Zhongyu WEI

CONTACT INFORMATION	Computer Science Department MS EC-31, 800 W. Campbell Rd. Richardson, TX 75080	*Voice:* (469) 616-8860 *E-mail:* zywei@hlt.utdallas.edu *Homepage:* http://www.hlt.utdallas.edu/~zywei

RESEARCH INTERESTS

Cross-media summarization, social media analysis, microblog search, natural language processing

CURRENT POSITION

University of Texas at Dallas (UTD), Richardson, TX, USA   Dec. 2014 - Present
Postdoctoral Associate, Human Language Technology Research Institute
Mentor: Yang Liu

EDUCATION

Chinese University of Hong Kong (CUHK), Hong Kong   Aug. 2010 - Dec. 2014
Ph.D., Systems Engineering and Engineering Management
Advisor: Kam-Fai Wong

Harbin Institute of Technology (HIT), Harbin, China
M.Eng., Computer Science and Technology   Aug. 2008 - Jul. 2010
B.Eng., Computer Science and Technology   Aug. 2004 - Jul. 2008

PROFESSIONAL EXPERIENCE

Qatar Computing Research Institute, Doha, Qatar   Jun. 2013 - Dec. 2013
Research Associate, Human Language Technology Group
Mentors: Stephan Vogel, Wei Gao

The Open University, Milton Keynes, UK   Apr. 2012 - Jun. 2012
Research Associate, Knowledge Media Institute
Mentors: Simon J. Buckingham Shum, Yulan He

PUBLICATIONS

- Conference Papers

- **Zhongyu Wei** and Wei Gao, *Gibberish, Assistant, or Master? Using Tweets Linking to News for Extractive Single-Document Summarization*, ACM SigIR 2015, accepted.

- Gaoyan Ou, Wei Chen, Dongqing Yang, **Zhongyu Wei**, Binyang Li, Tengjiao Wang and Kam-Fai Wong, *Exploiting Community Emotion for Microblog Event Detection*, EMNLP 2014.

- **Zhongyu Wei**, Wei Gao, *Utilizing Microblogs for Automatic News Highlights Extraction*, COLING 2014.

- Binyang Li, Lanjun Zhou, **Zhongyu Wei**, Kam-Fai Wong, Ruifeng Xu and Yunqing Xia, *Web Information Mining and Decision Support Platform for the Modern Service Industry*, ACL 2014, demo paper.

- Lanjun Zhou, Binyang Li, **Zhongyu Wei** and Kam-Fai Wong, *An Open Discourse Corpus for Chinese with Annotated Explicit Discourse Connectives*, LREC 2014.

- **Zhongyu Wei**, Junwen Chen, Wei Gao, Binyang Li, Lanjun Zhou, Yulan He and Kam-Fai Wong, *An Empirical Study on Uncertainty Identification in Social Media Context*, ACL 2013.

- **Zhongyu Wei**, Yulan He, Wei Gao, Lanjun Zhou, Binyang Li and Kam-Fai Wong, *Mainstream Media Behavior Analysis on Twitter: A Case Study on UK General Election*, ACM HyperText 2013.

- Rebecca Ferguson, **Zhongyu Wei**, Yulan He, Simon Buckingham Shum, *An Evaluation of Learning Analytics to Identify Exploratory Dialogue in Online Discussions*, ACM LAK 2013.

- Lanjun Zhou, Wei Gao, Binyang Li, **Zhongyu Wei**, and Kam-Fai Wong. *Cross-lingual Identification of Ambiguous Discourse Connectives for Resource-Poor Language*, COLING 2012.

- Yulan He, Hassan Saif, **Zhongyu Wei**, and Kam-Fai Wong, *Qunatising Opinions for Political Tweets Analysis*, LREC 2012.

- Binyang Li, Lanjun Zhou, Wei Gao, Kam-Fai Wong and **Zhongyu Wei**, *An Effective Approach for Topic-Specific Opinion Summarization*, AIRS 2011.

- Lanjun Zhou, Binyang Li, Wei Gao, **Zhongyu Wei** and Kam-Fai Wong, *Unsupervised Discovery of Discourse Relations for Eliminating Intra-sentence Polarity Ambiguities*, EMNLP 2011.

- Jun Xu, Qingcai Chen, Xiaolong Wang and **Zhongyu Wei**, *One-class classification models for financial industry information recommendation*, ICMLC 2010.

- Workshop Papers and Technical Reports

- Walid Magdy, Wei Gao, Tarek El-Ganainy and **Zhongyu Wei**, *QCRI at TREC 2014: Applying the KISS Principle for the TTG Task in the Microblog Track*, Microblog Track at TREC2014. (2nd place among 13 groups)

- **Zhongyu Wei**, Wei Gao, *Ranking Model Selection and Fusion for Effective Microblog Search*, SIGIR 2014 Workshop on Social Media Retrieval and Analysis(SoMeRA), Gold Coast, Australia, July, 2014.

- Tarek El-Ganainy, **Zhongyu Wei**, Walid Magdy, and Wei Gao, *QCRI at TREC 2013 Microblog Track*, Microblog Track at TREC2013. (2nd place among 65 automatic runs)

- Wei Gao, **Zhongyu Wei** and Kam-Fai Wong, *Microblog Search and Filtering with Time Sensitive Feedback and Thresholding Based on BM25*, Microblog Track at TREC2012.

- **Zhongyu Wei**, Wei Gao, Lanjun Zhou, Binyang Li, and Kam-Fai Wong, *Exploring Tweets Normalization and Query Time Sensitivity for Twitter Search*, Microblog Track at TREC2011.

- Journal Articles

- Binyang Li, Lanjun Zhou, Kam-Fai Wong, and **Zhongyu Wei**. *An Effective Information Representation for Opinion Retrieval*, Journal of the American Society for Information Science and Technology (JASIST), 2015, accepted for publication.

- **Zhongyu Wei**, Yulan He, Simon Buckingham Shum, Rebecca Ferguson, Wei Gao and Kam-Fai Wong, *A Self-Training Framework for Automatic Identification of Exploratory Dialogue*, International Journal of Computational Linguistics and Applications (IJCLA), 4(1): 111 - 126 (2013).

- Binyang Li, Kam-Fai Wong, Lanjun Zhou, **Zhongyu Wei**, and Jun Xu, *Pests Hidden in Your Fans: An Effective Approach for Opinion Leader Discovery*, In Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data, 227-237 (2013).

- **Zhongyu Wei**, Jun Xu, Xiaolong Wang. *One-class Classification based Finance News Story Recommendation*, Journal of Computational Information Systems (JCIS), 5(6): 1625-1631 (2009).

RESEARCH   **Chinese University of Hong Kong**
Research Assistant   Aug. 2010 - Dec. 2014

1. Utilizing Microblogs for Automatic News Highlights Generation
Collaboration project with Qatar Computing Institute of Technology

- OVERVIEW: This project explores to use relevant tweets of a news article to help generate better news summary. We proposed three strategies: (1) extract news sentence with the help of relevant tweets; (2) extract tweets as news highlights; (3) use relevant tweets to help select and compress the news sentence as highlights.
- CORPUS: We constructed two corpora for evaluation, including CNN news articles and their relevant tweets. One corpus includes human generated summary collected from Amazon Mechanical Turk.
- PUBLICATION: COLING2014, SIGIR2015, two papers under review

2. Emergent Rumor Detection and Credibility Ranking

- OVERVIEW: This project explores to identify rumor events from social media and generate tweets summary for a specific rumor event. We proposed a rumor detection framework from three aspects: (1) text uncertainty, (2) opinion controversy, (3) opinion based credibility.
- CORPUS: We constructed one corpus for rumor detection from a Chinese microblogging website. Human annotation provided for evaluation collected from Amazon Mechanical Turk.
- PUBLICATION: ACL2013, two papers under review

3. Web Information Mining and Decision Support Platform for the Modern Service Industry

- OVERVIEW: This project aims to provide enterprises with the services of retrieving news from websites, extracting commercial information, exploring customers' opinions, and analyzing collaborative/competitive social networks. We developed a demo system MODEST. The core technologies are applied to the pillar industries of Hong Kong, including innovative finance, modem logistics, information technology, etc.
- SYSTEM: [MODEST] (in Chinese)
- PUBLICATION: ACL2014

4. Microblog Search

- OVERVIEW: We propose to improve microblog search performance from two aspects: (1) utilize external information for pseudo relevance feedback (PRF) ; (2) re-rank search results by combing several state-of-the-art ranking algorithems.
- PUBLICATION: Technical reports on TREC2011, TREC2012, TREC2013, TREC2014, workshop paper on SIGIR2015
- IMPACT: We participated in microblog track in TREC and won second place for year 2013 and 2014.

5. Social Media Analysis

- OVERVIEW: This project aims to investigate the behavior of mainstream media on Twitter and study how they exert their influence to shape public opinion during the UK's 2010 General Election.
- PUBLICATION: ACM HyperText2013

Knowledge Media Institute
Research Associate   Apr. 2012 - Jun. 2012

1. Exploratory Discourse Detection

- OVERVIEW: This project aims to identify exploratory dialogues from online learning material. Exploratory dialogue is desirable in online learning environments, since training learners to use it has been shown to improve learning outcomes. We propose a self-training framework to identify exploratory dialogue and apply it to Social Learn Platform to assist distance education.
- CORPUS: We construct a corpus with annotation from two education experts. The annotation guideline is generated based on mutual input.
- SYSTEM: The technique has been Applied to Social Learn Platform to assist online course.
- PUBLICATION: IJCLA

CONFERENCE REVIEW	**Secondary Reviewer**

2015: NAACL, ACL
2014: SigIR, COLING, EMNLP
2013: ACL, EMNLP, IJCAI, RANLP, AIRS
2012: ACL, EMNLP, PACLIC, AIRS
2011: EMNLP

PROFESSIONAL SKILLS

General: Machine learning for NLP, fast modeling and prototyping, data analysis, familiar with NLP annotation task
Programming: Java, Python, Unix/Linux shell scripting, R, C/C++

TEACHING EXPERIENCE

Teaching Assistant at Systems Engineering and Engineering Management, CUHK

Undergraduate course "Engineering Entrepreneurship"	Spring 2011, 2012, 2013, 2014
Undergraduate course "Information Systems Analysis and Design"	Fall 2012
Undergraduate course "Information Technology Management"	Fall 2012, 2013
Undergraduate course "Information Systems Management"	Fall 2011

MENTORSHIP

Jing Ma, Research Assistant, Chinese University of Hong Kong	Spring 2015 - Present
Jing Li, PhD, Chinese University of Hong Kong	Spring 2014 - Present
Jun Wen, PhD, Chinese University of Hong Kong	Spring 2014 - Present

HONORS AND AWARDS

Excellent Graduates in Heilongjiang Province (top 1%)	2010
Excellent Postgraduate Students in HIT (top 1%)	2010
Excellent Graduates in Heilongjiang Province (top 1%)	2008
Excellent Student Leaders in Heilongjiang Province (top 1%)	2006

REFEREES

Kam-Fai Wong
PhD, Professor
The Chinese University of Hong Kong
Email: kfwong@se.cuhk.edu.hk

Wei Gao
PhD, Scientist
Qatar Computing Research Institute
Email: wgao@qf.org.qa

Yulan He
PhD, Senior lecturer
Aston University
Email: y.he9@aston.ac.uk

Yang Liu
PhD, Associate Processor
The University of Texas at Dallas
Email: yangl@hlt.utdallas.edu