

Yuhan Wei

Tel: +86 18702786858 Email: yuhan_wei@whu.edu.cn

EDUCATION

- Wuhan University**, M.S. in Data Science (Expected) Wuhan, China
■ Accumulative GPA: 3.51/4.00 Sep 2019-June 2022
■ Core Courses: Information System Engineering, Natural Language Processing, Principles of Machine Learning, Data Visualization, Mathematical Modeling
■ Honors: Second-class Scholarship for Academic Excellence (TOP 2/8, 2020); Excellent Student Leader (2020)
- Nanchang University**, B.S. in Information Management and Systems Nanchang, China
■ Accumulative GPA: 3.61/4.00; Major GPA: 3.71/4.00; Class Ranking: **1/40** Sep 2015-June 2019
■ Honors: Outstanding Student Scholarship (TOP 2/40, 2016/2017/2018); Merit Student & Excellent Student Leader (2016 & 2019)

PAPER & PATENT

- **Yuhan Wei**, Wei Lu, Qikai Cheng, Tingting Jiang, Shewei Liu. "How Humans Obtain Information from AI: Categorizing User Messages in Human-AI Collaborative Conversations", *Journal of the Association for Information Science and Technology*, Under Review
- Jinqing Yang, Shengzhi Huang, **Yuhan Wei**, Zhifeng Liu, Wei Lu. "An Improved Framework for Identifying Emerging Author Keywords", *Proceedings of the 10th Global TechMining Conference*, November 11, 2020 Virtual Event (Power Talk)
- Qian Guo, Tingting Jiang, **Yuhan Wei**, Wei Lu. "User Information Behavior in Dialogues between Users and Intelligent Customer Service", *Tenth National Academic Conference for Doctoral Students in Information Science*, Presented on December 6, 2020
- Wei Lu, Jinqing Yang, **Yuhan Wei**, Leyan Wu. "Science and Technology Information Analysis Combined with Time Lag Calculation in the Subject Evolutionary Reasoning Method", *National Invention Patent*, 201911081064.8

RESEARCH

- How Humans Obtain Information from AI: Categorizing User Messages in Human-AI Collaborative Conversations** June 2020-Mar 2021
Information Retrieval and Mining Institute of Wuhan University
Collaborated by WiscIR Lab of University of Wisconsin-Madison & Intelligent Insurance Lab of Tk.cn
Project Leader, Supervisor: Prof. Wei Lu/ Assistant Prof. Jiepu Jiang
- Established a database of 4235037 pieces of conversation data about insurance consulting, applied MySQL to process and clean the data, and used natural language processing to normalize conversations with Python, including word segmentation, part-of-speech tagging and key-phrase extraction.
 - Divided conversation into three types of user-AI, user-human, and mixed representatives, and concluded a categorization system for user messages based on the message intention; conducted data annotation experiment; summarized and reported the characteristics of different message types and compared their usage in sessions; visualized the experiment results by using Echarts.
 - Plan to study the structure of user intent in conversation and achieve the automatic prediction of user intent structure by deep-learning models, Logistic Regression and SVM have been applied already.
- Evolution of Research Topics and Identification of Emerging Keywords** Sep 2019-Dec 2020
Information Retrieval and Mining Institute of Wuhan University
Affiliated with "Literature Citation Recommendation Based on Multi-semantic Information Fusion" Project
Sponsored by National Natural Science Foundation of China, Supervisor: Prof. Wei Lu
- Proposed a dynamic novelty indicator method (DNIM) with a nested attenuation factor to measure the dynamic novelty values of different topics based on literature review.
 - Programed a web crawler with Python, obtained metadata of 157763 papers from ACM database from 1951 to 2018 which contained 219903 keywords.
 - Used Natural Language Toolkit to normalize the keywords, fitted the Top 500 keywords regarding use frequency by adopting logistic regression model, and validated the method by comparing the computation result of top 0.05%, 0.1%, 0.5%, 1%, and 1.5% keywords with benchmark model and DNIM.
- User Information Behavior in Dialogues between Users and Intelligent Customer Service** Apr 2020-Dec 2020
Center for Studies of Information Resources of Wuhan University, Supervisor: Prof. Wei Lu/Prof. Tingting Jiang
- Used Python to process and classify 251556 pieces of user messages from WeSure, an insurance agency platform

Yuhan Wei

Tel: +86 18702786858 Email: yuhan_wei@whu.edu.cn

and selected out 75918 rounds of question dialog.

- Marked off 147972 dialog and concluded top five topics, calculated the amount of user messages to decide whether the demand of users are satisfied.

COURSE PROJECTS

Model-based Global Language Trends Analysis Mar 2020-June 2020

- Obtained and computed the amount of native speakers from 2014 to 2016 and data of immigration rate, population growth rate, GDP, and cultural power for the thirteen language-speaking countries.
- Built population growth and improved-Markov models to predict population change and geographical distribution of these languages in the next five years.

Forecasting of Agricultural Hot Topics Based on Multi-source Data Analytics Sep 2019-Jan 2020

- Extracted 22657 pieces of fund project texts, 163778 patent documents, 135790 journal papers, and 15565 conference papers in agricultural area from 2008 to 2014.
- Divided the abstract texts into 32 datasets, processed word segmentation, word stemming, transformed the data into sparse vector, extracted topics through Latent Dirichlet Allocation, and visualized the relations among hot topics with Matplotlib.

Narrative Mode Analysis of CCTV News Based on Man's First View of Black Hole Report Sep 2019-Jan 2020

- Crawled 3032 pieces of CCTW news by using Python.TuShare and selected out 445 news about black hole.
- Applied Python to implement word segmentation and part-of-speech tagging, performed sentiment analysis of news text and concluded the narrative mode.

COMPETITIONS & ACTIVITIES

| | |
|--|--------------------|
| Teaching Assistant for Undergraduate Class "Classic Guide to Humanities and Social Science | Sep 2020-June 2021 |
| Volunteer at ACM/ IEEE Joint Conference on Digital Libraries 2020 | Aug 2020 |
| Participant in the 2018 Global Management Challenge (First Prize) | Jan 2019 |
| The 17 th National Enterprise Competition For MBA Training Colleges (Second Prize) | Apr 2018 |
| Honorable Mention for MCM: The Mathematical Contest in Modeling | Feb 2018 |
| Participant in the 2017 Global Management Challenge (Second Prize) | Jan 2018 |
| Participant in the 10 th National College Students Network Business Innovation and Application Competition (Outstanding Winner) | Dec 2017 |

SKILLS

- Computer Skills: Python, Java, SQL, HTML, SPSS, C++
- Language Proficiency: Native in Chinese, Fluent in English, Elementary in Spanish