

Lab One

Juan S. Vasquez

jvasquez1@marist.edu

February 4, 2018

1 EXERCISE 1.11

1.1 THE MEASURE OF SOFTWARE SIMILARITY (MOSS)

The MOSS is a very useful tool for detecting code plagiarism in programming-related CS courses. While the internal algorithms for how the tool actually works remains confidential, we can look at some techniques used by the MOSS to find similarity.

1.2 FINDING SIMILARITY

It is clear that the MOSS software is very sophisticated when it comes to catching plagiarism; more elaborate attempts at cheating, such as changing variable names and switching the order of functions in the program stand out just as much as duplicate code. This is mostly because there are three very important properties that MOSS uses and every plagiarism-detection program should have: *whitespace insensitivity*, *noise suppression*, and *position independence*.

1.2.1 WHITESPACE INSENSITIVITY

Matches cannot be affected by things like capitalization, punctuation, spacing, etc. In this case, variable names changes should also be ignored.

1.2.2 NOISE SUPPRESSION

Short matches between two documents, such as small functions, are too vague. The matches sections must be large and meaningful enough to warrant detection.

1.2.3 POSITION INDEPENDENCE

The rearranging, extending, and shortening of matches should not affect match level if the plagiarized areas remain significant.

2 EXERCISE 3.1

2.1 TOKEN SEQUENCE

Line 1 - ID(main), LPAREN, RPAREN, LBRACE
Line 2 - CONST, VAR_TYPE(float), ID(payment), ASSIGN, FLOATNUM(384.00), SEMICOLON
Line 3 - VAR_TYPE(float), ID(bal), SEMICOLON
Line 4 - VAR_TYPE(int), ID(month), ASSIGN, INTNUM(0), SEMICOLON
Line 5 - ID(bal), ASSIGN, INTNUM(15000), SEMICOLON
Line 6 - WHILE, LPAREN, ID(bal), GT, INTNUM(0), RPAREN, LBRACE
Line 7 - PRINT, LPAREN, QUOTE, CHAR(M), CHAR(o) ... CHAR(n), QUOTE, COMMA, ID(month),
COMMA, ID(bal), RPAREN, SEMICOLON
Line 8 - ID(bal), ASSIGN, ID(bal), MINUS, ID(payment), SUM, FLOATNUM(0.015), MULTIP, ID(bal),
SEMICOLON
Line 9 - ID(month), ASSIGN, ID(month), SUM, INTNUM(1), SEMICOLON
Line 10 - RBRACE
Line 11 - RBRACE

2.2 TOKENS WITH EXTRA INFORMATION

ID tokens require additional information, as well as INTNUM and FLOATNUM tokens.

3 EXERCISE 1.1.4

3.1 C AS A TARGET LANGUAGE

The advantage to using C as a target language for a compiler is that C is a very commonplace language.

4 EXERCISE 1.6.1

4.1 ASSIGNED VALUES

w is assigned 13
x is assigned 11
y is assigned 13
z is assigned 11