Lab One

Liam Healy

liam.healy1@Marist.edu

February 5, 2019

Crafting A Compiler: Chapter 3 Exercises

1 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. How does MOSS differ from other approaches for detecting possible plagiarism?

MOSS and other plagiarism tools are capable of catching some of the simplest plagiarism techniques, such as the renaming of variables and parameters. The way MOSS is different from other techniques is that it provides a side-by-side comparison of two programmers' code and highlights matching sections. There may be around a 10-15% match between the two which may not be the best proof of plagiarism, but if there is at least a 50% match between them, then it is very likely that the program being examined was plagiarised from the other.

2 Problem 3.1

Assume the following text is presented to a C scanner:

```
main(){
    const float payment = 384.00;
    float bal;
    int month = 0;
    bal=15000;
    while (bal>0){
        printf("Month: %2d Balance: %10.2f\n", month, bal);
        bal=bal-payment+0.015*bal;
        month=month+1;
    }
}
```

What token sequence is produced? For which tokens must extra information be returned in addition to the token code?

The token sequence produced would be: ('main', '(', ')', '{', 'const', 'float', 'payment', '=', '384.00', ';', 'float', 'bal', ';', 'int', 'month', '=', '0', ';', 'bal', '=', '15000', ';', 'while', '(', 'bal', '>', '0', ')', '{', 'printf', '(', '"Month: %2d Balance: %10.2fn"', 'month', 'bal', ')', ';', 'bal', '=', 'bal', '-', 'payment', '+', '0.015', '*', 'bal', ';', 'month', '=', 'month', '+', '1', ';', '}', '}')

When more than one pattern matches a lexeme, extra information must be returned for that token. In this case, bal, month and payment are created as id tokens, and since we can not have three ambiguous id tokens, each instance of the token must point to 'bal', 'month', or 'payment' on a symbol table.

"Dragon" textbook exercises:

3 Problem 1.1.4

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?**

The programming language C is unique because it can be used to program as a high level language and also a low level language. So in a case where the source program and target program are both high level languages, C is an ideal target language for a compiler because of its flexibility.

4 Problem 1.6.1

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

Fig. 1.13(a) : (on next page)

```
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;
```

The value assigned to 'w' is 13

The value assigned to 'x' is 11

The value assigned to 'y' is 13

The value assigned to 'z' is 11