CMPT 308 - Spring 2024 - Dr. Labouseur

Lab 1 – Lexer

Patrick Tyler
Patrick.Tyler1@marist.edu

February 1, 2024

1 Crafting A Compiler

1.1 Exercise 1.11 MOSS

MOSS, measure of software similarity, interrupts code for its semantic meaning rather than its raw contents. A neive approach to detect software similarity would be to to compare lines matching similar strings of text. This is very similar to how turnitin and text plagiarism checkers works. In the programming realm, avoiding similarity detection from that method would be simple: change variable names and code organization. MOSS, however, works to interrupt the code and extract its meaning. It achieves this with tokenization, creating an abstract syntax tree (AST), and semantic analysis. MOSS then compares the resulting structures with other code bases that are also interpreted in this way. Ultimately, this allows MOSS to find code which does the same thing and does it the same exact or similar way.

1.2 Exercise 3.1 Token Sequence

<id, main><close_paren><open_block>

<type, float><id, bal><end_statement>

<type, int><id, month><assignment><literal_int, 0><end_statement>

<id, bal><assignment><literal_int, 15000><end_statement>

<while><open_paren><id, bal><greater_than><literal_int, 0><open_paren><open_block>

<id printf><open_paren><literal_str, Month: %2d Balance: %10.2f n><separator><id, month><separator><id, bal><close_paren>

<id, month><assignment><id, month><add_operator><literal_int, 1><end_statement>

<close_block><close_block>

The tokens that carry extra information are identifiers (id) and all literals. Added information for observibility such as line and columns numbers for each token would also be important for a lexer to provide for future compiler steps.

2 Dragon

2.1 Exercise 1.1.4 C Target Language

The benefit of translating a language to C is that it is a well-established language with many existing compilers for a wide range of operating systems. This means that only the frontend of a compiler would need to be built out because the existing C compilers can translate it to every supported OS.

2.2 Exercise 1.6.1 Variables in Block Code

```
int w, x, y, z;
int i = 4; int j = 5;
{   int j = 7;
    i = 6;
    w = i + j;
}
x = i + j;
{   int i = 8;
    y = i + j;
}
z = i + j;
```

w is assigned i + j which is 6 + 7, so w = 13.
x is assigned i + j which is 6 + 5, so w = 11.
y is assigned i + j which is 8 + 5, so w = 13.
z is assigned i + j which is 6 + 5, so w = 11.