**INSTITUTO TECNOLÓGICO DE COSTA RICA II SEMESTRE 2017**

**INGENIERÍA EN COMPUTACIÓN**

**IC-8020 ELECTIVA TEXTO**

**TAREA PROGRAMADA 2**

**GUÍA DE DOCUMENTACIÓN**

1. Introducción

Presentar brevemente el problema. Puede “reciclar” partes del enunciado de la tarea programada.

1. Completar la siguiente tabla para describir el estado en que quedó la indización de la colección.

|  |  |  |
| --- | --- | --- |
| **Etapa** | **% de complet.** | **Comentario o aclaración** |
| **INDIZACIÓN** |  |  |
| **Campo "texto"** |  |  |
| Extrae adecuadamente del elemento <p> | 100 |  |
| Extrae adecuadamente del elemento <a> | 100 |  |
| Separación en palabras (letras incluyendo eñe) | 100 |  |
| Eliminación de stopwords | 100 |  |
| Extracción de raíces (stemming) | 100 |  |
| Eliminación de acentos, preservación eñe | 100 |  |
| **Campo "ref"** |  |  |
| Extrae adecuadamente del elemento <a> | 100 |  |
| Separación en palabras (letras incluyendo eñe) | 100 |  |
| Conversión a minúsculas | 100 |  |
| Eliminación de acentos (menos eñe= | 100 |  |
|  |  |  |
| **Indexado de la colección** |  |  |
| Indexa correctamente América y Asia |  |  |
| Agrega correctamente África, Europa y Oceanía |  |  |
|  |  |  |
| **Consultas** |  |  |
| Permite usar lenguaje de consultas de Lucene |  |  |

1. Completar la siguiente tabla para describir el resultado obtenido para algunas consultas de prueba. Los casos subrayados son de la segunda fase de indización (África, Asia y Oceanía) y no deben aparecer cuando se corren las consultas en la primera fase de indización (América y Asia).

| **Consulta** | **Resultado esperado** |  | **Comentarios sobre resultado obtenido** |
| --- | --- | --- | --- |
| texto:esfinge | Omán Egipto Estonia |  |  |
| ref:esfinge | Egipto Estonia |  |  |
| ref:guiza | Egipto |  |  |
| texto:tikal | Guatemala Nicaragua |  |  |
| ref:Uaxactún | Guatemala.htm |  |  |
| texto:Alajuela | Costa\_Rica Panama |  |  |
| **nilo** | |  |  | | --- | --- | | Bolivia  Israel  Jordania  Burundi  Egipto  Etiopía | Kenia  Nigeria  Níger  Sudán  Sudán\_del\_Sur | |  |  |
| **"río nilo"** | Egipto Sudán |  |  |
| **fiebre^1 amarilla^4** | |  |  | | --- | --- | | Cuba  Panamá  República\_Dominicana | Venezuela  Ghana  Togo | |  |  |
| **fiebre^4 amarilla^1** | |  |  | | --- | --- | | Cuba  Panamá  República\_Dominicana | Venezuela  Ghana  Togo | |  |  |
| **cuanacaste~0.8** | Costa\_Rica |  |  |
| **"fuga cerebros"~2** | |  |  | | --- | --- | | Argentina  Colombia | Haití  Nueva\_Zelanda | |  |  |
| **ref:"fiebre oro"~2** | |  |  | | --- | --- | | Aruba  Canadá  Estados\_Unidos | Panamá  Australia  Sudáfrica | |  |  |

1. Comentarios finales (estado del programa)

Indicar el estado final en que quedó el programa, problemas encontrados y limitaciones adicionales.

1. Directorio con documentación

La tarea debe ser entregada presentando un archivo comprimido (ZIP, RAR, TGZ) que incluya la siguiente estructura:  
 TP1-*ApellidosNombre1-ApellidosNombre2*  (directorio)  
 *Archivo\_de\_documentacion*  
 *Programas*  
 *Archivos\_adicionales*

*Pruebas*

1. Entrega

**Enviar la tarea y su documentacion a la direccion** [**josee.arayamonge@gmail.com**](mailto:josee.arayamonge@gmail.com)**.**