

Trabajo práctico integrador: diseño de lexer y parser

-Sintaxis y semántica de los lenguajes-

Resistencia, Chaco 30 de abril

Alumnos:

- Muñoz Nahuel Alan
- Un juan
- Bangher
- Castro

Ingeniería en sistemas de información

Universidad Tecnológica Nacional sede FRRe

Comisión 2

Primer Cuatrimestre

2023

Índice

Introducción:	3
Gramática	4
Vocabulario	4
Producciones	5

2023

Introducción:

El presente informe pretende brindar una descripción y explicación del funcionamiento de un analizador léxico y sintáctico, que será capaz de recibir tanto texto de forma interactiva como un archivo en formato xml con contenido DocBook.

DocBook es un lenguaje de marcado XML utilizado para la creación de documentos técnicos y de referencia, tales como manuales de usuario, libros de texto y especificaciones técnicas. DocBook proporciona una amplia gama de elementos y atributos que permiten a los autores estructurar y etiquetar el contenido de manera significativa. También incluye herramientas para la generación de varios formatos de salida, como HTML, PDF, y otros formatos de archivo de ayuda. Al ser un estándar abierto, DocBook es compatible con múltiples herramientas y sistemas de publicación.

HTML (HyperText Markup Language) es un lenguaje de marcado utilizado para crear y estructurar contenidos web. HTML utiliza una serie de etiquetas y atributos para definir la estructura, el estilo y la funcionalidad de una página web. Además, se puede utilizar CSS (Cascading Style Sheets) para establecer la presentación visual de los elementos HTML y JavaScript para añadir interactividad a la página web.

Mientras que ambos lenguajes de marcado se utilizan para estructurar y etiquetar contenidos, hay algunas diferencias importantes entre ellos. En primer lugar, HTML se centra en la presentación visual de una página web, mientras que DocBook se centra en la estructuración semántica de documentos técnicos y de referencia. En segundo lugar, mientras que HTML proporciona una amplia gama de elementos y atributos para crear páginas web interactivas, DocBook se centra más en la estructuración y organización de la información, con menos énfasis en la interactividad y la presentación visual. En tercer lugar, HTML es ampliamente utilizado en la creación de sitios web, mientras que DocBook se utiliza principalmente para la creación de documentos técnicos y de referencia.

2023

Gramática

Simbolos

No Terminales: { Σ , article, parrafo, Etiquetas_article, Secciones, Info, Abstract, TAbstract, Address, Autor, Copyright, Title, Link, SimPara, Emphatis, Coment, Para, Important, TImportant, FirstName, LastName, Street, City, State, Phone, Email, Date, TYear, Tholder, Cuerpo, MediaObject, Media, VideoObject, ImageObject, ItemizedList, ListItem, InformalTable, Tgroup, THead, Tfoot, TBody, Row, EntryTbl, Direccion, URL, ABC, Protocolo, TUrl, #texto }

Terminales: {<!DOCTYPE article>, <article>, </article>, <title>, </title>, <info>, </info>, <itemizedList>, </itemizedList>, <important>, </important>, <para>, </para>, <address>, </address>, <mediaobject>, </mediaobject>, <comment>, </comment>, <informaltable>, </informaltable>, <abstract>, </abstract>, <street>, </street>, <section>, </section>, <author>, </author>, <date>, </date>, <copyright>, </copyright>, <city>, </city>, <state>, </state>, <email>, </email>, <firstname>, </firstname>, <surname>, </surname>, <xlink:href=",""/>, </link>, <emphasis>, </emphasis>, <author>, </author>, <important>, </important>, <itemizedlist>, </itemizedlist>, <simpara>, </simpara>, <informaltable>, </informaltable>, <abstract>, </abstract>, <videoobject>, </videoobject>, <videodata fileref=","<imagedata fileref="," , <listitem>, </listitem>, <tbody>, </tbody>, <tfoot>, </tfoot>, <thead>, </thead>, <row>, </row>, <entry>, </entry>, <entrytbl>, </entrytbl>, a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z, , 1, 2, 3, 4, 5, 6, 7, 8, 9, 0, -, _, ., á, é, í, ó, ú, /, (,);, :, //, :, #, http, https, ftp, ftps}

2023

Producciones

$\Sigma \rightarrow \langle !DOCTYPE \text{ article} \rangle \langle \text{article} \rangle \text{ article } \langle / \text{article} \rangle$

$\text{article} \rightarrow \text{Etiquetas_article}$

$\text{article} \rightarrow \text{Parrafo } \text{Etiquetas_article}$

$\text{article} \rightarrow \text{Etiquetas_article } \text{Secciones}$

$\text{article} \rightarrow \text{Parrafo } \text{Etiquetas_article } \text{Secciones}$

$\text{Parrafo} \rightarrow \langle \text{title} \rangle \text{Title} \langle / \text{title} \rangle$

$\text{Parrafo} \rightarrow \langle \text{info} \rangle \text{Info} \langle / \text{info} \rangle \langle \text{title} \rangle \text{Title} \langle / \text{title} \rangle$

$\text{Parrafo} \rightarrow \langle \text{info} \rangle \text{Info} \langle / \text{info} \rangle$

$\text{Etiquetas_article} \rightarrow \text{Etiquetas_article } \text{Etiquetas_article}$

$\text{Etiquetas_article} \rightarrow$

$\langle \text{itemizedList} \rangle \text{ItemizedList} \langle / \text{itemizedList} \rangle | \langle \text{important} \rangle \text{Important} \langle / \text{important} \rangle | \langle \text{para} \rangle \text{Para} \langle / \text{para} \rangle | \langle \text{simpara} \rangle \text{SimPara} \langle / \text{simpara} \rangle | \langle \text{address} \rangle \text{Address} \langle / \text{address} \rangle | \langle \text{mediaobject} \rangle \text{MediaObject} \langle / \text{mediaobject} \rangle | \langle \text{informaltable} \rangle \text{InformalTable} \langle / \text{informaltable} \rangle | \langle \text{comment} \rangle \text{Comment} \langle / \text{comment} \rangle | \langle \text{abstract} \rangle \text{Abstract} \langle / \text{abstract} \rangle$

$\text{Secciones} \rightarrow \text{Secciones } \text{Secciones}$

$\text{Secciones} \rightarrow \langle \text{section} \rangle \text{Etiquetas_article} \langle / \text{section} \rangle$

$\text{Secciones} \rightarrow \langle \text{section} \rangle \text{Parrafo } \text{Etiquetas_article} \langle / \text{section} \rangle$

$\text{Secciones} \rightarrow \langle \text{section} \rangle \text{Etiquetas_article } \text{Secciones} \langle / \text{section} \rangle$

$\text{Secciones} \rightarrow \langle \text{section} \rangle \text{Parrafo } \text{Etiquetas_article}$

$\text{Secciones} \langle / \text{section} \rangle$

$\text{Info} \rightarrow \text{Info } \text{Info}$

$\text{Info} \rightarrow$

$\langle \text{mediaobject} \rangle \text{MediaObject} \langle / \text{mediaobject} \rangle | \langle \text{abstract} \rangle \text{Abstract} \langle /$

2023

abstract>|<address>Address</address>|<author>Author</author>
|<date>Date</date>|<copyright>Copyright</copyright>|<title>Titl
e</title>

Abstract → <title>Title</title> TAbstract

Abstract → TAbstract

TAbstract → TAbstract TAbstract

TAbstract →

<para>Para</para>|<simpara>SimPara</simpara>

Address → Address Address

Address →

#texto|<street>Street</street>|<city>City</city>|<state>State</st
ate>|<phone>Phone</phone>|<email>Email</email>

Author → Author Author

Author →

<firstname>FirstName</firstname>|<surname>Surname</surna
me>

Copyright → Copyright Copyright

Copyright → <year>Year</year> | <year>Year</year>

<holder>Holder</holder>

Title → Title - Title

Title →

#texto|<emphasis>Emphasis</emphasis>|<xlink:href="URL">Link<
/link>|<email>Email</email>

SimPara → SimPara SimPara

Emphasis → Emphasis Emphasis

Comment → Comment Comment

2023

Link → Link Link

SimPara →

```
#texto|<emphasis>Emphasis</emphasis>|<xlink:href="URL">Link</link>|<email>Email</email>|<author>Author</author>|<comment>Comment</comment>
```

Emphasis →

```
#texto|<emphasis>Emphasis</emphasis>|<xlink:href="URL">Link</link>|<email>Email</email>|<author>Author</author>|<comment>Comment</comment>
```

Comment →

```
#texto|<emphasis>Emphasis</emphasis>|<xlink:href="URL">Link</link>|<email>Email</email>|<author>Author</author>|<comment>Comment</comment>
```

Link →

```
#texto|<emphasis>Emphasis</emphasis>|<xlink:href="URL">Link</link>|<email>Email</email>|<author>Author</author>|<comment>Comment</comment>
```

Para → Para Para

Para →

```
#texto|<emphasis>Emphasis</emphasis>|<xlink:href="URL">Link</link>|<email>Email</email>|<author>Author</author>|<comment>Comment</comment>|<itemizedList>ItemizedList</itemizedList>|<important>Important</important>|<address>Address</address>|<mediaobject>MediaObject</mediaobject>|<informalTable>InformalTable</informalTable>
```

Important → <title>Title</title> TImportant

Important → TImportant

TImportant → TImportant TImportant

TImportant →

```
<itemizedList>ItemizedList</itemizedList>|<important>Important
```

2023

t</important>|<para>Para</para>|<simpara>SimPara</simpara>
|<address>Address</address>|<mediaobject>MediaObject</medi
aobject>|<informaltable>InformalTable</informaltable>|Commen
t|<abstract>Abstract</abstract>

FirstName→Cuerpo

Surname→Cuerpo

Street→Cuerpo

City→Cuerpo

State→Cuerpo

Phone→Cuerpo

Email→Cuerpo

Date→Cuerpo

TYear→Cuerpo

THolder→Cuerpo

Cuerpo → Cuerpo Cuerpo

Cuerpo →

#texto|<xlink:href="URL">Link</link>|<emphasis>Emphasis</emp
hasis>|<comment>Comment</comment>

MediaObject → Media

MediaObject → <info>Info</info> Media

Media → Media Media

Media →

<videoobject>VideoObject</videoobject>|<imageobject>ImageO
bject</imageobject>

VideoObject → <videodata fileref="Direccion"/>

VideoObject → <info>Info</info> <videodata
fileref="Direccion"/>

ImageObject → <imagedata fileref="Direccion"/>

2023

ImageObject → <info>Info</info> <imagedata
fileref="Direccion"/>

ItemizedList → ItemizedList ItemizedList

ItemizedList → <listitem>ListItem</listitem>

ListItem → ListItem ListItem

ListItem →

<itemizedList>ItemizedList</itemizedList>|<important>Importan
t</important>|<para>Para</para>|<simpara>SimPara</simpara>
|<address>Address</address>|<mediaobject>MediaObject</medi
aobject>|<informaltable>InformalTable</informaltable>|<comme
nt>Comment</comment>|<abstract>Abstract</abstract>

InformalTable → InformalTable - InformalTable

InformalTable →

<mediaobject>MediaObject</mediaobject>|TGroup

TGroup → <tbody>TBody</tbody> | <tbody>TBody</tbody>

<tfoot>TFoot</tfoot> | <thead>THead</thead>

<tbody>TBody</tbody> | <thead>THead</thead>

<tbody>TBody</tbody> <tfoot>TFoot</tfoot>

THead → <row>Row</row>

TFoot → <row>Row</row>

TBody → <row>Row</row>

Row → Row - Row

Row → <entry>Entry</entry>|<entrytbl>Entrytbl</entrytbl>

Entrytbl → <thead>THead</thead>

<tbody>TBody</tbody> | <tbody>TBody</tbody>

Entry → Entry Entry

Entry → #texto |

<itemizedList>ItemizedList</itemizedList> |

<important>Important</important> | <para>Para</para> |

2023

<simpara>SimPara</simpara> |
<mediaobject>MediaObject</mediaobject> |
<comment>Comment</comment> |
<abstract>Abstract</abstract>

Direccion → #texto/Direccion
Direccion → #texto.ABC

URL → Protocolo://TURL
URL → Protocolo://TURL:TURL
URL → Protocolo://TURL:TURL/TURL
URL → Protocolo://TURL:TURL/TURL#TURL

ABC → ABC ABC
ABC → a|b|c|d|e|f|g|h|i|j|k|l|m|n|o|p|q|r|s|t|u|
v|w|x|y|z

Protocolo → http|https|ftp|ftps
TURL → a|b|c|d|e|f|g|h|i|j|k|l|m|n|o|p|q|r|s|t|u|
v|w|x|y|z| |1|2|3|4|5|6|7|8|9|0|-|_|.

#texto → #texto #texto
#texto → a|b|c|d|e|f|g|h|i|j|k|l|m|n|o|p|q|r|s|t|
u|v|w|x|y|z| |1|2|3|4|5|6|7|8|9|0|-|_|.|á|é|í|ó
|ú|/|(|),

2023