

Crafting a Compiler

1.11

MOSS employs a technique called document fingerprinting, which divides a document or a body of text into substrings. The substrings are then hashed, before the tool selects some substrings to represent the document's fingerprint, an essential technique to detect copies. Additionally, MOSS automatically eliminates text that seems to be shared across every document, like libraries. In comparison to other plagiarism-detection tools, MOSS uses an efficient document fingerprinting algorithm ("wlnnowing") that is "33% of the lower bound" and supports a relatively greater number of languages through a web interface.

3.1

Token sequence:

1. T_main, main
2. T_leftParam, (
3. T_rightParam,)
4. T_openBrace, {
5. T_const, const
6. T_float, float
7. T_id, payment
8. T_assign, =
9. T_digit, 3
10. T_digit, 8
11. T_digit, 4
12. T_period, .
13. T_digit, 0
14. T_digit, 0
15. T_semi, ;
16. T_float, float
17. T_id, bal
18. T_semi, ;
19. T_int, int
20. T_id, month
21. T_assign, =
22. T_digit, 0
23. T_semi, ;
24. T_id, bal
25. T_assign, =
26. T_digit, 1
27. T_digit, 5
28. T_while, while
29. T_leftParam, (

30. T_id, bal
31. T_greaterThan, >
32. T_digit, 0
33. T_rightParam,)
34. T_openBrace, {
35. T_printf, printf
36. T_leftParam, (
37. T_quote, "
38. T_char, M
39. T_char, o
40. T_char, n
41. T_char, t
42. T_char, h
43. T_char, :
44. T_whitespace
45. T_char, %
46. T_char, 2
47. T_char, d
48. T_whitespace,
49. T_whitespace,
50. T_char, B
51. T_char, a
52. T_char, l
53. T_char, a
54. T_char, n
55. T_char, c
56. T_char, e
57. T_char, :
58. T_whitespace
59. T_char, %
60. T_char, 1
61. T_char, 0
62. T_char, .
63. T_char, 2
64. T_char, f
65. T_backSlash
66. T_char, n
67. T_quote, "
68. T_comma, ,
69. T_id, month
70. T_comma, ,
71. T_id, bal
72. T_rightParam,)

- 73. T_semi, ;
- 74. T_id, bal
- 75. T_assign, =
- 76. T_id, bal
- 77. T_minus, -
- 78. T_id, payment
- 79. T_plus, +
- 80. T_digit, 0
- 81. T_period, .
- 82. T_digit, 0
- 83. T_digit, 1
- 84. T_digit, 5
- 85. T_multiply, *
- 86. T_id, bal
- 87. T_semi, ;
- 88. T_id, month
- 89. T_assign, =
- 90. T_id, month
- 91. T_plus, +
- 92. T_digit, 1
- 93. T_semi, ;
- 94. T_closeBrace, }
- 95. T_closeBrace, }

Tokens that are identifiers, float constants, and int constants must have extra information be returned in addition to the token code.

Dragon

1.1.4

Many compilers today support C, and thus, they can compile to most hardware.

1.6.1

w = 13
x = 11
y = 13
z = 11