

Memoria Prácticas Compiladores

Autores:
<i>Marquina Meseguer, Alfredo - alfredo.m.m@um.es</i>
<i>Martínez Prior, Daniel - daniel.m.p@um.es</i>

ÍNDICE

Análisis Léxico	3
Macros	3
Palabras reservadas y Símbolos	4
Comentarios	4
IDs y Tipos	5
Errores	5
Análisis Sintáctico	5
Precedencia de operadores y conflicto	6
Análisis Semántico	6
Errores semánticos	6
Funciones añadidas a Lista Símbolos	7
Generación de Código	9
Pruebas	9
Ejecución de MiniC	12
Conclusión	13

Análisis Léxico

El análisis léxico lo podemos separar en las diferentes definiciones de lexemas a encontrar y sus tipos: macros, comentarios, palabras reservadas, id, tipos y símbolos.

El análisis léxico se encarga de devolver tokens para el análisis sintáctico, cada token a sido definido en un fichero generado por Bison llamado miniC.tab.h, aquí se encuentran como una enumeración yytokentype, donde cada uno tiene un número asignado a partir de doscientos cincuenta y ocho, con cuatro tokens creados por Bison.

```
enum yytokentype
{
    YYEMPTY = -2,
    YYEOF = 0,
    YYerror = 256,
    YYUNDEF = 257,
    EMPIEZA = 258,
    LEER = 259,
    ESCRIBIR = 260,
    A_PAREN = 261,
    C_PAREN = 262,
    SEMICOLON = 263,
    COMA = 264,
    ...
    /* "end of file" */
    /* error */
    /* "invalid token" */
    /* EMPIEZA */
    /* LEER */
    /* ESCRIBIR */
    /* A_PAREN */
    /* C_PAREN */
    /* SEMICOLON */
    /* COMA */
}
```

Macros

Las macros utilizadas son bastante simples y sus nombres resultan autoexplicativos, seleccionando en la mayor parte de casos simplemente la selección de los valores ASCII que corresponden a qué secciones. Por ejemplo, la macro dígito es el rango de los caracteres entre cero y nueve, y pánico es cualquier carácter que no se encuentra incluido en la gramática.

```
digito      [0-9]
letra       [a-zA-Z]
entero      {digito}+
panico      [^A-Za-z0-9();,=+-{}_~$]
```

Palabras reservadas y Símbolos

Las palabras reservadas también constituyen otra parte del léxico que resulta muy sencillo. Se trata de una serie de términos que deben tener un comportamiento propio, por esta razón las variables creadas más adelante no deberían poder compartir nombre con ninguna de estas palabras.

De la misma manera se obtienen los símbolos (multiplicación, suma, apertura de paréntesis, etc.) que cumplen con las mismas características de las palabras reservadas.

```
// Ejemplo de palabra reservada y de símbolo
void                                     return EMPIEZA;
"+"                                     return MAS_OP;
```

Comentarios

La identificación de líneas de comentario se ha realizado comprobando mediante una expresión regular. Esta comprueba que existen dos caracteres '/' seguidos y de hacerlos considera que el resto de la línea es un comentario.

```
/* Comentario: 1 línea */
"//"(.*)[\n]                ;
```

Para el comentario multilínea ha resultado un poco más difícil ya que requiere el uso de condiciones de contexto para poder detectarlos. Cuando se identifica los caracteres '/' y '*' juntos, se inicia la etiqueta COMENTARIO. A partir de este momento solo se comprobarán reglas que tengan incluida la etiqueta COMENTARIO hasta que se vuelva a iniciar el inicio que es de valor cero, esto llegará al encontrarse los mismo caracteres juntos pero en orden contrario. Con esto conseguimos que cualquier texto que se encuentre entre estos dos marcadores no sea compilado.

```
%x COMENTARIO
...

"/*"
<COMENTARIO>"*/"
<COMENTARIO>.
<COMENTARIO>\n

{yymore(); BEGIN COMENTARIO;}
BEGIN 0;
;
;
```

IDs y Tipos

Los tipos y los ids son elementos que se han conseguido de manera muy similar, ya que además de los tipos se debe identificar el lexema que este contiene, por lo que hay que devolverlo. Hemos pasado el valor del lexema o nombre de la ID a través de la variable `yyval` normalmente utilizada por Bison.

`yytext` es la variable que contiene la cadena de texto reconocida, por lo que al utilizar la función `strdup` con ella como parámetro estamos creando una copia de su contenido. Esta copia es asignada `yyval.lexema` para uso posterior del análisis sintáctico.

Al guardar el valor del id o lexema en la parte de lexema del struct `yyval`, cuando devolvemos uno de los tokens que debería devolver una cadena, el programa comprueba su valor en este sitio en específico.

```
(_|\\$|{letra})({letra}|{digito}|_|\\$){0,15}    {yyval.lexema = strdup(yytext);return ID;}
entero                                           {yyval.lexema = strdup(yytext);return NUM;}
"\"([^\n\\]|\\.)*\"                               {yyval.lexema = strdup(yytext);return STRING;}
```

Errores

Al igual que los comentarios, se ha necesitado del uso de etiquetas para implementar el modo pánico a este programa.

Primero se ha identificado el inicio de un error aceptando cualquier carácter que no se haya metido en ninguna de las reglas establecidas, es decir, la regla de error se encuentra al final de la lista de reglas. Una vez empezado el error léxico se continúan considerando todos los caracteres que no se entran dentro de los aceptados, como se ha mencionado antes, estos caracteres se encuentran dentro de la macro pánico. Finalmente, cuando se encuentra un carácter que sí está dentro de los aceptados se termina el error y se informa al sintáctico que ha ocurrido para que no compile.

```
<ERROR>{panico}  {yymore();}
<ERROR>.         {yyless(yylen-1);printf("Reconocido error en
                     linea %d: %s.\n",yylineno, yytext); BEGIN 0;
                     return ERR;}
.               {yymore(); BEGIN ERROR;}
```

snappify.com

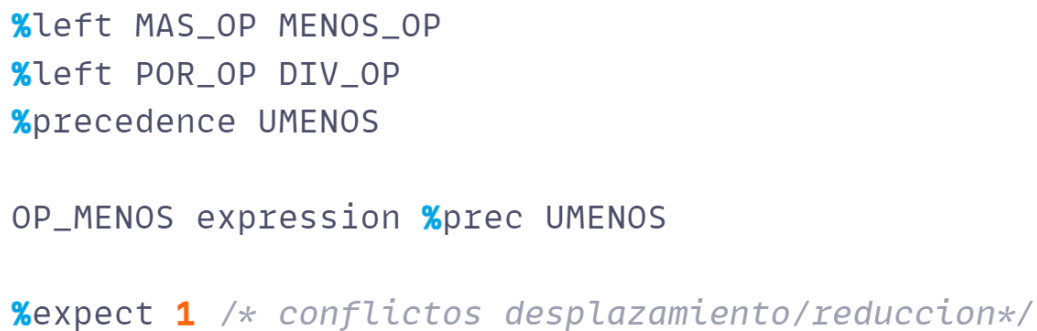
Análisis Sintáctico

El propósito de esta parte del compilador es validar y comprender la estructura gramatical del programa según las reglas de nuestro lenguaje de programación miniC y construir un árbol sintáctico que represente la estructura y las relaciones que existen en el programa.

Esta parte del programa ha sido en su mayor parte proporcionada por el enunciado así que los cambios requeridos para que sea funcional han sido pocos.

Precedencia de operadores y conflicto

Se evita la ambigüedad en el program a partir de expresar la relaciones de precedencia entre los operadores "+", "-", "*" y "/". Se quería especificar que los dos primeros operadores tienen menos preferencia que los dos segundos, con los miembros de los grupos compartiendo precedencia entre ellos. Para ello, se ha otorgado una asociatividad izquierda superior a "+" y "-" sobre "*" y "/". Sin embargo, seguía existiendo cierta ambigüedad pues el símbolo de resta y que indica la negatividad de un número son el mismo, por lo que se ha tenido que especificar precedencia ya no entre símbolos sino entre tokens del UMENOS sobre el resto, incluyendo OP_MENOS, para que se compruebe este sobre el resto. Finalmente se ha especificado a Bison el número de conflictos desplazamiento-reducción se puede encontrar en el fichero.



```
%left MAS_OP MENOS_OP
%left POR_OP DIV_OP
%precedence UMENOS

OP_MENOS expression %prec UMENOS

%expect 1 /* conflictos desplazamiento/reduccion*/
```

Análisis Semántico

El análisis semántico tiene como objetivo almacenar los diferentes tipos de datos en una tabla de símbolos y detectar errores semánticos. Para ello, utilizaremos una lista simple como estructura de datos. En las reglas de producción que involucren declaraciones, asignaciones, lecturas o impresiones, se realizará el procesamiento semántico. Se insertará cada variable, constante o cadena al final de la lista de símbolos, almacenando su nombre y tipo. Para las cadenas, se utilizará un campo adicional llamado "valor" para mantener el número de cadena en el código.

Errores semánticos

Los errores semánticos posibles ya se han sido definidos por el código proporcionado de miniC.y estos se tratan de los casos:

- Un identificador ya se ha declarado, por lo que no se puede volver a declarar.

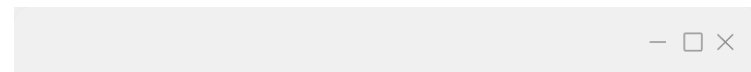
- Un identificador no ha sido declarado, por lo que no se puede llamar.
- Un identificador declarado como constante y se está intentando modificar.

daniel.martinez.prior@gmail.com Inserta foto aquí de los errores esos del if-else en el miniC.y

Funciones añadidas a Lista Símbolos

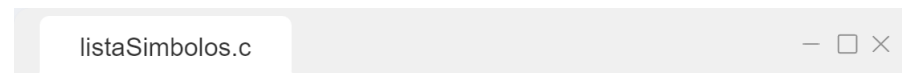
Se han implementado una serie de funciones para completar el código proporcionado listaSimbolos. Estas funciones son: perteneceTablaS, anyadeEntrada, esConstante e imprimeTabla.

Esta primera función trata de comprobar la pertenencia de un elemento a la tabla de símbolos. Para ello, basta con identificar su posición en la lista y comprobar que no se encuentra en la última, pues es el valor devuelto en caso de no estar presente en la estructura de datos.



```
bool perteneceTablaS(Lista lista, char *nombre) {
    PosicionLista pos = buscaLS(lista, nombre);
    PosicionLista final = finalLS(lista);
    return pos != final;
}
```

En la segunda, se añade una entrada a la lista de símbolos, para ello, debemos insertar un struct Simbolo al final de la lista.



```
void anyadeEntrada(Lista lista, char *nombre, Tipo tipo) {
    PosicionLista ultimaPosicion = finalLS(lista);
    Simbolo nuevoSimbolo = {nombre, tipo};
    insertaLS(lista, ultimaPosicion, nuevoSimbolo);
}
```

En la tercera comprobamos que un elemento sea constante, esta función se utiliza para que cada vez que se asigna o edita una variable podamos saber si se trata de una constante. La forma de comprobarlo es buscando en la lista la posición de este elemento. De encontrarse, se debería poder comprobar el valor del tipo el cual puede o no ser CONSTANTE.

listaSimbolos.c

— □ ×

```
bool esConstante(Lista lista, char *nombre) {
    Simbolo s;
    PosicionLista pos = buscaLS(lista, nombre);
    if (pos != finalLS(lista)) {
        s = recuperaLS(lista, pos);
        return (s.tipo == CONSTANTE);
    }
}
```

Finalmente la función que imprime la declaración de los datos a partir de la información guardada en una lista de símbolos. Se encuentra compuesta por la impresión de la cabecera y un par de bucles, donde el primero declara las cadenas y en el segundo, el resto de tipos que resultan ser solo números.

listaSimbolos.c

— □ ×

```
void imprimirTablaS(Lista lista) {
    printf("#####\n");
    printf("# Seccion de datos\n");
    printf("\t.data\n");
    printf("\n");
    // Obtenemos la posición del primer elemento
    PosicionLista pos = inicioLS(lista);
    Simbolo simbolo;

    // Recorremos la lista hasta llegar al final
    while (pos != finalLS(lista)) {
        // Obtenemos el simbolo de la posición actual
        simbolo = recuperaLS(lista, pos);
        /* Imprimimos la información del simbolo con
           formato según su tipo. */
    }
}
```



```

    if (simbolo.tipo == CADENA) {
        printf("$str%d:\n", simbolo.valor);
        printf("\t.asciiz %s\n", simbolo.nombre);
    }
    // Avanzamos a la siguiente posición
    pos = siguienteLS(lista, pos);
}

pos = inicioLS(lista);
while (pos != finalLS(lista)) {
    simbolo = recuperaLS(lista, pos);
    if (simbolo.tipo != CADENA) {
        printf("_%s:\n", simbolo.nombre);
        printf("\t.word 0\n");
    }
    pos = siguienteLS(lista, pos);
}
printf("\n")
}

```

Generación de Código

Finalmente terminamos generando el programa sin la sección de datos pues ha sido los último que se ha realizado durante el análisis semántico. Si este análisis ha transcurrido sin encontrar ningún error, se procederá a generar el código.

Para la representación de este código, se ha utilizado una lista simple. Cada uno de los nodos contendrá una operación a realizar en MIPS, cada nodo cuenta con cuatro campos, uno para la operación y tres más para los registros utilizados en la operación, tres es el número máximo de registros a utilizar y ciertas operaciones pueden simplemente utilizar uno de ellos. A su vez, cada operación realizada en el lenguaje miniC se traducirá a una o más operaciones de MIPS por lo que serán uno o varios nodos de esta lista.

Pruebas

Para la entrada:

```

void main() {
    var x = 30, y = -10;
    var b = 3;
    var z;
    z = x + y;
    print "hola \"esto es\" una prueba", "\n";
}

```

```
/* Esto es
un comentario
multilinea
*/
```

```
// Esto es un comentario de una sola linea
```

```
if (b) print "b es true porque su valor es distinto de 0\n";
while(b) {
    b = b-1;
    print"El valor de b va disminuyendo en", "cada iteración\n";
}
if(b) print "hola";
else print "El valor de b es 0\n";
}
```

La salida es:

```
#####
```

```
# Seccion de datos
```

```
.data
```

```
$str0:
    .asciiz "hola \"esto es\" una prueba"
$str2:
    .asciiz "hola \"esto es\" una prueba"
$str0:
    .asciiz "\n"
$str4:
    .asciiz "\n"
$str0:
    .asciiz "b es true porque su valor es distinto de 0\n"
$str6:
    .asciiz "b es true porque su valor es distinto de 0\n"
$str0:
    .asciiz "El valor de b va disminuyendo en"
$str8:
    .asciiz "El valor de b va disminuyendo en"
$str0:
    .asciiz "cada iteración\n"
$str10:
    .asciiz "cada iteración\n"
$str0:
    .asciiz "hola"
$str12:
    .asciiz "hola"
$str0:
    .asciiz "El valor de b es 0\n"
$str14:
    .asciiz "El valor de b es 0\n"
```

```

_x:
    .word 0
_y:
    .word 0
_b:
    .word 0
_z:
    .word 0

#####
# Seccion de codigo
    .text
    .globl main
main:
    li $t0, 30
    sw $t0, _x
    li $t0, 10
    neg $t1, $t0
    sw $t0, _y
    li $t0, 3
    sw $t0, _b
    lw $t0, _x
    lw $t2, _y
    add $t0, $t0, $t2
    sw $t0, _z
    li $v0, 4
    la $a0, $str2
    syscall
    li $v0, 4
    la $a0, $str4
    syscall
    lw $t0, _b
    beqz $t0, $l1
    li $v0, 4
    la $a0, $str6
    syscall
$l1:
$l2:
    lw $t0, _b
    beqz $t0, $l3
    lw $t3, _b
    li $t4, 1
    sub $t3, $t3, $t4
    sw $t3, _b
    li $v0, 4
    la $a0, $str8
    syscall
    li $v0, 4

```

```

        la $a0, $str10
        syscall
        b $l2
    $l3:
        lw $t0, _b
        beqz $t0, $l4
        li $v0, 4
        la $a0, $str12
        syscall
        b $l5
    $l4:
        li $v0, 4
        la $a0, $str14
        syscall
    $l5:

#####
# Fin
        jr $ra

```

Como podemos ver hemos encontrado un error por el que se nos duplican las cadenas de texto, pero no hemos tenido el suficiente tiempo para arreglarlo.

Ejecución de MiniC

A partir del fichero `makefile` y el código fuente se puede obtener el compilador de `miniC`. La forma sencilla de ejecutarlo es con el comando *make*, que ejecuta la primera instrucción que te devuelve una ejecutable llamado *miniC* que te permite compilar programas de `miniC` a MIPS si se pasa el fichero fuente como primer parámetro. Además, si se quiere ejecutar directamente, se puede ejecutar *make run* que intentará alcanzar y ejecutar la condición de etiqueta *run* la cual ejecuta *miniC* usando el fichero de prueba proporcionado llamado *p*.

makefile

```

miniC: lex.yy.c miniC.tab.c main.c listaSimbolos.c listaCodigo.c
    gcc -g lex.yy.c miniC.tab.c main.c listaSimbolos.c listaCodigo.c -lfl -o miniC
miniC.tab.h miniC.tab.c: miniC.y
    bison -d -v miniC.y
lex.yy.c: miniC.l miniC.tab.h
    flex miniC.l
limpia:
    rm -f lex.yy.c miniC.tab.* miniC.output miniC
run:    miniC p
        ./miniC p

```

Se ha de tener en cuenta que el programa vuelca toda la salida por la salida estándar, por lo que si se desea guardar en un fichero se debería redireccionar la salida. Teóricamente el código generado por `miniC`, si el programa es correcto, puede ser copiado y ejecutado en

un editor como Mars. Sin embargo, no se podría ejecutar directamente por una máquina que utilice este lenguaje ensamblador por los diversos mensajes de aviso lanzados por los compiladores.

Conclusión

En resumen, se ha logrado completar el desarrollo de un compilador capaz de traducir un lenguaje simplificado de C, miniC, al lenguaje ensamblador MIPS. Este proyecto ha supuesto un desafío interesante y nos ha permitido adquirir experiencia en el campo de la compilación y la traducción de lenguajes de programación.

El compilador implementado ha conseguido generar código MIPS válido. Se ha puesto un enfoque especial en el análisis semántico.

Aunque el compilador presenta algunas limitaciones, en su mayoría ha sido capaz de funcionar de manera adecuada en la traducción de programas de miniC a MIPS.

En definitiva, este trabajo ha sido una valiosa oportunidad para sumergirnos en el mundo de la compilación y la traducción de lenguajes de programación, adquiriendo habilidades y experiencia que nos servirán en futuros proyectos.