
BIOGRAPHICAL SKETCH

NAME: **Licong Cui**

POSITION TITLE: Associate Professor, McWilliams School of Biomedical Informatics, The University of Texas Health Science Center at Houston (UTHealth Houston)

EDUCATION/TRAINING

INSTITUTION AND LOCATION	DEGREE (if applicable)	Completion Date MM/YYYY	FIELD OF STUDY
Shaanxi Normal University, Xi'an, China	BS	07/2005	Information and Computing Science
Shaanxi Normal University, Xi'an, China	MS	07/2008	Pure Mathematics
Case Western Reserve University, Cleveland, OH	PhD	08/2014	Computer Science

A. Personal Statement

My research background in biomedical informatics includes ontologies and controlled vocabularies, knowledge representation and reasoning, large-scale clinical data integration and management, big data analytics, text mining, and neuroinformatics. As a well-trained Computer Scientist specialized in Biomedical Informatics, I have strong grounding in algorithms and computational methodologies. For instance, I have designed and developed scalable algorithms for analyzing biomedical data, including mining graphs for quality assurance of biomedical ontologies such as SNOMED CT, NCI Thesaurus and Gene Ontology, performing sleep stage classification and postictal generalized EEG suppression detection based on electrophysiological signals, and extracting epilepsy phenotypes from clinical narratives for patient cohort identification. I also have considerable experience in Applied Informatics, particularly the lifecycle of software development for integrating, exploring and analyzing big data in the National Sleep Research Resource (R24HL114473) and the Center for SUDEP Research (U01NS090408). I have served as a PI for two R01 and two R21 grants from NIH (R01NS116287; R01LM013335; R21AG068994; R21CA231904), and three grants from NSF (IIS1657306; IIS1931134; IIS2047001).

Ongoing and recently completed projects that I would like to highlight include:

R01 NS116287

Cui (PI)

05/15/20 – 04/30/25

“An informatics framework for SUDEP Risk Marker Identification and Risk Assessment”

The main goal of this project is to develop a SUDEP Risk Marker Extraction system for automated extraction of known and putative SUDEP risk markers from the multimodal patient data collected by the Center for SUDEP Research from Epilepsy Monitoring Units in multiple medical centers.

NSF IIS2047001

Cui (PI)

09/01/21 – 08/31/26

“CAREER: Advancing the Role of Ontologies for Data Science in Biomedicine”

The main goal of this project is to develop new methods for ontology matching and for ontology quality enhancement that directly impact data science practice in biomedicine, such as patient cohort identification.

R01 LM013335

Cui (PI)

08/01/20 – 07/31/23

NIH/NLM

“Biomedical Terminology Quality Assurance for Enhancing Clinical Queries over Electronic Health Records”

The main goal of this project is to develop a general automatic change-suggestion framework to systematically address quality issues in biomedical terminologies and to quantitatively assess the terminology quality impact on clinical queries over EHRs for patient cohort identification.

R21 AG068994

Cui (PI)

09/15/20 – 05/31/23

“An Interface Ontology for Alzheimer’s Disease Research”

The main goal of this project is to develop a novel interface ontology for Alzheimer's disease (AD) research and web-based data exploration tools for managing, querying, and exploring AD data resources.

Citations:

- a. **Cui L**, Zhu W, Tao S, Case JT, Bodenreider O, Zhang GQ. Mining Non-Lattice Subgraphs for Detecting Missing Hierarchical Relations and Concepts in SNOMED CT. *Journal of the American Medical Informatics Association* 24(4): 788-798, 2017. PMID: PMC6080685
- b. Zhang GQ, **Cui L**, Mueller R, Tao S, Kim M, Rueschman M, Mariani S, Mobley D, Redline S. The National Sleep Research Resource: towards a sleep data commons. *Journal of the American Medical Informatics Association* 25(10):1351-8, 2018. PMID: PMC6188513
- c. Abeysinghe R, Hinderer EW, Moseley HNB, **Cui L**. SSIF: Subsumption-based Sub-term Inference Framework to Audit Gene Ontology. *Bioinformatics* 36(10):3207-3214, 2020. PMID: PMC7214018
- d. Abeysinghe R, Zheng F, Bernstam EV, Shi J, Bodenreider O, **Cui L**. A deep learning approach to identify missing is-a relations in SNOMED CT. *Journal of the American Medical Informatics Association* 30(3):475-484, 2023. PMID: PMC9933066

B. Positions, Scientific Appointments, and Honors

Positions and Scientific Appointments

2023 – Present	Editorial Board Member, Journal of Biomedical Informatics
2023	Ad hoc reviewer, NIH Clinical Data Management and Analysis (CDMA) study section
2022 – Present	Associate Professor (Tenured), School of Biomedical Informatics, UTHealth Houston
2022	Ad hoc reviewer, NIH & NSF
2021 – Present	Section Editor, International Medical Informatics Association (IMIA) Yearbook of Medical Informatics - Knowledge Representation and Management Section
2020 – Present	Member, Center for Biomedical Semantics & Data Intelligence, UTHealth Houston
2019 – 2022	Assistant Professor (Tenure Track), School of Biomedical Informatics, UTHealth Houston
2019 – Present	Assistant Director, Texas Institute for Restorative Neurotechnologies, UTHealth Houston
2019 – 2020	Guest Editor, BMC Medical Informatics and Decision Making, Special Issue on Quality Assurance and Enrichment of Biological and Biomedical Ontologies and Terminologies
2016 – 2018	Assistant Professor (Tenure Track), Department of Computer Science, Institute for Biomedical Informatics, University of Kentucky, Lexington, KY
2017	Ad hoc reviewer, NSF
2017 – 2018	Guest Editor, Journal of Biomedical Informatics, Special Issue on Quality Assurance of Biomedical Terminologies and Ontologies
2015 – 2016	Assistant Professor (Research Track), Division of Biomedical Informatics, University of Kentucky, Lexington, KY
2014 – 2015	Research Assistant Professor, Department of Electrical Engineering and Computer Science, Case Western Reserve University, Cleveland, OH
2011 – Present	Member, American Medical Informatics Association (AMIA)

Honors

2022	AMIA New Investigator Award
2021	Distinguished Paper Award, AMIA 2021 Annual Symposium
2021	Dean's Excellence Award for Research, UTHealth School of Biomedical Informatics, 2020-2021

2021	Best Paper in Knowledge Representation and Management section, IMIA Yearbook of Medical Informatics 2021
2021	NSF CAREER Award
2020	Best Article, Best of Cell Metabolism 2020
2019	Best Student Paper Finalist (my PhD student), AMIA 2019 Annual Symposium
2018	Best Student Paper Finalist (my PhD student), AMIA 2018 Annual Symposium
2017	Outstanding Reviewer for Journal of Biomedical Informatics
2015	Best Article in Clinical Research Informatics, IMIA Yearbook of Medical Informatics

C. Contributions to Science

- 1. Quality Assurance of Biomedical Ontologies.** Biomedical ontologies serve as a key knowledge source in many biomedical applications including information extraction, information retrieval, data integration, data management and decision support. However, the large size and complexity of biomedical ontologies make quality assurance a time-consuming and difficult task. I have contributed to the design and development of a set of efficient algorithms for auditing biomedical ontologies with dramatically reduced time. These algorithms make it feasible not only to perform exhaustive structural analysis of large ontological hierarchies, but also to systematically mining lexical features of concepts to detect potential quality issues in biomedical ontologies, such as SNOMED CT, Gene Ontology, and National Cancer Institute Thesaurus.

 - Cui L**, Bodenreider O, Shi J, Zhang GQ. Auditing SNOMED CT hierarchical relations based on lexical features of concepts in non-lattice subgraphs. *Journal of biomedical informatics* 78:177-84, 2018. PMID: PMC5835197
 - Cui L**, Abeysinghe R, Zheng F, Tao S, Zeng N, Hands I, Durbin EB, Whiteman L, Remennik L, Sioutos N, Zhang GQ. Enhancing the Quality of Hierarchic Relations in the National Cancer Institute Thesaurus to Enable Faceted Query of Cancer Registry Data. *JCO Clinical Cancer Informatics* 4:392-8, 2020. PMID: PMC7265791
 - Zheng F, Shi J, Yang Y, Zheng WJ, **Cui L**. A transformation-based method for auditing the IS-A hierarchy of biomedical terminologies in the Unified Medical Language System. *Journal of the American Medical Informatics Association* 27(10):1568-75, 2020. PMID: PMC7566369
 - Abeysinghe R, Yang Y, Bartels M, Zheng WJ, **Cui L**. An evidence-based lexical pattern approach for quality assurance of Gene Ontology relations. *Briefings in Bioinformatics*;23(3):bbac12, 2022. PMID: PMC9116247
- 2. Large-scale Multimodal Data Integration and Analytics for Neurological Disorders.** Cross-institutional data sharing is crucial for developing and implementing large-scale clinical studies. Integration of patients across multiple institutions is required both for rare disease studies and other studies that need very large and diverse populations. I developed an ontology-based framework for integrating, querying, and analyzing multimodal patient data from multiple sources. This framework is adaptable and robust, and has been successfully deployed for two national research resource sharing projects related to neurological disorders: National Sleep Research Resource (NSRR) and Multi-Modality Epilepsy Data Capture and Integration System (MEDCIS).

 - Zhang GQ, **Cui L**, Lhatoo S, Schuele SU, Sahoo SS. MEDCIS: multi-modality epilepsy data capture and integration system. *In AMIA Annual Symposium Proceedings* 2014:1248-1257, 2014. PMID: PMC4420009
 - Cui L**, Zeng N, Kim M, Mueller R, Hankosky ER, Redline S, Zhang GQ. X-search: an open access interface for cross-cohort exploration of the National Sleep Research Resource. *BMC medical informatics and decision making* 18(1):99, 2018. PMID: PMC6234631
 - Tao S, **Cui L**, Chou WC, Lhatoo S, Zhang GQ. DaT3M: a data tracker for multi-faceted management of multi-site clinical research data submission, curation, master inventorying, and sharing. *In AMIA Annual Symposium Proceedings* 2022:466-475, 2022. PMID: PMC9285149
 - Li X, Huang Y, Lhatoo SD, Tao S, Bertran LV, Zhang GQ, **Cui L**. A hybrid unsupervised and supervised learning approach for postictal generalized EEG suppression detection. *Frontiers in Neuroinformatics* 16:1040084, 2022. PMID: PMC9806125

3. **Information Extraction from Biomedical Text.** The increased adoption of Electronic Health Records (EHRs) has generated an overwhelming amount of electronic information in unstructured or semi-structured form in healthcare, such as textual clinical narratives. They contain an extraordinary amount of underutilized biomedical knowledge. In order to take advantage of such knowledge to facilitate second use of EHRs for patient cohort discovery, I developed effective ontology-guided methods for automatic extraction of structured information from patient discharge summaries.
 - a. **Cui L**, Bozorgi A, Lhatoo SD, Zhang GQ, Sahoo SS. EpiDEA: Extracting structured epilepsy and seizure information from patient discharge summaries for cohort identification. *In AMIA Annual Symp Proc* 2012:1191-1200, 2012. PMID: PMC3540531
 - b. **Cui L**, Sahoo SS, Lhatoo SD, Garg G, Rai P, Bozorgi A, Zhang GQ. Complex epilepsy phenotype extraction from narrative clinical discharge summaries. *Journal of Biomedical Informatics* 51:272-279, 2014. PMID: PMC4464795
 - c. Sahoo SS, Lhatoo SD, Gupta DK, **Cui L**, Zhao M, Jayapadian C, Bozorgi A, Zhang GQ. Epilepsy and seizure ontology: towards an epilepsy informatics infrastructure for clinical research and patient care. *Journal of the American Medical Informatics Association* 21:82-89, 2014. PMID: PMC3912711
 - d. Tao S, Abeysinghe R, De La Esperanza BT, Lhatoo S, Zhang GQ, **Cui L**. Extracting Temporal Expressions of First Seizure Onset from Epilepsy Patient Discharge Summaries. *AMIA Informatics Summit*, 2023 (in press).

4. **Ontology-based Information Retrieval Interface and Crowdsourcing User Interface Evaluation.** The secondary use of healthcare and clinical data holds the key for decision support, hypothesis generation, acceleration of clinical trials and discovery of new knowledge. A bottleneck to the effective use of clinical data is the ineffective information retrieval interface. I have developed an ontology-driven query interface framework called VISAGE+ (Visual Aggregator and Explorer Plus), which has been used for exploring complex clinical data and Medicare data. On the other hand, the internet has become one of the most important sources for consumers to seek health-related information. A major factor contributing to the failure of health information retrieval is information overload (too much information was retrieved). I developed a Conjunctive Exploratory Navigation Interface (CENI) to facilitate effective health information retrieval by consumers. Anonymous testers recruited through crowdsourcing preferred CENI nearly two-to-one over a keyword-based lookup interface and the most commonly available lookup search interface using Google.
 - a. **Cui L**, Mueller R, Sahoo SS, Zhang GQ. Querying complex federated clinical data using ontological mapping and subsumption reasoning. *In IEEE International Conference on Healthcare Informatics 2013*: 351-360, 2013. DOI: 10.1109/ICHI.2013.49
 - b. **Cui L**, Tao S, Zhang GQ. A semantic-based approach for exploring consumer health questions using UMLS. *AMIA Annual Symp Proc* 2014:432-441, 2014. PMID: PMC4419919
 - c. **Cui L**, Carter RR, Zhang GQ. Evaluation of a novel conjunctive exploratory navigation interface for consumer health information: a crowdsourced comparative study. *Journal of Medical Internet Research* 16: e45, 2014. PMID: PMC3936301
 - d. Tao S, **Cui L**, Wu X, Zhang GQ. Facilitating Cohort Discovery by Enhancing Ontology Exploration, Query Management and Query Sharing for Large Clinical Data Repositories. *In AMIA Annual Symposium Proceedings* 2017:1685-1694, 2017. PMID: PMC5977665

Complete List of Published Work in MyBibliography:

<http://www.ncbi.nlm.nih.gov/myncbi/browse/collection/47518686/?sort=date&direction=descending>