**PROYECTO #2. ANALIZADOR LÉXICO**

|  |
| --- |
| **Desarrollo:** Individual.  **Fecha de Presentación:** Jue 28/junio/2018, 18:30 horas. |

Desarrollar un Analizador Léxico (Analex) para un lenguaje muy parecido al PASCAL, tomando en cuenta que el mismo **NO** es **case-sensitive**. Los token a reconocer son:

|  |  |
| --- | --- |
| **LEXEMA(S)** | **TOKEN** |
| “program” | <PROGRAM,\_> |
| “var” | <VAR,\_> |
| “procedure” | <PROCEDURE,\_> |
| “begin” | <BEGIN,\_> |
| “end” | <END,\_> |
| “if” | <IF,\_> |
| “then” | <THEN,\_> |
| “else” | <ELSE,\_> |
| “for” | <FOR,\_> |
| “to” | <TO,\_> |
| “do” | <DO,\_> |
| “downto” | <DOWNTO,\_> |
| “while” | <WHILE,\_> |
| “case” | <CASE,\_> |
| “of” | <OF,\_> |
| “repeat” | <REPEAT,\_> |
| “until” | <UNTIL,\_> |
| “read” | <READ,\_> |
| “readln” | <READLN,\_> |
| “write” | <WRITE,\_> |
| “writeln” | <WRITELN,\_> |

|  |  |
| --- | --- |
| **LEXEMA(S)** | **TOKEN** |
| “,” | <COMA,\_> |
| “:” | <DOSPTOS,\_> |
| “;” | <PTOCOMA,\_> |
| “(“ | <PA,\_> |
| “)” | <PC,\_> |
| “:=” | <ASSIGN,\_> |
| “not” y “!” | <NOT,\_> |
| “and” y “&” | <AND,\_> |
| “or” y “|” | <OR,\_> |
| “+” | <MAS,\_> |
| “-“ | <MENOS,\_> |
| “\*” | <POR,\_> |
| “mod” y “%” | <MOD,\_> |
| “div” y “/” | <DIV,\_> |
| (números enteros sin signo) | <NUM,-3> |
| (Empiezan con Letra y luego le sigue una combinación de Letras y Dígitos) | **<ID, -1>** |
| (chars encerrados entre comillas “..”, sin EOLN ni EOF) | **<STRINGctte, 0>** |
| “=”  “<”  “>”  “<=”  “>=”  “<>” | <OPREL,IGUAL>  <OPREL,MEN>  <OPREL,MAY>  <OPREL,MEI>  <OPREL,MAI>  <OPREL,DIS> |
| “boolean”  “integer” | <TIPO, BOOLEAN>  <TIPO, INTEGER> |

Puesto que su Analex no trabajará con la Tabla de Símbolos, para el token **ID** use como atributo –1 , y para el **STRINGctte** el 0: **<ID, -1>, <STRINGctte, 0>.**

|  |
| --- |
| **Comentarios**. Los comentarios que su Analex, debe ignorar son:   * Comentarios de línea: //….. * Comentarios multilínea: {….} y (\* …\*) |