Assignment One

Matt Farrell

Matthew.Farrell2@Marist.edu

February 14, 2022

1 Problem One

1.1 Crafting A Compiler (Problem 1.11)

The Measure Of Software Similarity (MOSS) [SWA03] tool can detect similarity of programs written in a variety of modern programming languages. Its main application has been in detecting similarity of programs submitted in computer science classes, where such similarity may indicate plagiarism (students, beware!). In theory, detecting equivalence of two programs is undecidable, but MOSS does a very good job of finding similarity in spite of that limitation. Investigate the techniques MOSS uses to find similarity. Howdoes MOSS differ from other approaches for detecting possible plagiarism?

MOSS is a software that uses an algorithm to detect similarities between two documents. In terms of programming, renaming of variables may seem like enough to make someone else's program your own to some people (don't be one of those people), but this is where MOSS comes in handy. It can detect that the new named variables are acting the same way as the original program, despite having different names. MOSS's algorithm ignores whitespace and makes sure there are enough instances of similarities so that certain words do not get detected that shouldn't be, even though they may have the same name. Another key feature is that it will pick up on if the user reorders large blocks of code, so moving things around will not free you from getting detected for plagiarism.

2 Problem TWO

2.1 Crafting A Compiler (Problem 3.1)

The token sequence that would be produced is as follows:

> [main] - Identifier, [(] - Left Parenthesis, [)] - Right Parenthesis, [] - Left$_B racket$, $[const] - Keyword$, $[float] - Keyword$, $[payment] - Identifier$, $[=] - AssignmentOp$, $[384.00] - constant$, $[;] - Symbol$, $[float] - keyword$, $[bal] - identifier$, $[;] - symbol$, $[int] - keyword$, $[month] - identifier [=] - assignmentop$, $[0] - const$, $[;] - symbol$, $[bal] - identifier$, $[=] - assignmentop$, $[15000] - const$, $[;] - symbol$, $[while] - keyword$, $[(] - leftparenthesis$, $[bal] - identifier$, $[>] - comparisonop$, $[0] - const$, $[)] - right_p arenthesis$, $[] - left_b racket$, $[printf] - keyu$
> $rightbracket$, $[] - rightbracket$

Identifier and constant tokens would require extra information when returned, such as the variable names and constant values.

3 Dragon

3.1 Exercise 1.1.4

A compiler that translates a high-level language into another high-level language is called a source-to-source translator. What advantages are there to using C as a target language for a compiler?

C has a few advantages to being used as a target language for a compiler, including the fact that it can be easily written and is very popular. Besides that, it is a pretty basic language which is why there are so many machine code generators out there for it. It is preferable to make something complicated out of something simple, rather than vice versa.

3.1.1 Exersice 1.6.1

For the block-structured C code of Fig. 1.13(a), indicate the values assigned to w, x, y, and z.

W = 13 (6 + 7); X = 11 (5 + 6); Y = 13 (8 + 5); Z = 11 (6 + 5);