Yanjie He

Email : heyanjie0@outlook.com
Website: https://yanjiehe.github.io/
Github: YanjieHe
LinkedIn: Yanjie He

Introduction

I am an engineer who is interested in Data Science and Machine Learning. Currently, I am a graduate student in the Data Analytics program in the George Washington University. My goal is to be a Data Scientist in the technology field.

Education

- **The George Washington University** — Washington D.C., USA
  Master of Science in Data Analytics, School of Engineering and Applied Science — *Aug. 2018 – Dec. 2020*
- **Shanghai University of International Business and Economics** — Shanghai, China
  Bachelor of Arts in Economics; GPA: 3.71/4.00 — *Sept. 2013 – June. 2017*

Work Experience

- **Kantar Media CIC** — Shanghai, China
  Data Engineer Intern — *July 2016 - Feb 2017, 8 months*
  - **Data Collection**: Designed data collection and data cleaning solution for Chanel APAC project.
  - **Text Mining**: Developed text mining system with team members. The system were used in making data analysis solution for L'Oréal, Chanel, Volkswagen and Dell.
    - The system contains a rule parser and an evaluator, which are able to let the data analysts to define the patterns that they want to match.
    - The system can provide basic-level functions for sentiment analysis
  - **Data Visualization**: Developed Data Visualization Solution for GroupM television show.

Projects

- **Compiler and Virtual Machine**: The project contains a compiler for a statically typed language, a byte code disassembler and a virtual machine. The compiler consist of following steps:
  - lexical analysis (I wrote DFA to extract tokens)
  - syntax analysis (I wrote a recursive descent parser by hand)
  - semantic analysis (I implemented the visitor pattern, which is quite useful in this situation)
  - code generation

  Besides, I designed a byte code instruction list, which is similar to the JVM instructions. And also, my implementation of the virtual machine taking JVM as reference. The compiler and disassembler are written in C++ and the virtual machine is written in ANSI C.
- **Matrix Library**: The matrix library is implemented in C#, which can be used for matrix computation and solving linear equations. Since the numerical calculation packages on the .Net platform are very limited, some developers used this library as a lightweight choice.
- **Reversi A.I.**: Reversi is a board game. Due to the huge amount of possible states of the game, it is next to impossible for computer to enumerate every situation. Therefore, I implemented the minimax algorithm with alpha-beta pruning. The project is written in C++ and has a GUI which is build using Qt5 Framework.

Programming Skills

- **Languages**: C, C++, C#, Java, Scala, Python, R, SQL, Scheme/Racket, Html/CSS
- **Technologies**: Data Analysis, Qt 5 Framework, Apache Tomcat 7, Linux, MySQL
- **Software**: Gephi, Minitab