

**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 at utexas dot edu  
Austin, TX 78701-1213 USA      <http://www.ischool.utexas.edu/~ml>

## Education

BROWN UNIVERSITY  
PROVIDENCE, RI      **Ph.D., Computer Science**, August 2009 (conferred May 2010)  
*Beyond keywords: finding information more accurately & easily using natural language*  
Committee: Eugene Charniak (chair), Mark Johnson (Macquarie University, Australia),  
& James Allan (University of Massachusetts, Amherst, MA)

**M.S., Computer Science**, May 2004  
*Parsing Biomedical Literature* [38]  
Committee: Eugene Charniak (advisor), Mark Johnson, & Thomas Dean

UNIVERSITY OF WASHINGTON  
SEATTLE, WA      **B.S., Computer Science**, June 1999  
[Washington Scholar](#) award recipient (full scholarship)

## Professional Experience

*Associate Professor* (08/15-present)  
*Assistant Professor* (08/09-08/15)  
School of Information, University of Texas at Austin

*Consulting Visiting Professor*, [CrowdFlower](#)  
San Francisco, CA (05/12-08/12)

*Technical Graduate Intern*, Intel Research Seattle (renamed the [Intel Science & Technology Center for Pervasive Computing](#)), Seattle, WA (10/01-8/02) [40, 41, 62]

*Research Engineer*, [LizardTech Inc.](#), Seattle, WA (06/99-08/01)

*Programmer*, [HyperBole Studios](#), Seattle, WA (01/96-12/98, part-time)  
Software developer on *The X-Files* game

## Additional Academic Appointments

*Courtesy Faculty Appointment* (08/09-present)  
*Graduate Studies Committee* (GSC) (08/16-present)  
Department of Computer Science

## University of Texas at Austin

*Associated Faculty*, [Department of Statistics and Data Sciences](#) (03/10-present)

*Associated Faculty*, [Computational Linguistics Laboratory](#) (08/09-present)

*Affiliated Faculty*, [Robotics Graduate Portfolio Program](#) (08/15-present)

## Previous Academic Appointments

*Research Assistant*, [Brown Laboratory for Linguistic Information Processing \(BLLIP\)](#),  
Brown University, Providence, RI (09/02-08/09)

*Visiting Student*, [Spoken Language Systems Laboratory \(LSV\)](#), Saarland  
University, Saarbrücken, Germany (06/09-08/09)

*Visiting Student*, [Center for Intelligent Information Retrieval \(CIIR\)](#),  
University of Massachusetts, Amherst, MA (06/08-05/09) [33, 32]

Visiting Student, [Institute of Formal and Applied Linguistics \(ÚFAL\)](#),  
Charles University, Prague, Czech Republic (01/07-06/07) [105]

Visiting Student, [Parsing and Spoken Structural Event Detection \(PASSED\) project](#),  
Johns Hopkins University Center for Language and Speech Processing (CLSP)  
Summer Workshop, Baltimore, MD (06/05-08/05) [34, 37, 36, 35]

Visiting Student, [Signal, Speech and Language Interpretation \(SSLI\) lab](#), University  
of Washington, Seattle, WA (06/04-08/04) [39]

Undergraduate Research Assistant, [Dept. of Radiation Oncology](#), University of Wash-  
ington, Seattle, WA (06/98-08/99) [42, 95]

**Research Areas** Crowdsourcing & Human Computation (HCOMP)  
Information Retrieval (IR)  
Natural Language Processing (NLP)

**Honors and  
Awards**

*Best Paper Award*, AAAI HCOMP 2016 [13]  
*Samsung HumanTech Paper Award*, 2015 [17]  
*Modeling Challenge Award*, 5th International Conference on Social Computing,  
Behavioral-Cultural Modeling, and Prediction (SBP), 2012  
*Plenary Keynote*, *Intl. Joint Conf. on Natural Language Processing (IJCNLP)*, 2011  
*Best Reviewer Award* (with Elias Ponvert), 2010 Conference on Empirical Methods on  
Natural Language Processing (EMNLP)  
*Best results*, TREC 2008 Relevance Feedback track [103]  
*Best Paper Award*, ACM CIKM Ph.D. Workshop, 2007 [74]  
*Best results*, NIST 2004 [Rich Transcription Fall Evaluation](#) [3] (6/6 participating tasks)

**Metrics** h-index: 29, i10-index: 55, Citations: 3296 ([Google Scholar](#), as of September 2018)  
Talk slides viewed over 15,000 times on [SlideShare](#) in 2016

**Publications** **Refereed Journal Papers**

- [1] Mucahid Kutlu, Tamer Elsayed, and Matthew Lease. Learning to Effectively Select Topics For Information Retrieval Test Collections. *Information Processing and Management (IPM)*, 54(1):37–59, 2018.
- [2] Kezban Dilek Onal, Ye Zhang, Ismail Sengor Altingovde, Md Mustafizur Rahman, Pinar Karagoz, Alex Braylan, Brandon Dang, Heng-Lu Chang, Henna Kim, Quinten McNamara, Aaron Angert, Edward Banner, Vivek Khetan, Tyler McDonnell, An Thanh Nguyen, Dan Xu, Byron C. Wallace, Maarten de Rijke, and Matthew Lease. Neural Information Retrieval: At the End of the Early Years. *Information Retrieval*, 21(2-3):111–182, 2018.
- [3] 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.

**Refereed Conference Full Papers**

- [4] Tanya Goyal, Tyler McDonnell, Mucahid Kutlu, Tamer Elsayed, and Matthew Lease. Your Behavior Signals Your Reliability: Modeling Crowd Behavioral Traces to Ensure Quality Relevance Annotations. In *6th AAAI Conference on Human Computation and Crowdsourcing (HCOMP)*, pages 41–49, 2018.

- [5] Mucahid Kutlu, Tamer Elsayed, Maram Hasanain, and Matthew Lease. When rank order isn't enough: New statistical-significance-aware correlation measures. In *Proceeding of the 27th ACM International Conference on Information and Knowledge Management (CIKM)*, 2018.
- [6] Mucahid Kutlu, Tyler McDonnell, Aashish Sheshadri, Tamer Elsayed, and Matthew Lease. Mix and Match: Collaborative Expert-Crowd Judging for Building Test Collections Accurately and Affordably. In *Proceedings of the 1st Biannual Conference on the Design of Experimental Search & Information REtrieval Systems (DESIRES)*, pages 42–46, 2018.
- [7] Mucahid Kutlu, Tyler McDonnell, Yasmine Barkallah, Tamer Elsayed, and Matthew Lease. What Can Rationales behind Relevance Judgments Tell Us About Assessor Disagreement? In *Proceedings of the 41st international ACM SIGIR conference on Research and development in Information Retrieval*, pages 805–814, 2018.
- [8] An Thanh Nguyen, Aditya Kharosekar, Aditya Kharosekar, Saumyaa Krishnan, Siddhesh Krishnan, Elizabeth Tate, Byron C. Wallace, and Matthew Lease. Believe it or not: Designing a human-ai partnership for mixed-initiative fact-checking. In *Proceedings of the 31st ACM User Interface Software and Technology Symposium (UIST)*, 2018.
- [9] An Thanh Nguyen, Aditya Kharosekar, Matthew Lease, and Byron C. Wallace. An interpretable joint graphical model for fact-checking from crowds. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*, pages 1511–1518, 2018.
- [10] An Thanh Nguyen, Junyi Jessy Li, Ani Nenkova, Byron C. Wallace, and Matthew Lease. Aggregating and Predicting Sequence Labels from Crowd Annotations. In *Proceedings of the 55th annual meeting of the Association for Computational Linguistics (ACL)*, pages 299–309, 2017.
- [11] Ye Zhang, Matthew Lease, and Byron C. Wallace. Exploiting Domain Knowledge via Grouped Weight Sharing with Application to Text Categorization. In *Proceedings of the 55th annual meeting of the Association for Computational Linguistics (ACL)*, pages 155–160, 2017.
- [12] Ye Zhang, Matthew Lease, and Byron Wallace. Active discriminative text representation learning. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI)*, pages 3386–3392, 2017.
- [13] Tyler McDonnell, Matthew Lease, Mucahid Kutlu, and Tamer Elsayed. Why Is That Relevant? Collecting Annotator Rationales for Relevance Judgments. In *Proceedings of the 4th AAAI Conference on Human Computation and Crowdsourcing (HCOMP)*, pages 139–148, 2016. **Best Paper Award** (*Selection made by vote of the Senior Program Committee among author-anonymized submissions*).
- [14] An Thanh Nguyen, Byron C. Wallace, and Matthew Lease. Probabilistic Modeling for Crowdsourcing Partially-Subjective Ratings. In *Proceedings of the 4th AAAI Conference on Human Computation and Crowdsourcing (HCOMP)*, pages 149–158, 2016.
- [15] An Thanh Nguyen, Byron C. Wallace, and Matthew Lease. A correlated worker model for grouped, imbalanced and multitask data. In *Proceedings of the 32nd Intl. Conference on Uncertainty in Artificial Intelligence (UAI)*, 2016. 10 pages.

- [16] James Cheng, Monisha Manoharan, Matthew Lease, and Yan Zhang. Is there a Doctor in the Crowd? Diagnosis Needed! (for less than \$5). In *Proceedings of the iConference*, 2015. 16 pages.
- [17] Hyun Joon Jung and Matthew Lease. A Discriminative Approach to Predicting Assessor Accuracy. In *Proceedings of the 37th European Conference on Information Retrieval (ECIR)*, pages 159–171, 2015. **Samsung Human-Tech Paper Award: Silver Prize in Computer Science.**
- [18] Hyun Joon Jung and Matthew Lease. Modeling Temporal Crowd Work Quality with Limited Supervision. In *Proceedings of the 3rd AAAI Conference on Human Computation and Crowdsourcing (HCOMP)*, pages 83–91, 2015.
- [19] An Thanh Nguyen, Byron C. Wallace, and Matthew Lease. Combining Crowd and Expert Labels using Decision Theoretic Active Learning. In *Proceedings of the 3rd AAAI Conference on Human Computation and Crowdsourcing (HCOMP)*, pages 120–129, 2015.
- [20] Donna Vakharia and Matthew Lease. Beyond Mechanical Turk: An Analysis of Paid Crowd Work Platforms. In *Proceedings of the iConference*, 2015. 17 pages.
- [21] Hyun Joon Jung, Yubin Park, and Matthew Lease. Predicting Next Label Quality: A Time-Series Model of Crowdsourcing. In *Proceedings of the 2nd AAAI Conference on Human Computation and Crowdsourcing (HCOMP)*, pages 87–95, 2014.
- [22] Yinglong Zhang, Jin Zhang, Matthew Lease, and Jacek Gwizdka. Multidimensional relevance modeling via psychometrics and crowdsourcing. In *Proceedings of the 37th international ACM SIGIR conference on Research and Development in Information Retrieval*, pages 435–444, 2014.
- [23] Aniket Kittur, Jeff Nickerson, Michael S. Bernstein, Elizabeth Gerber, Aaron Shaw, John Zimmerman, Matthew Lease, and John J. Horton. The Future of Crowd Work. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW)*, pages 1301–1318, 2013.
- [24] Ripon K. Saha, Matthew Lease, Sarfraz Khurshid, and Dewayne E. Perry. Improving Bug Localization using Structured Information Retrieval. In *Proceedings of the 28th IEEE/ACM International Conference on Automated Software Engineering (ASE)*, pages 345–355, 2013.
- [25] Aashish Sheshadri and Matthew Lease. SQUARE: A Benchmark for Research on Computing Crowd Consensus. In *Proceedings of the 1st AAAI Conference on Human Computation (HCOMP)*, pages 156–164, 2013.
- [26] Di Liu, Randolph Bias, Matthew Lease, and Rebecca Kuipers. Crowdsourcing for usability testing. In *Proceedings of the 75th Annual Meeting of the American Society for Information Science and Technology (ASIS&T)*, 2012. 10 pages.
- [27] Hohyon Ryu, Matthew Lease, and Nicholas Woodward. Finding and exploring memes in social media. In *Proceedings of the 23rd ACM Conference on Hypertext and Social Media*, pages 295–304, 2012.
- [28] Aibo Tian and Matthew Lease. Active learning to maximize accuracy vs. effort in interactive information retrieval. In *SIGIR '11: Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 145–154. ACM, 2011.
- [29] Stephen M. Wolfson and Matthew Lease. Look before you leap: Legal pitfalls of crowdsourcing. *Proceedings of the American Society for Information Science and Technology (ASIS&T)*, 2011. 10 pages.

- [30] Yongyi Zhou, Ramona Broussard, and Matthew Lease. Mobile options for online public access catalogs. In *Proceedings of the iConference*, pages 598–605, 2011.
- [31] Adriana Kovashka and Matthew Lease. Human and machine detection of stylistic similarity in art. In *Proceedings of the 1st Annual Conference on the Future of Distributed Work (CrowdConf)*, San Francisco, CA, 2010. 9 pages.
- [32] 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, 2009.
- [33] 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.
- [34] 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)*, pages 629–634, 2006. Selected for oral presentation.
- [35] 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 *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC)*, pages 333–338, 2006. Selected for oral presentation.
- [36] 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.
- [37] 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.
- [38] 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.
- [39] 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.
- [40] 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 (LNCS)*, pages 316–332. Springer-Verlag, 2002.
- [41] 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 (LNCS)*, pages 152–166. Springer-Verlag, 2002.

- [42] Ira J. Kalet, Jonn Wu, Matthew Lease, Mary M. Austin-Seymour, James F. Brinkley, and Cornelius Rosse. Anatomical information in radiation treatment planning. In *Proceedings of the American Medical Informatics Association (AMIA) Fall Symposium*, pages 291–295, 1999.

### Refereed Conference Short Papers

- [43] Brandon Dang, Martin J. Riedl, and Matthew Lease. Toward Safer Crowdsourced Content Moderation. In *6th ACM Collective Intelligence Conference*, 2018. 5 pages.
- [44] Brandon Dang, Martin J. Riedl, and Matthew Lease. But Who Protects the Moderators? The Case of Crowdsourced Image Moderation. In *6th AAAI Conference on Human Computation and Crowdsourcing (HCOMP): Works-in-Progress Track*, 2018. 5 pages.
- [45] Matthew Lease. Fact Checking and Information Retrieval. In *Proceedings of the 1st Biannual Conference on the Design of Experimental Search & Information REtrieval Systems (DESIREs)*, pages 97–98, 2018.
- [46] Xi Zheng, Akanksha Bansal, and Matthew Lease. Bullseye: Structured passage retrieval and document highlighting for scholarly search. In *The Thirteenth Asia-Pacific Conference on Conceptual Modelling (APCCM)*, 2017. 4 pages.
- [47] Brandon Dang, Miles Hutson, and Matthew Lease. MmmTurkey: A Crowdsourcing Framework for Deploying Tasks and Recording Worker Behavior on Amazon Mechanical Turk. In *4th AAAI Conference on Human Computation and Crowdsourcing (HCOMP): Demonstration Track*, 2016. <http://arxiv.org/abs/1609.00945>. 3 pages.
- [48] Yalin Sun, Pengxiang Cheng, Shengwei Wang, Hao Lyu, Matthew Lease, Iain Marshall, and Byron C. Wallace. Crowdsourcing Information Extraction for Biomedical Systematic Reviews. In *4th AAAI Conference on Human Computation and Crowdsourcing (HCOMP): Works-in-Progress Track*, 2016. <http://arxiv.org/abs/1609.01017>. 3 pages.
- [49] Reem Suwaileh, Mucahid Kutlu, Nihal Fathima, Tamer Elsayed, and Matthew Lease. ArabicWeb16: A New Crawl for Today’s Arabic Web. In *Proceedings of the 39th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 673–676, 2016.
- [50] Hyun Joon Jung and Matthew Lease. Inferring Missing Relevance Judgments from Crowd Workers via Probabilistic Matrix Factorization. In *Proceedings of the 35th international ACM SIGIR conference on Research and Development in Information Retrieval*, pages 1095–1096, 2012.
- [51] Shilpa Shukla, Matthew Lease, and Ambuj Tewari. Parallelizing ListNet Training using Spark. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 1127–1128. ACM, 2012.
- [52] Jorn Klinger and Matthew Lease. Enabling trust in crowd labor relations through identity sharing. In *Proceedings of the 74th Annual Meeting of the American Society for Information Science and Technology (ASIS&T)*, 2011. 4 pages.

- [53] Abhimanu Kumar, Matthew Lease, and Jason Baldrige. Supervised language modeling for temporal resolution of texts. In *CIKM '11: Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 2069–2072. ACM, 2011.
- [54] Abhimanu Kumar and Matthew Lease. Learning to rank from a noisy crowd. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 1221–1222, 2011. Additional refereed acceptance for encore presentation at the 3rd Human Computation Workshop (HCOMP) at the Twenty-Fifth AAAI Conference on Artificial Intelligence (AAAI). Appears in SIGIR proceedings only.
- [55] Hohyon Ryu and Matthew Lease. Crowdfworker filtering with support vector machine. In *Proceedings of the 74th Annual Meeting of the American Society for Information Science and Technology (ASIS&T)*, 2011. 4 pages.
- [56] Saeedeh Momtazi, Matthew Lease, and Dietrich Klakow. Effective term weighting for sentence retrieval. In *ECDL'10: Proceedings of the 14th European conference on Research and advanced technology for digital libraries*, pages 482–485, Berlin, Heidelberg, 2010. Springer-Verlag.
- [57] Ramona Broussard, Yongyi Zhou, and Matthew Lease. Mobile phone search for library catalogs. In *ASIS&T '10: Proceedings of the 73rd ASIS&T Annual Meeting on Navigating Streams in an Information Ecosystem*. American Society for Information Science, 2010. 4 pages.
- [58] Ramona Broussard, Yongyi Zhou, and Matthew Lease. University of Texas Mobile Library Search. In *ASIS&T '10: Proceedings of the 73rd ASIS&T Annual Meeting on Navigating Streams in an Information Ecosystem*. American Society for Information Science, 2010. Demo paper, 2 pages.
- [59] Matthew Lease and Mark Johnson. Early deletion of fillers in processing conversational speech. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT-EMNLP), Companion Volume: Short Papers*, pages 73–76, 2006.
- [60] Matthew Lease, Eugene Charniak, Mark Johnson, and David McClosky. A Look at Parsing and Its Applications. In *Proceedings of the 21st International Conference on Artificial Intelligence (AAAI): New Scientific and Technical Advances in Research (Nectar)*, pages 1642–1645, 2006.
- [61] Matthew Lease, Eugene Charniak, and Mark Johnson. Parsing and its applications for conversational speech. In *Proceedings of the IEEE 30th Intl. Conf. on Acoustics, Speech, & Signal Processing (ICASSP)*, volume 5, pages 961–964, 2005.
- [62] Matthew Lease. Plan-Aware Behavioral Modeling. In *Adjunct Proceedings of the 4th Intl. Conference on Ubiquitous Computing (UBICOMP)*, pages 35–36, 2002.

#### Refereed Workshop Papers

- [63] Matthew Lease, Gordon V. Cormack, An Thanh Nguyen, Thomas A. Trikalinos, and Byron C. Wallace. Systematic Review is e-Discovery in Doctor’s Clothing. In *Proc. of ACM SIGIR Workshop on Medical Information Retrieval (MedIR)*, 2016. 2 pages.
- [64] Hyun Joon Jung and Matthew Lease. Forecasting Crowd Work Quality via Multi-dimensional Features of Workers. In *ICML Workshop on Crowdsourcing and Machine Learning (CrowdML)*, 2015. 10 pages.

- [65] Hohyon Ryu and Matthew Lease. Generating automatic keywords for conversational speech asr transcripts. In *Proceedings of the 1st ACM SIGIR Workshop on the Exploration, Navigation and Retrieval of Information in Cultural Heritage (ENRICH)*, 2013. 4 pages.
- [66] Haofeng Zhou, Dennis Baskov, and Matthew Lease. Crowdsourcing transcription beyond mechanical turk. In *AAAI HCOMP Workshop on Scaling Speech, Language Understanding and Dialogue through Crowdsourcing*, 2013. 8 pages.
- [67] Hyun Joon Jung and Matthew Lease. Improving Quality of Crowdsourced Labels via Probabilistic Matrix Factorization. In *Proceedings of the 4th Human Computation Workshop (HCOMP) at the Twenty-Sixth AAAI Conference on Artificial Intelligence*, pages 101–106, 2012.
- [68] Lu Guo and Matthew Lease. Personalizing local search with twitter. In *Proceedings of the Workshop on Enriching Information Retrieval (ENIR) at the 34th Annual ACM SIGIR Conference*, 2011. 2 pages.
- [69] Hyun Joon Jung and Matthew Lease. Improving Consensus Accuracy via Z-score and Weighted Voting. In *Proc. of the 3rd Human Computation (HCOMP) Workshop at the AAAI Conference on Artificial Intelligence*, pages 88–90, 2011.
- [70] Abhimanu Kumar and Matthew Lease. Modeling annotator accuracies for supervised learning. In *Proceedings of the Workshop on Crowdsourcing for Search and Data Mining (CSDM) at the Fourth ACM International Conference on Web Search and Data Mining (WSDM)*, pages 19–22, 2011.
- [71] Matthew Lease. On Quality Control and Machine Learning in Crowdsourcing. In *Proceedings of the 3rd Human Computation (HCOMP) Workshop at the Twenty-Fifth AAAI Conference on Artificial Intelligence (AAAI)*, pages 97–102, 2011. Additional refereed acceptance for encore presentation at the [2012 AAAI Spring Symposium: Wisdom of the Crowd](#). Appears in HCOMP proceedings only.
- [72] Wei Tang and Matthew Lease. Semi-supervised consensus labeling for crowdsourcing. In *Proceedings of the ACM SIGIR Workshop on Crowdsourcing for Information Retrieval (CIR)*, pages 36–41, 2011.
- [73] 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. Association for Computational Linguistics, 2010.
- [74] 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**.
- [75] Matthew Lease and Guy Eddon. SmartElevator: Revitalizing a legacy device through inexpensive augmentation. In *Proceedings of the 3rd Intl. Workshop on Smart Appliances and Wearable Computing, held at the IEEE 23rd International Conference on Distributed Computing Systems (ICDCS)*, pages 254–259, 2003.

### Technical Reports & Online Working Papers

- [76] Mucahid Kutlu, Vivek Khetan, and Matthew Lease. Correlation and Prediction of Evaluation Metrics in Information Retrieval. Technical report, University of Texas at Austin and Qatar University, February 2018. arXiv:1802.00323.



- [77] Md Mustafizur Rahman, Mucahid Kutlu, Tamer Elsayed, and Matthew Lease. Efficient Test Collection Construction via Active Learning. Technical report, University of Texas at Austin and Qatar University, January 2018. arXiv:1801.05605.
- [78] Ye Zhang, Md Mustafizur Rahman, Alex Braylan, Brandon Dang, Heng-Lu Chang, Henna Kim, Quinten McNamara, Aaron Angert, Edward Banner, Vivek Khetan, Tyler McDonnell, An Thanh Nguyen, Dan Xu, Byron C. Wallace, and Matthew Lease. Neural Information Retrieval: A Literature Review. Technical report, U. of Texas at Austin, Nov. 2016. <https://arxiv.org/abs/1611.06792>.
- [79] Ethan Petuchowski and Matthew Lease. TurKPF: TurKontrol as a Particle Filter. Technical report, University of Texas at Austin, April 2014. arXiv:1404.5078, 8 pages.
- [80] Hyun Joon Jung and Matthew Lease. Crowdsourced task routing via matrix factorization. Technical report, University of Texas at Austin, October 2013. arXiv:1310.5142, 10 pages.
- [81] Matthew Lease, Jessica Hullman, Jeffrey P. Bigham, Michael S. Bernstein, Juho Kim, Walter S. Lasecki, Saeideh Bakhshi, Tanushree Mitra, and Robert C. Miller. Mechanical Turk is Not Anonymous. In *Social Science Research Network (SSRN) Online*, March 6, 2013. SSRN ID: 2228728, 15 pages.
- [82] Hyun Joon Jung and Matthew Lease. Evaluating classifiers without expert labels. Technical report, University of Texas at Austin, December 2012. arXiv:1212.0960, 21 pages.
- [83] Elben Shira and Matthew Lease. Expert search on code repositories. Technical Report TR-11-42, Department of Computer Science, University of Texas at Austin, December 2011. 8 pages.
- [84] Abhimanu Kumar, Jason Baldrige, Matthew Lease, and Joydeep Ghosh. Dating Texts without Explicit Temporal Cues. Technical report, University of Texas at Austin, November 2012. arXiv:1211.2290, 12 pages.
- [85] Di Liu, Matthew Lease, Rebecca Kuipers, and Randolph Bias. Crowdsourcing for usability testing. Technical report, School of Information, University of Texas at Austin, March 2012. arXiv:1203.1468, 10 pages.
- [86] Eunho Yang, Pradeep Ravikumar, and Matthew Lease. A new class of ranking functions for DCG-like evaluation metrics using conditional probability models. Technical Report AI14-02 (AI report), Department of Computer Science, University of Texas at Austin, October 29 2010. 8 pages.
- [87] Mary Harper, Bonnie Dorr, John Hale, Brian Roark, Izhak Shafran, Matthew Lease, Yang Liu, Matthew Snover, Lisa Yunge, Anna Krasnyanskayai, and Robin Stewart. 2005 Johns Hopkins Summer Workshop Final Report on Parsing and Spoken Structural Event Detection. Technical report, 2005. 116 pages.
- [88] Matthew Lease. Parsing and Disfluency Modeling. Technical Report CS-05-15, Brown University Department of Computer Science, 2005. 19 pages.

### Invited Papers

- [89] Jeffrey V. Nickerson and Matthew Lease. Work in the Age of Intelligent Machines: Workshop Report from the Sixth AAAI Conference on Human Computation and Crowdsourcing (HCOMP). *AI Magazine*, 39(3), 2018.

- [90] Tyler McDonnell, Mucahid Kutlu, Tamer Elsayed, and Matthew Lease. Beyond Dual-Supervision: the Many Benefits of Annotator Rationales for Relevance Judgments. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI): Sister Conference Best Paper Track*, pages 4909–4913, 2017.
- [91] Matthew Lease. Crowdsourcing for Success: Motivations, Design, & Ethics. In *Workshop on Novel Incentives and Engineering Unique Workflows (NIEUW), organized by the Linguistic Data Consortium (LDC)*, 2016. 2 pages.
- [92] Matthew Lease and Omar Alonso. Crowdsourcing and human computation, introduction. *Encyclopedia of Social Network Analysis and Mining (ESNAM)*, pages 304–315, September 2014.
- [93] Tatiana Josephy, Matthew Lease, Praveen Paritosh, Markus Krause, Mihai Georgescu, Michael Tjalve, and Daniela Braga. Workshops held at the first aaai conference on human computation and crowdsourcing: A report. *AI Magazine*, 35(2):75–78, 2014.
- [94] Matthew Lease and Emine Yilmaz. Crowdsourcing for Information Retrieval: Introduction to the Special Issue. *Information Retrieval*, 16(2):91–100, April 2013.
- [95] Ira J. Kalet, Robert 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: Lisp in the Mainstream*, November 16-18, 1998. Berkeley, California.

#### Shared-Task Evaluations: Organization & Participation

- [96] Mark D. Smucker, Gabriella Kazai, and Matthew Lease. Overview of the TREC 2013 Crowdsourcing Track. In *Proceedings of the 22nd National Institute of Standards & Technology (NIST) Text REtrieval Conference (TREC)*, 2014. 6 pages.
- [97] Mark D. Smucker, Gabriella Kazai, and Matthew Lease. Overview of the TREC 2012 Crowdsourcing Track. In *Proceedings of the 21st National Institute of Standards & Technology (NIST) Text REtrieval Conference (TREC)*, 2013. 12 pages.
- [98] Hyun Joon Jung and Matthew Lease. UT Austin in the TREC 2012 Crowdsourcing Track’s Image Relevance Assessment Task. In *Proceedings of the 21st National Institute of Standards & Technology (NIST) Text REtrieval Conference (TREC)*, 2013. 12 pages.
- [99] Aashish Sheshadri and Matthew Lease. SQUARE: Benchmarking Crowd Consensus at MediaEval. In *Proceedings of the MediaEval 2013 Multimedia Benchmark Workshop*, 2013. CEUR Workshop (cuer-ws.org) Proceedings Vol-1043, ISSN 1613-0073, 2 pages.
- [100] Matthew Lease and Gabriella Kazai. Overview of the TREC 2011 Crowdsourcing Track (Conference Notebook). In *The 20th National Institute of Standards & Technology (NIST) Text REtrieval Conference (TREC) Notebook*, 2012. 10 pages.
- [101] Chris Buckley, Matthew Lease, and Mark D. Smucker. Overview of the TREC 2010 Relevance Feedback Track (Notebook). In *The 19th National Institute of Standards & Technology (NIST) Text REtrieval Conference (TREC) Notebook*, 2010. 4 pages.
- [102] Marc Cartright, Jangwon Seo, and Matthew Lease. UMass Amherst and UT Austin at the TREC’09 Relevance Feedback Track. In *Proceedings of the 18th National Institute of Standards & Technology (NIST) Text REtrieval Conference (TREC)*, 2010. 10 pages.

- [103] 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.** 6 pages.
- [104] 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. Springer-Verlag, 2008.
- [105] Matthew Lease and Eugene Charniak. Brown at CL-SR07: Retrieving Conversational Speech in English and Czech. In *Working Notes of the Cross-Language Evaluation Forum (CLEF): Cross-Language Speech Retrieval (CL-SR) Track*, 2007. 6 pages.
- [106] 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.** 6 pages.

#### ACM SIGIR Forum

- [107] James Allan, Bruce Croft, Alistair Moffat, Mark Sanderson, et al. Frontiers, Challenges, and Opportunities for Information Retrieval – Report from SWIRL 2012, The Second Strategic Workshop on Information Retrieval in Lorne. *SIGIR Forum*, 46(1):2–32, June 2012.
- [108] Vitor R. Carvalho, Matthew Lease, and Emine Yilmaz. Crowdsourcing for search evaluation. *SIGIR Forum*, 44(2):17–22, December 2010.
- [109] Matthew Lease and Emine Yilmaz. Crowdsourcing for information retrieval. *SIGIR Forum*, 45(2):66–75, December 2011.
- [110] Matthew Lease, Vitor R. Carvalho, and Emine Yilmaz. Crowdsourcing for search and data mining. *SIGIR Forum*, 45(1):18–24, June 2011.

#### Edited Volumes

- [111] Matthew Lease and Arpita Ghosh, editors. *Proceedings of the 4th AAAI Conference on Human Computation and Crowdsourcing (HCOMP)*. 2016. Proceedings available [online](#).
- [112] Tatiana Josephy, Matthew Lease, and Praveen Paritosh, editors. *Proceedings of the AAAI HCOMP Workshop on Crowdsourcing at Scale (CrowdScale)*. 2013. Proceedings available [online](#).
- [113] Matthew Lease and Emine Yilmaz, editors. *Information Retrieval: Special Issue on Crowdsourcing*. Springer, April 2013. Volume 16 number 2.
- [114] Matthew Lease, Vaughn Hester, Alexander Sorokin, and Emine Yilmaz, editors. *Proceedings of the 2nd Workshop on Crowdsourcing for Information Retrieval at the 34th ACM International Conference on Information Retrieval (SIGIR 2011)*. Beijing, China, July 2011. Proceedings available [online](#).
- [115] Vitor Carvalho, Matthew Lease, and Emine Yilmaz, editors. *Proceedings of the Workshop on Crowdsourcing for Search and Data Mining (CSDM) at the Fourth ACM International Conference on Web Search and Data Mining (WSDM)*. Hong Kong, China, February 2011. Proceedings available [online](#).

- [116] Vitor Carvalho, Matthew Lease, and Emine Yilmaz, editors. *Proceedings of the ACM SIGIR 2010 Workshop on Crowdsourcing for Search Evaluation (CSE 2010)*. Geneva, Switzerland, July 2010. Proceedings available [online](#).

## Tutorials Presented

- [117] Matthew Lease. Crowdsourcing and Human Computation. In *SIAM International Conference on Data Mining*, May 2013.

PROPOSALS  
REFEREED UNLESS  
NOTED OTHERWISE

- [118] Matthew Lease and Omar Alonso. Crowdsourcing for search evaluation and social-algorithmic search. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, August 2012.

- [119] Omar Alonso and Matthew Lease. Crowdsourcing for research and engineering. In *CrowdConf 2011*, San Francisco, CA, November 2011. **Invited Tutorial**.

- [120] Omar Alonso and Matthew Lease. Crowdsourcing for information retrieval: Principles, methods, and applications. In *The 34th Annual ACM SIGIR Conference*, page 1299, Beijing, China, July 2011. ACM.

- [121] Omar Alonso and Matthew Lease. Crowdsourcing 101: Putting the wisdom of crowds to work for you. In *The Fourth ACM International Conference on Web Search and Data Mining (WSDM)*, pages 1–2, Hong Kong, China, February 2011. ACM.

## Invited Talks

### Notable Invited Talks

SLIDES ONLINE AT  
[SLIDESHARE.NET](#)

- [8th Meeting of the Forum for Information Retrieval Evaluation \(FIRE\)](#). Invited Talk: *Deep Learning for Information Retrieval: Models, Progress, and Opportunities*. Kolkata, India. December 10, 2016.
  - 2014 Microsoft Faculty Summit. Invited Talk: *Toward Effective & Sustainable Online Crowd Work*. Redmond, WA. July 15, 2014. Panel: [Meet the Crowd: Political Economies and Cultural Meanings of Digital Labor](#). Video [online](#).
  - [2012 Meeting on the Frontiers of Information Science & Technology](#). Invited Talk: *The Rise of Crowd Computing*. Shanghai, China. December 11, 2012.
  - [2012 Intl. Conf. on Social Computing, Behavioral-Cultural Modeling & Prediction](#). Invited Talk: *Finding and Exploring Memes in Social media*. College Park, MD. April 5, 2012.
- Intl. Joint Conf. on Natural Language Processing (IJCNLP). Plenary Keynote: [Crowd Computing: Opportunities and Challenges](#). Chiang Mai, Thailand. November 8, 2011.

### Additional Invited Talks

*What Can Machine Learning & Crowdsourcing Do for You? Exploring New Tools for Scalable Data Processing*. ACM JCDL 2018 [Workshop on Cyberinfrastructure and Machine Learning for Digital Libraries](#). Fort Worth, TX. June 3, 2018.

[Deep Learning for Information Retrieval: Models, Progress, and Opportunities](#).

- Qatar University, Department of Computer Science. (December 12, 2016)
- Qatar Computing Research Institute (QCRI). (December 15, 2016)

[The Rise of Crowd Computing](#)

- Qatar University, Department of Computer Science. (December 16, 2015)

- Qatar Computing Research Institute (QCRI). (December 14, 2015)

*Toward Better Crowdsourcing Science.* [ICML 2015 Workshop on Crowdsourcing and Machine Learning \(CrowdML\)](#). (July 10, 2015)

*The Search for Truth in Objective & Subject Crowdsourcing.* Crowdsourcing Lunch Seminar Series, School of Comp. Science, Carnegie Mellon University. (March 3, 2015)

*Finding and Exploring Memes in Social media.* [Workshop on Computational Journalism](#). Porto, Portugal (Presented via Skype). (April 22, 2014)

*Statistical Crowdsourcing: From Aggregating Judgments to Search Engine Evaluation*

- Information School, University of Washington (June 2, 2014).
- Computer Science & Engineering Dept., Texas A&M University (March 24, 2014)
- School of Communication & Information, Rutgers University (February 24, 2014)

*Crowdsourcing for Information Retrieval: From Statistics to Ethics*

- School of Information, UC Berkeley (October 21, 2013)
- School of Information, Syracuse University (October 28, 2013)

[IEEE Metrocon 2012](#). Title: *The Rise of Crowd Computing*. Forth Worth, TX. (October 11, 2012)

*Adventures in Crowdsourcing: Research at UT Austin & Beyond*

- Microsoft Inc., Mountain View, CA (August 23, 2012)
- CrowdFlower Inc., San Francisco, CA (August 22, 2012)
- LinkedIn Inc., Mountain View, CA (August 20, 2012)

*Shrinkwrapped Crowds*, [Crowdsourcing Work](#) Event, San Francisco, CA (July 24, 2012)

*Crowd Computing: Opportunities and Challenges.* Department of Computer Science, Brown University. (November 11, 2011)

## Panels

[Workshop on Explainable Recommendation and Search \(EARS\) at ACM SIGIR 2018](#). Title: *Benefits, Challenges, Opportunities, and Mechanisms of Explainable IR*. Ann Arbor, MI. July 12, 2018.

[Workshop on Learning from Limited/Noisy data for IR at ACM SIGIR 2018](#). Title: *The Future of Learning from Limited / Noisy Data and its Influence on IR Applications*. Ann Arbor, MI. July 12, 2018.

6th AAAI Conference on Human Computation and Crowdsourcing (HCOMP). Title: *Ethics in Human Computation and Crowdsourcing*. Zurich, Switzerland. July 6, 2018.

The 2015 Information Governance Conference (InfoGovCon). Title: [Information Governance Technologies and Ethics A Delicate Balance](#). Hartford, CT. October 1, 2015

## Local Invited Talks

*The Rise of Crowd Computing*

- [Austin Data Meetup - SXSW Edition](#). (March 11, 2016)
- [Student Undergraduate Research Group](#) (SURGe), College of Natural Sciences, UT Ausin. (November 1, 2013)

*Crowdsourcing & Human Computation: Labeling Data & Building Hybrid Systems*. [First Bytes](#) Workshop for High School Computer Science teachers, Department of Computer Science, University of Texas at Austin.

- June 30, 2016
- June 24, 2015

*Information Retrieval and Crowdsourcing*, [Wireless Networking and Communications Group](#). (WNCG) Colloquium, UT Austin (September 18, 2013)

[ID360: The Global Forum on Identity](#). Title: *Amazon's Mechanical Turk is Not Anonymous*. (May 1, 2013)

*Information Retrieval for Big Data*, [Computing Technologies Workshop for Data Enabled Science](#), Texas Advanced Computing Center, UT Austin (May 24, 2012)

*Finding and Exploring Memes in Social media*, ACM SIGKDD Chapter Meeting, Austin, TX (March 21, 2012)

*Crowdsourcing for Natural Language Processing*, Linguistics Department Colloquium, UT Austin (February 28, 2011)

*Before Fall 2009* *Beyond keywords: finding information more accurately and easily using natural language*

- Max Planck Institut fur Informatik, Saarbrucken, Germany (08/18/09)
- German Research Center for Artificial Intelligence (DFKI), Saarbrucken, Germany (06/24/09)
- [International Research Training Group \(IRTG\) Summer School in Computational Linguistics and Psycholinguistics](#), 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

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

**Professional Societies**

Association for the Advancement of Artificial Intelligence (AAAI)  
Association for Computing Machinery (ACM)  
ACM Special Interest Group on Information Retrieval (SIGIR)

**Professional Service**

**Conference Co-Chair:** Association for the Advancement of Artificial Intelligence (AAAI) Conference on Crowdsourcing and Human Computation (HCOMP), 2016

**Senior Associate Editor:** Journal of the Association for Information Science and Technology (JASIS&T), 2016-2017

**Area Chair / Senior Program Committee**

AAAI HCOMP 2014-2015 (single-tier structure with PC meeting), 2018  
ACM SIGIR Conference, 2013-2014, 2016-2018  
ACM CIKM Conference, 2014-2015  
Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), 2012

**Track Co-Chair, [Research Track](#)**, CrowdConf 2013 (with Paul Bennett)

**NIST TREC Track Organization** (Refereed Proposals)

*National Institute of Standards & Technology (NIST) Text REtrieval Conference (TREC)*  
2011-2013 [Crowdsourcing Track](#) (with Gabriella Kazai and Mark Smucker) [100, 97, 96]  
2010 [Relevance Feedback Track](#) (with Chris Buckley and Mark Smucker) [101]

**Workshop Co-Organizer** (Refereed Proposals)

AAAI HCOMP 2018, [Work in the Age of Intelligent Machines](#). July 5, 2018.  
AAAI HCOMP 2013, [Crowdsourcing at Scale](#) (CrowdScale)  
ACM SIGIR 2011 & 2015, [Crowdsourcing for Information Retrieval](#) [109, 114]  
ACM WSDM 2011, [Crowdsourcing for Search and Data Mining](#) [110, 115]  
ACM SIGIR 2010, [Crowdsourcing for Search Evaluation](#) [108, 116]  
iConference 2013 Alternative Event: *Calling All Computer Scientists and Social Scientists: Establishing a Research Agenda for Computational Social Science*

**Journal Guest Co-Editor** (Refereed Proposal)

2012 *Springer Information Retrieval: Special Issue on Crowdsourcing* [113]

**Doctoral Consortia**

Co-Chair, [AAAI HCOMP 2014](#)  
Program Committee & Mentor, AAAI HCOMP 2015  
Program Committee & Mentor, [ACM SIGIR 2014](#)

## Miscellaneous

*Industry Panel Chair*, AAAI HCOMP 2015  
*Proceedings Chair*, ACM CIKM, 2011  
*Conference Advisory Board*, [CrowdConf 2012](#)

## Conference Selection Committees

*Awards Committee*: European Conference on Information Retrieval (ECIR), 2011  
*Tutorial Committee*: ACM CIKM 2015, ACM SIGIR 2014, ECIR 2013  
*Workshop Committee*, ACM Web Search and Data Mining (WSDM), 2012

**Conference Program Committees (PCs)**: ACL-HLT 2011, ACM SIGIR 2010-2012, [CrowdConf 2010](#), ECIR 2011, EMNLP 2010-2011, IJCNLP 2011, WWW 2011

## Conference Reviewing

[ACM Collective Intelligence Conference](#), 2017-2018  
ACM Computer Supported Cooperative Work (CSCW), 2014 & 2016  
ACM Multimedia, 2013  
ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS), 2013  
ACM SIGCHI Conference on Human Factors in Computing Systems (CHI), 2013  
ASIS&T 2010, ASRU 2007 & 2009, COLING 2008, EMNLP 2009, ICASSP 2008-2009  
NAACL/HLT 2010, NIPS 2010

## Journal Reviewing

ACM Transactions on Information Systems (ToIS), 2017  
Computers in Human Behavior (Elsevier), 2015  
Communications of the ACM (CACM), 2012  
Artificial Intelligence (Elsevier), 2012  
MIT Sloan Management Review, 2012  
ACM Computing Surveys, 2009  
Information Processing & Management (Elsevier), 2009

## Workshop Program Committees (PCs)

Web Science 2016 Workshop on [Weaving Relations of Trust in Crowd Work: Transparency and Reputation across Platforms](#)  
[CrowdBench Workshop](#) on Benchmarks for Ubiquitous Crowdsourcing: Metrics, Methodologies, and Datasets (at IEEE PERCOM)  
*CrowdWork 2013: Workshop on Crowd Work and Human Computation*,  
3rd international conference on Social Computing and its Applications (SCA)  
*BPOE 2013: Workshop on Benchmarks, Performance Optimization, and Emerging hardware of Big Data Systems*, 2013 IEEE Conference on Big Data  
*Crowdsourcing and Data Mining*, ACM Conference on Knowledge Discovery and Data Mining (KDD), 2012  
*Crowdsourcing for Multimedia*, ACM Multimedia Conference, 2012  
*Crowdsearch*, World Wide Web (WWW) Conference, 2012  
*Query Representation and Understanding*, ACM SIGIR Conference, 2010

## Grant Reviewing - External

U.S. Army Research Office (ARO), Reviewer, 2016 & 2017  
U.S. National Science Foundation (NSF), Panelist, 2013, 2016; Reviewer 2015  
U.S. Institute of Museum and Library Science (IMLS), Panelist, 2013  
Swiss National Science Foundation (SNSF), Reviewer, 2012

## Grant Reviewing - Local

[UT Austin Longhorn Innovation Fund for Technology \(LIFT\)](#), Co-Organizer, 2013-2016



**External Tenure Case Referee**, Anonymized University, 2016

## Grants

### Active Projects

*CAREER: Achieving Quality Crowdsourcing Across Tasks, Data Scales, and Operational Settings*, National Science Foundation (NSF), March 2013-February 2018, PI, \$550K. 25.9% acceptance rate (101/390). *Research Experiences for Undergraduates (REU) supplements*, 2014 & 2017, \$32K.

*Neural Models for Text: Improving Efficiency, Interpretability and Accuracy*. Army Research Office (ARO), September 2018-May 2019, Subcontractor, \$53,796.

*Stampede 2: Operations and Maintenance for the Next Generation of Petascale Computing* (award #1663578), National Science Foundation (NSF), October 2017-September 2021, Project Personnel, \$133,300.

### Past Projects

*Efficient and Scalable Evaluation for Searching Massive Arabic Social Media and Web Collections*. Qatar National Research Fund (QNRF), January 2015-October 2018, co-PI, \$305K (of \$884K total). 20% acceptance rate.

*REMINDS: Relevance Mining and Detection System*. Science & Technology Foundation of Portugal (FCT). September 2016-August 2018, co-PI. 182,443€ to Portuguese PI, \$100K to UT Austin.

*Crowdsourcing using Intelligent Supervision to address Information Requirements in Crisis Situations (CRISIS)*, Office of Naval Research (ONR), September 2015-December 2017, Subcontractor, \$150K.

*Enabling, Enhancing, And Extending Petascale Computing For Science And Engineering* (award #1134872), National Science Foundation (NSF), September 2013-December 2017, Project Personnel, \$85K.

*Enabling a New Scale of Conversational Speech Archives via Crowdsourcing*. Institute of Museum and Library Services (IMLS), Early Career Award, June 2013-May 2016, PI, \$290K. **Sole IMLS Early Career Award made in 2013.**

*Blending Crowdsourcing with Automation for Fast, Cheap, & Accurate Analysis of Spontaneous Speech*. Young Faculty Award, Defense Advanced Research Projects Agency (DARPA), September 2012-August 2015. PI, \$300K. 9.1% acceptance rate (51/560).

*Spatial and Temporal Analysis of Multilingual Texts*, New York Community Trust, September 2012-August 2013, co-PI, \$73K.

*REACTION: Retrieval, Extraction and Aggregation Computing Technology for Integrating and Organizing News*. Science & Technology Foundation of Portugal. September 2010-August 2013, co-PI. 220,000€ to PIs and \$148.5K to iSchool (\$80K to my lab).

*Workshop: Doctoral Consortium for HCOMP 2014*. National Science Foundation (NSF), June 2014-May 2015, PI, \$25K.

## Corporate Gifts (Selected)

### Yahoo!

*Crowdsourcing as a methodology to investigate user engagement*. Faculty Research & Engagement Program, January 2013-May 2013, PI, \$20K.

2010-2012 [Yahoo! Speaker Series](#). \$10K in gifts, August 2011 & 2011.

#### **Amazon**

2011-2013 [NIST TREC Crowdsourcing Track](#) Sponsors. January 2012. \$5,000.

#### **University Competitive Awards**

**UT Austin** (university-wide selection)

[Longhorn Innovation Fund for Technology \(LIFT\)](#), \$94,000, September 2010-August 2010, PI, 16% acceptance rate for funding.

Graduate School [Mentoring Fellowship](#), (2013-2014, 2015-2016)

[Summer Research Assignment](#), (June 2011-July 2011)

#### **Teaching**

**Courses Taught** (see [Course Websites](#))

INF384H/CS395T: *Concepts of Information Retrieval*

Spring 2010, Fall 2010-2014, 2016 & 2018

INF385T/CS395T: *Human Computation & Crowdsourcing*

Spring 2011, 2012, 2014; Fall 2015 & 2017

INF385T/CS395T/LIN386M: *Data-Intensive Computing for Text Analysis* (Fall 2011)

INF383D: *Mathematical Foundations of Information Studies* (Spring 2010 & 2012)

INF384C: *Organizing Information* (Fall 2010, Spring 2011)

INF380E: *Perspectives on Information* (Fall 2013; Spring 2013, 2015-2018)

**Average Instructor Rating** (over 20 courses taught through Fall 2016): 4.2/5

#### **Student Mentoring**

##### **Post-doctoral Supervision**

[Mucahid Kutlu](#), Computer Science, Qatar University (Spring 2016-August 2018) [[49](#), [13](#), [90](#), [5](#), [6](#), [7](#), [76](#), [1](#), [4](#), [77](#)]

Co-supervised with Tamer Elsayed, Qatar University

##### **Ph.D. Dissertations Supervised**

Hyunjoon Jung, School of Information (*December 2015*) [[18](#), [64](#), [17](#), [21](#), [80](#), [82](#), [67](#), [50](#), [69](#)]

Title: [Temporal Modeling Crowd Work for Quality Assurance in Crowdsourcing](#)

Committee: Paul Bennett (Microsoft Research), Ken Fleischmann (School of Information), Ray Mooney (Computer Science), Byron Wallace (School of Information)

##### **Masters Theses Supervised**

Aashish Sheshadri, Computer Science (*May 2014*) [[25](#), [99](#)]

Title: [A Collaborative Approach to IR Evaluation](#)

Co-supervisor: Kristen Grauman, Computer Science

Stephen Wolfson, School of Information (*May 2012*) [[29](#)]

Title: [Crowdsourcing and the Law](#)

Reader: James Howison, School of Information

Haofeng Zhou, School of Information (*May 2015*) [[66](#)]

Title: [Crowdsourcing Construction of Information Retrieval Test Collections for Conversational Speech](#)

Reader: Byron Wallace, School of Information

#### **Student Committees**

##### **Ph.D.**

Chris Brown, Linguistics (August 2018)

Title: Analyzing group behavior from language use with Natural Language Processing and experimental methods: Three applications in political science and sociology

Supervisor: David Beaver

Suratna Budalakoti, Electrical & Computer Engineering, (May 2013)  
Title: Authority Identification In Online Communities And Social Networks  
Supervisor: Suzanne Barber

Adriana Kovashka, Department of Computer Science (May 2014) [31]  
Title: Interactive Image Search with Attributes  
Supervisor: Kristen Grauman

Ben Wing, Linguistics (December 2015)  
Title: Text-Based Document Geolocation and its Application to the Digital Humanities  
Supervisor: Jason Baldrige

#### **Masters Theses**

Abhimanu Kumar, Computer Science (*May 2012*) [70, 54, 53]  
Title: [Supervised Language Models for Temporal Resolution of Text in Absence of Explicit Temporal Cues](#)  
Supervisor: Joydeep Ghosh, Electrical & Computer Engineering

Sung Ju Hwang, M.A., Computer Science (May 2010)  
Title: Reading Between The Lines: Object Localization Using Implicit Cues from Image Tags  
Supervisor: Kristen Grauman, Computer Science

#### **Masters Reports**

Hao Lyu, School of Information (May 2016)  
Title: Sarcasm Detection in Twitter  
Supervisor: Byron Wallace, School of Information

#### **University Committees**

Research & Educational Technology (Member: 09/12-05/16)  
*Overhauled the process for review & selection of [LIFT](#) grant proposals*

#### **Departmental Committees**

##### **School of Information**

Awards and Admissions (Member: 2014-15, 2018-19)  
Doctoral Studies (Member: 2010, 2015-16)  
Extended Budget Council (Member: 2015-present)  
Graduate Studies (Chair: 2015-present, Member: 2009-present)  
*The GSC serves as the primary faculty governance structure for the iSchool, with all faculty as members. The chair is annually elected and leads faculty meetings.*  
Information Technology (Chair: 2013-14, Member: 2012-13)  
Research Environment (Chair: 2010-11, 2016-18, Member: 2011-12)  
Web (Member: 2015-16)

##### **Texas Advanced Computing Center (TACC)**

Stampede Technology Council (Member: 2013-present)

#### **Additional University Service**

*Faculty Advisor*, UT Austin ASIS&T Student Chapter (09/09-present)  
*Organizer*, School of Information Open House (Fall 2010 & Spring 2011)  
*Organizer*, School of Information / Yahoo! Speaker Series (2010-2012)