Forma

Descripción generada automáticamente con confianza media

**ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA INFORMÁTICA**

**DOBLE GRADO EN INGENIERÍA INFORMÁTICA E INGENIERÍA DEL SOFTWARE**

**Curso Académico 2023/2024**

**Procesadores de Lenguajes**

**8 de abril de 2024**

**Práctica Obligatoria PL**

**Equipo 6**:

Ángel Marqués García

Tarek Elshami Ahmed

Álvaro Serrano

# Índice

[Índice I](#_Toc162089216)

[1. Introducción 2](#_Toc162089217)

[2. Parte Aprobado 3](#_Toc162089218)

[2.1 Reducción de la gramática a LL(1) 3](#_Toc162089219)

[2.1.1 Eliminar recursividad por la izquierda 3](#_Toc162089220)

[2.1.2 Factorizar 3](#_Toc162089221)

[2.1.3 Generar los conjuntos de CAB, SIG 4](#_Toc162089222)

[2.2 Generar la gramática de ANTLR 5](#_Toc162089223)

[3. Parte Notable 6](#_Toc162089224)

[3.1 Sentencias de control de flujo 6](#_Toc162089225)

[3.2 Reconocimiento y notificación de errores 6](#_Toc162089226)

[4. Parte sobresaliente 6](#_Toc162089227)

[5. Probar el correcto funcionamiento del reconocedor 6](#_Toc162089228)

[5.1 Correctos 7](#_Toc162089229)

[5.1.1 Caso 1 7](#_Toc162089230)

[5.1.2 Caso 2 7](#_Toc162089231)

[5.1.3 Caso 3 7](#_Toc162089232)

[5.1.4 Caso 4 7](#_Toc162089233)

[5.2 Erróneos 7](#_Toc162089234)

[5.2.1 Caso 1 7](#_Toc162089235)

[5.2.2 Caso 2 7](#_Toc162089236)

[5.2.3 Caso 3 7](#_Toc162089237)

[5.2.4 Caso 4 7](#_Toc162089238)

# Introducción

Esta memoria redacta el proceso seguido por el grupo 6 en la realización de la práctica obligatoria de Procesadores de Lenguajes. El objetivo de esta práctica es diseñar e implementar un programa capaz de traducir código en un lenguaje fuente similar a C a un lenguaje final similar a PASCAL.

La práctica consta de tres sets de requisitos diferenciados, completar cada uno permite optar a una mejor calificación:

1. Requisitos Aprobado (parte obligatoria)
2. Requisitos Notable (sentencias de control de flujo y notificación de errores)
3. Requisitos Sobresaliente (recuperación de errores y toda la parte opcional)

A la misma vez, la práctica consta de dos entregas. La primera consta de los analizadores léxico y sintáctico, y la segunda de la práctica completa.

Por tanto, el primer paso que tomamos en el desarrollo de la práctica ha sido desarrollar los analizadores léxico y sintáctico. Este proceso consta de 3 fases principales:

1. Reducir la gramática a LL(1)
2. Trasladar la gramática a ANTLR
3. Probar su corrección y completitud

Y de nuevo este proceso se ha dividido entre los distintos sets de requisitos, de forma que primero ha sido aplicado a la parte de aprobado, y luego a las partes de aprobado y notable.

Dado que esta es la memoria de la primera entrega, las secciones pertinentes a la generación del lenguaje final a partir del lenguaje fuente no se han incluido.

# Parte Aprobado

## Reducción de la gramática a LL(1)

La gramática del lenguaje fuente es la siguiente:

program ::= defines partes

defines ::= ʎ | "#define" IDENT ctes defines

ctes ::= CONSTINT | CONSTFLOAT | CONSTLIT

partes ::= part partes | part

part ::= type restpart

restpart ::= IDENT "(" listparam ")" blq

| IDENT "(" "void" ")" blq

blq ::= "{" sentlist "}"

listparam ::= listparam "," type IDENT | type IDENT

type ::= "void" | "int" | "float"

sentlist ::= sentlist sent | sent

sent ::= type lid ";" | IDENT "=" exp ";" | IDENT "(" lexp ")" ";"

| IDENT "(" ")" ";" | "return" exp ";"

lid ::= IDENT | lid "," IDENT

lexp ::= exp | lexp "," exp

exp ::= exp op exp | factor

op ::= "+" | "-" | "\*" | "/" | "%"

factor ::= IDENT "(" lexp ")" | IDENT "(" ")"

| "(" exp ")" | IDENT | ctes

Para reducirla a LL(1) hemos seguido una serie de pasos:

### Eliminar recursividad por la izquierda

Hemos identificado que las siguientes reglas tenían recursividad por la izquierda:

* listparam
* sentlist
* lid
* lexp
* exp

Y hemos eliminado la recursividad por la izquierda modificando estas reglas y añadiendo nuevas, de forma que ahora son de la siguiente manera:

listparam: type IDENT listparam2;  
listparam2: ',' type IDENT listparam2 | ʎ;

sentlist: sent sentlist2;  
sentlist2: sent sentlist2| ʎ;  
lid: IDENT lid2;  
lid2: ',' IDENT lid2 | ʎ;  
lexp: exp lexp2;  
lexp2: ',' exp lexp2 | ʎ;  
exp: factor exp2;  
exp2: op exp exp2| ʎ;

Una vez la gramática no tiene recursividad por la izquierda, el siguiente paso es:

### Factorizar

Hemos identificado que las siguientes reglas podían ser factorizadas:

* restpart
* sent
* factor
* partes

Y las hemos factorizado modificándolas y añadiendo nuevas reglas, de forma que ahora son de la siguiente manera:

restpart: IDENT '(' restpart2;  
restpart2: listparam ')' blq | 'void' ')' blq;  
sent: type lid ';'  
 | 'return' exp ';'  
 | IDENT sent2  
sent2: '=' exp ';'| '(' sent3;  
sent3: ')'';'| lexp ')'';';

factor: '(' exp ')' | ctes | IDENT factor2;  
factor2: '(' factor3 | ;  
factor3: lexp ')' | ')';

partes: part partes2;  
partes2: partes | ;

Ahora que la gramática no tiene recursividad por la izquierda ni contiene reglas que puedan ser factorizadas a simple vista, el siguiente paso es:

### Generar los conjuntos de CAB, SIG

Para generar estos conjuntos utilizado la herramienta JFLAP. Dado que JFLAP sólo permite usar reglas que consten de una letra, hemos tenido que mapear cada regla a una letra del alfabeto. Así, la representación de las reglas en JFLAP es la siguiente:

A -> program  
B -> defines  
C -> ctes

D -> partes  
Ñ -> partes2  
E -> part  
F -> blq

G -> type  
H -> op  
I -> listparam  
J -> listparam2  
K -> sentlist  
L -> sentlist2  
M -> lid  
N -> lid2  
O -> lexp

P -> lexp2  
Q -> exp  
R -> exp2  
S -> factor  
T -> factor2  
U -> factor3  
V -> restpart  
W -> restpart2  
X -> sent   
Y -> sent2  
Z -> sent3

Y los símbolos terminales han sido codificados de la siguiente manera:

a -> '#defines'  
b -> IDENT  
c -> CONSTINT  
d -> CONSTFLOAT  
e -> CONSTLIT  
f -> '{'  
g -> '}'  
h -> 'void'  
i -> 'int'  
j -> 'float'  
k -> '+'  
l -> '-'  
m -> '\*'  
n -> '/'  
o -> '%'  
p -> ','  
q -> '('  
r -> ')'  
s -> ';'

t -> 'return'

Con esta traducción, JFLAP genera los siguientes conjuntos de Cabecera y Siguiente:  
Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Junto con la siguiente tabla de directores:

Imagen que contiene mucho, cubierto, llenado, luz

Descripción generada automáticamente

Una vez completado este paso, podemos proceder a generar la gramática en ANTLR.

## Generar la gramática de ANTLR

Para generar la gramática en ANTLR hemos copiado las reglas de la gramática final a ANTLR, ajustando el formato para que sea el adecuado, y creando además los fragmentos léxicos necesarios para el correcto funcionamiento del reconocedor. Se han creado los siguientes fragmentos léxicos:

IDENT: (LETTER | '$') (LETTER |DIGIT|'\_'| '$')\*;  
  
CONSTINT: DECIMAL\_INT | OCTAL\_INT | HEXA\_INT;  
CONSTFLOAT: DECIMAL\_REAL | OCTAL\_REAL | HEXA\_REAL;  
  
DECIMAL\_INT: ('+' | '-')? DIGIT+;  
DECIMAL\_REAL: ('+' | '-')? DIGIT+ '.' DIGIT+;  
  
OCTAL\_INT: '0' DECIMAL\_INT;  
OCTAL\_REAL: '0' DECIMAL\_REAL;  
  
HEXA\_INT: '0x' ('+' | '-')? HEX\_DIGIT+;  
HEXA\_REAL: '0x' ('+' | '-')? HEX\_DIGIT+ '.' HEX\_DIGIT+;  
  
CONSTLIT: '\'' (ESC\_SEQ | ~['])\* '\'' ;  
  
GPC: '//' ~[\r\n]\* -> skip ;GPC2: '/\*' .\*? '\*/' -> skip;  
  
JUMP: ('\r'? '\n' | ' ')+ -> skip;  
fragment ESC\_SEQ: '\\' '\'';  
fragment LETTER: [a-zA-Z];  
fragment HEX\_DIGIT: [0-9A-F] ;  
fragment DIGIT: [0-9] ;

# Parte Notable

Los requisitos de esta parte incluyen reconocer sentencias de control de flujo, y reconocer y notificar errores. Lo primero que hemos hecho ha sido modificar la gramática para añadir reconocimiento de las sentencias de control de flujo. Para ello hemos partido de una gramática de estas sentencias, que hemos reducido a LL(1), e incorporado al resto de la gramática.

## Sentencias de control de flujo

Partimos de las siguientes reglas:

lcond ::= lcond opl lcond | cond | "!" cond

opl ::= "||" | "&&"

cond ::= exp opr exp

opr "==" | "<" | ">" | ">=" | "<="

sent ::= …

| "if" "(" lcond ")" blq "else" blq

| "while" "(" lcond ")" blq

| "do" blq "until" "(" lcond ")"

| "for" "(" IDENT "=" exp ";" lcond ";" IDENT "=" exp ")" blq

Identificamos que la regla lcond tiene recursividad por la izquierda, y la eliminamos, obteniendo:

lcond ::= cond lcond2 | "!" cond lcond2;  
lcond2 ::= opl lcond lcond2 | ʎ;

Identificamos que ninguna regla puede ser factorizada a simple vista, y procedemos a generar los conjuntos de cabecera, siguiente, y directores. Cabe destacar que, al estar modificando la regla sent, será necesario recalcular los conjuntos de todas las reglas, pues al modificar la gramática podrían haber cambiado.

QUE LOS HAY QUE HACER A MANO PORQUE NO CABEN, WHAT THE FUCK DUDE?????

## Reconocimiento y notificación de errores

# Parte sobresaliente

# Probar el correcto funcionamiento del reconocedor

Por último, hemos realizado una serie de pruebas para comprobar que el reconocedor funciona de la forma esperada. Se ha probado su funcionamiento con 8 casos de prueba en total, cuatro de los cuales son correctos, y el resto erróneos. Los casos de prueba son los siguientes:

## Correctos

### Caso 1

### Caso 2

### Caso 3

### Caso 4

## Erróneos

### Caso 1

### Caso 2

### Caso 3

### Caso 4