Zhaohui Wu

Phone: +1-814-880-7807   Email: laowuz@gmail.com

Summary

➢ 10 years' experiences in building large scale machine learning systems in Web Search & Mining, NLP, Recommendation

Experiences

Pinterest   05/2018 - Current

  Software Engineer   San Francisco, CA, USA

  -- Homefeed recommendation with specialty on user action modeling, source generation and ranking

Microsoft   05/2016 - 05/2018

  Applied Scientist   Sunnyvale, CA, USA

  -- Bing relevance with specialty on question answering and information extraction

- Pioneered List & Table Answers for Bing and Cortana by building the answer generation pipelines from scratch
- Increased Bing List & Table answer coverage from 0% to 25% of all answers
- Co-developed the 2nd version of Bing Passage answer generation system

The Pennsylvania State University   08/2011 - 05/2016

  Research Assistant   State College, PA, USA

  -- Better understanding natural language using machine learning and web knowledge

  -- Information extraction and knowledge mining of scholarly and educational big data

Microsoft Research   05/2014 - 08/2014

  Research Intern   Redmond, WA, USA

  -- Novel/emerging entity detection from daily updated news streams and online knowledge repositories

Microsoft Research   05/2013 - 08/2013

  Research Intern   Mountain View, CA, USA

  -- Locally defined entity understanding by finding informative passages from long documents

Patents

1. Systems and Methods for News Event Organization. Publication Number: US20160188590 A1, Jun. 2016.
2. Searching Locally Defined Entities. Publication Number: WO2015168344 A1, Nov. 2015.
3. A Feature Dictionary Generating Method for Text Classification based on LZW Compression Algorithm. Publication Number: ZL200810232557, Jun. 2010.

Selected Publications

Jianbo Ye, Yanran Li, **Zhaohui Wu**, James Z. Wang, Wenjie Li, Jia Li. Determining Gains Acquired from Word Embedding Quantitatively Using Discrete Distribution Clustering. **ACL 2017**, Vancouver, Canada.

Zhaohui Wu, Yang Song, C. Lee Giles. Exploring Multiple Feature Spaces for Novel Entity Discovery. **AAAI 2016**, Phoenix AZ, USA.

Chen Liang, Shuting Wang, **Zhaohui Wu**, Kyle Williams, Bart Pursel, Benjamin Brautigam, Sherwyn Saul, Hannah Williams, Kyle Bowen, C. Lee Giles. BBookX: Building Online Open Books for Personalized Learning. **AAAI 2016**, Phoenix AZ, USA.

Shuting Wang, Alexander Ororbia, **Zhaohui Wu**, Kyle Williams, Chen Liang, C. Lee Giles. Using Prerequisites to Extract Concept Maps from Textbooks. **CIKM 2016**, Indianapolis, Indiana, USA.

Madian Khabsa, **Zhaohui Wu**, C. Lee Giles. Towards Better Understanding of Academic Search. **JCDL 2016**, Newark, New Jersey, USA.

Zhaohui Wu, Chen Liang, C. Lee Giles. Storybase: Towards Building a Knowledge Base for News Events. **ACL 2015**, Beijing, China.

Chen Liang, **Zhaohui Wu**, Wenyi Huang, and C. Lee Giles. Measuring Prerequisite Relations Among Concepts. **EMNLP 2015**, Lisbon, Portugal.

Zhaohui Wu, C. Lee Giles. Sense-aware Semantic Analysis: A Multi-prototype Word Representation Model using Wikipedia. **AAAI 2015**, Austin, TX, USA.

Wenyi Huang, **Zhaohui Wu**, Chen Liang, Prasenjit Mitra and C. Lee Giles. A Neural Probabilistic Model for Context Based Citation Recommendation. **AAAI 2015**, Austin, TX, USA.

Zhaohui Wu, Dayu Yuan, Pucktada Treeratpituk, C. Lee Giles. Science and Ethnicity: How Ethnicities Shape the Evolution of Computer Science Research Community. arXiv:1411.1129 (2014).

Zhaohui Wu, Yuanhua Lv, Ariel Fuxman. Searching Locally-Defined Entities. **CIKM 2014**, Shanghai, China.

Zhaohui Wu, Wenyi Huang, Chen Liang, C. Lee Giles. Crowd-sourcing Web Knowledge for Metadata Extraction. **JCDL 2014**, London, UK.

Zhaohui Wu, Jian Wu, Madian Khabsa, Kyle Williams, Hung-Hsuan Chen, Wenyi Huang, Suppawong Tuarob, Sagnik Ray Choudhury, Alexander Ororbia, Prasenjit Mitra and C. Lee Giles. Towards Building a Scholarly Big Data Platform: Challenges, Lessons and Opportunities. **JCDL 2014**, London, UK.

Wenyi Huang, **Zhaohui Wu**, Prasenjit Mitra and C. Lee Giles. RefSeer: A Citation Recommendation System. **JCDL 2014**, London, UK.

Zhaohui Wu, Zhenhui Li, Prasenjit Mitra, C. Lee Giles. Can Back-of-the-book Indexes be Automatically Created? **CIKM 2013**, San Francisco, CA, USA.

Zhaohui Wu, Sujatha Das, Zhenhui Li, Prasenjit Mitra, C. Lee Giles. Searching Online Book Documents and Analyzing Book Citations. **DocEng 2013**, Florence, Italy.

Zhaohui Wu, Prasenjit Mitra, C. Lee Giles. Table of Contents Recognition and Extraction for Heterogeneous Book Documents. **ICDAR 2013**, Washington, DC, USA.

Zhaohui Wu, C. Lee Giles. Measuring Term Informativeness in Context. **NAACL-HLT 2013**, Atlanta, GA, USA.

Qinghua Zheng, **Zhaohui Wu**, Xiaocheng Cheng, Lu Jiang, Jun Liu. Learning to Crawl Deep Web. *Information Systems*, 38(6):801-819, 2013.

Jun Liu, Lu Jiang, **Zhaohui Wu**, Qinghua Zheng, Yanan Qian. Mining Learning-Dependency between Knowledge Units from Text. **The** *VLDB Journal*, 20(3): 335-345, 2011.

Zhaohui Wu, Lu Jiang, Qinghua Zheng, Jun Liu. Learning to Surface Hidden Web content. **AAAI 2010**. Atlanta, GA, USA.

Zhaohui Wu, Lu Jiang, Qinghua Zheng, Jun Liu, Zhenhua Tian, Junzhou Zhao. A Peep at Pornography Web in China. **WebSci 2010**. Raleigh, NC, USA.

Lu Jiang, **Zhaohui Wu**, Qian Feng, Jun Liu, Qinghua Zheng. Efficient Deep Web Crawling Using Reinforcement Learning. **PAKDD 2010**. Hyderabad, India.

Education

The Pennsylvania State University 08/2011 - 05/2016
  -- Ph.D., Department of Computer Science and Engineering
  -- Advisor: Prof. C. Lee Giles
Xi'an Jiaotong University 09/2004 - 06/2011
  -- Master, Department of Computer Science and Technology, June 2011
  -- Bachelor, Department of Computer Science and Technology, June 2008

Professional/Community Services

Conference PC Member: SIGIR2018, EMNLP2018, ACL2018, AAAI2018, IEEE BigData2018, AAAI2017, WSDM2017, ScholarlyBigData@IJCAI2016, BigScholar2015; LTI-SRS2015
Journal Reviews: Plos One, JASIST, Journal of Intelligent Information Systems, ACM Transactions on Asian and Low-Resource Language Information Processing, IEEE Transactions on Services Computing, IEEE Transactions on Emerging Topics in Computing, Journal of the Association for Information Science and Technology
Organizer: Penn State SIG Computational Informatics Seminar; CVPR2013 workshop on Symmetry Detection
Reviewer: WWW2017, DocEng2016, TPDL2016, NAACL2016, WWW2016, AAAI2016, WWW2015, JIIS, LAS2015, WSDM2015, DocEng2015, EMNLP2014, CIKM2014, SIGIR2014, WWW2014, CIKM2013, WWW2013, SIGIR2013, JCDL2013, TPDL2013, SIGIR2012, JCDL2012, CIKM2012, TPDL2012