

UNIVERSIDAD TECNOLÓGICA DE LA MIXTECA
INSTITUTO DE COMPUTACIÓN
INGENIERÍA EN COMPUTACIÓN



BUSCADOR

PROGRAMACIÓN LÓGICA

SANTIAGO CRUZ HECTOR
TORRES BAUTISTA FERNANDO

INTRODUCCIÓN

El proyecto se basa en realizar un buscador, a partir de un documento XML basado en un grafo RDF (RDF/XML). Dicho buscador se tiene que adaptar para lo que originalmente fue diseñado PROLOG, para el procesamiento del lenguaje natural. Dicho lenguaje se traduce por medio de una gramática, ya que ésta indica cómo construir una frase de forma correcta. Con estas gramáticas y las cláusulas correctas, PROLOG realiza una búsqueda por los datos del documento XML, haciendo uso de mecanismos básicos del lenguaje como: backtracking, recursividad, unificación, verificación, instanciación y la negación.

El diseño de la aplicación se basa en los hechos de que las búsquedas que se pueden hacer son por Ficha, Título, Autor, Clasificación, Año, o por coincidencia en número de ejemplares en la base de información, se tenían desde el inicio y es práctico el hecho de tenerlas por separado de una búsqueda en lenguaje natural. Así quedan las búsquedas específicas y la que procesa el lenguaje natural, por medio de una oración, contempladas en el proyecto.

RECONOCIMIENTO DE FRASES

El reconocimiento de frases se realiza haciendo uso de una gramática que se desprende de oraciones comunes que se usan al pedir a cualquier persona que te ayude a buscar algo en específico o temas en general, como se dijo, en lenguaje natural. La gramática nos indica cómo se sigue la construcción de la oración hasta tener los elementos que busca el usuario.

Se puede decir que se desarrolló una gramática libre de contexto con un alfabeto que cuenta con símbolos terminales y no terminales, uno inicial y un conjunto de reglas de producción. Con la ayuda de la gramática y las reglas de producción como bases para crear una gramática de cláusulas definidas, que puede traducirse a predicados en PROLOG.

A continuación, se muestran las frases que se ocuparon para el desarrollo de la aplicación:

1. por favor muestrame libros de [tema]
2. por favor por favor quiero que me muestres libros de [tema]
3. por favor buscame libros acerca de [tema]
4. por favor muestrame libros del autor [autor]
5. por favor muestrame libros del tema [libro]
6. por favor busca el autor del libro [libro]
7. por favor buscame libros del autor [autor]
8. necesito que me muestres libros del año [anio]
9. necesito informacion sobre [tema]
10. podrias mostrarme informacion acerca de [tema]
11. podrias encontrar informacion sobre [tema]
12. existe un libro con ficha [ficha]
13. existe algun libro del autor [autor]
14. muestrame libros acerca de [tema]
15. ayudame a encontrar libros del autor [autor]

Lo siguiente son los predicados en Prolog que se basan en la una gramática. Parten de las anteriores oraciones. Su símbolo inicial, se podría decir es 'oración' :

oracion --> inicio, proposicion, fin.
fin --> nombre, preposicion.
fin --> nombre, preposicion, articulo, nombre.
fin --> nombre, contraccion, nombre.
fin --> cuantificador, nombre, preposicion.
fin --> articulo, nombre.
proposicion --> sintagma, proposicion.
proposicion --> sintagma, sintagma.
proposicion --> conjuncion, sintagma.
proposicion --> sintagma, nombre, sintagma.
proposicion --> cuantificador, sintagma.
proposicion --> [].
sintagma --> verbo, preposicion, verbo.
sintagma --> verbo, proposicion.
sintagma --> articulo, nombre, proposicion.
sintagma --> personal, verbo.
sintagma --> verbo.

sintagma --> relativo.
sintagma --> articulo.
sintagma --> verbo, articulo.
sintagma --> verbo, cuantificador.
inicio --> [por],[favor].
inicio --> [necesito].
inicio --> [podrias].
inicio --> [existe].
inicio --> [muestrame].
inicio --> [ayudame].
nombre --> [libro];[libros];[informacion];[autor];[ficha];[año].
preposicion --> [de];[sobre];[sobre],[el];[a];[con];[del].
preposicion --> [acerca],[de];[que],[hablen],preposicion.
personal --> [me].
verbo --> [busca];[muestres];[quiero];[dame];[tengan];[solicito];[mostrarme];
[busco];[podrias];[darme];[necesito];[ayudame];[decirme];[encontrar];
[encontrar];[necesito];[muestrame];[buscame];[saber].
verbo --> [estoy], [buscando].
conjuncion --> [que].
articulo --> [los];[el];[del];[a].
relativo --> [que].
cuantificador --> [algunos];[un];[todos];[algun].
contraccion --> [del].

REPOSITORIO

La categoría que se nos asignó al inicio del proyecto fue Mercadotecnia. Con un número total de 50 títulos.

La estructura de los documentos que componen el repositorio que se tiene es con un documento que describe de forma general el contenido de cada categoría, seguido por un documento que describe la categoría que se tiene, en este caso Mercadotecnia, y por último en carpetas se tienen el conjunto de archivos que describen a cada elemento que componen las categorías, en nuestro caso solo es una carpeta.

Los documentos son archivos RDF/XML, dicho de otra forma, documento XML con base en grafos RDF. En PROLOG se tiene la librería rdf ([library[rdf]]), para el manejo de este tipo de archivos.

El siguiente archivo, Descripcion_Mercadotecnia.rdf, que describe el contenido de las categorías:

```
<?xml version="1.0" ?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:lb="Biblioteca/">
  <rdf:Description rdf:about="Biblioteca/Mercadotecnia/1892.rdf">
    <lb:ficha>1892</lb:ficha>
    <lb:titulo>COMO PREPARAR UN PLAN DE MARKETING</lb:titulo>
    <lb:autor>STAPLETON, JOHN</lb:autor>
    <lb:clasificacion>HF5415.13 S85</lb:clasificacion>
    <lb:anio>[S.F.]</lb:anio>
    <lb:ejemplares>1</lb:ejemplares>
  </rdf:Description>
  ...
</rdf:RDF>
```

Como se puede observar el archivo muestra una descripción para un título, con referencia a un archivo 1892.rdf. Lo que sigue después de esta descripción es un total de 49 con títulos diferentes.

Por último, se tienen a los archivos que describen a cada título del documento anterior. Se identifican por medio de la ficha, ya que es única. La estructura o digamos la descripción del archivo es diferente a las anteriores, ya que describe en primera instancia una secuencia de temas, seguido por una descripción para cada tema que contenga cada título por archivo.

Como ejemplo de la descripción de los archivos, se muestra el referente a la ficha 775:

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:ct="Biblioteca/Mercadotecnia/">
  <rdf:Description rdf:about="DIRECCION DE MARKETING">
    <ct:tema>
      <rdf:Seq>
        <rdf:li>La dirección de marketing</rdf:li>
        <rdf:li>Identificación de las oportunidades de mercado</rdf:li>
        <rdf:li>Relación con clientes o consumidores</rdf:li>
        <rdf:li>Creación de marcas fuertes</rdf:li>
        <rdf:li>La definición de las ofertas de mercado</rdf:li>
        <rdf:li>La entrega de valor</rdf:li>
        <rdf:li>La comunicación del valor</rdf:li>
        <rdf:li>La generación de crecimiento rentable a largo plazo</rdf:li>
      </rdf:Seq>
    </ct:tema>
  </rdf:Description>
```

...

Como se observa en la primera parte, se muestra una serie de temas. A continuación se muestra la descripción del primer tema:

```
<rdf:Description rdf:about="La dirección de marketing">
  <ct:subtema>
    <rdf:Seq>
      <rdf:li>El marketing en el siglo XXI</rdf:li>
      <rdf:li>Desarrollo de estrategias y planes de marketing</rdf:li>
    </rdf:Seq>
  </ct:subtema>
</rdf:Description>
```

Cada tema se desprende en un subtema que puede contener simples líneas o párrafos que describan el contenido del tema.

TIPOS DE BÚSQUEDA

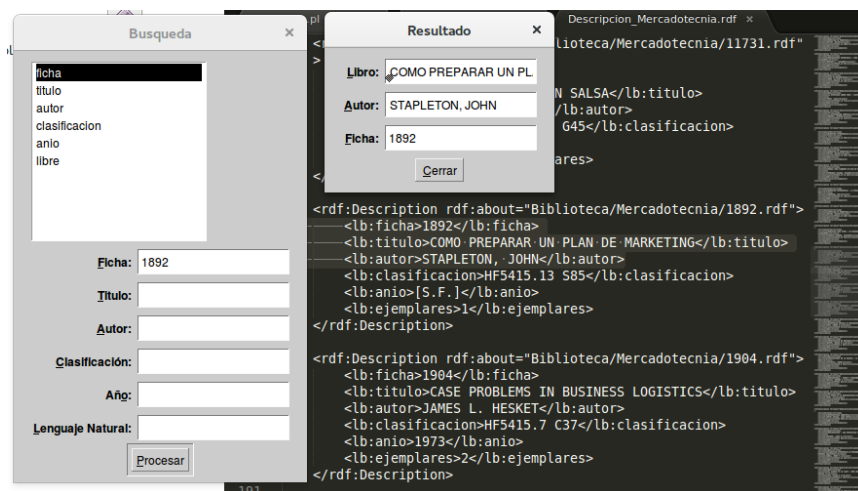
Se tienen seis tipos de búsqueda. Los primeros tipos de búsqueda, que son por Ficha, Título, Autor, Clasificación o Año, que son partes importantes en cuestión de búsquedas específicas. Después se tiene la búsqueda por procesamiento de lenguaje natural.

Los primeros tipos de búsqueda que son por algún parámetro de un libro, se realiza directamente sobre el documento rdf, por medio de la cláusula:

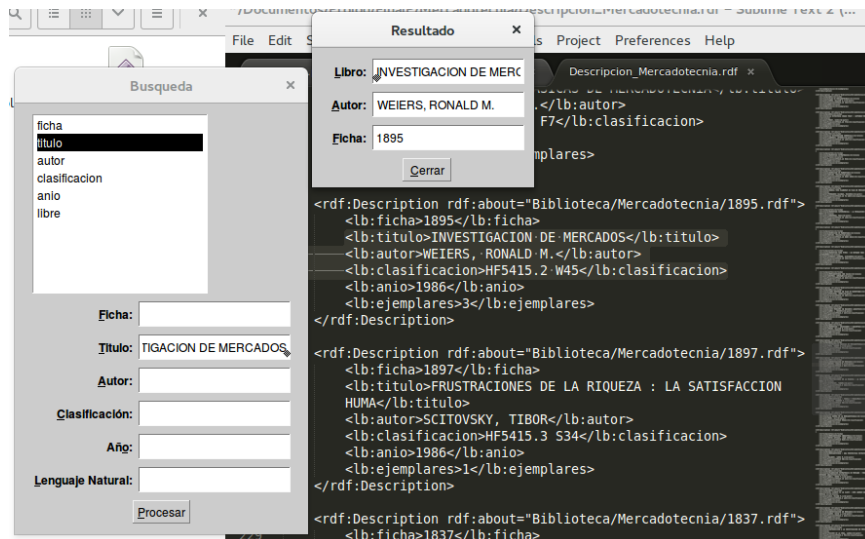
```
buscar(Ficha, Titulo, Autor, Clasif, Anio, Ejem):-  
    rdf_load('Descripcion_Mercadotecnia.rdf',[format(xml)]),  
    rdf(Archivo, 'Biblioteca/ficha',literal(Ficha)),  
    rdf(Archivo, 'Biblioteca/titulo',literal( Titulo)),  
    rdf(Archivo, 'Biblioteca/autor', literal(Autor)),  
    rdf(Archivo, 'Biblioteca/clasificacion',literal( Clasif)),  
    rdf(Archivo, 'Biblioteca/anio', literal(Anio)),  
    rdf(Archivo, 'Biblioteca/ejemplares',literal( Ejem)),  
    rdf_unload('Descripcion_Mercadotecnia.rdf').
```

Ejemplos de búsquedas simples:

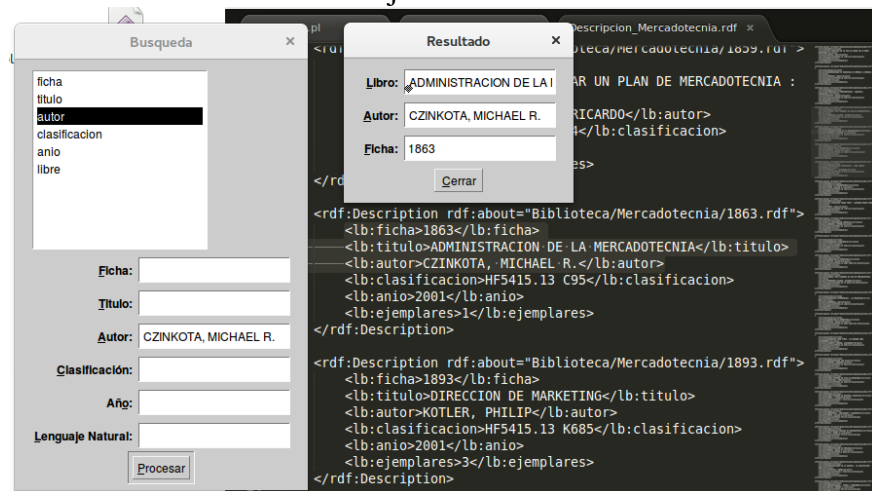
Ej. 1



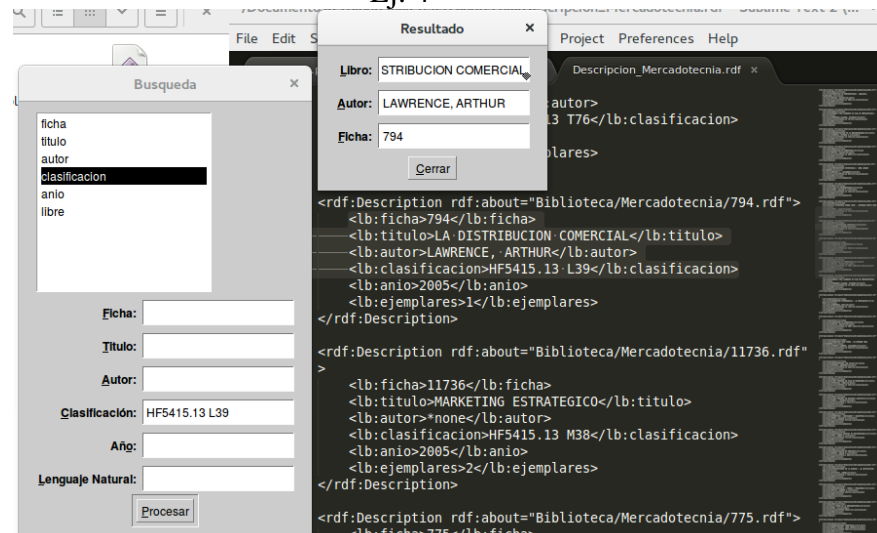
Ej. 2



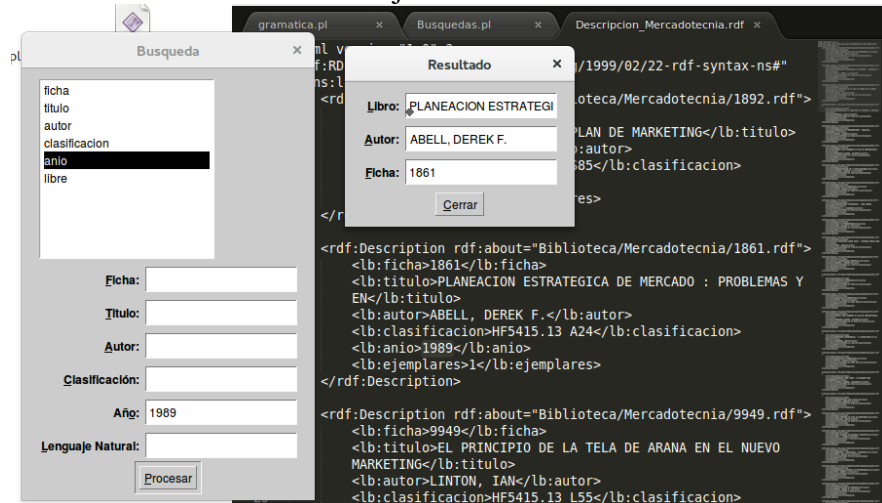
Ej. 3



Ej. 4



Ej. 5



En el código se reconoce el tipo de búsqueda, y después se manda un parámetro, ya sea la ficha el titulo o cualesquiera, para que obtenga los datos directamente sobre el rdf. Si la búsqueda se realiza de forma satisfactoria nos retornara una ventana con los datos obtenidos por a búsqueda.

El tipo de búsqueda con procesamiento de lenguaje natural se da por medio de la cláusula `busquedaAvanzada`. Para ello la gramática hace un trabajo excepcional, otorgando del procesamiento del lenguaje natural una lista con los parámetros de búsqueda.

```
busquedaAvanzada([X|R],_T,_A,_Ficha):-
    X==tema,
    buscaTema(R,Ficha,Res),
    Res==1,
    buscar(Ficha,T,A, B,N,M).
```

Se puede ver que se intenta verificar en un principio si la el primer parámetro de la lista que se envía es un tema, de ser así, hace una búsqueda por fuerza bruta a través de todos los archivos. Esta es la búsqueda más exhaustiva que se hace en el programa.

Esto se hace por medio de un `setoff()` aplicado a cada archivo para reconocer donde se tiene un tema en específico. Busca todo de la misma forma en que se encuentra escrito en los documentos RDF.

Cuando la búsqueda no es por tema, se unifica el tipo con la cláusula `busquedaAvanzada`:

```
busquedaAvanzada([X|R],T,A,Ficha):-
    X==autor,
    buscar(Ficha,T,A,B,R,N).
```

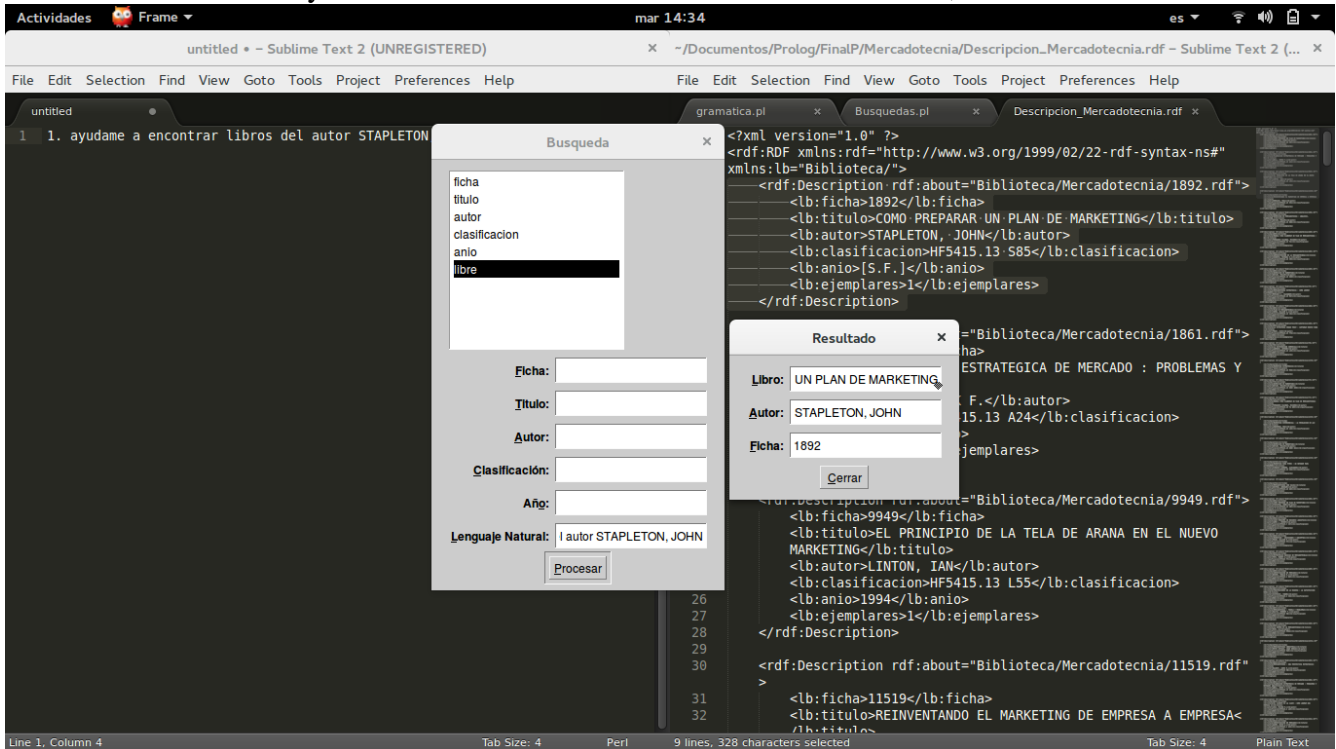
Una vez que obtiene que tipo de busqueda realizar, en este caso autor, lo manda directamente a buscar en los archivos rdf con la cláusula `buscar(Ficha,Titulo,Autor,Clasificacion,Año,Ejemplares)`.

PRUEBAS

Las pruebas que se realizaron corresponden a las siguientes oraciones:

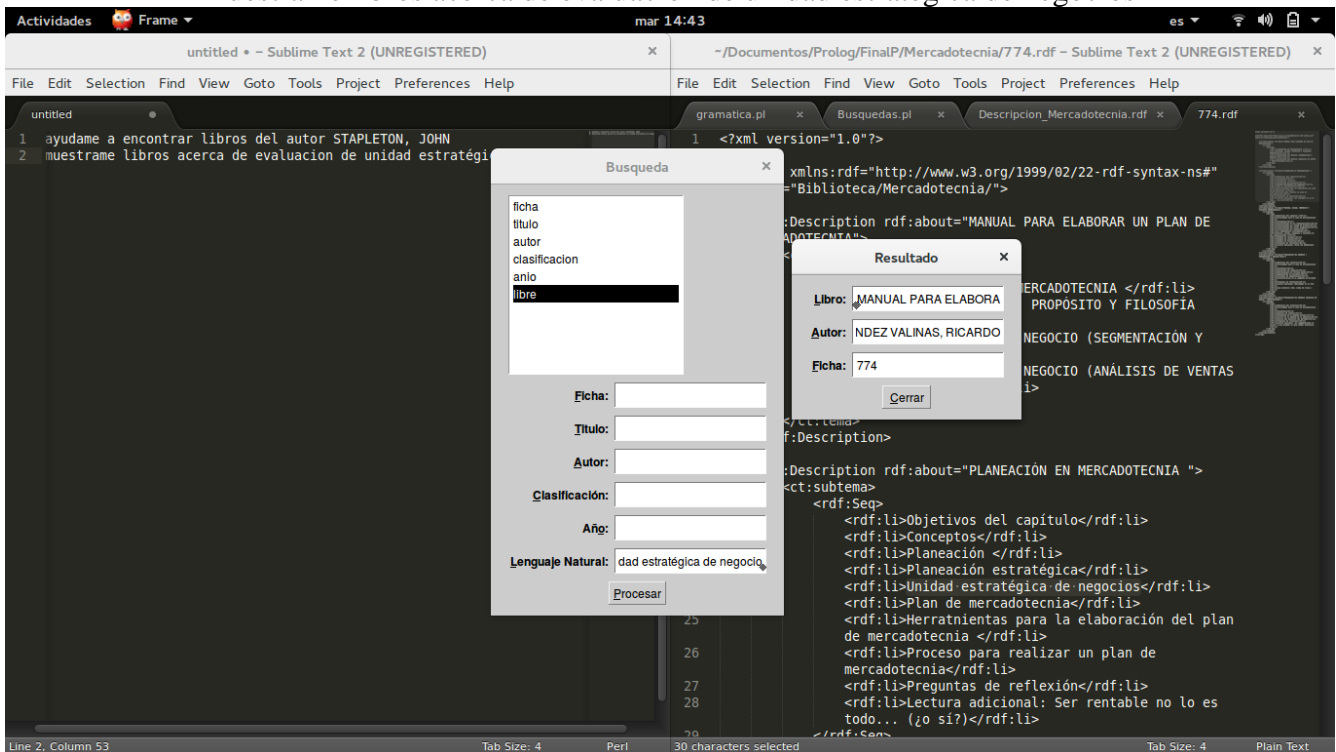
Ej.1

“ayudame a encontrar libros del autor STAPLETON, JOHN”



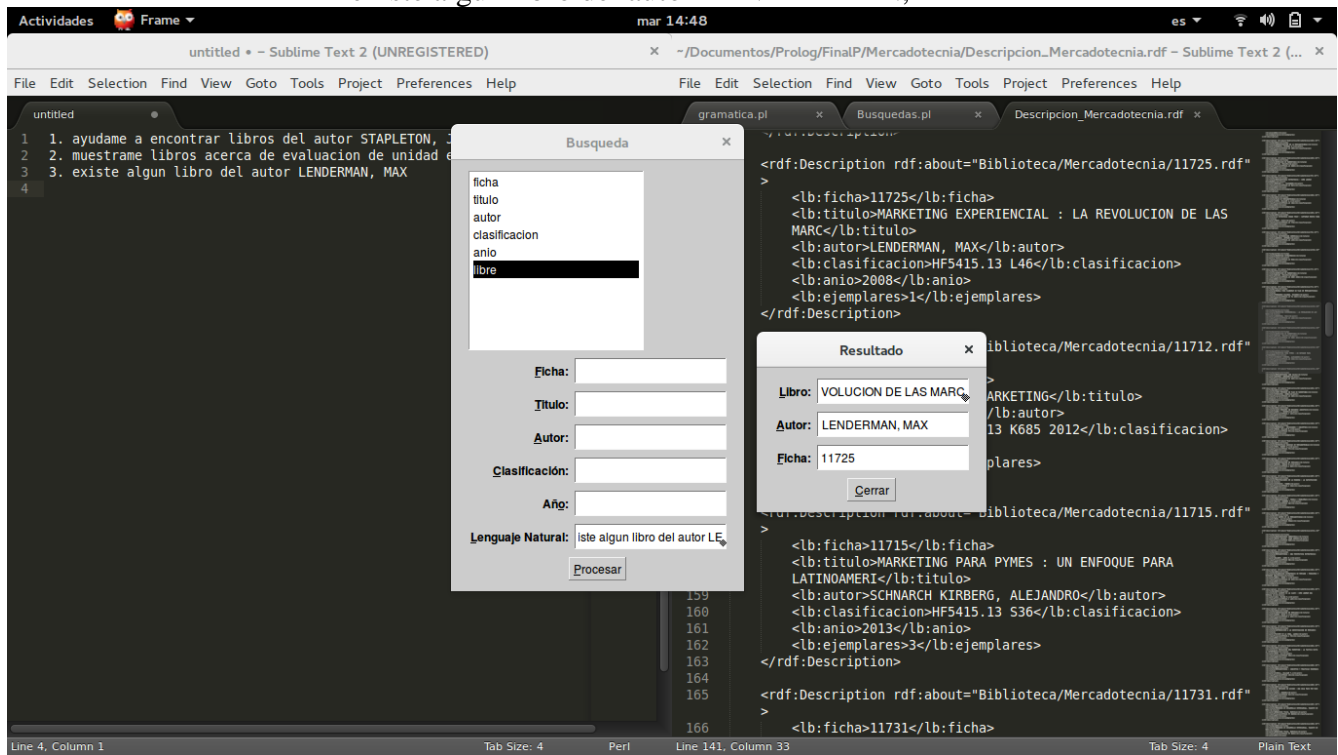
Ej. 2

“muestrame libros acerca de evaluacion de unidad estratégica de negocios”



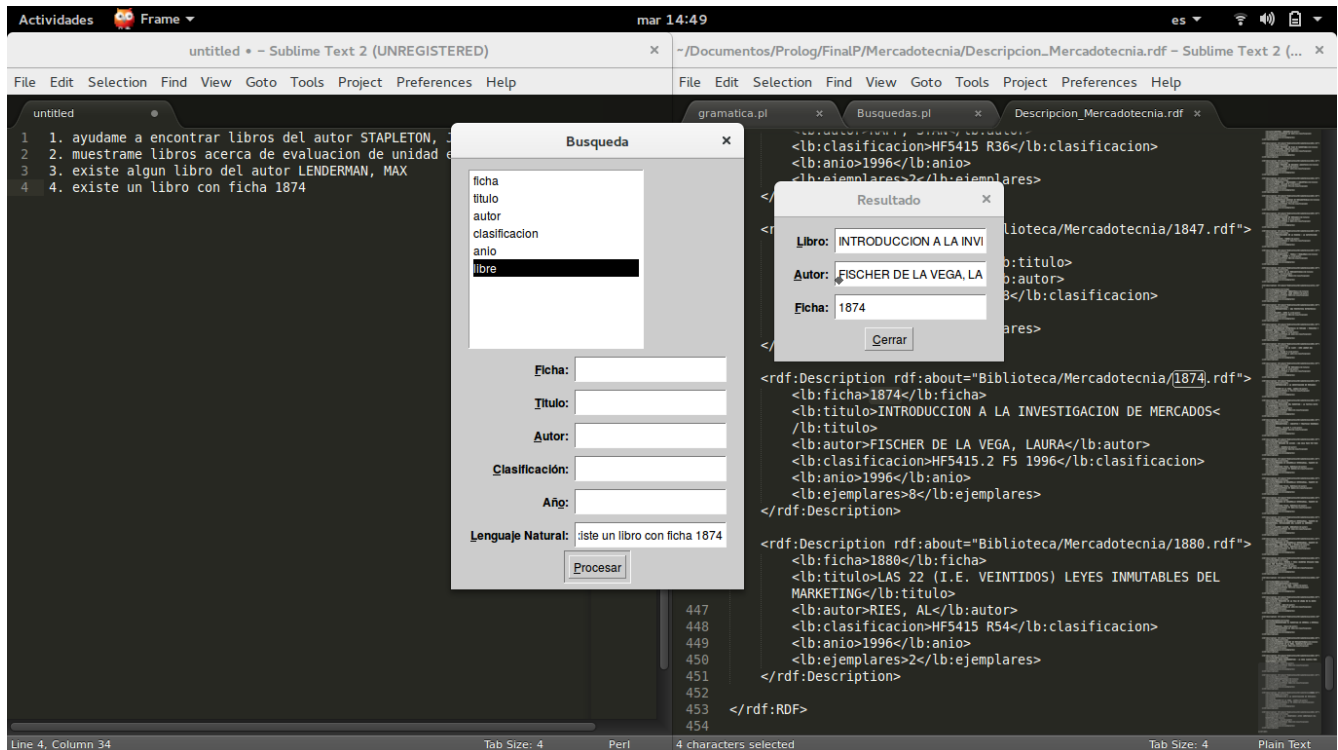
Ej. 3

“existe algun libro del autor LENDERMAN, MAX”



Ej. 4

“existe un libro con ficha 1874”



Ej. 5

“podrias encontrar informacion sobre el concepto de mercadotecnia y la razón de las utilidades”

The screenshot shows the Sublime Text 2 interface with two open files: 'untitled' and '1833.rdf'. The '1833.rdf' file contains XML data. A search window titled 'Busqueda' is open, showing a list of search criteria: 'ficha', 'titulo', 'autor', 'clasificacion', 'anio', and 'libro'. The 'Libro' field is selected. A results window titled 'Resultado' is also open, displaying the following information:

| Libro: | Autor: | Fecha: |
|----------------------|-------------------|--------|
| MERCADOTECNIA: UNA I | HULBERT, JAMES M. | 1833 |

The background XML data in '1833.rdf' includes the following structure:

```
<rdf:Description rdf:about="El concepto de mercadotecnia">
  <ct:subtema>
    <rdf:Seq>
      <rdf:li>Significados de la mercadotecnia</rdf:li>
      <rdf:li>Evolución del concepto de mercadotecnia</rdf:li>
      <rdf:li>El concepto de mercadotecnia y la razón de las utilidades</rdf:li>
      <rdf:li>La mercadotecnia y el entorno comercial contemporáneo</rdf:li>
      <rdf:li>Resumen y conclusiones</rdf:li>
    </rdf:Seq>
  </ct:subtema>
</rdf:Description>

<rdf:Description rdf:about="La labor de mercadotecnia">
  <ct:subtema>
    <rdf:Seq>
      <rdf:li>Responsabilidades de la administración de la mercadotecnia</rdf:li>
      <rdf:li>Principios de la estrategia de mercadotecnia</rdf:li>
      <rdf:li>Resumen y conclusiones</rdf:li>
    </rdf:Seq>
  </ct:subtema>
</rdf:Description>
```

Ej. 6

“podrias mostrarme informacion acerca de tipos de información que requiere la empresa, 22.”

The screenshot shows the Sublime Text 2 interface with two open files: 'untitled' and '12721.rdf'. The '12721.rdf' file contains XML data. A search window titled 'Busqueda' is open, showing a list of search criteria: 'ficha', 'titulo', 'autor', 'clasificacion', 'anio', and 'libro'. The 'Libro' field is selected. A results window titled 'Resultado' is also open, displaying the following information:

| Libro: | Autor: | Fecha: |
|---------------------|------------------------|--------|
| MERCADOTECNIA INDUS | REYES ARTIGAS, JUAN SI | 12721 |

The background XML data in '12721.rdf' includes the following structure:

```
<rdf:Description rdf:about="Tipos de información que requiere la empresa, 22.">
  <ct:subtema>
    <rdf:Seq>
      <rdf:li>Tipos de información que requiere la empresa, 22.</rdf:li>
      <rdf:li>Recopilación de información, 30.</rdf:li>
      <rdf:li>El análisis, la síntesis y la medición de la información, 38.</rdf:li>
      <rdf:li>El procesamiento de la información, 41.</rdf:li>
    </rdf:Seq>
  </ct:subtema>
</rdf:Description>

<rdf:Description rdf:about="La organización de las áreas mercadotécnicas y la mercadotecnia industrial">
  <ct:subtema>
    <rdf:Seq>
      <rdf:li>Organización de mercadotecnia en el área comercial, 46.</rdf:li>
      <rdf:li>Organización de mercadotecnia en el área de abastecimiento, 55.</rdf:li>
      <rdf:li>Los recursos humanos necesarios en las ventas industriales, 60.</rdf:li>
      <rdf:li>Descripción de puestos, 60.</rdf:li>
      <rdf:li>Sistemas de compensación, 64.</rdf:li>
      <rdf:li>Capacitación de los recursos humanos para la venta de productos industriales, 72.</rdf:li>
      <rdf:li>Selección y contratación de los recursos humanos, 74.</rdf:li>
    </rdf:Seq>
  </ct:subtema>
</rdf:Description>
```

Ej. 7

“necesito informacion sobre estrategia corporativa”

The screenshot shows a file manager window with a search bar and a list of files. The file 'Mercadotecnia' is selected. A search dialog box is open, showing the search criteria: 'Mercadotecnia'. The results dialog box shows the following information:

- Libro: ADMINISTRACION DE LA MERCADOTECNIA
- Autor: CZINKOTA, MICHAEL R.
- Fecha: 1863

The background shows a text editor window with RDF data. The selected text is:

```
<rdf:Description rdf:about="Biblioteca/Mercadotecnia/1863.rdf">
  <lb:ficha>1863</lb:ficha>
  <lb:titulo>ADMINISTRACION DE LA MERCADOTECNIA</lb:titulo>
  <lb:autor>CZINKOTA, MICHAEL R.</lb:autor>
  <lb:clasificacion>HF5415.13 C95</lb:clasificacion>
  <lb:anio>2001</lb:anio>
  <lb:ejemplares>1</lb:ejemplares>
</rdf:Description>
```

Ej. 8

“por favor buscame libros acerca de estrategia corporativa”

The screenshot shows a text editor window with a search bar and a list of files. The file 'Marketing Estrategico' is selected. A search dialog box is open, showing the search criteria: 'Marketing Estrategico'. The results dialog box shows the following information:

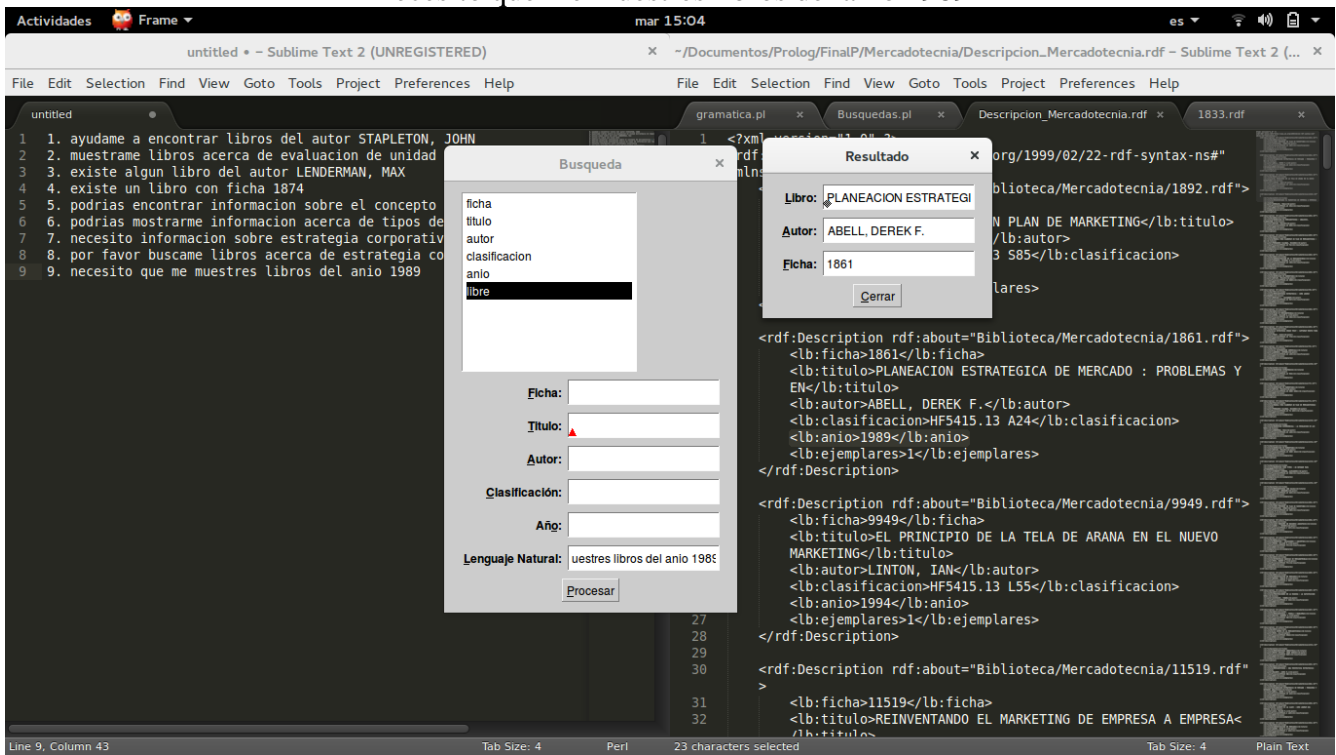
- Libro: MARKETING ESTRATEGICO
- Autor: none
- Fecha: 11736

The background shows a text editor window with RDF data. The selected text is:

```
<rdf:li>Marketing Estrategico</rdf:li>
<rdf:li>Estrategia corporativa</rdf:li>
<rdf:li>Cómo cambiar los criterios sobre las decisiones de inversión</rdf:li>
<rdf:li>Enfoques para la fijación del rumbo estratégico con base en la cartera</rdf:li>
<rdf:li>Resumen y conclusiones</rdf:li>
</rdf:Seq>
</ct:subtema>
</rdf:Description>
```

Ej. 9

“necesito que me muestres libros del año 1989”



The screenshot shows the Sublime Text 2 interface with a search dialog open. The search criteria are: **Ficha:** (empty), **Título:** (empty), **Autor:** (empty), **Clasificación:** (empty), **Año:** (empty), and **Lenguaje Natural:** "nuestros libros del año 1989". The results dialog shows a list of books with the following details:

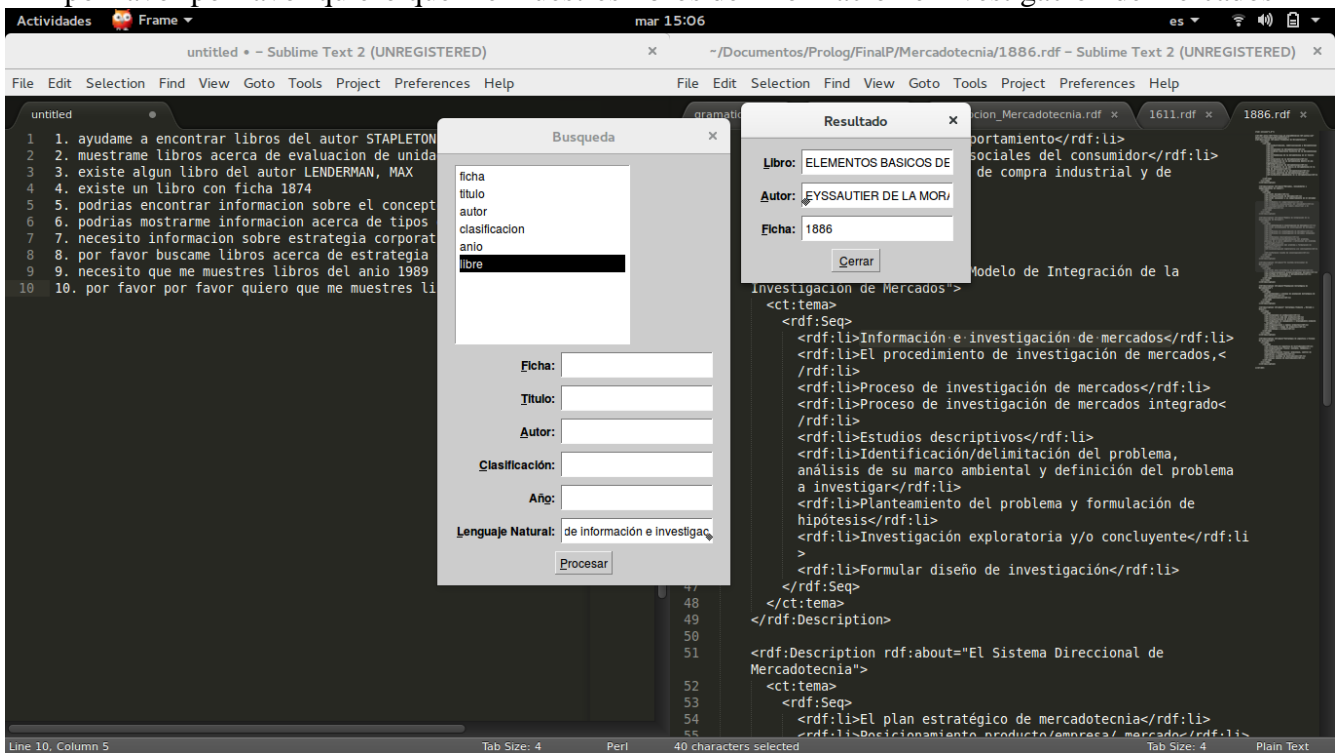
| Libro | Autor | Año |
|------------------------|-----------------|------|
| PLANEACION ESTRATEGICA | ABELL, DEREK F. | 1861 |

The background text in the editor shows an RDF description of a book:

```
<rdf:Description rdf:about="Biblioteca/Mercadotecnia/1861.rdf">
  <lb:ficha>1861</lb:ficha>
  <lb:título>PLANEACION ESTRATEGICA DE MERCADO : PROBLEMAS Y
  EN</lb:título>
  <lb:autor>ABELL, DEREK F.</lb:autor>
  <lb:clasificación>HF5415.13 A24</lb:clasificación>
  <lb:año>1989</lb:año>
  <lb:ejemplares>1</lb:ejemplares>
</rdf:Description>
```

Ej. 10

“por favor por favor quiero que me muestres libros de información e investigación de mercados”



The screenshot shows the Sublime Text 2 interface with a search dialog open. The search criteria are: **Ficha:** (empty), **Título:** (empty), **Autor:** (empty), **Clasificación:** (empty), **Año:** (empty), and **Lenguaje Natural:** "de información e investigación de mercados". The results dialog shows a list of books with the following details:

| Libro | Autor | Año |
|----------------------|----------------------|------|
| ELEMENTOS BASICOS DE | EYSSAUTIER DE LA MOR | 1886 |

The background text in the editor shows an RDF description of a book:

```
<rdf:Description rdf:about="El Sistema Direccional de
Mercadotecnia">
  <ct:tema>
    <rdf:Seq>
      <rdf:li>Información e investigación de mercados</rdf:li>
      <rdf:li>El procedimiento de investigación de mercados,<
      /rdf:li>
      <rdf:li>Proceso de investigación de mercados</rdf:li>
      <rdf:li>Proceso de investigación de mercados integrado<
      /rdf:li>
      <rdf:li>Estudios descriptivos</rdf:li>
      <rdf:li>Identificación/delimitación del problema,
      análisis de su marco ambiental y definición del problema
      a investigar</rdf:li>
      <rdf:li>Planteamiento del problema y formulación de
      hipótesis</rdf:li>
      <rdf:li>Investigación exploratoria y/o concluyente</rdf:li>
    </rdf:Seq>
  </ct:tema>
</rdf:Description>
```

CONCLUSIONES

PROLOG es un lenguaje muy útil, y con el procesamiento del lenguaje natural nos brinda la capacidad tener una interacción máquina-humano muy similar a una comunicación humano a humano. Durante este curso aprendimos un nuevo paradigma de programación el cual es muy distintos a lo que ya sabíamos pero que nos brinda un gran potencial al momento de realizar búsquedas con información, resolver problemas, etc.

En este proyecto final tratamos de utilizar los conocimientos adquiridos durante el curso al momento de implementar los distintos tipos de búsquedas, también aprendimos la forma de implementar interfaces en Prolog, la cual nos resultó fácil hasta cierto punto, pero la falta de tiempo nos afectó en cierta manera al momento de realizar las investigaciones sobre cómo usar las distintas librerías.

En cuanto a la gramática fue una de las partes más sencillas, esto se debe a que solo había que seguir las reglas establecidas a la estructuración de las oraciones, en contra parte una de las cosas que más se nos complicó durante este proyecto fue la búsqueda por temas pero al final la implementamos buscando por “fuerza bruta”, ya que vamos abriendo cada .rdf asociado a cada libro y comparamos el tema que ingreso el usuario con los temas de cada libro hasta que encuentre una coincidencia.