

Xiang Ren

2119, Siebel Center for Computer Science, 201 N. Goodwin Ave, Urbana, IL 61801, USA

☎ (+1) 217-377-2966 | ✉ xren7@illinois.edu | 🏠 xren7.web.engr.illinois.edu

Research Interests

Data mining, natural language processing, machine learning

Education

University of Illinois, Urbana-Champaign

Urbana, IL

PH.D. IN COMPUTER SCIENCE

2012 - PRESENT

- Thesis: *Constructing Structured Networks of Factual Knowledge from Massive Text Corpora*
- Advisor: Jiawei Han
- Expected conferral of degree: June 2017

M.SC. IN COMPUTER SCIENCE

April, 2016

- Thesis: *Effective Citation Recommendation by Information Network-Based Clustering*
- **David J. Kuck Outstanding M.S. Thesis Award**

Zhejiang University

Hangzhou, Zhejiang, China

B.ENG. IN COMPUTER SCIENCE, HONOR DEGREE (CHU KOCHEN HONORS COLLEGE)

2008 - 2012

- Senior Thesis: *A warm start SVD method for fast solving TILT* (Excellence Thesis Award)
- Committee: Yi Ma, Zhihua Zhang, Zhouchen Lin

Honors & Awards

2017	ACM Scholarship for ACM's Celebration of 50 Years of the Turning Award	ACM
2017	WWW 2017 Outstanding Reviewer Award	WWW
2016	Google PhD Fellowship (Link)	Google
	- Sole winner in the category of Structure Data and Database Management in the world	
2016	KDD Rising Star	Microsoft Academic Search
	- Rank No.3 among all the young researchers in the area of Data Mining	
2016	C. W. Gear Outstanding Graduate Student Award (Link)	UIUC
	- Highest honor given to one graduate student in CS@Illinois	
2015	Yahoo-DAIS Research Excellence Award	Yahoo
2015	Grand prize, Yelp Dataset Challenge (Link)	Yelp
2015	Runner-up, Prostate Cancer DREAM challenge	
	- One of the most prestige challenges on system biology and translational medicine	
2015	C. L. and Jane W.-S. Liu Award	UIUC
	- Given to one graduate student in CS@Illinois for his/her promising research	
2015	Richard T. Cheng Endowed Fellowship	UIUC
2014	Data Mining Research Award	UIUC
	- Voted by peers in Data Mining Research Group in CS@Illinois	
2011	Microsoft Young Fellowship	Microsoft Research Asia

Professional Experience

U.S. Army Research Laboratory

Adelphi, MD

VISITING RESEARCHER

Apr. 2016 - Jun. 2016

- Worked with Dr. Clare R. Voss and Dr. Taylor Cassidy on data-driven relation extraction from domain-specific text corpora.

Discovery Team, Pinterest

SOFTWARE ENGINEERING INTERN

- Designed and implemented machine learning pipelines for feed ranking. Shipped 3 projects in 12 weeks.

San Francisco, CA

May 2016 - Aug. 2016

Internet Services Research Center (ISRC), Microsoft Research

RESEARCH INTERN

- *Mentor:* Dr. Kuansan Wang
- Designed novel models to perform comparative analysis between documents to improve Microsoft Academic Search experience (results were published in WSDM 2017).
- Shipped data-driven entity recognition and phrase mining tools to Bing Ads productions.

Redmond, WA

May 2015 - Nov. 2015

Data Mining Research Group, University of Illinois at Urbana-Champaign

RESEARCH ASSISTANT

- *Advisor:* Prof. Jiawei Han
- Led a group of graduate students to work on NS CTA I1 project: *Constructing Unified, Structured Knowledge Networks by Integrating Text, Social Media and Multi-Genre Networks*.

Urbana, IL

Aug. 2013 - May. 2015

Data Management, Exploration and Mining (DMX) Group, Microsoft Research

RESEARCH INTERN

- *Mentor:* Dr. Surajit Chaudhuri
- Worked on synonym discovery for entities in structured knowledge bases to improve Bing synonym API (results were published in WWW 2015).

Redmond, WA

Apr. 2014 - Aug. 2014

StylePuzzle Inc.

CO-FOUNDER

- StylePuzzle (Link) is a personalized outfit recommendation system. It helps users find suitable fashion products.
- Built machine learning pipelines and performed systematic user studies and quantitative evaluations.

Champaign, IL

Aug. 2013 - May 2014

Machine Learning Group, Microsoft Research Asia

RESEARCH INTERN

- Modeled user search intents using Wikipedia and Bing query logs (results were published in WSDM 2014).

Beijing, China

May 2013 - Aug. 2013

Computer Vision Group, Microsoft Research Asia

RESEARCH INTERN

- *Mentor:* Prof. Yi Ma
- Developed efficient and effective optimization algorithms for low-rank matrix approximation (results were published in CVPR 2011, ECCV 2012, and International Journal of Computer Vision).

Beijing, China

Jul. 2011 - Aug. 2012

Professional Service

PROGRAM COMMITTEE

- ACM SIGKDD Conference on Knowledge Discovery and Data Mining (**KDD**): 2015-17
- Conference on Empirical Methods in Natural Language Processing (**EMNLP**): 2015-17
- Annual Meeting of the Association for Computational Linguistics (**ACL**): 2017
- International World Wide Web Conference (**WWW**): 2017
- ACM SIGIR Conference on Research and Development in Information Retrieval (**SIGIR**): 2017
- ACM SIGMOD Conference on Management of Data (**SIGMOD**), Student Research Competition: 2016
- SIAM Conference on Data Mining (**SDM**): 2017

CONFERENCE SESSION CHAIR

- “Applications”, ACM International Conference on Information and Knowledge Management (**CIKM 2016**)
- “Crowds and Users”, ACM SIGKDD Conference on Knowledge Discovery and Data Mining (**KDD 2015**)

JOURNAL REVIEWER

IEEE Transactions on Knowledge and Data Engineering (**TKDE**)
ACM Transactions on Knowledge Discovery from Data (**TKDD**)
ACM Transactions on Intelligent Systems and Technology (**TIST**)

CONFERENCE EXTERNAL REVIEWER

WSDM (2013, 2014, 2015, 2016), WWW (2013, 2014), ICDE (2014), ECML (2013), KDD (2013, 2014), SDM (2013)

Student Mentoring

Urvashi Khandelwal (B.S., class of 2014), CRA Outstanding Undergraduate Research Award in 2015, now a PhD student at Stanford University. Collaboration work were published in RecSys 2013, WSDM 2014 and KDD 2014.

Bradley Sturt (B.S., class of 2014), Chancellor's Scholar, now a PhD student at Massachusetts Institute of Technology. Collaboration work were published in RecSys 2013 and WSDM 2014.

Ahmed El-Kishky (Ph.D. student), NSF Graduate Research Fellowship and NDSEG Graduate Research Fellowship. Collaborator of a paper in KDD 2014, and tutorials in KDD 2015, WWW 2016 and SIGMOD 2016.

Tarique A. Siddiqui (M.Sc., class of 2016), Siebel Scholar in 2015, now a PhD student in University of Illinois at Urbana Champaign. Collaborator of a paper in CIKM 2016.

Wenqi He (M.Sc., class of 2017), Siebel Scholar in 2016. Collaborator of my papers in KDD 2016 and EMNLP 2016.

Meng Qu (M.Sc., class of 2017), Collaborator of my papers in KDD 2016 and EMNLP 2016.

Teaching Experience

2016	Guest Lecturer , CS 412: An Introduction to Data Warehousing and Data Mining. Gave lectures when the Instructor Prof. Jiawei Han was out of town.	UIUC
2014	Student Project Leader , CS 512: Data Mining: Principles and Algorithms. Supervised five students on their course projects (scientific facet extraction, relation extraction).	UIUC
2013	Guest Lecturer , CS 591: Advanced Topics on Data Mining. Gave lectures on recommender systems and entity recommendation in information networks.	UIUC
2013	Student Project Leader , CS 512: Data Mining: Principles and Algorithms. Supervised three students on their course projects (automatic entity recognition and typing).	UIUC

Invited Talks

- [7] Effort-Light StructMine: Turning Massive Text Corpora into Structures, **UCLA**, *Computer Science* March 17, 2017, Los Angeles, CA, USA.
- [6] Effort-Light StructMine: Turning Massive Text Corpora into Structures, **Georgia Tech**, *School of Computer Science* March 3, 2017, Atlanta, GA, USA.
- [5] Effort-Light StructMine: Turning Massive Text Corpora into Structures, **University of Southern California**, *Computer Science* Feb. 23, 2017, Los Angeles, CA, USA
- [4] Label Noise Reduction in Entity Typing, **Google Research** August 26, 2016, Mountain View, CA, USA.
- [3] Automatic Entity Recognition and Typing, **Microsoft Bing** October 25, 2015, Bellevue, WA, USA.
- [2] ClusType: Effective Entity Recognition and Typing by Relation Phrase-Based Clustering, **Microsoft Research Asia** September 10, 2015, Beijing, China.
- [1] Synonym Discovery for Structured Entities, **Microsoft Research** August 13, 2015, Redmond, WA, USA.

Grant Proposal Writing

- [3] **NS CTA I1:** *Constructing Unified, Structured Knowledge Networks by Integrating Text, Social Media and Multi-Genre Networks.*
 - U.S. Army Research Lab W911NF-09-2-0053 (NSCTA), \$100,000
 - Worked with PIs Prof. Heng Ji and Prof. Jiawei Han on writing subtask 1: Automated construction of robust multi-genre unified knowledge networks that will effectively adapt to new environments and streaming updates.

- [2] **NSF III Small:** *Multi-Dimensional Structuring, Summarizing and Mining of Social Media Data.*
 - NSF IIS 16-18481 (08-01-2016 – 07-31-2019), \$500,000
 - Worked with PI Prof. Jiawei Han on writing Subtasks 1 (Entity Extraction and Typing by Linking with News and Knowledge Bases) and Subtask 2 (Extraction and Clustering of Events and Subjects) sections in Theme 1.

- [1] **NSF III Medium:** *StructNet: Constructing and Mining Structure-Rich Information Networks for Scientific Research.*
 - Worked with PIs Prof. Heng Ji and Prof. Jiawei Han on writing Theme 1 (Construction of StructNet from Scientific Corpora).
 - Took charge of writing Subtask 2 (automatic relation extraction and typing with distant supervision) section of Theme 1.

Conference Tutorials

- [5] Building Structured Databases of Factual Knowledge from Massive Text Data
X. Ren, M. Jiang, J. Shang, J. Han
2017 International Conference on Management of Data (SIGMOD), 2017.

- [4] Constructing Structured Information Networks from Massive Text Corpora
X. Ren, M. Jiang, J. Shang, J. Han
26th International World Wide Web Conference (WWW), 2017.

- [3] Automatic Entity Recognition and Typing in Massive Text Data
X. Ren, A. El-Kishky, H. Ji, J. Han
2016 International Conference on Management of Data (SIGMOD), 2016.

- [2] Automatic Entity Recognition and Typing in Massive Text Corpora
X. Ren, A. El-Kishky, C. Wang, J. Han
25th International World Wide Web Conference (WWW), 2016.

- [1] Automatic Entity Recognition and Typing from Massive Text Corpora: A Phrase and Network Mining Approach
X. Ren, A. El-Kishky, C. Wang, J. Han
21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2015.

Publications

RESEARCH PAPERS UNDER REVIEW

- [30] Mining News Events from Comparable News Corpus: A Multi-Attribute Proximity Network Modeling Approach
H. Kim, X. Ren, A. El-Kishky, B. Cho, T. Hanratty and J. Han
under review, 2017

- [29] Learning Distributed Representations of Networks with Multiple Views
M. Qu, J. Tang, J. Shang, X. Ren, M. Zhang, J. Han
under review, 2017

- [28] AutoPhrase: A Unified Framework for Automated Quality Phrase Mining from Massive Text Corpora
J. Shang, J. Liu, M. Jiang, X. Ren, J. Han
under review, 2017

RESEARCH PAPERS IN JOURNALS

- [27] Linearized Alternating Direction Method with Adaptive Penalty and Warm Starts for Fast Solving Transform Invariant Low-Rank Textures
X. Ren and Z. Lin
International Journal of Computer Vision (IJCV), Volume 104 Issue 1, August 2013, Pages 1-14

- [26] Texture Repairing by Unified Low Rank Optimization
X. Liang, X. Ren, Z. Zhang, Y. Ma
Journal of Computer Science and Technology (JCST), Volume 31 Issue 3, May 2016, Pages 525-546

RESEARCH PAPERS IN CONFERENCE PROCEEDINGS

- [25] Automatic Synonym Discovery with Knowledge Bases
M. Qu, X. Ren, J. Han
23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2017, to appear
- [24] MetaPAD: Meta Pattern-driven Attribute Discovery in Massive Text Corpora
M. Jiang, J. Shang, T. Cassidy, X. Ren, L. M. Kaplan, T. P. Hanratty, J. Han
23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2017, to appear
- [23] CoType: Joint Extraction of Typed Entities and Relations with Knowledge Bases
X. Ren, Z. Qiu, W. He, M. Qu, C. R. Voss, H. Ji, T. F. Abdelzaher, J. Han
26th International Conference on World Wide Web (WWW), 2017.
(Acceptance rate: 164/966 = 17%) **SIGMOD 2017 Student Research Competition, Final List.**
- [22] Comparative Document Analysis for Large Text Corpora
X. Ren, Y. Lv, K. Wang, J. Han.
10th ACM International Conference on Web Search and Data Mining (WSDM), 2017
(Acceptance rate: 80/505 = 16%)
- [21] AFET: Automatic Fine-Grained Entity Typing by Hierarchical Partial-Label Embedding
X. Ren, W. He, M. Qu, H. Ji, C. R. Voss, J. Han.
2016 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2016, Pages 1369-1378
Long paper, oral presentation. (Acceptance rate: 109/1087 = 10%)
- [20] Label Noise Reduction in Entity Typing by Heterogeneous Partial-Label Embedding
X. Ren, W. He, M. Qu, H. Ji, C. R. Voss, J. Han.
22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2016, Pages 1825-1834
(Acceptance rate: 66/1115 = 6%)
- [19] FacetGist: Collective Extraction of Document Facets in Large Technical Corpora
T. A. Siddiqui*, X. Ren*, A. Parameswaran, J. Han
25th ACM International Conference on Information and Knowledge Management (CIKM), 2016, Pages 871-880
(Acceptance rate: 160/701 = 23%) *Equal contribution.
- [18] Representing Documents via Latent Keyphrase Inference
J. Liu, X. Ren, J. Shang, T. Cassidy, C. R. Voss, J. Han
25th International Conference on World Wide Web (WWW), 2016, Pages 1057-1067
(Acceptance rate: 115/727 = 16%)
- [17] Precision Matrix Estimation in High Dimensional Gaussian Graphical Models with Faster Rates
L. Wang, X. Ren, Q. Gu
19th International Conference on Artificial Intelligence and Statistics (AISTATS), 2016, Pages 177-185
(Acceptance rate: 129/537 = 24%)
- [16] ClusType: Effective Entity Recognition and Typing by Relation Phrase-Based Clustering
X. Ren, A. El-Kishky, C. Wang, F. Tao, C. R. Voss, H. Ji, J. Han
21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2015, Pages 995-1004
(Acceptance rate: 20%)
Presented as (an essential) part of ACL 2015 keynote “Construction and Mining of Heterogeneous Information Networks from Text data”
- [15] Synonym Discovery for Structured Entities on Heterogeneous Graphs
X. Ren, and T. Cheng
24th International Conference on World Wide Web (WWW), 2015, Pages 443-453
Industry track paper. (Acceptance rate: 10/929 = 1%).
- [14] Mining Quality Phrases from Massive Text Corpora
J. Liu, J. Shang, C. Wang, X. Ren, J. Han
2015 ACM SIGMOD International Conference on Management of Data (SIGMOD), 2015, Pages 1729-1744
(Acceptance rate: 26%) **Won grand prize of Yelp Dataset Challenges in 2015.**
- [13] ClusCite: Effective Citation Recommendation by Information Network-Based Clustering
X. Ren, J. Liu, X. Yu, U. Khandelwal, Q. Gu, L. Wang, J. Han
20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2014, Pages 821-830
(Acceptance rate: 14.7%)
- [12] Heterogeneous Graph-Based Intent Learning from Queries, Web Pages and Wikipedia Concepts
X. Ren, Y. Wang, X. Yu, J. Yan, Z. Chen, J. Han
7th ACM International Conference on Web Search and Data Mining (WSDM), 2014, Pages 23-32
(Acceptance rate: 64/335 = 18%)

- [11] Personalized Entity Recommendation: A Heterogeneous Information Network Approach
X. Yu, X. Ren, Y. Sun, Q. Gu, B. Sturt, U. Khandelwal, B. Norick, J. Han
7th ACM International Conference on Web Search and Data Mining (WSDM), 2014, Pages 283-292
(Acceptance rate: 18%, **citation: 63**)
- [10] Automatic Construction and Ranking of Topical Keyphrases on Collections of Short Documents
M. Danilevsky, C. Wang, N. Desai, X. Ren, J. Guo, J. Han
2014 SIAM International Conference on Data Mining (SDM), 2014, Pages 398-406
(Acceptance rate: 15%)
- [9] Semantic Frame-Based Document Representation for Comparable Corpora
H. Kim, X. Ren, Y. Sun, C. Wang, J. Han
IEEE 13th International Conference on Data Mining (ICDM), 2013, Pages 350-359
(Acceptance rate: 94/809 = 11.6%)
- [8] HeteRec: Entity Recommendation in Heterogeneous Information Networks with Implicit User Feedback
X. Yu, X. Ren, Y. Sun, B. Sturt, U. Khandelwal, Q. Gu, B. Norick, J. Han
7th ACM Recommender Systems Conference (RecSys), 2013, Pages 347-350
(Acceptance rate: 23%)
- [7] Repairing Sparse Low Rank Texture
X. Liang, X. Ren, Z. Zhang, Y. Ma
14th European Conference on Computer Vision (ECCV), 2012, Pages 482-495
(Acceptance rate: 28%)
- [6] A Non-convex Relaxation Approach to Sparse Dictionary Learning
J. Shi*, X. Ren*, G. Dai, J. Wang, Z. Zhang
2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011, Pages 1809-1816
Oral presentation. (Acceptance rate: **3.7%**) *Equal contribution.

RESEARCH PAPERS IN WORKSHOP PROCEEDINGS

- [5] Entity Typing: A Critical Step for Mining Structures from Massive Unstructured Text
X. Ren, W. He, A. El-Kishky, C. R. Voss, H. Ji, M. Qu, J. Han
12th International Workshop on Mining and Learning with Graphs (MLG), 2016. (**Invited Paper**)
- [4] Collaborative Filtering with Entity Similarity Regularization in Heterogeneous Information Networks
X. Yu, X. Ren, Q. Gu, Y. Sun, J. Han
2th IJCAI Workshop on Heterogeneous Information Network Analysis (IJCAI-HINA), 2013. (**Invited Paper**)

DEMONSTRATION PAPERS IN CONFERENCE PROCEEDINGS

- [3] LifeNet: A Structured Network-Based Knowledge Exploration and Analytics System for Life Sciences
X. Ren, J. Shen, M. Qu, X. Wang, Z. Wu, Q. Zhu, M. Jiang, F. Tao, S. Sinha, D. Liem, P. Ping, R. Weinshilboum, J. Han
2017 Annual Meeting of the Association for Computational Linguistics (ACL), 2016
- [2] Cross-media Event Extraction and Recommendation (**system demo**)
D. Lu, C. R. Voss, F. Tao, X. Ren, R. Guan, R. Korolov, T. Zhang, D. Wang, H. Li, T. Cassidy, H. Ji, S. Chang, J. Han, W. Wallace, J. Hendler, M. Si, L. Kaplan
2016 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2016
- [1] NewsNetExplorer: Automatic Construction and Exploration of News Information Networks
F. Tao, J. Han, H. Ji, G. Brova, C. Wang, A. El-Kishky, J. Liu, X. Ren, Y. Sun
2014 International Conference on Management of Data (SIGMOD), 2014