CMPUT 415 - Project Checkpoint 2
Program Documentation

Matthew Low
Anthony Galati
Stevan Clement
Mike Bujold

9 November 2011

The man page is available with the command *less ./documentation/pal.1*

1 Introduction

The following is the documentation for our PAL compiler. The most recent improvements are included at the end of this file.

2 Handling of Lexical Units

The lex program works in the usual way to match tokens, the only interesting part being how it handles comments that are allowed to spread across multiple lines. To do this, the lexer switches into a 'comment' state where it grabs everything until it sees the end of the comment. This state change was deemed necessary for the line counting to continue while observing comments.

So far we are not doing anything with strings between single quotes, just acknowledging them. Garbage characters are collected silently and lex calls an error function which then notifies the user.

3 Syntax Error Reporting Strategies

Syntactic errors were detected by placing the 'error' token in reasonable positions. The yyerror function then uses the information we store holding the line and column position to provide a reasonable message to the user.

Since line endings have little meaning in the PAL language, we don't use the yerrok macro, as there is no way to match an error to a whole line (to do so may mean an error matches the whole program). As such, if we encounter an error, we don't want to output a bunch of other errors that are immediately caused by the first.

4 Resolving Conflicts in the Bison Grammar

The conflicts in the bison grammar were mainly the result of ambiguities created by multiple similar paths to the grammar's atoms. The most serious example of this was how both "type"

and "expr", two similar rules, reduced along separate paths to simple_type. We were able to remove this ambiguity from the grammar by removing the rule for reducing to simple_type from those available to expr, and instead replacing it with the applicable atoms.

Another issue with the grammar was the possibility of declaring an array with type anonymous enum. We replaced this rule with two new rules, one of which allows for instantiation of strings and one which allows only previously declared variables as a type.

5 Symbol Table

The symbol table entries follow the class notes in their design. The symbol table itself is an array of linked lists of symbol table entries. The array is size 17 to correspond to the display register values in ASC and allows for a base level of predefined values which we have hard coded into the compiler.

The symbol table includes a few basic functions as an interface to it. PushLevel and PopLevel are included to change the scope of the program - these are called in Bison when we begin to parse inside a new procedure or function. If we go over the allowed number of levels the compiler will exit gracefully since we can't go any further with the display registers. Various functions have been created to add items to the symbol table - one for each of the main types (var, const, type, func, proc). We have included a symbol table printing function that is currently a work in progress.

The symbol table consists of sym_rec structures, which are at the main level of the symbol table. These structures have a class number (eg. OC_TYPE) which defines a field in a union containing more class specific information. The most interesting of these is the struct type_desc including the types of allowed types. This field is important to comparison later on, as we are able to pull out of it the address of a pointer that was created when a type as declared to compare it with other types. Every variable and constant is linked to a type, so through many levels of redirection we are able to get all of the information we require about an item in the symbol table.

6 Semantic Error Checking and Symbol Table Usage

All semantic error checking is handled inside of the Bison grammar - there are a few helper functions in semantics.c and symtab.c, but the Flex code does none, it simply returns tokens. Because of this, we were able to remove INT, REAL, BOOL, and CHAR from the grammar file, allowing Flex to pass Bison just the ID token with the identifier name that it parsed. All integer constants are handled with INT_CONST and real constants with REAL. In the grammar rules for declaration, most variables are handled by receiving an ID token and doing a local lookup with the string that was passed with it.

Populating the symbol table is straight forward for most classes. During function or procedure parsing the expected parameters are added to the next level of the symbol table as well as copied over to a linked list of expected parameters attached to the entry for the function or procedure at the current level. While parsing the parameters, if we find that one of them causes an error (if it has already been declared or is of an invalid type), we put it in the parameter list associated with the function or procedure and give it type OC_ERROR. Enumerated types also have a different handling. As the grammar parses them they are collected in a linked list and sent back through the grammar using $$ until a type needs to be defined or a variable needs a list of associated scalar types. Consider the following expression:

var room : (chair, table, door);

In this case, the scalar list would be parsed, and when it was finally realized that we were assigning a variable to a type of this list, a new anonymous type would be created with no name, only available by checking the type of the variable.

Type checking is currently done for expressions by simply grabbing the type of whatever constant or variable we have seen and passing it through $$ variable. If an error is seen at any point in an expression evaluation, an error is reported and $$ is set to NULL. We decided to do this because sending an arbitrary type through an expression would cause more errors than it would detect.

To check functions and procedures, since Bison sees the name before the parameter list, we save the important parts about the function in the struct plist_t and pass it through $$ as we parse the parameters. This way we can know about missing parameters and incorrect ones, and give functions a proper return type.

7 References Cited

Only the class notes, class discussions, Bison/Flex manuals, and a few helpful O'Reilly guides were used in completing this checkpoint.