Kun Qian
United States | +1-831-239-8201
kunqian.usa@gmail.com
personal website | linkedin |

Research Interests

Active Learning/Human-in-the-loop Machine Learning, Machine Learning with Small Data

Natural Language Processing, Weak Supervision, Explainable AI, Data Integration and Exchange.

Education

University of California, Santa Cruz *2012-2017 / USA*

Ph.D. in Computer Science

•Advisers: Balder ten Cate, Phokion Kolaitis, Wang-Chiew Tan

Beihang University *2007-2010 / China*

Master of Software Engineering

Kyushu University *2008-2009 / Japan*

Beihang University - Kyushu University Exchange Student Program

Chongqing University *2003-2007/ China*

Bachelor of Software Engineering

Experience

Amazon.com, Inc. *2021.03 - Now/Seattle*

Applied Scientist

•A member of Global Search Quality team at Search Science and AI department.

IBM Research AI *2017.02 - 2021.03/San Jose*

Research Staff Member

•A member of the Scalable Knowledge Intelligence Group

•Technical Lead of Entity Understanding and Normalization.

•Technical lead of Explainable AI in Natural Language Processing.

•Research focus:human-in-the-loop machine learning, weak supervision.

•Work published in AAAI, AACL, ACL, CIKM, COLING, DIS, EMNLP, ICDE, ISWC, IUI, KDD, PODS, VLDB, etc.

•1 patent granted and 4 patents under review

•Delivered Research Assets to Watson Discovery, Watson Compare and Comply.

IBM Almaden Research Center *summer 2013, summer 2015/San Jose*

Research Intern

Nanyang Technological University *2010-2011/Singapore*

Project Officer

•Research Project: Column-store vs. Row-store database systems. Worked with James Cheng, now at The Chinese University of Hong Kong (CUHK)

Publications

Click to see my **DBLP Profile** and **Google Scholar Profile**

2021

1. Marina Danilevsky, Shipi Dhanorkar, Yunyao Li, Lucian Popa, **Kun Qian**, and Anbang Xu.
Tutorial: Explainability for Natural Language Processing
(SIGKDD 2021, to appear) The 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining.

2. Shipi Dhanorkar, Christine T. Wolf, **Kun Qian**, Anbang Xu, Lucian Popa, and Yunyao Li.
Who needs to know what, when?: Broadening the Explainable AI (XAI) Design Space by Looking at Explanations Across the AI Lifecycle.
(DIS 2021, to appear) The 2021 ACM Designing Interactive Systems.

3. **Kun Qian**, Marina Danilevsky, Yannis Katsis, Ban Kawas, Erick Oduor, Lucian Popa, Yunyao Li
XNLP: A Living Survey for XAI Research in Natural Language Processing.
(IUI 2021) Annual Conference on Intelligent User Interfaces (demo track).

2020

4. **Kun Qian**, Poornima Chozhiyath Raman, Lucian Popa, and Yunyao Li
Learning Structured Representations of Entity Names using Active Learning and Weak Supervision.
(EMNLP 2020) The 2020 Conference on Empirical Methods in Natural Language Processing.
    - Acceptance rate: 16.7%

5. Marina Danilevsky, **Kun Qian**, Ranit Aharonov, Yannis Katasis, Ban Kawas, Prithviraj Sen
A Survey of the State of Explainable AI for Natural Language Processing.
(AACL-IJCNLP 2020) The 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics.

6. Domenico Lembo, Yunyao Li, Lucian Popa, **Kun Qian**, Federico Scafoglieri
Ontology Mediated Information Extraction with MASTRO SYSTEM-T.
(ISWC 2020) The 19th International Semantic Web Conference.
Best Demo Award

7. **Kun Qian**, Lucian Popa, and Yunyao Li
An Intuitive User Interface for Human-in-the-loop Entity Name Parsing and Entity Variant Generation.
(DaSH@SIGKDD) 1st Workshop on Data Science with Human-in-the-loop.

8. Nikita Bhutani, Xinyi Zheng, **Kun Qian**, Yunyao Li and H.V. Jagadish
Answering Complex Questions by Combining Information from Curatedand Extracted Knowledge Bases.
(ACL-NLI) 1st Workshop on Natural Language Interface @ACL 2020.

9. 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.

10. 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.

11. **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).

2019

12. 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.

13. **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 .

14. **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.

15. 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.

16. Phokion G. Kolaitis, Lucian Popa, and **Kun Qian***.
Knowledge Refinement via Rule Selection.
(AAAI-2019) The 33rd AAAI Conference on Artificial Intelligence .

2018

17. 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.
    - Included in "**IBM Research AI Selected Publications 2018**".

18. 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.

19. **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.

2015 - 2017

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

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

22. 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.

23. 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.

Before 2015

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

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

26. **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

27. **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

28. 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

29. Xiaoping Du, Lelai Deng, **Kun Qian**. *Current Market Top Business Scopes Trend—A Concurrent Text and Time Series Active Learning Study of NASDAQ and NYSE Stocks from 2012 to 2017.*
Applied Sciences. 2018; 8(5):751

PATENTS

1. Nikita Bhutani, Mauricio Hernandez-Sherrington, Yunyao Li, Min Li, and **Kun Qian**.
Entity Structured Representation and Variant Generation.
U.S. Patent 10,585,986
Status: **Granted**

2. **Kun Qian**, Yunyao Li, and Nikita Bhutani.
Resolving Queries using Structured and Unstructured Data.
Status: Filed

3. Jungo Kasai, **Kun Qian**, Sairam Gurajada, Yunyao Li, and Lucian Popa.
Low-resource Deep Entity Resolution with Transfer Learning.
Status: Filed

4. **Kun Qian**, Lucian Popa, Prithraj Sen, and Min Li.
Learning Models For Entity Resolution Using Active Learning.
Status: Filed

5. Sheshera Mysor, Sairam Gurajada, Lucian Popa, **Kun Qian**, and Prithraj Sen.
Deep Learning of Entity Resolution Rules (Filed, under review).
Status: Filed

Research interns I worked with

- Shipi Dhanorkar (2019) - PhD student at Pennsylvania State University
- Jungo Kasai (2018) - PhD student at University of Washington
- Nikita Bhutani (2017) - now at Research Scientist at Megagons Lab

Professional Affiliations and Services

Membership

- ACM, AAAI

Conference Program Committee Member

- EMNLP 2021, DaSH@NAACL 2021, IUI 2021 (demo), NAACL 2021, AAAI 2021
- ACL 2020, IJCAI 2020, ICDE 2020 (industry), AAAI 2020
- IEEE BigData 2019
- WebDB 2018, CIKM 2018
- CIKM 2017, KDD 2017, AAAI 2017, ADAMA 2017

Journal Referee

- ACM TODS (2018, 2019), IEEE TKDE (2019)

Awards

Best Demo Paper Award - ISWC 2020	*2020*
IBM Class-A Research Accomplishment	*2017*
UC Regents Fellowship	*2012*
Exceptional Mater Student - Beihang University	*2012*
Japan JASSO scholarship	*2008*
Exceptional Undergraduate Student - Chongqing University	*2007*

Skills

Programming Languages	Python | Java | Javascript
Deep Learning & Datea Science	PyTorch | Pytorch-Transformers | NLTK | Jupyter | Matplotlib | Numpy | Pandas
Web Programming	Angular | Angular Material | W3.css | Typescript | Django
Cloud & Distributed Computing	AWS | PySpark
Others	IBM SystemT | LaTeX