

# GABRIELE ROMANATO

## Dettagli di uno scanner lessicale

Uno scanner lessicale, noto anche come lexer o analizzatore lessicale, è un componente fondamentale di un compilatore o interprete. Il suo compito principale è quello di leggere una sequenza di caratteri di input e suddividerla in unità significative chiamate **token**.

## Funzionamento

Il processo di analisi lessicale si basa su una grammatica regolare, che definisce le regole per identificare i token. Lo scanner utilizza tipicamente *espressioni regolari* per riconoscere le sequenze di caratteri che costituiscono token validi. Ogni token è associato a una categoria, come parole chiave, identificatori, operatori, numeri o simboli di punteggiatura.

## Struttura di uno Scanner

Uno scanner è composto da diverse componenti principali:

- **Tabella dei token:** contiene la lista dei token riconosciuti.
- **Automatismi di riconoscimento:** spesso realizzati tramite automi a stati finiti.
- **Routine di gestione degli errori:** servono a trattare input non validi o imprevisti.

## Esempio di Tokenizzazione

Dato l'input:

```
if (x > 10) return x;
```

Lo scanner lo suddivide nei seguenti token:

- if – parola chiave
- ( – simbolo
- x – identificatore
- > – operatore
- 10 – numero
- ) – simbolo
- return – parola chiave
- x – identificatore
- ; – simbolo

## **Ruolo nello Sviluppo del Linguaggio**

Lo scanner lessicale è la prima fase del processo di compilazione e ha un impatto diretto sull'efficienza del parsing successivo. Una buona progettazione dello scanner migliora la chiarezza della sintassi e facilita l'individuazione di errori di scrittura nel codice sorgente.

## **Strumenti per la Generazione di Scanner**

Esistono strumenti automatici per la generazione di scanner, come:

- Lex (in ambiente Unix)
- Flex (versione migliorata di Lex)
- Generatori integrati nei moderni parser (come ANTLR)

Questi strumenti permettono di definire le regole lessicali in modo dichiarativo, semplificando la costruzione di compilatori.