

# **Sintaxis y semántica de los lenguajes**

## **Trabajo Práctico Integrador: Diseño e implementación de Lexer y Parser**

**Ciclo lectivo: 2023  
Primer cuatrimestre**

### **Integrantes Grupo 19 ISI B:**

Acosta Quintana, Lautaro

Galeano, Martín

Niveyro, Iván

Morel, Francisco

### **Localidad:**

Resistencia, Chaco

### **Fecha primera entrega:**

30/04/23

### **Universidad:**

Universidad Tecnológica Nacional, Facultad Regional Resistencia



## Contenido

<b>Contenido</b>	<b>1</b>
<b>1. Introducción</b>	<b>2</b>
<b>2. Pseudocódigo</b>	<b>2</b>
Pseudocódigo inicial:	5
Pseudocódigo final:	10
<b>3. Gramática</b>	<b>15</b>
Tabla de Terminales (Tokens):	15
Tabla de No terminales:	18
<b>4. Analizador Léxico</b>	<b>21</b>
<b>5. Analizador sintáctico</b>	<b>22</b>
<b>6. Funciones utilizadas</b>	<b>22</b>
<b>7. Modo de obtención del intérprete</b>	<b>24</b>
<b>8. Modo de ejecución del intérprete</b>	<b>24</b>
<b>9. Ejemplos presentados por el grupo:</b>	<b>25</b>
<b>10. Conclusión</b>	<b>33</b>
<b>11. Bibliografía</b>	<b>33</b>



## 1. Introducción

El trabajo consiste en la construcción de un analizador léxico y un analizador sintáctico con el fin de analizar un archivo de entrada, de formato y extensión .xml, y validarlo para posteriormente traducirlo a otro archivo con formato y extensión .html.

Para lograrlo, primero realizamos el diseño de las distintas reglas de construcción con las que se evaluará la entrada del programa. Este diseño está realizado en pseudocódigo y representa cómo se comportaría el analizador sintáctico.

A partir del análisis sobre el pseudocódigo generado para la gramática, definimos todos los Terminales (Tokens) y No terminales que usaremos para los ambos analizadores.

Para llevar a cabo la implementación del trabajo, elegimos el lenguaje de programación C (estándar C17), usando el compilador GCC (11.3.0) junto al generador de analizadores léxicos [Flex](#) (2.6.4) y el generador de analizadores sintácticos [Bison](#) (3.8.2).

## 2. Pseudocódigo

En esta sección se presenta el pseudocódigo [inicial](#) en el que se basará el analizador sintáctico y el pseudocódigo [final](#) con el cual se implementó el analizador sintáctico.

Los terminales están escritos en minúscula y los no terminales en *Camel/Case*. Con la intención de facilitar la lectura, los no terminales están coloreados en azul y las producciones gramaticales están divididas en secciones como en el pdf de guía.

Tabla de Tokens	
Token	Contenido
texto	Letras, números, signos de puntuación y caracteres especiales
url	Dirección web.
<article>; </article>	Etiqueta de inicio y final de un bloque articulo



<info>; </info>	Etiqueta de inicio y final de un bloque info
<title>; </title>	Etiqueta de inicio y final de un bloque title
<abstract>; </abstract>	Etiqueta de inicio y final de un bloque abstract
<para>; </para>	Etiqueta de inicio y final de un bloque para
<author>; </author>	Etiqueta de inicio y final de un author
<personname>; </personname>	Etiqueta de inicio y final de un nombre completo de persona
<firstname>; </firstname>	Etiqueta de inicio y final del primer nombre de una persona
<surname>; </surname>	Etiqueta de inicio y final del segundo nombre de una persona
<date>; </date>	Etiqueta de inicio y final para una fecha
<section>; </section>	Etiqueta de inicio y final de una sección
<simplesect>;</simplesect>	Etiqueta de inicio y final de una sección simple
<copyright>; </copyright>	Etiqueta de inicio y final para bloque de copyright
<address>; </address>	Etiqueta de inicio y final de una dirección
<city>; </city>	Etiqueta de inicio y final del nombre de una ciudad
<state>; </state>	Etiqueta de inicio y final del nombre de una provincia/estado
<postcode>; </postcode>	Etiqueta de inicio y final de un código postal



<street>; </street>	Etiqueta de inicio y final del nombre de una calle
<email>; </email>	Etiqueta de inicio y final de una dirección de correo
<phone>; </phone>	Etiqueta de inicio y final de un número de correo
<itemizedlist>; </itemizedlist>	Etiqueta de inicio y final para una lista itemizada
<listitem>; </listitem>	Etiqueta de inicio y final para una lista
<emphasis>; </emphasis>	Etiqueta de inicio y final de bloque emphasis
<holder>; </holder>	Etiqueta de inicio y final de bloque holder
<simpara>; </simpara>	Etiqueta de inicio y final de bloque simpara
<year>; </year>	Etiqueta de inicio y final para un número de año
<comment>; </comment>	Etiqueta de inicio y final para realizar comentarios
<important>; </important>	Etiqueta de inicio y final de bloque important
<link>; </link>	Etiqueta de inicio y final del bloque link
<videodata fileref ="; />	Etiqueta para adjuntar un archivo del tipo video
<imagedata fileref="; />	Etiqueta para adjuntar un archivo del tipo imagen
<informaltable>; </informaltable>	Etiqueta de inicio y final de bloque



	informaltable
<tgroup>; </tgroup>	Etiqueta de inicio y final de bloque tgroup
<row>; </row>	Etiqueta de inicio y final para indicar el número de filas de una tabla
<table>; </table>	Etiqueta de inicio y final para crear una tabla
<thead>; </thead> <tfoot>; </tfoot> <tbody>; </tbody>	Etiquetas de inicio y final para la dimensiones de una tabla
<entrytbl>; </entrytbl>	Etiqueta de inicio y final de bloque entrytbl
<entry>; </entry>	Etiqueta de inicio y final de bloque entry
<link xlink:href: ; />	Etiqueta para URL's

## Pseudocódigo inicial:

### Etiquetas estructurales:

```
Σ → ?xml version="1.0" encoding="UTF-8"?> Article
Article → <article>Info Title Content Section</article>
| <article>Info Title Content SimSection</article>
| <article>Info Title Content</article>
| <article>Info Content Section</article>
| <article>Info Content SimSection</article>
| <article>Info Content</article>
| <article>Title Content Section</article>
| <article>Title Content SimSection</article>
| <article>Title Cuerpo</article>
| <article>Content Section</article>
| <article>Content SimSection</article>
| <article>Content</article>

Content → ItemizedList Content | ItemizedList
| Important Content | Important
| Para Content | Para
```



SimPara Content	SimPara
Address Content	Address
MediaObject Content	MediaObject
InformalTable Content	InformalTable
Comment Content	Comment
Abstract Content	Abstract

### Secciones:

```
Section → <section> Info Title Content SimSection</section>
| <section>Info Title Content Section</section>
| <section>Info Content Section</section>
| <section>Info Content SimSection</section>
| <section>Title Content Section</section>
| <section>Title Content SimSection</section>
| <section>Content Section</section>
| <section>Content SimSection</section>
| <section>Content </section>
```

```
SimSection → <simplesect>Info Title Content</simplesect>
| <simplesect>Info Content</simplesect>
| <simplesect>Title Content</simplesect>
| <simplesect>Content</simplesect>
```

### Etiquetas básicas de párrafo:

InfoContent → Title InfoContent	Title
InfoContent MediaObject	MediaObject
InfoContent Abstract	Abstract
InfoContent Address	Address
InfoContent Author	Author
InfoContent Date	Date
InfoContent Copyright	Copyright

```
Info → <info> InfoContent </info>
```

```
AbstractContent → Para AbstractContent | Para
| SimPara AbstractContent | SimPara
```

```
Abstract → <abstract> Title AbstractContent </abstract>
| <abstract> AbstractContent </abstract>
```



```
AddressContent → texto AddressContent | texto
| Street AddressContent | Street
| City AddressContent | City
| State AddressContent | State
| Phone AddressContent | Phone
| Email AddressContent | Email
```

```
Address → <address>AddressContent</address>
```

```
AuthorContent → Firstname AuthorContent
| Surname AuthorContent
```

```
Author → <author> AuthorContent </author>
```

```
CopyrightYearContent → Year CopyrightYearContent
| Year CopyrightHolderContent
| Year
```

```
CopyrightHolderContent → Holder CopyrightHolderContent | Holder
```

```
Copyright → <copyright>CopyrightYearContent</copyright>
```

```
TitleContent → Emphasis TitleContent | Emphasis
| Link TitleContent | Link
| Email TitleContent | Email
| texto TitleContent | texto
```

```
Title → <title>TitleContent</title>
```

```
SimParaContent → Emphasis | texto
| Link | Email
| Author | Comment
```

```
SimPara → <simpara>SimPara SimParaContent</simpara>
| <simpara>SimParaContent</simpara>
```

```
Emphasis → <emphasis>Emphasis SimParaContent</emphasis>
| <emphasis>SimParaContent</emphasis>
```

```
Comment → <comment>Comment SimParaContent</comment>
| <comment>SimParaContent</comment>
```

```
Link → <link>Link SimParaContent</link>
```





| <link>SimParaContent</link>

ParaContent	→	Emphasis		Link
Email		Author		Comment
Important		Address		MediaObject
				ItemizedList
				InformalTable

Para → <para>Para ParaContent</para>  
| <para>ParaContent</para>

Important → <important>Title Content</important>  
| <important>Content</important>

SharedContent	→	SharedContent	Comment		Comment
SharedContent	Emphasis		Emphasis		
SharedContent	Link		Link		
SharedContent	texto		texto		

FirstName → <firstname>SharedContent</firstname>  
Surname → <surname>SharedContent</surname>  
Street → <street>SharedContent</street>  
City → <city>SharedContent</city>  
Phone → <phone>SharedContent</phone>  
Email → <email>SharedContent</email>  
Date → <date>SharedContent</date>  
Year → <year>SharedContent</year>  
Holder → <holder>SharedContent</holder>  
State → <state>SharedContent</state>

### Imágenes y multimedia:

MediaObjectContent → VideoObject MediaObjectContent | VideoObject  
| ImageObject MediaObjectContent | ImageObject

MediaObject → <mediaobject>Info VideoObject  
MediaObjectContent</mediaobject>  
| <mediaobject>Info ImageObject MediaObjectContent</mediaobject>  
| <mediaobject>ImageObject MediaObjectContent</mediaobject>  
| <mediaobject>VideoObject MediaObjectContent</mediaobject>  
| <mediaobject>VideoObject</mediaobject>  
| <mediaobject>VideoObject</mediaobject>

ImageObject→<imageobject> Info ImageData</imageobject>  
|<imageobject>ImageData</imageobject>



```
VideoObject→<videoobject>Info VideoData</videoobject>  
|<videoobject>VideoData</videoobject>
```

```
VideoData → <videodata fileref ="Ruta" />  
ImageData → <imagedata fileref="Ruta" />  
Ruta → /texto Ruta/ | texto.extension
```

#### Listas:

```
ItemizedList→ <itemizedlist>ListItem</itemizedlist>  
| <itemizedlist>ListItem ItemizedList</itemizedlist>  
  
ListItem → <listitem>ListItem Content</listitem>  
| <listitem>Content</listitem>
```

#### Tablas:

```
InformalTableContent → MediaObject InformalTableContent | MediaObject  
| Tgroup InformalTableContent | Tgroup
```

```
InformalTable →<informaltable>InformalTableContent</informaltable>
```

```
Tgroup → <tgroup>Thead Tbody Tfoot</tgroup>  
| <tgroup>Thead Tfoot</tgroup>  
| <tgroup>Tbody Tfoot</tgroup>  
| <tgroup>Tfoot </tgroup>
```

```
TableContent → Row TableContent | Row  
Thead→<thead>TableContent</thead>  
Tbody→<tbody>TableContent</tbody>  
Tfoot→<tfoot>TableContent</tfoot>
```

```
RowContent → Entry RowContent | Entry  
| EntryTbl RowContent | EntryTbl
```

```
Row → <row>RowContent</row>
```

```
EntryContent → texto EntryContent | texto  
| ItemizedList EntryContent | ItemizedList  
| Important EntryContent | Important  
| Para EntryContent | Para  
| SimPara EntryContent | SimPara  
| MediaObject EntryContent | MediaObject  
| Comment EntryContent | Comment  
| Abstract EntryContent | Abstract
```



```
Entry → <entry>EntryContent</entry>

Entrytbl → <entrytbl>Thead Tbody</entrytbl>
Entrytbl → <entrytbl>Tbody</entrytbl>
```

#### Enlaces:

```
Link → <link xlink:href: Url />
Url → Protocolo://Dominio:Puerto/Ruta#LocalizadorInterno
| Protocolo://Dominio:Puerto/Ruta
| Protocolo://Dominio:Puerto
| Protocolo://Dominio/Ruta # LocalizadorInterno
| Protocolo://Dominio/Ruta
| Protocolo://Dominio#LocalizadorInterno
| Protocolo://Dominio

Protocolo → http | https | ftp | ftps
Dominio → texto
Puerto → número
LocalizadorInterno → texto
```

#### Pseudocódigo final:

El pseudocódigo que utilizamos para la implementación del analizador sintáctico fue cambiando y modificándose a lo largo de la realización del trabajo. Por esto, nos pareció importante mostrar la versión final de la gramática a implementar.

#### Etiquetas estructurales:

```
Σ → <!DOCTYPE article> Article

Article → <article>Info Titledoc Content Recusection</article>

Recusection → λ
| Section Recusection
| Simsection Recusection

Content → ItemizedList Content
| Important Content
| Para Content
| SimPara Content
| Address Content
| MediaObject Content
| InformalTable Content
| ItemizedList
| Important
| Para
| SimPara
| Address
| MediaObject
| InformalTable
```



	Comment Content		Comment
	Abstract Content		Abstract

#### Secciones:

**Section** → <section> Info Titlesec Content Recusection</section>

**SimSection** → <simplesect>Info Titlesec Content</simplesect>

#### Etiquetas básicas de párrafo:

**InfoContent** →

MediaObject InfoContent		MediaObject
Abstract InfoContent		Abstract
Address InfoContent		Address
Author InfoContent		Author
Date InfoContent		Date
Copyright InfoContent		Copyright

**Info** →  $\lambda$   
| <info> Titlesec InfoContent </info>

**AbstractContent** → Para AbstractContent | Para  
| SimPara AbstractContent | SimPara

**Abstract** → <abstract> Titlesec AbstractContent </abstract>

**AddressContent** → texto AddressContent | texto

Street AddressContent		Street
City AddressContent		City
State AddressContent		State
Phone AddressContent		Phone
Email AddressContent		Email
Postcode AddressContent		Postcode

**Address** → <address>AddressContent</address>

**AuthorContent** → Firstname AuthorContent  
| Surname AuthorContent  
| Personname AuthorContent  
| Surname AuthorContent  
| Personname



**Author** → `<author> AuthorContent </author>`

**CopyrightYearContent** → `Year CopyrightYearContent`  
| `Year CopyrightHolderContent`  
| `Year`

**CopyrightHolderContent** → `Holder CopyrightHolderContent` | `Holder`

**Copyright** → `<copyright>CopyrightYearContent</copyright>`  
| `<copyright> texto </copyright>`

**TitleContent** → `Emphasis TitleContent` | `Emphasis`  
| `Link TitleContent` | `Link`  
| `Email TitleContent` | `Email`  
| `texto TitleContent` | `texto`

**Titlesec** → `λ`  
| `<title>TitleContent</title>`

**Titledoc** → `λ`  
| `<title>TitleContent</title>`

(↑ estos son iguales, pero varían en dónde aparecen y con qué etiqueta de html son reemplazados)

**SimParaContent** → `Emphasis` | `texto`  
| `Link` | `Email`  
| `Author` | `Comment`  
| `Personname`

**SimPara** → `<simpara>SimPara SimParaContent</simpara>`  
| `<simpara>SimParaContent</simpara>`

**Emphasis** → `<emphasis>SimParaContent</emphasis>`

**Comment** → `<comment>SimParaContent</comment>`

**Link** → `<link>Link SimParaContent</link>`  
| `<link>SimParaContent</link>`  
| `<link>Xlink</link>`

**ParaContent** → `Emphasis` | `Link`



Email	Author	Comment	ItemizedList
Important	Address	MediaObject	InformalTable   texto

Para → <para>Para ParaContent</para>  
| <para>ParaContent</para>

Important → <important>Titlesec Content</important>

SharedContent → Comment SharedContent | Comment  
| Emphasis SharedContent | Emphasis  
| Link SharedContent | Link  
| texto SharedContent | texto

PersonName → <personname>Fistname Surname</personname>  
FirstName → <firstname>SharedContent</firstname>  
Surname → <surname>SharedContent</surname>  
Street → <street>SharedContent</street>  
City → <city>SharedContent</city>  
Phone → <phone>SharedContent</phone>  
Email → <email>SharedContent</email>  
Date → <date>SharedContent</date>  
Year → <year>SharedContent</year>  
Holder → <holder>SharedContent</holder>  
State → <state>SharedContent</state>  
Postcode → <postcode>text</postcode>

### Imágenes y multimedia:

MediaObjectContent → λ  
| VideoObject MediaObjectContent  
| ImageObject MediaObjectContent

MediaObject →  
<mediaobject>Info VideoObject MediaObjectContent</mediaobject>  
| <mediaobject>Info ImageObject MediaObjectContent</mediaobject>

ImageObject→<imageobject> Info ImageData</imageoobject>

VideoObject→<videoobject>Info VideoData</videoobject>

VideoData → <videodata fileref ="Ruta" />  
| <videodata fileref ="URL" />  
ImageData → <imagedata fileref="Ruta" />  
| <imagedata fileref="URL" />



### Listas:

```
ItemizedList → <itemizedlist>ListItem</itemizedlist>
| <itemizedlist>ListItem ItemizedList</itemizedlist>

ListItem → <listitem> Content ListItem</listitem>
| <listitem>Content</listitem>
```

### Tablas:

```
InformalTableContent → MediaObject InformalTableContent | MediaObject
| Tgroup InformalTableContent | Tgroup

InformalTable → <informaltable>InformalTableContent</informaltable>

Tgroup → <tgroup>Thead Tbody Tfoot</tgroup>
| <tgroup>Thead Tbody</tgroup>
| <tgroup>Tbody Tfoot</tgroup>
| <tgroup>Tbody </tgroup>

Table → <table>TableContent</table>

TableContent → Row TableContent | Row
Thead → <thead>TableContent</thead>
Tbody → <tbody>TableContent</tbody>
Tfoot → <tfoot>TableContent</tfoot>

RowContent → Entry RowContent | Entry
| EntryTbl RowContent | EntryTbl

Row → <row>RowContent</row>

EntryContent → texto EntryContent | texto
| ItemizedList EntryContent | ItemizedList
| Important EntryContent | Important
| Para EntryContent | Para
| SimPara EntryContent | SimPara
| MediaObject EntryContent | MediaObject
| Comment EntryContent | Comment
| Abstract EntryContent | Abstract

Entry → <entry>EntryContent</entry>
```



```
Entrytbl→<entrytbl>Thead Tbody</entrytbl>  
Entrytbl→<entrytbl>Tbody</entrytbl>
```

**Enlaces:**

```
Xlink → <link xlink:href: Url />
```

### 3. Gramática

**Tabla de Terminales (Tokens):**

N°	Terminal	Descripción
1	DOCTYPE	-
2	A_ARTICLE	Etiqueta de apertura de article.
3	C_ARTICLE	Etiqueta de cierre de article.
4	A_INFO	Etiqueta de apertura de info.
5	C_INFO	Etiqueta de cierre de info.
6	A_TITLE	Etiqueta de apertura de title.
7	C_TITLE	Etiqueta de cierre de title.
8	A_ABSTRACT	Etiqueta de apertura de abstract.
9	C_ABSTRACT	Etiqueta de cierre de abstract.
10	A_PARA	Etiqueta de apertura de para.
11	C_PARA	Etiqueta de cierre de para.
12	A_AUTHOR	Etiqueta de apertura de author.
13	C_AUTHOR	Etiqueta de cierre de author.
14	A_PERSONNAME	Etiqueta de apertura de personname.
15	C_PERSONNAME	Etiqueta de cierre de personname.
16	A_FIRSTNAME	Etiqueta de apertura de firstname.
17	C_FIRSTNAME	Etiqueta de cierre de firstname.
18	A_SURNAME	Etiqueta de apertura de surname.





19	C_SURNAME	Etiqueta de cierre de surname.
20	A_DATE	Etiqueta de apertura de date.
21	C_DATE	Etiqueta de cierre de date.
22	A_SECTION	Etiqueta de apertura de section.
23	C_SECTION	Etiqueta de cierre de section.
24	A_SIMSECTION	Etiqueta de apertura de simsect.
25	C_SIMSECTION	Etiqueta de cierre de simsect.
26	A_COPYRIGHT	Etiqueta de apertura de copyright.
27	C_COPYRIGHT	Etiqueta de cierre de copyright.
28	A_ADDRESS	Etiqueta de apertura de address.
29	C_ADDRESS	Etiqueta de cierre de address.
30	A_CITY	Etiqueta de apertura de city.
31	C_CITY	Etiqueta de cierre de city.
32	A_STATE	Etiqueta de apertura de state.
33	C_STATE	Etiqueta de cierre de state.
34	A_POSTCODE	Etiqueta de apertura de postcode.
35	C_POSTCODE	Etiqueta de cierre de postcode.
36	A_STREET	Etiqueta de apertura de street.
37	C_STREET	Etiqueta de cierre de street.
38	A_EMAIL	Etiqueta de apertura de email.
39	C_EMAIL	Etiqueta de cierre de email.
40	A_PHONE	Etiqueta de apertura de phone.
41	C_PHONE	Etiqueta de cierre de phone.
42	A_ITEMIZEDLIST	Etiqueta de apertura de itemizedlist.
43	C_ITEMIZEDLIST	Etiqueta de cierre de itemizedlist.
44	A_LISTITEM	Etiqueta de apertura de listitem.
45	C_LISTITEM	Etiqueta de cierre de listitem.



46	A_EMPHASIS	Etiqueta de apertura de emphasis.
47	C_EMPHASIS	Etiqueta de cierre de emphasis.
48	A HOLDER	Etiqueta de apertura de holder.
49	C HOLDER	Etiqueta de cierre de holder.
50	A_SIMPARA	Etiqueta de apertura de simpara.
51	C_SIMPARA	Etiqueta de cierre de simpara.
52	A_YEAR	Etiqueta de apertura de year.
53	C_YEAR	Etiqueta de cierre de year.
54	A_COMMENT	Etiqueta de apertura de comment.
55	C_COMMENT	Etiqueta de cierre de comment.
56	A_IMPORTANT	Etiqueta de apertura de important.
57	C_IMPORTANT	Etiqueta de cierre de important.
58	A_LINK	Etiqueta de apertura de link.
59	C_LINK	Etiqueta de cierre de link.
60	A_MEDIAOBJECT	Etiqueta de apertura de mediaobject.
61	C_MEDIAOBJECT	Etiqueta de cierre de mediaobject.
62	A_VIDEOOBJECT	Etiqueta de apertura de videoobject.
63	C_VIDEOOBJECT	Etiqueta de cierre de videoobject.
64	A_IMAGEOBJECT	Etiqueta de apertura de imageobject.
65	C_IMAGEOBJECT	Etiqueta de cierre de imageobject.
66	A_INFORMALTAB LE	Etiqueta de apertura de informaltable.
67	C_INFORMALTAB LE	Etiqueta de cierre de informaltable.
68	A_TGROUP	Etiqueta de apertura de tgroup.
69	C_TGROUP	Etiqueta de cierre de tgroup.
70	A_ROW	Etiqueta de apertura de row.
71	C_ROW	Etiqueta de cierre de row.



72	A_TABLE	Etiqueta de apertura de table.
73	C_TABLE	Etiqueta de cierre de table.
74	A_THEAD	Etiqueta de apertura de thead.
75	C_THEAD	Etiqueta de cierre de thead.
76	A_TFOOT	Etiqueta de apertura de tfoot.
77	C_TFOOT	Etiqueta de cierre de tfoot.
78	A_TBODY	Etiqueta de apertura de tbody.
79	C_TBODY	Etiqueta de cierre de tbody.
80	A_ENTRYTBL	Etiqueta de apertura de entrytbl.
81	C_ENTRYTBL	Etiqueta de cierre de entrytbl.
82	A_ENTRY	Etiqueta de apertura de entry.
83	C_ENTRY	Etiqueta de cierre de entry.
84	TEXTO	Letras, números, signos de puntuación y caracteres especiales.
85	URL	Cadena que contiene una dirección URL.
86	RUTA	Cadena que contiene un camino absoluto o relativo.

**Tabla de No terminales:**

N°	No terminal	Descripción
1	sigma	indica por donde empieza el analizador sintáctico.
2	article	Representa un artículo completo que contiene toda la estructura y contenido del documento
3	info	Representa la sección de información del artículo, que puede contener datos como el título, los autores, la fecha, etc.
4	titledoc	Representa el título del documento
5	content	Representa el contenido principal del artículo, que puede incluir secciones, párrafos, listas, tablas, etc.



6	recusection	Representa una sección recursiva, lo que significa que puede contener otras secciones dentro de ella.
7	section	Representa una sección específica del artículo, que puede tener un título y un contenido asociado.
8	simsection	Representa una sección simple, que no tiene subsecciones pero puede contener contenido adicional.
9	itemizedlist	Representa una lista de elementos, donde cada elemento se representa con viñetas.
10	important	Representa un contenido importante o destacado en el artículo
11	para	Representa un párrafo de texto
12	simpara	Representa un párrafo simple, que no tiene una estructura compleja pero puede contener texto adicional
13	address	Representa una dirección o ubicación.
14	mediaobject	Representa un objeto multimedia, como una imagen o un video
15	informaltable	Representa una tabla informal sin una estructura rígida
16	table	Representa una tabla estructurada con filas y columnas
17	comment	Representa un comentario dentro del artículo
18	abstract	Representa un resumen del artículo
19	author	Representa al autor del artículo.
20	date	Representa la fecha asociada al artículo
21	copyright	Representa los derechos de autor del artículo
22	abstractcontent	Representa el contenido del resumen
23	infocontent	Representa el contenido de la sección de información
24	addresscontent	Representa el contenido de una dirección o ubicación
25	authorcontent	Representa el contenido asociado a un autor
26	copyrightyearcontent	Representa el contenido del año de los derechos de autor
27	copyrightholdercontent	Representa el contenido del titular de los derechos de autor



28	titlecontent	Representa el contenido asociado al título de una sección
29	titlesec	Representa una sección específica del título de una sección
30	paracontent	Representa el contenido de un párrafo
31	simparacontent	Representa el contenido de un párrafo simple
32	emphasis	Representa un resalte en el texto
33	sharedcontent	Utilizada para definir contenido compartido que puede ser referenciado y reutilizado en diferentes partes de un documento o en varios documentos
34	personname	Representa el nombre de una persona
35	firstname	Representa el primer nombre de una persona
36	surname	Representa el apellido de una persona
37	street	Representa una calle o dirección postal
38	postcode	Representa el código postal de una dirección
39	city	Representa una ciudad en una dirección
40	phone	Representa un número de teléfono
41	email	Representa una dirección de correo electrónico
42	holder	Representa un titular o poseedor de algo
43	year	Representa un año
44	state	Representa un estado o provincia en una dirección
45	mediaobjectcontent	Representa el contenido asociado a un objeto multimedia
46	imageobject	Representa un objeto de imagen
47	videobject	Representa un objeto de video
48	videodata	Representa datos asociados a un objeto de video
49	imagedata	Representa datos asociados a un objeto de imagen
50	listitem	Representa un elemento de una lista
51	informaltablecontent	Representa el contenido de una tabla informal
52	thead	Representa el encabezado de una tabla
53	tgroup	Representa un grupo de elementos en una tabla



54	tfoot	Representa el pie de una tabla
55	row	Representa una fila en una tabla
56	rowcontent	Representa el contenido de una fila en una tabla
57	entrycontent	Representa el contenido de una celda en una tabla
58	entry	Representa una celda en una tabla
59	entrytbl	Representa una tabla de entrada
60	xlink	Representa un enlace o referencia a un recurso externo

#### 4. Analizador Léxico

Un programa escrito usando Flex (2.6.4) consiste de tres secciones, separadas por '%%'. La primera sección contiene declaraciones de variables, librerías a utilizar y opciones del generador. La segunda contiene la lista de patrones (Las expresiones regulares que usaremos para reconocer los Tokens) y el código en C a ejecutar cuando un patrón es reconocido. Por último, la tercera sección es código C que es copiado al escáner generado.

```
%{
int chars = 0;
int words = 0;
int lines = 0;
}%

%%

[a-zA-Z]+ { words++; chars += strlen(yytext); }
\n      { chars++; lines++; }
.       { chars++; }

%%

main(int argc, char **argv)
{
    yylex();
    printf("%8d%8d%8d\n", lines, words, chars);
}

"Programa contador de palabras"(Levine J. R., 2009, p.3)
```



El programa desarrollado sigue esta misma estructura. En la primera sección definimos variables, funciones e incluimos librerías necesarias. También definimos opciones del analizador como:

- **“noyywrap”**: Se encarga de evitar que el analizador llame a la función `yywrap()`, la cual busca el próximo archivo a escanear.
- **“nodefault”**: Evita que el analizador incluya una regla que copie la input no aceptada a una variable `yyout`, optando por reportar un error si los patrones no reconocen todas las formas de input posibles.
- **“yylineno”**: Ordena que el analizador mantenga el número de la línea actual leída desde su entrada en la variable global `yylineno`.

En la segunda sección, definimos el conjunto de expresiones regulares que utilizaremos para reconocer los tokens que ingresen. Cuando un patrón sea reconocido devolverá el no terminal correspondiente.

## 5. Analizador sintáctico

Bison es un generador de analizadores sintácticos, que convierte una gramática libre de contexto en un analizador sintáctico LALR(1), estas siglas significan que realiza una derivación de izquierda a derecha (LR) con una búsqueda anticipada de un token (LA).

Un programa en Bison (3.8.2) está escrito en la misma estructura de tres secciones que Flex. En la primera sección se realiza la declaraciones de variables, librerías a utilizar y lo más importante, la declaración de los [terminales](#) (Tokens) a utilizar.

En la segunda sección es donde definimos, junto a los no terminales, todas las reglas sintácticas (producciones) que el analizador utilizará.

Bison permite la definición de acciones que acompañan a las reglas sintácticas. Estas acciones contienen código C que se ejecutará cada vez que esa regla es reconocida, de manera muy similar a como opera Flex cuando reconoce un patrón.

Hablamos de esta capacidad porque es a través de ella que implementamos la traducción de las etiquetas en XML a etiquetas HTML y también con ellas escribimos en el archivo de salida.

## 6. Funciones utilizadas

Las funciones utilizadas, junto a la rutina principal del programa, están incluidas en un tercer archivo `“main.c”`. Es allí donde ocurre el llamado al analizador sintáctico.



De manera auxiliar, se definieron una serie de funciones y controles:

- **getExtension:** Tiene el objetivo de devolver la extensión del archivo, si es que existe.
- **printWelcome:** Imprime por pantalla un mensaje de bienvenida al usuario.
- **errorControl:** Se encarga de controlar si el archivo enviado al analizador existe y, si existe, controla que la extensión sea la correcta. En cada caso, si no se cumplen las condiciones, termina la ejecución del programa e imprime por pantalla el error correspondiente.
- **yyerror:** Imprime por pantalla una descripción de un error encontrado durante la ejecución de ambos análisis.
- **main:** Punto de partida para la ejecución del programa.





## 7. Modo de obtención del intérprete

El programa está conformado por tres archivos:

- **scanner.l:** Es el archivo que contiene todo lo que respecta al analizador léxico. Es el archivo que flex utilizará para generar el analizador sintáctico (bajo el nombre *scanner.yy.c*).
- **grammar.y:** Es el archivo que contiene todo lo que respecta al analizador sintáctico. Es el archivo que Bison utilizará para generar tanto el analizador sintáctico con extensión *.c* y el archivo de cabecera *.h* (*grammar.tab.c* y *grammar.tab.h*).
- **main.c:** Es el archivo que contiene a la rutina main y demás funciones que se utilizaran de manera auxiliar a lo largo del programa.
- **Makefile:** Contiene el conjunto de comandos a ejecutar para compilar el ejecutable.

## 8. Modo de ejecución del intérprete

El intérprete tiene los siguientes modos de ejecución a través de la terminal:

- **Interactivo:** Ejecutamos el programa sin indicar un camino relativo o absoluto a un archivo, por lo que de manera predeterminada esperará a recibir una entrada a través del teclado.
- **A partir de un archivo:** Ejecutamos el programa indicando un camino relativo o absoluto a un archivo.

En ambos casos, independientemente de si el análisis es correcto o interrumpido por algún error, se generará en la carpeta desde la cual es ejecutado el programa, el archivo de salida.html con lo generado por el analizador. (Tener en cuenta que los paths en imágenes están hechos teniendo en cuenta que la salida.html estará en la carpeta bin).

Una consideración a la hora de ejecutar códigos de prueba es que el analizador léxico, a la hora de tokenizar un substring, cuando se quiera escribir un texto o número no se deben usar barras (/) y/o contrabarras (\), debido a que hemos tenido muchos inconvenientes a la hora de que el lexer tome un substring como URL o RUTA antes que TEXTO. Una alternativa sería usar la barra en horizontal (|) ó los guiones medios/bajos.



## 9. Ejemplos presentados por el grupo:

### Ejemplo 1

<!DOCTYPE article>

<article>

<info>

<title>Introducción a la Computación Cuántica</title>

<author>

<firstname>Lautaro</firstname>

<surname>Acosta</surname>

</author>

<address>

<street> French 414 </street>

<state>Chaco</state>

<city>Resistencia</city>

<postcode>3500</postcode>

</address>

<date>2 de Julio del 2023</date>

</info>

<title>Descripción General</title>

<para>

La computación cuántica es un revolucionario campo que combina principios de la mecánica cuántica y la ciencia de la computación para resolver complejos problemas de manera más eficiente que las computadoras clásicas.

</para>

<section>

<title>Bits Cuánticos (Cúbits)</title>

<mediaobject>

<imageobject>

<imagedata fileref="../ejemplos/imagenes/qubit.jpg"/>

</imageobject>

</mediaobject>

<para>

El bloque de construcción de la computación cuántica es el cúbit, el cual actúa como análogo cuántico del bit. A diferencia de estos últimos, que solamente pueden representar



uno de dos estados en un tiempo determinado, los cúbits pueden existir en múltiples estados simultáneamente debido a una propiedad denominada superposición

</para>

<para>

Cúbits pueden ser implementados usando varios sistemas físicos, como átomos, fotones o circuitos superconductores. Estos sistemas exhiben propiedades cuánticas que permite la manipulación y medición de cúbits.

</para>

</section>

<section>

<title>Compuertas Cúnticas y Algoritmos</title>

<para>

De manera similar a las computadoras clásicas que emplean el uso de compuertas lógicas para realizar operaciones sobre bits, las computadoras cuánticas emplean compuertas cuánticas para manipular cúbits. Las compuertas cuánticas son transformaciones unitarias que cambian el estado cuántico de un cúbit.

Tienen la característica de ser reversibles, a diferencia de las compuertas clásicas, y son representadas mediante matrices unitarias.

</para>

<mediaobject>

<imageobject>

<imagedata

fileref="https://thumbs.gfycat.com/AggressiveHighCats-size\_restricted.gif"/>

</imageobject>

</mediaobject>

<para>

Algoritmos Cúnticos, como el Algoritmo de Shor para descomponer en factores un número entero y el Algoritmo de Grover para la búsqueda de datos en una secuencia no ordenada, toman ventaja de las propiedades cuánticas para resolver problemas de manera exponencialmente más rápida que algoritmos clásicos.

</para>

</section>

<section>

<title>Desafíos y el futuro</title>

<para>La computación cuántica enfrenta varios desafíos: </para>



<itemizedlist>

<listitem>

<important><para>Decoherencia cuántica:</para></important>

<para>Fenómeno donde un sistema físico, bajo ciertas condiciones, deja de presentar un comportamiento previamente cuántico y pasa a exhibir un comportamiento clásico</para>

</listitem>

<listitem>

<important><para>Corrección de errores cuánticos:</para></important>

<para>Esta práctica trata de proteger la información cuántica de errores debidos a, por ejemplo, la decoherencia. Esto es esencial para que pueda ser implementada en condiciones realistas</para>

</listitem>

<listitem>

<important><para>Escalabilidad:</para></important>

<para>La computación Cuántica presenta múltiples problemas a la hora de escalar los sistemas, desde problemas de refrigeración, como de materiales, así como mantener la calidad de los cúbits</para>

</listitem>

</itemizedlist>

<para>

Investigadores están activamente trabajando para desarrollar códigos para corrección de errores y sistemas cuánticos robustos para superar estos desafíos. Adicionalmente, esfuerzos están siendo realizados para aumentar el número de cúbits y ,más importante, mejorar la calidad de los cúbits, para mejorar la estabilidad y confiabilidad de computadoras cuánticas.

</para>

</section>

<section>

<title>Conclusión</title>

<para>

La computación cuántica representa una tremenda promesa para la resolución de complejos problemas computacionales en muchos dominios. Aunque el campo esté en sus etapas tempranas, la investigación y los avances realizados están trazando el camino para



un futuro donde las computadoras cuánticas puedan resolver tareas más allá de las capacidad de sus contrapartes clásicas.

</para>

</section>

</article>

## Ejemplo 2

<!DOCTYPE article>

<article>

<info>

<title>Agujeros Negros</title>

<author>

<firstname>Ivan</firstname>

<surname>Niveyro</surname>

</author>

<date>02-07-2023</date>

<mediaobject>

<imageobject>

<imagedata

fileref="https://upload.wikimedia.org/wikipedia/commons/6/67/UTN\_logo.jpg"/>

</imageobject>

</mediaobject>

</info>

<title>AGUJEROS NEGROS</title>

<para>

<mediaobject>

<imageobject>

<imagedata

fileref="https://64.media.tumblr.com/6bcad2dfd7d34495536ce245ace698ce/tumblr\_n5k07zs

DuW1sp6e2vo1\_500.gif"/>

</imageobject>

</mediaobject>

</para>

<section>

<title>Introducción</title>



<para>Los agujeros negros son regiones del espacio-tiempo con una gravedad tan intensa que nada, ni siquiera la luz, puede escapar de su atracción. Estos objetos cósmicos fascinantes han capturado la imaginación de científicos y entusiastas de la astronomía durante décadas y continúan siendo un tema de investigación activo en la física teórica y la astrofísica.</para>

</section>

<section>

<title>Características de los agujeros negros</title>

<para>Las siguientes características son:</para>

<itemizedlist>

<listitem>

<important><para>Horizonte de eventos</para></important>

<para>El horizonte de eventos es la frontera imaginaria alrededor de un agujero negro a partir de la cual la velocidad de escape supera la velocidad de la luz. Una vez que un objeto cruza el horizonte de eventos, se considera atrapado dentro del agujero negro y su destino inevitable es colapsar en el punto singular.</para>

</listitem>

<listitem>

<important><para>Singularidad</para></important>

<para>La singularidad es el punto de densidad y curvatura infinitas en el centro de un agujero negro. Se cree que las leyes de la física tal como las conocemos dejan de aplicarse en la singularidad, lo que hace que sea un punto de gran interés y misterio para los físicos.</para>

</listitem>

<listitem>

<important><para>Eventos extremos</para></important>

<para>Los agujeros negros pueden dar lugar a fenómenos extremos, como los chorros de materia expulsados desde sus polos. Estos chorros pueden ser observados en forma de poderosos haces de radiación en diferentes longitudes de onda, desde ondas de radio hasta rayos gamma.</para>

</listitem>

</itemizedlist>

<mediaobject>

<imageobject>



```
<imagedata  
fileref="https://cdn.todamateria.com/imagenes/agujeronegro1-0-cke.jpg"/>  
</imageobject>  
</mediaobject>
```

```
</section>  
<section>  
<title>Historia</title>
```

<para>Los agujeros negros continúan siendo objeto de investigación activa en la física teórica y la astrofísica. Los científicos están interesados en comprender mejor la formación y evolución de los agujeros negros, así como en explorar las propiedades de los chorros de materia expulsados por ellos. Además, se están realizando esfuerzos para detectar ondas gravitacionales provenientes de la fusión de agujeros negros, lo que proporcionaría evidencia directa de su existencia y comportamiento.</para>

```
<informaltable>  
<tgroup>  
<thead>  
  <row>  
    <entry>Año</entry>  
    <entry>Cientificos - Entidad</entry>  
    <entry>Datos Relevantes</entry>  
  </row>  
</thead>  
<tbody>  
  <row>  
    <entry>1783</entry>  
    <entry>John Michell</entry>  
    <entry>Michell calculó que un cuerpo con una densidad 500 veces mayor a la  
    del Sol, pero con su mismo radio, tendría, en su superficie, una velocidad de escape igual a  
    la de la luz y sería invisible</entry>  
  </row>  
  <row>  
    <entry>1915</entry>  
    <entry>Karl Schwarzschild</entry>
```



<entry>Einstein desarrolló la relatividad general y demostró que la luz era influida por la interacción gravitatoria. Unos meses después, Karl Schwarzschild encontró una solución a las ecuaciones de Einstein, donde un cuerpo pesado absorbería la luz.</entry>

</row>

<row>

<entry>1930</entry>

<entry>Subrahmanyan Chandrasekhar</entry>

<entry>Demostró que un cuerpo con una masa crítica (ahora conocida como límite de Chandrasekhar) y que no emitiese radiación, colapsaría por su propia gravedad porque no habría nada que se conociera que pudiera frenarla</entry>

</row>

<row>

<entry>1930-1940</entry>

<entry>Eddington</entry>

<entry>Se opuso a la idea de que la estrella alcanzara un tamaño nulo, lo que implicaría una singularidad desnuda de materia, y que debería haber algo que inevitablemente pusiera freno al colapso, línea adoptada por la mayoría de los científicos</entry>

</row>

<row>

<entry>1939</entry>

<entry>Robert Oppenheimer</entry>

<entry>Predijo que una estrella masiva podría sufrir un colapso gravitatorio y, por tanto, los agujeros negros podrían ser formados en la naturaleza</entry>

</row>

<row>

<entry>1967</entry>

<entry>Stephen Hawking y Roger Penrose</entry>

<entry>Probaron que los agujeros negros son soluciones a las ecuaciones de Einstein y que en determinados casos no se podría impedir que se crease un agujero negro a partir de un colapso. La idea de agujero negro tomó fuerza con los avances científicos y experimentales que llevaron al descubrimiento de los púlsares</entry>

</row>

<row>





```
<entry>1969</entry>
<entry>John Wheeler</entry>
<entry>Acuñó la locución 'agujero negro' durante una reunión de cosmólogos
en Nueva York, para designar lo que anteriormente se llamó 'estrella en colapso gravitatorio
completo'</entry>
</row>
<row>
<entry>2019</entry>
<entry>Telescopio del Horizonte de Sucesos</entry>
<entry>Presentó la primera imagen jamás capturada de un agujero negro
supermasivo ubicado en el centro de la galaxia M87. Se esperaba a su vez el anuncio de
otra imagen real de un agujero negro en Sagitario A*</entry>
</row>
</tbody>
</tgroup>
</informaltable>
</section>
<section>
<title>Conclusiones</title>
<para>Los agujeros negros son objetos fascinantes y misteriosos que desafían
nuestra comprensión actual de la física. A través de la teoría de la relatividad y las
observaciones astronómicas, hemos obtenido información invaluable sobre su naturaleza y
propiedades. Sin embargo, todavía hay mucho por descubrir y comprender sobre estos
enigmáticos objetos cósmicos, lo que los convierte en un campo de investigación
emocionante y prometedor en la ciencia.</para>
<mediaobject>
<imageobject>
<imagedata fileref="https://i.giphy.com/media/5xsVPLFdyNa0w/giphy.webp"/>
</imageobject>
</mediaobject>
</section>
</article>
```



## 10. Conclusión

Consideramos que la experiencia resultó muy enriquecedora porque pudimos reconocer, en nuestra implementación del programa, muchos de los conceptos que damos a lo largo de la cursada. Además, la investigación posterior nos ayudó a expandir nuestro conocimiento sobre las aplicaciones prácticas de los analizadores, tanto el léxico como el sintáctico. En definitiva, poder transferir los conceptos que damos en clases a un programa que funciona nos resultó muy satisfactorio.

La parte más desafiante de la experiencia fue implementar la gramática en Bison debido a que desde un principio la gramática inicial era sumamente ambigua, esto sumado a que los analizadores sintácticos generados por Bison son muy susceptibles a la ambigüedad, resultaron en que tengamos muchos conflictos del tipo desplazamiento/reducción y tengamos que reconstruir en gran parte nuestra gramática inicial.

En conclusión, consideramos que pudimos implementar con éxito las funciones del programa.

## 11. Bibliografía

- Levine, J. R. (2009). Flex & Bison: Text Processing Tools. O'Reilly Media.
- Free Software Foundation. (s. f.). Bison 3.8.2. Recuperado de [https://www.gnu.org/software/bison/manual/html\\_node/index.html](https://www.gnu.org/software/bison/manual/html_node/index.html)