**TRABAJO PRÁCTICO INTEGRADOR**

**Sintaxis y Semántica de los Lenguajes**

**Ingeniería en Sistemas de Información**

**Primer Cuatrimestre**

****

**UNIVERSIDAD TECNOLÓGICA NACIONAL**

**Integrantes:**

* ***GONZALEZ, Micaela***
* ***MAZZUCHINI UBALDI, Guido Juliano***
* ***MOLINAS GONZALEZ, Victor Adan***
* ***TRESPALACIOS MARTINEZ, Carlos Daniel***

**Año lectivo: 2023**

**Fecha de entrega: 30/04/23**

**Grupo: G3**

**ÍNDICE**

[**INTRODUCCIÓN 2**](#_yjszcxpf4hxb)

[**GRAMÁTICA 3**](#_12046x67ogki)

[Terminales : 3](#_chyrdqo9rto1)

[No Terminales : 5](#_jnrf5n5cbbfy)

[**PRODUCCIONES 6**](#_iylp0u5uyl6w)

[**Analizador Léxico 13**](#_y16gzxsbd4zb)

# **INTRODUCCIÓN**

En este trabajo práctico integrador, nuestro objetivo es aplicar los conocimientos teóricos y técnicas aprendidas sobre gramática, lenguajes de programación y procesamiento léxico y sintáctico. En particular, nos centraremos en su aplicación en la creación, diseño y procesamiento de lenguajes.

Con este fin, vamos a analizar y verificar la gramática de un metalenguaje muy utilizado llamado XML, que se utiliza para almacenar datos de manera legible y crear documentación técnica en Internet. En concreto, nos enfocaremos en la versión de DocBook.

En este trabajo también desarrollaremos un lexer y parser utilizando Python junto a la librería PLY, con el objetivo de procesar la gramática. Como resultado, se obtendrá un archivo HTML, el cual reflejará el contenido del XML de DocBook.

# **GRAMÁTICA**

## **Terminales :**

* <?xml
* <!DOCTYPE article>
* version
* encoding
* <article>
* </article>
* <title>
* </title>
* <info>
* </info>
* <abstract>
* </abstract>
* <author>
* </author>
* <personname>
* </personname>
* <firstname>
* </firstname>
* <surname>
* </surname>
* <date>
* </date>
* <year>
* </year>
* <copyright>
* </copyright>
* <address>
* </address>
* <city>
* </city>
* <state>
* </state>
* <postcode>
* </postcode>
* <street>
* </street>
* <email>
* </email>
* <phone>
* </phone>
* <section>
* </section>
* <simplesect>
* </simplesect>
* <itemizedlist>
* </itemizedlist>
* <listitem>
* </listitem>
* <emphasis>
* </emphasis>
* <para>
* </para>
* <simpara>
* </simpara>
* <comment>
* </comment>
* <important>
* </important>
* <informaltable>
* </informaltable>
* <tgroup>
* </tgroup>
* <thead>
* </thead>
* <tfoot>
* </tfoot>
* <tbody>
* </tbody>
* <row>
* </row>
* <entry>
* </entry>
* <entrytbl>
* </entrytbl>
* <holder>
* </holder>
* <mediaobject>
* </mediaobject>
* <videoObject>
* </videoObject>
* <ImageObject>
* </ImageObject>
* <videodata

## **No Terminales :**

* TITLE
* ART
* INFO
* AUT
* PER
* FIR
* SUR
* SEC
* PARA

**Tokens:**

* URL: protocolo://dominio:puerto/ruta#fragmento
* TEXTO: Cadena de caracteres que no incluye los caracteres <,> ni &.
* NUMERO: Cadena numérica.
* CODING: Tipos de codificaciones (ISO-8859-1 /UTF-8 y otros tipos de codificaciones).
* RUTA: Directorio de archivos imagedata y videodata

# 

# **PRODUCCIONES**

**Bloque Principal:**

**TOKEN INICIO:**

**Σ** <?xml version=”NUMERO.NUMERO” encoding=”CODING”> DOC ART

**DOC** <!DOCTYPE article>

**ART** <article> ETIQUETAS </article>

**Casos generales** :

**ETIQUETAS**  INFO TITLE ETICOM SECT

**ETIQUETAS**  INFO TITLE ETICOM SSEC

**ETIQUETAS**  INFO TITLE ETICOM

**ETIQUETAS**  TITLE ETICOM SECT

**ETIQUETAS**  TITLE ETICOM SSEC

**ETIQUETAS**  TITLE ETICOM

**ETIQUETAS**  INFO ETICOM SECT

**ETIQUETAS**  INFO ETICOM SSEC

**ETIQUETAS**  INFO ETICOM

**ETIQUETAS**  ETICOM SECT

**ETIQUETAS**  ETICOM SSEC

**ETIQUETAS**  ETICOM

**INFO:**

**INFO** <info> ETIQUETAINFO </info>

**ETIQUETAINFO**  ETICOM2 ETIQUETAINFO

**ETIQUETAINFO**  ETICOM2

**ETIQUETAINFO**  AUT ETIQUETAINFO

**ETIQUETAINFO**  DATA ETIQUETAINFO

**ETIQUETAINFO**  COPY ETIQUETAINFO

**ETIQUETAINFO**  TITLE ETIQUETAINFO

**ETIQUETAINFO**  AUTHOR

**ETIQUETAINFO**  DATA

**ETIQUETAINFO**  COPY

**ETIQUETAINFO**  TITLE

**Section:**

**SECT** <section> INFO TITLE ETICOM </section> SECT

**SECT** <section> INFO ETICOM </section> SECT

**SECT** <section> TITLE ETICOM </section> SECT

**SECT** <section> INFO TITLE ETICOM </section>

**SECT** <section> INFO ETICOM </section>

**SECT** <section> TITLE ETICOM </section>

**BUCLE CON SIMPLE SECTIONS INTERNOS:**

**SECT** <section> INFO TITLE ETICOM SSEC </section> SECT

**SECT** <section> INFO ETICOM SSEC </section> SECT

**SECT** <section> TITLE ETICOM SSEC </section> SECT

**SECT** <section> INFO TITLE ETICOM SSEC </section>

**SECT** <section> INFO ETICOM SSEC </section>

**SECT** <section> TITLE ETICOM SSEC </section>

**BUCLE DE SECTIONS INTERNOS:**

**SECT** <section> INFO TITLE ETICOM SECT </section> SECT

**SECT** <section> INFO ETICOM SECT </section> SECT

**SECT** <section> TITLE ETICOM SECT </section> SECT

**SECT** <section> INFO TITLE ETICOM SECT </section>

**SECT** <section> INFO ETICOM SECT </section>

**SECT** <section> TITLE ETICOM SECT </section>

**SimpleSec:**

**SSEC**<simplesect> INFO TITLE </simplesect>

**SSEC** <simplesect> INFO </simplesect>

**SSEC**<simplesect> INFO TITLE ETISECT </simplesect>

**SSEC** <simplesect> INFO ETICOM </simplesect>

**SSEC**<simplesect> INFO TITLE </simplesect> SSEC

**SSEC** <simplesect> INFO </simplesect> SSEC

**SSEC**<simplesect> INFO TITLE ETICOM </simplesect> SSEC

**SSEC** <simplesect> INFO ETICOM </simplesect> SSEC

**ETIQUETAS COMPARTIDAS ART SECT SSEC IMPORTANT:**

**ETICOM**  PARA ETICOM

**ETICOM**  PARA

**ETICOM**  SIMPARA ETICOM

**ETICOM**  SIMPARA

**ETICOM**  COMMENT ETICOM

**ETICOM**  COMMENT

**ETICOM**  ETICOMP ETICOM

**ETICOM**  ETICOMP

—-----------------------------------

**ETICOMP**  INFTABLE ETICOMP

**ETICOMP**  INFTABLE

**ETICOMP**  ITEMLIST ETICOMP

**ETICOMP**  ITEMLIST

**ETICOMP**  IMPORTANT ETICOMP

**ETICOMP**  IMPORTANT

**ETICOMP**  ETICOM2 ETICOM

**ETICOMP**  ETICOM2

**ETIQUETAS COMPARTIDAS ART SECT SSEC INFO:**

**ETICOM2**  MEDIAOBJ ETICOM2

**ETICOM2** MEDIAOBJ

**ETICOM2**  ABSTRACT ETICOM2

**ETICOM2**  ABSTRACT

**ETICOM2**  ADR ETICOM2

**ETICOM2**  ADR

**ETIQUETAS COMPARTIDAS Street, City, State, Phone, Email, Date, Year ,Holder**

**ETICOM3**TEXTO ETICOM3

**ETICOM3**TEXTO

**ETICOM3**LINK ETICOM3

**ETICOM3**LINK

**ETICOM3**EMPHASIS ETICOM3

**ETICOM3**EMPHASIS

**ETICOM3**COMMENT ETICOM3

**ETICOM3**COMMENT

**ETIQUETAS COMPARTIDAS Para, SimPara, Emphasis, Comment, Link**

**ETICOM4**TEXTO ETICOM4

**ETICOM4**TEXTO

**ETICOM4**EMPHASIS ETICOM4

**ETICOM4**EMPHASIS

**ETICOM4**LINK ETICOM4

**ETICOM4**LINK

**ETICOM4**EMAIL ETICOM4

**ETICOM4**EMAIL

**ETICOM4**AUTHOR ETICOM4

**ETICOM4**AUTHOR

**ETICOM4**COMMENT ETICOM4

**ETICOM4**COMMENT

**ETIQUETAS COMPARTIDAS PersonName**

**ETIAUT**  FIRSTNAME ETIAUT

**ETIAUT**  FIRSTNAME

**ETIAUT**  SURNAME ETIAU

**ETIAUT**  SURNAME

**ETIQUETAS COMPARTIDAS Address**

**ETIADD**  TEXTO ETIADD

**ETIADD**  TEXTO

**ETIADD**  STREET ETIADD

**ETIADD**  STREET

**ETIADD**  CITY ETIADD

**ETIADD**  CITY

**ETIADD**  STATE ETIADD

**ETIADD**  STATE

**ETIADD**  PHONE ETIADD

**ETIADD**  PHONE

**ETIADD**  EMAIL ETIADD

**ETIADD**  EMAIL

**ETIQUETAS COMPARTIDAS entry**

**ETIENTRY**  TEXTO ETIENTRY

**ETIENTRY**  TEXTO

**ETIENTRY**  ITEMLIST ETIENTRY

**ETIENTRY**  ITEMLIST

**ETIENTRY**  IMPORTANT ETIENTRY

**ETIENTRY**  IMPORTANT

**ETIENTRY**  PARA ETIENTRY

**ETIENTRY**  PARA

**ETIENTRY**  SIMPARA ETIENTRY

**ETIENTRY**  SIMPARA

**ETIENTRY**  MEDIAOBJ ETIENTRY

**ETIENTRY**  MEDIAOBJ

**ETIENTRY**  COMMENT ETIENTRY

**ETIENTRY**  COMMENT

**ETIENTRY**  ABSTRACT ETIENTRY

**ETIENTRY**  ABSTRACT

**SIMPARA:**

**SIMPARA**<simpara> ETICOM4 </simpara>

**EMPHASIS:**

**EMPHASIS**<emphasis> ETICOM4 </emphasis>

**COMMENT:**

**COMMENT**  <comment> ETICOM4 </comment>

**LINK:**

**LINK**<link xlink:href = “URI”>ETICOM4</link>

**IMGD**  <imagedata fileref=”RUTA” />

**VIDEOD**  <videodata fileref=”RUTA” />

Será utilizado en una sola expresión regular VIDIMALINK

**PARA:**

**PARA** <para> ETICOM4 ETICOMP </para> PARA

**PARA** <para> ETICOM4 </para>

**PARA** <para> ETICOMP </para>

**TITLE:**

**TITLE**  <title> TEXTO </title>

**LIST:**

**ITEMLIST**  <itemizedlist> LISITEM </itemizedlist>

**LISTEM**  <listitem> ETICOM </listitem> LISITEM

**LISTEM** <listitem> ETICOM </listitem>

**IMPORTANT:**

**IMPORTANT** <important> TITLE ETICOM </important>

**IMPORTANT** <important> ETICOM </important>

**DATA:**

**FIRSTNAME**  <firstname> ETICOM3 </firstname>

**SURNAME**  <surname> ETICOM3 </surname>

**STREET**  <street> ETICOM3 </street>

**CITY**  <city> ETICOM3 </city>

**STATE**  <state> ETICOM3 </state>

**PHONE**  <phone> ETICOM3 </phone>

**EMAIL**  <email> ETICOM3 </email>

**DATE**  <date> ETICOM3 </date>

**YEAR**  <year> ETICOM3 </year> YEAR

**YEAR**  <year> ETICOM3 </year>

**HOLDER**  <holder> ETICOM3 </holder> HOLDER

**HOLDER**  <holder> ETICOM3 </holder>

**ADR**  <address> ETIADD </address>

**AUTHOR**  <author> PERSONNAME </author>

**PERSONNAME**  <personname> ETIAUT </personname>

**COPYRIGHT**  <copyright> YEAR HOLDER</copyright>

**COPYRIGHT**  <copyright> YEAR </copyright>

**POSTCODE**  <postcode>NUMERO</postcode>

**MULTIMEDIA:**

**MEDIAOBJ**  <mediaobject> INFO OBJ </mediaobject>

**MEDIAOBJ**  <mediaobject> OBJ </mediaobject>

**VIDOEOBJ**  <videoObject> INFO VIDEOD </videoObject>

**VIDOEOBJ**  <videoObject> VIDEOD </videoObject>

**IMGOBJ**  <ImageObject> INFO IMGD </ImageObject>

**IMGOBJ**  <ImageObject> IMGD </ImageObject>

**OBJ**  VIDOEOBJ OBJ | IMGOBJ OBJ

**OBJ**  VIDOEOBJ | IMGOBJ

**TABLA:**

**INFTABLE**  <informaltable> MEDIAOBJ INFTABLE </informaltable>

**INFTABLE**  <informaltable> MEDIAOBJ </informaltable>

**TGROUP**  <tgroup> THEAD TFOOT TBODY </tgroup>

**TGROUP**  <tgroup> THEAD TBODY </tgroup>

**TGROUP**  <tgroup> TFOOT TBODY </tgroup>

**TGROUP**  <tgroup> TBODY </tgroup>

**THEAD**  <thead> ROW </thead>

**THEAD**  <thead> ROW THEAD </thead>

**TFOOT**  <tfoot> ROW </tfoot>

**TFOOT**  <tfoot> ROW TFOOT </tfoot>

**TBODY**  <tbody> ROW </tbody>

**TBODY**  <tbody> ROW TBODY </tbody>

**ROW**  <row> ENTRY ENTRYTBL </row>

**ROW**  <row> ENTRY </row>

**ROW**  <row> ENTRYTBL </row>

**ENTRY**  <entry> ETIENTRY </entry> ENTRY

**ENTRY**  <entry> ETIENTRY </entry>

**ENTRYTBL**  <entrytbl> THEAD TBODY </entrytbl> ENTRYTBL

**ENTRYTBL**  <entrytbl> TBODY </entrytbl> ENTRYTBL

**ENTRYTBL**  <entrytbl> THEAD TBODY </entrytbl>

**ENTRYTBL**  <entrytbl> TBODY </entrytbl>

# **Analizador Léxico**

El siguiente analizador léxico está basado en el lenguaje Python con el uso de la librería **ply**. A su vez, trabaja con funciones con expresiones regulares tales como :

***t\_URI:***

Es utilizada para detectar URLs dentro de un documento DocBook.

Esta expresión regular busca coincidir con cadenas que contengan una URL compuesta de la siguiente manera : *protocolo://dominio:puerto/ruta#fragmento*.

***t\_VIDIMALINK:***

Es utilizada para detectar los tags de imagedata, videodata, y link respectivamente.

La expresión regular también contiene otra definición interna de una URL, debido a que no fue posible llamar una función dentro de la expresión regular.

Coincide con cadenas con los siguientes formatos:

*<videodata fileref="http://example.com/video.mp4" />*

*<imagedata fileref="http://example.com/image.jpg" />*

*<link xlink:href="http://example.com" >*

***t\_INICIO:***

Es utilizada para detectar el xml del docbook, con su respectiva versión, encoding o si se le quiere simplemente el tag: *<?xml version="1.0" encoding="UTF-8"?>*

***t\_DOCTYPE:***

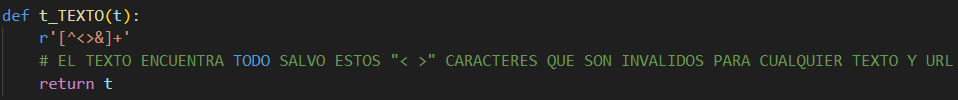
******

Esta función toma el tag de Doctype que contiene el docbook:

<!DOCTYPE article>

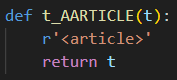
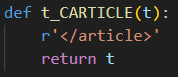
<!DOCTYPE>

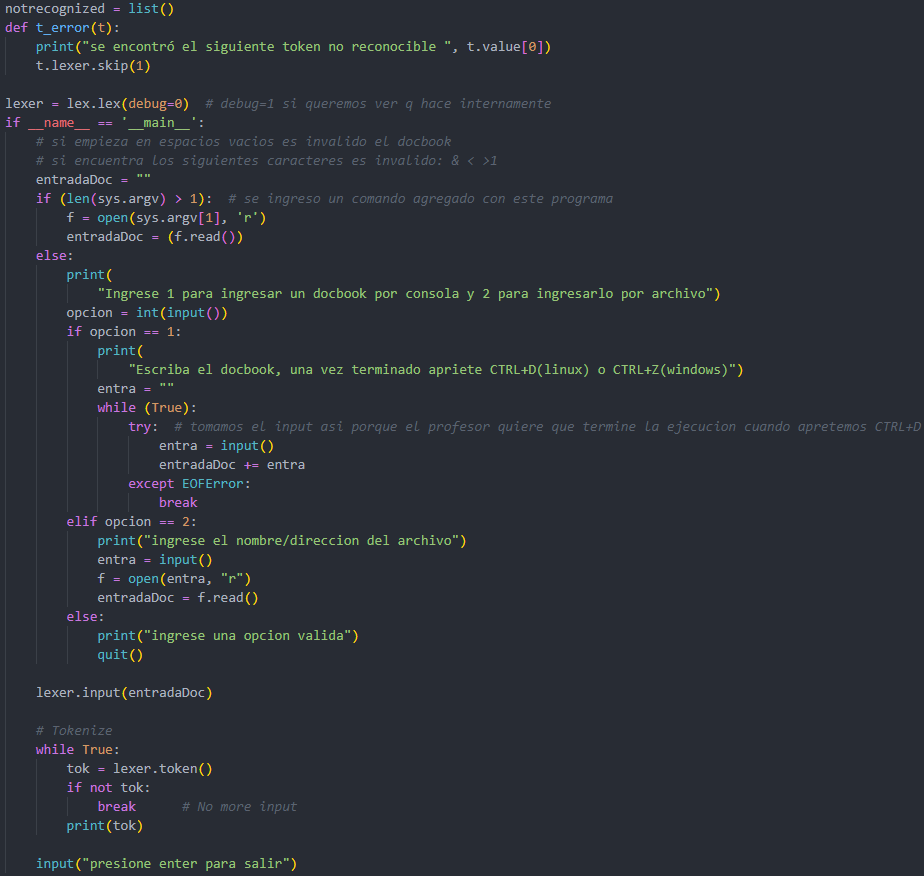
***t\_TEXTO:***

******

Las funciones restantes definen los tags de apertura y clausura, por ejemplo:

| Para la apertura: | Para la clausura: |
| --- | --- |





***EJEMPLOS DE DOCBOOK:***

<!DOCTYPE article PUBLIC "-//OASIS//DTD DocBook V4.1//EN">

<article>

<author>

<firstname>Jose</firstname>

<surname>Pérez</surname>

</author>

<title>Una aproximación analítica a <phrase>Hola Mundo</phrase> </title>

<date>13 de Junio de 2002<date>

<copyright>

<year>2002</year>

<holder>Jose Pérez</holder>

</copyright>

<date>15 de Mayo de 2002</date>

<revision>

<revnumber>1.1</revnumber>

<date>13 de Junio de 2002</date>

<revremark>Inclusión de capítulo Hello World</revremark>

</revision>

<section>

<title>Hola Mundo</title>

</section>

</article>

***EJEMPLO 2 DE DOCBOOK:***

<!DOCTYPE article>

<article>

<info>

<title>El titulo del articulo</title>

<author>

<firstname>Juan</firstname>

<surname>Perez</surname>

</author>

</info>

<sect1>

<title>Titulo para la seccion 1</title>

<para>

Esto es un parrafo

</para>

<para>

Otro parrafo.

</para>

</sect1>

</article>

***EJEMPLO 3 DE DOCBOOK:***

<!DOCTYPE article>

<article>

<section>

<info>

<title>El titulo del articulo</title>

</section>

<author>

<firstname>Juan</firstname>

<surname>Perez</surname>

</author>

</info>

<title>Titulo para la seccion 1</title>

<para>

Esto es un parrafo

<para>

</para>

Otro parrafo.

</para>

</article>