Universidad
Europea de Madrid
LAUREATE INTERNATIONAL UNIVERSITIES

ANÁLISIS SINTÁCTICO I

CONJUNTOS PRIMERO Y SIGUIENTE

Índice

Presentación 4

El conjunto PRIMERO 5

Ejemplo 1 de cálculo de conjunto PRIMERO 7

El conjunto SIGUIENTE 9

Reglas para el cálculo del conjunto SIGUIENTE 9

Detalles a tener en cuenta de la aplicación de las reglas 9

Ejemplo 1 de cálculo del conjunto SIGUIENTE 11

Cálculo del conjunto SIGUIENTE 11

Ejemplo 2 de cálculo del conjunto PRIMERO 13

Cálculo del conjunto PRIMERO 13

Ejemplo 2 de cálculo del conjunto SIGUIENTE 15

Cálculo de los conjuntos SIGUIENTE 15

Resumen 18

Presentación

El objetivo de este tema es entender el proceso de cálculo de los **conjuntos PRIMERO y SIGUIENTE,** tan necesarios para la construcción de los analizadores sintácticos, tanto ascendentes como descendentes.

Los objetivos a conseguir en este primer tema de la unidad tres son eminentemente prácticos y son:

- Entender el cálculo y la motivación del conjunto PRIMERO.
- Entender el cálculo y la motivación del conjunto SIGUIENTE.
- Familiarizarnos con el cálculo de ambos conjuntos a través de ejemplos.

El conjunto PRIMERO

Como se ha mencionado en la presentación, el cálculo de los conjuntos PRIMERO **y SIGUIENTE,** es necesario para la construcción de los analizadores sintácticos, tanto ascendentes como descendentes. Cuando decimos los primeros en las cadenas que se derivan de aquella que estamos analizando, se quiere decir que si hacemos el cálculo de PRIMERO de un símbolo de la gramática, por ejemplo X, se expresaría como PRIMERO(X) y se calcularía de la manera siguiente:

1. Si X es un símbolo que pertenece al conjunto de los terminales, entonces PRIMERO(X) = {X}.
2. Si X es λ, entonces PRIMERO(X) = {λ}.
3. Si X es un no terminal y X → Y1 Y2... Yn se incluirá los que haya en PRIMERO(Y1) y si de Y1 deriva λ, además de otros terminales, se incluirán estos terminales y en lugar de λ, lo que haya en PRIMERO(Y2), y así sucesivamente hasta llegar a Yn. Solo se incluiría λ, si esta estuviera en todas las producciones que se derivan de Y1 Y2... Yn.
4. Repetir hasta que no se puedan añadir más terminales o λ a ningún conjunto PRIMERO.

Un **ejemplo de esta regla** sería:

- X → Y1 Y2
- Y_1 → a B | λ
- Y_2 → b C | λ

PRIMERO(X) = PRIMERO(Y_1), y para calcular PRIMERO(Y_1) tenemos que calcular PRIMERO(aB) y PRIMERO(λ) que son las dos producciones que se derivan de Y_1.

Regla 1 y 2	Por la regla 1, puesto que **a** pertenece al conjunto de los terminales, PRIMERO(aB) = {a}. Por la regla 2, PRIMERO(λ) = {λ}. Por tanto obtenemos PRIMERO (Y_1) = {a, λ}.
Regla 3	Al aplicar la regla 3, y puesto que en PRIMERO(Y_1) se encuentra λ, entonces hay que calcular también PRIMERO(Y_2), y sacar del cálculo de PRIMERO(X) a λ. Esto sería: {a, λ} - {λ} U PRIMERO(Y_2) = {a} U {b, λ} = {a, b, λ}. Como λ, se encuentra en todas las producciones también se incluiría (si Y_2 fuera de la siguiente forma, Y_2 → b C, y por tanto no incluyera λ, no se incluiría). Por tanto PRIMERO(X) = {a, b, λ}.

Conjunto PRIMERO

El conjunto PRIMERO es el conjunto de **terminales** que pueden aparecer los primeros en las cadenas que se derivan de aquella que estamos analizando.

Ascendentes como descendentes

También tienen aplicación en la recuperación de errores puesto que permiten a partir de ellos crear el conjunto de *tokens* de sincronización.

Ejemplo 1 de cálculo de conjunto PRIMERO

Supongamos la siguiente gramática que responde a la de *if,* incluyendo el *else* y sin incluirlo:

S → i C t S | i C t S e S | a

C → b

```
i C t S, responde a "if <condición> then <sentencia>"
i C t S e S, responde a "if <condición> then <sentencia> else <sentencia>"
```

Como vemos esta gramática tiene un problema de ambigüedad, puesto que dos producciones derivadas de S comienzan igual, lo primero que tenemos que hacer es resolverla, siendo necesario **factorizar** para eliminar esta ambigüedad. Si aplicamos la regla de factorización que hemos aprendido; nos inventamos un símbolo no terminal que añadimos al final de la parte común (por ejemplo S´) y de este a su vez se obtendrá por un lado la cadena vacía y por otro lo que resta de cadena. En este ejemplo, la gramática sería:

- S → i C t S S´ |a
- S → e S | λ
- C → b

En detalle [Reglas para el cálculo del conjunto PRIMERO]

Calcular PRIMERO(S)	Empezamos por calcular PRIMERO(S) y tenemos dos producciones que comienzan ambas por un terminal, siendo que PRIMERO(S) será igual a la unión del cálculo de ambas. PRIMERO(S) = PRIMERO(i C t S S´) U PRIMERO(a). Aplicando la regla 1, tendríamos que **PRIMERO(S) = {i, a}.**
Calcular PRIMERO (S´)	Pasamos a calcular PRIMERO (S´), donde tenemos que PRIMERO(S´) = PRIMERO(e S) U PRIMERO(λ) y el primer término de la unión, aplicando la regla 1, nos da e, mientras que el segundo término aplicando la regla 2, nos da λ. Por tanto **PRIMERO(S´) = {e, λ}.**
Calcular PRIMERO(C)	Por último vamos a calcular PRIMERO(C), donde tenemos que PRIMERO(C) = PRIMERO(b), y aplicando la regla 1, obtenemos que **PRIMERO(C) = {b}.**

En detalle

Reglas para el cálculo del conjunto PRIMERO

1. Si X es un símbolo que pertenece al conjunto de los terminales, entonces PRIMERO(X) = {X}.
2. Si X es λ, entonces PRIMERO(X) = {λ}.
3. Si X es un no terminal y X → Y1 Y2... Yn se incluirá los que haya en PRIMERO(Y1) y si de Y1 deriva λ, además de otros terminales, se incluirán estos terminales y en lugar de λ, lo que haya en PRIMERO(Y2), y así sucesivamente hasta llegar a Yn. Solo se incluiría λ, si esta estuviera en todas las producciones que se derivan de Y1 Y2... Yn.
4. Repetir hasta que no se puedan añadir más terminales o λ a ningún conjunto PRIMERO.

El conjunto SIGUIENTE

El conjunto SIGUIENTE es el conjunto de terminales que pueden aparecer justo después del no terminal para el que estamos haciendo el cálculo, a partir de alguna forma sentencial derivada de este símbolo.

Reglas para el cálculo del conjunto SIGUIENTE

1. Si S es el símbolo inicial de la gramática, entonces $ está en SIGUIENTE(S).
2. Si tenemos la producción A → α B β, se añaden a SIGUIENTE(B) todos los terminales que haya en PRIMERO(β) excepto el símbolo de la cadena vacía (λ). En el caso de que apareciera λ en PRIMERO(β) entonces se elimina λ y se incluyen los terminales que se deriven de SIGUIENTE(A).
3. Si tenemos una producción A → α B añadir a SIGUIENTE(B) los terminales que se deriven de SIGUIENTE(A).
4. Repetir para todas las producciones en las que aparezca en el lado derecho el símbolo para el que estamos calculando el conjunto.

Detalles a tener en cuenta de la aplicación de las reglas

- Nunca estará λ en el conjunto SIGUIENTE del no terminal que estemos calculando.
- Para el cálculo del conjunto SIGUIENTE hay que fijarse en la aparición del símbolo en la parte derecha de la producción (al contrario que en los conjuntos PRIMERO que mirábamos el lado izquierdo).
- Hay que mirar en todas las producciones donde aparezca el símbolo en cuestión, no solo en la que aparezca primero.

Ejemplo

Regla 3

Los **símbolos α y β** representan a cadenas de terminales y no terminales incluyendo la cadena vacía λ, lo que indica que en una producción cualquiera no tiene por qué haber nada delante o detrás del símbolo para el que hacemos el cálculo.

El **símbolo λ** no puede aparecer nunca en el conjunto SIGUIENTE.

Conjunto SIGUIENTE

El conjunto SIGUIENTE se calcula para los no terminales de la gramática (mientras que el conjunto PRIMERO incluía también a los terminales).

Conjunto de terminales

Dentro del conjunto de terminales se incluye el símbolo de fin de cadena, $ como una nueva convención, es decir que se considera que al final de cualquier cadena de entrada está este símbolo terminal.

Ejemplo

Ejemplo regla 3

Veamos con un pequeño ejemplo algunos detalles de la regla 3, a partir de esta gramática ficticia:

- S → c A d D
- A → b B
- B → a C
- C → ...
- D → B a

Realicemos el cálculo de SIGUIENTE(B). Si nos fijamos en la gramática B aparece en dos producciones en el lado derecho (A → b B y D → B a), por tanto hay que mirar las dos para hacer el cálculo.

Si miramos la producción A → bB, se aplica la regla 3:

- SIGUIENTE(B) = SIGUIENTE(A) = {d}.

Si miramos la producción D → Ba, se aplica la regla 2, donde β equivaldría a a:

- SIGUIENTE(B) = PRIMERO(a) = {a}.

Por tanto SIGUIENTE(B) = SIGUIENTE(A) U PRIMERO(a) = {d, a}.

Ejemplo 1 de cálculo del conjunto SIGUIENTE

Vamos a partir de la gramática factorizada del ejemplo 1 de cálculo del conjunto PRIMERO:

- S → i C t S S´ |a
- S´ → e S | λ
- C → b

	PRIMERO
S	i, a
S´	e, λ
C	b

Tabla 1

Donde ya hemos calculado los conjuntos PRIMERO, obteniéndose los datos de la **tabla 1.**

Cálculo del conjunto SIGUIENTE

Reglas para el cálculo del conjunto SIGUIENTE

En detalle

SIGUIENTE(S)	Por la regla 1, al ser el axioma de la gramática se incorpora $. Por la regla 2 en la producción S → iCtSS´, y mirando el lado derecho de la producción, SIGUIENTE(S) = PRIMERO(S´), se incorporarían e y λ, pero sabemos que λ no puede estar y al eliminarlo se incorporaría SIGUIENTE(S) que es precisamente el conjunto que estamos calculando, por lo que aquí se pararía. Por tanto, paso a paso sería: • SIGUIENTE(S) = {$} U PRIMERO(S´) = {$, e, λ} y aplicando la regla 2,= {$, e, λ} - λ U SIGUIENTE(S) = {$,e}.
SIGUIENTE(S´)	Aplicando la regla 3 en la producción S → i C t S S´, SIGUIENTE(S´) = SIGUIENTE(S), que ya hemos calculado. Miramos el resto de producciones y no vuelve a aparecer S´en el lado derecho. • SIGUIENTE(S´) = SIGUIENTE(S) = {$, e}.

SIGUIENTE(C)	Por la regla 2 en la producción S → i C t S S´, que es donde aparece C en el lado derecho de una producción, se incluye lo que hay en PRIMERO(t) = {t}. • SIGUIENTE(C) = PRIMERO(t)= {t}.

Finalmente obtenemos los datos de la tabla 2.

Tabla 2

	SIGUIENTE
S	$, e
S´	$, e
C	t

Tabla 2

En detalle

Reglas para el cálculo del conjunto SIGUIENTE

1. Si S es el símbolo inicial de la gramática, entonces $ está en SIGUIENTE(S).
2. Si tenemos la producción A → α B β, se añaden a SIGUIENTE(B) todos los terminales que haya en PRIMERO(β) excepto el símbolo de la cadena vacía (λ). En el caso de que apareciera λ en PRIMERO(β) entonces se elimina λ y se incluyen los terminales que se deriven de SIGUIENTE(A).
3. Si tenemos una producción A → α B añadir a SIGUIENTE(B) los terminales que se deriven de SIGUIENTE(A).
4. Repetir para todas las producciones en las que aparezca en el lado derecho el símbolo para el que estamos calculando el conjunto.

Ejemplo 2 de cálculo del conjunto PRIMERO

Vamos a realizar un ejemplo con la siguiente gramática, donde un tipo de sentencia puede ser una lista de variables y a su vez la lista puede estar compuesta por otra lista seguida de sentencias o bien una sola sentencia:

S → (L) | id

L → L, S | S (NOTA: la coma es un símbolo terminal de la gramática).

Cálculo del conjunto PRIMERO

En detalle Reglas de cálculo del conjunto PRIMERO

PRIMERO(S)	Y para ello nos vamos a la producción S → (L) | id, donde por la regla 1 al comenzar ambas producciones por terminales, PRIMERO(S)= PRIMERO((L)) U PRIMERO(id) = **{(, id}.**
PRIMERO(L)	Nos fijamos en la producción L → S L´ y vemos que el primer símbolo de lado derecho es el no terminal S, por tanto aplicando la regla 3, PRIMERO(L) = PRIMERO(S) = **{(, id}.**
PRIMERO(L´)	Nos fijamos en las dos producciones que se derivan de L´ que son L´ →, S L´ y L´ → λ. Aplicando la regla 1 a la primera producción PRIMERO(L´) = PRIMERO(, S L´) = {,} y al aplicar la regla 2 a la segunda producción, obtenemos que PRIMERO(L´) = PRIMERO(λ) = λ. Si lo unimos todo, tenemos que PRIMERO(L´) = **{,, λ}.**

Finalmente nos quedan los datos de la **tabla 3.**

	PRIMERO
S	(, id
L	(, id
L´	, , λ

Tabla 3

Gramática

Al revisar la gramática, vemos que debido a L → L, S es recursiva por la izquierda y por tanto hay que eliminar esta ambigüedad. Si recordamos, este tipo de ambigüedades se resolvían de la siguiente forma:

- A → Aα | β, se resuelve → A → β A´
- A´ → α A´| λ

Aplicándoselo a L → L, S | S, nos queda resuelta la ambigüedad de la siguiente forma:

- L → S L´
- L´→ ,S L´| λ

Resultando la gramática sin ambigüedades:

- S → (L) | id
- L → S L´
- L´ → ,S L´| λ

En detalle

Reglas para el cálculo del conjunto PRIMERO

1. Si X es un símbolo que pertenece al conjunto de los terminales, entonces PRIMERO(X) = {X}.
2. Si X es λ, entonces PRIMERO(X) = {λ}.
3. Si X es un no terminal y X → Y1 Y2... Yn se incluirá los que haya en PRIMERO(Y1) y si de Y1 deriva λ, además de otros terminales, se incluirán estos terminales y en lugar de λ, lo que haya en PRIMERO(Y2), y así sucesivamente hasta llegar a Yn. Solo se incluiría λ, si esta estuviera en todas las producciones que se derivan de Y1 Y2... Yn.
4. Repetir hasta que no se puedan añadir más terminales o λ a ningún conjunto PRIMERO.

Ejemplo 2 de cálculo del conjunto SIGUIENTE

A partir de la gramática sin ambigüedades anterior, y del conjunto PRIMERO, vamos a hacer el cálculo de los conjuntos SIGUIENTE.

Cálculo de los conjuntos SIGUIENTE

SIGUIENTE(S)

SIGUIENTE(L)

SIGUIENTE(L´)

Gramática	Conjunto PRIMERO
S → (L) | id	
L → S L´	
L´→ , S L´| λ	

	PRIMERO
S	(, id
L	(, id
L´	, , λ

SIGUIENTE(S)

Puesto que S es el axioma de la gramática, por la regla 1, se incluye el símbolo $. Ahora nos fijamos en el resto de producciones que tienen S en el lado derecho y nos aparece L → S L´, donde aplicando la regla 2, SIGUIENTE(S) = PRIMERO(L´) = {, , λ} pero cuando está λ, hay que quitarla e incluir a SIGUIENTE(L).

Para calcular SIGUIENTE(L) miramos el lado derecho de las producciones donde L aparece y solo tenemos S → (L), donde el símbolo justo detrás de L es un terminal, por tanto PRIMERO de un terminal es ese terminal, en este caso el cierre de paréntesis). La otra producción donde aparece S en el lado derecho es L´→, S L´, donde aparece justo detrás de S el mismo símbolo, L´ al que acabamos de calcular su conjunto PRIMERO. Por tanto, de forma resumida nos queda:

- SIGUIENTE(S) = {$} U PRIMERO(L´) = {$} U {, , λ} – λ U SIGUIENTE(L) = {$, , ,)} U PRIMERO(L´) (que ya se ha calculado, con lo que no varía el resultado).

SIGUIENTE(L)

Hemos hecho el cálculo al ver SIGUIENTE(S). Fijándonos en la producción que tiene L en su lado derecho S → (L), nos aparece como símbolo terminal el cierre de paréntesis). Por tanto:

- SIGUIENTE(L) = {) }

SIGUIENTE(L)

L → S L´

L´→ , S L´| λ

S	(, id
L	(, id
L´	, , λ

SIGUIENTE(L´)

SIGUIENTE(L´)

También la hemos calculado ya, y aplicando la regla 3 en la producción L → S L´, SIGUIENTE(L´) = SIGUIENTE(L) = {) }.

Miramos también la otra producción donde aparece L´, L´→, S L´ y vemos que SIGUIENTE(L´) = SIGUIENTE(L´), con lo que el resultado no varía. Por tanto:

- SIGUIENTE(L´) = {) }

SIGUIENTE(L´)

Finalmente nos quedan los datos de la tabla 4.

Tabla 4

	SIGUIENTE
S	$, ,)
L	)
L´	)

Tabla 4

Reglas para el cálculo del conjunto SIGUIENTE

1. Si S es el símbolo inicial de la gramática, entonces $ está en SIGUIENTE(S).
2. Si tenemos la producción A → α B β, se añaden a SIGUIENTE(B) todos los terminales que haya en PRIMERO(β) excepto el símbolo de la cadena vacía (λ). En el caso de que apareciera λ en PRIMERO(β) entonces se elimina λ y se incluyen los terminales que se deriven de SIGUIENTE(A).
3. Si tenemos una producción A → α B añadir a SIGUIENTE(B) los terminales que se deriven de SIGUIENTE(A).
4. Repetir para todas las producciones en las que aparezca en el lado derecho el símbolo para el que estamos calculando el conjunto.

Resumen

En este tema hemos aprendido a calcular los conjuntos PRIMERO y SIGUIENTE utilizando las reglas para ello. Es importante recordar que para el cálculo del conjunto PRIMERO intervienen tanto terminales como no terminales, así como la cadena vacía (λ) mientras que para el conjunto SIGUIENTE solo intervienen no terminales y no puede aparecer en su resultado la cadena vacía.

El **cálculo del conjunto PRIMERO** sigue las siguientes reglas:

1. Si X es un símbolo que pertenece al conjunto de los terminales, entonces PRIMERO(X) = {X}.
2. Si X es λ, entonces PRIMERO(X) = {λ}.
3. Si X es un no terminal y X → Y_1 Y_2... Y_n se incluirá los que haya en PRIMERO(Y_1) y si de Y_1 deriva λ, además de otros terminales, se incluirán estos terminales y en lugar de λ, lo que haya en PRIMERO(Y_2), y así sucesivamente hasta llegar a Y_n. Solo se incluiría λ, si esta estuviera en todas las producciones que se derivan de Y_1 Y_2... Y_n.

El **cálculo del conjunto SIGUIENTE** sigue las siguientes reglas:

1. Si S es el símbolo inicial de la gramática, entonces $ está en SIGUIENTE(S).
2. Si tenemos la producción A → α B β, se añaden a SIGUIENTE(B) todos los terminales que haya en PRIMERO(β) excepto el símbolo de la cadena vacía (λ). En el caso de que apareciera λ en PRIMERO(β) entonces se elimina λ y se incluyen los terminales que se deriven de SIGUIENTE(A).
3. Si tenemos una producción A → α B añadir a SIGUIENTE(B) los terminales que se deriven de SIGUIENTE(A).