

# James Mayfield

The Johns Hopkins University  
Human Language Technology Center of Excellence  
810 Wyman Park Drive  
Baltimore MD 21211-2840 USA

Brief Curriculum Vitae  
20 March 2020

## Education:

A.B. 1979 Harvard College, Engineering (Cum Laude, general studies).  
Ph.D. 1989 University of California at Berkeley, Computer Science.

## Experience:

2007–present: Johns Hopkins University Human Language Technology Center of Excellence.  
Principal Computer Scientist.  
2004–present: Associate Research Professor, The Johns Hopkins University.  
1996–present: Johns Hopkins University Applied Physics Laboratory.  
Member, Principal Professional Staff.  
1996: University of Maryland, Baltimore County. Associate Professor (tenured).  
1989–1996: University of Maryland, Baltimore County. Assistant Professor.  
1987–1989: University of Maryland, Baltimore County. Instructor.  
1983–1987: University of California at Berkeley. Research Assistant.  
1981–1983: University of California at Berkeley. Teaching Assistant.  
1979–1981: Hewlett Packard Co. Development Engineer.

## Selected Software

- Cold Start code base <<https://github.com/jamesmayfield/ColdStart>>. About 20,000 lines of code used to run the NIST TAC evaluation's Cold Start Knowledge Base Population track 2012–2018. I created the Cold Start task and was the principal author of this code.
- KELVIN Knowledge Base Population system <<https://gitlab.hltcoe.jhu.edu/research/kelvin6>>. About 35,000 lines of code for extracting information from text and organizing it into a knowledge base. KELVIN was one of the top performing systems in the NIST TAC KBP track over the lifetime of the track. I was a contributing author.

Selected Data

- Renmin OCR/NER Collection <<https://github.com/hltcoe/cmn-renmin-ocr-ner-dataset>>. A collection to support the evaluation of named entity recognition (NER) applied to optical character recognition (OCR) output. It supports modification of the OCR system and the NER system without reannotation. It includes over 400,000 characters of NER-annotated Chinese newswire. I was one of the principal designers of this collection.
- Cross Language Entity Linking in 21 Languages Available from <<https://hltcoe.jhu.edu/research/datasets-and-resources/>>. This collection was developed to support the training and evaluation of cross-language linking of named entities from twenty-one non-English languages into an English knowledge base. It includes over 55,000 queries across twenty-one non-English languages, plus an English version of each query. I was one of the principal designers of this collection.

Publications—Selected Books, Book Chapters, and Journal Articles:

- Dawn Lawrie, James Mayfield, Paul McNamee and Douglas W. Oard, ‘Cross-language person-entity linking from twenty languages.’ *Journal of the Association for Information Science and Technology (JASIST)* 66:1106-1123. 2015.
- Paul McNamee and James Mayfield, ‘Character n-gram tokenization for European language text retrieval.’ *Information Retrieval* 7(1-2):73-97. 2004.
- James Mayfield, ‘Ontologies and text retrieval.’ *Knowledge Engineering Review* 17(1):71-75, 2002.
- James Mayfield, ‘Evaluating Plan Recognition Systems: Three Properties of a Good Explanation.’ *Artificial Intelligence Review* 14(4-5):351-376, 2000.
- Robert Wilensky, David N. Chin, Marc Luria, James Martin, James Mayfield, and Dekai Wu, ‘The Berkeley UNIX Consultant project.’ *Artificial Intelligence Review* 14(1-2):43-88, 2000.
- Charles Nicholas and James Mayfield, eds., *Intelligent Hypertext: Advanced Techniques for the World Wide Web*. Springer-Verlag, Lecture Notes in Computer Science series, 1997.
- James Mayfield, ‘Two-level models of hypertext.’ Charles Nicholas and James Mayfield, eds., *Intelligent Hypertext: Advanced Techniques for the World Wide Web*. Springer-Verlag, Lecture Notes in Computer Science series, 1997.
- Tim Finin, Yannis Labrou and James Mayfield, ‘KQML as an agent communication language.’ In Jeffrey Bradshaw, ed., *Software Agents*, MIT Press, 1997.
- Ralph D. Semmel and James Mayfield, ‘Automated query formulation using an entity-relationship conceptual schema.’ *Journal of Intelligent Information Systems* 8(3):267-290, 1997.
- James Mayfield and Charles Nicholas, ‘SNITCH: Augmenting hypertext documents with a semantic net.’ *International Journal of Intelligent and Cooperative Information Systems* 2(3):335-351, 1993.

Publications—Selected Conferences:

- Dawn Lawrie, James Mayfield and David Etter. ‘Building OCR/NER Test Collections.’ In *Proceedings of the Language Resources and Evaluation Conference (LREC 2020)*.
- Cash Costello, Shelby Anderson, Caitlyn Bishop, James Mayfield, and Paul McNamee. ‘Dragonfly: Advances in non-speaker annotation for low resource languages.’ In *Proceedings of the Language Resources and Evaluation Conference (LREC 2020)*.
- Paul McNamee, James Mayfield, Cash Costello, Caitlyn Bishop and Shelby Anderson. ‘Tagging location phrases in text.’ In *Proceedings of the Language Resources and Evaluation Conference (LREC 2020)*.
- Chan Hee Song, Dawn Lawrie, Tim Finin and James Mayfield. ‘Gazetteer Generation for Neural Named Entity Recognition.’ In *Proceedings of the Florida Artificial Intelligence Research Society Conference (FLAIRS 33)*, 2020.
- Yin Ling, Cash Costello, Boliang Zhang, Di Lu, Heng Ji, James Mayfield and Paul McNamee. ‘Platforms for non-speakers annotating names in any language.’ In *Proceedings of ACL 2019*. Melbourne, 2019.
- Tim Finin, Dawn Lawrie, James Mayfield, Paul McNamee and Cash Costello. ‘HLTCOE Participation in TAC KBP 2017: Cold Start TEDL and Low-resource EDL.’ In *Tenth Text Analysis Conference*. NIST, 2017.
- Heng Ji, Xiaoman Pan, Boliang Zhang, Joel Nothman, James Mayfield, Paul McNamee and Cash Costello. ‘Overview of TAC-KBP2017 13 Languages Entity Discovery and Linking.’ In *Tenth Text Analysis Conference*. NIST, 2017.
- James Mayfield, Paul McNamee, and Cash Costello. ‘Language-Independent Named Entity Analysis Using Parallel Projection and Rule-Based Disambiguation.’ BSNLP 2017, p. 92. 2017.
- Paul McNamee, James Mayfield, Samantha Y. Rowe, Alexander K. Rowe, Hannah L. Jackson, and Megan Baker. ‘High Recall Text Retrieval for Public Health Systematic Review.’ *Proceedings of the Florida Artificial Intelligence Research Society Conference (FLAIRS 30)*, Marco Island, Florida, May 2017.
- Tim Finin, Dawn Lawrie, James Mayfield, Paul McNamee, Jessa Laspesa, and Michael Latman. ‘HLTCOE Participation in TAC KBP 2016: Cold Start and EDL.’ *Proceedings of the Ninth Text Analysis Conference*, Gaithersburg, MD, November 2016.
- Tim Finin, Dawn Lawrie, Paul McNamee, James Mayfield, Douglas Oard, Nanyun Peng, Ning Gao, Yiu-Chang Lin, Joshi MacKin, and Tim Dowd. ‘HLTCOE participation in TAC KBP 2015: Cold start and TEDL.’ In *Eighth Text Analysis Conference*. NIST, 2015.
- James Mayfield, *Cold Start Knowledge Base Population at TAC 2014: Task Description*. [nist.gov/tac/2014/KBP/ColdStart/guidelines/KBP2014\\_ColdStartTaskDescription.1.3.pdf](http://nist.gov/tac/2014/KBP/ColdStart/guidelines/KBP2014_ColdStartTaskDescription.1.3.pdf). 2014.

Publications—Selected Conferences *cont.*:

- James Mayfield, ‘Cold Start Knowledge Base Population at TAC 2014: Task Overview.’ *Proceedings of the 2014 Text Analysis Conference*. 2014.
- Tim Finin, Paul McNamee, Dawn Lawrie, James Mayfield and Craig Harman, ‘Hot stuff at Cold Start: HLTCOE participation at TAC 2014.’ *Proceedings of the 2014 Text Analysis Conference*. 2014.
- James Mayfield, Paul McNamee, Craig Harman, Tim Finin and Dawn Lawrie, ‘KELVIN: Extracting knowledge from large text collections.’ *AAAI Fall Symposium on Natural Language Access to Big Data*. 2014.
- Lushan Han, Abhay L. Kashyap, Tim Finin, James Mayfield and Johnathan Weese, ‘UMBC EBILITY-CORE: Semantic Textual Similarity Systems.’ *Proceedings of the Second Joint Conference on Lexical and Computational Semantics*. 2013.
- James Mayfield and Tim Finin, ‘Evaluating the quality of a knowledge base populated from text.’ *Proceedings of AKBC-WEKEX*. 2012.
- Veselin Stoyanov, James Mayfield, Tan Xu, Douglas W. Oard, Dawn Lawrie, Tim Oates and Tim Finin, ‘A context-aware approach to entity linking.’ *Proceedings of AKBC-WEKEX*. 2012.
- Dawn Lawrie, James Mayfield, Paul McNamee and Douglas W. Oard, ‘Creating and Curating a Cross-Language Entity Linking Collection.’ To appear in *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC)*. 2012.
- Paul McNamee, James Mayfield, Douglas W. Oard, Tan Xu, Ke Wu, Veselin Stoyanov and David Doermann. ‘Cross-language entity linking in Maryland during a hurricane.’ *Proceedings of the 2011 Text Analysis Conference*. 2011.
- Paul McNamee, James Mayfield, Dawn Lawrie, Douglas W. Oard and David Doermann. ‘Cross-language entity linking.’ *Proceedings of the Fifth International Joint Conference on Natural Language Processing*, pp. 255–263 2011.
- James Mayfield, Dawn Lawrie, Paul McNamee and Douglas W. Oard. ‘Building a cross-language entity linking collection in twenty-one languages.’ Pamela Forner, Julio Gonzalo, Jaana Keklinen, Mounia Lalmas, and Maarten de Rijke (Eds.), *Proceedings of the Conference on Multilingual and Multimodal Information Access Evaluation*. Lecture Notes in Computer Science vol. 3491, pp. 3–13. 2011.
- Paul McNamee, Mark Dredze, Adam Gerber, Nikesh Garera, Tim Finin, James Mayfield, Christine Piatko, Delip Rao, David Yarowsky, and Markus Dreyer. ‘HLTCOE Approaches to Knowledge Base Population at TAC 2009.’ *Proceedings of the 2009 Text Analysis Conference*. 2009.
- Paul McNamee, James Mayfield and Charles Nicholas. ‘Addressing morphological variation in alphabetic languages.’ *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 75–82. 2009.
- Paul McNamee, James Mayfield and Charles Nicholas. ‘Translation corpus source and size in bilingual retrieval.’ *Proceedings of NAACL/HLT 2009*. pp. 25–28.

Publications—Selected Conferences cont.:

- Tim Finin, Zareen Syed, James Mayfield, Paul McNamee and Christine Piatko. ‘Wikilogy for cross-document entity coreference resolution.’ *AAAI Spring Symposium on Learning by Reading and Learning to Read*. 2009.
- James Mayfield, David Alexander, Bonnie Dorr, Jason Eisner, Tamer Elsayed, Tim Finin, Clay Fink, Marjorie Freedman, Nikeshe Garera, Paul McNamee, Saif Mohammad, Douglas Oard, Christine Piatko, Asad Sayeed, Zareen Syed, Ralph Weischedel, Tan Xu and David Yarowsky. ‘Cross-document coreference resolution: A key technology for learning by reading.’ *AAAI Spring Symposium on Learning by Reading and Learning to Read*. 2009.
- Clay Fink, Christine Piatko, James Mayfield, Tim Finin and Justin Martineau. ‘Geolocating blogs from their textual content.’ *AAAI Spring Symposium on Social Semantic Web: Where Web 2.0 meets Web 3.0*. 2009.
- Paul McNamee, Charles Nicholas and James Mayfield. ‘Don’t have a stemmer? Be Un+concern+ed.’ *Proceedings of the Thirty-First Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 813–814. 2008.
- Paul McNamee and James Mayfield. ‘N-Gram Morphemes for Retrieval.’ Carol Peters (ed.). *Cross-Language Information Retrieval and Evaluation: Workshop of the Cross-Language Evaluation Forum (CLEF 2007)*. 2007.
- R. Scott Cost, Christopher McCubbin, John Cole, Markus Dale, David Scheidt and James Mayfield. ‘Active metadata framework and emergent metadata distribution with scout agents.’ *Proceedings of the 2nd International Workshop on Engineering Emergence in Decentralised Autonomic Systems (EEDAS 2007)*. 2007.
- Akshay Java, Pranam Kolari, Tim Finin, James Mayfield, Anupam Joshi and Justin Martineau. ‘BlogVox: Separating blog wheat from blog chaff.’ *Proceedings of the Workshop on Analytics for Noisy Unstructured Text Data, 20th International Joint Conference on Artificial Intelligence (IJCAI-2007)*. 2007.
- Patrick Schone, Gary Ciany, Roy Cutts, Paul McNamee, James Mayfield and Tom Smith. ‘QACTIS enhancements in TREC QA 2006.’ In Ellen M. Voorhees and Lori P. Buckland, eds., *The Fifteenth Text Retrieval Conference, TREC-2006*. National Institute of Standards and Technology Special Publication. 2007.
- Akshay Java, Pranam Kolari, Tim Finin, Anupam Joshi, Justin Martineau and James Mayfield. ‘The BlogVox opinion retrieval system.’ In Ellen M. Voorhees and Lori P. Buckland, eds., *The Fifteenth Text Retrieval Conference, TREC 2006*. National Institute of Standards and Technology Special Publication. 2006.
- Paul McNamee and James Mayfield. ‘Translation of multiword expressions using parallel suffix arrays.’ *Proceedings of the Seventh Biennial Conference for the Association for Machine Translation in the Americas (AMTA 2006)*. pp. 100–109. 2006.
- James Mayfield and Paul McNamee. ‘Improving QA retrieval using document priors.’ *Proceedings of the Twenty-Ninth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2006)*. pp. 677–678. 2006.

Publications—Selected Conferences cont.:

- James Mayfield and Paul McNamee. ‘JHU/APL at TREC 2005: QA Retrieval and Robust tracks.’ Ellen M. Voorhees and Lori P. Buckland, eds., *The Fourteenth Text Retrieval Conference, TREC 2005*. National Institute of Standards and Technology Special Publication. 2006.
- Pat Schone, Gary Ciany, Roy Cutts, Paul McNamee, James Mayfield and Tom Smith. ‘QACTIS-based question answering at TREC-2005.’ Ellen M. Voorhees and Lori P. Buckland, eds., *The Fourteenth Text Retrieval Conference, TREC 2005*. National Institute of Standards and Technology Special Publication. 2006.
- James Mayfield and Paul McNamee. ‘Translating pieces of words.’ *Proceedings of the Twenty-Eighth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 643–644. 2005.
- James Mayfield. ‘DHS research challenges in cross-language information retrieval.’ *Proceedings of the DHS Text Processing Workshop*. 2005.
- Tim Finin, James Mayfield, Anupam Joshi, R. Scott Cost and Clay Fink. ‘Information retrieval and the Semantic Web.’ *Proceedings of the 38th Hawaii International Conference on System Sciences (HICSS-38)*. 2005.
- P. Schone, G. Ciany, P. McNamee, J. Mayfield, T. Bassi and A. Kulman. ‘Question Answering with QACTIS at TREC-2004.’ Ellen M. Voorhees and Lori P. Buckland, eds., *The Thirteenth Text Retrieval Conference, TREC-2004*. National Institute of Standards and Technology Special Publication 500–261. 2004.
- Paul McNamee and James Mayfield. ‘Cross-language retrieval using HAIRCUT for CLEF 2004.’ Carol Peters, ed., *Cross-Language Information Retrieval and Evaluation: Workshop of the Cross-Language Evaluation Forum, Clef 2004*. Lecture Notes in Computer Science vol. 3491, pp. 50–59. 2005.
- James Mayfield and Paul McNamee. ‘Triangulation without Translation.’ Kalervo Jarvelin, James Allan, Peter Bruzxa and Mark Sanderson, eds., *Proceedings of the Twenty-Seventh Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2004)*. pp. 490–491. 2004.
- Douglas W. Oard, Dagobert Soergel, G. Craig Murray, David Doermann, Jianqiang Wang, Bhuvana Ramabhadran, Martin Franz, James Mayfield, Samuel Gustman and Stephanie Strassel. ‘Building an information retrieval test collection for spontaneous conversational speech.’ *Proceedings of the Twenty-Seventh Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2004)*. pp. 41–48. 2004.
- James Mayfield, Paul McNamee, Christine Piatko and Claudia Pearce. ‘Lattice-based tagging using support vector machines.’ *Proceedings of the Twelfth International ACM Conference on Information and Knowledge Management (CIKM 2003)*, pp. 303–308. 2003.
- Paul McNamee and James Mayfield. ‘JHU/APL experiments in tokenization and non-word translation.’ Carol Peters (ed.). *Cross-Language Information Retrieval and Evaluation: Workshop of the Cross-Language Evaluation Forum (CLEF 2003)*. 2003.

Publications—Selected Conferences cont.:

- ‘N-grams for translation and retrieval in CL-SDR.’ Paul McNamee and James Mayfield. Carol Peters (ed.). *Cross-Language Information Retrieval and Evaluation: Workshop of the Cross-Language Evaluation Forum (CLEF 2003)*. 2003.
- James Mayfield and Paul McNamee. ‘Single n-gram stemming.’ *Proceedings of the Twenty-Sixth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, pp. 415–416. 2003.
- James Mayfield and Tim Finin. ‘Information retrieval on the Semantic Web: Integrating inference and retrieval.’ *Proceedings of the SIGIR 2003 Semantic Web Workshop*, 2003.
- James Mayfield, Paul McNamee and Christine Piatko. ‘Named entity recognition using hundreds of thousands of features.’ *Proceedings of the Seventh Conference on Natural Language Learning*, pp. 184–187. 2003.
- Urvi Shah, Tim Finin, Anupam Joshi, R. Scott Cost and James Mayfield. ‘Information retrieval on the Semantic Web.’ *Proceedings of the Eleventh International ACM Conference on Information and Knowledge Management (CIKM 2002)*, pp. 461–468. 2002.
- Paul McNamee and James Mayfield. ‘Scalable multilingual information access.’ Carol Peters (ed.). *Cross-Language Information Retrieval and Evaluation: Workshop of the Cross-Language Evaluation Forum, CLEF 2002, Revised Papers*.
- James Mayfield and Paul McNamee, ‘Three principles to guide CLIR research.’ *Proceedings of the SIGIR 2002 Workshop on Cross-Language Information Retrieval: A Research Roadmap*, pp. 78–80. 2002.
- Paul McNamee and James Mayfield, ‘Comparing Cross-Language Query Expansion Techniques by Degrading Translation Resources.’ *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2002)*. August, 2002.
- James Mayfield and Paul McNamee, ‘Converting bilingual dictionaries from human-readable to machine-readable form.’ *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2002)*. August, 2002.
- R. Scott Cost, Tim Finin, Yannis Labrou, Xiaochang Luan, Yun Peng, Ian Soboroff, James Mayfield and Akram Boughannam, ‘Jackal: A Java-based tool for agent development.’ *Proceedings of the AAAI Workshop on Software Tools for Developing Agents*, 1998.
- James Mayfield, Tim Finin, Chetan Shah, Rajkumar Narayanaswamy, William MacCartney and Keith Goolsbey, ‘Cyclic Friends Network: Getting Cyc Agents to Reason Together.’ *Proceedings of the ACM CIKM Intelligent Information Agents Workshop*, Baltimore MD. December, 1995.
- James Mayfield, Marty Hall, and Tim Finin, ‘Using automatic memoization as a software engineering tool in real-world AI systems.’ *Proceedings of the Eleventh IEEE Conference on Artificial Intelligence for Applications (CAIA ’95)*, pp. 87–93. Los Angeles, CA. IEEE Computer Society Press, February, 1995.