

Yizhong Wang

CONTACT INFORMATION	Science Bldg. No.1, 1812S School of EECS, Peking Univ. No.5 Yiheyuan Road, Haidian District Beijing, China 100871	Phone: (+86) 188-1821-2946 E-mail: yizhong@pku.edu.cn URL: eastonwang.github.io Github: github.com/eastonwang
RESEARCH INTERESTS	Natural Lanugage Processing and Machine Learning, with current special focus on Discourse Analysis and Summarization	
EDUCATION	<p>Peking University, Beijing, China September, 2016 - June, 2019 (expected)</p> <p>M.S. Candidate, Computer Science, MOE Key Lab of Computational Linguistics</p> <ul style="list-style-type: none">• Current research fields: Discourse Parsing, Scientific Paper Summarization, Deep Learning• Advisor: Prof. Sujian Li <p>Shanghai Jiao Tong University, Shanghai, China September, 2012 - June, 2016</p> <p>B.Eng., Computer Science and Technology (IEEE Pilot Class)</p> <ul style="list-style-type: none">• Dissertation topic: “Mining Cultural Differences between Terms and Relations in Text”• Advisor: Prof. Kenny Q. Zhu and Prof. Xinbing Wang	
EXPERIENCE AND INTERNSHIPS	<p>Algorithm Engineer Intern @ TouchPal Technology November, 2015 - March, 2016</p> <p>Working at Input Method Engine Group for TouchPal Keyboard, I'm mainly responsible for building new language models for Hindi and Brazil Portuguese. I also develop tools for crawling web corpus, analysing user data and model evaluation.</p> <p>Research Intern @ ADAPT Lab, SJTU April, 2014 - August, 2016</p> <p>I work under the supervision of Prof. Kenny Q. Zhu and deal with various NLP tasks. We write a paper about Dependency Parsing and I also lead the Cultural Difference Mining group that aims to discover different semantics of a term in different languages.</p> <p>Research Intern @ IIOT Center, SJTU January, 2015 - January, 2016</p> <p>I join the AceMap team that tries to build a system for academic search and recommendation. My main contribution is applying LDA model to find topics in papers and tag them with names. I also work on topic visualization by using jQuery and D3.js.</p>	
HONORS AND AWARDS	<p>Chun-Tsung Scholarship (Funded by Nobel Prize laureate T. D. Lee), 2016</p> <p>Excellent Graduate of Shanghai Jiao Tong Univ., 2016</p> <p>Meritorious Winner (First Prize) of Mathematical Contest in Modeling, 2015</p> <p>Xindong Scholarship, 2013 / 2015</p> <p>Academic Excellence Scholarship of Shanghai Jiao Tong Univ., 2013 / 2014 / 2015</p> <p>Three A Student of Shanghai Jiao Tong Univ., 2013</p>	
PUBLICATIONS	<p>Yizhong Wang, Sujian Li. A Two-stage Parsing Method for Text-level Discourse Analysis. Under the review of ACL, 2017</p> <p>Wenjing Fang, Kenny Q. Zhu, Yizhong Wang, Jia Tan. Towards Non-projective High-Order Dependency Parser. Accepted by COLING, 2016</p>	

PROFESSIONAL
SKILLS

Programming languages: Python (Preferred), C++, Java, PHP, JavaScript, Shell

Machine Learning packages: TensorFlow, PyTorch, XGBoost, scikit-learn

Natural Language Processing toolkits: NLTK, StanfordCoreNLP

Distributed systems: Hadoop, Spark, Hive

Operating systems: Unix-like (Preferred), Windows

Other frequently-used tools: Git, Latex, Vim, MS Office

SELECTED
PROJECTS

Research on Text-level Discourse Parsing

A typical document is organized in a coherent way that each text unit is relevant to its context and plays a role in the entire semantics. Text-level discourse analysis tries to identify such discourse structure of a document as a discourse tree and its success can benefit many downstream tasks, like sentiment analysis and summarization. In this ongoing project, we analyze the current prevailing frameworks in details and we have proposed some new approaches for parsing discourse structures given a document. One paper on this work has been submitted to ACL 2017.

Curral Difference Mining between Equivalent Terms in Different Languages

<http://adapt.seiee.sjtu.edu.cn/cdminer>

Equivalent terms on the same object or entity in English and Chinese may exhibit subtle cultural differences in their use scenarios or perceptions in the respective culture. Understanding such differences is useful for many multilingual applications, such as machine translation of culture-sensitive words, cross-lingual recommendation or educational uses. The team I lead explores four word-embedding empowered methods to automatically discover and measure such cultural differences. We try to solve two key problems: how to learn precise word representations from large corpus and how to compute similarity of word vectors across languages.

Sequence-based Dependency Parsing for Non-Projective Languages

<http://adapt.seiee.sjtu.edu.cn/BeanParser/>

We propose a novel high-order dependency parsing framework that targets nonprojective languages. It imitates how human parses sentences in an intuitive way. At every step of the parse, it determines which word is the easiest to process among all the remaining words, identifies its head word and then folds it under the head word. This greedy framework achieves competitive accuracy on WSJ evaluation set and shows additional advantage on the non-projective corpus. The paper on this work was accepted by COLING 2016.

AceMap: Academic Search Engine and Recommendation System

<http://acemap.sjtu.edu.cn>

This is a quite fancy project still under way. It aims to set up an academic system to help users find the most suitable papers under a topic and visualize the relations among authors, topics and affiliations as networks. I work on the topic model for papers and apply several variants of LDA model to cluster papers into different topics. Since LDA models learn the clusters in an unsupervised style, we also try several methods to tag each topic with a name automatically. Then the topic clusters are visualized using D3.js and users can browse them dynamically in the official website.

SELECTED
EXTRACURRICULAR
ACTIVITIES

- **Teaching Assistant** of Discrete Mathematics at PKU, 2016 - 2017
- Work as the **Monitor** of IEEE Pilot Class, 2013 - 2016
- Work as the **Director** of the Amateur Astronomers Association in SJTU, 2014 - 2015
- Strong interests in **History** and **Sports**