Justin Domingue – 260588454

Daniel Pham – 260526252

Seguei Nevarko – 260583807

Final Report

COMP 520

Work presented to Prof. Laurie Hendren

McGil University

Thursday, February 26, 2016

# Language and tools

We decided to use flex/bison because that’s what we used in the assignment and while SableCC seems like a promising avenue, we were already comfortable with flex/bison.

# Scanner

As discussed in the Language and Tools, we used flex to generate a lexical analysis of the input Golite code. Most rules are basic regex similar to what we have seen in class and done in the assignment. However, some intricacies of Go(lite) needed to be parsed at a finer level. To do so, we used a regex to capture the first few characters (“/\*” for comments, “ ’ ” and “ ` “ for comments, interpreted strings and raw strings). We then analyzed the input character by character, checking for allowed escapes and saving to *yytext* as needed. This implementation was suggested to us by the ANSI C grammar.

# Parser

The parser was completely handwritten from scratch in Bison following the Go(lite) specification document. It was rather similar to what we did in the assignments. It should be noted that while we used to allow optional semi-colon, this feature was turned off to comply with the Golite specification. Also, parenthesized types are allowed, to be inline with the Go specs (although the reference compiler does not allow them). <https://golang.org/ref/spec#Types>

To make the grammar simpler, we allowed over-generation to mitigate reduce/shift-reduce conflicts.

1. In function calls, the id is an expression to allow nested parentheses.
2. Function declaration with returns do not need to end with a terminating statement
3. In function declarations, the number of arguments in a return statement doesn’t have to match the number of return arguments defined in the function signature
4. Continue and break statement can be anywhere in the code
5. Expression statements can be anything.
6. Post conditions for for loops can contain variable declarations statements
7. In short variable declarations, a list of identifiers cannot be distinguished from a list of expression, leading to shift reduce conflicts. Therefore, we use an exp\_list on the left side instead of id\_list.
8. Left-hand side of assignment, increment and decrement statements doesn’t have to be an lvalue
9. A switch statement can have multiple default cases
10. In a variable declaration and assignments, the number of identifiers doesn’t have to match the number of expressions
11. Division by 0
12. ‘\_’ can be used as value

# Pretty Printer

The pretty printer follows has a standard implementation. Runes are escaped. It should be noted that we support indentation.

# Tree

Serguei built a tool that read the grammar file and automatically generated the code for the tree, the pretty printer and the updated grammar file (where the code for the tree generation was inserted). It didn’t work so well because the generated tree was following the grammar too closely. We ended up with a concrete syntax tree which was much more difficult to use than an abstract syntax tree. There was an idea of eliminating all the nodes that didn’t carry information to make it more like an AST but we ended up just writing it up by hand just like in the assignments.

Had we been able to write the tool to convert our CFG to an AST automatically (without specifying any node merging rules) we would have had the ability to rewrite the grammar with the greatest of ease without needing to worry about rewriting the code of the tree.

A particularity of our abstract syntax tree is that we have a node that is that our EXP (expression) node has a kind for every operation (binary, unary, or other) but there is a single structure for binary expressions and a single structure for unary expressions. It made the AST shorter and more readable.

# Weeding

Every weeding prospects identified in the Parser section were dealt with in this milestone. However, the following over-generations have a few particularities:

* 1. (Function calls) will be weeded out by the type checker
* 11. (Division by 0) used to be weeded out – but after talking with Vincent, we simply allow the possibility of having runtime errors
* 12. (Blank ids) can only be used on the left of assignments and short variable declarations. They also aren’t allowed as function ids since Golite does not support function literals. The same reasoning is applied to struct declarations.

Note that return statements in function declarations (2. and 3.) are disallowed although the reference compiler allows them. They follows the Golang spec: <https://golang.org/ref/spec#Terminating_statements>

# Symbol Table

The symbol table is slightly different than what was shown in class. The concept of layers is exactly the same and it has the same functions, but instead of using hash tables we are using red-black trees. The idea was to try something new. Unlike hash tables that need a good hash function to be efficient, a red-black tree is always fast because the runtime of all searches and inserts are at most O(log(n)) where n is the number of elements in the tree.

# Symbol Pass

Symbols are typed as: variables, type aliases, functions or inferred where inferred has its type determined at type checking. Variables and type alias pointers point directly to the type with with they are associated.

# Scoping Rules

These follow the GoLite specifications.

# Type Check

Again, these follow the GoLite specifications.

Some Rules:

* **Binary operations (particular to our implementation)**:

Types must be the same on both sides of the operator. We do not allow any constant coercion, meaning the integer 1 is never automatically converted to a float64 in `1.2 + 1`. This has a side effect that increments and decrements are only supported on integers.

bin\_op\_nocoercion.go

* Function types and returns:

The type of the return statement has to be assignable to the return type of the enclosing function.

return\_mismatch.go

* Variable assignments:

The expression has to be assignable to the *lvalue* on the left hand side.

assign\_string.go

* Type casts:

Expressions can only be converted to “int, float64, bool, rune, or a type alias that maps to one of those four”

conversion\_nonbasic.go

* Append:

First arguments must be a slice.

append\_array.go

* Structs:

Are considered equivalent if they have the same fields of the same type in the same order.

struct\_mismatch.go

* Printing:

We only allow printing of basic types (no arrays or slices)

print\_nonbasic.go

# Computation Intensive Programs

**primeNumbers.go:**

This program generates the **i**’th prime number where **i** is a number that is hardcoded in a variable. We will change this number to be able to choose the computation time that we want.

The algorithm is quite simple. For every number we check if it can be divided by the previous prime numbers that we found. We only check with prime numbers that are less or equal to the upper bound of sqrt(number). Since we don’t have the sqrt() function we just keep track of it using a variable and its square value.

**queenPuzzle.go:**

This is the famous queens problem where we have a chessboard of size **N**x**N** and we have to place **N** queens on it such that no two queens attack each other (horizontally, vertically or diagonally).

In our program this problem is solved using simple recursion. To have more control on the computation time we just set the variable **N** to the size that we want.

# **Code** Generation Status Report

**Target language:** C++

We wanted a fast, highly optimized target language. By choosing C++, the code generation is quite simple. In fact, Go-lite and C++ share many constructs.

The the code generation for the following subset of Go-lite has been implemented:

* C++ header
* Variable declarations (including multiple variable declarations with matching expressions)
* Type definitions
* Types (C++ equivalent: int, double, bool, char, std::string), structs, slices and arrays.
* Function declarations including signature, arguments and body
* The following statements: empty, block, return, if, if-else, for (including 3-part for loop, while loop and infinite loop), break, continue, assignment, print and println.
* No binary expressions.

# Team organization

We started working on the first milestone before our repository was set up. Here is a quick overview of the work done by each member (do note that peer programming was extensively used throughout this milestone):

Dan:

* Lexer
* Parser
* Weeder
* Type checker
* Symbol table printing
* Code generation

Justin

* Parser
* Tree
* Weeder
* Symbol pass
* Code generation

Serguei

* Investigated automating tree generation
* Pretty printer + AST tree printer (for debugging purposes)
* Symbol table using red-black tree
* Intensive programs written in Go