



Conference: Congreso Interdisciplinario de Energías Renovables -
Mantenimiento Industrial - Mecatrónica e Informática

Booklets



RENIECYT

Registro Nacional de Instituciones
y Empresas Científicas y Tecnológicas

2015-20795

CONACYT

RENIECYT - LATINDEX - Research Gate - DULCINEA - CLASE - Sudoc - HISPANA - SHERPA UNIVERSIA - E-Revistas - Google Scholar
DOI - REBID - Mendeley - DIALNET - ROAD - ORCID

Title: Análisis de datos utilizando web scraping para repertorio otomí educativo en dispositivos móviles android

Author: Erika López-González

Editorial label ECORFAN: 607-8324
BCIERMIMI Control Number: 2016-01
BCIERMIMI Classification(2016): 191016-0101

Pages: 14

Mail: Lz_erika_gz@yahoo.com.mx
RNA: 03-2010-032610115700-14

ECORFAN-México, S.C.

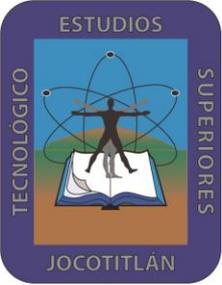
244 – 2 Itzopan Street
La Florida, Ecatepec Municipality
Mexico State, 55120 Zipcode
Phone: +52 1 55 6159 2296
Skype: ecorfan-mexico.s.c.
E-mail: contacto@ecorfan.org
Facebook: ECORFAN-México S. C.

Twitter: @EcorfanC

www.ecorfan.org

Holdings

Bolivia	Honduras	China	Nicaragua
Cameroon	Guatemala	France	Republic of the Congo
El Salvador	Colombia	Ecuador	Dominica
Peru	Spain	Cuba	Haití
Argentina	Paraguay	Costa Rica	Venezuela
Czech Republic			



Contenido

- ✓ Planteamiento del Problema
- ✓ Objetivos
- ✓ Antecedentes de la Investigación
- ✓ Marco Metodológico
- ✓ Pruebas y Resultados
- ✓ Conclusiones



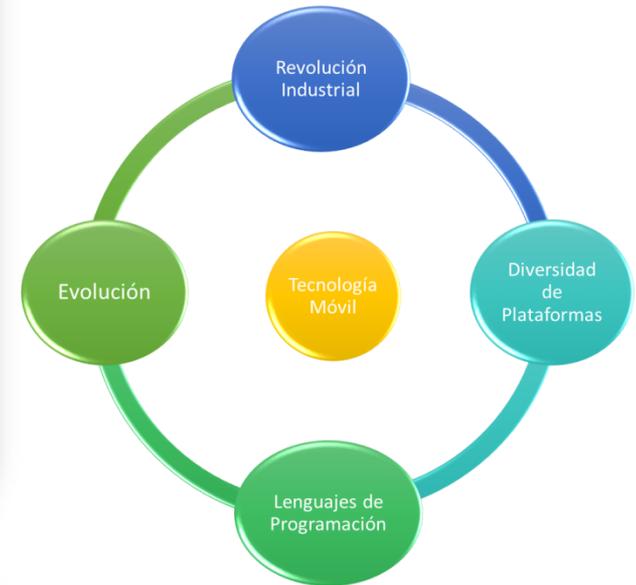
San Juan del Río, Qro. 19 al 21 de Octubre del 2016.

**Congreso Interdisciplinario de Energías Renovables,
Mantenimiento Industrial, Mecatrónica e Informática**



Planteamiento del problema

Integrar tecnología móvil con mejor funcionamiento, diseño, a su vez sea llamativa, práctica y de fácil uso para que pueda traducir palabras del español al Otomí y del Otomí al español ayudará a las personas a cultivar la fluidez y perseverancia del lenguaje para no debilitar su aprendizaje.



San Juan del Río, Qro. 19 al 21 de Octubre del 2016.

Objetivos

Objetivo General

Desarrollar una aplicación híbrida para dispositivos Android utilizando Apache Cordova como plataforma de desarrollo, aplicando Web Scraping para la recolección de información generando un diccionario español - otomí, otomí - español.

Objetivos Específicos

- ✓ Conocer, estudiar y buscar sitios con temática otomí.
- ✓ Analizar los sitios web de temática otomí para recolectar vocabulario aplicando scraping.
- ✓ Diseñar un algoritmo para aplicar el scraping en sitios con temática otomí.
- ✓ Crear una base de datos para almacenar el vocabulario obtenido.
- ✓ Generar una interfaz visualmente atractiva y fácil de utilizar.
- ✓ Implementar la aplicación para verificar el funcionamiento de la misma.

San Juan del Río, Qro. 19 al 21 de Octubre del 2016.

Antecedentes

Ewald Hekking

Diseño un programa por internet y en multimedia, para la enseñanza del lenguaje otomí que incluso cuenta con aplicaciones para telefonía móvil



Pedro D. Cardona Fuentes "YAAK"

Plataforma en línea para la enseñanza-aprendizaje y difusión de lenguas indígenas en México

San Juan del Río, Qro. 19 al 21 de Octubre del 2016.

Metodología

Diseñar una técnica para aplicar el scraping en sitios con temática otomí.

Extracción de datos con scraping

Crear una base de datos para almacenar el vocabulario obtenido.

Generar una interfaz visualmente atractiva y fácil de utilizar.

Implementar la aplicación para verificar el funcionamiento.

San Juan del Río, Gro. 19 al 21 de Octubre del 2016.

Técnica Scraping

$$\text{VAR} = [[a-zA-Z]^+ [a-z]^* [\lambda]]^+$$

$$\text{VAR}^3 [:] \text{VAR}^5$$

Lista de tareas

Búsqueda de sitios web con temática otomí.

Trabajar con URL'S para recolecta de código HTML "Contenido en sitios Web" mediante urllib2.0

```
var pal = [
    "abajo: ngati",
    "abandonado: xotsogi",
    "abandonar: tsogi, hëpü, hëgi",
    "abanicar: ts'üdi",
    "abanico: fuki, nthiti",
    "abaratar: k'ami",
    "abdicar: hiëgi",
    "abdomen: debi",
]
```

San Juan del Río, Qro. 19 al 21 de Octubre del 2016.

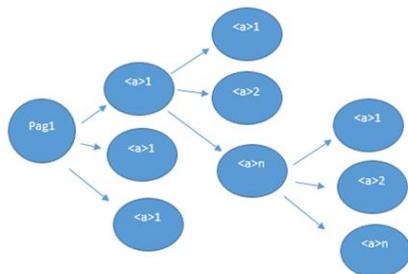
Extracción de datos

Lista de tareas

Analizar los sitios web de temática otomí para recolectar vocabulario aplicando scraping.

Aplicar screen- scraping mediante BeautifulSoup para el análisis del contenido del sitio Web.

Encontrando 52 para completar la coleccion 4541 palabras



```

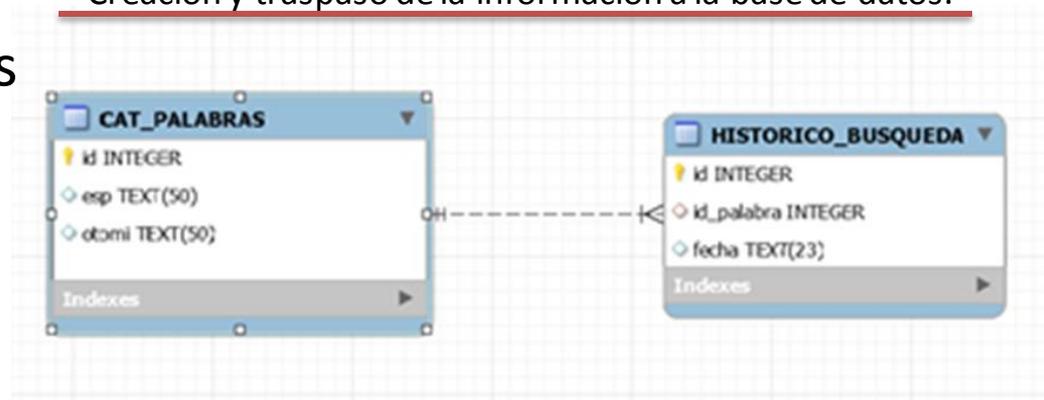
web.py  vocabulario.txt
5871 zo: aunque
5872 zobí: pelearon
5873 zobú: fermentar
5874 zodú: descomponerse
5875 zodye: enfermó
5876 zogú: dejar
5877 zogúbi: abandonaron
5878 zogúji: dejaron
5879 zogútsúji: te dejaron
5880 zoji: léanlo, háblales, probaron
5881 zokú: me habla
5882 zokúji: nos dejaron, nos hablaron
5883 zona, zonú: reflexiona
5884 zoo: despertó, sintió, bonito, probar
5885 zopjú: lee, háblale
5886 zopjúji: háblenle, lean
5887 zosú: caer algo en el techo
5888 zotjo, zotrjo: bonito
5889 zotsú: bella
5890 zoya: anciano
5891 zúji: tuvieron miedo
5892 zújmú: quedarse en un lugar para vivir
5893 zújnú: águila
5894 zúkúzúji: nos alcanzaron
5895 zúmbúru: retumbar
5896 Zúmi: Toluca
5897 zúnrrajma: remolino
5898 zúntú: sonar, repicar
5899 zúnú: suena
5900 zúru: agárrate
5901 zúruji: sosténganse, agárrense
5902 zúsú: chupar
5903 zútpú: chupar, besar
5904 zúú: mascar caña
5905 züü: miedo
    
```

San Juan del Río, Qro. 19 al 21 de Octubre del 2016.

Trabajo de base de datos

Lista de Tareas

- Diseño de la Base de datos
- Creación y traspaso de la información a la base de datos.



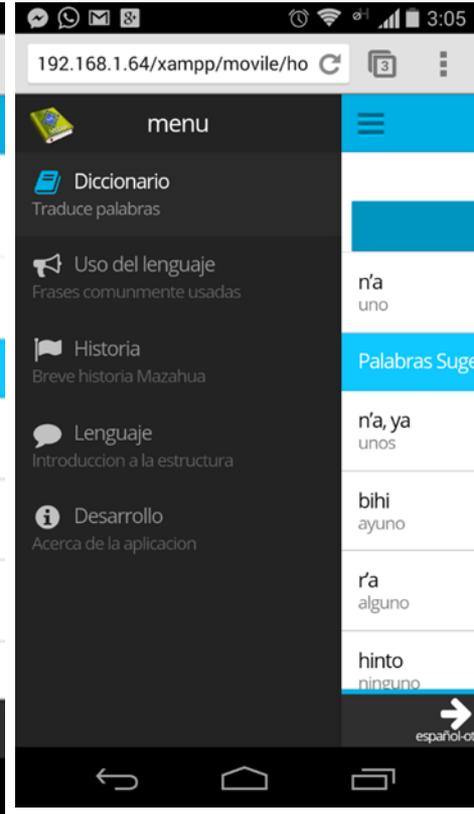
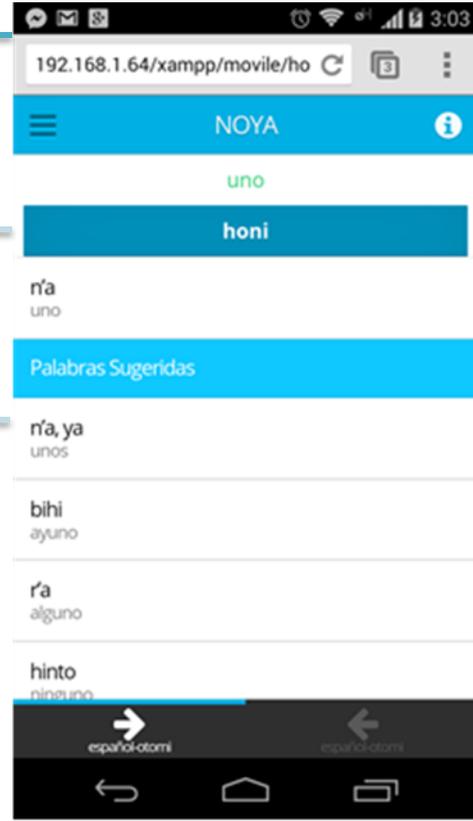
```
function c2(tx,rs){
  console.log(rs.rows.item(0));
}
function consulta(){
  db.transaction(function(tx){
    tx.executeSql('SELECT COUNT(*) as "Numero de Palabras" FROM palabras;',[],c2);
  });
}
```

Carga de 4541 palabras

San Juan del Río, Qro. 19 al 21 de Octubre del 2016.

Generación de la intefaz

- Lista de Tareas
- Diseño y creación de la interfaz mediante el framework Lungo.js
- Programación lógica para las interacciones de usuario.



```

<script>
  Lungo.init({
    name: "NOYA",
    version: "0.0.1"
  });
</script>

```

San Juan del Río, Qro. 19 al 21 de Octubre del 2016.

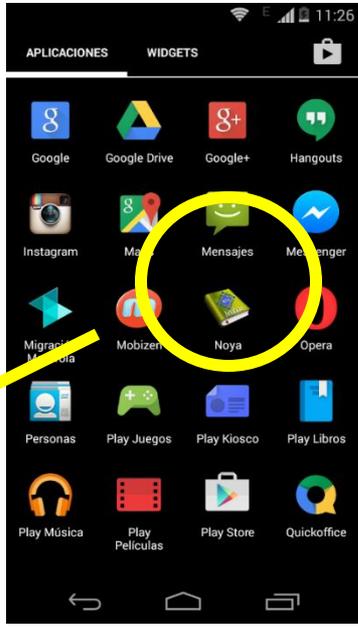
Generación de la interfaz

- Lista de Tareas**
- Documentación e instalación de componentes previos.
 - Instalación de Apache Cordova
 - Creación del proyecto con Apache Cordova

```
C:\Users\David>cd c:\Desarrollo
c:\Desarrollo>cordova create Noya com.example.noya "NOYA"
Creating a new cordova project.
c:\Desarrollo>_
```

```
C:\Users\David>cd c:\Desarrollo
c:\Desarrollo>cordova create Noya com.example.noya "NOYA"
Creating a new cordova project.
c:\Desarrollo>cd Noya
c:\Desarrollo\Noya>cordova platform add android
Adding android project...
Creating Cordova project for the Android platform:
```

```
BUILD SUCCESSFUL
Total time: 2 mins 7.837 secs
Built the following apk(s):
  c:\Desarrollo\Noya\platforms\android\build\outputs\apk\android-debug.apk
c:\Desarrollo\Noya>_
```



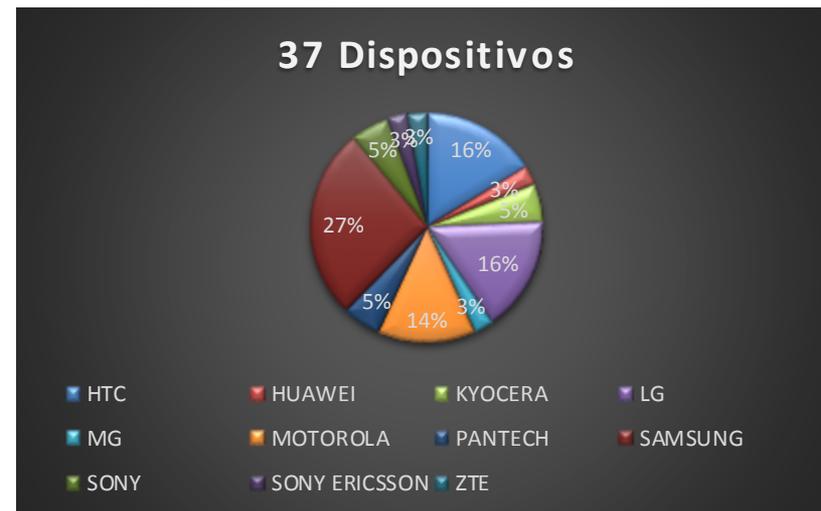
Apk Instalado

Nombre	Fecha de modifica...	Tipo	Tamaño
android-debug.apk	26/05/2015 10:58 ...	Archivo APK	2,151 KB
android-debug-unaligned.apk	26/05/2015 10:58 ...	Archivo APK	2,151 KB
manifest-merger-debug-report		ex...	3 KB

Tipo: Archivo APK
Tamaño: 2.09 MB
Fecha de modificación: 26/05/2015 10:58 p.m.

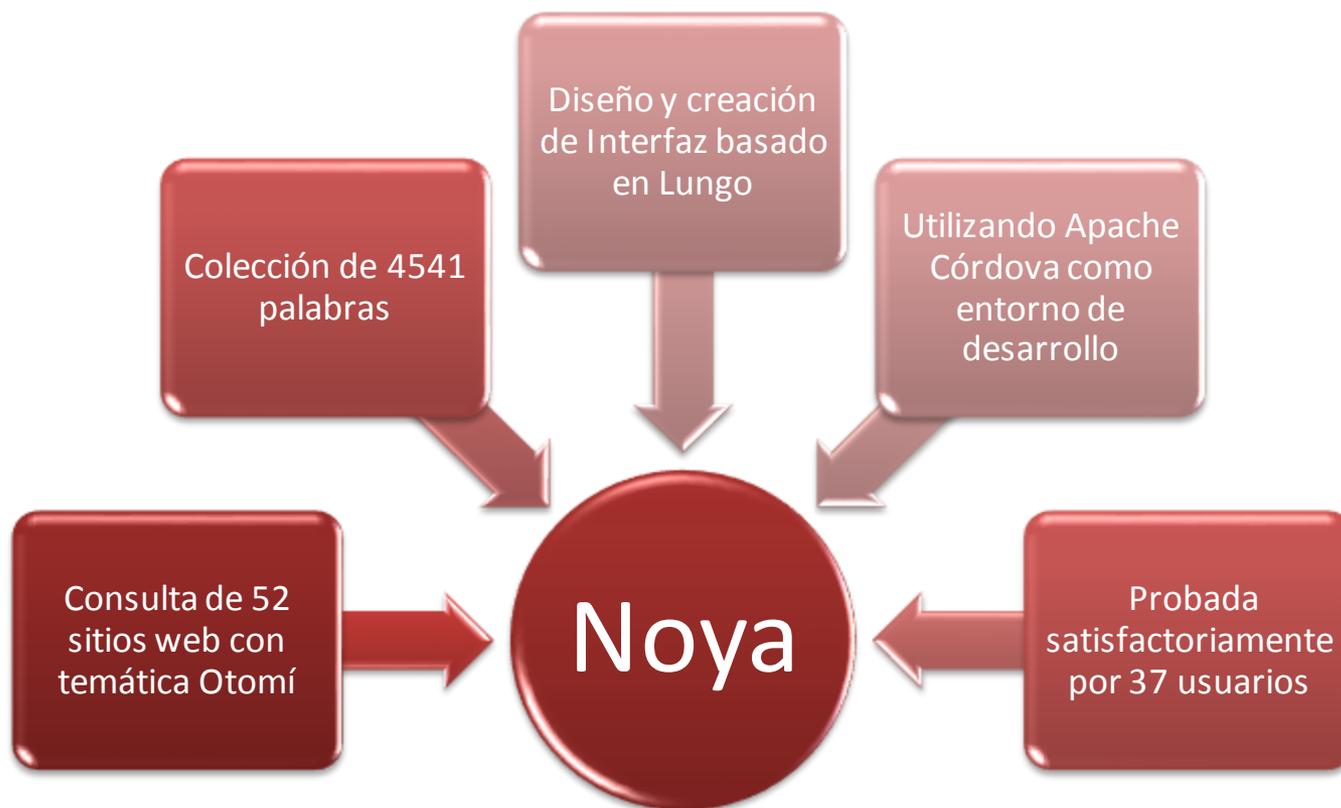
San Juan del Río, Qro. 19 al 21 de Octubre del 2016.

Pruebas y resultados



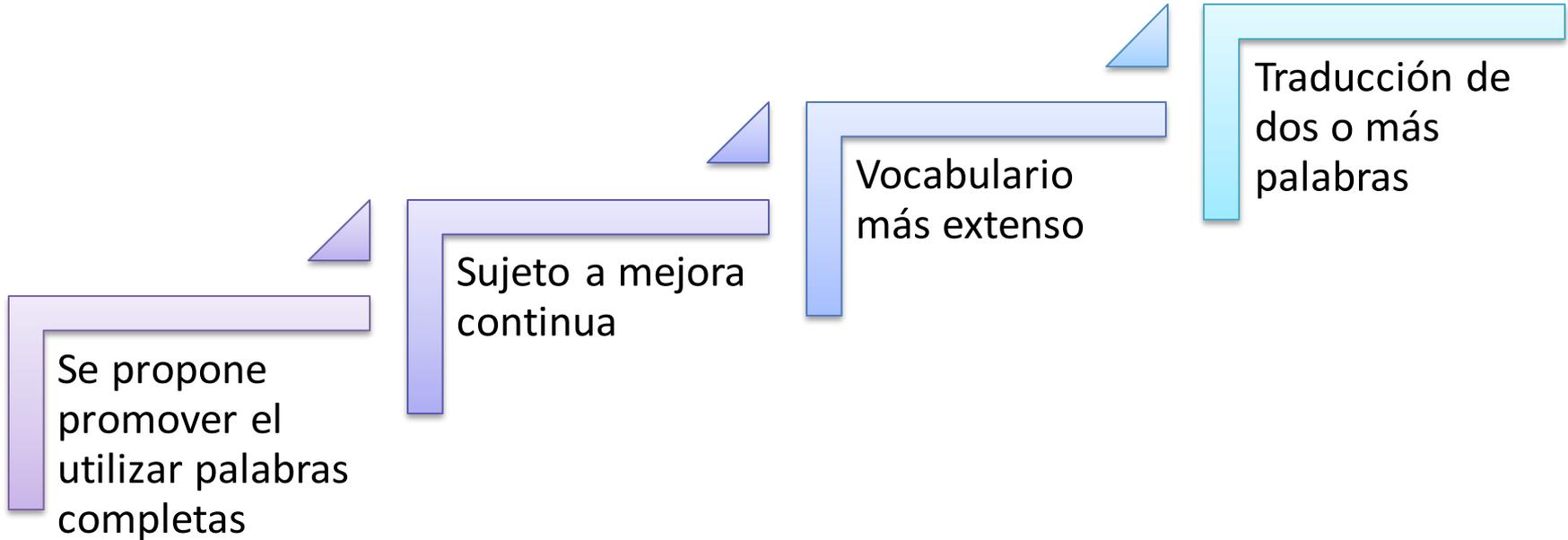
San Juan del Río, Qro. 19 al 21 de Octubre del 2016.

Conclusiones

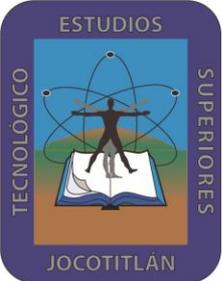


San Juan del Río, Gro. 19 al 21 de Octubre del 2016.

Recomendaciones



San Juan del Río, Qro. 19 al 21 de Octubre del 2016.



Referencias

Hekking, E. (24 de mayo de 2010). El universal .Recuperado el 17 de 02 de 2015, de <http://www.eluniversal.com.mx/articulos/58766.html>

INEGI. (2010). Censo de Población y Vivienda 2010,. Mexico: Instituto Nacional de Estadística.

Montoya-Casasola, M. Á., & Sandoval-Forero, E. A. (2013). Marginación sociodemográfica de los otomíes del Estado de México. Papeles de Población , 257-289.

Questa, R. A. (2006). Oyomies al norte del Estado de México y sur de Queretaro. México: Comision Nacional para el Desarrollo de los pueblos Indigenas.

Solis, A. (19 de 04 de 2015). FORBES. Recuperado el 19 de 04 de 2015, de <http://www.forbes.com.mx/las-15-apps-mas-utilizadas-del-mundo/>

San Juan del Río, Qro. 19 al 21 de Octubre del 2016.



**Congreso Interdisciplinario de Energías Renovables,
Mantenimiento Industrial, Mecatrónica e Informática** 2016



ECORFAN®

© ECORFAN-Mexico, S.C.

No part of this document covered by the Federal Copyright Law may be reproduced, transmitted or used in any form or medium, whether graphic, electronic or mechanical, including but not limited to the following: Citations in articles and comments Bibliographical, compilation of radio or electronic journalistic data. For the effects of articles 13, 162,163 fraction I, 164 fraction I, 168, 169,209 fraction III and other relative of the Federal Law of Copyright. Violations: Be forced to prosecute under Mexican copyright law. The use of general descriptive names, registered names, trademarks, in this publication do not imply, uniformly in the absence of a specific statement, that such names are exempt from the relevant protector in laws and regulations of Mexico and therefore free for General use of the international scientific community. BCIERMIMI is part of the media of ECORFAN-Mexico, S.C., E: 94-443.F: 008- (www.ecorfan.org/ booklets)