVITED TALKS

“Exploiting Document-Passage Relationships in Classification and Learning.”
University of Cambridge, Cambridge, UK (March 2007)

“Machine Learning and Biomedical Text Processing.”
National Library of Medicine, Bethesda, MD USA (July 2005)

PROFESSIONAL ACTIVITIES AND SERVICE

Journal Refereeing

Machine Learning, *Journal of Machine Learning Research (JMLR)*, *Information Retrieval (IR)*, *Bioinformatics*, *BMC Bioinformatics*, *Briefings in Bioinformatics*

Conference and Workshop Program Committees

Active Learning for NLP (2009), IJCNLP (2008), DTMBIO (2008)

Departmental Service

UW AI Seminar Coordinator (2008), Social Chair (2000-2001)

Memberships

ACL, ACM, UW Teaching Assistants Association

SOFTWARE PROJECTS

AMIL: Active Multiple-Instance Library

pages.cs.wisc.edu/~bsettles/amil/

Since 2007. Open source framework in Java for multiple-instance learning, with particular components for various active learning scenarios.

ABNER: A Biomedical Named Entity Recognizer

pages.cs.wisc.edu/~bsettles/abner/

Since 2005. Open source information extraction tool in Java (with API) for mining information from biomedical texts. 4,000+ downloads and 150+ citations to date.

OTHER SKILLS AND INTERESTS

Proficient en français écrit et parlé.

Gardening and environmental stewardship.

Extensive print and online media experience.

Music composition, songwriting, and performance.

Travel, backpacking, and language acquisition.

Anagrams, holorhymes, and other wordplay.

CITIZENSHIP

United States of America.

REFERENCES

Available upon request.