

Capítulo 4. Análisis semántico: Fase de Identificación

1 Objetivo

1.1 Objetivo del Análisis semántico

El tipo de los hijos de cada nodo (*expresión*, *sentencia*, etc) ya supone una restricción sobre qué nodos se pueden conectar con qué otros en el árbol (por ejemplo no se puede poner un objeto *DefVariable* como hijo de un nodo *Print*). Sin embargo quedan aún situaciones que, aunque los tipos de los hijos sean válidos, no deben ser aceptadas en un programa. Un ejemplo de esto sería la asignación de un real a una variable de tipo entero. Ambos nodos son de tipo *expresión* (tal y como se requiere a los hijos del nodo *asignación*), pero sin embargo el subárbol que forman esos tres nodos no debe aceptarse como válido en el AST (en este lenguaje solo se permiten asignaciones del mismo tipo).

Por tanto, una manera [muy] informal de explicar la misión del semántico es la de *validar todos los subárboles* del AST (que el padre y los hijos cumplan unas determinadas reglas). Si todos ellos son válidos, el AST será válido.

1.2 Fases del análisis semántico

El análisis semántico debe hacer distintas validaciones. Aunque se podría hacer todo el análisis semántico en una sola fase (de hecho en varios textos se presenta así), es más fácil de diseñar e implementar si se hace por partes. La mayor parte (no todas) de las validaciones que debe hacer el semántico se corresponden con las reglas de ámbito y las reglas de tipo. Por tanto el análisis semántico se hará en dos fases que se corresponden con la validación de cada uno de estos tipos reglas: las reglas de ámbito se comprueban en la Fase de Identificación y las reglas de tipo se comprobarán en el próximo capítulo en la fase de comprobación de tipos¹.

1.3 Objetivo de la Fase de Identificación

La Fase de Identificación es la primera de las dos fases en las que se va a dividir el análisis semántico. Esta fase tiene que realizar dos tareas.

1.3.1 Tarea 1. Validar

La primera tarea de esta fase es comprobar que:

- Todo símbolo que se usa en el programa ha sido definido y es accesible desde el lugar en el que se utiliza. En este lenguaje eso significa que las variables usadas hayan sido definidas).
- Que no haya definiciones inválidas. En este lenguaje eso significa que no haya más de una definición de una misma variable.

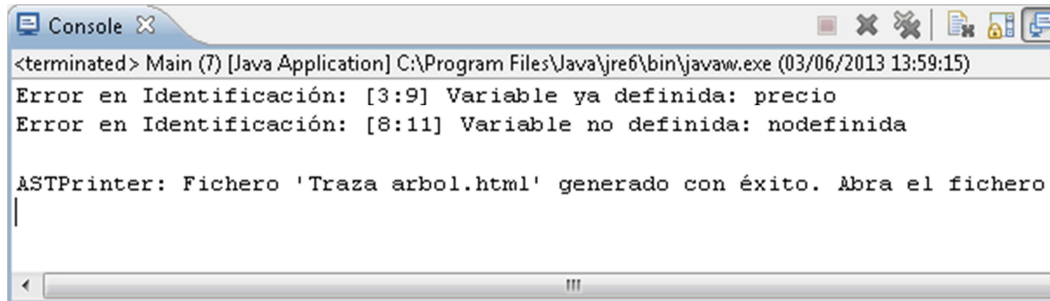
¹ Cabe preguntarse que si no todas las validaciones se corresponden con estas reglas, en qué fase se tratan las validaciones restantes que no son de ninguno de estos dos tipos. Aunque quizás fuera más coherente realizarlas en una tercera fase con “el resto” de las validaciones, en la práctica se suelen incluir en la segunda fase, la comprobación de tipos, debido a que suelen ser muy pocas y por tanto no suele merecer la pena hacerlas aparte.

Por ejemplo ante una entrada como:

```
DATA
    float precio;
    int precio;    // Error: variable ya definida
    int ancho;

CODE
    print precio;
    print nodefinida;    // Error: variable no definida
```

Debería notificar los siguientes errores:



1.3.2 Tarea 2. Enlazar

Aunque estrictamente la tarea anterior sería la única misión de esta fase, también debe realizar, como tarea adicional, el enlazado de las variables con su definición (que los objetos *Variable* apunten a su *DefVariable*). La razón de esta última tarea es que las fases posteriores del compilador van a necesitar de nuevo información sobre la definición de una variable (su tipo, su dirección, etc.). Si quedan ya enlazadas no habrá que volver a repetir en fases posteriores la búsqueda de los nodos *DefVariable* que ya se hace aquí.

2 Solución

2.1 Diseño. Creación de la Especificación

2.1.1 Extracción de Requisitos

En el apartado de objetivos se estableció que la función de esta fase es comprobar que:

1. Todo símbolo que se usa en el programa ha sido definido y es accesible desde el lugar en el que se utiliza.
2. Que no haya definiciones inválidas.

Los requisitos de esta fase son las reglas que responden a los dos puntos anteriores: las *Reglas de Ámbito* del lenguaje. Sin embargo en *Introducción.pdf* no se menciona ninguna regla de ámbito. Por tanto es de suponer que las reglas de ámbito sean las habituales:

1. Cada variable es accesible desde toda la sección CODE.
2. No se puede definir una variable con el mismo nombre que otra previamente definida (independientemente del tipo de ambas).

Es razonable suponer que si hubiera reglas de ámbito excepcionales se hubieran mencionado en la especificación (por ejemplo que se pudiera cambiar el tipo de las variables volviéndolas a definir de nuevo).

2.1.2 Metalenguaje elegido

Los requisitos semánticos anteriores se expresarán mediante una *Gramáticas Atribuida* (GAt). Este metalenguaje es una forma sencilla de expresar:

- Qué información se añade al árbol y en qué nodos (*atributos*).
- Qué condiciones deben cumplir los nodos de un árbol para que este sea válido (*predicados*).

Concretando para este caso:

- La información que se añade al árbol es en enlace entre las variables y su definición (un atributo *definición*).
- Las condiciones que debe cumplir el árbol es que no se defina dos veces una variable o que no se haya definido (predicados).

2.1.3 Especificación en el Metalenguaje

Una alternativa a comenzar la Gramática Atribuida desde cero es utilizar el esqueleto generado por *VGen*². En la carpeta *metalenguajes* generó un fichero '*Attribute Grammar.html*' en el cual ya aparece la estructura de una gramática atribuida. En la primera tabla aparece una primer columna con la gramática abstracta del lenguaje. Aparecen además una segunda y tercera columna vacías para que se pongan los predicados y las reglas semánticas respectivamente. Finalmente aparece una segunda tabla en el que se indicarán los atributos que se añadirán al árbol y en qué nodos.

Una vez abierto dicho esqueleto en Word, se traducen los requisitos a la notación del metalenguaje rellenando las dos columnas de la derecha y la tabla de atributos. El resultado de este proceso está en "*Identificación. Gramática Atribuida.pdf*"

² Si no se guardó la carpeta *metalenguajes* del capítulo 3 basta con volver a ejecutar *VGen* con el fichero *gramática.txt* que se creó en dicho capítulo.

Nodo	Predicados	Reglas Semánticas
programa → <i>definiciones</i> :defVariable* <i>sentencias</i> :sentencia*		
defVariable → <i>tipo</i> :tipo <i>nombre</i> :String	variables[nombre] == ∅	variables[nombre] = defVariable
intType :tipo → λ		
realType :tipo → λ		
print :sentencia → <i>expresion</i> :expresion		
asigna :sentencia → <i>left</i> :expresion <i>right</i> :expresion		
exprAritmetica :expresion → <i>left</i> :expresion <i>operador</i> :String <i>right</i> :expresion		
variable :expresion → <i>nombre</i> :String	variables[nombre] ≠ ∅	variable.definicion = variables[nombre]
literalInt :expresion → <i>valor</i> :String		
literalReal :expresion → <i>valor</i> :String		

Tabla de Atributos:

Categoría Sintáctica	Nombre	Tipo Java	H/S	Descripción
variable	definicion	DefVariable	Sintetizado	Nodo donde está definido la variable

Estructuras de datos auxiliares:

variables Map<String, DefVariable>

2.2 Implementación

2.2.1 Enlace con las definiciones

Tal y como se indica en la tabla de atributos de la gramática atribuida anterior, hay que añadir una propiedad *definición* a los objetos (nodos) de tipo *Variable*:

```
class Variable {
    private DefVariable definicion;

    public DefVariable getDefinicion() { return definicion; }

    public void setDefinicion(DefVariable definicion) {
        this.definicion = definicion;
    }

    // Resto igual
}
```

2.2.2 Fase de Identificación

Para implementar este *visitor*, en el proyecto de eclipse ya existe un fichero *semantico/Identificación.java* que sugiere donde colocar esta nueva clase. En dicho fichero se tiene un esqueleto al que solo falta añadirle los métodos *visit* para los nodos de nuestra gramática.

Esta tarea se puede hacer a mano, aunque es más rápido utilizar el fichero '*visitor/_PlantillaParaVisitors.txt*' que fue generado por *VGen*. Este fichero ya contiene todos

los métodos *visit* para los nodos de nuestra gramática y el código que recorre los hijos. Basta con copiar y pegar los métodos *visit* de la plantilla en *Identificación.java* de tal manera que el trabajo se podrá centrar en decidir qué se quiere hacer y en qué nodos.

En este caso, tal y como muestra la gramática atribuida, solo hay que actuar sobre dos nodos: *Variable* (para buscar la definición y enlazarla) y *DefVariable* (para comprobar que no esté repetida). Por tanto todos los demás métodos *visit* se pueden borrar³.

La clase *Identificación.java*, siguiendo la Gramática Atribuida anterior, quedaría así:

```
public class Identificacion extends DefaultVisitor {

    public Identificacion(GestorErrores gestor) {
        this.gestorErrores = gestor;
    }

    // class DefVariable { Tipo tipo; String nombre; }
    public Object visit(DefVariable node, Object param) {
        DefVariable definicion = variables.get(node.getNombre());
        predicado(definicion == null, "Variable ya definida: " + node.getNombre(), node.getStart());
        variables.put(node.getNombre(), node);
        return null;
    }

    // class Variable { String nombre; }
    public Object visit(Variable node, Object param) {
        DefVariable definicion = variables.get(node.getNombre());
        predicado(definicion != null, "Variable no definida: " + node.getNombre(), node.getStart());
        node.setDefinicion(definicion); // Enlazar referencia con definición
        return null;
    }

    /**
     * Método auxiliar opcional para ayudar a implementar los predicados de la Gramática Atribuida.
     *
     * Ejemplo de uso:
     * predicado(variables.get(nombre), "La variable no ha sido definida", expr.getStart());
     * predicado(variables.get(nombre), "La variable no ha sido definida", null);
     *
     * NOTA: El método getStart() indica la línea/columna del fichero fuente de donde se leyó el nodo.
     * Si se usa VGen dicho método será generado en todos los nodos AST. Si no se quiere usar getStart() se
     * puede pasar null.
     *
     * @param condicion Debe cumplirse para que no se produzca un error
     * @param mensajeError Se imprime si no se cumple la condición
     * @param posicionError Fila y columna del fichero donde se ha producido el error. Es opcional (acepta
     * null)
     */
    private void predicado(boolean condicion, String mensajeError, Position posicionError) {
        if (!condicion)
            gestorErrores.error("Identificación", mensajeError, posicionError);
    }

    private GestorErrores gestorErrores;
    private Map<String, DefVariable> variables = new HashMap<String, DefVariable>();
}
```

³ Dado que al final solo se han utilizado dos métodos *visit* de *_PlantillaParaVisitor.txt* podría plantearse el por qué copiar todos en vez de escribirlos esos dos directamente. En general se recomienda copiar todo y determinar método a método si hace falta actuar sobre dicho nodo o no para no olvidar ninguno. Además es más fácil borrar lo que sobra que comenzar con un *visitor* vacío.

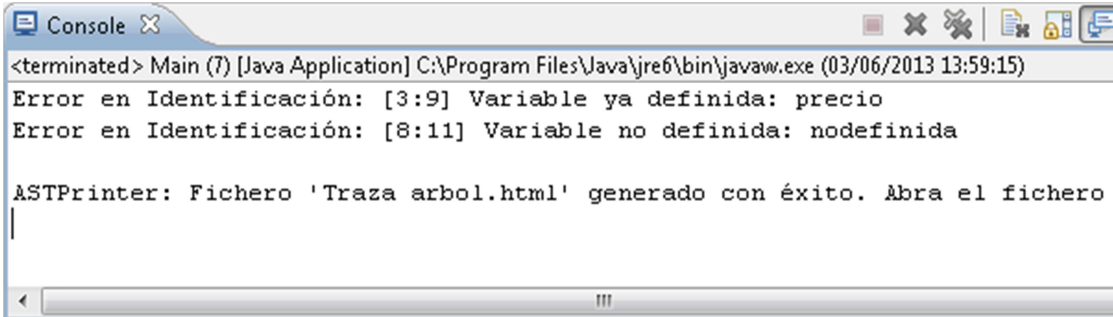
3 Ejecución

Se modifica el fichero *entrada.txt* para comprobar que el compilador detecta adecuadamente los errores de esta fase:

```
DATA
    float precio;
    int precio;    // Error: variable ya definida
    int ancho;

CODE
    print precio;
    print nodefinida;    // Error: variable no definida
```

Al ejecutar la clase *Main* se obtiene la siguiente salida tal y como se esperaba:



```
<terminated> Main (7) [Java Application] C:\Program Files\Java\jre6\bin\javaw.exe (03/06/2013 13:59:15)
Error en Identificación: [3:9] Variable ya definida: precio
Error en Identificación: [8:11] Variable no definida: nodefinida

ASTPrinter: Fichero 'Traza arbol.html' generado con éxito. Abra el fichero
```

4 Resumen de Cambios

Fichero	Acción	Descripción
Identificación. Gramática Atribuida.pdf	Creado	Metalenguaje en el que se han descrito los requisitos de esta fase
Variable.java	Modificado	Se añade una referencia a su definición
Identificacion.java	Modificado	Comprueba que toda variable ha sido definida y solo lo ha hecho una vez. Además enlaza las variables con sus definiciones
entrada.txt	Modificado	Contienen errores que deben detectarse en esta fase