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
https://www.jaypujara.org
@jay_mlr

RESEARCH
INTERESTS

Artificial Intelligence	Scalable Machine Learning	Probabilistic Graphical Models
Knowledge Graph Construction	Statistical Relational Learning	Natural Language Processing

EDUCATION

Ph.D. in Computer Science Fall 2010 – Spring 2016
University of Maryland, Computer Science Department, College Park, MD
Thesis: *Probabilistic Models for Scalable Knowledge Graph Construction*
Commitee: **Lise Getoor**, Hector Corrada Bravo, William Cohen, Hal Daumé III, Philip Resnik

Master of Science in Computer Science Summer 2004 – Spring 2005
Carnegie Mellon University, School of Computer Science, Pittsburgh, PA
Thesis: *Fundamental Properties of Feature Selection in fMRI Data*, advisor **Tom Mitchell**

Bachelor of Science in Computer Science Fall 2000 – Spring 2004
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: *Machine Learning Classification of fMRI Data*, advisor **Tom Mitchell**

KEY HONORS

SWSA Ten-Year Award, Knowledge Graph Identification (ISWC)	2023
Best Paper Award, Trustworthy NLP Workshop at NAACL	2021
Outstanding Paper Award, ACM Intelligent User Interfaces	2019
Top 10% Reviewer, Neural Information Processing Systems	2018
Best Paper Award, Statistical Relational AI Workshop	2016
Outstanding Reviewer, International Joint Conference on AI	2016
Best Student Paper Award, International Semantic Web Conference	2013
Best Paper Award, Collaboration, Electronic Messaging, Anti-abuse, and Spam	2011

ACADEMIC
EXPERIENCE

University of Southern California, Marina del Rey, CA	Fall 2017 - Present
Director, *Center on Knowledge Graphs*	Winter 2021 - Present
Principal Scientist, *Information Sciences Institute*	Summer 2022 - Present
Lead Scientist, *Information Sciences Institute*	Summer 2019 - Summer 2022
Research Assistant Professor, *Computer Science*	Summer 2018 - Present
Research Scientist, Keston Researcher-in-Residence, *ISI*	Fall 2017 - Fall 2019

Research, teaching, and mentoring on topics in artificial intelligence, and specializing in knowledge
graph construction.

University of California, Santa Cruz, CA Fall 2017 - Fall 2018
Research Consultant, *Jack Baskin School of Engineering*
Coordinating projects and advising students for the D3 Data Science Research Center.

University of California, Santa Cruz, CA   Spring 2016 - Fall 2017
Postdoctoral Scholar*, Jack Baskin School of Engineering*, Mentor: Lise Getoor
Statistical relational learning research, mentoring students, setting up the D3 Data Science Center.

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.

<table>
<tr><td>TEACHING
EXPERIENCE</td><td>

University of Southern California, Los Angeles, CA

</td></tr>
</table>

TEACHING
EXPERIENCE

University of Southern California, Los Angeles, CA
Instructor, *CSCI-563/INF-558, Department of Computer Science*   **Fall 2018**
Instructor, *CSCI-563/INF-558, Department of Computer Science*   **Spring 2020**
Instructor, *CSCI-563/DSCI-558, Department of Computer Science*   **Spring 2021**
Instructor, *CSCI-563/DSCI-558, Department of Computer Science*   **Fall 2021**
Instructor, *CSCI-563/DSCI-558, Department of Computer Science*   **Fall 2022**
Taught graduate-level course CSCI 563/DSCI 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.

FUNDING   **DARPA I2O**, Environment-driven Conceptual Learning (HR00112390061)   6/23-6/26
               PI, Artificial Domain-Understanding and Collaborative Agency $6,855,000 ($600,000 USC)
               ARL, Strengthening Teamwork Robust Operations Novel Groups (W911NF-23-2-0183) 7/23-6/24
               PI, Decision-making by Analogical Reasoning over Exotic Situations $100,000
               NSF, CISE Research Initiation Initiative (IIS-2153546)   6/23-6/26
               PI, CRII:III: Explainable and Robust AI Agents with Common Sense $175,000

DARPA DSO, Knowledge Management at Speed & Scale (HR00112220046)   9/22-9/25
PI, Knowledge Needed in Context $6,424,454
DARPA I2O, Machine Common Sense (N660011924033)   6/19-2/24
Co-PI (PI Szekely), MOWGLI: Multi-Modal Open World Grounded Learning and Inference $9,850,056
DARPA AIE, Modeling Influence Pathways (HR00112290106)   8/22-1/24
Co-PI (PI Lerman), EMPATH: Predicting Emergent Pathways of Information and Influence $1,000,000
US Air Force, Small Business Innovation Research (FA2384-23-P-0008)   3/23-12/23
PI, Knowledge Graphs for Intelligence, Surveillance and Reconnaissance $50,000
ISI, Exploratory Research Award   1/22-12/23
PI, Understanding Creativity & Collaboration through Temporal Knowledge Graph Learning $100,000
ISI, Exploratory Research Award   1/22-12/23
Co-PI (PI Lerman), Why Resilience In Innovation Is Necessary and How to Foster It $100,000
DARPA I2O, Science of AI and Learning for Open-world Novelty (W911NF2020003)   10/19-6/23
Co-PI (PI Kejriwal), GNOME: Generating Novelty in Open-world Multi-agent Environments $2,966,772
DARPA I2O, Modeling Adversarial Activity (FA8750-20-2-1002)   5/20-12/22
Co-PI (PI Szekely), KGTK - Knowledge Graph Toolkit $1,996,064
DARPA DSO, Systematizing Confidence in Open Research... (W911NF1920271)   10/19-12/22
PI, Macroscopic models for Reproducibility and Replicability $1,829,147
NSF, RAISE/Convergence Accelerator - Open Knowledge Network (OIA-1937153)   9/19-8/22
PI, Leveraging Financial Data – Business OKN $999,117
NSF, RAPID/Convergence Accelerator (89932-Z3582201)   6/20-9/21
Co-PI (PI Raschid), Supply Chain Portal for COVID19 $86,200 ($22,834 USC)
Amazon, Gift-Funded Research   3/20-10/24
PI, Using Temporal Patterns for Proactive KG Updates $50,000
Google, Focused Research Award   3/20-10/24
PI, Reasoning & Discovery Framework for Recommender Dialogs $400,000
JP Morgan, Faculty Research Award   5/20-8/22
PI, Supporting Cognitive Workflows with Hybrid Knowledge Graphs $240,000

MENTORING		
Muhao Chen, Research Scientist, USC, Supervisor	Fall 2020 - Present
Yifan Jiang, PhD student, USC, Advisor	Fall 2023 - Present
Eric Boxer, PhD student, USC, Advisor	Fall 2023 - Present
Kian Ahrabian, PhD student, USC, Advisor	Summer 2022 - Present
Dong-Ho Lee, PhD student, USC, Advisor	Fall 2021 - Present
Pegah Jandaghi, PhD student, USC, Advisor	Fall 2019 - Present
Pei Zhou, PhD student, USC, Advisor	Fall 2019 - Present
Avjit Thawani, PhD student, USC, Advisor	Fall 2019 - Present
Kexuan Sun, PhD student, USC, Advisor	Spring 2020 - Present
Basel Shbita, PhD student, USC, Research Mentor	Fall 2018 - Fall 2021
Binh Vu, PhD student, USC, Research Mentor	Winter 2018 - Fall 2021
Lee Kezar, PhD student, USC, Advisor	Fall 2019 - Spring 2022
Minh Pham, PhD student, USC, Research Mentor	Winter 2018 - Spring 2022
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 - Summer 2019
Molly Zhang, PhD student, UCSC, Research Mentor	Spring 2016 - Winter 2017
Abhinav Jindal, MS student, USC, Research Mentor	Fall 2023 - Present
Kiran Narahari, Volunteer, USC, Research Mentor	Spring 2023 - Present
Saurav Joshi, MS student, USC, Research Mentor	Fall 2023 - Present

Balaji Chidambaram, MS student, USC, Research Mentor	Fall 2022 - Spring 2023
Richard Myloth, MS student, USC, Research Mentor	Spring 2022 - Spring 2023
Arun Ananthan, MS student, USC, Research Mentor	Spring 2022 - Spring 2023
Xinwei Du, MS student, USC, Research Mentor	Spring 2022 - Winter 2023
Shuijing Zhang, MS student, USC, Research Mentor	Fall 2022
Aamir Miyajiwala, MS student, USC, Research Mentor	Fall 2022
Ronak Shah, MS student, USC, Research Mentor	Fall 2022
Hieu Nguyen, MS student, USC, Research Mentor	Fall 2022
Thiloshon Nagarajah, MS student, USC, Research Mentor	Spring 2022
Jae Young Kim, MS student, USC, Research Mentor	Spring 2022
Vineet Agarwal, MS student, USC, Research Mentor	Spring 2022
Kartik Shenoy, MS student, USC, Research Mentor	Summer 2021 - Spring 2022
Abhinav Thakur, MS student, USC, Research Mentor	Fall 2021 - Spring 2022
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 - Summer 2019
Prachi Agrawal, MS student, USC, Research Mentor	Fall 2017
Mei Zhang, undergraduate, USC, Research Mentor	Fall 2021 - Spring 2022
Maia Nkonabang, undergraduate, USC, Research Mentor	Fall 2021 - Spring 2022
Angela Steinmetz, undergraduate, USC, Research Mentor	Fall 2021 - Spring 2022
Timothy Wang, undergraduate, USC, Research Mentor	Fall 2021 - Spring 2022
Srivatsan Srinivasan, undergraduate, UMD, Research Mentor	Summer 2019
Harsha Rayudu, undergraduate student, USC, Research Mentor	Fall 2019 - Fall 2020
Pei Zhou, undergraduate student, USC, Research Mentor	Summer 2018
Anika Jain, MS student, USC, Research Mentor	Fall 2018
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 - Summer 2017
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
Ankur Goswami, undergraduate, UCSC, Research Mentor	Spring 2017 - Summer 2017
Connor Pryor, undergraduate, UCSC, Research Mentor	Summer 2017

Viterbi School of Engineering

Representative (CS) Engineering Faculty Committee,	Fall 2023, Spring 2024
Member, EFC Salary Benchmarking Committee,	Fall 2023, Spring 2024
Mentor for CKIDS projects,	Spring 2022, Fall 2022, Spring 2023
Chair for CKIDS program,	Spring 2019

CS Department

Research Faculty Appointment Committee,	Fall 2021
Research Faculty Appointment Committee,	Fall 2020
PhD Admissions Committee,	Spring 2019
ACM CS Undergrad Research Fair,	Spring 2019

Information Sciences Institute

Task Force on Agile Research Response,	Fall 2023
Mentor, ISI Scientist Mentoring Program	Spring 2023, Fall 2023
Co-Organizer, AI Seminar	Spring 2023 - Present
Speaker, ISI Research Development Seminar	2021, 2023

Member, ISI Space Management Committee		Fall 2022 - Present
Organizer, UnSeminar Series		Spring 2019
Co-Organizer, ISI Intelligent Systems Retreat		Spring 2018

External

Organizer, 9th Workshop on Statistical Relational AI at AAAI	2020
Organizer, Workshop on Reasoning for Complex QA at AAAI	2020
Organizer, Workshop on Data Science for Modeling with Financial Datasets at SIGMOD	2020
Organizer, Workshop on Neurosymbolic Representation and Reasoning at AKBC	2019
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
Participant, NSF Open Knowledge Network Advisory Committee	2022
Senior Program Committee, Conference on Artificial Intelligence	2020-2023
Area Chair, Empirical Methods in NLP	2023
Reviewer, Automated KB Construction	2020, 2022
Reviewer, ACM Computing Surveys	2022
Reviewer, VLDB Journal	2021
Guest Editor, Machine Learning Journal	2021
Program Committee, International Joint Conference on Artificial Intelligence	2016, 2020
Program Committee, Association for Computational Linguistics	2016, 2017, 2020
Reviewer, Transactions on Knowledge and Data Engineering	2016-2019
Area Chair, Automated KB Construction	2019
Program Committee, Empirical Methods in NLP	2017-2018, 2020
Reviewer, Neural Information Processing Systems	2014-2018
Reviewer, International Conference on Machine Learning	2018
Reviewer, Conference on Artificial Intelligence	2018
Editorial Board, Semantic Web Journal	2017
Program Committee, World Wide Web Conference	2017
Program Committee, Uncertainty in Artificial Intelligence	2017
Program Committee, International Semantic Web Conference	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		
	SWSA Ten-Year Award, Knowledge Graph Identification (ISWC)	2023
	ISI Exploratory Research Award, $100k	2022
	ISI Exploratory Research Award, $100k	2022
	Amazon Research Award, $50k gift	2022
	Google Focused Research Award, $80k gift,	2022
	Best Paper Award, TrustNLP Workshop at NAACL	2021
	JP Morgan Faculty Research Award, $90k gift,	2021
	Google Focused Research Award, $190k gift,	2021
	JP Morgan Faculty Research Award, $150k gift,	2020
	Google Focused Research Award, $130k gift,	2020
	Outstanding Paper Award, ACM Intelligent User Interfaces	2019
	Top 10% Reviewer, Neural Information Processing Systems	2018
	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

KGTK: User-friendly Manipulation of Large KGs, Conference on AI (AAAI)	2023
From Tables to Knowledge, Knowledge Discovery and Data Mining (KDD)	2021
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.

Book Chapters

Ilievski, F., Ma, K., Oltramari, A., Wang, P., **Pujara, J.,** (2023). "Building Robust and Explainable AI with Commonsense Knowledge Graphs and Neural Models". In: *Compendium of Neurosymbolic Artificial Intelligence*. Nieuwe Hemweg 6B, 1013 BG Amsterdam, NL: IOS Press, pp. 178–209.

Journal and Magazine Articles

Lerman, K., Yu, Y., Morstatter, F., **Pujara, J.,** (2022). Gendered citation patterns among the scientific elite. *Proceedings of the National Academy of Sciences* **119**(40). [Impact Factor: 12.779]

Gil, Y. (2021). Artificial intelligence for modeling complex systems: taming the complexity of expert models to improve decision making. *ACM Transactions on Interactive Intelligent Systems* **11**(2), 1–49. [Impact Factor: 1.89]

Ghasemi-Gol, M., **Pujara, J.**, Szekely, P., (2020). Learning Cell Embeddings for Understanding Table Layouts. *Knowledge and Information Systems* **1**(64), 39–64. [Impact Factor: 2.94]

Kouki, P., Schaffer, J., **Pujara, J.**, O'Donovan, J., Getoor, L., (2020). Generating and Understanding Personalized Explanations in Hybrid Recommender Systems. *ACM Transactions on Interactive Intelligent Systems* **10**(4), 1–40. [Impact Factor: 2.69]

Kouki, P., **Pujara, J.**, Marcum, C., Koehly, L., Getoor, L., (2018). Collective Entity Resolution in Multi-relational Familial Networks. *Knowledge and Information Systems* **61**(3), 1547–1581. [Impact Factor: 2.94]

Pujara, J., Miao, H., Getoor, L., Cohen, W. W., (2015). Using Semantics & Statistics to Turn Data into Knowledge. *AI Magazine* **36**(1), 65–74. [Impact Factor: 2.52]

Peer-Reviewed Conference Papers

Iglesias-Molina, A., Ahrabian, K., Ilievski, F., **Pujara, J.**, Corcho, O., (2023). Comparison of Knowledge Graph Representations for Consumer Scenarios. In: *International Semantic Web Conference*. [Acceptance Rate: 16%]

Lee, D.-H., Ahrabian, K., Jin, W., Morstatter, F., **Pujara, J.,** (2023). Temporal Knowledge Graph Forecasting Without Knowledge Using In-Context Learning. In: *Conference on Empirical Methods in Natural Language Processing*. [Acceptance Rate: 23.3%]

Lee, D.-H., Kadakia, A., Joshi, B., Chan, A., Liu, Z., Narahari, K., Shibuya, T., Mitani, R., Sekiya, T., **Pujara, J.**, Ren, X., (2023). XMD: An End-to-End Framework for Interactive Explanation-Based Debugging of NLP Models. In: *Association for Computational Linguistics*. [Acceptance Rate: 37%]

Lee, D.-H., **Pujara, J.**, Sewak, M., White, R., Jauhar, S., (2023). Making Large Language Models Better Data Creators. In: *Conference on Empirical Methods in Natural Language Processing*. [Acceptance Rate: 23.3%]

Lee, D.-H., Selvam, R. K., Sarwar, S. M., Lin, B. Y., Morstatter, F., **Pujara, J.**, Boschee, E., Allan, J., Ren, X., (2023). AutoTriggER: Label-Efficient and Robust Named Entity Recognition with Auxiliary Trigger Extraction. In: *European Chapter of the Association for Computational Linguistics*. [Acceptance Rate: 25%]

Lee, J.-H., Lee, D.-H., Sheen, E., Choi, T., **Pujara, J.**, Kim, J., (2023). seq2seq-SC: End-to-End Semantic Communication Systems with Pre-Trained Language Models. In: *Asilomar Conference on Signals, Systems, and Computers*.

Moon, J., Lee, D.-H., Cho, H., Jin, W., Park, C., Kim, M., May, J., **Pujara, J.**, Park, S., (2023). Analyzing Norm Violations in Live-Stream Chat. In: *Conference on Empirical Methods in Natural Language Processing*. [Acceptance Rate: 23.3%]

Thawani, A., Ghanekar, S., Zhu, X., **Pujara, J.,** (2023). Learn Your Tokens: Word-Pooled Tokenization for Language Modeling. In: *Findings of the Association for Computational Linguistics: EMNLP*. [Acceptance Rate: 46.2%]

Wang, Z., Ahrabian, K., Rusti, C., **Pujara, J.**, Lerman, K., (2023). Changes in Research Collaborations During the Pandemic. In: *International Society of Scientometrics and Informetrics Conference*.

Zhou, P., Zhu, A., Hu, J., **Pujara, J.**, Ren, X., Callison-Burch, C., Choi, Y., Ammanabrolu, P., (2023). I Cast Detect Thoughts: Learning to Converse and Guide with Intents and Theory-of-Mind in Dungeons and Dragons. In: *Association for Computational Linguistics*, pp.11136–11155. [Acceptance Rate: 25%]

Albalak, A., Tuan, Y.-L., Jandaghi, P., Pryor, C., Yoffe, L., Ramachandran, D., Getoor, L., **Pujara, J.**, Wang, W. Y., (2022). FETA: A benchmark for few-sample task transfer in open-domain dialogue. In: *Conference on Empirical Methods in Natural Language Processing*. [Acceptance Rate: 17%]

Ilievski, F., **Pujara, J.**, Shenoy, K., (2022). Does Wikidata Support Analogical Reasoning? In: *Iberoamerican Knowledge Graphs and Semantic Web Conference*. [Acceptance Rate: 39%]

Jin, W., Lee, D.-H., Zhu, C., **Pujara, J.**, Ren, X., (2022). Leveraging Visual Knowledge in Language Tasks: An Empirical Study on Intermediate Pre-training for Cross-Modal Knowledge Transfer. In: *Association for Computational Linguistics*. [Acceptance Rate: 25%]

Lee, D.-H., Kadakia, A., Tan, K., Agarwal, M., Feng, X., Shibuya, T., Mitani, R., Sekiya, T., **Pujara, J.**, Ren, X., (2022). Good Examples Make A Faster Learner: Simple Demonstration-based Learning for Low-resource NER. In: *Association for Computational Linguistics*. [Acceptance Rate: 25%]

Sun, K., Qiu, Z., Salinas, A., Huang, Y., Lee, D.-H., Benjamin, D., Morstatter, F., Ren, X., Lerman, K., **Pujara, J.,** (2022). Assessing Scientific Research Papers with Knowledge Graphs. In: *ACM Conference on Research and Development in Information Retrieval (SIGIR)*. [Acceptance Rate: 24.7%]

Zhou, P., Cho, H., Jandaghi, P., Lee, D.-H., Lin, B. Y., **Pujara, J.**, Ren, X., (2022). Reflect, Not Reflex: Inference-Based Common Ground Improves Dialogue Response Quality. In: *Conference on Empirical Methods in Natural Language Processing*. [Acceptance Rate: 17%]

Zhou, P., Gopalakrishnan, K., Hedayatnia, B., Kim, S., **Pujara, J.**, Ren, X., Liu, Y., Hakkani-Tur, D., (2022). Think Before You Speak: Explicitly Generating Implicit Commonsense Knowledge for Response Generation. In: *Association for Computational Linguistics*. [Acceptance Rate: 25%]

Mehrabi, N., Zhou, P., Morstatter, F., **Pujara, J.**, Ren, X., Galstyan, A., (2021). Lawyers are Dishonest? Quantifying Representational Harms in Commonsense Knowledge Resources. In: *Conference on Empirical Methods in Natural Language Processing*. [Acceptance Rate: 23%]

Pham, M., Knoblock, C., Chen, M., Vu, B., **Pujara, J.,** (2021). SPADE: A Semi-supervised Probabilistic Approach for Detecting Errors in Tables. In: *International Joint Conference on Aritificial Intelligence (IJCAI)*. [Acceptance Rate: 14%]

Sun, K., Rayudu, H., **Pujara, J.,** (2021). A Hybrid Probabilistic Approach for Table Understanding. In: *Conference on Artificial Intelligence (AAAI)*. [Acceptance Rate: 21%]

Sun, K., Wang, F., Chen, M., **Pujara, J.,** (2021). Tabular Functional Block Detection with Embedding-based Agglomerative Cell Clustering. In: *Conference on Information and Knowledge Management*. [Acceptance Rate: 21%]

Thawani, A., **Pujara, J.**, Ilievski, F., (2021). Numeracy enhances the Literacy of Language Models. In: *Conference on Empirical Methods in Natural Language Processing*. [Acceptance Rate: 23%]

Thawani, A., **Pujara, J.**, Szekely, P., Ilievski, F., (2021). Representing Numbers in NLP: a Survey and a Vision. In: *Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*. [Acceptance Rate: 28%]

Vu, B., Knoblock, C., Szekely, P., **Pujara, J.**, Pham, M., (2021). A Graph-based Approach for Inferring Semantic Descriptions of Wikipedia Tables. In: *International Semantic Web Conference*. [Acceptance Rate: 22%]

Wang, F., Sun, K., Chen, M., **Pujara, J.**, Szekely, P., (2021). Retrieving Complex Tables with Multi-Granular Graph Representation Learning. In: *ACM Conference on Research and Development in Information Retrieval (SIGIR)*. [Acceptance Rate: 21%]

Wang, F., Sun, K., **Pujara, J.**, Szekely, P., Chen, M., (2021). Table-based Fact Verification With Salience-aware Learning. In: *Findings of the Association for Computational Linguistics: EMNLP 2021*. [Acceptance Rate: 35%]

Zhou, P., Gopalakrishnan, K., Hedayatnia, B., Kim, S., **Pujara, J.**, Ren, X., Liu, Y., Hakkani-Tur, D., (2021). Commonsense-Focused Dialogues for Response Generation An Empirical Study. In: *Proceedings of the Special Interest Group on Discourse and Dialogue*. [Acceptance Rate: 39%]

Zhou, P., Jandaghi, P., Cho, H., Lin, B. Y., **Pujara, J.**, Ren, X., (2021). Probing Commonsense Explanation in Dialogue Response Generation. In: *Findings of the Association for Computational Linguistics: EMNLP 2021*. [Acceptance Rate: 35%]

Zhou, P., Khanna, R., Lee, S., Lin, B. Y., Ho, D., **Pujara, J.**, Ren, X., (2021). RICA: Evaluating Robust Inference Capabilities Based on Commonsense Axioms. In: *Conference on Empirical Methods in Natural Language Processing*. [Acceptance Rate: 23%]

Ghasemi-Gol, M., **Pujara, J.**, Szekely, P., (2019). Tabular Cell Classification Using Pre-Trained Cell Embeddings. In: *International Conference on Data Mining*. [Acceptance Rate: 9%]

Kouki, P., Schaffer, J., **Pujara, J.**, O'Donovan, J., Getoor, L., (2019). Personalized Explanations for Hybrid Recommender Systems. In: *ACM International Conference on Intelligent User Interfaces*. **Winner of Outstanding Paper award**. [Acceptance Rate: 25%]

Pham, M., Knoblock, C., **Pujara, J.,** (2019). Learning Data Transformations with Minimal User Effort. In: *IEEE BigData Conference*. [Acceptance Rate: 38%]

Pujara, J., Rajendran, A., Ghasemi-Gol, M., Szekely, P., (2019). A Common Framework for Developing Table Understanding Models. In: *International Semantic Web Conference - Posters*.

Szekely, P., Garijo, D., Bhatia, D., Wu, J., Yao, Y., **Pujara, J.,** (2019). T2WML: Table To Wikidata Mapping Langauge. In: *ACM International Conference on Knowledge Capture (K-CAP)*. [Acceptance Rate: 18%]

Vu, B., Knoblock, C., **Pujara, J.,** (2019). D-REPR: A Language for Describing and Mapping Diversely-Structured Data Sources to RDF. In: *ACM International Conference on Knowledge Capture (K-CAP)*. [Acceptance Rate: 18%]

Vu, B., Knoblock, C., **Pujara, J.,** (2019). Learning Semantic Models of Data Sources Using Probabilistic Graphical Models. In: *The Web Conference*. [Acceptance Rate: 18%]

Sridhar, D., **Pujara, J.**, Getoor, L., (2018). 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., (2017). 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., (2013). Knowledge Graph Identification. In: *International Semantic Web Conference (ISWC)*. **Winner of Best Student Paper award**. [Acceptance Rate: 21.5%]

Pujara, J., Daume 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

Du, X., Ahrabian, K., Ananthan, A. B. S., Myloth, R. D., **Pujara, J.,** (2023). Graph-Based Structure Aware Citation Intent Classification. In: *Workshop on Scientific Document Understanding at AAAI*.

Jandaghi, P., **Pujara, J.,** (2023). Identifying Quantifiably Verifiable Statements from Text. In: *ACL Workshop on Matching From Unstructured and Structured Data*.

Myloth, R. D., Ahrabian, K., Ananthan, A. B. S., Du, X., **Pujara, J.,** (2023). Is Dynamicity All You Need? In: *Workshop on Scientific Document Understanding at AAAI*.

Sun, K., **Pujara, J.,** (2023). Low-Resource Financial QA with Case-based Reasoning. In: *KDD Workshop on Robust NLP for Finance*.

Augustine, E., Pryor, C., Dickens, C., **Pujara, J.**, Wang, W. Y., Getoor, L., (2022). Visual Sudoku Puzzle Classification: A Suite of Collective Neuro-Symbolic Tasks. In: *Workshop on Neural-Symbolic Learning and Reasoning*.

Nagarajah, T., Ilievski, F., **Pujara, J.,** (2022). Understanding Narratives through Dimensions of Analogy. In: *IJCAI Workshop on Qualitative Reasoning*.

Thawani, A., **Pujara, J.**, Kalyan, A., (2022). Estimating Numbers Without Regression. In: *NeurIPS workshop on MathAI*.

Ilievski, F., **Pujara, J.**, Zhang, H., (2021). Story Generation with Commonsense Knowledge Graphs and Axioms. In: *AKBC Workshop on Commonsense Reasoning and Knowledge Bases*.

Kezar, L., **Pujara, J.,** (2021). Finding Pragmatic Differences Between Disciplines. In: *NAACL Workshop on Scholarly Document Processing*.

Lee, D.-H., Selvam, R. K., Sarwar, S. M., Lin, B. Y., Agarwal, M., Morstatter, F., **Pujara, J.**, Boschee, E., Allan, J., Ren, X., (2021). AutoTriggER: Named Entity Recognition with Auxiliary Trigger Extraction. In: *NAACL Workshop on Trustworthy Natural Langugage Processing*. **Winner of Best Paper award**.

Jandaghi, P., **Pujara, J.,** (2020). Human-like Time Series Summaries via Trend Utility Estimation. In: *Ninth International Workshop on Statistical Relational AI*.

Embar, V., **Pujara, J.**, Getoor, L., (2019). Collective Alignment of Large-scale Ontologies. In: *AKBC Workshop on Federated KBs and the Open Knowledge Network*.

Garijo, D. (2019). An Intelligent Interface for Integrating Climate, Hydrology, Agriculture, and Socioeconomic Models. In: *Proceedings of the 24th International Conference on Intelligent User Interfaces: Companion*.

Pujara, J., Raschid, L., Hoberg, G., Phillips, G., Knoblock, C., (2019). Enterprise OKN: A Federated Knowledge Graph for Financial Data. In: *AKBC Workshop on Federated KBs and the Open Knowledge Network*.

Shbita, B., Rajendran, A., **Pujara, J.**, Knoblock, C., (2019). Parsing, Representing, and Transforming Units of Measure. In: *Modeling the World's Systems*.

Thawani, A., Hu, M., Hu, E., Zafar, H., Divvala, N. T., Singh, A., Qasemi, E., **Pujara, J.,** (2019). Entity linking to knowledge graphs to infer column types and properties. In: *The Semantic Web Challenge on Tabular Data to Knowledge Graph Matching at ISWC*.

Yao, Y., Szekely, P., **Pujara, J.,** (2019). Extensible and Scalable Entity Resolution for Financial Datasets Using RLTK. In: *SIGMOD Workshop on Data Science for Macro-modeling with Financial and Economic Datasets*.

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., (2018). 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., (2013). 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., (2013). Large-Scale Knowledge Graph Identification using PSL. In: *Workshop on Structured Learning*. ICML.

Pujara, J., Miao, H., Getoor, L., Cohen, W. W., (2013). 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`	