



**DESARROLLO DE UN PROTOTIPO DE APLICACIÓN WEB QUE PERMITA LA
EXTRACCIÓN DE LAS OFERTAS LABORALES DE LAS PRINCIPALES
PLATAFORMAS QUE POSTULAN EMPLEOS EN LA REGIÓN CARIBE,
USANDO LA TÉCNICA WEB SCRAPING.**

PRESENTADO POR:

VÍCTOR MAURICIO OROZCO PUELLO

LUIS FERNANDO GÓMEZ ESTRADA

**UNIVERSIDAD DEL SINÚ ELÍAS BECHARÁ ZAINÚM SECCIONAL
CARTAGENA**

FACULTAD DE CIENCIAS EXACTAS E INGENIERÍAS

ESCUELA DE INGENIERÍA DE SISTEMAS

CARTAGENA-COLOMBIA

Junio 2019



**DESARROLLO DE UN PROTOTIPO DE APLICACIÓN WEB QUE PERMITA LA
EXTRACCIÓN DE LAS OFERTAS LABORALES DE LAS PRINCIPALES
PLATAFORMAS QUE POSTULAN EMPLEOS EN LA REGIÓN CARIBE,
USANDO LA TÉCNICA WEB SCRAPING.**

**Trabajo de grado presentado como requisito para optar el título de
INGENIERO DE SISTEMAS**

Asesor disciplinar

AISNER MARRUGO JULIAO

Asesor metodológico

EUGENIA ARRIETA RODRÍGUEZ

**UNIVERSIDAD DEL SINÚ ELÍAS BECHARÁ ZAINÚM SECCIONAL
CARTAGENA**

FACULTAD DE CIENCIAS EXACTAS E INGENIERÍAS

ESCUELA DE INGENIERÍA DE SISTEMAS

CARTAGENA-COLOMBIA

Junio 2019

ACTA DE CALIFICACIÓN Y APROBACIÓN

Nota de aceptación:


Director de Escuela

Director de Investigaciones

Firma del jurado

Firma del jurado

Cartagena de Indias, Junio de 2019

 <p>UNIVERSIDAD DEL SINÚ Elías Bechara Zainúm Seccional Cartagena</p>	<p>PROCESO: INVESTIGACIÓN TÍTULO: ACTA PROYECTO DE INVESTIGACIÓN CÓDIGO: R-INVE-021</p>
--	---

**EL DIRECTOR DE INVESTIGACIONES DE LA UNIVERSIDAD DEL SINU
“ELIAS BECHARA ZAINUM” SECCIONAL CARTAGENA**

HACE CONSTAR QUE:

En Cartagena, a los 13 días del mes de Junio del 2019, en la Oficina de la Dirección de Investigaciones de la Universidad, se aprobó por el jurado y se realizó la sustentación del Trabajo de Grado titulado **“DESARROLLO DE UN PROTOTIPO DE APLICACIÓN WEB QUE PERMITA LA EXTRACCIÓN DE LAS OFERTAS LABORALES DE LAS PRINCIPALES PLATAFORMAS QUE POSTULAN EMPLEOS EN LA REGIÓN CARIBE, USANDO LA TÉCNICA WEB SCRAPING “** que se desarrolló bajo la dirección del Ingeniero(a) **EUGENIA ARRIETA RODRÍGUEZ y AISNER MARRUGO JULIAO** , presentado por los egresados **VÍCTOR MAURICIO OROZCO PUELLO y LUIS FERNANDO GÓMEZ ESTRADA**

Los jurados designados fueron los ingenieros **RAFAEL MONTEROZA Y MARIA FERNANDA MEDINA**

Teniendo en cuenta la aprobación emitida, se encuentra que los egresados han cumplido con los requisitos de presentación y sustentación del trabajo de investigación, exigidos por el programa de **INGENIERÍA DE SISTEMAS**, Resolución 0178 de 15 de marzo de 2010.

Se expide esta constancia a los 13 del mes junio de 2019

DIRECCIÓN DE INVESTIGACIONES
Universidad del Sinú

COORDINADOR DE INVESTIGACIONES
Escuela de Ingeniería de Sistemas

TABLA DE CONTENIDO

Introducción	3
1. Diseño Metodológico	5
1.1. Planteamiento Del Problema	5
1.1.1. Descripción Del Problema.....	5
1.1.2. Justificación.....	6
1.1.3. Formulación Del Problema	7
1.1.4. Alcance.....	7
1.1.5. Objetivos	10
1.2. Estado Del Arte	11
1.3. Marcos De Referencia	13
1.3.1. Marco Teórico	13
1.3.2. Marco Conceptual.....	17
1.3.3. Marco Legal.....	20
1.4. Diseño Metodología	23
2. Análisis Del Sistema	30
2.1. Levantamiento De Requerimientos	30
Técnicas De Levantamiento De Requerimientos	30
2.2. Requisitos Funcionales.....	33
2.3. Requerisitos No Funcionales.....	39
3. Diseño Del Sistema	42
3.1. Diagrama De Flujo Del Prototipo	42
3.2. Diagrama De Casos De Uso	43
3.3. Modelo De Datos Lógico	44
3.5. Diseño De Interfaces De Usuario	46
Gestionar Empleos	46
Redirección Y Aplicar Empleos.....	47
4. Desarrollo	47
4.1. Desarrollo De La Plataforma Web	47
Interfaz Inicio De Sesión Administrador	47
Interfaz Seleccionar Script De Los Portal De Empleo.....	48

Interfaz Editar Etiquetas Script	48
Interfaz Actualizar Etiquetas Script.....	49
5. Pruebas Del Sistema	49
5.1. Pruebas De Interfaz Gráfica Para Requerimientos Web.....	49
5.2. Resultados De Las Pruebas De Interfaz.....	52
5.3. Pruebas De Funcionalidad Para Los Requerimientos.	54
5.4. Resultados De Las Pruebas De Funcionalidad.	56
Conclusiones.....	58
Recomendaciones	59
Bibliografía	60
Listas De Anexos	62
Anexo A. Pruebas Plataforma Web.....	62
Pruebas De Interfaz Gráfica Para Requerimientos Web.....	62
Pruebas De Funcionalidad Para Los Requerimientos	68
Anexo B. Diccionario De Datos.....	72

LISTA DE TABLAS

Tabla 1. Portales de Empleos.....	7
Tabla 2. Campos de búsqueda.....	8
Tabla 3. Marco Referencia BeautifulSoup Vs Scrapy.....	29
Tabla 4. Plantilla de Requerimientos Funcionales, Fuente Autor.....	31
Tabla 5. Plantilla de Requerimientos no Funcionales, Fuente Autor	32
Tabla 6. Requerimiento RF-01 Inicio Sesión.	33
Tabla 7. Requerimiento RF- 02. Selección portal Script.	34
Tabla 8. Requerimiento RF- 03. Editar Etiquetas Script.....	35
Tabla 9. Requerimiento RF- 04. Actualizar Etiquetas Script.	36
Tabla 10. Requerimiento RF- 05.Buscar por vacante, empresa y portal.	37
Tabla 11. Requerimiento RF- 06. Redirección de la Vacante hacia el portal externo.	38
Tabla 12. Requerimiento No Funcional – 01. Interfaz de la plataforma.	39
Tabla 13. Requerimiento No Funcional – 02. Seguridad en información.....	39
Tabla 14. Requerimiento No Funcional – 03. Eficiencia	40
Tabla 15. Requerimiento No Funcional – 04. Mantenimiento.	40
Tabla 16. Requerimiento No Funcional – 05. Usabilidad.	41
Tabla 17. Lista de chequeo para pruebas de interfaz, Fuente Autor	50
Tabla 18. Lista de chequeo para pruebas de Funcionalidad, Fuente Autor	54
Tabla 19. Prueba Interfaz inicio de sesión Administrador.....	62
Tabla 20. Prueba Interfaz Seleccionar Portal Empleo.	63
Tabla 21. Prueba Interfaz Actualizar Etiquetas Script.	65
Tabla 22. Prueba Interfaz Gestionar Empleos.	66
Tabla 23. Prueba de funcionalidad inicio sesión.	68
Tabla 24. Prueba de funcionalidad Interfaz Seleccionar Empleo.	69
Tabla 25. Prueba de funcionalidad Interfaz Actualizar Etiquetas Script.	70
Tabla 26. Prueba de funcionalidad Interfaz Gestionar Empleos.	71
Tabla 27. Cronograma	76

LISTA DE FIGURAS

Figura 1. Modelo de Prototipos, Fuente Autor Tomada de [20]	25
Figura 2. Diagrama de Flujo BPMM.	42
Figura 3. Diagrama casos de uso para Requerimiento RF-01, RF-02, RF-03, RF-04, RF-05, RF-06. ...	43
Figura 4. Diagrama casos de uso para Requerimiento RF-05, RF-06.....	44
Figura 5. Modelo Relacional Campos Tabla Vacantes.	45
Figura 6. Diagrama Entidad-Relación Tipo de Datos Tabla Vacantes.....	45
Figura 7. Diseño Gestionar Empleos.....	46
Figura 8. Redirección y Aplicar Empleos.....	47
Figura 9. Interfaz Inicio de Sesión Administrador.....	47
Figura 10. Interfaz Seleccionar Script de los Portales de Empleo.....	48
Figura 11. Interfaz Consultar Etiquetas Script.	48
Figura 12. Interfaz Actualizar Etiquetas Script.....	49
Figura 13. Gestionar Empleos.....	50
Figura 14. Resultados Prueba Interfaz RF- 01 Inicio de Sesion.....	52
Figura 15. Resultados Prueba Interfaz RF- 02 Selección Portal Empleo.....	52
Figura 16. Resultados Prueba Interfaz RF- 03 Actualizar Etiquetas Script.....	53
Figura 17. Resultados Prueba Interfaz RF- 04 Buscar por Vacante - Empresa y Portal.	53
Figura 18. Resultados Prueba Interfaz RF- 05 Aplicar Vacantes o Empleos	54
Figura 19. Resultados Prueba Funcionalidad RF- 01 Inicio de Sesión.....	56
Figura 20. Resultados Prueba Funcionalidad RF- 02 Seleccionar Portal Empleo.	56
Figura 21. Resultados Prueba Funcionalidad RF- 03 Editar Etiquetas Script.....	57
Figura 22. Resultados Prueba Funcionalidad RF- 04 Buscar por Vacante, Empresa y Portal.	57

INTRODUCCIÓN

Desde años anteriores la sociedad no contaba con internet lo que implicaba el uso de otros métodos de comunicación, tales como, el envío de cartas, fax y registro de hojas de vida en formatos, para poder comunicarse y gestionar la búsqueda de empleos. Para alcanzar el conocimiento se contaba con bibliotecas, museos y universidades, imponiendo y facilitando el aprendizaje en esta medida, teniendo en cuenta el avance de la ciencia y la cultura lo cual ha evolucionado mejorando la velocidad y la capacidad.

La llegada del internet y la implementación de muchos sitios web ha impactado de una manera única lo que la hace imprescindible en la vida de las personas, las empresas y universidades; globalizándolo como la más utilizada en cuestión de tiempo, en efecto, tener este servicio es cumplir con las condiciones de acceso a un computador, tabletas, dispositivo móvil (Celular) y acceso a internet.

En internet se puede acceder a múltiples servicios como: redes sociales, blogs, tiendas online, webs corporativas, foros, sitios de descargas, buscadores, sitios webs de noticias, wikis, sitios de consumo de video, sitios educativos, webs de juegos, etc.

Toda la gran variedad de páginas web albergan grandes cantidad de información, lo cual son expuestas por propietarios de cada cuenta, permitiendo el acceso de los usuarios y visualización de los datos e información que pueden ser relevantes y de ayuda para los usuarios.

Este proyecto tiene como desarrollo un enfoque de investigación y de análisis de patrones que identifique y centralicen la búsqueda de información en las bolsas de empleo, permitiendo la extracción de los datos, su clasificación y asegurarlos.

Como objetivo se identifica una técnica que aplica la Web Scraping para la búsqueda de ofertas de empleo, el cual tiene unas herramientas definida por

lenguajes de programación y aplicaciones que intervienen en el manejo de la información que se extrae y clasifica de una manera más organizada y limpia, para darle el uso adecuado en la búsqueda empleos que necesite el usuario.

El proceso de desarrollo tiene varias etapas definidas a través de estándares, que permiten la estructura requerida para el funcionamiento óptimo de la aplicación Web Scraping, involucrando el uso de lenguajes de programación y técnicas, las cuales permiten definir el diseño del modelo y la implementación general para consolidar los conceptos.

Sin embargo, se debe tener una previa investigación y estudio de los temas abordados para las metodologías a utilizar, conceptos, antecedentes e información que nos permita el desarrollo del proyecto y obtener el resultado esperado.

La información desarrollada en el presente documento presenta datos relevantes, relacionándolos con el punto de vista de nuestro entorno, dándole un contexto de integración, comunicación y captación del porqué y la necesidad en el desarrollo de una aplicación Web Scraping para la búsqueda de ofertas laborales en la región de Bolívar.

1. DISEÑO METODOLÓGICO

1.1. PLANTEAMIENTO DEL PROBLEMA

1.1.1. Descripción del Problema

Web Scraping es una técnica utilizada mediante programas de software y lenguajes de programación para extraer información de sitios web. Normalmente estos programas simulan la navegación de un humano en internet, ya sea utilizando el protocolo HTTP manualmente, o incrustando un navegador en una aplicación. Son herramientas que forman parte de las tecnologías Big Data y Analytics y se caracteriza por su capacidad para extraer información pública.

El scraping tiene mucho potencial en sectores como las fuerzas de seguridad del estado, financiero o los seguros, porque rastrea la red buscando toda la información pública que pueda haber en Internet sobre una persona, empresa, organización, ya sean en foros, organismos públicos, documentos oficiales, BOE, redes sociales, webs, etc.

El uso de esta tecnología va unido a otras técnicas de Big Data y Analytics. El Web Scraping tiene la misión de rastrear la web y obtener la información y los datos. Posteriormente con técnicas de Business Analytics, se convierten esos datos en información de valor. Según Vector ITC Group, este tipo de tecnología es muy útil para temas de investigación, ya sea por parte del gobierno, policía o detectives privados. No obstante, a nivel financiero también permite conocer la solvencia de los clientes, y en el caso del sector asegurador, permite calcular las primas de riesgo [1].

Con el pasar de los años es notorio ver como las redes han evolucionado y en Internet se ha vuelto una de las fuentes de información más importante para entidades y empresas que puedan mostrar sus productos y servicios en línea, puedan interactuar con el usuario sin necesidad de trasladarse a solicitar información personalmente.

El problema identificado en este proyecto se caracteriza principalmente por que en muchas de las páginas de empleos con ofertas de Bolívar están independizadas y postulan vacantes que en otras páginas web de empleos no se encuentran, lo que incurre a que un usuario tenga que estar buscando en diferentes portales web de empleos las vacantes que son de su interés, no se encuentran centralizados en base de datos la información de empleos, lo que hace que un usuario visitante tenga que destinar más tiempo buscando en las diferentes páginas web.

1.1.2. Justificación

La presente investigación se enfoca en desarrollar un prototipo web basado en Web Scraping que nos permita extraer, analizar, almacenar y visualizar información de la web que sirva para la búsqueda de ofertas laborales. La centralización de la información de varios portales de bolsas de empleo digitales permiten al usuario de manera ágil y rápida consultar las ofertas laborales más recientes en el mercado actual, y así permite almacenar información en bases de datos de ciertas tendencias que le sirvan a la toma de decisiones, así mismo la gestión y búsqueda de empleo.

El prototipo de aplicación Web Scraping cuenta con un factor diferencial de las demás técnicas ya que no está solamente aplicada para situaciones sociales, de mercadeo o redes sociales en la que siempre la enfocan, este prototipo permite la búsqueda, extracción, y centralización de los datos en un prototipo web de portales de vacantes; en conclusión lo primordial es aplicarlo como método de consulta y búsqueda de ofertas laborales en portales de empleo desde varias páginas de la web, y así permita tener más ofertas centralizadas en un solo prototipo web.

El otro enfoque con que cuenta este proyecto es para aprendizaje en la Universidad del Sinú a los futuros estudiantes de Ingeniería de sistemas, que puedan poner en práctica esta técnica conociéndola y aplicándola en su entorno

social, económico y cultural en la ciudad de Cartagena, y desarrollando nuevos módulos que se ajusten a las necesidades actuales en el prototipo web.

1.1.3. Formulación del Problema

¿Cómo desarrollar un prototipo para extraer, consultar y centralizar la información en la web de portales de empleos y búsquedas de vacantes en Bolívar?

1.1.4. Alcance

Este proyecto se centra en desarrollar un prototipo web con la técnica de Web Scraping que permita la optimización, análisis, extracción y visualización de empleos publicados en la web de Bolívar.

Con este prototipo se alcanza la extracción y centralización de la mayoría de las ofertas de empleo en Bolívar de algunos sitios web de empleo de mayor relevancia con el fin de hacer una búsqueda con mayores resultados y en una sola parte, para así acaparar la mayoría de ofertas en el departamento de Bolívar.

Teniendo en cuenta los diversos portales web de empleos que registran al departamento Bolívar, este prototipo cuenta con los siguientes portales:

Tabla 1. Portales de Empleos.

Portal De Empleo	URL (Localizador Uniforme de Recursos).
Computrabajo	https://www.computrabajo.com.co/empleos-en-bolivar [2]
Linkedin	https://co.linkedin.com/jobs/search?keywords=&location=Bol%C3%ADvar%2C%20Colombia&geo=105189166&pageNum=0&position=1 [3]

El código Script en python solo extrae la información web de los portales sin manipularla, aproximadamente 1 o 2 página; la información web que se extrae, se envía, para su posterior almacenamiento en un motor de bases de datos “Mysql”

de Xamp, con el nombre de “proyecto“. Los datos almacenados en el motor de bases se redirecciona al prototipo web, el cual centraliza las vacantes de los portales Computrabajo y LinkedIn que se mencionaron anteriormente en la *Tabla 1. Portales de Empleos.*

El prototipo web tiene campos de búsqueda para las vacantes donde se podrá consultar cualquier información con los siguientes datos:

Tabla 2. Campos de búsqueda.

Computrabajo / LinkedIn	Descripción
Vacante	Es el cargo o empleo que postula una empresa.
Empresa	Es una entidad económica – social que solicita servicios de trabajo o vacantes de empleo.
Descripción	Son las características o funciones de un cargo o empleo para un puesto de trabajo.
Municipio	Es una entidad administrativa que agrupa localidades que referencian a una ciudad o pueblo.
Fecha	Corresponde a la fecha en que se postuló la vacante de empleo.
Portal	Es el sitio web que integra u ofrece recursos o servicios de empleo.
Url	Corresponde a la dirección web donde se postuló la vacante de empleo.

El proyecto se divide en tres partes, la primera que es el análisis y extracción de las ofertas laborales de las páginas más relevantes en búsqueda de empleo en el departamento de Bolívar.

La segunda parte es el almacenamiento de toda la información en una base de datos que previamente es utilizada en las consultas de los usuarios que deseen ver la información extraída de las web seleccionadas.

En este último, para la visualización de la información por parte de los usuarios, se construye un prototipo sitio web en el cual las personas puedan ingresar para ver la información de los empleos, además tienen no solamente la información que se haya extraído, sino, que también se podrá hacer un filtro de las vacantes de empleo para que el usuario pueda tener más facilidad y acceso en los empleos para su mayor interés, este prototipo web permite la visualización en PC.

1.1.5. Objetivos

Objetivo general

Desarrollar un prototipo de aplicación web que permita la extracción de las ofertas laborales de las principales plataformas que ofertan empleos en la región caribe, usando la técnica Web Scraping.

Objetivos específicos

- Analizar los requerimientos funcionales y no funcionales con los que debe contar la aplicación web.
- Seleccionar la técnica más apropiada de Web Scraping, mediante un análisis comparativo de las diferentes técnicas y métodos de extracción y almacenamiento de los datos.
- Diseñar las interfaces gráficas de la aplicación web y el conjunto de diagramas que representen la solución de acuerdo a los requerimientos previamente definidos.
- Desarrollar el prototipo web que aplique la técnica de Web Scraping, respondiendo a los requerimientos establecidos, usando los lenguajes y librerías necesarias para implementar Web Scraping.
- Realizar la fase de pruebas de interfaz que permitan validar el funcionamiento correcto y la extracción de datos mediante la técnica Web Scraping.

1.2. ESTADO DEL ARTE

Dentro de este proyecto se puede definir como modalidad de la investigación documental que permita determinar la forma como se ha tratado este tema en el desarrollo de un prototipo de aplicación web que permita extracción de las ofertas laborales de las principales plataformas que postulan empleos en el departamento de Bolívar, usando la técnica Web Scraping, como se encuentra el avance del conocimiento en el momento de realizar la investigación y cuáles son las tendencias para el desarrollo de la temática abordaremos los siguientes proyectos que nos evidencian la técnica de Web Scraping en los proyectos mencionados:

El proyecto *“Análisis y diseño de un sistema para la extracción, análisis y comparación de precios de tiendas en la ciudad de Bogotá d.c.”* desarrollado por Yair López Poveda, 2017. Propone el diseño y desarrollo de una solución tecnológica que brinde a las personas una herramienta en la cual se pueda evidenciar una variación de precios de varias tiendas virtuales utilizando técnicas de Web Scraping en la cual se utilizaron los siguientes componentes MarkScrap: o Python, Scrapy, MongoDB, python y para la creación de scripts se utilizó Scrapy, un framework que permite recorrer páginas web y extraer datos esto con el fin de hacer una recolección total y así poder buscar, visualizar, comparar y categorizar precios con varias páginas web [4].

Después de validar la técnica Web Scraping implementada en el proyecto anterior para extraer y comparar precios en tiendas, se procede con el proyecto a continuación: *“Aplicación de Técnicas de Web Scraping al Boletín Oficial de Castilla y León (BOCyL)”* desarrollado por Cristina Hernández herrera, 2014. Se propone estudiar el concepto de Web Scraping y cómo mediante su uso, es posible extraer de forma estructurada la información contenida en un conjunto de páginas web se utilizaran técnicas fundamentales de scraping, que permite extraer la información de un sitio web de interés general como es el Boletín Oficial de

Castilla y León. (BOCyL), y almacenarlo en una base de datos el componente 'Herramienta de scraping', además se mostrará el uso de la información como es la creación de archivos open data será el lenguaje de programación Python, junto con las librerías propias de este programa que permiten realizar la actividad de Scraping [5].

Teniendo en cuenta la aplicabilidad con que se utilizó la Web Scraping en el proyecto anterior al boletín oficial de castilla y león, se enfoca en el siguiente proyecto que aborda un enfoque diferente a la técnica: El proyecto "*Extracción de información a partir de recursos web semi-estructurados*" Aitor Micaelo, 2009. Se basa en la obtención de datos de páginas web para la realización de minería de datos con esto el autor quiere obtener una búsqueda de repositorios web con contenido semi-estructurado un análisis de la estructura de esos repositorios para ver el tipo de información, implementar herramientas de parseado y procesado del contenido web y analizar herramientas de procesamiento del lenguaje natural en él se utilizaran las librerías HTMLParser es una librería de java que se utiliza para analizar HTML a tiempo real [6].

Otro ejemplo práctico y enfoque de la Web Scraping se presenta en el siguiente proyecto: El proyecto "*Recuperación de Información y Recomendaciones en la Web*", Ana Carrato y Ana Menéndez, 2014. Se aplicará Web Scraping a una selección de páginas de inmuebles una vez extraído los datos lograr la integración fueron analizados reconociendo ventajas y desventajas de estos procesos, desarrollando luego una aplicación que centralice el acceso a la información de diferentes fuentes, obtenidos previamente los datos con Web Scraping de esto se ofrece una aplicación web al usuario donde está centralizada toda la información de forma cómoda y rápida de acceder para la obtención de los datos utilizaron una extensión de google chrome web scraper esta aplicación requiere un plan de ruteo donde indica cómo se debe navegar [7].

Por último, le presentamos el proyecto donde lo enfocan en el ámbito de compras de computadores utilizando Web Scraping: El proyecto “*Software Asesor en la compra de computadores*”, Antonio Andrés Jiménez Navas Y Simón Santiago Soriano Pérez, 2013. Tiene como objetivo desarrollo de un sistema de recomendación en compra de computadoras basado en el análisis de los comentarios de los usuarios en blogs con miras a solucionar la problemática de realizar recomendaciones utilizando Sentiment Analysis y datos de páginas web extraídos por medio de Web Scraping utilizando para la extracción de datos la herramienta jsoup, una librería de java que permite la navegación a través de código HTML de una manera sencilla para el análisis de la información y así presentando al usuario final una base confiable para la recomendación de compra de computadores [8].

1.3. MARCOS DE REFERENCIA

1.3.1. Marco Teórico

Dentro de nuestro criterio y aprendizaje de esta técnica abarcaremos parte de la temática de Web Scraping [5].

Web Scraping

La recopilación de datos en internet es casi tan antigua como Internet en sí mismo aunque el raspado web no es un término nuevo en pasados años la práctica ha sido más conocida como raspado de pantalla, minería de datos, recolección de información en la web o variaciones similares en este sentido, cada vez que hemos realizado la acción de “copy&paste” por diferentes páginas de la web para obtener datos, y que posteriormente hemos utilizado para otra actividad diferente, lo que se ha realizado ha sido un “scrapeo” de datos en la web, pero esta práctica resulta ser muy tediosa al tener que gastar mucho tiempo en ello en el caso de querer obtener mucha información de páginas diferentes y posteriormente

organizar esos datos para ser utilizados. El raspado web es la práctica de recopilar información o datos por medio de un (programa que interactúa a través de un navegador usando HTML) esto se logra escribiendo un programa automatizado que consulta un servidor web solicita los datos generalmente en forma de HTML y luego analiza los datos para extraerlos y ser almacenados con el propósito de interpretar y transformar ese contenido de forma automatizada para su posterior utilización.

Funcionamiento de Web Scraping

El funcionamiento o extracción de la información consiste en simular en el navegador o por medio de programas denominados “bots” que crean una clonación de clickeo, lectura y “copy&paste” (copia y pega) automatizando la extracción de información del sitio elegido por el usuario. De modo que la Web Scraping se centra en transformar contenido no estructurado, que por lo general se encuentra en código HTML, en datos estructurados que pueden ser almacenados para su análisis o presentación.

Técnicas utilizadas para el Web Scraping

Existe una gran variedad de técnicas que permiten la realización de Web Scraping a continuación haremos una descripción de su proceso:

Recolección manual de datos: podríamos decir que es el método más arcaico de recopilación de información que es el famoso llamado “copy&paste” (copiar y pegar) copiar la información de un sitio web y pegarla. Esta ardua tarea puede resultar en un trabajo bastante tedioso ya que tendríamos q hacerlo de página en página esto sería demasiado difícil. A pesar de esto a veces este es el método efectivo contra algunas páginas debido a barreras para la realización de la extracción.

Algoritmos de minería de datos: la minería de datos se encarga de manera no trivial de extraer información implícita, previamente desconocida y potencialmente útil de bases de datos, transformando los datos en una estructura comprensible para su previo uso. Muchos sitios web tienen páginas similares ya que para agilizar la creación hacen uso de plantillas la minería de datos se encarga de detectar la plantilla y extraer la información.

Parsers de HTML: Algunos lenguajes, como XQuery y HTQL pueden ser utilizados para parsear documentos, recuperar y transformar el contenido de documentos HTML.

Protocolo HTTP: Páginas webs estáticas y dinámicas pueden ser obtenidas haciendo peticiones HTTP al servidor remoto utilizando sockets algoritmos de minería de datos: muchos sitios webs tienen grandes colecciones de páginas generadas dinámicamente a partir de una base de datos. Datos de la misma categoría aparecen usualmente en páginas similares mediante un script o una plantilla. En la minería de datos, un programa detecta estas plantillas en un contexto específico y extrae su contenido.

Herramienta es GNU Wget: Esta se usa desde la consola y es software libre 100% que permite la descarga de contenidos desde servidores web de una forma simple su nombre está compuesto por W de World Wide Web, y el “get” que significa obtener y estamos tan acostumbradas a leer en los sitios del software libre, actualmente soporta descargas mediante los protocolos HTTP, HTTPS y FTP. Con wget tenemos la posibilidad de bajar espejos enteros que son sitios complejos de forma recursiva (con todos sus directorios y subdirectorios), en conversión de enlaces para la visualización de contenidos HTML localmente.

Aplicaciones Web Scraping: Hay una gran variedad de aplicaciones y librerías disponibles que permiten la extracción de información para conseguir dicha extracción, se debe conocer la estructura de la página a la cual se le desea extraer la información o seleccionar los campos o datos que son de interés para el

usuario, esta es la técnica más recomendable ya que dispone de una gran variedad de librerías para la recolección de información.

Herramientas para el Web Scraping

Existe una gran variedad de herramientas destinadas para la obtención de información cada herramienta se encuentra escrita para uno o varios lenguajes de programación, como lo son Java, Python, PHP, C# en este segmento se va proceder a explicar alguna de ellas, estas herramientas que mencionaremos están destinadas para aquellos usuarios con conocimientos en programación.

Lxml: Es una biblioteca bastante extensa escrita para analizar documentos XML y HTML muy rápidamente, incluso manejando etiquetas desordenadas en el proceso Lxml tiene la ventaja de leer y escribir, en particular, el soporte para XPath facilita considerablemente archivos XML muy grandes además de una fácil programación para simplificar y agilizar procesos.

Scrapy: Es un marco de aplicación para rastrear sitios web y extraer datos estructurados que pueden utilizarse para una amplia gama de aplicaciones útiles, como minería de datos, procesamiento de información o archivo histórico. A pesar de que Scrapy fue diseñado originalmente para raspado web, también se puede utilizar para extraer datos mediante API (como Amazon Asociate Web Services) o como un rastreador web de propósito general.

Beautiful Soup: Es una biblioteca de Python para extraer datos de archivos HTML y XML. Funciona con su analizador favorito para proporcionar formas idiomáticas de navegar, buscar y modificar el árbol de análisis. Generalmente ahorra a los programadores horas o días de trabajo.

1.3.2. Marco Conceptual

En la actualidad cada día tenemos la necesidad de más acceso a la información, esto ha aumentado considerablemente con el transcurso del tiempo este crecimiento nos ha llevado a la llamada era de la información donde cada vez necesitamos más de ella y todo aquello que necesitamos saber y conocer lo encontramos en la web: últimas noticias, avances tecnológicos, cambios en el mercado, y mayormente lo que hacemos en las redes sociales.

Pero mucha de esta información que consultamos primero se identifica y después se extrae de sitios web para proporcionársela al usuario mediante una aplicación o sitio web. Para poder lograr esta extracción de datos se usaron frameworks, lenguajes de programación y técnicas de Web Scraping (raspado web).

A continuación se enunciarán y definirán cada uno de estos conceptos sobre los cuales se enmarcará el trabajo para facilitar la lectura y el entendimiento de esta.

Web Scraping: El scraping es una técnica informática que se utiliza para extraer datos de cualquier página de internet de forma automatizada. Es decir, extraemos a base de datos la información web mediante un programa. La traducción literal de scraping es raspado.

Copiar datos de una página web y pegarlos en una base de datos de Excel se consideraría extracción de datos. Si en lugar de hacerlo de forma manual, recurrimos a robots o bots que automatizan todo el trabajo, estaremos hablando de “Web Scraping” [9].

Python: Lenguaje de programación interpretado, interactivo y orientado a objetos. Combina la potencia con una sintaxis muy clara. Dispone de módulos, clases, excepciones, tipos de datos dinámicos de alto nivel, y creación de tipos. Posee interfaces para muchos sistemas y bibliotecas, así como con varios sistemas de interfaz gráfica (X11, Motif, Tk, Mac, MFC).

Es posible añadirle nuevos módulos escritos en C o C++. Fue creado a principios de los años 90 por Guido Van Rossum. Parece ser que Van Rossum escogió su nombre en homenaje al colectivo cómico británico Monty Python [10].

Html: HTML es el acrónimo en inglés de HyperText Markup Language (en español se traduce como lenguaje de marcado de hipertexto). Es un lenguaje abstracto que usan las aplicaciones para representar documentos (se les llama documentos a instancias completas, como lo son las páginas web) y que puede ser transmitido fácilmente por algún medio, como Internet. Los navegadores de Internet procesan e interpretan documentos descritos en HTML usando un analizador de HTML [11].

Beautiful Soup: BeautifulSoup es una librería en Python que sirve para la extracción sencilla de datos concretos de una página web en HTML sin excesiva programación. Es lo que técnicamente recibe el nombre de parsear HTML. Una de las ventajas de esta biblioteca en Python es que todos los documentos salientes de la extracción de datos lo hacen en UTF-8, lo cual es bastante interesante porque el problema típico de las codificaciones está totalmente resuelto.

Otras de las características potentes de BeautifulSoup es que utiliza analizadores de Python como lxml o html5lib, que permiten rastrear páginas web con estructura de árbol. Gracias a ellos se puede recorrer cada 'habitación' de una web, abrirla, extraer su información e imprimirla [12].

Librerías (Programación): una biblioteca o llamada por vicio del lenguaje librería (del inglés library) es un conjunto de implementaciones funcionales, codificadas en un lenguaje de programación, que ofrece una interfaz bien definida para la funcionalidad que se invoca. Las bibliotecas pueden vincularse a un programa (o a otra biblioteca) en distintos puntos del desarrollo o la ejecución, según el tipo de vínculo que se quiera establecer [13].

Anaconda (Python): Es una distribución gratuita y de código abierto de los lenguajes de programación Python y R para computación científica (ciencia de

datos, aplicaciones de aprendizaje automático, procesamiento de datos a gran escala, análisis predictivo, etc.), cuyo objetivo es simplificar la administración de paquetes y despliegue. Las versiones del paquete son gestionadas por el sistema de gestión de paquetes conda [14].

Mechanize (Librería): Es una distribución gratuita y de código abierto de los lenguajes de programación Python y R para computación científica (ciencia de datos, aplicaciones de aprendizaje automático, procesamiento de datos a gran escala, análisis predictivo, etc.), cuyo objetivo es simplificar la administración de paquetes y despliegue. Las versiones del paquete son gestionadas por el sistema de gestión de paquetes conda [15].

SSL (Librería): Este módulo proporciona una clase SSLSocket que se deriva del socket sockettipo, y proporciona una envoltura similar a un socket que permite encriptar y desencriptar los datos que pasan por el socket con SSL. Admite métodos adicionales como getpeercert (), por ejemplo, que recupera el certificado del otro lado de la conexión y cipher (), que recupera el cifrado que se utiliza para la conexión segura [16].

Csv (Librería): El denominado formato CSV (valores separados por comas) es el formato más común de importación y exportación para hojas de cálculo y bases de datos [17].

1.3.3. Marco Legal

A continuación, se identifica los puntos destacados que regula la práctica y por lo cual se identifica la legalidad o ilegalidad de la técnica Web Scraping particularmente, bajo el cual se desarrolla este proyecto:

Legalidad de Raspado Web: Si bien numerosas herramientas y tecnologías han estado disponibles para ayudar a los investigadores con Web raspado (Munzert et al. 2015), la legalidad de Web raspado es todavía una “zona gris” en el ámbito jurídico (Snell y Menaldo 2016). Aquí definimos la legalidad como el cumplimiento con las leyes y teorías legales. No hay una legislatura que se dirige directamente Web raspado. A partir de ahora, Web raspado es guiado por un conjunto de teorías relacionadas fundamentales legales y leyes, tales como “la infracción de derechos de autor”, “incumplimiento de contrato”, el Fraude y Abuso de Ley (CAFA), y “transgresión de muebles” (Dreyer y Stockton 2013; Snell y Menaldo 2016). Algunos detalles específicos de cómo estas teorías legales fundamentales se aplican a Web raspado se proporcionan a continuación.

Términos de Uso: A menudo se ha argumentado en el campo legal que un propietario de un sitio web puede prevenir con eficacia el acceso mediante programación a el sitio web mediante la prohibición, expresada en el término de “uso de política” publicada en la página web. El incumplimiento de estos términos puede dar lugar a un “incumplimiento de contrato” en el lado del usuario de la página web (secadora y Stockton 2013). Para procesar a alguien por violar las “condiciones de uso”, el usuario del sitio web tiene que entrar en un acuerdo explícito con el propietario del sitio para cumplir con las “condiciones de uso” política (por ejemplo, al hacer clic en una casilla de verificación). Por lo tanto, solo tiene que prohibir el raspado en el sitio web, desde un punto de vista legal no puede impedir a alguien rastreo de la página.

Material con Copyright: Raspar y volver a publicar datos o información que es propiedad y explícitamente derechos de autor por el propietario del sitio web puede dar lugar a un caso de “infracción de copyright” (Secadora y Stockton 2013). Sin embargo, un sitio web no significa necesariamente poseen los datos generados por sus usuarios. Por ejemplo, un sitio web dedicado al comentario no necesariamente poseen los comentarios generados por los usuarios de este sitio web. Por otra parte, las ideas no pueden ser propiedad - solo la forma específica o la representación de esas ideas. Así que uno puede utilizar los datos de derechos de autor para crear resúmenes de datos protegidos por copyright. Por último, todavía se puede usar material con copyright en una escala limitada de conformidad con el principio de “uso justo”.

Propósito de Raspado Web: Cualquier uso ilegal o fraudulento de los datos obtenidos a través de Web raspado está prohibida por varias leyes. Por ejemplo, una persona que acceda a los datos de la Web que se sabe que es confidencial y protegida puede ser procesada en virtud de la Ley de Fraude y Abuso de si el daño que se ha producido es mayor que \$ 5.000 (secadora y Stockton 2013). En la Web, esto a menudo se produce cuando alguien con conocimiento accede a “contenido Premium” y luego revende o continúa accediendo al contenido a través de un canal autorizado después de recibir una carta de “cese y desista” de titular del sitio web (Snell y Menaldo 2016).

Daños en la Web: Si Web raspado sobrecargas o daños de un sitio web o un servidor Web, a continuación, la persona responsable de los daños puede ser procesado bajo la “transgresión de muebles” ley (secadora y Stockton 2013). Sin embargo, los daños deben ser de material y fácil de probar en la corte para que el propietario del servidor web para tener derecho a una compensación económica.

Consideraciones Éticas de Raspado Web: Si bien las leyes y teorías legales se han aplicado a Web raspado en ambos tribunales y la doctrina, la ética de la Web raspado no ha sido abordado por la literatura anterior. Si bien hay muchos puntos de vista sobre la ética, a los efectos de este proyecto de investigación que vemos la ética como “un conjunto de conceptos y principios que nos guía para determinar qué comportamiento benefician o perjudican a las criaturas sensibles” (Paul y Elder 2006). Además de violar las leyes vigentes, raspado web puede resultar en un daño no intencionado a los “seres sensibles” que se asocian a un determinado sitio web, tales como los propietarios o clientes del sitio web. Estas consecuencias perjudiciales, son, por definición, difícil de predecir (Luz y McGrath 2010).

Caso: Sentencia del Tribunal Supremo de 9 de octubre, nº 572/2012, de RyanAir VS Atrápalo, en la cual se determinó que, en ese caso, el Web Scraping llevado a cabo por Atrápalo era legal [22].

En Colombia no hay una ley o reglamentación que rija esta técnica de raspado web o como Web Scraping, lo cual se debe considerar no dar mal uso de los datos que se obtiene al Scrapear o manipular la información, en este caso el proyecto Web Scraping no modifica ni altera la información que se extrajo de la web, solo se utiliza para almacenamiento, la búsqueda y visualización de ofertas laborales en Bolívar.

1.4. DISEÑO METODOLÓGIA

Línea de Investigación

Una línea de investigación es un enfoque que abarca conocimientos, inquietudes, prácticas y perspectivas de análisis que permitan el desarrollo de proyectos y productos construidos de manera sistemática alrededor de un tema de estudio. Adicionalmente, concibe el trabajo tanto interdisciplinario como multidisciplinario.

En la actualidad podemos encontrar diferentes líneas que abarcan un área específica de un proyecto, pero la que identifica la línea de investigación del prototipo web que permita la extracción de las ofertas laborales de las principales plataformas que ofertan empleos en el departamento de Bolívar, usando la técnica Web Scraping, centralizando varios portales de empleos en uno solo para una mejor optimalización y rapidez en la búsqueda de ofertas.

Podemos identificar que la línea de investigación está contemplada en: “Desarrollo de software” del grupo DEARTICA de la escuela de Ingeniería de Sistemas de la Universidad del Sinú, contribuyendo al fortalecimiento con este proyecto.

Tipo de Investigación

De acuerdo al enfoque con que cuenta el proyecto de Web Scraping como herramienta para realizar búsqueda de ofertas laborales, este tipo de investigación cuantitativa que se ha elaborado, pretende medir los patrones de datos a través de la extracción de información, almacenamiento, análisis y visualización.

La investigación o metodología cuantitativa es el procedimiento de decisión que pretende señalar, entre ciertas alternativas, usando magnitudes numéricas que pueden ser tratadas mediante herramientas del campo de la estadística. Por eso la investigación cuantitativa se produce por la causa y efecto de las cosas.

Para el desarrollo de un prototipo de aplicación web que permita extracción de las ofertas laborales de las principales plataformas que ofertan empleos en el departamento de Bolívar, usando la técnica Web Scraping. Se inició una investigación teniendo en cuenta el ciclo de vida del desarrollo de software y posteriormente aplicando la metodología Ágil XP en sus respectivas fases del proyecto.

Definición De La Metodología

Dentro de la metodología del proyecto Web Scraping como herramienta para desarrollar un prototipo para extraer, consultar y centralizar la información en la web de portales de empleos, en el inicio de sus procesos y como base a su desarrollo se enfoca a implementar la metodología de desarrollo ágil XP o programación extrema.

La programación extrema o XP es una metodología de desarrollo que se englobaría dentro de las denominadas metodologías ágiles en la que se da máxima prioridad a la obtención de resultados y reduce la burocracia que utiliza las metodologías tradicionales. XP es una metodología ágil para el desarrollo de software y consiste básicamente en ajustarse estrictamente a una serie de reglas que se centran en las necesidades del cliente para lograr un producto de buena calidad en poco tiempo, centrada en potenciar las relaciones interpersonales como clave para el éxito del desarrollo de software [20].

Otra ventaja es que se trabaja con mayor velocidad y eficiencia, ahorrar tiempo y costes a la hora de la entrega del proyecto, como también mejorar la motivación e implicación del equipo de desarrollo; permite que todo el equipo conozca el estado del proyecto y sus avances, como también los compromisos que son negociados y aceptados por los miembros.

Cabe destacar que esta metodología mejora la calidad del producto, a eso se debe que tanto los desarrolladores como los clientes tienen como objetivo que el

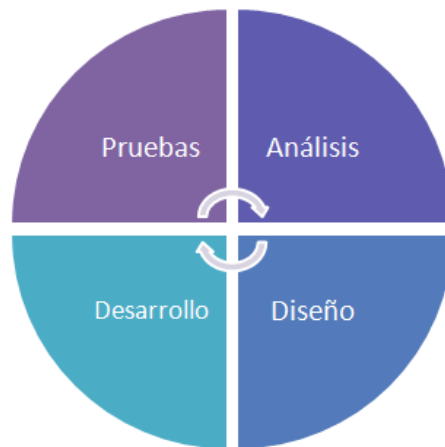
producto final sea exactamente lo que desea el cliente y busca, con esto es posible dar respuesta a aquellos incidentes o alertar de forma rápida tanto de errores o problemas que puedan darse desde el inicio, con lo que se mejora coste y entrega.

CICLO DE VIDA XP:

Como toda metodología en el desarrollo de un proyecto que cuanta con unas etapas o fases, el ciclo de vida XP como ágiles incluye los siguientes pasos:

- Entender lo que el cliente necesita > Fase de Exploración
- Estimar el esfuerzo > Fase de Planificación
- Crear la solución > Fase de Iteraciones
- Entregar el producto final al cliente > Fase de Puesta en Producción

Figura 1. Modelo de Prototipos, Fuente Autor Tomada de [20]



Lo que caracteriza a XP, al igual que al resto de métodos ágiles es un ciclo de vida dinámico. ¿Cómo lo logra XP? Mediante ciclos de desarrollo cortos (llamados iteraciones), al fin de los cuales se generan unos entregables funcionales.

En cada iteración se realiza un ciclo completo de análisis, diseño, desarrollo y pruebas, pero utilizando un conjunto de reglas y prácticas específicas de XP.

Seleccionar la técnica más apropiada de Web Scraping, mediante un análisis comparativo de las diferentes técnicas y métodos de extracción y almacenamiento de los datos.

En la actualidad existen muchas herramientas para el raspado y extracción de información en la web, existen diversas, pero para ello tomaremos las más relevantes para hacer un análisis comparativo de las diferentes técnicas de Web Scraping seleccionaremos las dos técnicas más relevante en la que nos enfocaremos que son BeautifulSoup y scrapy.

BeautifulSoup

Es una de las herramientas con una de las curvas de aprendizaje más rápido y fácil de aprender, esta herramienta ayuda al programador a recopilar información rápidamente de los datos que se desean extraer de la web seleccionada es muy amigable con desarrolladores novatos que desean aprender sobre Web Scraping esta herramienta de extracción permite analizar y extraer sin ninguna molestia el contenido a extraer, una de las ventajas de BeautifulSoup es su capacidad para detectar automáticamente las codificaciones. Esto le permite manejar con gracia los documentos HTML con caracteres especiales, esto hace que sea rápido e indoloro construir aplicaciones comunes. Por ejemplo, si desea encontrar todos los enlaces en la página web que bajamos antes, solo hay unas pocas líneas:

```
from bs4 import BeautifulSoup
Soup = BeautifulSoup (html_doc, 'html.parser')
for link in Soup.find_all('a):
```



```
print (link.get('href'))
# http://example.com/elsie
# http://example.com/lacie
# http://example.com/tillie
```

Es rápida sencilla fácil de aprender funciona muy bien y aparte BeautifulSoup funciona bien en Python 2 y Python 3, BeautifulSoup es una biblioteca de Python para extraer datos de archivos HTML y XML. Por lo que la compatibilidad no será un problema con esta herramienta podrá crear proyectos sin ningún problema

Scrapy

Es un marco de rastreo web para el desarrollador que puede extraer datos estructurados que pueden utilizar en una amplia gama de aplicaciones como extracción de información, minería de datos entre otras más. Fue originalmente para hacer raspado de información web, una de las características más importante es en la que está construida Twisted, una biblioteca de red asíncrona, por lo que Scrapy se implementa utilizando un código no bloqueante (también conocido como asíncrono).

Lo que hace que el rendimiento de la araña sea muy bueno, es que tiene soporte incorporado para extraer datos de fuentes HTML utilizando la expresión XPath y la expresión CSS y al igual que BeautifulSoup funciona muy bien Python 2 y Python 3, por lo que la compatibilidad no será un problema con scrapy, como ya lo es una herramienta más completa pero esta la hace que su curva de aprendizaje sea más compleja que todas las demás herramientas de extracción.

¿Entonces cual elegir?

Las dos herramientas de raspado son muy buenas en sus trabajos, BeautifulSoup solo se utiliza para analizar el HTML y extraer los datos y guardarlos, scrapy se

utiliza para descargar HTML, procesar y guardar los datos, para seleccionar la herramienta a utilizar en este proyecto se deben considerar muchos factores.

Curva de aprendizaje

BeautifulSoup

Es mucho más fácil de aprender se puede usar más rápidamente para extraer los datos que desea, claro en la mayor parte de la extracción se necesitarán complementos o librerías para ayudar a obtener la fuente HTML como Requests para que permita enviar solicitudes HTTP utilizando Python. Con él, puede agregar contenido como encabezados, datos de formulario, archivos de varias partes y parámetros a través de bibliotecas Python simples. Si usted es un desarrollador novato, no tiene mucha experiencia en programación y desea hacer un pequeño trabajo, BeautifulSoup puede ser su elección aquí porque es muy poco probable que no lo decepcione.

Scrapy

La curva de aprendizaje Scrapy es mucho más pronunciada y se debe leer algo de la documentación de Scrapy y tutoriales para comprender cómo funciona y trabajar duro para convertirse en un experto en Scrapy que no solo se ocupa de la extracción de contenido, sino también de muchas otras tareas, como descargar HTML.

Extensibilidad

La estructura scrapy es muy bien diseñada, si su proyecto si su producto tiende a ser muy extenso y robusto usted se beneficiará mucho de esta herramienta por su gran poder en la arquitectura y podrá para obtener mejores resultados. Pero por el

contrario si su proyecto es pequeño y su lógica no es muy compleja y pequeña y aparte de eso desea un trabajo rápido y fácil lo mejor es utilizar BeautifulSoup

Actuación

Con Scrapy puede enviar muchas solicitudes al mismo tiempo, por lo que es necesario configurar `download_delay` en la mayoría de los casos para evitar la prohibición, las páginas web se pueden rastrear rápidamente. Sin embargo, BeautifulSoup no tienen esta característica por lo que mucha gente dice que BeautifulSoup es lento. En realidad, esto es incorrecto, puede importar `multiprocessing` para acelerar todo el progreso. Entonces, si desea desarrollar una araña eficiente que pueda rastrear muchos conjuntos de datos en poco tiempo, Scrapy puede ahorrar mucho tiempo. Si no tiene experiencia con desarrolladores de Python, entonces BeautifulSoup debe ser su elección.

Tabla 3. Marco Referencia BeautifulSoup Vs Scrapy

Marco de referencia	BeautifulSoup	Scrapy
Curva de aprendizaje	Muy fácil de aprender, amigable para los principiantes.	La curva de aprendizaje de Scrapy es mucho más pronunciada, y trabajar duro para convertirse en un experto en Scrapy.
Extensibilidad	No es muy fácil extender el proyecto.	Puede desarrollar fácilmente middleware o canalización personalizados para agregar funciones personalizadas, fáciles de mantener.
Actuación	Necesitas importar <code>multiprocessing</code> para que funcione más rápido	Muy eficientes, las páginas web se pueden rastrear en poco tiempo.

Teniendo en cuenta las descripciones hecha de las temáticas anteriores y la tabla 3 de marco de referencia, se seleccionó la librería BeautifulSoup como la más apropiada y fácil uso de en python Spyder en la programación del código Web Scraping.

2. ANÁLISIS DEL SISTEMA

En este capítulo se hace la elicitación y especificación de requerimientos utilizando la norma IEEE-830 [21] que facilita la descripción general del sistema, con el fin de disminuir los problemas en el proceso de desarrollo y aplicando la técnica de entrevista con el usuario final.

2.1. LEVANTAMIENTO DE REQUERIMIENTOS

Técnicas de Levantamiento de Requerimientos

Para levantar los requerimientos del sistema, el autor de esta investigación se basó en las siguientes preguntas que fueron formuladas con la ayuda del docente AISNER MARRUGO Escuela de Ing. Sistema de la Universidad del Sinú Seccional Cartagena.

1. ¿Cuál es el proceso básico que la plataforma pretende resolver?
2. ¿Qué datos utiliza o produce este proceso?
3. ¿Qué eventos pueden ocurrir durante la ejecución del proceso?
4. ¿cuáles son las actividades que involucra este proceso?

Para especificar los requisitos funcionales y no funcionales se utilizarán las siguientes plantillas:

Tabla 4. *Plantilla de Requerimientos Funcionales, Fuente Autor*

Identificador:	<i>Entidad-Número Requerimiento</i>
Nombre del requerimiento:	<i>Nombre del requerimiento</i>
Actor:	<i>Usuarios involucrados en el requerimiento</i>
Indispensable/ deseable:	<p><i>Un Requerimiento No Funcional es Indispensable cuando es necesario tenerlo en cuenta para la realización de algún requerimiento funcional.</i></p> <p><i>Un Requerimiento No Funcional es Deseable cuando la funcionalidad que proporciona puede ser implementada de forma opcional, y ningún requerimiento funcional depende de ella. Ej.: Indispensable</i></p>
Prioridad:	<i>Nivel de prioridad del requerimiento (Alta, media, baja)</i>
Visible:	<i>Si será visto por el usuario</i>
Autor:	<i>Autor del sistema</i>
Fecha de elaboración:	<i>Fecha elaboración del sistema</i>
Revisión:	<i>Tutor metodológico</i>
Última revisión:	<i>Última revisión</i>
Resumen:	<i>Descripción del requerimiento</i>
Eventos:	<i>Eventos que se manifiestan en el desarrollo del requerimiento</i>
Caminos alternativos:	<i>Alternativas más complejas son descritas en esta sección por separado</i>
Excepción:	<p><i>Las acciones realizadas luego de presentarse las Excepciones son descritas en esta sección por separado, El criterio de "Crítico" implica evaluar el impacto de la ausencia y presencia del RNF, y cómo se ve afectado el sistema por estas condiciones. Ej.: Sí</i></p>
Precondiciones:	<i>Estado en el cual el sistema debe encontrarse antes de que la funcionalidad que expresa el requerimiento se lleve a cabo.</i>
Post-condiciones:	<i>Cosas que deben ocurrir después</i>

Tabla 5. *Plantilla de Requerimientos no Funcionales, Fuente Autor*

Identificador:	<i>Entidad – RNF – Número del RNF</i>
Nombre del requerimiento:	<i>Nombre del requerimiento</i>
Tipo:	<i>Tipo de Requerimiento No Funcional. Ej.: Restricción Tecnológica</i>
Indispensable/ deseable:	<i>Un Requerimiento No Funcional es Indispensable cuando es necesario tenerlo en cuenta para la realización de algún requerimiento funcional.</i> <i>Un Requerimiento No Funcional es Deseable cuando la funcionalidad que proporciona puede ser implementada de forma opcional, y ningún requerimiento funcional depende de ella. Ej.: Indispensable</i>
Crítico:	<i>El criterio de “Crítico” implica evaluar el impacto de la ausencia y presencia del RNF, y cómo se ve afectado el sistema por estas condiciones. Ej.: Sí</i>
Descripción:	<i>Comentario breve del por qué se necesita este RNF. Ej.: Para este proyecto estamos sujetos a utilizar tecnología .Net para el desarrollo debido a restricciones del cliente.</i>
Criterios de Aceptación:	<i>Criterios que validan este RNF frente a los interesados. Ej.: Todo el código que se implemente debe ser en lenguaje .Net.</i>

2.2. REQUISITOS FUNCIONALES

A continuación se representa el análisis de requerimientos necesarios para el desarrollo del prototipo web en cuanto a sus funcionalidades del sistema:

Tabla 6. Requerimiento RF-01 Inicio Sesión.

Identificador	RF-01
Nombre del requerimiento:	Inicio Sesión
Actor	Administrador
Indispensable/deseable	Indispensable.
Prioridad.	Alta
Visible	Si
Autor	Luis Fernando Gómez, Víctor Orozco
Fecha de elaboración:	29-03-2019
Revisión:	Aisner Marrugo
Última revisión:	
Resumen:	El administrador Inicia sesión con sus credenciales de Usuario y contraseña.
Eventos:	1. Podrá acceder a la interfaz portales de empleo para la selección al script a editar.
Caminos alternativos:	N/A
Excepción:	N/A
Precondiciones:	El administrador deberá estar creado en la base de datos del sistema.
Post-condiciones:	N/A

Tabla 7. Requerimiento RF- 02. Selección portal Script.

Identificador	RF-02
Nombre del requerimiento:	Selección del portal para la modificación del Script.
Actor	Administrador
Indispensable/ deseable	Indispensable.
Prioridad.	Alta
Visible	Si
Autor	Luis Fernando Gómez, Víctor Orozco
Fecha de elaboración:	29-03-2019
Revisión:	Aisner Marrugo
Última revisión:	
Resumen:	El administrador elegirá el portal donde desea actualizar o editar el script del código python.
Eventos:	1. se dirigirá a la presentación del código donde mostrara las variables y la etiqueta a editar.
Caminos alternativos:	N/A
Excepción:	N/A
Precondiciones:	El administrador deberá iniciar sesión para poder acceder a interfaz.
Post-condiciones:	N/A

Tabla 8. Requerimiento RF- 03. Editar Etiquetas Script.

Identificador	RF-03
Nombre del requerimiento:	Editar Etiquetas del Script.
Actor	Administrador
Indispensable/ deseable	indispensable
Prioridad.	Alta
Visible	Si
Autor	Luis Fernando Gómez, Víctor Orozco
Fecha de elaboración:	29-03-2019
Revisión:	Aisner Marrugo
Última revisión:	
Resumen:	El sistema permitirá al administrador consultar la etiqueta seleccionada en el script.
Eventos:	<ol style="list-style-type: none"> 1. visualizara la etiqueta de la variable (dato) del módulo de extracción. 2. Clic en el botón editar para ver la etiqueta del script que se encuentra en la base de datos.
Caminos alternativos:	N/A
Excepción:	N/A
Precondiciones:	N/A
Post-condiciones:	N/A

Tabla 9. Requerimiento RF- 04. Actualizar Etiquetas Script.

Identificador	RF-04
Nombre del requerimiento:	Actualizar etiquetas del Script.
Actor	Administrador
Indispensable/ deseable	indispensable
Prioridad.	Alta
Visible	Si
Autor	Luis Fernando Gómez, Víctor Orozco
Fecha de elaboración:	29-03-2019
Revisión:	Aisner Marrugo
Última revisión:	
Resumen:	El sistema permitirá al administrador Actualizar la etiqueta modificada en el script.
Eventos:	<ol style="list-style-type: none"> 1. El administrador actualiza la etiqueta de la variable (dato) del módulo de extracción. 2. Clic en el botón actualizar para cambiar el script que se encuentra en la base de datos.
Caminos alternativos:	N/A
Excepción:	N/A
Precondiciones:	N/A
Post-condiciones:	N/A

Tabla 10. Requerimiento RF- 05. *Buscar por vacante, empresa y portal.*

Identificador	RF-05
Nombre del requerimiento:	Buscar por Vacante, empresa y portal
Actor	Usuario
Indispensable/ deseable	Indispensable
Prioridad.	Alta
Visible	Si
Autor	Luis Fernando Gómez, Víctor Orozco
Fecha de elaboración:	29-03-2019
Revisión:	Aisner Marrugo
Última revisión:	
Resumen:	El sistema permitirá al usuario realizar consultas por nombre de la vacante, empresa y portal.
Eventos:	<ol style="list-style-type: none"> 1. El sistema permitirá en la en la ventada principal del portal de empleo mostrar la vacante, empresa o portal que se consultó. 2. El sistema permitirá consultar cualquier información de los campos de búsqueda de la Tabla 2.
Caminos alternativos:	N/A
Excepción:	N/A
Precondiciones:	N/A
Post-condiciones:	N/A

Tabla 11. Requerimiento RF- 06. Redirección de la Vacante hacia el portal externo.

Identificador	RF-06
Nombre del requerimiento:	Redireccionamiento de la vacante hacia el portal de empleo y aplicar.
Actor	Usuario
Indispensable/ deseable	Indispensable
Prioridad.	Alta
Visible	Si
Autor	Luis Fernando Gómez, Víctor Orozco
Fecha de elaboración:	29-03-2019
Revisión:	Aisner Marrugo
Última revisión:	
Resumen:	El sistema permitirá dar clic en la opción en la vacante que el usuario desee ver o aplicar.
Eventos:	3. El usuario consultará en la en la ventada principal de portal de empleo, podrá dar un clic en el botón aplicar en la vacante en el cual se dirigirá al portal de empleo para una posible postulación o consulta de la oferta.
Caminos alternativos:	N/A
Excepción:	N/A
Precondiciones:	N/A
Post-condiciones:	N/A

2.3. REQUERISITOS NO FUNCIONALES

Un requisito no funcional es aquel que permite evaluar el comportamiento de un sistema y especificar criterios para juzgar la operación de ella, entre estas tenemos:

Tabla 12. Requerimiento No Funcional – 01. Interfaz de la plataforma.

Identificador	RNF-01
Nombre del requerimiento:	Interfaz de la plataforma.
Tipo	Restricción de tecnología
Indispensable/deseable	Indispensable.
Critico	Si
Descripción:	La plataforma debe tener una interfaz de uso intuitiva y sencilla para que sea de fácil manejo a los usuarios del sistema.
Criterios de Aceptación:	La plataforma debe ser responsiva al usuario final y que sea usada en cualquier tipo de dispositivo móvil.

Tabla 13. Requerimiento No Funcional – 02. Seguridad en información.

Identificador	RNF-02
Nombre del requerimiento:	Seguridad en información
Tipo	Restricción de tecnología
Indispensable/deseable	Indispensable.
Critico	Si
Descripción:	Garantizar a los clientes la seguridad del sistema respecto a la información y datos que han sido recolectados
Criterios de Aceptación:	Proteger la privacidad de la información almacenada en la base de datos.

Tabla 14. Requerimiento No Funcional – 03. Eficiencia

Identificador	RNF-03
Nombre del requerimiento:	Eficiencia
Tipo	Restricción de tecnología
Indispensable/deseable	Indispensable.
Critico	Si
Descripción:	Alcanzar el objetivo fijado con el mínimo de esfuerzos y recursos del sistema.
Criterios de Aceptación:	El sistema debe ser capaz de brindar información actualizada y según las peticiones del usuario.

Tabla 15. Requerimiento No Funcional – 04. Mantenimiento.

Identificador	RNF-04
Nombre del requerimiento:	Mantenimiento
Tipo	Restricción de tecnología
Indispensable/deseable	Indispensable.
Critico	Si
Descripción:	Deberá ser mantenible, a través de módulos y herramientas de administración.
Criterios de Aceptación:	El sistema debe ser capaz de dar mantenimiento de una manera sencilla, localizando y corrigiendo errores del producto software.

Tabla 16. Requerimiento No Funcional – 05. Usabilidad.

Identificador	RNF-05
Nombre del requerimiento:	Usabilidad
Tipo	Restricción de tecnología
Indispensable/ deseable	Indispensable.
Critico	Si
Descripción:	La interfaz con el usuario será a través de una página web diseñada con la técnica responsive que busca una agradable y correcta visualización de una misma página en distintos dispositivos. Desde ordenadores de escritorio tablets y móviles. Esta será construida específicamente para el sistema propuesto.
Criterios de Aceptación:	Deberá ser mantenible, a través de módulos y herramientas de administración.

3. DISEÑO DEL SISTEMA

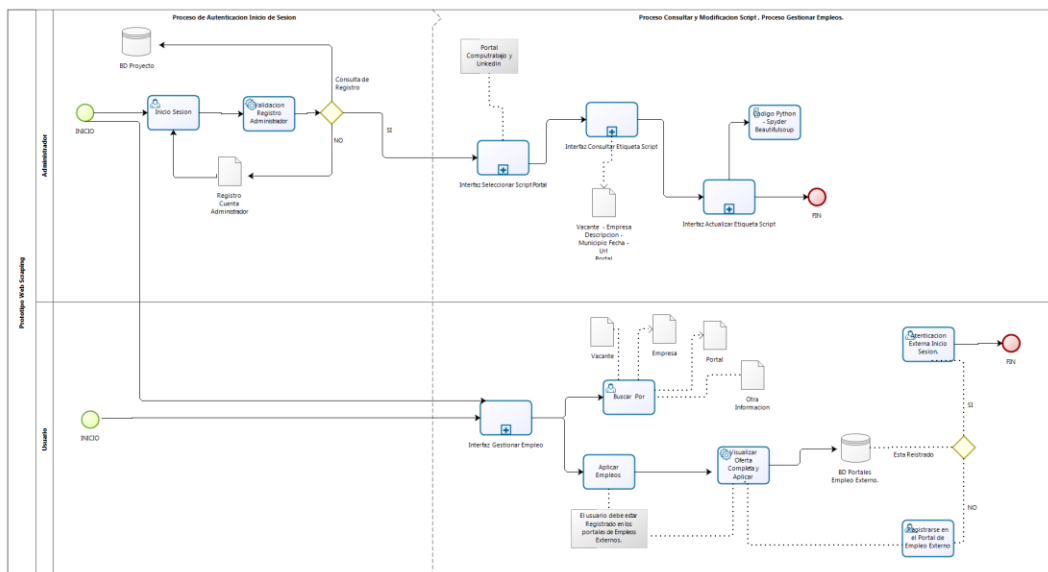
Este capítulo comprende el diseño de un prototipo web utilizando la técnica scraping para centrar los empleos más relevantes en un solo portal, caracterizando aquellos procesos que dan una mayor comprensión en el desarrollo de un software.

El modelo de datos proporciona una vista del modelo lógico y físico de datos, arquitectura general del software y las interfaces gráficas de los usuarios que interactuarán con el prototipo web. Estos modelos representan los requerimientos identificados en la fase anterior e implementando el uso de las técnicas apropiadas y adoptadas en la calidad de software.

3.1. DIAGRAMA DE FLUJO DEL PROTOTIPO

Este diagrama representa el flujo o secuencia de las diferentes acciones que se dan en el prototipo web, para este caso se representara diagramas de procesos, Ver en anexo B.

Figura 2. Diagrama de Flujo BPMN.



En la *Figura 2*. Se representa el diagrama de flujo o procesos BPMM de los actores: administrador y el usuario que corresponde a los requerimientos: RF-01, RF-02, RF-03, RF-04, RF-05, RF-06, RNF-01, RNF-02, RNF-03, RNF-04, RNF-05, funcionalidades del sistema.

3.2. DIAGRAMA DE CASOS DE USO

Otras de las maneras de representar las funcionalidades del sistema desde el punto de vista del usuario, utilizando actores y casos de uso, es a través de este diagrama permitiendo conocer las acciones previstas por el sistema para los usuarios. Para este prototipo web se definieron dos actores el administrador y los usuarios que interactúan con el sistema.

Figura 3. Diagrama casos de uso para Requerimiento RF-01, RF-02, RF-03, RF-04, RF-05, RF-06.

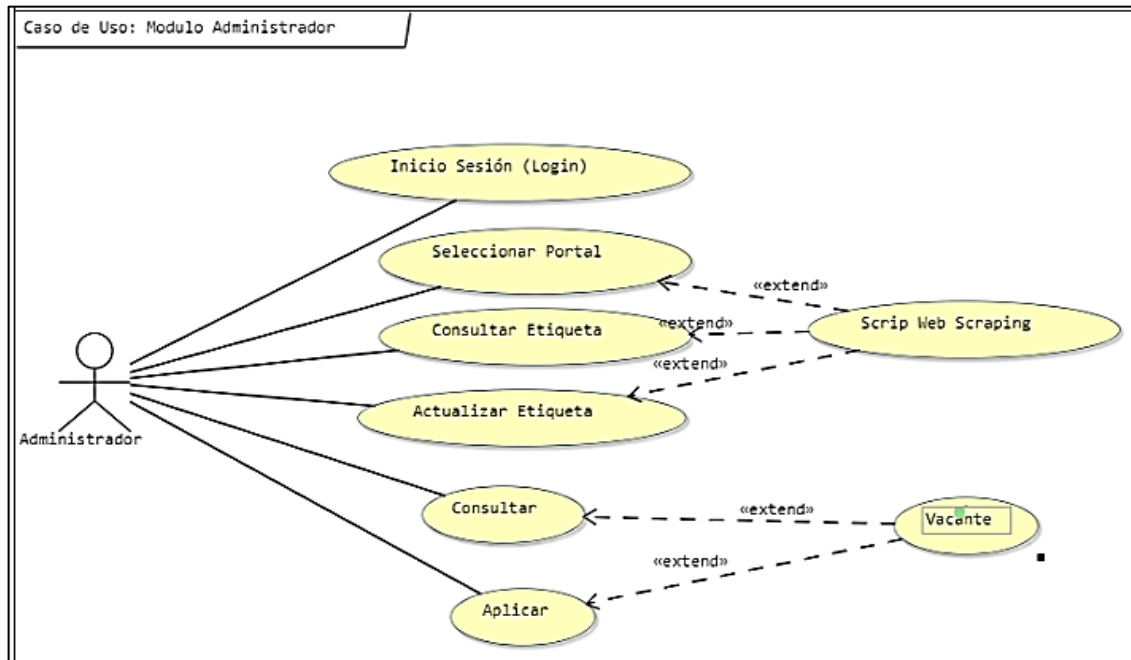
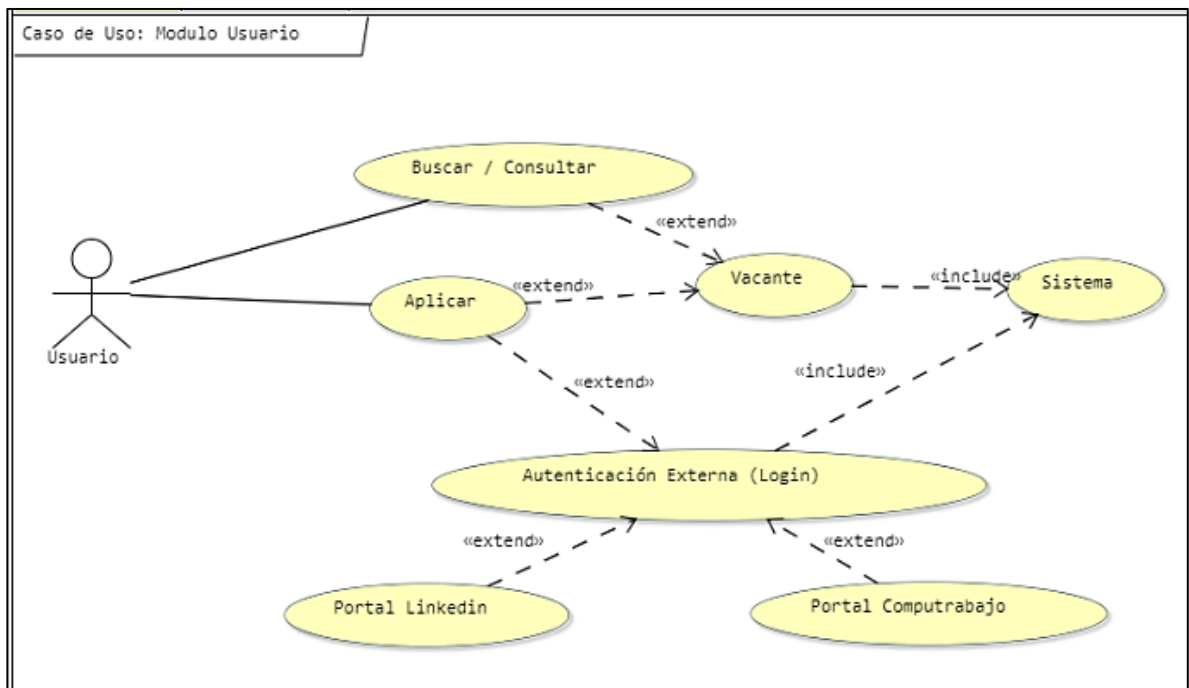


Figura 3. Se representa el diagrama de caso de uso con el módulo administrador y que corresponde a los requerimientos: RF-01, RF-02, RF-03, RF-04, RF-05, RF-06 funcionalidades del sistema.

Figura 4. Diagrama casos de uso para Requerimiento RF-05, RF-06.



En la Figura 4. Se representa el diagrama de caso de uso con el módulo Usuarios y que corresponde a los requerimientos: RF-05, RF-06, funcionalidades del sistema.

3.3. MODELO DE DATOS LÓGICO

A continuación, se muestra el modelo lógico de la base de datos que define las relaciones, los objetos, los datos y características representativas como entidades y atributos; para este prototipo no se identifica ninguna relación para la tabla Vacantes.

Figura 5. Modelo Relacional Campos Tabla Vacantes.



La Figura 5. La tabla vacantes no tiene una relación, aquí se encuentra los campos que identifican que tipo de información se almacena al extraer los datos con la técnica Web Scraping.

Figura 6. Diagrama Entidad-Relación Tipo de Datos Tabla Vacantes.

vacantes	
P *	vacante VARCHAR2 (45)
	empresa VARCHAR2 (45)
	descripcion CLOB
	municipio VARCHAR2 (45)
	fecha VARCHAR2 (45)
	url_vacante CLOB
	portal_web VARCHAR2 (45)
vacantes_PK (vacante)	

La Figura 6. La tabla vacantes no tiene una relación, aquí identificamos el tipo de datos que contiene la tabla, se especifican estos tipos para que no se genere error al insertar los datos directamente extraídos con Web Scraping a la tabla.

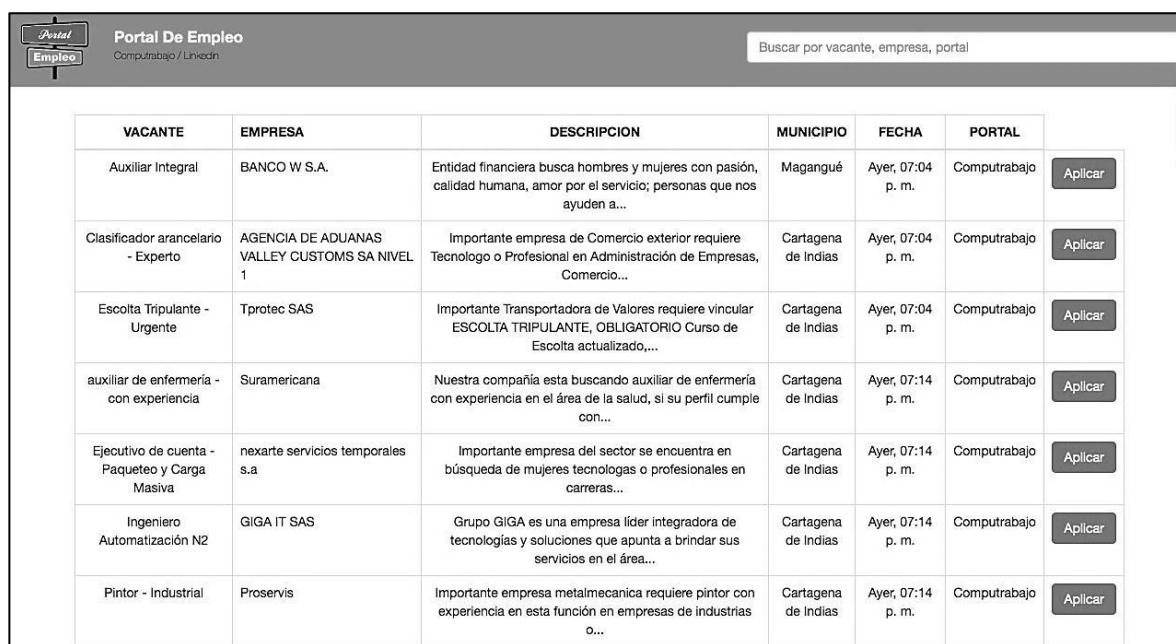
3.4. DISEÑO DE INTERFACES DE USUARIO

Gestionar Empleos

En la *Figura 7*. De la página principal se observa el listado de vacantes de empleos en Bolívar del portal computrabajo y linkedin tanto para los usuarios que visitan el portal de empleo, en esta interfaz no tiene un login como el administrador pues es visible sin necesidad de registrarse.

Los campos visualizados en la interfaz son un cuadro de texto para poder buscar por: vacante, empresa y el portal, pero también, se puede buscar por cualquier tipo de información. Otros campos que corresponden a los empleos que fueron extraídos son: la vacante, empresa, descripción, municipio, fecha y portal.

Figura 7. Diseño Gestionar Empleos.



VACANTE	EMPRESA	DESCRIPCION	MUNICIPIO	FECHA	PORTAL	
Auxiliar Integral	BANCO W S.A.	Entidad financiera busca hombres y mujeres con pasión, calidad humana, amor por el servicio; personas que nos ayuden a...	Magangué	Ayer, 07:04 p. m.	Computrabajo	Aplicar
Clasificador arancelario - Experto	AGENCIA DE ADUANAS VALLEY CUSTOMS SA NIVEL 1	Importante empresa de Comercio exterior requiere Tecnólogo o Profesional en Administración de Empresas, Comercio...	Cartagena de Indias	Ayer, 07:04 p. m.	Computrabajo	Aplicar
Escolta Tripulante - Urgente	Tprotec SAS	Importante Transportadora de Valores requiere vincular ESCOLTA TRIPULANTE, OBLIGATORIO Curso de Escolta actualizado,...	Cartagena de Indias	Ayer, 07:04 p. m.	Computrabajo	Aplicar
auxiliar de enfermería - con experiencia	Suramericana	Nuestra compañía esta buscando auxiliar de enfermería con experiencia en el área de la salud, si su perfil cumple con...	Cartagena de Indias	Ayer, 07:14 p. m.	Computrabajo	Aplicar
Ejecutivo de cuenta - Paqueteo y Carga Masiva	nexarte servicios temporales s.a	Importante empresa del sector se encuentra en búsqueda de mujeres tecnólogas o profesionales en carreras...	Cartagena de Indias	Ayer, 07:14 p. m.	Computrabajo	Aplicar
Ingeniero Automatización N2	GIGA IT SAS	Grupo GIGA es una empresa líder integradora de tecnologías y soluciones que apunta a brindar sus servicios en el área...	Cartagena de Indias	Ayer, 07:14 p. m.	Computrabajo	Aplicar
Pintor - Industrial	Proservis	Importante empresa metalmeccanica requiere pintor con experiencia en esta función en empresas de industrias o...	Cartagena de Indias	Ayer, 07:14 p. m.	Computrabajo	Aplicar

Redirección y Aplicar Empleos

En la Figura 8. Interfaz gestionar empleos, el usuario al dar aplicar a una vacante, se redireccionan a los portales donde aparece este empleo postulado que puede ser de computrabajo o linkedin, como requisito el usuario debe estar registrado en los portales externos.

Figura 8. Redirección y Aplicar Empleos.

VACANTE	EMPRESA	DESCRIPCION	MUNICIPIO	FECHA	PORTAL	
Aprendiz Mercadeo Y ventas - Cartagena	Eficacia	Importante empresa del sector requiere aprendicesTécnico en los programas de Mercadeo y Ventas, Merca-Impulso,...	Cartagena de Indias	Ayer, 09:59 p. m.	Computrabajo	<input type="button" value="Aplicar"/>

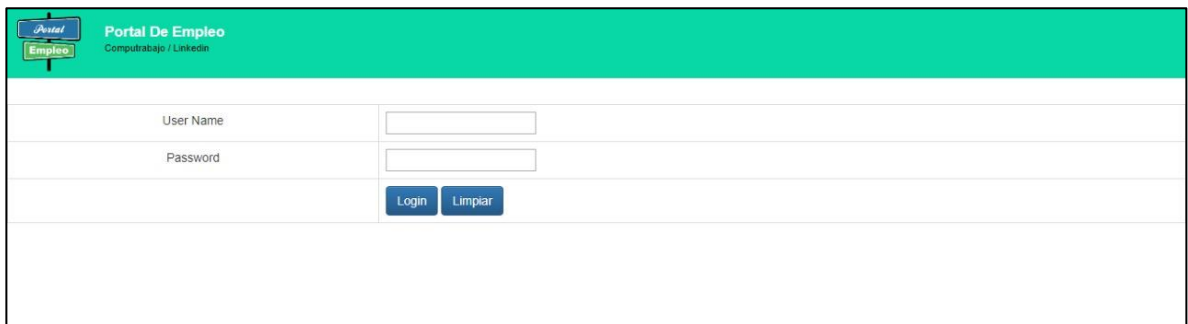
4. DESARROLLO

4.1. DESARROLLO DE LA PLATAFORMA WEB

Interfaz Inicio de Sesión Administrador

En la Figura 9. de la interfaz iniciar sesión permite la autenticación que se mantendrá durante la ejecución de la aplicación como “*administrador*”, se deberá ingresar su *usuario* “*User Name*” y *contraseña* “*Password*”, se verificará la existencia del usuario en los registros de la base de datos.

Figura 9. Interfaz Inicio de Sesión Administrador.



Portal De Empleo
Computrabajo / LinkedIn

User Name

Password

Interfaz Seleccionar Script del Portal de Empleo

La ventana principal del administrador Figura 10. Se muestran dos iconos que representan los portales de empleo donde se actualiza el script del prototipo web “Portal de Empleo”, solo el administrador tendrá acceso a esta interfaz.

Figura 10. Interfaz Seleccionar Script de los Portales de Empleo.



Interfaz Editar Etiquetas Script

Figura 11. Interfaz Consultar Etiquetas Script.

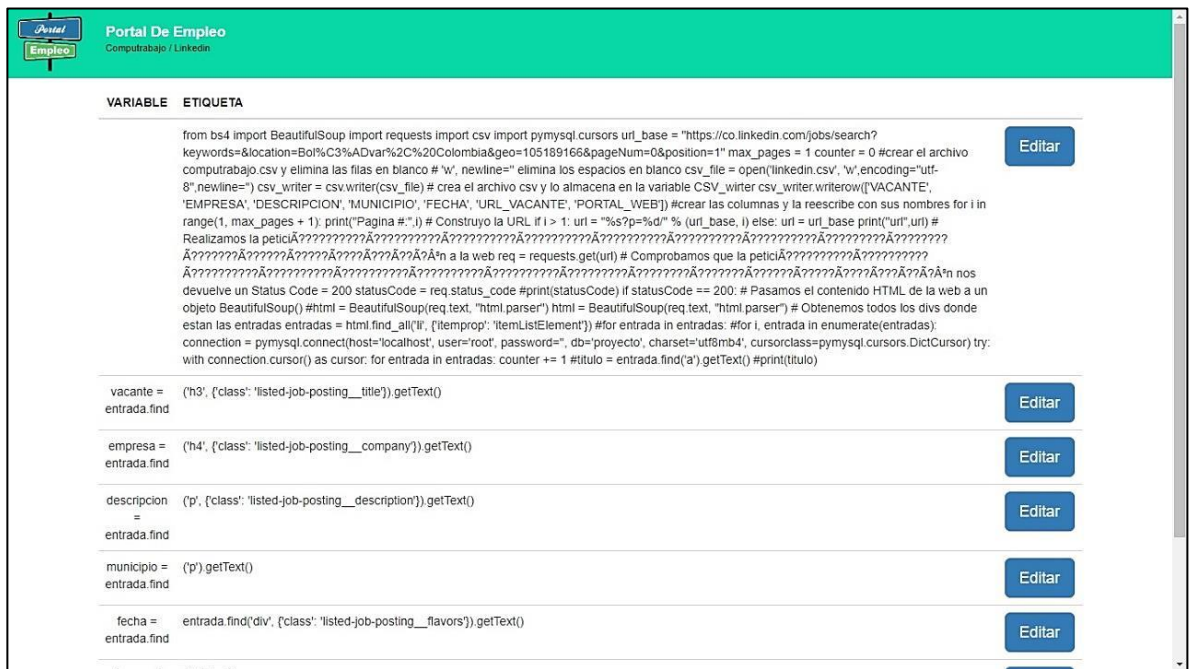


Figura 11. Cada vez que haya una actualización o cambio en los portales de empleo de computrabajo y linkedin, se valida la variable del dato correspondiente y se modifica la etiqueta del código html de los campos visualizados en el botón editar para su respectiva actualización [2] [3].

Interfaz Actualizar Etiquetas Script

Figura 12. Interfaz Actualizar Etiquetas Script.



Figura 12. Se visualiza la interfaz actualizar etiquetas script, el cual da una vista de la variable y la etiqueta que va actualizar el código.

5. PRUEBAS DEL SISTEMA

En este capítulo se presenta el proceso realizado con el Coordinador De Investigaciones de la Escuela de Contaduría pública de la Universidad del Sinú de Cartagena, quien es el cliente final. Este proceso es utilizado para validar y verificar el grado de cumplimiento de requerimientos levantados inicialmente para el desarrollo de la plataforma y así garantizar la calidad del producto.

Se realizaron dos tipos de pruebas: de interfaz gráfica y de funcionalidad.

5.1. Pruebas de interfaz gráfica para requerimientos Web

Las pruebas realizadas para cada una de las interfaces de la plataforma están basadas en el diseño de la *Tabla 18*. Lista de chequeo para pruebas de interfaz.

Figura 13. Gestionar Empleos.

VACANTE	EMPRESA	DESCRIPCION	MUNICIPIO	FECHA	PORTAL	
promotora de ventas - sector financiero	T&S TEMSERVICE S.A.S	Importante empresa bancaria solicita promotora de ventas, de libranza y en general de créditos, excelente presentación...	Cartagena de Indias	Hoy, 02:40 p. m.	Computrabajo	Aplicar
Asesor comercial Puntos de ventas		Importante multinacional requiere para su equipo de trabajo , candidatos con experiencia en servicio al...	promotora de ventas - sector financiero	Hoy, 02:45 p. m.	Computrabajo	Aplicar
Asesor De Repuestos - Motos		Se requiere tecnólogo en carreras de ingeniería o mecánica comerciales con experiencia mínima de 1 año en ventas...	Asesor comercial Puntos de ventas	Hoy, 02:45 p. m.	Computrabajo	Aplicar

Tabla 17. Lista de chequeo para pruebas de interfaz, Fuente Autor

PRUEBA DE INTERFAZ GRÁFICA PARA REQUERIMIENTOS WEB			
Seleccione la opción que aplique según el ítem			
Escala de calificación:			
Conforme	No Conforme	Parcialmente	No aplica
Sección	Calificación	Observación	
1. Interfaz General o Ventana Nueva			
Se visualiza los controles como activos cuando esto es necesario	Conforme		
En la barra de título aparece el nombre representativo de la interfaz	Conforme		
Primera letra en mayúsculas y las demás en minúscula			
	Conforme		
3. Cuadros de Texto			
¿Los cuadros de texto están alineados correctamente?	Conforme		
¿La cantidad de caracteres que permite ingresar, es la cantidad estipulada en el requerimiento o en la tabla de base de datos?	Conforme		
El ancho es correspondiente con la información que digitará el usuario.	Parcialmente		
Las cajas de texto ubican al lado derecho de las etiquetas.	Conforme		
Entre las etiquetas y los cuadros de texto se deja un espacio prudente	Conforme		
4. Campos de Fecha			
El ancho del campo deber ser correspondiente con la información que se va a ingresar.	No aplica		
Se ubica al lado derecho de las etiquetas	No aplica		
El formato manejado es de dd/mm/yyyy	No aplica		
5. Etiquetas			
¿Al final de estas ahí dos puntos?	No aplica		

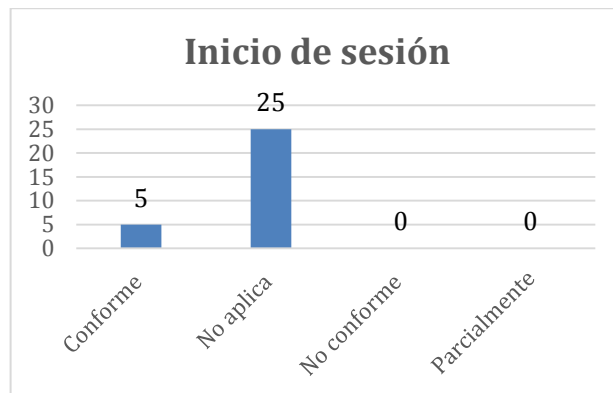
¿Las etiquetas están alineadas correctamente?	No aplica	
Las etiquetas están escritas con buena ortografía	No aplica	
El color de las etiquetas es legible	No aplica	
6. Combos		
La longitud del combo corresponde a la información a mostrar	No aplica	
Se ubican al lado derecho de las etiquetas	No aplica	
Los combos están alineados correctamente	No aplica	
7. Botones		
El nombre del botón primera letra en mayúscula y las demás en minúsculas	No Conforme	
Si tiene imagen este debe ir al lado izquierdo del texto	Conforme	
Si el botón tiene que abrir una nueva ventana lo hace correctamente	Conforme	
8. Grillas		
El tamaño de las grillas es el adecuado para la información que se desea mostrar	Conforme	
La repartición de las celdas que están dentro de la grilla está bien distribuida.	Conforme	
Los botones que se encuentran dentro de las grillas cumplen con la función necesaria.	No aplica	
9. Mensajes de Información		
¿Los mensajes brindan información conforme a lo requerido?	Conforme	
Se utiliza el icono correspondiente al momento de desplegar los mensajes de información	Conforme	
¿Se utilizan para alertar al usuario?	Conforme	
Se utilizan para brindar información cuando haya ocurrido un error	Parcialmente	
¿Al momento de brindar información cuando haya ocurrido un error, el mensaje mostrado es claro, explica dónde y porque se está generando el error o como arreglarlo?	Conforme	
¿Los mensajes desplegados están escritos con buena ortografía?	Conforme	

5.2. Resultados de las pruebas de Interfaz.

Requerimiento RF-01 Inicio de sesión.

Este requerimiento de inicio de sesión se obtuvieron resultado en no conforme y tampoco parcialmente, en no aplica 25 y conforme 5 si obtuvieron respuestas. Para mayor detalle revisar fase de pruebas Ver Anexo A.

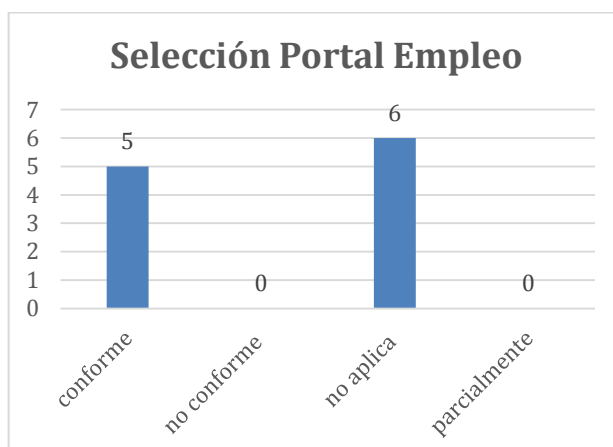
Figura 14. Resultados Prueba Interfaz RF- 01 Inicio de Sesion



Requerimiento RF-02 Selección Portal Empleo.

Este requerimiento no se obtuvo resultados en no conforme y parcialmente, en conforme 5 y no aplica en 6. Para mayor detalle revisar fase de pruebas Ver Anexo A.

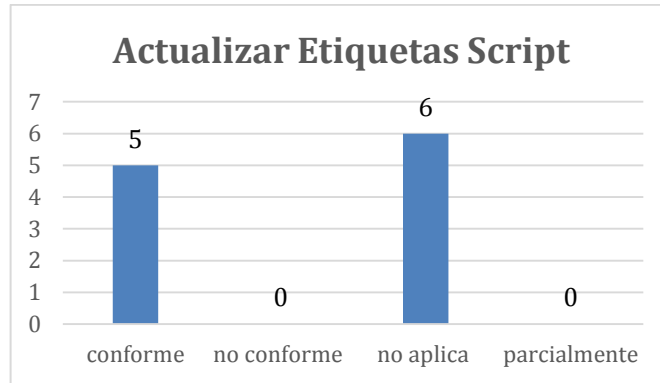
Figura 15. Resultados Prueba Interfaz RF- 02 Selección Portal Empleo.



Requerimiento RF-03 Actualizar Etiquetas Script

Este requerimiento no se obtuvo resultados en no conforme y parcialmente, en conforme 5 y en no aplica en 6. Para mayor detalle revisar fase de pruebas Ver Anexo A.

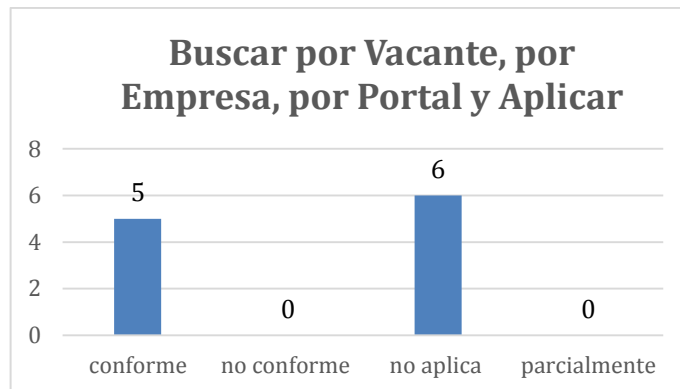
Figura 16. Resultados Prueba Interfaz RF- 03 Actualizar Etiquetas Script



Requerimiento RF-04 Buscar por Vacante, por Empresa, por Portal y Aplicar a la Vacante.

Este requerimiento no se obtuvo no conforme y tampoco parcialmente, en conforme 5 y en no aplica en 6. Para mayor detalle revisar fase de pruebas Ver Anexo A.

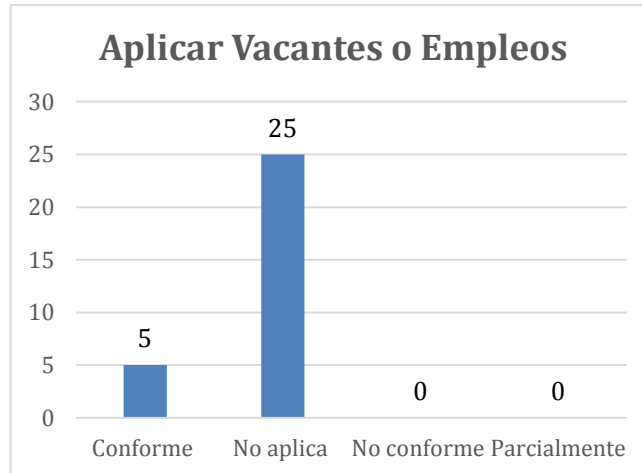
Figura 17. Resultados Prueba Interfaz RF- 04 Buscar por Vacante - Empresa y Portal.



Requerimiento RF-05 Aplicar Vacantes o Empleos

Este requerimiento no se obtuvo resultados en no conforme y tampoco parcialmente, conforme en 5 y en no aplica en 25. Para mayor detalle revisar fase de pruebas Ver Anexo A.

Figura 18. Resultados Prueba Interfaz RF- 05 Aplicar Vacantes o Empleos



5.3. Pruebas de funcionalidad para los requerimientos.

Esta prueba consta de 11 ítems en los que se revisan los requerimientos funcionales del prototipo, están basadas en el diseño de una lista de chequeo, detallada a continuación:

Tabla 18. Lista de chequeo para pruebas de Funcionalidad, Fuente Autor

PRUEBAS DE FUNCIONALIDAD PARA LOS REQUERIMIENTOS		
<i>Nombre del requerimiento</i>		
<i>Escala de Calificación</i>		
Conforme	No Conforme Parcialmente, No Aplica	
Ítems a Revisar	Calificación	Observación
Los botones funcionan adecuadamente.		

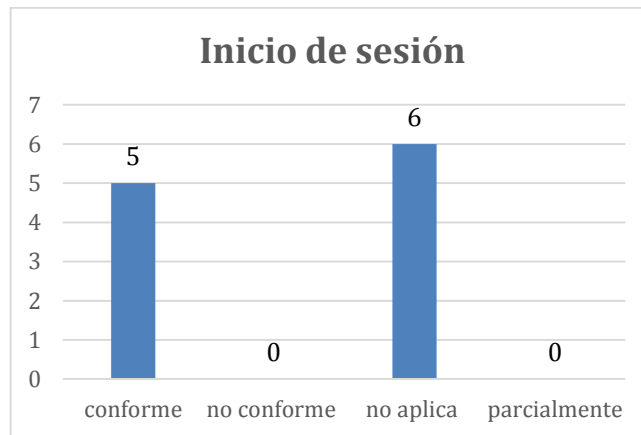
Los cuadros de Texto permiten ingresar solo el tipo de valor que está estipulado en el documento en donde se explica el requerimiento.		
Permite realizar el CRUD Correctamente.		
Las validaciones descritas en la documentación del requerimiento se están cumpliendo correctamente.		
Lo probado cumple con los requerimientos funcionales que están descritos en la documentación del requerimiento.		
Los combos cargan la información adecuada según el requerimiento.		
Los Check-Box cumplen con la función que les corresponde según la documentación del requerimiento		
Si los datos son de carácter obligatorio, se despliegan un mensaje de advertencia si el usuario no los ha digitado.		
Cuando se han ingresado datos incorrectos se muestra un mensaje de advertencia explicando que ha ocurrido un error.		
El requerimiento permite visualizar la ventana o el formulario con los datos precargados.		
Se muestra un mensaje cuando se ha cumplido una operación.		

5.4. Resultados de las pruebas de funcionalidad.

RF-01 Inicio de sesión.

En este requerimiento no se obtuvo resultados en no conforme y tampoco parcialmente, en conforme 5 y en no aplica en 6. Para mayor detalle Ver Anexo A.

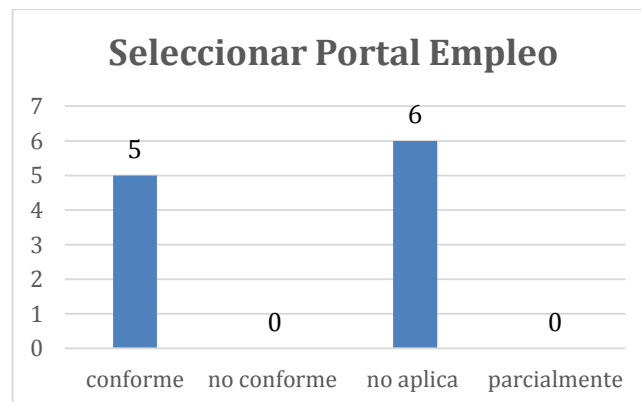
Figura 19. Resultados Prueba Funcionalidad RF- 01 Inicio de Sesión.



RF-02 Seleccionar Portal Empleo.

En este requerimiento no se obtuvo no conforme y tampoco parcialmente, en conforme 5 en no aplica en 6. Para mayor detalle ver Anexo A.

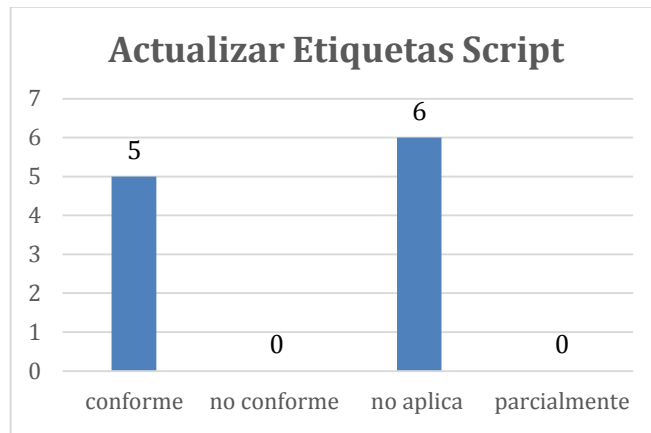
Figura 20. Resultados Prueba Funcionalidad RF- 02 Seleccionar Portal Empleo.



RF-03 Actualizar Etiquetas Script

En este requerimiento no se obtuvo no conforme y tampoco parcialmente, en conforme 5 en no aplica en 6. Para mayor detalle ver Anexo A.

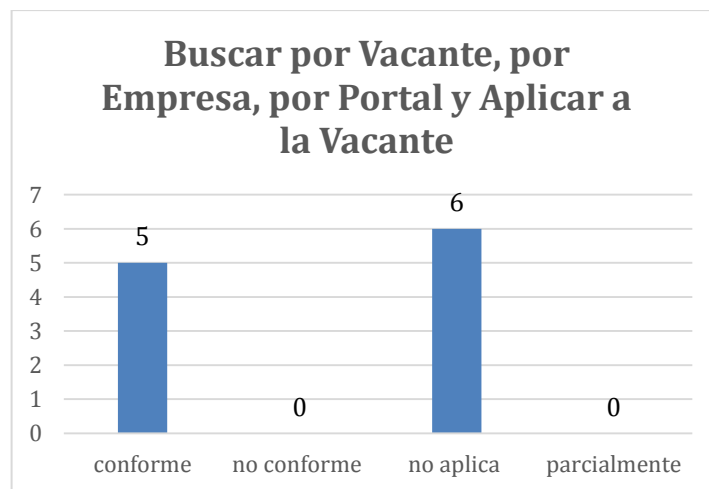
Figura 21. Resultados Prueba Funcionalidad RF- 03 Editar Etiquetas Script



RF-04 Buscar por Vacante, por Empresa, por Portal y Aplicar a la Vacante.

En este requerimiento no se obtuvo no conforme y tampoco parcialmente, en conforme 5 en no aplica en 6. Para mayor detalle ver Anexo A.

Figura 22. Resultados Prueba Funcionalidad RF- 04 Buscar por Vacante, Empresa y Portal.



CONCLUSIONES

Al finalizar este trabajo de investigación se logra construir un prototipo web que permite la visualización de las vacantes extraídas en los portales de empleo de (computrabajo y linkedin) por medio de la técnica de Web Scraping con el objetivo de centralizar la información en un solo portal web para la facilitación de los usuarios en la búsqueda de empleo en Bolívar.

Para llegar al cumplimiento de este objetivo se ejecutaron los objetivos específicos los cuales se realizaron de la siguiente manera:

- Para realizar el levantamiento de requerimientos funcionales y no funcionales de la aplicación web se hizo una investigación de la información a extraer de los distintos portales de empleo con la finalidad analizar y de obtener los requerimientos para la creación de prototipo web, los requisitos funcionales y no funcionales de acuerdo a la referencia de Standard IEEE 830 – 1998.
- Se realizó el diseño web atendiendo los requerimientos de los datos a extraer y haciendo uso de los diagramas adoptadas por la industria del software como diagramas de caso de usos modelo de datos lógico y se presenta el diseño de las interfaces graficas de la aplicación web.
- En esta capitulo se investigó las diversas técnicas y librerías para hacer Web Scraping en la cual se terminó eligiendo la librería más adecuada con respecto a facilidad y curva de aprendizaje corta para la realización de este proyecto teniendo así los resultados esperados en la extracción de información de los portales web elegidos.
- Para desarrollar este objetivo del desarrollo de la plataforma web se documentó la arquitectura del sistema, donde se explica las tecnologías utilizadas, su integración y se detallan cada una de las interfaces que conforman la aplicación y su funcionamiento.

- Finalmente se describen los resultados de las pruebas de interfaz y funcionales del sistema lo cual permitió validar los requerimientos identificados en la fase de análisis.

RECOMENDACIONES

Dentro de las pautas en este proyecto que se brindan para dar los lineamientos en la comprensión y entendimiento del tema al desarrollar un prototipo de aplicación web que permita la extracción de las ofertas laborales de las principales plataformas que postulan empleos en la región caribe, usando la técnica Web Scraping; entre las recomendaciones que realiza el autor de este proyecto Portal De Empleo para su mejora están:

- Módulo registro de usuarios, que permita inicio de sesión de usuarios registrados.
- Módulo de estadística, que visualice las vacantes más consultadas en el portal de empleo por los usuarios.
- Mejorar todas las interfaces gráficas de la aplicación implementando estándares de usabilidad y de accesibilidad web, para mejorar la interacción con los usuarios.
- Implementación del portal de empleo en red pública internet que permita el acceso a usuarios desde cualquier lugar.
- Agregar más plataformas de empleo para el prototipo portal de empleo utilizando la técnica Web Scraping.
- Implementar Mensaje de notificación al correo de las vacantes seleccionadas y aplicadas por el usuario.
- Que se ejecuten los script en un determinado tiempo al extraer e insertar los datos en la tabla.
- Que el prototipo se implemente en una red pública de internet (hosting web).

BIBLIOGRAFÍA

- [1] "Web Scraping, el gran aliado de la innovación analítica", *Computing*, 2018. [Online]. Available: <http://www.computing.es/analytics/noticias/1103589046201/web-scraping-gran-aliado-de-innovacion-analitica.1.html>.
- [2] "Bolívar Mayo 2019 | Computrabajo Colombia", *Computrabajo.com.co*, 2019. [Online]. Available: <https://www.computrabajo.com.co/empleos-en-bolivar>.
- [3] "Empleos en Bolívar, Colombia", *Linkedin EMPLEOS*, 2019. [Online]. Available: <https://co.linkedin.com/jobs/search?keywords=&location=Bol%C3%ADvar%2C%20Colombia&geo=105189166&pageNum=0&position=1>.
- [4] Y. LOPEZ POVEDA, "Análisis y diseño de un sistema para la extracción, análisis y comparación de precios de tiendas en la ciudad de Bogotá d.c.", *Repository.udistrital.edu.co*. Universidad Distrital Francisco José De Caldas. 2017.
- [5] C. Hernández Herrero, "Aplicación de Técnicas de Web Scraping al Boletín Oficial de Castilla y León (BOCyL)", *Uvadoc.uva.es*, 2014. [Online]. Available: <http://uvadoc.uva.es/handle/10324/5794>.
- [6] A. MICAELLO, "Extracción de información a partir de recursos web semi-estructurados - PDF", *Docplayer.es*, 2009. [Online]. Available: <https://docplayer.es/4561115-Extraccion-de-informacion-a-partir-de-recursos-web-semi-estructurados.html>.
- [7] A. CARRATO and A. MENENDEZ, "Proyecto Final. Recuperación de Información y Recomendaciones en la Web. Ana Carrato - C.I: Ana Menéndez - C. - PDF", *Docplayer.es*, 2014. [Online]. Available: <http://docplayer.es/19130060-Proyecto-final-recuperacion-de-informacion-y-recomendaciones-en-la-web-ana-carrato-c-i-4247054-2-ana-menendez-c.html>.
- [8] A. JIMENEZ and S. SORIANO, "Software Asesor en la compra de computadores", *Repository.javeriana.edu.co*, Pontificia Universidad Javeriana. BOGOTÁ, D.C 2013.
- [9] SALES ADSUARA, "Datstrats", *Datstrats.com*, 2019. [Online]. Available: <https://datstrats.com/blog/que-es-scraping-usos-aplicaciones>.
- [10] J. PRIETO, "python - GTI - Glosario Terminología Informática", *Tugurium.com*, 2014. [Online]. Available: <http://www.tugurium.com/gti/termino.php?Tr=python>.
- [11] "¿Qué es HTML y Para Qué Sirve? | Next University", *NextU LATAM*, 2019. [Online].

Available: <https://www.nextu.com/blog/que-es-html/>.

- [12] "Herramientas de extracción de datos: para principiantes y profesionales", BBVAOpen4U, 2016. [Online]. Available: <https://bbvaopen4u.com/es/actualidad/herramientas-de-extraccion-de-datos-para-principiantes-y-profesionales>.
- [13] "Biblioteca (informática)", *Es.wikipedia.org*, 2019. [Online]. Available: [https://es.wikipedia.org/wiki/Biblioteca_\(inform%C3%A1tica\)](https://es.wikipedia.org/wiki/Biblioteca_(inform%C3%A1tica)).
- [14] "Anaconda (Python distribution)", *En.wikipedia.org*, 2019. [Online]. Available: [https://en.wikipedia.org/wiki/Anaconda_\(Python_distribution\)](https://en.wikipedia.org/wiki/Anaconda_(Python_distribution)).
- [15] "Mechanize", PyPI, 2019. [Online]. Available: <https://pypi.org/project/mechanize/>.
- [16] "SSL", PyPI, 2019. [Online]. Available: <https://pypi.org/project/ssl/>.
- [17] "csv — CSV File Reading and Writing — Python 3.7.3 documentation", *Docs.python.org*, 2019. [Online]. Available: <https://docs.python.org/3/library/csv.html?highlight=csv#module-csv>.
- [18] Y. Borja López. (2017, Ago. 15) "Análisis y Diseño de un sistema para la Extracción, análisis y comparación de precios de tiendas en la ciudad de Bogotá D.C". [Online]. Available: <http://repository.udistrital.edu.co/bitstream/11349/6853/2/LopezPovedaYair2017.pdf>
- [19] D. HONORES CORREA, "Utilización de diagramas de Gantt para la programación y control de un proyecto en la empresa "DH".", *Repositorio.utmachala.edu.ec*, 2017. [Online]. Available: <http://repositorio.utmachala.edu.ec/handle/48000/10013>.
- [20] G. Fernández, "¿Sabes cómo funciona XP? - Management Plaza", *Management Plaza*, 2002. [Online]. Available: <https://managementplaza.es/blog/sabes-como-funciona-xp/>.
- [21] "Programación por capas", *Es.wikipedia.org*, 2018. [Online]. Available: https://es.wikipedia.org/wiki/Programaci%C3%B3n_por_capas.
- [22] F. Garau, "Jurisprudencia comentada - El denominado "Screen Scraping"", *Conflictuslegum.blogspot.com*, 2013. [Online]. Available: <http://conflictuslegum.blogspot.com/2013/02/jurisprudencia-comentada-el-denominado.html>.

ANEXOS

ANEXO A. Pruebas plataforma Web

Pruebas de interfaz gráfica para requerimientos web

Tabla 19. Prueba Interfaz inicio de sesión Administrador.

INTERFAZ DE INICIO DE SESIÓN ADMINISTRADOR			
Seleccione la opción que aplique según el ítem			
<i>Escala de calificación:</i>			
Conforme	No Conforme	Parcialmente	No aplica
<i>Sección</i>	<i>Calificación</i>	<i>Observación</i>	
1. Interfaz General o Ventana Nueva			
Se visualiza los controles como activos cuando esto es necesario	Conforme		
En la barra de título aparece el nombre representativo de la interfaz	Conforme		
Primera letra en mayúsculas y las demás en minúscula	Conforme		
3. Cuadros de Texto			
¿Los cuadros de texto están alineados correctamente?	Conforme		
¿La cantidad de caracteres que permite ingresar, es la cantidad estipulada en el requerimiento o en la tabla de base de datos?	Conforme		
El ancho es correspondiente con la información que digitará el usuario.	Parcialmente		
Las cajas de texto ubican al lado derecho de las etiquetas.	Conforme		
Entre las etiquetas y los cuadros de texto se deja un espacio prudente	Conforme		
4. Campos de Fecha			
El ancho del campo deber ser correspondiente con la información que se va a ingresar.	No aplica		
Se ubica al lado derecho de las etiquetas	No aplica		
El formato manejado es de dd/mm/yyyy	No aplica		
5. Etiquetas			
¿Al final de estas ahí dos puntos?	No aplica		
¿Las etiquetas están alineadas correctamente?	No aplica		
Las etiquetas están escritas con buena ortografía	No aplica		
El color de las etiquetas es legible	No aplica		
6. Combos			
La longitud del combo corresponde a la información a mostrar	No aplica		

Se ubican al lado derecho de las etiquetas	No aplica	
Los combos están alineados correctamente	No aplica	
7. Botones		
El nombre del botón primera letra en mayúscula y las demás en minúsculas	No Conforme	
Si tiene imagen este debe ir al lado izquierdo del texto	Conforme	
Si el botón tiene que abrir una nueva ventana lo hace correctamente	Conforme	
8. Grillas		
El tamaño de las grillas es el adecuado para la información que se desea mostrar	Conforme	
La repartición de las celdas que están dentro de la grilla está bien distribuidas.	Conforme	
Los botones que se encuentran dentro de las grillas cumplen con la función necesaria.	No aplica	
9. Mensajes de Información		
¿Los mensajes brindan información conforme a lo requerido?	Conforme	
Se utiliza el icono correspondiente al momento de desplegar los mensajes de información	Conforme	
¿Se utilizan para alertar al usuario?	Conforme	
Se utilizan para brindar información cuando haya ocurrido un error	Parcialmente	
¿Al momento de brindar información cuando haya ocurrido un error, el mensaje mostrado es claro, explica dónde y porque se está generando el error o como arreglarlo?	Conforme	
¿Los mensajes desplegados están escritos con buena ortografía?	Conforme	

Tabla 20. Prueba Interfaz Seleccionar Portal Empleo.

INTERFAZ SELECCIONAR PORTAL EMPLEO			
Seleccione la opción que aplique según el ítem			
Escala de calificación:			
Conforme	No Conforme	Parcialmente	No aplica
Sección	Calificación	Observación	
1. Interfaz General o Ventana Nueva			
Se visualiza los controles como activos cuando esto es necesario	Conforme		
En la barra de título aparece el nombre representativo de la interfaz	Conforme		
Primera letra en mayúsculas y las demás en minúscula	Conforme		
3. Cuadros de Texto			
¿Los cuadros de texto están alineados correctamente?	Conforme		

¿La cantidad de caracteres que permite ingresar, es la cantidad estipulada en el requerimiento o en la tabla de base de datos?	Conforme	
El ancho es correspondiente con la información que digitará el usuario.	Parcialmente	
Las cajas de texto ubican al lado derecho de las etiquetas.	Conforme	
Entre las etiquetas y los cuadros de texto se deja un espacio prudente	Conforme	
4. Campos de Fecha		
El ancho del campo deber ser correspondiente con la información que se va a ingresar.	No aplica	
Se ubica al lado derecho de las etiquetas	No aplica	
El formato manejado es de dd/mm/yyyy	No aplica	
5. Etiquetas		
¿Al final de estas ahí dos puntos?	No aplica	
¿Las etiquetas están alineadas correctamente?	No aplica	
Las etiquetas están escritas con buena ortografía	No aplica	
El color de las etiquetas es legible	No aplica	
6. Combos		
La longitud del combo corresponde a la información a mostrar	No aplica	
Se ubican al lado derecho de las etiquetas	No aplica	
Los combos están alineados correctamente	No aplica	
7. Botones		
El nombre del botón primera letra en mayúscula y las demás en minúsculas	No Conforme	
Si tiene imagen este debe ir al lado izquierdo del texto	Conforme	
Si el botón tiene que abrir una nueva ventana lo hace correctamente	Conforme	
8. Grillas		
El tamaño de las grillas es el adecuado para la información que se desea mostrar	Conforme	
La repartición de las celdas que están dentro de la grilla está bien distribuidas.	Conforme	
Los botones que se encuentran dentro de las grillas cumplen con la función necesaria.	No aplica	
9. Mensajes de Información		
¿Los mensajes brindan información conforme a lo requerido?	Conforme	
Se utiliza el icono correspondiente al momento de desplegar los mensajes de información	Conforme	
¿Se utilizan para alertar al usuario?	Conforme	
Se utilizan para brindar información cuando haya ocurrido un error	Parcialmente	
¿Al momento de brindar información cuando haya ocurrido un error, el mensaje mostrado es claro, explica dónde y porque se está generando el error o como arreglarlo?	Conforme	
¿Los mensajes desplegados están escritos con buena ortografía?	Conforme	

Tabla 21. Prueba Interfaz Actualizar Etiquetas Script.

INTERFAZ ACTUALIZAR ETIQUETAS SCRIPT			
Seleccione la opción que aplique según el ítem			
Escala de calificación:			
Conforme	No Conforme	Parcialmente	No aplica
Sección	Calificación	Observación	
1. Interfaz General o Ventana Nueva			
Se visualiza los controles como activos cuando esto es necesario	Conforme		
En la barra de título aparece el nombre representativo de la interfaz	Conforme		
Primera letra en mayúsculas y las demás en minúscula	Conforme		
3. Cuadros de Texto			
¿Los cuadros de texto están alineados correctamente?	Conforme		
¿La cantidad de caracteres que permite ingresar, es la cantidad estipulada en el requerimiento o en la tabla de base de datos?	Conforme		
El ancho es correspondiente con la información que digitará el usuario.	Parcialmente		
Las cajas de texto ubican al lado derecho de las etiquetas.	Conforme		
Entre las etiquetas y los cuadros de texto se deja un espacio prudente	Conforme		
4. Campos de Fecha			
El ancho del campo deber ser correspondiente con la información que se va a ingresar.	No aplica		
Se ubica al lado derecho de las etiquetas	No aplica		
El formato manejado es de dd/mm/yyyy	No aplica		
5. Etiquetas			
¿Al final de estas ahí dos puntos?	No aplica		
¿Las etiquetas están alineadas correctamente?	No aplica		
Las etiquetas están escritas con buena ortografía	No aplica		
El color de las etiquetas es legible	No aplica		
6. Combos			
La longitud del combo corresponde a la información a mostrar	No aplica		
Se ubican al lado derecho de las etiquetas	No aplica		
Los combos están alineados correctamente	No aplica		
7. Botones			
El nombre del botón primera letra en mayúscula y las demás en minúsculas	No Conforme		
Si tiene imagen este debe ir al lado izquierdo del texto	Conforme		
Si el botón tiene que abrir una nueva ventana lo hace correctamente	Conforme		

8.Grillas		
El tamaño de las grillas es el adecuado para la información que se desea mostrar	Conforme	
La repartición de las celdas que están dentro de la grilla está bien distribuidas.	Conforme	
Los botones que se encuentran dentro de las grillas cumplen con la función necesaria.	No aplica	
9.Mensajes de Información		
¿Los mensajes brindan información conforme a lo requerido?	Conforme	
Se utiliza el icono correspondiente al momento de desplegar los mensajes de información	Conforme	
¿Se utilizan para alertar al usuario?	Conforme	
Se utilizan para brindar información cuando haya ocurrido un error	Parcialmente	
¿Al momento de brindar información cuando haya ocurrido un error, el mensaje mostrado es claro, explica dónde y porque se está generando el error o como arreglarlo?	Conforme	
¿Los mensajes desplegados están escritos con buena ortografía?	Conforme	

Tabla 22. Prueba Interfaz Gestionar Empleos.

INTERFAZ GESTIONAR EMPLEOS			
Seleccione la opción que aplique según el ítem			
Escala de calificación:			
Conforme	No Conforme	Parcialmente	No aplica
Sección	Calificación	Observación	
1. Interfaz General o Ventana Nueva			
Se visualiza los controles como activos cuando esto es necesario	Conforme		
En la barra de título aparece el nombre representativo de la interfaz	Conforme		
Primera letra en mayúsculas y las demás en minúscula	Conforme		
3. Cuadros de Texto			
¿Los cuadros de texto están alineados correctamente?	Conforme		
¿La cantidad de caracteres que permite ingresar, es la cantidad estipulada en el requerimiento o en la tabla de base de datos?	Conforme		
El ancho es correspondiente con la información que digitará el usuario.	Parcialmente		
Las cajas de texto ubican al lado derecho de las etiquetas.	Conforme		
Entre las etiquetas y los cuadros de texto se deja un espacio prudente	Conforme		
4. Campos de Fecha			
El ancho del campo deber ser correspondiente con la información que se va a ingresar.	No aplica		

Se ubica al lado derecho de las etiquetas	No aplica	
El formato manejado es de dd/mm/yyyy	No aplica	
5. Etiquetas		
¿Al final de estas ahí dos puntos?	No aplica	
¿Las etiquetas están alineadas correctamente?	No aplica	
Las etiquetas están escritas con buena ortografía	No aplica	
El color de las etiquetas es legible	No aplica	
6. Combos		
La longitud del combo corresponde a la información a mostrar	No aplica	
Se ubican al lado derecho de las etiquetas	No aplica	
Los combos están alineados correctamente	No aplica	
7. Botones		
El nombre del botón primera letra en mayúscula y las demás en minúsculas	No Conforme	
Si tiene imagen este debe ir al lado izquierdo del texto	Conforme	
Si el botón tiene que abrir una nueva ventana lo hace correctamente	Conforme	
8. Grillas		
El tamaño de las grillas es el adecuado para la información que se desea mostrar	Conforme	
La repartición de las celdas que están dentro de la grilla está bien distribuida.	Conforme	
Los botones que se encuentran dentro de las grillas cumplen con la función necesaria.	No aplica	
9. Mensajes de Información		
¿Los mensajes brindan información conforme a lo requerido?	Conforme	
Se utiliza el icono correspondiente al momento de desplegar los mensajes de información	Conforme	
¿Se utilizan para alertar al usuario?	Conforme	
Se utilizan para brindar información cuando haya ocurrido un error	Parcialmente	
¿Al momento de brindar información cuando haya ocurrido un error, el mensaje mostrado es claro, explica dónde y porque se está generando el error o como arreglarlo?	Conforme	
¿Los mensajes desplegados están escritos con buena ortografía?	Conforme	

Pruebas de funcionalidad para los requerimientos

Tabla 23. Prueba de funcionalidad inicio sesión.

PRUEBAS DE FUNCIONALIDAD PARA LOS REQUERIMIENTOS		
RF-01 Inicio de sesión		
Escala de Calificación		
Conforme	No Conforme Parcialmente, No Aplica	
Ítems a Revisar	Calificación	Observación
Los botones funcionan adecuadamente.	Conforme	
Los cuadros de Texto permiten ingresar solo el tipo de valor que está estipulado en el documento en donde se explica el requerimiento.	Conforme	
Permite realizar el CRUD Correctamente.	No aplica	
Las validaciones descritas en la documentación del requerimiento se están cumpliendo correctamente.	Conforme	
Lo probado cumple con los requerimientos funcionales que están descritos en la documentación del requerimiento.	Conforme	
Los combos cargan la información adecuada según el requerimiento.	No aplica	
Los Check-Box cumplen con la función que les corresponde según la documentación del requerimiento	No aplica	
Si los datos son de carácter obligatorio, se despliegan un mensaje de advertencia si el usuario no los ha digitado.	Conforme	
Cuando se han ingresado datos incorrectos se muestra un mensaje de advertencia explicando que ha ocurrido un error.	Conforme	
El requerimiento permite visualizar la ventana o el formulario con los datos precargados.	Conforme	
Se muestra un mensaje cuando se ha cumplido una operación.	Conforme	

Tabla 24. Prueba de funcionalidad Interfaz Seleccionar Empleo.

PRUEBAS DE FUNCIONALIDAD PARA LOS REQUERIMIENTOS		
<i>RF-02 Seleccionar Portal Empleo</i>		
Escala de Calificación		
Conforme	No Conforme Parcialmente, No Aplica	
Ítems a Revisar	Calificación	Observación
Los botones funcionan adecuadamente.	Conforme	
Los cuadros de Texto permiten ingresar solo el tipo de valor que está estipulado en el documento en donde se explica el requerimiento.	Conforme	
Permite realizar el CRUD Correctamente.	Conforme	
Las validaciones descritas en la documentación del requerimiento se están cumpliendo correctamente.	Conforme	
Lo probado cumple con los requerimientos funcionales que están descritos en la documentación del requerimiento.	Conforme	
Los combos cargan la información adecuada según el requerimiento.	Conforme	
Los Check-Box cumplen con la función que les corresponde según la documentación del requerimiento	No aplica	
Si los datos son de carácter obligatorio, se despliegan un mensaje de advertencia si el usuario no los ha digitado.	Conforme	
Cuando se han ingresado datos incorrectos se muestra un mensaje de advertencia explicando que ha ocurrido un error.	Conforme	
El requerimiento permite visualizar la ventana o el formulario con los datos precargados.	Conforme	
Se muestra un mensaje cuando se ha cumplido una operación.	Conforme	

Tabla 25. Prueba de funcionalidad Interfaz Actualizar Etiquetas Script.

PRUEBAS DE FUNCIONALIDAD PARA LOS REQUERIMIENTOS		
<i>RF-03 Actualizar Etiquetas Script</i>		
Escala de Calificación		
Conforme	No Conforme Parcialmente, No Aplica	
Ítems a Revisar	Calificación	Observación
Los botones funcionan adecuadamente.	Conforme	
Los cuadros de Texto permiten ingresar solo el tipo de valor que está estipulado en el documento en donde se explica el requerimiento.	Conforme	
Permite realizar el CRUD Correctamente.	Conforme	
Las validaciones descritas en la documentación del requerimiento se están cumpliendo correctamente.	Conforme	
Lo probado cumple con los requerimientos funcionales que están descritos en la documentación del requerimiento.	Conforme	
Los combos cargan la información adecuada según el requerimiento.	Conforme	
Los Check-Box cumplen con la función que les corresponde según la documentación del requerimiento	No aplica	
Si los datos son de carácter obligatorio, se despliegan un mensaje de advertencia si el usuario no los ha digitado.	Conforme	
Cuando se han ingresado datos incorrectos se muestra un mensaje de advertencia explicando que ha ocurrido un error.	Conforme	
El requerimiento permite visualizar la ventana o el formulario con los datos precargados.	Conforme	
Se muestra un mensaje cuando se ha cumplido una operación.	Conforme	

Tabla 26. Prueba de funcionalidad Interfaz Gestionar Empleos.

PRUEBAS DE FUNCIONALIDAD PARA LOS REQUERIMIENTOS		
<i>RF-04 Buscar por Vacante, por Empresa, por Portal y Aplicar a la Vacante.</i>		
Escala de Calificación		
Conforme	No Conforme Parcialmente, No Aplica	
Ítems a Revisar	Calificación	Observación
Los botones funcionan adecuadamente.	Conforme	
Los cuadros de Texto permiten ingresar solo el tipo de valor que está estipulado en el documento en donde se explica el requerimiento.	Conforme	
Permite realizar el CRUD Correctamente.	Conforme	
Las validaciones descritas en la documentación del requerimiento se están cumpliendo correctamente.	Conforme	
Lo probado cumple con los requerimientos funcionales que están descritos en la documentación del requerimiento.	Conforme	
Los combos cargan la información adecuada según el requerimiento.	Conforme	
Los Check-Box cumplen con la función que les corresponde según la documentación del requerimiento	No aplica	
Si los datos son de carácter obligatorio, se despliegan un mensaje de advertencia si el usuario no los ha digitado.	Conforme	
Cuando se han ingresado datos incorrectos se muestra un mensaje de advertencia explicando que ha ocurrido un error.	Conforme	
El requerimiento permite visualizar la ventana o el formulario con los datos precargados.	Conforme	
Se muestra un mensaje cuando se ha cumplido una operación.	Conforme	

ANEXO B. Diccionario de datos

Un diccionario de datos es un diagrama o tabla que contiene todas las características lógicas de los datos de una tabla. A continuación, se muestran las descripciones de las tablas creadas en el sistema.

Tabla: scriptcomputrabajo

Descripción: Tabla donde se encuentra el script del portal computrabajo.

Llave Primaria: id

Estructuctura

#	Nombre	Tipo	Cotejamiento	Atributos	Nulo	Predeterminado	Comentarios	Extra	Acción
<input type="checkbox"/>	1 ID	varchar(10)	latin1_swedish_ci		No	Ninguna			Cambiar Eliminar Más
<input type="checkbox"/>	2 VARIABLE	varchar(105)	latin1_swedish_ci		Sí	NULL			Cambiar Eliminar Más
<input type="checkbox"/>	3 ETIQUETA	varchar(4700)	latin1_swedish_ci		Sí	NULL			Cambiar Eliminar Más

Seleccionar todo Para los elementos que están marcados: Examinar Cambiar Eliminar Primaria Único Agregar a columnas centrales Eliminar de las columnas centrales

+ Opciones

			ID	VARIABLE	ETIQUETA
<input type="checkbox"/>	Editar Copiar Borrar		1	primera parte	from bs4 import BeautifulSoup import requests im...
<input type="checkbox"/>	Editar Copiar Borrar		2	vacante = entrada.find	('h3', {'class': 'listed-job-posting__title'}).get...
<input type="checkbox"/>	Editar Copiar Borrar		3	empresa = entrada.find	('h4', {'class': 'listed-job-posting__company'}).g...
<input type="checkbox"/>	Editar Copiar Borrar		4	descripcion = entrada.find	('p', {'class': 'listed-job-posting__description'})...
<input type="checkbox"/>	Editar Copiar Borrar		5	municipio = entrada.find	('p').getText()
<input type="checkbox"/>	Editar Copiar Borrar		6	fecha = entrada.find	entrada.find('div', {'class': 'listed-job-posting_...
<input type="checkbox"/>	Editar Copiar Borrar		7	url_vacante = entrada.find	('a')['href']
<input type="checkbox"/>	Editar Copiar Borrar		8	segunda parte	portal_web = ("linkedin") ...

Tabla: scriptlinkedin

Descripción: Tabla donde se encuentra el script del portal linkedin.

Llave Primaria: id

Estructura

#	Nombre	Tipo	Cotejamiento	Atributos	Nulo	Predeterminado	Comentarios	Extra	Acción
<input type="checkbox"/>	1 ID	varchar(10)	latin1_swedish_ci		No	Ninguna			Cambiar Eliminar Más
<input type="checkbox"/>	2 VARIABLE	varchar(105)	latin1_swedish_ci		Sí	NULL			Cambiar Eliminar Más
<input type="checkbox"/>	3 ETIQUETA	varchar(4700)	latin1_swedish_ci		Sí	NULL			Cambiar Eliminar Más

Seleccionar todo *Para los elementos que están marcados:* Examinar Cambiar Eliminar Primaria Único

Agregar a columnas centrales Eliminar de las columnas centrales

+ Opciones

				ID	VARIABLE	ETIQUETA
<input type="checkbox"/>	Editar	Copiar	Borrar	1	primera parte	from bs4 import BeautifulSoup import requests im...
<input type="checkbox"/>	Editar	Copiar	Borrar	2	vacante = entrada.find	('h3', {class: 'listed-job-posting__title'}).get...
<input type="checkbox"/>	Editar	Copiar	Borrar	3	empresa = entrada.find	('h4', {class: 'listed-job-posting__company'}).g...
<input type="checkbox"/>	Editar	Copiar	Borrar	4	descripcion = entrada.find	('p', {class: 'listed-job-posting__description'})...
<input type="checkbox"/>	Editar	Copiar	Borrar	5	municipio = entrada.find	('p').getText()
<input type="checkbox"/>	Editar	Copiar	Borrar	6	fecha = entrada.find	entrada.find('div', {class: 'listed-job-posting_...
<input type="checkbox"/>	Editar	Copiar	Borrar	7	url_vacante = entrada.find	('a')['href']
<input type="checkbox"/>	Editar	Copiar	Borrar	8	segunda parte	portal_web = ("linkedin")

Tabla: Vacantes

Descripción: Tabla donde se insertaran los datos extraídos del portal Computrabajo por el script Web Scraping.

Llave Primaria: vacante

#	Nombre	Tipo	Cotejamiento	Atributos	Nulo	Predeterminado	Comentarios	Extra	Acción
<input type="checkbox"/>	1 VACANTE	varchar(45)	latin1_swedish_ci		Sí	NULL			Cambiar Eliminar Más
<input type="checkbox"/>	2 EMPRESA	varchar(45)	latin1_swedish_ci		Sí	NULL			Cambiar Eliminar Más
<input type="checkbox"/>	3 DESCRIPCION	text	latin1_swedish_ci		Sí	NULL			Cambiar Eliminar Más
<input type="checkbox"/>	4 MUNICIPIO	varchar(45)	latin1_swedish_ci		Sí	NULL			Cambiar Eliminar Más
<input type="checkbox"/>	5 FECHA	varchar(45)	latin1_swedish_ci		Sí	NULL			Cambiar Eliminar Más
<input type="checkbox"/>	6 URL_VACANTE	text	latin1_swedish_ci		Sí	NULL			Cambiar Eliminar Más
<input type="checkbox"/>	7 PORTAL_WEB	varchar(45)	latin1_swedish_ci		Sí	NULL			Cambiar Eliminar Más

Tabla: members

Descripción: Tabla donde estarán los usuarios administradores

Llave Primaria: id

#	Nombre	Tipo	Cotejamiento	Atributos	Nulo	Predeterminado	Comentarios	Extra	Acción
<input type="checkbox"/>	1 id	int(10)		UNSIGNED	No	Ninguna		AUTO_INCREMENT	Cambiar Eliminar Más
<input type="checkbox"/>	2 first_name	varchar(45)	latin1_swedish_ci		No	Ninguna			Cambiar Eliminar Más
<input type="checkbox"/>	3 last_name	varchar(45)	latin1_swedish_ci		No	Ninguna			Cambiar Eliminar Más
<input type="checkbox"/>	4 email	varchar(45)	latin1_swedish_ci		No	Ninguna			Cambiar Eliminar Más
<input type="checkbox"/>	5 uname	varchar(45)	latin1_swedish_ci		No	Ninguna			Cambiar Eliminar Más
<input type="checkbox"/>	6 pass	varchar(45)	latin1_swedish_ci		No	Ninguna			Cambiar Eliminar Más

Seleccionar todo *Para los elementos que están marcados:* Examinar Cambiar Eliminar Primaria Único Índice

Diagrama de Flujo Prototipo Web

Representa todos los procesos que realiza en el sistema tanto los actores del administrador como los usuarios que interactúan con el prototipo web.

