

HAL DAUMÉ III — CURRICULUM VITAE

(Last Updated: 04 September 2009)

Contact Information

University of Utah
School of Computing
50 South Central Campus Drive
Salt Lake City, UT 84112

phone: (801) 585-3586
fax: (801) 581-5843
email: me@hal3.name
web: <http://hal3.name/>
blog: <http://nlpers.blogspot.com/>

Employment History

Assistant Professor School of Computing, University of Utah	July 2006 – Present
Graduate Research Assistant Information Sciences Institute, University of Southern California	June 2001 – Present
Research Intern Machine Learning and Applied Statistics Group, Microsoft Research	Summer 2003

Education

University of Southern California Ph.D., Computer Science (advisor: Daniel Marcu) <i>Honors:</i> Outstanding Research Assistant Award, Dean's Doctoral Merit Fellowship	August 2006
University of Southern California M.S., Computer Science	May 2003
Carnegie Mellon University B.S., Mathematical Sciences (advisor: Rami Grossberg) <i>Honors:</i> Research Honors	August 1998 – May 2001

Research Interests

Natural Language Processing: linguistic typology, machine translation, summarization, discourse
Machine Learning: hierarchical Bayes, nonparametrics, structured prediction, domain adaptation

Student Committees

Current Students: Arvind Agarwal (PhD), Amit Goyal (PhD), Jagadeesh Jagaralamudi (PhD), Jiarong Jiang (PhD), Piyush Rai (PhD), Adam Teichert (MS).

Committee member: Venkat Anand (CS, PhD), Joshua Cates (CS, PhD), Jeff Ferraro (Biomedical Informatics, PhD), Siddharth Patwardhan (CS, PhD), Avishek Saha (PhD, CS), Shuying Shen (Health Sciences, PhD), Kannan U.V. (MS, CS), Andrew Wilder (CS, PhD), Zhou Yang (Business, MS)

Past Students: Scott Alfeld (BS, advisor; now at USC), Devyani Ghosh (MS, committee; now at Intel)

Professional Highlights

- Publications Chair, ICML** 2010
- Computational Thinking Olympiad** 2008-2009
Co-program chair of an NSF-sponsored Olympiad in *computational thinking*
- Executive Board Member, NAACL** 2007–2009
Executive Board Member (elected) of the North American Association for Computational Linguistics
- Chair, Undergraduate Curriculum Committee** 2009–Present
- Co-organizer, Prior knowledge in text and language processing** July 2008
Workshop at the International Conference on Machine Learning (ICML)
Jointly organized with Marc Dymetman, Guillaume Bouchard and Yee Whye Teh
- Research blog: Natural language processing blog** 2006 – Present
<http://nlpers.blogspot.com>
- Author: *Yet Another Haskell Tutorial*** Online Tutorial
An introduction to the Haskell programming language, used as a supplementary textbook in a functional programming course at Yale (taught by Paul Hudak), as well as several other universities (including Arizona and Waterloo); independently translated into Chinese and Portuguese. Downloaded over 100,000 times from my webpage, and mirrored extensively on the web.
- Bayesian Techniques for HLT Researchers** June 2006
Tutorial at the Human Language Technology Conference (HLT/NAACL)
- Co-organizer: Computationally Hard Problems in Speech & Language** June 2006
Workshop at the Human Language Technology Conference (HLT/NAACL)
Jointly organized with Ryan McDonald and Fernando Pereira (University of Pennsylvania)
- Co-organizer: Bayesian Methods in Natural Language Processing** December 2005
Workshop at the Conference on Neural Information Processing Systems (NIPS)
Jointly organized with Yee Whye Teh (National University of Singapore)
- Editorial Boards**
Machine Learning Journal (2008-2010)
IEEE Transactions on Speech and Language Processing (2008-2010)

Program Committees and Reviews

- Journals:** Journal of Machine Learning Research, Journal of Computational and Graphical Statistics, ACM Transactions on Speech and Language Processing, ACM Transactions on Pattern Analysis and Machine Intelligence, Journal of Artificial Intelligence Research, IBM Systems Journal;
- Conference Area Chair:** Conference on Artificial Intelligence and Statistics (2010), Association for Computational Linguistics (2008), Empirical Methods in Natural Language Processing (2007);
- Conference PCs:** American Association for Artificial Intelligence (2007, 2009), Association for Computational Linguistics (2006, 2007, 2008, 2009), Neural Information Processing Systems (2006, 2007, 2008, 2009), North American Chapter of the ACL (2006, 2007, 2008, 2009), European Conference on Machine Learning (2006), Empirical Methods in Natural Language Processing (2006, 2007, 2008), Conference on Uncertainty in Artificial Intelligence (2008, 2009), Computational Language Learning (2006, 2007, 2008, 2009), Student Workshop at ACL (2003, 2004, 2005, 2006, 2009);
- Conference Reviews:** Association for Computational Linguistics (2004, 2005, 2006), North American Chapter of the ACL (2004, 2005, 2006, 2007), International Conference on Machine Learning (2006, 2007, 2008, 2009), International Joint Conference on Natural Language Processing (2007), ACM

Special Interest Group on Information Retrieval (2006, 2007), American Association for Artificial Intelligence (2005, 2006, 2007), International Joint Conference for Artificial International (2005, 2006), World Wide Web (2005), Transactions on Asian Language Information Processing (2004);

Workshop PCs: Student Workshop at ACL (2003, 2004, 2005, 2006); Joint Inference in NLP Workshop at HLT-NAACL (2006), Analyzing Conversations in Text and Speech at HLT-NAACL (2006) Learning in Structured Outputs Workshop at ICML (2006), Task-focused Summarization and Q/A Workshop at ACL-COLING (2006), ELECTRA Workshop at SIGIR (2005).

Publications

My publications are listed by relevance to natural language processing (NLP) or machine learning (ML). Natural language processing papers are marked with a \circ and machine learning papers are marked with a \star .

Journal Publications

NLP ML

- \star **Hal Daumé III**, John Langford and Daniel Marcu. *Search-based Structured Prediction*. 2009. Machine Learning Journal 75(3).
- \star Pu Liu, Qiang Shi, **Hal Daumé III** and Gregory A. Voth. *A Bayesian statistics approach to multiscale coarse graining*. 2008. Journal of Chemical Physics 129(21), 214114.
- \star **Hal Daumé III** and Daniel Marcu. *Domain Adaptation for Statistical Classifiers*. 2006. Journal of Artificial Intelligence Research, 26(1), 101–126.
- \circ \star **Hal Daumé III** and Daniel Marcu. *Induction of Word and Phrase Alignments for Automatic Document Summarization*. 2005. Computational Linguistics, 31(4), December.
- \star **Hal Daumé III** and Daniel Marcu. *A Bayesian Model for Supervised Clustering with the Dirichlet Process Prior*. 2005. Journal of Machine Learning Research, 6(Sep), 1551–1577.

Invited Papers

NLP ML

- \circ \star **Hal Daumé III**. *Semi-supervised or semi-unsupervised?* Proceedings of the Workshop on Semi-supervised Learning in Natural Language Processing, 2009.

Conference Publications

NLP ML

- \circ \star **Hal Daumé III**. *Markov Topic Random Fields*. Proceedings of the Conference of the Association for Computational Linguistics, 2009.
- \star **Hal Daumé III**. *Bayesian Multitask Learning with Latent Hierarchies*. Proceedings of the Conference on Uncertainty in Artificial Intelligence, 2009.
- \circ \star **Hal Daumé III**. *Unsupervised Search-based Structured Prediction*. Proceedings of the International Conference on Machine Learning, 2009.
- \star Arvind Agarwal and **Hal Daumé III**. *Exponential Family Hybrid Semi-Supervised Learning*. Proceedings of the International Joint Conference on Artificial Intelligence, 2009.
- \star Piyush Rai, **Hal Daumé III** and Suresh Venkatasubramanian. *Streamed Learning: One-Pass SVMs*. Proceedings of the International Joint Conference on Artificial Intelligence, 2009.

- **Hal Daumé III.** *Non-Parametric Bayesian Areal Linguistics*. Proceedings of the Conference of the North American Association for Computational Linguistics, 2009.
- Amit Goyal, **Hal Daumé III** and Suresh Venkatasubramanian. *Streaming for Large Scale NLP: Language Modeling*. Proceedings of the Conference of the North American Association for Computational Linguistics, 2009.
- ★ Piyush Rai and **Hal Daumé III.** *The Infinite Hierarchical Latent Factor Model*. Proceedings of the Conference on Neural Information Processing Systems (NIPS) 2008.
- ★ **Hal Daumé III.** *Cross-Task Knowledge-Constrained Self Training*. Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2008.
- ★ Percy Liang, **Hal Daumé III** and Dan Klein. *Flattening Structured Classification: Trading Structure for Features*. Proceedings of the International Conference on Machine Learning (ICML), 2008.
- Ulf Hermjakob, Kevin Knight and **Hal Daumé III.** *Name Translation in Statistical Machine Translation: Learning When to Transliterate*. Proceedings of the Conference of the Association for Computational Linguistics (ACL), 2008.
- ★ Yee Whye Teh, **Hal Daumé III** and Daniel Roy. *Bayesian Agglomerative Clustering with Coalescents*. Proceedings of the Conference on Neural Information Processing Systems (NIPS) 2007.
- ★ **Hal Daumé III** and Lyle Campbell. *A Bayesian Model for Discovering Typological Implications*. Proceedings of the Conference of the Association for Computational Linguistics (ACL) 2007.
- ★ **Hal Daumé III.** *Frustratingly Easy Domain Adaptation*. Proceedings of the Conference of the Association for Computational Linguistics (ACL) 2007.
- ★ **Hal Daumé III.** *Fast Search for Dirichlet Process Mixture Models*. Proceedings of the Conference on Artificial Intelligence and Statistics (AISTats) 2007.
- ★ **Hal Daumé III** and Daniel Marcu. *Bayesian Query-Focused Summarization*. Proceedings of the Conference of the Association for Computational Linguistics (ACL) 2006.
- **Hal Daumé III** and Daniel Marcu. *A Large-Scale Exploration of Effective Global Features for a Joint Entity Detection and Tracking Model*. 2005. Proceedings of the Joint Conference on Human Language Technologies and Empirical Methods in Natural Language Processing (HLT/EMNLP), 97–104. Vancouver, Canada.
- ★ **Hal Daumé III** and Daniel Marcu. *Learning as Search Optimization: Approximate Large Margin Methods for Structured Prediction*. 2005. Proceedings of the International Conference on Machine Learning (ICML), 169–176. Bonn, Germany.
- ★ **Hal Daumé III** and Daniel Marcu. *A Phrase-Based HMM Approach to Document/Abstract Alignment*. 2004. Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 119–126. Barcelona, Spain.
- ★ **Hal Daumé III** and Daniel Marcu. *NP Bracketing by Maximum Entropy Tagging and SVM Reranking*. 2004. Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 254–261. Barcelona, Spain.

- **Hal Daumé III** and Eric Brill. *Web Search Intent Induction via Automatic Query Reformulation*. 2004. Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics (NAACL) Short Paper Session, 49–52. Boston, MA.
- **Hal Daumé III** and Daniel Marcu. *A Noisy-Channel Model for Document Compression*. 2002. Proceedings of the Conference of the Association for Computational Linguistics (ACL), 449–456, Philadelphia, PA.
- Eric Nyberg and **Hal Daumé III**. *Integrated Information Management: An Interactive, Extensible Architecture for Information Retrieval*. 2001. In the Proceedings of the Conference on Human Language Technologies (HLT). San Diego, CA.

Book Chapters

NLP ML

- **Hal Daumé III** and Daniel Marcu. *Learning as Search Optimization*. 2007. In *Predicting Structured Data*; G. Bakir, T. Hofman, B. Schölkopf, A. Smola, B. Taskar, S.V.N. Vishwanathan, eds. MIT Press.

Refereed Workshop Publications

NLP ML

- ★ Devyani Ghosh, John Carter and **Hal Daumé III**. *Perceptron-based Coherence Predictors*. 2008. Proceedings of the Workshop on Chip Multiprocessor Memory Systems and Interconnects (CMP-MSI) at the International Symposium on Computer Architecture (ISCA).
- ★ **Hal Daumé III**, John Langford and Daniel Marcu. *Search-Based Structured Prediction as Classification*. 2005. Proceedings of the Workshop on Advances in Structured Learning for Text and Speech Processing at NIPS 2005. Vancouver, Canada.
- ★ **Hal Daumé III** and Daniel Marcu. *Supervised Clustering with the Dirichlet Process*. 2004. Proceedings of the Conference on Neural Information Processing Systems Workshop on Structured Outputs. Whistler, Canada.
- **Hal Daumé III** and Daniel Marcu. *Generic Sentence Fusion is an Ill-Defined Summarization Task*. 2004. Proceedings of the Text Summarization Branches Out Workshop at ACL. Barcelona, Spain.
- **Hal Daumé III**, Kevin Knight, Irene Langkilde-Geary, Daniel Marcu and Kenji Yamada. *The Importance of Lexicalized Syntax Models for Natural Language Generation Tasks*. 2002. Proceedings of the International Conference on Natural Language Generation (INLG), 9–16. New York, NY.

Unrefereed Workshop Publications

NLP ML

- **Hal Daumé III** and Daniel Marcu. *Bayesian Summarization at DUC and a Suggestion for Extrinsic Evaluation*. 2005. Proceedings of the Document Understanding Conference (DUC) at HLT/EMNLP. Vancouver, Canada.
- ★ **Hal Daumé III** and Daniel Marcu. *Bayesian Multi-Document Summarization at MSE*. 2005. Proceedings of the Workshop on Multilingual Summarization Evaluation (MSE) at ACL. Ann Arbor, MI.
- ★ **Hal Daumé III** and Daniel Marcu. *A Tree-Position Kernel for Document Compression*. 2004. Proceedings of the Document Understanding Conference (DUC). Boston, MA.
- **Hal Daumé III**, Abdesammad Echihabi, Daniel Marcu, Dragos Stefan Munteanu and Radu Soricut. 2002. *GLEANS: A Generator of Logical Extracts and Abstracts for Nice Summaries*. Proceedings of the Document Understanding Conference (DUC), 9–14. Philadelphia, PA.

Book Reviews

NLP ML

- **Hal Daumé III.** *Automatic Summarization (by Inderjeet Mani)*. 2005. Appeared in Machine Translation.

Unrefereed Tutorials

NLP ML

- **Hal Daumé III.** *Yet Another Haskell Tutorial*. An introduction to the Haskell programming language, used at Yale, independently translated into Chinese and Portuguese. 200pp.
- ★ **Hal Daumé III.** *Support Vector Machines for Natural Language Processing*. Supplementary material for guest lectures at the University of Southern California. 25pp.

Publicly Available Software

MEGAM: Optimization software for maximum entropy models; implementation based on conjugate gradient and limited-memory BFGS methods. (Downloaded over 20,000 times.)
<http://hal3.name/megam/>

HBC: Software that compiles statistical models into executable C code. (Downloaded over 6,000 times.)
<http://hal3.name/HBC/>.

TAGCHUNK: Software for performing joint inference over part of speech tags and syntactic chunks. (Downloaded over 400 times.)
<http://hal3.name/TagChunk/>

SVMSEQUEL: Optimization software for support vector machines; implementation based on the kernel minover method. Supports linear, polynomial and RBF kernels, as well as string and tree kernels. (Downloaded over 900 times.)
<http://hal3.name/SVMsequel/>

FASTDT: Software for learning decision trees, supporting boosting and bagging. (Downloaded over 750 times.)
<http://hal3.name/FastDT/>

... many more, see <http://hal3.name/software.html>.

Teaching Experience

CS 5300/6300, Artificial Intelligence U Utah, CS
 Spring 2009 (enrollment: 78)
 Received Dean's letter for Excellence in Teaching

CS 5350/6350, Machine Learning U Utah, CS
 Spring 2007, Spring 2008, Fall 2008, Fall 2009 (enrollments: 28, 32, 36, 46)

CS 5964/6964, Applications of Natural Language Processing U Utah, CS
 Fall 2007 (enrollment: 28)

CS 7941, Topics in Machine Learning U Utah, CS
 Every semester, 2007–current (enrollments: about 15)
 Received Dean's letter for Excellence in Teaching

CSCI 544, Natural Language Processing (Teaching Assistant) USC, CS
 Spring 2005; *Professors:* Eduard Hovy and Daniel Marcu
 Received Outstanding TA Award

CSCI 544, Natural Language Processing (Invited Lecturer) Spring 2004; <i>Professors:</i> Eduard Hovy and Daniel Marcu	USC, CS
CSCI 599, Topics in Statistical Learning (Invited Lecturer) Spring 2004; <i>Professor:</i> Stefan Schaal	USC, CS
CSCI 562, Empirical Methods in NLP (Invited Lecturer) Fall 2003; <i>Professor:</i> Kevin Knight	USC, CS
21-229, Set Theory (Grader) Spring 2000; <i>Professor:</i> Rami Grossberg	CMU, Math
21-115/116, Calculus I & II (Teaching Assistant) Fall 1999; <i>Professor:</i> Russell Walker	CMU, Math

Invited Talks

Statistics, Typology and Language Processing Department of Linguistics, University of Utah	October, 2009
{ϵ-, Semi-, Un-}supervised Search-based Structured Prediction Yahoo! Research, New York	May, 2009
Non-parametric Models with Latent Hierarchies Columbia University, Department of Statistics	April, 2009
Domain Adaptation in Natural Language Johns Hopkins University, Center for Language and Speech Processing	March, 2009
Domain Adaptation in Natural Language University of Texas at Austin, Computer Science	November, 2008
Summer School in Computational- and Psycho-Linguistics University of Edinburgh	September 2008
Automatic Document Summarization Information Sciences Institute	January 2007
Search-based Structured Prediction University of Maryland, College Park	March 2007
Bayesian Techniques for Natural Language Processing Microsoft Research Asia, Beijing, China	July 2007
Bayesian Techniques for HLT Researchers Tutorial at the Human Language Technology Conference (HLT/NAACL)	June 2006
Statistical Learning Theory and SVMs for AI Researchers Tutorial at the ISI Intelligent Systems Division Retreat	February 2006
NLP for Smart People Who Know Nothing about NLP Toyota Technological Institute at Chicago	January 2006
Bayesian Methods for Natural Language Processing Tutorial at the Information Sciences Institute, AI seminar	March 2005

Funding

Statistical Linguistic Typology NSF-IIS Small Grant, \$550k over three years. Sole PI.	2009–2012
--	-----------

Graph Algorithms for Large-Data NLP Google Research Grant, \$55k; Co-PI with Suresh Venkatasubramanian.	2009–2010
Language Understanding for Research Papers U Utah URC Faculty Research & Creative Grant Proposal, \$6k. Sole PI.	2009
Advanced Learning Technology for Translation U Utah Seed Grant, \$35k. Sole PI.	2009
Computational Thinking Olympiad: Brainstorming Workshop NSF-IIS grant, \$20k over two years. Sole PI.	2008
Cross-task Learning for Natural Language Processing (REU) NSF-IIS Research Experience for an Undergraduate Supplement, \$6k over one year. Sole PI.	2008
Cross-task Learning for Natural Language Processing NSF-IIS Small Grant, \$404k over three years. Sole PI.	2007–2010

External Service

Publications Chair, ICML 2010

Publicity Chair, COLING 2010

Computational Thinking Olympiad

Co-program chair of an NSF-sponsored Olympiad in *computational thinking*

Executive Board Member, NAACL

Executive Board Member (elected) of the North American Association for Computational Linguistics

Publicity Chair, ACL 2008

Student Research Workshop Advisor, ACL

Research blog: Natural language processing blog

Since 2006, <http://nlpers.blogspot.com>

Co-organizer, Prior knowledge in text and language processing

Workshop at the International Conference on Machine Learning (ICML), 2008

Co-organizer, Computationally Hard Problems in Speech & Language

Workshop at the Human Language Technology Conference (HLT/NAACL), 2006

Co-organizer, Bayesian Methods in Natural Language Processing

Workshop at the Conference on Neural Information Processing Systems (NIPS) 2005

Internal Service

Curriculum Committee Chair, 2008 – Present

Overseen significant advances in math/science requirements, formation of undergraduate tracks, and development of four new undergraduate courses

Masters in Statistics Program, SoC Representative

Representative from the School of Computing to the Masters in Statistics Program (2008–present)

Committee Memberships

Admissions (2007, 2008); Curriculum (2007, 2008); College Council (2007, 2008)

Organizer, Natural Language Seminar Series

Weekly seminar at the Information Sciences Institute (January 2001 – May 2006)

Extracurricular Activities

Aikido — Aikido Schools of Ueshiba, sankyuu

Badminton (indoor – tournament level)

Skiing

Rock Climbing