

Shutian Ma

Email: mashutian0608@hotmail.com

Location: Sunnyvale, California

Brief Bio

- Doctoral Degree in Nanjing University of Science & Technology
- Conduct research using NLP, machine learning and deep learning models

Areas of Research

- Document clustering and classification
- Text embedding and node embedding
- Scientific software recommendation, Citation recommendation (Dissertation Topic)

Education

Ph.D. in Information Science

- Nanjing University of Science & Technology (2014-2021)
- Department of Information Science
- Supervisor: [Chengzhi Zhang](#)

Visiting scholar, Indiana university Bloomington, United States (10/2017-10/2018)

- Cyberinfrastructure for Network Science (CNS) Center, Department of Information and Library Science, School of Informatics, Computing, and Engineering
- Supervisor: [Katy Boner](#)
- Supervisor: [Xiaozhong Liu](#)

B.M. in Information Management and System

- Nanjing University of Science & Technology (2010-2014)
- Information Management and Information System, Department of Information Science

Exchange student in Economic Management

- Ajou University, South Korea (Spring 2012)
- Information Management and Information System, Department of Information Science

Experience

Visual Artist, FizzDragon, part-time, remote (07/2024-Now)

- Film visual optimization research and development using Midjourney

Academic Tutor, Woen, part-time, remote (05/2024-Now)

- Provide expert guidance and support to students in the field of Data science

NLP Engineer, Wuxi Lead Intelligent Equipment, Wuxi (10/2023-12/2023, Moving to US in 01/2023)

- AI Algorithm Group in Lithium Battery Business
- Question Answering System & Knowledge Graph

NLP Engineer, Tencent, Shenzhen (04/2021-08/2023)

NLP Engineer Intern, Tencent, Shenzhen (07/2020-04/2021)

- Customer Research & User Experience Design Center
- User Research & Text Mining

NLP Engineer Intern, Aegis, Nanjing (03/2020-04/2020)

- Data Preprocessing & Named Entity Recognition

NLP Engineer Intern, ByteDance, Beijing (11/2018-01/2019)

- Department of Data and NLP
- Data Preprocessing & Query Classification

Sub reviewer in Conference/Journal: WSDM 2023, SIGIR 2022, Scientometrics2022, PLOS One 2021, PACIS 2021, DLP-KDD 2020, IEEE BigData 2020, SDP workshop at EMNLP 2020, JIST 2019, PACLIC 33, BIRNDL-2017,

PC member in Conference/Journal: JCDL2023, JCDL2022, JCDL2021, JCDL 2020, JDIS 2020

Publications (citations 451)

1. **Ma, Shutian**, Heng Zhang, Chengzhi Zhang, and Xiaozhong Liu. Chronological citation recommendation with time preference. *Scientometrics*. 2021, 126(4): 2991-3010.
2. **Ma, Shutian**, Chengzhi Zhang*, and Xiaozhong Liu. A review of citation recommendation: from textual content to enriched context [J]. *Scientometrics*. 2020, 122(3): 1445-1472.
3. **Shutian Ma**, Jin Xu, Chengzhi Zhang*. Automatic identification of cited text spans: a multi-classifier approach over imbalanced dataset [J]. *Scientometrics*, 2018, 116(2): 1303-1330.
4. **Shutian Ma**, Yingyi Zhang, Chengzhi Zhang*. Using multiple Web resources and inference rules to classify Chinese word semantic relation [J]. *Information Discovery and Delivery*, 2018, 46(2): 120-126.
5. Börner, Katy, Olga Scrivner, Mike Gallant, **Shutian Ma**, Xiaozhong Liu, Keith Chewning, Lingfei Wu, and James A. Evans. Skill discrepancies between research, education, and jobs reveal the critical need to supply soft skills for the data economy [J]. *Proceedings of the National Academy of Sciences (PNAS)*, 2018, 115(50): 12630-12637.
6. **Ma, Shutian**, Heng Zhang, Tianxiang Xu, Jin Xu, Shaohu Hu, and Chengzhi Zhang. IR&TM-NJUST@ CLSciSumm-19 [C]. In: *Proceedings of the 4th Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL 2019)*, Paris, France, 2019:181-195.
7. **Shutian Ma**, Heng Zhang, Jin Xu, Chengzhi Zhang*. NJUST @ CLSciSumm-18 [C]. In: *Proceedings of the 3rd Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL 2018)*, Ann Arbor, MI, USA, 2018: 114-129
8. **Shutian Ma**, Jin Xu, Jie Wang and Chengzhi Zhang*. NJUST @ CLSciSumm-17 [C]. In: *Proceedings of the 2nd Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL 2017)*, Tokyo, Japan, 2017: 16-25.
9. **Shutian Ma**, Chengzhi Zhang*. Document Representation and Clustering Models for Bilingual Documents Clustering [C]. In: *Proceedings of 2017 Annual Meeting of the Association for Information Science and Technology (ASIST'2017)*, Washington, DC, USA, 2017: 499-502.
10. **Shutian Ma**, Chengzhi Zhang*. Using Full-text to Evaluate Impact of Different Software Groups [C]. In: *Proceedings of the 16th International Conference on Scientometrics and Informetrics (ISSI 2017)*, Wuhan, China, 2017: 1666-1667.
11. **Shutian Ma**, Chengzhi Zhang*. Documents Representation for Comparable Corpora Clustering: A Preliminary Study [C]. In: *Proceedings of iConference2017*, Wuhan, China, 2017: 876-880.
12. **Shutian Ma**, Chengzhi Zhang*, Daqing He. Document Representation Methods for Clustering Bilingual Documents [C]. In: *Processing of the 2016 Annual Meeting of the Association for Information Science and Technology (ASIST'2016)*, Copenhagen, Denmark, 2016: 1-10
13. **Shutian Ma**, Chengzhi Zhang*. Automatic Collection of the Parallel Corpus with Little Prior Knowledge [C]. In: *Proceedings of the 13th China National Conference on Computational Linguistics (CCL2014)*, Wuhan, China, 2014: 95-106.
14. **Shutian Ma**, Xiaoyong Zhang, Chengzhi Zhang*. NLPCC 2016 Shared Task: Chinese Words Similarity Measure via Ensemble Learning based on Multiple Resources [C]. In: *Proceedings of the Fifth Conference on Natural Language Processing and Chinese Computing & The Twenty Fourth International Conference on Computer Processing of Oriental Languages (NLPCC-ICCPOL 2016)*. Kunming, China, 2016: 862-869.

Skills and Abilities

- Software and Tools: Gensim, Scikit-Learn, Tensorflow, Pytorch, Gephi, Neo4j, Navicat
- Programming Languages: Java, Python, Linux command Line, SQL
- Languages: Chinese and English