Proyecto clasificador tweets

Integrantes: Jorge Miño

Sonia Jaguaco

Objetivo del proyecto

El proyecto se basa en la recolección tweets para procesamiento y limpieza, posterior a esto clasificarlos según los sentimientos y obtener datos de interés en función de su clasificación.

Resumen del proyecto

Recolectar tweets para el posterior análisis de sus sentimientos, posterior a la recolección filtrar los tweets con expresiones regulares, con los tweets filtrados extraer tweets y ponerles sentimientos a criterio de uno, una vez clasificados crear un train con los tweets y con el script Codigo Phyton clasificador.py y posterior a esto filtrar y obtener resultados.

Herramientas utilizadas: Couchdb, Pycharm, elasticsearch

Para instalar Couchdb ingresar a este link

https://www.digitalocean.com/community/tutorials/how-to-install-couchdb-and-futon-on-ubuntu-14-04

Para instalar Pycharm ingresar a este link

http://ubuntuhandbook.org/index.php/2015/07/install-pycharm-ubuntu-1404/

Elasticsearch

Para instalar elasticsearch seguir el siguiente enlace

https://www.digitalocean.com/community/tutorials/how-to-install-and-configure-elasticsearch-on-ubuntu-16-04

Para instalar river en elasticsearch

https://github.com/elastic/elasticsearch-river-couchdb

Instalación plugin head para visualizar elasticsearch

https://mobz.github.io/elasticsearch-head/

Desarrollo del proyecto

Proceso de recolección de tweets necesitamos registrarnos como desarrollador en la página de twitter y obtener credenciales.

En el archivo de Script/Cosechador Phyton.txt ponemos nuestras credenciales.

Para ejecutar el script utilizamos el siguiente comando este script nos permite recolectar tweets

python cosechador\_phyton.py localhost tweets

Creamos una vista en couchdb en el documento Documentacion/Proceso Clasificacion.docx ahí se explica cómo se crea una vista en couchdb y parte del análisis en elasticsearch y exportación utilizando curl y river.

Posterior a esto tenemos que clasificar los tweets lo que dentro del script /Script/Codigo Phyton clasificador.py que está en la

Carpeta del proyecto ejecutamos en pycharm dentro de este script esta explicado que hace cada parte del código.

Para el análisis de datos se creó vistas filtrando los sentimientos este documento de explicación se encuentra en

/Documentacion/Datos filtrados por sentimientos.docx