

KARINA HU

karinahusb@gmail.com | (732) 603-1361

EDUCATION

Stony Brook University, SUNY USA <i>Master of Arts in Computational Linguistic, GPA: 3.88/4.0</i>	Aug 2019 -Dec 2020
National Taiwan Normal University <i>Bachelor of Arts in English, GPA: 3.65/4.0</i>	Sep 2012-May 2016

SKILLS

Developer Tools: Git, NLTK, pandas, BeautifulSoup, matplotlib, PyTorch, Numpy, SpaCy
Programming Languages: Python, R, SQL
Linguistic Background: Syntax, Phonology, Computational Linguistic, Semantic, Phonetic

WORK EXPERIENCE

Glossika, Computational Linguistic Intern **Mar 2020- Present**

Portfolio: <https://viva.glossika.com/>

- Optimized the Syntax corpus and used the phrase structure rules to define the part of speech and analyzed the theta role of the text
- Developed torch.nn module for word embedding based on vocabulary size and the dimensionality of embeddings
- Applied knowledge of seq2seq model to develop NLP modeling tasks and handle great amount of preprocessing data (Arabic, Russian, Polish) with team
- Optimized datasets based on language and conducted dataset to be trained and built LSTM RNN (Seq2Seq), CBOW, Skip-gram and set up Docker container on AWS EC2 instance for Deep Learning

National Tsing Hua University Natural Language Processing Lab, Research Assistant **Jul 2020-Present**

Portfolio: <https://linggle.com/>

- Collaborated with python team members to optimize and develop English lesson plans for users to level up their grammar ability
- Developed the grammar detection system to detect Cambridge grammar rules by SpaCy, Regex, Pytorch
- Develop detection system API on Flask and collaborated with frontend (React) to show the level up graph to users on Linngle

iTutorGroup, Data Tagging Specialist **Sep 2016-Apr 2017**

- Successfully extracted suitable English-Mandarin teaching materials from the database and provided specific training session to call center members to hit great performance at request
- Collaborated with team members to create a motivate oriented culture

TECHNICAL PROJECTS

Deep Learning for Reddit Posts **Feb 2020-Present**

- Developed and designed a machine learning and data mining algorithm to scrape, clean large data and prepare unruly data using Python and BeautifulSoup
- Applied knowledge of tracking and indexing to design a web crawler algorithm to filter sentence structure and stored in a database by automatic scheduling
- Applied knowledge of CKY algorithm and Recursive Transition Network to parse tagging sentence structure
- Transformed vocabulary to a vector form and optimized datasets based on language and conducted dataset to be trained and built LSTM RNN to preprocess statistics and label which increased accuracy rate about 80%

Brill Tagger Project **Oct 2019-Dec 2019**

- Assigned the most common Part of Speech (POS) tagging in Syntax, and implemented a learning phase in which a pre-tagged text would be analyzed to determine which tag is most common for each word is in the corpus
- Learned Brill Tagger's set of rules to improve its performance by running the first step over another pre-tagged text and comparing its own output with the given tags
- Analyzed the rules that look at the context of a word and at the spelling of the word to capture the generalizations in the morphology and the capitalization rules

HONORS & ACTIVITIES

Extra-Curricular for Degree

- Exchange volunteer program: Bangladesh| 2014 (for teaching)
- Alumnus scholarships, Department of English, National Taiwan Normal University | 2016