



DOCUMENTO TÉCNICO PROYECTO MACROCOR

1. Especificaciones técnicas

La aplicación web de Macrocor ha sido diseñada usando el framework web de Django, junto con otras librerías de terceros que incorporan funcionalidades a la página como Chart.js, Bootstrap o FontAwesome. El servicio se aloja en un servidor web propiedad de la Universidad de Málaga.

Para la base de datos se utiliza un motor de base de datos relacional basado en MySQL en el que se recoge información sobre inicio de sesión, estructura de los corpus e historial de búsquedas realizadas en la página por los usuarios registrados. La información de los usuarios registrados se guarda en la base de datos de manera encriptada para que no pueda ser accedida por terceras personas ajenas al proyecto y a los que no se les ha otorgado permisos.

El algoritmo de búsqueda y explotación de corpus es de elaboración propia por el diseñador de la página web, y se ejecuta de manera concurrente en servidores de la Universidad de Málaga.

2. Personal

La aplicación web ha sido ideada, editada y diseñada por la Investigadora Principal del Proyecto Miriam Seghiri Dominguez y el editor web Antonio Lara Gutiérrez. Para preguntas sobre teoría o lingüística, contacte con Miriam Seghiri en seghiri@uma.es. Para preguntas o incidencias sobre la página web y su uso, contacte con Antonio Lara en antoniolaragutierrez2000@gmail.com.



3. Manual de uso y opciones

3.1. Esquema de vistas de la aplicación

Vista	Descripción	Capturas de pantalla
INICIO	<ul style="list-style-type: none"> ➤ Pantalla de bienvenida. ➤ Muestra una breve descripción de la página. ➤ Muestra estadísticas reales. ➤ Muestra preguntas/respuestas frecuentes. 	
PUBLICACIONES	<ul style="list-style-type: none"> ➤ Muestra enlaces a sitios web donde visualizar biografía/publicaciones de la I.P. del proyecto. 	
CONTACTO	<ul style="list-style-type: none"> ➤ Muestra información de contacto. ➤ Permite descargar el presente documento técnico (también disponible en el pie de página). 	
ACCESO/ REGISTRO	<ul style="list-style-type: none"> ➤ Permite acceder a la aplicación usando nombre de usuario y contraseña previamente solicitado. ➤ Permite acceder a la página de registro. En ella se rellenan los campos necesarios para registrarse en la aplicación durante 1 año. 	



MACROCOR

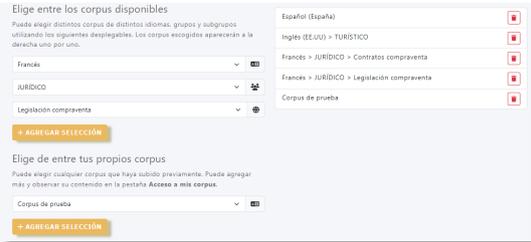
<p>MACROCOR PUS</p>	<ul style="list-style-type: none"> ➤ Permite buscar y explotar corpus disponibles y propios usando una gran variedad de opciones. ➤ Permite subir corpus privados para poder explotarlos individualmente o junto con otros corpus disponibles de la plataforma. 	
<p>RESULTADO</p>	<ul style="list-style-type: none"> ➤ Informa del análisis realizado incluyendo información estadística. ➤ Señala las coincidencias de la búsqueda, el contexto y un enlace al documento coincidente original. ➤ Permite obtener estadísticas de forma visual y en archivo de frecuencia de palabras. 	

3.2. Opciones de MacroCorpus

Opción	Descripción	Captura de pantalla
<p>Campo de búsqueda</p>	<p>Campo de texto que permite definir la búsqueda que se desea realizar en el/los corpus seleccionado/s. Se puede introducir:</p> <ul style="list-style-type: none"> ➤ Una palabra. Ej: casa. Busca coincidencias de la palabra literalmente en el corpus. ➤ Palabras independientes separadas por coma. Ej: casa, comida. Busca coincidencias de cada palabra literalmente en el corpus por separado. ➤ Conjunto de palabras entre comillas inglesas. Ej: "casa de campo". Busca coincidencias del conjunto de palabras completo en el corpus. ➤ Conjunto de palabras entre comillas inglesas con una colocación. Ej: "casa de *". Busca conjuntos de palabras que cumplan la colocación. Ej: "casa de campo", "casa de retiro", "casa de madera"... 	



MACROCOR

<p>Distinción mayúsculas y minúsculas</p>	<p>Checkbox que permite especificar si se van a distinguir entre mayúsculas y minúsculas al buscar coincidencias del término de búsqueda.</p> <ul style="list-style-type: none"> ➤ Si se marca Sí, distingue las mayúsculas, por lo que “Casa” y “casa” son coincidencias distintas, y se tiene en cuenta la especificada en el campo de búsqueda. ➤ Si se marca No, no distingue las mayúsculas y minúsculas, por lo que “Casa” y “casa” son ambas coincidencias del mismo término. 	
<p>Elección de corpus disponibles / corpus propios</p>	<p>Selector interactivo que permite elegir una combinación de corpus personalizada sobre la que realizar la búsqueda y establecer la frecuencia de palabras.</p> <ul style="list-style-type: none"> ➤ <i>Corpus disponibles</i>: selecciona los corpus ordenados por Idioma > Grupo general > Subgrupo específico y haga click sobre “Agregar selección” para visualizarlo en los corpus elegidos. Si solo selecciona un Idioma y hace click sobre “Agregar selección” se añaden TODOS los corpus en dicho idioma. Igual sucede al seleccionar un grupo de un Idioma. La jerarquía establecida es la siguiente: <div data-bbox="379 1279 890 1503" data-label="Diagram"> <pre> graph LR Idioma[Idioma] --> Grupo1[Grupo 1] Idioma --> Grupo2[Grupo 2] Grupo1 --> Subgrupo1[Subgrupo 1] Grupo1 --> Subgrupo2[Subgrupo 2] Grupo2 --> Subgrupo3[Subgrupo 3] </pre> </div> <ul style="list-style-type: none"> ➤ <i>Corpus propios</i>: selecciona el corpus propio que desea añadir (identificado por su nombre) y haga click sobre “Agregar selección” para visualizarlo en los corpus elegidos. <p>Puede eliminar los corpus elegidos en el visualizador de corpus a la derecha (haciendo click en el icono de la papelera) hasta obtener la combinación deseada sobre la que realizar la búsqueda/explotación.</p>	



MACROCOR

	<p><i>Nota: seleccionar muchos corpus o corpus extensos lleva consigo un mayor tiempo de procesamiento en el servidor</i></p>	
<p>Frecuencia de palabras</p>	<p>Campo numérico. El número indicado representa los grupos de palabras que se formarán y sobre los que se computarán los grupos de frecuencias de palabras.</p> <p>Este valor no modifica ni tiene influencia sobre el campo de búsqueda, únicamente indica el tamaño de los grupos de palabras para los que se calcula su frecuencia de aparición. Ej. Si selecciona 1, se buscan apariciones de grupos de palabras de tamaño 1 (casa, comida...). Si se selecciona 2, se buscan apariciones de grupos de palabras de tamaño 2 (casa de, comida sobre...).</p>	<p>Indica el número de palabras para agrupar y calcular la frecuencia de palabras en el corpus</p> <p>El número indicado representa los grupos de palabras que se formarán para buscar frecuencias de aparición en los corpus seleccionados. Este cálculo es independiente a la búsqueda que realizas.</p> <div style="border: 1px solid #ccc; padding: 5px; margin-top: 10px;"> <input type="text" value="1"/> <p>Número de palabras</p> </div>
<p>Stop words list</p>	<p>Selector de archivo que permite subir un archivo de texto temporal a la plataforma. En dicho archivo se van a especificar las “stop words”, las cuales no serán contabilizadas a la hora de computar la frecuencia de palabras. Este valor no modifica ni tiene influencia sobre el campo de búsqueda.</p> <p>El archivo debe estar en formato .txt e incluir una palabra por línea, del siguiente modo:</p> <div style="display: flex; justify-content: space-around; align-items: center; margin: 10px 0;"> <div style="text-align: center;">  <div style="border: 1px solid black; padding: 5px; width: 100px;"> <p>casa</p> <p>comida</p> <p>hogar</p> </div> </div> <div style="text-align: center;">  <div style="border: 1px solid black; padding: 5px; width: 100px;"> <p>casa,comida</p> </div> </div> </div> <p>Si no se selecciona archivo, se usarán las stop words list más comunes, como se pueden encontrar en la siguiente página: https://countwordsfree.com/stopwords para cada uno de los idiomas disponibles en la aplicación.</p>	<p>Establece la <i>stop words list</i> para la frecuencia de palabras</p> <p>Si no sube ningún fichero, se usarán los cargados por defecto en la aplicación, y disponibles en el siguiente enlace.</p> <p>La lista subida debe de estar en formato .txt incluyendo una palabra por línea. Las palabras que incluya no serán tenidas en cuenta para el cálculo de la frecuencia de palabras en el corpus o corpus escogidos.</p> <div style="border: 1px solid #ccc; padding: 5px; margin-top: 10px;"> <input type="button" value="Seleccionar archivo"/> <input type="button" value="Ninguno archivo selec."/> </div>



3.3. Subida y carga de Macrocorpus propios

La subida de archivos se realiza desde la sección de Macrocorpus, haciendo click sobre la pestaña “Acceso a mis corpus”:

Mis corpus

[AÑADIR CORPUS](#)

Nombre	Idioma	Grupo	Fecha de creación	Acciones
Corpus de prueba	Inglés (Reino Unido)	CIENTÍFICO	6 de julio de 2023 a las 11:02	

Puede añadir nuevos corpus o visualizar, editar o eliminar los ya creados. Para añadir un corpus seleccione un nombre, un idioma y un grupo, y seleccione los archivos que conformarán dicho corpus (puede subir los que considere necesarios, posteriormente podrá añadir más).

+ Añadir corpus a mi colección

Para añadir un corpus a tu colección, debes indicar una clasificación acorde a sus características y, posteriormente, confirmar si deseas que se incluya en el resto de corpus de la aplicación. Para que la aplicación funcione, debes subir archivos .txt (la aplicación se encarga de la posterior codificación).

Nombre del corpus: prueba

Idioma: Alemán

Grupo: CIENTÍFICO

Elegir archivos: 7 archivos

Edición del corpus Corpus de prueba

Nombre del fichero en el sistema	Descarga
NUEVO.txt	

Elegir archivos: Ninguno archivo selec.

[SUBIR](#)

[Cerrar](#)



3.4. Resultados de búsqueda

Al realizar una búsqueda obtiene la siguiente información

- Listado de corpus sobre los que ha realizado la búsqueda/explotación
- Análisis de resultados obtenidos con los siguientes indicadores:
 - Número total de ocurrencias individuales encontradas
 - Número total de ficheros analizados (ficheros de los corpus seleccionados)
 - Número total de ficheros con ocurrencias distintos: número mínimo de ficheros que han devuelto, al menos, una ocurrencia del campo de búsqueda
 - Valores tomados por la colocación: indica los valores que ha tomado la colocación marcada con un asterisco, siempre y cuando haya introducido Conjunto de palabras entre comillas inglesas con una colocación en el campo de búsqueda
 - Frecuencia de palabras: listado de frecuencia de palabras ordenado de mayor a menor frecuencia y ordenado alfabéticamente. Puede descargar un archivo para cada listado
- Resultados obtenidos: tabla con el archivo donde se ha encontrado la ocurrencia y el extracto de dicho archivo con la ocurrencia marcada. Si hace click sobre el nombre del archivo se le redirige a una vista donde puede acceder al texto íntegro de este.



4. Tratamiento de datos

- Responsable del Tratamiento:
 - o Identidad: Universidad de Málaga
 - o Dirección: Avda. Cervantes, 2. 29010, Málaga

- Encargado del tratamiento
 - o ¿Con qué finalidad tratamos sus datos?

El proyecto trata los datos personales recogidos en el formulario de registro de los usuario y los que facilite o se generen durante el uso de la aplicación con las siguientes finalidades:

- Permitir iniciar sesión y mantener un listado personal y privado de corpus.
- Enviar información de inicio de sesión la primera vez que se registra y cada vez que desee renovar su licencia (una vez al año).

- Crear estadísticas de búsquedas en los corpus disponibles de la aplicación.
- Mantener un listado de los corpus creados y archivos subidos a dichos corpus.

Salvo que se indique expresamente lo contrario, los datos solicitados son obligatorios para las finalidades anteriormente indicadas.

- o ¿Por cuánto tiempo conservaremos sus datos?

Los datos personales se conservarán mientras se mantenga la relación que vincula al usuario con la aplicación. No obstante lo anterior, una vez finalizada dicha relación, los datos se mantendrán bloqueados durante 3 años o durante los plazos de prescripción establecidos por las distintas normas que resulten de aplicación.



- ¿A qué destinatarios se comunicarán sus datos?

Los datos no se comunican a ninguna empresa ni organismo ajeno a la Universidad de Málaga, pues los datos recabados de la aplicación sólo sirven para poder extraer información para futuros proyectos o estudios lingüísticos.

- ¿Cuáles son sus derechos?

En cualquier momento, puede solicitar al webmaster o a la Investigadora Principal del proyecto confirmación sobre si sus datos personales están siendo tratados por ésta y, en caso afirmativo, acceder a los mismos. Asimismo, puede solicitar la rectificación de los datos inexactos o, en su caso, solicitar su supresión cuando, entre otros motivos, los datos ya no sean necesarios para los fines para los que se recogieron. En determinadas circunstancias, podrá solicitar la limitación del tratamiento de sus datos, en cuyo caso únicamente los conservaremos para el ejercicio o la defensa de reclamaciones. También, podrá oponerse al tratamiento de sus datos en determinadas circunstancias. En estos casos, se dejará de tratar los datos, salvo por motivos legítimos imperiosos o el ejercicio o la defensa de posibles reclamaciones. Por último, cuando sea procedente, podrá ejercitar el derecho de portabilidad para obtener los datos en formato electrónico o que se transmitan los mismos a otra entidad. A tales efectos, podrá solicitarlo a través de correo electrónico a seghiri@uma.es o antoniolaragutierrez2000@gmail.com.