

Informe de preentrega del Proyecto Final de la asignatura Sistemas de Recuperación de Información

Integrantes:

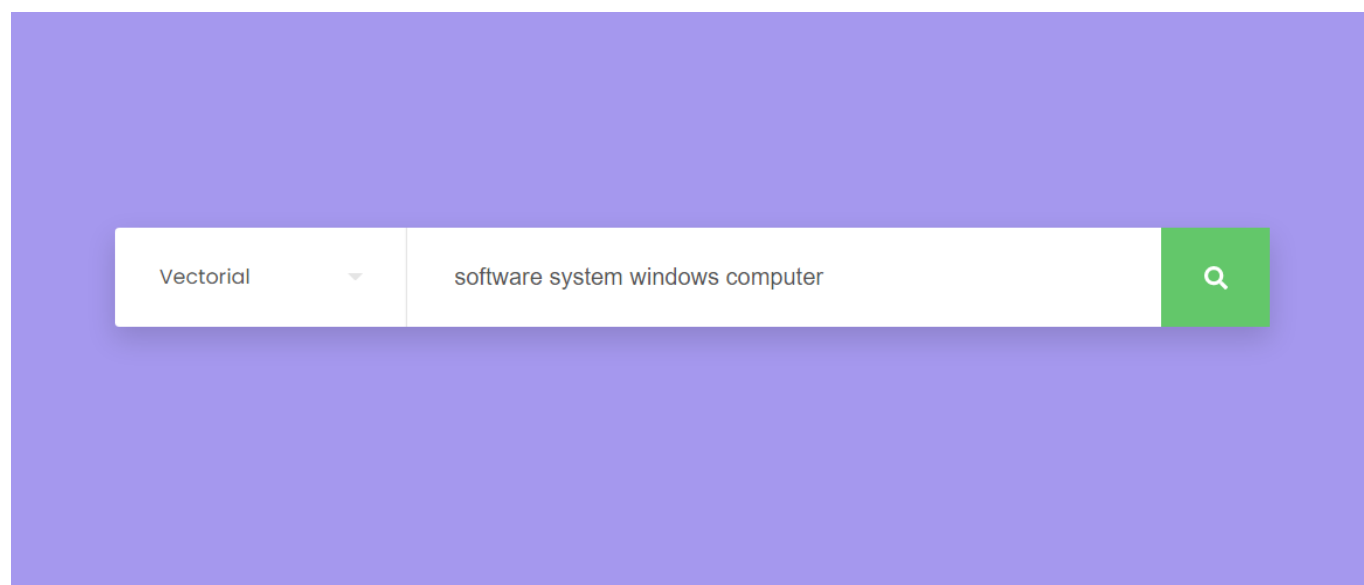
- Lázao A Castro Arango C-311
- Rainel Fernández Abreu C-312

Requerimientos de software:

```
Python 3.9  
nltk 3.0  
sklearn  
numpy  
django 3.2
```

Como primer modelo a implementar elegimos el modelo vectorial. El modelo ha sido testeado en en el corpus 20 Newsgroup con resultados satisfactorios.

Ejemplos que podemos mostrar:



Path	Descarga
C:/Users/lachy/Videos/LACHY/Proyecto Final SRI/corpus/comp.windows.x/67882	Muy pronto
C:/Users/lachy/Videos/LACHY/Proyecto Final SRI/corpus/sci.crypt/15590	Muy pronto
C:/Users/lachy/Videos/LACHY/Proyecto Final SRI/corpus/comp.graphics/38376	Muy pronto
C:/Users/lachy/Videos/LACHY/Proyecto Final SRI/corpus/comp.graphics/38375	Muy pronto
C:/Users/lachy/Videos/LACHY/Proyecto Final SRI/corpus/comp.graphics/38778	Muy pronto
C:/Users/lachy/Videos/LACHY/Proyecto Final SRI/corpus/comp.sys.ibm.pc.hardware/60361	Muy pronto
C:/Users/lachy/Videos/LACHY/Proyecto Final SRI/corpus/comp.windows.x/68232	Muy pronto
C:/Users/lachy/Videos/LACHY/Proyecto Final SRI/corpus/sci.space/59873	Muy pronto
C:/Users/lachy/Videos/LACHY/Proyecto Final SRI/corpus/comp.sys.mac.hardware/51576	Muy pronto
C:/Users/lachy/Videos/LACHY/Proyecto Final SRI/corpus/comp.os.ms-windows.misc/9956	Muy pronto

Como se muestra en la imagen, el proyecto en este punto cuenta con una interfaz visual desarrollada en django. Para hacer una consulta el usuario debe introducir el modelo que desea utilizar (hasta el momento solo el vectorial ha sido implementado) luego introduce su consulta y el sistema devolverá un conjunto de documentos relevantes a dicha consulta. Los resultados se muestran ordenados según el ranking y a cada documento relevante se le hace corresponder una url de su ubicación.

El corpus se analizó local en una de nuestras computadoras y los datos de cada documento se guardaron en diccionarios con extensión .txt .Los resultados que se muestran son las direcciones asociadas a los documentos relevantes con datos locales, próximamente se permitirá procesar cualquier ubicación y se mostrará el nombre del documento, su ruta y un enlace para abrirlo. Se limitó el número de resultados por el momento hasta que se agregue la paginación.

Para ejecutar el proyecto es preciso cumplir los requerimientos, ubicarse en el directorio raíz donde está ubicado y ejecutar:

```
python manage.py runserver
```