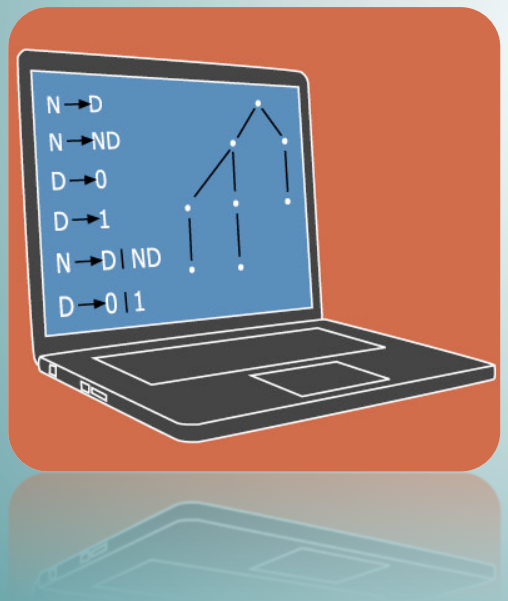
Procesadores de Lenguaje

Analizadores sintácticos ascendentes



Cristina Tirnauca

Domingo Gómez Pérez

Dpto. de Matemáticas,

Estadística y Computación

Análisis LR

En esta clase vamos a introducir el análisis ascendente, que se basa en como encontrar una derivación más a la derecha.

Análisis LR

Veamos un ejemplo de la idea que tenemos para hallar una derivación de la palabra (()) con la siguiente gramática: $S \to$ (S) | ().

- (())
- (S) $\to$ (())
- $S \to$ (S) $\to$ (())

Análisis LR (leyendo secuencialmente)

Veamos un ejemplo de la idea que tenemos para hallar una derivación de la palabra (()) con la siguiente gramática:
$S \to$ (S) | ().

- Leemos ((.
- Sustituimos ((por (S.
- Leemos el siguiente símbolo de la cinta y reducimos (S) por S.

Análisis LR (leyendo secuencialmente)

Cojamos esta otra gramática $S \to S\ (\ S\ ) \mid ()$,

- ¿Qué debemos hacer para aplicar esta idea a la palabra ()(())?
- ¿Cómo es el árbol de derivación de está palabra?
- ¿Qué pasa cuando recortamos algunas hojas del árbol de derivación? (empezando por la izquierda)

Pequeño repaso de derivaciones

Los analizadores LL recorren los datos de izquierda a derecha y construyen la derivación izquierda: en cada momento, la pila contiene lo que se espera "encontrar" en la cinta.

Los sucesivos pasos del analizador van leyendo la cinta y va modificando la pila para que tenga un resumen de lo que queda por ver.

Existe una estrategia alternativa: el análisis LR, que recorre los datos de izquierda a derecha pero construye la derivación derecha.

Sólo que la hemos de construir "al revés", empezando por el final: en cada momento, la concatenación del contenido de la pila con lo que queda por leer reconstruyen una derivación a la derecha de la palabra en órden inverso.

Analizadores ascendentes LR y autómatas con pila

Algunos convenios notacionales

Convenio:

Emplearemos un símbolo de "dólar" \$ para indicar el fin de datos; este convenio es histórico y tradicional.

Convenio:

Emplearemos un símbolo de "euro" € para indicar el fondo de la pila; este convenio no es ni histórico y ni tradicional, como es obvio.
Sólo lo emplearemos provisionalmente.

Operaciones básicas:

- Desplazar (*shift*) un símbolo terminal de la entrada a la pila;
- Reducir (*reduce*) una parte derecha de regla que hay en la cima de la pila, sustituyéndola por la correspondiente parte izquierda de la misma regla.

Analizadores ascendentes LR y autómatas con pila

Una visión intuitiva

Vamos a hacer "trampa" por ahora, nuestros autómatas con pila van a poder,

- Conocer el próximo símbolo de la cinta sin leerlo.
- Leer más de un elemento de la cinta sin perderlos.

Estos "superautómatas" son versiones de otros autómatas con pila.

Analizadores ascendentes LR y autómatas con pila

Una visión intuitiva

Ejemplo sencillo con la gramática de N y D:

N : D | N D

D : '0' | '1' | '2' | '3' | '4' | '5' | '6' | '7' | '8' | '9'

Colocar € en el fondo de la pila

Repetir:

1. si la cima de la pila es un dígito, reducir a D
2. si la cima de la pila es N D, reducir por N : N D
3. si la cima de la pila es € D, reducir por N : D
4. si la cima de la pila es € N y viene $
   (es decir, estamos al final de los datos), aceptar
5. si la cima de la pila es N y no viene $, desplazar
6. si la cima de la pila es € y no viene $, desplazar

Analizadores ascendentes LR y autómatas con pila

La historia se complica

Complicamos el ejemplo con la gramática de expresiones:

E : N | E OP E

N : D | N D

D : '0' | '1' | '2' | '3' | '4' | '5' | '6' | '7' | '8' | '9'

OP : '+' | '*'

Fijémonos que sólo tenemos dos operadores y además aún no hay paréntesis.

Analizadores ascendentes LR

Encontrando la forma general

El analizador sintáctico para esta gramática es:

Colocar € en el fondo de la pila

Repetir:

1. si la cima de la pila es un dígito, reducir a D
2. si la cima de la pila es '+' o '∗', reducir a OP
3. si la cima de la pila es N D, reducir por N : N D
4. si la cima de la pila es € D, reducir por N : D
5. si la cima de la pila es OP D, reducir por N : D
6. si la cima de la pila es € E y viene $, aceptar
7. si la cima de la pila es N y viene un operador, reducir por E : N
8. si la cima de la pila es N y viene $, reducir por E : N
9. si la cima de la pila es N y viene un dígito, desplazar
10. si la cima de la pila es OP y viene un dígito, desplazar
11. si la cima de la pila es € y viene un dígito, desplazar
12. si la cima de la pila es E OP E, reducir por E : E OP E
13. si la cima de la pila es E y viene un operador, desplazar

Analizadores ascendentes LR

Breve resumen de lo que sabemos

Características principales:

- Construyen un árbol sintáctico de las hojas hacia la raíz.
- Leen los datos de izquierda a derecha (*Left to right*) y construyen la derivación derecha (*Rightmost*); pero la construyen "al revés".
- Emplean una pila para mantener un "resumen" de lo que llevan visto hasta el momento: a cada vuelta del bucle principal, la parte ya leída de la entrada se puede derivar de lo que hay en la pila.
- Operaciones básicas (analizadores *shift-reduce*):
  - Desplazar (*shift*) un símbolo terminal de la entrada a la pila;
  - Reducir (*reduce*) una parte derecha de regla que hay en la cima de la pila, sustituyéndola por la correspondiente parte izquierda de la misma regla.
- En la pila hay tanto terminales como símbolos gramaticales.

Analizadores ascendentes LR

¿Por qué funcionan?

Propiedad fundamental que precisamos:

Un analizador sintáctico ha de aceptar si y solamente si la entrada se puede generar por la gramática.

En un analizador LR, aceptaremos cuando:

- Hayamos leído toda la entrada y
- en ese momento en la pila únicamente haya el símbolo inicial.

Por ello, cuando aceptamos, estamos seguros de que la gramática puede generar la entrada.

Pero... ¿Y viceversa?

Analizadores ascendentes LR

Ideas generales de lo que veremos a continuación

Cuestiones a tratar:

1. ¿Cómo queda el algoritmo de análisis LR en forma general?
   - Guiado por una tabla de análisis que indica si desplazar o reducir por una regla u otra, en función de la cima de la pila y el siguiente terminal.
   - Las dificultades estarán en cómo construir la tabla.
2. ¿No tendremos una piedra en el zapato si alguna regla tiene como parte derecha la palabra vacía?
3. ¿Qué va a ocurrir en caso de que la gramática (o, en general, el lenguaje a analizar) no admite autómatas a pila deterministas?

¡Jugarán roles estelares el FIRST y el FOLLOW...!

Tablas de análisis LR

Versión preliminar aproximada

El mismo ejemplo sencillo, en forma tabular:

	dígito	$
€	desplazar	
dígito	reducir a D	reducir a D
€ N	desplazar	aceptar
N D	reducir por N : N D	reducir por N : N D
€ D	reducir por N : D	reducir por N : D

N : D | N D

D : '0' | '1' | '2' | '3' | '4' | '5' | '6' | '7' | '8' | '9'

Tablas de análisis LR

Versión preliminar aproximada, requiere ajustar más

	dígito	operador	$
dígito	reducir a D	reducir a D	reducir a D
'+', '∗'	reducir a OP	reducir a OP	reducir a OP
N	desplazar	reducir N a E	reducir N a E
OP, €	desplazar		
N D	reducir a N	reducir a N	reducir a N
OP D	reducir D a N	reducir D a N	reducir D a N
€ D	reducir D a N	reducir D a N	reducir D a N
€ E			aceptar
E OP E	reducir a E	reducir a E	reducir a E
E		desplazar	

E : N | E OP E

N : D | N D

D : '0' | '1' | '2' | '3' | '4' | '5' | '6' | '7' | '8' | '9'

OP : '+' | '∗'