

## Matthew Lease

---

**CONTACT INFORMATION**      School of Information      Voice: (512) 471-9350  
University of Texas at Austin      Fax: (512) 471-3971  
1616 Guadalupe Ste. 5.202      E-mail: ml@ischool.utexas.edu  
Austin, TX 78701-1213 USA      <http://www.ischool.utexas.edu/~ml>

**INTERESTS**      Information Retrieval (IR), Natural Language Processing (NLP), Machine Learning

**EDUCATION**

**Brown University**, Providence, RI USA  
**Ph.D., Computer Science**, September 2009 (conferred May 2010)  
*Beyond keywords: finding information more accurately and easily using natural language*  
Committee: Eugene Charniak (advisor), Mark Johnson, and James Allan (University of Massachusetts Amherst)

**Brown University**, Providence, RI USA  
**M.S., Computer Science**, May 2004  
*Parsing Biomedical Literature* [15]  
Committee: Eugene Charniak (advisor), Mark Johnson, and Thomas Dean

**University of Washington**, Seattle, WA USA  
**B.S., Computer Science**, June 1999

**HIGHLIGHTS**

John P. Commons Teaching Fellowship, UT Austin, 2010-2011  
van Dam Fellowship, Brown University, 2009  
Best results, TREC 2008 Relevance Feedback track [8]  
Best Paper Award, CIKM 2007 Ph.D. Workshop [10]  
Post-workshop Research Award, Johns Hopkins University Center for Language and Speech Processing (CLSP), 2005 (1-year support) [17]  
Best results, NIST 2004 RTF Evaluation [12] (6/6 participating tasks)  
Best biomedical text parser [15] at publication  
University Fellowship, Brown University, 2002

**EXPERIENCE**

*Assistant Professor*, School of Information and Department of Computer Science, University of Texas at Austin, Austin, TX (08/09-present)

*Ph.D. Student*, Brown Laboratory for Linguistic Information Processing (BLLIP), Brown University, Providence, RI (09/02-08/09)

*Visiting Researcher*, Spoken Language Systems Laboratory (LSV), Saarland University, Saarbrücken, Germany (06/09-08/09)

*Visiting Researcher*, Center for Intelligent Information Retrieval (CIIR), University of Massachusetts, Amherst, MA (06/08-05/09) [7]

*Visiting Student*, Institute of Formal and Applied Linguistics (ÚFAL), Charles University, Prague, Czech Republic (01/07-06/07) [11]

*Group Member*, Parsing and Spoken Structural Event Detection (PASSED) group, Johns Hopkins University Center for Language and Speech Processing (CLSP) Summer Workshop, Baltimore, MD (06/05-08/05) [18, 19, 20, 21]

*Visiting Student*, Signal, Speech and Language Interpretation (SSLI) lab, University of Washington, Seattle, WA (06/04-08/04) [22]

*Technical Graduate Intern*, Intel Research Seattle, Seattle, WA (10/01-8/02) [28, 29]  
*Research Engineer*, LizardTech Inc., Seattle, WA (06/99-08/01) [? ? ] (lizardtech.com)  
*Undergraduate assistant*, University of Washington Medical Center,  
Seattle, WA (06/98-08/99) [30, 31]  
*Programmer*, HyperBole Studios, Seattle, WA (01/96-12/98)  
Credited work on *The X-Files* computer game, part-time job during college

## PUBLICATIONS

### Information Retrieval (IR)

- [1] Ramona Broussard, Yongyi Zhou, and Matthew Lease. Mobile phone search for library catalogs. In *Proceedings of the ASIS&T 2010 Annual Meeting*, 2010. Short paper, 24% acceptance rate.
- [2] Ramona Broussard, Yongyi Zhou, and Matthew Lease. University of texas mobile library search. In *Proceedings of the ASIS&T 2010 Annual Meeting*, 2010. Poster.
- [3] Catherine Grady and Matthew Lease. Crowdsourcing document relevance assessment with mechanical turk. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, pages 172–179, Los Angeles, June 2010. Association for Computational Linguistics.
- [4] Saeedeh Momtazi, Matthew Lease, and Dietrich Klakow. Effective Term Weighting for Sentence Retrieval. In *Proceedings of the 14th European Conference on Research and Advanced Technology for Digital Libraries (ECDL)*, 2010. Poster.
- [5] Marc Cartright, Jangwon Seo, and Matthew Lease. UMass Amherst and UT Austin at the TREC'09 Relevance Feedback Track. In *Proceedings of the Eighteenth Text Retrieval Conference (TREC 2009)*, 2010.
- [6] Matthew Lease. An improved markov random field model for supporting verbose queries. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pages 476–483. ACM, 2009.
- [7] Matthew Lease, James Allan, and W. Bruce Croft. Regression Rank: Learning to Meet the Opportunity of Descriptive Queries. In *Proceedings of the 31st European Conference on Information Retrieval (ECIR)*, pages 90–101, 2009.
- [8] Matthew Lease. Incorporating Relevance and Psuedo-relevance Feedback in the Markov Random Field Model: Brown at the TREC'08 Relevance Feedback Track. In *Proceedings of the Seventeenth Text Retrieval Conference (TREC 2008)*, 2009. Best results in track.
- [9] Matthew Lease and Eugene Charniak. A Dirichlet-smoothed Bigram Model for Retrieving Spontaneous Speech. In *Advances in Multilingual and Multimodal Information Retrieval: 8th Workshop of the Cross-Language Evaluation Forum (CLEF)*, volume 5152 of *Lecture Notes in Computer Science*, pages 687–694, 2008.
- [10] Matthew Lease. Natural Language Processing for Information Retrieval: the time is ripe (again). In *Proceedings of the 1st Ph.D. Workshop at the 16th ACM Conference on Information and Knowledge Management (CIKM)*, 2007. Best paper award.
- [11] Matthew Lease and Eugene Charniak. Brown at CLEF'07: Cross-language speech retrieval. In *Cross-Language Evaluation Forum (CLEF) Working Notes*, 2007.

## Natural Language Processing (NLP)

- [12] Matthew Lease, Mark Johnson, and Eugene Charniak. Recognizing disfluencies in conversational speech. *IEEE Transactions on Audio, Speech and Language Processing*, 14(5):1566–1573, September 2006.
- [13] Matthew Lease and Mark Johnson. Early deletion of fillers in processing conversational speech. In *Proceedings of Human Language Technology (HLT) Conference and Conference on Empirical Methods in Natural Language Processing (EMNLP), Companion Volume: Short Papers*, pages 73–76, 2006.
- [14] Matthew Lease, Eugene Charniak, Mark Johnson, and David McClosky. A Look at Parsing and Its Applications. In *Proceedings of the Twenty-First National Conference on Artificial Intelligence (AAAI): New Scientific and Technical Advances in Research (Nectar)*, 2006.
- [15] Matthew Lease and Eugene Charniak. Parsing biomedical literature. In R. Dale, K.-F. Wong, J. Su, and O. Kwong, editors, *Proceedings of the 2nd International Joint Conference on Natural Language Processing (IJCNLP)*, volume 3651 of *Lecture Notes in Computer Science*, pages 58 – 69. Springer-Verlag, 2005.
- [16] Matthew Lease, Eugene Charniak, and Mark Johnson. Parsing and its applications for conversational speech. In *Proceedings of the IEEE 30th Intl. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 5, pages 961–964, 2005.
- [17] Matthew Lease. Parsing and disfluency modeling. Technical Report CS-05-15, Brown University Department of Computer Science, 2005.
- [18] Ann Bies, Stephanie Strassel, Haejoong Lee, Kazuaki Maeda, Seth Kulick, Yang Liu, Mary Harper, and Matthew Lease. Linguistic resources for speech parsing. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC)*, 2006.
- [19] John Hale, Izhak Shafran, Lisa Yung, Bonnie J. Dorr, Mary Harper, Anna Krasnyanskaya, Matthew Lease, Yang Liu, Brian Roark, Matthew Snover, and Robin Stewart. PCFGs with syntactic and prosodic indicators of speech repairs. In *Proceedings of the 21st Intl. Computational Linguistics (COLING) Conference and 44th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 161–168, 2006.
- [20] Brian Roark, Yang Liu, Mary Harper, Robin Stewart, Matthew Lease, Matthew Snover, Izhak Shafran, Bonnie Dorr, John Hale, Anna Krasnyanskaya, and Lisa Yung. Reranking for sentence boundary detection in conversational speech. In *Proceedings of IEEE 31st International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 545–548, 2006.
- [21] Brian Roark, Mary Harper, Eugene Charniak, Bonnie Dorr, Mark Johnson, Jeremy G. Kahn, Yang Liu, Mari Ostendorf, John Hale, Anna Krasnyanskaya, Matthew Lease, Izhak Shafran, Matthew Snover, Robin Stewart, and Lisa Yung. Sparseval: Evaluation metrics for parsing speech. In *5th International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy, 2006.
- [22] Jeremy G. Kahn, Matthew Lease, Eugene Charniak, Mark Johnson, and Mari Ostendorf. Effective use of prosody in parsing conversational speech. In *Proceedings of Human Language Technology (HLT) Conference and Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 233–240, 2005.

- [23] Mary Harper, Bonnie Dorr, John Hale, Brian Roark, Izhak Shafran, Matthew Lease, Yang Liu, Matthew Snover, Lisa Yung, Anna Krasnyanskaya, and Robin Stewart. *Johns Hopkins Summer Workshop Final Report on Parsing and Spoken Structural Event Detection*, 2005.
- [24] Mark Johnson, Eugene Charniak, and Matthew Lease. An improved model for recognizing disfluencies in conversational speech. In *Proceedings of the Rich Transcription Fall Evaluation (RT-04F)*, 2004. Best results for all six tasks in which we participated. <http://www.nist.gov/speech/tests/rt/2004-fall>.

### Ubiquitous Computing

- [25] Matthew Lease and Guy Eddon. SmartElevator: Revitalizing a legacy device through inexpensive augmentation. In *Proceedings of the IEEE 23rd International Conference on Distributed Computing Systems (ICDCS): 3rd International Workshop on Smart Appliances and Wearable Computing*, pages 254 – 259, 2003.
- [26] Matthew Lease. How a smart elevator changed our way of life forever. *Conduit: Brown University Department of Computer Science Newsletter*, 12(1):7 – 9, 2003.
- [27] Matthew Lease. Plan-Aware Behavioral Modeling. In *Adjunct Proceedings of 4th Intl. Conference on Ubiquitous Computing (UBICOMP)*, pages 35–36, 2002. Poster presentation, [http://www.viktoria.se/ubicomp/ubicomp\\_adjunct\\_proceedings.pdf](http://www.viktoria.se/ubicomp/ubicomp_adjunct_proceedings.pdf).
- [28] A. LaMarca, W. Brunette, D. Koizumi, M. Lease, S.B. Sigurdsson, K. Sikorski, D. Fox, and G. Borriello. PlantCare: An Investigation in Practical Ubiquitous Systems. In *Proceedings of the 4th International Conference on Ubiquitous Computing (UBICOMP)*, volume 2498 of *LECTURE NOTES IN COMPUTER SCIENCE*, pages 316–332. Springer, 2002.
- [29] Anthony LaMarca, Waylon Brunette, David Koizumi, Matthew Lease, Stefan B. Sigurdsson, Kevin Sikorski, Dieter Fox, and Gaetano Borriello. Making sensor networks practical with robots. In *Pervasive '02: Proceedings of the First International Conference on Pervasive Computing*, volume 2414 of *LECTURE NOTES IN COMPUTER SCIENCE*, pages 152–166. Springer-Verlag, 2002.

### Medical Informatics

- [30] I.J. Kalet, J. Wu, M. Lease, M.M. Austin-Seymour, J.F. Brinkley, and C. Rosse. Anatomical information in radiation treatment planning. In *Proceedings of the American Medical Informatics Association (AMIA) Fall Symposium*, 1999.
- [31] I.J. Kalet, R.S. Giansiracusa, C. Wilcox, and M. Lease. Radiation Therapy Planning: an Uncommon Application of Lisp. In R. Gabriel, editor, *Proceedings of the Conference on the 40th Anniversary of Lisp*, 1998.

### INVITED TALKS

*Beyond keywords: finding information more accurately and easily using natural language*

- Max Planck Institut für Informatik, Saarbrücken, Germany (08/18/09)
- German Research Center for Artificial Intelligence (DFKI), Saarbrücken, Germany (06/24/09)
- NSF Partnerships for International Research and Education (PIRE) annual meeting, Irsee, Germany (06/11/09)
- School of Information, University of Texas at Austin, Austin, TX (05/08/09)
- AT&T Labs, Florham Park, NJ (05/05/09)

- Human Language Technology Center of Excellence, Johns Hopkins University, Baltimore, MD (04/23/09)
- Yahoo! Paris, France (04/10/09)
- Google Research, Zurich, Switzerland (04/06/09)
- University of Massachusetts, “Machine Learning and Friends” Computer Science Departmental Seminar, Amherst, MA (04/01/09)
- Yahoo! Labs, Montreal, Canada (03/20/09)
- University of Montreal, Montreal, Canada (03/19/09)

*Finding information more accurately and easily using natural language* (12/07/2008)  
NSF Partnerships for International Research and Education (PIRE) annual meeting, Institute of Formal and Applied Linguistics, Charles University, Prague, Czech Republic

*NLP + IR: Worth Another Look?* (6/18/08)  
Center for Intelligent Information Retrieval, University of Massachusetts, Amherst, MA

*Information Retrieval via Statistical Language Modeling* (6/6/07)  
International Research Training Group (IRTG) meeting, Language Technology and Cognitive Systems, Saarland University and University of Edinburgh, Trier, Germany

*Information Retrieval via Statistical Language Modeling* (3/5/07)  
Institute of Formal and Applied Linguistics, Charles University, Prague, Czech Republic

*Parsing and Disfluency Modeling for Conversational Speech* (3/13/06)  
2006 Brown-MIT-Harvard NLP summit, Brown University, Providence, RI

*Disfluencies and Parsing: English and Beyond* (8/17/05)  
Center for Language and Speech Processing (CLSP) Summer Workshop, Baltimore, MD

## GRANTS

Amazon, Web Services in Education program, *The 2010 Text REtrieval Conference meets MTurk: Achieving Scalable Search Engine Evaluation via Crowdsourcing*, \$4,000, August 2010-August 2011.

UT Austin Longhorn Innovation Fund for Technology (LIFT), *Enabling Data-Intensive Research and Education at UT Austin via Cloud Computing*, \$94,000, September 2010-August 2010, PI (co-PIs: J. Baldrige and W. Xu.)

Science and Technology Foundation of Portugal (FCT), *REACTION: Retrieval, Extraction and Aggregation Computing Technology for Integrating and Organizing News*, 220,000 €, September 2010-August 2013, co-PI (PI: M. Silva).

Google, Android for Education program, *Integrating the Android Platform Across IR, HCI, and Usability Courses*, \$6,000, April 2010, PI (co-PIs: R. Bias, L. Francisco-Revilla, and U. Karadkar).

Amazon, Web Services in Education program, *Discovering Patterns of Information Flow and Alteration across the Web*, \$3,500, March 2010-March 2011.

## PROFESSIONAL SOCIETIES

Association for Computing Machinery (ACM)  
ACM Special Interest Group on Information Retrieval (SIGIR)  
The American Society for Information Science & Technology (ASIS&T)

ASIS&T Special Interest Group on Information Needs, Seeking and Use (SIGUSE)  
Association for Computational Linguistics (ACL)

PROFESSIONAL  
SERVICE

**Organizer**

WSDM 2011 Workshop on Crowdsourcing for Search and Data Mining (CSDM 2011)  
SIGIR 2010 Workshop on Crowdsourcing for Search Evaluation (CSE 2010)  
TREC 2010 Relevance Feedback Track  
2010 UT Austin Workshop: Student Research in Information Retrieval and Web Search  
2006 Brown-MIT-Harvard NLP Summit

**Proceedings Chair:** CIKM 2011

**Program Committee**

COLING/ACL 2006 Student Research Workshop  
CrowdConf 2010  
EMNLP 2010 (multiple tracks)  
SIGIR 2010 (multiple tracks)  
SIGIR 2010 Workshop on Workshop on Query Representation and Understanding

**Reviewer**

ASRU 2007, 2009  
ASIS&T 2010 (Posters/Demos)  
COLING 2008  
EMNLP 2009  
ICASSP 2008-2009  
NAACL/HLT 2010  
NIPS 2010  
UBICOMP 2002  
ACM Computing Surveys (2009)  
Information Processing & Management (2009)

TEACHING

Academic Teaching Certificate, Sheridan Center for Teaching in Higher Education,  
Brown University (Spring 2005)

Two-semester training program including lectures, moderated break-out sessions, readings and exercises, and filmed observation/critique of my teaching by center staff

INF385T/CS395T: *Topics in Information Retrieval and Web Search* (Spring/Fall 2010)

INF383D: *Mathematical Foundations of Information Studies* (Spring 2010)

INF384C: *Organizing Information* (Fall 2010)

CS241: *Statistical Models in Natural-Language Processing* (Fall 2006)

Co-taught with Prof. Eugene Charniak, I gave the lectures, voluntary appointment

**Guest Lectures - UT Austin**

CS 371R: *Information Retrieval and Web Search* (04/15/10)

INF 397: *Research in Information Studies: Oral Narrative as History* (04/06/10)

INF 391D.9: *Introduction to Doctoral Research and Theory* (03/30/10)

INF 385E: *Information Architecture and Design* (03/22/10)

UGS 303: *Knowing, Acquiring and Sharing Information in the Information Age* (02/15/10)

**Guest Lectures - Brown University**

CS 253: *Internet and Web Algorithms* (04/02/09)

CG 116: *Introduction to Computational Linguistics* (11/16/04)

STUDENT  
MENTORING

**University of Texas at Austin, School of Information**

Catherine Grady, MSIS Program [3] (Fall 2009-present)

Hohyon Ryu, PhD Program (Fall 2010-present)

Ramona Broussard and Yonyi Zhou, MSIS Program [1, 2] (Spring 2010-present)

**University of Texas at Austin, Department of Computer Science**

Eunho Yang, PhD Program, co-advised with Pradeep Ravikumar (Spring 2010-present)

Karan Sangani, undergraduate, *Scientists for Tomorrow* program (Fall 2010-present)

Ekongi Yancy, undergraduate, *Scientists for Tomorrow* program (Fall 2010-present)

**Saarland University, Spoken Language Systems Laboratory (LSV)**

Saeedeh Momtazi, PhD Program, co-advised with Dietrich Klakow, research project [4]

**University of Massachusetts Amherst, Department of Computer Science**

Marc Cartright and Jangwon Seo, PhD Program, research project [5]

**Brown University, Department of Computer Science**

David Ellis, undergraduate, co-advised with Eugene Charniak (Spring 2005-Fall 2006)

STUDENT  
COMMITTEES

Reader for Sung Ju Hwang, M.A., Computer Science (May 2010)

*Reading Between The Lines: Object Localization Using Implicit Cues from Image Tags*

UNIVERSITY  
SERVICE

**University of Texas at Austin**

Faculty Advisor, UT Austin ASIS&T student chapter (09/09-present)

Chair, Research Environment Committee, School of Information (08/10-present)

Member, Doctoral Studies Committee, School of Information (12/09-08/10)

**Brown University**

Community Council, chair: President Ruth Simmons (03/05-11/06)

Graduate Student Council (01/03-12/06)