

Design of Compilers: Lab 1

Exercises 1.11 & 3.1

Andrew Mathew

January 31, 2021

CMPT 432 SPRING 2022

Crafting a Compiler: 1.11 How does MOSS differ from other approaches for detecting possible plagiarism?

MOSS has become a very popular tool, especially in academia, that teachers and professors alike use to scan student submitted code for potential plagiarism. In a hasty attempt to embellish code, a student may think to change variable names and add extra white spaces or even add new lines to make the code look different from the original copy. If a professor were to check the code manually by brute force, these shoddy attempts may fail to prevent the plagiarized work from slipping through. The main reason why MOSS is so effective at detecting plagiarism is because of 'Winnowing' (Sigmod 2003).

MOSS's is more efficient than other approaches because of its specific use of the winnowing algorithm. Winnowing is a local fingerprinting algorithm that uses a set of hashes to compare sub-strings between documents allows for easier and more efficient comparisons. Documents that contain a large number of matching fingerprints are flagged (Yang 2019).

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

```
main()
const float payment = 384.00;
float bal;
int month = 0;
bal=15000;
while (bal<0)
printf("Month: bal=bal-payment+0.015*bal;
month=month+1;
```

Token Sequence:

```
(FUNC-ID)
(FUNC-L-Paren)
(FUNC-R-Paren)
(FUNC-L-BRACE)
(CONST)
(FLOAT)
(ID)
(ASSIGN)
(FLOAT-CONST)
(SEMI-COL)
(FLOAT)
(ID)
(SEMI-COL)
```

(INT)
 (ID)
 (ASSIGN)
 (INT-CONST)
 (SEMI-COL)
 (ID)
 (ASSIGN)
 (INT-CONST)
 (SEMI-COL)
 (WHILE)
 (L-PAREN)
 (ID)
 (LOGICAL OP)
 (INT-CONST)
 (L-BRACE)
 (PRINT)
 (L-PAREN)
 (OPEN-QUOTE)
 (CHAR-M) (CHAR-o) (CHAR-n) (CHAR-t) (CHAR-h) (CHAR-.) (CHAR-) (CHAR-PRCNT)
 (CHAR-d) (CHAR-SPCE) (CHAR-B) (CHAR-a) (CHAR-l) (CHAR-a) (CHAR-n) (CHAR-
 c)
 (CHAR-e) (CHAR-.) (CHAR-) (CHAR-PRCNT) (CHAR-1) (CHAR-0)
 (CHAR-.) (CHAR-2) (CHAR-f) (CHAR-FWD-SL) (CHAR-n)
 (CLOSE-QUOTE)
 (COMMA) (ID) (COMMA) (ID) (R-PAREN) (SEMI-COL)
 (ID) (ASSIGN) (ID) (SUB-OP) (ID) (ADD-OP) (FLOAT-CONST) (MULT-OP) (ID) (SEMI-
 COL)
 (ID) (ASSIGN) (ID) (ADD-OP) (INT-CONST) (SEMI-COL)
 (R-BRACE)
 (R-BRACE)

Dragon: 1.1.4 What advantages are there to using C as a target language for a compiler?

Compilers for C are widely available, so choosing C as a target language is more feasible.

Dragon: 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 x = 11

y = 13 z = 11

Works Cited

<https://yangdanny97.github.io/blog/2019/05/03/MOSS>

<http://theory.stanford.edu/~aiken/publications/papers/sigmod03.pdf>