

Objetivo del módulo

El objetivo de este módulo es implementar un compilador muy simple. Realmente solo llegamos hasta el código intermedio, es decir, el front end.(análisis)

En este capítulo hay:

1. Un lenguaje fuente sencillo.
2. Un lenguaje destino cercano al fuente.
3. Sin optimización.
4. Sin back end dependiente de la máquina.
5. Sin herramientas.
6. Poca teoría.

El material se presentará demasiado rápido para una comprensión completa.

Introducción

Nos enfocaremos en el front end, es decir, la parte de análisis de un compilador.

- La **sintaxis** describe la forma de un programa en un lenguaje dado.
- La **semántica** describe el significado del programa.
- Usaremos la representación estándar para la sintaxis: **gramáticas libres de contexto (GLC)** o **BNF (Backus-Naur Form)**.
- Veremos **traducción dirigida por la sintaxis**: la gramática se complementa con atributos para guiar el front end.

El lenguaje fuente será un **pseudocódigo** con dígitos, operadores, comandos y estructuras de control.

El destino será el mismo, traducido a **Python con el módulo Turtle**.

Pasos:

- Tokenizar la entrada (scanner).
 - Modelar la sintaxis.
 - Usar la sintaxis para traducir a la representación intermedia (árbol).
-

Definición de sintaxis

Gramáticas

Una **GLC** consta de:

1. Conjunto de terminales (tokens del analizador léxico).
2. Conjunto de no terminales.
3. Conjunto de producciones ($LHS \rightarrow RHS$).
4. Símbolo inicial.

Ejemplo:

- Terminales: 0–9, +, -
- No terminales: lista, digito
- Producciones:
 - lista \rightarrow lista + digito
 - lista \rightarrow lista - digito
 - lista \rightarrow digito
 - digito \rightarrow 0 | 1 | ... | 9
- Símbolo inicial: lista

Si no se especifica, el primero es el símbolo inicial.

Derivaciones

Ejemplo: generar 7+4-5 a partir de lista.

Este proceso se llama **derivación**. El conjunto de todas las cadenas derivables es el **lenguaje generado por la GLC**.

Notas:

- Solo permite expresiones infijas con un dígito como operando.
 - Las expresiones posibles dependen de las producciones elegidas.
 - No hay espacios ni cadena vacía (a menos que se añada ε).
 - El lenguaje de entrada del compilador se aproxima al generado por la gramática.
-