

## El enum TokenType - Clasificando el Universo del Lenguaje

Objetivo: Entender que el primer paso en el análisis de un lenguaje es definir sus “categorías gramaticales”. Comprender el rol de las enumeraciones (enum) para crear código robusto y legible.

### La Filosofía de la Clasificación y la Analogía Lingüística

- ¿Cómo abordamos un problema complejo? Dividiéndolo. Antes de entender una frase completa, nuestro cerebro identifica sus partes: sustantivos, verbos, adjetivos, etc. No nos importa *qué* verbo es, solo que *es* un verbo.
- El análisis léxico es, en esencia, un problema de clasificación.
- **Analogía Central:**
  - **Lenguaje Natural:** “La niña amaba mucho a su papá” -> [ARTÍCULO, NOMBRE COMÚN, VERBO, ADVERBIO, PREPOSICIÓN, POSESIVO, NOMBRE COMÚN]
  - **Pseudocódigo:** AVANZA 100 -> [INSTRUCCION, NUMERO]
- **Evitar alto acoplamiento o las “Magic Strings” (Cadenas Mágicas)**
  - Ejemplo de código terrible: 

```
if (token.type == "INSTRUCCION") {  
    ...  
}
```
  - Debilidades:
    1. **Errores de Típo:** `if (token.type == "INSTRUCCION")` no da error de compilación, pero falla en ejecución.
    2. **Falta de Autocompletado:** El IDE no puede hacer sugerencias de autocompletado.
    3. **Refactorización Difícil:** Si se requiere cambiar el nombre “INSTRUCCION” a “KEYWORD”, se debe buscar y reemplazar en todo el proyecto, con riesgo de error.

### Introducción al enum como Solución

- **Concepto de Diseño: Tipado Fuerte (Strong Typing).** Un enum crea un nuevo tipo de dato. Una variable de tipo TokenType solo puede contener los valores definidos.
- **Análisis del Código TokenType.py:**
  - `class TokenType(Enum):` Proporciona una forma de crear enumeraciones, que son conjuntos de nombres simbólicos (miembros) vinculados a valores constantes únicos.
  - **INSTRUCCION:** La categoría para las acciones (AVANZA, GIRA\_DERECHA). Es el “verbo” del lenguaje.
  - **NUMERO, CADENA:** Categorías para datos literales. Son los “complementos” de la acción.
  - **PAREN\_IZQ, PAREN\_DER, COMA:** Delimitadores. Son los “signos de puntuación” que estructuran expresiones complejas.
  - **ID:** El identificador. Es el “sustantivo propio” (nombre de una variable).
  - **EOF (End of File):** De importancia crítica. ¿Cómo sabe el siguiente componente (el parser) que ha terminado de leer? EOF es la señal universal. Sin él, el parser podría entrar en un bucle infinito esperando más tokens.

### Actividad Práctica y Expansión del Lenguaje

- **Ejercicio Mental:** Añadir nueva funcionalidad al lenguaje: condicionales simples como `SI <condicion> ENTONCES ... FIN_SI`.
- **Preguntas para la Discusión:**
  1. ¿Qué nuevos valores se necesita añadir al enum TokenType?
  2. Si se requiere añadir los colores ROJO, VERDE, AZUL como palabras clave ¿serían un nuevo TokenType o simplemente lexemas del token INSTRUCCION?
- **Conclusión de la Actividad:** El enum es el ADN del lenguaje. Define los bloques de construcción fundamentales.

- **Resumen:** Hemos definido las categorías léxicas del lenguaje y entendido por qué un enum es la herramienta algorítmica y de diseño correcta para esta tarea, garantizando seguridad y legibilidad.
- **Avance a la siguiente sesión:** “Ya se tienen los ‘cajones’ para clasificar las palabras. En la próxima sesión, se construirá la ‘etiqueta’ que se asociará a cada palabra que se encuentre, conteniendo no solo su tipo, sino también su valor y dónde se encuentra: el Token.”