SHENGJIA YAN

(+86)15961887272 · sjyan1995@gmail.com · yanshengjia.com · https://github.com/yanshengjia

Education

- **Southeast University (SEU)** Nanjing, China
  B.E. in Computer Science and Technology; GPA: 3.56/4.0 *2013.08 - 2017.06*

Working Experience

- **AI Research Group, 17zuoye** Beijing, China
  NLP Engineer *2017.06 - Present*
  - Design and develop an automated essay enhancing system
  - Built an automated essay scoring system based on Feature Engineering, Logistic Regression and LSTM, which reached the best performance on the Kaggle ASAP dataset
  - Developed a grammar checker based on Convolutional Seq2Seq models and rules, which reached the best F0.5-score on CoNLL2014 datasets
- **Knowledge Science and Engineering Lab, Southeast University** Nanjing, China
  Research Assistant (advisor: Prof. Guilin Qi) *2014.10 - 2017.06*
  - Carried out data preprocessing using NLP approaches like spaCy to refine and analyze datasets
  - Presented and implemented a Random Walk algorithm in Python based on the Probabilistic Graphical Model to map the string mentions in web tables to their referent entities in a knowledge base
  - Achieved a 6% increase in F1-score compared to the latest published schemes. The result was **published in [1, 2]**

Selected Projects

- **Deep Learning Grammar Error Correction System** 2018.06
  - Designed and developed a GEC system based on Facebook well-known Convolutional Seq2Seq paper and rule-based proofreading software LanguageTool
  - Supported grammar checkings of more than 1000 tokens per second on a single Tesla P100 GPU by optimizing deep learning inference with Nvidia TensorRT and ONNX
- **Crowdsourcing NLP Annotation Platform** 2018.05
  - Designed a crowdsourcing annotation system with multiple quality control mechanisms based on annotation tool Brat
  - Developed the frontend with HTML, CSS, Bootstrap and JavaScript
  - Built the backend service using Tornado/Python, MongoDB and deployed on AWS
- **DNN-Based Face Recognition System** 2017.03
  - Implemented the neural network Backpropagation algorithm in C and Constructed a DNN to recognize human's faces, poses and emotions
  - 150+ stars and 180+ forks on GitHub
- **C-Minus Compiler** 2016.06
  - Implemented the Regular-Expression-to-NFA converter, LR(1) parser and semantic analysis module in Python
  - Visualized the compiling process by ploting NFA, DFA, GOTO graphs with GraphViz

Publications

1. "Entity Linking in Web Tables with Multiple Linked Knowledge Bases", In proceedings of *Joint International Semantic Technology Conference*. Springer, Cham, 2016: 239-253 [pdf]
2. "A Method of Entity Linking in Web Tables based on Multiple Linked Knowledge Bases", Chinese Patent, CN106503148A, 2017

Skills

- **Languages**: Python, C/C++, JavaScript, Markdown, LaTeX
- **Tools**: Git, MongoDB, Tornado, Bootstrap, Qt
- **Frameworks**: TensorFlow, PyTorch, Keras, Scikit-Learn, Gensim, spaCy

Honors

- Computer Programming Contest (Jiangsu Province), Third Prize, 2016.11
- SEU Computer Programming Contest, Fourth Place, 2016.10
- Outstanding Project, SEU Student Research Training Program, 2016.05

Extracurricular Activities

- International Student Leadership Program, California Polytechnic State University, CA, USA, 2016.01
- Nanjing Youth Olympic Games Volunteers, National Olympic Committee Assistant, Nanjing, China, 2014.07