Shu Shi

Undergraduate/Junior in Computer science

EDUCATION

09/2021– Now **University of Science and Technology of China**, *Undergraduate*, Hefei, China
School of the Gifted Young, Computer Science. GPA:3.49/4.3

SKILLS

Knowledge: Machine Learning, Operating System, Computer Organization, Foundations of Algorithms, Computer Network, Principles and Techniques of Compiler

Programming: C++[4yr], C[4yr], Python[2yr], LaTeX[2yr], Markdown[2yr], Pascal[2yr], Verilog[1yr], Golang[1Mth]

Packages: Numpy/Scipy[1yr], Plotly/Matplotlib[1yr], Pandas/Keras/Tensorflow/Scikit-Learn/PyTorch[1yr]

PROJECTS

10/2023 - 12/2023 **Compiler for Translating Cminusf into LoongArch(C++)**
- Develop a lexical analyzer based on Flex and a syntax analyzer based on Bison to automatically construct the syntax tree for the Cminus language.
- Utilize the Visitor pattern to traverse the syntax tree (invoke the Light IR C++ library), and achieve automated generation of Intermediate Representation (IR).
- Apply the Mem2Reg Pass to optimize Light IR code.
- Develop an automatic translation program from Light IR code to LoongArch code, utilizing a stack-based allocation strategy.

11/2023 - 12/2023 **Knowledge-aware Recommendation for Douban Movies(Python)**
- Retrieve and match entities in Freebase corresponding to a provided list of movie IDs. Starting from these entities, retrieve n-step reachable entities to form a subgraph.
- Filter relationships and entities in the graph with low occurrence frequency.
- Employ the Basic Matrix Factorization algorithm to establish a model, train it based on the filtered subgraph, and predict user ratings for items.

10/2023 - 11/2023 **Query System and Recommendation System based on Douban Data(Python)**
- Use a web crawler to scrape basic data for specified movies based on movie and book ID lists. (such as movie title, cast, synopsis, etc.)
- Perform tokenization on the basic information of movies and books (HanLP), then establish a keyword inverted index table based on this, and compress the index table.
- For a given keyword query, first convert it into a boolean expression query, and then return a set of books or movies that match the query based on the inverted index table.
- Utilize the lightGCN model to perform graph convolution operations on the user-item interaction graph for learning embedded representations of users and items.
- Predict user ratings for unrated items (books/movies), and evaluate the prediction results using NDCG (Normalized Discounted Cumulative Gain)

7/2023 - Now **Optimization of metadata management in CubeFS(Golang)**
- Performing cold and hot identification of metadata.
- Migrating metadata from in-memory B-tree to more cost-effective RocksDB storage.
- This project is still on-going.

5/2023 **Implementation of a Pipeline CPU(Verilog)**
- Realised a single-cycle CPU with the support of 19 types of instructions.
- Designed various sub-components, including the control unit, register file, program counter, immediate number expansion unit, ALU, and branch unit, as well as multiple multiplexers.
- Utilized IP cores to instantiate the Memory section, consisting of the Instruction Memory and Data Memory.
- Established interconnections between the CPU components and conducted comprehensive testing.
- Developed an assembly program for initializing the Instruction Memory and testing the single-cycle CPU.

School of the Gifted Young, USTC – Hefei, Anhui 230000
+86 177-6447-5075 • *sshi@mail.ustc.edu.cn* • *Timenumber*

6/2023 **Implementation of a Single-Cycle CPU(Verilog)**

- Realised a single-cycle CPU with the support of 19 types of instructions.
- Designed various sub-components, including the control unit, register file, program counter, immediate number expansion unit, ALU, and branch unit, as well as multiple multiplexers.
- Utilized IP cores to instantiate the Memory section, consisting of the Instruction Memory and Data Memory.
- Established interconnections between the CPU components and conducted comprehensive testing.
- Developed an assembly program for initializing the Instruction Memory and testing the single-cycle CPU.

5/2023 **Implementation of a Linux Shell(C)**

- Implemented a Linux Shell using the C language and system calls.
- Invocation of execute() for command executions excluding shell builtin.
- Incorporated the pipe operator and subcommand operator into the shell.
- Implemented three shell built-in commands: exit, cd, and kill.

10/2021 – 5/2022 **Deep Learning-Based Cat Face Recognition Technology(Python)**

- Gathered a real cat image dataset in USTC campus and categorized them with different labels.
- Conducted a comprehensive search in a large database for training and implemented transfer learning techniques in the model.
- Employed the ResNet model to recognize cat faces at USTC.
- Achieved superior performance of the ResNet model compared to a standard CNN. Wrote an essay summarizing the findings and delivered the final report.
- Personal Contribution: Collaborated with a partner to code the program, conducted a thorough search for the cat training dataset, organized and categorized the cat picture database, reshaped certain images to enhance cat identification for better training, and programmed the code for processing input data. Additionally, formulated conclusions and presented the final report on behalf of the team.

12/2022 **Huffman Tree-Based Compression(C)**

- Developed code for setting up the Huffman Tree, designed the data structure, implemented file compression and decompression.

HONORS AND AWARDS

2022 Outstanding Student Scholarship of USTC

2021 Freshman Excellence Scholarship of USTC

2020 Second Prize, Chinese Physics Olympiad, Anhui