KUN QIAN

Email: kunqian.usa@gmail.com, **Tel**:+1-(831)-239-8201

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

RESEARCH INTERESTS

Machine Learning/Deep Learning, Natural Language Processing, Active Learning/Human-in-the-loop Machine Learning, Weak Supervision, Explainable AI, Data Integration and Exchange.

EDUCATION

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

Beihang University, CHINA Beijing, China
Master in Software Engineering
Visited Kyushu University (Fukuoka, Japan) as 1-year exchange student.

Chongqing University, CHINA Chongqing, China
Bachelor in Software Engineering

WORK EXPERIENCE

IBM Research 2017-2021
Research Staff Member *San Jose, CA*

· Member of the Scalable Knowledge Intelligence Group at IBM Almaden Research Center. My work focuses on designing and developing human-in-the-loop machine learning systems for various entity-centric tasks including entity matching and entity normalization.

Projects Product Deliveries

- **Explainability for Natural Language Processing (Project Lead)**
  - Interactive website for XAI for NLP (demo paper in submission).
    - https://xainlp2020.github.io/xainlp/
  - Research Publications (AACL-IJCNLP 2020, ACM IUI'20 demo, AACL'20 tutorial)
- **Named Entity Normalization (Project Lead)**
  - 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).
    - Research publications: EMNLP 2020, AAAI 2020 demo
    - Video demo: https://youtu.be/6DDXARJezz4
  - 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.
    - Research publications: ICDE'18 demo, COLING'18 full paper
  - **Product Impact**:
    - Date/Time normalization for IBM Watson Discovery (delivered in Q2 2019)
    - Numeric entities (Currency, Complex DateTime, Numbers, etc.) normalization that requires complex reasoning planned for Q1 2021
- **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.
    - Research publications: VLDB'19 demo, DSMM@SIGMOD'19, CIKM'19 tutorial, CIKM'17 full paper
    - Video demo: https://youtu.be/5ENye9hg-UA
  - 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.
    - Research publication: ACL'19 full paper.
  - **Product Impact:** Assets generated by `SystemER` have been delivered to IBM Watson Health.
- **Pattern induction with very few human input**
  - **Product Impact**: delivered to IBM Watson Discovery in Q3 2020

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

Nanyang Technological University, Singapore 2010 - 2011
Project Officer

· 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

2021

1. **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**) IUI'21: Annual Conference on Intelligent User Interfaces (demo track).

2020

2. **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%

3. 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.

4. 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

5. **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.

6. 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.

7. 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.

8. 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.

9. **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

10. 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.

11. **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 .

12. **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.

13. 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.

14. 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%.

2018

15. 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**".

16. 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.

17. **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

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

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

20. 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.

21. 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

22. 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

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

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	IUI 2021 (demo), NAACL 2021, AAAI 2021 ACL 2020, IJCAI 2020, ICDE 2020 (industry), AAAI 2020 IEEE BigData 2019, WebDB 2018
External Reviewers	CIKM 2018, CIKM 2017, KDD 2017, AAAI 2017, ADAMA 2017
Membership	AAAI

AWARDS

- Best Demo Award @ The 19th International Semantic Web Conference (ISWC 2020)
- IBM Class-A Research Accomplishment 2017
- UC Regents Fellowship 2012
- Exceptional Mater Student - Beihang University
- Japan JASSO scholarship 2008
- Exceptional Undergraduate Student - Chongqing University

PROGRAMMING SKILLS

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