# KUN QIAN

**Email**: kunqian.usa@gmail.com, **Tel**:+1-(831)-239-8201
**Homepage**: https://kunqian-58.github.io

## RESEARCH INTERESTS

Human-in-the-loop machine learning (Active Learning), Deep Learning, Explainable Artificial Intelligence for NLP. Data Integration and Exchange.

## EDUCATION

**University of California, Santa Cruz, USA** 2012-2017
Advisers: Balder ten Cate, Phokion Kolaitis, and Wang-Chiew Tan
Ph.D. in Computer Science

**Beihang University, CHINA** 2007-2010
Master in Software Engineering

**Chongqing University, CHINA** 2003-2007
Bachelor in Software Engineering

## ACADEMIC AND WORK EXPERIENCE

**IBM Research** 2017 - present
*Software Designer and Research Scientist* *San Jose, CA*

· Part of the Scalable Knowledge Intelligence Group at IBM Almaden Research Center. My work focuses on human-in-the-loop machine learning for entity understanding.

**Main ongoing projects**

- **Explainability for Natural Language Processing**
  - Building a recommendation system for XAI for NLP (still ongoing).
  - Two Research Publications (ACM IUI'2020 demo, AACL'2020 tutorial)
- **Named Entity Normalization**
  - Built **PARTNER**, A Human-in-the-loop system for Entity Name Understanding with Deep Learning.
    - Designed and implemented both the front-end interface and the back-end learning algorithm (BiLSTM-CRF and BERT-CRF models).
    - One research publication (AAAI'20 demo)
  - Built **LUSTRE**, an active learning-based system for explainable entity name structure parsing.
    - Designed and implemented both the front-end interface and the back-end learning algorithm.
    - Two research publications (ICDE'18 demo, COLING'18)
  - Numeric entities normalization that requires complex reasoning.
- **Entity Resolution with Human-in-the-loop Machine Learning**;
  - Built **SystemER**, an active learning-based system for explainable entity resolution.
    - Designed and implemented both the front-end interface and back-end learning algorithm.
    - Four Research Publications (VLDB'19 demo, DSMM@SIGMOD'19, CIKM'19 tutorial, CIKM'17)
  - Low-resource Deep Entity Resolution with Transfer and Active Learning.
    - Designed a low-resource framework with active learning and transfer learning for neural entity resolution.
    - One Research Publication (ACL'19).

**University of California Santa Cruz** 2012 - 2016
*Teaching Assistant & Graduate Student Researcher* *Santa Cruz*

**IBM Research** Summer 2015, Summer 2013
*Summer intern* *San Jose*

**Nanyang Technological University** 2010 - 2011
*Project Officer* *Singapore*

· I worked with Prof. James Cheng (now at The Chinese University of Hong Kong (CUHK)) on a project that compares row-store database systems and column-store database systems.

## PUBLICATIONS

**DBLP Profile:** `https://dblp.uni-trier.de/pers/hd/q/Qian_0002:Kun`

(*Authors are ordered alphabetically if (1) it is a technical tutorial, and (2) the work was done with my Ph.D. adviser where we adopted the convention in theory community)

1. Eno Oduor, **Kun Qian**, Yunyao Li, Lucian Popa
   ***XAIT: An Interactive Website for Explainable AI for Text.***
   (**IUI 2020**) The 25th International Conference on Intelligent User Interfaces. To appear in March 2020.

2. Shipi Dhanorkar, Yunyao Li, Lucian Popa, **Kun Qian***, Christine T Wolf, and Anbang Xu.
   ***Explainability for Natural Language Processing.***
   **(AACL-IJCNLP 2020)** The 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics. To appear in December 2020.
   - Summer intern project that I mentored.

3. **Kun Qian**, Poornima Chozhiyath Raman, Yunyao Li, and Lucian Popa.
   ***PARTNER: Human-in-the-loop Entity Name Understanding with Deep Learning.***
   (**AAAI-2020**) The 34th AAAI Conference on Artificial Intelligence (demo).

4. Sairam Gurajada, Lucian Popa, **Kun Qian***, and Prithviraj Sen.
   ***Learning based Human-in-the-loop Methods for Entity Resolution.*** Tutorial.
   (**CIKM'19**) 28th ACM International Conference on Information and Knowledge Management.

5. **Kun Qian**, Douglas Burdick, Sairam Gurajada, and Lucian Popa.
   ***Learning Explainable Entity Resolution Algorithms for Small Business Data using SystemER.***
   (**DSMM'19@SIGMOD'19**) Data Science for Macro-modeling with Financial and Economic Datasets .

6. **Kun Qian**, Lucian Popa, and Prithviraj Sen.
   ***SystemER: A Human-in-the-loop System for Explainable Entity Resolution.***
   (**VLDB-2019**) The 45th International Conference on Very Large Data Bases.

7. Jungo Kasai, **Kun Qian**, Sairam Gurajada, Yunyao Li, Lucian Popa.
   ***Low-resource Deep Entity Resolution with Transfer and Active Learning.***
   (**ACL-2019**) The 57th Annual Meeting of The Association for Computational Linguitics.
   - Summer intern project that I mentored.

8. Phokion G. Kolaitis, Lucian Popa, and **Kun Qian***.
   ***Knowledge Refinement via Rule Selection.***
   (**AAAI-2019**) The 33rd AAAI Conference on Artificial Intelligence .
   - Oral and poster presentation. Acceptance rate: 16.2%.

9. Nikita Bhutani, **Kun Qian**, Yunyao Li, H.V. Jagadish, Mauricio A. Hernandez, Mitesh Vasa.
   ***Exploiting Structure in Representation of Named Entities using Active Learning.***
   (**COLING 2018**) The 27th International Conference on Computational Linguistics, pp. 687-699.
   - Summer intern project that I mentored.
   - Also included in "**IBM Research AI Selected Publications 2018**".

10. Balder ten Cate, Phokion Kolaitis, **Kun Qian***, and Wang-Chiew Tan.
    ***Active Learning of GAV Schema Mapppings.***
    **(PODS'18)** The 37th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems.

11. **Kun Qian**, Nikita Bhutani, Yunyao Li, H.V. Jagadish, Mauricio Hernandez.
    ***LUSTRE: An Interactive System for Entity Structured Representation and Variant Generation.***
    (**ICDE 2018**) 34th IEEE International Conference on Data Engineering. Paris, France. 2018, pp 1613-1616.

12. **Kun Qian**, Lucian Popa, Prithviraj Sen.
    ***Active Learning for Large-Scale Entity Resolution.***
    (**CIKM 2017**) 26th ACM International Conference on Information and Knowledge Management.

13. **Kun Qian**.
    ***Discovering Information Specifications from Data Examples.*** UCSC PhD dissertation. 2017

14. Balder ten Cate, Phokion G. Kolaitis, **Kun Qian***, and Wang-Chiew Tan.
***Approximation Algorithms for Schema-Mapping Discovery from Data Examples.***
(**ACM TODS**) ACM Transactions on Database Systems . Vol. 42, Issue 2, pp 12:1–12:41. 2017.

15. Balder ten Cate, Phokion G. Kolaitis, **Kun Qian***, and Wang-Chiew Tan.
***Approximation Algorithms for Schema-Mapping Discovery from Data Examples.***
(**AMW 2015**) Alberto Mendelzon International Workshop on Foundations of Data Management 2015.

**Granted Patents**

16. Nikita Bhutani, Mauricio Hernandez-Sherrington, Yunyao Li, Min Li, and **Kun Qian**.
***Entity Structured Representation and Variant Generation***. U.S. Patent 10,585,986, issued March 10, 2020.

**Filed Patents**

17. **Kun Qian**, Yunyao Li, and Nikita Bhutani.
***Resolving Queries using Structured and Unstructured Data.*** (Filed, under review).
18. Jungo Kasai, **Kun Qian**, Sairam Gurajada, Yunyao Li, and Lucian Popa.
***Low-resource Deep Entity Resolution with Transfer Learning***. (Filed, under review)
19. **Kun Qian**, Lucian Popa, Prithraj Sen, and Min Li.
***Learning Models For Entity Resolution Using Active Learning.*** (Filed, under review).

*Publications before 2015*

20. Xiaoping Du, Huamei Sun, **Kun Qian**, Yun Li, Liaotao Lu
*A Prediction Model for Vehicle Sideslip Angle Based on Neural Network.* IEEE ICIFE. 2010
21. Weiguo Li, Hanjie Zhang, Xiaoping Du, **Kun Qian**, Cuiying Li
*Data Analysis of Roadway Attributes through Partial Least Squares Regression.* IEEE ICIFE. 2010
22. **Kun Qian**, Sachio Hirokawa, Kenji Ejima, Xiaoping Du
*A Fast Associative Mining System Based on Search Engine and Concept Graph for Large-Scale Financial Report Texts.* IEEE ICIFE. 2010
23. **Kun Qian**, Xiaoping Du, Weiguo Li, Huamei Sun, Cuiying Li, Dezao Hou
*Data Analysis of Roadway Attributes' Influences upon Speed of Small Car on Mountain Highway through Clustering Algorithm.* IEEE ICIFE. 2010
24. Weiguo Li, Cuiying Li, Xiaoping Du, **Kun Qian**, Hanjie Zhang, and Dezao Hou
*A Traffic Flow Prediction Model based on Ordered Logistic Regression.* International Conference on Digital Content, Multimedia Technology and its Applications (IDC). 2010, pages 213-216
25. Xiaoping Du, Lelai Deng, **Kun Qian**. *Current Market Top Business Scopes TrendA Concurrent Text and Time Series Active Learning Study of NASDAQ and NYSE Stocks from 2012 to 2017.* Applied Sciences. 2018; 8(5):751

## INVITED TALKS

| | |
|---|---|
| **October 2019** | "Low-resource Deep Entity Resolution with Transfer and Active Learning". UCSC, California. |
| **Feb 2019** | "Human-in-the-loop Entity Resolution for Knowledge Curation". Stanford University, California. |
| **April 2018** | "Active Learning for Large-Scale Entity Resolution". Telecom ParisTech. Paris, France |
| **November 2017** | "Active Learning for Large-Scale Entity Resolution". Chongqing University. Chongqing, China |

## PROFESSIONAL AFFILIATIONS AND SERVICES

| | |
|---|---|
| **Journal Referee** | ACM TODS (2018, 2019), IEEE TKDE (2019) |
| **Conference PC** | ACL 2020, IJCAI 2020, ICDE 2020 (industry), AAAI 2020<br>IEEE BigData 2019, WebDB 2018 |
| **External Reviewers** | CIKM 2018, CIKM 2017, KDD 2017, AAAI 2017, ADAMA 2017 |
| **Membership** | AAAI |

## AWARDS

- IBM Class-A Research Accomplishment 2017
- UC Regents Fellowship
- Exceptional Mater Student - Beihang University
- Japan JASSO scholarship
- Exceptional Undergraduate Student - Chongqing University

## PROGRAMMING SKILLS

| | |
|---|---|
| **Programming** | Python, Java |
| **Web** | Angular, Angular Material, Django, HTML5, Javascript, CSS, W3.CSS, AngularJS, AngularJS Material |
| **Distributed Computing** | MapReduce, Spark, IBM Infosphere Streams |
| **Deep learning** | Pytorch, Pytorch-Transformers |