

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE INGENIERÍA DE SISTEMAS

Desarrollo de una herramienta para sugerir listas de actores y temas a ser usados en la fase de Targeting de Vigilancia Estratégica.

TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE INGENIERO EN SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

Cisneros Vásquez Paul Esteban

paul.cisneros@epn.edu.ec

Macías Santillán Luis Gabriel

luis.macias@epn.edu.ec

DIRECTOR: Edison Fernando Loza Aguirre

edison.loza@epn.edu.ec

CODIRECTOR: MSc. Carlos Montenegro Armas

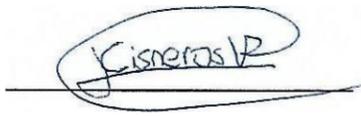
carlos.montenegro@epn.edu.ec

Quito, enero 2021

DECLARACIÓN

Nosotros, Luis Gabriel Macías Santillán y Paúl Esteban Cisneros Vásquez, declaramos bajo juramento que el trabajo aquí descrito es de nuestra autoría; que no ha sido previamente presentada para ningún grado o calificación profesional; y, que hemos consultado las referencias bibliográficas que se incluyen en este documento.

A través de la presente declaración cedemos nuestros derechos de propiedad intelectual correspondientes a este trabajo, a la Escuela Politécnica Nacional, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.



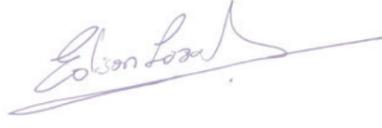
Cisneros Vásquez Paul Esteban



Macias Santillán Luis Gabriel

CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por Luis Gabriel Macías Santillán y Paul Esteban Cisneros Vásquez bajo mi supervisión.



Phd. Edison Loza Aguirre

DIRECTOR DE PROYECTO



MSc. Carlos Montenegro Armas

CODIRECTOR DE PROYECTO

ÍNDICE DE CONTENIDO

DECLARACIÓN	II
CERTIFICACIÓN	III
ÍNDICE DE CONTENIDO.....	IV
ÍNDICE DE TABLAS.....	VII
ÍNDICE DE FIGURAS	IX
RESUMEN	XI
ABSTRACT.....	XII
CAPÍTULO 1	1
1.1 INTRODUCCIÓN	1
1.1.1 Objetivos	2
1.1.2 Alcance	3
1.2 MARCO TEÓRICO	4
1.2.1 Vigilancia estratégica	4
1.2.2 Targeting.....	5
1.2.3 Trabajos relacionados.....	6
1.2.4 Bases de datos académicas.....	7
1.2.5 Web Scrapping.....	8
1.2.6 Procesamiento de lenguaje natural	8
1.3 TECNOLOGÍAS UTILIZADAS.....	9
1.3.1 Lenguajes	9
1.3.2 Entornos de ejecución.....	10
1.3.3 Repositorio de código.....	11
1.3.4 Herramientas utilizadas para desarrollar el sistema	11
1.3.5 Herramientas de procesamiento de lenguaje natural	12

CAPÍTULO 2	14
2.1 METODOLOGÍA DE DESARROLLO	14
2.1.1 Aplicación de la Scrum.....	14
2.1.2 Descripción de las necesidades en Scrum.....	15
2.2 Desarrollo de la solución	19
2.3 Metodología de evaluación	30
2.3.1 TAM.....	30
2.3.2 Detalles de evaluación.....	31
CAPÍTULO 3	33
3.1 RESULTADOS Y DISCUSIÓN	33
3.1.1 Producto final.....	33
3.1.2 Descripción de los Artefactos.....	34
3.1.3 Control de errores.....	38
3.1.4 Pseudocódigo de algoritmos para lematización y recolección de artículo.....	40
3.1.5 Pruebas y evaluación.....	41
3.1.6 Ejemplos de resultados:.....	41
3.1.7 Tiempo de respuesta.....	43
3.1.8 Análisis de rendimiento de página web.....	44
3.1.9 Resultados de la Evaluación.....	45
CAPÍTULO 4	49
4.1 CONCLUSIONES Y RECOMENDACIONES	49
4.1.1 Conclusiones.....	49
4.1.2 Recomendaciones.....	50
CAPÍTULO 5	51
5.1 REFERENCIAS BIBLIOGRÁFICAS	51
CAPÍTULO 6	55
6.1 ANEXO I. TABLAS RELACIONADAS AL PROCESO SCRUM	55

6.1.1	Sprint 0	55
6.1.2	Sprint 1	56
6.1.3	Sprint 2	57
6.1.4	Sprint 3	59
6.1.5	Sprint 4	61
6.1.6	Sprint 5	62
6.1.7	Sprint 6	63

ÍNDICE DE TABLAS

Tabla 1. Bases de datos académicas	7
Tabla 2. Roles de Scrum	16
Tabla 3. Historia Épica 01	16
Tabla 4. Historia Épica 02	16
Tabla 5. Historia Épica 03	17
Tabla 6. Historia Épica 04	17
Tabla 7. Product Backlog. Elaborado por autores.....	18
Tabla 8. Release planning. Elaborado por autores.	19
Tabla 9. Sprint Review Sprint 0. Elaborado por autores	20
Tabla 10. Sprint Review Sprint 1. Elaborado por autores	22
Tabla 11. Sprint Review Sprint 2. Elaborado por autores	23
Tabla 12. Sprint Review Sprint 3. Elaborado por autores	25
Tabla 13. Sprint Review Sprint 4. Elaborado por autores	27
Tabla 14. Sprint Review Sprint 5. Elaborado por autores	28
Tabla 15. Sprint Review Sprint 6. Elaborado por autores	30
Tabla 16. Procesadores Apple.	42
Tabla 17. Video Games.....	42
Tabla 18. Covid.	43
Tabla 19. Pregunta 6.....	48
Tabla 20. Fecha de las fases del Sprint 0. Elaborado por autores.....	55
Tabla 21. Historia del usuario Sprint 0.....	55
Tabla 22. Historia del usuario Sprint 0.....	55
Tabla 23. Sprint 0 Backlog. Elaborado por autores.....	56
Tabla 24. Fecha de las fases del Sprint 1. Elaborado por autores.....	56
Tabla 25. Historia del usuario Sprint 1. Elaborado por autores	56
Tabla 26. Historia del usuario Sprint 1. Elaborado por autores	57

Tabla 27. Historia del usuario Sprint 1. Elaborado por autores	57
Tabla 28. Sprint 1 Backlog. Elaborado por autores.....	57
Tabla 29. Fecha de las fases del Sprint 2. Elaborado por autores	57
Tabla 30. Historia del usuario Sprint 2.....	58
Tabla 31. Historia del usuario Sprint 2.....	58
Tabla 32. Historia del usuario Sprint 2. Elaborado por autores	58
Tabla 33. Sprint 2 Backlog. Elaborado por autores.....	58
Tabla 34. Fecha de las fases del Sprint 3. Elaborado por autores	59
Tabla 35. Historia del usuario Sprint 3. Elaborado por autores	59
Tabla 36. Historia del usuario Sprint 3.....	59
Tabla 37. Historia del usuario Sprint 3.....	60
Tabla 38. Historia del usuario Sprint 3.....	60
Tabla 39. Sprint 3 Backlog. Elaborado por autores.....	60
Tabla 40. Fecha de las fases del Sprint 4.....	61
Tabla 41. Historia del usuario Sprint 4.....	61
Tabla 42. Historia del usuario Sprint 4.....	61
Tabla 43. Sprint 4 Backlog.....	62
Tabla 44. Fecha de las fases del Sprint 5. Elaborado por autores	62
Tabla 45. Historia del usuario Sprint 5. Elaborado por autores	62
Tabla 46. Historia del usuario Sprint 5.....	63
Tabla 47. Historia del usuario Sprint 5.....	63
Tabla 48. Sprint 5 Backlog.....	63
Tabla 49. Fecha de las fases del Sprint 6.....	63
Tabla 50. Historia del usuario Sprint 6. Elaborado por autores	64
Tabla 51. Historia del usuario Sprint 6. Elaborado por autores	64
Tabla 52. Sprint 6 Backlog. Elaborado por autores.....	64

ÍNDICE DE FIGURAS

Figura 1. Proceso genérico de Vigilancia Estratégica modificado de [6].....	4
Figura 2. Logo Python.....	9
Figura 3. Logo Typescript.....	10
Figura 4. Logo NodeJS	10
Figura 5. Logo GitHub.....	11
Figura 6. Proceso Scrum	14
Figura 7. Ejecución del Sprint 0. Elaborado por autores.....	20
Figura 8. Ejecución del Sprint 1. Elaborado por autores.....	21
Figura 9. Ejecución del Sprint 2. Elaborado por autores.....	23
Figura 10. Ejecución del Sprint 3. Elaborado por autores.....	25
Figura 11. Ejecución del Sprint 4. Elaborado por autores.....	26
Figura 12. Ejecución del Sprint 5. Elaborado por autores.....	28
Figura 13. Ejecución del Sprint 6. Elaborado por autores.....	29
Figura 14. Proceso TAM	31
Figura 15. Arquitectura física	33
Figura 16. Arquitectura lógica	34
Figura 17. Pantalla de Bienvenida.....	34
Figura 18. Panel superior	35
Figura 19. Pantalla del Buscador del TargeBot	35
Figura 20. Temática	35
Figura 21. Fuentes de información en ingles.....	36
Figura 22. Fuente de información español	36
Figura 23. Pantallas palabras claves 1.....	36
Figura 24. Pantalla con dos palabras claves.....	37
Figura 25. Botón Generar lista	37
Figura 26. Pantalla de espera	37

Figura 27. Respuesta de TargeBot	38
Figura 28. Mensaje de error por falta de temática	38
Figura 29. Mensaje de error por falta de fuentes de información.....	39
Figura 30. Selección de una fuente de información.....	39
Figura 31. Tiempo de respuesta	44
Figura 32. Análisis en PageSpeed	44
Figura 33. Pregunta 1	45
Figura 34. Pregunta 2	46
Figura 35. Pregunta 2	46
Figura 36. Pregunta 4	47
Figura 37. Pregunta 5	47

RESUMEN

La Vigilancia Estratégica es un proceso informacional que ayuda a las compañías a adelantarse a problemas y oportunidades de su entorno socioeconómico. Dentro de este proceso, el Targeting es la fase en la cual se define el perímetro que guiará la recolección de información para su posterior análisis. Con el objetivo de acelerar la definición del perímetro de búsqueda, se ha propuesto el uso de listas de actores y temas relacionadas con una temática central. En este proyecto integrador se desarrolla una aplicación web que facilita la creación automática de estas listas de actores y temas. Nuestra aplicación obtiene información de bases de datos científicas y medios de prensa digitales usando técnicas de Web Scrapping. Luego, métodos de Procesamiento de Lenguaje Natural se encargan de estudiar toda la información obtenida y encontrar los temas y actores, comparando sus con diccionarios para así saber si las palabras tienen sentido. Para cumplir exitosamente con este proyecto se usó el marco de trabajo SCRUM que organiza el tiempo y las tareas de desarrollo eficientemente. Para evaluar la aplicación web, se realizaron pruebas en base a la metodología TAM, donde se verificó que los usuarios percibieron a esta aplicación web como útil y fácil de usar.

Palabras Clave: Targeting, Vigilancia estratégica, Web Scrapping, Procesamiento de Lenguaje Natural.

ABSTRACT

Strategic Scanning is an information process that helps companies get ahead of problems and opportunities in their socio-economic environment. Within this process, Targeting is the phase in which the perimeter is defined that will guide the collection of information for further analysis. To speed up the definition of the search perimeter, the use of lists of actors and topics related to a central theme has been proposed. This integrative project develops a web application that facilitates the automatic creation of these lists of actors and themes. Our application obtains information from scientific databases and digital media using Web Scrapping techniques. Then, Natural Language Processing methods oversee studying all the information obtained and finding the topics and actors, comparing them with dictionaries to know if the words make sense. To successfully accomplish this project, we used the SCRUM framework that organizes the time and development tasks efficiently. To evaluate the web application, tests were conducted based on the TAM methodology, where it was verified that users perceived this web application as useful and easy to use.

Keywords: Targeting, Strategic Scanning, Web Scrapping, Natural Language Processing

CAPÍTULO 1

1.1 INTRODUCCIÓN

Hoy por hoy, el internet se ha convertido en una poderosa herramienta que contiene cuantiosa información. Según el Instituto Nacional de Estadística y Censos (INEC), el acceso a internet crece en un 3% de la población ecuatoriana anualmente [1]. Esto quiere decir que para el 2020 más del 40% de los ecuatorianos tendrán información en la web. Así, la información existente en la red, puede resultar valiosa para muchas personas, empresas o entidades dado que hay mucho conocimiento oculto esperando ser descubierto: este es el desafío creado por la abundancia de datos de hoy [2].

La competitividad y la sostenibilidad de una organización son dependientes de su propia capacidad para reaccionar ante los cambios del exterior con la mayor rapidez y eficiencia posible al momento de definir sus ejes estratégicos [3]. Por esta razón, las organizaciones deben ser capaces de adquirir, difundir y principalmente analizar la información relacionada con su entorno socioeconómico [4]. Así, es vital que las organizaciones puedan hacer uso de la información del exterior, en su mayoría alojada en el internet, con el fin de ser usada como apoyo para los administradores en la planificación del futuro de dicha organización, a este proceso se lo llama Vigilancia Estratégica (VS) [5].

La VS es el proceso informacional que pretende obtener información relevante del exterior de las organizaciones para descubrir tendencias o cambios que se pueden dar en su entorno, este conocimiento sirve para que los administradores puedan planificar acertadamente el futuro de la organización [6].

A pesar de que, actualmente, se tiene abundante información disponible en el internet, y al igual que existen facilidades para automatizar tareas a través de un computador, algunas de las fases que comprenden la VS se las realiza manualmente y con poca eficiencia. Una de estas fases es el Targeting, fase sobre la cual versa este proyecto. Dado que ninguna organización tiene los recursos para escuchar absolutamente todo lo que pasa a su alrededor, el Targeting se convierte en un proceso necesario para focalizar el sistema de escucha de la VS, permitiendo así obtener información relevante y profundizar esfuerzos. Este proceso, debe confrontar las siguientes particularidades [7]:

- La definición del entorno a vigilar incluye la interacción de gerentes que provienen de diferentes unidades organizativas y que podrían tener una comprensión

fragmentaria y/o borrosa de un problema general. Además, dado que resulta complicado reunir a los gerentes en un mismo lugar, y a una misma hora, las reuniones deben centrarse en concretizar la definición del entorno a vigilar en VS.

- Las reuniones son realizadas utilizando, principalmente, una pizarra, papel y marcadores, lo que hace que este proceso sea lento e ineficiente.
- Se pierde mucho tiempo antes de iniciar un debate debido a él no saber por dónde comenzar. Para esto se pueden usar listas de actores y temas que permiten acelerar el arranque de las discusiones, pero que actualmente son generadas manualmente.

Para solucionar estos inconvenientes, en este trabajo se propone crear una aplicación que permita generar automáticamente una lista de actores y temas entorno a una problemática general, mediante el uso de bases de datos libres de artículos científicos y medios de comunicación, las cuales servirán como ayuda para desencadenar y facilitar las discusiones en la etapa del Targeting [4]. Así mismo, con la creación automática de listas de actores y temas se podrá alcanzar una representación consensuada del entorno que se está construyendo y cumplir con los objetivos de este proyecto [5].

1.1.1 Objetivos

Objetivo general

Desarrollar una herramienta que permita generar listas de actores y temas a ser usada en la fase de Targeting en la Visión Estratégica utilizando bases de datos enlazadas de artículos científicos.

Objetivos específicos

- Diseñar e implementar una herramienta que permita la recolección de información de una base de datos de conocimiento.
- Generar las listas de actores y temas usando técnicas de KDD para sugerir a gerentes relacionados al tema.
- Verificar la coherencia de la información obtenida con la ayuda de un diccionario léxico.
- Evaluar la aceptabilidad de la herramienta propuesta en términos de utilidad y facilidad de uso percibida mediante experimentos.

1.1.2 Alcance

La herramienta web que se va a implementar en el presente trabajo se encargará de recolectar artículos científicos o noticias alojados en el Internet, y mediante técnicas de procesamiento de lenguaje natural entregará al usuario una lista de actores involucrados y los temas que son tratados según el interés de usuario en torno a un tema general [8].

Así, el usuario escogerá una temática general, y con eso obtendrá una recomendación de actores y temas, con una previa selección del tipo de fuentes de información que se va a tomar en cuenta.

Esta herramienta se limitará a aportar en el proceso de targeting de la VS, comprenderá exclusivamente la creación de listas de actores y temas que optimizarán el tiempo de los gerentes o de un grupo de trabajo al momento de iniciar debates o reuniones de interés, mas no será un buscador de información en línea, no ofrecerá al usuario artículos o noticias en contexto y ninguna función semejante.

Esta aplicación es una REST API con una interfaz gráfica desarrollada para facilitar la interacción del usuario con los servicios que se ofrecerá. Por esta razón, se dividirá el desarrollo en dos grupos:

- El procesamiento de la información recibida desde un explorador web o backend: para el desarrollo del cual se hará uso de un framework de Python llamado Flask [9] . Con el uso de este framework se recolectarán los artículos científicos de fuentes como DOAJ [10] y Periódicos locales con el uso de un motor de búsqueda, a continuación, se procesará esta información realizando tareas de minería de texto para finalmente entregar resultados en un archivo JSON.
- Y la interfaz de usuario o frontend; que se encarga de la interacción del backend y el usuario final, mediante el uso de una interfaz gráfica basada en el framework Angular [11].

Mediante el desarrollo de esta herramienta se pretende aportar en el proceso de Targeting. Este servicio estará disponible en un servidor web para beneficio de cualquier aplicación web que pueda consumir un JSON y desee mostrar esta información resultante. Así mismo, se podrá hacer uso de esta aplicación para un posterior almacenamiento de los resultados en una base de datos.

1.2 MARCO TEÓRICO

1.2.1 Vigilancia estratégica

La VS es el proceso proactivo de búsqueda y procesamiento de información mediante el cual las organizaciones se anticipan a las tendencias o cambios en su entorno con el fin de crear o aumentar oportunidades, encontrar vulnerabilidades, reducir riesgos o mitigar efectos nocivos que se encuentren asociados a la falta de conocimientos [4]. Las expresiones "escaneo ambiental" y "sistema de alerta temprana" también se usan a menudo para hablar de VS . Este sistema consta de varios subprocesos, como se detallan en la Figura 1.

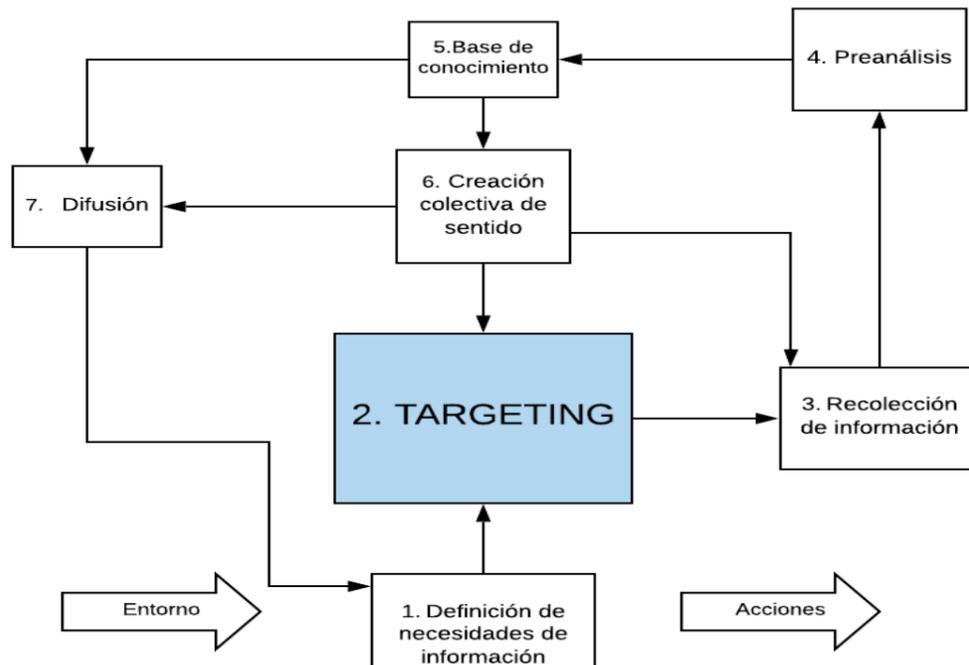


Figura 1. Proceso genérico de Vigilancia Estratégica modificado de [6]

En el proceso global de VS existen varios subprocesos que se pueden soportar con ayuda de la tecnología. Al igual que hay otros a los que ya se los ha automatizado completa o parcialmente; este proyecto integrador pretende automatizar la creación de listas y actores que pueden ser utilizadas para agilizar el proceso de Targeting [7].

La VS tiene dos formas que se complementan para obtener los datos [12], [13]: una es la "búsqueda enfocada", que como su nombre lo indica, se la usa cuando los gerentes ya

tienen alguna idea de las implicaciones de sus decisiones y ya tienen un enfoque de la información que se necesita [14]. El otro modo es el de “escaneo”, éste se usa cuando los gerentes no tienen preguntas o guías para buscar información [12]. El modo de escaneo es precisamente en el cual los gerentes tienen una idea vaga o nula de qué buscar y por consiguiente puede haber un exceso de información irrelevante y potencial sobrecarga de datos que sea producto de la falta de limitación en el espectro de búsqueda [15]. En consecuencia, la información puede terminar abrumándolos, dejando a un lado información importante o relevante para el análisis y posterior toma de decisiones [16],[17]. En la práctica, las organizaciones no tienen tiempo para realizar un escaneo general todos los temas interesantes respecto de sus competidores, para esto se hace uso del targeting [18].

1.2.2 Targeting

El targeting es un subproceso de la VS, que se lo realiza después de definir las necesidades de información de la organización, o lo que es lo mismo, las grandes temáticas de las cuales las organizaciones requieren estar informadas. Esta tarea consiste en la descripción de las perspectivas estratégicas y objetivos de la organización [18]. Como se puede esperar, en este proceso es indispensable para un correcto desarrollo de la VS, en el supuesto caso que se ignore o evite este proceso, se aumenta considerablemente el riesgo de fracaso del proyecto [19]. Esto se daría esencialmente porque en el target pone en consideración las prioridades de la organización que la guiarán a conseguir sus objetivos en VS, consecuentemente, sin realizar este procedimiento, se puede asumir que el proyecto no tiene una brújula.

A lo del tiempo, diversos autores han relacionado el targeting como un proceso difícil de abordar y por momentos difuso [20], por el mismo hecho de iniciar las discusiones para posterior comunicación entre gerentes; teniendo en cuenta que dentro del proceso los participantes se tienen que comunicar, y expresen sus puntos de vista (muchas veces sin usar el mismo lenguaje) tomando en cuenta la unidad de organización para llegar a consolidar un mismo ideal. Muchas veces, toda esta fase de target se debe realizar en el menor tiempo posible y abarcando puntos que pueden ser motivos de preocupación diferente para cada actor [19].

Así, en el modo de escaneo es necesario delimitar el alcance de acuerdo con las prioridades y objetivos de la empresa[6]. Por ello, se requiere definir una lista de actores y temas previo a iniciar la recolección de información.

La construcción del *target* es un procedimiento participativo que se utiliza en las reuniones cuando los gerentes conjuntamente desean instaurar las necesidades de información de la empresa y expresar sus insuficiencias en el conocimiento de su ambiente empresarial. Así, el Targeting facilita la construcción de la matriz objetivo (actores vs temas) que plasma las áreas de escaneo prioritarias para la VS [7]. Se entienden como actores y temas a lo siguiente:

- Un actor es aquel individuo que directa o indirectamente llega a alterar el desempeño de la organización, ya sea con sus decisiones o acciones. Estos actores pueden ser personas o entidades que no necesariamente sean competidores de la empresa, incluso pueden formar parte de ella.
- Un tema está relacionado con un centro de interés de la empresa que pueda afectar positiva o negativamente al futuro de la organización.

1.2.3 Trabajos relacionados

Existen unas pocas contribuciones relacionadas con el targeting en el VS, Zhang et al. [7] implementó un sistema que es capaz de obtener noticias en línea para monitoreo y clasificación en el entorno de la vigilancia sindrómica con el fin de mantener informado al personal de salud pública sobre brotes o información relevante de brotes de enfermedades infecciosas [21]. Por otro lado, en el ámbito empresarial se encuentra una contribución de Lau, Liao, Wong y Chiu [22] donde se implementó un sistema con inteligencia empresarial capaz de analizar factores de noticias textuales mediante el análisis de sentimientos y minería de relaciones comerciales para la toma de decisiones estratégicas en mercados determinados donde la información no es fácil de conseguir.

Entre otros proyectos, existe una herramienta desarrollada por Ochoa J. y Lema S. que promete socializar a actores que tengan los mismos temas de interés, los usuarios podrán compartir ideas por medio de una aplicación móvil o web para filtrar con eficiencia información de interés [23].

Las investigaciones mencionadas anteriormente, aportan al proceso de vigilancia estratégica, pero ninguna de ellas facilita el inicio del proceso del targeting que, como se ha mencionado en esta investigación, se requiere para que los líderes en el ambiente empresarial puedan iniciar una discusión productiva, donde los gerentes puedan ofrecer ideas, estrategias y compartir experiencias focalizadas en los temas sugeridos por la herramienta [7].

1.2.4 Bases de datos académicas

Parte de este proyecto involucra realizar un escaneo en para poder proporcionar información útil a los gerentes al momento de iniciar debates o reuniones de targeting sobre un tema concreto. Para este trabajo vamos a considerar una base de datos académica como un sistema de información que registra artículos de revistas, comunicaciones en congresos y libros (capítulos) que resultan de actividades de investigación.

Las bases de datos académicas de acceso abierto más conocidas [24] se muestran continuación en la Tabla 1:

Tabla 1. Bases de datos académicas

<p>SCIELO (Scientific Electronic Library Online o Biblioteca Científica Electrónica en Línea) proporciona el contenido de revistas científicas y su publicación electrónica [25].</p>	
<p>DOAJ (Directory of Open Access Journals) es un directorio que contiene más de un millón de artículos de libre acceso, todos los artículos pasan por un control de calidad previo [10].</p>	
<p>SPRINGER es la segunda base de datos científica más grande del mundo que proporciona libros, revistas y artículos científicos a investigadores. Springer también abarca a SpringerOpen que proporciona artículos de libre acceso [26].</p>	
<p>ERIC (Education Resources Information Center) permite el acceso a documentos, libros y artículos en texto completo [27].</p>	
<p>e-revist@s es un proyecto realizado en España para difundir artículos y revistas de libre acceso que igual que tienen estándares de calidad como la mayoría de base de datos científicas [28].</p>	
<p>arXiv es un repositorio de libre acceso que contiene más de 900000 libros, revistas y artículos científicos de la universidad Cornell [29].</p>	
<p>Redalyc es un repositorio de revistas científicas, libros y artículos, fue creada por la Universidad Autónoma de México, a la vez que fue uno de los primeros repositorios de libre acceso de Iberoamérica [30].</p>	

1.2.5 Web Scrapping

El término Web Scrapping, también conocido como raspado web, es casi tan antiguo como Internet y se refiere a la recopilación automática de datos de Internet. Otras formas en las que se puede encontrar este concepto son screen scraping, minería de datos web (data mining web), recolección web (web harvesting) [31], [32].

Podemos definir al Web Scrapping como la implementación de un agente o Bot con el propósito de descargar, analizar y organizar datos de la web de manera automatizada, realizando un trabajo eficiente comparado con el que un humano podría realizar [32],[33] .

Estas prácticas presentan un sin número de oportunidades en diferentes campos, por ejemplo:

- Monitorear un sitio de noticias en busca de nuevas historias sobre un tema particular de interés.
- Recuperar datos de una tabla interesante en una página de Wikipedia para realizar un análisis estadístico.
- Realizar minería de texto, crear un motor de recomendación o construir un predictivo modelo para detectar críticas falsas.
- Reunir características agregadas para mejorar el conjunto de datos basado en la información que se encuentra en la web.
- Analizar redes sociales utilizando datos de perfil encontrados en un foro web.

Es importante comprender que, si los datos obtenidos son tratados para uso personal y privado, y dentro del uso justo de las leyes de derechos de autor, representan actividades lícitas. No obstante, publicar los datos nuevamente, realizar un raspado violento capaz de poder causar un daño o eliminar un sitio, o con datos sujetos a derechos de autor violando sus términos de servicio, son actividades que no están exentas violaciones legales [33].

1.2.6 Procesamiento de lenguaje natural

Antes de definir el Procesamiento de Lenguaje Natural, es importante precisar el significado de lenguaje. El lenguaje es el proceso que realizan las personas para comunicarse entre sí. Esta comunicación se puede realizar por diferentes medios: escritos (cartas, textos, mensajes escritos), vocales (voz) y mediante señales. Estas formas de comunicación también se las puede combinar para mejorar la comprensión del emisor y receptor.

El emisor es el que envía la información por cualquiera de los tres métodos y el receptor es el que recibe la información [34].

Existen dos tipos de lenguaje :

- Lenguaje natural (inglés, alemán, español, francés, etc.)
- Lenguaje formal (programable, lógico, matemático, etc.)

El Procesamiento de Lenguaje Natural es un campo de la Inteligencia Artificial [35] que se orienta en el estudio de la comprensión del lenguaje que usamos las personas en el día a día. A pesar de que para las personas es fácil comunicarse entre ellas incluso de diferentes idiomas, la intervención de computadoras en estas comunicaciones constituye un verdadero reto. Y esto se complica aún más si se suma el lenguaje corporal, el sarcasmo, los sentimientos entre otras expresiones que los humanos usamos para describir lo que pensamos y sentimos .

1.3 TECNOLOGÍAS UTILIZADAS

Para el presente trabajo se utilizaron diversas tecnologías y herramientas, las cuales son detalladas a continuación.

1.3.1 Lenguajes

Python

Python es un lenguaje de programación, fácil de entender y a la vez aprender por la flexibilidad que otorga en el paradigma de programación orientado a objetos. Python es un lenguaje interpretado, y esto facilita el uso de principiantes, pero también existen varias comunidades que se dedican al desarrollo de librerías que no siempre son desarrolladas en Python sino también en C o C++ [36].



Figura 2. Logo Python

Tomando en cuenta lo mencionado anteriormente, para este proyecto se usará a Python como lenguaje principal en su desarrollo, debido principalmente al gran número de aportes

que tiene este lenguaje en el campo de la minería de datos y a la existencia de algoritmos de procesamiento de lenguaje natural.

Typescript

Typescript es un lenguaje de programación que engloba a JavaScript, tiene las características de JS, pero con modificaciones que hacen que permite crear nuevas variedades de objetos y tipos de variables. Este lenguaje es mantenido por Microsoft, sin embargo, es de código abierto [37]. En la actualidad hay varios framework que tienen a Typescript como el lenguaje base para desarrollo tanto del lado del cliente como del servidor.



Figura 3. Logo Typescript

1.3.2 Entornos de ejecución

Node Js

Node JS es un entorno que compila y ejecuta JavaScript trabajando en tiempo de ejecución y que permite el desarrollo de servidores Web orientados a eventos asíncronos con buen rendimiento [38].



Figura 4. Logo NodeJS

1.3.3 Repositorio de código

GitHub

GitHub es la plataforma de desarrollo colaborativo más grande del mundo, ofreciendo la posibilidad de alojar los proyectos en repositorios públicos o privados utilizando el sistema de control de versiones Git [39]. Permite controlar diferentes versiones de código y el alojamiento de estas.



Figura 5. Logo GitHub

1.3.4 Herramientas utilizadas para desarrollar el sistema

API de búsqueda DOAJ

El API de búsqueda de DOAJ ofrece acceso a revistas científicas, permitiendo extraer los metadatos y datos de artículos científicos de acceso libre del directorio. Puede ser consumida sin ninguna restricción de reutilización de su código [10].

API Springer de acceso abierto

El API Springer de acceso abierto es un API pública que facilita el acceso a contenido de libre uso no comercial de libros, libros electrónicos y publicaciones científicas relacionadas con ciencia, tecnología y medicina indexados por Springer [26].

BeautifulSoup

La librería BeautifulSoup [40] lleva el nombre de un poema de Lewis Carroll del mismo nombre en Las aventuras de Alicia en el País de las Maravillas. BeautifulSoup ayuda a formatear y organizar la web desordenada arreglando mal HTML y presentándonos con objetos Python fácilmente transitables de manera estructurada y exportar los resultados a una base de datos o archivo estructurado (JSON, CSV, XML, hojas de Excel) [31].

Con el fin de cumplir el proyecto, se usa esta librería para extraer y navegar por contenido HTML fácilmente, olvidando expresiones regulares demasiado complejas y análisis de texto. Debido a que la librería no se encuentra de forma predeterminada en Python, esta debe estar instalada.

1.3.5 Herramientas de procesamiento de lenguaje natural

Existen varias herramientas de procesamiento de lenguaje natural, pero en este proyecto se hará uso de NLTK [41] y Spacy [42]. A pesar de que estas dos librerías pueden llegar a realizar las mismas tareas, tienen ciertas diferencias que en este proyecto integrador se van a utilizar.

Spacy

Es una biblioteca que ayuda a simplificar y reutilizar el código de procesamiento de lenguaje natural, agilitando y facilitando la extracción de grandes cantidades de información. Esta herramienta también se caracteriza por ser una biblioteca optimizada en velocidad de respuesta sin sobrecarga de memoria RAM. Todo esto se debe a que usa un pipeline de procesamiento, según el idioma que se va a usar, para incluir las reglas específicas de dicho lenguaje y posteriormente realizar la tokenización, con esto convierte el texto en tokens con su respectiva puntuación.

El Pipeline consiste en una técnica para transformar el flujo de datos en un proceso secuencial, en otras palabras, hace que las salidas de cada proceso sea la entrada del siguiente. Spacy convierte un Sting en una lista de token a la que se puede acceder a todos sus atributos léxicos por medio de su índice.

Así mismo, Spacy tiene la opción de descarga de paquetes de modelos pre-entrenados. Uno de los modelos más importante en el filtrado de palabras es "es_core_web_sm", que se encarga de reconocer los diferentes tipos de palabras mediante un objeto nlp, el cual se encarga de obtener los parámetros binarios para la predicción, y usando part-of-speech para predecir las relaciones entre palabras que fundamentalmente sirve para diferenciar si una palabra es un sujeto u objeto de una oración.

Como se ha mencionado, Spacy es una librería de Python con muchas bondades, asimismo tiene debilidades que en este proyecto se van a solventar con el uso de otra librería llamada NLTK.

NLTK

Es una librería de Python que se usa para el procesamiento de lenguaje natural, pero con la diferencia de que esta está orientada netamente a la investigación. En NLTK se tiene más libertad, por así decirlo, para crear diferentes algoritmos, a diferencia de Spacy que limita el uso de sus funcionalidades. Para este proyecto, se usa esta librería con el fin de obtener grupos de palabras diferenciadas por su raíz, en un proceso que se lo llama stemming. Con esto se consigue agrupar varias palabras repetidas en diferentes conjugaciones.

CAPÍTULO 2

2.1 METODOLOGÍA DE DESARROLLO

2.1.1 Aplicación de la Scrum

Scrum es un framework formado de reglas y prácticas para el desarrollo ágil de software, basado en la teoría de control de procesos de productos tecnológicos, donde adapta enfoques desarrollo incremental y autoorganización en los procesos de los empleados. Es recomendable para proyectos complejos, donde surge de una visión general del resultado y a partir de detallar especificaciones y funcionalidades deseadas, donde no es posible anticipar todo lo que sucederá[43].

Scrum no es un proceso prescriptivo por lo que no indica la forma de realizar un evento determinado. Al contrario, Scrum ofrece varias técnicas y procesos para controlar el proyecto. Esto significa que es posible saber cómo va el proyecto en los bloques de tiempo y escoger decisiones para mantener el proyecto en marcha con el fin de lograr los objetivos planteados.

Scrum se destaca por ofrecer un proceso simple que sirve para gestionar proyectos complejos [44]. Se basa como un proceso iterativo e incremental (Figura 6). En cada iteración se resuelven una o varias funcionalidades del sistema deseado, y el resultado se conoce como un “incremento” del producto final.

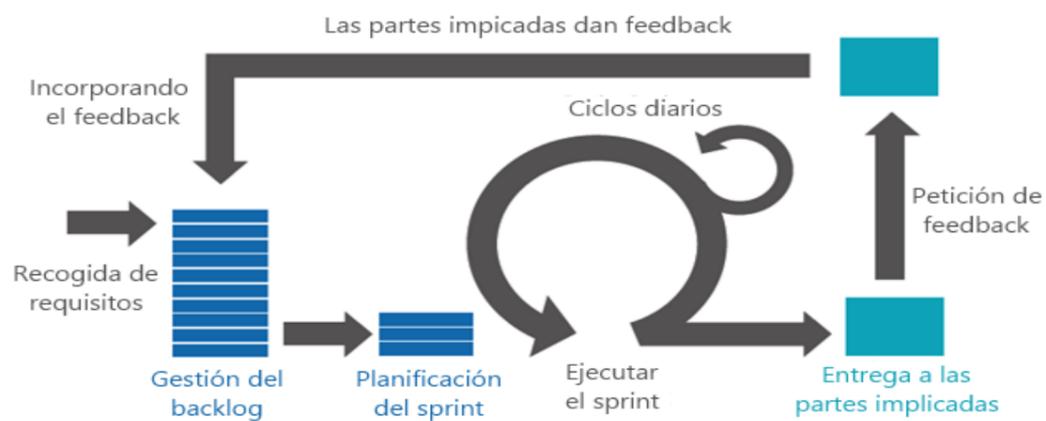


Figura 6. Proceso Scrum

Durante el transcurso de cada iteración se efectúa una tarea de inspección en donde el todos los implicados se reúnen para conocer cómo se va avanzando en el trabajo,

identificar problemas y tomar decisiones apropiadas para que la planificación de la iteración se cumpla. El proceso se lleva a cabo hasta tener el proyecto finalizado. Al final de cada iteración se prueba las funcionalidades resueltas y al usuario se presenta un producto funcionando para evitar planteamientos erróneos o alguna malinterpretación.

2.1.2 Descripción de las necesidades en Scrum

Los gerentes deben focalizar su trabajo en encontrar soluciones a los problemas estratégicos de la empresa u organización [7]. En consecuencia, es un hecho que los ellos disponen de una agenda apretada y por eso es necesario que las tareas de trabajo conjunto sean eficientes y eficaces.

En el caso específico del targeting de VS, se requiere preparar reuniones en las cuales participen los gerentes de las organizaciones para ayudar en el arranque de las discusiones. Por esto, es necesario que existan listas de actores y temas que guíen las discusiones que permitan identificar el perímetro de búsqueda de información, para consecutivamente poder conseguir información que permita anticipar cambios o revelar amenazas y oportunidades [45].

La generación automática de listas de actores y temas para el targeting reducirá el tiempo necesario para la definición del perímetro de búsqueda en la VS, lo cual es pertinente, considerando que el tiempo de los participantes es limitado.

La aplicación web que se desarrollara contara con un solo perfil de usuario. Este usuario corresponde al usuario final que maneja la página sería un profesor, estudiante, gerente de empresa o cualquier persona que busque contenido sobre un tema en específico.

Alcance

En el proyecto se diseñará e implementará una aplicación web, por lo tanto, El usuario será capaz de acceder a la página web permitiéndole generar una lista de actores y temas entorno a una problemática general mediante un buscador.

Como parte de la solución planteada se creará un componente capaz de recolectar datos de artículos científicos y de noticias, para de ese modo, enriquecer al módulo generador de la lista.

Al final, se realizará una evaluación de aceptabilidad de la solución elaborada por usuarios para comprobar su utilidad en el proceso de VS de una organización.

Roles

Mediante la siguiente tabla (Tabla 2) se mostrará los roles utilizados en el desarrollo del proyecto, los cuales son basados en el framework Scrum.

Tabla 2. Roles de Scrum

Rol	Encargado
Product Owner	PhD. Edison Loza Aguirre y MSc. Carlos Montenegro Armas
Scrum Master	PhD. Edison Loza Aguirre
Development Team	Paul Cisneros y Gabriel Macias

Requerimientos de los sistemas

TargetBot es una página Web pensada en asistir a cualquier usuario o gerente en los procesos de VS, específicamente en el proceso de Targeting. La página web constituye de un motor de búsqueda web tomando de entrada un tema específico y alimentado de noticias y artículos científicos para obtener generar una lista de actores y temas relacionados a la búsqueda.

La página web tendrá las siguientes capacidades:

- Motor de búsqueda de un tema específico
- Búsqueda avanzada por medio de palabras claves.
- Generador de una lista de actores y temas.

Historias de usuario épicas

En base al alcance definido del proyecto, se especifican las historias de usuario épicas que están descritas en las Tablas 3, 4, 5 y 6.

Tabla 3. Historia Épica 01

HISTORIA DE USUARIO ÉPICA	HUEB01
Título: Implementar backend de la aplicación	
Descripción: Se definirá el framework a utilizarse para el backend, controladores para la comunicación del servicio generador de las listas con el motor de búsqueda.	

Tabla 4. Historia Épica 02

HISTORIA DE USUARIO ÉPICA	HUEM02
Título: Implementar motor de búsqueda de un tema específico	
Descripción: Implementar una interfaz gráfica para el motor de búsqueda. Crear el módulo el para generar un corpus de noticias y artículos científicos.	

Tabla 5. Historia Épica 03

HISTORIA DE USUARIO ÉPICA	HUEL03
Título: Implementar generador de lista de actores y temas	
Descripción: Generar una lista de los actores y temas más mencionados relacionados a la búsqueda realizada del tema que escogió el usuario.	

Tabla 6. Historia Épica 04

HISTORIA DE USUARIO ÉPICA	HUEP04
Título: Pruebas de aceptabilidad	
Descripción: Se aplicarán pruebas de uso y rendimiento del consumo de red, memoria y se evaluara la disponibilidad. Se realizarán encuestas para verificar el uso.	

Product Backlog

El Product Backlog está definido en la Tabla 7, está tabla es creada a partir de historias de usuario en la realización de la página web TargetBot:

Tabla 7. Product Backlog. Elaborado por autores

ID	Como un(a)	Yo quiero	Con el fin de	Estimación en Horas
HUEL01-01	Usuario	Lematizar de un texto	Presentar como resultado el número de veces que se repite las palabras	25
HUEM02-01	Usuario	Obtener el texto de una noticia	Creación de un corpus de noticias filtrando el texto	25
HUEL01-02	Usuario	Añadir una funcionalidad de stemming en el módulo generador de listas	Reducir los grupos de palabras y obtener la raíz de cada palabra	10
HUEL01-03	Usuario	Validar la duplicidad de la lista de actores y temas	Filtrar los actores que salen en la lista de temas.	10
HUEL01-04	Usuario	Obtener un archivo de resultados con actores y temas	Visualizar los resultados obtenidos del generador de lista	25
HUEL01-05	Usuario	Optimización de código del módulo generador de lista	Obtener resultados más limpios y acertados a la realidad	2
HUEL01-06	Usuario	Ingresar un archivo de palabra claves	Optimizar la búsqueda de actores.	8
HUEM02-02	Usuario	Obtener el texto de un artículo científico de una base de conocimientos	Crear el corpus con los textos de sus artículos científicos	30
HUEF04-01	Usuario	Diseñar interfaces del Sistema	Creación de una interfaz de búsqueda de noticias y artículos por texto	8
HUEF04-02	Usuario	Implementar interfaces del Sistema	Realizar una búsqueda de un tema específico mediante una interfaz	22
HUEB03-01	Usuario	Diseñar el backend de la aplicación	Controlar todos los módulos y lógica de la página.	8
HUEB03-02	Usuario	Implementar el backend de la aplicación	Integrar el módulo generador del corpus y generador de lista	12
HUEB03-03	Usuario	Implementar integración de los servidores web	Comunicar el servidor backend y frontend y recibir respuestas entre ellos	65
HUEF04-04	Usuario	Visualizar un cuadro de dialogo mientras se realiza la consulta	Ver el progreso de la búsqueda realizada	15
HUEF04-05	Usuario	Modificar la interfaz de búsqueda	Mejorar la apariencia de la interfaz y tablas	15
HUEM05-01	Usuario	Recibir una lista validada de actores y temas.	Verificar la coherencia de la información obtenida con la ayuda de un diccionario léxico	45

Release planning

Después de haber definido el Product Backlog, se procede a definir las historias de usuario que se van a desarrollar en cada uno de los sprints (2 semanas), los cuales se mencionan en la Tabla 8:

Tabla 8. Release planning. Elaborado por autores.

Sprint 0	Sprint 1	Sprint 2	Sprint 3	Sprint 4	Sprint 5	Sprint 6
HUEL01-01	HUEL01-02	HUEL01-05	HUEF04-01	HUEF04-02	HUEB03-03	HUEM04-05
HUEM02-01	HUEL01-03	HUEL01-06	HUEF04-02	HUEB03-02	HUEF04-04	HUEM05-01
	HUEL01-04	HUEM02-02	HUEB03-01			
			HUEB03-02			

2.2 Desarrollo de la solución

Sprint 0

Sprint Planning

La planificación de la ejecución de las fases del Sprint 0 se muestra en la sección 6.1.1 (Anexo I).

Sprint Goal

Crear un script que reciba un texto quemado y que permita generar una lista de palabras lematizadas y la cantidad que se repiten en el texto. Se desea también implementar un script que mediante web scraping obtenga el texto de una noticia.

Sprint Backlog

En la sección 6.1.2, se muestra las actividades y en general el esfuerzo realizado en el Sprint 0.

Ejecución del Sprint

En la Figura 7 se puede apreciar el tiempo real empleado para ejecución de las actividades correspondientes para el sprint, el cual se compara con el tiempo estimado inicialmente para las mismas.

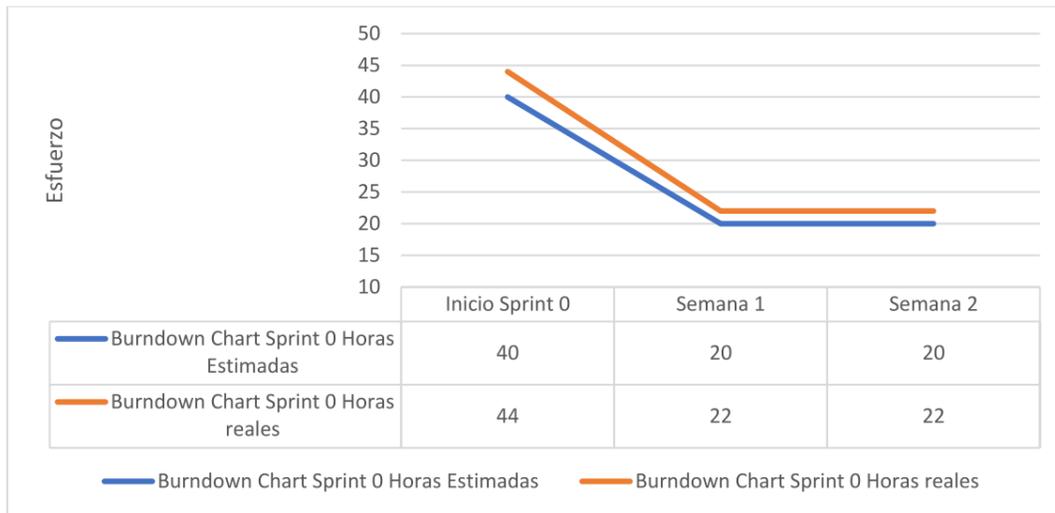


Figura 7. Ejecución del Sprint 0. Elaborado por autores

Inspección y adaptación

Sprint Review

Se procedió a realizar la revisión del Sprint 0 de acuerdo con los criterios establecidos de aceptación definidos para las historias de usuario. La Tabla 9 muestra los resultados del Sprint Review.

Tabla 9. Sprint Review Sprint 0. Elaborado por autores

Código	Historia de Usuario	Cumplimiento
HUEL01-01	Lematizar de un texto	✓
HUEM02-01	Obtener el texto de una noticia	✓

Actualización y refinamiento

Después de realizarse el Sprint 0, se pudo notar que no se presentó ninguna situación que traiga consigo incrementar nuevas Historias de Usuario, por lo tanto, no se realizaron cambios en Product Backlog o en el Release Planning

Sprint Retrospective

De acuerdo con el Burndown Chart del Sprint 0, se reflejó el esfuerzo que se ha realizado durante cada uno de los días del Sprint. Durante el mismo se obtuvieron los resultados de lematización. Así, se obtuvieron las palabras lematizadas, con el número de veces

que se repite cada ignorando conjunciones. Además, se usó un corpus de Wikipedia para obtener las abreviaciones u organizaciones que se nombran.

Se realizó la creación del corpus sobre el tema coronavirus (tomado como ejemplo práctico) con noticias de páginas como de diarios como El País, Metro, El Comercio y New York Times. El universo quedó descartado por problemas de configuración de la página web.

Sprint 1

Sprint Planning

La planificación de la ejecución de las fases del Sprint 1 se muestra en la sección 6.2.1 (Anexo I).

Sprint Goal

Realizar la entrega de un archivo de resultados con actores y temas. Mejorar la calidad de los resultados usando técnicas de minería de datos.

Sprint Backlog

En la sección 6.2.2, se muestra las actividades y en general el esfuerzo realizado en el Sprint 1.

Ejecución del Sprint

En la Figura 8 se puede apreciar el tiempo real que se ha definido para las actividades correspondientes para el sprint, de igual manera también se refleja el tiempo estimado.

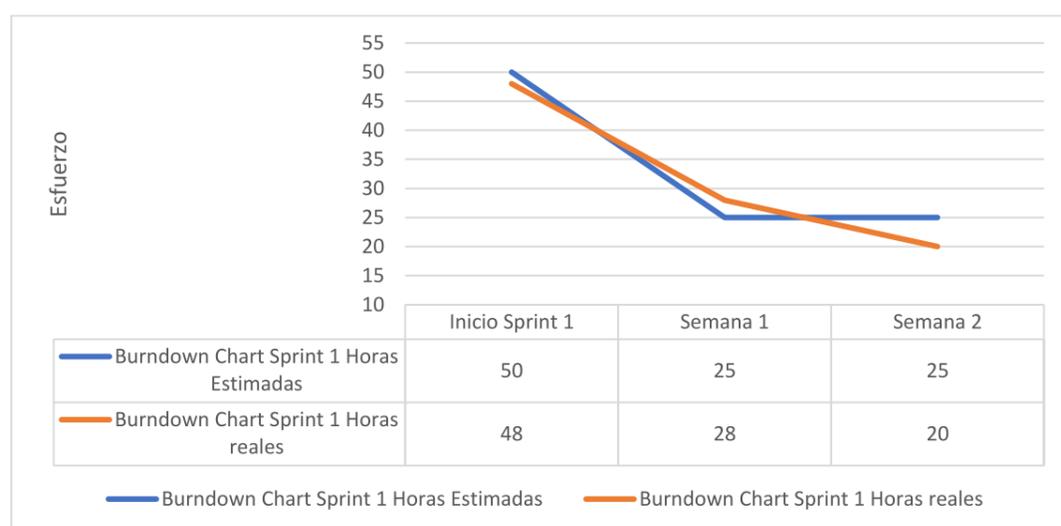


Figura 8. Ejecución del Sprint 1. Elaborado por autores.

Inspección y adaptación

Sprint Review

Se procedió a realizar la revisión del Sprint 1 de acuerdo con los criterios establecidos de aceptación definidos para las historias de usuario. La Tabla 10 a continuación muestra los resultados del Sprint Review de forma clara y detallada.

Tabla 10. Sprint Review Sprint 1. Elaborado por autores

Código	Historia de Usuario	Cumplimiento
HUEL01-02	Añadir una funcionalidad de stemming en el módulo generador de listas.	✓
HUEL01-03	Validar la duplicidad de la lista de actores y temas.	✓
HUEL01-04	Obtener un archivo de resultados con actores y temas.	✓

Actualización y refinamiento

Después de realizarse el Sprint 1, se pudo notar que no se presentó ninguna situación que traiga consigo incrementar nuevas Historias de Usuario, por lo tanto, no se realizaron cambios en Product Backlog o en el Release Planning.

Sprint Retrospective

De acuerdo con el Burndown Chart del Sprint 1, se reflejó el esfuerzo que se ha realizado durante cada uno de los días del Sprint. En el transcurso de estas semanas se realizó el stemming, que es obtener la raíz de cada palabra para poder agruparlas según su origen. También se realizó una clase con todos los métodos realizados. Por el lado de los diccionarios, se probó algunas Apis, controlando las frases que están como autores, ya no salga como tema. Se realizó la entrega de un archivo con la lista de actores y temas encontrados para un caso de estudio.

Sprint 2

Sprint Planning

La planificación de la ejecución de las fases del Sprint 2 se muestra en la sección 6.3.1 (Anexo I).

Sprint Goal

Crear la opción de que el usuario pueda ingresar palabras claves y ser usadas al generar la lista, así como también mejorar los resultados del módulo generador de lista.

Se agregó contenido de artículos científicos de la base de conocimientos Doaj al corpus que alimenta el generador de las listas.

Sprint Backlog

En la sección 6.3.2, se muestra las actividades y en general el esfuerzo realizado en el Sprint 2.

Ejecución del Sprint

En la Figura 9 se puede apreciar el tiempo real que se ha definido para las actividades correspondientes para el sprint, de igual manera también se refleja el tiempo estimado.

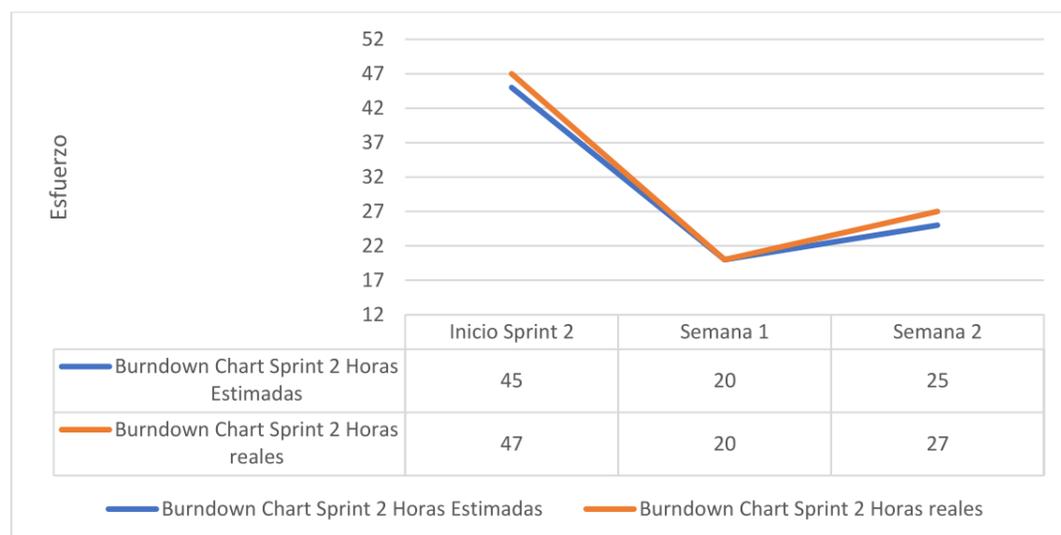


Figura 9. Ejecución del Sprint 2. Elaborado por autores

Inspección y adaptación

Sprint Review

Se procedió a realizar la revisión del Sprint 2 de acuerdo con los criterios establecidos de aceptación definidos para las historias de usuario. La Tabla 11 muestra los resultados del Sprint Review.

Tabla 11. Sprint Review Sprint 2. Elaborado por autores

Código	Historia de Usuario	Cumplimiento
HUEL01-05	Optimización de código del módulo generador de lista.	✓
HUEL01-06	Ingresar un archivo de palabra claves	✓
HUEM02-02	Obtener el texto de un artículo científico de una base de conocimientos.	✓

Actualización y refinamiento

Después de realizarse el Sprint 2, se pudo notar que no se presentó ninguna situación que traiga consigo incrementar nuevas Historias de Usuario, por lo tanto, no se realizaron cambios en Product Backlog o en el Release Planning.

Sprint Retrospective

De acuerdo con el Burndown Chart del Sprint 2, se reflejó el esfuerzo que se ha realizado durante cada uno de los días del Sprint. Se crea la funcionalidad de agregar un archivo con palabras clave para mejorar el filtrado de temas y actores de interés. Se utiliza la API de la web de Doaj para tener acceso a su base de datos de conocimiento de artículos. Se crea un script para buscar el texto de los artículos de un tema específico y devolver ese en un archivo Json para poder ser visualizado.

Sprint 3

Sprint Planning

La planificación de la ejecución de las fases del Sprint 1 se muestra en la sección 6.4.1 (Anexo I).

Sprint Goal

Llevar a cabo diseño e implementación de la interfaz gráfica, así como también del servidor backend.

Sprint Backlog

En la sección 6.4.2, se muestra las actividades y en general el esfuerzo realizado en el Sprint 3.

Ejecución del Sprint

En la Figura 10 se puede apreciar el tiempo real que se ha definido para las actividades correspondientes para el sprint, de igual manera también se refleja el tiempo estimado.

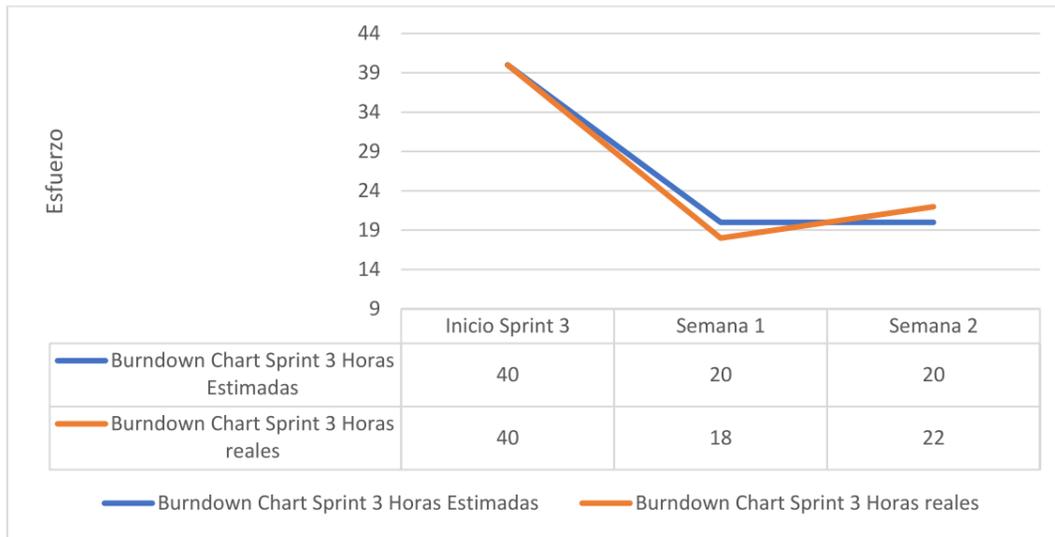


Figura 10. Ejecución del Sprint 3. Elaborado por autores

Inspección y adaptación

Sprint Review

Se procedió a realizar la revisión del Sprint 3 de acuerdo con los criterios establecidos de aceptación definidos para las historias de usuario. La Tabla 12 muestra los resultados del Sprint Review de forma clara y detallada.

Tabla 12. Sprint Review Sprint 3. Elaborado por autores

Código	Historia de Usuario	Cumplimiento
HUEF04-01	Diseñar interfaces del Sistema	✓
HUEF04-02	Implementar interfaces del Sistema	✓
HUEB03-01	Diseñar el backend de la aplicación	✓
HUEB03-02	Implementar el backend de la aplicación	○

Actualización y refinamiento

Después de realizarse el Sprint 3, se decidió mover HUEB03-02 para el siguiente Spring por no completar con los criterios de aceptación y presentar errores y no tener un prototipo presentable.

Sprint Retrospective

De acuerdo con el Burndown Chart del Sprint 3, se reflejó el esfuerzo que se ha realizado durante cada uno de los días del Sprint. Se analizó las mejoras para la búsqueda de

actores y temas. Se empezaron a tomar en cuenta actores identificables con dos palabras. Se empezaron a presentar los actores estén en singular. Se trabajó en hacer aproximaciones por sinónimos y en comprobar que los elementos que salen como actores, no sean temas. Además, se trabajó en la interfaz de búsqueda por palabras claves donde se podía seleccionar la fuente del corpus (artículos científicos y noticias).

Sprint 4

Sprint Planning

La planificación de la ejecución de las fases del Sprint 4 se muestra en la sección 6.5.1 (Anexo I).

Sprint Goal

Realizar la integración de los módulos generador de lista con el creador de corpus, así como también su integración al servidor backend.

Respecto a la interfaz gráfica se agregó un campo para poder agregar palabras claves recomendadas por el usuario para una búsqueda más avanzada.

Sprint Backlog

En la sección 6.5.2, se muestra las actividades y el esfuerzo realizado en el Sprint 4.

Ejecución del Sprint

En la Figura 11 se puede apreciar el tiempo real que se ha definido para las actividades correspondientes para el sprint, de igual manera también se refleja el tiempo estimado.

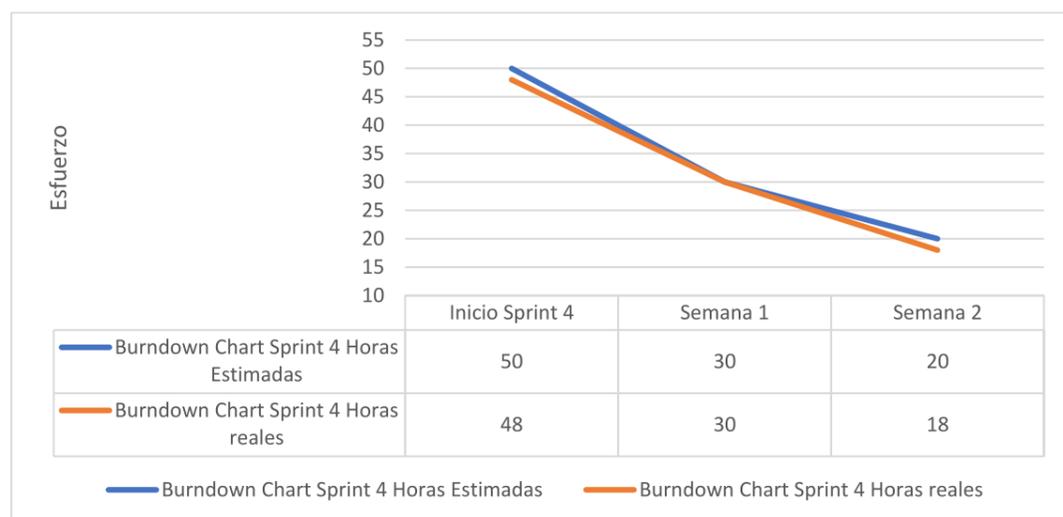


Figura 11. Ejecución del Sprint 4. Elaborado por autores

Inspección y adaptación

Sprint Review

Se procedió a realizar la revisión del Sprint 4 de acuerdo con los criterios establecidos de aceptación definidos para las historias de usuario. La Tabla 13 muestra los resultados del Sprint Review.

Tabla 13. Sprint Review Sprint 4. Elaborado por autores

Código	Historia de Usuario	Cumplimiento
HUEF04-02	Implementar interfaces del Sistema	✓
HUEB03-02	Implementar el backend de la aplicación	✓

Actualización y refinamiento

Después de realizarse el Sprint 4, se pudo notar que no se presentó ninguna situación que traiga consigo incrementar nuevas Historias de Usuario, por lo tanto, no se realizaron cambios en Product Backlog o en el Release Planning.

Sprint Retrospective

De acuerdo con el Burndown Chart del Sprint 4, se reflejó el esfuerzo que se ha realizado durante cada uno de los días del Sprint.

Se preparó los servidores del backend y frontend para una próxima tarea de comunicación.

Respecto a la interfaz gráfica se creó el un apartado para que el usuario agrega las palabras que más le interesa encontrar. A demás de las peticiones al servidor backend para realizar la búsqueda.

El servidor backend se integró el módulo generador de lista con el creador del corpus.

Sprint 5

Sprint Planning

La planificación de la ejecución de las fases del Sprint 5 se muestra en la sección 6.6.1 (Anexo I).

Sprint Goal

Llevar a cabo la comunicación de los servidores y su correcta funcionalidad, así como también implementar como ayuda visual un cuadro de carga para el cálculo de tiempo faltante de la búsqueda.

Sprint Backlog

En la sección 6.6.2, se muestra las actividades y el esfuerzo realizado en el Sprint 5.

Ejecución del Sprint

En la Figura 12 se puede apreciar el tiempo real que se ha definido para las actividades correspondientes para el sprint, de igual manera también se refleja el tiempo estimado.

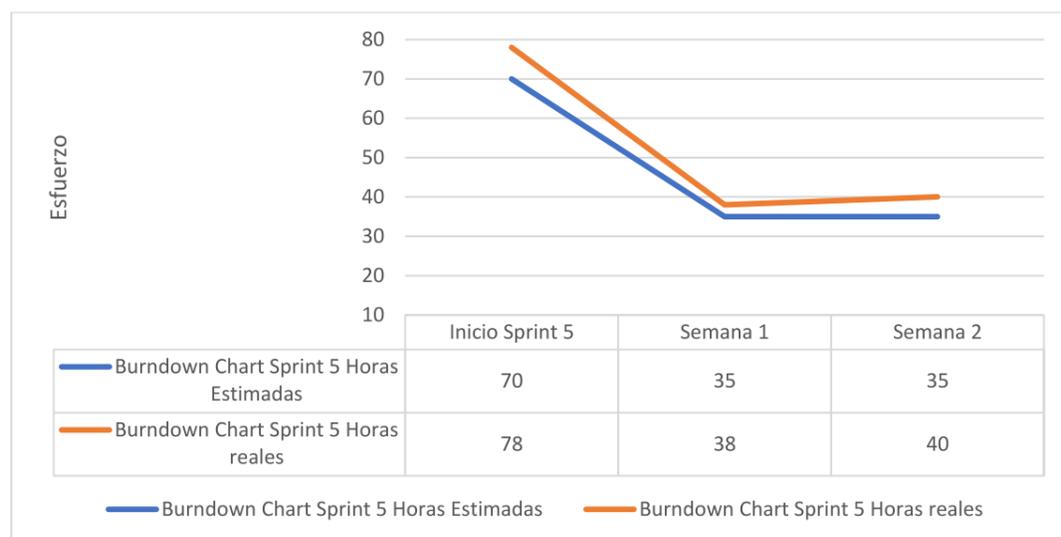


Figura 12. Ejecución del Sprint 5. Elaborado por autores

Inspección y adaptación

Sprint Review

Se procedió a realizar la revisión del Sprint 5 de acuerdo con los criterios establecidos de aceptación definidos para las historias de usuario. La Tabla 14 muestra los resultados del Sprint Review de forma clara y detallada.

Tabla 14. Sprint Review Sprint 5. Elaborado por autores

Código	Historia de Usuario	Cumplimiento
HUEB03-03	Implementar integración de los servidores web	✓
HUEF04-04	Visualizar un cuadro de dialogo mientras se realiza la consulta	✓

Actualización y refinamiento

Después de realizarse el Sprint 4, se pudo notar que no se presentó ninguna situación que traiga consigo incrementar nuevas Historias de Usuario, por lo tanto, no se realizaron cambios en Product Backlog o en el Release Planning.

Sprint Retrospective

De acuerdo con el Burndown Chart del Sprint 5, se reflejó el esfuerzo que se ha realizado durante cada uno de los días del Sprint.

Al realizar las primeras pruebas de funcionalidad de la conexión de los servidores, se pudo percatar que los tiempos de respuesta del programa son largos y se propuso para el siguiente Sprint la reducción de tiempos de respuesta del programa.

Sprint 6

Sprint Planning

La planificación de la ejecución de las fases del Sprint 6 se muestra en la sección 6.7.1 (Anexo I).

Sprint Goal

Realizar el proceso para poner en producción la página web

Sprint Backlog

En la sección 6.7.2, se muestra las actividades y en general el esfuerzo realizado en el Sprint 6.

Ejecución del Sprint

En la Figura 13 se puede apreciar el tiempo real que se ha definido para las actividades correspondientes para el sprint, de igual manera también se refleja el tiempo estimado.

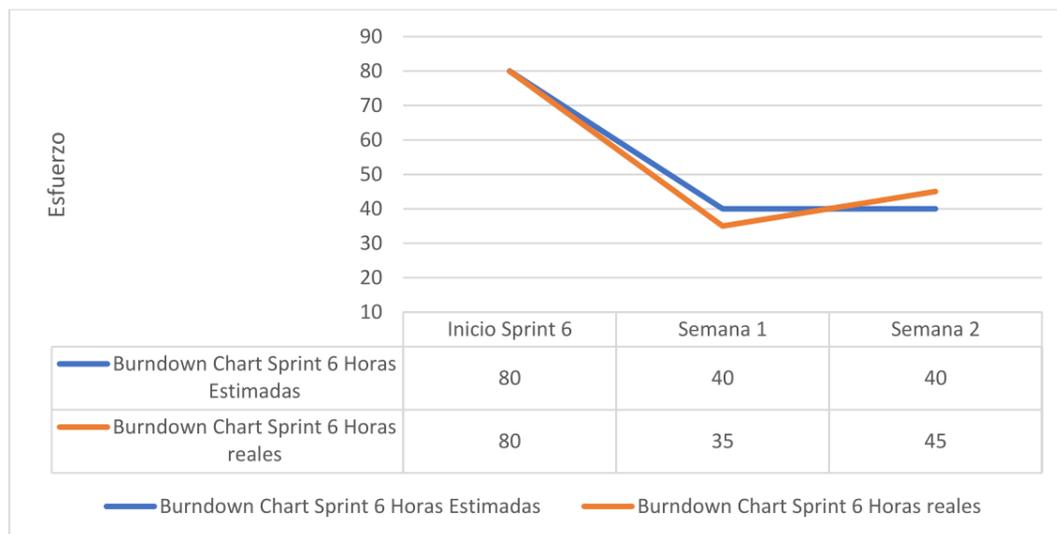


Figura 13. Ejecución del Sprint 6. Elaborado por autores.

Inspección y adaptación

Sprint Review

Se procedió a realizar la revisión del Sprint 6 de acuerdo con los criterios establecidos de aceptación definidos para las historias de usuario. La Tabla 15 muestra los resultados del Sprint Review de forma clara y detallada.

Tabla 15. Sprint Review Sprint 6. Elaborado por autores

Código	Historia de Usuario	Cumplimiento
HUEM04-05	Modificar la interfaz de búsqueda	✓
HUEM05-01	Recibir una lista validada de actores y temas.	✓

Actualización y refinamiento

Después de realizarse el Sprint 6, se pudo notar que no se presentó ninguna situación que traiga consigo incrementar nuevas Historias de Usuario, por lo tanto, no se realizaron cambios en Product Backlog o en el Release Planning.

Sprint Retrospective

De acuerdo con el Burndown Chart del Sprint 6, se reflejó el esfuerzo que se ha realizado durante cada uno de los días del Sprint.

Conforme a lo propuesto esta semana se procedió a retirar líneas de código no utilizadas y mejorar al código de la interfaz gráfica, . A su vez refinar la apariencia y calidad de resultados de las listas de actores y temas.

2.3 Metodología de evaluación

2.3.1 TAM

El Modelo de Aceptación de Tecnología (Technology Acceptance Model, TAM) [46] es una teoría de Sistemas de información utilizada para predecir, explicar y aumentar la aceptación del usuario final basado en la utilidad percibida y en la facilidad de uso percibida TAM se basa en la Teoría de la Acción Razonada (Theory of Reasoned Action, TRA), que hace referencia a un modelo de psicología social, mismo que aborda las causas que dan origen a los comportamientos intencionados. TRA menciona que el desempeño de las personas de un comportamiento específico está relacionado con su intención de comportamiento (Behavioral Intention), la cual está definida por el conjunto de la Actitud (Attitude Toward Behavior) que se fundamenta en las creencias y las

evaluaciones personales; y una Norma Subjetiva (Subjective Norm) que se basa en las creencias y la motivación para cumplir una tarea.

Por otro lado, la premisa de TAM muestra que el uso del sistema está determinado por el propósito de uso, el cual depende de la practicidad y facilidad de uso percibida por el mismo [40], como se muestra en la figura. Figura 14.

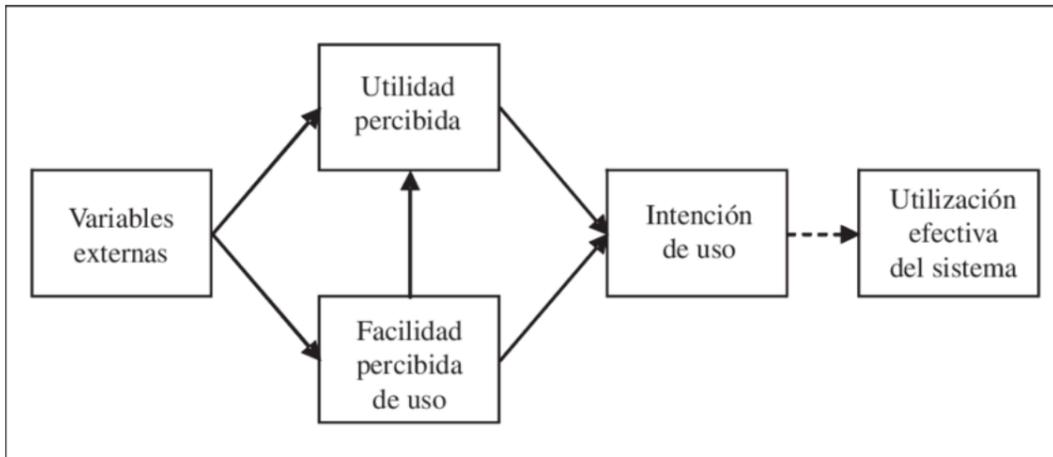


Figura 14. Proceso TAM

La motivación de TAM es valorar la aceptabilidad de sistemas de forma rápida en el proceso de implementación y diseño. Es relevante poder predecir si un determinado sistema informático será aceptado por el usuario final, poder diagnosticar las razones por las que el sistema puede no ser completamente aceptado y poder tomar decisiones correctivas y mejorar la aceptabilidad. Esto mejorará el impacto empresarial causado por la gran inversión en tiempo y dinero debido a la introducción de nuevas tecnologías de la información en la organización.

2.3.2 Detalles de evaluación

Para evaluar la herramienta web se realizó una reunión virtual con la ayuda de Zoom donde se testeó la aplicación web en tiempo real y se esperó a que cada uno de los participantes obtengan los resultados de la página.

En esta reunión se realizó una evaluación cuantitativa con 5 preguntas de escala Likert y una libre para acoger recomendaciones de los colaboradores. Las 6 preguntas fueron enviadas en un formulario de Google, esto con el fin de analizar las estadísticas de cada una de ellas. Las preguntas realizadas fueron las siguientes:

1. ¿Considera que la página web cumple su propósito de generar lista de actores y temas de un tema en específico?
2. ¿Qué tan satisfecho/a está con la facilidad de uso de la página web?
3. ¿Qué tan satisfecho/a está con la apariencia de la página web?
4. ¿Qué tan satisfecho/a está con el tiempo de respuesta de la búsqueda?
5. ¿Qué tan probable es que usted le recomiende esta página web a un/a amigo/a o miembro de la familia?
6. ¿Tiene alguna recomendación de mejora?

Se debe tomar en cuenta que la página web fue testada en su fase beta y a pesar de que la lógica del negocio estaba finalizada, se incorporaron posteriormente mejoras en el diseño gráfico de la página web.

CAPÍTULO 3

3.1 RESULTADOS Y DISCUSIÓN

3.1.1 Producto final

Arquitectura Física

La aplicación web se encuentra alojada en un servidor virtual privado rentado en Linode. Se decidió realizar el levantamiento de los dos servicios en la misma VPS durante el desarrollo y pruebas.

La VPS tiene las siguientes características:

- Sistema Operativo: Ubuntu 18.04 LTS,
- 1 CPU compartido
- 25 GB de almacenamiento SSD
- 1 GB de memoria RAM

La máquina virtual tiene dos servicios levantados en diferentes puertos. El frontend usa el puerto 80 y el backend el puerto 3000. El backend se compone de una API REST que es la encargada de la comunicación entre ambos servidores.

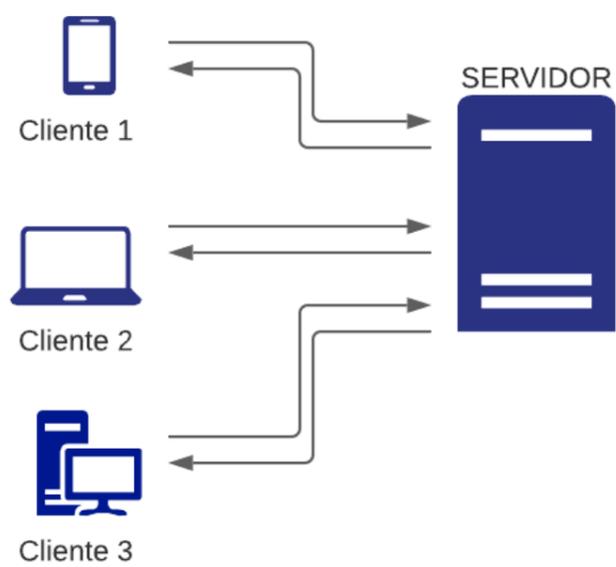


Figura 15. Arquitectura física

Arquitectura Lógica

Los clientes que van a usar la herramienta web ingresan en por medio de un navegador, el cual por medio de peticiones HTTP se comunica a la API Rest de Flask para posteriormente recibir un JSON de respuesta [9],[11].

El backend recibe la petición mediante una URI. En el backend se encuentra la lógica del negocio, la cual, con la información recibida, realiza las tareas de Web Scraping para buscar información en las diferentes fuentes, procesar los resultados y devolver la información separada en actores y temas (Figura 21).

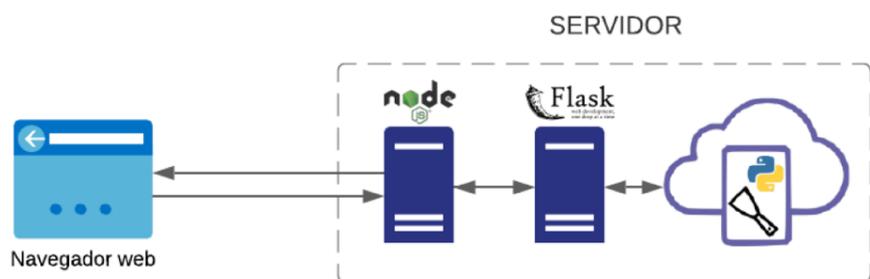


Figura 16. Arquitectura lógica

3.1.2 Descripción de los Artefactos

Aplicativo Web

La aplicación web se concibió para presentar una interfaz gráfica simple y de fácil acceso al público en general. Por lo tanto, para hacer uso de esta herramienta no es necesario registrarse ni mucho menos crear un usuario. En la pantalla inicial se puede ver un menú superior que corresponde a accesos directos a las diferentes partes de la página. Adicionalmente se muestra un resumen del objetivo de la aplicación (Figura 17).



Figura 17. Pantalla de Bienvenida

Para empezar a usar la herramienta, se debe dar clic o desplazar hacía abajo hasta el apartado de targeting (Figura 18).



Figura 18. Panel superior

Una vez que se dirija al apartado de Targeting, se requiere el ingreso de información en 3 partes importantes, de las cuales, solo las primeras dos son obligatorias (Criterios de búsqueda y Fuentes de Información) (Figura 19).



Figura 19. Pantalla del Buscador del TargeBot

Criterios de búsqueda

Los Criterios de búsqueda o temática se refieren al tema global o general en el que el sistema se va a enfocar. Entorno a esta temática el sistema intentará buscar información en las diferentes fuentes de información (Ej.: coronavirus, crisis bancaria).

Si la temática es genérica, el número de artículos aumentará, pero la precisión de los resultados disminuirá (Figura 20).



Figura 20. Temática

Fuentes de información

Como su título lo indica, esta sección permite al usuario escoger las fuentes de información que se va a usar para obtener artículos relacionados a la temática general de búsqueda. Estas fuentes solo pueden ser seleccionadas en un solo idioma. Si se selecciona alguna fuente en español, las fuentes en inglés se deshabilitarán y viceversa.

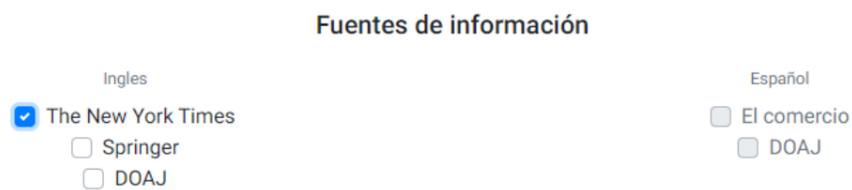


Figura 21. Fuentes de información en inglés

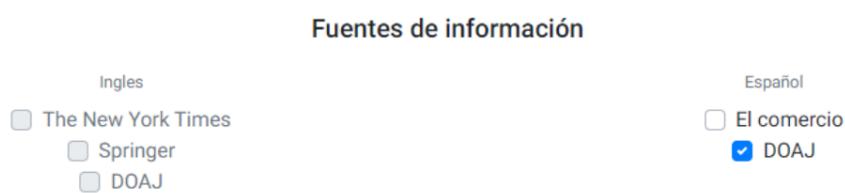


Figura 22. Fuente de información español

Esta particularidad de selección de un único idioma existe con el fin de obtener resultados más precisos. Como se explicó en el capítulo uno, la lematización y el stemming varían dependiendo el idioma.

Palabras clave

Las palabras clave no son un campo obligatorio, pero pueden llegar a ser esenciales para precisar los resultados. Estas palabras clave no tienen un límite. Sin embargo, si existe una sobrecarga de palabras clave, convergería a un resultado ilusorio (Figura 23).



Figura 23. Pantallas palabras claves 1

Las palabras agregadas se enlistarán entre el botón AGREGAR PALABRA y GENERAR LISTA. Y pueden ser eliminadas haciendo clic en la X después de la palabra clave como se muestra en la Figura 24.



Figura 24. Pantalla con dos palabras claves.

Para que la herramienta inicie la generación de listas de actores y temas, es necesario hacer clic en el botón de GENERAR LISTA (Figura 25).



Figura 25. Botón Generar lista

El sistema se encargará de informarnos del proceso de generación de listas mediante el avance de una barra que aparecerá en la parte superior (Figura 26).

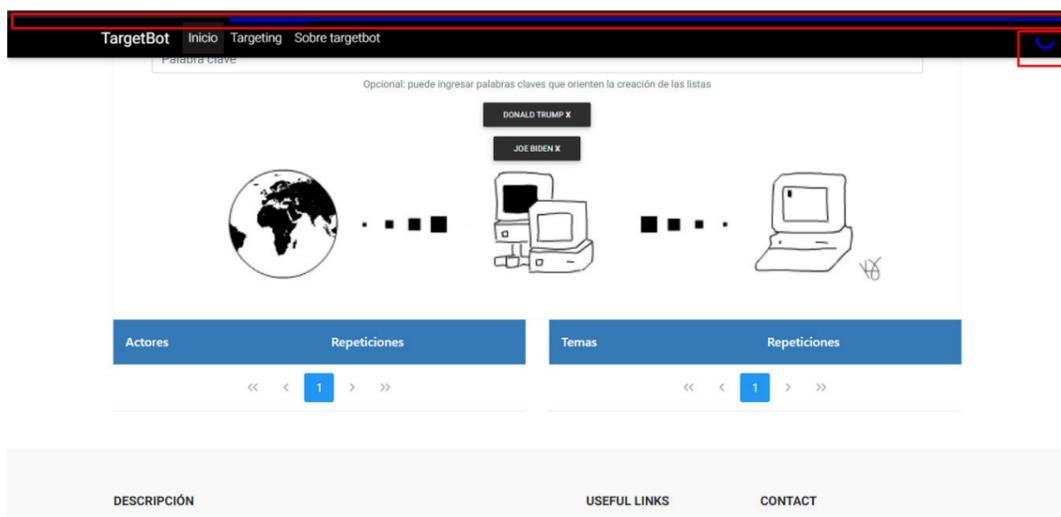


Figura 26. Pantalla de espera

Respuesta

Después de que el backend procese toda la información, se retornará las listas obtenidas en tablas de actores y temas con el número de repeticiones en la columna

derecha. En el ejemplo de la Figura 27 se muestran los resultados obtenidos al realizar la búsqueda de actores y temas relacionados con la pandemia de coronavirus.

Actores	Repeticiones	Temas	Repeticiones
RT-PCR	222	sars cov-2	2470
COVID	186	covid-19 pandemic	1429
IQR	128	covid-19 patient	958
the World Health Organization	92	coronavirus disease	802
RNA	92	disease covid-19	686
WHO	87	acute respiratory	636
PE	86	coronavirus disease covid-19	545
IgG	77	respiratory syndrome	514
PPE	70	acute respiratory syndrome	468
CXR	68	severe acute	450

Figura 27. Respuesta de TargeBot

3.1.3 Control de errores

La página web trata de evitar errores conocidos con la validación de los datos ingresados. Estas validaciones son 3 importantes:

1. Siempre debe existir una temática para iniciar la extracción de información caso contrario se visualizará un mensaje de error (Figura 28).

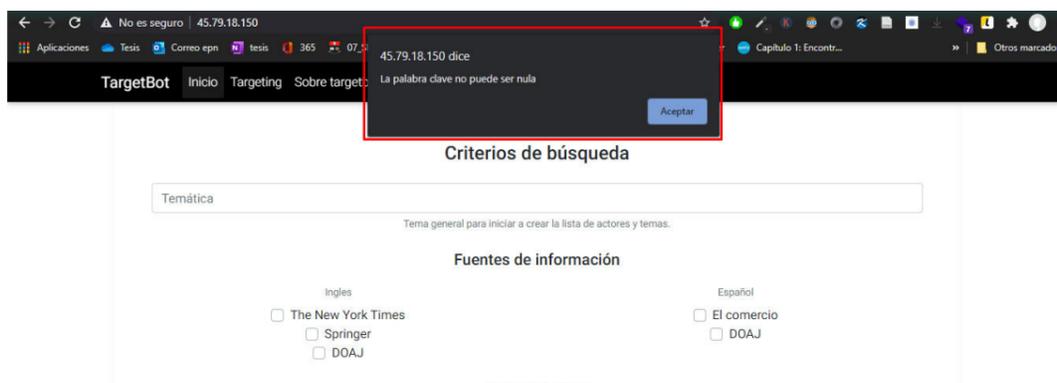


Figura 28. Mensaje de error por falta de temática

2. Siempre se debe seleccionar al menos una fuente de información para la obtención de artículos o noticias, caso contrario, se mostrará una notificación diciendo: "Debe seleccionar al menos una fuente de información", como se muestra en la Figura 29.

covid

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

Debe seleccionar al menos una fuente de información

GENERAR LISTA

Figura 29. Mensaje de error por falta de fuentes de información

3. Las fuentes de información deben seleccionarse en un mismo idioma debido a los procesos previos que se realizan para obtener las listas de actores y temas (Figura 30).

TargetBot Inicio Targeting Sobre targetbot

covid

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 30. Selección de una fuente de información

3.1.4 Pseudocódigo de algoritmos para lematización y recolección de artículo

Algoritmo para recolectar artículos científicos mediante una API

El pseudocódigo a continuación representa la secuencia ejecutada para recolectar artículos científicos de las diferentes bases a consultarse.

```
Función execute ( n,query )
//Link a la api de la base académica
  requestUrl = "http://api.springernature.com/meta/v2/json?q=(query) "
  request = requests.get(requestUrl)
//Obtener los resultados de la api
  Si request.status_code==200. Entonces
    Para documento Hasta documentos Con Paso paso Hacer
      //Guardar en el arreglo los abstract
      data['datasarticulo'].append({'text': documento['abstract']})
    Fin Para
  SiNo
    Escribir Error al obtener pagina
  Fin Si
Fin Función
Algoritmo ObtenerArticulos
  data['datasarticulo']<-[]
  Leer cantidad_articulo
  i<-0
  //recorrer
  Repetir
    execute ( i,"Covid" )
    i<-i+1
  Hasta Que cantidad_articulo < i
FinAlgoritmo
```

Algoritmo para lematizar artículos científicos

El pseudocódigo a continuación representa la secuencia ejecutada para el algoritmo que permite lematizar artículos la información obtenida de las bases consultadas.

```
Funcion ProcesarTexto ( texto )
  Si texto="h" Entonces
    doc = nlp(texto)
  Fin Si
Fin Funcion
Algoritmo Lematizacion
  Para w in doc Con Paso Hacer
    Si (w.is_stop != True and w.is_punct != True and w.pos_ != 'CONJ' and w.pos_
    != 'SPACE' and w.pos_ != 'SYM' and w.pos_ != 'SCONJ' and w.pos_ != 'PUNCT' and w.pos_
    != 'INTJ' and w.pos_ != 'NUM' and w.pos_ != 'ADP' and len(w) > 2) Entonces
      listaPalabrasLematizadas.append(w.lemma_.lower())
    Fin Si
  Fin Para
FinAlgoritmo
```

3.1.5 Pruebas y evaluación

Pruebas del sistema

Para las pruebas de rendimiento de la herramienta web, se usaron dos terminales diferentes: una computadora portátil conectada al internet por wifi y una computadora de escritorio conectada al internet por cable. Las características de los ordenadores son las siguientes:

- Computador portátil: Windows 10, Inter core i7. 500 GB de almacenamiento SSD, 8 GB de memoria RAM.
- Computador de escritorio: Windows 10, Inter core i7, 500 GB de almacenamiento SSD, 8 GB de memoria RAM.

3.1.6 Ejemplos de resultados:

Las siguientes tablas muestran resultados obtenidos del sistema relacionados con actores y temas de diferentes temáticas:

Temática: Elecciones en el Ecuador, Fuente: El Comercio.

Tabla 16. Elecciones en el Ecuador.

Actores	Rep.	Temas	Rep.
CNE	458	nacional electoral	161
TCE	153	consejo nacional	153
Consejo Nacional Electoral	115	consejo nacional electoral	153
Tribunal Contencioso Electoral	86	electoral cne	142
Asamblea	48	nacional electoral cne	137
FE	39	usd millón	107
Justicia Social	38	contencioso electoral	97
Fuerza Ecuador	35	electoral tce	82
Código de la Democracia	29	tribunal contencioso	78
FMI	29	tribunal contencioso electoral	78
Asamblea Nacional	28	lenín moreno	73
Alianza País	26	organización político	70
ONU	26	diana atamaint	67
El Consejo Nacional Electoral	24	justicia social	67
Parlamento Andino	21	contencioso electoral tce	65
Unes	20	elección general	63
Unión por la Esperanza	19	redar social	60
Partido Social Cristiano	19	plan gobernar	59
Fondo Monetario Internacional	18	binomio presidencial	56
JRV	17	asambleísta nacional	52

Temática: Procesadores Apple, Fuente: Doaj y El Comercio.

Tabla 16. Procesadores Apple.

Actores	Rep.	Temas	Rep.
Apple	80	sistema operativo	85
Microsoft	75	windows phone	35
Google	30	teléfono inteligente	25
Nokia	15	computador escritorio	25
Android	15	ios apple	20
Power Point	10	power point	15
Dell	10	operativo android	15
Android de Samsung	5	sistema operativo android	15
Stanford	5	dispositivo móvil	15
IBM	5	teléfono celular	15
Canales de Intel	5	android google	15
Point Technology	5	apple android	10
Point	5	usuario acceder	10
FDG Computer	5	word excel	10
Banco Pichincha	5	excel power	10
Club de Suscriptores	5	basar sistema	10
Samsung	5	estudiar consultor	10
Toshiba	5	ios apple android	10
RIM	5	word excel power	10
Gartner	5	excel power point	10

Temática: Video Games, Fuente: New York Times.

Tabla 17. Video Games.

Actores	Rep.	Temas	Rep.
Nintendo	54	video game	271
Minecraft	33	arcade game	37
Microsoft	32	animal crossing	36
Sony	29	pac man	35
Paramount	18	play game	32
Electronic Arts	16	game player	30
Animal Crossing	15	virtual reality	30
Second Life	12	mr. heisholt	30
Playstation	10	game publisher	27
NBA	9	game industry	26
Heisholt Inc.	9	new video	25
Atari	9	new york	25
the Nintendo Switch	7	new video game	25
Pac-Man	7	home video	24
Aladdin	7	game play	23
Nintendo Switch	6	united states	21
Rocket League	6	independent game	20
Luigi	6	game title	20
K Sports	6	electronic arts	20
XCOM	6	real world	19

Temática: Covid, Fuente: Springer y Doaj.

Tabla 18. Covid.

Actores	Rep.	Temas	Rep.
Pfizer	48	new york	54
BBC	21	health care	52
Twitter	20	test positive	37
Facebook	18	clinical trial	33
Covid	17	wear mask	32
Trump	17	united states	32
Regeneron	15	receive vaccine	28
Jessups	15	pfizer biontech	28
Times	12	second dose	28
Congress	11	coronavirus vaccine	28
United	11	care worker	26
I.C.U.	10	health care worker	26
the Alameda Midwifery Center	9	mr. hernandez	26
F.D.A.	9	general hospital	25
C.D.C.	9	los angeles	24
BioNTech	9	say dr.	23
The New York Times Opinion	9	immune system	23
Senate	9	ms. louie	22
United Airlines	9	public health	21
AstraZeneca	8	health authority	21

3.1.7 Tiempo de respuesta

Se realizaron 12 pruebas en dos computadores a diferentes horas. En las pruebas se pudo notar que el tiempo de demora de los resultados es directamente proporcional al número de peticiones al servidor. El tiempo de demora aumenta según el número de personas que realizan la petición al servidor.

Cuando se realizaron 7 peticiones al servidor simultáneamente con diferentes temáticas y de lugares físicamente remotos el tiempo promedio de respuesta del servidor fue de 17 min.

Por otro lado, se realizaron 5 pruebas en horas diferentes que ayudaron a corroborar que el tiempo de respuesta se reduce notablemente si no hay otras peticiones en cola.

El tiempo de respuesta con una petición de acuerdo a la fuente de información se muestra en la Figura 31.

Los tiempos de respuesta son independientes de las características de la computadora por la que se accede a la aplicación web, pero si dependen de la fuente de la que se va a extraer la información y del número de peticiones concurrentes.

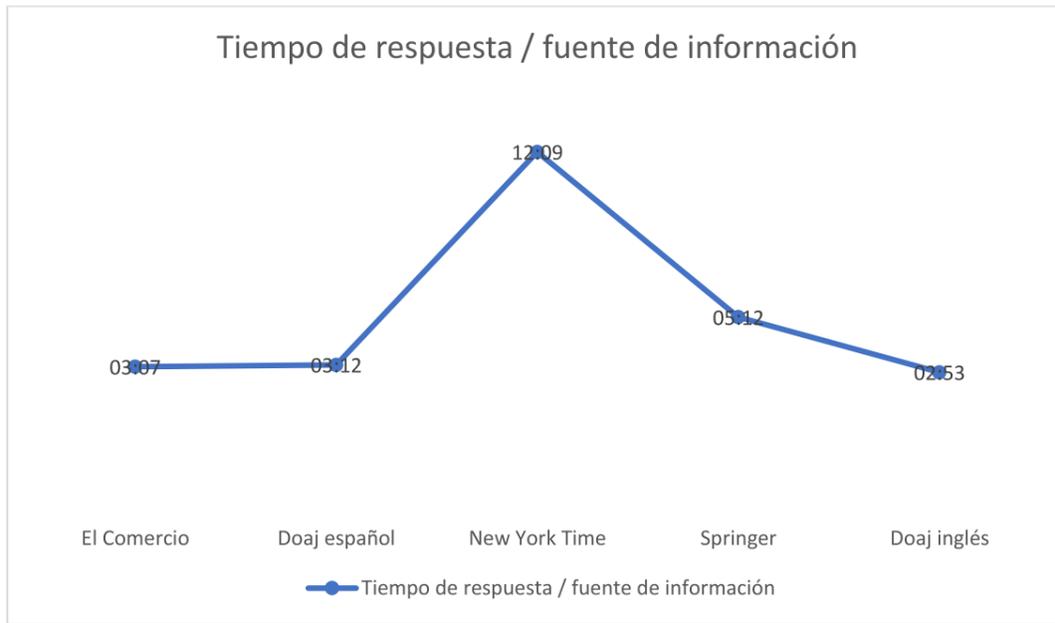


Figura 31. Tiempo de respuesta

3.1.8 Análisis de rendimiento de página web

Para realizar estas pruebas se usó la herramienta PageSpeed Insights (PSI) de Google. PSI entrega datos del rendimiento de la página web analizada al ingresar la IP pública o dominio de esta. El objetivo principal de la herramienta es ayudar a los CEO a mejorar sus páginas dándoles consejos (Figura 32).

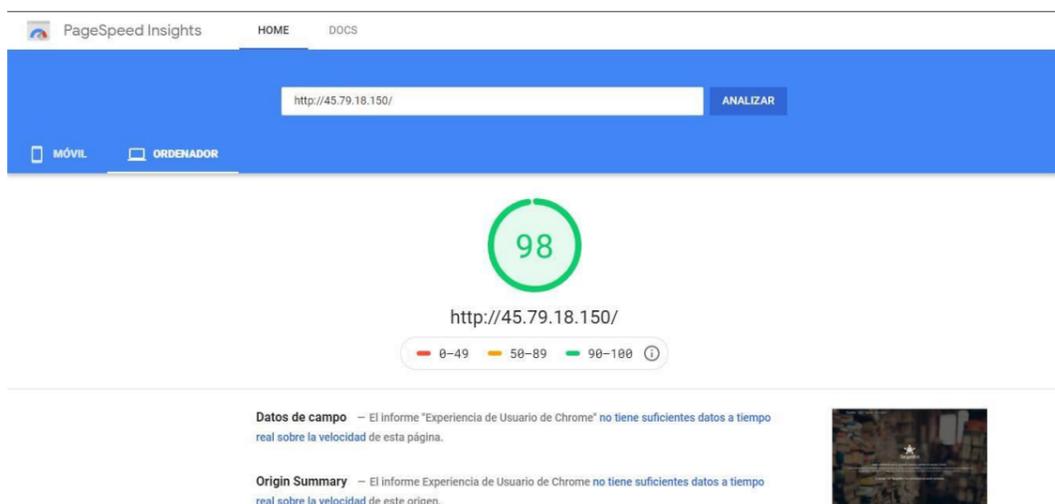


Figura 32. Análisis en PageSpeed

Como se muestra en la Figura 32 los resultados de las pruebas realizadas a la página web son buenos, y esto se debe a las herramientas usadas para el desarrollo del Front

End, no obstante, es importante tomar en cuenta que solo prueba la carga inicial de la página mas no el backend debido a que se debe llenar el formulario manualmente para comunicarse con el servidor.

3.1.9 Resultados de la Evaluación

Como se explicó en la Metodología de Evaluación, se realizó seis preguntas a siete estudiantes de la Facultad de Ingeniería en sistemas de la EPN.

Entre las personas que participaron en la evaluación se encontraban siete alumnos de la Facultad de Ingeniería en Sistemas de la Escuela Politécnica Nación y un profesor de la misma facultad. Todos los participantes habían recibido previamente formación o han trabajado en temas relacionados con VS. En base a las preguntas y respuestas de los involucrados, el formulario de Google proporcionó los siguientes resultados:

Pregunta 1:

Para esta pregunta, 6 de 7 los encuestados estuvieron de acuerdo en que la aplicación cumple el objetivo de generar una lista de actores y temas útil para VS (Figura 33).

¿Considera que la pagina web cumple su propósito de generar lista de actores y temas de un tema en especifico ?

7 respuestas

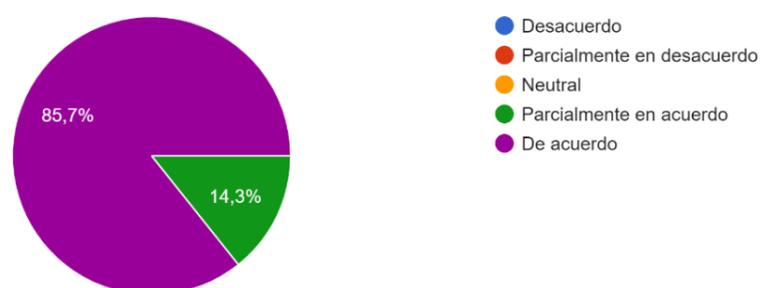


Figura 33. Pregunta 1

Pregunta 2:

En el caso de la facilidad de uso, seis de siete evaluadores se encontraron satisfecho o muy satisfechos al usar la aplicación web (Figura 34).

¿Qué tan satisfecho/a está con la facilidad de uso de la pagina web?

7 respuestas

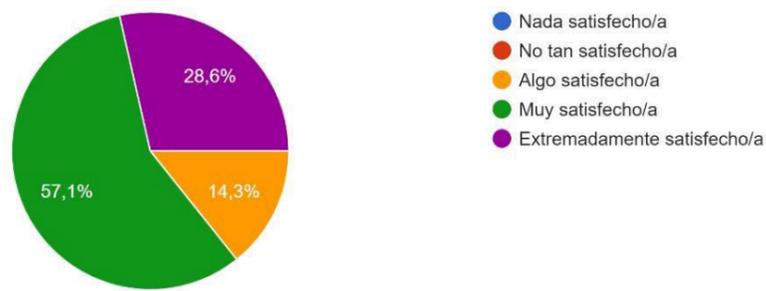


Figura 34. Pregunta 2

Pregunta 3:

Cinco de los siete evaluados declararon estar satisfechos o muy satisfechos con la apariencia de la página web. Los otros 2 percibieron que la página web puede mejorar en su apariencia (Figura 35).

¿Qué tan satisfecho/a está con la apariencia de la pagina web?

7 respuestas

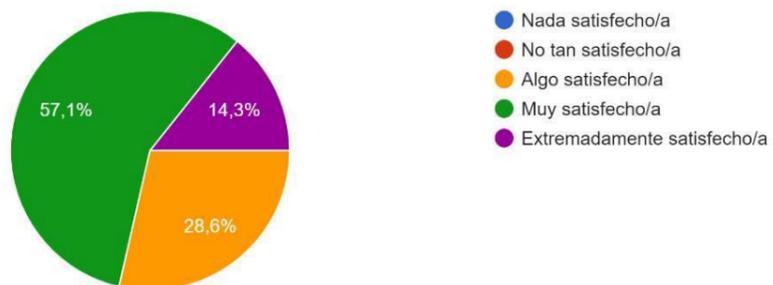


Figura 35. Pregunta 2

Pregunta 4:

Uno de los puntos bajos de la herramienta creada es el tiempo de demora en devolver los resultados al usuario y esto se pudo notar claramente en las respuestas de los encuestados (Figura 36).

¿Qué tan satisfecho/a está con el tiempo de respuesta de la búsqueda?

7 respuestas

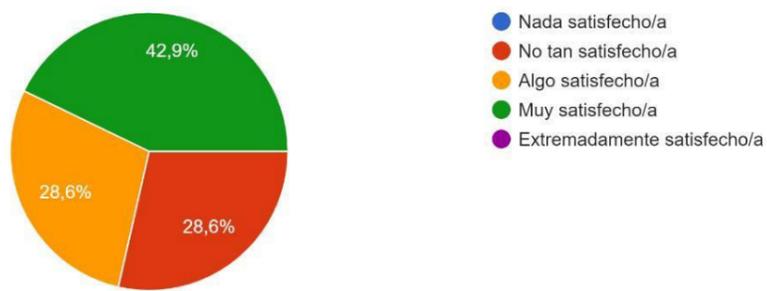


Figura 36. Pregunta 4

Pregunta 5:

Todas las personas encuestadas encontraron útil la solución que ofrece esta herramienta y recomendarían a otras personas usarla (Figura 37).

¿Qué tan probable es que usted le recomiende esta página web a un/a amigo/a o miembro de la familia?

7 respuestas

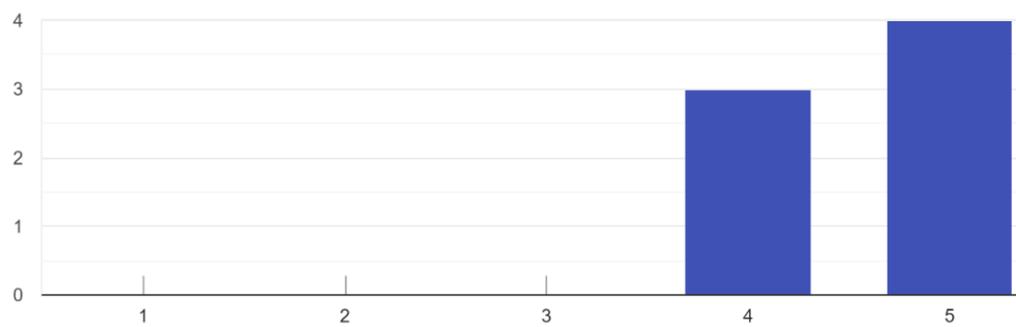


Figura 37. Pregunta 5

Pregunta 6:

Esta pregunta se la hizo con el fin de recibir recomendaciones u observaciones que pudieron no tomarse en cuenta en las preguntas anteriores. Estas recomendaciones serán usadas para una futura mejora del sistema de obtención de listas de actores y temas.

Tabla 19. Pregunta 6

N	Respuesta
1	Agregar un método u otra opción para cancelar al momento de buscar ya que cuando inicio el proceso no hay como parar el mismo, ya que pueden existir errores al momento de digitar de mala manera.
2	En la parte de Descripción al pie de página, dentro del diseño se puede mejorar.
3	El tiempo de búsqueda, creo que sería lo único que mencionar.
4	Existe una barra azul que indica el tiempo de respuesta, se podría colocar el porcentaje de búsqueda para que el usuario tenga una mejor visualización.
5	Respecto al diseño de la pantalla, en la parte de abajo de la pantalla se encuentra información que no se encuentra llenada, entiendo que se basaron en un template para la creación de la página web y pues eso está que le vean si le completan o si le quitan sin problema. De ahí todo lo demás está súper chévere.
6	Sería bueno agregar un tiempo promedio a la búsqueda si es posible. Dar una opción para una búsqueda rápida y una búsqueda a profundidad que el rápido saque los primeros 20 actores y temas que encuentre y el de profundidad si haga la búsqueda completa. También agregar un botón para cancelar la búsqueda por si acaso haya alguna equivocación en el criterio para no tener que refrescar toda la página en caso de equivocarse.
7	Existe una barra azul que indica el tiempo de respuesta, se podría colocar el porcentaje de búsqueda para que el usuario tenga una mejor visualización.

CAPÍTULO 4

4.1 CONCLUSIONES Y RECOMENDACIONES

4.1.1 Conclusiones

- Con la ayuda del uso del targeting se consigue la descripción de las perspectivas estratégicas y objetivos de la organización. Generalmente conllevaba a realizar reuniones para generar la lista de actores y temas de forma manual. Acorde a las sugerencias de los gerentes se desarrolló la herramienta para disparar las reuniones y obtener las listas de una forma más dinámica.
- Mediante el uso de la metodología Scrum se pudo trabajar en un equipo con 2 desarrolladores, las reuniones se llevaron a cabo de manera virtual con el Scrum Master, consiguiendo de esa manera tener un buen ambiente de trabajo y realizando entregas constantes de avances del producto.
- El beneficio de la metodología ágil Scrum consta del involucramiento del cliente en todo el proceso de desarrollo, permitiendo recibir retroalimentaciones del interesado del producto, cumpliendo las necesidades del cliente de una manera óptima.
- En este proyecto se desarrolló una herramienta capaz de ofrecer al usuario una lista de actores y temas que permita disparar discusiones para definir un perímetro de búsqueda en VS. Para ello, se contó con la ayuda de algunas técnicas de procesamiento de lenguaje natural, web scraping y desarrollo web que permitieron desplegar los resultados de forma amigable.
- Después de la evaluación realizada a este proyecto de titulación se ha determinado que la aplicación tiene una interfaz gráfica simple y que cumple con su objetivo. Esto a pesar del tiempo (en minutos) que se tarda en el levantamiento de estas listas a partir de varias fuentes bibliográficas, representa una ganancia importante frente a la construcción de semejantes listas de forma manual.
- Por otra parte, las pruebas automáticas realizadas al frontend mostraron resultados excelentes que intervienen directamente en la escala de resultado en las búsquedas de Google, esto facilita la exposición de la aplicación web y conjuntamente la divulgación de este proyecto de titulación.

- Los lenguajes de programación usados en este trabajo de titulación tanto en el frontend como el backend tienen muchos aportes que agilitaron el desarrollo del sistema además que el código es más limpio y ordenado.
- El targeting es solo una parte de todo la VS por lo que se debe continuar con trabajos futuros que hagan uso de este sistema para acelerar el proceso.

4.1.2 Recomendaciones

- El targeting es un proceso dependiente del procesamiento de la información de las diferentes fuentes, por lo que es importante entender que cada fuente de información no se comporta igual. Es decir, la forma de entregar la información no es un estándar y afecta directamente en el tiempo de desarrollo. Se recomienda hallar la manera de estandarizar primero la forma en la que se obtiene la información para acceder a muchas más fuentes de información.
- En español, es muy complicado reconocer todas las conjugaciones de todas las palabras. La utilización de redes neuronales convolucionales puede ayudar a mejorar los resultados de las búsquedas.
- En la actualidad hay muchas librerías enfocadas en el procesamiento de lenguaje natural, pero todas tienen sus mejores características funcionales en inglés, en trabajos futuros se podría realizar una librería con la ayuda de la inteligencia artificial y redes neuronales que etiquete las palabras con más precisión.
- Se recomienda el uso de Scrum en grupos de trabajo pequeños, debido a que permite un involucramiento del cliente en el desarrollo del producto, reduciendo el riesgo de no cumplir con las expectativas de este.
- En esta herramienta no se pudo erradicar todos los falsos positivos y falsos negativos de cada grupo de palabras, por lo que se recomienda realizar más herramientas que realicen el mismo trabajo, pero desde diferente enfoque. En tal virtud es necesario realizar más pruebas y con otras fuentes de información.

CAPÍTULO 5

5.1 REFERENCIAS BIBLIOGRÁFICAS

- [1] INEC, «Tecnologías de la Información y Comunicaciones (TICS) 2016», *Tecnologías de la Información y Comunicaciones (TIC'S) 2016*, 2016.
- [2] S. Wang y W. Shi, *Data mining and knowledge discovery*. 2012.
- [3] C. W. Choo, «The Art of Scanning the Environment. Bulletin of the American Society for Information Science». pp. 21-24, 1999.
- [4] Humbert LESCA, «Veille stratégique pour le management stratégique: Etat de la question et axes de recherche», *Economies et Sociétés*, vol. 20, n.º 5. pp. 31-50, 1994.
- [5] N. Lesca, M. L. Caron-Fasan, y S. Falcy, «How managers interpret scanning information», *Information and Management*, vol. 49, n.º 2, pp. 126-134, 2012.
- [6] H. Lesca y N. Lesca, *Strategic Decisions and Weak Signals*. 2014.
- [7] E. L. Aguirre, N. Lesca, H. Haddad, y 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, n.º 1, pp. 615-638, 2016.
- [8] X. Liu y D. Wu, «From natural language to programming language», en *Innovative Methods, User-Friendly Tools, Coding, and Design Approaches in People-Oriented Programming*, 2018.
- [9] A. Ronacher, «Welcome to Flask — Flask Documentation (1.1.x)», *Pallets Projects*, 2020. [En línea]. Disponible en: <https://flask.palletsprojects.com/en/1.1.x/>. [Accedido: 11-jul-2020].
- [10] «API – Directory of Open Access Journals (DOAJ)». [En línea]. Disponible en: <https://doaj.org/api/v2/docs>. [Accedido: 15-jul-2020].
- [11] «Angular - Introduction to the Angular Docs». [En línea]. Disponible en: <https://angular.io/docs>. [Accedido: 15-jul-2020].
- [12] N. Lesca y M. L. Caron-Fasan, «Strategic scanning project failure and

abandonment factors: Lessons learned», *European Journal of Information Systems*, 2008.

- [13] B. Vandenbosch y S. L. Huff, «Searching and scanning: How executives obtain information from executive information systems», *MIS Quarterly: Management Information Systems*, 1997.
- [14] V. Choudhury y J. L. Sampler, «Information specificity and environmental scanning: An economic perspective», *MIS Quarterly: Management Information Systems*, 1997.
- [15] M. Xu, V. Ong, Y. Duan, y B. Mathews, «Intelligent agent systems for executive information scanning, filtering and interpretation: Perceptions and challenges», *Information Processing and Management*, 2011.
- [16] T. Li y M. Li, «An Investigation and Analysis of Information Overload in Manager's Work», *iBusiness*, 2011.
- [17] A. J. Stanley y P. S. Clipsham, «Information overload - Myth or reality?», *IEE Colloquium (Digest)*, 1997.
- [18] E. L. Aguirre, M. L. Caron-Fasan, H. Haddad, y N. Lesca, «Using a meeting room system to improve targeting of Strategic Scanning», en *Proceedings of the 1st International Conference on Enterprise Systems, ES 2013*, 2013.
- [19] Peter Smith, *Professional Website Performance: Optimizing the Front End and the Back End*. 2013.
- [20] O. A. El Sawy, «PERSONAL INFORMATION SYSTEMS FOR STRATEGIC SCANNING IN TURBULENT ENVIRONMENTS: CAN THE CEO GO ON-LINE?», en *Proceedings of the International Conference on Information Systems*, 1984.
- [21] Y. Zhang, Y. Dang, H. Chen, M. Thurmond, y C. Larson, «Automatic online news monitoring and classification for syndromic surveillance», *Decision Support Systems*, 2009.
- [22] R. Y. K. Lau, S. S. Y. Liao, K. F. Wong, y D. K. W. Chiu, «Web 2.0 Environmental scanning and adaptive decision support for business mergers and acquisitions», *MIS Quarterly: Management Information Systems*, 2012.
- [23] S. Lema y J. Ochoa, *Desarrollo de una red social empresarial y de una aplicación*

móvil para recolectar y enriquecer información en un sistema de vigilancia estratégica. Quito: EPN, 2020.

- [24] «02.1 Bases de Datos de Acceso Abierto | Biblioteca Virtual @ Poli». [En línea]. Disponible en: <https://bibliopoli.wordpress.com/12-bases-de-datos-de-acceso-abierto/>. [Accedido: 18-may-2020].
- [25] «SciELO.org». [En línea]. Disponible en: <https://scielo.org/es/>. [Accedido: 18-may-2020].
- [26] «About Springer». [En línea]. Disponible en: <https://www.springer.com/la/about-springer>. [Accedido: 10-ene-2021].
- [27] «ERIC - Education Resources Information Center». [En línea]. Disponible en: <https://eric.ed.gov/>. [Accedido: 18-may-2020].
- [28] «e-Revistas. Plataforma Open Access de Revistas Científicas Electrónicas Españolas y Latinoamericanas». [En línea]. Disponible en: https://ddd.uab.cat/pub/ciencias/ciencias_a2012m3n21/suplement/index.html.4. [Accedido: 15-nov-2020].
- [29] «arXiv.org e-Print archive». [En línea]. Disponible en: <https://arxiv.org/>.
- [30] «Sistema de Información Científica Redalyc, Red de Revistas Científicas». [En línea]. Disponible en: <https://www.redalyc.org/>. [Accedido: 15-nov-2020].
- [31] R. Mitchell, *Web Scraping with Python, 2nd Edition*. 2018.
- [32] S. vanden Broucke y B. Baesens, *Practical Web Scraping for Data Science*. 2018.
- [33] D. K. Vishwakarma, D. Varshney, y A. Yadav, «Detection and veracity analysis of fake news via scrapping and authenticating the web search», *Cognitive Systems Research*, 2019.
- [34] «¿Qué es un lenguaje natural?» [En línea]. Disponible en: <https://www.greelane.com/es/humanidades/inglés/what-is-a-natural-language-1691422/>. [Accedido: 26-nov-2020].
- [35] «Procesamiento del lenguaje natural ¿qué es? - IIC». [En línea]. Disponible en: <https://www.iic.uam.es/inteligencia/que-es-procesamiento-del-lenguaje-natural/>. [Accedido: 26-nov-2020].

- [36] «Our Documentation | Python.org». [En línea]. Disponible en: <https://www.python.org/doc/>. [Accedido: 26-nov-2020].
- [37] «TypeScript: The starting point for learning TypeScript». [En línea]. Disponible en: <https://www.typescriptlang.org/docs>. [Accedido: 26-nov-2020].
- [38] «Documentación | Node.js». [En línea]. Disponible en: <https://nodejs.org/es/docs/>. [Accedido: 09-jul-2020].
- [39] «GitHub: Where the world builds software · GitHub». [En línea]. Disponible en: <https://github.com/>. [Accedido: 10-ene-2021].
- [40] «Beautiful Soup Documentation — Beautiful Soup 4.9.0 documentation». [En línea]. Disponible en: <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>. [Accedido: 10-ene-2021].
- [41] «Natural Language Toolkit — NLTK 3.5 documentation». [En línea]. Disponible en: <https://www.nltk.org/>. [Accedido: 10-ene-2021].
- [42] «spaCy · Industrial-strength Natural Language Processing in Python». [En línea]. Disponible en: <https://spacy.io/>. [Accedido: 10-ene-2021].
- [43] A. Menzinsky, G. López, y J. Palacio, *Scrum Manager: Guía de formación*. 2016.
- [44] K. Schwaber y J. Sutherland, «La Guía de Scrum. La Guía Definitiva de Scrum: Las Reglas del Juego.», *Scrum.Org*, 2017.
- [45] Y. Rodríguez-Cruz y M. Pinto, «Modelo de uso de información para la toma de decisiones estratégicas en organizaciones de información», *Transinformacao*. 2018.
- [46] L. G. Wallace y S. D. Sheetz, «The adoption of software measures: A technology acceptance model (TAM) perspective», *Information and Management*, 2014.
- [47] E. Loza-Aguirre y A. Buitrago-Hurtado, «Qualitative assessment of user acceptance within Action Design Research and Action Research: two case studies», *Latin American Journal of Computing*, 2014.

CAPÍTULO 6

6.1 ANEXO I. TABLAS RELACIONADAS AL PROCESO SCRUM

6.1.1 Sprint 0

Fechas de las fases

Tabla 20. Fecha de las fases del Sprint 0. Elaborado por autores

Fecha	Fase
03/09/2020	Sprint Planning.Sprint Goal Elaboración del Sprint Backlog
04/09/2020-15/09/2020	Ejecución del Sprint. Daily Scrum. Ejecución de las tareas. Trabajo pendiente (Burndown chart):
16/09/2020	Inspección y Adaptación Sprint Review Sprint Retrospective. Actualización y refinamiento del Product Backlog y Release Planning

Historia del usuario

Tabla 21. Historia del usuario Sprint 0.

Historia del Usuario	HUEL01-01
Título: Lematizar de un texto	
Descripción: Como un Usuario, quiero lematizar de un texto con el fin de presentar como resultado el número de veces que se repite las palabras	
Prioridad: Alta	Estimación: 25
Criterio de Aceptación: Retornar la palabra en su estado lematizado. La cantidad de veces que se repite en el texto.	

Tabla 22. Historia del usuario Sprint 0.

Historia del Usuario	HUEM02-01
Título: Obtener el texto de una noticia	
Descripción: Como un Usuario, quiero obtener el texto de una noticia con el fin de crear de un corpus de noticias filtrando el texto y titulo.	
Prioridad: Alta	Estimación: 25
Criterio de Aceptación: Recibir como salida un archivo formato Excel con el título y texto de la noticia. El texto debe estar sin tweets.	

Sprint 0 Backlog

Tabla 23. Sprint 0 Backlog. Elaborado por autores

Código	Historia de Usuario	Criterios de aceptación
HUEL01-01	Lematizar de un texto	Retornar la palabra en su estado lematizado. La cantidad de veces que se repite en el texto.
HUEM02-01	Obtener el texto de una noticia	Recibir como salida un archivo formato Excel con el título y texto de la noticia. El texto debe estar sin tweets.

6.1.2 Sprint 1

Fechas de las fases

Tabla 24. Fecha de las fases del Sprint 1. Elaborado por autores

Fecha	Fase
17/09/2020	Sprint Planning. Sprint Goal. Elaboración del Sprint Backlog
18/09/2020 – 29/09/2020	Ejecución del Sprint. Daily Scrum. Ejecución de las tareas Trabajo pendiente (Burndown chart):
30/09/2020	Inspección y Adaptación Sprint Review. Sprint Retrospective Actualización y refinamiento del Product Backlog y Release Planning

Historia del usuario

Tabla 25. Historia del usuario Sprint 1. Elaborado por autores

Historia del Usuario	HUEL01-02
Título: Añadir una funcionalidad de stemming en el módulo generador de listas.	
Descripción: Como un Usuario, quiero añadir una funcionalidad de stemming con el fin de reducir los grupos de palabras y obtener la raíz de cada palabra	
Prioridad: Media	Estimación: 10
Criterio de Aceptación: Reducir los grupos de palabras Obtener la raíz de cada palabra	

Tabla 26. Historia del usuario Sprint 1. Elaborado por autores

Historia del Usuario	HUEL01-03
Título: Validar la duplicidad de la lista de actores y temas.	
Descripción: Como un Usuario, quiero validar la duplicidad de la lista de actores y temas con el fin de filtrar los actores que salen en la lista de temas.	
Prioridad: Media	Estimación: 10
Criterio de Aceptación: No repetir ningún actor en la lista de temas.	

Tabla 27. Historia del usuario Sprint 1. Elaborado por autores

Historia del Usuario	HUEL01-04
Título: Obtener un archivo de resultados con actores y temas	
Descripción: Como un Usuario, quiero obtener un archivo de resultados con actores y temas con el fin de visualizar los resultados obtenidos del generador de lista.	
Prioridad: Alta	Estimación: 25
Criterio de Aceptación: Archivo formato json. Lista de actores y temas con las veces que se repiten.	

Sprint 1 Backlog

Tabla 28. Sprint 1 Backlog. Elaborado por autores

Código	Historia de Usuario	Criterios de aceptación
HUEL01-02	Añadir una funcionalidad de stemming en el módulo generador de listas.	Reducir los grupos de palabras Obtener la raíz de cada palabra
HUEL01-03	Validar la duplicidad de la lista de actores y temas.	No repetir ningún actor en la lista de temas.
HUEL01-04	Obtener un archivo de resultados con actores y temas.	Archivo formato json. Lista de actores y temas con las veces que se repiten.

6.1.3 Sprint 2

Fechas de las fases

Tabla 29. Fecha de las fases del Sprint 2. Elaborado por autores

Fecha	Fase
01/10/2020	Sprint Planning. Sprint Goal. Elaboración del Sprint Backlog
02/10/2020-13/10/2020	Ejecución del Sprint. Daily Scrum. Ejecución de las tareas. Trabajo pendiente (Burndown chart):
15/10/2020	Inspección y Adaptación. Sprint Review. Sprint Retrospective. Actualización y refinamiento del Product Backlog y Release Planning

Historia del usuario

Tabla 30. Historia del usuario Sprint 2.

Historia del Usuario	HUEL01-05
Título: Optimización de código del módulo generador de lista.	
Descripción: Como un Usuario, quiero obtener una optimización de código del módulo generador de lista con el fin de Obtener resultados más limpios y acertados a la realidad.	
Prioridad: Alta	Estimación: 2
Criterio de Aceptación: Reducción de las líneas del código.	

Tabla 31. Historia del usuario Sprint 2.

Historia del Usuario	HUEL01-06
Título: Ingresar un archivo de palabra claves	
Descripción: Como un Usuario, quiero ingresar un archivo de palabra claves con el fin de optimizar la búsqueda de actores.	
Prioridad: Alta	Estimación: 8
Criterio de Aceptación: EL generador de lista usara el archivo para buscar actores en específico. Se leerá el archivo ingresado por el usuario	

Tabla 32. Historia del usuario Sprint 2. Elaborado por autores

Historia del Usuario	HUEM02-02
Título: Obtener el texto de un artículo científico de una base de conocimientos.	
Descripción: Como un Usuario, quiero obtener el texto de un artículo científico de una base de conocimientos con el fin de crear le corpus con artículos científicos con su texto.	
Prioridad: Alta	Estimación: 8
Criterio de Aceptación: Recibir un conjunto de texto de artículos científicos.	

Sprint 2 Backlog

Tabla 33. Sprint 2 Backlog. Elaborado por autores

Código	Historia de Usuario	Criterios de aceptación
HUEL01-05	Optimización de código del módulo generador de lista.	Reducción de las líneas del código.
HUEL01-06	Ingresar un archivo de palabra claves	EL generador de lista usara el archivo para buscar actores en específico. Se leerá el archivo ingresado por el usuario.
HUEM02-02	Obtener el texto de un artículo científico de una base de conocimientos.	Recibir un conjunto de texto de artículos científicos.

6.1.4 Sprint 3

Fechas de las fases

Tabla 34. Fecha de las fases del Sprint 3. Elaborado por autores

Fecha	Fase
15/10/2020	Sprint Planning. Sprint Goal. Elaboración del Sprint Backlog
16/10/2020-27/10/2020	Ejecución del Sprint. Daily Scrum. Ejecución de las tareas. Trabajo pendiente (Burndown chart).
28/10/2020	Inspección y Adaptación. Sprint Review. Sprint Retrospective. Actualización y refinamiento del Product Backlog y Release Planning

Historia del usuario

Tabla 35. Historia del usuario Sprint 3. Elaborado por autores

Historia del Usuario	HUEF04-01
Título: Diseñar interfaces del Sistema	
Descripción: Como un Usuario, quiero diseñar interfaces del Sistema con el fin de crear una Interfaz de búsqueda de noticias y artículos por texto.	
Prioridad: Media	Estimación: 8
Criterio de Aceptación: El diseño deberá realizarse con la herramienta LucidChart. Visualizar la funcionalidad de la interfaz. Selección de framework a usar.	

Tabla 36. Historia del usuario Sprint 3.

Historia del Usuario	HUEF04-02
Título: Implementar interfaces del Sistema	
Descripción: Como un Usuario, quiero implementar interfaces del Sistema con el fin de realizar una búsqueda de un tema específico mediante una interfaz	
Prioridad: Alta	Estimación: 22
Criterio de Aceptación: Se ingresa el tema a buscar Se selecciona la fuente de los datos de la búsqueda Se ingresa un conjunto opcional de palabras claves Se despliega la tabla de resultados	

Tabla 37. Historia del usuario Sprint 3.

Historia del Usuario	HUEB03-01
Título: Diseñar el backend de la aplicación	
Descripción: Como un Usuario, quiero diseñar el backend de la aplicación con el fin de controlar todos los módulos y lógica de la página.	
Prioridad: Alta	Estimación: 8
Criterio de Aceptación: Selección del framework con el que se va a trabajar Deberá permitir un fácil mantenimiento	

Tabla 38. Historia del usuario Sprint 3.

Historia del Usuario	HUEB03-02
Título: Implementar el backend de la aplicación	
Descripción: Como un Usuario, quiero implementar el backend de la aplicación con el fin de integrar el módulo generador del corpus y generador de lista.	
Prioridad: Alta	Estimación: 12
Criterio de Aceptación: Se deberá contar con controladores para crear las rutas de los servicios web Implementado el módulo generador de la lista Integrado el creador de corpus	

Sprint 3 Backlog

Tabla 39. Sprint 3 Backlog. Elaborado por autores

Código	Historia de Usuario	Criterios de aceptación
HUEF04-01	Diseñar interfaces del Sistema	El diseño deberá realizarse con la herramienta LucidChart. Visualizar la funcionalidad de la interfaz. Selección de framework a usar.
HUEF04-02	Implementar interfaces del Sistema	Se ingresa el tema a buscar Se selecciona la fuente de los datos de la búsqueda Se ingresa un conjunto opcional de palabras claves Se despliega la tabla de resultados
HUEB03-01	Diseñar el backend de la aplicación	Selección del framework con el que se va a trabajar Deberá permitir un fácil mantenimiento
HUEB03-02	Implementar el backend de la aplicación	Se deberá contar con controladores para crear las rutas de los servicios web Implementado el módulo generador de la lista Integrado el creador de corpus

6.1.5 Sprint 4

Fechas de las fases

Tabla 40. Fecha de las fases del Sprint 4.

Fecha	Fase
29/10/2020	Sprint Planning. Sprint Goal Elaboración del Sprint Backlog
30/10/2020 – 10/11/2020	Ejecución del Sprint. Daily Scrum Ejecución de las tareas Trabajo pendiente (Burndown chart):
11/11/2020	Inspección y Adaptación. Sprint Review Sprint Retrospective Actualización y refinamiento del Product Backlog y Release Planning

Historia del usuario

Tabla 41. Historia del usuario Sprint 4.

Historia del Usuario	HUEF04-02
Título: Implementar interfaces del Sistema	
Descripción: Como un Usuario, quiero implementar interfaces del Sistema con el fin de realizar una búsqueda de un tema específico mediante una interfaz	
Prioridad: Alta	Estimación: 22
Criterio de Aceptación: Se ingresa el tema a buscar Se selecciona la fuente de los datos de la búsqueda Se ingresa un conjunto opcional de palabras claves Se despliega la tabla de resultados	

Tabla 42. Historia del usuario Sprint 4.

Historia del Usuario	HUEB03-02
Título: Implementar el backend de la aplicación	
Descripción: Como un Usuario, quiero implementar el backend de la aplicación con el fin de integrar el módulo generador del corpus y generador de lista.	
Prioridad: Alta	Estimación: 12
Criterio de Aceptación: Se deberá contar con controladores para crear las rutas de los servicios web Implementado el módulo generador de la lista Integrado el creador de corpus	

Sprint 4 Backlog

Tabla 43. Sprint 4 Backlog.

Código	Historia de Usuario	Criterios de aceptación
HUEF04-02	Implementar interfaces del Sistema	Se ingresa el tema a buscar Se selecciona la fuente de los datos de la búsqueda Se ingresa un conjunto opcional de palabras claves Se despliega la tabla de resultados
HUEB03-02	Implementar el backend de la aplicación	Se deberá contar con controladores para crear las rutas de los servicios web Implementado el módulo generador de la lista

6.1.6 Sprint 5

Fechas de las fases

Tabla 44. Fecha de las fases del Sprint 5. Elaborado por autores

Fecha	Fase
12/11/2020	Sprint Planning. Sprint Goal Elaboración del Sprint Backlog
13/11/2020 – 24/11/2020	Ejecución del Sprint. Daily Scrum. Ejecución de las tareas. Trabajo pendiente (Burndown chart):
25/11/2020	Inspección y Adaptación. Sprint Review. Sprint Retrospective. Actualización y refinamiento del Product Backlog y Release Planning

Historia del usuario

Tabla 45. Historia del usuario Sprint 5. Elaborado por autores

Historia del Usuario	HUEB03-03
Título: Implementar integración de los servidores web	
Descripción: Como un Usuario, quiero implementar integración de los servidores web con el fin de comunicar el servidor backend y frontend y mantener una comunicación estable entre ellos.	
Prioridad: Alta	Estimación: 65
Criterio de Aceptación: El backend recibe el query de búsqueda La interfaz recibe el json con las listas	

Tabla 46. Historia del usuario Sprint 5.

Historia del Usuario	HUEF04-04
Título: Visualizar un cuadro de dialogo mientras se realiza la consulta	
Descripción: Como un Usuario, quiero visualizar un cuadro de dialogo mientras se realiza la consulta con el fin de ver el progreso de la búsqueda realizada.	
Prioridad: Alta	Estimación: 5
Criterio de Aceptación: En la página web se visualiza al momento de realizar la búsqueda.	

Tabla 47. Historia del usuario Sprint 5.

Historia del Usuario	HUEF04-05
Título: Recibir una lista validada de actores y temas.	
Descripción: Como un Usuario, quiero recibir una lista validada de actores y temas con el fin de verificar la coherencia de la información obtenida con la ayuda de un diccionario léxico.	
Prioridad: Alta	Estimación: 15
Criterio de Aceptación: Actores y temas comparados con diccionarios	

Sprint 5 Backlog

Tabla 48. Sprint 5 Backlog.

Código	Historia de Usuario	Criterios de aceptación
HUEB03-03	Implementar integración de los servidores web	El backend recibe el query de búsqueda La interfaz recibe el json con las listas
HUEF04-04	Visualizar un cuadro de dialogo mientras se realiza la consulta	En la página web se visualiza al momento de realizar la búsqueda
HUEF04-05	Modificar la interfaz de búsqueda.	Actores y temas comparados con diccionarios

6.1.7 Sprint 6

Fechas de las fases

Tabla 49. Fecha de las fases del Sprint 6.

Fecha	Fase
26/11/2020	Sprint Planning. Sprint Goal Elaboración del Sprint Backlog
27/11/2020 - 08/12/2020	Ejecución del Sprint. Daily Scrum. Ejecución de las tareas Trabajo pendiente (Burndown chart):
09/12/2020	Inspección y Adaptación. Sprint Review Sprint Retrospective Actualización y refinamiento del Product Backlog y Release Planning

Historia del usuario

Tabla 50. Historia del usuario Sprint 6. Elaborado por autores

Historia del Usuario	HUEM04-05
Título: Optimización de código del programa,	
Descripción: Como un Usuario, quiero con el fin de mejor la herramienta optimizar los algoritmos utilizados.	
Prioridad: Media	Estimación: 15
Criterio de Aceptación: Reducción de código repetido	

Tabla 51. Historia del usuario Sprint 6. Elaborado por autores

Historia del Usuario	HUEM05-01
Título: Recibir una lista validada de actores y temas.	
Descripción: Como un Usuario, quiero con el fin de recibir una lista de actores y temas validar resultados entregados en la pantalla.	
Prioridad: Alta	Estimación: 45
Criterio de Aceptación: Reducción de resultados entre ambas listas <u>Eliminar palabras incompletas</u>	

Sprint 6 Backlog

Tabla 52. Sprint 6 Backlog. Elaborado por autores

Código	Historia de Usuario	Criterios de aceptación
HUEM04-05	Modificar la interfaz de búsqueda	Reducción de código repetido
HUEM05-01	Recibir una lista validada de actores y temas.	Reducción de resultados entre ambas listas <u>Eliminar palabras incompletas</u>