Universidad Autónoma Metropolitana Unidad Azcapotzalco
División de Ciencias Básicas e Ingeniería
Licenciatura en Ingeniería en Computación

Modalidad: Proyecto Tecnológico

Sistema web para identificar eventos y actores en textos periodísticos

Luis David Hernández Rojas
Matrícula 210203238
al210203238@alumnos.azc.uam.mx

Asesora:
Dra. Maricela Claudia Bravo Contreras
Profesor asociado "D"
Departamento de Sistemas
mcbc@correo.azc.uam.mx

Co asesor:
Dr. José Alejandro Reyes Ortiz
Profesor titular "A"
Departamento de Sistemas
jaro@correo.azc.uam.mx

Trimestre 2015 Invierno

Fecha de entrega:
8 de abril de 2015.

DECLARATORIA

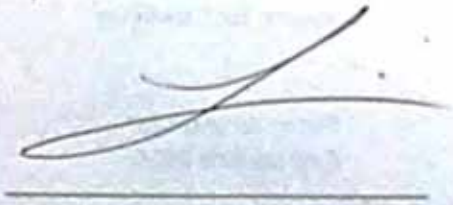
Yo, Dra. Maricela Claudia Bravo Contreras, *declaro que aprobé el contenido del presente Reporte de Proyecto de Integración y doy mi autorización para su* publicación en la Biblioteca Digital, así como en el Repositorio Institucional de UAM Azcapotzalco.

Dra. Maricela Claudia Bravo Contreras

Yo, Dr. José Alejandro Reyes Ortiz, *declaro que aprobé el contenido del presente Reporte de Proyecto de Integración y doy mi autorización para su* publicación en la Biblioteca Digital, así como en el Repositorio Institucional de UAM Azcapotzalco.

Dr. José Alejandro Reyes Ortiz

Yo, Luis David Hernández Rojas, *doy mi autorización* a la Coordinación de Servicios de Información de la Universidad Autónoma Metropolitana, Unidad Azcapotzalco, para publicar el presente documento en la Biblioteca Digital, así como en el Repositorio Institucional de UAM Azcapotzalco.



Luis David Hernández Rojas

RESUMEN

Hoy en día existe una gran cantidad de textos periodísticos dispersos en la red cuyo volumen es tan grande que sobrepasa la capacidad de una persona para obtener información útil que se convierta en conocimiento. A lo largo de la historia se han buscado posibles soluciones a dicho problema, siendo una de ellas los sistemas de extracción de información, que permiten estructurar datos relevantes a un dominio específico en los documentos.

De manera concisa, la extracción de información convierte el problema de analizar una amplia colección de textos a simplemente consultar una base de datos. Sin duda alguna, esto último es más rápido de realizar además de hacer más factible encontrar una relación entre los datos.

Por lo tanto, en este proyecto se propone diseñar e implementar un sistema que realice la extracción de información en textos periodísticos de medios nacionales mediante minería de textos, facilitando el acceso de grandes colecciones de datos textuales y mejorando así la productividad en las tareas de análisis y síntesis de información, que en este caso será basado en eventos y actores.

CONTENIDO

INTRODUCCIÓN

Hoy en día existe una gran cantidad de textos periodísticos dispersos en la red. De hecho, el volumen es tan grande que sobrepasa la capacidad de una persona para obtener información útil que se convierta en conocimiento. Una solución a este problema son los sistemas de extracción de información, que permiten estructurar datos relevantes a un dominio específico en los documentos.

Dicho de manera concisa, la extracción de información convierte el problema de analizar una amplia colección de textos a simplemente consultar una base de datos. Sin duda alguna, esto último es más rápido de realizar además de hacer más factible encontrar una relación entre los datos.

Para el sistema de extracción de información, primero un sistema de recuperación de la información obtiene documentos con información significativa y a continuación el sistema de extracción de información obtiene y organiza la información que sea de interés.

Por lo tanto, en este proyecto se propone diseñar e implementar un sistema que realice la extracción de información en textos periodísticos de medios nacionales mediante minería de textos, facilitando el acceso de grandes colecciones de datos textuales y mejorando así la productividad en las tareas de análisis y síntesis de información, que en este caso será basado en eventos y actores.

ANTECEDENTES

PROYECTOS DE INTEGRACIÓN

Sistema Configurable de Minería Web. [1]

En este proyecto de integración se aborda el problema dela basta cantidad de información que circula en la red y que muchas veces es imposible captar lo verdaderamente importante. En el proyecto se emplea minería de datos al igual que en el proyecto que se propone, sin embargo se busca aplicar a la extracción de eventos y actores de notas periodísticas de cinco de los medios más importantes de circulación nacional.

Esto permitirá que se puedan realizar diversos estudios a través de relaciones entre los artículos sin necesidad de invertir demasiado tiempo en estar checando los textos, debido a que solo se extraerán temas que sean de interés.

Sistema de recuperación de información de textos de investigación de la web. [2]

En este proyecto se desarrolló un sistema usando arañas web focalizadas que pueda encontrar documentos de tipo científico o de investigación en la web, aplicándoles minado de datos.

A diferencia de dicho proyecto de integración, se busca aplicar minado de datos a textos periodísticos de cinco medios de circulación nacional centrándonos en eventos y actores.

Sistema de almacenamiento semántico y recuperación de textos de investigación mediante ontologías. [3]

Este proyecto también aborda el problema del gran volumen de información que existe en la red y aplica la extracción y recuperación de información.

La diferencia esencial es que este proyecto aplicara minería de textos en artículos periodísticos.

TESIS

Extracción de información con algoritmos de clasificación. [4]

En esta tesis podemos localizar las bases teóricas para la extracción de información. Se muestran algoritmos que pueden emplearse en el análisis de textos.

El proyecto que se busca implementar se apoyará en esta tesis para el aprendizaje de reglas y aprendizaje estadístico, que servirá para poder llevar a la práctica la extracción de información. Cabe mencionar que no solo se hará la extracción, sino que también se va a clasificar.

ARTÍCULOS DE INVESTIGACIÓN

Minería de datos: Conceptos y tendencias. [5]

En este artículo, se aclaran diversas cuestiones mediante una introducción a la minería de datos como son: definición, ejemplificación de problemas que se pueden resolver con minería de datos, las tareas de la minería de datos, técnicas usadas y finalmente retos y tendencias en minería de datos.

SOFTWARE

TextStat. [6]

Es una aplicación libre para el análisis estadístico de textos. Esta aplicación puede leer un archivo de texto (ASCII, HTML, MS Word, OpenOffice.org) y generar una lista de palabras con la frecuencia de aparición y de concordancias a partir del texto. El proyecto propuesto obtendrá a partir de archivos de texto la comparación no sólo de palabras repetidas con la misma frecuencia, se obtendrá una medida de similitud semántica.

JUSTIFICACIÓN

La gran cantidad de información que existe en la red es fácil de acceder, sin embargo el gran volumen hace imposible que esa información sea convertida en conocimiento. El sistema de extracción de datos para eventos y actores en textos periodísticos utilizando técnicas de minera de textos, permitirá que las personas interesadas en información periodística puedan hacer más rápido y menos costoso el proceso de localizar temas e información de interés.

Es importante hacer énfasis en el hecho de que la minería de datos ha sido un campo muy importante dentro del estudio computacional en los últimos años. Su impacto económico y en la reducción de esfuerzo humano ha sido factor para el desarrollo tecnológico en el área de Ingeniería en Computación.

Este sistema será integrado en un proyecto de investigación del Departamento de sistemas en el Grupo de Investigación en Sistemas de Información Inteligentes que se llevará a cabo en la Universidad Autónoma Metropolitana Azcapotzalco

OBJETIVOS

OBJETIVO GENERAL

Diseñar e implementar un sistema Web para la anotación semántica de actores y eventos a partir de un corpus de textos periodísticos mexicanos aplicando técnicas de minería de textos haciendo uso de características sintácticas, semánticas y contextuales.

OBJETIVOS ESPECÍFICOS

• Diseñar e implementar un módulo de procesamiento de textos periódicos, el cual incluye las tareas de limpieza, etiquetado y segmentación.

• Diseñar e implementar un algoritmo basado en técnicas de minería de textos utilizando características sintácticas, semánticas y contextuales para la identificación de actores y

eventos en textos periodísticos escritos en lenguaje natural (español) a partir de un corpus pre procesado.

• Diseñar e implementar sistema web que integre el proceso completo de procesamiento e identificación de actores y eventos a partir de documentos con formato libre (leguaje natural).

MARCO TEÓRICO

MINERÍA DE DATOS

La minería de datos o exploración de datos es un campo de las ciencias de la computación referido al proceso que intenta descubrir patrones en grandes volúmenes de conjuntos de datos. Utiliza los métodos de la inteligencia artificial, aprendizaje automático, estadística y sistemas de bases de datos.

El objetivo general del proceso de minería de datos consiste en extraer información de un conjunto de datos y transformarla en una estructura comprensible para su uso posterior. Además de la etapa de análisis en bruto, que involucra aspectos de bases de datos y de gestión de datos, de procesamiento de datos, del modelo y de las consideraciones de inferencia, de métricas de Intereses, de consideraciones de la Teoría de la complejidad computacional, de post-procesamiento de las estructuras descubiertas, de la visualización y de la actualización en línea.

La tarea de minería de datos l es el análisis automático o semi-automático de grandes cantidades de datos para extraer patrones interesantes hasta ahora desconocidos, como los grupos de registros de datos (análisis clúster), registros poco usuales (la detección de anomalías) y dependencias (minería por reglas de asociación). Esto generalmente implica el uso de técnicas de bases de datos como los índices espaciales.

Estos patrones pueden entonces ser vistos como una especie de resumen de los datos de entrada, y pueden ser utilizados en el análisis adicional o, por ejemplo, en la máquina de aprendizaje y análisis predictivo.

Por ejemplo, el paso de minería de datos podría identificar varios grupos en los datos, que luego pueden ser utilizados para obtener resultados más precisos de predicción por un sistema de soporte de decisiones. Ni la recolección de datos, preparación de datos, ni la interpretación de los resultados y la información son parte de la etapa de minería de datos, pero que pertenecen a todo el proceso como pasos adicionales.

EXTRACCIÓN DE INFORMACIÓN

La extracción de información incluye aquellos algoritmos, métodos y procesos centrados en la identificación de información dentro de un texto. La posibilidad de localizar determinados elementos dentro del texto facilitará la representación de su contenido semántico. Los cuatro procesos que se describen a continuación proporcionan distintos datos de un texto que facilitan su interpretación:

Identificación de estructuras. Se trata de encontrar, dado un texto, informaciones muy concretas que suelen adoptar estructuras similares. Esto permite emplear patrones que combinan información de estructura con información lingüística.

Identificación de palabras clave. Aparte de reconocer estructuras, es interesante determinar de forma automática qué palabras de un texto resultan más adecuadas para caracterizarlo, es decir, qué palabras deben elegirse como posibles palabras clave. La correcta combinación de la frecuencia de aparición de una palabra en el texto junto con su frecuencia global, es decir, en la red, es un indicativo de la bondad de esa palabra para representar al texto completo

Reconocimiento de entidades con nombre. La posibilidad de reconocer automáticamente la aparición de un nombre propio en un texto es una de las aplicaciones más útiles de la extracción de información.

Elaboración de resúmenes. El procesamiento lingüístico del texto permite determinar qué partes del mismo resultan claves para interpretar su contenido. El proceso incorpora un conjunto de parámetros de configuración que permiten construir resúmenes de calidad para distintos tipos de documentos (noticias de prensa, textos legislativos, documentos internos de empresas, etc.).

REPRESENTACIÓN DE INFORMACIÓN

La información es todo aquello que puede ser manejado por un sistema, ya sea como entrada, como proceso, o bien como resultado. De esta forma, podemos clasificar a los sistemas informáticos como sistemas de flujo de información (si la información de entrada y salida es la misma) y sistemas de tratamiento de la información, en los que la información que entra y la que sale es distinta, ya que ha sufrido alguna manipulación.

Para poder procesar los distintos datos, la computadora debe convertirlos a un lenguaje numérico binario (0 y 1). Debido a la forma en que están construidos y al uso de los componentes electrónicos sólo dos valores pueden representarse. Para convertir los textos en números se utiliza un código de representación llamado ASCII (American Standard Code for Information Interchange) que es un estándar mundial. Una vez pasados a números se deben convertir esos números en valores binarios. Otro tipo de datos como sonidos o imágenes también deben convertirse en valores numéricos. En este apartado se verá cómo pasar de un sistema de numeración cualquiera (como el decimal) al sistema binario de los ordenadores.

DESARROLLO DEL PROYECTO

Para llevar a cabo la correcta implementación de este proyecto se utilizó el lenguaje de programación java en NetBeans como ambiente de desarrollo. En la *Fig.1* se puede observar los módulos que conforman el Sistema de extracciòn.

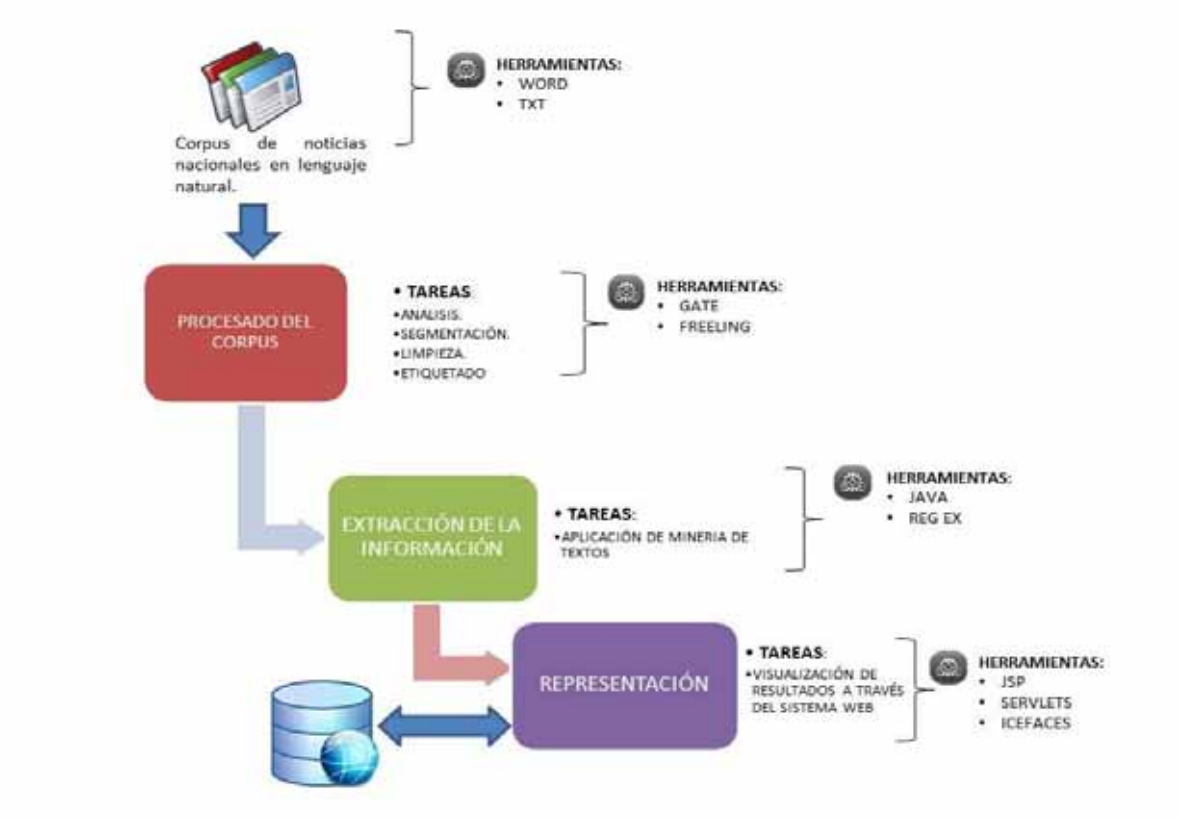


Fig1. Diagrama de módulos del sistema de información para la extracción de eventos y actores en textos periodísticos utilizando técnicas de minera de textos

MÓDULO DE PROCESADO DEL CORPUS

En este módulo se lleva a cabo el análisis, segmentación, limpieza y etiquetado de los textos periodísticos nacionales. Para poder realizarlos aplicaremos el uso de *TreeTagger* como herramienta para las dependencias gramaticales y para el análisis léxico.

El TreeTagger es una herramienta para anotar textos con información de part-of-speech y lema, desarrollado dentro del proyecto TC en el Institute for Computational Linguistics of the University of Stuttgart. Ha sido utilizado con éxito para taggear textos en alemán, inglés, francés, italiano, español, griego, y francés antiguo, y es fácilmente adaptable a otros lenguajes si se dispone de un lexicon y corpus marcado manualmente.

TreeTagger nos regresará un *ArrayList* que contiene: lema, token y pos, que serán trabajados en el modulo de extracción de acuerdo a las reglas que se implementaron. *Ver Anexo A1.* En la *fig.2* se muestra un ejemplo de la aplicación de este módulo.

Texto de entrada:

NOTA SIN PROCESAR

El volcamiento de un bus de turismo, ocurrido en el kilómetro 8 de la vía Bogotá-Choachí, dejó un saldo de 38 personas lesionadas, de las cuales 30 tuvieron que ser remitidas a hospitales . De las personas que iban en el bus y fueron valoradas en el sitio del accidente, 8 no ameritaron traslado a centro asistencial; el resto de los heridos fueron llevados a hospitales.

Texto de salida:

Se encontraron un total de: 1 notas. Iniciando procesado...

Procesado nota: 1

Nota Limpia: El volcamiento de un bus de turismo, ocurrido en el kilómetro 8 de la vía Bogotá-Choachí, dejó un saldo de 38 personas lesionadas, de las cuales 30 tuvieron que ser remitidas a hospitales. De las personas que iban en el bus y fueron valoradas en el sitio del accidente, 8 no ameritaron traslado a centro asistencial; el resto de los heridos fueron llevados a hospitales.

Fig2. Aplicación de módulo de preprocesado de textos.

MÓDULO DE EXTRACCIÓN DE LA INFORMACIÓN.

Este módulo será el encargado de extraer la información de eventos y actores en textos periodísticos. Para llevar a cabo esta tarea se implementarán algoritmos de inteligencia artificial basados en minera de textos en Java.

El módulo se divide en dos clases java que se encargan de la extracción de actores y eventos a partir de la recepción de un arreglo que contiene la información del corpus periodístico segmentado por el TreeTagger.

Para la extracción de actores se implementaron las siguientes reglas ayudándonos del etiquetado POS:

(ART)? (NP|NC)* (ADJ)?

Donde:

* → cero o más veces {3} [Hasta 3 veces]

? → cero o una vez

Para la extracción de actores se implementaron las siguientes reglas ayudándonos del etiquetado POS:

Regla genérica

(V*)|(V*V*)|(V*V*V*)

Donde:

V* → Todas las etiquetas POS que comienzan con V, haciendo relación a apariciones verbales.

Reglas de extracción de lista contemplada

A continuación se muestra una lista detallada con las reglas de extracción sugeridas basándonos en verbos en infinitivo.

acabar de + infinitivo	pasar de + infinitivo
haber que + infinitivo	echarse a + infinitivo
acabar por + infinitivo	ponerse a + infinitivo
ir a + infinitivo	romper a + infinitivo
comenzar a + infinitivo	tener que + infinitivo
llegar a + infinitivo	terminar por + infinitivo
comenzar por + infinitivo	venir a + infinitivo
	estar para + infinitivo
llevar a + infinitivo	venir (de) + infinitivo
deber (de) + infinitivo	estar por + infinitivo
parar de + infinitivo	volver a + infinitivo
dejar de + infinitivo	haber de + infinitivo

Se implementaron las clases descritas en el *Anexo A2* . En la *fig.3* se muestra un ejemplo de la aplicación de este módulo con la información obtenida del módulo anterior.

Texto de entrada:

Entrada al módulo de extracción

Nota Limpia: El volcamiento de un bus de turismo , ocurrido en el kilómetro 8 de la vía Bogotá-Choachí , dejó un saldo de 38 personas lesionadas , de las cuales 30 tuvieron que ser remitidas a hospitales . De las personas que iban en el bus y fueron valoradas en el sitio del accidente , 8 no ameritaron traslado a centro asistencial; el resto de los heridos fueron llevados a hospitales .

Texto de salida:

Iniciando procesado...

Procesado nota: 1

Actores:[El volcamiento de , turismo , Bogotá-Choachí , personas , hospitales , accidente , traslado , centro , hospitales]

Eventos: [ocurrido, dejó, lesionadas, tuvieron, ser remitidas, remitidas, iban, fueron valoradas, valoradas, ameritaron, asistencial;, heridos fueron llevados, fueron llevados, llevados]

Fig3. Aplicación de módulo de extracción de información.

MÓDULO DE REPRESENTACIÓN.

Para la visualización se utilizó java y html para la programación de una página web por medio de tecnologías Web (*JSP)* .

JavaServer Pages (JSP) es una tecnología que ayuda a los desarrolladores de software a crear páginas web dinámicas basadas en HTML, XML, entre otros tipos de documentos. JSP es similar a PHP, pero usa el lenguaje de programación Java.

Para desplegar y correr JavaServer Pages, se requiere un servidor web compatible con contenedores servlet como Apache Tomcat.

La principal ventaja de JSP frente a otros lenguajes es que el lenguaje Java es un lenguaje de propósito general que excede el mundo web y que es apto para crear clases que manejen lógica de negocio y acceso a datos de una manera prolija. Esto permite separar en niveles las aplicaciones web, dejando la parte encargada de generar el documento HTML en el archivo JSP.

Otra ventaja es que JSP hereda la portabilidad de Java, y es posible ejecutar las aplicaciones en múltiples plataformas sin cambios. Es común incluso que los desarrolladores trabajen en una plataforma y que la aplicación termine siendo ejecutada en otra. *Ver Anexo A3*

En la *fig.4* se muestra un ejemplo de la aplicación de este módulo ejecutando la aplicación desde la url http://aisii.azc.uam.mx:8080/PIIC_LuisDavidHernandezRojas

Fig4. Aplicación de módulo de representación de información.

RESULTADOS

Al finalizar el desarrollo del sistema web se realizó una prueba con un corpus de textos periodísticos con un total de 324 notas.

A continuación desplegamos el porcentaje de éxito a cada uno de los módulos:

$$Efec_{notas} = \frac{notas_{ext}}{notas_{total}} * 100$$

$$Efec_{actor} = \frac{actor_{ext}}{actor_{total}} * 100$$

$$Efec_{evento} = \frac{evento_{ext}}{evento_{total}} * 100$$

En las *Tabla 1* se muestran los resultados de extracción de la información en la prueba realizada sobre el corpus de 324 notas.

Total de notas	Notas correctas	Notas Incorrectas	Efecnotas
324	324	0	100%

Total de actores	Actores correctos	Actores Incorrectos	Efecactor
4223	3856	367	91.3%

Total de eventos	Eventos correctos	Eventos Incorrectos	Efecevento
6782	5876	906	86.5%

Tabla 1. Resultado de la ejecución del sistema web.

El porcentaje obtenido para el corpus de prueba es el siguiente:

$Efec_{notas} = 100\%$

$Efec_{actor} = 91.3\%$

$Efec_{eventos} = 86.5\%$

ANALISIS Y DISCUSIÓN DE LOS RESULTADOS

Como se mostró en la sección de resultados, podemos observar que la efectividad en la extracción de actores y eventos es variable y no logra obtenerse un 100%. Esto se debe a diversos factores, entre los cuales se listan los siguientes:

Para actores.

1.- El etiquetador TreeTagger en ocasiones desconoce el contexto de las palabras y puede cambiar la etiqueta pertinente por otra que considera adecuada.

2.- Existen ciertos errores en las reglas de extracción extracción de actores, lo que implica que puede existir duplicidad en los ítems obtenidos.

Para eventos.

1.- El etiquetador TreeTagger en ocasiones desconoce el contexto de las palabras y puede cambiar la etiqueta pertinente por otra que considera adecuada.

2.- Existen ciertos errores en las reglas de extracción de eventos, lo que implica que puede existir duplicidad en los ítems obtenidos.

Estos errores pueden corregirse con la adecuación de las reglas de extracción de actores y eventos para lograr obtener resultados más precisos.

CONCLUSIONES

Al término del desarrollo del Proyecto de Integración, podemos concluir que se cumplió satisfactoriamente con los objetivos propuestos. Se impelmentará un sistema web de extracción de actores y eventos de textos periodísticos a través de minería de datos que facilitará el análisis de grandes cantidades de información para su futura manipulación.

Así mismo, se da un precedente para que el sistema pueda ser manipulado y adaptado en un futuro para poder obtener mejores resultados y que la información extraída pueda ser almacenada en una base de datos para poder realizar un análisis estadístico.

REFERENCIAS BIBLIOGRÁFICAS

[1] A. Urquiza Pérez ,"Sistema Configurable de Minería Web", proyecto terminal, División de Ciencias Básicas e Ingeniería, Universidad Autónoma Metropolitana Azcapotzalco, México, 2012.

[2] L.Y. García Jurado, "Sistema de recuperación de información de textos de investigación de la web", proyecto terminal, División de Ciencias Básicas e Ingeniería, Universidad Autónoma Metropolitana Azcapotzalco, México, 2013.

[3] F. Tebar Martínez, "Sistema de almacenamiento semántico y recuperación de textos de investigación mediante ontologías", proyecto terminal, División de Ciencias Básicas e Ingeniería, Universidad Autónoma Metropolitana Azcapotzalco, México, 2013.

[4] A. Téllez Valera, "Extracción de información con algoritmos de clasificación", tesis de maestría, Instituto Nacional de Astro física Óptica y Electrónica, México, Puebla, 2005.

[5] J.C. Riquelme, R. Ruiz, K. Gilbert, "Minería de datos: Conceptos y tendencias". Revista Iberoamericana de Inteligencia Artificial, vol. 10, no 29, p. 11-18., 2006.

[6] A. Benini, "Text Analysis under Time Pressure Tools for humanitarian and development worker" Washington, DC, 2009

ANEXO A1: PROCESADO DEL CORPUS

INICIO DEL PROCESO INVOCADO DESDE LAS JSP WEB

```
/* Proceso.java */
package pt;
import java.io.*;
import org.annolab.tt4j.TreeTaggerException;
import java.util.ArrayList;
import java.util.HashSet;
import java.util.List;
import java.util.Set;
import javax.servlet.http.HttpSession;
import org.apache.commons.io.FilenameUtils;
import org.apache.commons.io.IOUtils;
import javax.servlet.ServletContextEvent;

public class Proceso  {


   public  ArrayList<ArrayList<String>> inicio() throws  TreeTaggerException, IOException {

    ArrayList <ArrayList <String>> informacion = new ArrayList <ArrayList <String>> () ;
   ArrayList <String> noticias= new ArrayList <String> ();
   ArrayList <String> even= new ArrayList<String>();
   ArrayList <String> actor=new ArrayList<String>();

      /* AQUI SE HACE LA LLAMADA A LA CLASE QUE UBICA LAS PALABRAS QUE SERÁN ELIMINADAS*/
   vacias l= new vacias();
   String [] re= l.pvacias();


    String nota="";
      try{
          // Abrimos el archivo
          FileInputStream fstream = new FileInputStream("C:/LDHR/temp.txt");
          // Creamos el objeto de entrada
          DataInputStream entrada = new DataInputStream(fstream);
          // Creamos el Buffer de Lectura
          BufferedReader buffer = new BufferedReader(new InputStreamReader(entrada));
          String strLinea;
          // Leer el archivo linea por linea
          while ((strLinea = buffer.readLine()) != null)
          {
```

```
            nota = strLinea;

            /* AQUI SE HACE LA LLAMADA A LA CLASE QUE LIMPIA EL CORPUS*/
    limpieza b= new limpieza();
    String cadlimp= b.cadenalimpia(re, nota);
    noticias.add(cadlimp);

    etiquetador n= new etiquetador();
    ArrayList <ArrayList <String>> aR = new ArrayList <ArrayList <String>> () ;
    aR=n.retorno(cadlimp);

exactores a2= new exactores();
String Sactor= a2.act2(aR);
actor.add(Sactor);

exeventos e2=new exeventos();
String Seven=e2.ev2(aR);
even.add(Seven);

          }
          // Cerramos el archivo e introducimos la información que se va a rertornar
          entrada.close();
          informacion.add(noticias);
          informacion.add(actor);
          informacion.add(even);


       }
       catch (Exception e){ //Catch de excepciones
          System.err.println("Ocurrio un error: " + e.getMessage());
       }

    return informacion;

    }
}
```

Módulo de procesado del corpus

/* vacias.java*/

```
package pt;
import java.io.*;
import java.io.File;
import java.util.ArrayList;
import java.util.Arrays;
import java.util.logging.Level;
import java.util.logging.Logger;

public class vacias {
   String [] vacias ;
   public String []  pvacias ()
   {
      File f = new File("C:/LDHR/PVacias.txt");
      BufferedReader entrada;

      try {
         entrada = new BufferedReader(new FileReader(f));
         String linea;
         while ((linea =entrada.readLine()) != null)
         {
          vacias = linea.split(",");
          }
      }
      catch (IOException e) {
         e.printStackTrace();
      }
            return vacias;

  }
}
```

/*limpieza.java*/

```
package pt;
import java.io.*;

public class limpieza {
   String cl=null;
   public String cadenalimpia (String [] palabras, String noticia) throws IOException
   {
      cl=noticia;
      for (int i = 0; i < palabras.length; i++)
       {
         cl=cl.replaceAll(" " + palabras[i] + " " ," ");
```

```
        }
      return cl;
   }
 }
```

/* etiquetador.java*/

```
package pt;
import java.io.IOException;
import java.util.ArrayList;
import org.annolab.tt4j.*;
import static java.util.Arrays.asList;

public class etiquetador {

   public ArrayList<ArrayList <String>>   retorno (String cadenalimpia) throws IOException, TreeTaggerException
   {
ArrayList <ArrayList <String>> arreglo = new ArrayList <ArrayList <String>> () ;
final ArrayList <String> p=new ArrayList<String> ();
final ArrayList <String>l =new ArrayList<String> ();
final ArrayList <String>t =new ArrayList<String> ();

 String [] c= cadenalimpia.split(" ");
      System.setProperty("treetagger.home", "C:/TreeTagger");
      TreeTaggerWrapper<String> tt = new TreeTaggerWrapper<String>();

      try {
            tt.setModel("C:/TreeTagger/modelos/spanish.par:iso8859-1");
            tt.setHandler(new TokenHandler<String>()
            {
                  public void token(String token, String pos, String lemma)
                  {
                   p.add(pos);
                   l.add(lemma);
                   t.add(token);
                  }
            });

           tt.process(asList(c));

      }
      finally

      {
         tt.destroy();
      }
      arreglo.add(p);
      arreglo.add(t);
```

```
      arreglo.add(l);

      return arreglo;
   }
}
```

ANEXO A2: EXTRACCIÓN DE LA INFORMACIÓN

Módulo de extracción de la información.

/*exactores.java */

```
package pt;
import java.io.IOException;
import java.util.ArrayList;

public class exactores {

   public String act2 (ArrayList<ArrayList<String>> arr) throws IOException

   {
      ArrayList<String> npos= new ArrayList<String> ();
      ArrayList<String> ntok= new ArrayList<String> ();
      ArrayList<String> nlem= new ArrayList<String> ();
      ArrayList <String> act=new ArrayList<String> ();

         npos.addAll(arr.get(0));
         ntok.addAll(arr.get(1));
         nlem.addAll(arr.get(2));
        int c= npos.size();
        String h="";

        for (int i = 0; i < c ; i++)
        {
          if(i==0) { //IF1
             // Caso: ART+(NP|NC)+(NP|NC)+(NP|NC)+ADJ
           if ( ( npos.get(i).equalsIgnoreCase("ART")) && ((npos.get(i+1).equalsIgnoreCase("NP") || npos.get(i+1).equalsIgnoreCase("NC"))) &&
((npos.get(i+2).equalsIgnoreCase("NP") || npos.get(i+2).equalsIgnoreCase("NC")))&& ((npos.get(i+3).equalsIgnoreCase("NP") ||
npos.get(i+3).equalsIgnoreCase("NC"))) && (npos.get(i+4).equalsIgnoreCase("ADJ") ) )
           {
           h = ntok.get(i)+ " "+ ntok.get(i+1)+ " "+ ntok.get(i+2)+ " "+ ntok.get(i+3)+ " "+ ntok.get(i+4);
                        act.add(h);
           }
           // Caso: ART+(NP|NC)+(NP|NC)+ADJ
```

```
            else  if ( ( npos.get(i).equalsIgnoreCase("ART")) && ((npos.get(i+1).equalsIgnoreCase("NP") ||
npos.get(i+1).equalsIgnoreCase("NC"))) && ((npos.get(i+2).equalsIgnoreCase("NP") || npos.get(i+2).equalsIgnoreCase("NC"))) &&
(npos.get(i+3).equalsIgnoreCase("ADJ") ) )
            {
            h = ntok.get(i)+ " "+ ntok.get(i+1)+ " "+ ntok.get(i+2)+ " "+ ntok.get(i+3);
                        act.add(h);
            }
            // Caso: ART+(NP|NC)+ADJ
            else
            if ( ( npos.get(i).equalsIgnoreCase("ART")) && ((npos.get(i+1).equalsIgnoreCase("NP") || npos.get(i+1).equalsIgnoreCase("NC")))
&& (npos.get(i+2).equalsIgnoreCase("ADJ") ) )
            {
            h = ntok.get(i)+ " "+ ntok.get(i+1)+ " "+ ntok.get(i+2);
                        act.add(h);
            }

            else if ( (npos.get(i).equalsIgnoreCase("ART")) && ((npos.get(i+1).equalsIgnoreCase("NP") || npos.get(i+1).equalsIgnoreCase("NC")))
&& ((npos.get(i+2).equalsIgnoreCase("NP") || npos.get(i+2).equalsIgnoreCase("NC")))&& ((npos.get(i+3).equalsIgnoreCase("NP") ||
npos.get(i+3).equalsIgnoreCase("NC"))) && (!npos.get(i+4).equalsIgnoreCase("ADJ") )  )
            {
            h = ntok.get(i)+ " "+ ntok.get(i+1)+ " "+ ntok.get(i+2)+ " "+ ntok.get(i+3);
                        act.add(h);
            }
            else
            if ( ( npos.get(i).equalsIgnoreCase("ART")) && ((npos.get(i+1).equalsIgnoreCase("NP") || npos.get(i+1).equalsIgnoreCase("NC"))) &&
((npos.get(i+2).equalsIgnoreCase("NP") || npos.get(i+2).equalsIgnoreCase("NC"))) && (!npos.get(i+3).equalsIgnoreCase("ADJ") )  )
            {
            h = ntok.get(i)+ " "+ ntok.get(i+1)+ " "+ ntok.get(i+2);
                        act.add(h);
            }
            else
            if ( ( npos.get(i).equalsIgnoreCase("ART")) && ((npos.get(i+1).equalsIgnoreCase("NP") || npos.get(i+1).equalsIgnoreCase("NC"))) &&
(!npos.get(i+2).equalsIgnoreCase("ADJ") )  )
            {
            h = ntok.get(i)+ " "+ ntok.get(i+1);
                        act.add(h);
            }
            else
  // Caso: (NP|NC)+(NP|NC)+(NP|NC)+ADJ
            if ( ((npos.get(i).equalsIgnoreCase("NP") || npos.get(i).equalsIgnoreCase("NC"))) && ((npos.get(i+1).equalsIgnoreCase("NP") ||
npos.get(i+1).equalsIgnoreCase("NC")))&& ((npos.get(i+2).equalsIgnoreCase("NP") || npos.get(i+2).equalsIgnoreCase("NC"))) &&
(npos.get(i+3).equalsIgnoreCase("ADJ") ) )
            {
            h = ntok.get(i)+ " "+ ntok.get(i+1)+ " "+ ntok.get(i+2)+ " "+ ntok.get(i+3)+ " ";
                        act.add(h);
            }
            // Caso: (NP|NC)+(NP|NC)+ADJ
```

```
        else   if ( ((npos.get(i).equalsIgnoreCase("NP") || npos.get(i).equalsIgnoreCase("NC"))) && ((npos.get(i+1).equalsIgnoreCase("NP") ||
npos.get(i+1).equalsIgnoreCase("NC"))) && (npos.get(i+2).equalsIgnoreCase("ADJ") ) )
            {
            h = ntok.get(i)+ " "+ ntok.get(i+1)+ " "+ ntok.get(i+2)+ " ";
                        act.add(h);
            }
            // Caso: (NP|NC)+ADJ
            else
            if (  ((npos.get(i).equalsIgnoreCase("NP") || npos.get(i).equalsIgnoreCase("NC")))  && (npos.get(i+1).equalsIgnoreCase("ADJ") ) )
            {
            h = ntok.get(i)+ " "+ ntok.get(i+1)+ " ";
                        act.add(h);
            }
              // Caso: (NP|NC)+(NP|NC)+(NP|NC)
        else   if ( ((npos.get(i).equalsIgnoreCase("NP") || npos.get(i).equalsIgnoreCase("NC"))) && ((npos.get(i+1).equalsIgnoreCase("NP") ||
npos.get(i+1).equalsIgnoreCase("NC")))&& ((npos.get(i+2).equalsIgnoreCase("NP") || npos.get(i+2).equalsIgnoreCase("NC"))) &&
(!npos.get(i+3).equalsIgnoreCase("ADJ") ) )
            {
            h = ntok.get(i)+ " "+ ntok.get(i+1)+ " "+ ntok.get(i+2)+ " ";
                        act.add(h);
            }
            // Caso: (NP|NC)+(NP|NC)+ADJ
          else if ( ((npos.get(i).equalsIgnoreCase("NP") || npos.get(i).equalsIgnoreCase("NC"))) && ((npos.get(i+1).equalsIgnoreCase("NP") ||
npos.get(i+1).equalsIgnoreCase("NC"))) && (!npos.get(i+2).equalsIgnoreCase("ADJ") ) )
            {
            h = ntok.get(i)+ " "+ ntok.get(i+1)+ " "+ ntok.get(i+2)+ " ";
                        act.add(h);
            }
            // Caso: (NP|NC)+ADJ
            else
            if (  ((npos.get(i).equalsIgnoreCase("NP") || npos.get(i).equalsIgnoreCase("NC")))  && (!npos.get(i+1).equalsIgnoreCase("ADJ") ) )
            {
            h = ntok.get(i)+ " ";
                        act.add(h);
            }
          }
         else
          if(i>1)
            {
               if ( ( !npos.get(i-1).equalsIgnoreCase("ART")) && ((npos.get(i).equalsIgnoreCase("NP") || npos.get(i).equalsIgnoreCase("NC")))
&& ((npos.get(i+1).equalsIgnoreCase("NP") || npos.get(i+1).equalsIgnoreCase("NC")))&& ((npos.get(i+2).equalsIgnoreCase("NP") ||
npos.get(i+2).equalsIgnoreCase("NC"))) && (!npos.get(i+3).equalsIgnoreCase("ADJ") )  )
            {
            h =ntok.get(i)+ " "+ ntok.get(i+1)+ " "+ ntok.get(i+2);
                        act.add(h);
            }
            else
```

```
            if ( ( !npos.get(i-1).equalsIgnoreCase("ART")) && ((npos.get(i).equalsIgnoreCase("NP") || npos.get(i).equalsIgnoreCase("NC"))) &&
((npos.get(i+1).equalsIgnoreCase("NP") || npos.get(i+1).equalsIgnoreCase("NC"))) && (!npos.get(i+2).equalsIgnoreCase("ADJ") )  )
            {
            h = ntok.get(i)+ " "+ ntok.get(i+1);
                        act.add(h);
            }
            else
            if ( ( !npos.get(i-1).equalsIgnoreCase("ART")) && ((npos.get(i).equalsIgnoreCase("NP") || npos.get(i).equalsIgnoreCase("NC"))) &&
(!npos.get(i+1).equalsIgnoreCase("ADJ") )  )
            {
            h = ntok.get(i)+ " ";
                        act.add(h);
            }
            else
            if ( ( !npos.get(i-1).equalsIgnoreCase("ART")) && ((npos.get(i).equalsIgnoreCase("NP") || npos.get(i).equalsIgnoreCase("NC"))) &&
((npos.get(i+1).equalsIgnoreCase("NP") || npos.get(i+1).equalsIgnoreCase("NC")))&& ((npos.get(i+2).equalsIgnoreCase("NP") ||
npos.get(i+2).equalsIgnoreCase("NC"))) && (npos.get(i+3).equalsIgnoreCase("ADJ") ) )
            {
            h =ntok.get(i)+ " "+ ntok.get(i+1)+ " "+ ntok.get(i+2)+ " "+ ntok.get(i+3);
                        act.add(h);
            }
            else
            if ( ( !npos.get(i-1).equalsIgnoreCase("ART")) && ((npos.get(i).equalsIgnoreCase("NP") || npos.get(i).equalsIgnoreCase("NC"))) &&
((npos.get(i+1).equalsIgnoreCase("NP") || npos.get(i+1).equalsIgnoreCase("NC"))) && (npos.get(i+2).equalsIgnoreCase("ADJ") ) )
            {
            h = ntok.get(i)+ " "+ ntok.get(i+1)+ " "+ ntok.get(i+2);
                        act.add(h);
            }
           else     if ( ( !npos.get(i-1).equalsIgnoreCase("ART")) && ((npos.get(i).equalsIgnoreCase("NP") ||
npos.get(i).equalsIgnoreCase("NC"))) && (npos.get(i+1).equalsIgnoreCase("ADJ") ) )
           {
            h = ntok.get(i)+ " "+ ntok.get(i+1);
                        act.add(h);
            }
            }
         }
  return act.toString();
   }
}
```

/*exeventos.java */

```
package pt;
import java.io.IOException;
import java.util.ArrayList;
```

```
public class exeventos {

   public String ev2 (ArrayList<ArrayList<String>> arr) throws IOException

   {
       ArrayList<String> npos= new ArrayList<String> ();
       ArrayList<String> ntok= new ArrayList<String> ();
       ArrayList<String> nlem= new ArrayList<String> ();
       ArrayList <String> eve=new ArrayList<String> ();

    npos.addAll(arr.get(0));
    ntok.addAll(arr.get(1));
    nlem.addAll(arr.get(2));
      int c= npos.size();
      String h="";
          for (int i = 0; i < c ; i++) {

             /* ***** REGLAS GENERICAS DE EXTRACCI�N ***** */
             if (( npos.get(i).startsWith("V") ) && ( npos.get(i+1).startsWith("V") ) && ( npos.get(i+2).startsWith("V") ) )
               {
                     h = ntok.get(i)+ " "+ ntok.get(i+1)+ " "+ ntok.get(i+2);
                                 eve.add(h);
           }
             else
                if (( npos.get(i).startsWith("V") ) && ( npos.get(i+1).startsWith("V") )   )
               {
                     h = ntok.get(i)+ " "+ ntok.get(i+1);
                                 eve.add(h);
           }
             else
                  if (( npos.get(i).startsWith("V") ) )
               {
                     h = ntok.get(i);
                                 eve.add(h);
           }

    /* ****** FIN DE LAS REGLAS GENERICAS ******* */

            else

  /* +++++ REGLAS DE EXTRACCION DE LA LISTA CONTEMPLADA +++++ */

            if ( nlem.get(i).equalsIgnoreCase("acabar") && (( nlem.get(i+1).equalsIgnoreCase("de"))||( nlem.get(i+1).equalsIgnoreCase("por")) )
&& ( npos.get(i+2).endsWith("inf")) )
                {
                     h = ntok.get(i)+ " "+ ntok.get(i+1)+ " "+ ntok.get(i+2);
                                 eve.add(h);
```

```
        }
        else
            if ( (nlem.get(i).equalsIgnoreCase("haber")) && (( nlem.get(i+1).equalsIgnoreCase("de"))||(
nlem.get(i+1).equalsIgnoreCase("que"))) && ( npos.get(i+2).endsWith("inf")) )
            {

                h = ntok.get(i)+ " "+ ntok.get(i+1)+ " "+ ntok.get(i+2);
                        eve.add(h);
              }

        else
            if (( nlem.get(i).equalsIgnoreCase("ir")) && ( nlem.get(i+1).equalsIgnoreCase("a")) && ( npos.get(i+2).endsWith("inf"))  )
            {

                h = ntok.get(i)+ " "+ ntok.get(i+1)+ " "+ ntok.get(i+2);
                        eve.add(h);
              }
      else
            if (( nlem.get(i).equalsIgnoreCase("comenzar")) && (( nlem.get(i+1).equalsIgnoreCase("a"))||(
nlem.get(i+1).equalsIgnoreCase("por"))) &&( npos.get(i+2).endsWith("inf")) )
            {

                h = ntok.get(i)+ " "+ ntok.get(i+1)+ " "+ ntok.get(i+2);
                        eve.add(h);
              }

    else
            if (( nlem.get(i).equalsIgnoreCase("llegar")) && ( nlem.get(i+1).equalsIgnoreCase("a")) && ( npos.get(i+2).endsWith("inf"))  )
            {

                h = ntok.get(i)+ " "+ ntok.get(i+1)+ " "+ ntok.get(i+2);
                        eve.add(h);
              }

        else
            if (( nlem.get(i).equalsIgnoreCase("llevar")) && ( nlem.get(i+1).equalsIgnoreCase("a")) && ( npos.get(i+2).endsWith("inf"))  )
            {

                h = ntok.get(i)+ " "+ ntok.get(i+1)+ " "+ ntok.get(i+2);
                        eve.add(h);
              }

        else
            if (( nlem.get(i).equalsIgnoreCase("deber")) && ( nlem.get(i+1).equalsIgnoreCase("de"))  && ( npos.get(i+2).endsWith("inf"))  )
            {

                 h = ntok.get(i)+ " "+ ntok.get(i+1)+ " "+ ntok.get(i+2);
```

```
                    eve.add(h);
            }
        else
              if (( nlem.get(i).equalsIgnoreCase("deber")) && ( npos.get(i+1).endsWith("inf"))  )
         {

              h = ntok.get(i)+ " "+ ntok.get(i+1);
                    eve.add(h);
            }

        else
         if (( nlem.get(i).equalsIgnoreCase("parar")) && ( nlem.get(i+1).equalsIgnoreCase("de")) && ( npos.get(i+2).endsWith("inf")) )
         {

              h = ntok.get(i)+ " "+ ntok.get(i+1)+ " "+ ntok.get(i+2);
                    eve.add(h);
                 }

        else
         if (( nlem.get(i).equalsIgnoreCase("pasar")) && ( nlem.get(i+1).equalsIgnoreCase("de")) && ( npos.get(i+2).endsWith("inf")) )
         {

              h = ntok.get(i)+ " "+ ntok.get(i+1)+ " "+ ntok.get(i+2);
                    eve.add(h);
            }

    else
         if (( nlem.get(i).equalsIgnoreCase("dejar")) && ( nlem.get(i+1).equalsIgnoreCase("de")) && ( npos.get(i+2).endsWith("inf")) )
         {

              h = ntok.get(i)+ " "+ ntok.get(i+1)+ " "+ ntok.get(i+2);
                    eve.add(h);
         }

  else
         if (( nlem.get(i).equalsIgnoreCase("echarse")) && ( nlem.get(i+1).equalsIgnoreCase("a")) && ( npos.get(i+2).endsWith("inf")) )
         {

              h = ntok.get(i)+ " "+ ntok.get(i+1)+ " "+ ntok.get(i+2);
                    eve.add(h);
         }

         else

         if (( nlem.get(i).equalsIgnoreCase("ponerse")) && ( nlem.get(i+1).equalsIgnoreCase("a")) && ( npos.get(i+2).endsWith("inf")) )
         {
```

```
                h = ntok.get(i)+ " "+ ntok.get(i+1)+ " "+ ntok.get(i+2);
                        eve.add(h);
            }

        else
            if (( nlem.get(i).equalsIgnoreCase("romper")) && ( nlem.get(i+1).equalsIgnoreCase("a")) && ( npos.get(i+2).endsWith("inf")) )
            {

                h = ntok.get(i)+ " "+ ntok.get(i+1)+ " "+ ntok.get(i+2);
                        eve.add(h);
            }

             else
            if (( nlem.get(i).equalsIgnoreCase("tener")) && ( nlem.get(i+1).equalsIgnoreCase("que")) && ( npos.get(i+2).endsWith("inf")) )
            {

                h = ntok.get(i)+ " "+ ntok.get(i+1)+ " "+ ntok.get(i+2);
                        eve.add(h);
            }

             else
            if (( nlem.get(i).equalsIgnoreCase("terminar")) && ( nlem.get(i+1).equalsIgnoreCase("por")) && ( npos.get(i+2).endsWith("inf")) )
            {

                h = ntok.get(i)+ " "+ ntok.get(i+1)+ " "+ ntok.get(i+2);
                        eve.add(h);
            }

              else
            if (( nlem.get(i).equalsIgnoreCase("venir")) && (( nlem.get(i+1).equalsIgnoreCase("a"))||( nlem.get(i+1).equalsIgnoreCase("de")
)) && ( npos.get(i+2).endsWith("inf")) )
            {

                h = ntok.get(i)+ " "+ ntok.get(i+1)+ " "+ ntok.get(i+2);
                        eve.add(h);
            }

            else
            if (( nlem.get(i).equalsIgnoreCase("estar")) && (( nlem.get(i+1).equalsIgnoreCase("para"))||(
nlem.get(i+1).equalsIgnoreCase("por") )) && ( npos.get(i+2).endsWith("inf")) )
            {

                h = ntok.get(i)+ " "+ ntok.get(i+1)+ " "+ ntok.get(i+2);
                        eve.add(h);
            }

            else
```

```
            if (( nlem.get(i).equalsIgnoreCase("volver")) && ( nlem.get(i+1).equalsIgnoreCase("por")) && ( npos.get(i+2).endsWith("inf")) )
            {

                h = ntok.get(i)+ " "+ ntok.get(i+1)+ " "+ ntok.get(i+2);
                        eve.add(h);
            }
        }
return eve.toString();
          }
}
```

ANEXO A3: REPRESENTACIÓN DE LA INFORMACIÓN

```
<%-- index.jsp --%>
<%@page import="java.io.File"%>
<%@page import="pt.Proceso"%>
<%@ page contentType="text/html; charset=utf-8"
      import="pt.Principal"
      import="pt.exactores"
      errorPage=""%>

<!DOCTYPE html>
<html>

  <head>

  <meta charset="utf-8">

  <!-- Use the .htaccess and remove these lines to avoid edge case issues.
       More info: h5bp.com/b/378 -->

  <title>Sistema de Extracción </title>
          <meta http-equiv="X-UA-Compatible" content="IE=edge" />
          <meta name="viewport" content="width=device-width, initial-scale=1.0, user-scalable=no" />
          <!-- Icono del navegador -->
          <link rel="shortcut icon" href="img/favicon.png"/>
          <!-- Bootstrap Core CSS -->
          <link href="css/bootstrap.css" rel="stylesheet" type="text/css" />
          <link href="css/freelancer.css" rel="stylesheet" type="text/css" />
          <!-- Fonts -->
          <link href="font-awesome/css/font-awesome.min.css" rel="stylesheet" type="text/css" />
          <link href='http://fonts.googleapis.com/css?family=Montserrat:400,700' rel='stylesheet' type='text/css' />
          <!-- Js -->
          <script src="js/jquery-1.10.2.js"></script>
```

```
        <script src="js/bootstrap.min.js"></script>
        <script src="http://cdnjs.cloudflare.com/ajax/libs/jquery-easing/1.3/jquery.easing.min.js"></script>
        <script src="js/classie.js"></script>
        <script src="js/cbpAnimatedHeader.js"></script>
        <script src="js/freelancer.js"></script>
    <style type="text/css">
        .titulo_contenido h1, h2, h3, h4, h5, h6{
            text-align: center;
        }
        #resultados{
            text-align: left;
        }
        .resultados ul{
            text-decoration: none;
        }
        .nav_cabecera a:hover{
            color: #fff;
        }
        .img_help{
            width: 250px;
        }
        .img_logo{
            width: 80px;
            margin-top: -10px;
            margin-right: 15px;
        }
        .align_item{
            text-align: left;
        }
        .text_help{
            text-align: justify;
        }
        th {
            background-color: #00003b;
            color: white;
            text-align:justify;
        }
        hr {
            display: block;
            margin-top: 0.5em;
            margin-bottom: 0.5em;
            margin-left: auto;
            margin-right: auto;
            border-style: inset;
            border-width: 1px;
        }
    </style>
```

```
    <!-- Encabezado -->

    <div class="container">
            <!-- Brand and toggle get grouped for better mobile display -->
            <div class="navbar-header page-scroll">
              <button type="button" class="navbar-toggle" data-toggle="collapse" data-target="#bs-example-navbar-collapse-1">
                <span class="sr-only">Toggle navigation</span>
                <span class="icon-bar"></span>
                <span class="icon-bar"></span>
                <span class="icon-bar"></span>
              </button>
              <a class="navbar-brand" href="#page-top" ><h3>SISTEMA DE EXTRACCIÓN WEB DE ACTORES Y EVENTOS EN TEXTOS
PERIODISTICOS</h3><br /></a>
            </div>
            <!-- Coleccion de nav links -->
            <br><br><br><br>
            <div class="collapse navbar-collapse" id="bs-example-navbar-collapse-1">
              <ul class="nav navbar-nav navbar-right">
                <li class="hidden">
                  <a href="#page-top"></a>
                </li>
                <li class="page-scroll">
                  <a href="#portfolioModal1" class="portfolio-link" data-toggle="modal">Guía</a>
                </li>
                <li class="page-scroll">
                  <a href="#portfolioModal2" class="portfolio-link" data-toggle="modal">Acerca de</a>
                </li>
                <li class="page-scroll">
                  <a href="#portfolioModal3" class="portfolio-link" data-toggle="modal">Minería de datos</a>
                </li>
              </ul>
            </div>
            <!-- /.navbar-collapse -->
          </div>
</head>


<BODY BACKGROUND="img/fondo.jpg">

   <!-- DESDE AQUI ESTA EL CUERPO -->

<!-- GUIA DE USO -->
        <div class="portfolio-modal modal fade" id="portfolioModal1" tabindex="-1" role="dialog" aria-hidden="true">
          <div class="modal-content">
            <div class="close-modal" data-dismiss="modal">
              <div class="lr">
```

```
                    <div class="rl"></div>
                  </div>
                </div>
                <div class="container">
                  <div class="row">
                    <div class="col-lg-8 col-lg-offset-2">
                      <div class="modal-body">
                        <h2>Guía de uso</h2>
                        <hr /><br />
                        <img src="img/logomcfblanco.jpg" /><br/>
                        <div class="text_help">
                          <ul>
                            <li type="square">NOTAS IMPORTANTES
                              <ol>
                                <li>El corpus debe cargarse en archivo .txt
                                  <ol>
                                    <li><strong>Corpus de textos periodísticos: </strong> Se puede llamar corpus a cualquier colección que contenga más de un texto periodístico.</li>
                                    <li><strong>Lenguaje: </strong>Es importante que las noticias esten escritas en español.</li>
                                  </ol>

                                </li>
                              </ol>
                            </li>
                            <li type="square">CARGAR
                              <ol>
                                <li>Dar click en el botón <strong>Examinar</strong> para buscar el archivos a procesar. Este corpus será cargado temporalmente en el servidor con el fin de que el proceso se realice en el servidor.</li>
                                <li>Se abrirá un explorador de archivos, seleccione el archivo del corpus con extensión <strong>.txt</strong> y de clic en el botón <strong>Aceptar</strong>.</li>
                                <li>Si desea cancelar la subida de archivos, puede dar click en el botón <strong>Cancelar</strong> para eliminar los archivos seleccionados y/o limpiar la lista de archivos a subir.</li>
                                <li>Dar click en el botón <strong>Comenzar extraccion</strong> para comenzar el proceso</li>

                              </ol>
                            </li>
                            <br />

                          </ul>
                          <br />
                          <br />
                        </div>
                        <img src="img/logoarxfblanco.jpg" /><br/>
                        <button type="button" class="btn btn-default" data-dismiss="modal"><i class="fa fa-times"></i> Cerrar</button>
                      </div>
                    </div>
                  </div>
```

```
            </div>
          </div>
        </div>

<!-- ACERCA DE  -->
        <div class="portfolio-modal modal fade" id="portfolioModal2" tabindex="-1" role="dialog" aria-hidden="true">
          <div class="modal-content">
            <div class="close-modal" data-dismiss="modal">
              <div class="lr">
                <div class="rl">
                </div>
              </div>
            </div>
            <div class="container">
              <div class="row">
                <div class="col-lg-8 col-lg-offset-2">
                  <div class="text_help">
                    <h2>Acerca de</h2>
                    <hr /><br />
                    <img src="img/logomcfblanco.jpg" />
                    <p>El <a>Sistema de extracción de eventos y actores en textos periodísticos utilizando técnicas de minera de
textos</a> fue desarrollado como Proyecto de Integración para la carrera de Ingeniería en computación.</p>
                    <p>El sistema fue desarrollado con las siguientes tecnologías:
                      <p><a target="_blank" href="http://tomcat.apache.org/">Apache Tomcat</a>
                      <p><a target="_blank" href="http://www.oracle.com/technetwork/java/javase/downloads/index.html">Oracle
JDK</a>
                      <p><a target="_blank" href="https://netbeans.org">Netbeans IDE</a>


                      <br/>

                      <br/>
                      Nombre oficial del proyecto: <strong> <a href="#"> Sistema de extracción de eventos y actores en textos
periodísticos utilizando técnicas de minera de textos</a></strong>
                      <br />
                      Año: <strong><a href="#"> Abril 2015</a></strong>
                      <br />
                      Alumno: <strong><a href="#"> Luis David Hernández Rojas</a></strong>
                      <br />
                      Asesora: <strong><a href="#"> Dra. Maricela Claudia Bravo Contreras</a></strong>
                      <br/>
                      Asesor: <strong><a href="#"> Dra. José Alejandro Reyes Órtiz</a></strong>
                      <br/>
                      <p><img src="img/logoarxfblanco.jpg" />

                    <br/><br/>
```

```
                        <button type="button" class="btn btn-default" data-dismiss="modal"><i class="fa fa-times"></i> Cerrar</button>
                    </div>
                </div>
            </div>
        </div>
    </div>

<!-- MINERIA DE DATOS  -->
    <div class="portfolio-modal modal fade" id="portfolioModal3" tabindex="-1" role="dialog" aria-hidden="true">
        <div class="modal-content">
            <div class="close-modal" data-dismiss="modal">
                <div class="lr">
                    <div class="rl">
                    </div>
                </div>
            </div>
            <div class="container">
                <div class="row">
                    <div class="col-lg-8 col-lg-offset-2">
                        <div class="text_help">
                            <h2>Minería de Datos</h2>
                            <hr /><br />
                            <img src="img/logomcfblanco.jpg" />
                            <p> La <a href="#">Minería de Datos</a> tiene como objetivo  extraer información de un conjunto de datos y transformarla en una estructura comprensible para su uso posterior. </p>
                            </br>
                            <p>La tarea de minería de datos real es el análisis automático o semi-automático de grandes cantidades de datos para extraer patrones interesantes hasta ahora desconocidos, como los grupos de registros de datos (análisis clúster), registros poco usuales (la detección de anomalías) y dependencias (minería por reglas de asociación). </p>
                        </br>
                            <p>Estos patrones pueden entonces ser vistos como una especie de resumen de los datos de entrada, y pueden ser utilizados en el análisis adicional o, por ejemplo, en la máquina de aprendizaje y análisis predictivo. Por ejemplo, el paso de minería de datos podría identificar varios grupos en los datos, que luego pueden ser utilizados para obtener resultados más precisos de predicción por un sistema de soporte de decisiones.</p>
                            <button type="button" class="btn btn-default" data-dismiss="modal"><i class="fa fa-times"></i> Cerrar</button>
                        </div>
                    </div>
                </div>
            </div>
        </div>
    </div>



<p>  <div class="container">
```

```
                    <div class="row">
                      <div class="col-lg-8 col-lg-offset-2">

<h2>comenzar</h2>

<h4>Recuerda cargar un corpus periodístico con extensión <b>".txt"</b>. Si tienes dudas consulta la pestaña <b>Guia de Uso</b></h4>

</p>
    <form action="procesado.jsp" enctype="MULTIPART/FORM-DATA" method="post">
       <pre> <input type="file" name="file" /><br/>
          <input type="submit" value="Comenzar Extraccion" />
       </pre>
      </form>

<form name="ejemplo" action="ejemplo.jsp"  method="post" >
    <pre>
            <input type="submit" value="Ver ejemplo"  />
    </pre>
   </form>
</div></div></div>
  </body>
</html>

<%-- procesado.jsp --%>

<%@page import="pt.Proceso"%>
<%@page contentType="text/html" pageEncoding="UTF-8"%>
<%@ page import="java.util.*" %>
<%@ page import="org.apache.commons.fileupload.*" %>
<%@ page import="org.apache.commons.fileupload.disk.*" %>
<%@ page import="org.apache.commons.fileupload.servlet.*" %>
<%@ page import="org.apache.commons.io.*" %>
<%@ page import="java.io.*" %>
<!DOCTYPE html>
<html>

    <head>

    <meta charset="utf-8">

    <!-- Use the .htaccess and remove these lines to avoid edge case issues.
           More info: h5bp.com/b/378 -->

    <title>Sistema de Extracción </title>
            <meta http-equiv="X-UA-Compatible" content="IE=edge" />
            <meta name="viewport" content="width=device-width, initial-scale=1.0, user-scalable=no" />
            <!-- Icono del navegador -->
```

```
        <link rel="shortcut icon" href="img/favicon.png"/>
        <!-- Bootstrap Core CSS -->
        <link href="css/bootstrap.css" rel="stylesheet" type="text/css" />
        <link href="css/freelancer.css" rel="stylesheet" type="text/css" />
        <!-- Fonts -->
        <link href="font-awesome/css/font-awesome.min.css" rel="stylesheet" type="text/css" />
        <link href='http://fonts.googleapis.com/css?family=Montserrat:400,700' rel='stylesheet' type='text/css' />
        <!-- Js -->
        <script src="js/jquery-1.10.2.js"></script>
        <script src="js/bootstrap.min.js"></script>
        <script src="http://cdnjs.cloudflare.com/ajax/libs/jquery-easing/1.3/jquery.easing.min.js"></script>
        <script src="js/classie.js"></script>
        <script src="js/cbpAnimatedHeader.js"></script>
        <script src="js/freelancer.js"></script>
    <style type="text/css">
        .titulo_contenido h1, h2, h3, h4, h5, h6{
            text-align: center;
        }
        #resultados{
            text-align: left;
        }
        .resultados ul{
            text-decoration: none;
        }
        .nav_cabecera a:hover{
            color: #fff;
        }
        .img_help{
            width: 250px;
        }
        .img_logo{
            width: 80px;
            margin-top: -10px;
            margin-right: 15px;
        }
        .align_item{
            text-align: left;
        }
        .text_help{
            text-align: justify;
        }
        th {
            background-color: #00003b;
            color: white;
            text-align:justify;
        }
        hr {
```

```
            display: block;
            margin-top: 0.5em;
            margin-bottom: 0.5em;
            margin-left: auto;
            margin-right: auto;
            border-style: inset;
            border-width: 1px;
          }
        </style>

    <!-- Encabezado -->

     <div class="container">
            <!-- Brand and toggle get grouped for better mobile display -->
            <div class="navbar-header page-scroll">
              <button type="button" class="navbar-toggle" data-toggle="collapse" data-target="#bs-example-navbar-collapse-1">
                <span class="sr-only">Toggle navigation</span>
                <span class="icon-bar"></span>
                <span class="icon-bar"></span>
                <span class="icon-bar"></span>
              </button>
              <a class="navbar-brand" href="#page-top" ><h3>SISTEMA DE EXTRACCIÓN WEB DE ACTORES Y EVENTOS EN TEXTOS
PERIODISTICOS</h3><br /></a>
            </div>
            <!-- Coleccion de nav links -->
            <br><br><br><br>
            <div class="collapse navbar-collapse" id="bs-example-navbar-collapse-1">
              <ul class="nav navbar-nav navbar-right">
                <li class="hidden">
                  <a href="#page-top"></a>
                </li>
                <li class="page-scroll">
                  <a href="#portfolioModal1" class="portfolio-link" data-toggle="modal">Guía</a>
                </li>
                <li class="page-scroll">
                  <a href="#portfolioModal2" class="portfolio-link" data-toggle="modal">Acerca de</a>
                </li>
                <li class="page-scroll">
                  <a href="#portfolioModal3" class="portfolio-link" data-toggle="modal">Minería de datos</a>
                </li>
              </ul>
            </div>
            <!-- /.navbar-collapse -->
          </div>
</head>
```

```
<BODY BACKGROUND="img/fondo.jpg">

   <!-- DESDE AQUI ESTA EL CUERPO -->

<!-- GUIA DE USO -->
            <div class="portfolio-modal modal fade" id="portfolioModal1" tabindex="-1" role="dialog" aria-hidden="true">
              <div class="modal-content">
                <div class="close-modal" data-dismiss="modal">
                  <div class="lr">
                    <div class="rl"></div>
                  </div>
                </div>
                <div class="container">
                  <div class="row">
                    <div class="col-lg-8 col-lg-offset-2">
                      <div class="modal-body">
                        <h2>Guía de uso</h2>
                        <hr /><br />
                        <img src="img/logomcfblanco.jpg" /><br/>
                        <div class="text_help">
                          <ul>
                            <li type="square">NOTAS IMPORTANTES
                              <ol>
                                <li>El corpus debe cargarse en archivo .txt
                                  <ol>
                                    <li><strong>Corpus de textos periodísticos: </strong> Se puede llamar corpus a cualquier colección que contenga más de un texto periodístico.</li>
                                    <li><strong>Lenguaje: </strong>Es importante que las noticias esten escritas en español.</li>
                                  </ol>

                                </li>
                              </ol>
                            </li>
                            <li type="square">CARGAR
                              <ol>
                                <li>Dar click en el botón <strong>Examinar</strong> para buscar el archivos a procesar. Este corpus será cargado temporalmente en el servidor con el fin de que el proceso se realice en el servidor.</li>
                                <li>Se abrirá un explorador de archivos, seleccione el archivo del corpus con extensión <strong>.txt</strong> y de clic en el botón <strong>Aceptar</strong>.</li>
                                <li>Si desea cancelar la subida de archivos, puede dar click en el botón <strong>Cancelar</strong> para eliminar los archivos seleccionados y/o limpiar la lista de archivos a subir.</li>
                                <li>Dar click en el botón <strong>Comenzar extraccion</strong> para comenzar el proceso</li>

                              </ol>
                            </li>
                            <br />
```

```
                        </ul>
                        <br />
                        <br />
                      </div>
                      <img src="img/logoarxfblanco.jpg" /><br/>
                      <button type="button" class="btn btn-default" data-dismiss="modal"><i class="fa fa-times"></i> Cerrar</button>
                    </div>
                  </div>
                </div>
              </div>
            </div>

<!-- ACERCA DE  -->
        <div class="portfolio-modal modal fade" id="portfolioModal2" tabindex="-1" role="dialog" aria-hidden="true">
          <div class="modal-content">
            <div class="close-modal" data-dismiss="modal">
              <div class="lr">
                <div class="rl">
                </div>
              </div>
            </div>
            <div class="container">
              <div class="row">
                <div class="col-lg-8 col-lg-offset-2">
                  <div class="text_help">
                    <h2>Acerca de</h2>
                    <hr /><br />
                    <img src="img/logomcfblanco.jpg" />
                    <p>El <a>Sistema de extracción de eventos y actores en textos periodísticos utilizando técnicas de minera de textos</a> fue desarrollado como Proyecto de Integración para la carrera de Ingeniería en computación.</p>
                    <p>El sistema fue desarrollado con las siguientes tecnologías:
                      <p><a target="_blank" href="http://tomcat.apache.org/">Apache Tomcat</a>
                      <p><a target="_blank" href="http://www.oracle.com/technetwork/java/javase/downloads/index.html">Oracle JDK</a>
                      <p><a target="_blank" href="https://netbeans.org">Netbeans IDE</a>


                      <br/>

                      <br/>
                      Nombre oficial del proyecto: <strong> <a href="#"> Sistema de extracción de eventos y actores en textos periodísticos utilizando técnicas de minera de textos</a></strong>
                      <br />
                      Año: <strong><a href="#"> Abril 2015</a></strong>
                      <br />
                      Alumno: <strong><a href="#"> Luis David Hernández Rojas</a></strong>
```

```
                        <br />
                        Asesora: <strong><a href="#"> Dra. Maricela Claudia Bravo Contreras</a></strong>
                        <br/>
                        Asesor: <strong><a href="#"> Dra. José Alejandro Reyes Órtiz</a></strong>
                        <br/>
                        <p><img src="img/logoarxfblanco.jpg" />

                      <br/><br/>


                        <button type="button" class="btn btn-default" data-dismiss="modal"><i class="fa fa-times"></i> Cerrar</button>
                    </div>
                  </div>
                </div>
              </div>
            </div>

<!-- MINERIA DE DATOS  -->
        <div class="portfolio-modal modal fade" id="portfolioModal3" tabindex="-1" role="dialog" aria-hidden="true">
          <div class="modal-content">
            <div class="close-modal" data-dismiss="modal">
              <div class="lr">
                <div class="rl">
                </div>
              </div>
            </div>
            <div class="container">
              <div class="row">
                <div class="col-lg-8 col-lg-offset-2">
                  <div class="text_help">
                    <h2>Minería de Datos</h2>
                    <hr /><br />
                    <img src="img/logomcfblanco.jpg" />
                    <p> La <a href="#">Minería de Datos</a> tiene como objetivo  extraer información de un conjunto de datos y transformarla en una estructura comprensible para su uso posterior. </p>
                    </br>
                    <p>La tarea de minería de datos real es el análisis automático o semi-automático de grandes cantidades de datos para extraer patrones interesantes hasta ahora desconocidos, como los grupos de registros de datos (análisis clúster), registros poco usuales (la detección de anomalías) y dependencias (minería por reglas de asociación). </p>
                    </br>
                    <p>Estos patrones pueden entonces ser vistos como una especie de resumen de los datos de entrada, y pueden ser utilizados en el análisis adicional o, por ejemplo, en la máquina de aprendizaje y análisis predictivo. Por ejemplo, el paso de minería de datos podría identificar varios grupos en los datos, que luego pueden ser utilizados para obtener resultados más precisos de predicción por un sistema de soporte de decisiones.</p>
                    <button type="button" class="btn btn-default" data-dismiss="modal"><i class="fa fa-times"></i> Cerrar</button>
                  </div>
```

```
                    </div>
                  </div>
                </div>
              </div>
            </div>



<p>   <div class="container">
                    <div class="row">
                      <div class="col-lg-8 col-lg-offset-2">

<h2>RESULTADOS</h2>


 <%
      /*FileItemFactory es una interfaz para crear FileItem*/
      FileItemFactory file_factory = new DiskFileItemFactory();

      /*ServletFileUpload esta clase convierte los input file a FileItem*/
      ServletFileUpload servlet_up = new ServletFileUpload(file_factory);
      /*sacando los FileItem del ServletFileUpload en una lista */
      List items = servlet_up.parseRequest(request);

      for(int i=0;i<items.size();i++){
         /*FileItem representa un archivo en memoria que puede ser pasado al disco duro*/
         FileItem item = (FileItem) items.get(i);
         /*item.isFormField() false=input file; true=text field*/
         if (! item.isFormField()){
            /*cual sera la ruta al archivo en el servidor*/
            File archivo_server = new File("C:/LDHR/temp.txt");
            /*y lo escribimos en el servido*/
            item.write(archivo_server);
            out.print("Se cargo el archivo con el nombre Nombre --> " + item.getName() );
            out.print("<br>");
         }
      }

Proceso comienza=new Proceso();
ArrayList <ArrayList <String>> recibido = new ArrayList <ArrayList <String>> () ;
recibido=comienza.inicio();

 ArrayList <String> notas= new ArrayList<String> ();
 ArrayList <String> actores = new ArrayList<String> ();
 ArrayList <String> eventos = new ArrayList<String> ();
```

```
 notas.addAll(recibido.get(0));
 actores.addAll(recibido.get(1));
 eventos.addAll(recibido.get(2));

 int tam=notas.size();
 out.println("<pre>" );
 out.println("Se encontraron un total de "+ tam +" notas. Mostrando resultados obtenidos... \n" );
 out.println("<br>" );
 out.println("</pre>" );
%>


<textarea name="res" cols="100" rows="10" readonly="readonly">
<% for(int i=0;i<=(tam-1);i++)
 {
 out.println(" ---- RESULTADO NUMERO " + (i+1) +" ---- \n");

 out.println(" Nota :  " + notas.get(i)+ "\n");

 out.println(" Actores : " + actores.get(i)+"\n");

 out.println(" Eventos: " + eventos.get(i)+ "\n");

 }
%>
 </textarea>

<form  action="index.jsp"  method="post" >
<pre>
<input type="submit" value="Ir a Inicio" />
</pre></form>
            </div></div></div>
   </body>
</html>

<%-- ejemplo.jsp --%>
<%@page import="java.util.ArrayList"%>
<%@page import="pt.ejemplo"%>
<%@page contentType="text/html" pageEncoding="UTF-8"%>

<!DOCTYPE html>
<html>

   <head>

   <meta charset="utf-8">
```

```
<title>Sistema de Extracción </title>
        <meta http-equiv="X-UA-Compatible" content="IE=edge" />
        <meta name="viewport" content="width=device-width, initial-scale=1.0, user-scalable=no" />
        <!-- Icono del navegador -->
        <link rel="shortcut icon" href="img/favicon.png"/>
        <!-- Bootstrap Core CSS -->
        <link href="css/bootstrap.css" rel="stylesheet" type="text/css" />
        <link href="css/freelancer.css" rel="stylesheet" type="text/css" />
        <!-- Fonts -->
        <link href="font-awesome/css/font-awesome.min.css" rel="stylesheet" type="text/css" />
        <link href='http://fonts.googleapis.com/css?family=Montserrat:400,700' rel='stylesheet' type='text/css' />
        <!-- Js -->
        <script src="js/jquery-1.10.2.js"></script>
        <script src="js/bootstrap.min.js"></script>
        <script src="http://cdnjs.cloudflare.com/ajax/libs/jquery-easing/1.3/jquery.easing.min.js"></script>
        <script src="js/classie.js"></script>
        <script src="js/cbpAnimatedHeader.js"></script>
        <script src="js/freelancer.js"></script>
    <style type="text/css">
        .titulo_contenido h1, h2, h3, h4, h5, h6{
            text-align: center;
        }
        #resultados{
            text-align: left;
        }
        .resultados ul{
            text-decoration: none;
        }
        .nav_cabecera a:hover{
            color: #fff;
        }
        .img_help{
            width: 250px;
        }
        .img_logo{
            width: 80px;
            margin-top: -10px;
            margin-right: 15px;
        }
        .align_item{
            text-align: left;
        }
        .text_help{
            text-align: justify;
        }
        th {
            background-color: #00003b;
```

```
            color: white;
            text-align:justify;
        }
        hr {
            display: block;
            margin-top: 0.5em;
            margin-bottom: 0.5em;
            margin-left: auto;
            margin-right: auto;
            border-style: inset;
            border-width: 1px;
        }
    </style>

<!-- Encabezado -->

<div class="container">
        <!-- Brand and toggle get grouped for better mobile display -->
        <div class="navbar-header page-scroll">
            <button type="button" class="navbar-toggle" data-toggle="collapse" data-target="#bs-example-navbar-collapse-1">
                <span class="sr-only">Toggle navigation</span>
                <span class="icon-bar"></span>
                <span class="icon-bar"></span>
                <span class="icon-bar"></span>
            </button>
            <a class="navbar-brand" href="#page-top" ><h3>SISTEMA DE EXTRACCIÓN WEB DE ACTORES Y EVENTOS EN TEXTOS
PERIODISTICOS</h3><br /></a>
        </div>
        <!-- Coleccion de nav links -->
        <br><br><br><br>
        <div class="collapse navbar-collapse" id="bs-example-navbar-collapse-1">
            <ul class="nav navbar-nav navbar-right">
                <li class="hidden">
                    <a href="#page-top"></a>
                </li>
                <li class="page-scroll">
                    <a href="#portfolioModal1" class="portfolio-link" data-toggle="modal">Guía</a>
                </li>
                <li class="page-scroll">
                    <a href="#portfolioModal2" class="portfolio-link" data-toggle="modal">Acerca de</a>
                </li>
                <li class="page-scroll">
                    <a href="#portfolioModal3" class="portfolio-link" data-toggle="modal">Minería de datos</a>
                </li>
            </ul>
        </div>
        <!-- /.navbar-collapse -->
```

```
            </div>
</head>


<BODY BACKGROUND="img/fondo.jpg">

   <!-- DESDE AQUI ESTA EL CUERPO -->

<!-- GUIA DE USO -->
            <div class="portfolio-modal modal fade" id="portfolioModal1" tabindex="-1" role="dialog" aria-hidden="true">
               <div class="modal-content">
                  <div class="close-modal" data-dismiss="modal">
                     <div class="lr">
                        <div class="rl"></div>
                     </div>
                  </div>
                  <div class="container">
                     <div class="row">
                        <div class="col-lg-8 col-lg-offset-2">
                           <div class="modal-body">
                              <h2>Guía de uso</h2>
                              <hr /><br />
                              <img src="img/logomcfblanco.jpg" /><br/>
                              <div class="text_help">
                                 <ul>
                                    <li type="square">NOTAS IMPORTANTES
                                       <ol>
                                          <li>El corpus debe cargarse en archivo .txt
                                             <ol>
                                                <li><strong>Corpus de textos periodísticos: </strong> Se puede llamar corpus a cualquier colección que contenga más de un texto periodístico.</li>
                                                <li><strong>Lenguaje: </strong>Es importante que las noticias esten escritas en español.</li>
                                                 </ol>

                                          </li>
                                       </ol>
                                    </li>
                                    <li type="square">CARGAR
                                       <ol>
                                          <li>Dar click en el botón <strong>Examinar</strong> para buscar el archivos a procesar. Este corpus será cargado temporalmente en el servidor con el fin de que el proceso se realice en el servidor.</li>
                                          <li>Se abrirá un explorador de archivos, seleccione el archivo del corpus con extensión <strong>.txt</strong> y de clic en el botón <strong>Aceptar</strong>.</li>
                                          <li>Si desea cancelar la subida de archivos, puede dar click en el botón <strong>Cancelar</strong> para eliminar los archivos seleccionados y/o limpiar la lista de archivos a subir.</li>
                                          <li>Dar click en el botón <strong>Comenzar extraccion</strong> para comenzar el proceso</li>
```

```
                                </ol>
                            </li>
                            <br />

                        </ul>
                        <br />
                        <br />
                    </div>
                    <img src="img/logoarxfblanco.jpg" /><br/>
                    <button type="button" class="btn btn-default" data-dismiss="modal"><i class="fa fa-times"></i> Cerrar</button>
                </div>
            </div>
        </div>
    </div>
</div>

<!-- ACERCA DE  -->
    <div class="portfolio-modal modal fade" id="portfolioModal2" tabindex="-1" role="dialog" aria-hidden="true">
        <div class="modal-content">
            <div class="close-modal" data-dismiss="modal">
                <div class="lr">
                    <div class="rl">
                    </div>
                </div>
            </div>
            <div class="container">
                <div class="row">
                    <div class="col-lg-8 col-lg-offset-2">
                        <div class="text_help">
                            <h2>Acerca de</h2>
                            <hr /><br />
                            <img src="img/logomcfblanco.jpg" />
                            <p>El <a>Sistema de extracción de eventos y actores en textos periodísticos utilizando técnicas de minera de textos</a> fue desarrollado como Proyecto de Integración para la carrera de Ingeniería en computación.</p>
                            <p>El sistema fue desarrollado con las siguientes tecnologías:
                                <p><a target="_blank" href="http://tomcat.apache.org/">Apache Tomcat</a>
                                <p><a target="_blank" href="http://www.oracle.com/technetwork/java/javase/downloads/index.html">Oracle JDK</a>
                                <p><a target="_blank" href="https://netbeans.org">Netbeans IDE</a>


                                <br/>

                                <br/>
                                Nombre oficial del proyecto: <strong> <a href="#"> Sistema de extracción de eventos y actores en textos periodísticos utilizando técnicas de minera de textos</a></strong>
```

```
                        <br />
                        Año: <strong><a href="#"> Abril 2015</a></strong>
                        <br />
                        Alumno: <strong><a href="#"> Luis David Hernández Rojas</a></strong>
                        <br />
                        Asesora: <strong><a href="#"> Dra. Maricela Claudia Bravo Contreras</a></strong>
                        <br/>
                        Asesor: <strong><a href="#"> Dra. José Alejandro Reyes Órtiz</a></strong>
                        <br/>
                        <p><img src="img/logoarxfblanco.jpg" />

                      <br/><br/>


                        <button type="button" class="btn btn-default" data-dismiss="modal"><i class="fa fa-times"></i> Cerrar</button>
                    </div>
                  </div>
                </div>
              </div>
            </div>
          </div>

<!-- MINERIA DE DATOS  -->
          <div class="portfolio-modal modal fade" id="portfolioModal3" tabindex="-1" role="dialog" aria-hidden="true">
            <div class="modal-content">
              <div class="close-modal" data-dismiss="modal">
                <div class="lr">
                  <div class="rl">
                  </div>
                </div>
              </div>
              <div class="container">
                <div class="row">
                  <div class="col-lg-8 col-lg-offset-2">
                    <div class="text_help">
                      <h2>Minería de Datos</h2>
                      <hr /><br />
                      <img src="img/logomcfblanco.jpg" />
                      <p> La <a href="#">Minería de Datos</a> tiene como objetivo  extraer información de un conjunto de datos y transformarla en una estructura comprensible para su uso posterior. </p>
                      </br>
                      <p>La tarea de minería de datos real es el análisis automático o semi-automático de grandes cantidades de datos para extraer patrones interesantes hasta ahora desconocidos, como los grupos de registros de datos (análisis clúster), registros poco usuales (la detección de anomalías) y dependencias (minería por reglas de asociación). </p>
                    </br>
                      <p>Estos patrones pueden entonces ser vistos como una especie de resumen de los datos de entrada, y pueden ser utilizados en el análisis adicional o, por ejemplo, en la máquina de aprendizaje y análisis predictivo. Por ejemplo, el paso de minería de
```

datos podría identificar varios grupos en los datos, que luego pueden ser utilizados para obtener resultados más precisos de predicción por un sistema de soporte de decisiones.</p>

```
                        <button type="button" class="btn btn-default" data-dismiss="modal"><i class="fa fa-times"></i> Cerrar</button>
                    </div>
                </div>
            </div>
        </div>
    </div>
</div>
```

```
<p>   <div class="container">
            <div class="row">
                <div class="col-lg-8 col-lg-offset-2">

<h2>Ejemplo</h2>
<h4>Mostrando resultados</h4>
 <br></p><br>

 <%
   ejemplo comienza=new ejemplo();
   ArrayList <ArrayList <String>> recibido = new ArrayList <ArrayList <String>> () ;
   recibido=comienza.inicio();


  ArrayList <String> notas= new ArrayList<String> ();
  ArrayList <String> actores = new ArrayList<String> ();
  ArrayList <String> eventos = new ArrayList<String> ();


  notas.addAll(recibido.get(0));
  actores.addAll(recibido.get(1));
  eventos.addAll(recibido.get(2));

  int tam=notas.size();
  out.println("<pre>" );
  out.println("Se encontraron un total de "+ tam +" notas. Mostrando resultados obtenidos... \n" );
  out.println("<br>" );
  out.println("</pre>" );
 %>

<textarea name="res" cols="100" rows="10" readonly="readonly">
<% for(int i=0;i<=(tam-1);i++)
  {
 out.println(" ---- RESULTADO NUMERO " +(i+1)+" ---- \n");
```

```
 out.println(" Nota: " + notas.get(i)+ "\n");

 out.println(" Actores : " + actores.get(i)+"\n");

 out.println(" Eventos: " + eventos.get(i)+ "\n");

  }
%>

</textarea>

 <form  action="index.jsp"  method="post" >
<pre>
<input type="submit" value="Ir a Inicio" />
</pre></form>
  </div></div></div>
  </body>
</html>
```