

## Jiajie Peng

---

CONTACT INFORMATION	School of Computer Science and Technology 315 Zonghe Building, 92 West Dazhi Street , Nan Gang District, Harbin, China	Voice: +86-187 4513 7124 E-mail: <a href="mailto:jiajiepeng@hit.edu.cn">jiajiepeng@hit.edu.cn</a>
RESEARCH INTERESTS	Bioinformatics, data mining, data-driven ontology construction, gene ontology based semantic similarity, biological network reconstruction	
EDUCATION	<b>Harbin Institute of Technology</b> , Harbin, China Ph.D., Computer Science and Technology Advisor: Yadong Wang	Aug. 2010 - now
	<b>Michigan State University</b> , East Lansing, MI, USA Visiting Scholar, MSU-DOE Plant Research Laboratory Mentor: Jin Chen	Sep. 2010 - Sep. 2012
	<b>Harbin Institute of Technology</b> , Harbin, China M.Phil., Computer Science and Technology B.Sc., Computer Science and Technology	Aug. 2008 - Jul. 2010 Aug. 2004 - Jul. 2008
PUBLICATIONS	<b>- Conference Papers</b> <ul style="list-style-type: none"><li>• <b>Zhongyu Wei</b> and Wei Gao, <i>Gibberish, Assistant, or Master? Using Tweets Linking to News for Extractive Single-Document Summarization</i>, ACM SigIR 2015, accepted.</li><li>• Gaoyan Ou, Wei Chen, Dongqing Yang, <b>Zhongyu Wei</b>, Binyang Li, Tengjiao Wang and Kam-Fai Wong, <i>Exploiting Community Emotion for Microblog Event Detection</i>, EMNLP 2014.</li><li>• <b>Zhongyu Wei</b>, Wei Gao, <i>Utilizing Microblogs for Automatic News Highlights Extraction</i>, COLING 2014.</li><li>• Binyang Li, Lanjun Zhou, <b>Zhongyu Wei</b>, Kam-Fai Wong, Ruifeng Xu and Yunqing Xia, <i>Web Information Mining and Decision Support Platform for the Modern Service Industry</i>, ACL 2014, demo paper.</li><li>• Lanjun Zhou, Binyang Li, <b>Zhongyu Wei</b> and Kam-Fai Wong, <i>An Open Discourse Corpus for Chinese with Annotated Explicit Discourse Connectives</i>, LREC 2014.</li><li>• <b>Zhongyu Wei</b>, Junwen Chen, Wei Gao, Binyang Li, Lanjun Zhou, Yulan He and Kam-Fai Wong, <i>An Empirical Study on Uncertainty Identification in Social Media Context</i>, ACL 2013.</li><li>• <b>Zhongyu Wei</b>, Yulan He, Wei Gao, Lanjun Zhou, Binyang Li and Kam-Fai Wong, <i>Mainstream Media Behavior Analysis on Twitter: A Case Study on UK General Election</i>, ACM HyperText 2013.</li><li>• Rebecca Ferguson, <b>Zhongyu Wei</b>, Yulan He, Simon Buckingham Shum, <i>An Evaluation of Learning Analytics to Identify Exploratory Dialogue in Online Discussions</i>, ACM LAK 2013.</li><li>• Lanjun Zhou, Wei Gao, Binyang Li, <b>Zhongyu Wei</b>, and Kam-Fai Wong. <i>Cross-lingual Identification of Ambiguous Discourse Connectives for Resource-Poor Language</i>, COLING 2012.</li><li>• Yulan He, Hassan Saif, <b>Zhongyu Wei</b>, and Kam-Fai Wong, <i>Quantifying Opinions for Political Tweets Analysis</i>, LREC 2012.</li><li>• Binyang Li, Lanjun Zhou, Wei Gao, Kam-Fai Wong and <b>Zhongyu Wei</b>, <i>An Effective Approach for Topic-Specific Opinion Summarization</i>, AIRS 2011.</li></ul>	

- 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).

#### - Under Review

- **Zhongyu Wei**, Yang Liu, Chen Li, *An Empirical Study on Using Topic Structure for News Highlights Generation*.
- **Zhongyu Wei**, Yang Liu, Chen Li, *Using Tweets to Help Sentence Compression for News Highlights Generation*.
- **Zhongyu Wei**, Wei Gao, Kam-Fai Wong, Yang Liu, *An Unsupervised Method of Using Tweets for News Highlights Generation*.
- Jing Li, Wei Gao, **Zhongyu Wei**, Kam-Fai Wong, *Summarizing Microblog Repost Trees with Structure Information*.
- Jing Ma, Wei Gao, **Zhongyu Wei**, Kam-Fai Wong, *Detect Rumors Using Time Series of Social Context Information on Microblogging Websites*.

## 1. Utilizing Microblogs for Automatic News Highlights Generation

*Collaboration project with Qatar Computing Institute of Technology*

- OVERVIEW: This project aims 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 aims to identify rumor events from social media and generate tweets summary for each identified rumor event. We proposed a rumor detection framework considering three factors: (1) text uncertainty, (2) temporal characteristics (3) opinion controversy.
- CORPUS: We constructed one corpus for rumor detection, including events consisting of related microblogs from a Chinese microblogging website. Human annotations were 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. The core technologies have been applied to the pillar industries of Hong Kong, including innovative finance, modern logistics, information technology, etc.
- SYSTEM: [MODEST] (in Chinese)
- PUBLICATION: ACL2014

## 4. Microblog Search

- OVERVIEW: We proposed to improve microblog search performance from two aspects: (1) utilize external information for pseudo relevance feedback (PRF) ; (2) re-rank search results by combining several state-of-the-art ranking algorithms.
- PUBLICATION: Technical reports on TREC2011, TREC2012, TREC2013, TREC2014 and SoMeRA2015
- 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. We proposed a self-training framework to identify exploratory dialogue.
- CORPUS: We constructed a corpus including online meeting text with human annotations collected from two education experts. The annotation guideline was generated based on collaboration with annotators.
- SYSTEM: The technique has been applied to [Social Learn Platform](#).
- PUBLICATION: IJCLA, ACM LAK2013

CONFERENCE REVIEW	<b>Secondary Reviewer</b>	
	2015: NAACL, ACL 2014: SigIR, COLING, EMNLP 2013: ACL, EMNLP, IJCAI, RANLP, AIRS 2012: ACL, EMNLP, PACLIC, AIRS 2011: EMNLP	
PROFESSIONAL SKILLS	<b>General:</b> Machine learning for NLP, fast modeling and prototyping, data analysis, familiar with human annotation collection for NLP task <b>Programming:</b> Java, Python, Unix/Linux shell scripting, R, C/C++	
TEACHING EXPERIENCE	<b>Teaching Assistant</b> 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 Graduate in Heilongjiang Province (top 1%)                      2010 Excellent Postgraduate Student in HIT (top 1%)                      2010 Excellent Graduate in Heilongjiang Province (top 1%)                      2008 Excellent Student Leader in Heilongjiang Province (top 1%)                      2006	
REFEREES	<div> <b>Kam-Fai Wong</b>  PhD, Professor  The Chinese University of Hong Kong  Email: kfwong@se.cuhk.edu.hk </div> <div> <b>Wei Gao</b>  PhD, Scientist  Qatar Computing Research Institute  Email: wgao@qf.org.qa </div> <div> <b>Yulan He</b>  PhD, Senior lecturer  Aston University  Email: y.he9@aston.ac.uk </div>	