

# **Geographical Text Analysis**

**MANUAL DE USUARIO**

# Derechos de autor

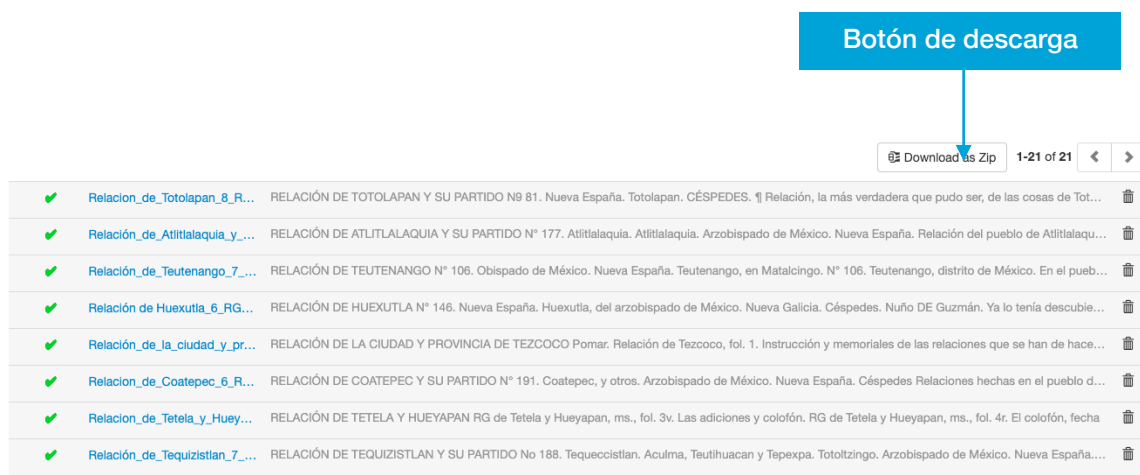
## Plataforma GTA

El objetivo del presente documento es dar a conocer las características y funcionalidades de la herramienta, antes de comenzar, se debe mencionar que GTA hace uso de archivos que se pueden generar en la plataforma Tagtog. Esta es una herramienta de anotación de texto eficiente lista para entrenar IA, la cual genera archivos con anotaciones de textos planos.

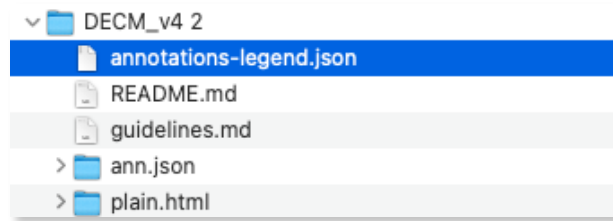
## Descarga de archivos de Tagtog

GTA trabaja con archivos que genera la plataforma Tagtog, cuando contenido de documentos con esta plataforma es catalogado, se pueden descargar y user en GTA, estos documentos se descargan de la siguiente forma:

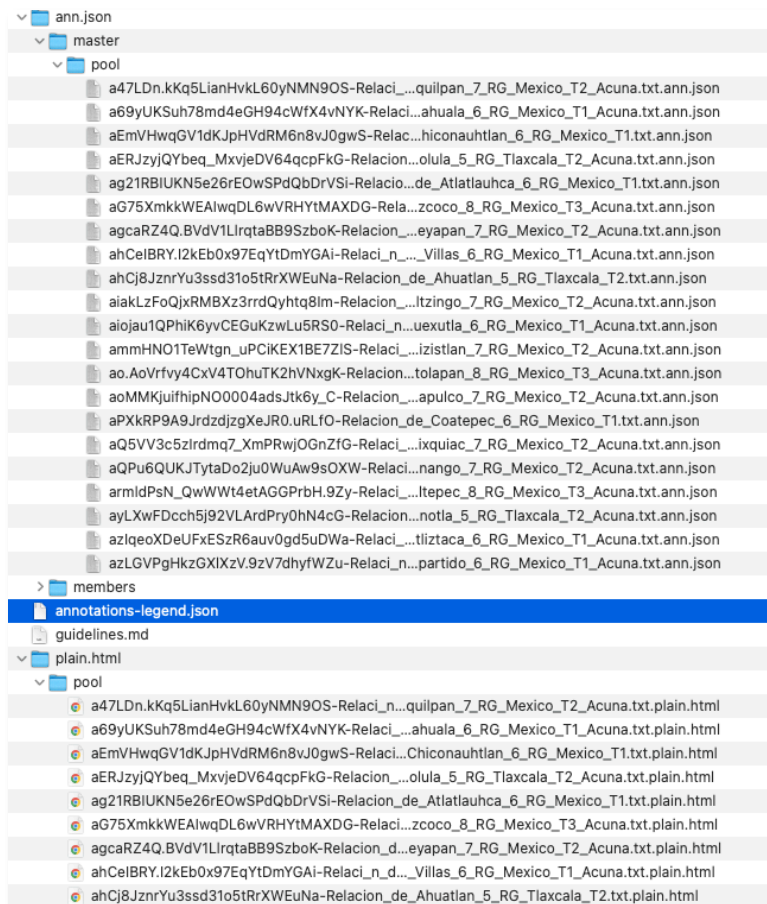
- ❖ Al seleccionar un proyecto se enlistan los documentos trabajados, en la parte superior derecha hay un botón para descargar todo el proyecto.



- ❖ El archivo descargado en raíz tiene un json llamado “annotations-legend”, este archivo contiene las claves de entidades, este el uno de los archivos que se necesitan. **No se debe cambiar el nombre de este archivo nunca.**



- ❖ La carpeta ann.json contienen los archivos con las anotaciones que se hicieron en los documentos del corpus y la carpeta plan.html contiene los documentos con los textos planos. Los pares deben tener el mismo nombre con la diferencia de sus extensiones.



## Pantalla de Inicio

## Area para subir los archivos

## Archivo de entidades

Upload your files here

## Listado de documentos (pares de html y json)

## Botón de inicio

### Process corpus

- **Area de subir archivos:** En esta area se arrastran los archivos o se da click para poder subir los archivos que se quieren procesar, en esta area se suben los 3 tipos de archivos ya mencionados.
- **Archivo de entidades:** Contiene la lista de entidades que se utilizaron con la plataforma Tagtog, esta lista tienen las claves con las que están catalogadas las anotaciones en cada documento. Se puede reemplazar presionando el botón a su derecha.
- **Listado de documentos:** La información debe venir en pares el texto plano (html) y las anotaciones del documento (json).

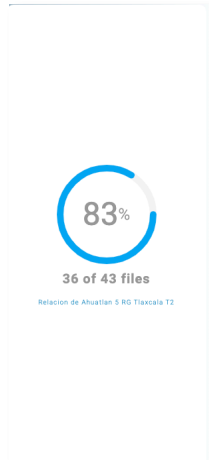
Relaci n de Huexutla 6 R... [json](#) [html](#) [X](#)

Los dos documentos deben tener el mismo nombre Por ejemplo:

tagtogldFile-documentName.txt.ann.json

tagtogldFile-documentName.txt.plain.html

- **Botón de inicio:** Se inicia el procesamiento de los archivos y se muestra el progreso del mismo.



## Pantalla del Dashboard

Pantalla principal de la plataforma en la que se muestran 5 secciones principales: Filtros, Corpus, Gazetteer, Información de anotación y Contexto.

The dashboard is titled "Geographical Text Analysis" and contains five main sections:

- Filtros:** Located on the left, it includes a "Filter by keyword" section with a text input and an "Apply filter" button. Below it is a "Filter by entity and labels" section with a list of categories (Architecture, Ethnic\_Group, Activity, Language, Natural\_Resource, Cultural\_Artifact, Location, toponym, settlement type, geographic feature type, ecclesiastic jurisdiction, civil jurisdiction, Not Labelled, Person, Mobility) and checkboxes. At the bottom is a "Built Complex Query" section with an "Import Query" button and a "Build Query" button.
- Corpus:** Located in the top center, it displays a "Documents index" table with columns for Document, Entity, Paragraph, Index, and Labels. It also shows a "Corpus" view with a list of documents and a "Text" view with a list of documents.
- Gazetteer:** Located on the right, it shows a map of Mexico with a list of locations and a "Map view" button.
- Información de anotación:** Located in the bottom center, it displays a table with columns for Document, Entity, Paragraph, Index, and Labels. It also shows a "Context" view with a list of documents.
- Contexto:** Located in the bottom right, it displays a table with columns for Document, Left, Term, and Right. It also shows a "Context" view with a list of documents.

Conformada por las siguientes herramientas:

- **Filtros:** Contiene 3 formas de filtrar las anotaciones por palabra clave, por categorías y por el constructor de consultas. Por palabra clave se buscan anotaciones exactas o similares a lo que se escribió en el campo. La búsqueda por categorías depende si las anotaciones están clasificadas en alguna entidad o etiquetas, se puede realizar varias selecciones dependiendo de las categorías seleccionadas.

Filtro por palabra clave

The image shows a vertical interface for filtering annotations. It is divided into three main sections, each with a title and a corresponding filter type. The first section, 'Filter by keyword', has a text input field labeled 'Type here' and an 'Apply filter' button. The second section, 'Filter by entity and labels', lists various categories with expandable/collapsible arrows and checkboxes. The 'Location' category is expanded, showing sub-categories like 'toponym', 'settlement type', 'geographic feature type', 'ecclesiastic jurisdiction', 'civil jurisdiction', and 'Not Labelled', all of which are checked. The third section, 'Built Complex Query', has an 'Import Query' text input field and a 'Build Query' button. A vertical color bar is visible on the right side of the 'Filter by entity and labels' section.

Filter by keyword

Type here

Apply filter

Filter by entity and labels

Architecture ☐

Ethnic\_Group ☐

Activity ☐

Language ☐

Natural\_Resource ☐

Cultural\_Artefact ☐

Location ☒

toponym ☒

settlement type ☒

geographic feature type ☒

ecclesiastic jurisdiction ☒

civil jurisdiction ☒

Not Labelled ☒

Person ☐

Mobility ☐

Apply filter

Built Complex Query

Import Query

Build Query

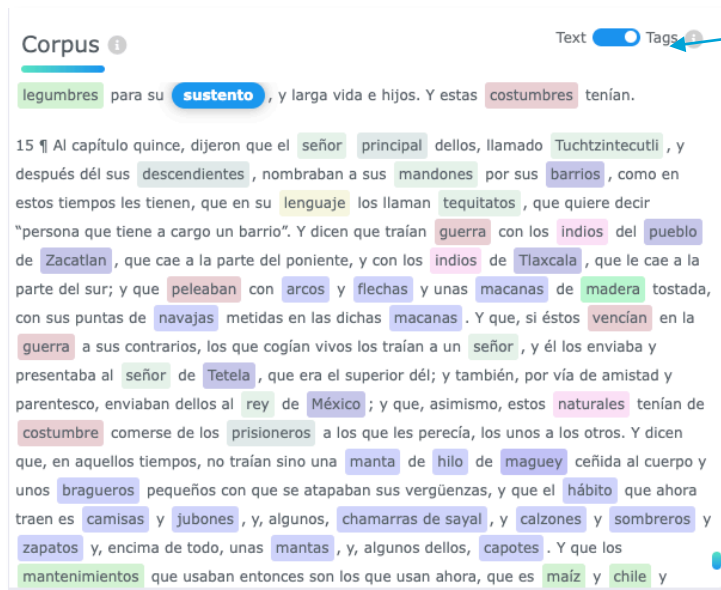
Filtro por entidades y etiquetas

Constructor de consulta

- **Index:** Es una lista de los documentos cargados y cada elemento cuenta con un botón para ocultar o mostrar el corpus del documento.



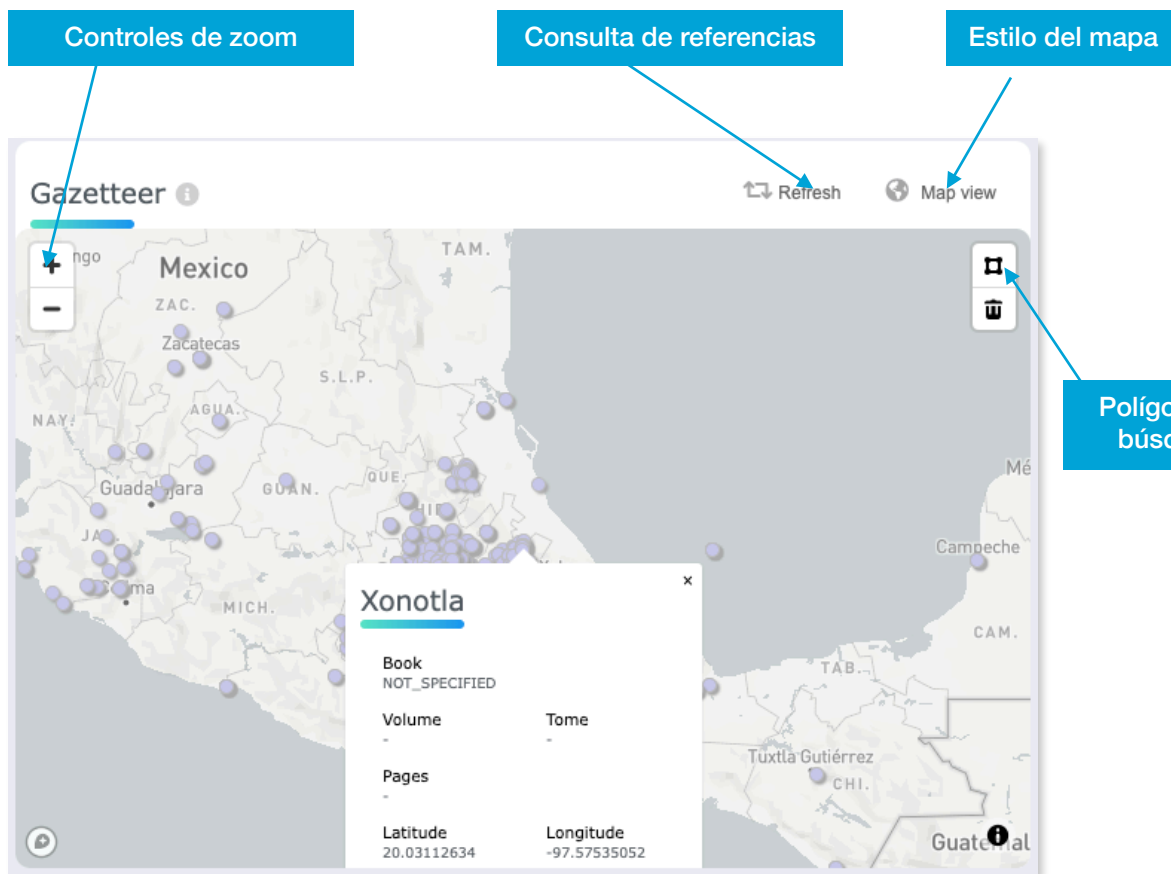
- **Corpus:** Contiene el texto plano de los documentos que conforman el corpus, en él se muestran las anotaciones resaltadas de distintos colores dependiente de las entidades a las que pertenecen. Desde esta sección se puede explorar el corpus, así como seleccionar alguna palabra clave y obtener su información (ocurrencias en el tomo o corpus completo, contexto de las ocurrencias, su ubicación en el mapa si es que se cuenta con ella las referencias del gazetteer).



### Forma de búsqueda

El control en la parte superior derecha controla la forma de interactuar con la sección del corpus. Apuntando a "Tags" se puede dar click en las anotaciones resaltadas y realizar búsqueda de la anotación. En el otro caso "Text" señala que se puede seleccionar texto con el puntero y buscar el texto seleccionado.

- **Gazetteer:** Mapa interactivo que muestra información geográfica de palabras clave. La información que se muestra en esta sección se obtiene de un API que busca referencias de topónimos con datos de georeferenciación del topónimo y su bibliografía.





- ➡ El mapa tiene varias herramientas, la primera es consultar las referencias de los topónimos que están en las anotaciones de los documentos. La segunda es cambiar el estilo del mapa para poder apreciar los topónimos en mapa con diferentes estilos.



- ➡ Para el uso de polígonos de búsqueda basta con seleccionar el botón de polígono y dibujar el polígono deseado, se dará doble click para terminar el polígono y comenzar a buscar coincidencias de polígonos en los documentos que están dentro de los polígonos. Se pueden dibujar varios polígonos y además borrarlos solo con seleccionar un polígono y presionar el botón de basura.
- **Información de anotación:** Brinda información detallada de la anotación seleccionada de la sección del corpus o del contexto. Además se despliegan las referencias encontradas para esa anotación.

Detalles de la anotación

### Annotation information ⓘ

**Document**  
Relaci\_n\_de\_Atltitlaquia\_y\_partido\_6\_RG\_Mexico\_T1\_Acuña

<b>Entity</b> Location	<b>Paragraph</b> s1v1	<b>Index</b> 2933	<b>Labels</b> toponym
---------------------------	--------------------------	----------------------	--------------------------

**Context**  
la de los magueyes, de que son muy aprovechados, por los particulares y muchos aprovechamientos que dellos tienen. Destos magueyes, que en alguna manera aluden a la raíz que en España se dice zabila, sino que son mucho mayores, tienen por principales aprovechamientos el hilo, más o menos delgado, según que lo benefician, y tienen las mantas con que se cubren,

---

**Book**  
Relaciones Geográficas del Siglo XVI: Antequera

<b>Toponym</b> España	<b>Volume</b> 2	<b>Tome</b> 1	<b>Author</b> Acuña, René
<b>Latitude</b> 40.58901942		<b>Longitude</b> -4.148009986	
<b>Country</b> UNKNOWN		<b>State</b> UNKNOWN	

**Pages**  
35, 36, 37, 38, 56, 57, 72, 74, 75, 105, 106, 121, 122, 123, 124, 134, 147, 158, 167, 179, 181, 190, 195, 200, 205, 215, 234, 240, 272, 289, 296, 316, 332, 351, 355, 356, 357

---

**Book**  
NOT\_SPECIFIED

<b>Toponym</b> España	<b>Volume</b> -	<b>Tome</b> -	<b>Author</b> UNKNOWN
<b>Latitude</b>		<b>Longitude</b>	

Referencia de la anotación

- **Contexto:** Herramienta que permite explorar el contexto de las anotaciones del corpus, dependiendo de los filtros que se apliquen son los resultados que desplegarán. La ventana de contexto contiene 4 controles principales, que son los siguientes:
  - \* **Tamaño de contexto:** Define la cantidad de palabras que se mostrará en el contexto a la izquierda y a la derecha.
  - \* **Contexto izquierdo y derecho:** Una vez cargados los resultados se puede filtrar con la búsqueda en su contexto, escribiendo en los campos correspondientes y dar enter se inicia la búsqueda de anotaciones que contenga la ingresado.
  - \* **Exportar:** Se puede descargar los resultados en un archivo csv o json, con los detalles de cada anotación y sus referencias georeferenciadas.

Control de tamaño de contexto

Campos de filtrar por el contexto

Botón de exportación

Context

Context size

30

Left context

Right context

200 de 207 Matches

EXPORT

Document	Left	Term	Right
Relaci_n_de_Atengo_y_Mizquihuala_6_RG_M...	oblación fue el gran cocolizte que hubo en esta	Nueva España	, de mortandad. La doctrina e iglesia la rige ...
Relaci_n_de_Atengo_y_Mizquihuala_6_RG_M...	y, después que entraron los españoles en esta	Nueva España	, ha habido tres veces pestilencia, en que no ...
Relaci_n_de_Atltitlaquia_y_partido_6_RG_Me...	itlalaquia. Atltitlaquia. Arzobispado de México.	Nueva España	. Relación del pueblo de Atltitlaquia, y sus s...
Relaci_n_de_Atltitlaquia_y_partido_6_RG_Me...	, que en alguna manera aluden a la raíz que en	España	se dice zabila, sino que son mucho mayores, ...
Relaci_n_de_Cempoala_Epazoyuca_y_Tetliztac...	lo de Cempoala de la nueva congregación de la	Nueva España	, en primero de noviembre de mil y quiniento...
Relaci_n_de_Cempoala_Epazoyuca_y_Tetliztac...	vechos destos magueyes. 23 Las arboledas de	España	que tienen estos cuatro pueblos y sus sujeto...
Relaci_n_de_Cempoala_Epazoyuca_y_Tetliztac...	). EPАЗOYUCA En el pueblo de Epazoyuca de la	Nueva España	, en primero de noviembre de mil y quiniento...
Relaci_n_de_Cempoala_Epazoyuca_y_Tetliztac...	il más provechoso. Y, de los árboles venidos de	España	, se dan muy bien los duraznos, y membrillo...
Relaci_n_de_Cempoala_Epazoyuca_y_Tetliztac...	uachtli. Y, en general, toda la verdura traída de	España	se da muy bien. 25 El trigo de España se da ...
Relaci_n_de_Cempoala_Epazoyuca_y_Tetliztac...	raída de España se da muy bien. 25 El trigo de	España	se da muy bien. En este pueblo se coge muy...
Relaci_n_de_Cempoala_Epazoyuca_y_Tetliztac...	a). TETLIZTACA En el pueblo de Tetlitzaca de la	Nueva España	, a quince días del mes de noviembre de mil ...
Relaci_n_de_Cuatro_Villas_6_RG_Mexico_T1_...	RO VILLAS Nº 215. Del arzobispado de México.	Nueva España	. Tepuztlan. Céspedes. La Villa de Tepuztlan ...
Relaci_n_de_Cuatro_Villas_6_RG_Mexico_T1_...	que, para ello, le envió el señor visorrey desta	Nueva España	don Martín Enríquez, gobernador en ella, en l...
Relaci_n_de_Cuatro_Villas_6_RG_Mexico_T1_...	de México, donde está la Real Audiencia desta	Nueva España	; aunque otros dicen que hay once leguas. Tl...
Relaci_n_de_Cuatro_Villas_6_RG_Mexico_T1_...	ervo" y es de la manera de la flor del adelfa de	España	. XXIII Al veinticuatro capítulo: Parece que la...
Relaci_n_de_Cuatro_Villas_6_RG_Mexico_T1_...	a los perrillos. Y dicen que, de los animales de	España	, no hay ningunos, si no es puercos y perros ...
Relaci_n_de_Cuatro_Villas_6_RG_Mexico_T1_...	nonte; y que las gallinas de Castilla vinieron de	España	, y que la manera de criarlas es como en Cas...
Relaci_n_de_Cuatro_Villas_6_RG_Mexico_T1_...	en el arzobispado de México, gobernación de la	Nueva España	, y es villa del Marquesado del Valle y cabece...
Relaci_n_de_Cuatro_Villas_6_RG_Mexico_T1_...	ento, 152. 2º Nº 200. Céspedes. Huaxtepeque.	Nueva España	. Arzobispado de México La Villa de Huaxtepe...
Relaci_n_de_Cuatro_Villas_6_RG_Mexico_T1_...	s, que para ello le envió el señor visorrey desta	Nueva España	, don Martín Enríquez, gobernador en ella, en...
Relaci_n_de_Cuatro_Villas_6_RG_Mexico_T1_...	2 Anacitla. Anacitla. Arzobispado de México	Nueva España	Céspedes. La Villa de Anacitla y sus estanc...