MAYANK KEJRIWAL

4676 Admiralty Way, Ste. 1001, Marina Del Rey, CA 90292 | 1-217-819-6696 | kejriwal@isi.edu | kejriwalresearch.azurewebsites.net | https://www.linkedin.com/in/mayankkejriwal/ | https://github.com/mayankkejriwal

CURRENT POSITION

Computer Scientist **Since June 2016**
Information Sciences Institute
University of Southern California (USC) Viterbi School of Engineering

EDUCATION

University of Texas at Austin
M.Sc. and Ph.D. in Computer Science **Aug. 2012-May 2016**
Current G.P.A: *3.83*
Ph.D. thesis: *Populating a Linked Data Entity Name System*
Adviser: Daniel P. Miranker
Key Courses: Natural Language Processing, Information Retrieval, Neural Networks, Semantic Web, Algorithms, Graphical Models

University of Illinois at Urbana-Champaign
B.Sc. in Computer Engineering, Highest Honors **Aug. 2009-Aug. 2012**
Graduation G.P.A: *3.89*
Honors thesis: *Machine Learning Techniques in Offline Handwriting Transcription*
Advisers: P. S. Carney and Kenton McHenry
Upper-level Courses: Artificial Intelligence, Scientific Computation, Machine Learning, Computer Graphics, Theory of Computation, Computer Organization and Design, Digital Systems Laboratory, Analog Signal Processing, Solid State Electronic Devices, Computer Systems Engineering, Senior Research Project, Applied Linear Algebra, Probability with Engineering Applications, Differential Equations, Calculus III, Quantum Physics I, Classical Mechanics I and II, Electromagnetic Fields, Thermal and Statistical Physics, Classical Physics Lab

National University of Singapore
Study Abroad **Jan. 2012-May 2012**
Key Courses: Introduction to Economics, Introduction to Japanese Studies, and Japan in the 20th Century

CURRENTLY FUNDED PROJECTS

Data Scientist in a Box (DSBox)
Funded under DARPA D3M
Key Responsibilities: Performing research and development on, and building, an automated sequence modeling-based pipeline to minimize data cleaning effort by data scientists in the field.

Domain-specific Insight Graphs (DIG)
Funded under DARPA MEMEX
Key Responsibilities: Researching and developing robust information extraction, search, link prediction and entity resolution algorithms to assist investigators and law enforcement in complex domains such as human trafficking and securities fraud.

Text-enabled Humanitarian Operations in Real-time (THOR)
Funded under DARPA LORELEI
Key Responsibilities: Building robust entity linking and network analysis systems to provide situational awareness to humanitarian and military personnel.

SCHOLARSHIPS, AWARDS AND TRAVEL GRANTS

Amazon Cloud Credits for Research Grant	2017
Microsoft Azure for Research Grant	2017
AAAI Doctoral Consortium Travel Award	2015
Microsoft Azure for Research Grant	2014
National Science Foundation Travel Grant	2014
Department Nominee: Microsoft, Google fellowships	2013
Department Travel Grant	2013
MCD Fellowship	2012-2014
Daniel and Carol Dobberpuhl Award	2012
Senior 100 Honorary	2012

Henry O. Koehler Scholarship	2011
Chancellor's Scholar	2010-2012
Edmund J. James Scholar	2009-2012
Dean's List	2009-2012

EXPERIENCE

Information Sciences Institute, USC
Computer Scientist **June 2016-Current**
Information Integration group

Capsenta, Inc.
Semantic Web Consultant **May 2016-June 2016**
Worked on integrating advanced ontology matching solutions in the company's current products

University of Texas at Austin
Graduate Research Assistant **June 2014-May 2016**
Research group of Daniel P. Miranker

CareerBuilder LLC, Norcross, Georgia
Data Scientist Intern **May 2015-Aug. 2015**
Worked in the R&D team on the Recruitment Edge and Company Normalization products

University of Texas at Austin
Teaching Assistant **Aug. 2012-June 2014**
Contemporary Issues in Computer Science (Spring 2014)
Automata Theory (Spring 2013)
Data Management (Falls 2013, 2012)
Artificial Intelligence (Fall 2012)

Rackspace, the Open Cloud Company, San Antonio, Texas
Summer Intern **May 2013-July 2013**
Worked on the Big Data team helping with a new Hadoop-based product roll-out

National Center for Supercomputing Applications, Urbana, Illinois
Undergraduate Research Intern **May 2011-Aug. 2011**
Developed scalable machine learning techniques to automate digital handwriting transcription of terabyte-level US census data

University of Illinois at Urbana-Champaign
Camp Counselor **May 2010-Aug. 2010**
Designed STEM exercises for high-school students, and supervised them thereof

University of Illinois at Urbana-Champaign
Freshman Research Intern **May 2010-Aug. 2010**
Helped to automate biomedical image processing tasks using the ImageJ software tool

SKILLS

- **Specialties** –Machine Learning, Data Integration, Semantic Web, Natural Language Processing
- **Languages**– Java, C/C++, Python, R, MATLAB, SQL, SPARQL, Scala, Latex
- **Systems/Tools**– Hadoop, Microsoft Azure, Eclipse IDE**,** Weka, LibSVM, SecondString, FEBRL, Lucene, Mallet

LANGUAGES

- **Native** – English, Hindi
- **Fluent** – Bengali

PROFESSIONAL SERVICE

- **Co-Chair,** Hybrid Statistical Semantic Understanding and Emerging Semantics Workshop (HSSUES) full-day workshop, to be held at International Semantic Web Conference (ISWC), 2017.
- **Program Committee Member,** Data-driven Discovery of Models (D3M) half-day workshop, to be held at International Conference on Data Mining (ICDM), 2017.

- **Program Committee Member,** Broadening Participation in Data Mining (BPDM) workshop, to be held at KDD 2017.
- **Program Committee Member (Research; Posters and demos),** International Semantic Web Conference (ISWC), 2017.
- **Sub-reviewer,** ACM World Wide Web Conference (WWW), 2017.
- **Program Committee Member (Posters and demos),** International Semantic Web Conference (ISWC), 2016.
- **Program Committee Member,** International Joint Conference on Artificial Intelligence (IJCAI), 2016.
- **Reviewer,** ACM Journal of Data and Information Quality (JDIQ), 2016.
- **Sub-reviewer,** International Joint Conference on Artificial Intelligence (IJCAI), 2015.
- **Sub-reviewer,** Very Large Databases (VLDB), 2013.

OUTREACH

- **AI Topics for K-12,** Editor-in-chief (March 2017-Current)
- **AAAI Connections,** Volunteer (February 2017)
- **Present your research to a 12-year old**, Invited speaker at multiple venues (December 2015-May 2016)

EXTRACURRICULAR SERVICE

- **AI Seminar Coordinator:** Information Sciences Institute, 2017
- **National Science Foundation Data Science Workshop (2015):** an NSF-funded workshop attended by Ph.D. level data scientists from across the United States
- **Heidelberg Laureate Forum (2015):** Participant, one of among 200 globally selected Math and CS students
- **Professional Societies:** International Neural Network Society (INNS), Association for the Advancement of Artificial Intelligence (AAAI), Institute of Electrical and Electronics Engineers (IEEE), Association for Computing Machinery (ACM), Society for Industrial and Applied Mathematics (SIAM)
- **Honors Societies:** Phi Kappa Phi, Tau Beta Pi, Alpha Lambda Delta, Phi Eta Sigma
- **Volunteering and Hackathons:** Engineering Open House, 2010 (University of Illinois at Urbana-Champaign); dataHackUT, 2014 (University of Texas at Austin); Kaggle (2014-current); Internal hackathon, 2015 (CareerBuilder), End Human Trafficking Hackathon, 2016 (Cornell)

RESEARCH MENTOR

- **Haotian Zhang (THOR)**
- **Rahul Kapoor (DIG)**
- **Yixiang Yao (Record Linkage Toolkit)**
- **Jing Peng (THOR)**
- **Qiaozhi Song (THOR)**
- **Daye Nam (THOR)**

TUTORIALS

Constructing Domain-specific Knowledge Graphs (KGC) **October 2017**
Mayank Kejriwal, Craig Knoblock, and Pedro Szekely
Full-day tutorial at International Semantic Web Conference (ISWC), 2017, held in Vienna, Austria

Data Mining in Unusual Domains with Information-rich Knowledge Graph Construction, Inference and Search **August 2017**
Mayank Kejriwal and Pedro Szekely
Conventional Tutorial at KDD, 2017, held in Halifax, Canada

Information Integration **July 2017**
Mayank Kejriwal
Tutorial at IS-GEO Summer Institute, 2017 held in Austin, TX

BOOKS

Knowledge Graphs: Theory, Techniques and Applications **2018**
Mayank Kejriwal, Craig Knoblock, and Pedro Szekely
MIT Press

Populating a Linked Data Entity Name System: A Big Data Solution for Unsupervised Instance Matching **2017**
Mayank Kejriwal
IOS Press, Studies in the Semantic Web Series
ISBN 978-3-89838-717-0 2017

PUBLICATIONS

Adaptive Candidate Generation for Scalable Edge-discovery Tasks on Data Graphs **May 2017**
Mayank Kejriwal
MLG Workshop at ACM KDD, 2017, held in Halifax, Nova Scotia, Canada

Investigative Knowledge Discovery for Combating Illicit Activities **TBD**
Mayank Kejriwal, Pedro Szekely, Craig Knoblock
IEEE Intelligent Systems Magazine (To Appear)

Predicting Role Relevance with Minimal Domain Expertise in a Financial Domain **May 2017**
Mayank Kejriwal, Pedro Szekely
DSMM Workshop at ACM SIGMOD, 2017, held in Chicago, Illinois

Using Contexts and Constraints for Improved Geotagging of Human Trafficking Webpages **May 2017**
Rahul Kapoor, **Mayank Kejriwal**, Pedro Szekely
GeoRich Workshop at ACM SIGMOD, 2017, held in Chicago, Illinois

Supervised Typing of Big Graphs using Semantic Embeddings **May 2017**
Mayank Kejriwal, Pedro Szekely
Semantic Big Data (SBD) Workshop at ACM SIGMOD, 2017, held in Chicago, Illinois

Information Extraction in Illicit Domains **April 2017**
Mayank Kejriwal, Pedro Szekely
ACM World Wide Web Conference, 2017, held in Perth, Australia

Local, Domain-independent Heuristics for the FEIII Challenge: Lessons and Observations **June 2016**
Mayank Kejriwal, Daniel P. Miranker
DSMM Workshop at ACM SIGMOD, 2016, held in San Francisco, California

A Pipeline for Extracting and Deduplicating Domain-Specific Knowledge Bases **October 2015**
Mayank Kejriwal, Qiaoling Liu, Ferosh Jacob and Faizan Javed
Industry track in the IEEE International Conference on Big Data, held in Santa Clara, California

Decision-making Bias in Instance Matching Model Selection **October 2015**
Mayank Kejriwal, Daniel P. Miranker
The 14th International Semantic Web Conference, to be held in Bethlehem, Pennsylvania

An Unsupervised Instance Matcher for Schema-free RDF Data **July 2015**
Mayank Kejriwal, Daniel P. Miranker
The Journal of Web Semantics

Sorted Neighborhood for Schema-free RDF Data **May 2015**
Mayank Kejriwal, Daniel P. Miranker
Winner of best paper award
The 4th Knowledge Discovery and Data Mining meets Linked Open Data workshop,
The 12th European Semantic Web Conference, held in Portoroz, Slovenia

Semi-supervised Instance Matching using Boosted Classifiers **May 2015**
Mayank Kejriwal, Daniel P. Miranker
The 12th European Semantic Web Conference, held in Portoroz, Slovenia

Entity Resolution in a Big Data Framework **January 2015**
Mayank Kejriwal
Doctoral consortium, the 29th Conference on Artificial Intelligence (AAAI), held in Austin, TX

Populating Entity Name Systems for Big Data Integration **October 2014**
Mayank Kejriwal
The 13th International Semantic Web Conference, held in Riva Del Garda, Italy

A Two-Step Blocking Scheme Learner for Scalable Link Discovery **October 2014**
Mayank Kejriwal and Daniel P. Miranker

The 9th International Workshop on Ontology Matching at the 13th International Semantic Web Conference, held in Riva Del Garda, Italy

Schema Matching over Relations, Attributes and Data Values **July 2014**
Aibo Tian, **Mayank Kejriwal** and Daniel P. Miranker
The 26th International Conference on Scientific and Statistical Database Management, held in Aalborg, Denmark

An Unsupervised Algorithm for Learning Blocking Schemes **December 2013**
Mayank Kejriwal and Daniel P. Miranker
The 13th IEEE International Conference on Data Mining, held in Dallas, Texas

Extended Scaled Neural Predictor for Improved Branch Prediction **August 2013**
Zihao Zhou, **Mayank Kejriwal** and Risto Miikkulainen
The IEEE International Conference on Neural Networks, held in Dallas, Texas

A Framework to Access Handwritten Information within Large Digitized Paper Collections **October 2012**
Liana Diesendruck, Luigi Marini, Rob Kooper, **Mayank Kejriwal** and Kenton McHenry
The 8th IEEE International Conference on eScience, held in Chicago, Illinois

ABSTRACTS, POSTERS AND PRESS

Populating a Linked Data Entity Name System **June 2017**
Dissertation abstract in AI Matters

The Summer of Data Science **October 2015**
Quoted in blog post by Kristin Tolle, Director of Data Science at Microsoft Research

Populating a Linked Data Entity Name System **August 2015**
Mayank Kejriwal, Daniel P. Miranker
Poster at the Heidelberg Laureate Forum, held in Heidelberg, Germany

Unsupervised Instance Matching on Schema-free Linked Data **August 2015**
Mayank Kejriwal, Daniel P. Miranker
Poster at the National Science Foundation Data Science Workshop, held in Seattle, Washington

Minimally Supervised Instance Matching: An Alternate Approach **May 2015**
Mayank Kejriwal, Daniel P. Miranker
Poster and abstract at the 12th European Semantic Web Conference, held in Portoroz, Slovenia

The Math Behind…The Web of Linked Data **March 2015**
Mayank Kejriwal
Poster for the *SIAM: Math Matters, Apply It!* program

A Sorted Neighborhood Workflow for the Semantic Web **January 2015**
Mayank Kejriwal and Daniel P. Miranker
Poster and abstract at the 29th Conference on Artificial Intelligence (AAAI)

On Linking Heterogeneous Dataset Collections **October 2014**
Mayank Kejriwal and Daniel P. Miranker
Poster and abstract at the 13th International Semantic Web Conference, held in Riva Del Garda, Italy

Populating Entity Name Systems for Big Data Integration **October 2014**
Mayank Kejriwal and Daniel P. Miranker
Doctoral Consortium Poster at the 13th International Semantic Web Conference, held in Riva Del Garda, Italy

An Unsupervised Algorithm for Learning Blocking Schemes **December 2013**
Mayank Kejriwal and Daniel P. Miranker
Poster at the 13th IEEE International Conference on Data Mining, held in Dallas, Texas

Extended Scaled Neural Predictor for Improved Branch Prediction **August 2013**
Zihao Zhou, **Mayank Kejriwal** and Risto Miikkulainen
Poster at the IEEE International Conference on Neural Networks, held in Dallas, Texas

Digitization and Search: A Non-Traditional Use of HPC **October 2012**
Liana Diesendruck, Luigi Marini, Rob Kooper, **Mayank Kejriwal** and Kenton McHenry
Poster and abstract at the 8th IEEE International Conference on eScience, held in Chicago, Illinois

Toward Free and Searchable Historical Census Images **September 2011**
Kenton McHenry, Luigi Marini, **Mayank Kejriwal**, Rob Kooper and Peter Bajcsy
- Published as an article in the Electronic Imaging & Signal Processing section by the International Society for Optics and Photonics
- Published as a technical article in infoDOCKET

TECHNICAL REPORTS

Experience: Type alignment on DBpedia and Freebase **August 2016**
Mayank Kejriwal and Daniel P. Miranker
Technical Report (arXiv:1608.04442)

Self-contained NoSQL Resources for Cross-Domain RDF **August 2016**
Mayank Kejriwal and Daniel P. Miranker
Technical Report (arXiv:1608.04437)

On the Complexity of Sorted Neighborhood **January 2015**
Mayank Kejriwal and Daniel P. Miranker
Technical Report (arXiv:1501.01696)

A DNF Blocking Scheme Learner for Heterogeneous Datasets **January 2015**
Mayank Kejriwal and Daniel P. Miranker
Technical Report (arXiv:1501.01694)

N-Way Heterogeneous Blocking **February 2014**
Mayank Kejriwal and Daniel P. Miranker
Regular Technical Report (TR-14-06), University of Texas at Austin

TALKS

From Noisy Information Extraction to Rich Information Retrieval in Unusual Domains **June 2017**
NLP Seminar
Information Sciences Institute, USC

Predicting Role Relevance with Minimal Domain Expertise in a Financial Domain **May 2017**
DSMM Workshop at ACM SIGMOD, 2017, held in Chicago, Illinois

Using Contexts and Constraints for Improved Geotagging of Human Trafficking Webpages **May 2017**
GeoRich Workshop at ACM SIGMOD, 2017, held in Chicago, Illinois

Populating a Linked Data Entity Name System **April 2016**
Ph.D. Final Defense, University of Texas at Austin
Committee: Daniel P. Miranker (chair), Raymond Mooney, Joydeep Ghosh, Risto Miikkulainen and Eric Price

Populating a Linked Data Entity Name System **April 2016**
AI Seminar: Interview Talk
Information Sciences Institute, USC

Decision-making Bias in Instance Matching Model Selection **October 2015**
The 14th International Semantic Web Conference, to be held in Bethlehem, Pennsylvania

Decision-making Bias in Instance Matching Model Selection **October 2015**
The 14th International Semantic Web Conference, to be held in Bethlehem, Pennsylvania

Sorted Neighborhood for Schema-free RDF Data **May 2015**
The 4th Knowledge Discovery and Data Mining meets Linked Open Data workshop,
The 12th European Semantic Web Conference, held in Portoroz, Slovenia

Semi-supervised Instance Matching using Boosted Classifiers **May 2015**
The 12th European Semantic Web Conference, held in Portoroz, Slovenia

Entity Resolution in a Big Data Framework **January 2015**
Doctoral consortium, the 29th Conference on Artificial Intelligence (AAAI)

Populating a Linked Data-based Entity Name System **November 2014**
Ph.D. Oral Proposal, University of Texas at Austin
Committee: Daniel P. Miranker (chair), Raymond Mooney, Joydeep Ghosh, Risto Miikkulainen and Eric Price

Populating Entity Name Systems for Big Data Integration **October 2014**
The 13th International Semantic Web Conference, held in Riva Del Garda, Italy

A Two-Step Blocking Scheme Learner for Scalable Link Discovery **October 2014**
The 9th International Workshop on Ontology Matching at the 13th International Semantic Web Conference, held in Riva Del Garda, Italy

Populating an Entity Name System **March 2014**
Research Preparation Exam, University of Texas at Austin
Committee: Daniel P. Miranker (chair), Greg Plaxton and Pradeep Ravikumar

An Unsupervised Algorithm for Learning Blocking Schemes **December 2013**
The 13th IEEE International Conference on Data Mining, held in Dallas, Texas

Machine Learning Techniques for Offline Handwriting Transcription **December 2012**
Undergraduate honors thesis presentation at the University of Illinois at Urbana-Champaign