

Lenguajes de Programación 2022-1

Nota de clase 15: Featherweight Java

Orientación a Objetos

Javier Enríquez Mendoza

15 de enero de 2022

El objetivo de esta nota de clase es modelar algunos aspectos del paradigma de programación orientado a objetos, mediante el estudio de un lenguaje de estudio llamado **Featherweight Java** que es un subconjunto del lenguaje Java.

Como pudimos observar con el lenguaje **TinyC**, la formalización de lenguajes de programación imperativos resulta una tarea más compleja de lo que es en el estilo funcional, esto se debe principalmente a lo separados que se encuentran los lenguajes imperativos de sistemas matemáticos en contraste con los lenguajes funcionales que tienen una estrecha relación con la teoría que los fundamenta.

Para lograr estudiar los conceptos de Java no existe un modelo perfecto, en su mayoría los modelos teóricos de lenguajes orientados a objetos son malos para abstraer el comportamiento completo del lenguaje. Sin embargo, estos modelos son útiles si lo que se busca es abstraer un conjunto específico de propiedades del lenguaje.

En nuestro caso lo que buscamos es modelar las características principales del paradigma orientado a objetos así como su sistema de tipos. Para este propósito **Featherweight Java** resulta bastante útil. Las principales características que modela **Featherweight Java** son:

- Clases.
- Objetos.
- Métodos.
- Atributos y acceso a ellos.
- Herencia.
- Polimorfismo de subtipado.
- Recursión abierta.

Sin embargo se dejan fuera del modelo otros conceptos interesantes como lo son:

- Concurrencia.
- Interfaces.

- Manejo de accesos.
- Referencias.
- Clases internas.
- Excepciones.

Los cuales se pueden estudiar con otros modelos del paradigma orientado a objetos, que abstraen específicamente estas características.

1 Sintaxis

En esta sección se presenta la sintaxis concreta de Featherweight Java . Así como algunas convenciones sintácticas importantes que se deben tomar en cuenta al escribir un programa en el lenguaje.

Definición 1.1 (Sintaxis de Featherweight Java). Se presenta la sintaxis del lenguaje separado en categorías y usando las siguientes metavariables:

- Nombres de variables: x, y, z
- Nombres de atributos: f, g
- Nombres de clases: C, A, B
- Nombres de métodos: m

Expresiones

$e ::= x$	Variables
$e.f$	Acceso a atributo
$e.m(\vec{e})$	Invocación de método
$\text{new } C(\vec{e})$	Instancia de objeto
$(C) e$	Casting

Valores

$v ::= \text{new } C(\vec{v})$	Instancia de objeto
--------------------------------	---------------------

Métodos y Clases

$K ::= C(\vec{C} \vec{x}) \{ \text{super}(\vec{x}); \text{this}.\vec{f} = \vec{x} \}$	Constructores
$M ::= C m(\vec{C} \vec{x}) \{ \text{return } e; \}$	Métodos
$L ::= \text{class } C \text{ extends } B \{ \vec{A} \vec{f}; K \vec{M} \}$	Clases

En la definición anterior se usa constantemente la notación \vec{t} que no habíamos empleado anteriormente. Ésta recibe el nombre de notación vectorial y es una forma de simplificar la notación multiparamétrica que usamos en notas anteriores. Un vector denota la sucesión de expresiones

$$\vec{t} =_{def} t_1, t_2, \dots, t_n$$

También se pueden presentar dos vectores juntos sin comas o puntos y coma separándolos, se entiende que ambos vectores tienen la misma longitud y en tal caso esta combinación denota a la sucesión obtenida tomando un elemento de cada vector a la vez, es decir

$$\vec{C} \vec{x} =_{def} C_1 x_1, C_2 x_2, \dots, C_n x_n$$

En el caso de la definición de el constructor el vector empleado significa:

$$\text{this}.\vec{f} = \vec{x} =_{def} \text{this}.f_1 = x_1; \text{this}.f_2 = x_2; \dots; \text{this}.f_n = x_n$$

de igual forma se entiende que ambos vectores tienen la misma longitud.

De la definición de la sintaxis podemos abstraer algunas convenciones importantes que se deben seguir en los programas en **Featherweight Java** y que no necesariamente se siguen en los programas en Java. A continuación las listamos:

- En la definición de una clase se incluye siempre la superclase, aún cuando ésta es **Object**.
- Los constructores siempre se declaran explícitamente aún cuando estos son triviales o cuando el comportamiento es el esperado por Java y tienen el mismo nombre que la clase.
- Dentro del constructor siempre se llama al constructor de la superclase mediante la instrucción **super**.
- Siempre se menciona el objeto al invocar un método o acceder a un atributo aún cuando éste es **this**.
- Los métodos consisten únicamente de una expresión **return**.
- Todos los constructores funcionan de forma genérica, es decir, reciben el mismo número de parámetros como atributos tenga la clase, hacen la asignación de los mismos y no pueden hacer nada más.
- Se puede definir un único constructor por clase.
- En la definición de las clases los atributos solo son declarados, es decir, no podemos asignar valores por defecto. Solo toman valor en el constructor.

Veamos a continuación algunos ejemplos de programas escritos en **Featherweight Java**.

Ejemplo 1.1 (La clase **Pair**). Para ejemplificar la sintaxis se define un programa para manejo de pares.

```
class A extends Object { A () { super(); } }

class B extends Object { B () { super(); } }

class Pair extends Object {
    Object fst;
    Object snd;

    Pair (Object fst, Object snd){
        super();
        this.fst = fst;
        this.snd = snd;
    }

    Pair setfst (Object newfst){
        return new Pair(newfst, this.snd);
    }
}
```

```
}
```

Con las definiciones anteriores la expresión:

```
new Pair(new A (), new B ()).setfst(new B ())
```

se evaluaría intuitivamente a la expresión:

```
new Pair(new B (), new B ())
```

Ejemplo 1.2 (Árboles binarios en Featherweight Java). Se define un programa para manejo de árboles binarios con información solo en las hojas.

```
class Node extends Object {
    Object left;
    Object right;

    Node (Object left, Object right) {
        super();
        this.left = left;
        this.right = right;
    }

    Nat sum () {
        return (this.left.sum() + this.right.sum());
    }
}

class Leaf extends Object {
    Nat value;

    Leaf (Nat value) {
        super();
        this.value = value;
    }

    Nat sum() {
        return this.value;
    }
}
```

Para este ejemplo se supone definido el tipo `Nat` que modela los números naturales y la operación de suma para los elementos de este tipo.

1.1 Tablas de Clases

Un programa en JPP consta de una expresión e junto con una tabla de clases T la cual generalmente no se menciona explícitamente.

Definición 1.2 (Tabla de Clases). Una tabla de clases es una función finita que asigna clases a nombres de clase, en otras palabras, T es una sucesión finita de declaraciones de clase de la forma

$$T(C) = \text{class } C \text{ extends } B \{ \vec{A} \vec{f}; K \vec{M} \}$$

Veamos ahora como acceder a la información almacenada en esta tabla.

Búsqueda de Atributos

$$\frac{}{\text{fields}(\text{Object}) = \emptyset}$$

$$\frac{T(C) = \text{class } C \text{ extends } B \{ \vec{C} \vec{f}; K \vec{M} \} \quad \text{fields}(B) = \vec{B} \vec{g}}{\text{fields}(C) = \vec{B} \vec{g}, \vec{C} \vec{f}}$$

Búsqueda del tipo de un método

$$\frac{T(C) = \text{class } C \text{ extends } D \{ \vec{C} \vec{f}; K \vec{M} \} \quad B \text{ m}(\vec{B}, \vec{x}) \{ \text{return } e; \} \text{ figura en } \vec{M}}{\text{mtype}(\text{m}, C) = \vec{B} \rightarrow B}$$

$$\frac{T(C) = \text{class } C \text{ extends } D \{ \vec{C} \vec{f}; K \vec{M} \} \quad \text{m no figura en } \vec{M}}{\text{mtype}(\text{m}, C) = \text{mtype}(\text{m}, D)}$$

En estas reglas la notación vectorial $\vec{B} \rightarrow B$ se entiende como $B_1 \rightarrow B_2 \rightarrow \dots \rightarrow B$.

Búsqueda del cuerpo de un método

$$\frac{T(C) = \text{class } C \text{ extends } D \{ \vec{C} \vec{f}; K \vec{M} \} \quad B \text{ m}(\vec{B}, \vec{x}) \{ \text{return } e; \} \text{ figura en } \vec{M}}{\text{mbody}(\text{m}, C) = (\vec{x}, e)}$$

$$\frac{T(C) = \text{class } C \text{ extends } D \{ \vec{C} \vec{f}; K \vec{M} \} \quad \text{m no figura en } \vec{M}}{\text{mbody}(\text{m}, C) = \text{mbody}(\text{m}, D)}$$

2 Semántica Dinámica

En esta sección se presenta la semántica dinámica de Featherweight Java mediante una semántica operacional estructural.

Definición 2.1 (Semántica Operacional de Featherweight Java). Se define la semántica como un sistema de transición con función de transición \rightarrow_{FJ} que depende de la tabla de clases T de un programa, esta tabla se omite en las reglas del sistema, sin embargo siempre está presente.

Recordemos que los únicos valores del lenguaje están definidos por la instrucción **new**.

Selección de Atributos Como los parámetros del constructor corresponden exactamente con

los atributos definidos en las clases, podemos obtener directamente el valor de cada uno a partir de la instancia del objeto. Modelado en la siguiente regla:

$$\frac{\text{fields}(\mathbf{C}) = \vec{\mathbf{C}} \vec{\mathbf{f}}}{(\text{new } \mathbf{C}(\vec{\mathbf{v}})).\mathbf{f}_i \longrightarrow_{\text{FJ}} \mathbf{v}_i}$$

Invocación de métodos

$$\frac{\text{mbody}(\mathbf{m}, \mathbf{C}) = (\vec{\mathbf{x}}, \mathbf{e})}{(\text{new } \mathbf{C}(\vec{\mathbf{v}})).\mathbf{m}(\vec{\mathbf{w}}) \longrightarrow_{\text{FJ}} \mathbf{e}[\vec{\mathbf{x}}, \text{this} := \vec{\mathbf{w}}, (\text{new } \mathbf{C}(\vec{\mathbf{v}}))]}$$

En la regla anterior para la evaluación de una invocación de un método, se evalúa el cuerpo del método en donde se sustituyen todos los parámetros de éste por los argumentos de la llamada y se sustituye `this` por la instancia misma que hace la invocación.

Casting

$$\frac{\mathbf{C} \leq \mathbf{D}}{(\mathbf{D}) (\text{new } \mathbf{C}(\vec{\mathbf{v}})) \longrightarrow_{\text{FJ}} \text{new } \mathbf{C}(\vec{\mathbf{v}})}$$

Reglas de congruencia Como es usual la forma de evaluar es de derecha a izquierda lo que se modela con las siguientes reglas burocráticas:

$$\frac{e \longrightarrow_{\text{FJ}} e'}{e.f \longrightarrow_{\text{FJ}} e'.f}$$

$$\frac{e \longrightarrow_{\text{FJ}} e'}{e.m(\vec{e}) \longrightarrow_{\text{FJ}} e'.m(\vec{e})}$$

$$\frac{e_i \longrightarrow_{\text{FJ}} e'_i}{e.m(\dots, e_i, \dots) \longrightarrow_{\text{FJ}} e.m(\dots, e'_i, \dots)}$$

$$\frac{e_i \longrightarrow_{\text{FJ}} e'_i}{\text{new } \mathbf{C}(\dots, e_i, \dots) \longrightarrow_{\text{FJ}} \text{new } \mathbf{C}(\dots, e'_i, \dots)}$$

$$\frac{e \longrightarrow_{\text{FJ}} e'}{(\mathbf{C}) e \longrightarrow_{\text{FJ}} (\mathbf{C}) e'}$$

3 Semántica Estática

Para la definición de la semántica estática se define un sistema de tipos para el lenguaje, en este caso los tipos son los nombres de las clases y no se tienen tipos primitivos, es decir, todos los tipos que se necesiten para un programa deben ser definidos dentro del mismo, a excepción claro de `Object`.

Definición 3.1 (Semántica Estática de Featherweight Java). Para la definición de su sistema de tipos deben definirse las reglas de tipado de las expresiones, las reglas de subtipado y *casting* así como reglas que modelen cuando un método y una clase están bien formadas. A continuación se presentan todas estas definiciones.

Reglas de Subtipado

$$\frac{T(\mathbf{C}) = \text{class } \mathbf{C} \text{ extends } \mathbf{D} \{ \vec{\mathbf{C}} \vec{\mathbf{f}}; \mathbf{K} \vec{\mathbf{M}} \}}{\mathbf{C} \leq \mathbf{D}}$$

Es decir, si la clase \mathbf{C} extiende a la clase \mathbf{D} entonces \mathbf{C} es subtipo de \mathbf{D} .

Adicional a esta regla el subtipado debe cumplir las propiedades de transitividad y reflexividad presentadas en la nota de clase anterior.

Reglas de Tipado

$$\frac{}{\Gamma, x : \mathbf{C} \vdash x : \mathbf{C}}$$

Variables

$$\frac{\Gamma \vdash e : \mathbf{C} \quad \text{fields}(\mathbf{C}) = \vec{\mathbf{C}} \vec{\mathbf{f}}}{\Gamma \vdash e.\mathbf{f}_1 : \mathbf{C}_1}$$

Acceso a atributos

$$\frac{\Gamma \vdash e_0 : \mathbf{C}_0 \quad \Gamma \vdash \vec{e} : \vec{\mathbf{C}} \quad \text{mtype}(\mathbf{m}, \mathbf{C}_0) = \vec{\mathbf{D}} \rightarrow \mathbf{C} \quad \vec{\mathbf{C}} \leq \vec{\mathbf{D}}}{\Gamma \vdash e_0.\mathbf{m}(\vec{e}) : \mathbf{C}}$$

Invocación de métodos

$$\frac{\Gamma \vdash \vec{e} : \vec{\mathbf{C}} \quad \text{fields}(\mathbf{C}) = \vec{\mathbf{D}} \vec{\mathbf{f}} \quad \vec{\mathbf{C}} \leq \vec{\mathbf{D}}}{\Gamma \vdash \text{new } \mathbf{C}(\vec{e}) : \mathbf{C}}$$

Creación de objetos

Reglas de casting

$$\frac{\Gamma \vdash e : \mathbf{D} \quad \mathbf{D} \leq \mathbf{C}}{\Gamma \vdash (\mathbf{C}) e : \mathbf{C}}$$

Upcasting

$$\frac{\Gamma \vdash e : \mathbf{D} \quad \mathbf{C} \leq \mathbf{D} \quad \mathbf{C} \neq \mathbf{D}}{\Gamma \vdash (\mathbf{C}) e : \mathbf{C}}$$

Downcasting

$$\frac{\Gamma \vdash e : \mathbf{D} \quad \mathbf{C} \not\leq \mathbf{D} \quad \mathbf{D} \not\leq \mathbf{C} \quad \text{stupid warning}}{\Gamma \vdash (\mathbf{C}) e : \mathbf{C}}$$

Casting estúpido

Esta última regla es un tecnicismo necesario para poder probar la preservación de tipos con la semántica operacional estructural.

Formación de Clases Para definir las reglas que modelan cuándo una clase está bien formada, se definen los nuevos juicios $\mathbf{M} \text{ ok in } \mathbf{C}$ para indicar que el método \mathbf{M} está bien formado en la clase \mathbf{C} , $\mathbf{C} \text{ ok}$ que indica que la clase \mathbf{C} está bien formada y por último el juicio $T \text{ ok}$ para decir que la tabla de clases T está bien formada.

$$\frac{\begin{array}{l} \mathbf{K} = \mathbf{C}(\vec{\mathbf{D}} \vec{\mathbf{y}}, \vec{\mathbf{C}} \vec{\mathbf{x}}) \{ \text{super}(\vec{\mathbf{y}}); \text{this}.\vec{\mathbf{f}} = \vec{\mathbf{x}} \} \\ \text{fields}(\mathbf{D}) = \vec{\mathbf{D}} \vec{\mathbf{g}} \\ \vec{\mathbf{M}} \text{ ok in } \mathbf{C} \end{array}}{\text{class } \mathbf{C} \text{ extends } \mathbf{D} \{ \vec{\mathbf{C}} \vec{\mathbf{f}}; \mathbf{K} \vec{\mathbf{M}} \} \text{ ok}}$$

Formación de métodos

$$\frac{\begin{array}{l} T(\mathbf{C}) = \text{class } \mathbf{C} \text{ extends } \mathbf{D} \{ \dots \} \\ \text{mtype}(\mathbf{m}, \mathbf{D}) = \vec{\mathbf{C}} \rightarrow \mathbf{C}_0 \\ \vec{x} : \vec{\mathbf{C}}, \text{this} : \mathbf{C} \vdash e : \mathbf{C}'_0 \\ \mathbf{C}'_0 \leq \mathbf{C}_0 \end{array}}{\mathbf{C}_0 \mathbf{m}(\vec{\mathbf{C}} \vec{x}) \{ \text{return } e; \} \text{ ok in } \mathbf{C}}$$

En las premisas se está considerando la posibilidad de que el método m se haya redefinido (*overriding*) verificando que se respete el tipo dado en la superclase. Es decir, el cuerpo del método en la subclase debe devolver un subtipo del tipo del resultado del mismo método en la superclase. También se observa que los tipos de los parámetros del método deben ser exactamente los mismos que los de la superclase.

Es importante observar que si $D = \text{Object}$ entonces el método no se está redefiniendo puesto que Object no tiene métodos. En tal caso la premisa acerca de mtype debe ignorarse, siendo aceptadas cualesquiera C_0 y \vec{C} .

Formación de Tablas Decimos que una tabla de clases está bien formada si todas las clases declaradas en ella están bien formadas, modelado con la regla:

$$\frac{\forall C \in \text{dom}(T), T(C) \text{ ok}}{T \text{ ok}}$$

4 Seguridad de Featherweight Java

Al igual que con los sistemas anteriores estudiaremos la seguridad del lenguaje **Featherweight Java** con las propiedades de preservación de tipos y progreso de la función de transición, en este caso estas propiedades deben considerar la herencia así como el uso de *castings* inválidos.

Proposición 4.1 (Preservación de tipos). Si T es una tabla de clases bien formada, $\Gamma \vdash e : C$ y $e \rightarrow_{\text{FJ}} e'$, entonces existe C' tal que $C' \leq C$ y $\Gamma \vdash e' : C'$

Demostración. Consultar 1 □

Respecto a la propiedad de progreso esta vez debemos observar que un programa bien tipado podría bloquearse debido a un uso indebido del casting, pero en general podemos probar que la única forma en la que un programa bien tipado puede bloquearse es si llega a un punto en el que no puede computar un *downcasting*.

Proposición 4.2 (Progreso de la relación \rightarrow_{FJ}). Sea T una tabla de clases bien formada, si $\emptyset \vdash e : C$ entonces sucede una y solo una de las siguientes condiciones:

- e es un valor.
- e contiene una expresión de la forma $(C) \text{ new } D(\vec{v})$ en donde $D \leq C$, es decir, no se puede realizar el *downcast*.
- Existe e' tal que $e \rightarrow_{\text{FJ}} e'$.

Demostración. Consultar 1 □

Teorema 4.3 (Seguridad de Featherweight Java). Si T es una tabla de clases bien formada, $\emptyset \vdash e : C$ y $e \rightarrow_{\text{FJ}}^* e'$ con e' en forma normal, entonces se cumple una y solo una de las siguientes condiciones:

- e' es un valor v tal que $\emptyset \vdash v : D$ y $D \leq C$.
- e' contiene como subexpresión $(C) \text{ new } D(\vec{v})$ en donde $D \leq C$.

En otras palabras un programa bien tipado siempre termina o bien tiene un *downcasting* que no se puede resolver.

Demostración. Directo de las proposiciones 4.1 y 4.2 y la propiedad de transitividad de la relación de subtipado. \square

5 Correspondencia con Java

Como se menciona al inicio de esta nota, **Featherweight Java** es un modelo de algunas de las características principales del lenguaje de programación Java, principalmente diseñada para el estudio de la seguridad de su sistema de tipos. Sin embargo otras de las características de la programación orientada a objetos así como mucho del poder de expresividad de Java quedan fuera de este modelo.

Por esta razón es importante hacer explícita la correspondencia deseada entre un programa en **Featherweight Java** y su versión en Java. Esta correspondencia se da con los tres lemas siguientes:

Lema 5.1. Cada programa sintácticamente correcto en **Featherweight Java** es también sintácticamente correcto en Java.

Lema 5.2. Un programa sintácticamente correcto es tipable en **Featherweight Java** si y sólo si es tipable en Java.

Lema 5.3. La ejecución de un programa bien tipado en **Featherweight Java** se comporta de la misma forma en Java.

Como consecuencia del lema 5.3 tenemos el siguiente corolario.

Corolario 5.4. La evaluación de un programa en **Featherweight Java** no termina si y sólo si compilarlo y ejecutarlo en Java causa no terminación.

Por supuesto los lemas anteriores no pueden demostrarse ya que no existe una formalización completa del lenguaje Java. Sin embargo es muy útil enunciar la correspondencia de manera precisa para dejar en claro el objetivo que estamos tratando de alcanzar.

Referencias

- [1] Igarashi A., Pierce B., Wadler P., Featherweight Java: A Minimal Core Calculus for Java and GJ, ACM-TRANSACTION, Enero 23 2002. <https://www.cis.upenn.edu/~bcpierce/papers/fj-toplas.pdf>
- [2] Miranda Perea F., González Huesca L., Nota de Clase del curso de Lenguajes de Programación, Facultad de Ciencias UNAM, Semestre 2021-1.