UNIVERSIDAD NACIONAL DE EDUCACIÓN A DISTANCIA

Escuela Técnica Superior de Ingeniería Informática

Procesadores de Lenguajes

Tema 2

Análisis Léxico

Javier Vélez Reyes
jvelez@lsi.uned.es

Objetivos del Tema

- Conocer el funcionamiento de un analizador léxico
- Entender las relaciones de éste con
  - La tabla de símbolos
  - La gestión de errores
- Aprender a especificar formalmente un analizador
- Conocer las distintas técnicas de implementación

Javier Vélez Reyes jvelez@lsi.uned.es

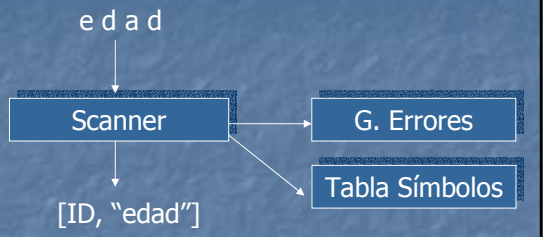
Índice General

- Introducción
- Especificación de un Analizador Léxico
- Implementación de un Analizador Léxico

Javier Vélez Reyes jvelez@lsi.uned.es

Introducción

- Análisis Léxico
  - Scanner
  - Tabla de Símbolos
  - Gestión de Errores
- Funciones
  - Tratar con la tabla de símbolos
  - Generar tokens bajo demanda del analizador sintáctico
  - Manejar el fichero fuente
  - Ignorar comentarios
  - Contabilizar posición de tokens
  - Preprocesar macros, constantes, includes…



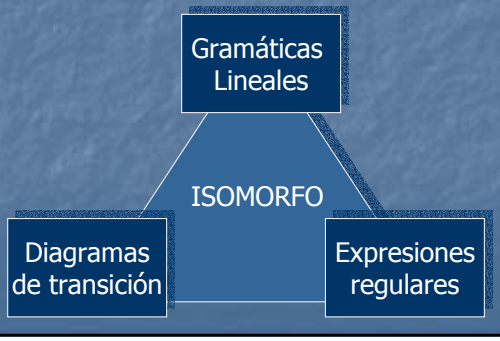
Javier Vélez Reyes jvelez@lsi.uned.es

Índice General

Javier Vélez Reyes jvelez@lsi.uned.es

Especificación de analizador léxico I

- Especificación formal
  - Gramáticas Lineales
    - Recursivas a izquierdas
    - Recursivas a derechas
  - Lenguajes (expresiones) regulares
  - Autómatas finitos (Diagramas de Transición)



Javier Vélez Reyes jvelez@lsi.uned.es

Gramáticas Linealmente Recursivas

- Gramáticas Linealmente Recursivas
  - Alfabeto terminal
  - Alfabeto no terminal
  - Axioma
  - Reglas de producción
- Tipos
  - Recursividad a izquierdas

```
S  :=  A letra | A digito
A  :=  A letra | A digito | letra
```

  - Recursividad a derechas

```
S  :=  letra A
A  :=  letra A | digito A | letra | digito
```

Javier Vélez Reyes jvelez@lsi.uned.es

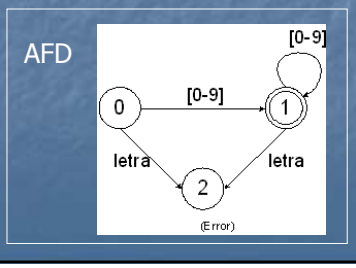
Expresiones regulares

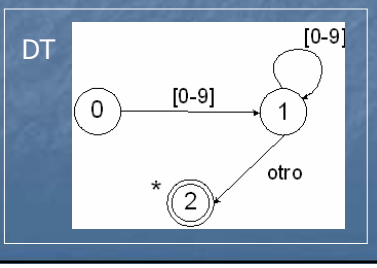
- Expresiones regulares
  - Alfabeto
  - Cierre simétrico + (1 o más)
  - Cierre transitivo * (0 o más)
- Ejemplos
  - $a^*(b \mid c)^+$
    - bccbcbccbc
    - abbbccc
    - abbcbcccbbc
  - $(0\text{-}9)^*.(0\text{-}9)^+$
    - 0.236425
    - 3567.45627
    - .758478
    - 5.0

Javier Vélez Reyes jvelez@lsi.uned.es

Diagramas de transición

- Diferencias con autómatas finitos
  - Acciones asociadas a ciertos estados
    - Estados de Aceptación
    - Estados con Retroceso (*)
  - No tienen estados de absorción (se omiten)
  - Los estados de aceptación no tienen transiciones
  - Transición especial **otro**





Javier Vélez Reyes jvelez@lsi.uned.es

Especificación de analizador léxico II

- Términos utilizados
  - Token
    - Elemento léxico del lenguaje
    - Símbolo No Terminal de las fases siguientes
  - Patrón
    - Expresión regular que define el lenguaje
    - Letra (Letra | Dígito)*
  - Lexema
    - Secuencia de caracteres que concuerda con un patrón
    - numeroUsuarios
  - Atributos
    - Estructura de datos de cada token para almacenarse en la TS
    - Depende del tipo de token
    - [ID, Lexema, Tipo, Valor, línea]

Javier Vélez Reyes jvelez@lsi.uned.es

Especificación de analizador léxico III

- Pasos para especificar un analizador léxico
  - Identificar la colección de tokens
  - Estructurar la colección de tokens
  - Describir el lenguaje como expresiones regulares
  - Especificar un Diagrama de Transición
  - Traducir el Diagrama a una tabla de transición
- Ejemplo

Especificar un analizador léxico que reconozca

- Números enteros
- Operadores aritméticos y de incremento (+, - *, /, ++,--)
- Identificadores
- WHILE

Javier Vélez Reyes jvelez@lsi.uned.es

Índice General

Javier Vélez Reyes jvelez@lsi.uned.es

Implementación del analizador léxico I

- Estrategias de implementación
  - Implementación automática
    - Especificación de los patrones de la gramática
    - LEX
  - Implementación manual del diagrama de transiciones
    - Simulación de transiciones del diagrama
  - Implementación manual directa
    - Codificación con estructuras condicionales (if, case, …)
  - Implementación híbrida
    - Análisis directo de las estructuras más sencillas
      - Operadores
    - Análisis mediante diagrama de estructuras complejas
      - Cadenas no específicas
      - Prefijos comunes

Javier Vélez Reyes jvelez@lsi.uned.es

Implementación del analizador léxico II

- Prioridad de tokens
  - Dar prioridad al token con lexema más largo
    - DO / DOT
    - > / >=
  - En generadores automáticos (LEX)
    - Anteponer el patrón para el token más largo
    - Después el más corto

Javier Vélez Reyes jvelez@lsi.uned.es

Implementación del analizador léxico III

- Reconocimiento de palabras reservadas
  - Resolución explícita
    - Se indican todos los patrones de cada palabra reservada
    - Se integran en el diagrama de transiciones global
    - Complejidad mayor
    - LEX
  - Resolución implícita
    - Considerar las palabras reservadas como identificadores
    - Insertar las palabras clave como tal en la tabla de símbolos
    - Buscar identificadores en la tabla de símbolos

Javier Vélez Reyes jvelez@lsi.uned.es

Implementación del analizador léxico III

- Gestión de errores
  - Visión muy local del programa
- Tipos de Errores
  - Caracteres inválidos ñ, ç,…
  - Ausencia de concordancia con patrones
- Recuperación de errores
  - Ignorar caracteres no válidos hasta formar token
  - Borrar caracteres extraños
  - Insertar carácter que falta
  - Reemplazar un carácter por otro
  - Conmutar posición de dos caracteres adyacentes

Javier Vélez Reyes jvelez@lsi.uned.es

Bibliografía

[AJO] AHO, SETHI, ULLMAN: *Compiladores: Principios, técnicas y herramientas*,: Addison-Wesley Iberoamericana, 1990

[GARRIDO] A. Garrido, J. Iñesta, F. Moreno y J. Pérez. 2002. *Diseño de compiladores*. Universidad de Alicante.