

<TopNews/>

Guía de uso

Tabla de contenidos

Características del programa	3
Requisitos del sistema	3
Funcionalidades del programa.....	3
Descarga.....	3
Componentes importantes:	4
Base de datos	4
Bitácora	4
Gestor principal.....	4
Log.....	4
Parablas clave	4
Subproceso principal.....	4
Subproceso webscraping	4
Iniciando el programa	5
Ventana inicial.....	5
Ventanas del programa.....	7
Ventana Principal	7
Opciones de la ventana Principal	7
Ventana Fuentes de información.....	9
Ventana Palabras clave	10
Consejos importantes para añadir valores	12
Ventana Destinatarios	13
Ejecución del Subproceso principal	14
Fase 2: Subproceso webscraping	15
Almacenamiento de datos	17
Envío de mensajería electrónica	19
Office365 vs Outlook.....	19
Manejo de errores	21
Errores conocidos	22
Reportar errores	23
Desarrollador	23

Características del programa

Este programa fue desarrollado utilizando el lenguaje Python3.7 y el framework Kivy, complementado con HTML y CSS para las visualizaciones y reportería.



Requisitos del sistema

Los siguientes son los requerimientos deseables para el buen funcionamiento del programa:

- Sistema operativo Windows 10.
- 2.5 GB disponibles de espacio en disco.
- Acceso a internet y a las fuentes de información contempladas por el programa.
- Acceso a cuenta de Microsoft Office 365 y/o software Microsoft Outlook instalado en el equipo y funcional.

Funcionalidades del programa

- 1) **Recopilar** información de sitios de internet. El programa implementa elementos avanzados de webscraping para buscar y obtener información de sitios web de interés.
- 2) **Analizar** dicha información para determinar su relevancia según los parámetros definidos por el usuario (Palabras clave).
- 3) **Consolidar y almacenar** la información que resulte ser relevante de manera que quede disponible para aprovechamiento del usuario en cualquier momento.
- 4) **Distribuir** la información relevante a los destinatarios interesados de una forma clara y organizada.

Descarga

Puede descargar la última versión del programa en: <https://github.com/Ferd656/TopNews>

O bien, póngase en contacto con el [Desarrollador](#).

Componentes importantes:

Base de datos

El programa utilizará una base de datos SQLite, ubicada en el directorio del archivo ejecutable, para almacenar información relacionada con su funcionamiento.

Bitácora

El programa llevará un registro de los ítems que ha procesado durante la ejecución del Subproceso webscraping con el fin de que pueda tomar decisiones sobre la información que ya fue procesada. Este registro, llamado Bitácora se encontrará almacenado en la Base de datos del programa.

Los ítems almacenados en bitácora pueden tener los siguientes estados:

- REV: Indica que el ítem ya fue revisado por el programa y no lo clasificó como relevante.
- PEND: Indica que el ítem revisado fue clasificado como relevante, pero que aún no se ha concretado el proceso de envío por correo electrónico.
- ENV: Indica que el ítem revisado fue clasificado como relevante, y se concretó el respectivo proceso de envío por correo electrónico.

Gestor principal

El gestor principal es el elemento medular del programa, consiste de toda la interfaz gráfica y sus diversas interacciones, además se encarga de la inicialización del Subproceso webscraping y la distribución de mensajería electrónica. En otras palabras, se trata del proceso principal.

Log

Consiste en un fichero de texto, ubicado en el directorio del archivo ejecutable, que almacenará detalles sobre los procesos ejecutados por el programa. El Log se reiniciará cada vez que el programa se inicialice, se recomienda realizar una revisión periódica del Log cada vez que el programa se encuentre en ejecución.

Parabras clave

Palabras definidas por el usuario utilizadas como parámetro por el programa para determinar la relevancia de la información obtenida mediante el análisis de texto.

Subproceso principal

El subproceso principal consiste de un ciclo de 3 fases:

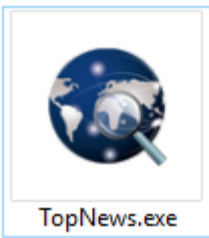
1. Cuenta regresiva sobre la frecuencia de escaneo.
2. Ejecución del Subproceso webscraping.
3. Generación y envío de la correspondencia electrónica con los insumos generados por el Subproceso webscraping. Posteriormente inicia de nuevo desde la fase 1.

Subproceso webscraping

Se encarga de ejecutar la búsqueda de información a través de internet, analizarla y almacenarla para transferirla al Gestor principal mediante el Subproceso principal, el cual se encargará de su debida distribución.

Iniciando el programa

Para iniciar el programa TopNews haga sobre click en el ícono del ejecutable o acceso directo:

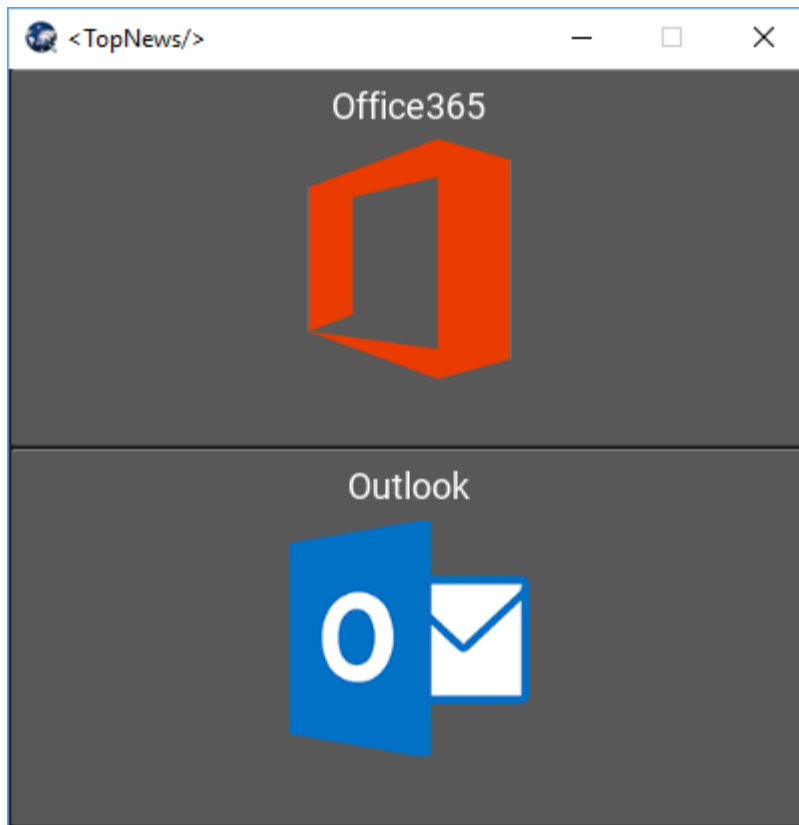


Cada vez que el programa inicia, se eliminarán los elementos en Bitácora clasificados como PEND.

Recomendación: Almacene el archivo ejecutable en el directorio que considere apropiado para el mismo, considerando que en dicho directorio serán almacenados también otros elementos como la Base de datos del programa.

Ventana inicial

Una vez el software ha concretado su inicialización, se mostrará la siguiente ventana inicial:



Haciendo click en la sección que corresponda, usted podrá indicar la plataforma mediante la cual desea que el programa gestione la generación y envío de mensajería electrónica: Office365(SMTP) o Outlook.

The screenshot shows a web browser window titled "<TopNews/>". The login form has a dark background. The "Usuario" field contains the email "114420864@bancobcr.com". The "Contraseña" field is empty. Below the fields, the text "Validando . . ." is displayed in yellow. At the bottom, there are two buttons: "Ingresar" (blue) and "Volver" (grey).

Si usted selecciona la plataforma Office365, será dirigido a una pantalla donde el programa validará sus credenciales de acceso y conexión a internet.

The screenshot shows the same login form as above, but with an error message. The "Usuario" field still contains "114420864@bancobcr.com" and the "Contraseña" field is empty. The error message, in red, reads: "(535, b'5.7.3 Authentication unsuccessful [BN6PR22CA0043.namprd22.prod.outlook.com']) Autenticación Office365 no fue exitosa. Verifique credenciales y conexión a internet." Below the error message, it says "Intentos restantes = 2" in red. The "Ingresar" and "Volver" buttons are still present at the bottom.

En caso de ocurrir un problema durante el inicio de sesión, el programa le mostrará una etiqueta con la descripción del error.

Para cierto tipo de errores el programa establecerá un contador que le regresará automáticamente a la página anterior al llegar a cero.

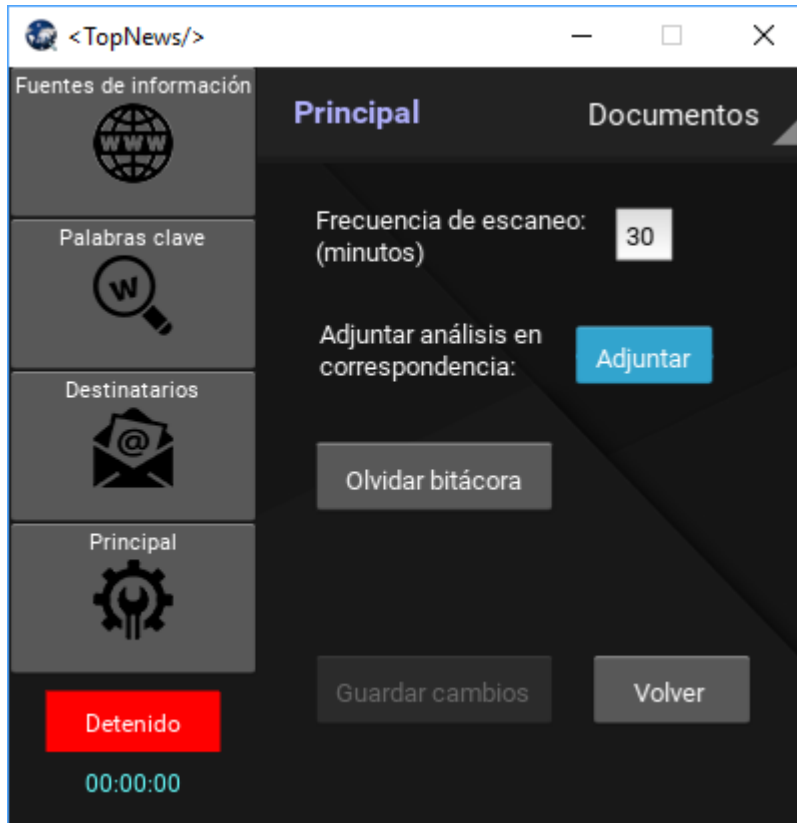
Los errores más comunes son:

- Credenciales incorrectas.
- Problemas con la conexión a internet.

Ventanas del programa

Ventana Principal

Una vez haya seleccionado su plataforma de preferencia en la ventana inicial(y haber validado sus credenciales para el caso de Office365) se le llevará a una nueva sección del programa. La sección consiste en un conjunto de ventanas a las cuales podrá acceder con ayuda de los botones ubicados en la parte izquierda del programa.



La primera ventana que se mostrará será la ventana “Principal”.

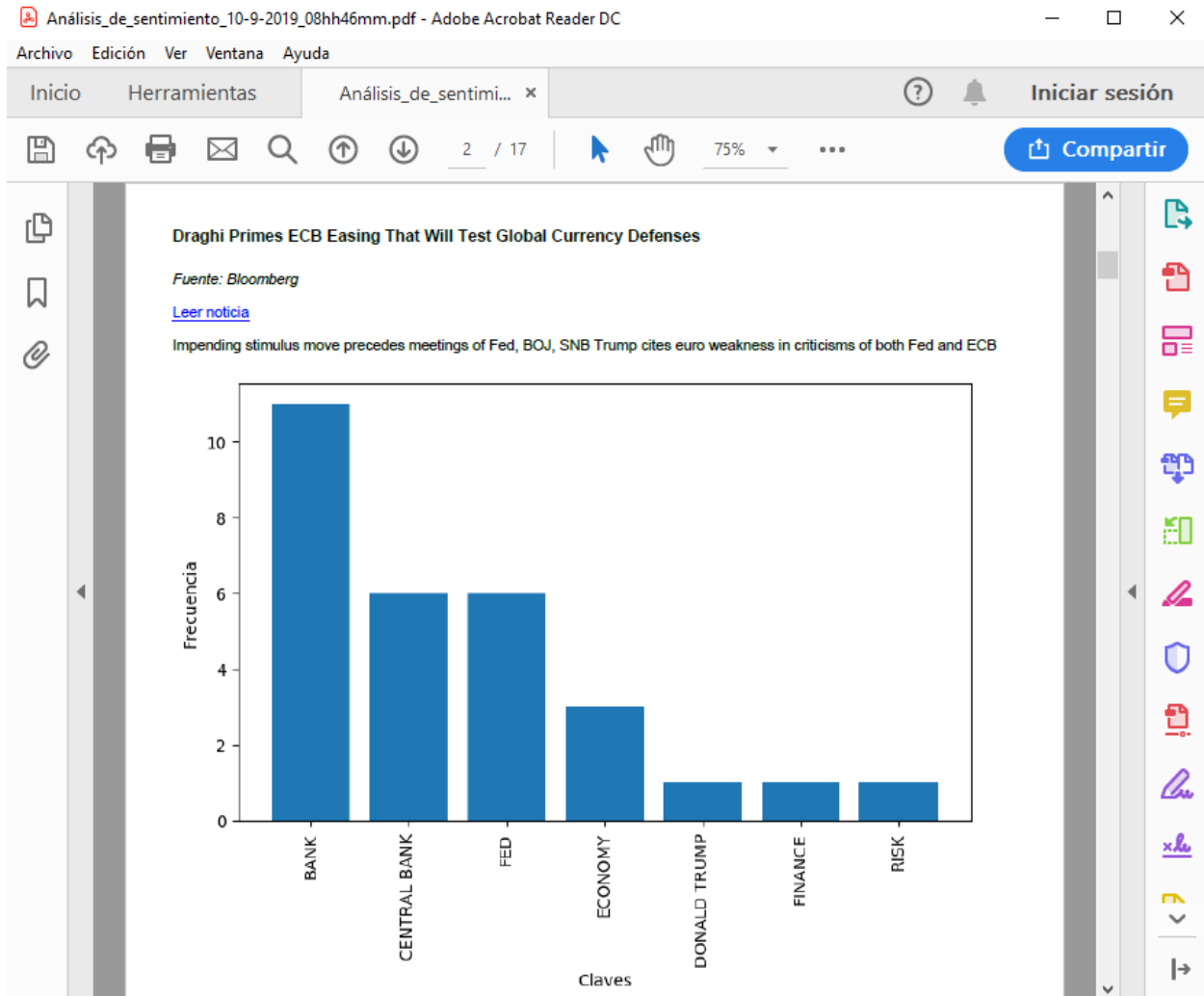
Opciones de la ventana Principal

En la parte superior derecha de esta ventana tendrá una lista desplegable con la etiqueta “Documentos”, la cual contiene una serie de botones que le ayudarán a acceder a las carpetas o documentos relacionados con el programa.

La ventana contiene a demás diversas opciones referentes a la ejecución del Subproceso webscraping:

- **Frecuencia de escaneo:** El programa realizará un escaneo de los sitios web establecidos periódicamente. Esta periodicidad, en minutos, es definida en este campo. El campo acepta un mínimo de 1 minuto y un máximo de 99 minutos.

- Adjuntar análisis: Cuando el programa analiza los datos de los sitios web, este documentará los resultados para las noticias relevantes en un reporte en formato pdf.



Si lo desea, dicho reporte puede ser enviado como elemento adjunto en la mensajería de correo electrónico, para ello debe presionar el botón “Adjuntar” hasta que el mismo se muestre en color **celeste**. Seguidamente, de estar habilitado, hacer click en el botón “Guardar cambios”.

- Olvidar bitácora: Al presionar este botón y confirmar la acción, serán eliminados todos los registro de la Bitácora del programa.

Otras opciones:

- Guardar cambios: El botón Guardar cambios se encontrará siempre inicialmente inhabilitado. Cuando haya un cambio en las configuraciones de esta ventana el botón se habilitará.
Importante: Los cambios en las configuraciones de la ventana Principal solamente tendrán efecto tras haber presionado el botón Guardar cambios. El mismo volverá entonces a su estado original (inhabilitado).
- Volver: Regresa a la Ventana inicial.

Ventana Fuentes de información

En la ventana Fuentes de información se le mostrará un listado de etiquetas con los nombres de las fuentes información disponibles para ser procesadas.



Al lado derecho de cada una de estas etiquetas habrá un botón que le permitirá activar o desactivar el escaneo de la correspondiente fuente de información.

Añadir más fuentes de información:

Cada fuente de información requiere un procesamiento distinto, debido a que no todos los sitios web están organizados bajo la misma estructura HTML ni implementan los mismos controles anti-robots. En otras palabras, cada una de las fuentes de información disponibles en el programa representa una programación

diferente. Por lo tanto, para añadir una nueva, debe ponerse en contacto con el Desarrollador quien se encargará de la implementación de una nueva fuente de información disponible en el programa.

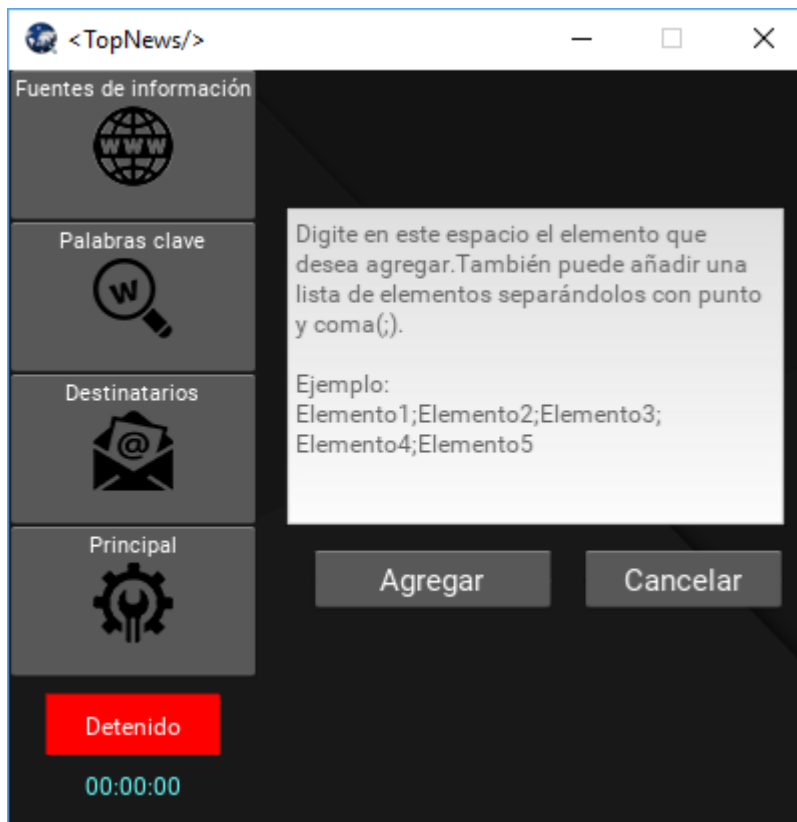
Ventana Palabras clave



La venta Palabras clave le permite visualizar y administrar (agregar o eliminar) las claves que utilizará el programa para determinar la relevancia de los items escaneados durante el Subproceso webscraping. En la parte central de la ventana se mostrará una lista con las palabras clave definidas por el usuario en orden alfabético.

Añadir más palabras clave:

Para añadir una nueva palabra clave haga click en el botón de color verde con el símbolo de adición(+) ubicado en la parte derecha de la ventana.

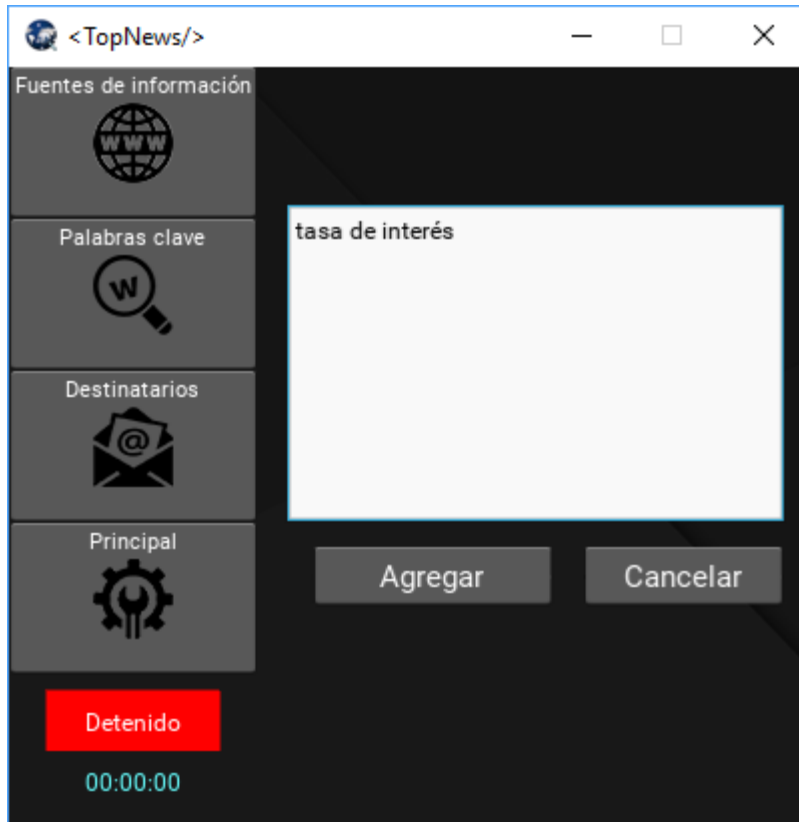


Tras hacer click en el botón mencionado en el párrafo anterior, se le llevará a una nueva ventana en donde podrá introducir los valores que desee añadir.

Restricciones:

- La longitud del valor(palabra clave) que esté añadiendo no debe ser inferior a 3 caracteres ni superior a 40.
- Los valores que ya se encuentren almacenados no serán añadidos nuevamente.

Las palabras clave que sean añadidas serán procesadas por un servicio de *GoogleTranslate* de manera que toda palabra en idioma español será almacenada junto a su contraparte en inglés y viceversa. Esto permite al programa analizar la información de los sitios web que se encuentren en alguno de estos dos idiomas. Este proceso requiere conexión a internet.



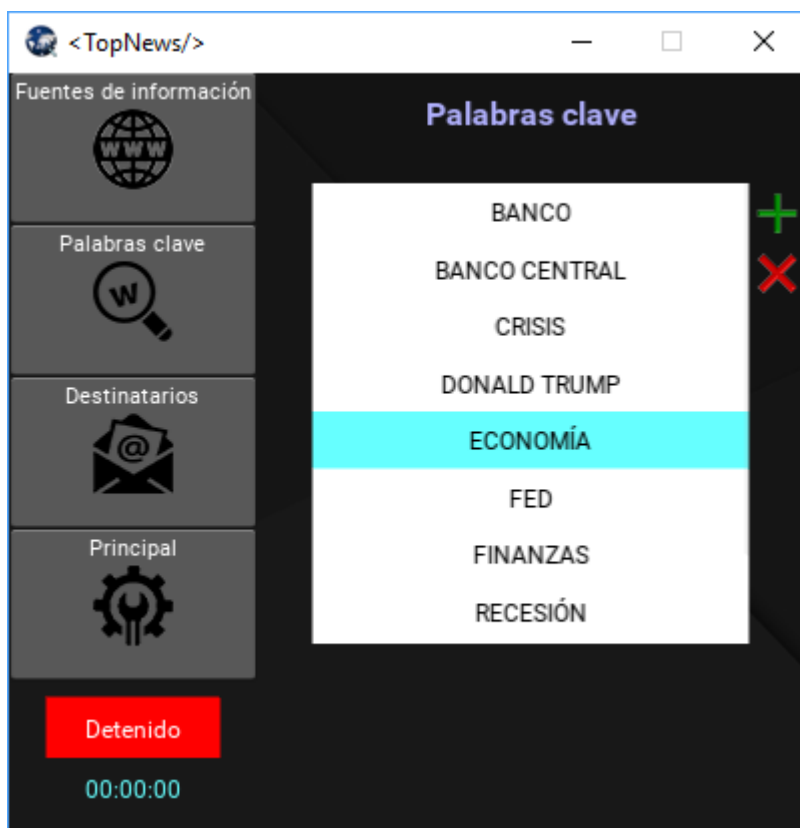
Para verificar la traducción de las palabras clave durante el proceso de añadido puede consultar el Log del programa. Ejemplo:

```
Log.txt - Notepad
File Edit Format View Help
08/09/2019 21:26:07|INICIO DE APLICACIÓN
08/09/2019 21:26:16|Validando login Office365 . . .
08/09/2019 21:26:17|Servidor Office365 abierto.
08/09/2019 21:26:23|(535, b'5.7.3 Authentication unsuccessful [BN6PR1201CA0005.namprd12.prod.outlook.com]')
08/09/2019 21:26:23|Cerrando servidor Office365.
08/09/2019 21:26:23|Servidor Office365 cerrado.
08/09/2019 21:27:17|Validando login Office365 . . .
08/09/2019 21:27:18|Servidor Office365 abierto.
08/09/2019 21:29:59|Cambios guardados en configuración.
08/09/2019 21:30:03|Cambios guardados en configuración.
08/09/2019 21:34:57|MÓDULO DE TRADUCCIÓN: Palabra='TASA DE INTERÉS' Resultado= Detected(lang=es, confidence=1.0)
08/09/2019 21:34:59|ELEMENTO AÑADIDO A CLAVES: 'TASA DE INTERÉS','INTEREST RATE'
08/09/2019 21:34:59|ELEMENTOS 'REV' ELIMINADOS DE BITÁCORA
```

Consejos importantes para añadir valores

- Puede añadir un valor individual o un conjunto de valores separados por punto y coma(;).
- Si no desea traducir un valor, puede añadirlo “en su estado puro”, para ello coloque el símbolo numeral(#) a la parte izquierda del valor. Por ejemplo: si usted desea añadir las siglas FED, refiriéndose a la Reserva Federal de los Estados Unidos, usted deberá incluir el ítem de la siguiente forma: #FED. De lo contrario la traducción interpretará FED como una palabra en inglés que se refiere a la conjugación del verbo “FEED”, que significa “ALIMENTAR”.

De ocurrir algún error durante la inclusión de un valor debido al incumplimiento de las restricciones previamente mencionadas, o por causas ajenas al programa (por ejemplo fallo en la conexión a internet), el mismo será registrado en el Log del programa y, en algunos casos, aparecerá una notificación en pantalla indicando el detalle.



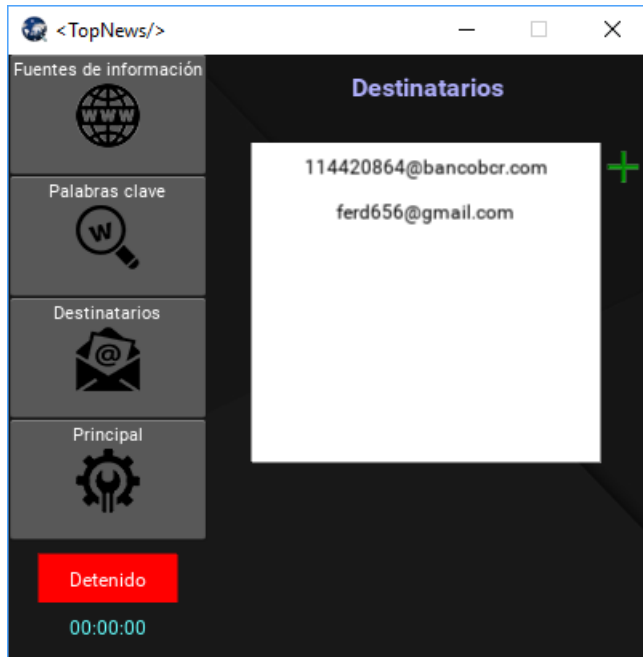
Eliminar palabras clave:

Para eliminar alguna de las palabras clave registradas, haga click en la lista sobre el valor que desea eliminar. Seguidamente aparecerá al lado derecho de la lista un nuevo botón de color rojo con la letra equis(**X**). Al presionar dicho botón y confirmar la acción el valor será eliminado.

Cada vez que se añade o elimina una palabra clave, el programa interpreta que cambia la forma en la que se evalúa la relevancia del contenido en las fuentes de información y por lo tanto elimina todos los elementos con estado REV de la Bitácora.

Ventana Destinatarios

En la ventana Destinatarios usted podrá visualizar y administrar (agregar o eliminar) la lista de direcciones de correo electrónico a las cuales el programa distribuirá la información de los diferentes elementos analizados durante el Subproceso webscraping.



Tanto el proceso de añadir como el de eliminar se manejan de la misma forma que en la ventana Palabras clave. Con la excepción de que los elementos de correo electrónico no son sujetos a traducción.

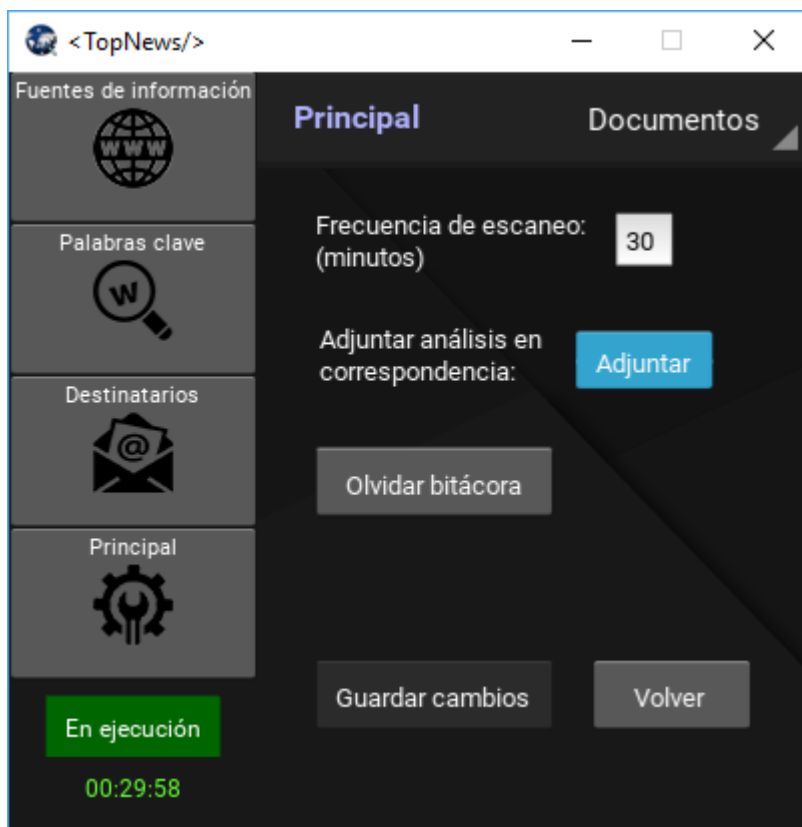
Restricciones:

- La longitud del valor no debe superar los 40 caracteres.
- El valor será corroborados para cumplir con el estándar de una dirección de correo electrónico válida. De incumplir esta validación el valor no será añadido.

Puede verificar los detalles del proceso de añadido o eliminación en el Log del programa. Este proceso no tendrá efecto en Bitácora.

Ejecución del Subproceso principal

En la esquina inferior izquierda del programa podrá visualizar el botón que ejecuta el Subproceso principal. Al iniciar el programa, este botón se muestra en color rojo y con la leyenda “Detenido”, al hacer click en el botón este intercambiará propiedades entre: color rojo con leyenda de “Detenido” y color verde y leyenda “En ejecución”.



Mientras el botón muestre la leyenda “En ejecución” se estará ejecutando el ciclo correspondiente al Subproceso principal.

Durante la fase 1 del Subproceso principal, se actualizará un contador regresivo de tiempo que se ubica bajo el botón. El contador iniciará con los minutos definidos como frecuencia de escaneo y finalizará en cero. Una vez finalizado el contador, dará inicio a la fase 2 del Subproceso principal, que consiste en la ejecución del Subproceso webscraping. En cualquier momento durante la fase 1 del Subproceso principal usted

puede detener la ejecución haciendo click en el botón y regresándolo a su estado original (con leyenda “Detenido”).

Durante la fase 2 del Subproceso principal, el programa excluirá del análisis todos aquellos elementos que se encuentren registrados en Bitácora. Al inicio de la fase 2 se eliminarán de la Bitácora todos aquellos elementos con estado PEND, de esa forma podrán ser reprocesados aquellos elementos que quedaron pendientes de finalizar alguna ejecución anterior.

Durante la fase 2 y 3 del Subproceso principal se ocultará la ventana del programa.

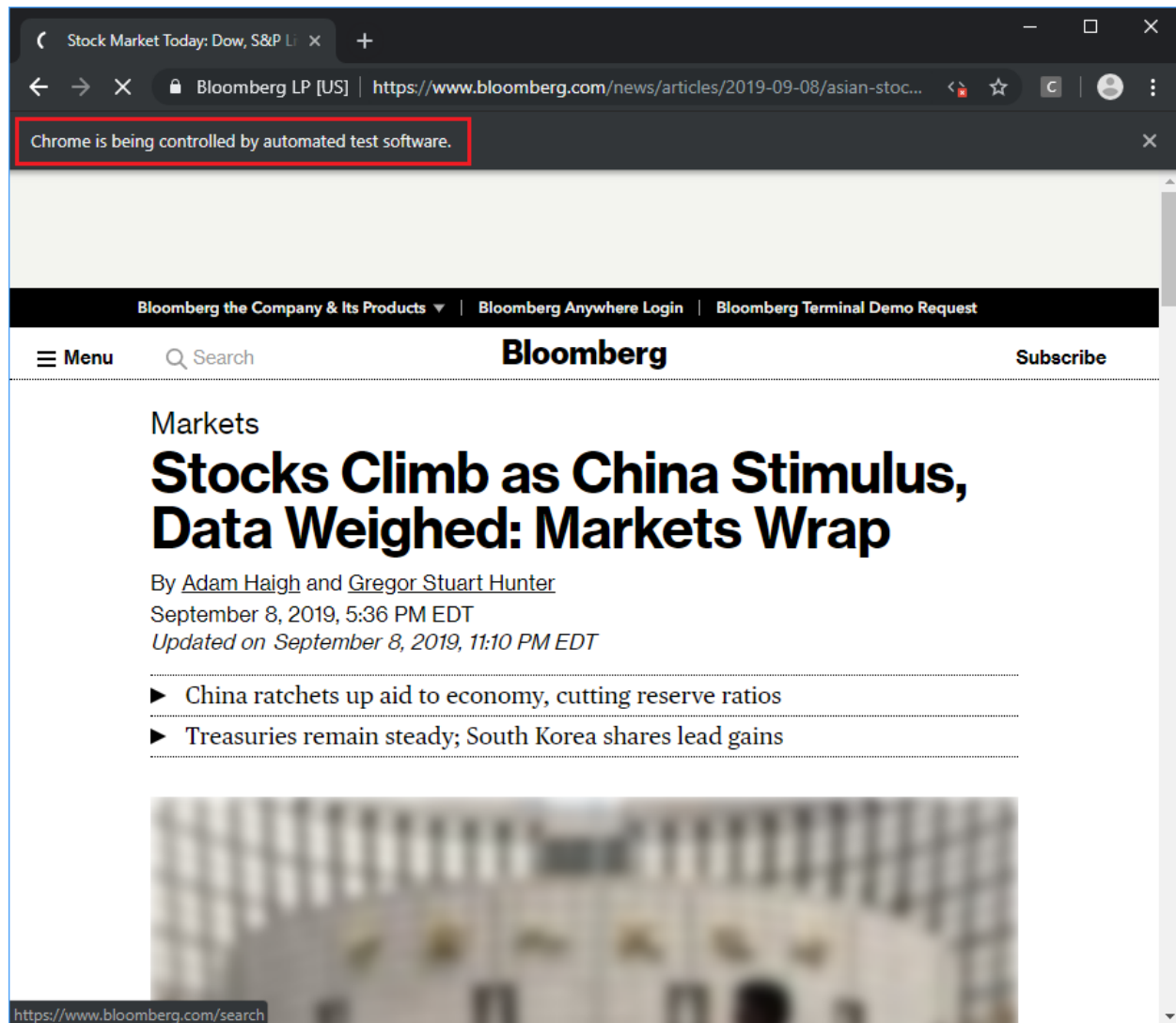
Fase 2: Subproceso webscraping

El Subproceso webscraping se encargará de la búsqueda, extracción y análisis de la información.

Es posible que el subproceso requiera un webdriver para acceder a la información de algunas fuentes. Por defecto los webdriver necesarios estarán incluidos en el programa, sin embargo los mismos tienden a sufrir actualizaciones a través del tiempo. Un webdriver desactualizado puede ocasionar que el escaneo de una o varias fuentes de información deje de funcionar.

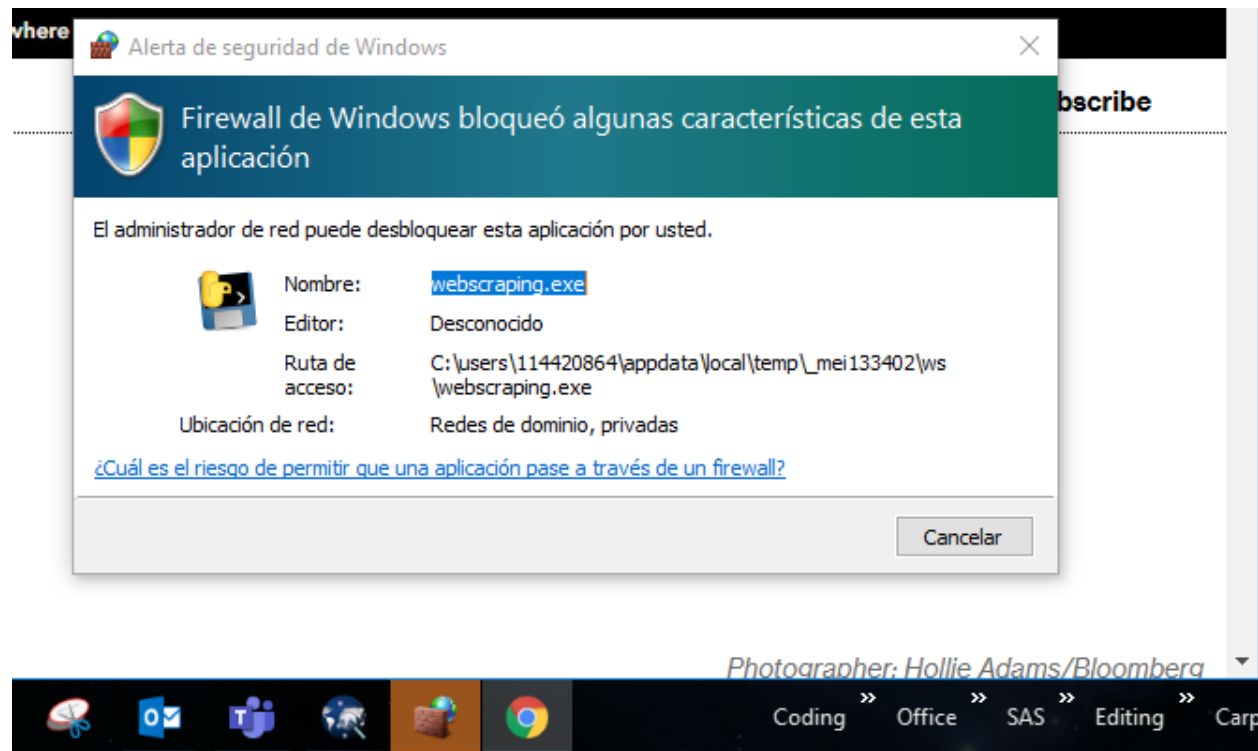
Es por esa razón, que el programa siempre buscará el webdriver en el directorio de su ejecutable. Esto permite que el usuario descargue la última versión del webdriver y lo coloque en el directorio del ejecutable del programa, y de esta forma evitar problemas de obsolescencia. Si el webdriver no es encontrado en el directorio del ejecutable del programa, este utilizará el webdriver incorporado por defecto.

Durante la fase 2, es posible que aparezca una ventana adicional de navegador en la que se puede observar una interacción por parte del programa.



Es recomendable no cerrar dicha ventana en caso de que aparezca, ya que eso interrumpiría el escaneo del contenido de esas páginas y por lo tanto no se analizará la información completa. Sin embargo esto no interrumpirá la ejecución del Subproceso principal.

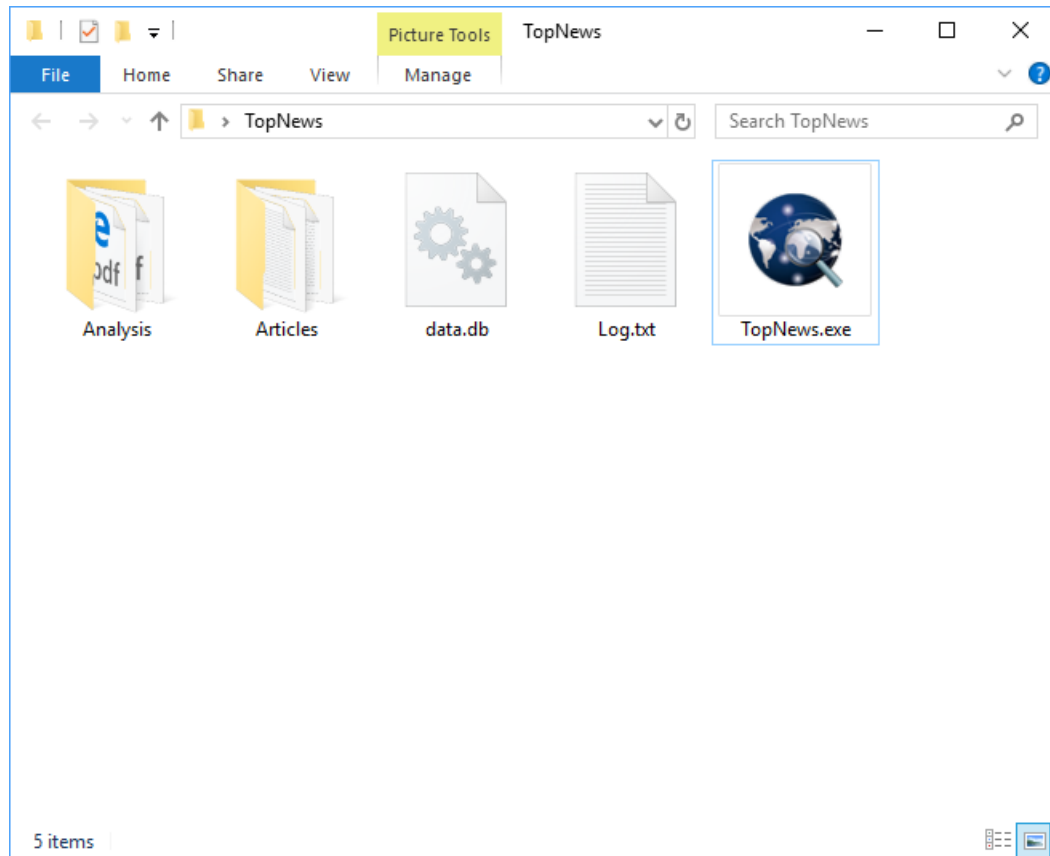
También es posible que durante la ejecución aparezca una Alerta de seguridad de Windows, como la que se muestra a continuación:



En ese caso, usted puede ignorar la advertencia y simplemente cerrarla, esto no afectará el funcionamiento del programa.

Almacenamiento de datos

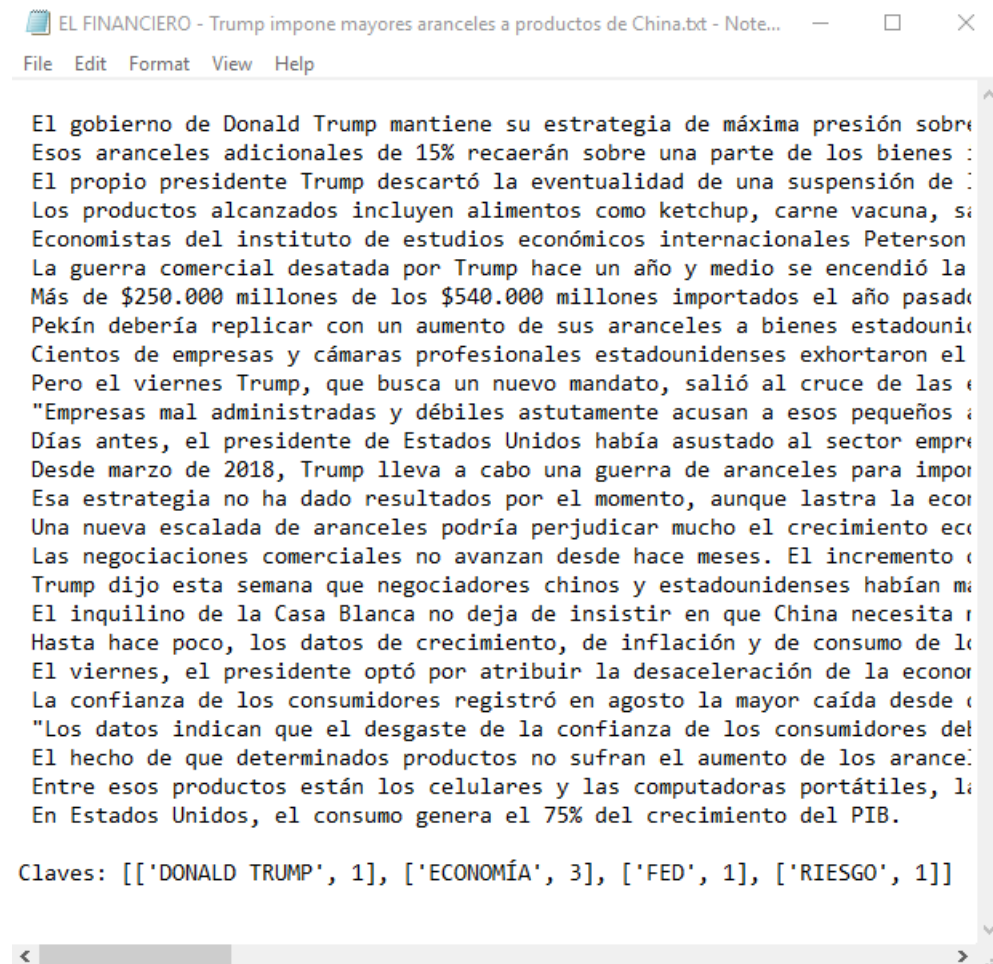
Durante la ejecución del Subproceso principal, el programa almacenará los reportes de análisis en formato pdf en la carpeta Analysis, y el contenido de cada uno de los elementos relevantes analizados en la carpeta Articles.



Los ficheros de texto contenidos en la carpeta Articles tendrán un nombre compuesto por: nombre de la fuente de la información y, separado por un guión, el título del ítem analizado.

El contenido de dichos ficheros consiste en:

- El texto del ítem analizado.
- El resultado de la frecuencia por palabra clave encontrada.



```
EL FINANCIERO - Trump impone mayores aranceles a productos de China.txt - Note...
File Edit Format View Help

El gobierno de Donald Trump mantiene su estrategia de máxima presión sobre
Esos aranceles adicionales de 15% recaerán sobre una parte de los bienes :
El propio presidente Trump descartó la eventualidad de una suspensión de :
Los productos alcanzados incluyen alimentos como ketchup, carne vacuna, sa
Economistas del instituto de estudios económicos internacionales Peterson
La guerra comercial desatada por Trump hace un año y medio se encendió la
Más de $250.000 millones de los $540.000 millones importados el año pasado
Pekín debería replicar con un aumento de sus aranceles a bienes estadouni
Cientos de empresas y cámaras profesionales estadounidenses exhortaron el
Pero el viernes Trump, que busca un nuevo mandato, salió al cruce de las «
"Empresas mal administradas y débiles astutamente acusan a esos pequeños «
Días antes, el presidente de Estados Unidos había asustado al sector empre
Desde marzo de 2018, Trump lleva a cabo una guerra de aranceles para impor
Esa estrategia no ha dado resultados por el momento, aunque lastra la econ
Una nueva escalada de aranceles podría perjudicar mucho el crecimiento eco
Las negociaciones comerciales no avanzan desde hace meses. El incremento «
Trump dijo esta semana que negociadores chinos y estadounidenses habían ma
El inquilino de la Casa Blanca no deja de insistir en que China necesita r
Hasta hace poco, los datos de crecimiento, de inflación y de consumo de la
El viernes, el presidente optó por atribuir la desaceleración de la econor
La confianza de los consumidores registró en agosto la mayor caída desde «
"Los datos indican que el desgaste de la confianza de los consumidores del
El hecho de que determinados productos no sufran el aumento de los arance
Entre esos productos están los celulares y las computadoras portátiles, la
En Estados Unidos, el consumo genera el 75% del crecimiento del PIB.

Claves: [['DONALD TRUMP', 1], ['ECONOMÍA', 3], ['FED', 1], ['RIESGO', 1]]
```

Tip: Estos ficheros pueden ser de utilidad al usuario para ejecutar otros análisis de texto.

Envío de mensajería electrónica

En la fase 3 del Subproceso principal se generará la mensajería electrónica. Esta característica tendrá un comportamiento distinto dependiendo de la plataforma que haya seleccionado en el Ventana inicial.

Office365 vs Outlook

Envío por Office365

En caso de que haya seleccionado Office365 como plataforma de envío, la mensajería se envía de inmediato.

Ventajas:

- No necesitaría ninguna interacción manual, por lo que no sería necesario estar pendiente del programa.

Desventajas:

- No se le facilitaría revisar el contenido del mensaje de correo electrónico antes que se envíe.

Envío por Outlook

Para el caso de Outlook, el mensaje de correo electrónico se despliega y el programa esperará hasta que dicho mensaje sea enviado o cerrado para poder continuar.

Ventajas:

- Le permitiría revisar fácilmente el contenido del mensaje de correo electrónico antes que el mismo sea enviado.

Desventajas:

- El proceso quedaría detenido hasta que el mensaje sea enviado o cerrado, por lo que el tiempo de ejecución del ciclo Subproceso webscraping se extendería hasta recibir la respectiva interacción del usuario.

A continuación se muestra un ejemplo de un mensaje de correo electrónico generado por el programa:



Manejo de errores

El programa cuenta con rigurosos mecanismos de control sobre sus propios procesos, abarcando casi cualquier posible inconveniente que se pueda dar durante la ejecución de los mismos.

Para la mayoría de los casos, el programa mostrará mensajes por pantalla detallando sobre la eventual incidencia.

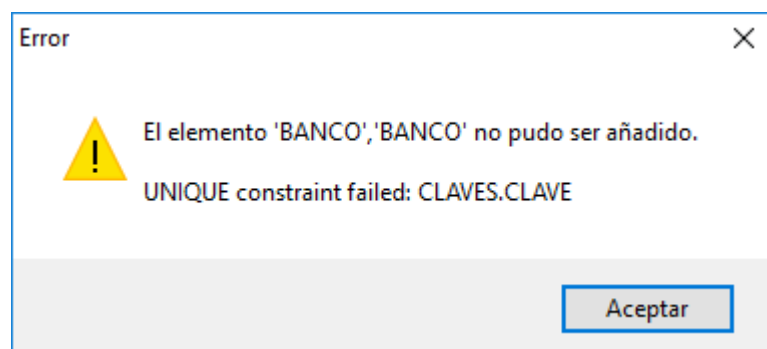
Los errores críticos podrán ser identificados mediante el ícono similar al siguiente:



Los errores leves o advertencias (“warnings”), pueden ser identificados mediante el ícono similar al siguiente:

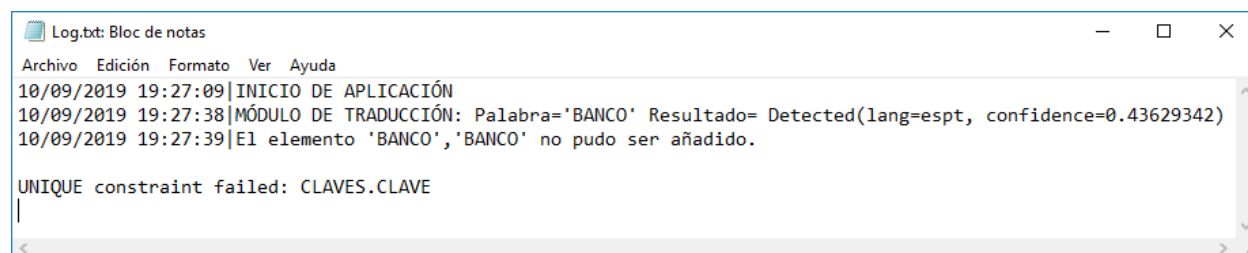


Un ejemplo de un error leve es el siguiente:



En este caso se indica que se está intentando añadir la palabra clave BANCO pero la misma ya está perviamente añadida.

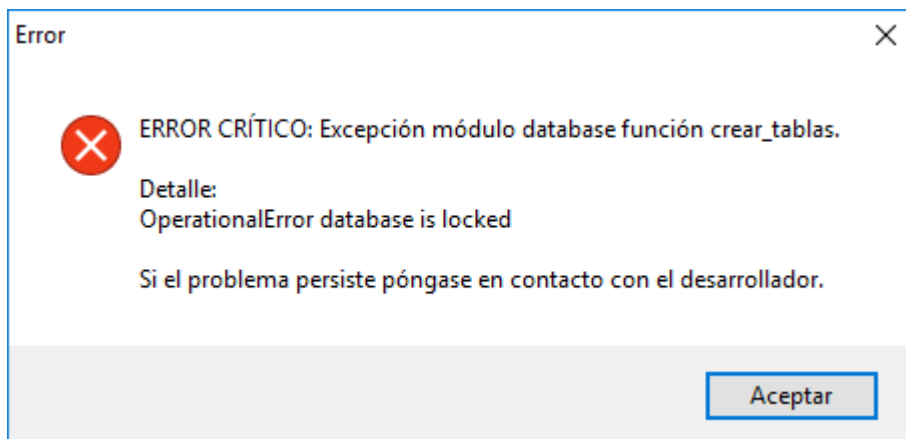
En el Log se guardará registro del error:



Hay casos en los que el programa no muestra el mensaje en pantalla, principalmente con el fin de no interrumpir la ejecución de algún subproceso. Por lo tanto estos errores serán más difíciles de detectar en comparación con aquellos que muestran mensaje en pantalla. Se recuerda entonces que el programa dejará un registro de lo ocurrido durante los procesos en el Log, por lo tanto se recomienda realizar una revisión periódica del Log durante cada ejecución del programa.

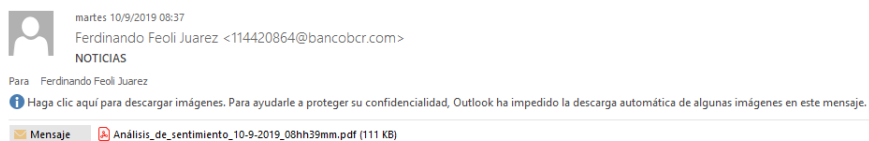
Errores conocidos

- **OperationalError database is locked**

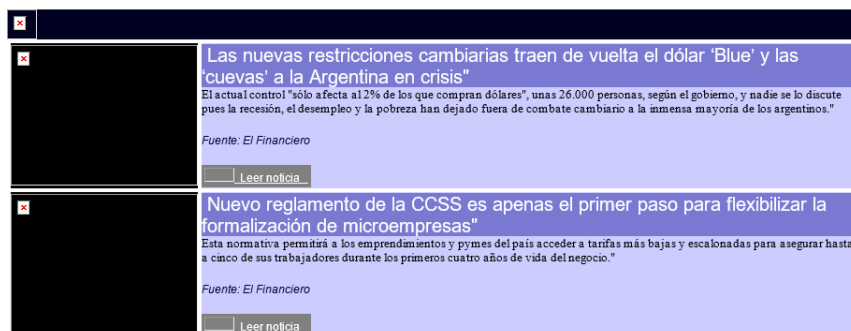


Este error se produce porque la Base de datos está siendo utilizada por otro programa. Para solucionarlo cierre la conexión a la base de datos desde el programa que esté utilizando. Si el problema persiste cierre todos los programas (incluyendo TopNews) y vuelva a intentar.

- **Errores en la visualización de información**



Puede darse el caso en el que las imágenes del cuerpo HTML en los mensajes de correo electrónico no carguen debidamente.



Estos errores se producen por temas de redes o seguridad de sus aplicaciones Outlook/Office365. Generalmente se produce cuando tiene bloqueada la descarga automática de imágenes para los cuerpos de los correos electrónicos o se pierde la

conexión a internet durante la ejecución del las fases 2 y/o 3 del Subproceso principal.

Para solucionarlo, en primera instancia verifique que la descarga de imágenes no se encuentre bloqueada en su plataforma, para ello puede ver [artículo de soporte de office](#). Si esto no soluciona el problema, reinicie el programa y compruebe su conexión a internet durante la próxima ejecución. Verifique además que posea acceso a las fuentes de información y a Microsoft OneDrive. Si no tuviera acceso a una de ellas póngase en contacto con su administrador de red para que le facilite el acceso.

Nota: Deberá borrar la Bitácora mediante la correspondiente opción en la página Principal para volver a enviar ítems que ya fueron registrados en Bitácora con estado ENV.

Reportar errores

Para los casos en que los errores sean persistentes o el programa tenga un comportamiento que no sea el esperado puede ponerse en contacto con el Desarrollador. Para ello se pide:

- Enviar captura de pantalla del mensaje emitido o guardado en Log.
- Describir a detalle la situación y cómo/cuándo se produce.

Desarrollador

Para ponerse en contacto con el Desarrollador, visite: <https://www.linkedin.com/in/ferdinandfeoli>

```
_+88_
_+880_
_++88_
_++88_
_+880_++_
_+888_+88_
_++880_+88_
_++888_+++88_+++8_
_++8888_+++8880++88_+++88_
_+++8888+++8880++8888_++888_
_++888++8888+++888888++888_
_++88++8888++888888++888_
_++++++8888888888888888_
_++++++8888888888888888_
_++++++0008888888888888_
_++++++0000888888888888_
_++++++0008888888888888_
_++++++08888888888888_
_++++++08888888888888_
_++++++888888888888_
_++++++008888888888_
_++++++008888888888_
_++++++000888888888_
_<//////////>_
```