Marist College
Compilers

Department of Computer Science
Mason Nakamura

Lab 1.0

CRAFTING A COMPILER

Exercise 1.11 (Info from here)

The MOSS algorithm implements techniques to ignore white space, be unaffected by the renaming of variables, and ignore the position of the code. It uses a technique called document fingerprinting which uses a set of hashes to compare documents for similarity rather than comparing the substrings within the documents which saves a lot of time. Specifically, MOSS uses a fingerprinting algorithm called winnowing which produces a sequence of hashes from a stream of substrings. More specifically, where ϕ is the hash function and H_0 is the set of characters to be hashed such that $\phi : h \longrightarrow \mathbb{R}$, where $h \in H_0$. The structure of H_0 is up to the user. For example, one may want to have $|h| = 5$ while iterating over every other character of the string, $\cup_{h \in H_0} h_i$. Even more specifically, let H_1 be the set of hashed substrings such that

$$H_1 = \{\phi(h)\} \ \forall h \in H_0 \ s.t. \ |h| \text{ is predetermined.}$$

After H_1 is found, we then stochastically collect n number of fingerprints from H_1 and compare that to the n fingerprints of other documents.

The difference between the MOSS algorithm and other plagiarism detection approaches is that MOSS uses a function to output a stream of real numbers for comparison of other documents' fingerprints. Contrarily, other techniques use the words or sentences themselves as a means of comparison for similarity without applying a transform or function before comparison.

Exercise 3.1 Token Sequence:

[keyword, separator, separator,separator,
type, type, id, assign_op, number, separator,
type, id, separator,
type, id, assign_op, number, separator,
id, assign_op, number,
keyword, separator, id, greater_op, number, separator, separator,
keyword, string, separator, id, separator, id, separator, separator, separator,
id, assign_op, id, minus_op, id, add_op, number, mult_op, id, separator,
id, operator, id, operator, number, separator,
separator,
separator]

Extra information must be appended to the tokens identified as id, number, keyword, and separator which will be used after the parser.

DRAGON

Exercise 1.1.4 Cited from here.

Since there is a C compiler on almost every device and it is heavily optimised, translating the source language into C allows for compatibility across most devices. C is also low level and is easy to generate with minimal interpretation errors.

Exercise 1.6.1

$w := 13$
$x := 9$
$y := 13$
$z := 9$