

Jay Pujara

CONTACT INFORMATION	University of Southern California Information Sciences Institute 4676 Admiralty Way, Ste 1001 Marina del Rey, CA 90292	202 567 7885 jay@cs.umd.edu @jay_mlr	
RESEARCH INTERESTS	Artificial Intelligence Knowledge Graph Construction	Scalable Machine Learning Statistical Relational Learning	Probabilistic Graphical Models Natural Language Processing
EDUCATION	Ph.D. in Computer Science University of Maryland , Computer Science Department, College Park, MD Thesis: <i>Probabilistic Models for Scalable Knowledge Graph Construction</i> Committee: Lise Getoor , Hector Corrada Bravo, William Cohen, Hal Daumé III, Philip Resnik	Fall 2010 – Spring 2016	
	Master of Science in Computer Science Carnegie Mellon University , School of Computer Science, Pittsburgh, PA Thesis: <i>Fundamental Properties of Feature Selection in fMRI Data</i> , advisor Tom Mitchell	Summer 2004 – Spring 2005	
	Bachelor of Science in Computer Science Carnegie Mellon University , School of Computer Science, Pittsburgh, PA Additional Degrees: Cognitive Science, Electrical and Computer Engineering Minors: Robotics, Mathematical Sciences, and Logic and Computation Distinction: Graduated with University Honors and College Honors Thesis: <i>Machine Learning Classification of fMRI Data</i> , advisor Tom Mitchell	Fall 2000 – Spring 2004	
KEY HONORS	Best Paper Award , Statistical Relational AI Workshop Outstanding Reviewer , International Joint Conference on AI Best Student Paper Award , International Semantic Web Conference Best Paper Award , Collaboration, Electronic Messaging, Anti-abuse, and Spam Yahoo! FREP Award , “Active Feature Acquisition”, Advisor: Martin Zinkevich Lemonade Stand Multi-agent Competition	2016 2016 2013 2011 2010-2012 2nd place, 2009; 3rd place, 2010	
ACADEMIC EXPERIENCE	University of Southern California , Marina del Rey, CA Research Assistant Professor , <i>Computer Science Department</i> Research, teaching, and mentoring on topics in artificial intelligence	Summer 2018 - Present	
	University of Southern California , Marina del Rey, CA Research Scientist, Keston Researcher-in-Residence , <i>Information Sciences Institute</i> Part of the Center on Knowledge Graphs working on knowledge graph construction.	Fall 2017 - Present	
	University of California , Santa Cruz, CA Research Consultant , <i>Jack Baskin School of Engineering</i> Coordinating projects and advising students for the D3 Data Science Research Center.	Fall 2017 - Present	
	University of California , Santa Cruz, CA Postdoctoral Scholar , <i>Jack Baskin School of Engineering</i> , Mentor: Lise Getoor Statistical relational learning research, mentoring students, setting up the D3 Data Science Center.	Spring 2016 - Fall 2017	

University of Maryland, College Park, MD Fall 2010 - Spring 2016
Research Assistant, *Computer Science Department*, Mentor: Lise Getoor
Research on probabilistic models for scalable knowledge graph construction.

University of California, Santa Cruz, CA Winter 2016
Lecturer, *Technology and Information Management*
Taught TIM 245, a graduate-level course on Data Mining.

University of California, Santa Cruz, CA Spring 2014 - Winter 2016
Visiting Student, *Jack Baskin School of Engineering*, Mentor: Lise Getoor
Research on streaming inference in probabilistic graphical models.

Carnegie Mellon University, Pittsburgh, PA Fall 2014
Visiting Scholar, *Machine Learning Department*, Mentor: William Cohen
Integrating knowledge graph identification and entity resolution with NELL.

Carnegie Mellon University, Pittsburgh, PA Summer 2004
Research Assistant, *School of Computer Science*, Mentor: Tom Mitchell
Research on fMRI data analysis to engineer more reliable feature selection methods.

University of Pittsburgh, Pittsburgh, PA Summer 2003
Research Programmer, *Learning R&D Center*, Mentor: Walt Schneider
Implemented and validated neural attention models for cognition in Matlab.

Carnegie Mellon University, Pittsburgh, PA Summer 2001
Research Programmer, *Robotics Institute*, Mentor: Henry Schneiderman
Developed a web demo of face detection algorithms and an image ground-truthing application.

PROFESSIONAL
EXPERIENCE

Google Inc, Mountain View, CA Summer 2014
Research Intern, *Google Knowledge Vault*, Mentors: Kevin Murphy and Luna Dong
Scalable collective entity reconciliation for the Knowledge Graph and structured knowledge sources.

LinkedIn Corp., Mountain View, CA Summer 2012
Data Science Intern, *LinkedIn Skills*, Mentor: Peter Skomoroch
Hadoop implementation of hierarchical topic models to discover the structure of LinkedIn skills.

Yahoo! Inc, Sunnyvale, CA [remote] Fall 2010 – Spring 2012
Data Researcher, *Yahoo! Mail*, Mentors: Martin Zinkevich, Gareth Shue
Research on cost-sensitive active feature acquisition for e-mail classification.

Yahoo! Inc, Sunnyvale, CA Fall 2006 – Fall 2010
Senior Software Engineer, *Yahoo! Mail* Fall 2007 – Fall 2010
Software Engineer, *Yahoo! Mail* Fall 2006 – Fall 2007
Team lead for trusted user identification across Yahoo!, spearheaded the migration of spam analysis infrastructure to Hadoop, and overhauled the user feedback processing system. Other projects include defining contextual user experiences, designing reputation systems for URL, IP, and content features, and engineering distributed systems for robust data access.

Oracle Corp, Redwood Shores, CA Fall 2005 – Fall 2006
Member of Technical Staff, *Business Intelligence*
Implemented intelligent caching systems allowing queries to be fulfilled by middleware caches.

TEACHING
EXPERIENCE

University of Southern California, Los Angeles, CA **Fall 2018**
Instructor, *Department of Computer Science*
 Taught graduate-level course CSCI 563/INF 558, "Building Knowledge Graphs."

University of Southern California, Los Angeles, CA **Fall 2017**
Invited Presenter, *CSCI 563: Building Knowledge Graphs*
 Presented a lecture and hands-on demonstration of knowledge graph construction.

University of California, Santa Cruz, CA **Winter 2016**
Lecturer, *Technology and Information Management* Average evaluation (overall): 3.9/5
 Taught graduate-level course TIM 245, "Data Mining."

University of California, Santa Cruz, CA **Spring 2014**
Invited Presenter, *CMPS 290C: Advanced Analytics for Heterogeneous Information Networks*
 Presented a lecture and hands-on demonstration of knowledge graph construction.

National Youth Science Camp, Barstow, WV
Invited Presenter Average evaluation: 3.1/5 **Summer 2013**
Invited Presenter Average evaluation: 4.3/5 **Summer 2012**
Invited Presenter Average evaluation: 3.9/5 **Summer 2011**
 Developed and presented lectures "How to Think Like a Computer Scientist" and "The Mysteries of Computer Science," and three-day interactive seminars on "A Brief, Yet Helpful Guide to Machine Learning" and "Game Theory and Artificial Intelligence."

University of Maryland, College Park, MD **Fall 2011**
Teaching Assistant, *CMSC 421: Artificial Intelligence*
Guest Lecturer, *CMSC 421: Artificial Intelligence*
 Designed and evaluated course examinations, written assignments and hands-on projects. Presented two lectures on "Game Playing and Search" and an alpha-beta pruning activity.

InternalDrive Corporation, Stanford, CA **Summer 2002**
Camp Instructor, *Game Programming and C++*
 Taught courses on Game Programming and C++ to children ages 8-18.

George Washington Community Education Center, Charleston, WV **Fall 1999**
Course Instructor, *Computer Skills, and Introduction to the Internet*
 Developed and taught two, 10-week lab courses, including a free class for senior citizens.

MENTORING

Binh Vu , PhD student, USC, Research Mentor	Winter 2018 - Present
Dhanya Sridhar , PhD student, UCSC, Research Mentor	Fall 2015 - Summer 2018
Pigi Kouki , PhD student, UCSC, Research Mentor	Fall 2015 - Spring 2018
Sabina Tomkins , PhD student, UCSC, Research Mentor	Fall 2015 - Summer 2017
Eriq Augustine , PhD student, UCSC, Research Mentor	Fall 2016 - Summer 2017
Varun Embar , PhD student, UCSC, Research Mentor	Fall 2016 - Present
Molly Zhang , PhD student, UCSC, Research Mentor	Spring 2016 - Winter 2017
Arun Rajendran , MS student, USC, Research Mentor	Winter 2018 - Summer 2018
Bharat Pulvarti , MS student, USC, Research Mentor	Winter 2018 - Present
Pegah Jandaghi , MS student, USC, Research Mentor	Fall 2017 - Present
Prachi Agrawal , MS student, USC, Research Mentor	Fall 2017
Nikhil Kini , MS student, UCSC, Research Mentor	Spring 2016 - Spring 2017
Shachi Kumar , MS student, UCSC, Research Mentor	Fall 2015 - Fall 2016
Hung-Ju Chen , MS student, UCSC, Research Mentor	Fall 2016 - Spring 2017

Johnnie Chang , MS student, UCSC, Research Mentor	Fall 2016 - Spring 2017
Ankit Gupta , MS Project Advisor, UCSC	Spring 2016 - Present
Vedashree Bagade , MS Project Advisor, UCSC	Summer 2016
Anirudh Challa , MS Project Advisor, UCSC	Spring 2016
Stan Thornhill , MS Thesis Advisor, UCSC & Chair	Spring 2016
Pei Zhou , undergraduate student, USC, Research Mentor	Summer 2018
Ankur Goswami , undergraduate, UCSC, Research Mentor	Spring 2017 - Present
Connor Pryor , undergraduate, UCSC, Research Mentor	Summer 2017

ACADEMIC
SERVICE

Organizer , Intelligent Systems Division Retreat at USC ISI	2018
Organizer , 8th Workshop on Statistical Relational AI at IJCAI	2018
Organizer , 6th Workshop on Automated Knowledge Base Construction at NIPS	2017
Organizer , 7th Workshop on Statistical Relational AI at UAI	2017
Organizer , 5th Workshop on Automated Knowledge Base Construction at NAACL	2016
Organizer , Data Science Afternoon, University of California, Santa Cruz	2015
Organizer , Yahoo! Machine Learning Symposium, University of Maryland	2012
Elected Representative , Graduate Student Government, University of Maryland	2012-2014
Elected Representative , CS Department Council, University of Maryland	2012-2013
President , CS Executive Council, University of Maryland	2011-2013
Member , CS Executive Council, University of Maryland	2010-2015
Presenter , National Youth Science Camp, Barstow, WV	2011-2013
Reviewer , Transactions on Knowledge and Data Engineering	2016-2018
Program Committee , Empirical Methods in NLP	2017-2018
Reviewer , Neural Information Processing Systems	2014-2018
Reviewer , International Conference on Machine Learning	2018
Editorial Board , Semantic Web Journal	2017
Program Committee , World Wide Web Conference	2017
Program Committee , Uncertainty in Artificial Intelligence	2017
Program Committee , Association for Computational Linguists	2017
Program Committee , International Semantic Web Conference	2016
Program Committee , International Joint Conference on Artificial Intelligence	2016
Program Committee , Conference on Computational Linguistics	2016
Reviewer , Transactions on Knowledge Discovery from Data	2016
Reviewer , International Semantic Web Conference	2014
Program Committee , Workshop on Knowledge Base Construction (KBCOM)	2018
Program Committee , Workshop on Linked Data for Information Extraction	2014-2017
Program Committee , Workshop on Statistical Relational Artificial Intelligence	2016
Program Committee , Workshop on Automated Knowledge Base Construction	2016

HONORS
AND
AWARDS

Best Paper Award , Statistical Relational AI Workshop	2016
Outstanding Reviewer , International Joint Conference on AI	2016
John D. Gannon Travel Fellowship , University of Maryland	2014
Best Student Paper Award , International Semantic Web Conference	2013
Student Travel Award , International Semantic Web Conference	2013
Travel Scholarship , International Conference on Machine Learning	2011
Best Paper Award , Collaboration, Electronic Messaging, Anti-abuse, and Spam	2011
Dean's Fellowship Award , University of Maryland, College Park	2010-2012
Yahoo! FREP Award , "Active Feature Acquisition", Advisor: Martin Zinkevich	2010-2012

Lemonade Stand Multi-agent Competition

2nd place, 2009; 3rd place, 2010

TUTORIALS

Mining Knowledge Graphs From Text, Web Search and Data Mining (WSDM) 2018
Knowledge Graph Construction From Text, Conference on AI (AAAI) 2017

PUBLICATIONS

Theses

Pujara, J. (2016). "Probabilistic Models for Scalable Knowledge Graph Construction". PhD thesis. University of Maryland, College Park.

Pujara, J. (2005). "Fundamental Properties of Feature Selection in fMRI Data". MS thesis. Carnegie Mellon University.

Journal and Magazine Articles

Kouki, P., **Pujara, J.**, Marcum, C., Koehly, L., Getoor, L., (2018). Collective Entity Resolution in Multi-relational Familial Networks. *Knowledge and Information Systems*.

Pujara, J., Miao, H., Getoor, L., Cohen, W. W., (2015). Using Semantics & Statistics to Turn Data into Knowledge. *AI Magazine* **36**(1), 65–74.

Peer-Reviewed Conference Papers

Chang, J., Chen, R., **Pujara, J.**, Getoor, L., (2018). Clustering System Data using Aggregate Measures. In: *SysML Conference*. [Acceptance Rate: 57.1%]

Sridhar, D., **Pujara, J.**, Getoor, L., (2018a). Scalable Probabilistic Causal Structure Discovery. In: *International Joint Conference on Artificial Intelligence*. [Acceptance Rate: 20.5%]

Kim, S., Kini, N., **Pujara, J.**, Koh, E., Getoor, L., (2017). Probabilistic Visitor Stitching on Cross-Device Web Logs. In: *World Wide Web Conference*. [Acceptance Rate: 17%]

Kouki, P., **Pujara, J.**, Marcum, C., Koehly, L., Getoor, L., (2017a). Collective Entity Resolution in Familial Networks. In: *IEEE International Conference on Data Mining*. [Acceptance Rate: 9.3%]

Kouki, P., Schaffer, J., **Pujara, J.**, O'Donovan, J., Getoor, L., (2017). User Preferences for Hybrid Explanations. In: *ACM Conference on Recommender Systems*. [Acceptance Rate: 16.4%]

Pujara, J., Augustine, E., Getoor, L., (2017). Sparsity and Noise: Where Knowledge Graph Embeddings Fall Short. In: *Conference on Empirical Methods in Natural Language Processing (EMNLP)*. [Acceptance Rate: 18.4%]

Tomkins, S., **Pujara, J.**, Getoor, L., (2017). Disambiguating Energy Disaggregation: A Collective Probabilistic Approach. In: *International Joint Conference on Artificial Intelligence*. [Acceptance Rate: 26%]

Kumar, S., **Pujara, J.**, Getoor, L., Mares, D., Gupta, D., Riloff, E., (2016). Unsupervised Models for Predicting Strategic Relations between Organizations. In: *International Conference on Advances in Social Networks Analysis and Mining*. [Acceptance Rate: 13.6%]

Grycner, A., Weikum, G., **Pujara, J.**, Foulds, J., Getoor, L., (2015). RELLY: Inferring Hypernym Relationships Between Relational Phrases. In: *Conference on Empirical Methods in Natural Language Processing*. [Acceptance Rate: 24%]

Pujara, J., London, B., Getoor, L., (2015). Budgeted Online Collective Inference. In: *Uncertainty and Artificial Intelligence (UAI)*. [Acceptance Rate: 29%]

Pujara, J., Miao, H., Getoor, L., Cohen, W. W., (2013b). Knowledge Graph Identification. In: *International Semantic Web Conference (ISWC)*. **Winner of Best Student Paper award**. [Acceptance Rate: 21.5%]

Pujara, J., Daumé III, H., Getoor, L., (2011). Using Classifier Cascades for Scalable E-Mail Classification. In: *Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference*. **Winner of Best Paper Award**. [Acceptance Rate: 49%]

Refereed Workshop Papers

- Embar, V., Farnadi, G., **Pujara, J.**, Getoor, L., (2018). Aligning Product Categories using Anchor Products. In: *WSDM Workshop on Knowledge Base Construction, Reasoning and Mining*.
- Gupta, R., **Pujara, J.**, Knoblock, C. A., Sharanappa, S. M., Pulavarti, B., Hoberg, G., Phillips, G., (2018). Feature Selection Methods For Understanding Business Competitor Relationships. In: *Fourth International Workshop on Data Science for Macro-Modeling with Financial and Economic Datasets*. ACM SIGMOD.
- Pujara, J.** (2018). Hybrid Link Prediction for Competitor Relationships. In: *Fourth International Workshop on Data Science for Macro-Modeling with Financial and Economic Datasets*. ACM SIGMOD.
- Sridhar, D., **Pujara, J.**, Getoor, L., (2018b). Using Noisy Extractions to Discover Causal Knowledge. In: *Sixth Workshop on Automated Knowledge Base Construction*. NIPS.
- Pujara, J.** (2017). Extracting Knowledge Graphs from Financial Filings. In: *Third International Workshop on Data Science for Macro-Modeling with Financial and Economic Datasets*. ACM SIGMOD.
- Fakhraei, S., Sridhar, D., **Pujara, J.**, Getoor, L., (2016). Adaptive Neighborhood Graph Construction for Inference in Multi-Relational Networks. In: *12th International Workshop on Mining and Learning with Graphs (MLG)*. ACM SIGKDD.
- Pujara, J.**, Getoor, L., (2016). Generic Statistical Relational Entity Resolution in Knowledge Graphs. In: *Sixth International Workshop on Statistical Relational AI. Winner of Best Paper Award*. IJCAI.
- Pujara, J.**, London, B., Getoor, L., Cohen, W. W., (2015). Online Inference for Knowledge Graph Construction. In: *Fifth International Workshop on Statistical Relational AI*. AUAI.
- Grycner, A., Weikum, G., **Pujara, J.**, Foulds, J., Getoor, L., (2014). A Unified Probabilistic Approach for Semantic Clustering of Relational Phrases. In: *Fourth Workshop on Automated Knowledge Base Construction*. NIPS.
- Pujara, J.**, Getoor, L., (2014). Building Dynamic Knowledge Graphs. In: *Fourth Workshop on Automated Knowledge Base Construction*. NIPS.
- Pujara, J.**, Murphy, K., Dong, X. L., Janssen, C., (2014). Probabilistic Models for Collective Entity Resolution Between Knowledge Graphs. In: *Bay Area Machine Learning Symposium*.
- Pujara, J.**, Miao, H., Getoor, L., (2013). Joint Judgments with a Budget: Strategies for Reducing the Cost of Inference. In: *Workshop on Machine Learning with Test-Time Budgets*. ICML.
- Pujara, J.**, Miao, H., Getoor, L., Cohen, W. W., (2013a). Extended Abstract: Large-Scale Knowledge Graph Identification using PSL. In: *AAAI Fall Symposium on Semantics for Big Data*.
- Pujara, J.**, Miao, H., Getoor, L., Cohen, W. W., (2013c). Large-Scale Knowledge Graph Identification using PSL. In: *Workshop on Structured Learning*. ICML.
- Pujara, J.**, Miao, H., Getoor, L., Cohen, W. W., (2013d). Ontology-Aware Partitioning for Knowledge Graph Identification. In: *Third Workshop on Automatic Knowledge Base Construction*. CIKM.
- Huang, B., Bach, S. H., Norris, E., **Pujara, J.**, Getoor, L., (2012). Social Group Modeling with Probabilistic Soft Logic. In: *Workshop on Social Network and Social Media Analysis: Methods, Models, and Applications*. NIPS.
- Pujara, J.**, Skomoroch, P., (2012). Large-Scale Hierarchical Topic Models. In: *Workshop on Big Learning*. NIPS.
- Claudino, L., Khamis, S., Liu, R., London, B., **Pujara, J.**, Plaisant, C., Shneiderman, B., (2011). Facilitating Medication Reconciliation with Animation and Spatial Layout. In: *Workshop on Interactive Healthcare Systems*.
- Pujara, J.**, London, B., Getoor, L., (2011). Reducing Label Cost by Combining Feature Labels and Crowdsourcing. In: *Workshop on Combining Learning Strategies to Reduce Label Cost*. ICML.
- Pujara, J.**, Getoor, L., (2010). Coarse-to-Fine, Cost-Sensitive Classification of E-Mail. In: *Workshop on Coarse-to-Fine Processing*. NIPS.

Patents

- Pujara, J.**, Ramarao, V., Xi, X., Zinkevich, M., Dasgupta, A., Tseng, B., Chu, W., Shue, G., (2016). "User Trustworthiness". 9519682.
- Pujara, J.** (2011). "Real-Time Ad-Hoc Spam Filtering of E-Mail". 8069128.
- Wei, K., Zheng, H., **Pujara, J.**, (2011). "Employing Pixel Density to Detect a Spam Image". 7882177.

Choi, J., **Pujara, J.**, Ramarao, V., Wei, K., (2010). "Identifying IP Addresses for Spammers". 7849146.

INVITED TALKS

What Do Probabilistic Models Know?	
UCLA, Dept of Computer Science	Winter 2018
UC Irvine, Dept of Computer Science	Winter 2018
Probabilistic Models for Large, Noisy, Dynamic Data	
University of Southern California, Dept of Computer Science	Winter 2018
University of Pittsburgh, School of Computing and Information	Winter 2018
University of Alberta, Dept of Computer Science	Fall 2017
Johns Hopkins University, Human Language Technologies Center of Excellence	Winter 2017
University of Southern California, Information Sciences Institute	Winter 2017
University of Iowa, Dept of Computer Science	Winter 2017
Rochester Institute of Technology, Dept of Computer Science	Winter 2017
Analytics for big, noisy, dynamic data	
University of Maryland, Robert H. Smith School of Business	Winter 2017
Probabilistic Soft Logic , UC Santa Cruz Games and Playable Media Group	Fall 2016
Knowledge Graph Construction	
Allen Institute for AI	Summer 2016
Max Planck Institut Informatik	Summer 2015
Karlsruhe Institute of Technology	Summer 2015
UC Santa Cruz, CMPS 290C	Spring 2014
Efficient Online Collective Inference for Graphical Models	Spring 2015
Banff International Research Station, Workshop on New Perspectives for Relational Learning	
Knowledge Graph Identification	
Carnegie Mellon University, ReadTheWeb Group	Fall 2014
University of Maryland, CLIP Colloquium	Spring 2012
Using Classifier Cascades for Scalable E-mail Classification	
University of Maryland, Computer Vision Student Seminar	Winter 2012
Using Hadoop to Fight Spam , Yahoo! Developer network	Spring 2009

SOFTWARE

Knowledge Graph Identification , Lead Developer	Fall 2012 - Summer 2016
https://github.com/linqs/KnowledgeGraphIdentification	
Probabilistic Soft Logic , Frequent Contributor	Spring 2013 - Summer 2016
https://github.com/linqs/psl	
Streaming Inference for PSL , Lead Developer	Spring 2015 - Summer 2015
https://github.com/puuj/uai15-boci-code	
Map-Reduce Latent Dirichlet Allocation , Contributor - Hierarchical LDA	Summer 2012
https://github.com/lintool/Mr.LDA	