

## Douglas C Downey

Department of Computer Science and Engineering  
University of Washington, Box 352350  
Seattle, WA 98195-2350

ddowney@cs.washington.edu  
(206) 320-7749  
<http://www.cs.washington.edu/homes/ddowney>

Research Interests	Natural Language Processing, Machine Learning, Artificial Intelligence	
Education	<p><b>University of Washington</b> Ph.D. Computer Science and Engineering expected 2008 M.S. awarded June 2004. GPA: 3.92/4.0</p> <p><b>Case Western Reserve University</b> B.S./M.S. in Computer Science Minors in Mathematics and Economics GPA: 4.0/4.0</p>	<p>September 2002 – present</p> <p>August 1996 – May 2000</p>
Research Experience	<p><b>University of Washington</b>, Research Assistant <i>Advisor: Oren Etzioni</i></p> <p>My Ph.D. research is focused on investigating ways to exploit the redundancy present in large text collections (like the Web) to perform autonomous information extraction and classification. This work has involved both experimental and theoretical components. As an architect on the KnowItAll project, I helped develop the original design of a large-scale, autonomous information extraction system. I implemented several core components, and demonstrated the system's effectiveness experimentally (published in WWW 2004, AAAI 2004, and AIJ 2005). I also developed a formal model characterizing the redundancy found in large text collections. The resulting combinatorial balls-and-urns model makes precise how redundancy can be exploited to determine which extracted information is correct, based on the number of times and the number of different ways in which it is expressed. In experiments, the model was shown to provide substantially more accurate probabilities than previous techniques (Distinguished Paper Award, IJCAI 2005). Further, I developed a redundancy-based technique for named entity delimitation that outperforms previous machine learning techniques for complex entity names (presented at IJCAI 2007). Although redundancy-based techniques are ineffective on sparse data, I developed a bootstrapping approach based on unsupervised language models which does extend to sparse data. The technique was shown to be more accurate and substantially more scalable than previous work (ACL 2007). Lastly, I generalized the notion of bootstrapping from redundant mentions for general classification tasks, and showed the efficacy of the approach in document classification (to appear, 2008).</p> <p><b>Case Western Reserve University</b>, Research Assistant <i>Advisor: Randall D. Beer</i></p> <p>Investigated cognitive behavior using simple models. Programmed genetic algorithms to evolve neural network controllers for simulated agents exhibiting simple forms of short-term memory and selective attention. Investigated the resulting agents' behavior using tools from dynamical systems theory.</p>	<p>June 2003 – present</p> <p>October 1998 – September 2000</p>

Professional Experience	<p><b>Microsoft Corporation</b>, Research Intern  Mentor: Susan Dumais Manager: Eric Horvitz</p> <p>Performed several studies of Web search and browsing behavior, with a focus on using machine learning techniques to build novel predictive models. I developed tools for processing and summarizing search activity logs; the tools were later re-used by a number of different groups within Microsoft Research and Microsoft at large. The models showed promising predictive accuracy, and also helped to identify which factors are most important for characterizing user search activity. I also demonstrated the model's utility in a sample application, aimed at the principled prefetching of content. The research findings led to a paper selected for oral presentation at IJCAI 2007, and a poster appearing at SIGIR 2007.</p> <p><b>Intel Corporation</b>, Internet Software Engineer</p> <p>Developed heuristics to improve the ranking of documents returned by the intel.com search engine. Constructed a tool to automatically render web forms for gathering customer data from sets of arbitrary questions authored by marketing. Experience developing scalable web applications, and extensive use of .NET and SQL Server.</p>	<p>March 2006 – June 2006</p> <p>October 2000 – September 2002</p>
Teaching Experience	<p><b>University of Washington</b>, Teaching Assistant  Data Mining (CSE 546). Instructor: Oren Etzioni.</p> <p><b>University of Washington</b>, Guest Lecturer  Advanced Internet and Web Services (CSE 454). Instructor: Dan Weld.  Artificial Intelligence (CSE 573). Instructor: Oren Etzioni.</p>	<p>Spring 2005</p> <p>Autumn 2005, Autumn 2007</p>
Awards	<p>2006 Microsoft Research Fellowship, sponsored by Microsoft Live Labs  2005 IJCAI Distinguished Paper Award  2001 NSF Graduate Fellowship  2001 NDSEG Graduate Fellowship (awarded)  1996 United States Presidential Scholar</p>	
Professional Service	<p>Reviewer for:  User Modeling Conference, 2007  Data Mining and Knowledge Discovery Journal, 2007  AAAI Conference, 2005  ACM Transactions on Internet Technology, 2004</p>	
Publications	<ol style="list-style-type: none"> <li>1. Doug Downey, Stefan Schoenmackers, Oren Etzioni. Sparse Information Extraction: Unsupervised Language Models to the Rescue. <i>Proceedings of the 45<sup>th</sup> Annual Meeting of the Association of Computational Linguistics (ACL)</i>, 2007.</li> <li>2. Doug Downey, Matthew Broadhead, Oren Etzioni. Locating Complex Named Entities in Web Text. <i>Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI)</i>, 2007.</li> <li>3. Doug Downey, Susan Dumais, Eric Horvitz. Models of Searching and Browsing: Languages, Studies, and Applications. <i>Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI)</i>, 2007.</li> <li>4. Doug Downey, Susan Dumais, and Eric Horvitz. Heads and Tails: Studies of Web Search with Common and Rare Queries. <i>Proceedings of the 30th Annual International ACM Conference on Research and Development in Information Retrieval (SIGIR)</i>, 2007.</li> <li>5. Doug Downey, Oren Etzioni, and Stephen Soderland. A Probabilistic Model of Redundancy in Information Extraction. <i>Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence (IJCAI)</i>, 2005. <b>Distinguished Paper Award.</b></li> <li>6. Michael J. Cafarella, Doug Downey, Oren Etzioni, Stephen Soderland.</li> </ol>	

KnowItNow: Fast, Scalable Information Extraction from the Web. *Proceedings of the 2005 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2005.

7. Oren Etzioni, Michael Cafarella, Doug Downey, Ana-Maria Popescu, Tal Shaked, Stephen Soderland, Daniel S. Weld, and Alexander Yates. Unsupervised Named-Entity Extraction from the Web: An Experimental Study. *Artificial Intelligence* 165:91-134, 2005.
8. Doug Downey, Oren Etzioni, Daniel S. Weld, and Stephen Soderland. Learning Text Patterns for Web Information Extraction and Assessment. *Proceedings of the AAAI-04 Workshop on Adaptive Text Extraction and Mining*, 2004.
9. Oren Etzioni, Michael Cafarella, Doug Downey, Ana-Maria Popescu, Tal Shaked, Stephen Soderland, Daniel S. Weld, and Alexander Yates. Methods for Domain-Independent Information Extraction from the Web: An Experimental Comparison. *Proceedings of the Nineteenth National Conference on Artificial Intelligence (AAAI)*, 2004.
10. Oren Etzioni, Michael Cafarella, Doug Downey, Stanley Kok, Ana-Maria Popescu, Tal Shaked, Stephen Soderland, Daniel S. Weld, and Alexander Yates. Web-scale Information Extraction in KnowItAll. *Proceedings of the Thirteenth International World Wide Web Conference (WWW)*, 2004.
11. Andrew C. Slocum, Douglas C. Downey, Randall D. Beer. Further Experiments in the Evolution of Minimally Cognitive Behavior: From Perceiving Affordances to Selective Attention. *Proceedings of the Sixth International Conference on the Simulation of Adaptive Behavior*, 2000.

References

Professor Oren Etzioni  
University of Washington  
Department of Computer Science and Engineering  
Box 352350  
Seattle, WA 98195-2350  
(206) 685-3035  
[etzioni@cs.washington.edu](mailto:etzioni@cs.washington.edu)

Dr. Eric Horvitz  
Microsoft Research  
One Microsoft Way  
Redmond, WA 98052-6399  
(425) 706-2127  
[horvitz@microsoft.com](mailto:horvitz@microsoft.com)

Professor Daniel S. Weld  
University of Washington  
Department of Computer Science and Engineering  
Box 352350  
Seattle, WA 98195-2350  
(206) 543-9196  
[weld@cs.washington.edu](mailto:weld@cs.washington.edu)

Dr. Susan Dumais  
Microsoft Research  
One Microsoft Way  
Redmond, WA 98052-6399  
(425) 706-8049  
[sdumais@microsoft.com](mailto:sdumais@microsoft.com)

Professor Randall D. Beer  
Cognitive Science Program  
840 Eigenmann  
1910 E. 10th St.  
Indiana University  
Bloomington, IN 47406  
(812) 856-0873  
[rdbeer@indiana.edu](mailto:rdbeer@indiana.edu)