KUN QIAN

Email: qian.kun@ibm.com   **Tel**:+1-(831)-239-8201   **LinkedIn**:https://www.linkedin.com/in/kun-qian/

Homepage: https://kunqian-58.github.io

RESEARCH INTERESTS

Entity resolution and knowledge creation, data integration, active learning, and artificial intelligence.

EDUCATION

University of California, Santa Cruz, USA　　　　　　　Graduation date: March 24th, 2017
Advisers: Balder ten Cate, Phokion Kolaitis, and Wang-Chiew Tan
Ph.D. of Computer Science

Beihang University, CHINA　　　　　　　Graduation date: July 30th, 2010
Master of Engineering in Software Engineering

Chongqing University, CHINA　　　　　　　Graduation date: June 30th, 2007
Bachelor of Engineering in Software Engineering

ACADEMIC AND WORK EXPERIENCE

IBM Research　　　　　　　February 14th 2017 - present
Software Designer and Researcher　　　　　　　*San Jose, CA*

· I am a member of the Scalable Knowledge Intelligence Group at IBM Almaden Research Center. My work at IBM focuses on designing and implementing human-in-the-loop machine learning (or, active learning-based) systems for entity resolution and knowledge creation. My duty includes publishing papers in top research conferences, mentoring summer research interns, transferring advanced technologies to IBM products, as well as producing patents.

Concrete duties include:
- designing and implementing machine learning algorithms for entity understanding and knowledge creation;
- conduct experimental evaluation of designed algorithms on concrete datasets;
- publishing research papers as well as system demos at academic journals and conferences;
- interacting with IBM development teams and other researchers within the IBM Research division;
- mentoring summer interns who work on projects related to entity resolution and knowledge creation.

University of California Santa Cruz　　　　　　　January 6th, 2012 - December 31st, 2016
Teaching Assistant & Graduate Student Researcher　　　　　　　*Santa Cruz*

· Worked as teaching assistant and research assistant with faculty members at UC Santa Cruz. Research topic: establish theoretical foundations for data exchange and data integration.

Concrete duties include:
- Reading and presenting research papers relating to data exchange and data integration;
- Designing, implementing, and evaluating novel algorithms for information integration specifications discovery;
- writing research papers and technical reports.
- helping design the course, preparing materials, grading assignments, and teaching lab sessions;

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

· I worked on a research project that focused on designing and implementing an active learning-based system for large-scale entity resolution.

Nanyang Technological University　　　　　　　October 21st, 2010 - August 15th, 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

Google Scholar Profile: `https://scholar.google.ca/citations?user=T-9ljuEAAAAJ&hl=en`

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

(* Authors are ordered alphabetically in two cases: (1) it is a technical tutorial, and (2) the work done with my Ph.D. adviser Phokion G. Kolaitis and we adopted the convention in theory community)

1. Shipi Dhanorkar, Yunyao Li, Lucian Popa, **Kun Qian***, Christine T Wolf, and Anbang Xu. *Explainability for Natural Language Processing.* The 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics. (AACL-IJCNLP) 2020.

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

3. Sairam Gurajada, Lucian Popa, **Kun Qian**, and Prithviraj Sen. *Learning based Human-in-the-loop Methods for Entity Resolution.* Tutorial. 28th ACM International Conference on Information and Knowledge Management (CIKM 2019). Beijing.

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

5. **Kun Qian**, Lucian Popa, and Prithviraj Sen. *SystemER: A Human-in-the-loop System for Explainable Entity Resolution.* The 45th International Conference on Very Large Data Bases (VLDB-19).

6. Jungo Kasai, **Kun Qian**, Sairam Gurajada, Yunyao Li, Lucian Popa. *Low-resource Deep Entity Resolution with Transfer and Active Learning.* The 57th Annual Meeting of The Association for Computational Linguistics (ACL-19).

7. Phokion G. Kolaitis, Lucian Popa, and **Kun Qian***. *Knowledge Refinement via Rule Selection.* AAAI Conference on Artificial Intelligence (AAAI-19).
   - Oral and poster presentation. Acceptance rate: 16.2%.

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

9. Balder ten Cate, Phokion Kolaitis, **Kun Qian***, and Wang-Chiew Tan. Active Learning of GAV Schema Mapppings. In *Proceedings of the 37th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems*, SIGMOD/PODS'18, pp. 355-368, 2018. ACM

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

11. **Kun Qian**, Lucian Popa, Prithviraj Sen. *Active Learning for Large-Scale Entity Resolution.* 26th ACM International Conference on Information and Knowledge Management (CIKM) 2017. Singapore. pp. 1379-1388

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

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

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

Patents

15. Jungo Kasai, **Kun Qian**, Sairam Gurajada, Yunyao Li, and Lucian Popa. *Low-resource Deep Entity Resolution with Transfer Learning.* (filed, under review)

16. **Kun Qian**, Lucian Popa, Prithraj Sen, and Min Li. *Learning Models For Entity Resolution Using Active Learning. Filed, under review.*

17. *Nikita Bhutani, Mauricio Hernandez, Yunyao Li, Min Li, and **Kun Qian**. Entity Structured Representation and Variant Generation. (filed, under review)*

Publications before 2015

18. *Xiaoping Du, Huamei Sun, **Kun Qian**, Yun Li, Liaotao Lu*
    A Prediction Model for Vehicle Sideslip Angle Based on Neural Network. *IEEE ICIFE. 2010*

19. *Weiguo Li, Hanjie Zhang, Xiaoping Du, **Kun Qian**, Cuiying Li*
    Data Analysis of Roadway Attributes through Partial Least Squares Regression. *IEEE ICIFE. 2010*

20. ***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*

21. **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*

22. *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*

23. *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 program committee	ICDE 2020 (industry), AAAI 2020, IEEE BigData 2019, WebDB 2018
Conference 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	Java, Python, C#, C/C++
Web	Angular, Angular Material, Django, HTML5, Javascript, CSS, W3.CSS, AngularJS, AngularJS Material
Distributed Computing	MapReduce, Spark, IBM Infosphere Streams
Deep learning	Pytorch