Yuze Gao

Email: yuze.gao@outlook.com
Tel: +0065 90553268

Last updated: Jun. 2017

Education

Sept. 2010 - Jul. 2014

Northeastern University **Shenyang, China**

Bachelor of Engineering in Computer Science and Technology

Thesis: The Design and Implementation of a Simple Teaching Compiler for C

Research Experience

Oct.2016-Present

NLP Lab **Singapore University of Technology and Design** **Singapore**

Position: Research Assistant

Project: 1. WP2. MINDEF grant. Cross-functional information systems for decision making

Aug. 2014 - Feb. 2016

Natural Language Processing Lab, Northeastern University **Shenyang, China**

Position: Research Assistant

Project: Improving Phrase-based Statistical Machine Translation

Responsibility: This project mainly aims to speed up the decoding process in a phrase-based statistical machine translation system while maintaining its BLUE, which is an index of the translation accuracy. One solution is to apply pruning in the decoding process. Since natural language translation systems share some principles with compilers, thus we can use some similar hard or soft constrains, such as derivation rules or grammatical structures, to direct its decoding progress. In particular, I successfully applied syntactic parse trees and boundary information into the optimization of pruning in the decoding process of natural languages. This gained significant decoding speed-up with acceptable loss of translation accuracy.

Mar. 2014 - Jun. 2014

Natural Language Processing Lab, Northeastern University **Shenyang, China**

Project: The Design and Implementation of a Simple Teaching Compiler for C *(Thesis Project)*

Responsibility: This project aims to develop a teaching compiler for the course "Compiler Theory." The envisaged tool aims to demonstrate the compilation process clearly by giving the detailed intermediate result of each stage. Current implementation supports simple C constructs (like while, for, and if) and data types (like int, float, and array). The supported intermediate results include lexical analysis results (token sequence, lexical analysis errors and identifier list), syntax analysis (first set, follow set, select set, terminal symbol, non-terminal symbol, analysis steps and forecasting analysis table), and translation results (intermediate language like quart expression, symbol table and the target assembly language).

Publication

2015

1. **A Comparison of Pruning Methods for CYK-based Decoding in Machine Translation**
   Yuze Gao, Tong Xiao

Academic Service

Sept. 2015 Publishing Editor of Proceedings of the 11th China Workshop on Machine Translation (CWMT 2015)

Internships

Dec. 2013 - Feb. 2014 **Institute of Information Management and Integration, Northeastern University** **Shenyang, China**

Project: A Programming Experience Sharing Forum

Responsibility: The goal of this project is to develop a web-based programming experience sharing platform using the SSH (Struts, Spring, Hibernate) framework for Java web applications. As leader of a four-person student team, apart from adjusting the progress of the project, I was also responsible for the design and implementation of the database layer of the forum.

Apr. 2013 - May 2013 **Institute of Computer Technology Application, Northeastern University** **Shenyang, China**

Project: A Web-based Management System for Motorcycle Shops

Responsibility: The project aims to build a web-based management system for a motorcycle company with many nation-wide branches. In particular, I was responsible for the frontend UI design.

Technical Skills

Natural Language Processing

Familiar with statistical and syntax-based machine translation (especially the phrase-based and hierarchy phrase-based) and related machine learning algorithms, Neural Network related Parsing, Sentiment analysis.

NLP Directions: Text Mining, Opinion and Sentiment Mining, Language Parsing

Compilers

Familiar with compiler concepts and programming language processing

Have some experience in applying compiler techniques to Statistical Machine Translation Systems

Programming Languages

Major: C/ C++, Python

Minor: Perl, Assembly Language

Programming Toolkits

Familiar with UNIX programming environment and Linux system programing

Platform: Ubuntu and other Debian-based distributions

Tools: GCC, GDB, GNU Make, Eclipse