

TRABAJO PRÁCTICO GRUPAL

Consigna - Primera parte:

Generar un **analizador léxico o lexer con ANTLR** llamado ***LexerConsumos.g4*** que permita reconocer los tokens de un archivo de entrada como el presentado (Consumos.xml). Para ello deberán definirse una serie de reglas léxicas (expresiones regulares) con el nombre indicado a continuación.

- Reglas para reconocer tokens que son cadenas “fijas” o literales, es decir, para cualquier consulta de consumos de línea telefónica móvil, son siempre iguales:
 1. CONSULTA_ABRE: representa la cadena “<consulta_consumos>”
 2. CONSULTA_CIERRA: representa la cadena “</consulta_consumos>”
 3. LINEA_ABRE: representa la cadena “<num_linea>”
 4. LINEA_CIERRA: representa la cadena “</num_linea>”
 5. PLAN_ABRE: representa la cadena “<plan>”
 6. PLAN_CIERRA: representa la cadena “</plan>”
 7. CODCLI_ABRE: representa la cadena “<codigo_cliente>”
 8. CODCLI_CIERRA: representa la cadena “</codigo_cliente>”
 9. NOMCLI_ABRE: representa la cadena “</nombre_cliente>”
 10. NOMCLI_CIERRA: representa la cadena “</nombre_cliente>”
 11. DESDE_ABRE: representa la cadena “<desde>”
 12. DESDE_CIERRA: representa la cadena “</desde>”
 13. HASTA_ABRE: representa la cadena “<hasta>”
 14. HASTA_CIERRA: representa la cadena “</hasta>”
 15. CONS_ABRE: representa la cadena “<consumos>”
 16. CONS_CIERRA: representa la cadena “</consumos>”
 17. REGCON_ABRE: representa la cadena “<registro_consumo>”
 18. REGCON_CIERRA: representa la cadena “</registro_consumo>”
 19. NUMDEST_ABRE representa la cadena “<numero_destino>”
 20. NUMDEST_CIERRA representa la cadena “</numero_destino>”
 21. TARIFA_ABRE representa la cadena “<tarifa>”
 22. TARIFA_CIERRA representa la cadena “</tarifa>”
 23. FECHA_ABRE representa la cadena “<fecha>”
 24. FECHA_CIERRA representa la cadena “</fecha>”
 25. HORA_ABRE representa la cadena “<hora>”
 26. HORA_CIERRA representa la cadena “</hora>”
 27. DURACION_ABRE representa la cadena “<duracion>”
 28. DURACION_CIERRA representa la cadena “</duracion>”
 29. IMPORTE_ABRE representa la cadena “<importe>”
 30. IMPORTE_CIERRA representa la cadena “</importe>”
 31. TOTAL_ABRE representa la cadena “<total_consumos>”
 32. TOTAL_CIERRA representa la cadena “</total_consumos>”
- Reglas que reconocen tokens que tienen un formato dado y pueden estar conformados por diferentes cadenas o secuencias de símbolos.
 33. NUM_LINEA: es una cadena de símbolos numéricos que corresponde al número de línea del cliente, formada por 4 secuencias de 3 dígitos separadas por guiones medios “-” entre éstas (999-999-999-999).
 34. PLAN: es una de las siguientes cadenas de texto “PLAN 3G”, “PLAN 5G”, “PLAN 8G” o “PLAN EMPRESA”.
 35. CODIGO_CLI: es una cadena de símbolos numéricos y “-”, cuyo formato es XXXXXX-XX.
 36. NOMBRE_CLI: es una cadena de símbolos alfabéticos que corresponden al nombre del cliente formado por el apellido, seguido de coma (,) como separador, seguido de espacio en blanco, y

el/los nombres, todo en mayúscula. Pueden existir otros espacios en blanco si posee más nombres.

37. FECHA: es una cadena de texto de longitud 10 que representa una fecha, con formato dd/mm/aaaa. Considerar que el año debe estar comprendido entre 1900 y 2100
38. TARIFA: es una de las siguientes cadenas de texto "LLAMADA NACIONAL", "LLAMADA INTERNACIONAL", "MENSAJE DE TEXTO" o "DESCARGA DE CONTENIDO".
39. HORA: es una cadena de texto de 8 dígitos que representa una hora, con el formato hh:mm:ss.
40. DURACION: es una cadena numérica entre 0 y 999.
41. IMPORTE: es una cadena numérica, que posee el "." como separador de miles y la "," como separador de decimales. Siempre debe tener dos decimales y al menos un cero en la parte entera.

Se debe tener en cuenta que:

- El archivo de entrada no contiene palabras con acento o que contengan la letra ñ.
- Se pueden ignorar los saltos de línea y las tabulaciones. Sólo se incluyen en el archivo de entrada por cuestiones de legibilidad.

El analizador léxico deberá imprimir un mensaje por cada token reconocido y mostrarlo siguiendo las siguientes reglas:

- Para los tokens enumerados del 1 al 32, se deberá imprimir el nombre del token identificado.
- Para los tokens enumerados del 33 al 41, se deberá imprimir el nombre del token identificado seguido por el valor del token (separado por dos puntos).

Por ejemplo, para el siguiente fragmento del archivo de entrada

```
...
<registro_consumo>
  <numero_destino>342-155-677-992</numero_destino>
  <tarifa>LLAMADA NACIONAL</tarifa>
  <fecha>06/03/2020</fecha>
  <hora>10:45:04</hora>
  <duracion>326</duracion>
  <importe>37,49</importe>
...
```

se imprimirán los mensajes que se listan a continuación (se incluyen saltos de línea y espacios para mejorar la visualización de la salida):

```
...
REGCON_ABRE
NUMDEST_ABRE NUM_LINEA: 342-155-677-992 NUMDEST_CIERRA
TARIFA_ABRE TARIFA: LLAMADA NACIONAL TARIFA_CIERRA
FECHA_ABRE FECHA: 06/03/2020 FECHA_CIERRA
HORA_ABRE HORA: 10:45:04 HORA_CIERRA
DURACION_ABRE DURACION: 326 DURACION_CIERRA
IMPORTE_ABRE IMPORTE: 37,49 IMPORTE_CIERRA
...
```

Si le resulta más fácil, puede trabajar con las herramientas de ANTLRWorks (Run->Interpret Lexer y Windows->Lexer Debugger Controller) de manera de controlar que los tokens se reconocen correctamente.