I. CURRICULUM VITAE

William W. Cohen

EDUCATION

- Ph.D. in Computer Science, Rutgers University, 1990
- M.S. in Computer Science, Rutgers University, 1988
- B.S. in Computer Science, Duke University, 1984.

EMPLOYMENT

- Professor, Machine Learning Department, Carnegie Mellon University, July 2013-present.
    - Also member of CMU's Language Technology Institute
- Research Professor, Machine Learning Department, Carnegie Mellon University, July 2011-July 2013.
- Associate Research Professor, Machine Learning Department, Carnegie Mellon University, July 2003-July 2011.
- Visiting Researcher, Google, June 2008-July 2009.
- Visiting Associate Profession, Center for Automated Learning and Discovery, Carnegie Mellon University, July 2002-May 2003.
- Distinguished Research Scientist, WhizBang! Labs, Pittsburgh, PA, April 2000-May 2002. (Also Adjunct Faculty, Center for Automated Learning and Discovery, Carnegie Mellon University, Nov 2000-July 2002)
- Technology Consultant, AT&T Labs-Research, Shannon Laboratories, Florham Park NJ, July 1999-April 2000. (This was a distinguished/senior position, otherwise comparable to a "Research Staff Member" appointment).
- Principle Research Staff Member, AT&T Labs-Research, Shannon Laboratories, Florham Park NJ, Feb 1996-July 1999.
- Member Technical Staff, AT&T Bell Laboratories, Murray Hill NJ, Sep 1990-Feb 1996.
- Member Technical Staff, Computer Science Corporation, under contract to the Space Telescope, Science Institute, Baltimore MD, Nov 1985-Aug 1986.
- Computer Aided Design Research & Development Specialist, General Electric Microelectronics Center, Research Triangle Park, NC, June 1984-Sep 1985.

PERSONAL

- United States citizen.
- Married, two children, one dog, seven guitars, two mandolins.

II. PUBLICATION LIST

BOOKS

1. William W. Cohen & Samuel Gosling (Eds.): **Proceedings of the Fourth International Conference on Weblogs and Social Media, ICWSM 2010,** The AAAI Press.
2. William W. Cohen, Andrew McCallum, Sam T. Roweis (Eds.): **Machine Learning, Proceedings of the Twenty-Fifth International Conference (ICML 2008)**, 2008 ACM Press.
3. William W. Cohen: **A Computer Scientist's Guide to Biology,** 2007. Springer.
4. William W. Cohen & Andrew Moore (Eds.): **ICML 2006, Proceedings of the 23rd International Conference on Machine Learning,** Carnegie Mellon University, Pittsburgh, PA, USA, June 25-29, 2006. Omnipress.
5. William W. Cohen & Haym Hirsh (Eds.): **Machine Learning, Proceedings of the Eleventh International Conference**, Rutgers University, New Brunswick, NJ, USA, July 10-13, 1994. Morgan Kaufmann.

CHAPTERS IN BOOKS

6. William W. Cohen, Matthew Hurst & Lee S. Jensen (2003): A Flexible Learning System for Wrapping Tables and Lists in HTML Documents. In **Web Document Analysis: Challenges and Opportunities**, ed. Antonacopoulos & Hu, Word Scientific Publishing. (Versions also published as: William W. Cohen, Matthew Hurst & Lee S. Jensen (2002): A Flexible Learning System for Wrapping Tables and Lists in HTML Documents. **WWW 2002**: 232-241; Lee S. Jensen & William W. Cohen (2001): A Structured Wrapper Induction System for Extracting Information from Semi-Structured Documents. **Proc. of the IJCAI-2001 Workshop on Adaptive Text Extraction and Mining**)
7. William W. Cohen (1995): Learning to Classify English Text with ILP Methods. **Advances in ILP**, ed. L.uc de Readt, IOS Press.
8. Haym Hirsh and William W. Cohen (1994): Learning from data with bounded inconsistency: Theoretical and experimental results. **Computational learning theory and natural learning systems (Volume I)**, MIT Press.
9. William W. Cohen, Russell Greiner, and Dale Schuurmans (1994): Probabilistic hill-climbing**. Computational learning theory and natural learning systems (Volume II)**, MIT Press.

REFEREED JOURNAL PAPERS – PUBLISHED

10. Nan Li, William W. Cohen, and Ken Koedinger (2013): Problem Order Implications for Learning. *International Journal of Artificial Intelligence in Education,* to appear.
11. Einat Minkov and William W. Cohen (2011): Improving Graph-Walk Based Similarity with Reranking: Case Studies for Personal Information Management. *ACM Transactions on Information Systems* 29(1). (Also published as: Einat Minkov, Andrew Ng, and William Cohen: Contextual Search and Name Disambiguation in Email using Graphs. **SIGIR 2006**)
12. Ni Lao and William W. Cohen (2010): Relational Retrieval Using a Combination of Path-Constrained Random Walks. *Machine Learning* 81(1): 53-67. (Also published as: No Lao and William Cohen: Relational Retrieval Using a Combination of Path-Constrained Random Walks, **ECML-2010.**)
13. Amr Ahmed, Andrew Arnold, Luis. P. Coelho, J. Kangas, A.-S. Sheikh, Eris Xing, William W. Cohen, and Robert F. Murphy (in press). Structured Literature Image Finder: Parsing Text and Figures in Biomedical Literature. *Journal of Web Semantics*.

14. William W. Cohen & Einat Minkov (2006): A Graph-Search Framework for Associating Gene Identifiers with Documents. *BMC Bioinformatics*, 7:440
15. Sarah Zelikovitz, Haym Hirsh, William W. Cohen: Extending WHIRL with background knowledge for improved text classification. *Information Retrieval* 10(1):35-67 (2006).
16. Zhenzhen Kou, William W. Cohen, Robert F. Murphy: High-recall protein entity recognition using a dictionary. *Bioinformatics* 21(Suppl 1):266-273 (2005). (Also published as: Zhenzhen Kou, William W. Cohen, Robert F. Murphy: High-recall protein entity recognition using a dictionary, **ISMB 2005**: 255-273.)
17. William W. Cohen (2003): Learning and Discovering Structure in Web Pages. *IEEE Data Eng. Bull.* 26(3): 3-10 (2003)
18. Mikael Bilenko, Ray Mooney, William W. Cohen, Pradeep Ravikumar & Steve Fienberg (2003): Adaptive Name-Matching in Information Integration. *IEEE Intelligent Systems* 18(5): 16-23 (2003)
19. Chumki Basu, Haym Hirsh, William W. Cohen & Craig Neville-Manning (2001): Technical Paper Recommendation: A Study in Combining Multiple Information Sources. *J. Artif. Intell. Res. (JAIR)* 14: 231-252 (2001). (Originally published as: Chumki Basu, Haym Hirsh, William W. Cohen (1998): Recommendation as Classification: Using Social and Content-Based Information in Recommendation.. **AAAI/IAAI 1998**: 714-720)
20. William W. Cohen, Andrew McCallum, Dallan Quass (2000): Learning to Understand the Web. *IEEE Data Eng. Bull.* 23(3): 17-24 (2000)
21. William W. Cohen (2000): Data Integration using Similarity Joins and a Word-based Information Representation Language. *ACM Trans. Inf. Syst.* 18(3): 288-321 (2000). (Versions also published as: William W. Cohen (1998): Integration of Heterogeneous Databases without Common Domains Using Queries Based on Textual Similarity. **SIGMOD 1998**: 201-212; William W. Cohen (1997): Knowledge Integration for Structured Information Sources Containing Text (Extended Abstract). **Proc. of SIGIR 1997 Workshop on Networked IR**)
22. William W. Cohen and Wei Fan (2000): Web-Collaborative Filtering: Recommending Music by Crawling the Web. *Computer Networks* 33(1-6): 685-698 (2000). (Originally published as: William W. Cohen and Wei Fan (2000): Web-Collaborative Filtering: Recommending Music by Crawling The Web. **WWW 2000**)
23. William W. Cohen & Prem Devanbu (2000): Automatically Exploring Hypotheses about Fault Prediction: a Comparative Study of Inductive Logic Programming Methods. *International Journal of Software Engineering and Knowledge Engineering* 9(5): 519-546 (1999). (Originally published as: William W. Cohen and Prem Devanbu (1997): A Comparative Study of Inductive Logic Programming Methods for Software Fault Prediction**. ICML 1997**: 66-74)
24. William W. Cohen (2000): WHIRL: A Word-based Information Representation Language. *Artif. Intell. 118(1-2)*: 163-196 (2000)
25. William W. Cohen & Yoram Singer (1999): Context-sensitive learning methods for text categorization. *ACM Trans. Inf. Syst.* 17(2): 141-173 (1999). (Originally published as: William W. Cohen and Yoram Singer (1996): Context-sensitive learning methods for text categorization. **SIGIR 1996**: 307-315)
26. William W. Cohen (1999): Reasoning about Textual Similarity in a Web-Based Information Access System. *Autonomous Agents and Multi-Agent Systems* 2(1): 65-86 (1999)
27. William W. Cohen and Wei Fan (1999): Learning Page-Independent Heuristics for Extracting Data from Web Pages. *Computer Networks* 31(11-16): 1641-1652 (1999). (Originally published as: William W. Cohen and Wei Fan (1999): Learning Page-Independent Heuristics for Extracting Data from Web Pages. **WWW 1999**)
28. William W. Cohen, Rob Schapire, Yoram Singer (1999): Learning to Order Things. *J. Artif. Intell. Res. (JAIR)* 10: 243-270 (1999). (Originally published as: William W. Cohen, Robert E. Schapire, Yoram Singer (1997): Learning to Order Things. **NIPS 1997**)

29. William W. Cohen (1998): The WHIRL Approach to Information Integration. *IEEE Intelligent Systems*, Sept/Oct 1998, pp 20--23
30. William W. Cohen (1998): Hardness Results for Learning First-Order Representations and Programming by Demonstration. *Machine Learning* 30(1): 57-87 (1998). (Originally published as: William W. Cohen (1996): The Dual DFA Learning Problem: Hardness Results for Programming by Demonstration and Learning First-Order Representations (Extended Abstract). **COLT 1996**: 29-40)
31. William W. Cohen (1995): PAC-Learning Non-Recursive Prolog Clauses. *Artif. Intell.* 79(1): 1-38 (1995)
32. William W. Cohen (1996): Adaptive Mapping and Navigation by Teams of Simple Robots. *Robotics and Autonomous Systems*, 18: 411-434 (1996)
33. William W. Cohen (1995): Inductive Specification Recovery: Understanding Software by Learning from Example Behaviors. *Autom. Softw. Eng.* 2(2): 107-129 (1995). (Originally published as: William W. Cohen (1994): Recovering Software Specifications with Inductive Logic Programming. **AAAI 1994**: 142-148)
34. William W. Cohen and C. David Page Jr (1995): Polynomial Learnability and Inductive Logic Programming: Methods and Results. *New Generation Comput.* 13(3&4): 369-409 (1995)
35. William W. Cohen (1995): PAC-Learning Recursive Logic Programs: Efficient algorithms. *J. Artif. Intell. Res. (JAIR*) 2: 501-539 (1995)
36. William W. Cohen (1995): PAC-learning recursive Logic Programs: Negative results. J. *Artif. Intell. Res. (JAIR)* 2: 541-573 (1995)
37. L. Thorn McCarty and William W. Cohen (1994): The Case for Explicit Exceptions. *Methods of Logic in Computer Science*, 1(1)
38. William W. Cohen (1994): Incremental abductive EBL. *Machine Learning* 15(1): 5-24 (1994)
39. William W. Cohen (1994): Grammatically Biased Learning: Learning Logic Programs using an Explicit Antecedent Eescription Language. *Artif. Intell.* 68(2): 303-366 (1994)
40. William W. Cohen and Haym Hirsh (1994): Learnability of Description Logics with Equality Constraints. *Machine Learning* 17(2-3): 169-199 (1994). (Originally published as: William W. Cohen and Haym Hirsh (1992): Learnability of description logics. **COLT 1992**: 116-127)
41. William W. Cohen (1992): Using Distribution-free Learning Theory to Analyze Solution Path Caching Mechanisms. *Computational Intelligence* 8: 336-375 (1992)
42. William W. Cohen (1992): Abductive Explanation Based Learning: A solution to the multiple inconsistent explanation problem. *Machine Learning* 8: 167-219 (1992)
43. A. De Geus and W. Cohen (1985): Optimization of Combinational Logic using a Rule-based Expert Eystem. *IEEE Design and Test of Computers*

REFEREED CONFERENCE/WORKSHOP PAPERS

44. On Modeling Community Behaviors and Sentiments in Microblogging (2014): Tuan-Ahn Hoang and William W. Cohen and Ee-Peng Lim. **SDM-2014.**
45. Partha Pratim Talukdar and William W. Cohen (2014): Scaling Graph-based Semi Supervised Learning to Large Number of Labels Using Count-Min Sketch. **AI-Stats 2014.**
46. Jay Pujara, Hui Miao, Lise Getoor, and William W. Cohen (2013): Ontology-Aware Partitioning for Knowledge Graph Identification. **AKBC-2013.**
47. Bhavana Dalvi, William W. Cohen, and Jamie Callan (2013): Classifying Entities into an Incomplete Ontology. **AKBC-2013.**
48. Douglas Pierce, David P. Redlawsk, and William W. Cohen (2013): Social Influences on Political Information Search and Evaluation. **APSA-2013.**
49. William Yang Wang, Kathryn Mazaitis, William W. Cohen (2013): Programming with Personalized PageRank: A Locally Groundable First-Order Probabilistic Logic. **CIKM-2013.** (Honorable Mention for Best Student Paper).

50. Jay Pujara, Hui Miao, Lise Getoor, and William W. Cohen (2013): Knowledge Graph Identification. **ISWC-2013** (Best Student Paper).
51. Ramnath Balasubramanyan, Bhavana Dalvi and William W. Cohen (2013).): From Topic Models to Semi-Supervised Learning: Biasing Mixed-membership Models to Exploit Topic-Indicative Features in Entity Clustering. **ECML/PKDD 2013.**
52. Bhavana Dalvi and William W. Cohen and Jamie Callan (2013): Exploratory Learning. **ECML/PKDD 2013.**
53. Nan Li, William W. Cohen, and Ken Koedinger (2013): Discovering Student Models With A Clustering Algorithm Using Problem Content. **EDM-2013.**
54. Nan Li, Yanduong Tian, William W. Cohen, and Ken Koedinger (2013): Integrating Perceptual Learning With External World Knowledge In A Simulated Student. **AIED-2013.**
55. Nan Li, Eliane Stampfer, William W. Cohen, and Ken Koedinger (2013): Efficient Cross-Domain Cognitive Model Discovery Using A Simulated Student. **CogSci-2013.**
56. Tuan-Ahn Hoang, William W. Cohen, Ee-Peng Lim, Doug Pierce, David Redlawsk (2013): Politics, Sharing and Emotion in Microblogs. **ASONAM-2013.**
57. Dana Movshovitz-Attias and William W. Cohen (2013): Natural Language Models for Predicting Programming Comments. **ACL-2013** (short paper).
58. Ramnath Balasubramanyan and William W. Cohen (2013): Regularization of Latent Variable Models to Obtain Sparsity. SDM-2013Bhavana Dalvi and William W. Cohen (2013): Very Fast Similarity Queries on Semi-Structured Data from the Web. **SDM-2013.**
59. Frank Lin and William W. Cohen (2012): A General and Scalable Approach to Mixed Membership Clustering. **ICDM-2012.**
60. Nan Li, William W. Cohen, Kenneth R. Koedinger (2012): Learning to Perceive Two-Dimensional Displays Using Probabilistic Grammars. **ECML-2012.**
61. Ramnath Balasubramanyan and William W. Cohen (2012): Entropic Regularization of Mixed-membership Network Models using Pseudo-observations. **MLG-2012.**
62. Nan Li, Abraham Schreiber, William W. Cohen, Kenneth R. Koedinger (2012): Creating Features from a Learned Grammar in a Simulated Student. **ECAI-2012.**
63. Nan Li, William W. Cohen, Kenneth R. Koedinger (2012): Learning to Perceive Two-Dimensional Displays Using Probabilistic Grammars. **ECAI-2012.**
64. Mahesh Joshi, Mark Dredze, William Cohen and Carolyn Rose (2012): Multi-Domain Learning: When Do Domains Matter? **EMNLP-CoNLL-2012.**
65. Ni Lao, Amar Subramanya, Fernando Pereira and William W. Cohen (2012): Reading The Web with Learned Syntactic-Semantic Inference Rules. **EMNLP-CoNLL-2012.**
66. Einat Minkov and William W. Cohen (2012): Graph Based Similarity Measures for Synonym Extraction from Parsed Text. **TextGraphs-2012.**
67. Noboru Matsuda, Evelyn Yarzebinski, Victoria Keiser, Rohan Raizada, William W. Cohen, Gabriel Stylianides, Kenneth R. Koedinger (2012): Shallow learning as a pathway for successful learning both for tutors and tutees. **CogSci-2012.**
68. Nan Li, William W. Cohen, and Ken Koedinger (2012): Problem Order Implications for Learning Transfer. **ITS-2012.**
69. Nan Li, William W. Cohen, and Ken Koedinger (2012): Efficient Cross-Domain Learning of Complex Skills. **ITS-2012** (short paper).
70. Ramnath Balasubramanyan, William W. Cohen, Doug Pierce, and David P. Redlawsk (2012): Modeling Polarizing Topics: When Do Different Political Communities Respond Differently to the Same News? **ICWSM-2012.**
71. Bhavana Dalvi, William W. Cohen, and Jamie Callan (2012): WebSets: Extracting Sets of Entities from the Web Using Unsupervised Information Extraction. **WSDM-2012.**
72. Ni Lao, Tom Mitchell, and William W. Cohen (2011): Random Walk Inference and Learning in A Large Scale Knowledge Base. **EMNLP-2011.**
73. Frank Lin and William W. Cohen (2011): Adaptation of Graph-Based Semi-Supervised Methods to Large-Scale Text Data. **MLG-2011.**
74. Nan Li and William W. Cohen and Kenneth R. Koedinger and Noboru Matsuda (2011): A Machine Learning Approach for Automatic Student Model Discovery. **EDM-2011.**

75. Noboru Matsuda, Evelyn Yarzebinski, Victoria Keiser, Rohan Raizada, Gabriel J. Stylianides, William W. Cohen, Kenneth R. Koedinger (2011): Learning by Teaching SimStudent - An Initial Classroom Baseline Study comparing with Cognitive Tutors. **AIED-2011.**
76. Ramnath Balasubramanyan and William W. Cohen (2011): Block-LDA: Jointly Modeling Entity-Annotated Text and Entity-Entity Links. **SDM-2011.**
77. Philip Stutz, Abraham Bernstein and William W. Cohen (2010): Signal/Collect: Graph Algorithms for the (Semantic) Web. **ISWC-2010.**
78. Frank Lin and William W. Cohen (2010): Semi-Supervised Classification of Network Data Using Very Few Labels**. ASONAM-2010**.
79. Frank Lin and William W. Cohen (2010): Power Iteration Clustering. **ICML-2010.**
80. Frank Lin and William W. Cohen (2010): A Very Fast Method for Clustering Big Text Datasets. **ECAI-2010.**
81. Ni Lao and William W. Cohen (2010): Fast Query Execution for Retrieval Models based on Path Constrained Random Walks. **KDD-2010.**
82. Noboru Matsuda, Victoria Keiser, Rohan Raizada, Arthur Tu, Gabriel Stylianides, William W. Cohen, and Kenneth R. Koedinger (2010). Learning by Teaching SimStudent: Technical Accomplishments and an Initial Use with Students. **ITS-2010.**
83. Vitor R. Carvalho, Ramnath Balasubramanyan and William W. Cohen. Information Leaks and Suggestions (2009): A Case Study using Mozilla Thunderbird. **CEAS-2009.**
84. Richard Wang and William W. Cohen (2009). Character-level Analysis of Semi-Structured Documents for Set Expansion. **EMNLP 2009.**
85. Noboru Matsuda, Andrew Lee, William W. Cohen, and Ken Koedinger (2009). A Computational Model of How Learner Errors Arise from Weak Prior Knowledge. **CogSci 2009**.
86. Amr Ahmed, Eric P. Xing, William W. Cohen, and Robert F. Murphy (2009). Structured Correspondence Topic Models for Mining Captioned Figures in Biological Literature. **KDD 2009**.
87. Tae Yano, Noah A. Smith, and William W. Cohen (2009): Predicting Response to Political Blog Posts with Topic Models. **NAACL 2009**.
88. Richard Wang and William W. Cohen (2008): Automatic Set Instance Extraction using the Web. **ACL/HLT 2009**.
89. Richard Wang and William W. Cohen (2008): Iterative Set Expansion of Named Entities Using the Web. **ICDM 2008.**
90. Andrew Arnold and William W. Cohen (2008): Intra-document Structural Frequency Features for Semi-Supervised Domain Adaptation. **CIKM 2008.**
91. Richard Wang, Nico Schlaefer, William W. Cohen, and Eric Nyberg (2008): Automatic Set Expansion for List Question Answering. **EMNLP 2008**.
92. Einat Minkov and William W. Cohen (2008): Learning Graph Walk Based Similarity Measures for Parsed Text. **EMNLP 2008.**
93. Andrew Arnold, Ramesh Nallapati and William W. Cohen (2008): Exploiting Feature Hierarchy for Transfer Learning in Named Entity Recognition. **ACL 2008**.
94. Ramesh Nallapati, Amr Ahmed, Eric Xing, and William W. Cohen (2008): Joint Latent Topic Models for Text and Citations. **KDD 2008.**
95. Noboru Matsuda, William W. Cohen, Jonathan Sewall, Gustavo Lacerda, and Kenneth R. Koedinger (2008): Why Tutored Problem Solving may be better than Example Study: Theoretical Implications from a Simulated-Student Study. **ITS 2008.**
96. Yi-Chia Wang, Mahesh Joshi, William Cohen, and Carolyn Rose (2008): Recovering Implicit Thread Structure in Newsgroup Style Conversations. **ICWSM 2008.**
97. Ramesh Nallapati and William W. Cohen (2008): Link-PLSA-LDA: A New Unsupervised Model for Topics and Influence of Blogs**. ICWSM 2008.**
98. Vitor Carvalho and William W. Cohen (2008): Ranking Users for Intelligent Message Addressing. **ECIR 2008.**
99. Noboru Matsuda and William Cohen and Ken Koedinger (2007): Evaluating a simulated student using real students data for training and testing. **UM 2007**.

100. Noboru Matsuda and William Cohen and Ken Koedinger (2007): Predicting students performance with a SimStudent that learns cognitive skills from observation. **AIED 2007.**
101. Vitor Carvalho, Wen Wu and William Cohen (2007): Discovering Leadership Roles in Email Workgroups. **CEAS 2007.**
102. Ramesh Nallapati, William Cohen, Susan Ditmore, John Lafferty and Kin Ung (2007): Multiscale Topic Tomography. **KDD 2007**.
103. Richard Wang and William Cohen (2007): Language-Independent Set Expansion of Named Entities using the Web. **ICDM 2007.**
104. Einat Minkov and William Cohen (2007): Learning to Rank Typed Graph Walks: Local and Global Approaches. **WebKDD 2007.**
105. Juchang Hua, Orhan Ayasli, William Cohen and Robert Murphy (2007): Identifying Fluorescence Microscope Images in Online Journal Publications using Both Image and Text Features. **ISBI 2007.**
106. Vitor Carvalho and William W. Cohen (2007): Preventing Information Leaks in Email in SDM 2007, **SDM 2007.**
107. Zhenzhen Kou and William W. Cohen (2007): Stacked Graphical Models for Efficient Inference in Markov Random Fields, **SDM 2007.**
108. Zhenzhen Kou, William W. Cohen, and Robert F. Murphy (2007): A Stacked Graphical Model for Associating Information from Text And Images In Figures. **PSB 2007.**
109. Einat Minkov and William W. Cohen (2006): An Email and Meeting Assistant using Graph Walks. **CEAS 2006.**
110. Vitor Carvalho and William W. Cohen (2006): Single-Pass Online Learning: Performance, Voting Schemes and Online Feature Selection. **KDD 2006** (forthcoming)
111. Einat Minkov, Richard C. Wang, and William W. Cohen (2005): Extracting Personal Names from Email: Applying Named Entity Recognition to Informal Text. **EMNLP/HLT 2005**
112. Edoardo M. Airoldi, William W. Cohen, Stephen E. Fienberg (2005): Bayesian methods for frequent terms in text: Models of contagion and the Delta square statistic. **CSNA 2005**
113. William W. Cohen, Einat Minkov & Anthony Tomasic (2005): Learning to Understand Web Site Update Requests. **IJCAI 2005**
114. William W. Cohen & Vitor Carvalho (2005): Stacked Sequential Learning. **IJCAI 2005**
115. Vitor Carvalho & William W. Cohen (2005): On the Collective Classification of Email Speech Acts. **SIGIR 2005**
116. Carolyn Rose, Pinar Donmez, G. Gweon, A. Knight, B. Junker, W. Cohen, K. Koedinger, N. Hefferman (2005): Automatic and Semi-Automatic Skill Coding with a View Towards Supporting On-Line Assessment.**AIED 2005**
117. Sunita Sarawagi & William W. Cohen (2004): Semi-Markov Conditional Random Fields for Information Extraction**. NIPS 2004**
118. William W. Cohen, Vitor R. Carvalho & Tom Mitchell (2004): Learning to Classify Email into "Speech Acts". **EMNLP 2004**
119. Vitor R. Carvalho & William W. Cohen (2004): Learning to Extract Signature and Reply Lines from Email. **CEAS 2004**
120. Pradeep Ravikumar & William W. Cohen (2004): A Hierarchical Graphical Model for Record Linkage. **UAI 2004**
121. William W. Cohen & Sunita Sarawagi (2004): Exploiting Dictionaries in Named Entity Extraction: Combining Semi-Markov Extraction Processes and Data Integration Methods. **KDD 2004**: 89-98
122. William W. Cohen, Richard Wang & Robert Murphy (2003): Understanding Captions in Biomedical Publications. **KDD 2003**: 499-504
123. William W. Cohen (2003): Infrastructure Components for Large-Scale Information Extraction Systems. **IAAI 2003**: 71-78
124. Cheng Zhai, William W. Cohen & John Lafferty (2003): Beyond Independent Topical Relevance: Methods and Evaluation Metrics for Subtopic Retrieval. **SIGIR 2003**: 10-17

125. William W. Cohen (2002): Improving A Page Classifier with Anchor Extraction and Link Analysis. **NIPS 2002**
126. William W. Cohen & Jacob Richman (2002): Learning to Match and Cluster Large High-Dimensional Data Sets For Data Integration. **KDD 2002**: 475-480
127. William W. Cohen, David McAllester, and Henry Kautz (2000): Hardening Soft Information Sources. **KDD 2000**: 255-259
128. William W. Cohen (2000): Automatically extracting features for concept learning from the Web. **ICML 2000**: 159-166
129. William W. Cohen and Yoram Singer (1999): Simple, Fast, and Effective Rule Learner. **AAAI/IAAI 1999**: 335-342
130. William W. Cohen (1999): Recognizing Structure in Web Pages using Similarity Queries. **AAAI/IAAI 1999**: 59-66
131. William W. Cohen (1999): A Demonstration of WHIRL (demonstration abstract). **SIGIR 1999**: 327
132. William W. Cohen & Haym Hirsh (1998): Joins that Generalize: Text Classification Using WHIRL. **KDD 1998:** 169-173
133. William W. Cohen (1998): Providing Database-like Access to the Web Using Queries Based on Textual Similarity. **SIGMOD 1998**: 558-560
134. William W. Cohen (1998): A Web-based Information System that Reasons with Structured Collections of Text. **Agents 1998**: 400-407
135. William W. Cohen and Daniel Kudenko (1997): Transferring and Retraining Learned Information Filters. **AAAI/IAAI 1997**: 583-590
136. William W. Cohen (1996): Learning Trees and Rules with Set-valued Features. **AAAI/IAAI**, Vol. 1 1996: 709-716
137. William W. Cohen (1995): Fast Effective Rule Induction. **ICML 1995**: 115-123
138. William W. Cohen (1995): Text Categorization and Relational Learning. **ICML 1995**: 124-132
139. William W. Cohen (1994): Pac-learning nondeterminate Clauses. **AAAI 1994:** 676-681
140. William W. Cohen and Haym Hirsh (1994): Learning the CLASSIC description logic: Theoretical and experimental results. **KR 1994:** 121-133
141. William W. Cohen (1993): Cryptographic limitations on learning one-clause logic programs. **AAAI 1993**: 80-85
142. William W. Cohen (1993): Pac-learning a restricted class of recursive logic programs. **AAAI 1993**: 86-92
143. William W. Cohen (1993): Efficient pruning methods for separate-and-conquer rule learning systems. **IJCAI 1993:** 988-994
144. William W. Cohen, Alex Borgida, and Haym Hirsh (1992): Computing least common subsumers in description logics. **AAAI 1992**: 754-760
145. William W. Cohen (1992): Compiling prior knowledge into an explicit bias. **ICML 1992**: 102-110
146. William W. Cohen (1991): The generality of overgenerality. **ICML 1991**: 490-494
147. William W. Cohen (1990): Learning from textbook knowledge: A case study. **AAAI 1990**: 743-748
148. William W. Cohen (1990): Learning approximate control rules of high utility.**ICML 1990**: 268-276
149. William W. Cohen (1990): An analysis of representation shift in concept learning. **ICML 1990**: 104-112
150. William W. Cohen (1988): Generalizing number and learning from multiple examples in explanation-based learning. **ICML 1988**: 256-269

UNREFEREED CONFERENCE/WORKSHOP PAPERS

151. David Redlawsk, Douglas Pierce, William Cohen, Ramnath Balasubramanyan, Tae Yano (2012): Emotional Convergence Among Members of Online Social Networks. **APSA-2012** (Meeting of the American Political Science Association).

152. David Redlawsk, Douglas Pierce, William Cohen, Ramnath Balasubramanyan, Tae Yano (2012): Rational Updating in the Face of Incongruent Candidate Information. **EPSA-2012** (Meeting of the European Political Science Association).
153. David Redlawsk, Douglas Pierce, William Cohen, Ramnath Balasubramanyan, Tae Yano (2012): Rational Updating in the Face of Incongruent Candidate Information. **MPSA-2012** (Meeting of the Midwest Political Science Association).
154. Bhavana Dalvi, William W. Cohen, and Jamie Callan (2012): Collectively Representing Semi-Structured Data from the Web. **KBC-2012.**
155. Dana Movshovitz-Attias and William W. Cohen (2012): Alignment-based Extraction of Abbreviations from Biomedical Text. **BioNLP-2012.**
156. Dana Movshovitz-Attias and William W. Cohen (2012): Bootstrapping Biomedical Ontologies for Scientific Text using NELL. **BioNLP-2012.**
157. Ramnath Balasubramanyan, Kathryn Rivard, William W. Cohen, Jelena Jakovljevic and John Woolford (2012): Evaluating Joint Modeling of Yeast Biology Literature and Protein-Protein Interaction Networks. **BioNLP-2012.**
158. Partha Pratim Talukdar and William W. Cohen (2012): Crowdsourced Comprehension: Predicting Prerequisite Structure in Wikipedia. **BEA-2012.**
159. Jacob Eisenstein, Tae Yano, William W. Cohen, Noah A. Smith, and Eric P. Xing (2011): Structured Databases of Named Entities from Bayesian Nonparametrics. **UNSUP-2011**.
160. Bhavana Dalvi, Jamie Callan, and William W. Cohen (2011): Entity List Completion Using Set Expansion Techniques. **TREC 2011.**
161. Ramnath Balasubramanyan and William W. Cohen and Doug Pierce and David P. Redlawsk. What pushes their buttons? Predicting comment polarity from the content of political blog posts. **ACL/HLT Workshop on Language in Social Media (LSM 2011).**
162. Ramnath Balasubramanyan and William W. Cohen (2010): Block-LDA: Jointly modeling entity-annotated text and entity-entity links**. ICML-2010 Workshop on Topic Modeling.**
163. Nan Li, William W. Cohen, and Kenneth R. Koedinger. A Computational Model of Accelerated Future Learning through Feature Recognition. **ITS-2010** (poster).
164. Amr Ahmed, Andrew O. Arnold, Luis Pedro Coelho, Joshua Kangas, Abdul-Saboor Sheikk, Eric P. Xing, William W. Cohen, and Robert F. Murphy. Structured Literature Image Finder. **Biolink** 2009.
165. Andrew Arnold and William W. Cohen (2009): Information Extraction as Link Prediction: Using Curated Citation Networks to Improve Gene Detection. **ICWSM 2009** (poster).
166. Andrew Arnold and William W. Cohen (2009): Information Extraction as Link Prediction: Using Curated Citation Networks to Improve Gene Detection. **WASA 2009**.
167. Ramnath Balasubramanyan, Frank Lin, William W. Cohen, Noah A. Smith, and Matthew Hurst (2009): From Episodes to Sagas: Understanding the News by Identifying Temporally Related Story Sequence. **ICWSM 2009** (poster)
168. Einat Minkov, Ramnath Balasubramanyan, and William W. Cohen (2008): Activity-centric Search in Email. **AAAI 2008 Workshop on Enhanced Messaging**.
169. Ramnath Balasubramanyan, Vitor Carvalho, and William W. Cohen (2008): CutOnce - Recipient Recommendation and Leak Detection in Action. **AAAI 2008 Workshop on Enhanced Messaging.**
170. Andrew Arnold, Ramesh Nallapati and William W. Cohen (2007): A Comparative Study of Methods for Transductive Transfer Learning**. ICDM 2007 Workshop on Mining and Management of Biological Data.**
171. Ramesh Nallapati, William W. Cohen, and John Lafferty (2007): Parallelized Variational EM for Latent Dirichlet Allocation: An Experimental Evaluation of Speed and Scalability. **ICDM 2007 Workshop on High Performance Data Mining.**
172. Ramesh Nallapati, Amr Ahmed, William Cohen and Eric Xing (2007): Sparse Word Graphs: A Scalable Algorithm for Capturing Word Correlations in Topic Models. **ICDM 2007 Workshop on High Performance Data Mining.**

173. Vitor Carvalho and William W. Cohen (2006): Improving Email Speech Act Analysis via N-gram Selection. **HLT/NAACL ACTS Workshop 2006**.
174. Einat Minkov, Richard C.Wang, Anthony Tomasic and William W. Cohen (2006): NER Systems that Suit Users Preferences: Adjusting the Recall-Precision Trade-off for Entity Extraction. **HLT/NAACL 2006** (short paper).
175. William W. Cohen (2006): A Graph-Search Framework for GeneId Ranking (Extended Abstract). **BioNLP'06.**
176. Noboru Matsuda, William Cohen & Ken Koedinger (2005): An Intelligent Authoring System with Programming by Demonstration. **Proceedings of the Japan National Conference on Information and Systems in Education**.
177. Noboru Matsuda, William Cohen & Ken Koedinger (2005): Building Cognitive Tutors with Programming by Demonstration. **ILP 2005** (late-breaking paper).
178. Noboru Matsuda, William Cohen & Ken Koedinger (2005): Applying Programming by Demonstration in an Intelligent Authoring Tool for Cognitive Tutors. **AAAI Workshop on Human Comprehensible Machine Learning.**
179. Yifen Huang, Dinesh Govindaraju, Tom Mitchell, Vitor Rocha de Carvalho & William W. Cohen (2004): Inferring Ongoing Activities of Workstation Users by Clustering Email. **CEAS 20042** (short paper)
180. Lee S. Jensen & William W. Cohen (2001): Grouping Extracted Fields. **Proc. of the IJCAI 2001 Workshop on Adaptive Text Extraction and Mining**
181. G. Miller, D. Rosenthal, W. Cohen, and M. Johnston (1987): Expert systems tools for hubble space telescope scheduling. **Proc. of the Goddard Conference on Space Applications of Artificial Intelligence and Robotics**
182. T. Hornick, W. Cohen, and G. Miller (1987): A natural language query system for hubble space telescope proposal selection. **Proc. of the Goddard Conference on Space Applications of Artificial Intelligence and Robotics**
183. K. Bartlett, W. Cohen, A. De Geus, and G. Hachtel (1986): Synthesis and optimization of multi-level logic under timing constraints. **Proc. of the IEEE International Conference on Computer-Aided Design**
184. W. Cohen, K. Bartlett, and A. De Geus (1985): Impact of metarules in a rule-based expert system for gate level optimization. **Proc. of the IEEE Int'l Symp. on Circuits and Systems**
185. W. Cohen, K. Bartlett, and A. De Geus (1984): Impact of metarules in a rule-based expert system for gate level optimization. **Proc. of the IEEE Int'l Symp. on Circuits and Systems**
186. Karl Garrison, David Gregory, William W. Cohen & Aart De Geus, (1984): Automatic Area and Performance Optimization of Combinatorial Logic. **Proc. of the IEEE International Conference on Computer-Aided Design**

OTHER PUBLICATIONS

187. William W. Cohen (2001): Issues in Extracting Information from the Web (Extended abstract, submitted by invitation). **IWPT 2001**
188. William W. Cohen (2000): Extracting Information from the Web for Concept Learning and Collaborative Filtering. (Submitted by invitation) **ALT 2000**: 1-12
189. Jaime G. Carbonell, Yiming Yang, William W. Cohen (2000): Special Issue of Machine Learning on Information Retrieval - Introduction. *Machine Learning* 39(2/3): 99-101 (2000)
190. William W. Cohen (1993): A Review of `Creating a Memory of Causal Relationships' by Michael Pazzani. *Machine Learning* 10(2) (1993)

PATENTS AND INVENTION DISCLOSURES

191. Context-dependent Similarity Measurements. Provisional patent #61/224,757.

192. Method and apparatus for extracting data from data sources on a network. Patent #6,516,308.
193. A system and method for accessing heterogeneous databases. Patent #6,295,533.
194. A system and method for finding information in a distributed information system using query learning and meta search. Patent # 5,347,623.
195. Rule induction on large noisy data sets. Patent # 5,719,692.
196. Software discovery system. Patent # 5,642,472.
197. Biased learning system. Patents # 5,481,650 and # 5,627,945.

SOFTWARE ARTIFACTS

198. RIPPER: Rule learning system, distributed for research purposes since 1995. Over 200 papers in ResearchIndex cite the archival description of RIPPER (Cohen, ICML-1995, citation #53 above) , many of which are applications of RIPPER to real problems.
199. SLIPPER: Boosting-based rule learning system, distributed for research purposes by Rutgers University since 2001.
200. WHIRL: Database/information-retrieval system, distributed for research purposes by Rutgers University since 2001. Used in publications by research groups at Rutgers, AT&T, Columbia, University College/Dublin, and elsewhere.
201. SecondString: Open-source JAVA package of string-distance metrics for use in record-linkage problems, distributed since 2002. Used in publications by research groups from U Penn, U Mass, U Washington, U Illinois, and elsewhere. Over 1000 downloads to date.
202. MinorThird: Open-source JAVA package for text classification and extraction, distributed since spring 2004. Actively used by several research groups at CMU (other than my own students!). Over 200 downloads to date.

III. EVIDENCE OF EXTERNAL REPUTATION

INVITED TALKS AND AWARDS

- "Learning to Reason with Extracted Information", keynote talk for Google's invitation-only **Natural Language Understanding Workshop**, Zurich, Switzerland, March 2014.
- "Learning to Construct and Reason with a Large Knowledge Base of Extracted Information", invited talk at the **23rd International Conference on Inductive Logic Programming (ILP-2103)**, Rio de Janeiro, Brazil, August 2013.
- "Unifying Personalized PageRank and Prolog", invited talk at the **ICML 2013 Workshop on Structured Learning (SLG 2013)**, Atlanta, Georgia, June 2013.
- "Learning Similarity Measures Based on Random Walks", invited talk at the **21st ACM International Conference on Information and Knowledge Management (CIKM 2012),** at Maui, Hawaii, October 2012.
- "Reasoning with Data Extracted from the Scientific Literature", invited talk given at a joint session of the **AAAI Fall Symposium on Discovery Informatics** and the **AAAI Fall Symposium on Information Retrieval and Knowledge Discovery in Biomedical Text,** Arlington, Virginia, October 2012.
- "Learning Relationships Defined by Linear Combinations of Constrained Random Walks", invited talk given at the **Ninth International Workshop on Mining and Learning with Graphs (MLG-2011),** San Diego, CA, August 2011.
- **"**Learning to Extract a Broad-Coverage Knowledge Base from the Web**",** invited talk given at the **Symposium on Data-Intensive Analysis, Analytics, and Informatics,** Pittsburgh, PA, April 2011.
- "Open Information Extraction Methods: Computers that Learn to Read", invited talk to be given at the **2011 Annual Conference of the National Federation of Advanced Information Services (NFAIS 2011),** Philadelphia, PA, February 2011.
- "Predictively Modeling Social Media", invited talk given at **the 1st International Workshop on Mining Social Media**, co-located with **13th Conference of the Spanish Association for Artificial Intelligence (CAEPIA-TTIA 2009)**, Sevilla, Spain, November 2009.
- "Matching and clustering product descriptions using learned similarity metrics", invited talk given at the **IJCAI 2009 Workshop on Information Integration on the Web**, Pasadena, CA, July 2009.
- "Graph-Based Methods for Open Information Extraction", invited talk at **NIPS 2009 Workshop on Graph Learning,** Whistler, BC, Canada, December 2009.
- "Using Machine Learning to Discover and Understand Structured Data", invited talk given at **LinkedData 2008,** New York, NY, July 2009.
- "Embodied Cognition and Knowledge: Integration of Heterogeneous Databases without Common Domains Using Queries Based on Textual Similarity", 10-year "Test of Time" Award talk at **SIGMOD 2008,** Vancouver, BC, August 2008.

The "Test of Time" award in 2008 is a ten-year retrospective best paper award given to the most influential paper published in SIGMOD 1998.

- "Machine Learning for Information Management: Some Promising Directions", at **The Sixth International Conference on Machine Learning and Applications (ICMLA'07),** Cincinatti, OH December 2007.
- "A Framework for Learning to Query Heterogeneous Data", at **The Third International ACM SIGMOD Workshop on Information Quality in Information Systems (IQIS 2006)**, Chicago, IL, June 2006.

- “On Beyond Hypertext: Searching in Graphs Containing Documents, Words, and Actual Data”, at **the Greater New York Area DB/IR Day 2006 (DB/IR Day 2006),** Piscataway, NJ, April, 2006.
- “A Century Of Progress On Information Integration: A Mid-Term Report”, at **the 8th International Workshop on the Web and Databases (WebDB 2005),** Baltimore, MD, June 2005.
- "Issues in Extracting Information from the Web", at the **7th International Workshop on Parsing Technologies (IWPT 2001)**, Sponsored by ACL/SIGPARSE, Beijing, China, October 2001.
- "Learning Using the Web as Background Knowledge", at **the Eleventh International Conference on Algorithmic Learning Theory (ALT-2000)**, Sydney, Australia, December 2000.
- "What can we learn from the Web?", at the **16th International Conference on Machine Learning (ICML-1999)**, Bled, Slovenia, June 1999.
- "What the Well-Informed IR Researcher Should Know About Machine Learning", at the **1996 AAAI Spring Symposium on Machine Learning and Information Access**, Palo Alto, CA, March 1996.
- "Learning to Classify English Text with ILP Methods", at the **Fifth International Workshop on Inductive Logic Programming (ILP-1995),** Leuven, Belgium, September 1995.

SEMINARS, COLLOQUIA, & TUTORIALS

- "Collaborative Filtering" (invited tutorial), at the **DIMACS Tutorial on Social Choice and Computer Science**, Piscataway, New Jersey, May 2004.
- "Information Extraction from the World Wide Web" (invited tutorial, joint with Andrew McCallum), at **The Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2003)**, August 2003; also presented at **Neural Information Processing Systems 2002 (NIPS 2002)**, Vancouver, Canada, December 2002.
- "Exploiting Document Structure in Information Extraction and Document Classification" (invited seminar), McKay Distinguished Lecture, University of California, Berkeley, October 2002.

CONFERENCE AND WORKSHOP COMMITTEES

- ICWSM-2010: Co-program chair.
- ICWSM-2009: Co-chair.
- ICML-2008; General chair.
- ICML-2006; Co-program chair.
- ICML-1994: Co-chair.
- ICML-1998, ICML-1997, ICML-1996, ICML-1995: Advisory board member.
- ICWSM-2011, AAAI-2004, SIGIR-2002, SIGIR-2001, and ICML-2000: Area chair or equivalent.
- ICML-2009, AAAI-2008, WWW-2008, NAACL-2006, KDD-2003, SIGMOD-2003, WWW-2003, NIPS-2002, ICML-2002, ICML-2001, SIGIR-2001, WWW-2000, ILP-2000, SIGIR-99, WWW-99, ILP-99, COLT-98, ICML-97, ILP-97, AAAI-96, ALT-96, ILP-95, ILP-94, AAAI-93, and ICML-93: program committee member.

MEMBERSHIPS IN PROFESSIONAL SOCIETIES

- President of the International Machine Learning Society, June 2011—present
- Founding member of the International Machine Learning Society, and Director, 1998—2008.

- Fellow of the American Association for Artificial Intelligence.

EDITORIAL POSITIONS & ADVISORY BOARDS

- *Journal of Machine Learning Research*, Sep 2001-Sep 2005; Sep 2009-present (Action Editor).
- *Transactions on Knowledge Discovery from Data*, Sep 2009-present (Action Editor).
- *The Open Systems Biology Journal,* March 2009-present (Editorial board member).
- *Artificial Intelligence*, Dec 2006-Dec 2010 (Associate Editor)
- *Machine Learning,* Jan 1997-Sep 2001, May 2005-May 2008 (Action Editor)
- *Journal of AI Research,* Jan 1995-Dec 1997 (Associate Editor)*;* Jan 1998-present (Advisory Board Member)