

Jose Luis Ribero Duque

Ingeniero de Datos

Formación Educativa

- Universidad de Antioquia: Físico
- Universidad de San Diego California: Especialista en Big Data

Proyecto (E, T, L)

Descripción de la Necesidad a Solucionar:

Para las elecciones locales regionales 2018 se necesitaba crear una base de datos unificada con los contactos donde esté la información mínima de cada habitante de un municipio en particular. Esta base sería utilizada para contactar a cada persona y entregar un mensaje político con la idea de ganar adeptos a la campaña política. Se tenían bases de datos de diferentes fuentes, hechas a mano y con estructuras diferentes

Objetivo: Entregar un mensaje vía WhatsApp

Objetivos Específicos:

- Obtener de diferentes fuentes de datos la mayor cantidad de datos de los habitantes del municipio
- Filtrar las bases de datos y hacer un proceso de preparación donde se mejora la estructura de la base de datos y se limpian los campos erróneos o vacíos
- Unificar las diferentes bases de datos en una única base
- Filtrar la base de datos eliminando los contactos que no tengan cuenta de WhatsApp ACTIVA
- Enviar un mensaje por medio de WhatsApp Web

Desarrollo de la Solución:

Extracción

Se hace recolección de las bases de datos, descargando de páginas web y a través de diferentes contactos, también utilizando métodos rudimentarios. Al revisar la información las bases de datos están en su mayoría en Excel y en hojas de papel.

Transformación

Se procede a digitalizar toda la información que estaba en formato físico en un base consolidada quedando así con tres bases de datos no uniformes con datos erróneos y campos vacíos.

- db_cocorna_enlaces.csv
- db_cocorna_people.csv
- db_cocorna_consolidate.csv
-

Para filtrar esta información y darle a las tres bases un formato unificado se hizo el archivo de Python *filtering.py*

Una vez estructuradas las tres bases de datos y eliminados los campos erróneos y vacíos se procede a unificar con el archivo *data_integration.py* las tres bases de datos eliminando los contactos repetidos, obteniendo así la base de datos
Cocorna_phone_numbers_integrated.csv

Carga

Una vez se obtiene la base de datos con la mejor estructura posible se carga a la aplicación de WhatsApp para detectar los teléfonos con cuentas ACTIVAS y se descarga esta base de nuevo para obtener la base de datos filtrada *WhatsApp_contacts.py* esta será la base de datos que se utilizará para enviar los mensajes a través de WhatsApp. El proceso de envío de los mensajes se hace a través de un código que toma los datos de la base de contactos, abre una ventana de Google Chrome con la url de WhatsApp Web busca el contacto y le envía el mensaje deseado.