Bachelor's thesis

TINYC COMPILER FRONTEND

Mykhailo Anisimov

Faculty of Information Technology
Department of Software Engineering
Supervisor: Ing. Petr Máj, Ph.D.
February 21, 2025

FACULTY
OF INFORMATION
TECHNOLOGY
CTU IN PRAGUE

Assignment of bachelor's thesis

Title:	TinyC Compiler Frontend
Student:	Mykhailo Anisimov
Supervisor:	Ing. Petr Máj, Ph.D.
Study program:	Informatics
Branch / specialization:	Software Engineering 2021
Department:	Department of Software Engineering
Validity:	until the end of summer semester 2025/2026

Instructions

The aim of the project is to design universal compiler frontend for the TinyC programming language as used in the NI-GEN course that can be given to its students so they can focus on the middle- and back-end work. The frontend should be implemented in C++. It should parse the TinyC language into an abstract syntax tree whose representation should follow established Object Oriented Programming principles. It should be available either as a library with the AST classes directly usable by students, or as a standalone executable that will output the parsed AST in a standardized JSON format (including source location information).

The thesis should:
1) Analyze the landscape of language parsers and language agnostic AST representations (such as babel/parser for JavaScript)
2) Design and document AST representation for TinyC and its JSON format.
3) Design, document, implement and test the TinyC parser.
4) Discuss further development of the project.

Electronically approved by Ing. Michal Valenta, Ph.D. on 12 February 2025 in Prague.

Czech Technical University in Prague
Faculty of Information Technology

Citation of this thesis: Anisimov Mykhailo. *TinyC Compiler Frontend*. Bachelor's thesis. Czech Technical University in Prague, Faculty of Information Technology, 2025.

Here goes the acknowledgment part...

Declaration

FILL IN ACCORDING TO THE INSTRUCTIONS. VYPLŇTE V SOULADU S POKYNY. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Curabitur sagittis hendrerit ante. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Cras pede libero, dapibus nec, pretium sit amet, tempor quis. Sed vel lectus. Donec odio tempus molestie, porttitor ut, iaculis quis, sem. Suspendisse sagittis ultrices augue. Donec ipsum massa, ullamcorper in, auctor et, scelerisque sed, est. In sem justo, commodo ut, suscipit at, pharetra vitae, orci. Pellentesque pretium lectus id turpis.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Curabitur sagittis hendrerit ante. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Cras pede libero, dapibus nec, pretium sit amet, tempor quis. Sed vel lectus. Donec odio tempus molestie, porttitor ut, iaculis quis, sem. Suspendisse sagittis ultrices augue. Donec ipsum massa, ullamcorper in, auctor et, scelerisque sed, est. In sem justo, commodo ut, suscipit at, pharetra vitae, orci. Pellentesque pretium lectus id turpis.

In Prague on February 21, 2025

Abstract

Fill in the abstract of this thesis in English. Lorem ipsum dolor sit amet. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Cras pede libero, dapibus nec, pretium sit amet, tempor quis. Sed vel lectus. Donec odio tempus molestie, porttitor ut, iaculis quis, sem. Suspendisse sagittis ultrices augue.

Keywords enter, comma, separated, list, of, keywords, in, ENGLISH

Abstrakt

Fill in the abstract of this thesis in Czech. Lorem ipsum dolor sit amet. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Cras pede libero, dapibus nec, pretium sit amet, tempor quis. Sed vel lectus. Donec odio tempus molestie, porttitor ut, iaculis quis, sem. Suspendisse sagittis ultrices augue.

Klíčová slova enter, comma, separated, list, of, keywords, in, CZECH

Contents

1 2nd exercise **1**

1.1 Microtypography 1

1.2 Source code 1

A TinyC Grammar **2**

A.1 Program Structure 2

A.2 Statements 3

A.3 Expressions and Variable Declarations 4

Contents of the attachments **6**

List of Figures

List of Tables

List of code listings

List of abbreviations

DFA	Deterministic Finite Automaton
FA	Finite Automaton
LPS	Labelled Prüfer Sequence
NFA	Nondeterministic Finite Automaton
NPS	Numbered Prüfer Sequence
XML	Extensible Markup Language
XPath	XML Path Language
XSLT	eXtensible Stylesheet Language Transformations
W3C	World Wide Web Consortium

Chapter 1
2nd exercise

Each (sub)chapter should have some introductory text.

1.1 Microtypography

A text – especially a professional one such as this work - must be divided into paragraphs. Each paragraph should relate to one topic or idea. . . Paragraphs must be visually separated from each other. There are several suitable styles for this, which we described in the last lecture. Paragraphs can be set in different ways. In professional texts, the “block” typesetting is common. It is necessary to change the interword spaces appropriately. Their recommended size is 0.25–0.33 square.

1.2 Source code

Appendix A
TinyC Grammar

A.1 Program Structure

⟨*PROGRAM*⟩ ::= ⟨*PROGRAM_ITEM*⟩ ⟨*PROGRAM*⟩ | ϵ

⟨*PROGRAM_ITEM*⟩ ::= ⟨*NON_VOID_TYPE*⟩ identifier ⟨*NOT_VOID_FUNC_OR_VAR*⟩
| void ⟨*VOID_DECL_TAIL*⟩
| ⟨*STRUCT_DECL*⟩
| ⟨*FUNPTR_DECL*⟩

⟨*NOT_VOID_FUNC_OR_VAR*⟩ ::= ⟨*VARIABLE_TAIL*⟩ | ⟨*FUNC_DECL_TAIL*⟩

⟨*VOID_DECL_TAIL*⟩ ::= identifier ⟨*FUNC_DECL_TAIL*⟩
| ⟨*STAR_PLUS*⟩ identifier ⟨*FUNC_OR_VAR_TAIL*⟩

⟨*FUNC_OR_VAR_TAIL*⟩ ::= ⟨*VARIABLE_TAIL*⟩ | ⟨*FUNC_DECL_TAIL*⟩

⟨*VARIABLE_TAIL*⟩ ::= ⟨*OPT_ARRAY_SIZE*⟩ ⟨*OPT_INIT*⟩ ⟨*MORE_GLOBAL_VARS*⟩
;

⟨*FUNC_DECL_TAIL*⟩ ::= `(` ⟨*OPT_FUN_ARGS*⟩ `)` ⟨*FUNC_TAIL*⟩

⟨*FUNC_TAIL*⟩ ::= ⟨*BLOCK_STMT*⟩ | ;

⟨MORE_GLOBAL_VARS⟩ ::= , identifier ⟨OPT_ARRAY_SIZE⟩ ⟨OPT_INIT⟩
⟨MORE_GLOBAL_VARS⟩ | ϵ

⟨OPT_FUN_ARGS⟩ ::= ⟨FUN_ARG⟩ ⟨FUN_ARG_TAIL⟩ | ϵ

⟨FUN_ARG_TAIL⟩ ::= , ⟨FUN_ARG⟩ ⟨FUN_ARG_TAIL⟩ | ϵ

⟨FUN_ARG⟩ ::= ⟨TYPE⟩ identifier

A.2 Statements

⟨STATEMENT⟩ ::= ⟨BLOCK_STMT⟩ | ⟨IF_STMT⟩ | ⟨SWITCH_STMT⟩
| ⟨WHILE_STMT⟩ | ⟨DO_WHILE_STMT⟩ | ⟨FOR_STMT⟩
| ⟨BREAK_STMT⟩ | ⟨CONTINUE_STMT⟩ | ⟨RETURN_STMT⟩
| ⟨EXPR_STMT⟩

⟨BLOCK_STMT⟩ ::= `{` ⟨STATEMENT_STAR⟩ `}`

⟨STATEMENT_STAR⟩ ::= ⟨STATEMENT⟩ ⟨STATEMENT_STAR⟩ | ϵ

⟨IF_STMT⟩ ::= `if` `(` ⟨EXPR⟩ `)` ⟨STATEMENT⟩ ⟨ELSE_PART⟩

⟨ELSE_PART⟩ ::= `else` ⟨STATEMENT⟩ | ϵ

⟨SWITCH_STMT⟩ ::= `switch` `(` ⟨EXPR⟩ `)` `{` ⟨CASE_DEF_STMT_STAR⟩
`}`

⟨CASE_DEF_STMT_STAR⟩ ::= ⟨CASE_STMT⟩ ⟨CASE_DEF_STMT_STAR⟩
| ϵ | ⟨DEFAULT_CASE⟩ ⟨CASE_STMT_STAR⟩

⟨CASE_STMT⟩ ::= `case` integer_literal `:` ⟨CASE_BODY⟩

⟨CASE_BODY⟩ ::= ⟨STATEMENT_STAR⟩

⟨DEFAULT_CASE⟩ ::= `default :` ⟨CASE_BODY⟩

⟨WHILE_STMT⟩ ::= `while (` ⟨EXPR⟩ `)` ⟨STATEMENT⟩

⟨DO_WHILE_STMT⟩ ::= `do` ⟨STATEMENT⟩ `while (` ⟨EXPR⟩ `) ;`

⟨FOR_STMT⟩ ::= `for (` ⟨OPT_EXPR_OR_VAR_DECL⟩ `;` ⟨OPT_EXPR⟩ `;` ⟨OPT_EXPR⟩ `)` ⟨STATEMENT⟩

⟨OPT_EXPR_OR_VAR_DECL⟩ ::= ⟨EXPR_OR_VAR_DECL⟩ | ϵ

⟨OPT_EXPR⟩ ::= ⟨EXPR⟩ | ϵ

⟨BREAK_STMT⟩ ::= `break ;`

⟨CONTINUE_STMT⟩ ::= `continue ;`

⟨RETURN_STMT⟩ ::= `return` ⟨OPT_EXPR⟩ `;`

⟨EXPR_STMT⟩ ::= ⟨EXPR_OR_VAR_DECL⟩ `;`

A.3 Expressions and Variable Declarations

⟨EXPR_OR_VAR_DECL⟩ ::= ⟨VAR_DECLS⟩ | ⟨EXPRS⟩

⟨VAR_DECLS⟩ ::= ⟨VAR_DECL⟩ ⟨VAR_DECLS_TAIL⟩

⟨VAR_DECLS_TAIL⟩ ::= `,` ⟨VAR_DECL⟩ ⟨VAR_DECLS_TAIL⟩ | ϵ

⟨VAR_DECL⟩ ::= ⟨TYPE⟩ identifier ⟨OPT_ARRAY_SIZE⟩ ⟨OPT_INIT⟩

⟨OPT_ARRAY_SIZE⟩ ::= `[` ⟨E9⟩ `]` | ϵ

$\langle \mathit{OPT\_INIT} \rangle ::= \texttt{=} \; \langle \mathit{EXPR} \rangle \mid \epsilon$

$\langle \mathit{EXPRS} \rangle ::= \langle \mathit{EXPR} \rangle \; \langle \mathit{EXPRS\_TAIL} \rangle$

$\langle \mathit{EXPRS\_TAIL} \rangle ::= \texttt{,} \; \langle \mathit{EXPR} \rangle \; \langle \mathit{EXPRS\_TAIL} \rangle \mid \epsilon$

$\langle \mathit{EXPR} \rangle ::= \langle \mathit{E9} \rangle \; \langle \mathit{EXPR\_TAIL} \rangle$

$\langle \mathit{EXPR\_TAIL} \rangle ::= \texttt{=} \; \langle \mathit{EXPR} \rangle \mid \epsilon$

$\langle \mathit{E9} \rangle ::= \langle \mathit{E8} \rangle \; \langle \mathit{E9\_Prime} \rangle$
$\langle \mathit{E9\_Prime} \rangle ::= \texttt{||} \; \langle \mathit{E8} \rangle \; \langle \mathit{E9\_Prime} \rangle \mid \epsilon$

$\langle \mathit{E8} \rangle ::= \langle \mathit{E7} \rangle \; \langle \mathit{E8\_Prime} \rangle$
$\langle \mathit{E8\_Prime} \rangle ::= \texttt{\&\&} \; \langle \mathit{E7} \rangle \; \langle \mathit{E8\_Prime} \rangle \mid \epsilon$

$\langle \mathit{E7} \rangle ::= \langle \mathit{E6} \rangle \; \langle \mathit{E7\_Prime} \rangle$
$\langle \mathit{E7\_Prime} \rangle ::= \texttt{|} \; \langle \mathit{E6} \rangle \; \langle \mathit{E7\_Prime} \rangle \mid \epsilon$

$\langle \mathit{E6} \rangle ::= \langle \mathit{E5} \rangle \; \langle \mathit{E6\_Prime} \rangle$
$\langle \mathit{E6\_Prime} \rangle ::= \texttt{\&} \; \langle \mathit{E5} \rangle \; \langle \mathit{E6\_Prime} \rangle \mid \epsilon$

$\langle \mathit{E5} \rangle ::= \langle \mathit{E4} \rangle \; \langle \mathit{E5\_Prime} \rangle$
$\langle \mathit{E5\_Prime} \rangle ::= \texttt{==} \; \langle \mathit{E4} \rangle \; \langle \mathit{E5\_Prime} \rangle \mid \texttt{!=} \; \langle \mathit{E4} \rangle \; \langle \mathit{E5\_Prime} \rangle \mid \epsilon$

$\langle \mathit{E4} \rangle ::= \langle \mathit{E3} \rangle \; \langle \mathit{E4\_Prime} \rangle$
$\langle \mathit{E4\_Prime} \rangle ::= \texttt{<} \; \langle \mathit{E3} \rangle \; \langle \mathit{E4\_Prime} \rangle \mid \texttt{<=} \; \langle \mathit{E3} \rangle \; \langle \mathit{E4\_Prime} \rangle \mid \texttt{>} \; \langle \mathit{E3} \rangle \; \langle \mathit{E4\_Prime} \rangle \mid \texttt{>=} \; \langle \mathit{E3} \rangle \; \langle \mathit{E4\_Prime} \rangle \mid \epsilon$

$\langle \mathit{E3} \rangle ::= \langle \mathit{E2} \rangle \; \langle \mathit{E3\_Prime} \rangle$
$\langle \mathit{E3\_Prime} \rangle ::= \texttt{<<} \; \langle \mathit{E2} \rangle \; \langle \mathit{E3\_Prime} \rangle \mid \texttt{>>} \; \langle \mathit{E2} \rangle \; \langle \mathit{E3\_Prime} \rangle \mid \epsilon$

$\langle \mathit{E2} \rangle ::= \langle \mathit{E1} \rangle \; \langle \mathit{E2\_Prime} \rangle$
$\langle \mathit{E2\_Prime} \rangle ::= \texttt{+} \; \langle \mathit{E1} \rangle \; \langle \mathit{E2\_Prime} \rangle \mid \texttt{-} \; \langle \mathit{E1} \rangle \; \langle \mathit{E2\_Prime} \rangle \mid \epsilon$

Contents of the attachments

```
/
├── readme.txt ............................ stručný popis obsahu média
├── exe .................... adresář se spustitelnou formou implementace
├── src
│   ├── impl ................................. zdrojové kódy implementace
│   └── thesis ................... zdrojová forma práce ve formátu LaTeX
└── text ................................................... text práce
    └── thesis.pdf ............................ text práce ve formátu PDF
```