

Especificación de requerimientos de software:

Herramienta de análisis automático de texto para usuarios inexpertos

Buendia, Pablo Agustín
Fernández, Valentina

Versión 1.0

Índice

Alcance del proyecto	3
Actores del sistema	3
Glosario	3
Requerimientos Funcionales	4
Requerimientos No Funcionales	7
Casos de uso	8
CU1: Procesar texto con aplicación de detección de entidades.	8
CU2: Procesar texto con aplicación de detección de violencia de género en mensajes de Whatsapp.	9
CU3: Crear carpeta	9
CU4: Eliminar carpeta	10
CU5: Descargar resultados	10
CU6: Cambiar esquema de colores	11
CU7: Registrar usuario	11
CU8: Modificar contraseña	11
CU9: Iniciar Sesión	12
Modelo de Dominio	13
Diagrama de Componentes	14

Alcance del proyecto

Este sistema está pensado como una herramienta para el análisis de evidencia textual en investigaciones judiciales. Servirá como una interfaz orientada a usuarios no expertos, que pueda incorporar modelos de análisis de texto con diferentes objetivos.

Actores del sistema

Actor	Descripción
Empleado judicial	Es aquel que utiliza el sistema, cargando lotes de texto, busca que sean analizados, según sus necesidades y luego poder visualizar los resultados de ese análisis.
Administrador	Este usuario tiene acceso a la administración de Django, puede cargar nuevos modelos de análisis al sistema y crear usuarios.

Glosario

Término	Definición
Inteligencia artificial	Se refiere a sistemas o máquinas que imitan la inteligencia humana para realizar tareas y pueden mejorar iterativamente a partir de la información que recopilan.
NLP	Es el área de estudio centrada en cómo los ordenadores entienden el lenguaje humano, lo interpretan y procesan. Involucra lingüística, ciencias de la información e inteligencia artificial. ¹
Tokenización	Segmentar texto en palabras, signos de puntuación, etc. ¹
Lematización	Asignar a las palabras el infinitivo correspondiente, es decir, su forma base. ¹
Entidad	Objeto del mundo real, el cual tiene un nombre asignado. ¹
Violencia sexual	Cualquier acción que implique la vulneración en todas sus formas, con o sin acceso genital, del derecho de la mujer de decidir voluntariamente acerca de su vida sexual o reproductiva a través de amenazas, coerción, uso de la fuerza o intimidación,

¹ Montani, I., Van Landeghem, S., y Gomez, R., Honnibal, M. spaCy 101: Features. Spacy. <https://spacy.io/usage/spacy-101#features>

	incluyendo la violación dentro del matrimonio o de otras relaciones vinculares o de parentesco, exista o no convivencia, así como la prostitución forzada, explotación, esclavitud, acoso, abuso sexual y trata de mujeres. ²
Violencia física	La que se emplea contra el cuerpo de la mujer, produciendo dolor, daño o riesgo de producirlo y cualquier otra forma de maltrato, agresión que afecte su integridad física. ²
Violencia económica	
Violencia psicológica	La que daña las emociones y la autoestima. Afecta el desarrollo personal y perjudica la salud psicológica. Son formas de violencia psicológica las amenazas, las humillaciones, los insultos, el abandono, la indiferencia y las burlas
Violencia simbólica	La que a través de patrones estereotipados, mensajes, valores, íconos o signos transmita y reproduzca dominación, desigualdad y discriminación en las relaciones sociales, naturalizando la subordinación de la mujer en la sociedad. ²

Requerimientos funcionales

1. El sistema permitirá al usuario interactuar a través de una interfaz web.

Nro.	Descripción
1.1	El sistema permitirá la carga de lotes de textos para su análisis, en los formatos .docx .txt .zip
1.2	El sistema contará con aplicaciones para la clasificación según violencia de mensajes de WhatsApp y para la detección de entidades en cualquier tipo de texto.
1.3	El sistema mostrará al usuario las distintas aplicaciones de análisis que puede elegir aplicar.
1.4	El sistema ejecutará el análisis solicitado por el usuario y mostrará los resultados de dicho análisis.
1.5	El sistema permitirá al usuario descargar un documento detallando los resultados del análisis. (resultados completos, estadísticas, entre otros.)

2. El sistema contará con un sistema de registro de usuarios:

² Ley 26485. *Ley de Protección Integral para Prevenir, Sancionar y Erradicar la Violencia contra las Mujeres en los ámbitos en que desarrollen sus relaciones interpersonales*. 1 de abril de 2009. D.O. No. 31632.

Nro.	Descripción
2.1	El sistema permitirá solo a los usuarios de tipo administrador crear nuevos usuarios.
2.2	El sistema permitirá a los usuarios generar carpetas para guardar los archivos que desea procesar en cada análisis.
2.3	El sistema permitirá a los usuarios eliminar carpetas generadas.
2.4	El sistema podría incluir búsquedas o filtros de los resultados de distintos análisis.
2.5	El sistema permitirá a los usuarios modificar sus contraseñas.

3. El sistema contará con una aplicación de detección de entidades:

Nro.	Descripción
3.1	El sistema permitirá al usuario cargar un documento de texto o lote de textos, e indicar que quiere conocer las entidades presentes en ellos.
3.2	El sistema será capaz de verificar si un archivo que se subió al sistema ya existe en la misma carpeta.
3.3	El sistema permitirá realizar análisis múltiples veces sobre la misma carpeta o conjunto de archivos.
3.4	Según el modelo que se utilice se define el tipo de entidades a detectar. Por ejemplo, un modelo podrá detectar: organizaciones, personas, lugares, fechas.
3.5	El sistema mostrará cada entidad junto con el texto, o el extracto de texto, donde fue encontrada.
3.6	El sistema permitirá al usuario configurar sus preferencias de esquema de colores para la visualización de entidades.
3.7	El sistema incluirá esquemas de colores para personas con daltonismo.
3.8	El sistema será capaz de mostrar solo las frases que contengan entidades de una o varias categorías especificadas.
3.9	El sistema mostrará estadísticas de los resultados de las entidades encontradas. Algunas de estas podrían ser: <ul style="list-style-type: none"> • Número de entidades de cada categoría.

	<ul style="list-style-type: none"> ● Porcentaje de entidades de cada categoría. ● Entidades repetidas.
--	--

4. El sistema contará con una aplicación de detección de violencia de género en mensajes de WhatsApp:

Nro.	Descripción
4.1	El sistema permitirá al usuario cargar un lote de textos, e indicar que quiere conocer cuáles de ellos pueden indicar la existencia de violencia de género.
4.2	El sistema proveerá una lista de los textos marcados como ‘relevantes’ o posibles indicadores de violencia.
4.3	El sistema clasificará cada texto en el lote cargado en una categoría según el tipo de violencia que puede detectar en él: <ul style="list-style-type: none"> ● Violencia sexual. ● Violencia física. ● Violencia económica. ● Violencia psicológica. ● Violencia simbólica. ● No violento.
4.4	El sistema realizará un análisis de la información adicional (fechas, archivos adjuntos, etc.) de los mensajes relevantes si fuera necesario.
4.5	El sistema mostrará como resultado los conjuntos de frases encontradas en cada categoría.
4.6	El sistema mostrará el score de violencia de cada frase clasificada.
4.7	El sistema permitirá descargar los resultados completos del análisis.
4.8	El sistema elaborará estadísticas referentes a los resultados. Presencia de textos violentos, número de mensajes identificados en cada categoría y que porcentaje representan en el total de frases y en el total de frases violentas.
4.9	El sistema elaborará un informe al finalizar el análisis, resumiendo los resultados obtenidos.
4.10	El sistema permitirá generar un <i>wordcloud</i> de visualización personalizado al usuario, permitiendo filtrar palabras específicas, preposiciones, y palabras provenientes de frases no violentas.

Requerimientos no funcionales

Tecnologías a aplicar:

- Python 3.10 en adelante.
- Spacy v3 en adelante
- Label Studio
- Django 3.2 en adelante

Adecuación Funcional

- Los modelos entrenados para el sistema deben ser analizados en cuanto a su rendimiento y la calidad de sus resultados.

Usabilidad:

- El *front-end* del sistema debe correr en cualquier computadora con un navegador web moderno.
- El *back-end* del sistema tiene que estar instalado en un servidor Linux.
- El sistema advertirá en su reporte al finalizar el análisis que los resultados obtenidos por medio del análisis no garantizan un 100% de precisión y dará las recomendaciones apropiadas.

Extensibilidad:

- El sistema debe estar abierto para añadir nuevas tablas de estadística y procesar nuevos tipos de formatos de texto. Se busca que a futuro sea posible incorporar nuevos módulos de NLP sin que esto represente un gran cambio al sistema existente.

Mantenibilidad:

- El sistema deberá ser desarrollado y configurado empleando tecnologías que faciliten la mantenibilidad del sistema, como Django y Python.

Seguridad:

- El sistema deberá ser desarrollado y configurado utilizando tecnologías que faciliten la seguridad del sistema, como Django que provee características de seguridad por defecto como protección de “*cross site scripting*” (XSS), protección de “*Cross site request forgery*” (CSRF), protección contra inyección de SQL, entre otras.

Casos de uso

CU1: Registrar usuario

Actor Principal: Administrador

Precondiciones:

- El Administrador ha iniciado sesión en su cuenta
- El usuario a registrar no existe en el sistema

1. El administrador va a la página de registro del sistema.

2. El administrador toma los datos del Empleado Judicial que quiere registrar (nombre de usuario, mail y contraseña)
3. El administrador escribe estos datos en el formulario presentado por el sistema y presiona 'aceptar' para completar el registro.

CU2: Iniciar Sesión

Actor Principal: Empleado Judicial

Precondiciones:

- El Empleado Judicial tiene un usuario en el sistema.
1. El empleado Judicial ingresa al sistema.
 2. El Empleado Judicial completa su información de inicio de sesión (nombre de usuario y contraseña.) y presiona el botón de inicio.
 3. El Sistema inicia la sesión del Empleado Judicial.

Extensiones:

2a. El Empleado Judicial ingresa una contraseña incorrecta:

1. El sistema indica que la contraseña es incorrecta.
2. El Empleado Judicial reingresa la contraseña.

CU10: Borrar análisis

Actor Principal: Empleado Judicial

Precondiciones:

- El Empleado Judicial tiene un usuario en el sistema.
 - Existe un análisis en el sistema que fue realizado por el usuario que intenta borrarlo
1. El Empleado Judicial entra a la sección 'Mis análisis' del sistema.
 2. El Empleado Judicial indica que quiere eliminar un análisis.
 3. El sistema muestra un mensaje de confirmación.
 4. El Empleado Judicial confirma la acción.
 5. El sistema borra el análisis.

Extensiones:

1a.

1. El Empleado Judicial entra a los resultados de un análisis particular.
2. El Empleado Judicial presiona el botón de "borrar".

CU3: Modificar contraseña

Actor Principal: Empleado Judicial

Precondiciones:

- El Empleado Judicial ha iniciado sesión en su cuenta
- 1. El Empleado Judicial entra a sus preferencias de usuario.
- 2. El Empleado Judicial indica que quiere hacer un cambio a su contraseña
- 3. El sistema pide al Empleado Judicial su contraseña actual, y la confirma.
- 4. El sistema pide al Empleado Judicial ingresar la nueva contraseña.
- 5. El sistema guarda la nueva contraseña.

Extensiones:

3a. El Empleado Judicial ingresa una contraseña incorrecta:

1. El sistema indica que la contraseña es incorrecta.
2. El Empleado Judicial reingresa la contraseña actual.

CU4: Crear carpeta y cargar archivos

Actor Principal: Empleado judicial.

Precondiciones:

- El Empleado Judicial ha iniciado sesión en su cuenta
- 1. El Empleado judicial indica que quiere crear una nueva carpeta.
- 2. El Empleado judicial nombra la carpeta.
- 3. El sistema crea la carpeta.
- 4. El Empleado judicial elige qué archivos quiere subir y los sube a la carpeta.

Extensiones:

2a. El Empleado Judicial intenta nombrar la carpeta con un nombre ya existente.

1. El sistema indica que ya existe una carpeta con ese nombre.
2. El sistema da la posibilidad de cambiar el nombre o cancelar la operación.
3. El Empleado elige una de las opciones.

4a. El Empleado judicial intenta subir archivos en formatos no soportados.

1. El sistema indica que esos archivos no son válidos y cuales son las extensiones permitidas.
2. El sistema da la posibilidad de subir archivos diferentes.

CU5: Eliminar carpeta

Actor Principal: Empleado judicial.

Precondiciones:

- El Empleado Judicial ha iniciado sesión en su cuenta
- Existe la carpeta a ser eliminada.

1. El Empleado judicial indica que eliminar una de sus carpetas.
2. El sistema le pide una confirmación.
3. El Empleado judicial acepta la confirmación.
4. El sistema elimina la carpeta.

CU6: Procesar texto con aplicación de detección de entidades.

Actor Principal: Empleado Judicial

Precondiciones:

- El Empleado Judicial ha iniciado sesión en su cuenta.
- Los textos están en un formato aceptado por el sistema.

1. El Empleado Judicial ingresa en la página principal del sistema.
2. El Sistema muestra un menú, incluyendo un botón que permite ir a sus carpetas para cargar los textos.
3. El Empleado Judicial elige una carpeta donde sube los textos y vuelve a la página principal.
4. El Sistema muestra un panel que permite seleccionar una aplicación para analizar texto.
5. El Empleado Judicial elige la aplicación de detección de entidades para analizar los textos.
6. El Empleado Judicial selecciona la carpeta con los archivos que busca analizar.
7. El Empleado Judicial elige un modelo para procesar los datos.
8. El Sistema procesa los datos y muestra los resultados, con las entidades y estadísticas recolectadas.

Extensiones:

5a. El Empleado Judicial elige cargar los archivos en una nueva carpeta.

1. CU3

7a. El Empleado Judicial quiere cambiar la paleta de colores en la que se mostraran los resultados

1. El Empleado Judicial presiona un botón que indica el cambio del esquema de colores.
2. El Empleado Judicial elige una de las opciones presentadas y presiona el botón de 'aceptar'.
3. El sistema cambia los colores en los que se mostraran los resultados.

CU7: Procesar texto con aplicación de detección de violencia de género en mensajes de Whatsapp.

Actor Principal: Empleado Judicial

Precondiciones:

- El Empleado Judicial ha iniciado sesión en su cuenta
- Los textos están en un formato aceptado por el sistema.

1. El Empleado Judicial ingresa en la página principal del sistema.
2. El Sistema muestra un menú, incluyendo un botón que permite ir a sus carpetas para cargar los textos.
3. El Empleado Judicial elige una carpeta donde sube los textos.
4. El Sistema muestra un panel que permite seleccionar una aplicación para analizar texto.
5. El Empleado Judicial elige la aplicación de detección de violencia de género en mensajes de Whatsapp.
6. El Empleado Judicial selecciona la carpeta con los archivos que busca analizar.
7. El Empleado Judicial elige un modelo para procesar los datos.
8. El Sistema procesa los datos y muestra los resultados, con las entidades y estadísticas recolectadas.

Extensiones:

- 5a. El Empleado Judicial elige cargar los archivos en una nueva carpeta.
 1. CU3

CU8: Descargar resultados

Actor Principal: Empleado judicial.

Precondiciones:

- El Empleado Judicial ha iniciado sesión en su cuenta
- El sistema realizó un análisis sobre textos subidos por el Empleado Judicial.

1. El Empleado Judicial indica que quiere descargar los resultados del análisis.
2. El sistema presenta la elección de descargar el informe de resumen de los resultados, los resultados completos o ambos.
3. El Empleado Judicial elige los archivos que quiere descargar y persona 'aceptar'

4. Los archivos son descargados a la computadora del Empleado Judicial.

Extensiones:

1a. El Empleado no se encuentra visualizando los resultados de un análisis actual.

1. El Empleado Judicial entra a la carpeta con el análisis que quiere descargar y entra a la visualización de ese análisis.

CU9: Filtrar Resultados

Actor Principal: Empleado judicial.

Precondiciones:

- El Empleado Judicial ha iniciado sesión en su cuenta
- El Empleado Judicial ha realizado previamente un análisis

1. El Empleado Judicial entra a uno de sus análisis.
2. El Empleado Judicial elige los filtros a aplicar en los resultados completos.
3. El sistema muestra los resultados según estos filtros.

CU10: Borrar Análisis

Actor Principal: Empleado judicial.

Precondiciones:

- El Empleado Judicial ha iniciado sesión en su cuenta
- El Empleado Judicial ha realizado previamente un análisis

1. El Empleado Judicial entra a uno de sus análisis.
2. El Empleado Judicial presiona un botón para borrar el análisis.
3. El sistema pide una confirmación y si el Empleado judicial la acepta, borra el análisis.

CU11: Cambiar esquema de colores

Actor Principal: Empleado judicial.

Precondiciones:

- El Empleado Judicial ha iniciado sesión en su cuenta

1. El Empleado Judicial entra a sus preferencias de usuario.
2. El Empleado Judicial presiona el botón que indica el cambio del esquema de colores.

3. El Empleado Judicial elige una de las opciones presentadas y presiona el botón de 'aceptar'.
4. El sistema cambia los colores en los que se mostraran los resultados de la aplicación de detección de entidades.

CU12: Crear nueva nube de palabras

Actor Principal: Empleado judicial.

Precondiciones:

- El Empleado Judicial ha iniciado sesión en su cuenta
- El Empleado Judicial ha realizado previamente un análisis de detección de violencia.

1. El Empleado Judicial entra a uno de sus análisis.
2. El Empleado Judicial presiona un botón indicando que quiere crear un nuevo wordcloud.
3. El Empleado Judicial indica que palabras quiere omitir.
4. El sistema crea el wordcloud y este es descargado directamente al dispositivo del Empleado judicial.

Modelo de Dominio

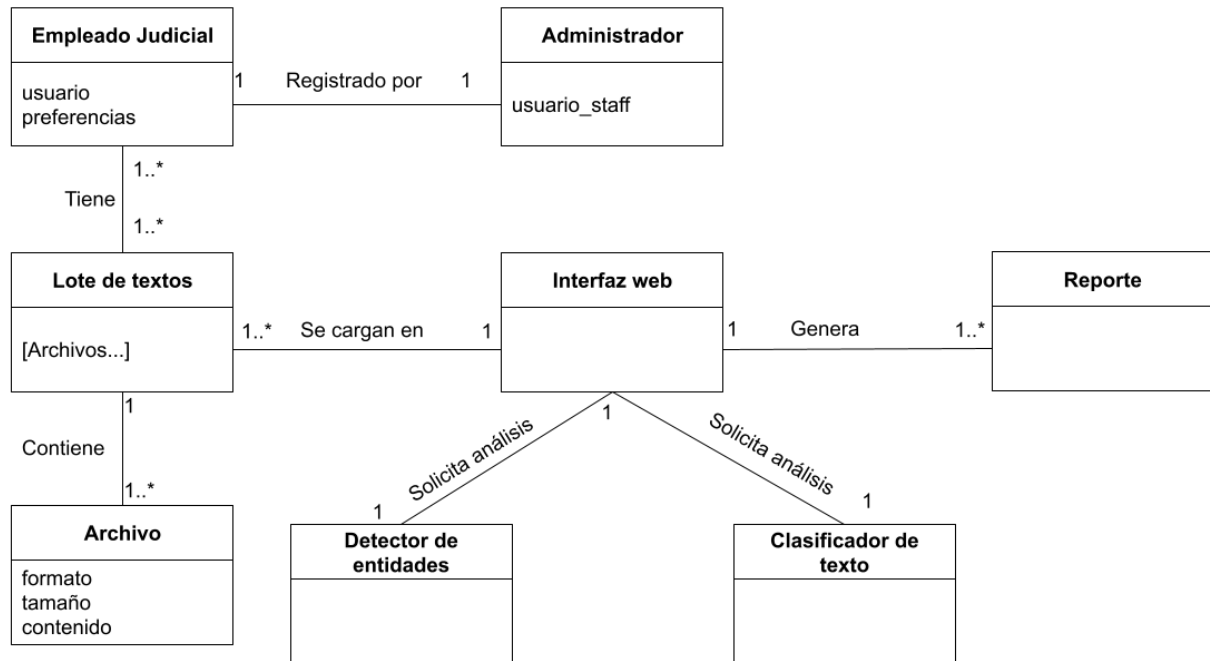


Diagrama de Componentes

