

NewsAPI-AR

Un servicio para análisis y visualización de noticias de Argentina

72.45 - Proyecto Final

Tutor: Lic. Diego Ariel Aizemberg

Integrantes:

Oliver Balfour (55177)

Marcos Ariel Atar (57352)

Fecha: Mayo 2022

Introducción

El objetivo del proyecto es desarrollar un servicio web gratuito que **pone a disposición información acerca de noticias de Argentina.**

Para ello, se creó una aplicación que **interpreta archivos RSS de noticias de los medios más importantes y almacena en una base de datos la información más relevante.**

Las noticias almacenadas son enriquecidas utilizando un servicio de **reconocimiento de entidades nombradas** y se encuentran **disponibles a través de una API para realizar consultas.**

Origen del proyecto

El proyecto está basado en dos proyectos anteriores:

XDATA: portal web focalizado en la visualización de noticias.

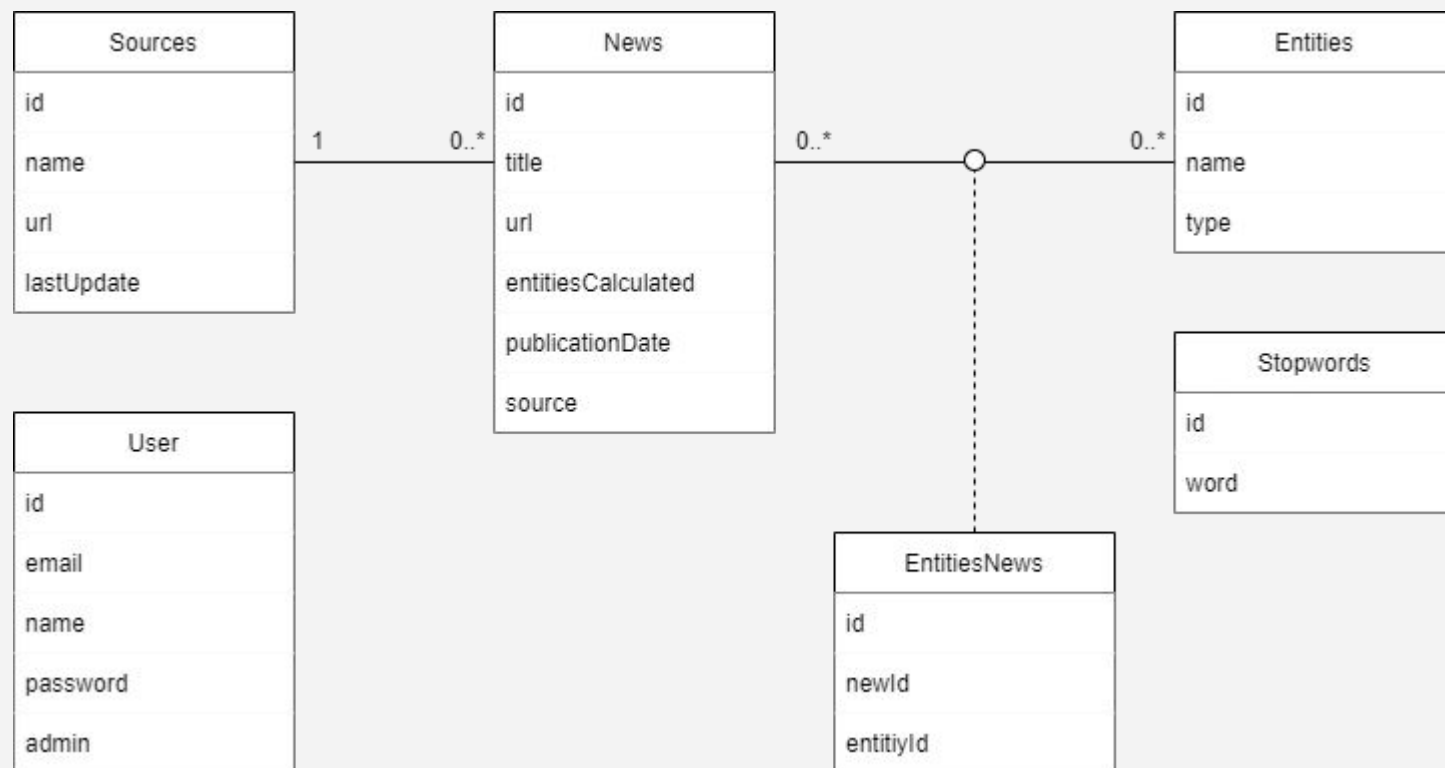
NERd API: cuenta con una herramienta para el reconocimiento de entidades en textos utilizando su propio modelo entrenado de procesamiento natural del lenguaje.

Desarrollamos nuestra propia API, similar a la utilizada en el proyecto XDATA pero utilizando diferentes tecnologías e implementando nuevos endpoints que incluyen el reconocimiento de entidades en los titulares utilizando la NERd API.

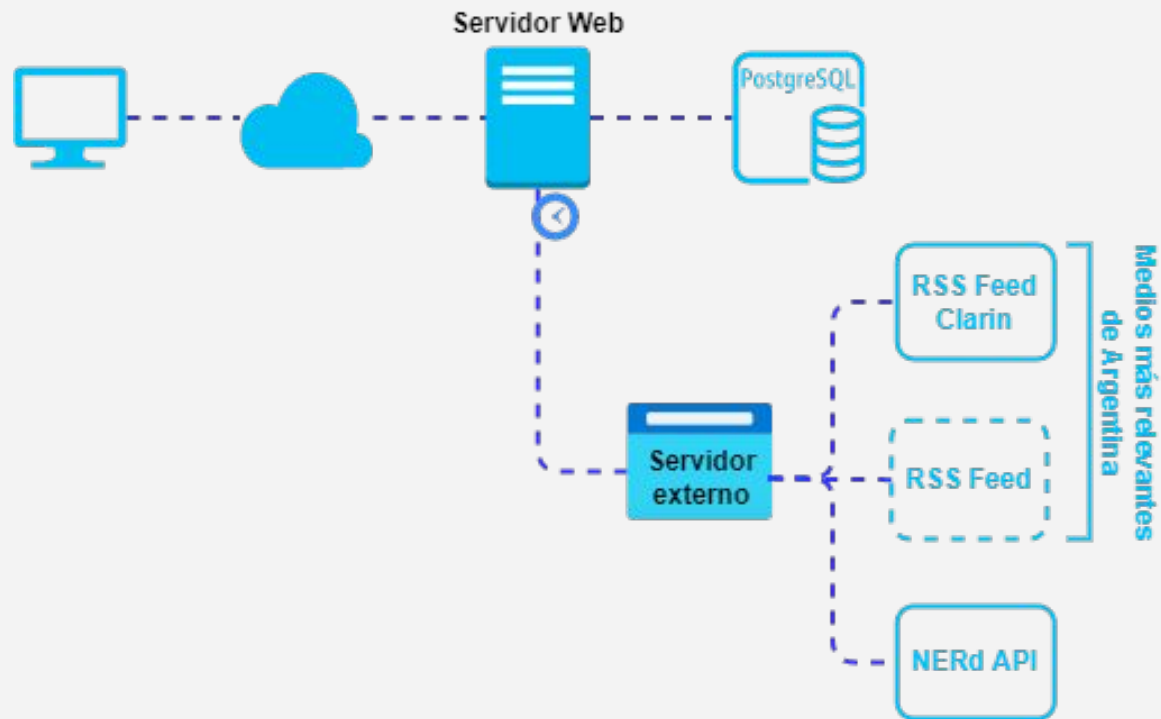
Tecnologias utilizadas



Diagrama UML



Arquitectura



Generación de entidades

Periódicamente, se corre un proceso que se comunica con la NERd API para conocer las entidades presentes en todos los nuevos titulares.

La fortuna más allá de la gloria: ¿cuántos millones recaudó el Real Madrid por ganar la Champions League?

Desde 2018 que no juega en Argentina: El nuevo refuerzo que podría sumar el equipo de Battaglia

PER LOC ORG MISC DATE

Endpoints

GET

nube-de-palabras: genera un conjunto de palabras y para cada palabra cuenta su frecuencia absoluta.

cantidad-de-noticias: devuelve la cantidad de noticias publicadas en una fecha.

heatmap: genera un conjunto de medios por fecha y para cada medio cuenta la cantidad de noticias publicadas.

tendencias: dado un conjunto de palabras y un rango de fechas, para cada palabra cuenta su frecuencia relativa al máximo en dicho periodo.

busqueda: devuelve el resultado de una consulta personalizada.

wordtree: genera un conjunto de titulares.

medios: genera un conjunto con todos los medios que se encuentran registrados para consultar.

Endpoints

GET

stopwords: genera un conjunto con todas las palabras registradas como stopwords.

entidades: genera un conjunto con todas las entidades encontradas.

POST

login: recibe las credenciales para el acceso a la aplicación como usuario.

insert-stopword: recibe la stopword a registrar en la aplicación.

insert-medio: recibe el medio a registrar en la aplicación.

DELETE

stopword: elimina una stopword registrada.

medio: elimina un medio registrado.

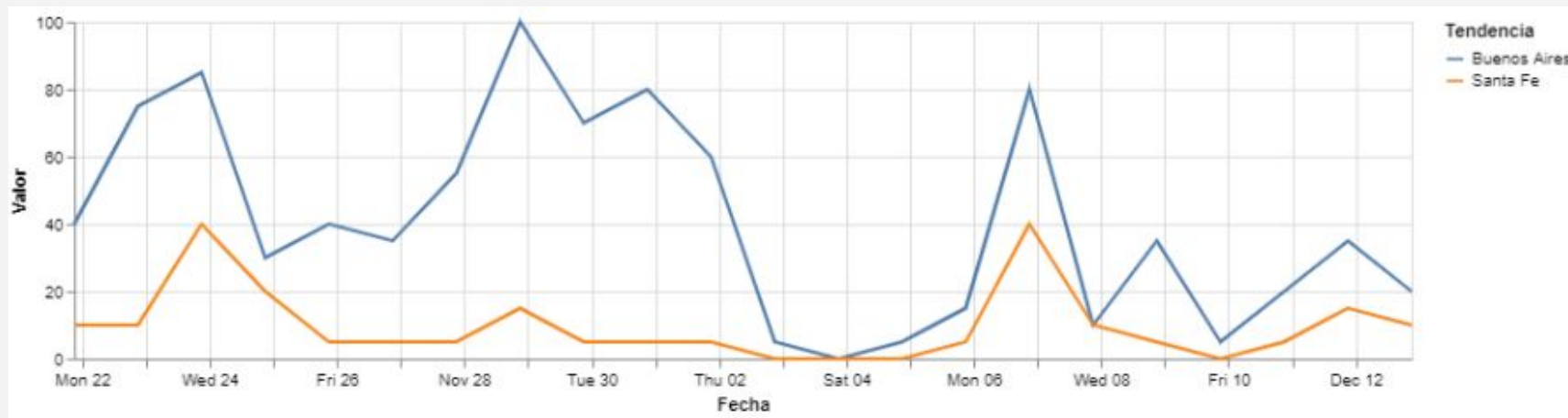
Visualizaciones

Entidades: Devuelve todas las entidades encontradas en un rango de fechas. Si no se especifica rango de fechas, devuelve la cantidad de noticias del día.



Visualizaciones

Tendencias: Tendencias de las palabras por fecha, relativas al máximo en el período de fechas indicado.



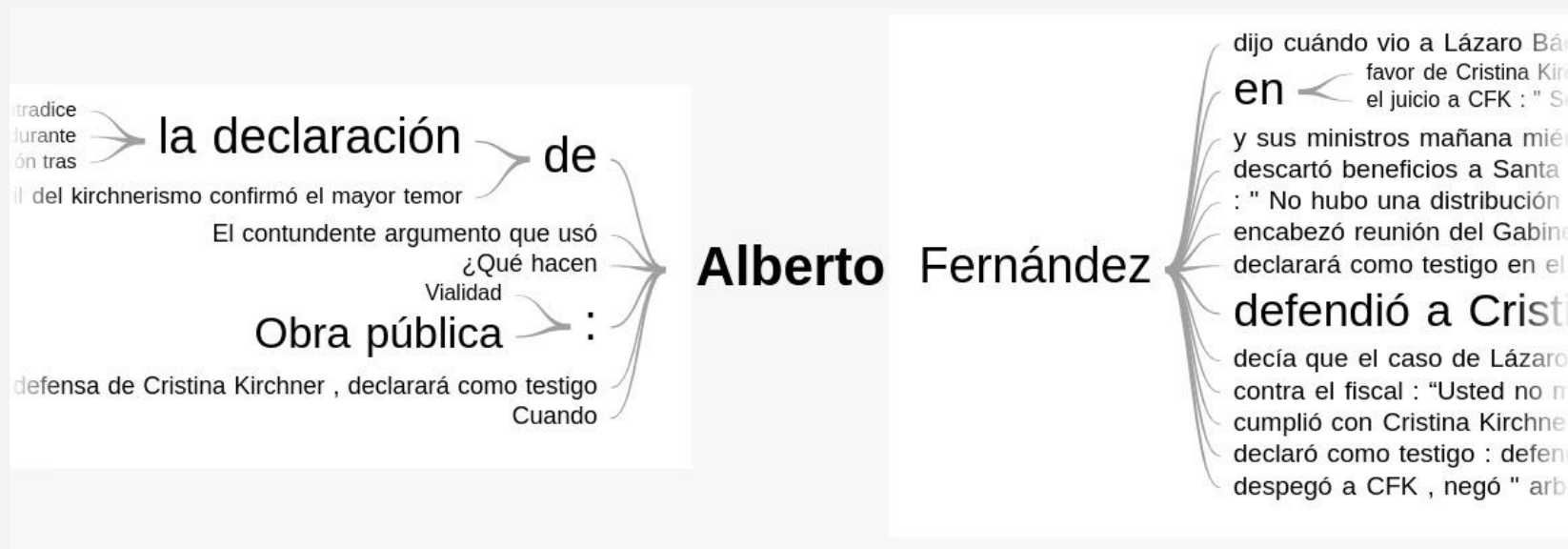
Visualizaciones

Nube de palabras: Devuelve las palabras con más apariciones en los títulos de las noticias.



Visualizaciones

Wordtree: Devuelve los títulos de las noticias en la base de datos para el día de la fecha.



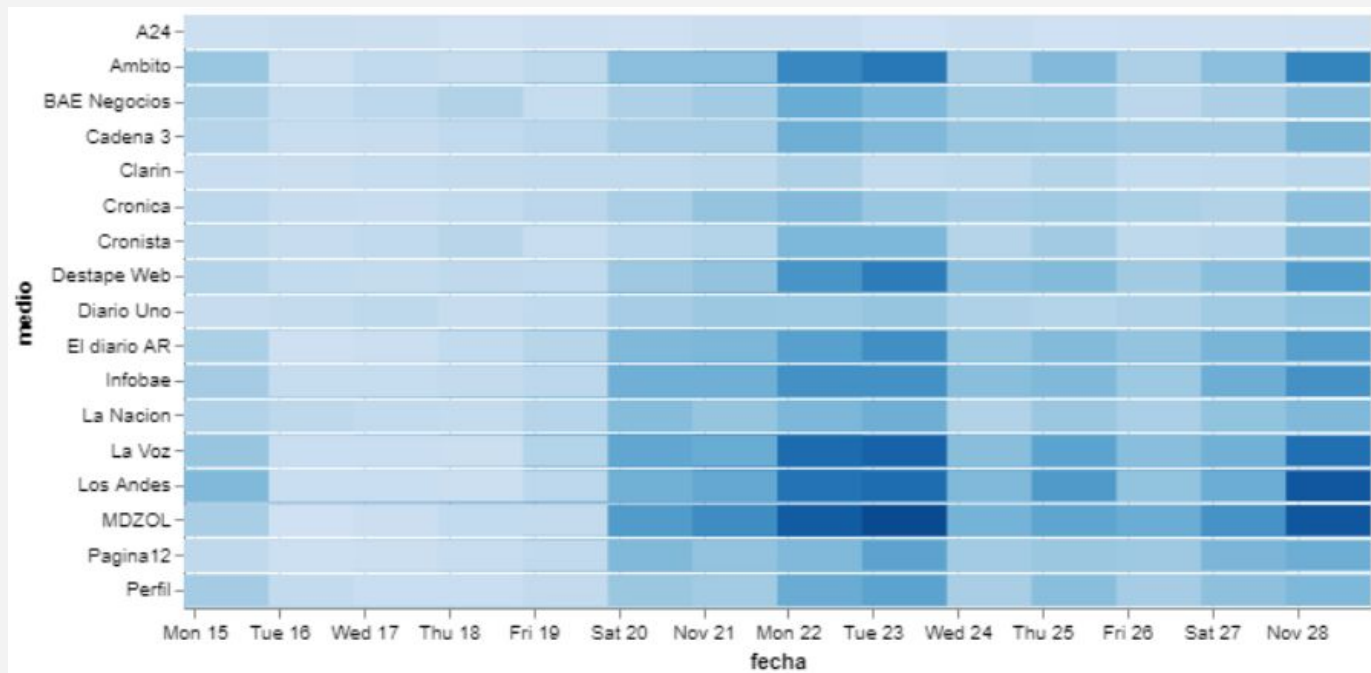
Visualizaciones

Búsqueda: Muestra una tabla con todos los links a las noticias que coinciden con los parámetros de búsqueda ingresados.

Titulo
Video: lo encontró robando y lo obligó con un machete a que muestre su cara a una cámara de seguridad
El FMI trata las sobretasas que cobra por deuda excesiva, pero nadie cree que las vaya a eliminar
Se filtro el mensaje que Antonio Laje le envió a su equipo de trabajo tras el escándalo
El BCRA acelera el ritmo de devaluación diario, pero no logra detener la caída de las reservas
El complicado método para descargar y actualizar la app Cuidar por el pase sanitario que instauró el Gobierno
La frase de Joan Manuel Serrat que provocó la emoción de Gerardo Rozín
El INE se ajustó al presupuesto para 2022: la Revocación de Mandato costará 3,830 millones de pesos
Sneed ausente en duelo de Chiefs tras asesinato de hermano
Muerte de Vicente Fernández minuto a minuto: Hijos del charro montan la Guardia de honor junto a su féretro

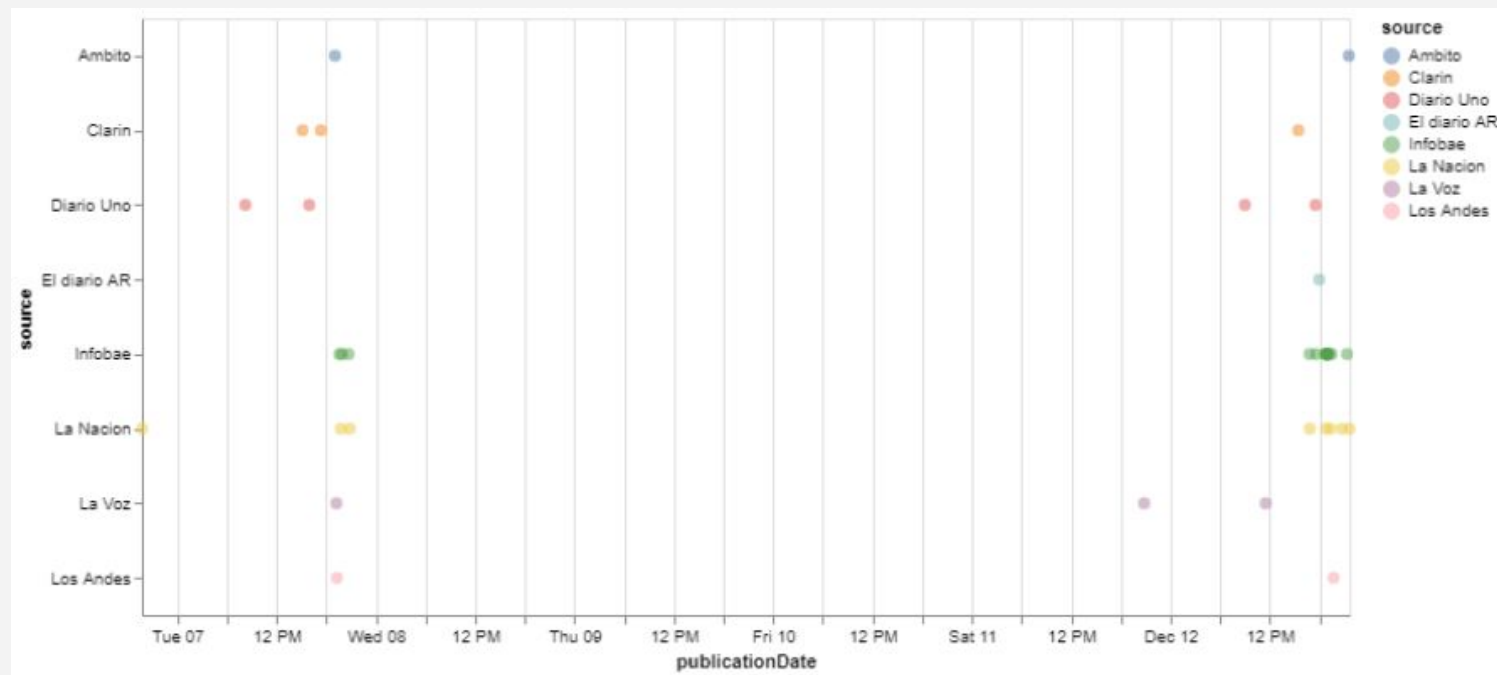
Visualizaciones

Noticias por fecha y medio: Heatmap que visualiza la cantidad de noticias recolectadas en la API por cada medio disponible.



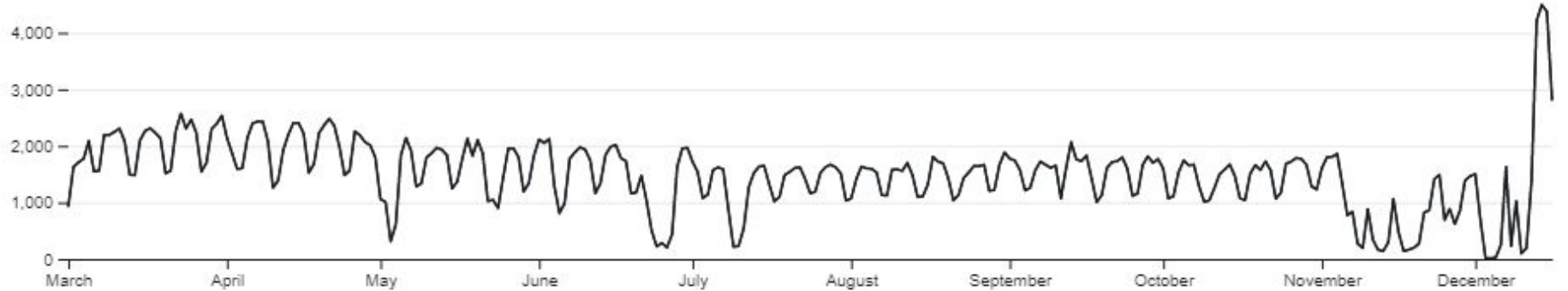
Visualizaciones

Primicia: Dado un tema y un intervalo temporal, quien arranca con la noticia y quien se copia.



Visualizaciones

Cantidad de noticias: Permite visualizar la cantidad de noticias por fecha.



Demo

[Swagger](#) / [Observable](#)

Noticias			▼
GET	/cantidad-de-noticias		🔒
GET	/heatmap		🔒
GET	/busqueda		🔒
GET	/wordtree		🔒
GET	/nube-de-palabras		🔒
GET	/tendencias		🔒

Preguntas

