

ARTICLE HISTORY

Received 25 October 2021
Accepted 23 November 2021

Carol Oña

Departamento de Informática y Ciencias de
Computación
Escuela Politécnica Nacional
Quito, Ecuador
carol.ona@epn.edu.ec

Nicole Pirca

Departamento de Informática y Ciencias de
Computación
Escuela Politécnica Nacional
Quito, Ecuador
nicole.pirca@epn.edu.ec

Lorena Recalde

Departamento de Informática y Ciencias de
Computación
Escuela Politécnica Nacional
Quito, Ecuador
lorena.recalde@epn.edu.ec
ORCID: 0000-0002-4949-3278

Edison Loza-Aguirre

Departamento de Informática y Ciencias de
Computación
Escuela Politécnica Nacional
Quito, Ecuador
edison.loza@epn.edu.ec
ORCID: 0000-0003-1481-8105

Development of a System for Strategic Surveillance of Social Networks and Digital Press Media using Web Scraping

Desarrollo de Sistema para Vigilancia Estratégica de Redes Sociales y Medios de Prensa Digitales mediante Web Scraping

Desarrollo de Sistema para Vigilancia Estratégica de Redes Sociales y Medios de Prensa Digitales mediante Web Scraping

Development of a System for Strategic Surveillance of Social Networks and Digital Press Media using Web Scraping

Carol Oña
Departamento de Informática y
Ciencias de Computación
Escuela Politécnica Nacional
Quito, Ecuador
carol.ona@epn.edu.ec

Nicole Pirca
Departamento de Informática y
Ciencias de Computación
Escuela Politécnica Nacional
Quito, Ecuador
nicole.pirca@epn.edu.ec

Lorena Recalde
Departamento de Informática y
Ciencias de Computación
Escuela Politécnica Nacional
Quito, Ecuador
lorena.recalde@epn.edu.ec
ORCID: 0000-0002-4949-3278

Edison Loza-Aguirre
Departamento de Informática y
Ciencias de Computación
Escuela Politécnica Nacional
Quito, Ecuador
edison.loza@epn.edu.ec
ORCID: 0000-0003-1481-8105

Resumen—Desde el punto de vista de los sistemas de información empresariales, ha surgido la necesidad de crear herramientas que faciliten las tareas de Vigilancia Estratégica. En la actualidad, el laboratorio ADA de la Facultad de Ingeniería de Sistemas de la Escuela Politécnica Nacional está ejecutando varios proyectos distribuidos en diferentes fases cuya meta final es desarrollar esta suite de Vigilancia Estratégica. El presente proyecto pertenece a la (tercera) fase denominada “Recolección de Información”, cuyo objetivo es desarrollar un sistema de recolección de información automatizado, optimizando los tiempos de búsqueda y reduciendo la sobrecarga de información. La herramienta desarrollada permite búsquedas sobre medios de prensa digitales relevantes y pre-establecidos y las redes sociales Twitter, Facebook, Instagram y LinkedIn.

El proyecto se ha desarrollado siguiendo los lineamientos definidos por Scrum y se basa en el patrón Modelo – Vista – Controlador. Además, el sistema ha sido evaluado por el Modelo de Aceptación Tecnológica (TAM). Los resultados de evaluación muestran que más del 80% de los encuestados consideran que el sistema posee facilidad de uso. Sin embargo, el 100% coincide respecto a la utilidad percibida del sistema en cuanto a la recolección de información: éste cumple con su objetivo orientado a recopilar información de las redes sociales y medios de prensa digitales que sea útil en el proceso de toma de decisiones basado en la Vigilancia Estratégica.

Palabras clave—Vigilancia Estratégica, Recolección de Información, Redes Sociales, Medios Digitales.

Abstract—From the point of view of business information systems, the need to create tools that facilitate the tasks of Strategic

Surveillance has arisen. Currently, the ADA laboratory of the Faculty of Systems Engineering of the National Polytechnic School is executing several projects distributed in different phases whose final goal is to develop this Strategic Surveillance suite. This project belongs to the (third) phase called "Information Collection". Thus, the objective of the project is to develop an automated information collection system which optimizes search times and reduces information overload. The developed tool allows the search to be carried out on relevant and pre-established digital press media and the social networks Twitter, Facebook, Instagram, and LinkedIn.

The project has been developed following the guidelines defined by Scrum and is based on the Model - View - Controller pattern. In addition, the system has been evaluated by the Technological Acceptance Model (TAM). The evaluation results show that more than 80% of the respondents consider that the system is easy to use. However, 100% of them coincide with respect to the perceived usefulness of the system in terms of collecting information: it fulfills its objective of collecting information that is the input in the decision-making process based on Strategic Scanning.

Keywords—Strategic Scanning, Information Collection, Social Networks, Digital Media.

I. INTRODUCCIÓN

Una organización debe resistir a los cambios y evoluciones que se producen en su entorno. Es por esto por lo que, hoy en día, la mayoría de las organizaciones realizan, en mayor o menor grado, actividades de Vigilancia Estratégica (conocida también como Strategic Scanning). La Vigilancia Estratégica es una alternativa que permite a la organización mantenerse al corriente con las tendencias y evoluciones de su entorno,

identificar amenazas, oportunidades, anticipar cambios y soportar los procesos de toma de decisiones [1].

La Vigilancia Estratégica es un proceso organizacional que implica la recolección y análisis de información con el objetivo de soportar las actividades de toma de decisiones estratégicas en una organización [1]. Dicha información puede ser obtenida de noticias, informes, comentarios, redes sociales y medios digitales relacionados al ámbito de la organización. Una vez recopilados los datos, éstos son analizados para ser convertidos en conocimiento que permita a las gerencias tomar decisiones que sean de beneficio para la organización [2].

Resulta paradójico, sin embargo, que en la era en la cual el Internet facilita el acceso a cantidades infinitas de información, las fuentes más útiles y que alimentan los sistemas de Vigilancia Estratégica son más bien directas: observaciones de terreno, comentarios recabados en conferencias donde participa la organización, “chismes” de clientes y proveedores, entre otras. De hecho, resulta necesario entender que, a pesar de los grandes avances en informática (Big Data, Smart Data, Procesamiento Natural de Lenguaje, etc.), la mayor parte del proceso de recolección de información reposa aún en las capacidades de las personas que colaboran en la institución. Es por este motivo que, el rol de la tecnología es sobre todo de soporte y no de remplazo al personal que trabaja en este proceso [2].

En el entorno económico de una organización existe una gran cantidad de información, pero no toda puede ser recolectada y analizada, ni tampoco toda esa información posee un valor estratégico útil sobre el devenir de la empresa. Debido a estos antecedentes, y con el fin de prevenir problemas de sobrecarga de información, se debe limitar el entorno o perímetro de vigilancia con la identificación de la lista de actores a vigilar, las interrelaciones entre los actores, los temas, la lista de fuentes de información a escudriñar y las palabras clave que permitirán buscar información de una problemática general en Ecuador [2].

Una vez definido el alcance de Vigilancia Estratégica, se procede a recolectar información con el fin de obtener y almacenar datos relevantes que agilicen el proceso de toma de decisiones de la organización. Dicha información puede ser obtenida de diversas fuentes como: bases de datos, Internet, personas, noticias u opiniones [3]. En lo que respecta a fuentes provenientes del Internet, esta búsqueda puede ser automatizada gracias a la utilización de herramientas y algoritmos de recolección de información como APIs y Web Spiders.

El Departamento de Informática y Ciencias de la Computación (DICC) de la Facultad de Ingeniería en Sistemas de la Escuela Politécnica Nacional, tiene como línea de investigación la Computación Aplicada a los Sistemas de Información. Dentro del DICC se está ejecutando el programa de investigación Tyche (Fig. 1). Este proyecto está organizado en varias fases que están siendo apoyadas por el desarrollo de proyectos de titulación. En el presente proyecto, que es parte de la tercera fase, se propone desarrollar una herramienta que, dado el perímetro proporcionado por la segunda fase del programa de investigación “Tyche”, permitirá la recolección de información relevante para organizaciones ecuatorianas como las de tipo pymes de manufactura de Quito.

La información será recolectada de redes sociales en línea como: Twitter, Facebook, Instagram y LinkedIn; y también se recolectará información de 5 medios de prensa en idioma español que sean digitales e influyentes en Latinoamérica. Este proceso es continuo ya que esta información debe ser siempre actualizada, con el fin de ser útil a la toma de decisiones. Esto agilizará la recopilación de información relevante, disminuyendo el esfuerzo humano. La información de Vigilancia Estratégica de este proyecto servirá para organizaciones de cualquier tamaño y que operen en cualquier dominio de actividad, como las pymes de manufactura de Quito que están interesadas en usar información que les permita reducir la incertidumbre de su toma de decisiones estratégicas. Así, la solución propuesta permitirá optimizar la búsqueda de esta información a la vez que se toman precauciones para reducir la sobrecarga de la misma.

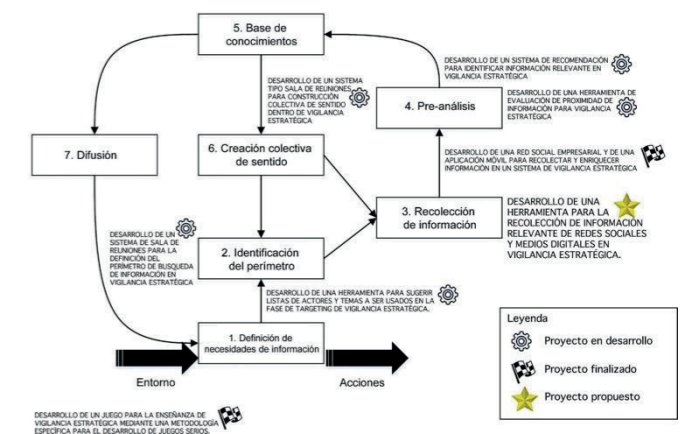


Fig. 1. Proyectos de investigación del programa “Tyche”

II. FUNDAMENTO TEÓRICO

A. Vigilancia Estratégica

En la actualidad, se accede a un exceso de información que, por su naturaleza, requiere ser procesada y filtrada para dejar de lado la que sea de baja calidad. La información a nivel comercial, o que trate temas sobre tecnología, competencia o entorno es absolutamente ineludible para tomar decisiones estratégicas y para anticiparse a los hechos en las organizaciones [4]. Sin embargo, el identificar la información de mejor calidad y la más relevante no es tarea sencilla. Entonces, la Vigilancia Estratégica entra aquí como un conjunto de acciones que una organización tiene que gestionar y poner en marcha para acceder a la información apropiada en el tiempo adecuado y que le permita tomar las decisiones convenientes [5].

El sistema de Vigilancia Estratégica se compone de tres tareas fundamentales:

- 1) Recoger la información que la empresa necesita.
- 2) Transformar toda esa información en conocimiento.
- 3) Valorar ese conocimiento y hacerlo llegar a las personas adecuadas de la organización.

De acuerdo con [6], al implementar un sistema de Vigilancia Estratégica en una organización, el objetivo que se quiere alcanzar se centra principalmente en los siguientes aspectos:

- *Anticipar los cambios*: detectar con la máxima antelación posible las eventualidades más importantes que se puedan producir en el entorno de la empresa.
- *Minimizar riesgos*: identificar amenazas que aparezcan como consecuencia de nuevos productos, normativas o competidores, y tomar las decisiones adecuadas respecto a la elección de nuevas tecnologías.
- *Comparar*: se trata de valorar tanto los puntos fuertes como las posibles debilidades contrastando frente a los de la competencia y en relación con las necesidades de los clientes.
- *Innovar*: localizar y definir oportunidades de mejora, así como ideas revolucionarias en el mercado.
- *Cooperar*: identificar oportunidades para colaborar con otras empresas y encontrar socios apropiados.

Entonces, el proceso de Vigilancia Estratégica que debe seguirse es [7]:

- Establecer las necesidades de información que tiene la organización e identificar los factores cruciales de vigilancia.
- Definir las fuentes y tipos de información relevantes, comprender la factibilidad de acceso y elegir los medios de extracción y seguimiento de la información.
- Obtener la información.

Una vez extraída la información, se la debe analizar y ampliar cuando sea necesario; además, las personas correspondientes deben tener fácil acceso a esta información para que finalmente puedan utilizarla en la toma de decisiones.

Es así como, mediante la Vigilancia Estratégica en una organización, la gerencia es capaz de identificar y monitorizar las tendencias en la cadena de valor de sus clientes, proveedores y entorno. Asimismo, se puede obtener información de valor estratégico de manera focalizada y continuada para reducir el riesgo en la toma de decisiones [7].

El aspecto más importante para su implementación es lograr que el sistema de vigilancia se adapte a la organización de forma particular, exclusiva y “personalizada”. En otras palabras, se debe focalizar en las necesidades reales de la empresa de forma que se responda a lo que realmente se quiere y se necesita.

B. Web Scraping

Actualmente, al poseer una vasta cantidad de información sobre diversos temas y tratar de obtener datos relevantes para una organización que trabaje con Vigilancia Estratégica, llega a ser necesario implementar técnicas de recopilación, organización y filtrado de datos que puedan controlar mejor la cantidad inmensa de información y además ayuden a que las organizaciones puedan procesarla de manera más eficiente.

Web Scraping es un proceso que usa bots, spiders o crawlers para extraer contenido y datos de un sitio web. De esta forma, se extrae el código HTML y junto con él, los datos almacenados en la base de datos [8].

Durante el Web Scraping se extraen y almacenan datos de páginas web para analizarlos o utilizarlos después. Gracias a este raspado web, se almacenan diversos tipos de información,

por ejemplo, datos de contacto, tales como direcciones de correo electrónico, números de teléfono, también términos de búsqueda o URL [9]. El Web Scraping se aplica en diversas tareas, por ejemplo, para recopilar datos de consumo, interacciones o información especial con gran rapidez. En el ámbito profesional, el scraping se utiliza frecuentemente para obtener ventajas respecto a la competencia [10].

El lenguaje de programación más utilizado para scrapear es Python [11], ya que tiene varias librerías para este fin: Scrapy, BeautifulSoup y Selenium. La más conocida y utilizada es probablemente Scrapy, que es un framework, y permite administrar peticiones, preservar sesiones de usuario y seguir redirecciones. Una de las mayores ventajas de Scrapy es que es muy eficiente y es capaz de scrapear más cantidad, más rápido y a menos coste de CPU que las otras alternativas [12].

C. Modelo de Aceptación Tecnológica

El Modelo de Aceptación Tecnológica, definido por sus siglas como TAM, fue propuesto por Davis en 1989 y se desarrolló con base en la Teoría de Acción Razonada (TRA) [13]. El objetivo es predecir la aceptación de los sistemas de información por los usuarios en las organizaciones que permita garantizar el éxito de su implementación [14]. En TAM se buscó establecer los criterios para comprender la intención conductual del uso. Entre los hallazgos se encontró que esta intención está influenciada más bien por una actitud individual que tiene dos factores determinantes: la utilidad percibida y la facilidad de uso percibida [14].

En total, hay cuatro variables principales en TAM que determinan el uso efectivo de la tecnología, como se explica en la Fig. 2 [15].

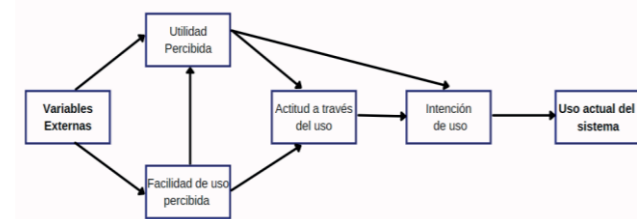


Fig. 2. Modelo de Aceptación de la Tecnología (TAM)[12]

De acuerdo con TAM, las variables externas influyen de manera directa en la Utilidad Percibida y en la Facilidad de Uso Percibida. Es así como las variables externas participan de forma indirecta en la actitud a través del uso y la intención de uso. La Facilidad de Uso Percibida tiene un efecto causal en la Utilidad Percibida, además del efecto significativo de esta variable en la actitud del usuario (un sentimiento en favor o en contra) hacia el uso del sistema [15]. En investigaciones posteriores se elimina Actitud a través del uso del modelo debido a que las medidas experimentales de esta variable no entregan evidencia suficiente para mantenerla.

III. ARQUITECTURA DEL SISTEMA

La arquitectura que se usó para la realización del sistema es Modelo – Vista – Controlador (MVC) la cual tiene como base el esquema descrito en la Fig. 3.

Como primer paso, el usuario realiza una solicitud al controlador con la información sobre lo que desea realizar. Entonces el Controlador llama y decide a quien delegar la tarea, como siguiente paso, el Modelo se encarga de realizar

las operaciones para cumplir con la solicitud. Una vez terminadas las operaciones, se regresa al Controlador los resultados el cual procede a la Vista. Finalmente, la Vista procesa y envía los datos de manera que sea visualmente entendible para el usuario.

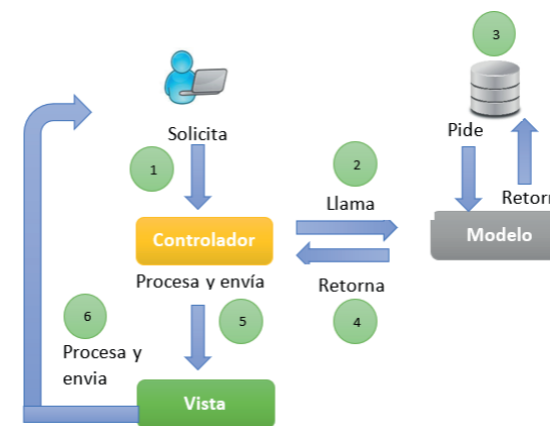


Fig. 3. Modelo MVC

IV. METODOLOGÍA SCRUM

Scrum es la metodología a ser utilizada en el desarrollo del sistema, el cual es un marco de trabajo que facilita el proceso de desarrollo de software iterativo y utilizado, comúnmente, en entornos basados en desarrollo ágil de software [16]. La realización del proyecto bajo la metodología Scrum permitirá realizar entregas de software funcional y útil para la organización. Además, se prioriza la respuesta a cambios y nuevos requerimientos, lo que genera una retroalimentación continua y una comunicación eficaz con el cliente. El desarrollo del proyecto en iteraciones facilita la división de tareas y asignación de tiempos de trabajo para los desarrolladores [17].

Dado este escenario para el proyecto y la metodología a aplicarse, las fases con sus actividades de acuerdo con el proceso planteado por Scrum son [18]:

- *Inicio*: Identificar al Scrum Máster, Stakeholder(s) y Equipo Scrum [19], desarrollar las épicas, crear la lista priorizada del Backlog y realizar el plan de lanzamiento.
- *Planificación*: Crear las Historias de Usuario, comprometer Historias de Usuario, identificar tareas, aplicar el proceso de estimación de Historias de Usuario y tareas y crear el Sprint Backlog.
- *Implementación*: crear entregables, realizar Daily Standup y refinar el Backlog priorizado del proyecto.
- *Revisión y Retrospectiva*: Demostrar y validar el Sprint, realizar la retrospectiva del Sprint.
- *Lanzamiento*: Preparar y enviar entregables y realizar retrospectiva del Proyecto.

A. Product Backlog

Las historias de usuario se definieron con criterios de aceptación, mismos que son considerados como criterios para comprobar si una historia de usuario fue desarrollada según la expectativa del dueño del producto y si esto se cumple.

Definidas las épicas y features se presenta el Product Backlog priorizado, el cual está detallado en la Tabla I.

Notación:

- HU: Historias de Usuario
- ID: Identificador
- Ez: Esfuerzo estimado

TABLA I. PRODUCT BACKLOG PRIORIZADO

HU (ID)	ENUNCIADO DE LA HISTORIA DE USUARIO	EZ
HU024	Como equipo Scrum de desarrollo se desea concretar los acuerdos en roles y disponibilidad de tiempo del Equipo Scrum con la finalidad de tener una estructura organizativa establecida dentro del proyecto.	3
HU022	Como equipo Scrum de desarrollo se desea conocer los estándares de codificación y documentación del código con el fin de realizar un desarrollo homogéneo del sistema.	3
HU092	Como equipo Scrum de desarrollo se desea instalar herramientas, programas y librerías que van a contribuir con el desarrollo del sistema web con el fin de tener un ambiente de trabajo listo para empezar a implementar el sistema web.	8
HU023	Como equipo Scrum de desarrollo se desea definir la arquitectura final del sistema con la finalidad de construir el sistema de forma organizada.	13
HU037	Yo como usuario autorizado, deseo poder obtener los rankings de los 5 medios de prensa digitales más influyentes de Latinoamérica con la finalidad de poder extraer información relevante para el proceso de recopilación de información.	8
HU0136	Yo como usuario autorizado, deseo poder extraer información de los medios de prensa elegidos con la finalidad de obtener información útil dentro del sistema de Vigilancia Estratégica.	34
HU0137	Yo como usuario autorizado, deseo ver la información obtenida de la ejecución de los spiders en la web con la finalidad de que sea accesible en todo momento por el Equipo desarrollador Scrum.	13
HU040	Yo como usuario autorizado, deseo extraer información de usuarios o páginas públicas de las Redes Sociales mediante sus APIs o librerías con la finalidad de obtener información relevante para el proceso de Vigilancia Estratégica.	21
HU0138	Yo como usuario autorizado deseo poder ver la información del usuario de la red social en la Web con la finalidad de que esta información sea accesible por el Equipo Scrum.	13
HU0135	Yo como usuario autorizado, deseo poder ingresar al sistema web con la finalidad de acceder y utilizar todas sus funcionalidades.	13
HU058	Yo como usuario autorizado, deseo poder tener la posibilidad de modificar los datos de registro de un Tema con la finalidad de poder actualizar su información.	8
HU068	Yo como usuario autorizado, deseo tener la posibilidad de eliminar un Tema registrado en caso de ser necesario.	5

HU (ID)	ENUNCIADO DE LA HISTORIA DE USUARIO	EZ
HU063	Yo como usuario autorizado, deseo registrar un Tema en el sistema con la finalidad de guardar la información y utilizarla durante el proceso de recopilación de información.	13
HU0185	Yo como usuario autorizado, deseo registrar un Actor en el sistema con la finalidad de guardar la información y utilizarla durante el proceso de recopilación de información.	13
HU025	Yo como usuario autorizado, deseo tener la posibilidad de eliminar un Actor registrado en caso de ser necesario.	5
HU0114	Yo como usuario autorizado, deseo tener la posibilidad de modificar los datos de registro de un Actor con la finalidad de poder actualizar su información.	13
HU0117	Yo como usuario autorizado, deseo tener la posibilidad de eliminar un usuario Actor registrado en caso de ser necesario.	8
HU064	Yo como usuario autorizado, deseo poder buscar a todos los usuarios relacionados con el Actor en las distintas redes sociales con la finalidad de seleccionar a los usuarios más adecuados para continuar con el proceso de registrar un Actor.	13
HU0118	Yo como usuario autorizado, deseo poder visualizar y seleccionar uno o más usuarios de las diferentes redes sociales con la finalidad de agregar usuarios al Actor y continuar con el proceso de registro del Actor.	8
HU027	Yo como usuario autorizado, deseo poder seleccionar uno o más medios de prensa digitales en español más influyentes de Latinoamérica disponibles con la finalidad de poder realizar la recolección de información en el medio más adecuado para el Actor ingresado.	13
HU067	Yo como usuario autorizado, deseo ingresar entidades similares al nombre del Actor, con la finalidad de utilizar estas entidades como parámetros de búsqueda en el proceso de recolección de información.	13
HU0119	Yo como usuario autorizado, deseo agregar usuarios de redes sociales al Actor con la finalidad de incrementar mi lista de usuarios y obtener más información de estos al momento de empezar el proceso de recolección de información.	13
HU0120	Yo como usuario autorizado, deseo poder tener la posibilidad de eliminar los medios de prensa digital del Actor registrado en caso de ser necesario.	8
HU0121	Yo como usuario autorizado, deseo poder tener la posibilidad de modificar las entidades registradas dentro del medio digital de un Actor con la finalidad de actualizar mi parámetro de búsqueda en el proceso de recolección de información.	8
HU086	Yo como usuario autorizado, deseo modificar los parámetros de consulta de Fuentes de la Celda con la finalidad de actualizar la información registrada y continuar con el proceso de recolección de información.	21
HU026	Yo como usuario autorizado, deseo visualizar la información obtenida de noticias según el criterio de búsqueda ingresado con la finalidad de identificar factores claves a la Vigilancia Estratégica.	5

HU (ID)	ENUNCIADO DE LA HISTORIA DE USUARIO	EZ
HU0126	Yo como usuario autorizado, deseo modificar el nivel de prioridad al cruce de Actor & Tema con la finalidad de indicar los cruces más relevantes a tomar en cuenta en el proceso de Vigilancia Estratégica.	13
HU0122	Yo como usuario autorizado, deseo poder crear una Matriz con el cruce de Actor/Tema con la finalidad de observar gráficamente cuales son los cruces que se pueden relacionar.	21
HU0123	Yo como usuario autorizado, deseo poder visualizar todas las Matrices Actor-Tema creadas por los colaboradores en tiempo real con la finalidad de observar el trabajo de los demás usuarios.	21
HU0124	Yo como usuario autorizado, deseo poder tener la posibilidad de eliminar la Matriz Actor-Tema creada en caso de ser necesario.	13
HU0125	Yo como usuario autorizado, deseo invertir las filas y columnas de la Matriz Actor-Tema con la finalidad de permitir otra visualización de la Matriz si fuera necesario.	13
HU087	Yo como usuario autorizado, deseo eliminar resultados obtenidos con la finalidad de descartar información irrelevante.	13
HU057	Yo como usuario autorizado, deseo calificar los resultados obtenidos con la finalidad de agregar valor a la información obtenida.	5
HU0128	Yo como usuario autorizado, deseo exportar los resultados obtenidos de la recolección de información con la finalidad de compartir estos datos con la comunidad de Vigilancia Estratégica.	13

B. Planificación de Lanzamiento

Con el fin de permitir al equipo Scrum tener una visión global de las características a implementarse y cuándo éstas se completarán, se planificó, qué Historias de Usuario debían desarrollarse para cada Sprint, tomando en cuenta los puntos de esfuerzo (Tabla II). La duración de cada Sprint fue fijada en 2 semanas.

TABLA II. PLANIFICACIÓN DE LANZAMIENTO

Sprint #	Sprint 0	Sprint 1	Sprint 2	Sprint 3	Sprint 4	Sprint 5
HU (ID)	HU022 HU023 HU024 HU092	HU037 HU040 HU0136 HU0137 HU0138	HU025 HU058 HU063 HU068 HU0114 HU0135 HU0185	HU027 HU064 HU067 HU0117 HU0118 HU0119 HU0120 HU0121	HU0122 HU0123 HU0124 HU0125	HU026 HU057 HU086 HU087 HU0126 HU0128
Esfuerzo total	27	89	70	84	68	70

C. Ejecución de Sprints

La ejecución de los Sprints se realizó con base al plan de entregas definido previamente, la misma que se actualizó y adaptó al final de cada Sprint para cubrir los nuevos requerimientos y cambios en el producto. Como ejemplo, se

muestra la ejecución del Sprint 1, que da inicio al proceso de recolección de información con técnicas de web Scraping.

• *Ejemplo: Ejecución de Sprint 1*

En este Sprint se definió realizar una investigación sobre el acceso a las APIs de redes sociales y también la aplicación de web Scraping en 5 medios de prensa digitales en español más influyentes de Latinoamérica.

• *Planificación del Sprint 1*

Equipo Scrum se reunió para definir la lista del Product Backlog a realizarse en este Sprint, indicando cuáles serán los entregables junto con los criterios de aceptación de las Historias de Usuario.

• *Objetivos del Sprint 1*

- Elegir los 5 medios de prensa digitales más influyentes de Latinoamérica.
- Extraer información de los medios de prensa elegidos.
- Visualizar la información obtenida de la ejecución de los spiders en la web.
- Extraer información de usuarios o páginas públicas de las redes sociales mediante sus APIs o librerías.

• *Revisión del Sprint 1*

El Sprint 1 fue revisado por el Equipo Scrum de acuerdo con los criterios de aceptación de las Historias de Usuario descritos en la Tabla III alcanzando con éxito la meta del Sprint 1 y cumpliendo con cada funcionalidad planteada.

TABLA III. CRITERIOS DE ACEPTACIÓN DE LAS HU DEL SPRINT 1

ID: DESCRIPCIÓN	CRITERIOS DE ACEPTACIÓN	ESTADO
HU037: Yo como usuario autorizado, deseo poder obtener los rankings de los medios 5 medios de prensa digitales más influyentes de Latinoamérica con la finalidad de poder extraer información relevante para el proceso de recopilación de información.	Estos rankings deben basarse en: <ul style="list-style-type: none"> • El número de visitas al medio digital. • Rankings de varios medios de la prensa local. • Estos medios deben ser de distintos países de la región. • Los medios digitales deben ser en idioma español. • Los medios elegidos deben permitir aplicar las técnicas de web Scraping. 	DONE
HU040: Yo como usuario autorizado, deseo extraer información de usuarios o páginas públicas de las redes sociales mediante sus APIs o librerías con la finalidad de obtener información relevante para el proceso de Vigilancia Estratégica.	La información que se requiere obtener de la API de la red social Twitter es: <ul style="list-style-type: none"> • Para la información del usuario: <ul style="list-style-type: none"> ○ Nombre del usuario. ○ Descripción del usuario. ○ Fotografía o avatar del usuario. ○ Nombre de la red social a la que pertenece. • Para la información del tuit del usuario: 	DONE

ID: DESCRIPCIÓN	CRITERIOS DE ACEPTACIÓN	ESTADO
	<ul style="list-style-type: none"> ○ Descripción del tuit del usuario. ○ Fecha de publicación de los Tuits en su formato día/mes/año. ○ Enlace del tuit. ○ Nombre de la red social a la que pertenece. 	
	La información que se requiere obtener de la red social Facebook es: <ul style="list-style-type: none"> • Descripción de la publicación reciente del usuario Facebook. • Fecha de la publicación en su formato día/mes/año. • Enlace de la publicación. • Nombre de la red social a la que pertenece. 	DONE
	La información que se requiere obtener de la red social LinkedIn es: <ul style="list-style-type: none"> • Descripción de la publicación reciente del usuario LinkedIn. • Fecha de la publicación en su formato día/mes/año. • Enlace de la publicación. • Nombre de la red social a la que pertenece. 	DONE
HU0136: Yo como usuario autorizado, deseo extraer información de los medios de prensa elegidos con la finalidad de obtener información útil para el sistema de Vigilancia Estratégica.	La información que se requiere obtener de los medios de prensa es: <ul style="list-style-type: none"> • El título de la noticia. • La descripción de la noticia. • Fecha de la noticia publicada en su formato día/mes/año. • Enlace de la noticia. 	DONE
HU0137: Yo como usuario autorizado, deseo ver la información obtenida de la ejecución de los spiders en la web con la finalidad de que sea accesible en todo momento por el Equipo desarrollador Scrum.	<ul style="list-style-type: none"> • Si el Equipo desarrollador Scrum desea acceder a la data obtenida luego de la ejecución del spider, esta debe ser mostrada en la web y estar disponible en formato JSON. • Si se desea actualizar la data obtenida luego de la ejecución del spider esta acción se debe poder realizar en tiempo real. 	DONE
HU0138: Yo como usuario autorizado, deseo poder ver la información del usuario de la red social en la Web con la finalidad de que esta información sea accesible por el equipo Scrum.	<ul style="list-style-type: none"> • Si el Equipo Scrum desea acceder a la información del usuario esta deberá estar disponible en todo momento. • Si el Equipo de desarrollo Scrum desea obtener la información del usuario, el servidor deberá permitir ingresar el nombre del usuario y obtener dicha información. 	DONE

En el Burndown Chart que corresponde al Sprint 1 se resume el desempeño del Equipo de desarrollo realizado durante cada uno de los días del Sprint (Fig. 4).

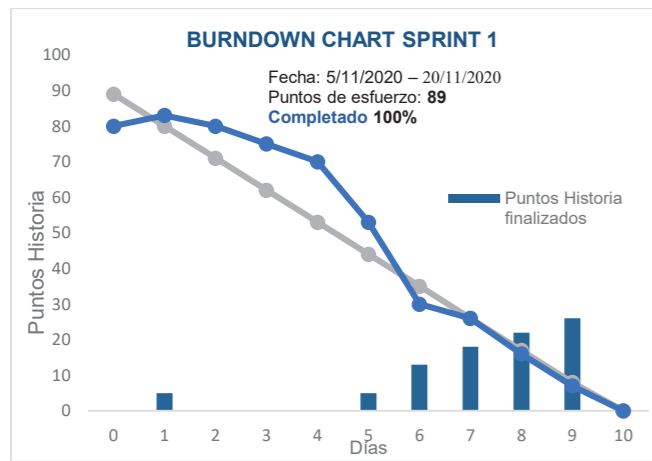


Fig. 4. Bumdown chart del Sprint 1.

En Burndown chart, se indica la fecha de inicio y fecha de finalización del Sprint. El equipo de desarrollo Scrum ha completado el 100% de las Historias de Usuario asignadas. Como se observa, se presentó retrasos en los primeros días del sprint debido a que el equipo necesitó un Spike (tiempo orientado a la investigación o experimentación). Este fue necesario para realizar la investigación sobre el manejo de limitaciones de las APIs de redes sociales, sin embargo, se encontró soluciones óptimas como el replanteamiento de los criterios de aceptación de la HU040 para así lograr el objetivo planteado del Sprint 1.

• *Retrospectiva del Sprint 1*

Con el objetivo de seguir mejorando la productividad del Equipo Scrum en el desarrollo del presente proyecto, se plantea en la Tabla IV un resumen general del Sprint 1, además de las acciones de mejora para aplicar en el siguiente Sprint.

TABLA IV. SPRINT 1 EN RESUMEN

QUÉ SE HIZO BIEN	QUÉ SE HIZO MAL
<ul style="list-style-type: none"> Cumplir los objetivos planteados del Sprint 1. Cumplir con las tareas asignadas en el tiempo esperado. Encontrar alternativas eficientes para afrontar los inconvenientes de limitaciones de las APIs de las redes sociales. Virtudes que demostró el Equipo Scrum al solventar problemas en corto tiempo para lograr el cumplimiento del Sprint 1. Reasignar el criterio de aceptación del manejo de la red social de Instagram a un siguiente sprint. 	<ul style="list-style-type: none"> Falta de comunicación por parte del Equipo Scrum para definir los criterios sobre los cuales se elegirían los medios de prensa digitales más influyentes en Latinoamérica. Mala estimación de la HU040, no se consideró el Spike que necesita el equipo para continuar con el desarrollo del proyecto por lo que provocó una sobrecarga de trabajo para el equipo.
IDEAS	ACCIONES
<ul style="list-style-type: none"> Investigar posibles alternativas para trabajar con la red social de Instagram. 	<ul style="list-style-type: none"> Buscar información más detallada y actualizada sobre la API de Instagram.

- Definir los posibles Spike que el equipo necesita para continuar con el desarrollo del proyecto.
- Mantener chat activo con el intercambio de criterios y reporte de avances.
- Actualizar los puntos de esfuerzo en las siguientes Historias de Usuario.
- Actualizar el Product Backlog.

V. DISCUSIÓN

El sistema fue sometido a pruebas funcionales y de aceptación (TAM). Para validar la funcionalidad de la aplicación se establecieron casos de prueba que permitan demostrar si se cumplen o no los requerimientos de desarrollo predefinidos. Por otra parte, para las pruebas TAM, se realizaron pruebas al sistema con la colaboración de cinco grupos de seis participantes cada uno, los participantes fueron estudiantes que tenían conocimientos relacionados al proceso de SScan los cuales, al finalizar, respondieron una encuesta que utilizó preguntas relacionadas con los parámetros en los que se basa TAM.

Con base en los resultados obtenidos se concluye que el sistema es intuitivo. Sin embargo, se puede recomendar que los usuarios que manejen el sistema deben tener un conocimiento previo de definiciones clave de Vigilancia Estratégica. Además, al momento del despliegue del sistema, los usuarios finales deben cumplir con un proceso de *training* para el uso óptimo de las funcionalidades que éste ofrece, dado que los resultados de evaluación reportan que el sistema cuenta con una aprobación del 83% en cuanto a la facilidad de uso.

Para la recolección de información a partir de las distintas fuentes disponibles del sistema, es importante mencionar que, debido a las diferentes técnicas de extracción de información, es un reto solventar todos los desafíos que se presentan al trabajar con Web Scraping. Por ejemplo, uno de ellos es el tiempo excesivo de carga de los sitios web de donde se desea extraer información, inclusive existe la posibilidad de que no se cargue las páginas debido a que reciben demasiadas solicitudes de acceso. Por un lado, esto no es un problema cuando personalmente navegamos por el sitio ya que solo se necesita volver a cargar la página web y esperar que el sitio responda; pero, en contraste, un scraper no conoce cómo lidiar con tal inconveniente por lo cual dentro del sistema se controla dichas peticiones generando un tiempo considerable de espera para los usuarios que trabajen al mismo tiempo con el sistema.

VI. CONCLUSIONES

La aplicación de la metodología Scrum en el presente proyecto permitió el rápido avance en su realización, ya que al tener una comunicación diaria con todo el Equipo Scrum se pudo solventar de manera eficiente cualquier inconveniente que se presentara, así también obtener retroalimentaciones al final de cada Sprint con el fin de asegurar el cumplimiento de cada funcionalidad requerida por el Product Owner.

El trabajo con las APIs de las redes sociales fue un reto en el presente proyecto, ya que, en medida del crecimiento de la mala práctica del uso de la data, las APIs de redes sociales actualmente han impuesto restricciones para los desarrolladores y analistas de datos. Por este motivo, para el cumplimiento de las metas del proyecto, se trabajó con librerías actualizadas al tiempo de desarrollo de este sistema,

año 2021, respetando las políticas de seguridad de cada red social.

El sistema web ayuda a disminuir tiempo y esfuerzo para encontrar información precisa y útil para compartir con la comunidad de Vigilancia Estratégica, ya que aprovecha la ventaja de trabajar con Web Scraping al permitir recolectar información y contenido de forma rápida y actual. Este antecedente es valorado en gran medida considerando la sobrecarga de información a la que están sometidas las organizaciones.

Los resultados de las pruebas muestran que de manera general, el trabajo desarrollado logra cumplir con los objetivos propuestos como lo es el proceso de recolección de información de las redes sociales y medio digitales propuestos.

Para el sistema es importante el principio de usabilidad con el fin de que los usuarios finales puedan emplearlo de manera fácil, intuitiva y sin complicaciones que afecten la recolección de información que alimenta al proceso de Vigilancia Estratégica. Entonces, de acuerdo con los resultados recopilados después de la evaluación, el sistema presenta una aceptación del 83% de parte de los participantes respecto a su facilidad de uso.

REFERENCIAS

- F. J. Aguilar. Scanning the business environment. New York, USA: Macmillan, 1967.
- H. Lesca, and N. Lesca. Weak signals for strategic intelligence: anticipation tool for managers, London, UK: ISTE-Wiley, 2011.
- E. Loza-Aguirre, and A. Buitrago, "Qualitative assessment of user acceptance within Action Design Research and Action Research: two case studies," Latin American Journal of Computing, vol. 1, no. 1, pp. 12-12, 2014.
- A. Carbonell Martínez, "Vigilancia tecnológica e inteligencia competitiva al servicio de la innovación," Universitat Politècnica de València, España, 2019.
- CMI Gestión, "Vigilancia Estratégica como ventaja competitiva," [Online]. Available: <https://cmigestion.es/2017/03/07/vigilancia-estrategica-como-ventaja-competitiva/>. [Accessed: Mar.16, 2021].
- CIDEI, "Vigilancia Estratégica," [Online]. Available: <http://cidei.net/vigilancia-estrategica/> [Accessed: Mar.07, 2021].
- A. Carbonell Martínez, "Vigilancia tecnológica e inteligencia competitiva al servicio de la innovación. 3C Tecnología,," Glosas de innovación aplicadas a la pyme, vol. 8(4), pp. 61-69, 2019.
- Antevenio, "Qué es el web scraping, para qué sirve y como suele utilizarse," . (2021). [Online]. Available: <https://www.antevenio.com/blog/2019/03/que-es-el-web-scraping-y-para-que-sirve/> [Accessed: Mar.20, 2021].
- IONOS Digitalguide, "¿Qué es el web scraping?," [Online]. Available: <https://www.ionos.es/digitalguide/paginas-web/desarrollo-web/que-es-el-web-scraping/> [Accessed: Jul.26, 2021].
- MINTIC, "¿Qué es web scraping? Curso de Python," [Online]. Available: <https://mintic.platzi.com/clases/1698-mintic-python/7118-que-es-web-scraping-6/> [Accessed: Abr.04, 2021].
- A. Visus, "¿Para qué sirve Python? Razones para utilizar este lenguaje de programación," Esic Business & Marketing School. [Online]. Available: <https://www.esic.edu/rethink/tecnologia/para-que-sirve-python> [Accessed: Abr.10, 2021].
- A. Lafuente, "Qué es el web scraping," Aukera. [Online]. Available: <https://aukera.es/blog/web-scraping/> [Accessed: Abr.18, 2021].
- D. Leyton, "Extensión al Modelo de Aceptación de Tecnología TAM," Repositorio uchile.cl, 2013.
- V. Del Campo, and F. Scholborgh, "Modelo de Aceptación Tecnológica (TAM)," Revista Espacios, vol. 41, no. 37, pp. 159-171, 2020.
- J. Yong, J. Rivas and J. Chaparro, "Modelo de aceptación tecnológica (TAM): un estudio de la influencia de la cultura nacional y del perfil del usuario en el uso de las TIC," Sistema de Información Científica Redalyc, Red de Revistas Científicas, Colombia, 2009.
- J. J. Diaz Ortiz, and M. A. Romero Suarez, "Desarrollo e implementación de un aplicativo web, utilizando la metodología Scrum, para mejorar el proceso de atención al cliente en la empresa Z Aditivos SA," 2017.
- M. Trigás Gallego, "Metodología Scrum," 2012
- G. Mancuzo, "Fases de la Metodología Scrum". ComparaSoftware Blog, 2020.
- Softeng, "Proceso y Roles de Scrum," [Online]. Available: <https://www.softeng.es/es-es/empresa/metodologias-de-trabajo/metodologia-scrum/proceso-roles-de-scrum.html> [Accessed: May.05, 2020].

AUTHORS



Carol Oña

Ingeniera en Sistemas Informáticos y de Computación de la Escuela Politécnica Nacional. Con conocimientos en desarrollo Web; backend (PHP, Javascript, JAVA, Python) y frontend (HTML5, CSS3 y Javascript). Participó en varios proyectos de desarrollo, uno de ellos fue el desarrollo de la primera fase del Sistema Integrado de Centros de Formación Artesanal (SICFA) para la Junta Nacional de Defensa del Artesano. Actualmente se encuentra en el campo de diseño web enfocada en la planificación, diseño, implementación y mantenimiento de sitios web, con el objetivo de crecer, aprender y mejorar cada día.



Nicole Pirca

Ingeniera en Sistemas Informáticos y de Computación, graduada en la Escuela Politécnica Nacional apasionada en el campo de ciencia de datos (Data Science). Recibió diploma de Bachillerato de Servicios Aplicaciones Informáticas en 2015. Actualmente trabaja en análisis de datos y procesamiento de información aplicado en el área comercial y participa activamente en la creación de sistemas de software con la tecnología de Blockchain y Big Data.



Lorena Recalde

Lorena Recalde obtuvo su Ph.D. en la Universidad Pompeu Fabra, Barcelona, España, 2018, donde colaboró en el grupo de investigación Web Science and Social Computing. Realizó su maestría en Suiza, en la Universidad de Fribourg y obtuvo el Master en Ciencias de Computación con mención en Procesamiento Avanzado de la Información, 2014. Sus principales temas de investigación son los Sistemas de Recomendación, con un enfoque especial en el modelado de interés del usuario de redes sociales, minería Web, ML y NLP. Además de las publicaciones en conferencias y en dos revistas Q1, Lorena ha presentado su trabajo en tres sesiones de posters, un simposio de doctorado, dos tutoriales académicos y fue invitada como keynote en GEITEC2019, Lima, Perú. Actualmente es investigadora y docente a tiempo completo de pregrado y maestría en el Departamento de Informática y Ciencias de la Computación de la Escuela Politécnica Nacional, Quito, Ecuador.



Edison Loza-Aguirre

Edison Loza-Aguirre trabaja como profesor agregado de la Escuela Politécnica Nacional del Ecuador (EPN) y es un investigador asociado al Laboratorio CERAG CNRS FRE 3748 de la Universidad Grenoble Alpes (UGA). Actualmente es el Coordinador del Laboratorio en Analítica de Datos y Sistemas de Información de la EPN. Obtuvo su título de PhD en Gestión de Sistemas de Información en la UGA. Tiene un Máster en Investigación de la Universidad Pierre-Mendès-France y una Maestría en Gestión de TICs de la EPN. Realizó sus investigaciones posdoctorales en el Instituto de Administración de Empresas de Grenoble, en Francia. Sus investigaciones se centran en vigilancia estratégica, analítica de datos y sistemas de soporte para toma de decisiones. Ha presentado sus trabajos en diversas conferencias y ha publicado artículos en revistas como Communications of the Association Information Systems, IET Software y el French Journal of Management Information Systems. A nivel profesional, trabaja como consultor en Sistemas de Soporte para la Toma de Decisiones, Analítica de Datos, Estrategia de Datos, y Gobernanza de Datos.