**ANALIZADOR LÉXICO Y SINTÁCTICO**

Diego Jiménez Fernández-Pacheco GII + GIS Asier Ruano Peñas   GII + GIS

Índice

[Introducción 2](#_Toc479104947)

[Analizador léxico con JFlex 2](#_Toc479104948)

[Analizador sintáctico con Cup 3](#_Toc479104949)

[Conexión entre el Analizador Léxico y el Analizador Sintáctico 4](#_Toc479104950)

[Casos de Prueba 4](#_Toc479104951)

[Corrección de errores con soluciones propuestas 4](#_Toc479104952)

## Introducción

En esta memoria se recogen los aspectos y cuestiones que se consideran más relevantes para la comprensión del trabajo realizado. Consiste en las especificaciones léxico y sintáctico para el diseño e implementación de un visualizador de código de un lenguaje de programación similar a C, así como el traductor dirigido por sintaxis que genera una página web con el programa proporcionado.

En esta entrega se incluyen las partes de código que analizan los elementos del lenguaje (analizador léxico) y su correcta combinación (analizador sintáctico). Además se incluye la explicación de la traducción dirigida por sintaxis usada para generar el HTML.

La presente memoria se divide en tres partes: una en la que explicaremos el desarrollo del analizador léxico, otra en la que se explicará el del analizador sintáctico y por último el analizador semántico. Además se explicará la unión de los analizadores léxico y sintáctico para conseguir su colaboración. En un apartado final, se explicará cómo se ha abordado la gestión de errores y se incluirán los casos de prueba utilizados para comprobar el correcto comportamiento del analizador.

Los lenguajes y herramientas que se han empleado para este trabajo han sido JFlex en la parte del analizador léxico y Cup en la parte del analizador sintáctico. Con estas dos herramientas se han construido las expresiones regulares que indican al analizador léxico qué *tokens* debe reconocer y la gramática que debe ser reconocida por el programa.

Todo el trabajo se encuentra en un repositorio de GitHub. Este repositorio es público, por lo que si el profesor desea tener acceso al mismo le enviaremos el enlace.

## Analizador léxico con JFlex

Para poder reconocer los lexemas que cumplían con los patrones de nuestro lenguaje de programación similar a C y generar los *tokens* que se le pasarían posteriormente al analizador sintáctico, necesitábamos generar una especificación léxica a partir de la cual, mediante expresiones regulares, nuestro programa entendiese que una determinada cadena de entrada se correspondía a un determinado patrón.

Para ello nos valimos de la herramienta JFlex, que nos permite generar el autómata que reconocerá nuestros *tokens* a partir de las expresiones regulares apropiadas.

Así, basándonos en el enunciado de la práctica, diseñamos en un fichero *.flex* las expresiones regulares que darían lugar al autómata deseado.

Mediante esta construcción léxica conseguimos que nuestro analizador genere todos los *tokens* correspondientes al programa para que puedan ser posteriormente analizados por el analizador sintáctico.

Se ha decidido poner en primer lugar los tokens de palabras reservadas y símbolos ya que en caso contrario detectaba como identificador una palabra reservada.

Se han añadido en los terminales sentencias del tipo debug para poder mostrar la ejecución del programa cuando se encuentra algún error.

## Analizador sintáctico con Cup

El analizador sintáctico debe comprobar que los elementos reconocidos por el léxico están correctamente combinados.

Para la realización del analizador sintáctico y conseguir la función buscada, hicimos uso de la herramienta Cup. Esta herramienta construye un autómata a pila basándose en la gramática de nuestro lenguaje a reconocer. Así, lo que hicimos para llevar a cabo esta parte de la práctica fue especificar en un archivo .cup la gramática que debían cumplir los diferentes elementos del programa para ser correctos.

Al principio tuvimos algunos problemas con la recursividad en las gramáticas que hacían que el programa no se ejecutara de forma correcta.

La solución fue sencilla: Cup cuenta con una directiva que permite precisamente indicar precedencia entre los símbolos de la gramática (*precedence left [lista de símbolos]*) de modo que el analizador ya podía generar correctamente la tabla de símbolos de la gramática.

Además, las gramáticas de la parte opcional han sido simplificadas, para evitar repeticiones y producciones unitarias.

## Conexión entre el Analizador Léxico y el Analizador Sintáctico

Una vez generados ambos analizadores, debíamos conseguir que se comunicasen entre sí para poder colaborar.

Existen varias formas de hacer comunicarse estos dos componentes, pero nosotros, siguiendo con las pautas marcadas en clase, elegimos que el analizador léxico, al detectar un token, devolviese un elemento de la clase *symbol* (la cual es generada por cup, igual que la clase *parser*) para que de esta manera pudiese ser utilizada por la clase *parser*. Creamos una clase principal en la cual creamos una instancia del analizador léxico, otra del parser y las unimos de tal manera que al hacer la llamada al método *parser.parse* se procesa todo el archivo que hayamos puesto.

Elegimos que el analizador léxico fuese una subrutina del analizador sintáctico, de manera que cada vez que se procesa un *token*, el analizador sintáctico llama al léxico mediante *nextToken()* para procesar el siguiente.

Esto se consigue en Java mediante *Parser,* que se encarga precisamente de conectar ambos analizadores.

El programa básicamente lee de un fichero de texto las diferentes cadenas de entrada, se las pasa al analizador léxico para que genere un token o, en su defecto, detecte los errores y posteriormente hace que el sintáctico se inicie y compruebe la combinación de los tokens de una línea, llamando al léxico cuando sea necesario.

## Analizador Semántico

Se ha utilizado una estructura jerárquica para mostrar el HTML, además para mostrar la página web se usa un String.

Cada elemento del documento HTML es un String formado por tres partes: cabeza, cuerpo y fin. Dichos String se van uniendo conforme se crean y especifican, hasta que al final solo queda un único String que contiene la página web completa.

Un ejemplo de las partes del String sería un div. Cabeza contendría <div>, cuerpo el contenido del div y por último fin seria la etiqueta </div>

Se ha utilizado un nuevo axioma de la gramática para que el título del Documento solo aparezca una vez. La nueva regla es del tipo S’::= S, siendo S el axioma original.

ESPECIFICAR SI SE HA USADO JERARQUIA DE CLASES Y PONER ALGUN EJEMPLO MÁS.

ESPECIFICAR LOS ATRIBUTOS USADOS: SINTETIZADOS Y/O HEREDADOS

DECIR SI ES DEFINICION DIRIGIDA POR SINTAXIS O ESQUEMA DE TRADUCCIÓN

## Casos de Prueba MODIFICARLOS CON SEMANTICO

Casos de prueba correctos de léxico:

* Correcta1l: Dos bucles For anidados que realizan una operación de suma.
* Correcta2l: Un programa de resta en el que dados dos valores, comprueba que el primero sea mayor que el segundo para restar, sino devuelve el segundo valor.

Casos de prueba incorrectos de léxico:

* Incorrecta1l: El mismo ejercicio que correcta1l pero tiene un error léxico al tener un nombre de variable con un guion.
* Incorrecta2l: Usando el ejemplo de correcta2l, los errores son returnn (tiene 2n), i-1, j-1 ya que como no tiene espacios a los lados del – detecta un identificador y una constante.

Casos de prueba correctos de sintáctico:

* Correcta1s: Un programa de multiplicar valores usando una función auxiliar.
* Correcta2s: Un programa de suma de vectores.

Casos de prueba incorrectos de sintáctico:

* Incorrecta1s: El error lo encontramos en cada declaración e inicialización de variable, ya que nuestro analizador sintáctico no permite que se declare y a la vez inicialice una variable.
* Incorrecta2s: El error está en la sentencia suma = suma + m1[i] + m2[i]; ya que no se permite acceder a matrices con variables.

## Corrección de errores con soluciones propuestas

Léxico

Cuando entra un número con formato octal pero con un nueve, lo transforma en un número de base decimal.

Cuando aparecen dos operadores seguidos (++, --) devuelve el tipo de número y el + lo elimina para que el analizador sintáctico pueda continuar sin problemas; ejemplo ++255 devuelve <decint>.

Sintáctico

Si al ejecutar el código se encuentra un error, el programa muestra por pantalla la secuencia de terminales aceptados y en qué terminal está el error que hace fallar la aplicación.

Se ha añadido modificado los métodos report\_error y report\_fatal\_error, para que en el momento en el que se detecte un error informe de él de manera detallada y continúe la ejecución o, en caso de fatal\_error; lo informe y luego termine la ejecución.

Semántico