

SISTEMA RECOLECTOR DE NOTICIAS NACIONALES

1st David Santiago Velasco Daza
Universidad Pontificia
Bolivariana Bucaramanga,
Colombia
david.velasco.2020@upb.edu.co

2st David Leonardo Sepulveda
ChávezUniversidad Pontificia
Bolivariana Bucaramanga,
Colombia
david.sepulveda.2020@upb.edu.co

3st Gleisson Stik Perez Rivera
Universidad Pontificia Bolivariana
Bucaramanga,
Colombia
gleisson.perez.2020@upb.edu.co

4st Javier Andres Guevara Macias
Universidad Pontificia
Bolivariana Bucaramanga,
Colombia
javier.guevara.2020@upb.edu.co

Resumen - En el presente documento, se evidencia el trabajo realizado por la empresa INT_NEWS por ende se evidencian distintitos elementos de gran importancia sobre el proyecto propuesto, además del desarrollo de objetivos para hallar una solución a la problemática planteada, el manejo y velocidad de un grupo el cual adopto el marco de trabajo Scrum de la mano con la metodología Espiral, con el fin de entregar resultados y trabajar en mejoras con ayuda del cliente.

Abstract - In this document, the work carried out by the company INT_NEWS is finally evidenced, different elements of great importance are evidenced about the proposed project, in addition to the development of objectives to find a solution to the problem raised, the handling and speed of a group which adopts the Scrum framework hand in hand with the Spiral methodology, in order to deliver results and work on improvements with the help of the client.

I. INTRODUCCIÓN

En la actualidad debido a la mala redacción de algunos medios de comunicación, cierta parte de la comunidad tiende a estar desinformada por ende la empresa INT_NEWS opto por desarrollar un aplicativo web que permita realizar una recolección diaria de titulares

informativos de cuatro medios de comunicación confiables en Colombia (El Tiempo, Pulzo, El Herald, El Colombiano). Este sistema contara con algunos requisitos establecido por la parte interesada, tales como: contar con un entorno tecnológico común, soportar un algoritmo de web scraping (Python) encargado de hacer un barrido diario con el fin de recolectar los principales titulares de los medios anteriormente mencionado y así almacenarlos en una base de datos no relacional (mongoDB), para finalmente ser mostrada al cliente mediante un prototipo web funcional (HTML, CSS, JS).

Teniendo en cuenta lo planteado anteriormente, el presente informe se expone el avance del trabajo realizado por el grupo Empresa INT_NEWS. Se presenta en la sección II el estado del arte, seguido de la sección III marco teórico y tecnológico del proyecto, continuando con la definición de los objetivos en la sección IV, así mismo se propone en la sección V una metodología acorde con el marco de trabajo Scrum, además de la sección VI en donde se abordan los resultados y se detalla el trabajo realizado, teniendo en cuenta la sección VII el cual presenta el cronograma propuesto previamente en donde se identificó las tareas y tiempos a cumplir. Se realiza el análisis de retrospectiva del proyecto en la sección VIII de acuerdo con el trabajo y a los resultados obtenidos, creando un espacio de reflexión con lecciones aprendidas en las fases de desarrollo del

proyecto, así mismo se manifiestan las conclusiones y el trabajo a futuro en la sección IX, dando paso a las secciones finales X y XI en donde se encuentra el trabajo futuro y los agradecimientos respectivamente.

II. ESTADO DEL ARTE

En base al desarrollo que se planteó para este proyecto se realizó un análisis en la etapa inicial el cual se llegaron a identificar los 4 factores más importantes a realizar los cuales son: el más relevante en los requisitos es el sistema web scraping, el aplicativo en el que estará soportado (frontend), El backend que comprende la base de datos de usuarios y del aplicativo y su respectiva integración con el servicio en nube. Respecto al análisis hecho se lograron identificar algunos casos de aplicaciones similares a la que se tiene planteada realizar, las cuales se tomaran como guías para este desarrollo.

En general de las fuentes y aplicaciones visitadas que tienen una vista similar o un objetivo muy allegado al planteamiento, el equipo se percató de la falta simplicidad y mayor atención al usuario, es decir, mientras que unas páginas eran unas

extensiones de las cuales todo tenía que ser manual, otras eran paginas estáticas las cuales no contaban con la personalización del usuario y estudio de este mediante IA. [1] De la información obtenida a partir de lo consultado se utilizarán los aspectos relacionados con las aplicaciones halladas en el análisis previo, para así obtener resultados concretos y que sean coherentes con las necesidades expuestas por el cliente.

III. MARCO CONCEPTUAL

Debido a que se hizo un análisis de páginas que ofrecieran un servicio similar al que se

planteó desarrollar, se encontraron diseños complejos para la comprensión del usuario, a partir de esto se desarrolló un prototipo claro y familiar para el usuario con la ayuda de la herramienta FIGMA, este prototipo fue plasmado y desarrollado en las tecnologías HTML para el maquetado, estructura de la página web, luego a esta estructura se realizó su diseño con SCSS posteriormente compilado a CSS, y para su parte lógica se implementó JS, para dar funcionalidad a la página. En la parte de seguridad para la web se implementó el protocolo HTTPS.

La tecnología encargada de realizar el barrido por los medios mencionado y posteriormente realizar la recolección de datos (noticias), se realizó un algoritmo en Python, donde en el mismo algoritmo se encarga de enviar y almacenar la información recolectada a la Base de Datos MONGODB, el cual es un sistema gestor de BD no relacional (requisito por la parte interesada).

La estructura anteriormente descrita y desarrollada fue almacenada en la nube, esto de la mano con el servicio de computación en la nube AZURE, donde se encuentra el servidor WEB implementado en Windows Server donde se administran los recursos de la página WEB y así ser accesible para el usuario. Así mismo en AZURE se instalará el servidor FTPS para la transferencia de archivos contenedores de actualizaciones y cambios que se irán realizando a la web, este servidor estará en la misma maquina donde reposa el servidor WEB.

IV. OBJETIVOS.

Desarrollar un sistema de información que cumpla con los requisitos de Infraestructura, seguridad, integración y servicios de red, mediante el uso de tecnologías web siguiendo las especificaciones solicitadas con el fin de poder exponer noticias del día a día y así facilitar funciones a los usuarios.

Objetivos específicos.

- Analizar los requerimientos funcionales y no funcionales para determinar el alcance del proyecto.
- Diseñar el sistema e interfaz de usuario que conforma el sistema de información.
- Diseñar e implementar la base de datos que usará el sistema para el almacenamiento de la información.
- Diseñar la arquitectura del sistema el cual servirá de base durante el desarrollo y despliegue de los servicios solicitados por la parte interesada.
- Desarrollar el aplicativo WEB recolector de noticias implementando el algoritmo.
- Diseñar e implementar la infraestructura tecnológica para los requerimientos de la empresa.
- Validar la herramienta web, servicios y servidores mediante pruebas funcionales y no funcionales.

V. METODOLOGIA.

La metodología que se llevó a cabo para desarrollar el proyecto fue la espiral, la cual es el que modela el proceso del software de manera evolutiva, el cual de manera iterativa se basa en creación de prototipos, pero aun así conserva propiedades que son implementadas en la cascada. Con esto se quiere decir que es producir un producto que tenga una mejora de manera continua.

La fase que se encuentra en primer lugar es la planificación, en donde de acuerdo con lo

pactado con el cliente, se traza un plan de trabajo, la comunicación con el cliente es el principal objetivo en esta fase, en la cual el equipo se encarga de entender y despejar cualquier duda con los interesados.

En la segunda fase, se encuentra el análisis de requerimientos, esto con el fin de priorizar y organizar las tareas, definir la solución clara y los pasos para lograr el cumplimiento de las

probar su correcto funcionamiento, por medio de pruebas funcionales y no funcionales. En la cuarta fase se encuentra el despliegue, en donde lo desarrollado anteriormente pasa a un plano de producción. En la quinta fase llamada incremento, el equipo tendría como terminada la tarea con su respectiva documentación la cual será valorada en la fase final de comunicación con el cliente, en donde aceptara o brindara información al grupo para mejorar o direccionar el trabajo del equipo.



Imagen I realizada por el grupo de trabajo tomando como la Imagen de: H. S. Corvo, «Modelo espiral: historia, características, etapas, ejemplo.» [En línea]. Available: <https://www.lifeder.com/modelo-espiral/>.

Esta metodología al permitirnos iterar sobre estas fases a lo largo del proyecto nos permite trabajar de la mano con el marco de trabajo SCRUM, al darnos la posibilidad de relacionarnos con el cliente y hacerlo parte del proceso, así mismo entregar incrementos que aportan valor al proyecto.

VI. RESULTADOS.

A continuación, se presentan tablas de los resultados obtenidos a lo largo del proyecto, de igual manera se evidencia gráficamente los resultados.

TABLA I
Resultado de LogIn de usuarios

Definición de tareas	Realizado
Definición de campos	✓
Diseño con ayuda de herramienta grafica	✓
Implementación en HTML	✓
Creación de base de datos	✓
Conexión con la base de datos	✓
Diseño e implementación de la base de datos	✓
Validación de campos	✓
Documentación de pruebas	✓
Subir al servidor	✓

Figura 1 – LogIn al sistema.

TABLA II
Resultado de registro de usuario.

Definición de tareas	Realizado
Definición de campos	✓
Diseño con ayuda de herramienta grafica	✓
Implementación en HTML	✓

Conexión con la base de datos	✓
Documentación de pruebas	✓
Subir al servidor	✓

Figura 2 – Registro en el sistema.

TABLA II
Resultado de documentación.

Definición de las tareas	Realizado
Definición de los usuarios	✓
Página web	✓
Story Mapping	✓
Base de datos	✓
Diseño del Sprint	✓

INFRAESTRUCTURA	APLICATIVO WEB	SEGURIDAD	DISPONIBILIDAD	BASES DE DATOS	DOCUMENTACIÓN
Instalacion de sistemas operativos	Soporte al proceso de recoleccion de noticias	Control de permisos	Informacion accesible a los usuarios con permisos	Creacion de las bases de datos	Diagrama de Infraestructura y comportamiento.
Servidor de la base de datos	Mantenimiento a los servidores y sistemas operativos	Ley 1581 de 2012: tratamiento de datos	Uso de los servicios en el area de categorias	Conexión de las bases de datos	Modelo relacional de las bases de datos
Desarrollo del sistema			Servicio en la seleccion de noticias destacadas	Registro y actualizacion de las noticias cada dia	Diagrama de arquitectura del sistema
				Mantenimiento de la bases de datos	SLOs, SLIs, SLAs, KPIs del sistema

Figura 2 – Story Mapping.

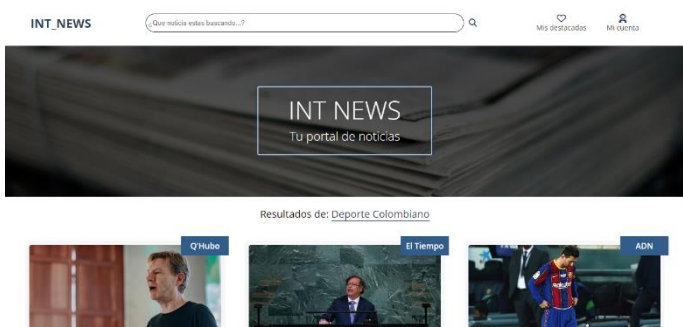


Figura 3 – Index de la página web.

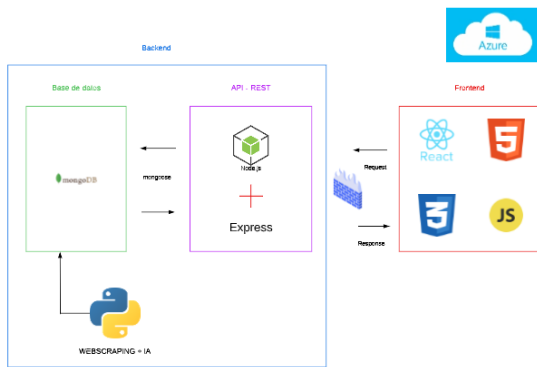


Figura 4 – Arquitectura del proyecto.

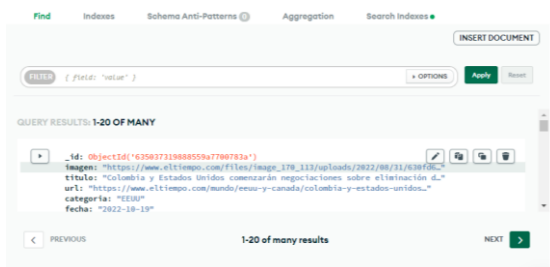


Figura 5 – Base de datos en MongoDB atlas.

```

fecha = str(datetime.date.today())

mongoHost = 'mongodb://localhost'
client = MongoClient(mongoHost)

db = client['integrador2']
collection = db['noticias']

# urlElTiempo = 'https://www.eltiempo.com/'
elTiempo = requests.get('https://www.eltiempo.com/')

# urlPulzo = 'https://www.pulzo.com/'
pulzo = requests.get('https://www.pulzo.com/')

# urlElColombiano = 'https://www.elcolombiano.com/'
elColombiano = requests.get('https://www.elcolombiano.com/')

# urlElHeraldo = 'https://www.elheraldo.co/'
elHeraldo = requests.get('https://www.elheraldo.co/')

```

Figura 6 – Fuentes WebScraping.

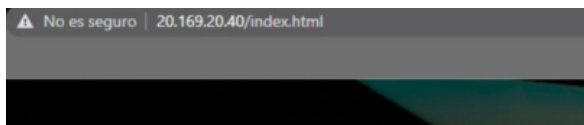


Figura 7 – aplicativo en el servidor web.

VII. AVANCE DEL CONOGRAMA.

El avance del cronograma se evidenciará en las siguientes imágenes, también será anexado el documento donde se plasma más a detalle.



OBJETIVOS ESPECIFICOS	ACTIVIDADES	ID RACIONAL	SEMANAS					
			1	2	3	4	5	6
Objetivo 1: Analizar los requerimientos funcionales y no funcionales para determinar el alcance del proyecto.	Actividad 1.1: Analizar de la situación problema.	0.4642						
	Actividad 1.2: Investigar conceptos, tecnologías y herramientas relacionadas con el proyecto.	0.4642						
	Actividad 1.3: Documentar los requerimientos funcionales y no funcionales del proyecto.	0.4642						
	Actividad 1.4: Documentar la especificación de requerimientos.	0.4642						
	Actividad 1.5: Priorizar los requerimientos y distribuirlos.	0.4642						
Objetivo 2: Diseñar el sistema e interfaz de usuario que confirme el sistema de información.	Actividad 2.1: Aprobación del documento de especificación de requerimientos y de los funcionalidades a desarrollar.	0.4642						
	Actividad 2.2: Diseñar la especificación de la aplicación web.	0.4642						
	Actividad 2.3: Diseñar diagrama UML para el desarrollo del proyecto (UML de casos de uso, flujos de actividades, diagramas, etc).	0.4642						
	Actividad 2.4: Diseñar un prototipo del front mediante el uso de las herramientas adecuadamente investigadas.	0.4642						
	Actividad 2.5: Diseñar un prototipo del proyecto mediante el uso de las herramientas adecuadamente investigadas.	0.4642						
Objetivo 3: Diseñar e implementar la base de datos que sustentará la empresa.	Actividad 3.1: Diseñar el modelo conceptual de la base de datos.	0.4642						
	Actividad 3.2: Implementar la base de datos propuesta.	0.4642						
	Actividad 3.3: Realizar pruebas de funcionamiento y correcciones en la base de datos.	0.4642						
Objetivo 4: Diseñar la arquitectura del sistema el cual servirá de base durante el desarrollo y despliegue de los servicios solicitados.	Actividad 4.1: Indagar sobre las diferentes arquitecturas que puedan llegar a utilizarse en el proyecto.	0.4642						
	Actividad 4.2: Definir la arquitectura que más se adapte a las necesidades de la empresa.	0.4642						
	Actividad 4.3: Generar el diseño de la posible implementación en el proyecto.	0.4642						
Objetivo 5: Desarrollar el aplicativo WEB recolector de noticias implementando el algoritmo.	Actividad 5.1: Analizar sobre los parámetros, objetivos de la empresa.	0.4642						
	Actividad 5.2: Planear, analizar y desarrollo del algoritmo de web scraping.	0.4642						
	Actividad 5.3: Implementar la configuración del repositorio adecuado para garantizar la recuperación de datos de noticias.	0.4642						
Objetivo 6: Diseñar e implementar la infraestructura tecnológica para los requerimientos de la empresa.	Actividad 6.1: Diseñar el prototipo de red y a su vez identificar los riesgos y documentar los modos para el manejo de los riesgos.	0.4642						
	Actividad 6.2: Implementar los servidores WEB y BD en la nube.	0.4642						
	Actividad 6.3: Implementar el sistema de control y validación, administrar el acceso a internet para los otros puntos del sistema.	0.4642						
Objetivo 7: Validar la herramienta web, servicios y servidores mediante pruebas funcionales y no funcionales.	Actividad 7.1: Desarrollar acciones de prueba del software con el usuario final.	0.4642						
	Actividad 7.2: Apoyar al software como resultado de la validación.	0.4642						
	Actividad 7.3: Validar los módulos instalados.	0.4642						
Objetivo 8: Analizar y socializar los resultados obtenidos en las etapas del proyecto.	Actividad 8.1: Desarrollo de acta de entrega del proyecto con ayuda de herramientas como: software de texto, hojas de cálculo, diagramas, base de datos, tablas, entre otros.	1.5						
	Actividad 8.2: Presentación de los resultados.	1.5						

Figura 4 – avance del cronograma.

VIII. RETROESPECTIVA DEL PROYECTO.

El grupo tuvo un error de planeación y definición de historias de usuario. A pesar de cumplir con varias tareas, este error no permitió mostrar la culminación de estas las cuales son fundamentales para el funcionamiento del aplicativo, donde se cumplieron los requisitos mínimos establecidos por el cliente.

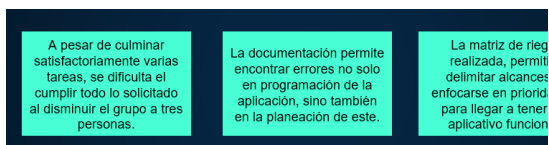
SPRINT I



SPRINT II



SPRINT III



CONCLUSIONES.

- Se debe mejorar la comunicación entre las partes interesadas como también dentro del equipo de desarrollo para obtener mejores resultados en un futuro.
- Se dificultó el manejo del tiempo ya que se tienen otras responsabilidades académicas.
- El uso del marco teórico scrum y metodología espiral permitió que durante el proyecto se entregaran resultados en periodos cortos así mismo trabajar de la mano con el cliente, esto debido a la alta interacción con el cliente y el cronograma establecido para los sprint.
- El uso de nuevas tecnologías no solo implicó nuevo conocimiento, también crearon retos y permitieron que el grupo se sintiera en un ambiente laboral real, esto por la responsabilidad en cada sprint de tener un resultado.

LECCIONES APRENDIDAS.

- Es de gran ayuda comenzar con el desarrollo del proyecto así sean aportes mínimos para tener un buen rendimiento.
- Planificar y definir de manera acertada las historias de usuario.
- Debido a la ausencia de un compañero en el equipo de trabajo, se aumentó la deuda técnica y el trabajo por hacer, esto debido al diseño mismo del proyecto, el cual se planteó para un grupo mayor de personas.

IX. TRABAJO A FUTURO

Con proyecto oficialmente finalizado junto con muchas experiencias y lecciones adquiridas, nos encaminamos a mejorar grupalmente e individual, para dar paso a nuestros siguientes proyectos que se nos presentarán educativa y laboralmente en un futuro no muy lejano, se debe trabajar en aprendizaje propio de temáticas alusivas a la carrera y en gestión de proyectos que desencadena a su vez en trabajo en grupo.

Después de los resultados obtenidos, se espera tomar todo lo aprendido en este proyecto tanto del marco de trabajo Scrum, como en materia de programación para perfeccionar trabajo futuro, mejorar el trabajo en equipo buscando estrategias para la comunicación en el entorno de trabajo.

X. AGRADECIMIENTOS.

Damos gracias a nuestros profesores y todas aquellas personas que forman parte del grupo académico de la facultad de Ingeniería de Sistemas e Informática de la Universidad Pontificia Bolivariana seccional Bucaramanga, quienes contribuyeron con sus conocimientos y demás virtudes que fueron necesarias para el desarrollo total del Proyecto y para ser unos excelentes profesionales y seres humanos de calidad.

XI. REFERENCIAS.

[1] developers.google.com, «HTTPS,» [En línea]. Available: <https://developers.google.com/search/docs/advanced/security/https?hl=es>.

[2] javatpoint.com, «What is Windows?,» [En línea]. Available: <https://www.javatpoint.com/windows>.

[3] scrum.org, «WHAT IS SCRUM?,» [En línea]. Available: <https://www.scrum.org/resources/what-is-scrum>.

[4] kanbanize.com, «¿Qué es un tablero Kanban?,» [En línea]. Available: <https://kanbanize.com/es/recursos-de-kanban/primeros-pasos/que-es-tablero-kanban>.

[5] digital.gov, «An Introduction to GitHub,» [En línea]. Available: <https://digital.gov/resources/an-introduction-github/>.

[6] www.java.com, «What is Java technology and why do I need it?,» [En línea]. Available: https://www.java.com/en/download/help/whatis_java.html.

[7] A. Sharma. "Express JS Tutorial: What is Express in Node JS?" Simplilearn.com. <https://www.simplilearn.com/tutorials/nodejs-tutorial/what-is-express-js> (accedido el 17 de octubre de 2022).

[8] B. Botelho y J. Vaughan. "What is MongoDB? A definition from WhatIs.com". SearchDataManagement. <https://www.techtarget.com/searchdatamanagement/definition/MongoDB> (accedido el 17 de octubre de 2022).

[9] M. Perez. "What is Web Scraping and What is it Used For? | ParseHub". Web Scraping Blog (Tips, Guides + Tutorials) | ParseHub. <https://www.parsehub.com/blog/what-is-web-scraping/> (accedido el 17 de octubre de 2022).

[10] S. J. Bigelow. "What is Microsoft Azure and How Does It Work?" SearchCloudComputing. <https://www.techtarget.com/searchcloudcomputing/definition/Windows-Azure> (accedido el 17 de octubre de 2022).