

Manual de usuario



Búsqueda de entidades nombradas en documentos PDF encontrados en una
página web

TABLA DE CONTENIDO

1	OBJETIVO	3
2	DEFINICIONES	3
3	DESARROLLO DEL MANUAL DE USUARIO	4
3.1	INGRESAR A MASK'AY	4
3.2	INICIAR BÚSQUEDA	4
3.3	RESULTADOS DE BÚSQUEDA	5

1 OBJETIVO

Dar a conocer los pasos que se deben tener en cuenta para una búsqueda efectiva de palabras específicas (entidades nombradas) dentro de documentos PDF encontrados en el URL de interés.

2 DEFINICIONES

Mask'ay es una herramienta de búsqueda, que permite la extracción de información dentro de documentos tipo PDF que se encuentran alojados en páginas web.

Entidades nombradas son categorías predefinidas, como lo son personas, organizaciones, expresiones de tiempo y cantidades.

Archivos generados a partir de procesos ópticos: Son aquellos documentos PDF que han sido creados a partir de una o varias fotografías o imágenes que han sido procesadas mediante un escáner.

3 DESARROLLO DEL MANUAL DE USUARIO

3.1 INGRESAR A MASK'AY

Para iniciar una búsqueda en Mask'ay se debe digitar la entidad nombrada y el URL en el formulario, estos campos son obligatorios (Ver Ilustración 1).

El usuario puede escoger si desea que la búsqueda contenga archivos PDF generados a partir de procesos ópticos cuando se habilita la opción “Incluir documentos PDF generados a partir de imágenes”, sin embargo, esto ocasiona que la búsqueda sea más demorada por el procesamiento de imágenes.

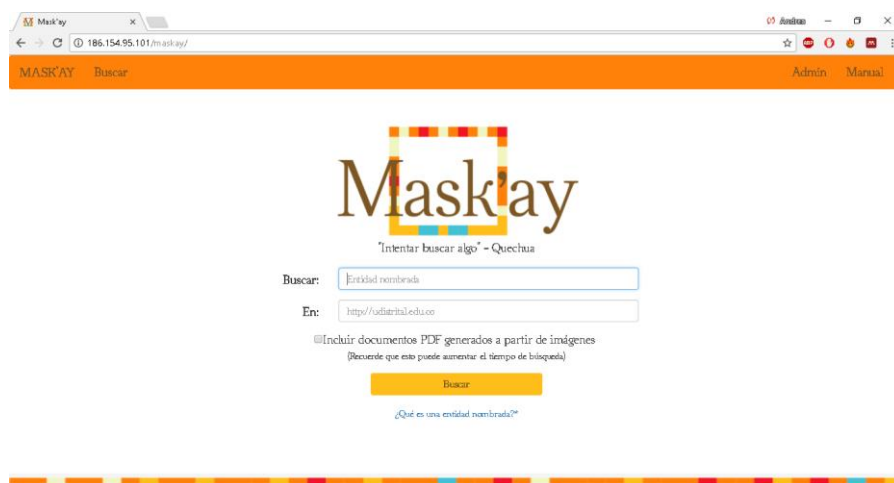


Ilustración 1 Página principal

3.2 INICIAR BÚSQUEDA

Al terminar de ingresar la información, se debe oprimir el botón buscar, esto hará que el software comience todo el procesamiento y mostrará una pantalla de carga.

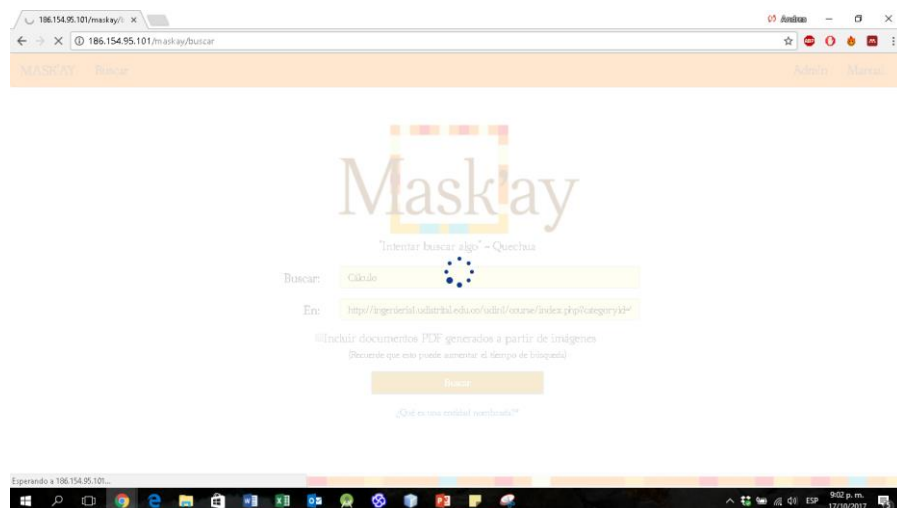


Ilustración 2 Procesando información

3.3 RESULTADOS DE BÚSQUEDA

Una vez terminada la búsqueda se pueden observar los resultados (Ver Ilustración 3):

- El total de documentos PDF encontrados.
- Un párrafo de cada documento PDF que contiene la entidad nombrada solicitada incluyendo el número de veces que se encuentra en el documento.
- El link para abrir el documento PDF.

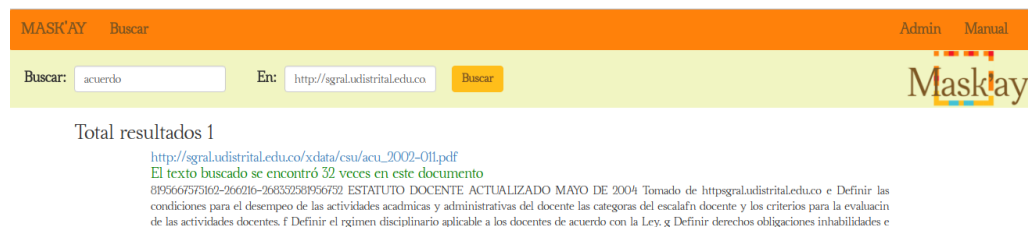


Ilustración 3 Resultado de búsqueda

Para iniciar una nueva búsqueda, en la parte superior, se encuentra el formulario donde se pueden ingresar los nuevos valores de búsqueda (Ver Ilustración 4).



En caso, de que el sitio solicitado nunca haya sido rastreado por la herramienta, los resultados podrían tardar más de tres horas, por esta razón se le solicitará al usuario registrar un correo electrónico para enviarle una notificación cuando el rastreo del enlace haya finalizado y de esta manera, el usuario pueda volver a ingresar la búsqueda (Ver Ilustración 5).

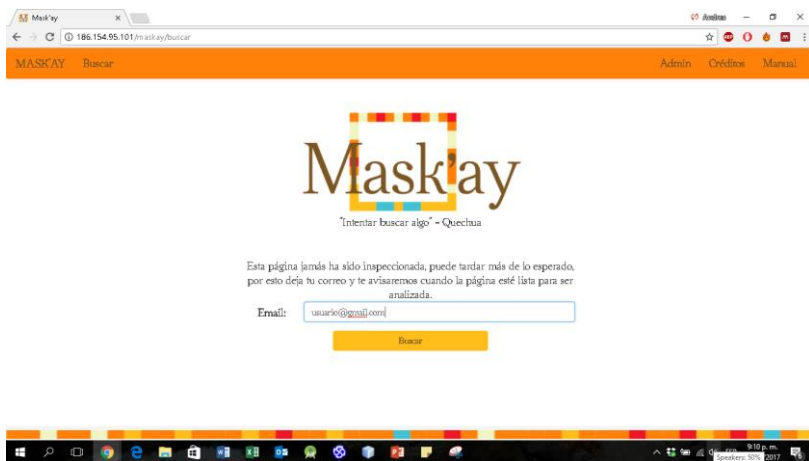


Ilustración 4 Registrar correo para envío de notificación