

GUÍA DE USUARIO *VERBUM*



<https://github.com/nuxa17/Verbum>

Índice

1. Introducción.....	2
2. Descripción general del sistema.....	2
3. Material necesario y requisitos previos.....	2
4. Funcionalidades del Sistema.....	3
4.1. Pantalla principal.....	3
4.1.1. Sección de información (Information).....	4
4.1.2. Sección de Configuración de búsqueda (Settings).....	4
4.1.3. Sección de Análisis (Analysis).....	5
4.1.4. Sección cargar archivo (Load text file).....	5
4.1.5. Sección Menú Superior (View).....	6
4.1.6. Sección Menú Superior (Help).....	7
4.2. Pantalla de análisis.....	8
4.2.1. Pestaña Criteria.....	9
4.2.2. Pestaña Collocations.....	9
4.2.2.1. Sección de Configuración de búsqueda (Settings):.....	10
4.2.2.2. Sección Consultas (Queries).....	11
4.2.3. Pestaña Sentiment.....	12
5. Ventana de guardado.....	13
6. Preguntas frecuentes.....	14
Descripción de las tags.....	15
Criterios Manipulativos.....	16

1. Introducción

En este manual se describen las funcionalidades de la aplicación, con el objetivo de que el usuario que interactúa con ella, tenga un material donde resolver sus dudas en cuanto al funcionamiento de la misma. El objetivo del presente es que pueda sacar el mayor provecho posible a la herramienta descrita.

2. Descripción general del sistema

Verbum es una herramienta para el análisis de textos en inglés que hace uso de técnicas de Natural Language Processing (NLP). Su principal objetivo es analizar el grado de manipulación que presentan dichos textos. El usuario puede cargar textos en distintos formatos y configurar ciertos parámetros para que el programa los analice de acuerdo a diferentes criterios.

El usuario dispone además de una serie de herramientas adicionales como un buscador de n-gramas y dos analizadores de sentimiento en el texto.

3. Material necesario y requisitos previos

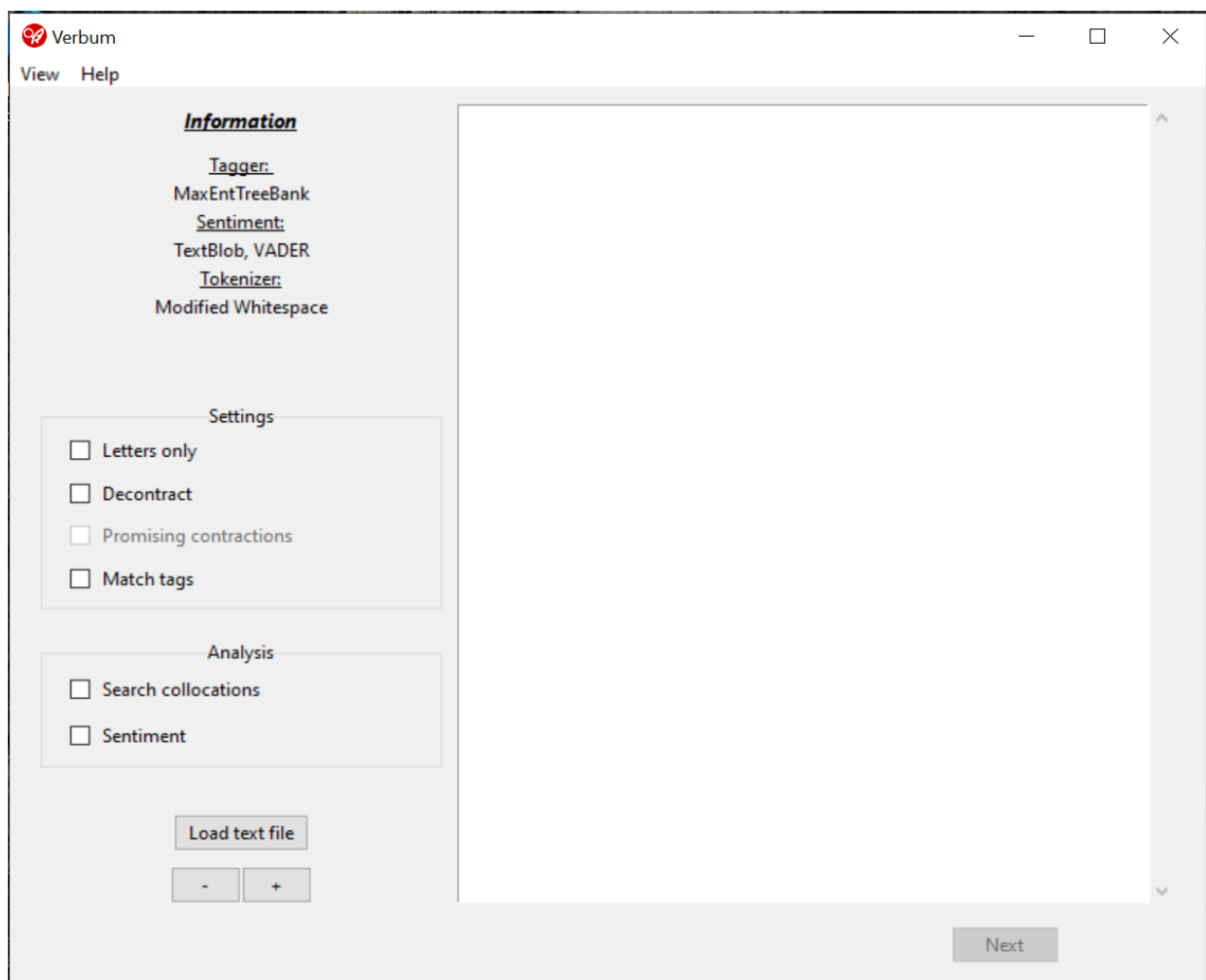
Para el uso de Verbum no se requiere instalar ningún programa adicional. La aplicación se presenta de forma autocontenida, de manera que todo lo necesario está incluido dentro del propio ejecutable. Puede ejecutarse en los sistemas operativos Windows y macOS X.

4. Funcionalidades del Sistema

A continuación se describen las funcionalidades del sistema, divididas en las distintas pantallas con las que puede interactuar el usuario.

4.1. Pantalla principal

Esta es la pantalla principal de la aplicación, en la que podrá cargar un texto y modificar los distintos parámetros del análisis.



Pantalla Principal

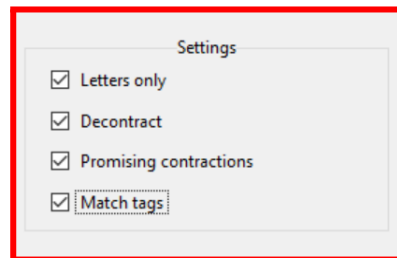
4.1.1. Sección de información (Information)

Aquí se pueden ver las distintas herramientas utilizadas para los análisis.



Panel de información.

4.1.2. Sección de Configuración de búsqueda (Settings)



Opciones de configuración de la búsqueda.

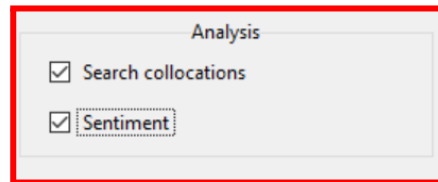
- **Letters only:** Recomendado. Elimina del texto los caracteres no alfanuméricos, a excepción de apóstrofes. En caso contrario se separarán las palabras por espacios, algo que puede influir mucho en el resultado.
Nota: para poder reconstruir palabras correctamente, todos los tipos de comillas quedan normalizados, y los guiones de palabras compuestas no son eliminados.
- **Decontract:** Recomendado. Expandir las contracciones Se puede configurar la “agresividad” en Promising contractions.
- **Promising contractions:** Si está marcado marcando, toma la mejor suposición para una contracción si hay varias opciones para ella. En caso contrario, se recuperan las partes "seguras".

Marcado: *he's* -> *he is* , pero podría ser *has* o *was*.

Sin marcar: *he's* -> *he*.

- **Match tags** : Utilizar clases de palabras ajustadas para mejorar la concordancia. Recomendado, pero puede dar lugar a algunos falsos positivos/negativos debido a la naturaleza de los etiquetadores y afectar así a las tasas de manipulación. **Nota:** para corregir un poco el error de los *taggers*, se suele buscar la categoría gramatical más general. Puedes consultar los distintos tipos de *tags* [aquí](#).

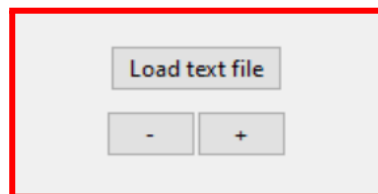
4.1.3. Sección de Análisis (Analysis)



Herramientas extras para el análisis.

- **Search collocations:** Activar la pestaña de n-gramas después del análisis.
- **Sentiment:** Activar la pestaña Sentiment después del análisis.

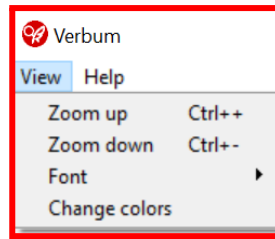
4.1.4. Sección cargar archivo (Load text file)



Cargar Archivo.

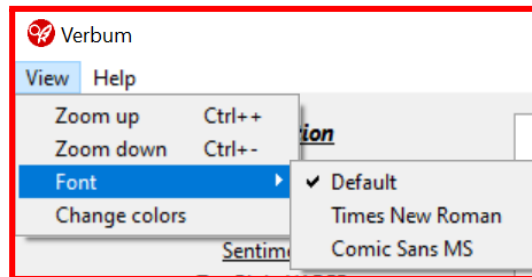
- **Load text file:** Cargar un archivo de texto. Los formatos soportados son: *.txt*, *.doc/docx*, *.rtf* y *.pdf*.
- **+/-:** Aumentar/disminuir el tamaño de visualización de la letra al texto cargado.

4.1.5. Sección Menú Superior (View)



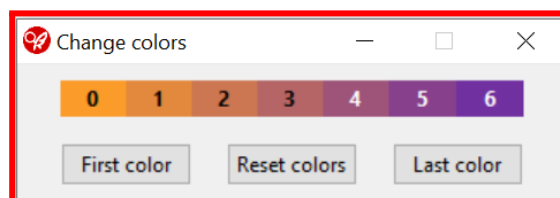
Opciones de View.

- **Zoom up/down:** Aumenta/disminuye el tamaño de la letra de los componentes literales de la aplicación.
- **Font:** Cambia el tipo de letra de los componentes literales de la aplicación. Se mostrará el siguiente panel para seleccionar el tipo de letra preferido.



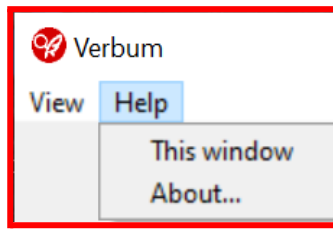
Selector de fuente.

- **Change colors:** Cambia el gradiente de colores para el resultado del análisis. Pulsando en *First color* se elige el color del extremo izquierdo. Pulsando en *Last color* se elige el color del extremo derecho. Si se pulsa en *Reset colors* se regresa a la configuración por defecto, que es la que se muestra en la figura. Las letras cambiarán en función de su color de fondo.



Configuración del gradiente

4.1.6. Sección Menú Superior (Help)



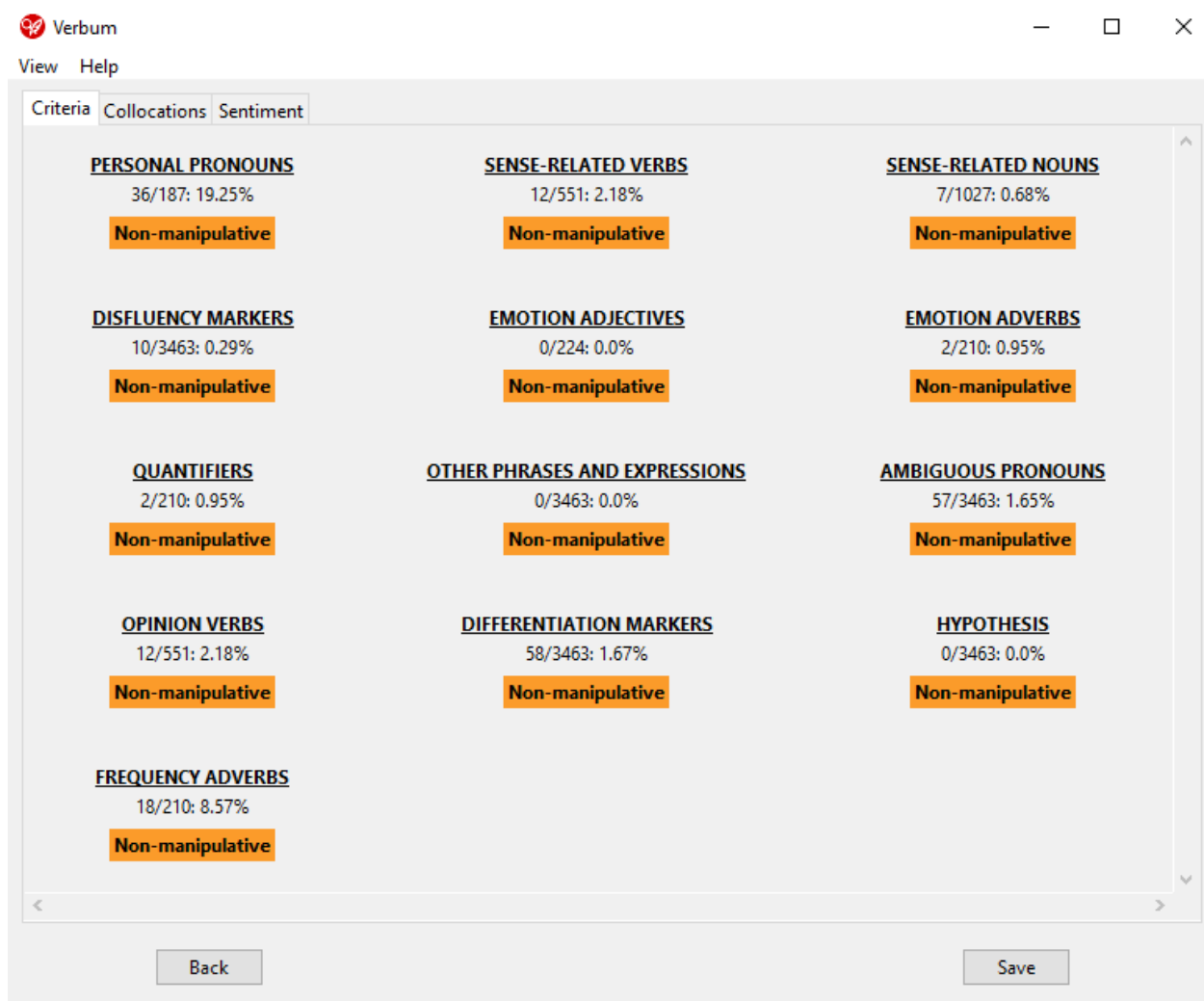
Ventanas de información

- **This window:** Información acerca de la pantalla actual, con una referencia a su página correspondiente en el manual.
- **About:** Información acerca de la aplicación, con referencias al manual completo y al repositorio de código.

Para realizar el análisis del texto tras ajustar los parámetros, se pulsa el botón *Next*, situado en la parte inferior derecha.

4.2. Pantalla de análisis

En esta pantalla se encuentran varias pestañas en dependencia de los checks seleccionados en la sección *Analysis* de la primera pantalla. A propósito de este manual se asumen marcados ambos. La pestaña *Criteria* siempre aparecerá, dado que es el objetivo principal de la aplicación.

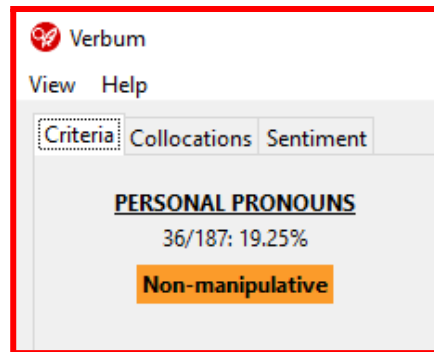


Pantalla de análisis.

Los botones inferiores permiten volver a la [pantalla principal](#) o [guardar](#) los resultados en un reporte final.

4.2.1. Pestaña Criteria

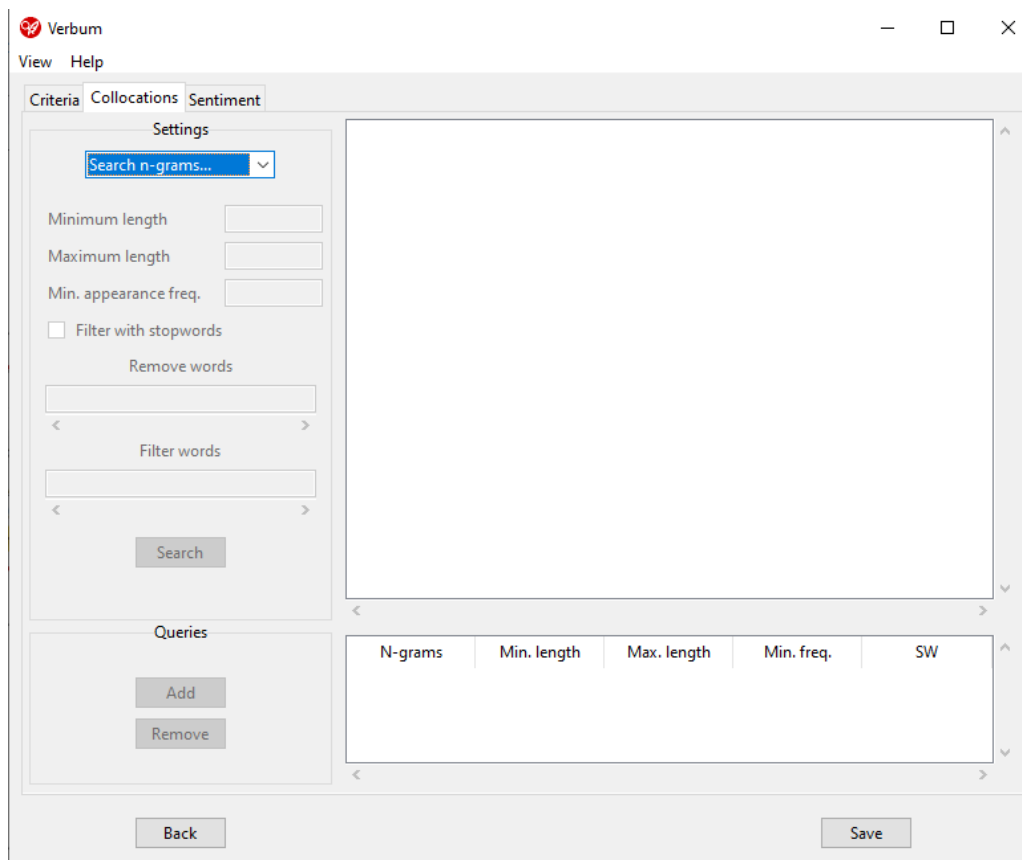
Es la pantalla completa mostrada anteriormente. Se muestra el grado de manipulación del texto analizado según los distintos criterios, por ejemplo *Personal Pronouns*. Puede consultar los diferentes criterios en esta [página](#).



Componentes del análisis manipulativo

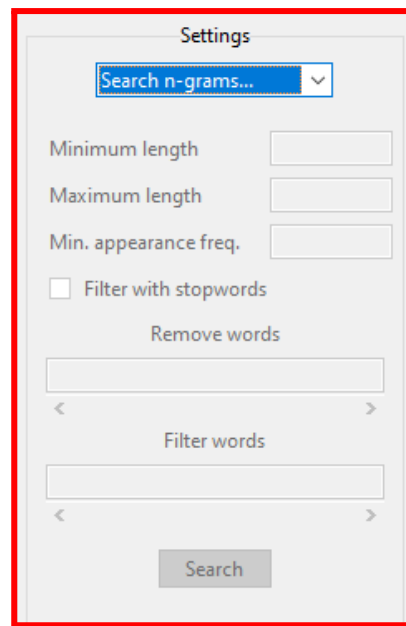
4.2.2. Pestaña Collocations

En esta pantalla se permite la búsqueda de collocations a partir de los n-gramas.



Pantalla de Collocations.

4.2.2.1. Sección de Configuración de búsqueda (Settings):

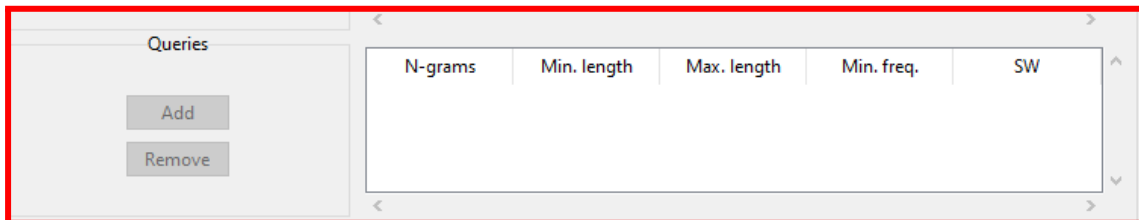
The image shows a 'Settings' dialog box with a red border. At the top is a dropdown menu labeled 'Search n-grams...' with a downward arrow. Below this are three input fields: 'Minimum length', 'Maximum length', and 'Min. appearance freq.'. There is a checkbox labeled 'Filter with stopwords'. Below the checkbox is a button labeled 'Remove words' and an input field. Below that is a button labeled 'Filter words' and another input field. At the bottom is a 'Search' button. The entire dialog box is enclosed in a red rectangular border.

Configuración de los n-gramas

- **Search n-grams:** Selecciona distintos tipos de n-gramas que quiera buscar. Pueden elegirse bigramas, trigramas y cuatrigramas.
- **Mínimum length:** Elige la longitud mínima de las palabras en los n-gramas. Ayuda a eliminar conectores lingüísticos y palabras sin significado útil. Solo admite valores enteros positivos.
- **Maximum length:** Elige la longitud máxima de las palabras en los n-gramas. Solo admite valores enteros positivos.
- **Min. appearance freq.:** Impone una frecuencia mínima de aparición de los n-gramas buscados. Solo admite valores enteros positivos.
- **Remove Stopwords:** Elimina las palabras contenidas en el corpus [Stopwords](#) de nltk. **Nota:** Esto excluirá palabras relativamente relevantes como pronombres. Se puede usar la opción *Remove words* como alternativa.
- **Remove Words:** Excluye n-gramas que contengan ciertas palabras. Deben separarse por espacios.
- **Filter Words:** Muestra los n-gramas que contienen alguna de estas palabras. Deben separarse por espacios.

4.2.2.2. Sección Consultas (Queries)

En este panel se pueden gestionar las distintas consultas realizadas y sus resultados para incluirlas dentro del reporte final. Para cada consulta se muestran varias de sus opciones de configuración.

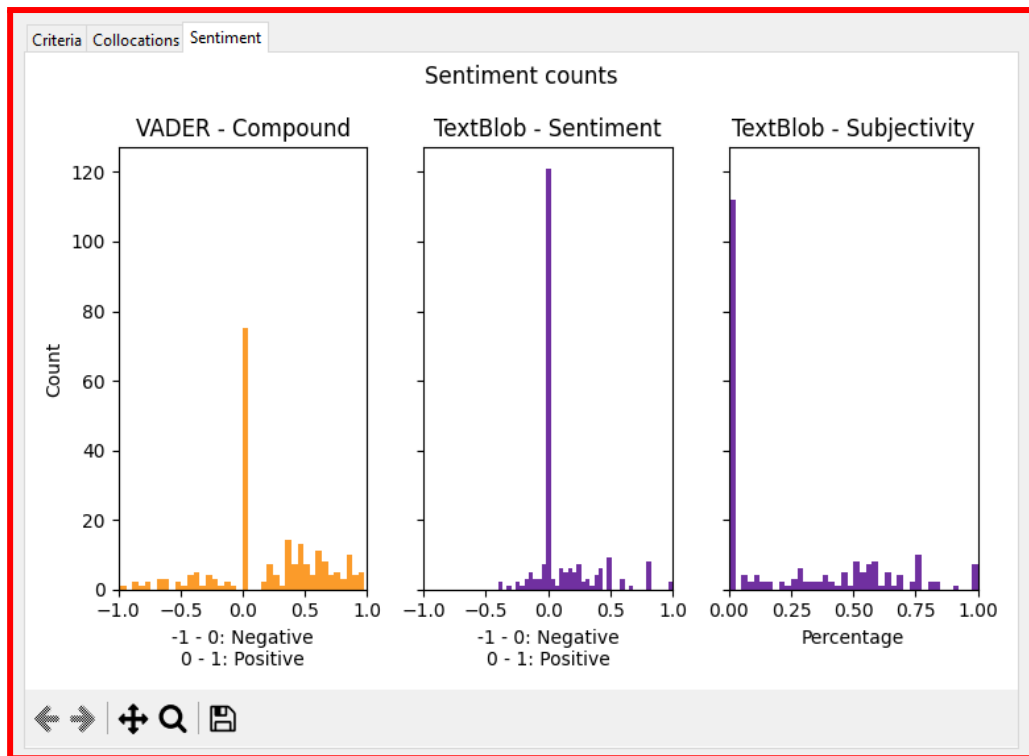


Panel de consultas.

- **Add:** Almacena la última búsqueda y sus resultados para incluirlos en el reporte final.
- **Remove:** Borra las búsquedas seleccionadas.

4.2.3. Pestaña Sentiment

En esta pestaña se encuentran las gráficas resultado de los análisis de sentimiento. Estos análisis se realizan para cada frase con dos herramientas por separado, *VADER* y *TextBlob*. De esta última también se incluye un reporte de subjetividad. Estas gráficas pueden incluirse en el reporte final y utilizan los colores principales.



Pantalla Sentiment

Las gráficas cuentan con las siguientes opciones por orden de aparición:

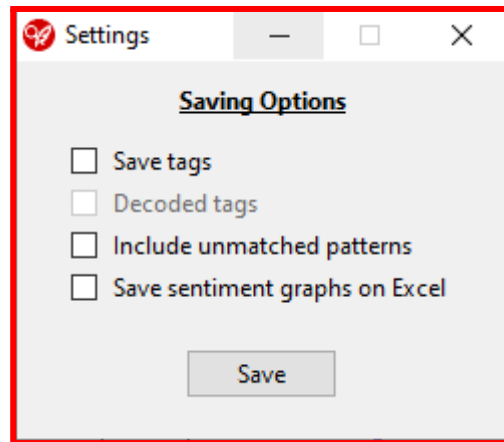


Opciones de navegación

- **Volver atrás/avanzar:** Moverse entre modificaciones.
- **Mover la gráfica :** Navegar dentro de las gráficas.
- **Zoom:** Hacer zoom a una sección de las gráficas.
- **Guardar resultado:** Guardar las gráficas como imagen PNG.

5. Ventana de guardado

Esta ventana dispone de varias opciones de configuración para el reporte final.



Opciones de guardado

- **Save tags:** Incluir una columna con las *tags* en el archivo de resultados.
- **Decode tags:** Mostrar la descripción de las *tags* en vez de su forma codificada. Disponible con *Save tags* marcado.
- **Include unmatched patterns:** Incluir los patrones sin ocurrencias en el texto (añadirá muchas filas vacías).
- **Save sentiment graphs on Excel:** Incluir las gráficas de los análisis de sentimiento.

6. Preguntas frecuentes

¿Por qué en las gráficas de sentimiento el número de ceros es tan alto?

En las gráficas se pueden observar que aparecen muchos resultados con valor cero. Esto se debe a que los analizadores buscan palabras concretas en las frases, y la mayoría de estas no incluyen dichas palabras. Por tanto, su sentimiento pasa a ser nulo.

¿Por qué se han elegido esos colores por defecto?

Los colores predeterminados tienen un alto contraste entre ellos y son fácilmente distinguibles para personas con cualquier tipo de daltonismo. Sin embargo, estos colores pueden cambiarse y un nuevo gradiente se creará a partir de ellos. Basados en este [artículo](#) de Lluís Montoliu.

¿Por qué algunas fuentes no cambian de tamaño?

Para poder ofrecer una experiencia adecuada a cada plataforma, se hace uso de diálogos nativos del sistema, cuya apariencia no se puede modificar.

¿Por qué no puedo analizar un archivo de Pages?

Pages es un formato de archivo propietario, por lo que dar soporte fuera de las aplicaciones propias de Apple es muy complicado. Sin embargo, puede convertir el archivo a un formato compatible con la aplicación. Los formatos soportados son: *.txt*, *.doc/docx*, *.rtf* y *.pdf*.

Descripción de las *tags*

CC: Conjunción coordinada.

CD: Cardinal.

DT: Determinante.

EX: ‘*there*’ existencial.

FW: Extranjerismo.

IN: Preposición o
conjunción subordinada.

JJ: Adjetivo.

JJR: Adjetivo comparativo.
Padre: JJ.

JJS: Adjetivo superlativo.
Padre: JJ.

LS: Marcador en una lista.

MD: Modal.

NN: Sustantivo en singular o ‘*mass*’.

NNS: Sustantivo en plural.
Padre: NN.

NNP: Nombre propio, singular.
Padre: NN.

NNPS: Nombre propio, plural.
Padre: NNS.

PDT: Predeterminante.
Padre: DT.

POS: Terminación posesiva; p.ej. ‘*s*’.

PRP: Pronombre personal.

PRP\$: Pronombre posesivo.

RB: Adverbio.

RBR: Adverbio, comparativo.
Padre: RB.

RBS: Adverbio, superlativo.
Padre: RB.

RP: *Particle*.

SYM: Símbolo.

TO: ‘*to*’.

UH: Interjección.

VB: Verbo, infinitivo.

VBD: Verbo, pasado.

VBG: Verbo, gerundio o
present participle.

VCN: Verbo, *past participle*.

VBP: Verbo, presente singular
(no 3ª persona).

VBZ: Verbo, presente singular
(3ª persona).

WDT: *Wh*- determinante.
Padre: DT.

WP: *Wh*- pronombre.

WP\$: *Wh*- pronombre posesivo.
Padre: WP.

WRB: *Wh*- adverbio.
Padre: RB.

***:** comodín, coincide con cualquiera.

-NONE-: categoría desconocida.

Los taggers suelen utilizar la composición de las frases para encuadrar mejor las categorías gramaticales, clasificando peor ciertas palabras que otras con frecuencia. Asimismo, no se detectan estructuras como los phrasal verbs, por lo que cada palabra individual necesitaría una tag específica. Se han añadido las que mejores resultados han dado, lo que ayuda a reducir el error sin desviarse demasiado.

Criterios Manipulativos

Criterio	Descripción	Porcentajes
PERSONAL PRONOUNS	Pronombres personales en 3ª persona frente al resto.	≥ 30%: Muy manipulativo ≥ 35%: Altamente manipulativo
SENSE-RELATED VERBS	Verbos relacionados con los sentidos frente al resto.	≥ 30%: Ligeramente manipulativo ≥ 40%: Muy manipulativo ≥ 50%: Altamente manipulativo
SENSE-RELATED NOUNS	Sustantivos relacionados con los sentidos frente al resto.	≥ 20%: Manipulativo ≥ 30%: Altamente manipulativo ≥ 40%: Extremadamente manipulativo
DISFLUENCY MARKERS	Densidad de palabras y expresiones de disfluencia.	≥ 15%: Manipulativo ≥ 20%: Muy manipulativo ≥ 30%: Altamente manipulativo
EMOTION ADJECTIVES	Adjetivos relacionados con las emociones frente al resto.	≥ 20%: Ligeramente manipulativo ≥ 30%: Manipulativo ≥ 35%: Altamente manipulativo
EMOTION ADVERBS	Adverbios relacionados con las emociones frente al resto.	≥ 20%: Ligeramente manipulativo ≥ 30%: Manipulativo ≥ 35%: Altamente manipulativo
QUANTIFIERS	Adverbios de cantidad frente al resto.	≥ 20%: Ligeramente manipulativo ≥ 30%: Manipulativo ≥ 35%: Altamente manipulativo
OTHER PHRASES AND EXPRESSIONS	Densidad de expresiones para desviar la atención.	≥ 10%: Manipulativo ≥ 20%: Altamente manipulativo
AMBIGUOUS PRONOUNS	Densidad de pronombres ambiguos.	≥ 25%: Subjetivo ≥ 30%: Muy manipulativo ≥ 35%: Altamente manipulativo
OPINION VERBS	Verbos de opinión frente al resto.	≥ 20%: Subjetivo ≥ 30%: Ligeramente manipulativo ≥ 40%: Muy manipulativo ≥ 50%: Altamente manipulativo
DIFFERENTIATION MARKERS	Densidad de marcadores de diferenciación.	≥ 50%: Ligeramente manipulativo ≥ 60%: Altamente manipulativo
HYPOTHESIS	Densidad de expresiones de hipótesis.	≥ 30%: Subjetivo ≥ 35%: Manipulativo ≥ 40%: Altamente manipulativo
FREQUENCY ADVERBS	Adverbios de frecuencia absolutos frente al resto.	≥ 40%: Manipulativo ≥ 50%: Altamente manipulativo