

Sijia Ge

E-mail: Sijia.Ge@colorado.edu
Mobile: (720)5307370

Smiley Court, 1300 30th St.
Boulder, Colorado, 80303

EDUCATION

Department of Linguistics, University of Colorado-Boulder	Aug.2021-Now
M.S Computational Linguistics(CLASIC)	Expected:Jun.2023
School of Chinese Language and Literature, Nanjing Normal University	Sep.2016-Jun.2019
M.A Linguistics&Applied Linguistics (with Concentration on Computational Linguistics)	GPA:86.3/100(overall)
School of Chinese Language and Literature, Shanxi University	Sep.2012-Jul.2016
B.A Chinese Language and Literature	85.8/100(overall)

PROFESSIONAL EXPERIENCE

Beijing Lingosail Tech Co., Ltd.	
<i>Product Manager Intern, Beijing</i>	Mar.2021-June.2021
<ul style="list-style-type: none">■ Design the prototype of the new products including basic UI, function and interaction■ Schedule the overall progress of development as well as promotion model■ Test the function of new products with black box testing	

PROJECT&RESEARCH

A Joint Model of Automatic Sentence Segmentation and Lexical Analysis for Ancient Chinese	
<i>Research Assistant, Nanjing Normal University</i>	Sep.2018-Mar.2019
<ul style="list-style-type: none">■ Established a seq2seq model for sentence segmentation and lexical analysis based on Bi-LSTM-CRF and TensorFlow■ Transferred two separated tasks into one task, the F1-score of sentence segmentation, word segmentation, and POS reached 78.95,85.73% and 72.65% separately, with an average increase of 3.5%, 0.18%, and 0.35% respectively	
Annotated Imagery Corpus of Three Hundred Tang Poems	
<i>Research Leader, Nanjing Normal University</i>	Jul.2018-Jan.2019
<ul style="list-style-type: none">■ Annotated 4496 imageries occurring in the Three Hundred Tang Poems based on HowNet■ Designed the semantic annotation system including the literal meaning and the metaphorical meaning	
Named Entity Recognition on Chinese Classics Based on Bi-LSTM-CRF	
<i>Research Assistant, Beijing Gulian Corporation</i>	Oct.2017-Mar.2018
<ul style="list-style-type: none">■ Established an interface for named entity recognition on Chinese classics based on Bi-LSTM-CRF and utilized trie-tree for correcting with an F-score of 82.3% on <i>ZuanZhuan</i>■ Deployed the environment for deep learning, stress testing, and function testing	
Chinese Abstract Meaning Representation Corpus	
<i>Team Member, Nanjing Normal University</i>	Jul.2017-Mar.2018
<ul style="list-style-type: none">■ Counted the distribution of semantic roles for the predicate in 5000 sentences automatically■ Compared the data with semantic roles labeling based on Chinese PropBank to verify the efficiency of AMR for solving the conflicts between core and non-core roles, representation of multi-functional roles, and solution to dropped roles	
The Platform for monitoring the popularity of Mandarin Chinese	
<i>Team Member, The Education Department of Jiangsu Province</i>	Dec.2016-Dec.2017
<ul style="list-style-type: none">■ Developed a web platform based on spring framework for recording the nationwide result of the survey for Mandarin Chinese popularity■ Implemented the import and export data of Excel and audio files	
Chinese FrameNet corpus	
<i>Team Member, Shanxi University</i>	Nov.2015-Mar.2016
<ul style="list-style-type: none">■ Extended the core word of six frames and annotated about 300 sentences for these frames■ Participated in the discussion of the annotation rules	

RELATED SKILLS

Technical: Python, Genism, Nltk, Scikit-learn, Keras, Tensorflow, SQL, Django, HTML, Axure, Java, CentOS, SPSS, Latex
Language: Mandarin Chinese(Native), English(Fluent)

SELECTED PUBLICATION

Paper

Ning Cheng, Bin Li, Liming Xiao, Changwei Xu, **Sijia Ge**, Xingyue Hao, Minxuan Feng. Integration of Automatic Sentence Segmentation and Lexical Analysis of Ancient Chinese based on BiLSTM-CRF Model. *Proceedings of LT4HALA 2020 - 1st Workshop on Language Technologies for Historical and Ancient Languages*. Marseille, France, 2020:52-58.

Xingyue Hao, **Sijia Ge***, Yang Zhang, Yuling Dai, Peiyi Yan, and Bin Li. The Construction and Analysis of Annotated Imagery Corpus of Three Hundred Tang Poems. J.-F. Hong et al. (Eds.): *Proceedings of the 20th Chinese Lexical Semantics Workshop (CLSW 2019)*, LNAI 11831, pp. 517–524, 2020.

Li Song, Yuan Wen, **Sijia Ge**, Bin Li, Junsheng Zhou, Weiguang Qu and Nianwen Xue. An Easier and Efficient Framework to Annotate Semantic Roles: Evidence from the Chinese AMR Corpus. *The 13th Workshop on Asian Language Resources on LREC 2018*. Miyazaki, Japan, May 07, 2018:29-35.

Patent

A Method and System of Automatic Lexical Analysis for Ancient Chinese. (the 3rd applicant). 2019. No. CN201910085019.3

OTHER EXPERIENCE

A Representative of 19th Graduate Congress May.2018

Volunteer for the China National Conference on Computational Linguistics 2017(CCL 2017) Oct.2017

- guided the guests, in charge of the dormitory for over 400 attendees

Volunteer for the Museum of Shanxi Province Nov.2012-May.2013

- worked as a tour guide at the showroom about the history of Chinese coins