

Tecnológico de Costa Rica

Escuela de Computación

Tarea Programada 2

Documentación

IC-8020 - Recuperación de Información Textual

Profesor:

José Enrique Araya

Elaborado por:

José Ricardo Brenes Camacho - 201236179

Kevin Alonso Escobar Miranda - 201265615

31 de Octubre de 2015

1. Introducción

El presente proyecto es el segundo proyecto del curso IC-8020 Recuperación de la Información Textual que tiene como objetivo de realizar búsquedas en un conjunto de archivos en un directorio , según el diseño e implementación de los autores.

El funcionamiento posee características como el uso del modelo vectorial y archivos invertidos para representar y procesar la información de los documentos y las consultas ingresadas por el usuario. En este caso la colección de documentos se relaciona al tema de Geografía, la información del directorio y cada archivo que contiene es provista por el profesor y esta consiste en un conjunto de archivos HTML que contienen la descripción de los diferentes países y dependencias del mundo.

A continuación se presenta las pruebas y evidencias del funcionamiento del proyecto, comparado con los resultados esperados según la guía de documentación.

2. Pruebas

Etapas	% de complet.	Comentario o aclaración
Indexa toda la colección	100%	Se cumple con la funcionalidad
Transformaciones implementadas	100%	Se cumple con la funcionalidad
Separación en palabras	100%	
Conversión a minúsculas	100%	
Extracción de raíces (dar referencia herramienta usada)	100%	Snowball - http://snowball.tartarus.org/
Eliminación de acentos, preservación eñe	100%	
Eliminación de stopwords	100%	
Indexa números	100%	
Estructura archivo invertido	100%	Se cumple con la funcionalidad
Archivo documentos (dar # total docs.)	100%	266 documentos
Diccionario (dar # total términos)	100%	67377 términos

Postings (tamaño fijo, dar # total de entradas)	100%	465566 entradas
Implementa consultas básicas	100%	Se cumple con la funcionalidad
Incluye pesos 123 en consultas	100%	
Crea archivo escalafón correctamente	100%	
Implementa consultas avanzadas	100%	Se cumple con la funcionalidad
Procesa expresión regular	100%	
Crea archivo escalafón correctamente	100%	
Documento salida	100%	Se cumple con la funcionalidad
Genera archivo HTML de salida	100%	
Incluye: fecha y hora; rango, ruta colección, texto consulta vectorial, texto consulta avanzada, primeros 200 caracteres.	100%	
Interfaz Indicar estado de la implementación de los siguientes aspectos:	100%	Se cumple con la funcionalidad
Indexar una colección	100%	
Realizar consulta vectorial	100%	
Refinar consulta	100%	
Mostrar resultados	100%	

3. Completar la siguiente tabla para describir el resultado obtenido para algunas consultas de prueba.

Se deben adjuntar los archivos generados por sus pruebas.

En la sección Pruebas se detallan las operaciones que se deben hacer.

Pruebas	Resultados esperados	Comentarios sobre resultado obtenido
Prueba 1		
a) río	178 archivos (Prueba 1 - río.txt)	140 archivos
b) nilo	12 archivos (Prueba 1 – nilo.txt)	13 archivos
c) río 1; nilo 3	Los 12 archivos que tienen río y nilo deben aparecer mejor en el escalafón (Prueba 1 – río AND nilo.txt)	144 archivos dentro de los cuales si se encuentran los archivos de río y nilo
d) río +nilo	Solo deben recuperarse los 3 archivo indicados (Prueba 1 – refinación – río nilo.txt)	3 archivos
Prueba 2		
a) fiebre	15 archivos (Prueba 2 – fiebre.txt)	14 archivos
b) amarilla	14 archivos (Prueba 2 – amarilla.txt)	37 archivos
c) fiebre; amarilla	Los 6 archivos que tienen río y nilo deben aparecer mejor en el escalafón (Prueba 2 – fiebra AND amarilla.txt)	50 archivos
d) fiebre +amarilla	Solo deben recuperarse los 6 archivo indicados (Prueba 2 - refinación - fiebre amarilla.txt)	6 archivos los mismos documentos indicados
Prueba 3		
a) liberia 1	13 archivos (Prueba 3 – liberia.txt) Costa Rica probablemente quede de tercero o más abajo debido a la mayor frecuencia en Liberia y Sierra Leona.	12 archivos
b) nicoya 2	2 archivos (Prueba 3 – nicoya.txt) Costa Rica debe aparecer de primero en el escalafón.	2 archivos
c) liberia 1; nicoya 2	Los archivos de las dos pruebas anteriores. Costa Rica debe estar de primera en el escalafón: tiene ambos términos y especialmente el segundo que es más importante en la consulta. Nicaragua debe estar alto en el escalafón debido a que tiene varias veces el término nicoya que es el más importante.	Costa Rica se obtuvo en la posición 12 y el país Liberia en la posición 1
Prueba 4		
a) 3.000.000 1; 4.000.000 3	7 archivos (Prueba 4 – 3M 4M.txt) Se espera que Etiopía aparezca de primero en el escalafón ya que es el único que posee 3M y 4M. Los que solo tienen 4M (Afganistán, Laos, Nueva Zelanda) deben aparecer mejor que los	5 archivos

	que solo tienen 3M (Siria, Egipto, Sudán del Sur). Aunque Siria puede subir un poco porque tiene 3M dos veces.	
--	--	--

PRUEBAS

Los resultados que aquí se muestran se obtuvieron usando la facilidad de "Find in files" de Notepad++. Se utilizaron expresiones regulares en búsquedas que incluían recorrer los subdirectorios. ***Debido a limitaciones como no poder omitir el texto en las etiquetas HTML, los resultados presentados son una buena aproximación, pero en algunos casos podrían aparecer documentos escogidos por texto en etiquetas HTML que su proyecto debe omitir.***

Llamar a los archivos generados con el nombre y parte de prueba. Por ejemplo **P1a.esc** para el escalafón resultante de correr la consulta en Prueba 1 (a), **P2b.html** para la salida de la consulta en Prueba 2 (b). Puede agregar información extra en los nombres de los archivos para hacerlos más comprensibles:

P1a-rios.esc, P2b-amarilla.html.

Generar los archivos HTML para las primeras 20 posiciones del escalafón.

Prueba 0

- a) Indexar colección Geografía.
- b) Anotar:
 - número de documentos procesados
265
 - número de términos (palabras) en el diccionario
53107
 - número de términos (números) en el diccionario
14270
 - número de entradas en el archivo de postings
465566

Prueba 1

- a) Correr consulta vectorial:
 - río**
 - Se espera que recuperen 178 archivos listados en "Prueba 1 - río.txt".
 - Muchos de esos archivos tienen el plural "ríos".
- b) Correr la consulta vectorial:

nilo

Se espera que recuperen los 12 archivos siguientes:

```
Search "<nilo>" (108 hits in 12 files)
Geografia\América\Estados_soberanos\Bolivia.htm (2 hits)
Geografia\Asia\Estados_soberanos\Israel.htm (2 hits)
Geografia\Asia\Estados_soberanos\Jordania.htm (3 hits)
Geografia\África\Estados_soberanos\Burundi.htm (3 hits)
Geografia\África\Estados_soberanos\Egipto.htm (55 hits)
Geografia\África\Estados_soberanos\Eritrea.htm (3 hits)
Geografia\África\Estados_soberanos\Etiopía.htm (3 hits)
Geografia\África\Estados_soberanos\Kenia.htm (1 hit)
Geografia\África\Estados_soberanos\Nigeria.htm (2 hits)
Geografia\África\Estados_soberanos\Níger.htm (3 hits)
Geografia\África\Estados_soberanos\Sudán.htm (19 hits)
Geografia\África\Estados_soberanos\Sudán_del_Sur.htm (12 hits)
```

El archivo "Prueba 1 - nilo.txt" se detalla el resultado.

Se espera que Egipto, Sudán y Sudán del sur aparezcan arriba en el escalafón debido a su mayor frecuencia. Pero si los documentos son grandes puede que no.

c) Correr la consulta vectorial:

río 1; nilo 3

Van a salir todos los documentos de río junto con los de nilo. Pero **los siguientes casos deben aparecer muy arriba en el escalafón ya que tienen ambos:**

```
Search "<ríos?>.*<nilo>|<nilo>.*<ríos?>" (12 hits in 12 files)
Geografia\América\Estados_soberanos\Bolivia.htm (1 hit)
Geografia\Asia\Estados_soberanos\Israel.htm (1 hit)
Geografia\Asia\Estados_soberanos\Jordania.htm (1 hit)
Geografia\África\Estados_soberanos\Burundi.htm (1 hit)
Geografia\África\Estados_soberanos\Egipto.htm (1 hit)
Geografia\África\Estados_soberanos\Eritrea.htm (1 hit)
Geografia\África\Estados_soberanos\Etiopía.htm (1 hit)
Geografia\África\Estados_soberanos\Kenia.htm (1 hit)
Geografia\África\Estados_soberanos\Nigeria.htm (1 hit)
Geografia\África\Estados_soberanos\Níger.htm (1 hit)
Geografia\África\Estados_soberanos\Sudán.htm (1 hit)
Geografia\África\Estados_soberanos\Sudán_del_Sur.htm (1 hit)
```

En el archivo "Prueba 1 - río AND nilo.txt" se detalla el resultado.

d) Refinar resultado de consulta (c) usando:

río +nilo

Se espera se recuperen únicamente los siguientes 3 archivos:

```
Search "río +nilo" (16 hits in 3 files)
Geografia\Asia\Estados_soberanos\Israel.htm (1 hit)
Geografia\África\Estados_soberanos\Egipto.htm (11 hits)
```

Geografia\África\Estados_soberanos\Sudán.htm (4 hits)

En el archivo "Resultado - refinación - río nilo.txt" se detalla el resultado.

Prueba 2

a) Correr la consulta vectorial

fiebre

Se espera recuperar los siguientes 15 archivos:

```
Search "<fiebres?>" (38 hits in 15 files)
Geografia\América\Dependencias\Aruba.htm (1 hit)
Geografia\América\Estados_soberanos\Canadá.htm (2 hits)
Geografia\América\Estados_soberanos\Cuba.htm (2 hits)
Geografia\América\Estados_soberanos\Estados_Unidos.htm (2 hits)
Geografia\América\Estados_soberanos\México.htm (1 hit)
Geografia\América\Estados_soberanos\Panamá.htm (3 hits)
Geografia\América\Estados_soberanos\República_Dominicana.htm (2
hits)
Geografia\América\Estados_soberanos\Venezuela.htm (2 hits)
Geografia\Asia\Estados_soberanos\Siria.htm (5 hits)
Geografia\Oceanía\Estados_soberanos\Australia.htm (5 hits)
Geografia\África\Estados_soberanos\Angola.htm (5 hits)
Geografia\África\Estados_soberanos\Ghana.htm (2 hits)
Geografia\África\Estados_soberanos\Malawi.htm (2 hits)
Geografia\África\Estados_soberanos\Sudáfrica.htm (2 hits)
Geografia\África\Estados_soberanos\Togo.htm (2 hits)
```

En el archivo "Prueba 2 - fiebre.txt" se detalla el resultado.

b) Correr la consulta vectorial

amarilla

Se espera recuperar los siguientes 14 archivos:

```
Search "<amarilla?>" (20 hits in 14 files)
Geografia\América\Estados_soberanos\Costa_Rica.htm (1 hit)
Geografia\América\Estados_soberanos\Cuba.htm (2 hits)
Geografia\América\Estados_soberanos\Panamá.htm (1 hit)
Geografia\América\Estados_soberanos\Perú.htm (2 hits)
Geografia\América\Estados_soberanos\República_Dominicana.htm (2
hits)
Geografia\América\Estados_soberanos\Uruguay.htm (1 hit)
Geografia\América\Estados_soberanos\Venezuela.htm (2 hits)
Geografia\Europa\Estados_soberanos\Andorra.htm (1 hit)
Geografia\Europa\Estados_soberanos\Estonia.htm (1 hit)
Geografia\Europa\Estados_soberanos\Portugal.htm (1 hit)
Geografia\Oceanía\Dependencias\Isla_Sala_y_Gómez.htm (1 hit)
Geografia\Oceanía\Dependencias\Pitcairn.htm (1 hit)
Geografia\África\Estados_soberanos\Ghana.htm (2 hits)
Geografia\África\Estados_soberanos\Togo.htm (2 hits)
```

En el archivo "Prueba 2 - amarilla.txt" se detalla el resultado.

c) Correr la consulta vectorial:

fiebre; amarilla

Se deben obtener todos los archivos de (a) y (b).

Los siguientes archivos deben estar arriba en el escalafón puesto que tienen ambos términos:

```
Search "<fiebres?>.*<amarillas?>|<amarillas?><fiebres?>"
(6 hits in 6 files)
Geografia\América\Estados soberanos\Cuba.htm (1 hit)
Geografia\América\Estados soberanos\Panamá.htm (1 hit)
Geografia\América\Estados soberanos\República Dominicana.htm (1 hit)
Geografia\América\Estados soberanos\Venezuela.htm (1 hit)
Geografia\África\Estados soberanos\Ghana.htm (1 hit)
Geografia\África\Estados soberanos\Togo.htm (1 hit)
```

En el archivo "Prueba 2 - fiebre AND amarilla.txt" se detalla el resultado.

Los siguientes archivos tienen fiebre pero no amarilla:

```
Geografia\América\Dependencias\Aruba.htm (1 hit)
Geografia\América\Estados soberanos\Canadá.htm (2 hits)
Geografia\América\Estados soberanos\Estados Unidos.htm (2 hits)
Geografia\América\Estados soberanos\México.htm (1 hit)
Geografia\Asia\Estados soberanos\Siria.htm (5 hits)
Geografia\Oceanía\Estados soberanos\Australia.htm (5 hits)
Geografia\África\Estados soberanos\Angola.htm (5 hits)
Geografia\África\Estados soberanos\Malawi.htm (2 hits)
Geografia\África\Estados soberanos\Sudáfrica.htm (2 hits)
```

Los siguientes casos tienen amarilla pero no fiebre:

```
Geografia\América\Estados soberanos\Costa Rica.htm (1 hit)
Geografia\América\Estados soberanos\Perú.htm (2 hits)
Geografia\América\Estados soberanos\Uruguay.htm (1 hit)
Geografia\Europa\Estados soberanos\Andorra.htm (1 hit)
Geografia\Europa\Estados soberanos\Estonia.htm (1 hit)
Geografia\Europa\Estados soberanos\Portugal.htm (1 hit)
Geografia\Oceanía\Dependencias\Isla Sala y Gómez.htm (1 hit)
Geografia\Oceanía\Dependencias\Pitcairn.htm (1 hit)
```

En total son $6 + 9 + 8 = 23$ archivos.

d) Refinar resultado de consulta (c) usando fiebre y amarilla separadas por uno o más espacios:

fiebre +amarilla

Se espera se recuperen únicamente los siguientes 6 archivos:

```
Search "<fiebres? +amarillas?>" (11 hits in 6 files)
  Geografia\América\Estados_soberanos\Cuba.htm (2 hits)
  Geografia\América\Estados_soberanos\Panamá.htm (1 hit)
  Geografia\América\Estados_soberanos\República_Dominicana.htm (2
hits)
  Geografia\América\Estados_soberanos\Venezuela.htm (2 hits)
  Geografia\África\Estados_soberanos\Ghana.htm (2 hits)
  Geografia\África\Estados_soberanos\Togo.htm (2 hits)
```

En el archivo "Prueba 2 - refinación - fiebre amarilla.txt" se detalla el resultado.

Prueba 3

a) Correr la consulta vectorial

liberia 1

Se espera recuperar los siguientes 13 archivos:

```
Search "<liberia>" (187 hits in 13 files)
  Geografia\América\Estados_soberanos\Costa_Rica.htm (6 hits)
  Geografia\Asia\Estados_soberanos\Bangladés.htm (3 hits)
  Geografia\Europa\Con_reconocimiento_limitado\Kósovo.htm (3 hits)
  Geografia\Europa\Estados_soberanos\Mónaco.htm (3 hits)
  Geografia\Europa\Estados_soberanos\Serbia.htm (4 hits)
  Geografia\Europa\Estados_soberanos\Suecia.htm (3 hits)
  Geografia\Oceanía\Estados_soberanos\Fiyi.htm (3 hits)
  Geografia\África\Estados_soberanos\Costa_de_Marfil.htm (7 hits)
  Geografia\África\Estados_soberanos\Guinea.htm (8 hits)
  Geografia\África\Estados_soberanos\Guinea_Ecuatorial.htm (3 hits)
  Geografia\África\Estados_soberanos\Liberia.htm (126 hits)
  Geografia\África\Estados_soberanos\Malí.htm (3 hits)
  Geografia\África\Estados_soberanos\Sierra_Leona.htm (15 hits)
```

Costa Rica probablemente quede de tercero o más abajo debido a la mayor frecuencia en Liberia y Sierra Leona.

En el archivo "Prueba 3 - liberia.txt" se detalla el resultado.

b) Correr la consulta vectorial

nicoya 2

Se espera recuperar los siguientes 3 archivos.

```
Search "<nicoya>" (38 hits in 3 files)
  Geografia\América\Estados_soberanos\Costa_Rica.htm (31 hits)
  Geografia\América\Estados_soberanos\Nicaragua.htm (6 hits)
  Geografia\América\Estados_soberanos\Panamá.htm (1 hit)
  Costa Rica debería aparecer de primero en el escalafón.
```

En el archivo "Prueba 3 - nicoya.txt" se detalla el resultado.

c) Correr la consulta vectorial

liberia 1; nicoya 2

Se espera recuperar la combinación de archivos de las consultas (a) y (b):

Geografia\América\Estados soberanos\Costa_Rica.htm (6 hits)
 Geografia\Asia\Estados soberanos\Bangladés.htm (3 hits)
 Geografia\Europa\Con_reconocimiento_limitado\Kósovo.htm (3 hits)
 Geografia\Europa\Estados soberanos\Mónaco.htm (3 hits)
 Geografia\Europa\Estados soberanos\Serbia.htm (4 hits)
 Geografia\Europa\Estados soberanos\Suecia.htm (3 hits)
 Geografia\Oceanía\Estados soberanos\Fiyi.htm (3 hits)
 Geografia\África\Estados soberanos\Costa_de_Marfil.htm (7 hits)
 Geografia\África\Estados soberanos\Guinea.htm (8 hits)
 Geografia\África\Estados soberanos\Guinea_Ecuatorial.htm (3 hits)
 Geografia\África\Estados soberanos\Liberia.htm (126 hits)
 Geografia\África\Estados soberanos\Malí.htm (3 hits)
 Geografia\África\Estados soberanos\Sierra_Leona.htm (15 hits)

Geografia\América\Estados soberanos\Costa_Rica.htm (31 hits)
 Geografia\América\Estados soberanos\Nicaragua.htm (6 hits)
 Geografia\América\Estados soberanos\Panamá.htm (1 hit)

Costa Rica debe estar de primera en el escalafón: tiene ambos términos y especialmente el segundo que es más importante en la consulta.

Nicaragua debe estar alto en el escalafón debido a que tiene varias veces el término más importante.

Prueba 4

a) Correr la consulta vectorial

3.000.000 1; 4.000.000 3

Se espera recuperar los siguientes 13 archivos:

Search "\D[34]\.000\.000" (9 hits in 7 files)
 Geografia\Asia\Estados soberanos\Afganistán.htm (1 hit)
 Geografia\Asia\Estados soberanos\Laos.htm (1 hit)
Geografia\Asia\Estados soberanos\Siria.htm (2 hits)
 Geografia\Oceanía\Estados soberanos\Nueva_Zelanda.htm (1 hit)
 Geografia\África\Estados soberanos\Egipto.htm (1 hit)
Geografia\África\Estados soberanos\Etiopía.htm (2 hits)
 Geografia\África\Estados soberanos\Sudán_del_Sur.htm (1 hit)

En el archivo "Prueba 4 - 3M 4M.txt" se detalla el resultado.

Se espera que Etiopía aparezca de primero en el escalafón ya que es el único que posee 3M y 4M.

Los que solo tienen 4M (Afganistán, Laos, Nueva Zelanda) deben aparecer mejor que los que solo tienen 3M (Siria, Egipto, Sudán del Sur). Aunque Siria puede subir un poco porque tiene 3M dos veces.

3. Evidencias y comentarios generales

En general, el proyecto cumple con todas las funcionalidades y con el comportamiento esperado.

Como observaciones o aclaraciones, se puede destacar y a la vez recalcar que los resultados en la cantidad de archivos no es exactamente igual a la sugerida por el profesor ya que se omiten las apariciones del texto en los tags del archivo Html lo cual reduce la cantidad de apariciones y puede cambiar un poco el comportamiento de la consulta, pero sigue siendo un dato aproximado al estimado.

En este apartado se presentan algunos screenshots para evidenciar la ejecución del programa y algunas de sus pruebas.

Escribe una de las siguientes acciones:

- indexar RutaDirectorioColeccion RutaArchivoStopwords RutaDirectorioIndices
- consultar RutaDirectorioIndices RutaEscalafon "Consulta"
- mostrar RutaEscalafon RangoInferior RangoSuperior RutaArchivoSalidaHTML
- refinar RutaEscalafon RutaNuevoEscalafon "ExpresionRegular"
- salir

```
indexar Geografia stopwords.txt indices/
indexar
indices/documentos.csv
obteniendo rutas de archivos...
generando tuplas de rutas de archivos...
creando archivo csv indices/documentos.csv desde diccionario...
extrayendo texto de los archivos...
extrayendo palabras de archivos...
eliminando stopwords...
aplicando stemming...
eliminando tildes de las raices...
contando terminos por documento
calculando pesos de los terminos
creando archivo csv normas.csv desde diccionario...
creando archivo csv diccionario.csv desde diccionario...
creando archivo csv postings.csv...
```

Escribe una de las siguientes acciones:

- indexar RutaDirectorioColeccion RutaArchivoStopwords RutaDirectorioIndices
- consultar RutaDirectorioIndices RutaEscalafon "Consulta"
- mostrar RutaEscalafon RangoInferior RangoSuperior RutaArchivoSalidaHTML
- refinar RutaEscalafon RutaNuevoEscalafon "ExpresionRegular"
- salir

```
consultar indices/ escalafon/ "liberia 1 ; nicoya 2 "
consultar
Ejecutando consulta vectorial...
aplicando stemming...
eliminando tildes de las raices...
creando archivo csv escalafon/escalafon.csv...
```

```

Escribe una de las siguientes acciones:
• indexar RutaDirectorioColeccion RutaArchivoStopwords RutaDirectorioIndices
• consultar RutaDirectorioIndices RutaEscalafon "Consulta"
• mostrar RutaEscalafon RangoInferior RangoSuperior RutaArchivoSalidaHTML
• refinar RutaEscalafon RutaNuevoEscalafon "ExpresionRegular"
• salir

mostrar escalafon/escalafon.csv 0 13 pruebas/
mostrar
Creando reporte HTML...
Nombre del archivo de reporte: bf418718
Escribe una de las siguientes acciones:
• indexar RutaDirectorioColeccion RutaArchivoStopwords RutaDirectorioIndices
• consultar RutaDirectorioIndices RutaEscalafon "Consulta"
• mostrar RutaEscalafon RangoInferior RangoSuperior RutaArchivoSalidaHTML
• refinar RutaEscalafon RutaNuevoEscalafon "ExpresionRegular"
• salir

mostrar escalafon/escalafon.csv 0 144 pruebas/
mostrar
Creando reporte HTML...
Nombre del archivo de reporte: 4d81efbd
Escribe una de las siguientes acciones:
• indexar RutaDirectorioColeccion RutaArchivoStopwords RutaDirectorioIndices
• consultar RutaDirectorioIndices RutaEscalafon "Consulta"
• mostrar RutaEscalafon RangoInferior RangoSuperior RutaArchivoSalidaHTML
• refinar RutaEscalafon RutaNuevoEscalafon "ExpresionRegular"
• salir

refinar escalafon/ escalafon/ río +nilo
refinar
extrayendo texto de los archivos...
creando archivo csv escalafon/escalafonNuevo.csv...
Escribe una de las siguientes acciones:
• indexar RutaDirectorioColeccion RutaArchivoStopwords RutaDirectorioIndices
• consultar RutaDirectorioIndices RutaEscalafon "Consulta"
• mostrar RutaEscalafon RangoInferior RangoSuperior RutaArchivoSalidaHTML
• refinar RutaEscalafon RutaNuevoEscalafon "ExpresionRegular"
• salir

mostrar escalafon/escalafonNuevo.csv 0 3 pruebas/
mostrar
Creando reporte HTML...
Nombre del archivo de reporte: d785096c
Escribe una de las siguientes acciones:
• indexar RutaDirectorioColeccion RutaArchivoStopwords RutaDirectorioIndices
• consultar RutaDirectorioIndices RutaEscalafon "Consulta"
• mostrar RutaEscalafon RangoInferior RangoSuperior RutaArchivoSalidaHTML
• refinar RutaEscalafon RutaNuevoEscalafon "ExpresionRegular"
• salir

```

4. Referencias

1. Markup.py - an HTML and XML generator in Python (s. f.). Recuperado de <http://markup.sourceforge.net/>
2. Snowball (s. f.). Recuperado de <http://snowball.tartarus.org/>