

**ESCUELA POLITÉCNICA NACIONAL**

**FACULTAD DE INGENIERÍA DE SISTEMAS**

**MIGRACIÓN A UN ENTORNO DE PRODUCCIÓN ÚNICO DE  
HERRAMIENTAS DE VIGILANCIA ESTRATÉGICA  
DESARROLLADOS COMO TRABAJOS DE TITULACIÓN EN LA  
FIS – PARTE II**

**COMPONENTE PRÁCTICO E IMPLEMENTACIÓN**

**TRABAJO DE INTEGRACIÓN CURRICULAR PRESENTADO COMO  
REQUISITO PARA LA OBTENCIÓN DEL TÍTULO DE INGENIERO DE  
SOFTWARE**

**BRYAN FERNANDO CEDEÑO MENDOZA**

**bryan.cedeno01@epn.edu.ec**

**DIRECTOR: EDISON FERNANDO LOZA AGUIRRE**

**edison.loza@epn.edu.ec**

**DMQ, agosto 2024**

## **CERTIFICACIONES**

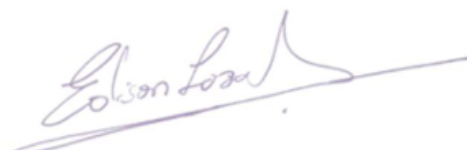
Yo, BRYAN FERNANDO CEDEÑO MENDOZA declaro que el trabajo de integración curricular aquí descrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.



---

**BRYAN CEDEÑO**

Certifico que el presente trabajo de integración curricular fue desarrollado por BRYAN FERNANDO CEDEÑO MENDOZA, bajo mi supervisión.




---

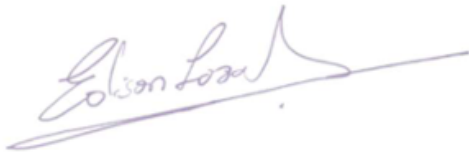
**EDISON LOZA**  
**DIRECTOR**

## DECLARACIÓN DE AUTORÍA

A través de la presente declaración, afirmamos que el trabajo de integración curricular aquí descrito, así como el (los) producto(s) resultante(s) del mismo, son públicos y estarán a disposición de la comunidad a través del repositorio institucional de la Escuela Politécnica Nacional; sin embargo, la titularidad de los derechos patrimoniales nos corresponde a los autores que hemos contribuido en el desarrollo del presente trabajo; observando para el efecto las disposiciones establecidas por el órgano competente en propiedad intelectual, la normativa interna y demás normas.



BRYAN FERNANDO CEDEÑO MENDOZA



EDISON FERNANDO LOZA AGUIRRE

## **DEDICATORIA**

Con profundo agradecimiento dedico este trabajo a mis padres, por creer siempre en mí y brindarme la oportunidad de desarrollar mis capacidades y habilidades. Siempre han sido mi fuente de inspiración y me han enseñado que hasta el más pequeño trabajo dignifica. A mis hermanos quienes han sido amigos y confidentes, siempre apoyándome en los momentos críticos, siendo admiradores incondicionales de mis avances y fracasos. Por su apoyo y amor incondicional que han sido fundamental en la culminación de este trabajo.

## **AGRADECIMIENTO**

Con gran satisfacción, me gustaría expresar mi agradecimiento a mi familia y con mención especial a mi madre, mujer que a pesar de enfrentarse a un sinnúmero de dificultades logro hacer de mi un hombre de bien, su apoyo, amor, convicciones, valores han sido un pilar fundamental en mi vida.

Agradezco a mi abuela y mis tías con quienes me críe, que fueron como madre y hermanas, con quienes compartí de los momentos más felices de pequeño y desde donde se sembró en mí el sentimiento de curiosidad y compasión. Sentimientos que hoy impulsan también mi vida profesional.

Agradezco a Cesar que a pesar de que ya no estás aquí me enseñaste a vivir en un momento de mi vida en el que no era tan fácil hacerlo. Me diste un camino que seguir y me enseñaste a reconocer las cosas que quiero en mi vida y cuáles no.

Mi gratitud todos mis amigos de carrera y aquellos de fuera que siempre han estado ahí para sacarme una sonrisa, a Verónica, una chica valiente, amorosa, aventurera que durante tantos años has sido un gran apoyo en mi vida, que me has brindado tu amistad incondicional.

Por otro lado expreso mi gratitud profunda a la Universidad EPN que me acogió en sus clases, me formo académicamente. A Edison Loza, mi tutor, con confiar en mis capacidades para llevar a cabo este desafiante proyecto. Le agradezco por toda la atención, apoyo y motivación que durante estos últimos meses que me han permitido cumplir con la finalización del proyecto y por consecuencia una etapa de mi vida.

# ÍNDICE DE CONTENIDO

CERTIFICACIONES.....	I
DECLARACIÓN DE AUTORÍA.....	II
DEDICATORIA.....	III
AGRADECIMIENTO.....	IV
ÍNDICE DE CONTENIDO.....	V
ÍNDICE DE FIGURAS .....	VI
ÍNDICE DE TABLAS .....	VIII
RESUMEN .....	IX
ABSTRACT .....	X
1 DESCRIPCIÓN DEL COMPONENTE DESARROLLADO .....	1
1.1 Objetivo general.....	2
1.2 Objetivos específicos .....	2
1.3 Alcance .....	2
1.4 Marco teórico .....	3
1.4.1 Vigilancia estratégica.....	3
1.4.2 Fase de Targeting .....	4
1.4.3 Web Scraping.....	5
1.4.4 Procesamiento de Lenguaje Natural .....	7
1.4.5 Estado la aplicación web Anterior.....	7
1.5 Herramientas y tecnologías utilizadas.....	8
1.5.1 Backend .....	8
1.5.2 Frontend.....	10
2 METODOLOGÍA .....	11
2.1 Desarrollo Incremental .....	11
2.2 Implementación de Metodología .....	13
2.2.1 Definición de Requisitos Generales.....	13
2.2.2 Asignación de requerimientos a los incrementos .....	15
2.2.3 Diseño de arquitectura del Módulo .....	18
2.2.4 Incrementos del sistema.....	19
3 RESULTADOS .....	31
3.1 Arquitectura.....	31

3.2	Descripción de la Aplicación .....	33
3.3	Pruebas.....	37
3.3.1	Resultados de Pruebas Unitarias .....	42
3.4	Recopilación de Esfuerzo .....	45
4	CONCLUSIONES Y RECOMENDACIONES.....	46
4.1	CONCLUSIONES .....	46
4.2	RECOMENDACIONES .....	47
5	REFERENCIAS BIBLIOGRÁFICAS .....	48
	ANEXOS .....	51

## ÍNDICE DE FIGURAS

Figura 1.	Proceso de Web Scraping .....	6
Figura 2.	Fases de la Metodología Incremental. ....	12
Figura 3.	Historia de Usuario asignadas a Incremento 1.....	16
Figura 4.	Historia de Usuario asignadas a Incremento 2.....	16
Figura 5.	Historia de Usuario asignadas a Incremento 3.....	17
Figura 6.	Historia de Usuario asignadas a Incremento 4.....	17
Figura 7.	Historia de Usuario asignadas a Incremento 5. ....	18
Figura 8.	Arquitectura Lógica del Módulo. ....	18
Figura 9.	Algoritmo de Lematización. ....	20
Figura 10.	Estructura HTML de diario Primicias. ....	20
Figura 11.	Archivo JSON de Noticias.....	21
Figura 12.	Algoritmo de Stemming.....	22
Figura 13.	Lista de Actores y Temas en formato JSON. ....	23
Figura 14.	Métodos HTTP de la API DOAJ. ....	25
Figura 15.	Estructura de archivo JSON de respuesta del servidor DOAJ.....	25
Figura 16.	Archivo JSON de los corpus de artículos científicos. ....	26
Figura 17.	Interfaz del Motor de Búsqueda. ....	28
Figura 18.	Servicios disponibles del aplicativo y su conexión con el cliente Apollo . ....	29
Figura 19.	Interfaz Actores.....	30
Figura 20.	Arquitectura Lógica del Sistema.....	31
Figura 21.	Modelo Lógica de la BD.....	33
Figura 22.	Interfase del Motor de Búsqueda. ....	33
Figura 23.	Ingreso de palabras claves. ....	34
Figura 24.	Lista de Actores y Temas.....	35
Figura 25.	Errores en el motor de búsqueda.....	35
Figura 26.	Creación de Actor desde la lista de actores. ....	36
Figura 27.	Lista de Actores en el módulo Actores.....	36

Figura 28. Lista de Temas en el módulo Temas. ....	37
Figura 29. Matriz de Actores y Temas. ....	37
Figura 30. Lista de actores de la prueba 1. ....	38
Figura 31. Lista de temas de la prueba 1. ....	38
Figura 32. Lista de actores de la prueba 2. ....	39
Figura 33. Lista de temas de la prueba 2. ....	39
Figura 34. Lista de actores de la prueba 3. ....	40
Figura 35. Lista de temas de la prueba 3. ....	40
Figura 36. Lista de actores de la prueba 4. ....	41
Figura 37. Lista de temas de la prueba 4. ....	41
Figura 38. Lista de actores de la prueba 5. ....	42
Figura 39. Lista de temas de la prueba 5. ....	42
Figura 40. Prueba Unitaria abstracción de corpus de noticias. ....	43
Figura 41. Prueba Unitaria algoritmo de Lematización. ....	44
Figura 42. Prueba Unitaria algoritmo de Stemming. ....	44
Figura 43. Menú principal de Sala de Reuniones Original. ....	56
Figura 44. Interfaz de Actores de Sala de Reuniones Original. ....	56
Figura 45. Interfaz de Temas de Sala de Reuniones Original. ....	56
Figura 46. Interfaz de Inicio de herramienta Targetbot Original. ....	57
Figura 47. Motor de búsqueda de herramienta Targetbot Original. ....	57
Figura 48. Listado de Actores y Temas de la Herramienta Targetbot original. ....	58



## ÍNDICE DE TABLAS

Tabla 1. Tecnologías Backend utilizadas.....	8
Tabla 2. Tecnologías Frontend utilizadas.....	10
Tabla 3. Historia de usuario épica MO-HUE01.....	13
Tabla 4. Historia de usuario épica LI-HUE02.....	14
Tabla 5. Historia de usuario épica IN-HUE03.....	14
Tabla 6. Historias de usuarios.....	14
Tabla 7. Tareas propuestas para el incremento 1.....	19
Tabla 8. Estado de HU para el incremento 1.....	20
Tabla 9. Tareas propuestas para el incremento 2.....	21
Tabla 10. Estado de HU para el incremento 2.....	22
Tabla 11. Tareas propuestas para el incremento 3.....	23
Tabla 12. Estado de HU para el incremento 3.....	26
Tabla 13. Tareas propuestas para el incremento 4.....	27
Tabla 14. Estado de HU para el incremento 4.....	28
Tabla 15. Tareas propuestas para el incremento 5.....	28
Tabla 16. Estado de HU para el incremento 5.....	30
Tabla 17. Evaluaciones de Funcionamiento: Consulta 1.....	38
Tabla 18. Evaluaciones de Funcionamiento: Consulta 2.....	38
Tabla 19. Evaluaciones de Funcionamiento: Consulta 3.....	39
Tabla 20. Evaluaciones de Funcionamiento: Consulta 4.....	40
Tabla 21. Evaluaciones de Funcionamiento: Consulta 5.....	41
Tabla 22. Historia de Usuario1.....	52
Tabla 23. Historia de Usuario 2.....	52
Tabla 24. Historia de Usuario 3.....	52
Tabla 25. Historia de Usuario 4.....	53
Tabla 26. Historia de Usuario 5.....	53
Tabla 27. Historia de Usuario 6.....	53
Tabla 28. Historia de Usuario 7.....	54
Tabla 29. Historia de Usuario 8.....	54
Tabla 30. Historia de Usuario 9.....	54
Tabla 31. Historia de Usuario 10.....	55
Tabla 32. Historia de Usuario 11.....	55

## RESUMEN

En cualquier organización, la necesidad de estar informados de los factores que puedan afectar positiva o negativamente los intereses de la organización es fundamental y la vigilancia estratégica ayuda a suplir esta necesidad. En este contexto, el presente estudio aborda la migración y actualización de una herramienta denominada 'Targetbot', con el objetivo de optimizar su integración en un sistema de sala de reuniones y potenciar la fase de targeting.

Targetbot fue desarrollada empleando técnicas de procesamiento del lenguaje natural y web Scraping para generar de manera automatizada una lista de actores y temas relevantes para una organización. Con el fin de garantizar una transición fluida y mantener las funcionalidades clave, se optó por una metodología de desarrollo incremental. Esta metodología permitió identificar y abordar de manera gradual las áreas que requerían atención. Se realizaron modificaciones significativas en la estructura, el código y la interfaz de usuario de la herramienta, con el fin de adaptarla a las nuevas necesidades y mejorar su rendimiento. Finalmente, los resultados obtenidos demuestran la exitosa integración de la aplicación con el sistema de sala de reuniones.

**PALABRAS CLAVE:** Vigilancia Estratégica, Web Scraping, Migración, Desarrollo Incremental, Targeting, Procesamiento de Lenguaje Natural, Actualización Tecnológica.

## **ABSTRACT**

In any organization, the need to be informed of factors that can positively or negatively impact the organization's interests is fundamental, and strategic surveillance helps to fulfill this need. In this context, this study addresses the migration and update of a tool called 'Targetbot', with the objective of optimizing its integration into a meeting room system and enhancing the targeting phase.

"Targetbot" was developed using natural language processing and web scraping techniques to automatically generate a list of relevant actors and topics for an organization. To ensure a smooth transition and maintain key functionalities, an incremental development methodology was adopted. This methodology allowed for the gradual identification and addressing of areas requiring attention. Significant modifications were made to the tool's structure, code, and user interface to adapt it to new needs and improve its performance. Finally, the results obtained demonstrate the successful integration of the application with the meeting room system.

**KEYWORDS:** Strategic surveillance, Web Scraping, Migration, Incremental development, Targeting, Natural language processing , Technology upgrade.

# 1 DESCRIPCIÓN DEL COMPONENTE DESARROLLADO

En un mundo empresarial dinámico y en constante evolución, la capacidad de las organizaciones para anticipar, comprender y adaptarse a los cambios en su entorno se convierte en un diferenciador estratégico fundamental [1], [2]. La Vigilancia Estratégica (VS) emerge como un proceso esencial mediante el cual las organizaciones no solo buscan información, sino que anticipan tendencias, identifican oportunidades y reducen riesgos asociados a la falta de conocimiento [3].

La VS se estructura en varias fases. El Targeting en el contexto de VS es el momento clave en el que se identifican las áreas prioritarias de interés para una empresa. Implica la detallada descripción y definición de los actores significativos, aquellos individuos o entidades con potencial para influir en la organización, junto con los temas que podrían tener un impacto positivo o negativo en su futuro [4]. Esta fase resulta crucial al establecer la dirección de la búsqueda de información, permitiendo concentrar los esfuerzos de análisis en áreas estratégicas específicas. La elaboración de listas de actores y temas marca el inicio de un análisis detallado, guiando la recolección de datos hacia fuentes pertinentes y significativas, facilitando así la toma de decisiones fundamentadas y estratégicas [2].

Aquí es donde una herramienta de generación de listas de actores y temas desempeña un papel crucial al delinear los puntos de enfoque para la búsqueda de información y así facilitar una toma de decisiones informadas y estratégicas [5]. Dado que, si bien se cuenta con una gran disponibilidad de información en Internet, las organizaciones por sí solas no tienen la capacidad para procesar toda esa información y analizar el impacto de los factores externos.

Entre las carencias en el proceso de VS destaca la manualidad con la que se lo realiza, es decir: las reuniones son poco estructuradas, existen a veces poca sincronía de información entre los participantes o personal gerencial, no se tiene una visión concreta que se desea discutir, existen complicaciones geográficas u horarias [6].

El presente proyecto propone la migración de “una herramienta de sugerencia de listas de actores y temas” hacia una plataforma integral permitiendo la actualización tecnológica y la integración en un entorno común de producción y así proporcionar un procesamiento más ágil y eficiente de los datos. Esta migración no solo implica una mejora en la operatividad, sino que también busca consolidar y mejorar el proceso de VS al integrar las

salidas de datos del aplicativo existente en un repositorio de datos centralizado. Esto proporcionará una visión más holística y accesible de la información recolectada, permitiendo una toma de decisiones más informada, precisa y más colectiva.

## **1.1 Objetivo general**

Migrar la herramienta de sugerencia de lista de actores y temas hacia una plataforma integral, utilizada durante la fase de Targeting de VS con base en fuentes de información académicas, generales y/o significativas.

## **1.2 Objetivos específicos**

1. Actualizar y homogeneizar las tecnologías empleadas en el front-end y back-end de la herramienta de sugerencia de actores y el entorno de producción Integral.
2. Identificar y resolver los errores y deficiencias de la herramienta, incluyendo la verificación de la coherencia de la información presentes en la herramienta de sugerencia de actores y temas.
3. Incorporar la salida de datos de la herramienta en un repositorio integrado que permita una visión global y accesible de la información recolectada.
4. Mejorar y estandarizar la interfaz de la herramienta para sugerir listas de actores y temas, asegurando su coherencia con el nuevo ambiente Integral.

## **1.3 Alcance**

El presente proyecto toma como punto referencia el proyecto: “Desarrollo de una herramienta para sugerir listas de actores y temas a ser usados en la fase de Targeting de Vigilancia Estratégica” desarrollado por Paul Cisneros y Gabriel Macias en su trabajo de titulación de la Escuela politécnica Nacional [5]. El alcance de este componente es la migración y actualización de dicha herramienta a un nuevo sistema centralizado donde cumplirá con los requerimientos iniciales del sistema e integrará su funcionalidad con la herramienta de sala de reuniones desarrollada en “Migración a un entorno de producción único de herramientas de vigilancia estratégica desarrollados como trabajos de titulación en la FIS”, trabajo desarrollado por Fernando Ponce como su trabajo de titulación de la Escuela Politécnica nacional [7].

Este componente se limitará a realizar una actualización tecnológica tanto de la interfaz gráfica como de REST API, realizar mejoras a nivel de código y habilitar funcionalidades que dejaron de funcionar. Bajo ningún concepto se pretende agregar funcionalidades nuevas a las del proyecto base.

El componente contara con una interfaz gráfica cuyo diseño estará en sintonía con la plataforma integral de sala de reuniones, permitirá la interacción con la REST API y habilitara el uso de la lista de actores y temas hacia el resto de la plataforma.

La REST API o backend permitirá generar una lista de actores y temas median el uso de fuentes de información preseleccionadas. Esta hará uso del procesamiento de lenguaje natural para generar sugerencias respecto a una temática que se desea investigar, el igual que el proyecto base su funcionalidad es ayudar al proceso de VS con la generación de la lista, pero no se debe considerar como un buscador de información.

## **1.4 Marco teórico**

### **1.4.1 Vigilancia estratégica**

El proceso de VS se refiere a la adquisición y utilización de información sobre eventos, tendencias y relaciones en el entorno externo de una organización. Este conocimiento asiste a la dirección en la planificación del curso futuro de la organización. La exploración del entorno permite comprender las fuerzas y factores relevantes de cambio externos y desarrollar respuestas efectivas para asegurar o mejorar la posición de la organización frente a dichos cambios [1].

La VS, aunque se puede confundir con actividades de obtención de información como inteligencia competitiva y de negocios, es una actividad complementaria a éstas ya que amplía su alcance para analizar información de todos los sectores del entorno externo. La VS está vinculada a un mejor rendimiento organizacional a condición de integrarse de manera efectiva con la estrategia organizacional, utilizando la información recopilada de manera pertinente en el proceso de planificación y así provocando un impacto significativo en la mejora de la comunicación y la discusión dentro de la organización sobre temas orientados al futuro. Es un método fundamental de aprendizaje organizacional, ya que la capacidad de adaptación de una organización depende de conocer e interpretar los cambios externos [3].

La actividad de VS no es un proceso monolítico, sino que implica tanto observar información (visualización) como buscar información activamente (búsqueda). Por tanto es útil distinguir entre cuatro modos de vigilancia organizativa :

- La “vigilancia no dirigida” o exposición a información sin una necesidad específica, explorando ampliamente el entorno para detectar señales de cambio temprano [6].
- La “vigilancia condicionada” o enfoque en información sobre temas seleccionados para evaluar su impacto en la organización, con el cambio a búsqueda si se considera significativo [6].
- La “búsqueda informal” o búsqueda activa de información para profundizar en el conocimiento de un problema, con el objetivo de determinar la necesidad de acción [8].
- La “búsqueda formal” o esfuerzo deliberado y planificado para obtener información específica sobre un problema, estructurado y enfocado en información detallada para respaldar decisiones o acciones [8] .

Cada modalidad se adapta a diversas necesidades y niveles de estructuración al obtener información para decisiones estratégicas. En este estudio, nos enfocamos en dos enfoques: "búsqueda enfocada" y "exploración". La "búsqueda enfocada" o el modo “formal” se utiliza cuando los gerentes están inmersos en decisiones y requieren información precisa para comprender el contexto, opciones e implicaciones. Este modo se define por una pregunta específica que limita la búsqueda general de información [2].

En contraste, el modo de "exploración" relacionada con la búsqueda “no dirigida” y “condicionada”, que se emplea cuando los gerentes carecen de preguntas específicas que orienten una búsqueda de información. Así, siguen de manera constante la información que podría anticipar cambios en el entorno empresarial o identificar amenazas y oportunidades [2]. Este proyecto pretende partir de este modo para su implementación dado que se tendrá una idea superficial de qué información se debería tomar en consideración, pero sin una definición del espectro de búsqueda.

#### **1.4.2 Fase de Targeting**

El "Targeting" en la VS implica delimitar el alcance, definiendo la parte del entorno que se relaciona con los objetivos estratégicos de la organización. A pesar de sugerencias de explorar todo el entorno, las limitaciones de recursos hacen que un escaneo completo sea impráctico y propenso a la sobrecarga de información. Diferentes métodos se han

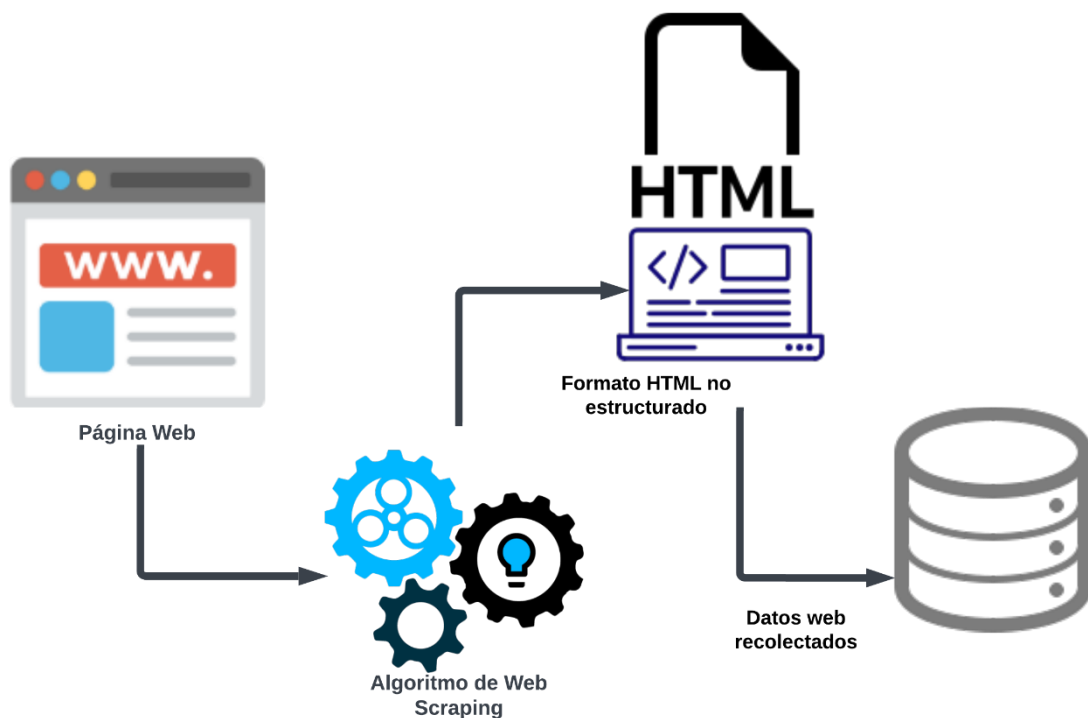
propuesto para el "Targeting", desde limitar fuentes de información hasta vigilar tendencias clave o eventos críticos. Un enfoque más integral, llamado "Target matrix" se centra en identificar y relacionar actores y temas relevantes para la VS [4]. Este método define actores como personas o entidades que podrían influir en el futuro de la organización y temas como centros de interés en consideración al futuro de esta. Al aplicar el "Target matrix", se obtiene una lista de actores y temas, junto con una matriz que indica las interrelaciones clave entre ellos.

### **1.4.3 Web Scraping**

La obtención de información automatizada desde el internet es una práctica frecuentemente realizada. El termino de "Web Scraping" es la designación más popular para esta práctica, que también es conocida como: minería web, web harvesting o screem scraping [9].

Web Scraping es la práctica de extracción o recuperación de información estructurada o no estructurada de cualquier manera alternativa a un usuario, interactuando con su navegador con el objetivo de investigación, análisis, colección personal, descubrimiento de conocimiento, entre otros. Normalmente consiste en un programa que automatiza consultas hacia un servidor y la información obtenida es dividida para su procesamiento y extracción de valor [10] como se lo observa en la Figura 1. Esta práctica incluye varias componentes de Informática como seguridad de información, análisis de lenguaje natural y análisis de datos.





**Figura 1. Proceso de Web Scraping**

El Web scraping permite obtener y procesar grandes cantidades de información ya que nos permite acceder a varios recursos a la vez y no solo una página web como lo haría un explorador normal [10]. Facilita además una extracción de información más personalizada que con un explorador normal o una búsqueda cotidiana.

Una API cumple la función de proveer información estructurada sobre un programa o sitio web, que en teoría reemplazaría la necesidad de utilizar web scraping o crear bots para la obtención de dicha información. Sin embargo, no todos los programas disponen de una API o estas no satisface totalmente las necesidades del investigador, el formato de datos no es el requerido, el límite en volumen de datos es insuficiente, etc. Por tanto la práctica de web scripting es necesaria ya que se puede ver la información del explorador, acceder mediante script, almacenarla y procesarla convenientemente.

Web scraping es una herramienta muy útil en ámbitos personales, corporativos e investigativos pero se debe tener en consideración que cuenta con varias variables legales. Para realizar web scraping sobre un sitio web se debe prestar atención a las políticas de privacidad, términos y condiciones para no incurrir en prácticas ilegales de difusión y acceso restringido de información [9].

#### **1.4.4 Procesamiento de Lenguaje Natural**

Dentro de la Inteligencia Artificial, el procesamiento del lenguaje natural (NLP) hace referencia a un sistema que permiten analizar, producir o tratar de entender un lenguaje como el español, ruso, inglés, entre otros. Hace uso de la lingüística computacional y modelos estadísticos que permite procesar el lenguaje y entender el significado de un extracto junto la intención y sentimientos de quien lo creo [11].

Dado que el lenguaje que manejamos es ambiguo, el NLP tiene varias limitaciones. El lenguaje generalmente es irregular, posee homónimos, metáforas, expresiones idiomáticas y gramaticales propias, que se deben reconocer desde el principio para un procesamiento más fino [11].

El NLP incluye componentes fundamentales como lo son el preprocesamiento de datos y la selección del algoritmo de aprendizaje. La primera hace referencia a la limpieza y preparación de datos donde se utilizan técnicas como tokenización, eliminación de palabras vacías, lematización, etiquetado de categoría gramatical o parceo para preprocesar los datos. Por otra parte, el algoritmo de aprendizaje permite extraer la información de valor con procesamientos como: análisis de sentimientos, modelado de lenguaje, reconocimiento de entidades, traducción automática, modelado de temas [12].

Los sistemas NLP tienen como entrada un texto, lenguaje hablado o entradas por teclado y su finalidad es traducir, comprender o representar el contenido de la entrada para construir DB o generar resúmenes o generar diálogos con los usuarios. En este proyecto se utiliza NLP para realizar el análisis de la información adquirida por Web scraping y mediante lematización abstraer los lemas más repetidos en y generar la lista de actores y temas.

#### **1.4.5 Estado la aplicación web Anterior**

El presente proyecto de migración tiene como punto de partida el aplicativo “Desarrollo de una herramienta para sugerir listas de actores y temas a ser usados en la fase de Targeting de Vigilancia Estratégica” [5]. Esta herramienta tiene como objetivo principal: “generar una lista de actores y temas a partir de noticias y artículos científicos y finalmente ser utilizada en la fase de targeting en la visión estratégica”.

La herramienta dispone de varios módulos entre los cuales se integra un motor de búsqueda que permite recolectar corpus de noticias o corpus de artículos científicos. Por otra parte, se cuenta con un módulo de NLP que permite realizar técnicas de lematización




y Stemming sobre los corpus de noticias, con el objetivo de reducirlos y extraer las raíces de las palabras contenidas en un fragmento de texto. Finalmente se cuenta con una interfase gráfica que me permite realizar la búsqueda y generar una lista de actores y temas respecto al tema de interés. Se bien se trata de un aplicativo útil en el proceso de vigilancia estratégica, la falta de mantenimiento ha producido que la aplicación deje de funcionar correctamente y necesite de una migración correctiva, tecnológica y de plataforma y ante recurrentes fallos de funcionamiento.







## 1.5 Herramientas y tecnologías utilizadas.


El presente trabajo está desarrollado utilizando una gama de herramientas y tecnologías que se describen a continuación.

### 1.5.1 Backend

**Tabla 1. Tecnologías Backend utilizadas**




Nombre	Descripción	Utilizado en	Versión Previa	Versión Migración
 Python	Lenguaje de programación de alto nivel, conocido por su sintaxis sencilla y legibilidad. Es ampliamente utilizado en ciencia de datos, aprendizaje automático, desarrollo web y automatización [13].	Esta tecnología es utilizada principalmente en el desarrollo del servidor del motor de búsqueda y la implementación de los algoritmos de Web Scraping, procesamiento de textos.	<2.7	3.12
 API de búsqueda DOAJ	Es una API provista por una de las editoriales académicas más importantes. Esta ofrece una fuente de información sobre revistas y artículos científicos al público de manera accesible [14].	Es una de las fuentes de información académica utilizada en la extracción de información.	N/A	N/A
 API Springer	El API ofrece acceso a una amplia gama de libros, revistas y artículos científicos mediante meta data entregada por el servidor [15].	Es una de las fuentes de información académica utilizada en la extracción de información.	N/A	N/A




 BeautifulSoup BeautifulSoup	Biblioteca de Python para el parsing de documentos HTML y XML. Se utiliza comúnmente para extraer datos de páginas web, construir rastreadores web, analizar el contenido de sitios web y recopilar datos para análisis [16].	Se utiliza para analizar el archivo HTML y extraer información de los elementos.	N/R	4.12.3
 Spacy	Biblioteca de Python para el procesamiento del lenguaje natural (NLP). Ofrece herramientas para tareas como tokenización, lematización, reconocimiento de entidades nombradas y análisis de dependencias [17].	Utilizada para desarrollar el algoritmo de lematización y Stemming.	N/R	3.7.5
 Apollo	Es un cliente de GraphQL que facilita la integración de GraphQL en aplicaciones Frontend. Permite el manejo de los datos es decir, su uso abraça fetching de datos, modificación de información y cacheo [18].	Permite realizar la conexión del Frontend con el servicio de base de datos.	1.9.1	4.9.4
 .Docker	Es una plataforma de contenedores que permite empaquetar aplicaciones y sus dependencias en contenedores aislados. Esto facilita la implementación, escalado y distribución de aplicaciones [19].	Permite empaquetar en contenedor el servidor GraphQL y la base de datos. Además del Servidor de motor de búsqueda.	19.03.8	24.0.2
 GraphQL GraphQL	Lenguaje de consulta para APIs que permite a los clientes solicitar exactamente los datos que necesitan. A diferencia de REST, GraphQL proporciona una forma más flexible y eficiente de obtener datos de una API [20].	Nos permite realizar consultas y gestión de datos en la base de datos.	15.0.0	16.8.1
 PostgreSQL PostgreSQL	Sistema de gestión de bases de datos relacionales (RDBMS) objeto-relacional avanzado. Utiliza SQL como lenguaje de consulta [21].	Utilizado para consultas a nuestra base de datos.	13.3	15.0

 Prisma	ORM (Object-Relational Mapper) para Node.js y TypeScript que permite interactuar con bases de datos de manera más intuitiva y productiva. Genera automáticamente código TypeScript a partir de un esquema de base de datos [22].	Desplegar la BD.	1.34.10	5.4.2
---	--	------------------	---------	-------

## 1.5.2 Frontend

**Tabla 2. Tecnologías Frontend utilizadas**

Nombre	Descripción	Utilizado en	Versión Previa	Versión Migración
 Angular	Un framework de desarrollo Frontend de código abierto que se utiliza para crear aplicaciones web de una sola página (SPA) Proporciona una estructura y un conjunto de herramientas para la creación de interfaces [23].	Framework del aplicativo	9.1.9	17.0.0
 Angular Material	Una biblioteca de componentes de interfaz de usuario basada en Angular proporciona un conjunto de componentes prediseñados como botones, formularios, tablas, tarjetas [24].	Creación de componentes de la interfaz.	9.1.3	17.0.3
 <b>Bootstrap</b> Bootstrap	Framework de desarrollo Frontend que proporciona una colección de clases CSS y componentes de JavaScript. Facilita la creación de diseños responsivos, grids, formularios, componentes de navegación y otros elementos sin mucho css [25].	Aplicar diseño a formularios y botones.	4.5.3	N/A

 <p>NPM</p>	<p>Un administrador de paquetes para el lenguaje de programación JavaScript, utilizado para instalar, actualizar y gestionar dependencias de proyectos Node.js [26].</p>	<p>Instalador de paquetes de librerías utilizadas</p>	<p>N/E</p>	<p>10.2.0</p>
 <p>PrimeNG</p>	<p>Una biblioteca de componentes de interfaz de usuario para Angular, similar a Angular Material, que ofrece una amplia gama de componentes como cuadros de diálogo, menús, tablas de datos, gráficos [27].</p>	<p>Creación de componentes como: barra de navegación, menús, etc.</p>	<p>9.1.0</p>	<p>17.0.0</p>
 <p>Typescript</p>	<p>Lenguaje de programación inspirado en JavaScript que agrega características de tipado estático, clases, interfaces y otras estructuras de programación orientadas a objetos [28].</p>	<p>Lenguaje de desarrollo de la lógica de los componentes.</p>	<p>3.7.5</p>	<p>5.2.2</p>

## 2 METODOLOGÍA

Para el desarrollo de la presente aplicación web se tomó en consideración algunos factores importantes como: la naturaleza del proyecto, recursos y objetivos planteados para escoger la metodología. A continuación, se detalla la metodología utilizada y su beneficio en el desarrollo y finalmente las distintas fases de su aplicación.

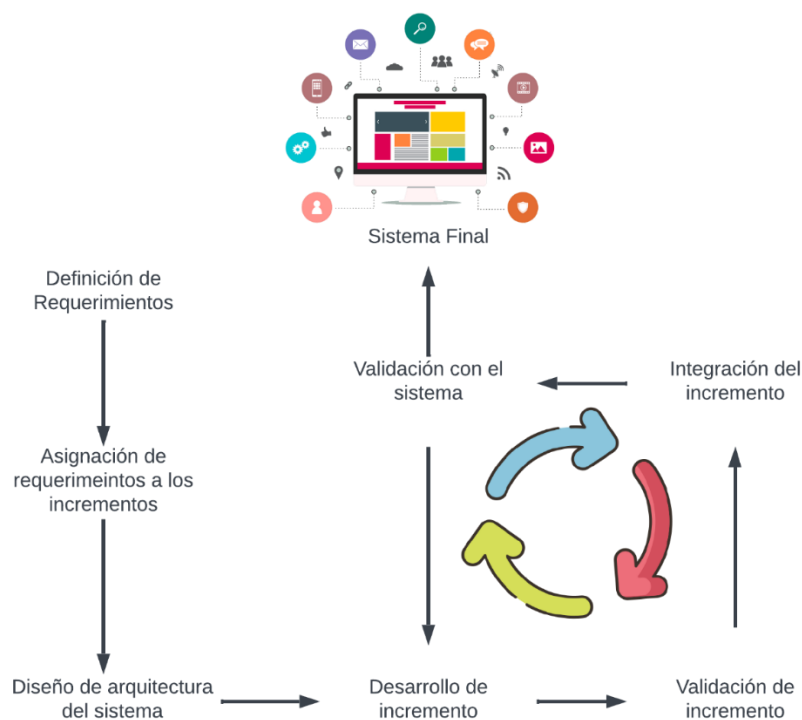
### 2.1 Desarrollo Incremental

El desarrollo incremental consiste en separar el proyecto en fracciones pequeñas o servicios en los que se puede separar el sistema. En un principio se tiene una idea general de cuáles son los objetivos principales del sistema. Entonces, se identifican cuáles son los más importantes o requieren una atención temprana y se procede con su desarrollo en distintas fases o incrementos. Cada incremento suple o abarca una o un conjunto de funcionalidades del sistema. Al finalizar un incremento se realiza la integración con los anteriores permitiendo aumentar la funcionalidad con cada entrega y permitiendo que el sistema evolucione y mejore con cada entrega [29].

Al ser una metodología que permite realizar entregas funcionales de software no es necesario que el proyecto esté completamente finalizado para que el cliente pueda hacer uso de este. Además, los requerimientos pueden ser modificados para aquellos incrementos posteriores que aún están en espera ya que a medida que cada incremento es desplegado nos permite clarificar los requerimientos faltantes [30].

Entre otras ventajas de este del tipo de desarrollo incremental, podemos citar la seguridad contra fallos en las principales funcionalidades del sistema ya que estas son las que se entregan al principio y que conllevan más pruebas de funcionamiento [30].

La metodología de desarrollo incremental empieza con una definición de los requerimientos del sistema de forma general y se identificaran las funcionalidades más importantes o las cuales se desea poner a disposición tempranamente. A continuación, se agrupa cada funcionalidad o grupo de funcionalidades en incrementos considerando su importancia [29]. Como se explica más adelante del documento, el sistema deberá regirse por una arquitectura que debe plantearse acorde a las necesidades del sistema. Entonces se procede iterativamente con el desarrollo del incremento del sistema, validando la o las funcionalidades del incremento y su validación con el resto del sistema en cada entrega. Este proceso lo podemos evidenciar en la Figura 2.



**Figura 2. Fases de la Metodología Incremental.**

Al tratarse de un sistema heredado, se combina el desarrollo incremental con la metodología Butterfly [31], que proporciona una guía para la migración de sistemas heredados a un sistema objetivo de manera simple, rápida y segura. La principal premisa de esta metodología es que durante el proceso de migración se elimina la necesidad de que tanto el sistema heredado y el sistema objetivo accedan al sistema de información simultáneamente.

Las fases de esta metodología se describen a continuación y se implementaran paralelamente al desarrollo incremental.

- Preparación para la migración.
- Comprender la semántica del sistema heredado y desarrollar los esquemas de datos del sistema de destino.
- Crear un almacén de datos de muestra basado en los datos del sistema de destino.
- Migrar de forma incremental todos los componentes del sistema heredado a la arquitectura de destino.
- Migración gradual de datos heredados al sistema de destino.
- Transición al sistema de destino.

## **2.2 Implementación de Metodología**

### **2.2.1 Definición de Requisitos Generales**

El proyecto consiste en la migración de la “Herramienta para sugerir listas de actores y temas a ser usados en la fase de Targeting de Vigilancia Estratégica” [5]. por tanto se parte de un conjunto de requerimientos bien definidos que satisfacen las necesidades del sistema. El aplicativo usa un motor de búsqueda que analiza noticias y artículos científicos para generar una lista personalizada de actores y temas vinculados a un tema de interés específico. Este constara del “motor de búsqueda”, el “generador de lista de actores y temas” y “filtros por palabras claves”.

Tomando en consideración a los requerimientos del sistema se pueden generar las siguientes historias de usuarios Épicas. Detalladas en las Tablas 3, 4, 5.

**Tabla 3. Historia de usuario épica MO-HUE01**



Identificador: MO-HUE01	Historia de Usuario Épica
Título de Historia: Desarrollar un motor de búsqueda de temáticas relacionadas a los intereses de las organizaciones.	
Descripción: Yo como usuario quiero disponer de un motor de búsqueda para generar un corpus basado en noticias y artículos científicos.	

**Tabla 4. Historia de usuario épica LI-HUE02**

Identificador: LI-HUE02	Historia de Usuario Épica
Título de Historia: Desarrollar un módulo de extracción de entidades para generar listas de actores y temas.	
Descripción: Yo como usuario quiero producir una lista de los principales actores y temas recurrentes en la búsqueda de una temática relacionada a la organización.	

**Tabla 5. Historia de usuario épica IN-HUE03**

Identificador: IN-HUE03	Historia de Usuario Épica
Título de Historia: Desarrollo de un entorno de producción único de herramientas de vigilancia estratégica.	
Descripción: Yo como usuario quiero disponer de un entorno de producción único que permita la interacción de las herramientas de generación de lista de actores y temas y la sala de reuniones "Targeting" para complementar el proceso de vigilancia estratégica.	

A partir de las historias de usuario épicas se desglosan las siguientes historias de usuarios que permiten cumplir con los requerimientos funcionales. Cada historia de usuario consta con una importancia o prioridad que ayuda a inferir el orden de desarrollo, el grupo de funcionalidad al que pertenece y si forma parte de la funcionalidad principal o complementaria del sistema, además de un marcador de estimación relacionado a la complejidad. En la Tabla 6 se observa las historias de usuarios recabadas.

**Tabla 6. Historias de usuarios**

Identificador	Historia de usuario	Prioridad	Estimación
LI-HU01	Yo como usuario quiero lematizar un texto para identificar las recurrencias en las que una palabra aparece en un texto	A	8
MO-HU01	Yo como usuario quiero adquirir o sustraer el texto de una noticia en internet para crear un corpus de la noticia.	A	8
LI-HU02	Yo como usuario quiero aplicar Stemming a un texto para reducir el vocabulario del corpus y adquirir las raíces de las palabras que conforman.	A	5
LI-HU03	Yo como usuario deseo verificar la duplicación en la lista de actores y temas para evitar actores en la lista de temas.	M	3

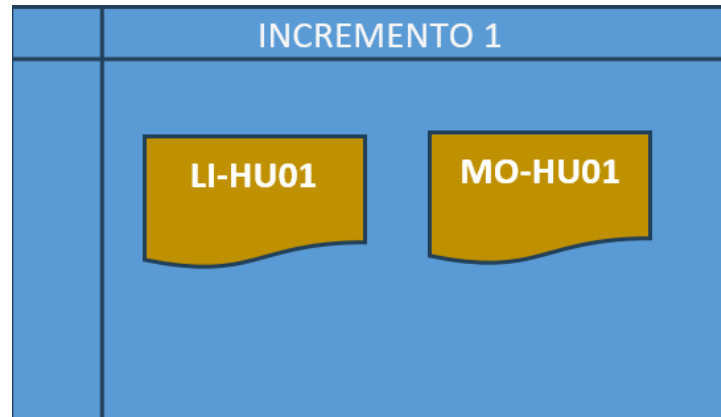
LI-HU04	Yo como usuario quiero obtener una lista de actores y temas para posteriormente ser analizados con el equipo de trabajo.	M	5
MO-HU02	Yo como usuario quiero ingresar un conjunto de palabras claves para mejorar la búsqueda en los corpus de texto.	B	2
MO-HU03	Yo como usuario quiero adquirir o sustraer texto de artículos científicos para crear un corpus de texto científico.	M	5
IN-HU01	Yo como usuario quiero seleccionar el tipo de fuente de consulta para filtrar entre noticias y artículos científicos.	M	1
IN-HU02	Yo como usuario quiero ingresar una temática de interés para mi organización para enterarme de los actores y temas relacionados.	M	2
IN-HU03	Yo como usuario quiero visualizar un indicador para enterarme del estado actual de la búsqueda.	B	1
IN-HU04	Yo como usuario quiero poder guardar los actores y temas generados a partir de la consulta para ser utilizados en la creación de la matriz de actores y temas.	A	5

## 2.2.2 Asignación de requerimientos a los incrementos

Una vez establecidas las historias de usuario e identificadas las funcionalidades más importantes para el sistema, se procede a describir los incrementos en los que se entrega el proyecto. A continuación se detalla la funcionalidad de cada incremento y las historias de usuario asociadas.

### Incremento 1

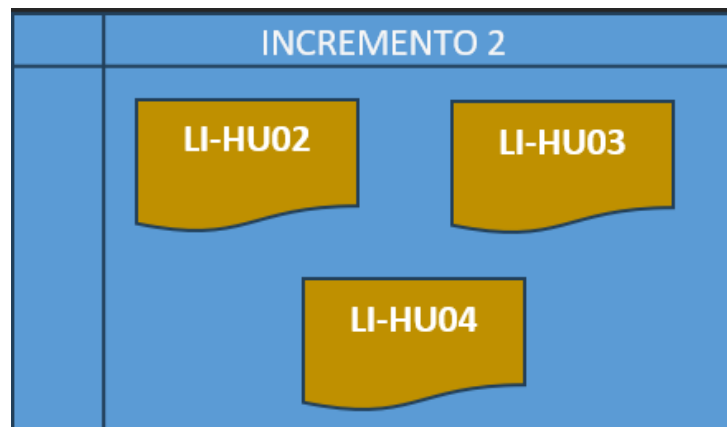
Construir un módulo de procesamiento de texto que permita realizar la lematización de texto quemado y conteo de frecuencia de palabras. Además, desarrollar un mecanismo de extracción de texto de fuentes en línea mediante web scraping. Para este incremento se toman en consideración las siguientes historias de usuario en la Figura 3.



**Figura 3. Historia de Usuario asignadas a Incremento 1.**

### **Incremento 2**

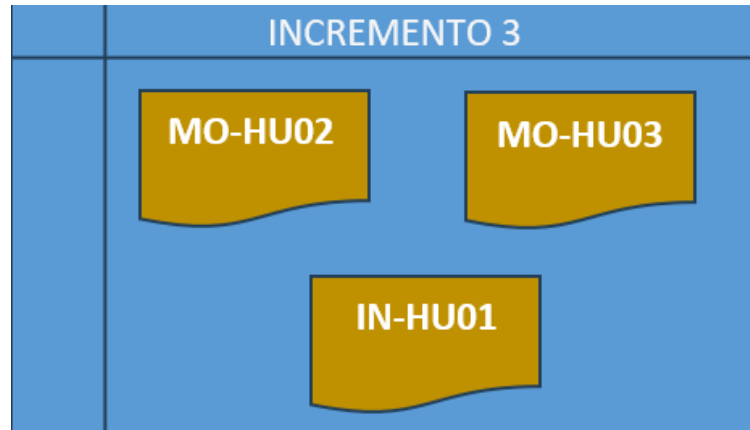
Generar un reporte que, mediante técnicas de minería de datos, identifique los temas predominantes y actores más relevantes de un corpus de texto. Así generar una lista evitando duplicidad. Para este incremento se toman en consideración las siguientes historias de usuario en la Figura 4.



**Figura 4. Historia de Usuario asignadas a Incremento 2.**

### **Incremento 3**

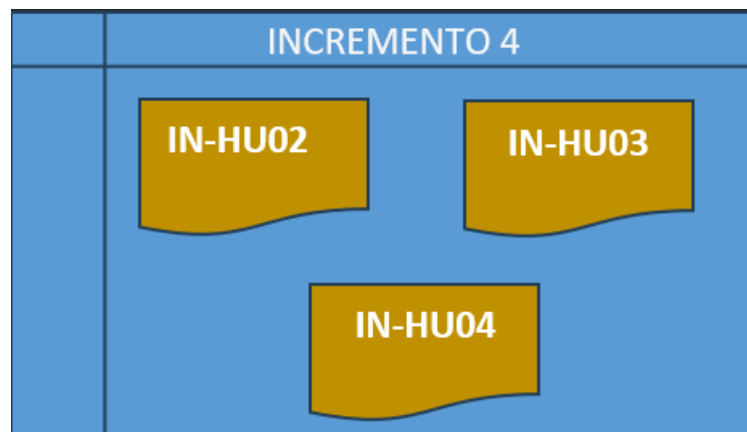
Implementar la función de búsqueda personalizada mediante el uso de palabras clave para poder refinar la exactitud de los resultados. Habilitar la búsqueda en artículos científicos como Doaj para enriquecer las fuentes desde donde se obtiene la información. Para este incremento se toman en consideración las siguientes historias de usuario en la Figura 5.



**Figura 5. Historia de Usuario asignadas a Incremento 3.**

#### **Incremento 4**

Implementar una interfaz dentro de la sala de reuniones “targeting” que me permita interactuar con el motor de búsqueda y se visualice la lista de actores y temas generados. El sistema debe indicar el estado de las búsquedas y tiempo estimado de espera del proceso. Adicionalmente se podrán seleccionar y guardar los actores y temas relevantes, según el criterio del usuario, dentro del sistema para ser utilizados en la generación de matriz de actores y temas. Para este incremento de toman en consideración las siguientes historias de usuario en la Figura 6.



**Figura 6. Historia de Usuario asignadas a Incremento 4.**

#### **Incremento 5**

Realizar la integración del módulo generación de actores y temas con la arquitectura de sala de reuniones “targeting”, verificar la conexión entre servidores y su correcta funcionalidad con el resto del sistema. Para este incremento de toman en consideración las siguientes historias de usuario en la Figura 7.

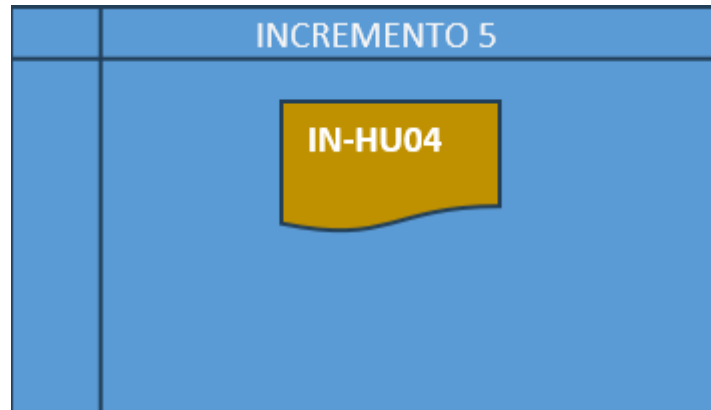


Figura 7. Historia de Usuario asignadas a Incremento 5.

### 2.2.3 Diseño de arquitectura del Módulo

Para la migración se propone la siguiente arquitectura cliente servidor. En la Figura 8 se observa la disposición de la arquitectura.

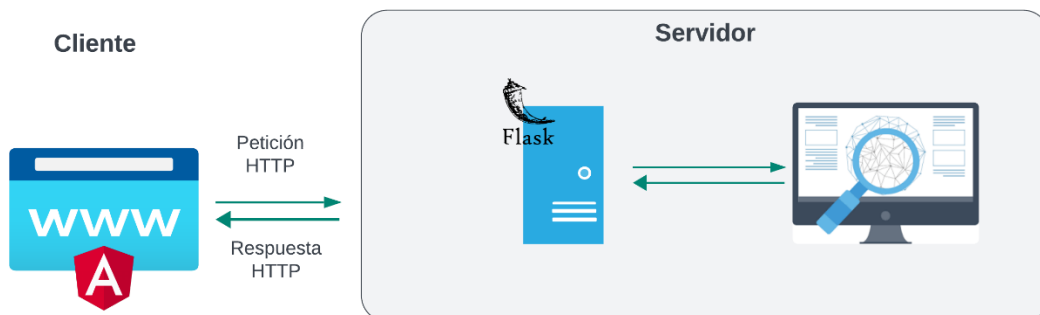


Figura 8. Arquitectura Lógica del Módulo.

La integración de las herramientas de VS está formada por dos componentes principalmente. A continuación, se describe la arquitectura del módulo generador de lista de actores y temas mientras que en el capítulo 3 del documento se describe la arquitectura resultante de la migración en un entorno de producción único.

En esta arquitectura podemos observar que el módulo encargado de realizar la recolección, análisis y procesamiento de información utilizado en el proceso de targeting limita la información importante alineada con los objetivos estratégicos y genera una lista de actores y temas encontrados en Internet.

El flujo empieza con una petición HTTP hacia la API Rest realizada en Flask. En este, la lógica de negocio procesará la información recibida y procede a realizar WebScraping en las fuentes de información. El servidor devuelve la información en forma de actores y temas que luego es utilizada en el lado de Targeting.

## 2.2.4 Incrementos del sistema

### Incremento 1

Este incremento tiene como función principal proveer al sistema de un mecanismo de lematización de texto y la extracción de noticias desde páginas web utilizando web scraping como herramienta de extracción. Para esto se establece un esfuerzo de 30 horas para la entrega del incremento y se establecen las siguientes tareas que se observan en la Tabla 7.

**Tabla 7. Tareas propuestas para el incremento 1**

Identificador de HU	Historia de usuario	Tareas	Estimación de Horas
LI-HU01	Yo como usuario quiero lematizar un texto para identificar las recurrencias en las que una palabra aparece en un texto	<ul style="list-style-type: none"> <li>Entender los fundamentos de Lematización de un texto.</li> <li>Implementar un algoritmo con la librería Spacy que permita realizar la lematización de un texto.</li> <li>Realizar un conteo de recurrencias de cada palabra del texto para identificar posibles actores y temas.</li> <li>Formar grupo de frases de 1, 2 ,3 palabras y evaluar su recurrencia.</li> </ul>	15
MO-HU01	Yo como usuario quiero adquirir o sustraer el texto de una noticia en internet para crear un corpus de la noticia.	<ul style="list-style-type: none"> <li>Crear peticiones http hacia paginas objetivos de los principales periódicos.</li> <li>Identificar cuáles son los elementos http que cargan el texto correspondiente a una noticia dependiendo del web objetivo.</li> <li>Realizar la extracción de texto de las etiquetas http mediante la librería BeautifulSoup.</li> <li>Concatenar cada fragmento de texto de las etiquetas en un único corpus de noticia.</li> </ul>	15

Durante el desarrollo del incremento se implementa un algoritmo de lematización de texto, como se observar en la Figura 9. Este permite procesar el texto mediante la librería Spacy que asigna un lema a cada palabra, este lema es la forma base o lematizada de la palabra en el diccionario [32]. Estas palabras a su vez, serán filtradas, siendo desechadas aquellas que correspondan a fechas, números, porcentajes, entre otras entidades que no se

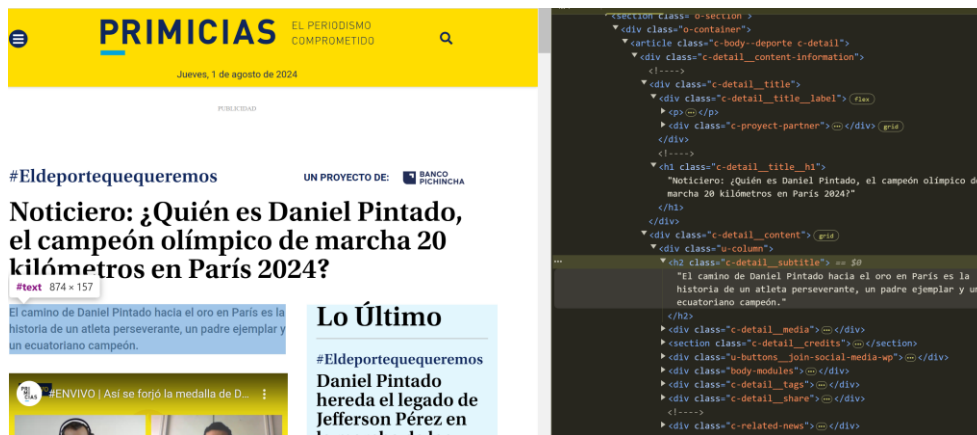
deberían considerar como futuros actores o temas y se cuenta la frecuencia del texto de la entidad y su etiqueta.

```

doc = nlp(texto)
print("\n\n****Entidades que pueden ser stakeholders (partes interesadas) AUTORES****\n")
ubicacionesTemasTotales = []
autoresTotales = Counter(palabra.text+" "+palabra.label_ for palabra in doc.ents
                          if palabra.label_ != 'MISC' and
                          palabra.label_ != 'DATE' and palabra.label_ != 'CARDINAL' and
                          palabra.label_ != 'ORDINAL' and palabra.label_ != 'PERCENT' and
                          palabra.label_ != 'QUANTITY')
for nombre in autoresTotales:
    if(autoresTotales[nombre] > 2):
        ubicacionesTemasTotales.append(nombre)
    
```

**Figura 9. Algoritmo de Lematización.**

Para el desarrollo de la herramienta de web scraping se utiliza como fuente principal el diario primicias, diario de noticias digital, se puede observar en la Figura 10 que se identifica un patrón de cuáles son los elementos HTML que contiene el texto correspondiente a la noticia. Nuestros elementos de interés serán: el elemento “div” con clase “body-modules” que es el contenedor principal de la noticia y los respectivos elementos “p” que tienen el texto de la noticia.



**Figura 10. Estructura HTML de diario Primicias.**

Durante la validación del incremento se revisa que se hayan cumplido cada una de las tareas establecidas como se observa en la Tabla 8. Cada una de las tareas y por consiguiente la funcionalidad del incremento se realizaron correctamente.

**Tabla 8. Estado de HU para el incremento 1**

Identificador	Estado
LI-HU01	Completo
MO-HU01	Completo

Para validar este incremento se dispuso de un texto propositivo y se procesa mediante el algoritmo propuesto de lematización, el texto se procesa y se identifican las concurrencias de la palabra en el texto. Adicionalmente, como se observa en la Figura 11 el estado actual del sistema nos permite realizar la extracción de un corpus de noticia del sitio web de “Diario Primicias” y almacenarlo en un archivo JSON.

```
{
  "Diario": "Primicias",
  "Pais": "Ecuador",
  "Fecha": "2020-2-28",
  "Titulo": "El coronavirus habría llegado a Italia a mediados de enero, según científicos",
  "Texto": "Investigadores de un hospital especializado en Milán, en el norte de Italia, lograron aislar la versión",
},
{
  "Diario": "Primicias",
  "Pais": "Ecuador",
  "Fecha": "2020-2-28",
  "Titulo": "Los ciclistas del Tour EAU completaron los controles médicos para detectar casos de coronavirus",
  "Texto": "Los miembros de la organización del Tour de los Emiratos Árabes Unidos (EAU) y representantes de los me
```

**Figura 11. Archivo JSON de Noticias.**

Al ser el primer incremento, la integración con el resto de las componentes se lo evidencia en las entregas posteriores. Por otra parte, las fuentes de noticias que se procederá a agregar en futuros incrementos deben cumplir con cierto requerimiento. Se observa que algunas fuentes como “The New York Times” tienen mecanismo de protección contra web scraping y otros como El Comercio tienen un motor de búsqueda de noticias interno que está desactualizado provocando errores de búsqueda. Por tanto, se debe seleccionar nuevas fuentes de noticias que permitan hacer la minería de datos necesaria.

**Incremento 2**

Este incremento tiene como función identificar los temas predominantes y actores más relevantes de un corpus de texto y producir una lista evitando duplicidad. Para esto se establece un esfuerzo de 22 horas para la entrega del incremento y se establecen las siguientes tareas que se observan en la Tabla 9.

**Tabla 9. Tareas propuestas para el incremento 2**

Identificador de HU	Historia de usuario	Tareas	Estimación de Horas
LI-HU02	Yo como usuario quiero aplicar Stemming a un texto para reducir el vocabulario del corpus y adquirir las	<ul style="list-style-type: none"> <li>Entender y comprender la técnica de Stemming.</li> <li>Implementar un algoritmo con la librería nltk que permita realizar la Stemming de un texto.</li> </ul>	12



	raíces de las palabras que conforman.	<ul style="list-style-type: none"> <li>Identificar o clasificar palabras que correspondan a actores y entidades que pertenezcan a Temas.</li> </ul>	
LI-HU03	Yo como usuario deseo verificar la duplicación en la lista de actores y temas para evitar actores en la lista de temas.	<ul style="list-style-type: none"> <li>Realizar un conteo de recurrencias de cada palabra del texto para identificar la duplicidad de actores y temas.</li> <li>Desarrollar un mecanismo que elimine los elementos duplicados tanto en la lista de actores y la lista de temas.</li> </ul>	
LI-HU04	Yo como usuario quiero obtener una lista de actores y temas para posteriormente ser analizados con el equipo de trabajo.	<ul style="list-style-type: none"> <li>Extraer los actores y temas recabados del corpus de noticias.</li> <li>Disponer de la lista de actores y temas en formato JSON y así facilitar el uso en los siguientes procesos.</li> </ul>	10

Para el desarrollo de este incremento se implementa el algoritmo de Stemming como se observa en la Figura 12. En el incremento anterior se obtuvo un listado de palabras posibles a ser actores y temas, a partir de este listado se aplica el “Stemmer” para adquirir la raíz de las palabras eliminando sufijo y prefijos [33]. Posteriormente se filtra aquellas raíces repetidas y la lista de actores y temas correspondiente a las raíces únicas son almacenadas.

```

autoresStemming = []
raicesPalabras = [spanishstemmer.stem(token) for token in nombres]
for i, rp in enumerate(raicesPalabras):
    for pc in palabrasClave:
        if rp == pc:
            autoresStemming.append(
                nombres[i]+" "+nombres[i+1]+" "+nombres[i+2])
print (autoresStemming)

```

**Figura 12. Algoritmo de Stemming.**

Durante la validación se determina que el cumplimiento de cada tarea establecidas fue completado, como se observa en la Tablas 10. El algoritmo de “Stemming” da un resultado aceptable, pero se busca mejorar la detección de temas repetitivos en los siguientes incrementos.

**Tabla 10. Estado de HU para el incremento 2**

Identificador	Estado
LI-HU02	Completo
LI-HU03	Completo

Como se puede observar en la Figura 13 se genera una lista de temas y actores con una cantidad equivalente al número de ocurrencias encontradas en el corpus de noticias.

```

Lematizacion > {} ActoresTemasjson > ...
1 {
2   "Actores y temas": [
3     {
4       "Actores": [
5         [
6           "Google",
7           9
8         ],
9         [
10          "OMS",
11          9
12        ],
13        [
14         "FAO",
15         9
16        ],
17        [
18         "NHS",
19         9
20        ],
21        [
22         "BCE",
23         9
24        ],
25        [
26         "UIDE",
27         6
28        ],
29        [
30         "Babahoyo",
31         6
32        ],
206      "Temas": [
207        [
208          "emergencia sanitario",
209          29
210        ],
211        [
212          "arroz azúcar",
213          24
214        ],
215        [
216          "caña azúcar",
217          19
218        ],
219        [
220          "actividad físico",
221          16
222        ],
223        [
224          "redar social",
225          16
226        ],
227        [
228          "producto lácteo",
229          14
230        ],
231        [
232          "dieta diario",
233          13
234        ],
235        [
236          "fruto seco",
237          12
238        ],
239        [
240          "nutricionista recomendar",
241          12
242        ],
243        [
244          "langosta centroamericano",
245          12

```

**Figura 13. Lista de Actores y Temas en formato JSON.**

Por otra parte, se desarrolla la integración con el incremento inicial. Permitted al algoritmo de procesamiento de lenguaje natural tomar como fuente de alimentación el corpus de noticias generado por la herramienta de web scraping y no solo usar texto quemado en código. Esto nos permite generar o hacer un análisis del lenguaje a partir de un sin número de texto disponibles en la web.

**Incremento 3**

Este incremento tiene como función personalizar la búsqueda mediante palabras claves y adicionar fuentes de artículos científicos a la opción de fuentes de búsqueda. Para esto se establece un esfuerzo de 30 horas para la entrega del incremento y se establecen las siguientes tareas que se observan en la Tabla 11.

**Tabla 11. Tareas propuestas para el incremento 3**

Identificador de HU	Historia de usuario	Tareas	Estimación de Horas
MO-HU02	Yo como usuario quiero ingresar un conjunto de palabras claves para mejorar la búsqueda en los corpus de texto.	<ul style="list-style-type: none"> <li>• Agregar un mecanismo de ingreso de palabras claves utilizadas para mejorar la búsqueda.</li> <li>• Interpolar el resultado de procesos anteriores (Lematización y Stemming) con el listado de palabras claves.</li> </ul>	8
MO-HU03	Yo como usuario quiero adquirir o sustraer texto de artículos científicos para crear un corpus de texto científico.	<ul style="list-style-type: none"> <li>• Identificar cuáles son las fuentes de artículos científicos disponibles.</li> <li>• Revisar la disponibilidad de Apis disponibles de las fuentes de artículos científicos.</li> <li>• Revisar la documentación disponible de las Apis.</li> <li>• Realizar peticiones a la Api y verificar el tipo de respuesta que entrega el servidor.</li> <li>• Desarrollar un mecanismo de petición hacia la Api y Recopilación de texto.</li> <li>• Almacenar en texto recabado en un formato disponible a ser usado por el resto del sistema.</li> </ul>	15
IN-HU01	Yo como usuario quiero seleccionar el tipo de fuente de consulta para filtrar entre noticias y artículos científicos.	<ul style="list-style-type: none"> <li>• Desarrollar un flujo de respuesta dependiendo de la selección de la fuente de información seleccionada.</li> <li>• Habilitar y deshabilitar las fuentes de consulta dependiente de la selección del usuario.</li> </ul>	7

Durante el desarrollo de este incremento se implementa un mecanismo que mejora la adquisición de actores y temas realizando coincidencias con un grupo de palabras claves. En el incremento anterior se obtuvieron la raíz de las palabras abstraídas de un texto y como se observa en la Figura 12 , se realiza una comparación entre la raíz de las palabras con una lista de palabras claves y si coinciden se tomará la palabra, no la raíz, para así capturar posibles nombres compuestos y frases relacionadas con las palabras claves.

Para el desarrollo de abstracción de texto a partir de fuentes académicas se consideraron fuentes como Doaj y Springer. Como se puede observar en la Figura 14 para el caso de la Api proporcionada por la fuente DOAJ, se cuenta con varios métodos o peticiones que se pueden realizar hacia el servidor y recabar información sobre sus artículos y revistas [14].

CRUD Journals		^
GET	/api/journals/{journal_id}	Retrieve a journal by ID
Search		^
GET	/api/search/applications/{search_query}	Search your applications [Authenticated, not public]
GET	/api/search/articles/{search_query}	Search articles
GET	/api/search/journals/{search_query}	Search journals

**Figura 14. Métodos HTTP de la API DOAJ.**

Dependiendo del tipo de petición se obtiene determinado resultado. A continuación, se realiza una petición al servidor y esta entrega un archivo JSON como se observa en la Figura 15. El archivo contiene claves o parámetros que albergan información relevante del artículo. Los parámetros necesarios para extraer nuestro texto de artículo son “bibjson” que contine todos los artículos encontrados y el “abstract” que contiene el fragmento de texto de interés de cada artículo.

```

{
  "page": 0,
  "pageSize": 0,
  "query": "string",
  "results": [
    {
      "bibjson": {
        "alternative_title": "string",
        "apc": {
          "currency": "string",
          "max_price": "string"
        },
        "apc_url": "string",
        "author_pays": "string",
        "author_pays_url": "string",
        "copyright": {
          "author_retains": "string",
          "url": "string"
        }
      }
    }
  ]
}

```

**Figura 15. Estructura de archivo JSON de respuesta del servidor DOAJ.**

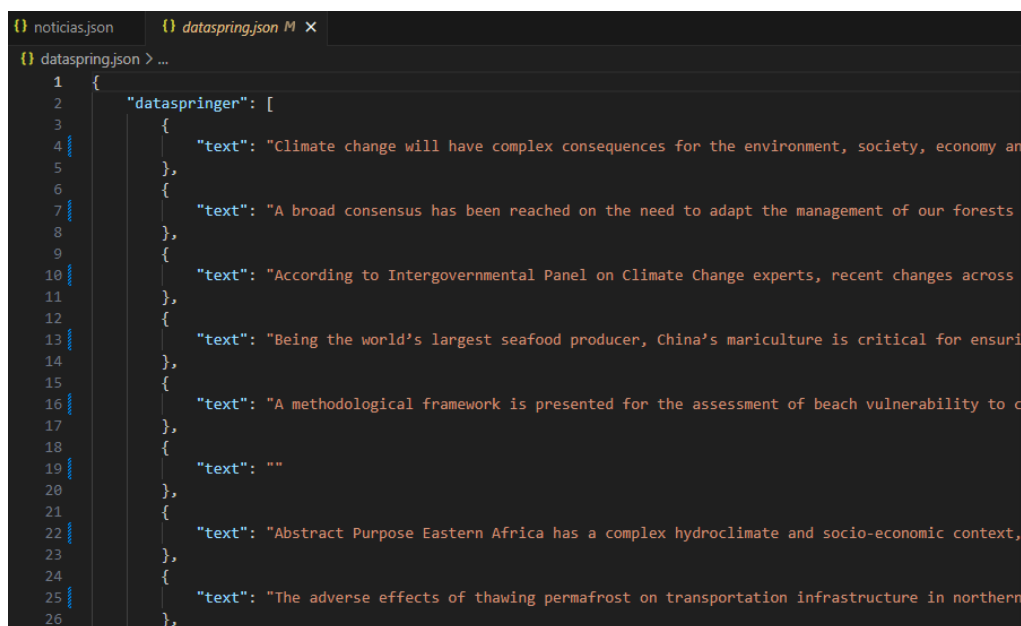
El incremento se valida correctamente, con un cumplimiento total de las tareas como se puede observar en la Tabla 12. El sistema es capaz de trabajar en base a fuentes de información académica y la búsqueda de temas han mejorado con el ingreso de palabras claves.

**Tabla 12. Estado de HU para el incremento 3**

Identificador	Estado
MO-HU02	Completo
MO-HU03	Completo
IN-HU01	Completo

Al realizar la validación de los resultados obtenidos se puede observar que en la figura 16 se almacenan los “abstract” correspondientes a cada uno de los artículos encontrados en la búsqueda realizada. Estos servirán como el corpus de artículo o el texto que posteriormente será analizado.

Por otra parte, la integración con el resto del sistema se cumple y permite al sistema evolucionar y ahora permite seleccionar entre distintas fuentes de información; académicas y de noticias, para elaborar corpus de noticias y textos científicos, además con la introducción de palabras claves se interpolan la lista de palabras claves con la lista de actores y temas permitiendo enriquecer nuestra búsqueda.



```
1 {
2   "dataspringer": [
3     {
4       "text": "Climate change will have complex consequences for the environment, society, economy an
5     },
6     {
7       "text": "A broad consensus has been reached on the need to adapt the management of our forests
8     },
9     {
10      "text": "According to Intergovernmental Panel on Climate Change experts, recent changes across
11    },
12    {
13      "text": "Being the world's largest seafood producer, China's mariculture is critical for ensuri
14    },
15    {
16      "text": "A methodological framework is presented for the assessment of beach vulnerability to c
17    },
18    {
19      "text": ""
20    },
21    {
22      "text": "Abstract Purpose Eastern Africa has a complex hydroclimate and socio-economic context,
23    },
24    {
25      "text": "The adverse effects of thawing permafrost on transportation infrastructure in northern
26    },
27  ]
28 }
```

**Figura 16. Archivo JSON de los corpus de artículos científicos.**

#### **Incremento 4**

Este incremento tiene como objetivo la implementación de la interfaz del módulo trabajado dentro de la sala de reuniones “targeting” que permitirá interactuar con el motor de búsqueda, la visualización de la lista de actores y temas generados y el sistema de reuniones. Además, se agregan indicadores el estado y tiempo estimado de espera del

proceso. Adicionalmente se establece el flujo de almacenamiento de los actores y temas relevantes generados, dentro del sistema y posteriormente ser utilizados en la generación de matriz de actores y temas. Para esto se establece un esfuerzo de 30 horas para la entrega del incremento y se establecen las siguientes tareas que se observan en la Tabla 13.

Tabla 13. Tareas propuestas para el incremento 4

Identificador de HU	Historia de usuario	Tareas	Estimación de Horas
IN-HU02	Yo como usuario quiero ingresar una temática de interés para mi organización para enterarme de los actores y temas relacionados.	<ul style="list-style-type: none"> <li>Implementar una barra de búsqueda de temas de interés.</li> <li>Implementar un mecanismo de selección de fuentes de información.</li> <li>Implementar la función de búsqueda por palabras claves.</li> <li>Establecer las peticiones al servidor del motor de búsqueda y parametrizarlas.</li> <li>Implementar una tabla que me permita desplegar la lista de actores y temas.</li> </ul>	10
IN-HU03	Yo como usuario quiero visualizar un indicador para enterarme del estado actual de la búsqueda.	<ul style="list-style-type: none"> <li>Construir indicadores que representen el estado del proceso de búsqueda.</li> <li>Utilizar técnicas de interpolación para escalar los valores de los indicadores y generar visualizaciones.</li> </ul>	10
IN-HU04	Yo como usuario quiero poder guardar los actores y temas generados a partir de la consulta para ser utilizados en la creación de la matriz de actores y temas.	<ul style="list-style-type: none"> <li>Crear peticiones hacia la base de datos del entorno unificado.</li> <li>Seleccionar el o los actores que se desean almacenar.</li> <li>Seleccionar el o los temas que se desean almacenar.</li> </ul>	10

Para el desarrollo de la interfaz se toma de partida los estilos de diseño que tiene el sistema unificado para acoplar el motor de búsqueda. Como se observa en la Figura 17, se dispone de una barra de búsqueda, ingreso de palabras claves, indicadores y un filtro de fuentes. Que permitirán interactuar el módulo desarrollado con el entorno de producción unificado.

**Criterios de búsqueda**

Temática

Tema general para iniciar a crear la lista de actores y temas.

---

**Fuentes de información**

inglés
Español

The New York Times  
 Springer  
 DOAJ

El comercio  
 DOAJ

---

**Palabras clave**

Palabra clave

**Figura 17. Interfaz del Motor de Búsqueda.**

Para la selección de fuentes de información se diseñaron casillas de selección múltiple y se habilita la opción de escoger entre una o más fuentes siempre y cuando sean en un idioma común, esto para evitar problemas al aplicar los algoritmos de lematización y Stemming ya que estos trabajan en función del idioma.

Como se observa en la Tabla 14. La realización de las tareas se llevó satisfactoriamente y se dispone de una interfase que coincide con la estética de “Sala de reuniones Targeting”. Por otra parte, la integración con el resto de los módulos se cumple correctamente, ahora se cuenta con una interfaz que interactúa y permite visualizar de mejor manera los procesos que realizábamos en los anteriores incrementos.

**Tabla 14. Estado de HU para el incremento 4**

Identificador	Estado
IN-HU02	Completo
IN-HU03	Completo
IN-HU04	Completo

### Incremento 5

Finalmente, este incremento tiene como objetivo realizar la integración del módulo generación de actores y temas con la arquitectura de sala de reuniones “targeting” en términos de conexión entre servidores y su correcta funcionalidad con el resto del sistema.

Para esto se establece un esfuerzo de 15 horas para la entrega del incremento y se establecen las siguientes tareas que se observan en la Tabla 15.

**Tabla 15. Tareas propuestas para el incremento 5**

Identificador de HU	Historia de usuario	Tareas	Estimación de Horas

IN-HU04	Yo como usuario quiero poder guardar los actores y temas generados a partir de la consulta para ser utilizados en la creación de la matriz de actores y temas.	<ul style="list-style-type: none"> <li>• Realizar una conexión con la base de datos del entorno unificado</li> <li>• Crear peticiones hacia la base de datos del entorno unificado.</li> <li>• Verificar la lógica de negocio.</li> <li>• Validar el correcto funcionamiento del sistema.</li> </ul>	15
---------	--	--	----

Durante el desarrollo de este incremento se hace uso del mecanismo que posee el entorno de producción unificado para realizar la conexión a base de datos GraphQL, como se observa en la Figura 18. Se implementa un archivo de servicios que funge como intermediario del componente “tabla de actores” con la base de datos de GraphQL para realizar la gestión de datos. Inicialmente se tendrá una instancia de “Apollo” que es el cliente de GraphQL y es el que se encarga de realizar las solicitudes al servidor.

Para Guardar los actores dentro de GraphQL se establece el servicio “Create” el cual ejecuta una mutación y pasa el objeto actor como variable que posteriormente crea un nuevo actor en la Base de datos. Este mismo servicio se lo realiza para el componente “tabla de temas”.

```

@Injectable({
  providedIn: 'root',
})
export class ActorHttpService {
  constructor(private apollo: Apollo) {}
  findAll(): Observable<ActorResponseModel> {
    return this.apollo
      .watchQuery<ActorResponseModel>({
        query: ALL_ACTORS,
        errorPolicy: 'all',
      })
      .valueChanges.pipe(map((actors) => actors.data));
  }

  create(actor: Actor): Observable<MutationResult<Actor>> {
    return this.apollo.mutate<Actor>({
      mutation: CREATE_ACTOR,
      //variables : formas en las que se pasa la informacion
      variables: {
        actor,
      },
      errorPolicy: 'all',
    });
  }
}

```

Figura 18. Servicios disponibles del aplicativo y su conexión con el cliente Apollo .



La ejecución de las actividades para este incremento se llevó a cabo sin inconvenientes. El estado de finalización se puede verificar en la Tabla 16, lo que indica que el incremento se realizó sin problemas.

**Tabla 16. Estado de HU para el incremento 5**

Identificador	Estado
IN-HU04	Completo

Para verificar el estado del incremento, se utiliza como referencia la lista generada por el motor de búsqueda, eligiendo varios actores para ser almacenados en la base de datos del sistema. Tal como se muestra en la Figura 19, los actores seleccionados y posteriormente guardados ya están registrados en el sistema y están disponibles para su uso en otros módulos.



**Figura 19. Interfaz Actores.**

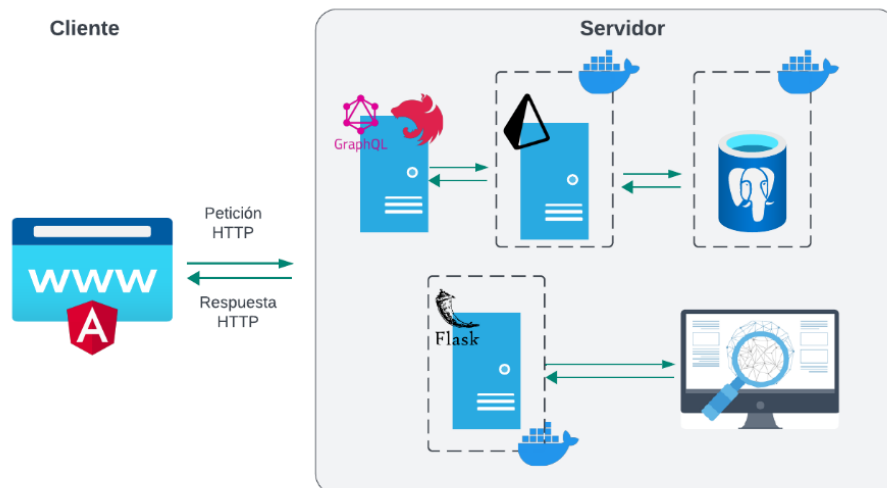
Finalmente, la integración con el resto del módulo completa toda la funcionalidad del sistema permitiendo realizar búsquedas, generar lista de temas y actores que los interesados pueden discutir de su importancia. Posteriormente estos son utilizados en la realización de la matriz de actores y temas concluyendo así, esta pequeña fase dentro de la vigilancia estratégica.

### 3 RESULTADOS

Tras la implementación de la metodología de desarrollo propuesta, se ha logrado construir una infraestructura tecnológica adecuada a los requerimientos del sistema, un sistema web integrado y un conjunto robusto de pruebas que validan el cumplimiento de las funcionalidades propuestas al principio de la migración. En esta sección se detallan los resultados alcanzados, incluyendo la migración e integración con el entorno de producción único, la configuración de la infraestructura y los resultados de las evaluaciones realizadas.

#### 3.1 Arquitectura

Para la integración de las herramientas utilizadas en vigilancia estratégica se creó la siguiente arquitectura cliente servidor. En la Figura 20 se observa la disposición de la arquitectura del módulo migrado y su integración en el entorno de producción unificado .



**Figura 20. Arquitectura Lógica del Sistema.**

El servidor está formado por dos componentes principalmente:

#### **Sala de reuniones de VS**

Este módulo es el encargado de representar las relaciones entre actores y temas en forma de una matriz de targeting alineada con los objetivos y prioridades de la compañía.

El flujo comienza con las peticiones HTTP hacia el servidor GraphQL y estas son transformadas a consultas de PostgreSQL mediante el servidor prisma. Por su parte el servidor entrega la respuesta y esta será transformada en cada una de las etapas y finalmente entregada al usuario.

Este módulo utiliza tecnología de WebSocket que permitirá tener un canal bidireccional de comunicación, esto con el objetivo de tener una conexión activa cliente servidor que permitirá comunicación constante entre estos y permitir que varios usuarios estén sincronizados con la información y decisiones tomadas.

### **Generador de Lista de Actores y Temas**

Módulo encargado de realizar la recolección, análisis y procesamiento de información utilizado en el proceso de targeting, analiza la información importante alineada con los objetivos estratégicos y genera una lista de actores y temas a partir de información encontrada en Internet.

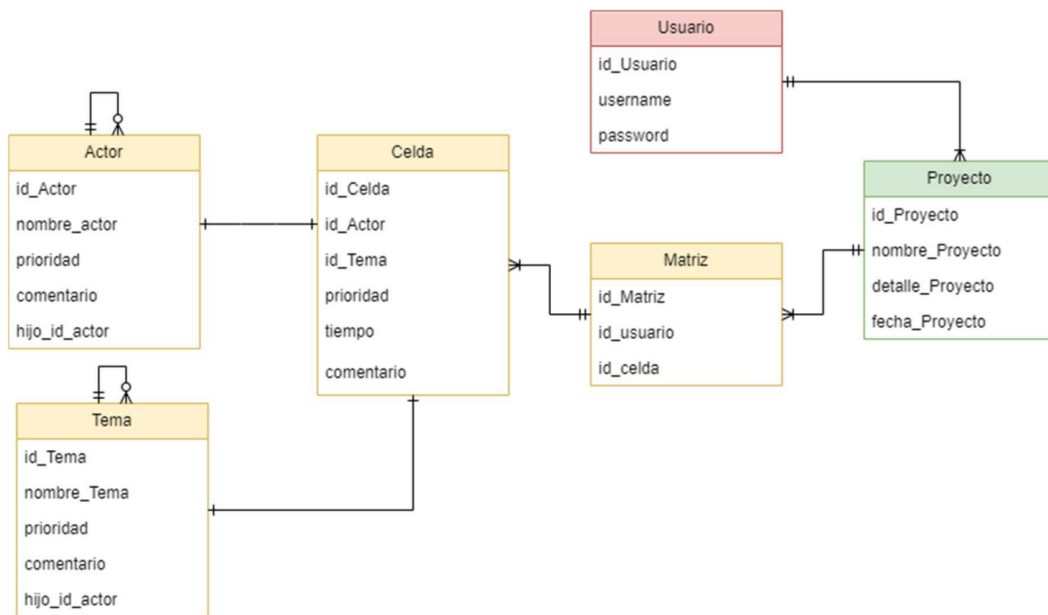
El flujo empieza con una petición HTTP hacia la API Rest realizada en Flask. En este, la lógica de negocio procesa la información recibida y procede a realizar WebScraping en las fuentes de información. El servidor procesa la información y aplica técnicas de procesamiento de lenguaje natural. Posteriormente se devuelve la información en forma de actores y temas que luego es utilizada para completar el proceso de Targeting en el sistema.

La arquitectura propuesta se encuentra alojada en un servidor privado cuyas características son las siguientes.

- SO: Linux
- RAM: mínimo 4 GB de memoria (depende de ocurrencia de usuarios)
- Almacenamiento: 100GB.

### **Modelo Lógico de base de datos**

Partiendo de un modelo lógico de datos, se desarrolló una representación gráfica en GraphQL mediante la definición de tipos. Estos tipos, análogos a las tablas en el modelo relacional, sirvieron como base para construir la base de datos. Utilizando Prisma y PostgreSQL, se generó tanto el esquema de la base de datos como el modelo GraphQL, ambos alineados con la lógica del aplicativo original, aunque con la adición de la entidad 'proyectos'. La Figura 21 ilustra visualmente este modelo lógico.

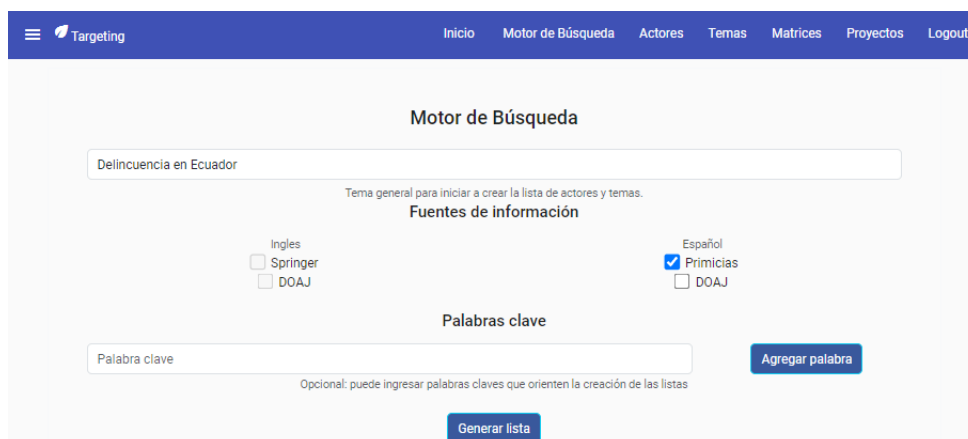


**Figura 21. Modelo Lógico de la BD.**

### 3.2 Descripción de la Aplicación

Esta sección explica el funcionamiento normal que tiene el aplicativo en su uso regular. A continuación, se detalla las partes del módulo migrado, validaciones, comportamiento, y el uso de la lista de actores y temas generado en el resto del entorno unificado.

El módulo migrado cuenta con un componente llamado motor de búsqueda que es el encargado de permitirnos definir la temática de la cual se hará el proceso de targeting. La Figura 22 ilustra los componentes que conforman el motor además que se observa el cambio en estética que sufrió el módulo respecto a su contraparte original.



**Figura 22. Interfase del Motor de Búsqueda.**

El componente nos permite realizar una búsqueda de la temática de interés. En el ejemplo de la Figura 22, se trató de buscar resultados sobre la temática “Delincuencia en Ecuador”.

Cave recalcar que es importante reconocer la generalidad de la temática y cuáles serán las fuentes bases para que el resultado sea más exacto. Es decir, es recomendable que temas de interés científico tengan como base fuentes científicas o que temáticas de interés general como por ejemplo: situación del país se las realice consultando un diario nacional. Para el ejemplo se tomará en consideración el diario Nacional Primicias.

Es posible seleccionar múltiples fuentes de información, la única condición restrictiva es que las fuentes que se escojan deben ser del mismo idioma, recordemos que al realizar el análisis de lenguaje natural las librerías toman como base un idioma en específico. Para asegurar esta condición, al escoger una fuente de información, la interfaz deshabilita las fuentes del otro idioma.

El aplicativo nos permite ingresar palabras claves que, si bien no son obligatoria durante la búsqueda, remarcaran la relación entre la temática y los resultados obtenidos. Como se observa en la Figura 23, se ingresó las palabras “Quito” y “Noboa” para ayudar en nuestra búsqueda.

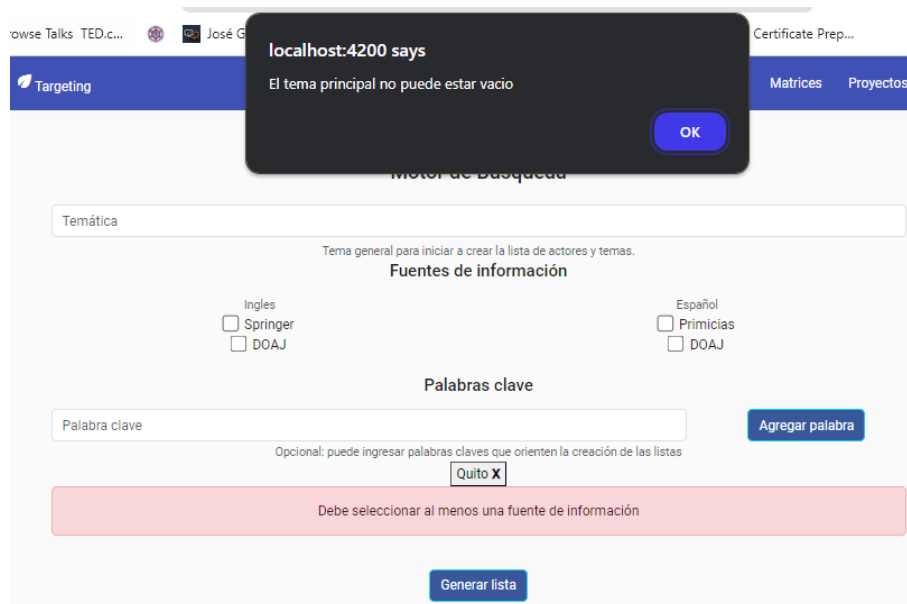
Figura 23. Ingreso de palabras claves.

Una vez ingresados los datos relevantes para la búsqueda, la aplicación web hace una petición al servidor y nos responderá con la lista de actores y temas para VS. La Figura 24 muestra los resultados correspondientes. Entidades como “choneKillers”, “Rasquiña” están entre los actores más representativos y temas como: “grupos delictivos”, “robo de vehículos”, “lavado de activos” son los temas que destacan. Si podemos corroborar el funcionamiento del Motor de búsqueda y su eficiencia.

Lista de Actores		Lista de Temas	
Nombres	Frecuencia	Nombres	Frecuencia
Inteligencia	3	lavado activo	10
Chone Killers	3	prisión preventivo	9
'Rasquiña'	3	grupo delictivo	8
Pipo	3	delito persona	8
Delincuencia Organizada Transnacional	2	robo vehiculo	8

**Figura 24. Lista de Actores y Temas.**

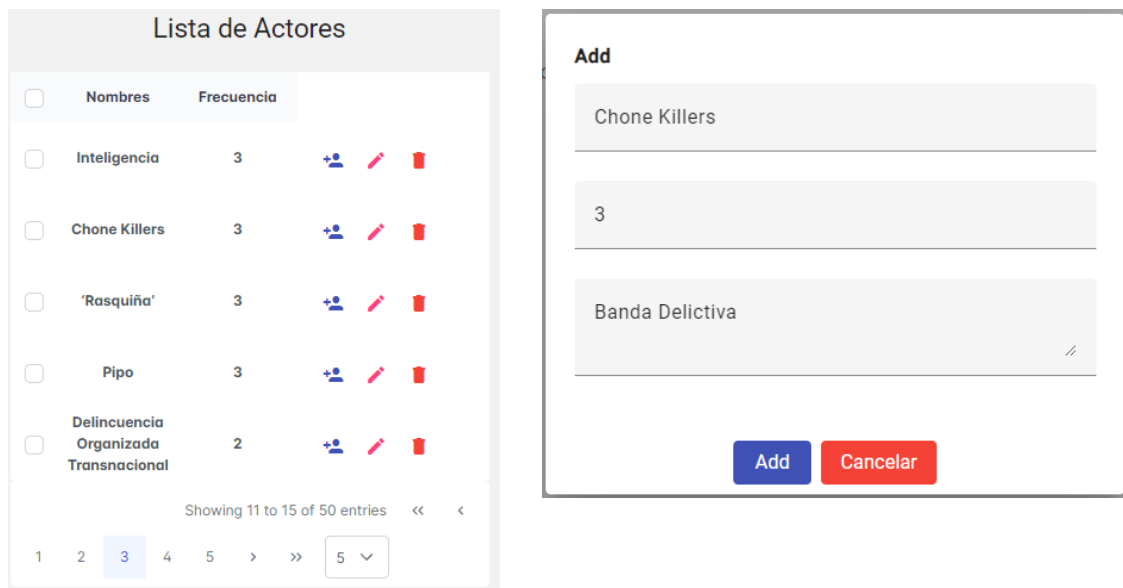
Ahora bien el sistema tiene algunos mecanismos de comunicación con el usuario durante el proceso de búsqueda. Como se observa en la Figura 25, si se desea realizar una exploración sin poner la temática o si existen problemas con el servidor, el aplicativo nos muestra una imagen de error. La falta de selección de fuente de información también provoca errores.



**Figura 25. Errores en el motor de búsqueda.**

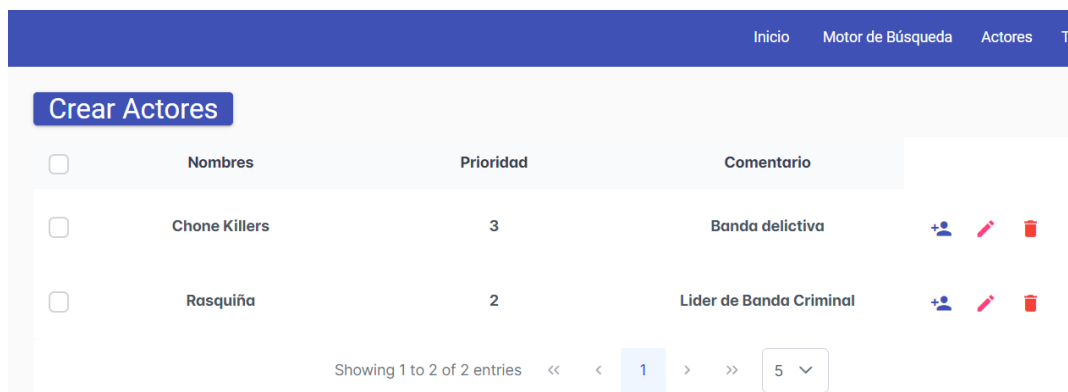
Una vez generada la lista de actores y temas se evidencio su integración con el resto de los módulos del sistema. La Figura 26 ilustra el proceso en el que agregamos un actor de la lista dentro de nuestro sistema. Para hacerlo seleccionamos el botón añadir del actor en la lista desplegada de actores y este nos redirige al formulario de creación de Actores

donde asignamos un comentario y una prioridad una vez discutido con nuestro equipo de trabajo.

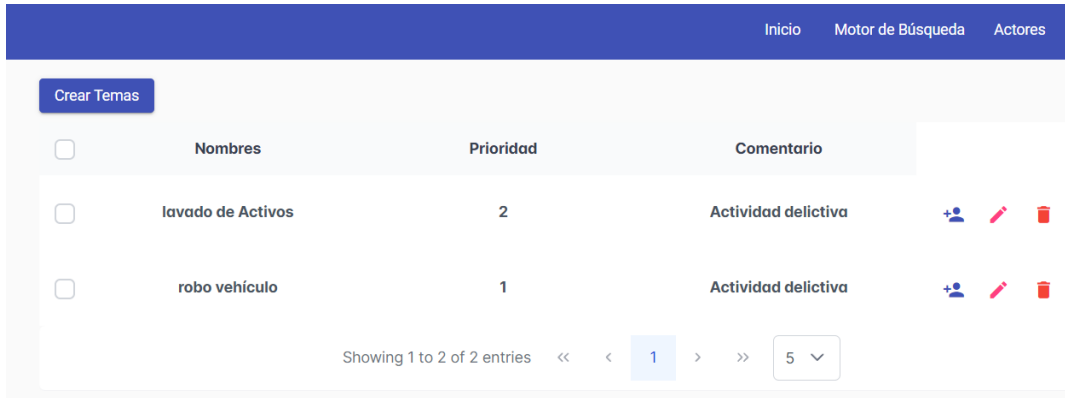


**Figura 26. Creación de Actor desde la lista de actores.**

Una vez llenos los atributos del actor podemos guardarlo y evidenciar su registro dentro del módulo de actores. Como se observa en las Figura 27 y Figura 28, "Chone Killers" se registró dentro de la lista de actores del sistema. Este mismo procedimiento se lo realizó para varios elementos de la lista de temas que fueron de interés para el equipo de trabajo.

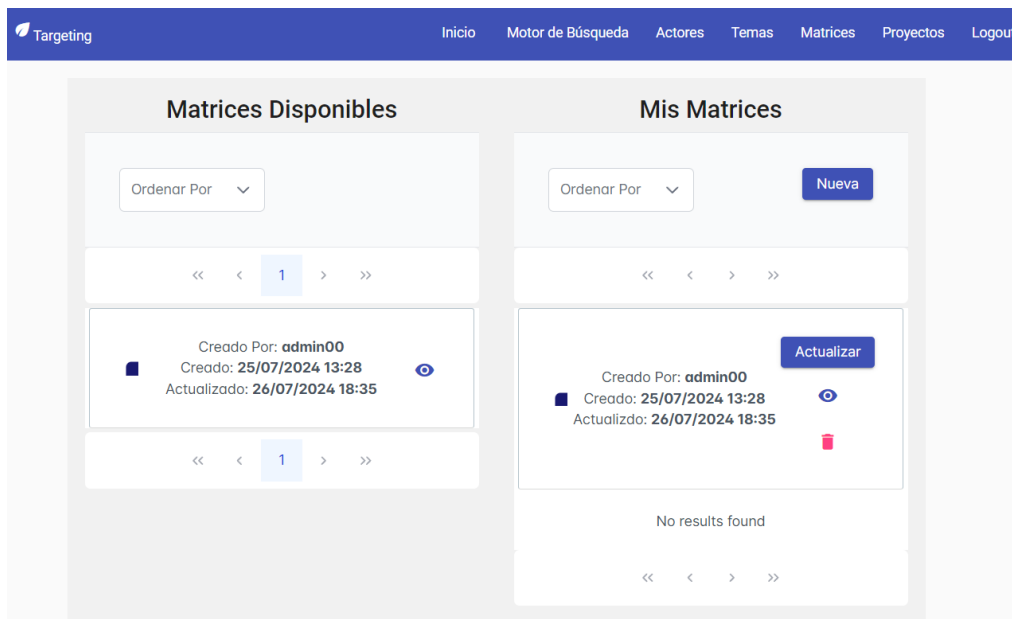


**Figura 27. Lista de Actores en el módulo Actores.**



**Figura 28. Lista de Temas en el módulo Temas.**

Finalmente, con este listado de actores y temas ingresados, como muestra la Figura 29, el sistema genero una matriz de actores y temas.



**Figura 29. Matriz de Actores y Temas.**

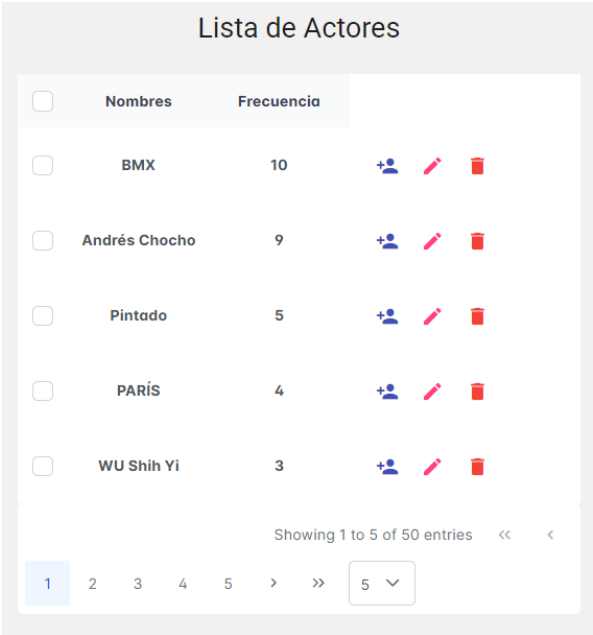
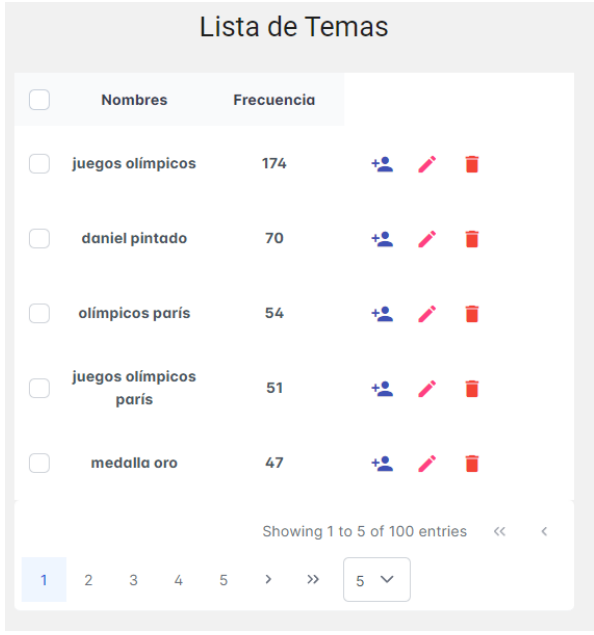
### 3.3 Pruebas

Para evaluar el funcionamiento de la aplicación web se desarrollaron varias pruebas, permitiendo identificar la eficiencia del aplicativo y corregir inconvenientes.

Dado que la funcionalidad más importante del módulo es generar la lista de actores y temas se consideraron los siguientes escenarios y sus respectivos resultados.



**Tabla 17. Evaluaciones de Funcionamiento: Consulta 1.**

PRUEBA 1																																																			
Temática:	Juegos Olímpicos																																																		
Fuentes de información:	<ul style="list-style-type: none"> <li>• Primicias (español)</li> </ul>	Palabras Claves	<ul style="list-style-type: none"> <li>• Paris</li> </ul>																																																
RESULTADO																																																			
Lista de actores		Lista de temas																																																	
 <p><b>Lista de Actores</b></p> <table border="1"> <thead> <tr> <th><input type="checkbox"/></th> <th>Nombres</th> <th>Frecuencia</th> <th></th> </tr> </thead> <tbody> <tr> <td><input type="checkbox"/></td> <td>BMX</td> <td>10</td> <td>+ ✎ 🗑</td> </tr> <tr> <td><input type="checkbox"/></td> <td>Andrés Chocho</td> <td>9</td> <td>+ ✎ 🗑</td> </tr> <tr> <td><input type="checkbox"/></td> <td>Pintado</td> <td>5</td> <td>+ ✎ 🗑</td> </tr> <tr> <td><input type="checkbox"/></td> <td>PARÍS</td> <td>4</td> <td>+ ✎ 🗑</td> </tr> <tr> <td><input type="checkbox"/></td> <td>WU Shih Yi</td> <td>3</td> <td>+ ✎ 🗑</td> </tr> </tbody> </table> <p>Showing 1 to 5 of 50 entries &lt;&lt; &lt; 1 2 3 4 5 &gt; &gt;&gt; 5 ▾</p>		<input type="checkbox"/>	Nombres	Frecuencia		<input type="checkbox"/>	BMX	10	+ ✎ 🗑	<input type="checkbox"/>	Andrés Chocho	9	+ ✎ 🗑	<input type="checkbox"/>	Pintado	5	+ ✎ 🗑	<input type="checkbox"/>	PARÍS	4	+ ✎ 🗑	<input type="checkbox"/>	WU Shih Yi	3	+ ✎ 🗑	 <p><b>Lista de Temas</b></p> <table border="1"> <thead> <tr> <th><input type="checkbox"/></th> <th>Nombres</th> <th>Frecuencia</th> <th></th> </tr> </thead> <tbody> <tr> <td><input type="checkbox"/></td> <td>juegos olímpicos</td> <td>174</td> <td>+ ✎ 🗑</td> </tr> <tr> <td><input type="checkbox"/></td> <td>daniel pintado</td> <td>70</td> <td>+ ✎ 🗑</td> </tr> <tr> <td><input type="checkbox"/></td> <td>olímpicos paris</td> <td>54</td> <td>+ ✎ 🗑</td> </tr> <tr> <td><input type="checkbox"/></td> <td>juegos olímpicos paris</td> <td>51</td> <td>+ ✎ 🗑</td> </tr> <tr> <td><input type="checkbox"/></td> <td>medalla oro</td> <td>47</td> <td>+ ✎ 🗑</td> </tr> </tbody> </table> <p>Showing 1 to 5 of 100 entries &lt;&lt; &lt; 1 2 3 4 5 &gt; &gt;&gt; 5 ▾</p>		<input type="checkbox"/>	Nombres	Frecuencia		<input type="checkbox"/>	juegos olímpicos	174	+ ✎ 🗑	<input type="checkbox"/>	daniel pintado	70	+ ✎ 🗑	<input type="checkbox"/>	olímpicos paris	54	+ ✎ 🗑	<input type="checkbox"/>	juegos olímpicos paris	51	+ ✎ 🗑	<input type="checkbox"/>	medalla oro	47	+ ✎ 🗑
<input type="checkbox"/>	Nombres	Frecuencia																																																	
<input type="checkbox"/>	BMX	10	+ ✎ 🗑																																																
<input type="checkbox"/>	Andrés Chocho	9	+ ✎ 🗑																																																
<input type="checkbox"/>	Pintado	5	+ ✎ 🗑																																																
<input type="checkbox"/>	PARÍS	4	+ ✎ 🗑																																																
<input type="checkbox"/>	WU Shih Yi	3	+ ✎ 🗑																																																
<input type="checkbox"/>	Nombres	Frecuencia																																																	
<input type="checkbox"/>	juegos olímpicos	174	+ ✎ 🗑																																																
<input type="checkbox"/>	daniel pintado	70	+ ✎ 🗑																																																
<input type="checkbox"/>	olímpicos paris	54	+ ✎ 🗑																																																
<input type="checkbox"/>	juegos olímpicos paris	51	+ ✎ 🗑																																																
<input type="checkbox"/>	medalla oro	47	+ ✎ 🗑																																																
<b>Figura 30. Lista de actores de la prueba 1.</b>		<b>Figura 31. Lista de temas de la prueba 1.</b>																																																	

**Tabla 18. Evaluaciones de Funcionamiento: Consulta 2.**

PRUEBA 2			
Temática:	Energías Solar		
Fuentes de información:	<ul style="list-style-type: none"> <li>• Doaj(español)</li> </ul>	Palabras Claves	
RESULTADO			
Lista de actores		Lista de temas	

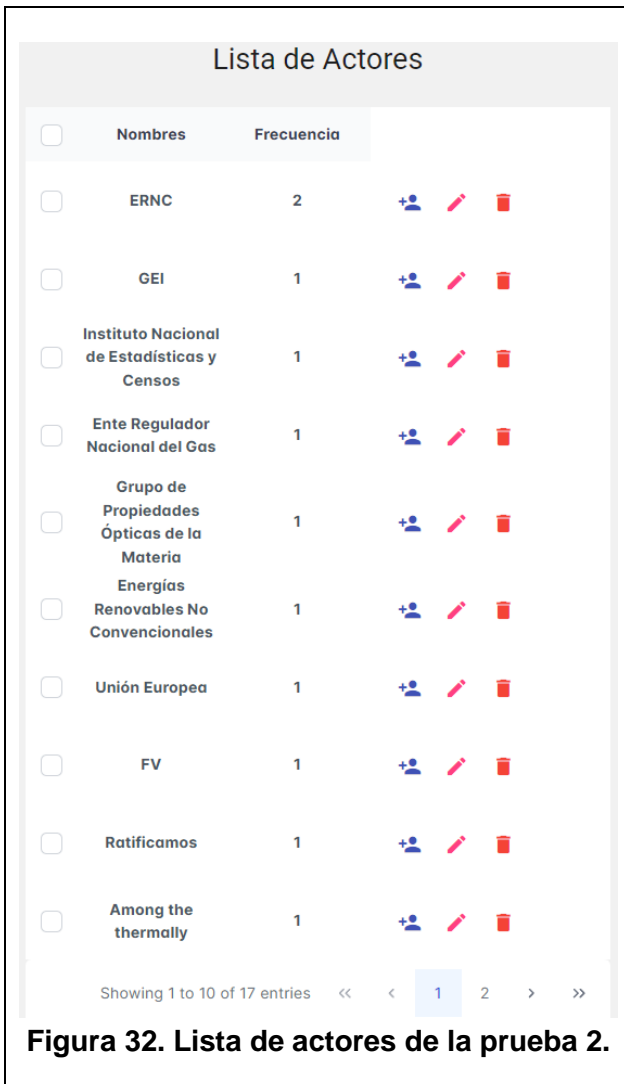


Figura 32. Lista de actores de la prueba 2.

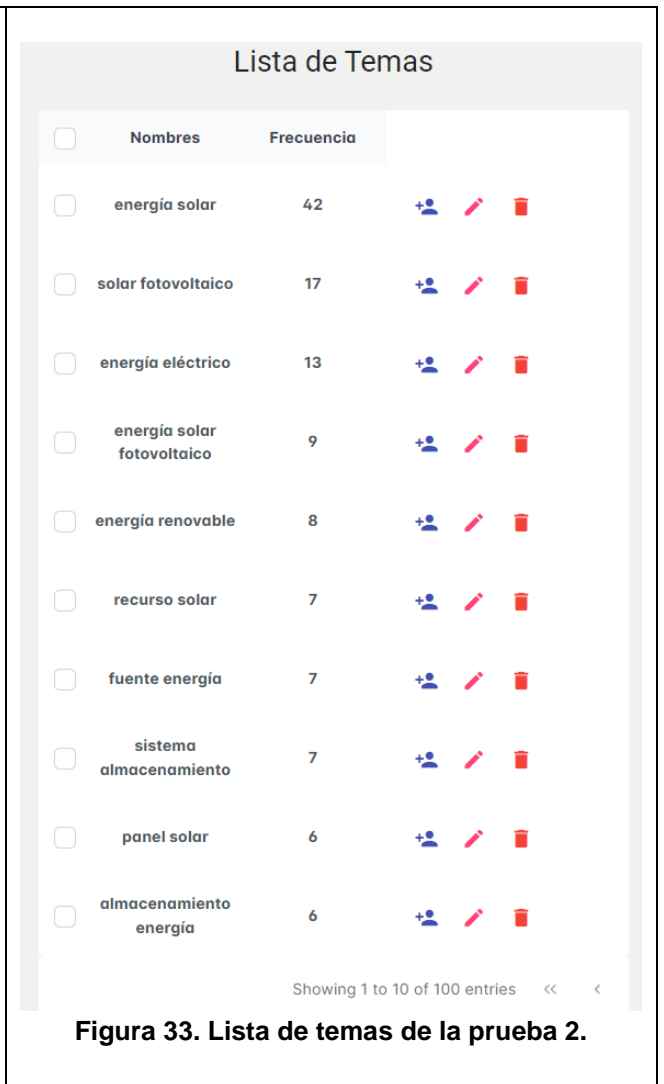






















Figura 33. Lista de temas de la prueba 2.





















Tabla 19. Evaluaciones de Funcionamiento: Consulta 3.

PRUEBA 3			
Temática:	Ambiente		
Fuentes de información:	<ul style="list-style-type: none"> <li>• Primicias (español)</li> <li>• Doaj (español)</li> </ul>	Palabras Claves	
RESULTADO			
Lista de actores	Lista de temas		

Lista de Actores			
<input type="checkbox"/>	Nombres	Frecuencia	
<input type="checkbox"/>	Fritschi	7	+  
<input type="checkbox"/>	Informe	6	+  
<input type="checkbox"/>	UE	5	+  
<input type="checkbox"/>	Naciones Unidas	5	+  
<input type="checkbox"/>	Fuerzas Armadas	5	+  
<input type="checkbox"/>	Foram	5	+  
<input type="checkbox"/>	Superintendencia de Compañías	4	+  
<input type="checkbox"/>	WWF	4	+  
<input type="checkbox"/>	Maate	4	+  
<input type="checkbox"/>	UV	4	+  

Showing 11 to 20 of 50 entries << <

Figura 34. Lista de actores de la prueba 3.





















Lista de Temas			
<input type="checkbox"/>	Nombres	Frecuencia	
<input type="checkbox"/>	ministerio ambiente	54	+  
<input type="checkbox"/>	daniel noboa	30	+  
<input type="checkbox"/>	área protegido	29	+  
<input type="checkbox"/>	cambio climático	28	+  
<input type="checkbox"/>	palo quemado	25	+  
<input type="checkbox"/>	sade fritschi	24	+  
<input type="checkbox"/>	consulta ambiental	23	+  
<input type="checkbox"/>	red social	22	+  
<input type="checkbox"/>	parque nacional	19	+  
<input type="checkbox"/>	proyecto inmobiliario	19	+  

Showing 1 to 10 of 100 entries << <

Figura 35. Lista de temas de la prueba 3.





















Tabla 20. Evaluaciones de Funcionamiento: Consulta 4.

PRUEBA 4			
Temática:	Brain diseases		
Fuentes de información:	• Doaj (English)	Palabras Claves	• Health
RESULTADO			
Lista de actores	Lista de temas		

Lista de Actores			
<input type="checkbox"/>	Nombres	Frecuencia	
<input type="checkbox"/>	BBB	61	+  
<input type="checkbox"/>	EEG	18	+  
<input type="checkbox"/>	CNS	14	+  
<input type="checkbox"/>	MIF	12	+  
<input type="checkbox"/>	CSF	12	+  
<input type="checkbox"/>	EOG	9	+  
<input type="checkbox"/>	Brain	7	+  
<input type="checkbox"/>	NET	7	+  
<input type="checkbox"/>	LPS	6	+  
<input type="checkbox"/>	GBM	6	+  

Showing 1 to 10 of 50 entries << <

Figura 36. Lista de actores de la prueba 4.





















Lista de Temas			
<input type="checkbox"/>	Nombres	Frecuencia	
<input type="checkbox"/>	brain disease	156	+  
<input type="checkbox"/>	blood brain	32	+  
<input type="checkbox"/>	brain barrier	32	+  
<input type="checkbox"/>	brain disorder	31	+  
<input type="checkbox"/>	blood brain barrier	30	+  
<input type="checkbox"/>	drug delivery	25	+  
<input type="checkbox"/>	treatment brain	24	+  
<input type="checkbox"/>	disease brain	23	+  
<input type="checkbox"/>	brain tissue	22	+  
<input type="checkbox"/>	alzheimer disease	20	+  

Showing 1 to 10 of 100 entries << <

Figura 37. Lista de temas de la prueba 4.




















Tabla 21. Evaluaciones de Funcionamiento: Consulta 5.

PRUEBA 5			
Temática:	French Revolution		
Fuentes de información:	• Springer(English)	Palabras Claves	• Politics
RESULTADO			
Lista de actores	Lista de temas		

Lista de Actores			
<input type="checkbox"/>	Nombres	Frecuencia	
<input type="checkbox"/>	Napoleon	106	+  
<input type="checkbox"/>	Burke	97	+  
<input type="checkbox"/>	Coleridge	62	+  
<input type="checkbox"/>	Wollstonecraft	47	+  
<input type="checkbox"/>	Napoleonic	44	+  
<input type="checkbox"/>	Paine	30	+  
<input type="checkbox"/>	Wordsworth	26	+  
<input type="checkbox"/>	Byron	24	+  
<input type="checkbox"/>	State	23	+  
<input type="checkbox"/>	Priestley	23	+  

Showing 1 to 10 of 50 entries << <

**Figura 38. Lista de actores de la prueba 5.**

Lista de Temas			
<input type="checkbox"/>	Nombres	Frecuencia	
<input type="checkbox"/>	seventeenth century	37	+  
<input type="checkbox"/>	american revolution	36	+  
<input type="checkbox"/>	late eighteenth	35	+  
<input type="checkbox"/>	revolution france	33	+  
<input type="checkbox"/>	louis xvi	32	+  
<input type="checkbox"/>	french revolutionary	31	+  
<input type="checkbox"/>	political social	30	+  
<input type="checkbox"/>	ancien régime	29	+  
<input type="checkbox"/>	year later	28	+  
<input type="checkbox"/>	mary wollstonecraft	28	+  

Showing 11 to 20 of 100 entries << <

**Figura 39. Lista de temas de la prueba 5.**

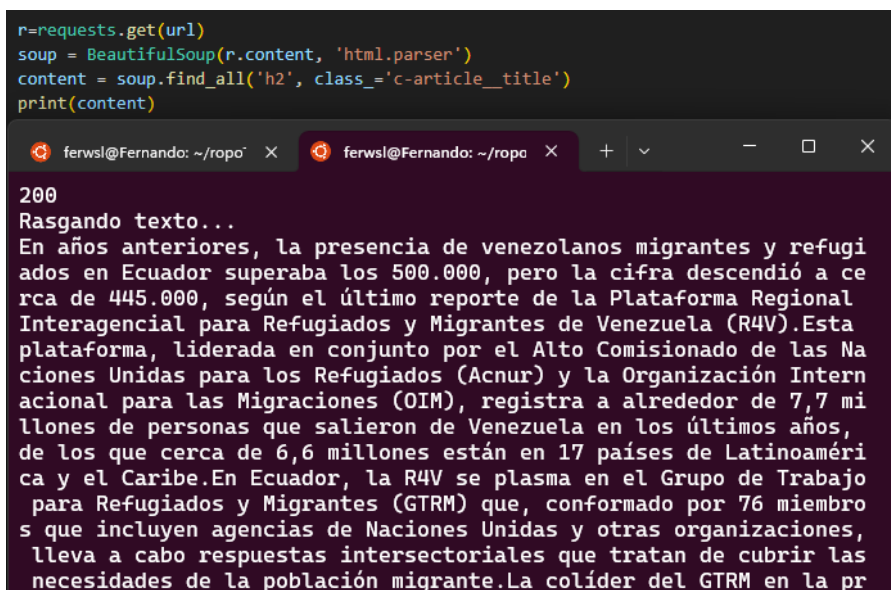
### 3.3.1 Resultados de Pruebas Unitarias

Las pruebas unitarias es un mecanismo que nos permite evaluar nuestro código durante el desarrollo, gracias a esta práctica se puede analizar el funcionamiento esperado que tiene los métodos o algoritmos dentro de nuestra aplicación y corregir errores de manera temprana [34] . Durante el desarrollo del aplicativo se desarrollaron diferentes pruebas unitarias de manera manual. A continuación, se detallan las pruebas unitarias y los resultados obtenidos durante el proceso de evaluación.

La aplicación cuenta con un sin número de funcionalidades que lo integran sin embargo se consideró que varias funcionalidades eran más críticas que otras o eran la base fundamental de varios procesos. Por tanto, durante el desarrollo se adoptaron pruebas continuas enfocadas a asegurar que dicho componente o parte del código cumpliera con los requisitos.

El algoritmo de construcción de corpus de noticias y artículos científicos es una de las bases para el funcionamiento del sistema por tanto era prioritario el buen funcionamiento de este. Como se observa en la Figura 40. La prueba consistía en realizar una petición al servidor de noticias o de artículos científicos enviando una URL con los parámetros correspondientes según la consulta. El resultado esperado era un fragmento de texto que describiera la noticia alojada en internet y como se puede observar el resultado era satisfactorio.

```
r=requests.get(url)
soup = BeautifulSoup(r.content, 'html.parser')
content = soup.find_all('h2', class_='c-article_title')
print(content)
```



```
200
Rasgando texto...
En años anteriores, la presencia de venezolanos migrantes y refugiados en Ecuador superaba los 500.000, pero la cifra descendió a cerca de 445.000, según el último reporte de la Plataforma Regional Interagencial para Refugiados y Migrantes de Venezuela (R4V). Esta plataforma, liderada en conjunto por el Alto Comisionado de las Naciones Unidas para los Refugiados (Acnur) y la Organización Internacional para las Migraciones (OIM), registra a alrededor de 7,7 millones de personas que salieron de Venezuela en los últimos años, de los que cerca de 6,6 millones están en 17 países de Latinoamérica y el Caribe. En Ecuador, la R4V se plasma en el Grupo de Trabajo para Refugiados y Migrantes (GTRM) que, conformado por 76 miembros que incluyen agencias de Naciones Unidas y otras organizaciones, lleva a cabo respuestas intersectoriales que tratan de cubrir las necesidades de la población migrante. La colider del GTRM en la pr
```

**Figura 40. Prueba Unitaria abstracción de corpus de noticias.**

Durante varias semanas la ocurrencia de éxitos al evaluar esta funcionalidad era del 100% sin embargo eventualmente se detectó que para la fuente “Primicias” el corpus no se creaba. Tras realizar un análisis se identificó que el problema era causado por el servidor web y no por el algoritmo en sí. Los identificadores de los elementos HTML que se utilizaban para hacer web scraping habían cambiado por tanto se modificaron en el código.

El algoritmo de lematización de texto también es un factor muy importante en el desarrollo de este proyecto. Como se observa en la Figura 41 la prueba consistía en alimentar el algoritmo con un corpus de noticias y recibir la forma lematizada de las palabras excluyendo conjunciones, espacios, palabras sin interés. El resultado siempre fue satisfactorio al ejecutar las pruebas, consecuentemente por que el algoritmo evaluado era bastante simple. Sin embargo, fue de gran ayuda ya que permitía entender fácilmente como se realizaba el procesamiento de texto y escalarlo a funciones más complejas.

```

nlp = spacy.load("es_core_news_sm")
doc = nlp(texto)

print("\n\n****Entidades que pueden ser stakeholders (partes interesadas) AUTORES****\n\n")
ubicacionesTemasTotales = []
autoresTotales = Counter(palabra.text+" "+palabra.label_ for palabra in doc.ents if palabra.label_ != 'MISC' and palabra.label_ != 'DATE' and palabra.label_ != 'CARDINAL' and palabra.label_ != 'ORDINAL' and palabra.label_ != 'PROPN')
for nombre in autoresTotales:
    if(autoresTotales[nombre] > 2):
        ubicacionesTemasTotales.append(nombre)
        print(nombre+"/t")

```

```

****Entidades que pueden ser stakeholders (partes interesadas) AUTORES****

Ecuador LOC/t
China LOC/t
Andramuño PER/t

```

Figura 41. Prueba Unitaria algoritmo de Lematización.

Finalmente se realizaron pruebas sobre el algoritmo de Stemming. Otro algoritmo simple pero de vital importancia. La Figura 42 muestra que el objetivo de evaluar este algoritmo era obtener las raíces de las palabras dentro del texto y durante la ejecución de las pruebas no se presentaba ningún problema de ejecución concluyendo que el algoritmo estaba bien implementado. Las pruebas realizadas y el entendimiento del algoritmo ayudaron a crear algoritmos más complejos. Esto permitió tener un mecanismo de comparación de raíces de las palabras y contar las ocurrencias de aparición y así poder sacar las palabras que más se repetían y que posteriormente podrían ser considerados como actores o temas.

```

raicesPalabras = [spanishstemmer.stem(token) for token in nombres]
for i, rp in enumerate(raicesPalabras):
    for pc in palabrasClave:
        if rp == pc:
            autoresStemming.append(
                nombres[i]+" "+nombres[i+1]+" "+nombres[i+2])

```

```

**** stemming: palabras en raices ****
['ministerio salud pública', 'gobierno china oms', 'ministerio salud pública']
['cerc', 'person', 'vigil', 'epidemiolog', 'pais', 'cas', 'sospech', 'coronavirus', 'autor', 'minis
teri', 'salud', 'public', 'confirm', 'ayer', 'doming', 'ener', 'pasajer', 'compart', 'vuel', 'ciuda
dan', 'chin', 'año', 'edad', 'sal', 'hong', 'kong', 'destin', 'quit', 'coronavirus', 'ecuador', 'in
vestig', 'cas', 'sospech', 'aclar', 'lun', 'ener', 'ministr', 'salud', 'catalin', 'adramuñ', 'entre
v', 'estacion', 'radial', 'guayaquil', 'hor', 'result', 'prueb', 'diagnost', 'aplic', 'patient', 'a
nalisis', 'descart', 'trat', 'influenz', 'virus', 'respiratori', 'patient', 'asil', 'hospital', 'pu
blic', 'moment', 'establ', 'orient', 'tiemp', 'espaci', 'ministr', 'famili', 'afect', 'seguimient',
'present', 'sintom', 'ciudadan', 'chin', 'originari', 'fujian-fuquing', 'chin', 'arrib', 'ecuador',
'ener', 'present', 'to', 'expector', 'temperatur', 'dolor', 'torac', 'sign', 'insuficient', 'resp
iratori', 'renal', 'ministr', 'posibil', 'patient', 'ten', 'comorbil', 'enfermed', 'bas', 'andramuñ
', 'reforz', 'control', 'puert', 'aeropuert', 'pas', 'fronteriz', 'pais', 'aeropuert', 'guayaquil',

```

Figura 42. Prueba Unitaria algoritmo de Stemming.

### 3.4 Recopilación de Esfuerzo

Durante el desarrollo de la migración e integración del aplicativo “Targetbot” al sistema unificado se realiza el siguiente análisis de actividades necesarias que permitieron el cumplimiento del objetivo.

#### Front-End

- Se realizaron la actualización de dependencias deprecadas o que no disponían de soporte con las versiones actuales.
- Se migraron los módulos de la plataforma anterior a la plataforma actual para esto se realizó los siguientes subprocesos:
  - Cambio de rutas a los servicios.
  - Creación de nuevos servicios.
  - Exportación de módulos.
  - Cambios de estilos de los módulos.
  - Implementación de funcionalidades que comparte con el Font-end de la herramienta Integrada.

#### Back-End

- Se realizó la actualización de las librerías utilizadas en el proceso de Webscraping y procesamiento de lenguaje natural ya que en algunos casos estaban deprecadas.
- Se realizó un análisis para corregir la lógica de WebScraping para algunas de las fuentes de información. Este proceso tuvo como objetivo cambiar y rectificar el cómo se utilizan las APIs de extracción de información y su procesamiento.
- El código presentaba malas prácticas, se reestructuró de mejor manera el código y se implementó buenas practica con el objetivo de que mejore la eficiencia de procesamiento y el código sea más legible para futuras modificaciones.
- De manera general se implementó un entorno digital para la ejecución del backend, se eliminaron scripts no utilizados y se mejoró la documentación en código.[35]



## 4 CONCLUSIONES Y RECOMENDACIONES

### 4.1 CONCLUSIONES

La migración de la herramienta de sugerencia de lista de actores y temas se realizó exitosamente y ha permitido recuperar la funcionalidad de un componente muy importante dentro del proceso de Targeting, provocado por la falta de mantenimiento y descuido. Este proyecto ha permitido que la herramienta sea integrada en un sistema más completo de análisis que permite a las organizaciones realizar la matriz de relación de actores y temas y por consecuencia mejorar el proceso de VS.

Se identifica la importancia de crear sistemas robustos y el uso de tecnologías maduras con soporte a largo plazo. El empleo de tecnologías como bibliotecas de componentes ayudan a facilitar el desarrollo sin embargo, debido a su rápida actualización, muchas funcionalidades dejan de funcionar en proyectos antiguos al no recibir mantenimiento de código. Por otra parte existen librerías que en la actualidad ya no cuentan con soporte produciendo que el código deje de funcionar. La herramienta migrada sufría de ambos escenarios, pero se recalca que uno de los inconvenientes más grandes encontrado fue la incapacidad para recobrar la utilidad de la fuente de información “The new York Times” provocando una pérdida considerable de tiempo y encontrando alternativas accesibles.

El uso de la metodología Incremental fue un pilar importante al momento de asegurar el cumplimiento de funcionalidades de la herramienta de origen. Dado que se identificaron cuáles son las funcionalidades más representativas y con cada incremento se respondía a las necesidades del software. Además de permitirnos tener una visión más clara de los futuros incrementos y cambios que no se tuvieron en consideración al principio de la implementación. Finalmente se logró completar el objetivo de la migración. Se generó una herramienta capaz de producir una lista de actores y temas en base a diferentes fuentes de información (generales y académicas) y su integración en un sistema unificado que complementa la fase de targeting.

## 4.2 RECOMENDACIONES

Es recomendable que el aplicativo se integre otros módulos que complemente las fuentes de información de donde se recaban los datos. Actualmente existen tesis de desarrollo de redes sociales corporativas. Estas podrían ser integras o interconectadas al sistema para recabar información relacionadas a la temática empresarial. Al ser dirigidas a compartir información de mercado, el enfoque de la información manejada podría contener un valor más elevado.

Es recomendable integrar otros mecanismos de procesamiento de Lenguaje Natural, si bien las técnicas de Lematización y Stemming presentan un resultado aceptable, la integración con tecnologías más maduras como los LLM utilizados por ChatGPT, Gemini, entre otros podrían impulsar aún más la generación de lista e incorporar más funcionalidades que ayuden a la organización a identificar los actores y temas de interés alineados a la organización.

Al realizar la migración e integración de varias herramientas es recomendable definir una arquitectura y tecnologías compatible entre todas las herramientas ya que esto podría provocar errores de integración y retrasos innecesarios por otra parte se podría optar por un enfoque de microservicios que facilitaría enormemente la integración.

## 5 REFERENCIAS BIBLIOGRÁFICAS

- [1] J. Fasquelle, M.-L. Caron-Fasan, and N. Lesca, "Strategic Scanning and innovative design: fuel the C/K method through Strategic Scanning information," *Latin American Journal of Computing*, vol. IV, pp. 37–45, Nov. 2017.
- [2] E. Loza-Aguirre, N. Lesca, M.-L. Caron-Fasan, and H. Haddad, *Using a meeting room system to improve targeting of Strategic Scanning*. 2013. doi: 10.1109/ES.2013.6690085.
- [3] P. Ikebujo, "Environmental Scanning as a Process of Strategic Decision-Making – A Review," Jun. 2020.
- [4] E. Loza-Aguirre, N. Lesca, H. Haddad, and M.-L. Caron-Fasan, "Getting a Clean Shot on a Blurred Target: Improving Targeting for Strategic Scanning through Action Research in 10 French Organizations," *Communications of the Association for Information Systems*, vol. 39, pp. 615–638, Dec. 2016, doi: 10.17705/1CAIS.03927.
- [5] Cisneros Paúl and Luis Macías, "Desarrollo de una herramienta para sugerir listas de actores y temas a ser usados en la fase de Targeting de Vigilancia Estratégica," Escuela Politécnica Nacional, Quito, 2021. Accessed: May 04, 2024. [Online]. Available: <https://bibdigital.epn.edu.ec/handle/15000/21517>
- [6] B. Vandenbosch and S. L. Huff, "Searching and Scanning: How Executives Obtain Information from Executive Information Systems," *MIS Quarterly*, vol. 21, no. 1, pp. 81–107, 1997, doi: 10.2307/249743.
- [7] F. Ponce, "Migración a un entorno de producción único de herramientas de vigilancia estratégica desarrollados como trabajos de titulación en la FIS," Escuela Politécnica Nacional, Quito, 2024. Accessed: Jun. 14, 2024. [Online]. Available: <https://bibdigital.epn.edu.ec/handle/15000/25671>
- [8] M. Guechtouli, "Management des activités de veille stratégique :. entre une organisation formelle et informelle," *La Revue des Sciences de Gestion*, vol. 266, no. 2, pp. 23–31, 2014, doi: 10.3917/rsg.266.0023.
- [9] A. Chapagain, *Web Scraping With Python*, First. Birmingham: Packt Publishing Ltd., 2019.
- [10] Ionos, "¿Qué es el web scraping?" Accessed: Apr. 05, 2024. [Online]. Available: <https://www.ionos.com/es-us/digitalguide/paginas-web/desarrollo-web/que-es-el-web-scraping/>
- [11] Decide4AI, "Procesamiento del Lenguaje Natural (PLN o NLP)." Accessed: Mar. 11, 2024. [Online]. Available: <https://decidesoluciones.es/2019/09/12/>
- [12] Elastic, "Procesamiento de lenguaje natural (NLP)." Accessed: Mar. 11, 2024. [Online]. Available: <https://www.elastic.co/es/what-is/natural-language-processing>
- [13] Python, "¿Qué es Python? Resumen Ejecutivo." Accessed: May 12, 2024. [Online]. Available: <https://www.python.org/doc/essays/blurb/>

- [14] DOAJ, "DOAJ Documentation," <https://doaj.org/api/v4/docs>. Accessed: Apr. 05, 2024. [Online]. Available: <https://doaj.org/api/v4/docs>
- [15] Springer, "Springer Link API details." Accessed: Apr. 05, 2024. [Online]. Available: <https://support.springer.com/en/support/solutions/articles/6000195668-springerlink-api-details>
- [16] L. Richardson, "Beautiful Soup Documentation." Accessed: May 06, 2024. [Online]. Available: <https://beautiful-soup-4.readthedocs.io/en/latest/#>
- [17] spaCY, "spaCy: Industrial-strength NLP." Accessed: May 06, 2024. [Online]. Available: <https://pypi.org/project/spacy/#description>
- [18] Apollo docs, "Introduction to Apollo Client." Accessed: May 12, 2024. [Online]. Available: <https://www.apollographql.com/docs/react/>
- [19] Red Hat, "What is Docker?" Accessed: May 12, 2024. [Online]. Available: <https://www.redhat.com/en/topics/containers/what-is-docker>
- [20] Red Hat, "What is GraphQL?" Accessed: May 14, 2024. [Online]. Available: <https://www.redhat.com/en/topics/api/what-is-graphql>
- [21] PostgreSQL Docs, "PostgreSQL." Accessed: May 14, 2024. [Online]. Available: <https://www.postgresql.org/about/>
- [22] Prisma Docs, "Prisma ORM." Accessed: May 14, 2024. [Online]. Available: <https://www.prisma.io/docs/orm/overview/introduction/what-is-prisma>
- [23] Angular Docs, "Introduction to Angular concepts." Accessed: May 14, 2024. [Online]. Available: <https://v17.angular.io/guide/architecture>
- [24] D. Paredes, "Angular Basics: What Is Angular Material, When To Use It, When To Look Elsewhere." Accessed: May 14, 2024. [Online]. Available: <https://www.telerik.com/blogs/angular-basics-what-is-angular-material-when-use-when-look-elsewhere>
- [25] Bootstrap Docs, "Get started with Bootstrap." Accessed: May 14, 2024. [Online]. Available: <https://getbootstrap.com/docs/5.3/getting-started/introduction/>
- [26] L. Karrys, "About npm." Accessed: May 14, 2024. [Online]. Available: <https://docs.npmjs.com/about-npm>
- [27] D. Nguyen, "Introduction to PrimeNG: A Rich UI Component Library for Angular." Accessed: Jun. 11, 2024. [Online]. Available: <https://dalenguyen.medium.com/introduction-to-primeng-a-rich-ui-component-library-for-angular-6b15a1e26202>
- [28] M. Shubel, "What Is TypeScript?" Accessed: Jun. 11, 2024. [Online]. Available: <https://thenewstack.io/what-is-typescript/>
- [29] I. Sommerville, *Ingeniería de Software*, Séptima Edición. Madrid: Pearson Educación S.A, 2005.
- [30] G. Mancuzo, "Modelo incremental fases." Accessed: Mar. 22, 2024. [Online]. Available: <https://blog.comparasoftware.com/modelo-incremental-fases/>

- [31] B. Wu *et al.*, “The Butterfly Methodology: a gateway-free approach for migrating legacy information systems,” in *Proceedings. Third IEEE International Conference on Engineering of Complex Computer Systems (Cat. No.97TB100168)*, 1997, pp. 200–205. doi: 10.1109/ICECCS.1997.622311.
- [32] UDC-DPN, “Lematización de textos.” Accessed: Jun. 19, 2024. [Online]. Available: [https://ucd-dnp.github.io/ConTexto/versiones/master/ejemplos/05\\_lematizacion\\_de\\_textos.html](https://ucd-dnp.github.io/ConTexto/versiones/master/ejemplos/05_lematizacion_de_textos.html)
- [33] K. Pykes, “Stemming y lematización en Python.” Accessed: Jun. 19, 2024. [Online]. Available: <https://www.datacamp.com/es/tutorial/stemming-lemmatization-python>
- [34] Parasoft, “¿Qué son las pruebas unitarias? Una guía completa,” <https://es.parasoft.com/learning-center/unit-testing-guide/>. Accessed: Apr. 09, 2024. [Online]. Available: <https://es.parasoft.com/learning-center/unit-testing-guide/>
- [35] aws, “¿Qué son las pruebas unitarias?” Accessed: Apr. 11, 2024. [Online]. Available: <https://aws.amazon.com/es/what-is/unit-testing/#:~:text=Una%20prueba%20unitaria%20es%20un,la%20!%C3%B3gica%20te%20B3rica%20del%20desarrollador.>

## **ANEXOS**

ANEXO I. Historias de Usuario.

ANEXO II. Interfaces de Herramientas Originales.

## ANEXO I

**Tabla 22. Historia de Usuario1**

Historia de Usuario	1	Identificador	LI-HU01
Título:	Aplicar lematización a un texto		
Prioridad:	Alta	Estimación:	8
Descripción:	Yo como usuario quiero lematizar un texto para identificar las recurrencias en las que una palabra aparece en un texto.		
Criterios de Aceptacion:			
<ul style="list-style-type: none"> <li>• Dado que el usuario ha ingresado un fragmento de texto.</li> <li>• Cuando el usuario aplique Lematización al texto, entonces el algoritmo genera un listado de las palabras juntamente con su lema eliminando palabras sin peso a la búsqueda.</li> </ul>			

**Tabla 23. Historia de Usuario 2**

Historia de Usuario	2	Identificador	MO-HU01
Título:	Extraer texto de noticias a partir de fuentes de Internet		
Prioridad:	Alta	Estimación:	8
Descripción:	Yo como usuario quiero adquirir o sustraer el texto de una noticia en internet para crear un corpus de la noticia.		
Criterios de Aceptacion:			
<ul style="list-style-type: none"> <li>• Dado que el usuario ha ingresado una URL del archivo correspondiente a una noticia del que se quiere extraer el texto.</li> <li>• Cuando el usuario realice Web Scraping, entonces se realiza una iteración de los elementos HTML que conforman el corpus de noticias y se extrae el texto.</li> </ul>			

**Tabla 24. Historia de Usuario 3**

Historia de Usuario	3	Identificador	LI-HU02
Título:	Aplicar Stemming a un texto		
Prioridad:	Alta	Estimación:	5
Descripción:	Yo como usuario quiero aplicar Stemming a un texto para reducir el vocabulario del corpus y adquirir las raíces de las palabras que conforman.		
Criterios de Aceptacion:			
<ul style="list-style-type: none"> <li>• Dado que el usuario ha ingresado un fragmento de texto.</li> </ul>			

- Cuando el usuario aplique Stemming al texto, entonces el algoritmo elimina los prefijos y sufijos de las palabras generadas luego de la lematización.

**Tabla 25. Historia de Usuario 4**

Historia de Usuario	4	Identificador	LI-HU03
Título:	Verificar duplicación de entidades de la lista actores y temas		
Prioridad:	Media	Estimación:	3
Descripción:	Yo como usuario deseo verificar la duplicación en la lista de actores y temas para evitar actores en la lista de temas.		
Criterios de Aceptacion:			
<ul style="list-style-type: none"> <li>• Dado que el usuario ha generado una lista de palabras eliminando prefijos y sufijos.</li> <li>• Cuando el usuario desee verificar las ocurrencias de las palabras, entonces el sistema hará una comparativa de las palabras sin prefijo y sufijo y devolverá las palabras correspondientes de la lista de lematización.</li> </ul>			

**Tabla 26. Historia de Usuario 5**

Historia de Usuario	5	Identificador	LI-HU04
Título:	Obtener lista de actores y temas		
Prioridad:	Media	Estimación:	5
Descripción:	Yo como usuario quiero obtener una lista de actores y temas para posteriormente ser analizados con el equipo de trabajo.		
Criterios de Aceptacion:			
<ul style="list-style-type: none"> <li>• Dado que el usuario dispone de un corpus de noticia</li> <li>• Cuando el usuario procese el texto, entonces el algoritmo genera un listado de actores y temas y se los almacena en un archivo JSON de manera accesible para el resto del sistema.</li> </ul>			

**Tabla 27. Historia de Usuario 6**

Historia de Usuario	6	Identificador	MO-HU02
Título:	Filtrar búsqueda en base a palabras claves		
Prioridad:	Baja	Estimación:	2
Descripción:	Yo como usuario quiero ingresar un conjunto de palabras claves para mejorar la búsqueda en los corpus de texto.		
Criterios de Aceptacion:			



- Dado que el usuario quiere mejorar la búsqueda.
- Cuando el usuario ingrese un listado de palabras claves, entonces el algoritmo realiza una comparación entre la lista de palabras claves y la lista de palabras con lema y se toma en consideración las palabras que tengan relación con las palabras claves.

**Tabla 28. Historia de Usuario 7**

Historia de Usuario	7	Identificador	MO-HU03
Título:	Extraer texto de artículos a partir de fuentes académicas		
Prioridad:	Media	Estimación:	5
Descripción:	Yo como usuario quiero adquirir o sustraer texto de artículos científicos para crear un corpus de texto científico.		
Criterios de Aceptacion:			
<ul style="list-style-type: none"> <li>• Dado que el usuario quiere usar como fuente de consulta bases de información académica.</li> <li>• Cuando el usuario indique que tipo de fuente científica desea consultar, entonces el algoritmo realiza la petición a una fuente de información académica y extrae los “Abstract” de los artículos. .</li> </ul>			

**Tabla 29. Historia de Usuario 8**

Historia de Usuario	8	Identificador	IN-HU01
Título:	Filtrar búsqueda en base a fuentes de información		
Prioridad:	Media	Estimación:	1
Descripción:	Yo como usuario quiero seleccionar el tipo de fuente de consulta para filtrar entre noticias y artículos científicos.		
Criterios de Aceptacion:			
<ul style="list-style-type: none"> <li>• Dado que el usuario quiere seleccionar como fuente de información entre fuentes académicas o generales.</li> <li>• Cuando el usuario seleccione la casilla de n fuente, se utiliza uno de los métodos WebScraping desarrollados y se extrae el corpus de noticia.</li> </ul>			

**Tabla 30. Historia de Usuario 9**

Historia de Usuario	9	Identificador	IN-HU02
Título:	Ingresar temática de consulta		
Prioridad:	Media	Estimación:	2

Descripción:	Yo como usuario quiero ingresar una temática de interés para mi organización para enterarme de los actores y temas relacionados.
Criterios de Aceptacion:	<ul style="list-style-type: none"> <li>Dado que el usuario quiere investigar sobre una temática de interés.</li> <li>Cuando el usuario ingrese un tema en la barra de búsqueda, entonces se realiza una petición con el tema de interés a las diferentes fuentes de información.</li> </ul>

**Tabla 31. Historia de Usuario 10**

Historia de Usuario	10	Identificador	IN-HU03
Título:	Visualizar estado de búsqueda		
Prioridad:	Baja	Estimación:	1
Descripción:	Yo como usuario quiero visualizar un indicador para enterarme del estado actual de la búsqueda.		
Criterios de Aceptacion:	<ul style="list-style-type: none"> <li>Dado que el usuario ha realizado una búsqueda.</li> <li>Cuando el usuario quiera verificar el estado de la búsqueda, entonces sistema muestra una barra de carga que simboliza el porcentaje de búsqueda que falta.</li> </ul>		

**Tabla 32. Historia de Usuario 11**

Historia de Usuario	11	Identificador	IN-HU04
Título:	Almacenar lista de actores y temas		
Prioridad:	Alta	Estimación:	5
Descripción:	Yo como usuario quiero poder guardar los actores y temas generados a partir de la consulta para ser utilizados en la creación de la matriz de actores y temas.		
Criterios de Aceptacion:	<ul style="list-style-type: none"> <li>Dado que el usuario ha generado un listado de actores y temas.</li> </ul> <p>Cuando el usuario quiera utilizar los actores y temas generados para la creación de la matriz relación, entonces el usuario puede seleccionar los actores y temas que se desea almacenar en el sistema y estos aparecerán en la interfaz de actores y temas.</p>		

## ANEXO II

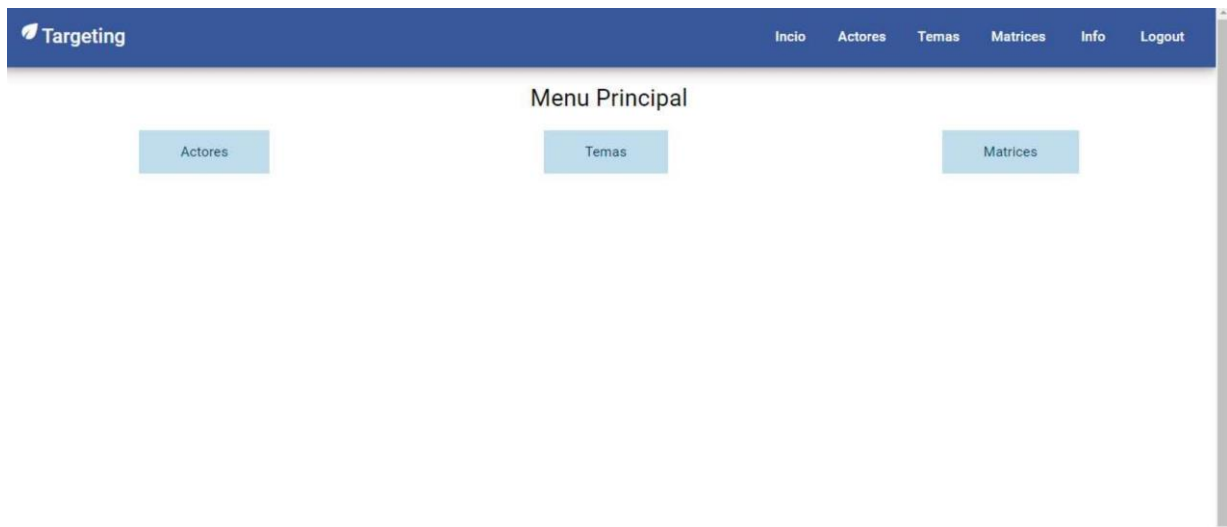


Figura 43. Menú principal de Sala de Reuniones Original

Actor	Prioridad	Comentario	
Competitors	★ ★ ★	Los competidores principales	+ 👤 ✎ 🗑
▼ Customers	★ ★ ★		+ 👤 ✎ 🗑
↳ B2B customer	★ ★ ★		+ 👤 ✎ 🗑
↳ Customers of our customers	★ ★ ★		+ 👤 ✎ 🗑
▼ Suppliers	★ ★ ★		+ 👤 ✎ 🗑
↳ Tier 1	★ ★ ★		+ 👤 ✎ 🗑
↳ Tier 2	★ ☆ ☆		+ 👤 ✎ 🗑
↳ Tier3	★ ★ ★		+ 👤 ✎ 🗑

Navigation: 1 of 1 pages (8 items)

Figura 44. Interfaz de Actores de Sala de Reuniones Original

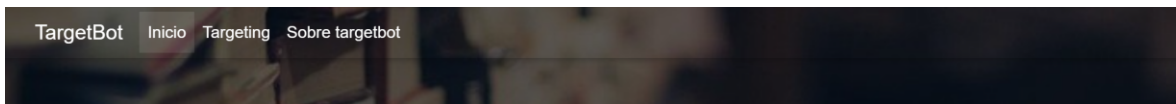
Tema	Prioridad	Comentario	
▼ Environmental	★ ★ ★		+ 👤 ✎ 🗑
↳ Products and services	★ ★ ★		+ 👤 ✎ 🗑

Navigation: 1 of 1 pages (2 items)

Figura 45. Interfaz de Temas de Sala de Reuniones Original.



**Figura 46. Interfaz de Inicio de herramienta Targetbot Original**



## Targeting

El Targeting es un proceso necesario para focalizar el sistema de escucha de la VS, permitiendo así obtener información relevante. Sin embargo, este proceso, necesita de información previa.

**Criterios de búsqueda**

Temática

Tema general para iniciar a crear la lista de actores y temas.

**Fuentes de información**

Inglés

The New York Times

Springer

DOAJ

Español

El comercio

DOAJ

**Palabras clave**

Palabra clave

Opcional: puede ingresar palabras claves que orienten la creación de las listas

**AGREGAR PALABRA**

**GENERAR LISTA**

**Figura 47. Motor de búsqueda de herramienta Targetbot Original.**

Actores	Repeticiones	Temas	Repeticiones
Consejo Nacional Electoral	59	nicolás maduro	162
PUD	21	gonzález urrutia	115
OEA	15	edmundo gonzález	99
Plataforma Unitaria Democrática	14	maría corina	74
Unión Europea	12	nacional electoral	70
oposición	11	corina machado	70
Fuerza Armada	9	maría corina machado	69
Foro Penal	7	consejo nacional	67
Centro Carter	7	consejo nacional electoral	65
República Bolivariana de Venezuela	7	edmundo gonzález urrutia	61

<< < 1 2 3 4 5 > >>

**Figura 48. Listado de Actores y Temas de la Herramienta Targetbot original**