

# Yingcheng Sun

University of North Carolina at Greensboro  
Department of Computer Science  
Greensboro, NC 27412

Phone: (216) 302-9575  
Email: y\_sun4@uncg.edu  
Homepage: [yingchengsun.github.io/academic](https://yingchengsun.github.io/academic)

## Research Interest

Information Retrieval, Natural Language Processing and Machine Learning, with their applications in clinical informatics that help solving important healthcare problems

## Academic Experience

**Assistant Professor, University of North Carolina at Greensboro** Aug. 2022 - Present  
Department of Computer Science

**Postdoctoral Research Scientist, Columbia University** Jan. 2020 - Jun. 2022  
Department of Biomedical Informatics  
Mentor: prof. Chunhua Weng

## Education

**Ph.D. in Computer Science, Case Western Reserve University** Dec. 2019  
Advisor: prof. Kenneth Loparo

**M.E. in Software Engineering, Xidian University, China** Apr. 2013  
Advisor: prof. Qingshan Li

**B.E. in Software Engineering, Xidian University, China** Jul. 2010

## Teaching Experience

### University of North Carolina at Greensboro

CSC 454/654 Algorithm Analysis and Design	Fall 2023
CSC 490 Senior Capstone	Fall 2023
CSC 495/693 Natural Language Processing with Deep Learning	Spring 2023
CSC 410/610 Big Data and Machine Learning	Spring 2023
CSC 410 Big Data and Machine Learning	Fall 2022

## Publications

### Journal Papers

\* Note that **JAMIA** and **JBIM** are leading health informatics journals with impact factor 10.2 and 7.942

1. Kang, T., **Sun, Y.\***, Kim, J.H., Ta, C., Perotte, A., Schiffer, K., Wu, M., Zhao, Y., Moustafa-Fahmy, N., Peng, Y. and Weng, C., 2023. EvidenceMap: a three-level knowledge representation for medical evidence computation and comprehension. *Journal of the American Medical Informatics Association (JAMIA)*, p.ocado36.(\*co-first author)
2. **Sun, Y.**, Butler, A., Diallo, I., Kim, J.H., Ta, C., Rogers, J.R., Liu, H. and Weng, C., A Framework for Systematic Assessment of Clinical Trial Population Representativeness Using Electronic Health Records Data. *Applied Clinical Informatics*, 12(04), pp.816-825, Sep. 2021.
3. **Sun, Y.**, Butler, A., Lin, F., Liu, H., Stewart, L.A., Kim, J.H., Idnay, B.R.S., Ge, Q., Wei, X., Liu, C. and Yuan, C., The COVID-19 Trial Finder. *Journal of the American Medical Informatics Association (JAMIA)*, 28(3), pp.616-621, Mar. 2021.
4. **Sun, Y.**, Butler, A., Stewart, L.A., Liu, H., Yuan, C., Southard, C.T., Kim, J.H. and Weng, C., Building an OMOP common data model-compliant annotated corpus for COVID-19 clinical trials. *Journal of biomedical informatics (JBI)*, 118, p.103790, Jun. 2021.
5. **Sun, Y.**, Guo, F., Kaffashi, F., Jacono, F.J., DeGeorgia, M. and Loparo, K.A., INSMA: An integrated system for multimodal data acquisition and analysis in the intensive care unit. *Journal of biomedical informatics (JBI)*, 106, p.103434, Jun. 2020.
6. **Sun, Y.**, Kolacinski, R. and Loparo, K., Transitive Topic Modeling with Conversational Structure Context: Discovering Topics that are Most Popular in Online Discussions. *International Journal of Semantic Computing*, 14(02), pp.273-293, 2020.
7. Fang, Y., Idnay, B., **Sun, Y.**, Liu, H., Chen, Z., Marder, K., Xu, H., Schnall, R., and Weng, C., Combining Human and Machine Intelligence for Clinical Trial Eligibility Querying, *Journal of the American Medical Informatics Association (JAMIA)*, 2022.
8. Liu, H., Chi, Y., Butler, A., **Sun, Y.** and Weng, C., A knowledge base of clinical trial eligibility criteria. *Journal of Biomedical Informatics (JBI)*, 117, p.103771, Apr. 2021.
9. Lee, J., Liu, C., Kim J, Chen, Z., **Sun, Y.**, Rogers, J. and Weng, C., Deep Learning for Rare Disease: A Scoping Review. *Journal of Biomedical Informatics (JBI)*, 2022 Oct 17:104227.
10. Kim, J.H., Butler, A.M., Ta, C.N., **Sun, Y.**, Maurer, M.S. and Weng, C. The potential role of EHR data in optimizing eligibility criteria definition for cardiovascular outcome trials. *International Journal of Medical Informatics*, 156, p.104587, Dec. 2021.
11. Kury, F., Butler, A., Yuan, C., Fu, L.H., **Sun, Y.**, Liu, H., Sim, I., Carini, S. and Weng, C., Chia, a large annotated corpus of clinical trial eligibility criteria. *Scientific data*, 7(1), pp.1-11, Aug. 2020.
12. Li, Q., **Sun, Y.** and Xue, B., Complex query recognition based on dynamic learning mechanism. *Journal of Computational Information Systems*, 8(20), pp.8333-8340, 2012.
13. Li, Q., Zou, Y. and **Sun, Y.**, User Personalization Mechanism in Agent-based Meta Search Engine. *Journal of Computational Information Systems*, 8(20), pp. 8325-8332, 2012.

## Conference Papers

1. **Sun, Y.**, Tang, J., Butler, A., Liu, C., Fang, Y. and Weng, C., Interactive Similarity-Based Search of Clinical Trials. *World Congress on Health and Biomedical Informatics (MEDINFO' 21)*, October 2-4, 2021
2. Fang, Y., Kim, J.H., Idnay, B.R., Garcia, R.A., Castillo, C.E., **Sun, Y.**, Liu, H., Liu, C., Yuan, C. and Weng, C., Participatory Design of a Clinical Trial Eligibility Criteria Simplification Method. In *31st Medical Informatics Europe Conference (MIE'21)*, May 29-31, 2021.
3. Li, X., Liu, H., Kury, F., Yuan, C., Butler, A., **Sun, Y.**, Ostropolets, A., Xu, H. and Weng, C., A Comparison between Human and NLP-based Annotation of Clinical Trial Eligibility Criteria Text Using The OMOP Common Data Model. In *AMIA Annual Symposium Proceedings (AMIA'21)*, November 14 - 18, 2021.
4. **Sun, Y.**, Kolacinski, R. and Loparo, K., Eliminating search intent bias in learning to rank. In *2020 IEEE 14th International Conference on Semantic Computing (ICSC'20)*, IEEE, February 3-5, 2020
5. **Sun, Y.**, Loparo, K. and Kolacinski, R., Conversational structure aware and context sensitive topic model for online discussions. In *2020 IEEE 14th International Conference on Semantic Computing (ICSC'20)*, IEEE, , February 3-5, 2020. (Best paper)
6. **Sun, Y.** and Loparo, K., Opinion Spam Detection Based on Heterogeneous Information Network. In *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI'19)*, IEEE, Nov 4-6, 2019. (Acceptance Ratio =  $160/617 = 26\%$ )
7. **Sun, Y.**, Cai, X. and Loparo, K., Learning - based Adaptation Framework for Elastic Software Systems. In *2019 IEEE 31st International Conference on Software Engineering and Knowledge Engineering (SEKE'19)*, IEEE, Jul 10-12, 2019.
8. **Sun, Y.** and Loparo, K., Topic Shift Detection in Online Discussions using Structural Context. In *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC'19)*, IEEE, Jul 15-19, 2019.
9. **Sun, Y.** and Loparo, K., Information Extraction from Free Text in Clinical Trials with Knowledge-Based Distant Supervision. In *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC'19)*, IEEE, Jul 15-19, 2019.
10. **Sun, Y.** and Loparo, K., A Clicked-URL Feature for Transactional Query Identification. In *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC'19)*, IEEE, Jul 15-19.
11. **Sun, Y.** and Loparo, K., Context Aware Image Annotation in Active Learning with Batch Mode. In *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC'19)*, IEEE, Jul. 2019.
12. **Sun, Y.** and Loparo, K., A Common Gene Expression Signature Analysis Method for Multiple Types of Cancer. In *2019 19th Industrial Conference on Data Mining*, Springer, Jul 17-21, 2019.
13. **Sun, Y.** and Loparo, K., 2019, July. Knowledge-guided Text Structuring in Clinical Trials. In *2019 19th Industrial Conference on Data Mining*, Springer, Jul 17-21, 2019.
14. **Sun, Y.** and Loparo, K., Context Aware Image Annotation in Active Learning. In *2019 19th Industrial Conference on Data Mining*, Springer, Jul 17-21, 2019.
15. **Sun, Y.** and Li, Q., The research situation and prospect analysis of meta-search engines. In *2012 2nd International Conference on Uncertainty Reasoning and Knowledge Engineering*, IEEE, Aug. 2012.
16. Li, Q. and **Sun, Y.**, An agent based intelligent meta search engine. In *2012 4th International Conference on Web Information Systems and Mining (WISM)*, Springer, Oct. 2012.

17. Li, Q., Zou, Y. and **Sun, Y.**, Ontology based user personalization mechanism in meta search engine. In *2012 2nd International Conference on Uncertainty Reasoning and Knowledge Engineering*, IEEE, Aug. 2012.

### *Papers in Press or Submission*

1. Jared M., Regis, K. and **Sun, Y.**, ROSA: Mixed-methods formative evaluation of Rural Online Shopping Assistant. *Journal of Nutrition Education and Behavior*, (in press).
2. Sony A., and **Sun, Y.**, MVDB: A Medical Vector Database for Semantic Concept Retrieval and Normalization *Journal of the American Medical Informatics Association (JAMIA)*, (in submission).
3. Sony A., and **Sun, Y.**, An Interactive Interface for Medical Multi-Document Evidence Synthesis and Visualization, *Journal of the American Medical Informatics Association (JAMIA)*, (in submission).
4. **Sun, Y.**, Tian, K., Perotte, A., Kim, J., Liu, H., Turfah, A. and Weng, C. EvidenceBase: A Dynamic Knowledge Base for Evidence from Randomized Clinical Trials (JAMIA), (in submission).

### *Patents*

1. Li, Q., **Sun, Y.**, Zou, Y., Xu, J. and Xi, B., 2012, Agent Based Intelligent Meta Search Engines System, CN Patent No. 102902800 A.
2. Li, Q., Zou, Y., **Sun, Y.**, Xu, J. and Xi, B., 2013. A Personalized Method of Agent Based Meta Search Engines, CN Patent No.103593413 A.

## Grants and Contracts

### **As Principal Investigator**

1. NIH/NSF Smart Health and Biomedical Research in the Era of Artificial Intelligence and Advanced Data Science (SCH) Grant : \$1,100,000 (in submission)
2. NIH/FDA Limited Competition: Clinical and Translational Science Award (CTSA) Program: Collaborative and Innovative Acceleration Award (UG3/UH3 Clinical Trial Optional) (in submission)
3. UNCG Research Award: Structured Clinical Evidence Extraction and Synthesis \$5,000, 2023-2024
4. Dorothy Levis Munroe Student Research Grant : \$1,000, 2023-2024
5. Google Cloud Credit Grant: A Dynamic Knowledge Base for Evidence from Randomized Clinical Trials: \$5,000, 2022-2023
6. Google Cloud Credit Grant: Interactive Search Models for Clinical Trials: \$5,000, 2021-2022

### **As Co-Principal Investigator**

1. NIH R34: "Friend for Health: A Contextualized and Artificial Intelligence-based Conversational Virtual Reality Avatar Tool to Promote Cardiometabolic Health among Adolescents Vulnerable to Health Disparities" : \$450,000 (submitted)
2. NIH R34: "Rural Online Shopping Assistant (ROSA)": \$500,000 (submitted)

### **Involved in Proposal Writing**

1. NIH R01: "RESCUE: Rare Disease Detection and Escalation Support via a Learning Health System", 2021 (Funded: 5,000,000)

## Press Coverage

"COVID-19 trial finder provides simplified search process for COVID-related clinical trials" (*Eurekalert*, 05, 2020; *Medical Press*, 05, 2020; *Columbia DBMI*, 05, 2020)

## Presentations

### Conference and workshop keynotes

Keynote speaker in annual Health Informatics Conference (JIS Go Live 2021), "Towards Improved Information Retrieval for Evidence-based Medicine", Nov 15, 2021.

Keynote speaker in AI-Enhanced Clinical Trials Summit 2022, "Augmented Intelligence for Evidence-based Medicine: Optimizing Clinical Trial Eligibility Criteria Design", Boston, Dec. 6, 2022.

### Invited talks and university presentations

1. Invited talk at seminar series in the Department of Biomedical Informatics at University of California San Diego, "Improving Clinical Information Retrieval for Medical-based Evidence", Nov 13, 2021.
2. Invited talk at seminar series in the Department of Computer Science at Xidian University, "Clinical Text Mining and Retrieval for Evidence-based Medicine", Sep 8, 2021.

## Professional Activities

### Scientific Program Committee

International Conference on Software Engineering Advances (ICSEA'20, ICSEA'21)

International Conference on Big Data, Small Data, Linked Data and Open Data (ALLDATA'21, ALLDATA'22)

### Reviewer for journals

Journal of Biomedical Informatics (JBI)

Journal of the American Medical Informatics Association Open (JAMIA Open)

Applied Clinical Informatics (ACI)

International Journal of Medical Informatics (IJMI)

### Reviewer for conferences

American Medical Informatics Association Annual Symposium (AMIA'20, AMIA'21, AMIA'22)

The IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC'20)