

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: jiajiepeng@hit.edu.cn
RESEARCH INTERESTS	Bioinformatics, data mining, data-driven ontology construction, gene ontology based semantic similarity, biological network reconstruction	
EDUCATION	Harbin Institute of Technology , Harbin, China Ph.D., Computer Science and Technology Advisor: Yadong Wang	Aug. 2010 - now
	Michigan State University , East Lansing, MI, USA Visiting Scholar, MSU-DOE Plant Research Laboratory Mentor: Jin Chen	Sep. 2010 - Sep. 2012
	Harbin Institute of Technology , Harbin, China M.Phil., Computer Science and Technology B.Sc., Computer Science and Technology	Aug. 2008 - Jul. 2010 Aug. 2004 - Jul. 2008

PUBLICATIONS

- **Peng J**, Li H, Wang Y and Chen J, *A web tool for measuring gene semantic similarities by combining Gene Ontology and gene co-function networks*, proceedings of the 6th ACM Conference on Bioinformatics, Computational Biology (ACM BCB15), 2015, in press.
- **Peng J**, Li H, Liu Y, Juan L, Jiang Q, Wang Y and Chen J, *InteGO2: a Web Tool for Measuring and Visualizing Gene Semantic Similarities using Gene Ontology*, BMC Genomics, 2015, in press
- **Peng J**, Wang T, Hu J, Wang Y and Chen J, *Constructing organelle association network in Arabidopsis thaliana*, Current Genomics, 2015, in press.
- Liu Y, Liu J, Lu J, **Peng J**, Juan L, Zhu X, Li B and Wang Y, *Joint detection of copy number variations in parent-offspring trios*, Bioinformatics, 2015, in press.
- **Peng J**, Uygun S, Kim T, Wang Y, Rhee SY and Chen J, *Measuring semantic similarities by combining gene ontology annotations and gene co-function networks*, BMC Bioinformatics, 2015, 16:44.
- Jiang Q, Ma R, Wang J, Wu X, Jin S, **Peng J**, Tan R, Zhang T, Li Y and Wang Y, *LncRNA2Function: a comprehensive resource for functional investigation of human lncRNAs based on RNA-seq data*, BMC genomics, 2015, 16:S2.
- Jiang Q, Wang J, Wu X, Ma R, Jin S, Han Z, Tan R, **Peng J**, Liu G, Li Y and Wang Y, *LncRNA2Target: a database for differentially expressed genes after lncRNA knockdown or overexpression*, Nucleic Acids Res, 2015, 43:D193-D196.
- **Peng J**, Wang Y, and Chen J, *Towards integrative gene functional similarity measurement*, BMC Bioinformatics, 2014, 15:S5.
- Cheng L, Li J, Ju P, **Peng J** and Wang Y, *SemFunSim: A New Method for Measuring Disease Similarity by Integrating Semantic and Gene Functional Association*, PLoS ONE, 2014, 9:e99415.
- **Peng J**, Li H, Jiang Q, Wang Y and Chen J, *An Integrative Approach for Measuring Semantic Similarities using Gene Ontology*, BMC Systems Biology, 2014, 8:S8.
- Jin S, Tan R, Jiang Q, Xu L, **Peng J**, Wang Yo and Wang Y, *A Generalized Topological Entropy for Analyzing the Complexity of DNA Sequences*, PLoS ONE, 2014, 9:e88519.

- **Peng J**, Chen J and Wang Y, *Identifying cross-category relations in gene ontology and constructing genome-specific term association networks*, BMC Bioinformatics, 2013, 14:S15.
- **Peng J**, Wang T, Wang Y and Chen J, *Extending Gene Ontology with Gene Network Data*, Bioinformatics, Major Revision.

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 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 combing 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

PROFESSIONAL SKILLS **General:** Data mining for Bioinformatics, multi-omics data analysis, fast modeling and prototyping, familiar with network construction and analysis, familiar with ontology based data analysis
Programming: Java, C/C++, R, Unix/Linux shell scripting, Python

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 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

<p>Kam-Fai Wong PhD, Professor The Chinese University of Hong Kong Email: kfwong@se.cuhk.edu.hk</p>	<p>Wei Gao PhD, Scientist Qatar Computing Research Institute Email: wgao@qf.org.qa</p>
<p>Yulan He PhD, Senior lecturer Aston University Email: y.he9@aston.ac.uk</p>	