



8/14/2024

Manual Técnico

IntegrityAI



Jeison Stiven Sabogal Varela
UNIVERSIDAD CENTRAL

Tabla de Contenido

Table of Contents

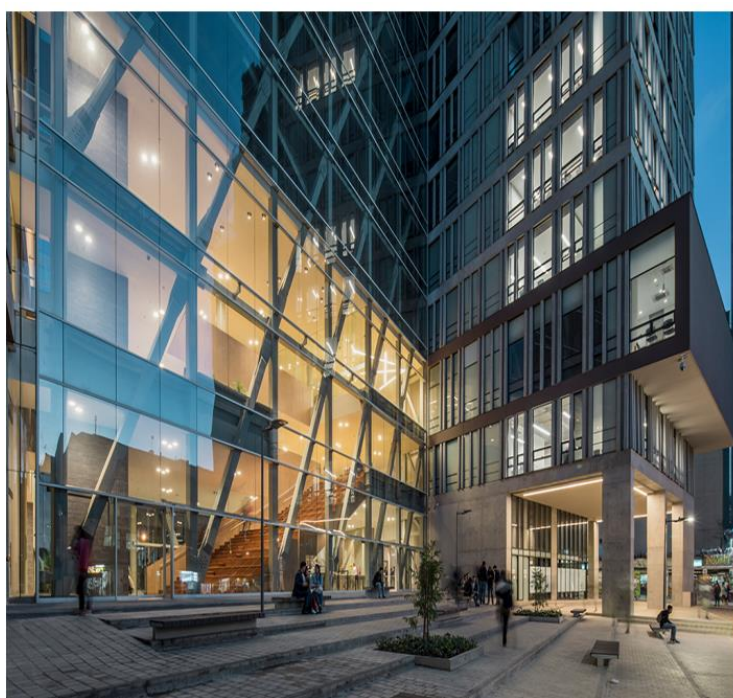
Inicio	2
Creación de Proyecto.....	3
Datos a Validar.....	3
Fuentes Externas Estándar	6
Libre Consulta:.....	6
Plataformas Especializadas:.....	7
Banderas Rojas:.....	8
Dash SECOP II	8
Bases de datos	10
Tablas de Referencia	10
DataBase	11
Banderas Rojas	11
Estructura Proyecto	12
APIs	12
Custom Search API	12
Red Flags Unión Europea.....	13
NOMIC Atlas.....	14
Software.....	15
Spyder (Anaconda)	15
GPT4All	15
Ollama	16
Hardware.....	16
Estructura Lógica:	18
Estructura Proyecto:	18
Anexos.....	20
Anexo 1 – Interfaz de Inicio	20
Anexo 2 – Creación de Proyecto	21
Anexo 3 – Interfaz Graficas SECOP II	23
Anexo 4 – Bases de Datos	23
Anexo 5 – Registro	24
Anexo 6- Banderas Rojas.....	26
Anexo 7 – Interfaz Fechas.....	28
Anexo 8 – Interfaz Texto	29

Interfaz

Inicio

Pantalla de inicio ([Anexo 1](#)) que contiene 3 funciones básicas, las cuales son:

1. **Nuevo Proyecto:** Permite ir a la interfaz de “Crear Nuevo Proyecto”.
2. **Cargar Proyecto:** Lista todos los proyectos creados, abrir su ubicación y abrir el dash de visualización para aquellos proyectos que sean enfocados al SECOP II.
3. **Actualizar Tablas Maestros:** Este botón ejecuta de forma automática una herramienta de “Web Scraping” que descarga todos los datos en **\\IntegrityAI\\resources\\Master\\DB** de las páginas y URLs reportadas dentro del archivo [DataBase](#).



En los casos para los cuales se cargue un proyecto creado previamente, al darle abrir siempre que haya sido un proyecto del SECOP II, abrirá automáticamente el dash del proyecto o caso contrario solo abrirá la carpeta donde esta almacenado, con el fin que el usuario tenga acceso rápido a los datos cargados durante la creación del mismos.

	Nombre de Carpeta	Acción
1	AA	Abrr
2	AAA	Abrr
3	BB	Abrr
4	BBB	Abrr
5	CASA	Abrr
6	CC	Abrr
7	CCC	Abrr
8	DD	Abrr
9	DDO	Abrr
10	FFF	Abrr
11	GGG	Abrr
12	GUAJIRA	Abrr
13	HHH	Abrr
14	III	Abrr
15	KKK	Abrr
16	MAESTRIA_FINAL	Abrr
17	MMM	Abrr
18	NNN	Abrr
19	OOO	Abrr
20	PPP	Abrr
21	QQQ	Abrr

Creación de Proyecto

Datos a Validar

Esta sección condensara todos los archivos que serán objetivo particular de validación y cuenta con una interfaz que condensa de manera sistemática y ordenada una gran cantidad de bases de datos utilizadas de forma regular en procesos de *Integrity Dilifence* en Auditoría Forense, facilitando a los usuarios especializados y casuales de procesos de Auditoria la aplicación sistematizada de la validación general o particular de diferentes fuentes de datos y proyectos de manera asistida por herramientas de NLP y de visualización. Cada una de las secciones tiene un objetivo y enfoque particular ([Anexo 2](#)), cada una de las cuales son:

Seleccionar Datos Objetivo

Tipo de Proyecto

☐ Forensic
 ☐ SECOPII

Identificación del Tercero

☐ Natural
 ☐ Jurídica
 ☐ General

Información a Validar

Seleccionar Archivo
 Seleccionar Carpeta

Ingresar URL

Enfoque Forense

Ingresar el detalle del enfoque, pruebas o validaciones que considere importantes realizar dentro de la base de datos.

Tipo de Proyecto:

- **Forensic:** Indica que es un proyecto general y el origen de datos puede ser de cualquier fuente.
- **SECOP II:** Realiza una consulta directa del SECOP II y activa un segundo dash para la visualización de los datos de la base de datos y consultas posteriores requeridas por el usuario.

Seleccionar Datos Objetivo

Tipo de Proyecto

☐ Forensic

☒ SECOP II

Opciones SECOP II

Departamento:

Colombia

Código de Identificación:

Cota Inferior:

25/08/2024

Cota Superior:

25/08/2024

Dentro de las opciones desplegadas permite seleccionar un departamento específico, ingresar el número de Cédula de Ciudadanía o NIT del tercero, así como el rango de fecha que emplea el campo de “Fecha de Inicio” del SECOP II como filtro de la información, para poder efectuar su descarga.

Identificación del Tercero:

Esta sección es utilizada para capturar los datos de la empresa o tercero que se desea validar de forma particular, particularmente capturando su Nombre/Razón Social y/o CC/NIT en los cuales el usuario seleccione las opciones “**Natural**” o “**Jurica**”. Si se marca la opción de “Realizar Búsqueda WEB”, se hará uso de la API “[Custom Search API](#)” la cual es una herramienta alojada en “*Google Cloud*”, la cual mediante IA analizara noticias, documentos o información en la WEB de la combinación del nombre y/o identificación con una base de datos predeterminadas en español e inglés para este tipo de proyectos, las cuales se pueden agregar o modificar directamente en el código “**etapa2.py**”. Las palabras combinadas con el nombre indicado por el usuario inicialmente son:

"malversación", "fraude", "gobernar", "escándalo", "demanda", "corrupción",
"esquema", "soborno", "bancarrota", "ilegal", "sanción", "lavado de dinero",
"Investigación", "crimen", "detención", "terror", "contrabando", "evasión", "violación",
"embezzle", "fraud", "govern", "scandal", "lawsuit", "corrupt", "scheme", "brib",
"bankrupt", "illegal", "sanction", "money launder", "investigat", "crim", "arrest", "terror",
"smuggl", "evasion", "violat"

Identificación del Tercero

☒ Natural
 ☐ Jurídica
 ☐ General

Ingrese el Nombre

Cédula de Ciudadanía

☐ Realizar Búsqueda WEB

Identificación del Tercero

☐ Natural
 ☒ Jurídica
 ☐ General

Razón Social

NIT

☐ Realizar Búsqueda WEB

Al seleccionar **"General"** no se activarán dichos módulos, puesto que se daría por entendido que se desea realizar una validación global de los datos para la identificación de patrones dentro del total de la población.

Identificación del Tercero

☐ Natural
 ☐ Jurídica
 ☒ General

Nota: Todos los datos aquí seleccionados son capturados y almacenados para consultados de forma posterior o entrenar futuros modelos en el archivo de input.xlsx que se encuentra en **"IntegrityAI\Projects\NombreProyecto\inputs.xlsx"**

Información a Validar:

Esta sección permite seleccionar un archivo en particular, una carpeta o seleccionar una URL que se desee analizar. Posteriormente estos datos capturados serán normalizados en archivo **.parquet** y para el caso de la URL adicionalmente será descargada cualquier página web ingresada en el mejor formato posible identificado.

Información a Validar

Seleccionar Archivo
 Seleccionar Carpeta

Ingresar URL

Ingrese la URL

Enfoque Forense:

Esta sección permite que el usuario escriba que el usuario ingrese en lenguaje natural el enfoque, validaciones o requerimientos de búsqueda dentro de la base de datos

Enfoque Forense

Ingresa el detalle del enfoque, pruebas o validaciones que considere importantes realizar dentro de la base de datos.

Fuentes Externas Estándar

Libre Consulta:

Datos Libres

Todas las fuentes de datos aquí relacionadas realizan una búsqueda de datos específica para cada plataforma, puesto los parámetros de seguridad y estructura de cada una de ellas. Estas bases de datos se alimentan de forma automática desde un Excel que almacenado el nombre del grupo, la URL y el nombre estándar que se le dará al momento de la descarga "**E:\IntegrityAI\resources\inputs_opc\DataBase.xlsx**", todas las cuales son de libre consulta y no presentan ningún costo adicional. Ver detalle en el [Anexo 4](#).

Libre Consulta

Datos Libres

Información de Bases de Datos Libres

- ☐ Superintendencia Financiera (Sanciones)
- ☐ Partidos y Movimientos Políticos
- ☐ Offshore Leaks Database (ICIJ)
- ☐ ONU
- ☐ Departamento de Estado de EEUU (Manual)
- ☐ DIAN
- ☐ Corte Suprema de Justicia (Magistrados)
- ☐ Consejo Superior de la Judicatura
- ☐ Banco Mundial
- ☐ Banco Inglaterra
- ☐ BID (Banco Interamericano de Desarrollo)

Datos con restricción de Consulta:

Esta sección le permite al usuario tener una lista detallada de diferentes plataformas, clasificadas para aquellas que solo aplican para personas naturales y otra para personas Jurídicas.

Datos con Restricciones de Consulta	
Plataformas Personas Naturales	Plataformas Objetivo General
Requiere Realizar Descarga Manual	
<ul style="list-style-type: none">InterpolPolicía NacionalRegistraduría Nacional - Consulta DefunciónADRESConsejo Nacional ElectoralPartidos y Movimientos PolíticosMigración ColombiaCSE HomeRUP RUES (Registro Único de Proponentes)Portal Anticorrupción de ColombiaCorte Constitucional (Magistrados)Corte Suprema de Justicia (Busqueda Expendientes)Registraduria Nacional del Estado CivilRama JudicialCorte Constitucional (Expedientes)Comisión Nacional de la Disciplina JudicialDirectorio de Servidores Públicos (SIGEP)Consulta PEP Función PúblicaConsejo de EstadoConsejo Superior de la JudicaturaXing (Perfil Profesional)LinkedIn (Perfil Profesional)Rama Judicial Consejo Superior de la JuridicaCorte Suprema de Justicia (Magistrados)Fiscalía General de la NaciónConsejo Nacional Electoral	<ul style="list-style-type: none">Cámara de Comercio BogotáCámara de Comercio Bogotá EspecíficoCámara de Comercio MedellínCámara de Comercio CaliCámara de Comercio CartagenaCámara de Comercio PereiraCámara de Comercio BarranquillaCámara de Comercio MonteríaDIAN (Validar NIT)OFACBureau of Industry and SecurityBanco MundialBID (Banco Interamericano de Desarrollo)Offshore Leaks Database (ICIJ)Contraloría General de la RepúblicaSuperintendencia Financiera (Buscar)Superintendencia Financiera (Tablas de Busqueda)DEA (Drug Enforcement Administration)FBI (Federal Bureau of Investigation)Procuraduría General de la NaciónDepartamento de Estado de EEUU (Report)Departamento de Estado de EEUU (List)Superintendencia Financiera (Autoliquidación)

Estas plataformas se abren de forma automática al darle clic encima de cualquiera de las opciones disponibles, con el fin de que el usuario pueda completar el formulario, registro o identificarse dentro de la plataforma con el fin de descargar la información y poder utilizarla como formato de entrada o de validación. Algunas de estas plataformas son gratuitas, pero otras ocupan tener un usuario o pagar una suscripción, pero a diferencia de las plataformas de la sección de “Datos Libres”, sus políticas de seguridad o el requisito de realizar un *login* no permite realizar su descarga automática. El detalle de las URL se encuentra almacenado en "**IntegrityAllresources\inputs_opc\Registro.xlsx**" mismo que puede ser utilizado para actualizar la interfaz de la herramienta o actualizar cualquier URL que haya presentado cambios. Ver detalle de las consultas realizadas en [Anexo 5](#).

Plataformas Especializadas:

El módulo de plataformas especializada, permite a usuarios avanzados con suscripción a alguna de ellas, poder acceder y realizar consultas de la información capturadas en los parámetros de entrada. Puesto que su uso representa costos importantes, no será activado de forma inicial, pero queda abierta su implantación puesto que permitirá:

- Consultar información privada de la composición accionaria de cualquier empresa, con lo cual se podrían identificar relaciones o intereses particulares de un usuario específico.
- Fuentes seguras de información en cuanto a noticias y reportes, al ser filtrados cualquier tipo de FakeNew.
- Acceso a actas y estados financieros.

- Identificación de conflictos de interés y/o caracterización de otros terceros relacionados.
- Búsquedas en bases de datos nacionales e internacionales certificadas.

Suscripción de Pago

Datos de Suscripción

☐ Informa

☐ Factiva

☐ Expedientes

☐ EMIS

☐ D&B Hoovers

Banderas Rojas:

Las banderas rojas son un compendio de casos identificados con un alto factor de riesgo que son de libre consulta y permiten al usuario y al modelo NLP conocer en qué factores particulares debería enfocarse. Las fuentes formuladas pueden agregarse o quitarse según sea necesario, puesto que las mismas son indicadas dentro del archivo

"**IntegrityAI\resources\inputs_opc\Banderas Rojas.xlsx**". Inicialmente para efectos académicos fueron consideradas 3 fuentes de datos, ver detalle en [Anexo 6](#) :

- **Matriz de Riegos de Cundinamarca:** Realiza un compendio de todos los casos de riesgo de fraude, corrupción o colusión identificados por la entidad, desde el año 2021 hasta el año 2024.
- **Casos de fraude en Colombia:** Lista los casos más famosos en Colombia, realizando un resumen de estos, partes involucradas y la forma en la cual se realizaron.
- **Banderas Rojas EU:** Utiliza una API del proyecto Banderas Rojas que mejorar la transparencia de la contratación pública y apoyar la lucha contra la contratación corrupta. Proporciona una herramienta interactiva que permite el seguimiento de los procesos de contratación por parte de ciudadanos, periodistas o incluso funcionarios públicos y detectar fraudes riesgos en las diferentes etapas del proceso de adquisición. La herramienta Banderas rojas verifica automáticamente los documentos de contratación desde el Tenders Electronic Daily (TED) y filtra las contrataciones riesgosas a través de un algoritmo. Aunque riesgoso no significa corrupto, los documentos de adquisición marcados vale la pena comprobarlos.

Dash SECOP II

Sección activada de forma automática cuando se selecciona la opción de SECOP II en la sección de parámetros del proyecto. Permite analizar visualmente las bases de datos generada desde el endpoint <https://www.datos.gov.co/resource/ibiy-vk9h.json>.

El dash cuenta con 3 módulos (Ver [Anexo 3](#)) los cuales el primero contiene todos los campos llave o primarios que pueden ser utilizados de filtro con el fin de depurar los reportes generados por la herramienta siempre que exista la necesidad de ello.

Aunque en la imagen solo son apreciables los primeros campos, es posible escoger entre los 30 campos disponibles, cada uno de los cuales cuenta con un filtro para facilitar su identificación, así como el uso del botón de “Reiniciar Filtros” para eliminar todas las selecciones previas en caso de no ser ya requerido aplicar los mismos.

Todos los filtros que se apliquen en esta sección repercutirán en todo el dash, con lo cual es importante entender cuales de ellos son requeridos o se desean aplicar durante el proceso de análisis.

En el modulo central se disponen de 3 pestañas que presentan todos los datos descargados del SECOPII filtrados en el modulo de la izquierda. La primera pestaña genera 7 tipos de gráficos como son “Línea”, “Barras”, “Dispersión”, “Área”, “Histograma” y “Torta”, usando como información del eje X un campo de texto seleccionable y en el eje Y uno de los campos de importe, donde este ultimo puede ser expresado como valor “Total”, el “Promedio”, “Máximo” y “Mínimo” según se desee analizar.

La pestaña de “Fechas” realiza un conteo de los días entre dos fechas seleccionadas y retorna un diagrama de dispersión y una tabla de resumen de la cantidad de resultados identificados ([Ver Anexo 7](#)).

Dentro de la pestaña de “Texto” se encuentra una herramienta que permite realizar comparar e identificar líneas con un porcentaje de similitud indicado por el usuario y muy útil, al momento de la identificación de bandearas rojas ([Ver Anexo 8](#)).

El modulo de la derecha presenta de forma sistematizada todos los parámetros seleccionados por el usuario durante el proceso de creación de la herramienta, con el fin de poder soportar en que sea necesario dicha información al usuario.

Lista de Hiperparámetros		
#	Parámetro	
1	Tipo de Proyecto - SECOPII	Tru
2	Departamento	Co
3	Cota Inferior	1/0
4	Cota Superior	1/0
5	Enfoque Forense	VA
6	Datos Libres - ONU	Tru
7	Banderas Rojas - Matriz de Riesgos Cundinamarca	Tru

Parámetro: Datos Libres - ONU
Valor: True

Bases de datos

Esta sección contiene el detalle de todas las tablas, bases de datos e información de consulta utilizada por la herramienta. Es importante anotar que estas entradas son dinámicas y pueden agregarse o quitarse actualizando los archivos maestros que contienen las URL, almacenados en **IntegrityAI\resources\inputs_opc**. Es importante anotar que, dada la naturaleza de los datos, algunas URL pueden variar en el tiempo o puede ser necesario agregar o quitar algunas de ellas, lo cual se hace directamente cambiando los archivos de Excel según se desee, puesto que los mismos actualizan el dash de creación de proyectos y utiliza las herramientas de Web Scraping adecuadas sin importar el nuevo enlace que sea seleccionado

Tablas de Referencia

Las tablas de referencia contienen aquellas fuentes de datos públicas y privadas, de pago o gratuitas en las cuales se consignan los nombres, identificaciones o alias de personas

Naturales o Jurídicas de todo el mundo, así como matrices de riesgo. De estas tablas saldrán los datos que utiliza el dash de creación de proyecto y en algunos casos utilizan filtros para su clasificación o agrupación:

- **Anexo 4 DataBase.xlsx:** Contiene la información con las etiquetas para el Dash columna **A**, una descripción genérica en la columna **B**, la URL usada en la columna **C** y el nombre genérico aplicado para el archivo descargado en la columna **D**.
- **Anexo 5: Registro.xlsx:** Condimenta las etiquetas de los datos en la columna **A**, la URL sugerida en la columna **B** y la etiqueta de clasificación de si es aplicable solo para personas naturales o personas jurídicas.
- **Anexo 6: Banderas Rojas.xlsx:** Listado de las opciones que tomará el dash, almacenarlas en la columna **A**, así como su respectiva URL en la columna **B** y nombre del archivo en la columna **C**.
- **ConPago.xlsx:** Listado de plataformas de pago asociadas a auditorías forense especializadas.

DataBase

Esta carpeta almacenada en "**\\IntegrityAI\\resources\\Master\\DB**" guarda la información que es descargada de forma automática utilizando las URL del archivo

"**\\IntegrityAI\\resources\\inputs_opc\\DataBase.xlsx**", toma la información de la columna "**A**", para mostrar las opciones dentro del dash, la columna "**C**" contiene la URL que descargara la herramienta y la columna "**E**" da nombre al archivo y la "**F**" señala con una X si se realizará la descarga o no (X que si 0 que no). El nombre de los archivos está dado por la inicial "**M**" para indicar que hace parte del dataset Maestro.

Para el caso de la sección de "**\\IntegrityAI\\resources\\inputs_opc\\Registro.xlsx**" se pueden agregar la cantidad de plataformas que se consideren necesarias o eliminar las existentes siempre que se completen todos los campos, pero en ningún caso pueden utilizarse etiquetas repetidas.

Banderas Rojas

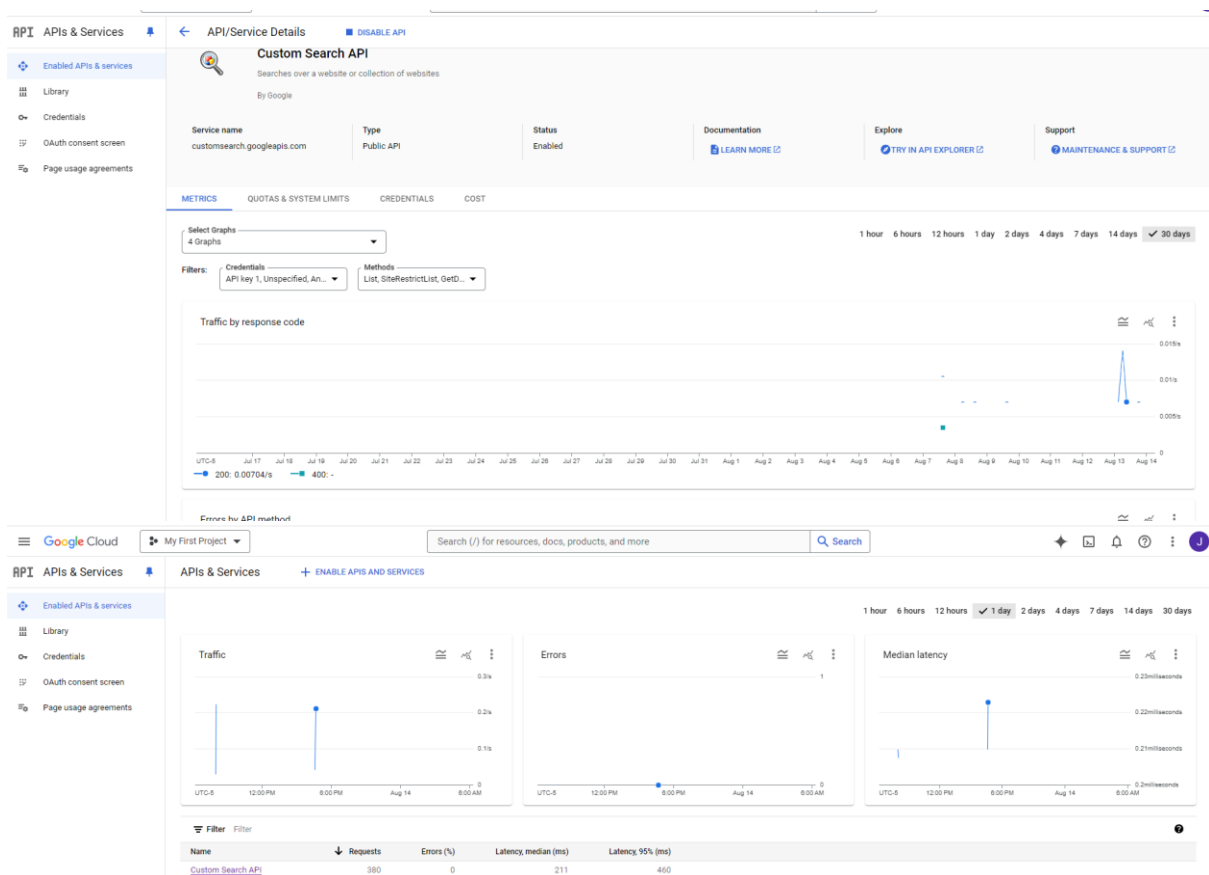
Utiliza el mismo archivo de "**\\IntegrityAI\\resources\\inputs_opc\\DataBase.xlsx**" para realizar las consultas, pero almacena los archivos con la inicial "**B**" en la ruta "**E:\\IntegrityAI\\resources\\Master\\BR**". Existe una sección dentro de esa ruta llamada Bibliografía que almacena documentos técnicos asociados a corrupción, fraude y colusión en Colombia, que ayuda al modelo a entrenarse un poco mejor respecto a qué y cómo buscar "**\\IntegrityAI\\resources\\Master\\BR\\Bibliografia**".

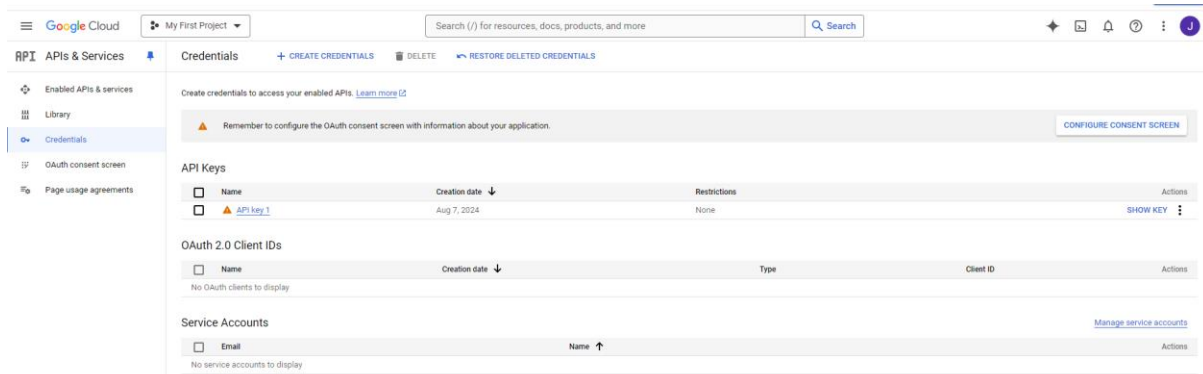
Estructura Proyecto

APIs

Custom Search API


Herramienta de suscripción, que tras creado dentro del servidor WEB de “Google Cloud” con búsqueda sobre toda la WEB, retorna un archivo .json, el cual luego de filtrar la información de la web, retorna la URL de la página en la cual encontró un match y el resumen de la noticia. Para efectos académicos, el código solo solicita 1 consulta, pero puede ser aumentando directamente dentro del código para aumentar la cantidad de búsquedas requeridas a la API dentro del código “**etapa2.py**”. La conexión se realiza creando una API KEY dentro de la plataforma y cada búsqueda tiene un costo de 0.005 USD.





Red Flags Unión Europea

Herramienta de consulta libre para desarrolladores implementada por la Unión Europea la cual reporta la siguiente información consignada en su plataforma oficial [Red Flags API](#). La misma permite realiza la búsqueda, control y seguimientos de contratos mediante el uso de la plataforma Tenders Electronic Daily (TED) que mediante el usa de una IA, clasifica y retorna aquellos contratos o personas identificadas con un alto riesgo asociado.

[ABOUT THE PROJECT](#)[REGISTRATION](#)[DOCUMENTATION](#)

RedFlags API

The purpose of the **Red Flags API** is to **make the data collected by the RedFlags project available** in a structured and automatically processable format outside the RedFlags web page.

The API provides information about the **collected notices and organizations in JSON format**. You can access the service through HTTP requests, and the query parameters must be placed in the GET parameters of each request. You can find more information about the specific requests and their usage in the Red Flags API documentation.

In order to use **Red Flags API** you need to **register** by filling the registration form. As a result you will receive an e-mail containing your unique API key. You must **place this key into each request as a GET parameter**. The API key provides your personal access to the service, do not share it with others.

[REGISTRATION](#)[DOCUMENTATION](#)

About the project

The **Red Flags** project aims to enhance the transparency of public procurements and **support the fight against corrupt procurements**. It provides an interactive tool that allows the monitoring of procurement processes and their implementation by citizens, journalists or even public officials and catch fraud risks at different stages of the procurement process.

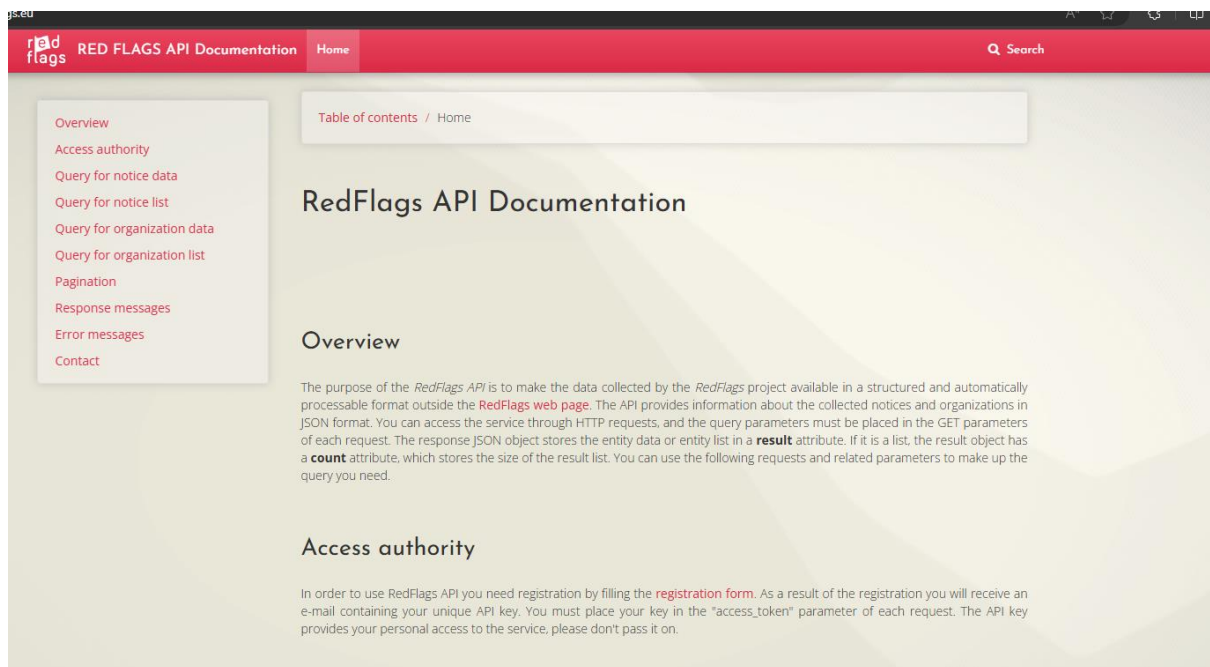
The **Red Flags** tool automatically checks procurement documents from the Tenders Electronic Daily (TED) and **filters risky procurements** through a special algorithm. Although risky does not mean corrupt, flagged procurement documents are worth checking.

[READ MORE](#)

Contact

If you find a mistake or you have any questions, contact us here:

info@redflags.eu



NOMIC Atlas

La API, aunque de uso opcional, aporta grandes resultados en el proceso de *embeddings* de la información previo a la carga del modelo de NLP. La KEY es generada dentro de la plataforma y es cargada directamente dentro de la configuración de la herramienta de GPT4all, su activación es completamente opcional, pero mejora sustancialmente todo el proceso. Cabe anotar que igualmente puede usar dentro de spyder, con el mismo fin cuando se opte por integrar la herramienta. Se detalla el coto del uso de la API.

Current Monthly Usage			
Your usage and charges for the current billing period: from Aug 24, 2024 to Sep 24, 2024			
ITEM	COUNT	UNIT PRICE	TOTAL
Public Data Points (over 250,000)	0	\$0.00005	\$0.00
Private Data Points	0	\$0.0001	\$0.00
Text Embedding Tokens	8498211	\$0.0000001	\$0.75
Image Embeddings	0	\$0.0001500	\$0.00
Seats	1	\$10	\$10.00
Total			\$10.00

Usage Limit	
Hard Data Limit	
To control your billing, you can set a maximum amount of data points that can be billed monthly. Once this threshold is reached, you won't be able to index more data points.	
Data Points Maximum	
	1,000,000

Software

Dentro de los softwares utilizados y por la capacidad de recursos disponibles de Hardware, se optó por la utilización de Spyder (Anaconda) y GPT4ALL, puesto que este último permite ejecutar de forma local modelos de NLP de una manera más eficiente que Spyder, aun en cuanto no exista ninguna restricción de ejecutarlo de forma conjunta o directamente todo dentro del mismo proyecto de Spyder, para obtener el mismo resultado.

Nota: Es posible descargar modelos como Ollama 3.1:405B (215 GB) que fue entrenado con 405 Billones de parámetros que mejorarían substancialmente los resultados de la herramienta, pero por restricciones a nivel de hardware ha sido utilizado el modelo de Ollama 3.1:8B (4.15 GB), sin que eso implique que no se puede utilizar cualquier otro modelo con el mismo fin.

Spyder (**Anaconda**)

El detalle total de las librerías instaladas en el ambiente sobre el cual se ejecuta la herramienta puede ser validado en el Anexo 8, lo cual es importante, puesto que algunas de ellas son APIs o extensiones especializadas para poder realizar correctamente el proceso de transformación o descarga de los datos. Las librerías utilizadas dentro del desarrollo de la herramienta han sido, con varias o todas sus utilidades:

PIL	mimetypes	sys	io	pyarrow
PyQt5	openai	threading	json	requests
docx	os	time	matplotlib	selenium
glob	pandas	tkinter	subprocess	shutil
gpt4all	pdfplumber	webdriver_manager	zipfile	webbrowser

Dash

GPT4All

Herramienta de software libre optimizado para utilizar modelos entrenados de NLP, con la funcionalidad de carga documentación externa, la cual, al ser capturada con la herramienta desarrollada, permite aplicar el proceso de auditoria forense de una forma más eficiente y con un menor consumo de recursos de hardware. Consultar en: <https://docs.gpt4all.io/>

GPT4All

[GPT4All Documentation](#)[Quickstart](#)[Chats](#)[Models](#)[LocalDocs](#)[Settings](#)[Cookbook](#)[Local AI Chat with your Google Drive](#)[Local AI Chat with your Obsidian Vault](#)[Local AI Chat with your OneDrive](#)[Python SDK](#)[GPT4All Python SDK](#)[Monitoring](#)[SDK Reference](#)[Help](#)[FAQ](#)[Troubleshooting](#)

GPT4All Documentation

GPT4All runs large language models (LLMs) privately on everyday desktops & laptops.

No API calls or GPUs required - you can just download the application and [get started](#).

Desktop Application

GPT4All runs LLMs as an application on your computer. Nomic's embedding models can bring information from your local documents and files into your chats. It's fast, on-device, and completely **private**.

[Download for Windows](#) [Download for Mac](#) [Download for Linux](#)

Python SDK

Use GPT4All in Python to program with LLMs implemented with the [llama.cpp](#) backend and Nomic's C backend. Nomic contributes to open source software like [llama.cpp](#) to make LLMs accessible and efficient **for all**.

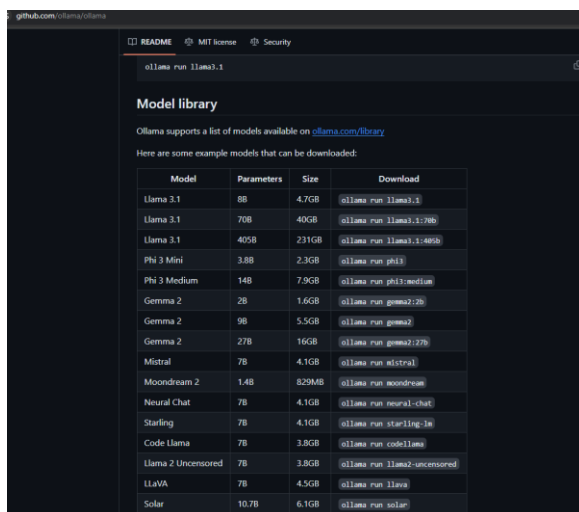
```
pip install gpt4all
```

```
from gpt4all import GPT4All
model = GPT4All("Meta-Llama-3-8B-Instruct.Q4_0.gguf") # downloads / loads a 4.66GB LLM
with model.chat_session():
    print(model.generate("How can I run LLMs efficiently on my laptop?", max_tokens=1024))
```

Ollama

La herramienta permite descargar diferentes modelos de lenguaje de NLP entrenados, con el fin de poder utilizarlos dentro del proyecto de forma local, con resultados muy próximos a los obtenidos con Chat GPT4o. Consultar documentación en:

<https://github.com/ollama/ollama>



The screenshot shows the GitHub repository for Ollama, specifically the 'Model library' section. It lists various models available for download, including Llama 3.1, Phi 3 Mini, Gemma 2, Mistral, and others. Each entry includes the model name, parameters, size, and a download link.

Model	Parameters	Size	Download
Llama 3.1	8B	4.7GB	ollama run llama3.1
Llama 3.1	70B	40GB	ollama run llama3.1:70b
Llama 3.1	405B	231GB	ollama run llama3.1:405b
Phi 3 Mini	3.8B	2.3GB	ollama run phi3
Phi 3 Medium	14B	7.9GB	ollama run phi3-medium
Gemma 2	2B	1.6GB	ollama run gemma2:2b
Gemma 2	9B	5.5GB	ollama run gemma2
Gemma 2	27B	16GB	ollama run gemma2:27b
Mistral	7B	4.1GB	ollama run mistral
Moonstream 2	1.4B	829MB	ollama run moonstream
Neural Chat	7B	4.1GB	ollama run neural-chat
Starling	7B	4.1GB	ollama run starling-lm
Code Llama	7B	3.8GB	ollama run codellama
Llama 2 Uncensored	7B	3.8GB	ollama run llama2-uncensored
LLaVA	7B	4.5GB	ollama run llava
Solar	10.7B	6.1GB	ollama run solar

Hardware

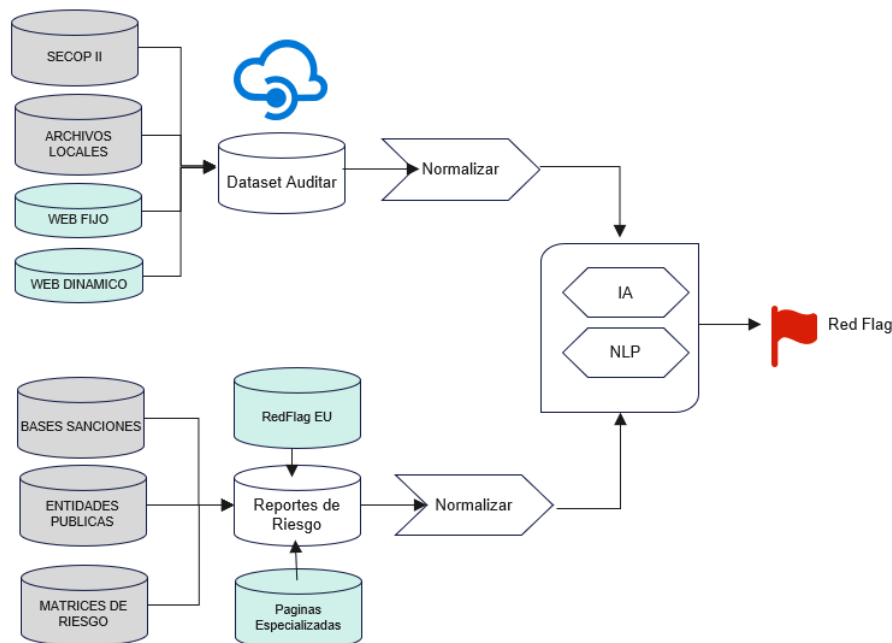
La herramienta ha sido programada de forma con un equipo con las siguientes características:

Item	Value
System Type	x64-based PC
System SKU	SKU
Processor	12th Gen Intel(R) Core(TM) i9-12900K, 3200 Mhz, 16 Core(s), 24 Logical Proce...
BIOS Version/Date	American Megatrends Inc. 1620, 12/08/2022
SMBIOS Version	3.4
Embedded Controller Version	255.255
BIOS Mode	UEFI
BaseBoard Manufacturer	ASUSTeK COMPUTER INC.
BaseBoard Product	PRIME B660-PLUS D4
BaseBoard Version	Rev 1.xx
Platform Role	Desktop
Secure Boot State	Off
PCR7 Configuration	Elevation Required to View
Windows Directory	C:\WINDOWS
System Directory	C:\WINDOWS\system32
Boot Device	\Device\HarddiskVolume1
Locale	Mexico
Hardware Abstraction Layer	Version = "10.0.22621.2506"
User Name	DESKTOP-GUDB9A1\GJ121KC
Time Zone	SA Pacific Standard Time
Installed Physical Memory (RAM)	32,0 GB
Total Physical Memory	31,7 GB
Available Physical Memory	16,7 GB
Total Virtual Memory	33,7 GB
Available Virtual Memory	8,87 GB
Page File Space	2,00 GB
..	...

Estructura Lógica:

Para la ejecución del proyecto se optó por segmentar la información en dos grandes grupos los cuales son procesados dentro de una herramienta de NLP con el fin de poder mediante el uso de esta identificar posibles banderas rojas o indicadores de alto riesgo. Los dos grupos son:

- **Datos Auditar:** Serán en todo caso el foco de revisión toda vez que se componen de los archivos suministrados o descargados dentro de los cuales se propende por encontrar cualquier actividad inusual o posible indicador de fraude, corrupción o colusión que deba ser sujeto de una revisión más detallada.
- **Datos Referencia:** Son todos los ejemplos o casos que le permiten a la herramienta de NLP conocer casos específicos, indicadores o potenciales hallazgos en bases de datos de riesgo con el fin de poder mejorar el proceso de revisión de la base de datos a auditar.



Una vez realizado el proceso de captura de la información se procede a crear bases de datos normalizadas en archivo. parquet, la cual facilita su uso dentro de la herramienta de NLP y la identificación de potenciales banderas rojas, tal como se detalla en la gráfica anterior.

Estructura Proyecto:

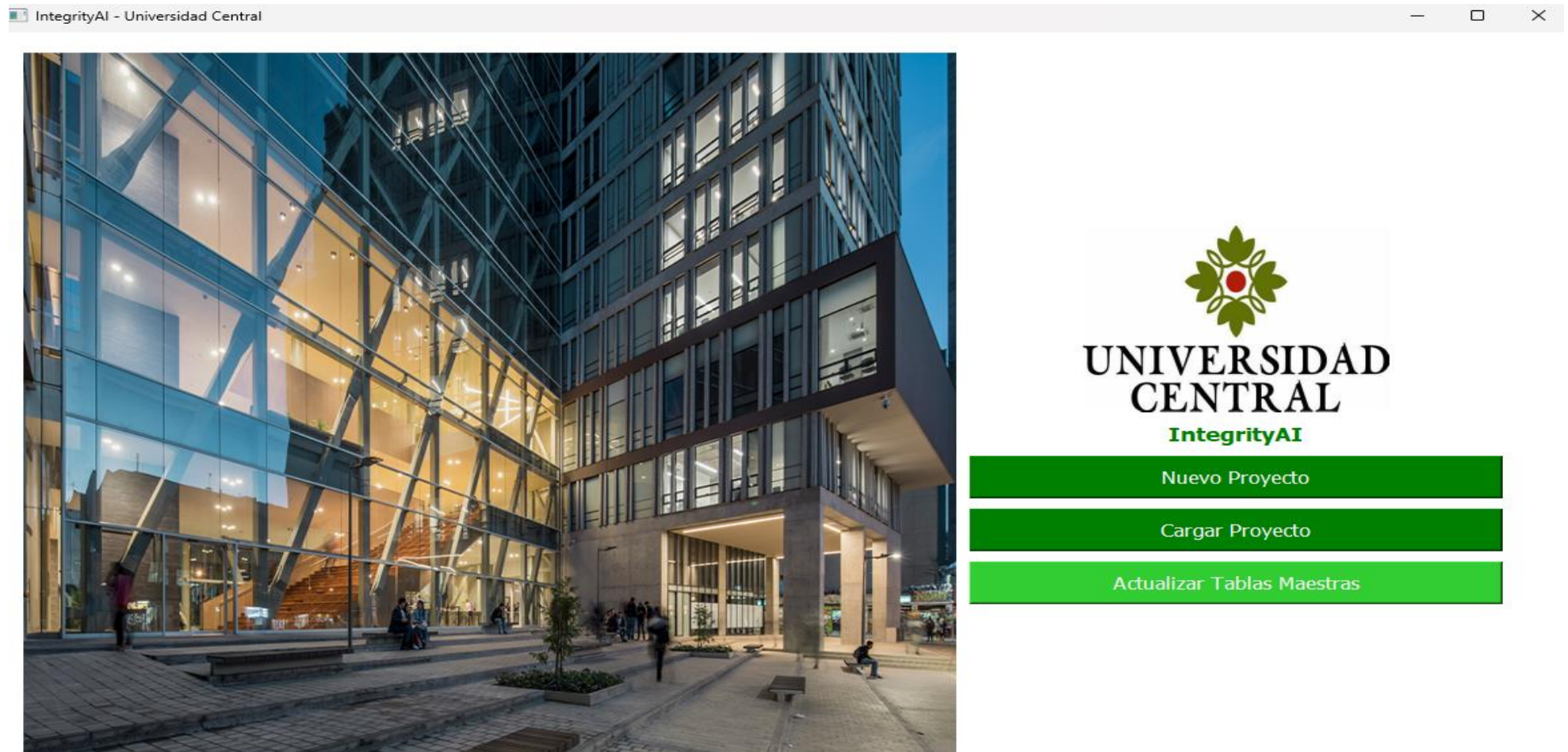
A continuación, se detalle el orden dado al proyecto para su almacenamiento:

```
/ IntegrityAI
|-- / READMRE          # Almacena al manual de usuario y estructura del proyecto.
|-- /.spyproject       # Carpeta del sistema, generada por el ambiente Spyder.
|-- / resources        # Información maestra utilizada para crear proyecto.
    |-- /inputs_opc    # Contiene los archivos Excel maestros.
    |-- /Master        # Almacena los datos maestros de descarga automática.
    |-- /Bibliografía  # Documentos sobre la identificación de fraude.
```

--/BR	# Bases de datos de Banderas Rojas. (Originales)
--/DB	# Bases de datos de Libre Consulta. (Originales)
-- /scr	# Contiene todos los códigos utilizados.
-- /classes	# Contiene el código de interfaz principal de la herramienta.
-- /utilities	# Contiene los códigos de detalle y transformación de datos.
-- /web_utilities	# Herramientas desarrolladas para extracción web.
-- /template	# Template utilizado para la creación de los proyectos.
-- /Projects	# La carpeta almacena todos los proyectos generados.

Anexos

Anexo 1 – Interfaz de Inicio



Anexo 2 – Creación de Proyecto

Nombre del Proyecto:

Generar Proyecto

INICIO

Seleccionar Datos Objetivo		
Tipo de Proyecto		
<input checked="" type="radio"/> Forensic	<input checked="" type="radio"/> SECOPII	
Identificación del Tercero		
<input checked="" type="radio"/> Natural	<input checked="" type="radio"/> Jurídica	<input checked="" type="radio"/> General
Información a Validar		
Seleccionar Archivo	Seleccionar Carpeta	
Ingresar URL		
Enfoque Forense		
Ingresar el detalle del enfoque, pruebas o validaciones que considere importantes realizar dentro de la base de datos.		

Fuentes Externas Estándar	
Libre Consulta	Datos con Restricciones de Consulta
Datos Libres	Plataformas Personas Naturales Plataformas Objetivo General
Información de Bases de Datos Libres	Requiere Realizar Descarga Manual
<input type="checkbox"/> Superintendencia Financiera (Sanciones) <input type="checkbox"/> Partidos y Movimientos Políticos <input type="checkbox"/> Offshore Leaks Database (ICIJ) <input type="checkbox"/> ONU <input type="checkbox"/> Departamento de Estado de EEUU (Manual) <input type="checkbox"/> DIAN <input type="checkbox"/> Corte Suprema de Justicia (Magistrados) <input type="checkbox"/> Consejo Superior de la Judicatura <input type="checkbox"/> Banco Mundial <input type="checkbox"/> Banco Inglaterra <input type="checkbox"/> BID (Banco Interamericano de Desarrollo)	Interpol Policía Nacional Registraduría Nacional - Consulta Defunción ADRES Consejo Nacional Electoral Partidos y Movimientos Políticos Migración Colombia CSE Home RUP RUES (Registro Único de Proponentes) Portal Anticorrupción de Colombia Corte Constitucional (Magistrados) Corte Suprema de Justicia (Busqueda Expedientes) Registraduría Nacional del Estado Civil Rama Judicial Corte Constitucional (Expedientes) Comisión Nacional de la Disciplina Judicial Directorio de Servidores Públicos (SIGEP) Consulta PEP Función Pública Consejo de Estado Consejo Superior de la Judicatura Xing (Perfil Profesional) LinkedIn (Perfil Profesional) Rama Judicial Consejo Superior de la Judicatura Corte Suprema de Justicia (Magistrados) Fiscalía General de la Nación Consejo Nacional Electoral

Especializados
Suscripción de Pago
Datos de Suscripción
<input type="checkbox"/> Informa <input type="checkbox"/> Factiva <input type="checkbox"/> Expedientes <input type="checkbox"/> EMIS <input type="checkbox"/> D&B Hoovers
Banderas Rojas
Fuentes de Indicadores de Banderas Rojas
<input type="checkbox"/> Matriz de Riesgos Cundinamarca <input type="checkbox"/> Casos de Fraude en Colombia <input type="checkbox"/> API RedFlags EU

Nombre del Proyecto:

Generar Proyecto

INICIO

Seleccionar Datos Objetivo

Tipo de Proyecto

☒ Forensic

☒ SECOPII

Identificación del Tercero

☒ Natural

☒ Jurídica

☒ General

Información a Validar

Seleccionar Archivo

Seleccionar Carpeta

Ingresar URL

Enfoque Forense

Ingresar el detalle del enfoque, pruebas o validaciones que considere importantes realizar dentro de la base de datos.

Fuentes Externas Estándar

Libre Consulta

Datos Libres

Información de Bases de Datos Libres

- ☐ Superintendencia Financiera (Sanciones)
- ☐ Partidos y Movimientos Políticos
- ☐ Offshore Leaks Database (ICIJ)
- ☐ ONU
- ☐ Departamento de Estado de EEUU (Manual)
- ☐ DIAN
- ☐ Corte Suprema de Justicia (Magistrados)
- ☐ Consejo Superior de la Judicatura
- ☐ Banco Mundial
- ☐ Banco Inglaterra
- ☐ BID (Banco Interamericano de Desarrollo)

Datos con Restricciones de Consulta

Plataformas Personas Naturales

Plataformas Objetivo General

Requiere Realizar Descarga Manual

Cámara de Comercio Bogotá
Cámara de Comercio Bogotá Especifico
Cámara de Comercio Medellín
Cámara de Comercio Cali
Cámara de Comercio Cartagena
Cámara de Comercio Pereira
Cámara de Comercio Barranquilla
Cámara de Comercio Montería
DIAN (Validar NIT)
OFAC
Bureau of Industry and Security
Banco Mundial
BID (Banco Interamericano de Desarrollo)
Offshore Leaks Database (ICIJ)
Contraloría General de la República
Superintendencia Financiera (Buscar)
Superintendencia Financiera (Tablas de Búsqueda)
DEA (Drug Enforcement Administration)
FBI (Federal Bureau of Investigation)
Procuraduría General de la Nación
Departamento de Estado de EEUU (Report)
Departamento de Estado de EEUU (List)
Superintendencia Financiera (Autoliquidación)

Especializados

Suscripción de Pago

Datos de Suscripción

- ☐ Informa
- ☐ Factiva
- ☐ Expedientes
- ☐ EMIS
- ☐ D&B Hoovers

Banderas Rojas

Fuentes de Indicadores de Banderas Rojas

- ☐ Matriz de Riesgos Cundinamarca
- ☐ Casos de Fraude en Colombia
- ☐ API RedFlags EU

Anexo 3 – Interfaz Graficas SECOP II



PROYECTO SECOPII - CASA

Reiniciar Filtros

Filtros

Nombre entidad

(Secretaría Distrital de Integración Social)
ACI Medellín
ACIDC

Nit entidad

80000118
800002916
800003935
800005042

Departamento

Amazonas
Antioquia
Arauca
Atlántico

Ciudad

Abejorral
Abrego
Acacías
Acevedo

Localización

Colombia, Antioquia, Abejorral
Colombia, Antioquia, Alejandría
Colombia, Antioquia, Amagá

Orden

Corporación Autónoma
Nacional
Territorial

General Fechas Texto

Eje X: Eje Y: Tipo de Visualización: Tipo de Agregación: Actualizar Visualización

Cantidad de registros: 82753

Lista de Hiperparámetros

#	Parámetro
1	Tipo de Proyecto - SECOPII
2	Departamento
3	Cota Inferior
4	Cota Superior
5	Identificación del Tercero - General
6	Datos Libres - ONU
7	Datos Libres - DIAN
8	Datos Libres - Corte Suprema de Justicia (Magistrados)
9	Datos Libres - Banco Mundial
10	Banderas Rojas - Matriz de Riesgos Cundinamarca
11	Banderas Rojas - Casos de Fraude en Colombia

Activate Windows

Go to Settings to activate Windows.

Anexo 4 – Bases de Datos

ETIQUETA BASE DE DATOS

URL

Clasificación

Nombre

DIAN	Proveedores Ficticios 2023	https://www.dian.gov.co/Proveedores_Ficticios/Proveedores-Ficticios-04-10-2023.pdf#search=proveedores%20ficticios	General	M01
DIAN	Proveedores Ficticios 2024	https://www.dian.gov.co/Proveedores_Ficticios/Proveedores-Ficticios-01-03-2024.pdf#search=proveedores%20ficticios	General	M02
DIAN	Proveedores Ficticios 2022	https://www.dian.gov.co/Proveedores_Ficticios/Proveedores_Ficticios_181220202.pdf#search=proveedores%20ficticios	General	M03
DIAN	Proveedores Ficticios 2021	https://www.dian.gov.co/Proveedores_Ficticios/Proveedores-Ficticios-31-05-2021.pdf#search=proveedores%20ficticios	General	M04
ONU	DataBase	https://scsanctions.un.org/consolidated-sp/	General	M05
Banco Inglaterra	Banco Inglaterra	https://ofsistorage.blob.core.windows.net/publishlive/2022format/ConList.csv	General	M06
Departamento de Estado de EEUU (Manual)	Base de Datos	https://www.state.gov/wp-content/uploads/2021/07/Congressional-Report-Section-353-Names.pdf	General	M07
Departamento de Estado de EEUU (Manual)	Solo Datos Pagina WEB	https://www.state.gov/terrorist-exclusion-list/	General	M08
Banco Mundial	Descarga Archivo	https://projects.worldbank.org/en/projects-operations/procurement/debarred-firms	General	M09
BID (Banco Interamericano de Desarrollo)	Descarga Archivo	https://mydata.iadb.org/resource/vsju-qrtz.json	General	M10
Offshore Leaks Database (ICIJ)	Deja descargar DB	https://offshoreleaks-data.icij.org/offshoreleaks/csv/full-olddb.LATEST.zip	General	M11
Superintendencia Financiera (Sanciones)	Tabla en WEB	https://www.datos.gov.co/resource/iaeu-rcn6.json	General	M12
Partidos y Movimientos Políticos	Página WEB	https://www.senado.gov.co/index.php/el-senado/senadores	Persona	M13
Corte Suprema de Justicia (Magistrados)	Página WEB	https://www.ramajudicial.gov.co/portal/sobre-la-rama/magistrados-de-las-altas-cortes/corte-suprema-de-justicia	Persona	M14
Consejo Superior de la Judicatura	Página WEB	https://www.ramajudicial.gov.co/portal/sobre-la-rama/magistrados-de-las-altas-cortes/consejo-superior-de-la-judicatura	Persona	M15

Anexo 5 – Registro

Etiqueta	URL	Clasificación
Cámara de Comercio Bogotá	https://sico.ccb.org.co/Sico/Paginas/frmLogin.aspx	General

<i>Cámara de Comercio Bogotá Específico</i>	https://linea.ccb.org.co/gestionexpedientes/matriculas/terminos-y-condiciones	General
<i>Cámara de Comercio Medellín</i>	https://virtuales.camaramedellin.com.co/Consultaexpedientes/#!/login	General
<i>Cámara de Comercio Cali</i>	https://www.ccc.org.co/sedevirtual/consulta-expedientes-inicio/	General
<i>Cámara de Comercio Cartagena</i>	https://www.cccartagena.org.co/tramites-en-linea/	General
<i>Cámara de Comercio Pereira</i>	https://docxflow.camarapereira.org.co/public/sucursal/expedientes	General
<i>Cámara de Comercio Barranquilla</i>	https://virtuales.camarabaq.org.co/Paginas/Inicio.aspx	General
<i>Cámara de Comercio Montería</i>	https://ccmonteria.org.co/servicios-virtuales/consultar-expedientes	General
<i>DIAN (Validar NIT)</i>	https://muisca.dian.gov.co/WebGestionmasiva/DefSelPublicacionesExterna.faces;jsessionid=0BE2B04BB6AFA2B69E76A08F3B216E64	General
<i>OFAC</i>	https://sanctionssearch.ofac.treas.gov/	General
<i>Bureau of Industry and Security</i>	https://www.ecfr.gov/current/title-15/subtitle-B/chapter-VII/subchapter-C/part-744	General
<i>Banco Mundial</i>	https://projects.worldbank.org/en/projects-operations/procurement/debarred-firms	General
<i>BID (Banco Interamericano de Desarrollo)</i>	https://www.iadb.org/es/quienes-somos/transparencia/sistema-de-sanciones/empresas-e-individuos-sancionados	General
<i>Offshore Leaks Database (ICIJ)</i>	https://offshoreleaks.icij.org/	General
<i>Contraloría General de la República</i>	https://www.contraloria.gov.co/es/web/guest/control-fiscal/responsabilidad-fiscal/certificado-de-antecedentes-fiscales	General
<i>Superintendencia Financiera (Buscar)</i>	https://wl.superfinanciera.gov.co/SiriWeb/publico/sancion/rep_sanciones_general_par.jsf	General
<i>Superintendencia Financiera (Tablas de Búsqueda)</i>	https://www.superfinanciera.gov.co/publicaciones/20708/sanciones-20708/	General
<i>DEA (Drug Enforcement Administration)</i>	https://www.dea.gov/	General
<i>FBI (Federal Bureau of Investigation)</i>	https://www.fbi.gov/wanted	General
<i>Interpol</i>	https://www.interpol.int/es/Como-trabajamos/Notificaciones/Notificaciones-rojas/Ver-las-notificaciones-rojas	Persona
<i>Procuraduría General de la Nación</i>	https://apps.procuraduria.gov.co/webcert/Certificado.aspx?t=dAylAkFT/gSkkvpDol89aORiq2C8LI3z9uHAnBFaF08/32nPrGQhH4HhIkyJHgMD30HMsssetl+/G8W9Pv8tRCAPwrekO2TVffPpwgClddzRipvVlkwZR+TABjraDbp0hlrwP2GCLMX8RI+aw1u6GwgGEkvY16WZ4qm8sL17YyY5FLVGWSy+Eys6FEeWWKx6G+Rm/XaHa4L9eayTZkhzvlL4bJrDB501fmA/wv3NJ74I=&tpo=2	General
<i>Policía Nacional</i>	https://antecedentes.policia.gov.co:7005/WebJudicial/antecedentes.xhtml	Persona
<i>Registraduría Nacional - Consulta Defunción</i>	https://defunciones.registraduria.gov.co/	Persona
<i>ADRES</i>	https://www.adres.gov.co/consulte-su-eps	Persona
<i>Consejo Nacional Electoral</i>	https://www.cne.gov.co/la-entidad/magistrados	Persona
<i>Partidos y Movimientos Políticos</i>	https://www.senado.gov.co/index.php/el-senado/senadores	Persona
<i>Migración Colombia</i>	https://www.migracioncolombia.gov.co/buscar?qn	Persona

CSE Home	https://cse.google.com/cse?cx=001883528321410408823:kiduwzotqdm	Persona
RUP RUES (Registro Único de Proponentes)	https://www.rues.org.co/	Persona
Portal Anticorrupción de Colombia	https://portal.paco.gov.co/index.php?pagina=contratista&identificacion=0	Persona
Corte Constitucional (Magistrados)	https://www.ramajudicial.gov.co/portal/sobre-la-rama/magistrados-de-las-altas-cortes/corte-constitucional	Persona
Corte Suprema de Justicia (Búsqueda Expedientes)	http://consultajurisprudencial.ramajudicial.gov.co:8080/WebRelatoria/csj/index.xhtml	Persona
Registraduría Nacional del Estado Civil	https://registraduria.gov.co/-Notificaciones-de-tutelas-.html	Persona
Rama Judicial	https://consultaprocesos.ramajudicial.gov.co/Procesos/NumeroRadicacion#DetalleProceso	Persona
Corte Constitucional (Expedientes)	http://www.corteconstitucional.gov.co/secretaria/	Persona
Comisión Nacional de la Disciplina Judicial	https://www.ramajudicial.gov.co/comision-nacional-de-disciplina-judicial	Persona
Directorio de Servidores Públicos (SIGEP)	https://www1.funcionpublica.gov.co/web/sigep2/directorio	Persona
Consulta PEP Función Pública	https://www.funcionpublica.gov.co/fdci/consultaCiudadana/consultaPEP	Persona
Consejo de Estado	https://servicios.consejodeestado.gov.co/testmaster/nue_consua.asp	Persona
Consejo Superior de la Judicatura	https://www.ramajudicial.gov.co/portal/sobre-la-rama/magistrados-de-las-altas-cortes/consejo-superior-de-la-judicatura	Persona
Xing (Perfil Profesional)	https://www.xing.com/es	Persona
LinkedIn (Perfil Profesional)	https://www.linkedin.com/login?fromSignIn=true&trk=quest_homepage-basic_nav-header-signin	Persona
Rama Judicial Consejo Superior de la Jurídica	sirna.ramajudicial.gov.co/Paginas/Certificado.aspx	Persona
Corte Suprema de Justicia (Magistrados)	https://www.ramajudicial.gov.co/portal/sobre-la-rama/magistrados-de-las-altas-cortes/corte-suprema-de-justicia	Persona
Fiscalía General de la Nación	https://www.fiscalia.gov.co/colombia/	Persona
Departamento de Estado de EEUU (Report)	https://www.state.gov/wp-content/uploads/2021/07/Congressional-Report-Section-353-Names.pdf	General
Departamento de Estado de EEUU (List)	https://www.state.gov/terrorist-exclusion-list/	General
Superintendencia Financiera (Autoliquidación)	https://www.superfinanciera.gov.co/publicaciones/10083550/reporte-autoliquidaciones-10083550/	General
Consejo Nacional Electoral	https://www.cne.gov.co/la-entidad/magistrados	Persona

Anexo 6- Banderas Rojas.

ETIQUETA	URL	Nombre	Descripción
----------	-----	--------	-------------

<i>Casos de Fraude en Colombia</i>	https://es.wikipedia.org/wiki/Anexo:Casos_de_corrupci%C3%B3n_pol%C3%ADtica_en_Colombia_por_a%C3%B1o	BR01	Wikipedia
<i>Matriz de Riesgos Cundinamarca</i>	https://mapas1.cundinamarca.gov.co/wcm/connect/51836e86-0a88-4206-bf55-363907cd6f90/MAPA+DE+RIESGOS+DE+CORRUPCI%C3%93N+ACTUALIZADO+Ago+sto+03+2022.xlsx?MOD=AJPERES&CONVERT_TO=url&CACHEID=ROOTWORKSPACE-51836e86-0a88-4206-bf55-363907cd6f90-oVodXZM	BR02	Tabla del 2021
<i>Matriz de Riesgos Cundinamarca</i>	https://mapas1.cundinamarca.gov.co/wcm/connect/c1ef1aad-9bb7-4e6b-99ed-378b70c473b6/Mapa+de+Riesgos+de+Corrupci%C3%B3n.xlsx?MOD=AJPERES&CONVERT_TO=url&CACHEID=ROOTWORKSPACE-c1ef1aad-9bb7-4e6b-99ed-378b70c473b6-oVo8cKh	BR03	Tabla del 2022
<i>Matriz de Riesgos Cundinamarca</i>	https://mapas1.cundinamarca.gov.co/wcm/connect/5ae8311e-a5db-45f1-bf11-675a54de3156/MAPA+RIESGOS+CORRUPCI%C3%93N+Y+FRAUDE+%2B+ACTUALIZACI%C3%93N+1+2023.xlsx?MOD=AJPERES&CONVERT_TO=url&CACHEID=ROOTWORKSPACE-5ae8311e-a5db-45f1-bf11-675a54de3156-oVo9VIV	BR04	Tabla del 2023
<i>Matriz de Riesgos Cundinamarca</i>	https://mapas1.cundinamarca.gov.co/wcm/connect/cc73f10c-2a36-4627-bf58-596ffebbbbfc/Mapa+Riesgos+de+Corrupci%C3%B3n+y+Fraude+2024.xlsx?MOD=AJPERES&CONVERT_TO=url&CACHEID=ROOTWORKSPACE-cc73f10c-2a36-4627-bf58-596ffebbbbfc-p44Sz8A	BR05	Tabla del 2024
<i>API RedFlags EU</i>	UTILIZA API	BR06	Es una API

Anexo 7 – Interfaz Fechas



PROYECTO SECOPII - CASA

Reiniciar Filtros

Filtros

Nit entidad

Buscar en Nit entidad

800000118
800002916
800003935
800005042

Departamento

Buscar en Departamento

Tolima
Valle del Cauca
Vaupés
Vichada

Ciudad

Buscar en Ciudad

Abejorral
Abrego
Acacias
Acevedo

Localización

Buscar en Localización

Colombia, Antioquia, Abejorral
Colombia, Antioquia, Alejandría
Colombia, Antioquia, Amagá

Orden

Buscar en Orden

Corporación Autónoma
Nacional
Territorial

Sector

Buscar en Sector

ente y Desarrollo Sostenible
ia Tecnología

General Fechas Texto

Fecha base: fecha de inicio del contrato Fecha comparar: fecha de fin del contrato Actualizar Visualización

Diferencia de días entre fecha_de_inicio_del_contrato y fecha_de_fin_del_contrato

days_diff	count
0	5
1	5
2	8
3	5
4	7

Cantidad de registros: 7532

Lista de Hiperparámetros

#	Parámetro
1	Tipo de Proyecto - SECOPII
2	Departamento
3	Cota Inferior
4	Cota Superior
5	Identificación del Tercero - General
6	Datos Libres - ONU
7	Datos Libres - DIAN
8	Datos Libres - Corte Suprema de Justicia (Magistrados)
9	Datos Libres - Banco Mundial
10	Banderas Rojas - Matriz de Riesgos Cundinamarca
11	Banderas Rojas - Casos de Fraude en Colombia

Parámetro: Tipo de Proyecto - SECOPII
Valor: True

Anexo 8 – Interfaz Texto



PROYECTO SECOPII - MAESTRIA_FINAL

Reiniciar Filtros

Filtros

Nombre entidad

Buscar en Nombre entidad

(Secretaría Distrital de Integración Social)

ADMINISTRADORA DE LOS RECURSOS DEL SIS

AEROCIVIL

Nit entidad

Buscar en Nit entidad

800000118

800002916

8000068503

800007652

Departamento

Buscar en Departamento

Amazonas

Antioquia

Arauca

Atlántico

Ciudad

Buscar en Ciudad

Acacías

Acevedo

Agua De Dios

Agustín Codazzi

Localización

Buscar en Localización

Colombia, Antioquia, Amagá

Colombia, Antioquia, Amalfi

Colombia, Antioquia, Apartadó

Orden

Buscar en Orden

Corporación Autónoma

Nacional

Territorial

General Fechas Texto

Campo a Validar: domicilio representante legal Similitud: 76% Buscar Similitudes

ID Contrato 1	umento Proveedi	Texto 1	Similitud (%)	Texto 2	umento Proveedi	ID Contrato 2
CO1.PCCNTR....	1042444114	CARRERA 35 26 ...	76%	CARRERA 20 ...	1143451729	CO1.PCCNTR.6586860
CO1.PCCNTR....	1042444114	CARRERA 35 26 ...	79%	Carrera 9 11a-16	1067809219	CO1.PCCNTR.6566618
CO1.PCCNTR....	1129567170	Calle 81 # 74 - 07	76%	Calle 58 #38-107	1047360592	CO1.PCCNTR.6589848
CO1.PCCNTR....	33108174	CALLE 135 ...	100%	CALLE 135 ...	1234094778	CO1.PCCNTR.6550136
CO1.PCCNTR....	63542029	CALLE 16 78 29	83%	CALLE 73 # ...	1047367280	CO1.PCCNTR.6582258
CO1.PCCNTR....	1143463867	Calle 49 # 2-99	77%	CALLE 9 12-30	1048287387	CO1.PCCNTR.6570968
CO1.PCCNTR....	32853656	calle 18 numer...	76%	CALLE 14 N ...	1048280540	CO1.PCCNTR.6508855
CO1.PCCNTR....	22507924	Carrera 14 ...	79%	carrera 19 # 25-64	1043012316	CO1.PCCNTR.6591970
CO1.PCCNTR....	22507924	Carrera 14 ...	81%	Carrera 40 # 4 77	1192729902	CO1.PCCNTR.6573688

Cantidad de registros: 273

Lista de Hiperparámetros

#	Parámetro	
1	Tipo de Proyecto - SECOPII	Tru
2	Departamento	Co
3	Cota Inferior	1/C
4	Cota Superior	1/C
5	Enfoque Forense	VA
6	Datos Libres - ONU	Tru
7	Banderas Rojas - Matriz de Riesgos Cundinamarca	Tru