

Desarrollo de una Herramienta Software para Generación, Actualización y Consulta de Hojas  
de Vida de Investigadores Mediante la Obtención de Datos.

Johan Sebastián Duran Caro  
Daniel Rojas Cáceres

Trabajo de Grado para optar al título de Ingeniero de Sistemas e Informática.

Director  
PhD. SONIA CRISTINA GAMBOA SARMIENTO  
Doctora en Educación.

UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERÍAS FÍSICO MECÁNICAS  
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA  
BUCARAMANGA  
2019

## Agradecimientos

*Agradezco enormemente a todas las personas que estuvieron apoyándome en este proceso, a mis padres, a mis amigos, a mi novia. Pero sobre todo mi directora de proyecto la PhD. Sonia Cristina Gamboa, ya que sin su guía y paciencia no se hubiese podido completar este proyecto, agradezco también a mi alma mater la Universidad Industrial de Santander ya que gracias a la UIS soy quien soy ahora.*  
*Daniel Rojas C.*

*Agradezco a mis padres y hermanos que estuvieron apoyándome cada día para permitirme ser quien soy actualmente. A mis profesores y colegas cuyos consejos y sabiduría me permitieron alcanzar esta meta. Finalmente agradezco a todos mis amigos; si no hubiera sido por ellos, me hubiese graduado hace 5 años.*  
*Johan Sebastián Durán Caro.*

“El éxito es la capacidad de ir de fracaso en fracaso sin perder el entusiasmo.”

Winston Churchill

## Dedicatoria

Dedico este proyecto a todas las personas que confiaron en mí, a todas la que siempre creyeron que podría lograrlo, se lo dedico a mis padres gracias a sus valores que me inculcaron me enseñaron a no rendirme, este logro es tanto mío como de ustedes a mi querida madre Julia Cáceres y mi Padre Wilson Rojas.

Se lo dedico a Dios porque sin su guía posiblemente no hubiera aguantado el camino, ya que ante todo él es quien guía mis pasos.

A mis amigos también, porque ellos me dieron ánimos cuando la esperanza se desvanecía.

Se lo dedico también a esa persona maravillosa que Dios me puso en el camino y que a pesar de todos los tropiezos sigue a mi lado, dándome esperanzas y motivación para seguir adelante, a mi querida novia.

Dedico este proyecto a todas las personas que siempre me preguntaron ¿Ya te graduaste?, ¿Pa' cuando el grado?, ¿Puedo activar mi carta trampa?, ¿Mija q?, y todos aquellos que pusieron su granito en darme la fuerza y voluntad necesaria para culminar esta carrera tan anhelada (Padres, hermanos, tíos, abuelos, profesores, personal administrativo, amigos, compañeros de trabajo, Daniel, novias, exnovias) y a quienes lamentablemente ya no nos acompañan (Sergio, Don Hugo, Marcos, Dolores) Q.E.P.D.

Gracias.

## Tabla de Contenido

	Pág.
<b>Introducción .....</b>	<b>14</b>
<b>1. Planteamiento del Problema.....</b>	<b>15</b>
1.1 Antecedentes.....	16
<b>2. Justificación.....</b>	<b>17</b>
<b>3. Objetivos.....</b>	<b>19</b>
3.1 Objetivo General.....	19
3.2 Objetivos Específicos.....	19
<b>4. Marco Teórico.....</b>	<b>20</b>
<b>5. Tecnologías Involucradas en el Desarrollo del Proyecto .....</b>	<b>24</b>
5.1 Bases de Datos Mysql.....	24
5.2 Python.....	26
5.2.1 Peewee.....	26
5.2.2 Web Scrappy.....	27
5.3 Java.....	31
5.3.1 EJB.....	33
5.3.2 JPA.....	34
5.3.3 JSF.....	35
5.3.4 Hibernate.....	35
5.3.5 PrimeFaces.....	36
5.3.6 iText.....	37
5.3.7 iReport Designer.....	38
5.4 Gitlab.....	38
<b>6. Metodología.....</b>	<b>39</b>
<b>7. Análisis y Diseño del Sistema de Información .....</b>	<b>41</b>
7.1 Especificación de Requerimientos.....	41
7.1.1 Requisitos Funcionales.....	41
7.1.2 Requisitos No Funcionales.....	42
7.2 Casos de Uso.....	42
7.2.1 Identificación de Actores.....	42
7.2.2 Diagrama Modelo Casos de Usos.....	43
7.2.2.1 Descripción Diagrama Modelo Casos de Uso.....	44
7.2.3 Diagrama de Caso de Uso – Gestión de Información por Perfil.....	48
7.2.3.1 Descripción del Diagrama de Caso de Uso – Gestión Información por Perfil.....	49
7.2.4 Diagrama de Caso de Uso – Gestión Información por Administrar.....	53

7.2.4.1	Descripción del Diagrama de Caso de Uso – Gestión Información por Administrador .....	54
<b>8.</b>	<b>Diseño Componentes .....</b>	<b>58</b>
8.1	<i>Modelo Entidad Relación.</i> .....	58
8.1.1	Diagrama Entidad Relación.....	58
8.1.2	Entidades de bases de datos.....	60
8.2	<i>Diseño de Componentes del Sistema</i> .....	75
8.3	<i>Diagrama de Secuencias de la Extracción de Datos.</i> .....	76
8.4	<i>Técnicas de Programación y Patrones Utilizados.</i> .....	77
<b>9.</b>	<b>Codificación del Componente.....</b>	<b>79</b>
9.1	<i>Estudio de los Datos que Serán Extraídos</i> .....	79
9.1.1	Análisis de CvLac.....	79
9.1.2	Análisis de ResearchGate.....	80
9.2	<i>Extracción de Datos</i> .....	81
9.2.1	Desarrollo de Módulo de Extracción .....	81
9.3	<i>Desarrollo de la Interfaz de Presentación de Resultados.</i> .....	88
9.4	<i>Componente de generación de PDF</i> .....	95
<b>10.</b>	<b>Pruebas.....</b>	<b>96</b>
<b>11.</b>	<b>Conclusiones .....</b>	<b>98</b>
<b>12.</b>	<b>Recomendaciones .....</b>	<b>99</b>
	<b>Referencias bibliográficas. ....</b>	<b>100</b>

## Lista de Figuras

	<b>Pág.</b>
<b>Figura 1.</b> Estructura de ORM.....	27
<b>Figura 2.</b> Estructura Scrapy.....	28
<b>Figura 3.</b> Arquitectura JEE.....	33
<b>Figura 4.</b> Arquitectura EJB .....	34
<b>Figura 5.</b> Arquitectura Básica de Hibernate.....	36
<b>Figura 6.</b> Modelo de RUP .....	40
<b>Figura 7.</b> Identificación de actores.....	42
<b>Figura 8.</b> Modelo de casos de uso.....	44
<b>Figura 9.</b> Diagrama de caso de uso - Gestión información por perfil.....	49
<b>Figura 10.</b> Diagrama de casos de uso - Gestión información por administrar.....	53
<b>Figura 11.</b> Diagrama Entidad Relación.....	59
<b>Figura 12.</b> Diagrama de componentes.....	75
<b>Figura 13.</b> Diagrama de secuencia .....	77
<b>Figura 14.</b> Representación Modelo Vista Controlador .....	78
<b>Figura 15.</b> Item Scrapy.....	82
<b>Figura 16.</b> Código Fuente CvLac.....	83
<b>Figura 17.</b> Spider Scrapy.....	84
<b>Figura 18.</b> Paquete de utilidades .....	85
<b>Figura 19.</b> JSON de salida.....	85
<b>Figura 20.</b> Mapeo modelo Peewee.....	86
<b>Figura 21.</b> Paquete DAO.....	87
<b>Figura 22.</b> Pipeline.py .....	87
<b>Figura 23.</b> Validación de ítem bibliográficos.....	88
<b>Figura 24.</b> Formulario Registro.....	89
<b>Figura 25.</b> Mi Perfil.....	89
<b>Figura 26.</b> Información personal de idiomas .....	90
<b>Figura 27.</b> Información Personal de Condecoraciones .....	90
<b>Figura 28.</b> Gestión de Perfil .....	91
<b>Figura 29.</b> Menú de Idiomas .....	91
<b>Figura 30.</b> Añadir Idioma.....	92
<b>Figura 31.</b> Eliminar Idioma.....	92
<b>Figura 32.</b> Menú estudio .....	93
<b>Figura 33.</b> Añadir Estudio.....	93
<b>Figura 34.</b> Menú Condecoraciones .....	94
<b>Figura 35.</b> Gestión Proyectos .....	94
<b>Figura 36.</b> Experiencia profesional .....	95
<b>Figura 37.</b> Reporte fase de diseño.....	96

## Lista de Tablas

	<b>Pág.</b>
<b>Tabla 1.</b> Descripción del Actor Administrador.....	43
<b>Tabla 2.</b> Descripción del Actor Usuario Autenticado.....	43
<b>Tabla 3.</b> Descripción Diagrama Modelo Casos de Uso CU01 .....	44
<b>Tabla 4.</b> Descripción Diagrama Modelo Casos de Uso CU02 .....	45
<b>Tabla 5.</b> Descripción Diagrama Modelo Casos de Uso CU03 .....	47
<b>Tabla 6.</b> Descripción Diagrama Modelo Casos de Uso CU04 .....	47
<b>Tabla 7.</b> Descripción del Diagrama de Caso de Uso – Gestión Información por Perfil CU05 .....	49
<b>Tabla 8.</b> Descripción del Diagrama de Caso de Uso – Gestión Información por Perfil CU06 .....	50
<b>Tabla 9.</b> Descripción del Diagrama de Caso de Uso – Gestión Información por Perfil CU07 .....	51
<b>Tabla 10.</b> Descripción del Diagrama de Caso de Uso – Gestión Información por Perfil CU08 .....	52
<b>Tabla 11.</b> Descripción del Diagrama de caso de uso – Gestión información por administrador CU09.....	54
<b>Tabla 12.</b> Descripción del Diagrama de caso de uso – Gestión información por administrador CU10.....	55
<b>Tabla 13.</b> Descripción del Diagrama de caso de uso – Gestión información por administrador CU11.....	56
<b>Tabla 14.</b> Descripción del Diagrama de caso de uso – Gestión información por administrador CU12.....	57
<b>Tabla 15.</b> Entidad : language_ domain .....	60
<b>Tabla 16.</b> Entidad: language_ domain Key.....	61
<b>Tabla 17.</b> Entidad: experience .....	61
<b>Tabla 18.</b> Entidad: experience Key.....	62
<b>Tabla 19.</b> Entidad: user .....	62
<b>Tabla 20.</b> Entidad: study .....	63
<b>Tabla 21.</b> Entidad: study key .....	64
<b>Tabla 22.</b> Entidad: distinction.....	64
<b>Tabla 23.</b> Entidad: distinction key .....	64
<b>Tabla 24.</b> Entidad: Projects.....	65
<b>Tabla 25.</b> Entidad: projects key .....	65
<b>Tabla 26.</b> Entidad: projects_person .....	66
<b>Tabla 27.</b> Entidad: projects_person key obj persona .....	66
<b>Tabla 28.</b> Entidad: projects_person key id Project .....	66
<b>Tabla 29.</b> Entidad: language .....	66
<b>Tabla 30.</b> Entidad: obj_author .....	67
<b>Tabla 31.</b> Entidad: obj_author key rol .....	67

<b>Tabla 32.</b>	Entidad: obj_author id bibliographic .....	67
<b>Tabla 33.</b>	Entidad : obj_author key usuario .....	68
<b>Tabla 34.</b>	Entidad : obj_bibliographic .....	68
<b>Tabla 35.</b>	Entidad : obj_bibliographic key father .....	69
<b>Tabla 36.</b>	Entidad : obj_bibliographic key id .....	70
<b>Tabla 37.</b>	Entidad : obj_bibliographic ISBN .....	70
<b>Tabla 38.</b>	Entidad : obj_bibliographic Index .....	70
<b>Tabla 39.</b>	Entidad : obj_bibliographic key language .....	70
<b>Tabla 40.</b>	Entidad : obj_type .....	71
<b>Tabla 41.</b>	Entidad : investigation_lines_projects .....	71
<b>Tabla 42.</b>	Entidad : investigation_lines_projects index .....	71
<b>Tabla 43.</b>	Entidad : investigation_lines_projects Key id .....	72
<b>Tabla 44.</b>	Entidad : investigation_lines_groups .....	72
<b>Tabla 45.</b>	Entidad : investigation_lines_groups Grupos .....	72
<b>Tabla 46.</b>	Entidad : investigation_lines .....	73
<b>Tabla 47.</b>	Entidad : investigation_group .....	73
<b>Tabla 48.</b>	Entidad : user investigation_group .....	74
<b>Tabla 49.</b>	Entidad : user_investigation_group key id .....	74
<b>Tabla 50.</b>	Entidad : user_investigation_group key id usuario .....	74
<b>Tabla 51.</b>	Especificaciones Equipo Pruebas .....	96
<b>Tabla 52.</b>	Pruebas de Extracción de datos Scrapy .....	97
<b>Tabla 53.</b>	Prueba Componente Interfaz de Usuario .....	97

## Resumen en Español

**TÍTULO:** DESARROLLO DE UNA HERRAMIENTA SOFTWARE PARA GENERACIÓN, ACTUALIZACIÓN Y CONSULTA DE HOJAS DE VIDA DE INVESTIGADORES MEDIANTE LA OBTENCIÓN DE DATOS\*

**AUTORES:** JOHAN SEBASTIÁN DURAN CARO Y DANIEL ROJAS CÁCERES\*\*

**PALABRAS CLAVES:** HERRAMIENTA WEB, EXTRACCIÓN DE DATOS, HOJAS DE VIDA, INVESTIGACIÓN, SCRAPY.

### DESCRIPCIÓN:

En el desarrollo activo para mejorar la autogestión de la Universidad Industrial de Santander, se evidencia que los grupos de investigación que precisan obtener un registro de las actividades e investigaciones que contribuyan a procesos de autoevaluación en los programas académicos y actualización en la información de los usuarios. Gracias a las TIC (tecnologías de la información y comunicación) las cuales nos proporcionan herramientas que permitan agilizar estos procesos y estimar de manera más precisa los procesos dichos anteriormente surge la elaboración de este proyecto.

Partiendo de la necesidad de tener una base de datos que permita gestionar la información de los investigadores, sus publicaciones, libros y artículos en revistas científicas con el fin de mantener íntegras y actualizadas las hojas de vida, se propone el desarrollo de una herramienta web que lo facilite, trayendo la información de las hojas de vida de los investigadores, ya diligenciadas por ellos en los otros portales como Scienti, de Colciencias y ResearchGate.

En este proyecto se busca integrar en el grupo Filosofía y Enseñanza de la Filosofía (FiloEn) un complemento a la gestión de información de usuario, mediante procesos de extracción de información mediante la técnica scrapy de dos sitios web: CvLac y ResearchGate.

---

\* Trabajo de grado.

\*\*Facultad de Ingenierías Físico Mecánicas, Escuela de Ingeniería de Sistemas e Informática, Director: PhD. Sonia Cristina Gamboa Sarmiento, Doctora en Educación.

## Resumen en Ingles

**TITLE:** DEVELOPMENT OF A SOFTWARE TOOL FOR THAT GENERATION, UPDATE AND CONSULTATION OF RESEARCHERS CURRICULA BY OBTAINING DATA.\*

**AUTHORS:** JOHAN SEBASTIAN DURAN CARO Y DANIEL ROJAS CACERES.\*\*

**KEYWORDS:** TOOL, DATA EXTRACTION, WEB, RESUMES, INVESTIGATION, SCRAPY.

### DESCRIPTION:

In the active development to improve the self-management of the Industrial University of Santander, it is evident that the research groups that need to obtain a record of the activities and research that contribute to self-evaluation processes in the academic programs and update in the information of the users. Thanks to the ICT (information and communication technologies) which provide us with tools that allow these processes to be speeded up and more accurately estimate the processes mentioned above, the development of this project arises.

Starting from the need to have a database that allows managing the information of researchers, their publications, studies, acknowledgments, directed projects, books and articles in scientific journals in order to keep resumes integrated and updated, the development of a web tool that facilitate it, bringing the information of the resumes of the researchers, already filled out by them in the other portals such as Scienti, from Colciencias and ResearchGate.

This project seeks to integrate a complement to the management of user information into the Philosophy and Teaching of Philosophy (PhiloEn) group, through information extraction processes using the scrapy technique which will provide the necessary tools to integrate the information extracted to our database in two websites: CvLac and ResearchGate.

---

\* Bachelor thesis.

\*\*Faculty of Physical-Mechanical Engineering, School of System Engineering and Informatics. Director: PhD. Sonia Cristina Gamboa Sarmiento.

## Introducción

El constante desarrollo de las Tecnologías de la Información y la Comunicación (TIC) hace necesario el continuo avance de las herramientas que facilitan su uso, gestión y organización en la vida, tanto académica como laboralmente; con el fin de dejar a la comunidad universitaria una herramienta que permita gestionar información sobre los investigadores y su producción intelectual en el área de filosofía surge este proyecto, el cual tiene como propósito mantener al día y de forma íntegra las hojas de vida de los investigadores.

En esta información se evidencia los resultados de investigación y la memoria colectiva de los grupos, Para definir el concepto de memoria colectiva se debe considerar todas las representaciones a través de las diferentes épocas que han sido producidas por diferentes grupos o colectivos, estas se conservan y se difunden a través de la interacción (Jedlowski, citado por Páez., Techio., Marques., & Beristain, 2007). Sin embargo, para Páez, Et al (2007) no se puede limitar la memoria a una consideración del pasado o recuerdo solamente, sino como la forma de llevar al presente lo que se ha producido y avanzar en las generaciones siguientes.

Partiendo de lo anterior y llevándolo al entorno de la investigación, se hace necesario la implementación de nuevas herramientas de las TIC que permitan el desarrollo y la calidad de la investigación, agregando de esta manera transparencia al proceso educativo, facilitando de este modo la identificación de patrones, tendencias u otros conocimientos que favorezcan la toma de decisiones, planificación y autoevaluación.

En el presente proyecto se diseñó e implementó una herramienta web basada en la extracción de datos mediante la técnica Scrapy. Scrapy es un Framework que se encuentra diseñado el lenguaje de programación Python, y la técnica Scrapy a la que se refiere el documento consiste en el rastreo y extracción de datos estructurados, que pueden usarse en

diferentes entornos como procesamiento de información y archivos históricos (Scrapy, 2018), para este caso la información extraída permite la producción de hojas de vida, organizándolas en una base de datos relacional que contiene toda la información pertinente al ámbito personal, académico y de propiedad intelectual de la comunidad en general, siendo gestionadas como perfiles de usuario, haciendo uso de la programación orientada a objetos de Java.

Lo dicho hasta aquí supone que, en el presente proyecto el cual ha sido dividido en cuatro capítulos, encontrara el desarrollo de una herramienta software para generación, actualización y consulta de hojas de vida de investigadores mediante la obtención de datos. En el primer capítulo encontrarán el planteamiento del problema y la justificación del proyecto donde se expondrán algunos antecedentes y la importancia de las herramientas web en la investigación, en el segundo capítulo se encontrarán los objetivos a los que responde esta investigación; luego, el marco teórico y las herramientas usadas para el desarrollo de esta herramienta las encontrara en el capítulo tres y en el capítulo cuatro se encuentra la metodología, para concluir, en el capítulo cinco se encuentra lo que corresponde a análisis del sistema de información, diseño, codificación del componente y pruebas.

## **1. Planteamiento del Problema**

La Universidad Industrial de Santander ha demostrado ser una de las mejores universidades de Colombia, siendo posicionada dentro de las primeras 10 más destacadas, por la firma *Sapiens Research* en su más reciente clasificación de universidades en 2019 (citado por Univalle, 2019).

La UIS se destaca por una formación de alto nivel y calidad, generando ventajas permanentes para un desarrollo sostenible, por tanto, la autoevaluación en la universidad es

una necesidad que se extiende cada vez más; por esta razón se llevó a cabo un proyecto que busco desarrollar un sistema de información que genere hojas de vida de investigadores aportando ventajas a la hora de apoyar los procesos de tanto de autoevaluación y gestión en el programa académico.

## 1.1 Antecedentes

Salazar y Chaparro (2014); crean un *sistema de información para generación de hojas de vida de investigadores*, este sistema fue creado para operar sobre bases de datos no relacionales, teniendo como fin almacenar la información académica de los investigadores, sin embargo, el sistema posee limitaciones, tales como no almacenar las publicaciones de los investigadores, como lo son los libros, artículos, colaboraciones y demás producción intelectual.

Para cubrir las falencias del sistema de información existente, Es necesario desarrollar una nueva herramienta software que opere sobre bases de datos relacionales, que asegure mantener la información actualizada y veraz de los investigadores, lo cual servirá para medir los niveles de producción intelectual.

Por otra parte, Galeano y Salamanca (2016), desarrollaron un portal que se planteó como complemento a la catedra docente, permitiendo a los estudiantes practicar y afianzar los conocimientos adquiridos en la clase presencial, dicho proyecto fue denominado *Diseño e implementación de un portal web para soportar la enseñanza de la filosofía*.

Dicho proyecto integró las herramientas en un portal donde los profesores encontraron otro medio de comunicación con sus estudiantes, asimismo, dispone de herramientas que les permiten organizar sus propios grupos de estudios virtuales.

Por tanto, teniendo en cuenta los aportes anteriores, la necesidad de una constante actualización de los sistemas académicos y el constante crecimiento de la producción intelectual, es menester desarrollar una herramienta que permita compilar la información de los diferentes autores presentes en la comunidad UIS, esto para que se brinde un mejor seguimiento a los diferentes trabajos de diversas autorías y mejore la autoevaluación de la institución.

## **2. Justificación**

La incorporación de las TIC supone mejorar la calidad de vida de los ciudadanos, favoreciendo el desarrollo en áreas como la investigación, gestión, planeación e información. La tecnología se debe contemplar como la herramienta y no como el fin en sí misma, esto según las autoras Martínez & Parra (2016). Su valor está en la capacidad de construir soluciones a problemas prácticos y reales haciendo llegar a la sociedad mejoras tangibles.

Carreño & Gamboa (2014), resaltan la necesidad de mantener los procesos y tecnologías a la vanguardia de los avances del mundo, destacan las ciencias de la información y la ingeniería de sistemas, ya que los avances en los sistemas computacionales a través de la implementación de entornos virtuales interactúan en la gestión del conocimiento, por medio de proyectos de investigación en diferentes enfoques se busca nutrir las herramientas TIC que se van desarrollando, así mismo los diferentes investigadores bien sean expertos o en formación puedan construir con sus aportes la memoria colectiva de las organizaciones.

Para Cabral (2016), el desarrollo, la calidad y la viabilidad en la sociedad actual depende de la incorporación efectiva de las nuevas tecnologías.

Así mismo, debemos partir de que el conocimiento científico se genera por las necesidades de la humanidad y la forma en la que se puede dar respuestas a los cuestionamientos de la

sociedad es la investigación, de manera que por medio de esta se brinda una solución a los problemas, al mismo tiempo que se tiene la posibilidad de aumentar la eficiencia y rentabilidad de los diferentes sectores (Patiño, 2013)

Por tanto, la implementación de un sistema que pueda mantener el *currículum vitae* actualizado, les permitirá a los miembros de comunidad educativa, realizar un proceso de seguimiento y autoevaluación, para la mejorar la calidad de la investigación.

### 3. Objetivos

#### 3.1 Objetivo General

Desarrollar una herramienta software para la web que permita recuperar de bases de datos de investigadores la información correspondiente a hojas de vida y permitir almacenar, administrar y gestionarla como perfiles de usuario.

#### 3.2 Objetivos Específicos

- ✓ Analizar los requerimientos necesarios para el desarrollo de la herramienta software.
- ✓ Establecer una estructura de datos correspondiente a la información de usuarios y su relación con la disposición de almacenamiento planteada.
- ✓ Diseñar e implementar una herramienta software para la web que permita realizar las siguientes funciones: búsqueda y recuperación de información de los usuarios registrados en las bases de datos, almacenar información recuperada; como lo son las hojas de vida y gestionar los usuarios.
- ✓ Desarrollar para la herramienta Software los siguientes módulos: Extraer de las plataformas CvLAv y ResearchGate la información para actualizar hojas, permitir que las hojas de vida puedan ser gestionadas por los usuarios, consultar hojas de vida de usuarios, permitir la impresión de las hojas de vida de los usuarios.
- ✓ Realizar las pruebas de software de usuarios, así como también de sus funciones.

#### 4. Marco Teórico

En los últimos años existe un especial interés de las universidades públicas por diagnosticar el estado en el que se encuentran sus programas educativos, la preocupación de los interesados por ingresar a programas de calidad, la necesidad de demostrar su nivel académico e identificar el estado actual de sus estudiantes, egresados e investigadores.

“La propiedad intelectual se relaciona con las creaciones de la mente: invenciones, obras literarias y artísticas, así como símbolos, nombres e imágenes utilizados en el comercio” (Wipo, 2004), en el caso de las universidades es una evidencia de la calidad y aporte al renombre de la entidad, según la Wipo (2004), la propiedad intelectual se divide en dos grandes categorías, como lo son la propiedad industrial, abarcando marcas, diseños e indicaciones de georreferencia y el derecho de autor donde se encuentran las propiedades intelectuales alrededor del arte ( música, teatro, literatura, entre otras), programas de radio, televisión y arquitectura.

Por tanto, el diseño de las herramientas web para llevar un seguimiento de esta producción intelectual se hace imprescindible, teniendo en cuenta la investigación realizada por Arbeláez (2014) en su artículo *Las tecnologías de la información y la comunicación (TIC) un instrumento para la investigación*. Donde expone la importancia de la implementación de los sistemas de información en la investigación, puesto que suponen un proceso más rápido y dinámico en la recolección y análisis de la información, ya que estos elementos aportan también una amplia gama de beneficios como las publicaciones en la web o bases de datos educativas, la accesibilidad a las comunidades en general y a grupos científicos particulares, entre otros. Esto permite que, al mismo tiempo que va evolucionando la investigación, se realicen desarrollos en la tecnología, como los diferentes programas existentes en la

búsqueda, gestión y exposición de datos, aumentado los niveles de investigación y su calidad en sí.

El permanente avance en las tecnologías y el crecimiento constante de la producción intelectual de los individuos, colectivos o instituciones, pone en consideración el desarrollo de una herramienta la cual permita contener esos datos, por medio de la utilización de un software al servicio de los grupos de investigación y comunidad en general, de la Universidad Industrial de Santander.

El desarrollo de dicho software se elabora en el paradigma de la programación orientada a objetos, el cual utiliza objetos para crear aplicaciones. Su uso se popularizó a principios de la década de 1990.

### **Programación Orientada a Objetos (POO)**

La POO “se puede definir como un paradigma que busca realizar una abstracción de la realidad, que se puede implementar en una aplicación de software con el fin de resolver problemas mediante el uso de un lenguaje de programación” (Flórez, 2012, Pg.23). La POO busca a establecer una armonía entre la importancia que existe en los procesos y los datos.

Existen 4 bases fundamentales en la POO según nos expone Bernal & Fernández (2012) que son:

- **Abstracción:** Hace referencia a las características de un objeto, aquella que pueden diferenciarlo de los demás, puede ser subjetiva respecto al observador que realiza la abstracción del objeto.

- **Encapsulación:** Es la característica que indica que visibilidad dependiendo de la clase que tienen los atributos, de esta manera se determina la seguridad información puesta en dichos atributos.

- **Modularidad:** Descompone por partes un conjunto, dando relevancia a los términos de acoplamiento y cohesión, donde se busca una asociación, para tener módulos acoplados y un alto índice de cohesión.

- **Jerarquía:** Estructuración por niveles de distintos elementos.

Es necesario comprender algunos conceptos dentro de la POO como lo son:

**Objetos:** es un ejemplar específico de una clase, dependiendo como se definió es su estructura y comportamiento, es particular e independiente del resto de ejemplares (Bernal & Fernández, 2012).

**Atributos:** son las características que se incluyen a la clase, es cada uno de los datos que la describen.

**Clases:** contienen en sí mismas atributos y métodos, en esta se describen, las estructuras que lo conforman y las funciones asociadas a ella, “una clase es el modelo en el que se construyen objetos” (Bernal & Fernández, 2012; pg. 7). Implementa conceptos que se abstraen de la realidad (Flórez, 2012).

**Método:** son los servicios que se incluyen dentro de la clase, en este se implementa el código necesario para un servicio, se conforma de los elementos; visibilidad (private, public o protected; retorno, los métodos puede retornar información o no, si no lo tiene se usa la palabra reservada “void” y si tiene retorno en su puesta en marcha debe estar presente la palabra “return”, el nombre; funge como identificador del método en la clase y los parámetros; este puede ser un tipo primitivo de dato o una declaración de clase y un método puede recibir de 0 a  $n$  parámetros (Flórez, 2012).

Existen beneficios significativos al usar POO como lo es la reusabilidad del código por medio de la herencia y el polimorfismo, donde la herencia permite a las clases que se diferencien unas de otras y el polimorfismo de define como la capacidad de que un mismo

mensaje funcione con diferentes objetos, operando sobre un conjunto posible de objetos compatibles (Rios, 2005)

Se hace necesario uso del aporte teórico de Usabilidad Web, debido a que al ser un software al servicio de la comunidad su implementación debe ser intuitiva, eficaz y eficiente en su contexto de uso.

### **Usabilidad Web.**

La usabilidad, es un anglicismo cuyo origen según Bevan (citado por Carvajal y Saab, 2010), surge sobre los años 80 con el fin de sustituir el término “amigable con el usuario”. La definición más formal existente en la actualidad es la del estándar internacional ISO 9241-11 (citada por Carvajal y Saab, 2010. Pg., 12) “El grado en que un producto puede ser usado por determinados usuarios para lograr sus propósitos con eficacia, eficiencia y satisfacción en un contexto de uso específico”.

Se define lo anterior para poder abarcar el termino usabilidad web definido por Bevan (citado por Carvajal y Saab, 2010) siendo un sistema o producto cuya simplicidad de uso y la aceptabilidad para una clase particular de usuarios, que se desarrolla para realizar tareas específicas.

### **Investigación en Colombia**

A partir de 2009, Colciencias por medio de la ley 1286, se convierte en el Departamento Administrativo de Ciencia (SNCTI), Tecnología e Innovación, ente que funciona como organismo de administración pública, se entiende por SNCTI:

“Un sistema abierto del cual forman parte las políticas, estrategias, programas, metodologías y mecanismos para la gestión, promoción, financiación, protección y divulgación de la investigación científica y la innovación tecnológica, así como las

organizaciones públicas, privadas o mixtas que realicen o promuevan el desarrollo de actividades científicas tecnológicas y de innovación” (Colciencias, s.f).

Con el fin de mantener actualizados los grupos de investigación, se realizan convocatorias donde se miden, los resultados que se han logrado en los distintos proyectos y el uso de las distintas estrategias implementadas, siendo este un estímulo y fortalecimiento para los investigadores (Colciencias, s.f.). Por este mismo propósito se implementa la herramienta web que busca mantener actualizados los aportes de los diferentes autores en la comunidad UIS.

Por medio de la hoja de vida se realiza esta actualización, este es un documento que presenta las habilidades, formación y vida laboral de una persona, con el fin de optar a un puesto de trabajo designa “la relación de datos personales, formación académica, actividad laboral y méritos de una persona” (Real Academia Española, 2010). En el ámbito universitario es de vital importancia el registro, seguimiento y control ya que realza los méritos de cada actor (docentes, estudiantes, investigadores).

## **5. Tecnologías Involucradas en el Desarrollo del Proyecto**

### **5.1 Bases de Datos Mysql**

Las bases de datos constituyen uno de los elementos clave sobre los que se cimientan los sistemas de información de empresas e instituciones, podríamos decir que una base de datos es una representación de la información en forma de datos interrelacionados que se almacenan formando grandes volúmenes de información (Cobo, Gómez, Pérez & Rocha, 2005).

Arias (2014) expone que MySQL es un sistema de gestión de bases de datos relacional que se base en el lenguaje SQL (Structured Query Language), lanzado en 1995 y adquirido por Sun Microsystems en el 2008(más tarde, en 2009, Oracle compró Sun).

Thibaud (2006, Pg. 10) presenta en la siguiente lista a las principales características de - MySQL

- Está escrito en C/C++ y probado con numerosos compiladores.
- Funciona en muchos sistemas operativos
- Soporta muchas API como C/C++, PHP, Perl, Eiffel, Python, Java, Tcl.
- Las utilidades cliente y administración utilizan los sockets TCP/IP, los Sockets Unix o los canales con nombre NT (Named Pipes)
- Dispone de un driver ODBC (Open DataBase Connectivity) para Windows lo que le aporta compatibilidad con la mayoría de lenguajes disponible en este sistema operativo.
- Dispone de un driver ADO vía OLEDB en Windows para un mejor rendimiento que ODBC
- Es totalmente multithread, lo que aumenta enormemente el rendimiento del servidor. MySQL está previsto para funcionar en entornos multiprocesador
- Dispone de un sistema de administración de la memoria extremadamente potente. La cantidad de memoria tiene una incidencia directa sobre el rendimiento de MySQL. Si un servidor presenta fallos de rendimiento, suele bastar con añadir memoria RAM para mejorar sensiblemente su rendimiento.
- Se puede interrogar con el lenguaje SQL, lo que garantiza una mejor compatibilidad con otros SGBD, Todos los scripts SQL escritos para otro SGBD que respeten el sql estándar pueden utilizarse con MySQL

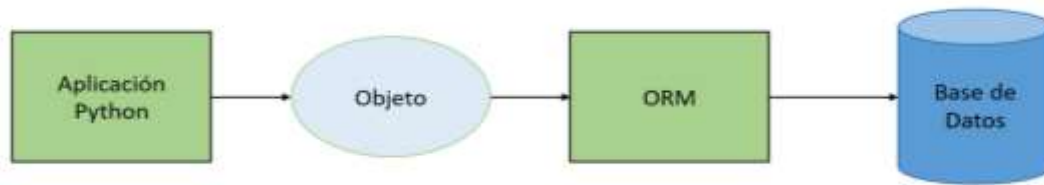
## 5.2 Python

Es un lenguaje desarrollado como proyecto de código abierto y lo administra, Python software Foundation, su creador es Guido van Rossum, se trata de un lenguaje de programación en scripts, este permite dividir el programa en módulos reutilizables desde otros programas Python. Es un lenguaje interpretado, lo que permite ahorrar el proceso de compilado (Python, 2019).

### Características generales de Python

- Lenguaje de programación de alto nivel del tipo scripting.
- Diseñado para ser fácil de leer y simple de implementar.
- Es código abierto (de libre uso).
- Puede ejecutarse en Mac, Windows y sistemas Unix; también ha sido portado a máquinas virtual JAVA y .NET.
- Es a menudo usado para desarrollar aplicaciones web y contenido web dinámico.
- Se utiliza para crear extensiones tipo plug-ins para programas de 2d y 3d como Autodesk Maya, GIMP, Blender, Inkscape, etc.
- Los scripts de Python tienen la extensión de archivo .PY, que pueden ser interpretados y ejecutados inmediatamente.
- Permite grabar programas compilados con extensión de archivo PYC, los cuales suelen ser usados como módulo que pueden ser referenciados por otros programas Python (Alegsa, 2018).

**5.2.1 Peewee.** Es una aplicación de correlación objeto-relacional (ORM) para puentear datos almacenados en tablas de bases de datos relacionales con objetos Python.



**Figura 1.** Estructura de ORM

Peewee se puede utilizar con prácticamente cualquier framework web, además es bueno para extraer datos de una base de datos relacional mediante scripts en Jupyter Notebook.

Se admiten todos los backends de bases de datos relacionales comunes como PostgreSQL, MySQL o SQLite, aunque todavía se requiere un controlador de base de datos (Python, 2018).

**5.2.2 Web Scrapy.** Scrapy es un framework 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.

Aunque Scrapy fue diseñado originalmente para el raspado web (Scrapy, 2016), también se puede utilizar para extraer datos mediante API (como Amazon Associates Web Services) o como rastreador web de propósito general.

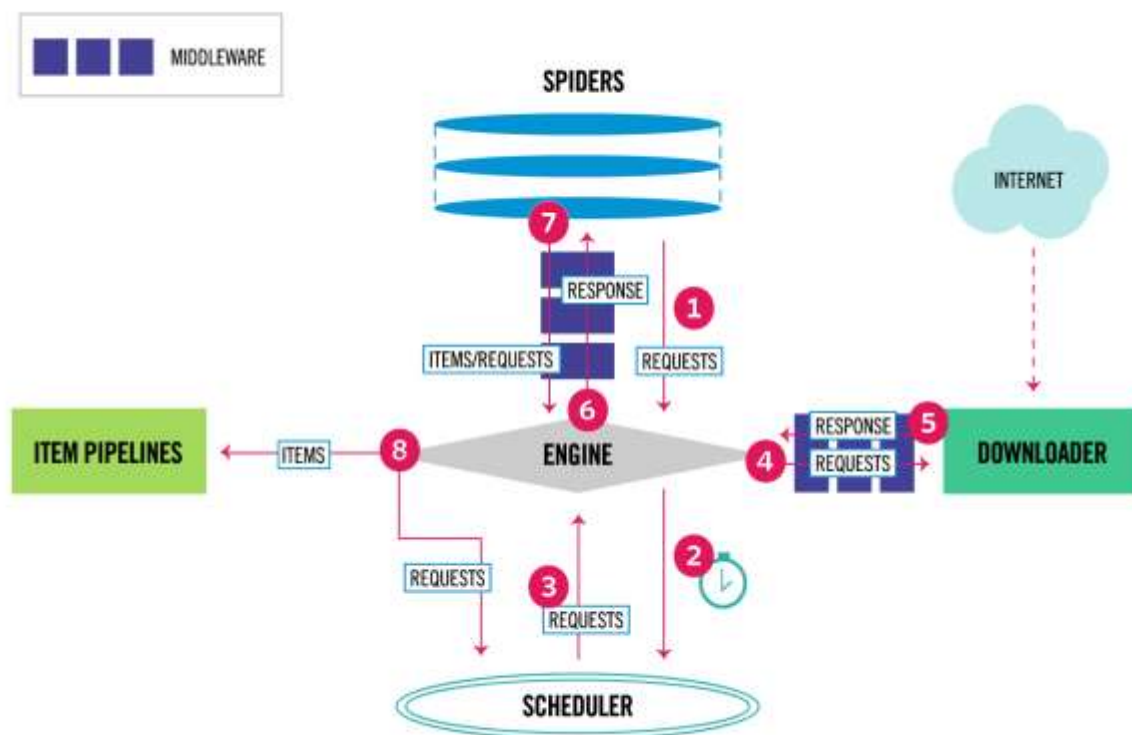
Scrapy tiene una serie de herramientas poderosas para hacer el "scraping" o extraer información de webs de manera fácil y eficiente. Estas herramientas comprenden:

- Soporte para extraer y seleccionar datos de fuentes HTML/XML usando selectores CSS y expresiones XPath, con métodos de ayuda para extraer usando expresiones regulares.
- Una consola interactiva en IPython para ensayar los CSS y expresiones XPath para extraer datos, muy útil cuando se construyen métodos propios.

- Soporte para exportar los registros en formatos múltiples como JSON, CSV y XML.
- Soporte para manejar declaraciones foráneas, no estándares y códigos rotos.
- Fuerte extensibilidad, ya que permite conectar tu propia funcionalidad usando señales, extensions y pipelines (Gidahatari, 2018).

Scrapy está escrito con Twisted, una estructura de redes de eventos impulsada por Python. Por lo tanto, se implementa mediante un código no bloqueador (también asíncrono) para la concurrencia.

### Estructura de Scrapy



**Figura 2.** Estructura Scrapy. Fuente: Scrapy.org (s.f). [https://docs.scrapy.org/en/latest/\\_images/scrapy\\_architecture\\_02.png](https://docs.scrapy.org/en/latest/_images/scrapy_architecture_02.png)

El flujo de datos en Scrapy es controlado por el motor de ejecución, y va así:

- El motor obtiene las solicitudes iniciales para arrastrarse desde la araña.
- El motor programa las Solicitudes en el Programador y pide las siguientes Solicitudes de rastreo.
- El planificador devuelve las siguientes solicitudes al motor.
- El Motor envía las Solicitudes al Descargador, pasando a través de los Middlewares del Descargador.
- Una vez que la página termina de descargar, el Downloader genera una respuesta (con esa página) y la envía al motor, pasando a través de los middlewares del descargador.
- El Motor recibe la Respuesta del Descargador y la envía al Spider para su procesamiento, pasando a través del Spider Middleware.
- La Araña procesa la Respuesta y devuelve los elementos raspados y las nuevas Solicitudes (para seguir) al Motor, pasando a través del Spider Middleware.
- El motor envía los elementos procesados a Pipelines de elementos, a continuación, envía solicitudes procesadas al Programador y solicita las siguientes solicitudes de rastreo.
- El proceso se repite (desde el paso 1) hasta que no haya más peticiones desde el Programador.

## Componentes

### **Motor Scrapy (Engine)**

El motor es responsable de controlar el flujo de datos entre todos los componentes del sistema y de activar eventos cuando se producen ciertas acciones. Consulte la sección Flujo de datos anterior para obtener más detalles.

### **Programador (Scheduler)**

El programador recibe las solicitudes del motor y las en cola para alimentarlas más tarde (también al motor) cuando el motor las solicite.

### **Descargador (Downloader)**

El descargador es responsable de buscar páginas web y alimentarlos al motor que, a su vez, los alimenta a las arañas.

### **Arañas (Spiders)**

Las arañas son clases personalizadas escritas por los usuarios de Scrapy para analizar las respuestas y extraer elementos (aka elementos raspados) de ellos o solicitudes adicionales a seguir. Para más información ver Arañas.

### **Pipeline del artículo (Item Pipeline)**

El Pipeline del artículo es responsable de procesar los artículos una vez que han sido extraídos (o raspados) por las arañas. Las tareas típicas incluyen limpieza, validación y persistencia (como almacenar el elemento en una base de datos). Para obtener más información, consulte Pipeline de elementos.

### **Descargas de middlewares**

Las descargas de middlewares son ganchos específicos que se sitúan entre el motor y el descargador y procesan las solicitudes cuando pasan del motor al descargador y las respuestas que pasan del descargador al motor.

Utilice una descarga de middlewares si necesita realizar una de las siguientes acciones:

- Procesar una solicitud justo antes de que se envíe al Descargador (es decir, justo antes de que Scrapy envíe la solicitud al sitio web).
- Cambiar la respuesta recibida antes de pasarla a una araña.
- Enviar una nueva solicitud en lugar de pasar respuesta recibida a una araña.
- Pasar la respuesta a una araña sin buscar una página web.
- Silenciosamente soltar algunas solicitudes.

### Spider middlewares

Spider middlewares son ganchos específicos que se sitúan entre el motor y las arañas y son capaces de procesar araña de entrada (respuestas) y de salida (elementos y solicitudes).

Utilice un middleware de *Spider* si necesita:

- Salida post-proceso de devoluciones de llamada araña - cambiar / agregar / quitar solicitudes o artículos.
- Post-proceso *start\_requests*
- Manejar excepciones araña
- *Call errback* en lugar de *callback* para algunas de las solicitudes basadas en el contenido de la respuesta. (Scrapy et al, 2016).

### 5.3 Java

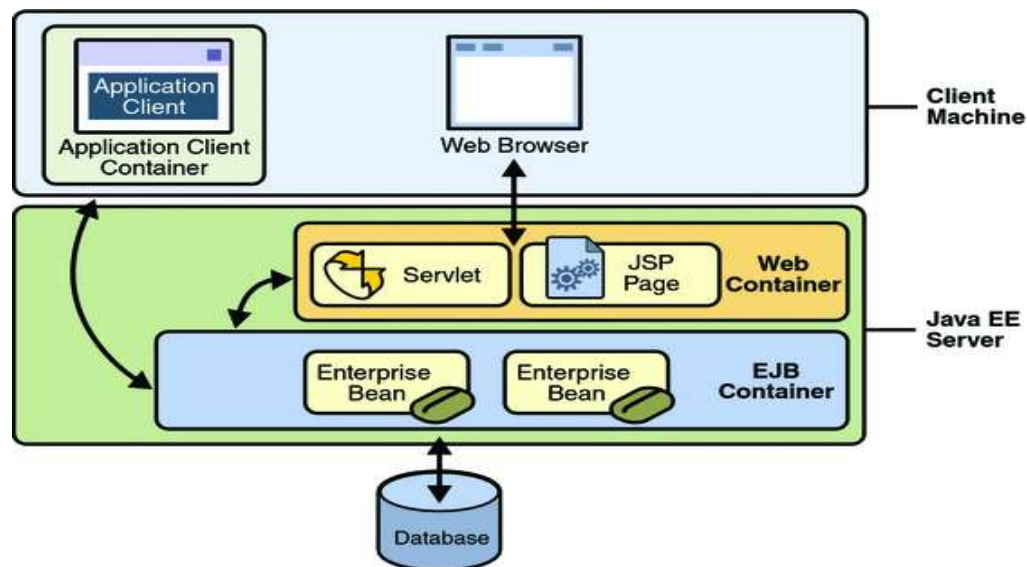
Java es un lenguaje de programación de propósito general, es decir que se puede utilizar para el desarrollo de todo desde aplicaciones web, empresariales, programas de escritorio hasta aplicaciones para teléfonos, fue desarrollado en 1990 y su creación se le atribuye a James Gosling.

Fue creado para ser ligero y potente omitiendo algunas características superficiales de otros lenguajes de programación y reduciendo el tamaño a lo esencial, se obtiene un lenguaje de programación con todas las características necesarias combinadas de manera elegante y fácil de interpretar, Java nos ofrece múltiples características deseables en cualquier proyecto como las mencionan Bell & Parr (2003):

- Es orientado a objetos (soporta POO)
- Java es muy bueno para la web y probablemente sea la característica que más lo popularizo en un principio.
- Es independiente de la plataforma, esto hace que el mismo código que se escribe una vez pueda ser ejecutado en múltiples plataformas gracias a su Java Virtual Machine (JVM) la cual compila el java bytecode haciendo que las instrucciones sean interpretadas fácilmente por cualquier sistema, es decir cumple la función de un intérprete. (Bell & Parr, 2003)

### **Arquitectura JEE**

Java Enterprise Edition (JEE) es una propuesta para el desarrollo de aplicaciones distribuidas que se sustenta en el lenguaje Java, sin embargo JEE no debe ser considerada como una mejora a JSE (Java Standard Edition) más bien debe considerarse como una normativa que define cómo se deben desarrollar los diferentes componentes, como debería ser la comunicación entre ellos y las diferentes restricciones a considerar para tener con ello un desarrollo robusto y fácil de escalar y sustentar. (Groussard, 2010)

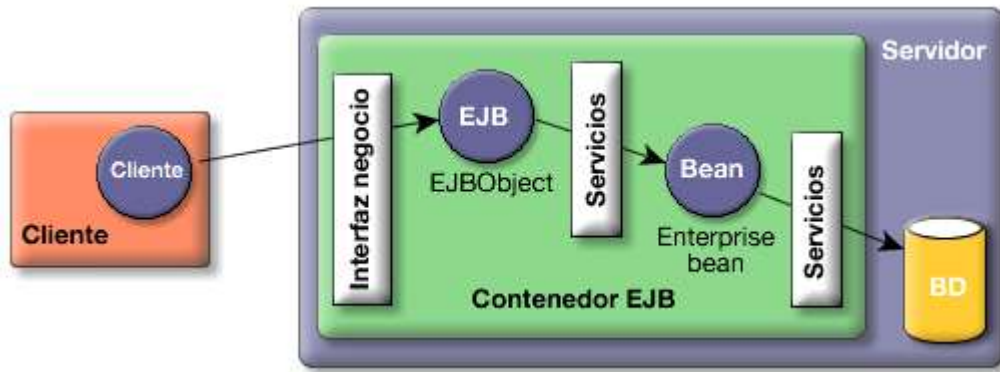


**Figura 3.** Arquitectura JEE. Fuente: Oracle (s.f). <https://docs.oracle.com/cd/E19575-01/819-3669/images/overview-serverAndContainers.gif>

Según nos expone. Groussard, (2010, Pg.19) Una aplicación web generalmente se compone de los siguientes elementos:

- “-Recursos estáticos: Páginas HTML, imágenes, sonidos, hojas de estilo
- Recursos Dinámicos: Servlets, jsp, Java Bean.
- Librerías de clases utilizadas por los diferentes componentes dinámicos.
- Gestor de despliegue que permite definir los parámetros de funcionamiento de la aplicación en el servidor, los enlaces entre las URL y los recursos dinámicos de la aplicación, la página por defecto, error de la aplicación, la seguridad, etc.”

**5.3.1 EJB.** El funcionamiento de los componentes EJB se basa fundamentalmente en el trabajo del contenedor EJB. El contenedor EJB es un programa Java que corre en el servidor y que contiene todas las clases y objetos necesarios para el correcto funcionamiento de los Enterprise bean.



**Figura 4.** Arquitectura EJB. Fuente: Universidad de Alicante (s.f). <http://www.jtech.ua.es/j2ee/2003-2004/abierto-j2ee-2003-2004/ejb/imagenes/arquitecturaEJB.png>

En la ilustración 4 se puede ver una representación de muy alto nivel del funcionamiento básico de los Enterprise beans. En primer lugar, se puede ver que el cliente que realiza peticiones al bean y el servidor que contiene el bean está ejecutándose en máquinas virtuales Java distintas. Incluso pueden estar en distintos hosts. Otra cosa a resaltar es que el cliente nunca se comunica directamente con el Enterprise bean, sino que el contenedor EJB proporciona un EJBOBJECT que hace de interfaz. Cualquier petición del cliente (una llamada a un método de negocio del enterprise bean) se debe hacer a través del objeto EJB, el cual solicita al contenedor EJB una serie de servicios y se comunica con el Enterprise bean. Por último, el bean realiza las peticiones correspondientes a la base de datos (Introducción a la tecnología, 2004).

**5.3.2 JPA.** JPA o Java Persistence API es el standard de Java encargado de automatizar dentro de lo posible la persistencia de nuestros objetos en base de datos. El cual es un framework del lenguaje de programación Java, que maneja datos relacionales en aplicaciones usando la plataforma Java en sus ediciones Standard (Java SE) y Enterprise (Java EE).

**5.3.3 JSF.** JSF es un framework MVC (Modelo-Vista-Controlador) basado en el API de Servlets que proporciona un conjunto de componentes en forma de etiquetas definidas en páginas XHTML mediante el framework Facelets. Facelets se define en la especificación 2 de JSF como un elemento fundamental de JSF que proporciona características de plantillas y de creación de componentes compuestos. JSF utiliza las páginas Facelets como vista, objetos Javabeen como modelos y métodos de esos objetos como controladores. El servlet FacesServlet realiza toda la tediosa tarea de procesar las peticiones HTTP, obtener los datos de entrada, validarlos y convertirlos, colocarlos en los objetos del modelo, invocar las acciones del controlador y renderizar la respuesta utilizando el árbol de componentes.

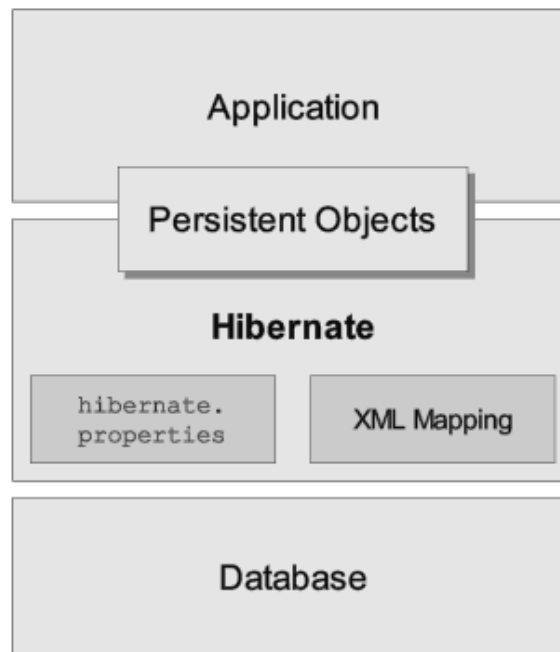
**5.3.4 Hibernate.** Hibernate ORM permite a los desarrolladores escribir aplicaciones con mayor facilidad, cuyos datos sobreviven al proceso de solicitud. Como marco de Object / Relational Mapping (ORM), Hibernate se preocupa por la persistencia de los datos tal como se aplica a las bases de datos relacionales (a través de JDBC)

### **¿Por qué usar un Framework ORM?**

Cuando desarrollamos aplicaciones en muchos casos todo termina siendo un conjunto de ABM (alta, baja y modificaciones de datos) que luego consultamos. Para ello se utiliza una base de datos donde hay muchas tareas repetidas: por cada objeto que quiero persistir debo crear una clase que me permita insertarlo, eliminarlo, modificarlo y consultarlo. Con excepción de consultas especiales, el resto es siempre lo mismo. Este es el momento dónde un ORM tiene una importancia fundamental. Con solo configurarlo todas estas tareas se ejecutan automáticamente y solo tendrás que preocuparte por las consultas especiales.

### ¿Cómo funciona Hibernate?

Se define en un archivo XML o mediante anotaciones los atributos correspondientes a una clase, con una columna de una tabla. Según Hibernate (2018); Es una tarea simple donde existen herramientas que lo hacen por nosotros.



**Figura 5.** Arquitectura Básica de Hibernate. Fuente: Hibernate Community Documentation (s.f). <http://docs.jboss.org/hibernate/core/3.5/reference/es-ES/html/images/overview.png>

**5.3.5 PrimeFaces.** “PrimeFaces es una colección de componentes de User Interface(UI) ricos para JavaServer Faces, Todos los widgets son de código abierto y de uso gratuito bajo la Licencia Apache. PrimeFaces es desarrollado por PrimeTek Informatics, un proveedor con años de experiencia en el desarrollo de soluciones de UI de código abierto” (PrimeFaces, 2019).

Çivici, Ç. (2014) explica que: PrimeFaces es un conjunto de componentes JSF de código abierto con varias extensiones las cuales son:

- Un amplio conjunto de componentes (Editor HTML, Diálogo, Autocompletar y gráficos, entre otros). Asynchronous JavaScript And

XML (Ajax) incorporado basado en las API de Ajax JSF 2.0 estándar, cero configuraciones y no se requieren dependencias.

- Soporte push a través de *Atmosphere Framework*
- Kit de interfaz de usuario móvil para crear aplicaciones web móviles.
- Skinning Framework con más de 35 temas incorporados y soporte para herramientas de diseño de temas visuales.
- Amplia documentación.
- Comunidad de usuarios grande, vibrante y activa.
- Desarrollado con "pasión" de los desarrolladores de aplicaciones a los desarrolladores de aplicaciones. Çivici, Ç. (2014).

**5.3.6 iText.** iText es una biblioteca PDF que permite crear, adaptar, inspeccionar y mantener documentos en el formato de documento portátil (PDF), además de esto permite:

- Generar documentos e informes basados en datos de un archivo XML o una base de datos.
- Crear mapas y libros, explotando numerosas funciones interactivas disponibles en PDF.
- Añadir marcadores, números de página, marcas de agua y otras funciones a documentos PDF existentes.
- Dividir o concatenar páginas de archivos PDF existentes.
- Completar formularios interactivos.
- Servir documentos PDF generados dinámicamente o manipulados a un navegador web.

iText es utilizado por los desarrolladores de Java, .NET, Android y GAE para mejorar sus aplicaciones con funcionalidad PDF (Sourceforge, 2018).

**5.3.7 iReport Designer.** La herramienta iReport es un diseñador de informes JasperReports escrito en Java. Este instrumento permite que los usuarios corrijan visualmente informes complejos con cartas, imágenes, subinformes, etc. iReport está además integrado con JFreeChart, una de la biblioteca gráficas OpenSource más difundida para Java. Los datos para imprimir pueden ser recuperados por varios caminos incluso múltiples uniones JDBC, TableModels, JavaBeans, XML, etc.

Características de iReport:

- 100% escrito en JAVA y además OPENSOURCE y gratuito.
- Maneja el 98% de las etiquetas de JasperReports.
- Permite diseñar con sus propias herramientas: rectángulos, líneas, elipses, campos de los textfields, cartas, subreportes.
- Soporta internacionalización nativamente.
- Browser de la estructura del documento.
- Recopilador y exportador integrados.
- Soporta JDBC.
- Soporta JavaBeans como orígenes de datos
- Incluye asistentes para crear automáticamente informes.
- Tiene asistentes para generar los subreportes.
- Tiene asistentes para las plantillas.
- Facilidad de instalación. Ampliado desde (informes et al., 2018).

## 5.4 Gitlab

GitLab es un proyecto de software libre de la compañía del mismo nombre que está programado en Ruby. Es una suite completa que nos permite:

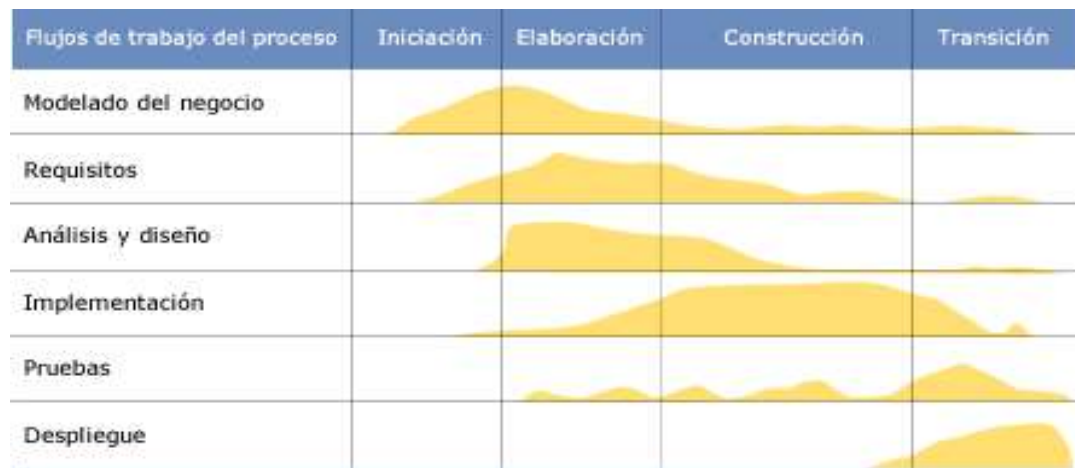
- Gestionar, administrar, crear y conectar nuestros repositorios con diferentes aplicaciones y hacer todo tipo de integraciones con ellas.
- Revisar diffs, los commits, y ver dónde se producen los cambios. Podemos crear y gestionar merge requests, comentar sobre ellos y hacerlo de manera colaborativa.
- Revisar toda la red de cambios del repositorio, viendo cómo se despliegan las ramas y a quién corresponde la responsabilidad de cada una de ellas.
- Incluye un completo sistema de integración continua que cubre todas las fases, según (Qué es GitLab 2018) desde el test hasta el despliegue y que además es compatible con Docker, tanto como para correr todos los tests como para crear máquinas Docker dentro del registro de contenedores que provee GitLab.

## 6. Metodología

### **Metodología Rational Unified Process - RUP**

La metodología RUP utiliza el enfoque de la orientación a objetos en su diseño, está diseñado y documentado el uso de la notación UML (Unified Modeling Language) para ilustrar los procesos en acción. Utiliza técnicas y prácticas probadas comercialmente (Aguilar, Cancino & González, 2018, pg. 3).

Es un proceso considerado pesado y preferentemente aplicable a grandes equipos de desarrollo y grandes proyectos, pero el hecho de que es ampliamente personalizable permite adaptarse a proyectos de cualquier escala (Aguilar, Cancino & González, 2018, pg. 3).



**Figura 6.** Modelo de RUP. Fuente: Wikipedia (s.f). [https://upload.wikimedia.org/wikipedia/commons/thumb/4/4d/Rup\\_espanol.gif/300px-Rup\\_espanol.gif](https://upload.wikimedia.org/wikipedia/commons/thumb/4/4d/Rup_espanol.gif/300px-Rup_espanol.gif)

### Etapa de Análisis

En esta etapa se busca definir el propósito y acordar el alcance del proyecto, identificando los riesgos asociados a este; proponiendo una visión muy general de la arquitectura del software.

### Etapa de Diseño

En esta etapa, los requisitos esenciales del sistema se transforman en los casos de uso. Además, se diseña el modelo de datos, bases de datos e interfaces documentándose mediante diagramas UML.

### Etapa de Implementación

En esta etapa se inicia el desarrollo del código fuente en la herramienta web. Debido a la metodología planteada se pueden realizar cambios en los requerimientos del sistema en esta etapa si es necesario.

## **Etapa de Documentación**

En esta última fase se elaboran los diferentes documentos que soporten el sistema realizado, y se hará en paralelo junto con la fase de implementación para tener una mejor organización y calidad en el manejo del sistema. La documentación en proyectos como este es sumamente importante porque de ésta depende en gran medida el éxito y uso masivo del sistema.

## **7. Análisis y Diseño del Sistema de Información**

Para el diseño e implementación del desarrollo de una herramienta web para generación, actualización y consulta de hojas de vida de usuarios mediante la obtención de datos, se realizó la toma de requerimientos y especificación de requisitos para los módulos a implementar en la herramienta web.

### **7.1 Especificación de Requerimientos**

**7.1.1 Requisitos Funcionales.** Especificaciones destinadas a cubrir los siguientes aspectos:

- El sistema debe permitir actualizar y editar las hojas de vida por los usuarios.
- El sistema puede ser consultado por el usuario y administrador del sistema que este registrado y autenticado.
- El sistema debe permitir al usuario extraer de manera automatizada la información del CvLac o del ResearchGate.
- El sistema debe permitir que el usuario avale la información que fue actualizada por el sistema.

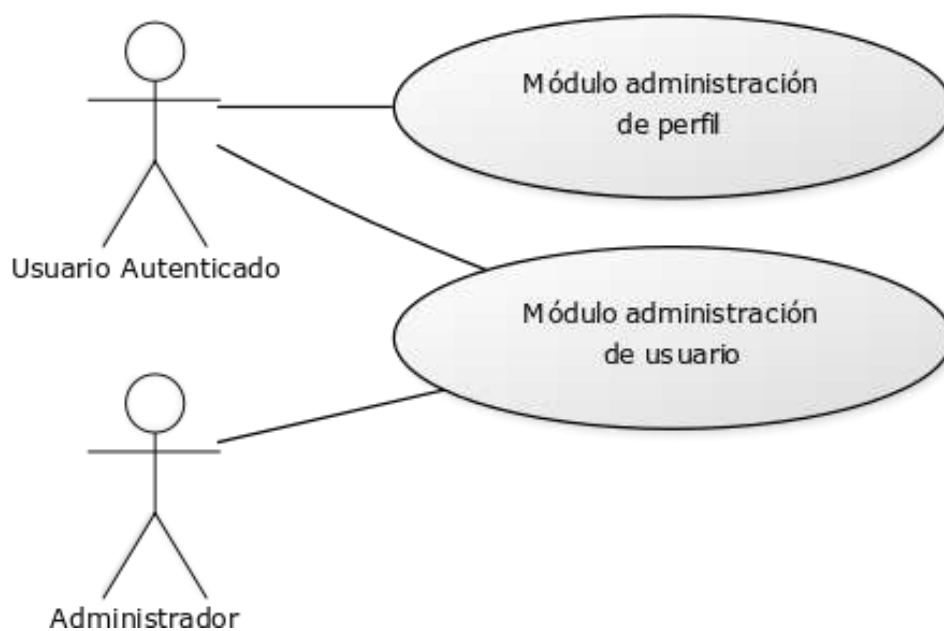
**7.1.2 Requisitos No Funcionales.** Especificaciones destinadas a los aspectos de calidad del sistema:

- El sistema debe visualizarse y funcionar correctamente en cualquier navegador.
- El sistema no debe tardar más de 30 segundos en mostrar los resultados requeridos por el usuario.

## 7.2 Casos de Uso

Los casos de uso describen la interacción entre el usuario y el sistema.

**7.2.1 Identificación de Actores.** [Usuario Autenticado]- (Módulo administración de perfil),[Usuario Autenticado]- (Módulo administración de usuario),[Administrador] -(Módulo administración de usuario).



**Figura 7.** Identificación de actores.

- **Descripción del Actor Administrador.**

**Tabla 1.***Descripción del Actor Administrador*

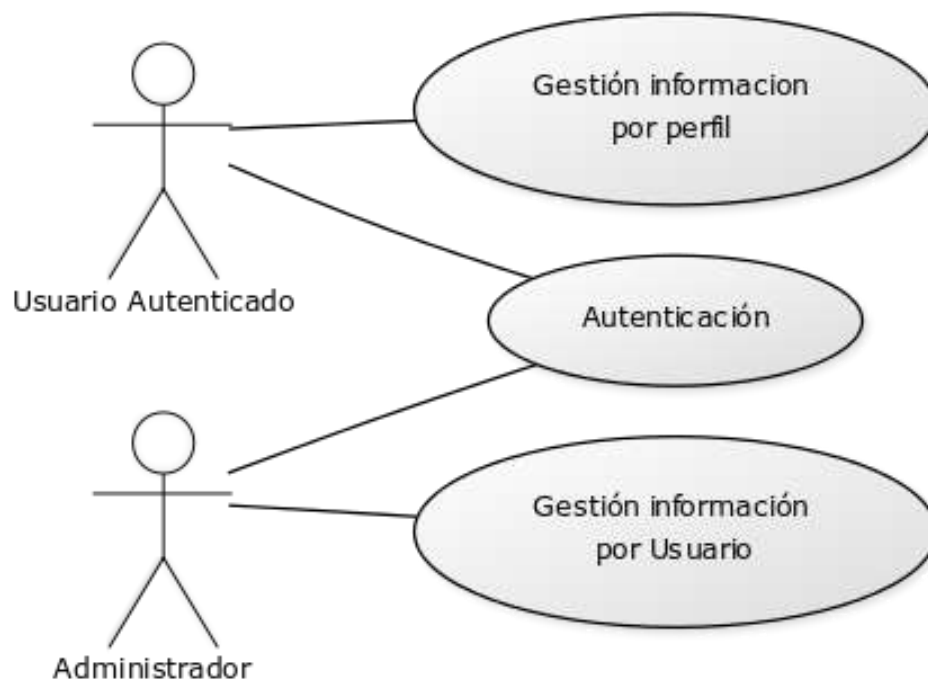
<b>Actor:</b>	<b>Administrador</b>
<b>Casos de Uso asociados:</b>	Autenticación, gestión información por usuarios.
<b>Descripción:</b>	El administrador es el encargado de gestionar la información de los usuarios seleccionando la página web a extraer la información (CvLac o ResearchGate).
<b>Tipo</b>	Principal

- **Descripción del Actor Usuario Autenticado.**

**Tabla 2.***Descripción del Actor Usuario Autenticado*

<b>Actor:</b>	<b>Usuario autenticado</b>
<b>Casos de Uso asociados:</b>	Autenticación, gestión información por perfil, representar relaciones.
<b>Descripción:</b>	El usuario autenticado tiene la capacidad de gestionar la información de su perfil seleccionando la página web a extraer la información (CvLac o ResearchGate), editarla, consultar su hoja de vida y visualizar la relación entre los usuarios.
<b>Tipo</b>	Principal

**7.2.2 Diagrama Modelo Casos de Usos.** [Usuario Autenticado]-(Autenticación), [Administrador] -(Autenticación),[Usuario Autenticado]-(Gestión información por perfil),[Administrador] -(Gestión información por Usuario)



**Figura 8.** Modelo de casos de uso.

### 7.2.2.1 Descripción Diagrama Modelo Casos de Uso.

**Tabla 3.**

Descripción Diagrama Modelo Casos de Uso CU01

Título del Caso de uso	Autenticación	Id del Caso de Uso	CU01
Entidades/Actores	Usuario autenticado, Administrador.		
Propósito	Permite al usuario o administrador ingresar y tener permisos de creación o edición.		
Descripción del caso de Uso	Este caso de uso está diseñado para ser utilizado para usuarios con permisos administrativos.		
Condiciones de entrada	<ul style="list-style-type: none"> <li>El usuario debe autenticarse.</li> </ul>		

<b>Flujo principal</b>	El caso de uso se inicia cuando se muestran dos cuadros de texto donde se debe ingresar dato de usuario y contraseña, además de un botón de inicio de sesión que envía los datos para la autenticación.
<b>Sub-flujo</b>	<b>SF-1 Datos</b> Se muestra dos cuadros de texto para el ingreso de datos del usuario. En la primera caja se ingresa el nombre del usuario y en la segunda se ingresa la contraseña.
<b>Condiciones de Salida</b>	<ul style="list-style-type: none"> <li>• Desplegar el menú correspondiente.</li> </ul>
<b>Excepciones</b>	<p><b>EX-1</b> <i>Debe especificar usuario</i>, el campo usuario se encuentra vacío debe ingresar su nombre de usuario en el campo.</p> <p><b>EX-2</b> <i>Debe especificar contraseña</i>, el campo contraseña se encuentra vacío debe ingresar su contraseña en el campo.</p> <p><b>EX-3</b> <i>Verifique su usuario y contraseña</i>, usuario o contraseña incorrectos, se solicita al usuario verificar e ingresar sus datos correctamente.</p>

**Tabla 4.**

Descripción Diagrama Modelo Casos de Uso CU02

<b>Título del Caso de uso</b>	<b>Gestión Información por perfil</b>	<b>Id del Caso de Uso</b>	<b>CU02</b>
<b>Entidades/Actores</b>	Usuario autenticado.		
<b>Propósito</b>	Permite al usuario, consultar, editar, extraer información y generar PDF de su hoja de vida.		
<b>Descripción del</b>	En este caso de uso se permite extraer información del		

<b>caso de Uso</b>	investigador que debe estar registrado en las páginas del CvLac o del ResearchGate, esta información se puede editar y consultar, además, generar PDF de su hoja de vida.
<b>Condiciones de entrada</b>	<ul style="list-style-type: none"> <li>• El usuario ingresa el enlace donde se encuentra su hoja de vida (CvLac o del ResearchGate).</li> <li>• El usuario da clic en actualizar y accede a la plataforma donde puede generar el PDF de su hoja de vida, consultar o editarla.</li> </ul>
<b>Flujo principal</b>	El caso de uso se inicia cuando el usuario se ha autenticado correctamente. Seguido ingresa el tipo de usuario que puede ser docente o estudiante, luego de esto ingresa el enlace del CvLac o ResearchGate donde se va a extraer su información. Cuando ingresa puede ver una ventana con información sugerida en las pestañas de estudios, distinciones y experiencia profesional. En la última pestaña se puede ver un resumen de la hoja de vida y la opción de generar el PDF.
<b>Sub-flujo</b>	<b>SF-1</b> Se muestra un cuadro de texto donde el usuario ingresa el enlace el cual se extraerá la información.
<b>Condiciones de Salida</b>	<ul style="list-style-type: none"> <li>• Guarda la Información registrada.</li> </ul>
<b>Excepciones</b>	<b>EX-1</b> <i>Se debe ingresar uno de los enlaces de donde se desea extraer.</i>

**Tabla 5.**

Descripción Diagrama Modelo Casos de Uso CU03

<b>Título del Caso de uso</b>	<b>Gestión Información por usuarios.</b>	<b>Id del Caso de Uso</b>	<b>CU03</b>
<b>Entidades/Actores</b>	Administrador.		
<b>Propósito</b>	Permite al administrador, consultar y generar PDF de la hoja de vida que desee.		
<b>Descripción del caso de uso</b>	En este caso de uso se permite extraer información de los usuarios registradas en las páginas del CvLac o del ResearchGate,		
<b>Condiciones de entrada</b>	<ul style="list-style-type: none"> <li>• El administrador ingresa el enlace (CvLac o del ResearchGate) donde se encuentra la hoja de vida del usuario.</li> <li>• El administrador hace clic en actualizar y accede a la plataforma donde puede consultar la hoja de vida.</li> </ul>		
<b>Flujo principal:</b>	El caso de uso se inicia cuando el Administrador se ha autenticado correctamente. En la última pestaña se puede ver un resumen de la hoja de vida y la opción de generar el PDF.		
<b>Sub-flujo</b>	<b>SF-1</b> Se muestra un cuadro de texto donde el usuario ingresa el enlace de donde se extraerá la información.		
<b>Excepciones</b>	<b>EX-1</b> <i>Se debe ingresar un enlace válido de donde se desea extraer(CvLac o ResearchGate).</i>		

**Tabla 6.**

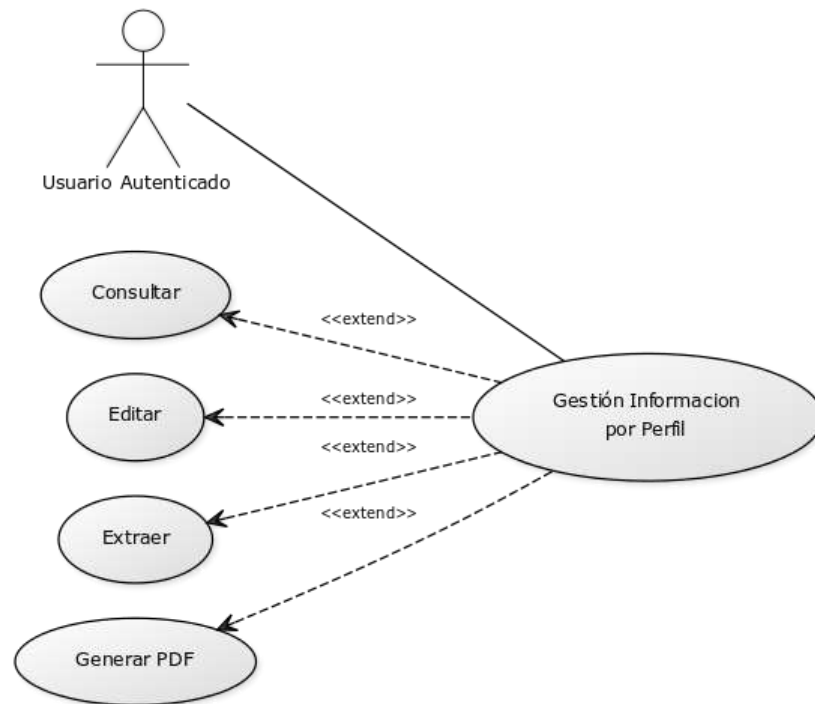
Descripción Diagrama Modelo Casos de Uso CU04

<b>Título del Caso de uso</b>	<b>Representar Relaciones.</b>	<b>Id del Caso de Uso</b>	<b>CU04</b>
<b>Entidades/Actores</b>	Usuario autenticado, visitante		
<b>Propósito</b>	Permite al usuario representar relaciones entre los usuarios.		

<b>Descripción del caso de uso</b>	En este caso de uso se permite representar relaciones entre los usuarios registrados en el portal FiloEn a partir del grupo de investigación que pertenece.
<b>Condiciones de entrada</b>	<ul style="list-style-type: none"> <li>• El usuario debe dar clic en representar relaciones para poder visualizarlas.</li> </ul>
<b>Flujo principal</b>	El caso de uso se inicia cuando el usuario da clic en el botón ver relaciones que está en el portal de FiloEn.
<b>Sub-flujo</b>	<b>SF-1</b> Se muestra un botón representar relaciones donde se le puede dar clic para visualizar las relaciones en forma de grafo de los grupos de investigación y de los usuarios.
<b>Condiciones de Salida</b>	<ul style="list-style-type: none"> <li>• Presenta un gráfico que representa las Relaciones entre los Investigadores y los grupos de investigación asociados.</li> </ul>
<b>Excepciones</b>	<b>E-1</b> .El usuario debe dar clic en el botón representar relaciones para poder visualizar el grafo.

### 7.2.3 Diagrama de Caso de Uso – Gestión de Información por Perfil.

(Consultar)<(Gestión Información por Perfil),(Editar)<(Gestión Información por Perfil),(Extraer)<(Gestión Información por Perfil),(Generar PDF)<(Gestión Información por Perfil),[Usuario Autenticado]-(Gestión Información por Perfil).



**Figura 9.** Diagrama de caso de uso - Gestión información por perfil

### 7.2.3.1 Descripción del Diagrama de Caso de Uso – Gestión Información por Perfil

**Tabla 7.**

*Descripción del Diagrama de Caso de Uso – Gestión Información por Perfil CU05*

Título del Caso de uso	Consultar	Id del Caso de Uso	CU05
<b>Entidades/Actores</b>	Usuario autenticado		
<b>Propósito</b>	Permite consultar la información deseada del usuario.		
<b>Descripción del caso de uso</b>	Este caso de uso permite consultar la información de los investigadores que han agregado la hoja de vida.		
<b>Condiciones de entrada</b>	<ul style="list-style-type: none"> <li>El usuario debe tener registrada la hoja de vida en el sitio.</li> </ul>		
<b>Flujo principal</b>	El caso de uso inicia cuando ingresa el enlace del CvLac del usuario y enseguida se da clic en el botón actualizar.		

<b>Sub-flujo</b>	<p><b>SF-1</b> Se muestra un cuadro donde el usuario debe ingresar el link del CvLac del usuario.</p> <p><b>SF-2</b> El usuario da clic en el botón actualizar enseguida puede consultar su información personal, idiomas, estudios, distinciones, experiencia profesional y un resumen de su hoja de vida</p>
<b>Condiciones de Salida</b>	<ul style="list-style-type: none"> <li>• Presenta un gráfico que representa las Relaciones entre los usuarios y los grupos de investigación asociados.</li> </ul>
<b>Excepciones</b>	<b>E-1</b> Debe dar clic en actualizar para visualizar la información.

**Tabla 8.**

*Descripción del Diagrama de Caso de Uso – Gestión Información por Perfil CU06*

<b>Título del Caso de uso</b>	<b>Editar</b>	<b>Id del Caso de Uso</b>	<b>CU06</b>
<b>Entidades/Actores</b>	Usuario autenticado		
<b>Propósito</b>	Permite editar la información deseada de los usuarios.		
<b>Descripción del caso de uso</b>	En este caso de uso permite editar la información de los usuarios que está registrada en el sitio.		
<b>Condiciones de entrada</b>	<ul style="list-style-type: none"> <li>• El usuario debe tener registrada la hoja de vida en el sitio.</li> </ul>		
<b>Flujo principal</b>	El caso de uso inicia cuando ingresa el link del CvLac del usuario y enseguida se da clic en el botón actualizar.		
<b>Sub-flujo</b>	<p><b>SF-1</b> Se muestra un cuadro donde el usuario debe ingresar el link del CvLac del Usuario.</p> <p><b>SF-2</b> El usuario da clic en el botón actualizar para poder acceder a la plataforma y editar la información que desee ya sea</p>		

	información personal, idiomas, formación académica, distinciones, experiencia laboral, producción bibliográfica (libros, artículos, colaboraciones, proyectos dirigidos, reconocimientos).
<b>Condiciones de Salida</b>	<b>E-1</b> Debe dar clic en actualizar para visualizar su información y poder editar los diferentes campos.
<b>Excepciones</b>	

**Tabla 9.**

*Descripción del Diagrama de Caso de Uso – Gestión Información por Perfil CU07*

<b>Título del Caso de uso</b>	<b>Extraer</b>	<b>Id del Caso de Uso</b>	<b>CU07</b>
<b>Entidades/Actores</b>	Usuario autenticado		
<b>Propósito</b>	Permite extraer la información deseada de los usuarios.		
<b>Descripción del caso de uso</b>	En este caso de uso permite extraer información de los usuarios que está registrada en el sitio web de Colciencias-CvLac o del ResearchGate.		
<b>Condiciones de entrada</b>	<ul style="list-style-type: none"> <li>El usuario debe tener registrada la hoja de vida en el CvLac o en ResearchGate.</li> </ul>		
<b>Flujo principal</b>	El caso de uso inicia cuando ingresa el enlace del CvLac o del ResearchGate del usuario y enseguida se da clic en el botón actualizar, paso siguiente en estudios, distinciones y experiencia profesional aparece una ventana de información sugerida con la información que se extrajo de alguna de las páginas y que nos servirá para llenar los campos.		

<b>Sub-flujo</b>	<p><b>SF-1</b> Se muestra un cuadro donde el usuario debe ingresar el enlace del CvLac del usuario.</p> <p><b>SF-2</b> El usuario da clic en el botón actualizar para extraer la información.</p> <p><b>SF-3</b> En estudios, distinciones y experiencia profesional el usuario puede visualizar la información.</p>
<b>Condiciones de Salida</b>	<p><b>E-1</b> Debe dar clic en actualizar para visualizar la información extraída.</p> <p><b>E-2</b> El usuario debe tener su hoja de vida registrada en el CvLac o en ResearchGate.</p>
<b>Excepciones</b>	

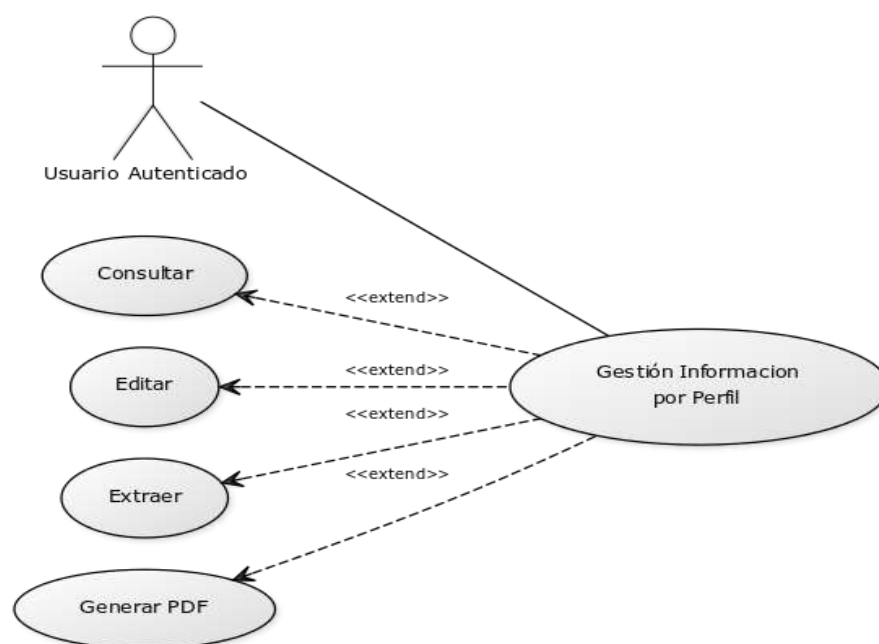
**Tabla 10.**

*Descripción del Diagrama de Caso de Uso – Gestión Información por Perfil CU08*

<b>Título del Caso de uso</b>	<b>Generar PDF</b>	<b>Id del Caso de Uso</b>	<b>CU08</b>
<b>Entidades/Actores</b>	Usuario autenticado		
<b>Propósito</b>	Permite generar PDF de su hoja de vida.		
<b>Descripción del caso de uso</b>	En este caso de uso permite generar el PDF de su hoja de vida.		
<b>Condiciones de entrada</b>	<ul style="list-style-type: none"> <li>El usuario debe agregar su información a la hoja de vida.</li> </ul>		
<b>Flujo principal</b>	El caso de uso inicia cuando el usuario ingresa el enlace del CvLac y enseguida se da clic en el botón actualizar, paso seguido puede ver el resumen de su hoja de vida si ya la tiene registrada y		

	generar el PDF; de lo contrario agrega la información antes de generar el PDF.
<b>Sub-flujo</b>	
<b>Condiciones de Salida</b>	<p><b>SF-1</b> Se muestra un cuadro donde el usuario debe ingresar el enlace del CvLac del mismo.</p> <p><b>SF-2</b> El usuario da clic en el botón actualizar para extraer la información.</p> <p><b>SF-3</b> el usuario da clic en resumen para visualizar el resumen de su hoja de vida.</p> <p><b>SF-4</b> el usuario da clic en PDF para generar el PDF de su hoja de vida.</p>
<b>Excepciones</b>	<b>E-1</b> Debe tener la hoja de vida registrada en la plataforma.

#### 7.2.4 Diagrama de Caso de Uso – Gestión Información por Administrar.



**Figura 10.** Diagrama de casos de uso - Gestión información por administrar

### 7.2.4.1 Descripción del Diagrama de Caso de Uso – Gestión Información por Administrador.

**Tabla 11.**

*Descripción del Diagrama de caso de uso – Gestión información por administrador CU09*

<b>Título del Caso de uso</b>	<b>Consultar.</b>	<b>Id del Caso de Uso</b>	<b>CU09</b>
<b>Entidades/Actores</b>	Administrador		
<b>Propósito</b>	Permite consultar la información deseada de cualquier usuario.		
<b>Descripción del caso de uso</b>	Este caso de uso permite consultar la información de los usuarios que han agregado la hoja de vida.		
<b>Condiciones de entrada</b>	<ul style="list-style-type: none"> <li>El usuario debe tener registrada la hoja de vida en el sitio.</li> </ul>		
<b>Flujo principal</b>	El caso de uso inicia cuando ingresa el enlace del CvLac o del ResearchGate del usuario y a continuación se da clic en el botón actualizar.		
<b>Sub-flujo</b>	<p><b>SF-1</b> Se muestra un cuadro donde el usuario debe ingresar el enlace del CvLac del investigador.</p> <p><b>SF-2</b> El usuario da clic en el botón actualizar para poder acceder a la plataforma y editar la información que desee ya sea información personal, idiomas, formación académica, distinciones, experiencia laboral, producción bibliográfica (libros, artículos, colaboraciones, proyectos dirigidos, reconocimientos).</p>		
<b>Condiciones de Salida</b>			
<b>Excepciones</b>	<b>E-1</b> Debe dar clic en actualizar para visualizar la información.		

**Tabla 12.***Descripción del Diagrama de caso de uso – Gestión información por administrador CU10*

<b>Título del Caso de uso</b>	<b>Editar</b>	<b>Id del Caso de Uso</b>	<b>CU10</b>
<b>Entidades/Actores</b>	Administrador		
<b>Propósito</b>	Permite editar la información deseada de los usuarios.		
<b>Descripción del caso de uso</b>	En este caso de uso permite editar la información de los usuarios que está registrada en el sitio.		
<b>Condiciones de entrada</b>	<ul style="list-style-type: none"> <li>• El usuario debe tener registrada la hoja de vida en el sitio.</li> </ul>		
<b>Flujo principal</b>	El caso de uso inicia cuando ingresa el enlace del CvLac o del ResearchGate del usuario y enseguida se da clic en el botón actualizar.		
<b>Sub-flujo</b>	<p><b>SF-1</b> Se muestra un cuadro donde el usuario debe ingresar el enlace del CvLac.</p> <p><b>SF-2</b> El usuario da clic en el botón actualizar para poder acceder a la plataforma y editar la información que desee ya sea información personal, idiomas, formación académica, distinciones, experiencia laboral, producción bibliográfica (libros, artículos, colaboraciones, proyectos dirigidos, reconocimientos).</p>		
<b>Condiciones de Salida</b>			
<b>Excepciones</b>	<b>E-1</b> Debe dar clic en actualizar para visualizar su información y poder editar los diferentes campos.		

**Tabla 13.***Descripción del Diagrama de caso de uso – Gestión información por administrador CU11*

<b>Título del Caso de uso</b>	<b>Extraer</b>	<b>Id del Caso de Uso</b>	<b>CU11</b>
<b>Entidades/Actores</b>	Administrador		
<b>Propósito</b>	Permite extraer la información deseada de los usuarios.		
<b>Descripción del caso de uso</b>	En este caso de uso permite extraer información de los usuarios que está registrada en el sitio web de Colciencias-CvLac o del ResearchGate.		
<b>Condiciones de entrada</b>	<ul style="list-style-type: none"> <li>El usuario debe tener registrada la hoja de vida en el CvLac o en ResearchGate.</li> </ul>		
<b>Flujo principal</b>	El caso de uso inicia cuando ingresa el enlace del CvLac o del ResearchGate del usuario y enseguida se da clic en el botón actualizar, a continuación en estudios, distinciones y experiencia profesional aparece una ventana de información sugerida con la información que se extrajo de alguna de las páginas y que nos servirá para llenar los campos.		
<b>Sub-flujo</b>	<p><b>SF-1</b> Se muestra un cuadro donde el usuario debe ingresar el enlace del CvLac.</p> <p><b>SF-2</b> El usuario da clic en el botón actualizar para extraer la información.</p> <p><b>SF-3</b> En estudios, distinciones y experiencia profesional el usuario puede visualizar la información extraída en una ventana de información sugerida.</p>		
<b>Condiciones de Salida</b>			
<b>Excepciones</b>	<p><b>E-1</b> Debe dar clic en actualizar para visualizar la información extraída.</p> <p><b>E-2</b> el usuario debe tener su hoja de vida registrada en el CvLac o en ResearchGate.</p>		

**Tabla 14.**

*Descripción del Diagrama de caso de uso – Gestión información por administrador CU12*

<b>Título del Caso de uso</b>	<b>Generar PDF</b>	<b>Id del Caso de Uso</b>	<b>CU12</b>
<b>Entidades/Actores</b>	Administrador		
<b>Propósito</b>	Permite generar PDF de la hoja de vida de los usuarios.		
<b>Descripción del caso de uso</b>	En este caso de uso permite generar el PDF de la hoja de vida de los usuarios.		
<b>Condiciones de entrada</b>	<ul style="list-style-type: none"> <li>• El administrador debe agregar su información a la hoja de vida.</li> </ul>		
<b>Flujo principal</b>	El caso de uso inicia cuando ingresa el enlace del CvLac del usuario y enseguida se da clic en el botón actualizar, a continuación puede ver el resumen de su hoja de vida si ya la tiene registrada y generar el PDF; de lo contrario agrega la información antes de generar el PDF.		
<b>Sub-flujo</b>	<p><b>SF-1</b> Se muestra un cuadro donde el usuario debe ingresar el link del CvLac del usuario.</p> <p><b>SF-2</b> El usuario da clic en el botón actualizar para extraer la información.</p> <p><b>SF-3</b> el usuario da clic en resumen para visualizar el resumen de su hoja de vida.</p> <p><b>SF-4</b> el usuario da clic en PDF para generar el PDF de su hoja de vida.</p>		
<b>Condiciones de Salida</b>			
<b>Excepciones</b>	<b>E-1</b> Debe tener la hoja de vida registrada en la plataforma.		

## **8. Diseño Componentes**

### **8.1 Modelo Entidad Relación.**

#### **8.1.1 Diagrama Entidad Relación.** Diligenciamiento opcional:

Se diligencia cuando se adicionan Entidades o cuando se adicionan o eliminan relaciones entre entidades. Se debe diagramar el modelo entidad relación mostrando las nuevas tablas o las relaciones creadas a partir de las nuevas llaves foráneas.

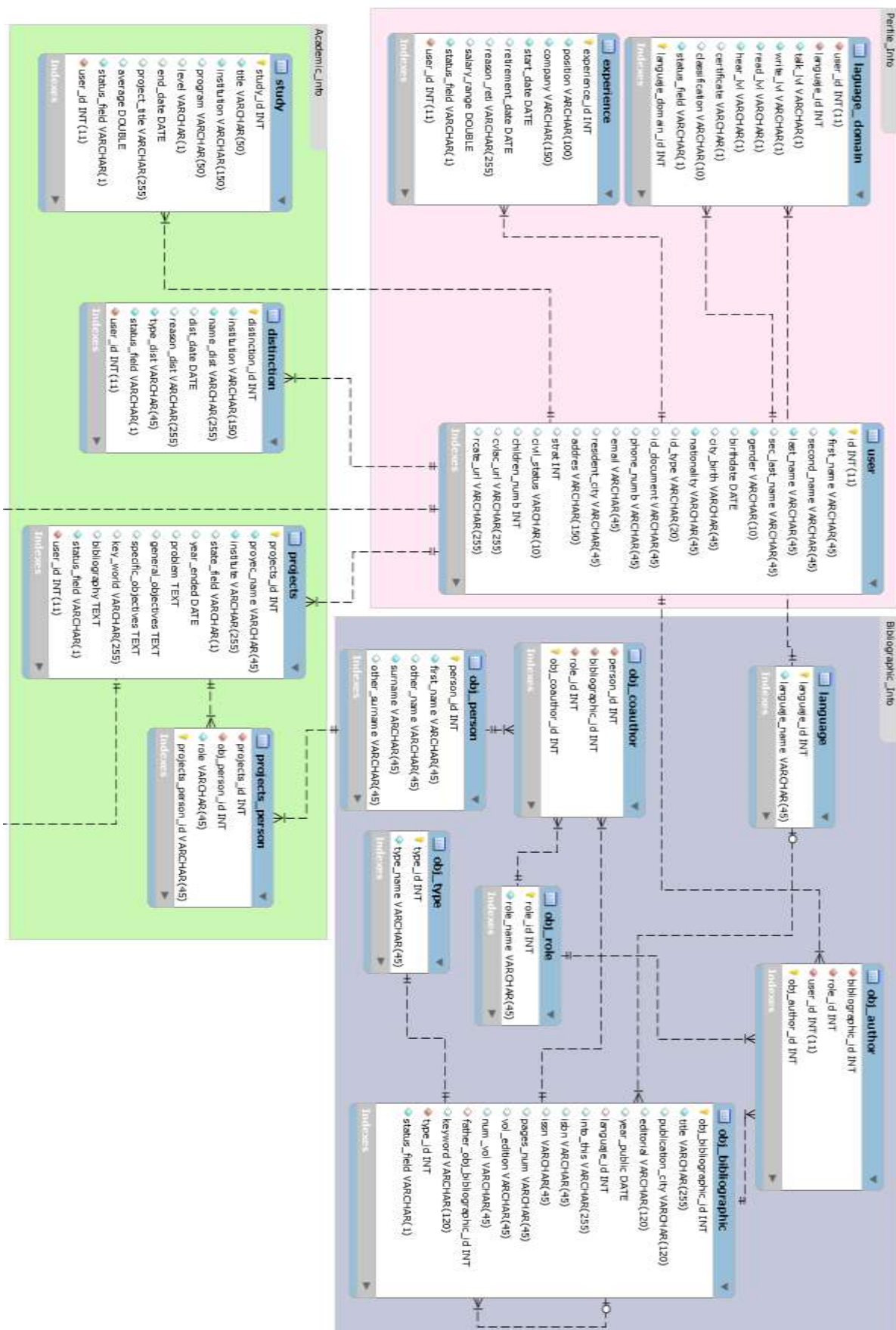


Figura 11. Diagrama Entidad Relación

### 8.1.2 Entidades de bases de datos. Diligenciamiento Opcional:

Se debe diligenciar cuando:

- Se adiciona o elimina una Entidad.
- Se adicionan, eliminan o modifican columnas de una entidad existente.
- Se adicionan, eliminan o modifican constraints (llaves primarias, foráneas o check).
- Se adicionan, eliminan o modifican índices.

A continuación, se detalla los campos y características de cada entidad

**Tabla 15.**

*Entidad : language\_domain*

<b>Entidad : language_domain</b>						
<b>Atributos</b>	<b>Tipo Dato</b>	<b>Precisión</b>	<b>Escala</b>	<b>Nulidad</b>	<b>Llave</b>	<b>Descripción</b>
user_id	INT	11	0	N	PK	Id Usuario
language_id	INT		0	N	PK	Id Lenguaje
talk_lvl	VARCHAR	1	0	N		Nivel de Conversación
write_lvl	VARCHAR	1	0	N		Nivel de Escritura
read_lvl	VARCHAR	1	0	N		Nivel de Lectura
hear_lvl	VARCHAR	1	0	N		Nivel de Escucha
Certificate	VARCHAR	1	0	S		Lenguaje Certificado
classification	VARCHAR	10	0	S		Clasificación Lenguaje
status_field	VARCHAR	1	0	N		Estado

**Tabla 16.***Entidad: language\_domain Key*

<b>Entidad : language_domain</b>		
<b>Indice : fk_language_domain_user1_idx user</b>		<b>Tipo de índice : PRIMARY KEY</b>
<b>Atributos</b>	<b>Nulidad</b>	<b>Descripción</b>
user_id	N	Id del usuario.

**Tabla 17.***Entidad: experience*

<b>Entidad : experience</b>						
<b>Atributos</b>	<b>Tipo Dato</b>	<b>Precisión</b>	<b>Escala</b>	<b>Nulidad</b>	<b>Llave</b>	<b>Descripción</b>
user_id	INT	11	0	N	PK	Id Usuario
experience_id	INT		0	N	PK	Id experiencia Laboral
position	VARCHAR	100	0	N		Cargo
company	VARCHAR	150	0	N		Compañía
start_date	DATE		0	N		Fecha de Inicio
retirement_date	DATE		0	S		Fecha de Retiro
reason_reti	VARCHAR	255	0	S		Razón Retiro
salary_range	DOUBLE		0	S		Salario Promedio
status_field	VARCHAR	1	0	N		Estado

**Tabla 18.***Entidad: experience Key*

<b>Entidad : experience</b>		
<b>Indice : fk_experience_user1_idx user</b>		<b>Tipo de índice : PRIMARY KEY</b>
<b>Atributos</b>	<b>Nulidad</b>	<b>Descripción</b>
User_id	N	Id del usuario.

**Tabla 19.***Entidad: user*

<b>Entidad : user</b>						
<b>Atributos</b>	<b>Tipo Dato</b>	<b>Precisión</b>	<b>Escala</b>	<b>Nulidad</b>	<b>Llave</b>	<b>Descripción</b>
user_id	INT	11	0	N	PK	Id Usuario
first_name	INT	45	0	N		Primer Nombre
second_name	VARCHAR	45	0	S		Segundo Nombre
last_name	VARCHAR	45	0	N		Apellido
sec_last_name	VARCHAR	45	0	S		Segundo Apellido
Gender	VARCHAR	10	0	N		Género
birthdate	DATE	0	0	S		Fecha de Nacimiento
city_birth	VARCHAR	45	0	S		Ciudad de Nacimiento
nationality	VARCHAR	45	0	N		Nacionalidad
id_type	VARCHAR	20	0	S		Tipo de Documento de Identificación
id_document	VARCHAR	45	0	S		Número de Documento de Identificación
phone_num	VARCHAR	45	0	S		Número telefónico
Email	VARCHAR	45	0	S		Correo Electrónico

resident_city	VARCHAR	45	0	S		Ciudad de Residencia
address	VARCHAR	150	0	S		Dirección
Strat	INT		0	S		Estrato
civil_status	VARCHAR	10	0	S		Estado civil
children_numb	INT		0	S		Número de Hijos
cvlac_url	VARCHAR	255	0	S		Dirección enlace CVLAC
rcate_url	VARCHAR	255	0	S		Dirección enlace Research Gate

**Tabla 20.***Entidad: study*

Entidad : study						
Atributos	Tipo Dato	Precisión	Escala	Nulidad	Llave	Descripción
user_id	INT	11	0	N		Id Usuario
study_id	INT		0	N	PK	Id de estudio
Title	VARCHAR	50	0	N		Título Obtenido
institution	VARCHAR	150	0	N		Nombre Institución
program	VARCHAR	50	0	S		Programa Académico
Level	VARCHAR	1	0	S		Nivel Académico
end_date	DATE		0	S		Fecha de Finalización
project_title	VARCHAR	255	0	S		Título del proyecto
average	DOUBLE		0	S		Promedio
status_field	VARCHAR	1	0	N		Estado

**Tabla 21.***Entidad: study key*

<b>Entidad : study</b>		
<b>Indice : fk_study_user1_idx user</b>		<b>Tipo de índice : PRIMARY KEY</b>
<b>Atributos</b>	<b>Nulidad</b>	<b>Descripción</b>
User_id	N	Id del usuario.

**Tabla 22.***Entidad: distinction*

<b>Entidad : distinction</b>						
<b>Atributos</b>	<b>Tipo Dato</b>	<b>Precisión</b>	<b>Escala</b>	<b>Nulidad</b>	<b>Llave</b>	<b>Descripción</b>
user_id	INT	11	0	N		Id Usuario
Distinction_id	INT		0	N	PK	Id de estudio
institution	VARCHAR	150	0	N		Título Obtenido
name_dist	VARCHAR	255	0	N		Nombre Institución
dist_date	DATE		0	S		Programa Académico
reason_dist	VARCHAR	255	0	S		Nivel Académico
type_dist	VARCHAR	45	0	N		Fecha de Finalización
status_field	VARCHAR	1	0	N		Estado

**Tabla 23.***Entidad: distinction key*

<b>Entidad : distinction</b>		
<b>Indice : fk_distinction_user1_idx user</b>		<b>Tipo de índice : PRIMARY KEY</b>
<b>Atributos</b>	<b>Nulidad</b>	<b>Descripción</b>
User_id	N	Id del usuario.

**Tabla 24.***Entidad: Projects*

<b>Entidad : projects</b>						
<b>Atributos</b>	<b>Tipo Dato</b>	<b>Precisión</b>	<b>Escala</b>	<b>Nulidad</b>	<b>Llave</b>	<b>Descripción</b>
user_id	INT	11	0	N		Id Usuario
projects_id	INT		0	N	PK	Id de proyecto
proyec_name	VARCHAR	45	0	N		Nombre de Proyecto
institute	VARCHAR	255	0	S		Nombre Institución
state_field	VARCHAR	1	0	S		Estado del campo
year_ended	DATE		0	S		Año finalización
problem	TEXT		0	S		Problemática
general_objectives	TEXT		0	S		Objetivos generales
specific_objectives	TEXT		0	S		Objetivos Específicos
key_word	VARCHAR	255	0	S		Palabras Claves
Bibliography	TEXT		0	S		Bibliografía
status_field	VARCHAR	1	0	N		Estado

**Tabla 25.***Entidad: projects key*

<b>Entidad : projects</b>		
<b>Indice: fk_projects_user1_idx user</b>		<b>Tipo de índice : PRIMARY KEY</b>
<b>Atributos</b>	<b>Nulidad</b>	<b>Descripción</b>
User_id	N	Id del usuario.

**Tabla 26.***Entidad: projects\_person*

Entidad : projects_person						
Atributos	Tipo Dato	Precisión	Escala	Nulidad	Llave	Descripción
projects_id	INT		0	N	PK	Id de proyecto
obj_person_id	INT		0	N	PK	Id Objeto Persona
Role	VARCHAR	45	0	N		Rol

**Tabla 27.***Entidad: projects\_person key obj persona*

Entidad : projects_person		
<b>Indice:</b>	<b>fk_projects_obj_person_obj_person1_idx</b>	<b>Tipo de índice : PRIMARY KEY</b>
<b>obj_person</b>		
Atributos	Nulidad	Descripción
obj_person_id	N	Id Objeto Persona.

**Tabla 28.***Entidad: projects\_person key id Project*

Entidad : projects_person		
<b>Indice:</b>	<b>fk_projects_obj_person_projects1_idx projects</b>	<b>Tipo de índice : PRIMARY KEY</b>
<b>project_id</b>		
Atributos	Nulidad	Descripción
project_id	N	Id del proyecto.

**Tabla 29.***Entidad: language*

Entidad : language						
Atributos	Tipo Dato	Precisión	Escala	Nulidad	Llave	Descripción
language_id	INT		0	N	PK	Id Lenguaje
language_name	VARCHAR		0	N		Id de proyecto

**Tabla 30.***Entidad: obj\_author*

<b>Entidad : obj_author</b>						
<b>Atributos</b>	<b>Tipo Dato</b>	<b>Precisión</b>	<b>Escala</b>	<b>Nulidad</b>	<b>Llave</b>	<b>Descripción</b>
bibliographic_id	INT		0	N	PK	Id bibliográfica
role_id	INT		0	N	PK	Id Rol
user_id	INT	11	0	N	PK	Id usuario

**Tabla 31.***Entidad: obj\_author key rol*

<b>Entidad : obj_author</b>		
<b>Indice: fk_obj_author_obj_role1_idx obj_role</b>		<b>Tipo de índice : PRIMARY KEY</b>
<b>Atributos</b>	<b>Nulidad</b>	<b>Descripción</b>
rol_id	N	Id Rol

**Tabla 32.***Entidad: obj\_author id bibliographic*

<b>Entidad : obj_author</b>		
<b>Indice: fk_obj_bibliographic_profile_obj_bibliographic_idx obj_bibliographic</b>		<b>Tipo de índice : PRIMARY KEY</b>
<b>Atributos</b>	<b>Nulidad</b>	<b>Descripción</b>
bibliographic_id	N	Id bibliográfico

**Tabla 33.***Entidad : obj\_author key usuario*

<b>Entidad : obj_author</b>		
<b>Indice: fk_obj_author_user1_idx user</b>		<b>Tipo de índice : PRIMARY KEY</b>
<b>Atributos</b>	<b>Nulidad</b>	<b>Descripción</b>
user_id	N	Id Usuario

**Tabla 34.***Entidad : obj\_bibliographic*

<b>Entidad : obj_bibliographic</b>						
<b>Atributos</b>	<b>Tipo Dato</b>	<b>Precisión</b>	<b>Escala</b>	<b>Nulidad</b>	<b>Llave</b>	<b>Descripción</b>
obj_bibliographic_id	INT	11	0	N		Id objeto bibliográfico
Title	VARCHAR	255	0	N	PK	Título
publication_city	VARCHAR	120	0	S		Ciudad de Publicación
editorial	VARCHAR	120	0	S		Editorial
year_public	DATE		0	S		Año de Publicación
lenguaje_id	INT		0	S		Id Lenguaje
into_this	VARCHAR	255	0	S		En este libro
Isbn	VARCHAR	45	0	S		International Standard Book Number

Issn	VARCHAR	45	0	S		International Standard Serial Number
pages_num	VARCHAR	45	0	S		Número de páginas
vol_edition	VARCHAR	45	0	S		Edición del volumen
num_vol	VARCHAR	45	0	S		Número del volumen
father_obj_bibliographic_id	INT		0	S		Id Objeto bibliográfico padre
keyword	VARCHAR	120	0	S		Palabras clave
type_id	INT		0	N		Id Tipo
status_field	VARCHAR	1	0	N		Estado

**Tabla 35.***Entidad : obj\_bibliographic key father*

<b>Entidad : obj_bibliographic</b>		
<b>Indice: fk_cv_obj_bibliographic_cv_obj_bibliographic1_idx</b>	<b>Tipo de índice :</b>	
<b>obj_bibliographic</b>	<b>PRIMARY KEY</b>	
<b>Atributos</b>	<b>Nulidad</b>	<b>Descripción</b>
father_obj_bibliographic_id	S	Objeto bibliográfico Padre

**Tabla 36.***Entidad : obj\_bibliographic key id*

Entidad : obj_bibliographic		
Indice: fk_cv_obj_bibliographic_cv_obj_type1_idx obj_type		Tipo de índice : PRIMARY KEY
Atributos	Nulidad	Descripción
Type_id	N	Tipo ID

**Tabla 37.***Entidad : obj\_bibliographic ISBN*

Entidad : obj_bibliographic		
Indice: isbn_UNIQUE obj_bibliographic		Tipo de índice : PRIMARY KEY
Atributos	Nulidad	Descripción
isbn_UNIQUE	S	International Standard Book Number

**Tabla 38.***Entidad : obj\_bibliographic Index*

Entidad : obj_bibliographic		
Indice: issn_UNIQUE obj_bibliographic		Tipo de índice : UNIQUE
Atributos	Nulidad	Descripción
issn_UNIQUE	S	International Standard Serial Number

**Tabla 39.***Entidad : obj\_bibliographic key language*

Entidad : obj_bibliographic		
Indice: fk_obj_bibliographic_language1_idx language		Tipo de índice : UNIQUE
Atributos	Nulidad	Descripción
language_id	N	ID lenguaje

**Tabla 40.***Entidad : obj\_type*

<b>Entidad : obj_type</b>						
<b>Atributos</b>	<b>Tipo Dato</b>	<b>Precisión</b>	<b>Escala</b>	<b>Nulidad</b>	<b>Llave</b>	<b>Descripción</b>
type_id	INT		0	N	PK	Id tipo
type_name	VARCHAR	45	0	N		Nombre tipo

**Tabla 41.***Entidad : investigation\_lines\_projects*

<b>Entidad : investigation_lines_projects</b>						
<b>Atributos</b>	<b>Tipo Dato</b>	<b>Precisión</b>	<b>Escala</b>	<b>Nulidad</b>	<b>Llave</b>	<b>Descripción</b>
investigation_lines_id	INT		0	N	PK	id líneas de investigación
project_id	INT		0	N	PK	id proyecto

**Tabla 42.***Entidad : investigation\_lines\_projects index*

<b>Entidad : investigation_lines_projects</b>		
<b>Indice:</b>	<b>fk_investigation_lines_projects_projects1_idx</b>	<b>Tipo de índice :</b>
<b>projects</b>		<b>PRIMARY KEY</b>
<b>Atributos</b>	<b>Nulidad</b>	<b>Descripción</b>
project_id	N	id proyecto

**Tabla 43.***Entidad : investigation\_lines\_projects Key id*

<b>Entidad : investigation_lines_projects</b>		
<b>Indice:</b> <b>fk_investigation_lines_projects_investigation_lines1_idx</b> <b>investigation_lines_groups</b>		<b>Tipo de índice :</b> <b>PRIMARY KEY</b>
<b>Atributos</b>	<b>Nulidad</b>	<b>Descripción</b>
investigation_lines_id	N	id líneas de investigación

**Tabla 44.***Entidad : investigation\_lines\_groups*

<b>Entidad : investigation_lines_groups</b>						
<b>Atributos</b>	<b>Tipo</b>	<b>Precisión</b>	<b>Escala</b>	<b>Nulidad</b>	<b>Llave</b>	<b>Descripción</b>
	<b>Dato</b>					
investigation_lines_id	INT		0	N	PK	id líneas de investigación
investigation group	INT		0	N	PK	grupo de investigación

**Tabla 45.***Entidad : investigation\_lines\_groups Grupos*

<b>Entidad : investigation_lines_groups</b>		
<b>Indice:</b> <b>fk_investigation_lines_has_investigation_group_investigatio_idx</b> <b>investigation_group</b>		<b>Tipo de índice :</b> <b>PRIMARY KEY</b>
<b>Atributos</b>	<b>Nulidad</b>	<b>Descripción</b>
investigation group	N	grupo de investigación

**Tabla 46.***Entidad : investigation\_lines*

<b>Entidad : investigation_lines</b>						
<b>Atributos</b>	<b>Tipo Dato</b>	<b>Precisión</b>	<b>Escala</b>	<b>Nulidad</b>	<b>Llave</b>	<b>Descripción</b>
investigation_lines_id	INT		0	N	PK	id líneas de investigación
line_name	VARCHAR	45	0	N		nombre de línea
description	VARCHAR	255	0	S		descripción

**Tabla 47.***Entidad : investigation\_group*

<b>Entidad : investigation_group</b>						
<b>Atributos</b>	<b>Tipo Dato</b>	<b>Precisión</b>	<b>Escala</b>	<b>Nulidad</b>	<b>Llave</b>	<b>Descripción</b>
group_id	INT		0	N	PK	id grupo
name_group	VARCHAR	128	0	S		nombre de grupo
director	VARCHAR	45	0	S		director
institution	VARCHAR	150	0	S		Institución
creation_year	DATE		0	S		Año de Creación
knowledge_area	VARCHAR	45	0	S		Área de Conocimiento
category	VARCHAR	45	0	S		Categoría
logo	MEDIUMBLOB		0	S		Logo
web_page	VARCHAR	128	0	S		Página Web
email_group	VARCHAR	99	0	S		Correo Grupo

**Tabla 48.***Entidad : user investigation\_group*

<b>Entidad : user investigation_group</b>						
<b>Atributos</b>	<b>Tipo Dato</b>	<b>Precisión</b>	<b>Escala</b>	<b>Nulidad</b>	<b>Llave</b>	<b>Descripción</b>
user_id	INT		0	N	PK	id usuario
group_id	VARCHAR	45	0	N	PK	id Grupo

**Tabla 49.***Entidad : user\_investigation\_group key id*

<b>Entidad : user_investigation_group</b>		
<b>Atributos</b>	<b>Nulidad</b>	<b>Descripción</b>
group_id	N	id Grupo

<b>Entidad : user_investigation_group</b>	
<b>Indice:</b>	<b>Tipo de índice :</b>
<b>fk_user_has_investigation_group_investigation_group1_idx</b> <b>investigation_group</b>	<b>PRIMARY KEY</b>

**Tabla 50.***Entidad : user\_investigation\_group key id usuario*

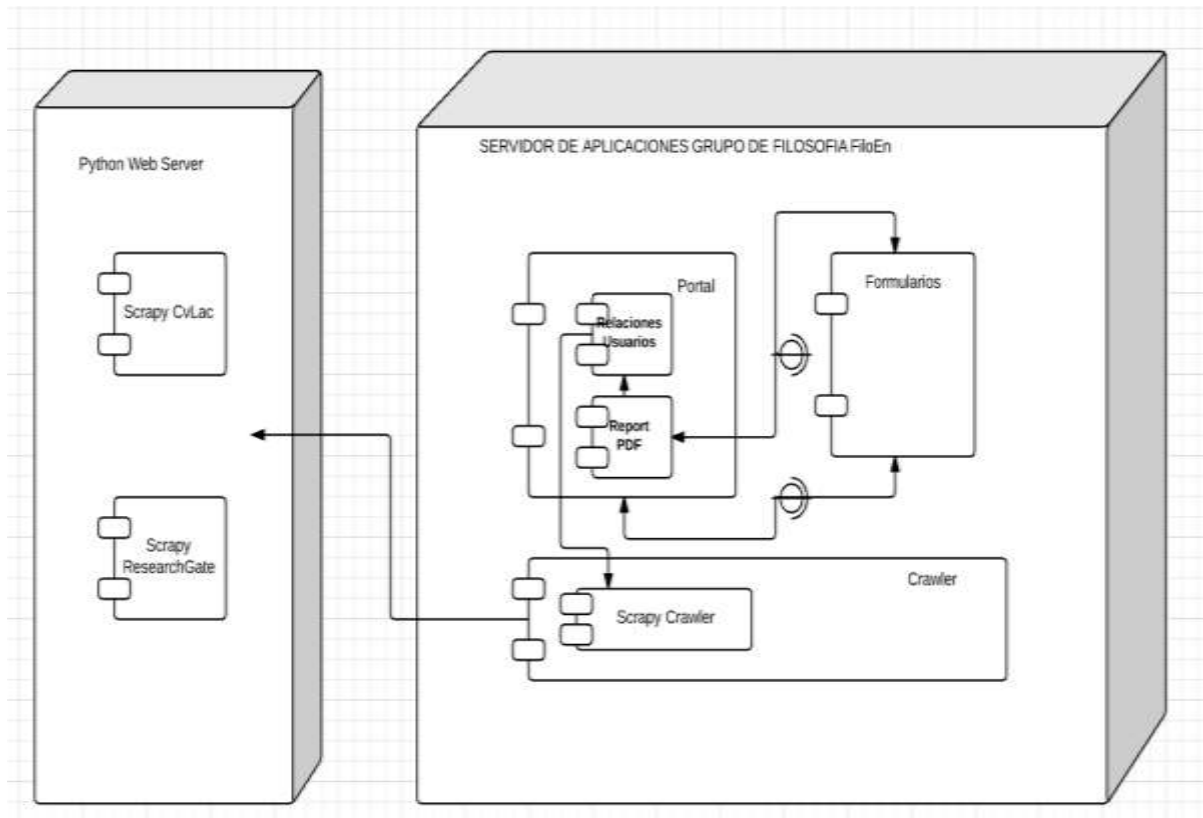
<b>Entidad: user_investigation_group</b>		
<b>Atributos</b>	<b>Nulidad</b>	<b>Descripción</b>
user_id	N	id usuario

<b>Entidad: user_investigation_group</b>	
<b>Indice:</b>	<b>Tipo de índice :</b>
<b>fk_user_has_investigation_group_user1_idx</b> user	<b>PRIMARY KEY</b>

## 8.2 Diseño de Componentes del Sistema

Para facilitar la comprensión se realizó el desarrollo basado en componentes que interactúan entre sí, a continuación, se muestra el sistema del grupo FiloEn.



**Figura 12.** Diagrama de componentes

Podemos Observar los siguientes componentes:

- Nodo Python Web Server: Contiene los componentes Scrapy CvLAC y Scrapy ResearchGate cuyas funciones es la extracción de los datos.
- Nodo Servidor de aplicaciones grupo de filosofía FiloEn: Contiene los Módulos Crawler, Portal y Formularios.
  - Módulo Crawler: Contiene el módulo Scrapy Crawler el cual se comunica con el Python Web Server.

- Módulo Portal: Posee dos componentes; Relaciones usuarios donde se procesan las relaciones de los usuarios; Report PDF encargado de generar la información que se aceptaron en los formularios.
- Módulo Formularios: Representa la información procesada para la generación de PDF, generada en el portal.

### **8.3 Diagrama de Secuencias de la Extracción de Datos.**

Inicialmente Web Page Interacciona con el Bean de usuario (userBean) en donde se proporciona el enlace URL el cual ejecuta el proceso buscar URL, el userBean entonces interacciona con Scrapy Service el cual extrae la información de la plataforma CvLAC o ResearchGate, dicha información es validada y asentada en la BD mediante un modelo Peewee producto de una introspección del modelo relacional.

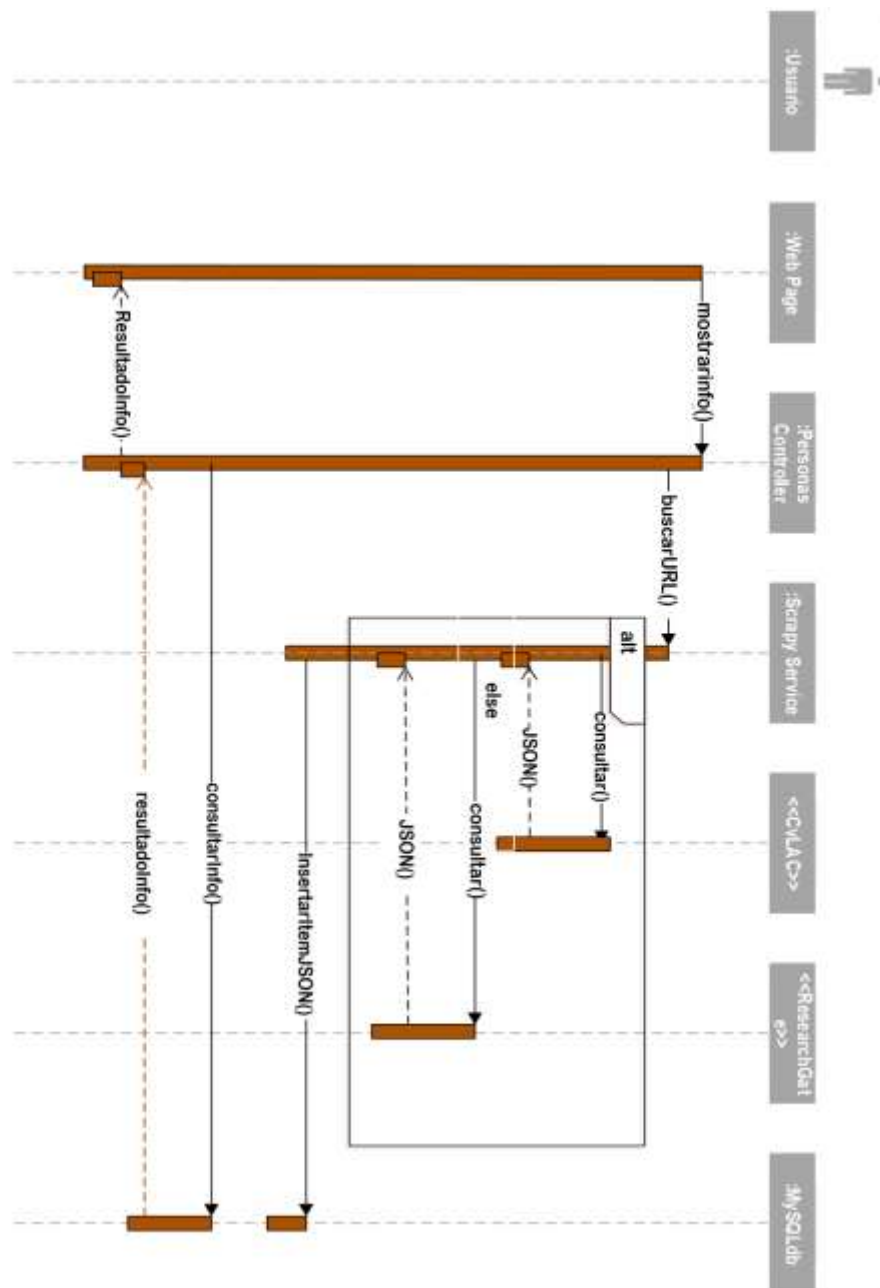
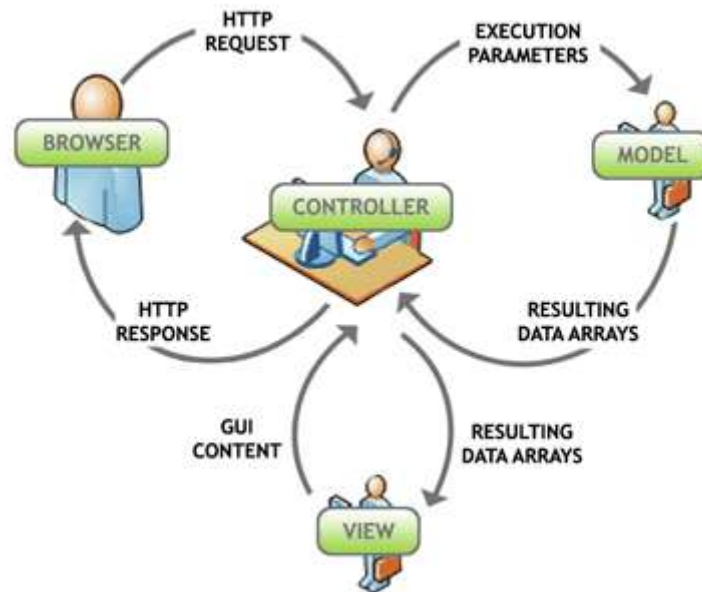


Figura 13. Diagrama de secuencia

#### 8.4 Técnicas de Programación y Patrones Utilizados.

Modelo Vista Controlador (MVC) es un estilo de arquitectura de software que separa los datos de una aplicación, la interfaz de usuario, y la lógica de control en tres componentes

distintos. Se trata de un modelo muy maduro y que ha demostrado su validez a lo largo de los años en todo tipo de aplicaciones, y sobre multitud de lenguajes y plataformas de desarrollo.



**Figura 14.** Representación Modelo Vista Controlador. Fuente: <https://www.troopsf.com/mvc/>

**El modelo** es el responsable de:

Acceder a la capa de almacenamiento de datos. Lo ideal es que el modelo sea independiente del sistema de almacenamiento. Define las reglas de negocio (la funcionalidad del sistema).

**El controlador** es responsable de:

Recibe los eventos de entrada (un clic, un cambio en un campo de texto, etc.). Contiene reglas de gestión de eventos, del tipo "SI Evento Z, entonces Acción W". Estas acciones pueden suponer peticiones al modelo o a las vistas.

**Las vistas** son responsables de:

Recibir datos del modelo y las muestras al usuario. Tienen un registro de su controlador asociado (normalmente porque además lo instancia). Según la Universidad de Alicante (2018).

## 9. Codificación del Componente

En este numeral encontrara información detalla respecto al análisis, creación y desarrollo de cada componente.

### 9.1 Estudio de los Datos que Serán Extraídos

Para la estación de datos es importante conocer su procedencia de tal manera que los datos extraídos sean relevantes, veraces y de preferencias verificados por terceros, por tal razón elegimos dos páginas relevantes en el ámbito académico y profesional como lo son CvLac (Currículum Vitae de Latinoamérica y el Caribe) y Researchgate.

**9.1.1 Análisis de CvLac.** CvLac es una plataforma administrada por COLCIENCIAS, dedicada a la gestión de Cv (Currículum Vitae), este servicio permanente de hoja de vida en el cual los investigadores, innovadores, académicos, expertos y grupos que así lo deseen pueden actualizar su información para los usos del Sistema Nacional de Ciencia y Tecnología(SNCTeI). Es el instrumento o formato electrónico de hoja de vida básico utilizado por COLCIENCIAS para:

- El registro permanente de hojas de vida de investigadores, innovadores, académicos y expertos del SNCTeI. El registro de hojas de vida de candidatos a ingresar al servicio de información de pares evaluadores reconocidos del SNCTeI. Las convocatorias para reconocimiento y medición de Grupos Colombianos de Investigación Científica o Tecnológica.
- Convocatorias de becas para formación de recurso humano, apoyo a programas doctorales a nivel nacional y jóvenes investigadores.

La información consignada en CvLAC es individual y debe ser diligenciada personalmente por cada investigador quien realiza la actualización a través de una clave única de acceso al sistema (Colciencias,2016). Después de una revisión de los diferentes tipos de información relevantes para la realización de este trabajo encontramos que en la página de CvLac podemos de extraer de cada usuario los siguientes datos:

- Formación Académica
- Experiencia Profesional
- Idiomas
- Líneas de Investigación
- Reconocimientos
- Trabajos Dirigidos/Tutorías
- Artículos
- Libros
- Capítulos de libros
- Traducciones

**9.1.2 Análisis de ResearchGate.** ResearchGate es una red social y una herramienta de colaboración dirigida a personas que hacen ciencias de cualquier disciplina, el punto de partida para cada persona investigadora es la creación de un perfil personal, a través del cual, y mediante información proporcionada por ella misma la plataforma pondrá su disposición grupos de interés, personas y literatura afines al perfil, tendencias de investigación que puedan despertar interés respetando en todo momento los derechos de autor, algunas funciones que pueden realizar son :

- Permite hacer búsqueda de artículos de revistas científicas foros y grupos de investigación

- Posee un motor de búsqueda semántica que navega por recursos internos y externos de investigación de las bases de datos más importantes

La revisión de esta herramienta web nos permitió darnos cuenta que la información relevante que se puede extraer para la realización de este proyecto es en resumen la producción bibliográfica de cada persona.

## 9.2 Extracción de Datos.

Teniendo claro la información a extraer de cada página, pasamos a un análisis para determinar la integridad de cada enlace web que sea proporcionado de estas dos páginas (CvLac y ResearchGate) , para llevar acabo la extracción de datos mediante scrapy es necesario un enlace valido para que se ejecute correctamente encada uno de los dos datos, por tal razón solo serán procesados enlaces con la siguiente estructura:

- Para CvLac el enlace debe tener la siguiente estructura:  
[http://190.216.132.131:8081/cvlac/visualizador/generarCurriculoCv.do?cod\\_rh=CODIGO\\_INVESTIGADOR](http://190.216.132.131:8081/cvlac/visualizador/generarCurriculoCv.do?cod_rh=CODIGO_INVESTIGADOR)
- En el caso de ResearchGate el enlace debe cumplir con el siguiente formato:  
[http://www.researchgate.net/profile/NOMBRE\\_DEL\\_INVESTIGADOR](http://www.researchgate.net/profile/NOMBRE_DEL_INVESTIGADOR)

**9.2.1 Desarrollo de Módulo de Extracción.** Para comenzar es necesario la creación una clase “ítem” que define la información que se va a guardar y en este caso debe ser acorde a la información que consideramos relevante da cada sitio web.

```

class CvCvIacItem(scrapy.Item):
#-----Tipo
tipo = scrapy.Field()
#-----info Basica
url = scrapy.Field()#<-----
nombre_completo = scrapy.Field()
nacionalidad = scrapy.Field()
sexo = scrapy.Field()
userid = scrapy.Field()
#-----Estudios
titulo_estudio = scrapy.Field()
estudio = scrapy.Field()
eduniversidad = scrapy.Field()
carrera = scrapy.Field()
edufecha = scrapy.Field()
#-----Experiencia Laboral
cargo = scrapy.Field()
empresa = scrapy.Field()
#-----Idioma
idioma = scrapy.Field()
habla = scrapy.Field()
escribe = scrapy.Field()
lee = scrapy.Field()
escucha = scrapy.Field()
#-----Reconocimientos (Distinciones)
reconocimiento = scrapy.Field()
#-----Proyectos Dirigidos
tipo_py = scrapy.Field()
bloque = scrapy.Field()

```

**Figura 15.** Item Scrapy

Una vez tengamos configurado el ítem en este caso class CvCvIacItem(scrapy.item) pasamos a la configuración de la araña en el cual decidimos utilizar expresiones de xpath basándonos en la estructura de etiquetas html de la información a extraer.

```

<td colspan="2"><h3>Formación Académica</h3></td>
</tr>
<tr>
<td><li>&nbsp;</li></td>
<td><strong>Doctorado</strong>
STATE UNIVERSITY OF NEW YORK AT STONY BROOK<br />
Filosofía<br />
Agosto 2008 - Diciembre 2012<br />
Postmetaphysical Reason and Postsecular Consciousness: Habermas; Analysis
in the Public Sphere</td>
</tr>
<tr>
<td><li>&nbsp;</li></td>
<td><strong>Maestría/Magister</strong>
STATE UNIVERSITY OF NEW YORK AT STONY BROOK<br />
Maestría en Filosofía<br />
Agosto 2008 - Mayo 2010<br />
Postmetaphysical Reason and Postsecular Consciousness: Habermas; Analysis of Religion in the Public Sphere</td>
</tr>
<tr>
<td><li>&nbsp;</li></td>
<td><strong>Especialización</strong>
UNIVERSIDAD INDUSTRIAL DE SANTANDER UIS<br />
Especialización en Docencia Universitaria<br />
Febrero 2007 - Marzo 2008<br />
El Desarrollo de la competencia lectora como estrategia de enseñanza y aprendizaje de la asignatura de Lógica: Una apuesta
</td>
</tr>
<tr>
<td><li>&nbsp;</li></td>
<td><strong>Pregrado/Universitario</strong>
UNIVERSIDAD INDUSTRIAL DE SANTANDER UIS<br />
Filosofía<br />
Febrero 2000 - Septiembre 2005<br />
La relación lenguaje derecho en la filosofía de Habermas</td>
</tr>

```

**Figura 16.** Código Fuente CvLac

Para cada segmento de la información analizada se definieron funciones con expresiones de xpaths para cada tipo de información (académicas, bibliográficas, de idiomas, etc).

```

88 def parse_estudios(self, response, idx, items):
89     idxEstudio = idx.index(u'Formaci\xf3n Acad\xe9mica') + 2
90     estudios = response.xpath("//table[@width='100%'][" + str(idxEstudio) + "]/tr")
91     estudios.extract()
92     for index, estudio in enumerate(estudios):
93         if (index > 0):
94             item = CvCvIacItem()
95             item["tipo"] = "estudio"
96             item["titulo_estudio"] = estudio.xpath("td[2]//strong/text()").extract()
97             item["estudio"] = estudio.xpath("td[2]/text()").extract()
98             item["carrera"] = item["estudio"][1].replace("\r\n", "").replace(" ", "")
99             item["eduniversidad"] = item["estudio"][0].replace("\r\n", "").replace(" ", "")
100            item["edufecha"] = item["estudio"][2].replace("\r\n", "").replace(" ", "")
101            item["estudio"] = item["estudio"][3].replace("\r\n", "").replace(" ", "")
102            item["titulo_estudio"] = item["titulo_estudio"][0].replace("\r\n", "").replace(" ", "")
103            item["userid"] = self.listAwaitCvIac.paritylist[response.url]
104            items.append(item)
105     return items

```

**Figura 17.** Spider Scrapy

Algunas expresiones de xpath no fueron suficientes para separar de manera adecuada la información, por este motivo fue necesario crear un paquete para el parseo de expresiones regulares, el cual se llamó Utilidades.py.

```

1  # -*- coding: utf-8 -*-\
2
3  #paquete de utilidades, en su mayoría manejo de str
4  #Facilita la extracción de datos para el registro en la BD
5  from Crawler.Data_Access.Dao import *
6  from Crawler.Data_Access.Model import User
7
8  #Clase que genera las llaves de usuario junto a la lista de pendientes
9  class listAwaitCvIac():...
23
24  #Repertorio de funciones para el Manejo de USER
25  class parseNombre():...
71
72  # Repertorio de funciones para el Manejo de estudio
73  class parseStudy():...
82
83  #Clase para el manejo de FECHAS
84  class parseDate():...
22
23  #Clase para el parseo de niveles del lenguaje
24  class parseLanguage():...
30
31  #Clase para el manejo de distinciones
32  class parseDistinction():...
42
43  #Clase para el parseo de proyectos
44  class parseProject():...
~^

```

Figura 18. Paquete de utilidades

Hasta el momento tenemos como salida un JSON con la estructura ítems que contiene toda la información raspada.

```

1  {"url": "http://scienti.colciencias.gov.co:8081/cvIac/visualizador/generarCurrículoCv.do?cod_rh=0000380059", "nombre_com
2  "estudio": "Formaci\u00f3n y argumentaci\u00f3n. Automatizaci\u00f3n de procesos argumentativos", "titulo_estudio": "Do
3  "estudio": "Dise\u00f1o de un entorno virtual para favorecer la independencia cognoscitiva de los estudiantes de educac
4  "estudio": "Planeaci\u00f3n estrat\u00e9gica de sistemas para laboratorio de gen\u00e9tica UIS", "titulo_estudio": "Pre
5  "cargo": "Actividades de administraci\u00f3n", "empresa": "Universidad Industrial de Santander - UIS", "tipo": "experie
6  "cargo": "Actividades de administraci\u00f3n", "empresa": "Departamento Administrativo de Ciencia, Tecnolog\u00eda e In
7  "cargo": "Actividades de investigaci\u00f3n", "empresa": "Ciup", "tipo": "experiencia"},
8  "cargo": "Actividades de administraci\u00f3n", "empresa": "Universidad Pedag\u00f3gica Nacional - U.P.N.", "tipo": "exp
9  "cargo": "Actividades de investigaci\u00f3n", "empresa": "Facultad de Humanidades", "tipo": "experiencia"},
10 "cargo": "Actividades de administraci\u00f3n", "empresa": "Universidad Aut\u00f3noma De Bucaramanga - Unab", "tipo": "e
11 "cargo": "Actividades de administraci\u00f3n", "empresa": "Insed", "tipo": "experiencia"},
12 "cargo": "Actividades de administraci\u00f3n", "empresa": "Universidad Industrial de Santander - UIS", "tipo": "experie
13 "cargo": "Actividades de administraci\u00f3n", "empresa": "Universidad Pontificia Bolivariana Seccional Bucaramanga", "
14 "cargo": "Actividades de administraci\u00f3n", "empresa": "Universidad Santo Tom\u00e1s Seccional Bucaramanga", "tipo":
15 "habla": "Bueno", "Lee": "Bueno", "tipo": "idioma", "escribe": "Bueno", "escucha": "Bueno", "idioma": "\u00a0Ingl\u00e9s
16 "tipo": "reconocimiento", "reconocimiento": "Tesis de doctorado laureada,Universidad Pedag\u00f3gica Nacional - U.P.N.
17 "tipo": "reconocimiento", "reconocimiento": "Visiting Scholar,University Of Texas - Agostode 2008"},
18 "tipo": "reconocimiento", "reconocimiento": "Miembro Colaborador,C\u00e1tedra Latinoamericana de Fenomenolog\u00eda - d
19 "tipo": "reconocimiento", "reconocimiento": "Beneficiario de cr\u00e9dito condonable,Departamento Administrativo de Cie

```

Figura 19. JSON de salida

Teniendo toda la información es necesario asentarla en la base de datos, para tal fin, nos apoyamos en el ORM Peewee el cual nos generó un modelo de la base de datos.

```
566 class Study(BaseModel):
567     average = FloatField(null=True)
568     end_date = DateField(null=True)
569     institution = CharField()
570     level = CharField(null=True)#char 1
571     program = CharField(null=True) #estudio
572     project_title = CharField(null=True)
573     status_field = CharField()
574     study = PrimaryKeyField(db_column='study_id')
575     title = CharField()#carrera
576     user = ForeignKeyField(db_column='user_id', rel_model=User, to_field='id')
577
578 class Meta:
579     db_table = 'study'
580
581 class Theory(BaseModel):
582     figure_type = ForeignKeyField(db_column='figure_type_id', rel_model=FigureType, to_field='id')
583     link = TextField()
584     text = TextField()
585
586 class Meta:
587     db_table = 'theory'
588
```

**Figura 20.** Mapeo modelo Peewee

Una vez teniendo el modelo desarrollamos un paquete DAO (Data Access Object) para el manejo del modelo relacional con el fin de facilitar la inserción y validación de datos en el pipeline de Scrapy.

```

3     from Model import *
4
5     class userDao():
6         def insert(self, usuario):
7             usuario.save()
8
9         def findByUrl(self, url):
10            try:
11                return User.select().where((User.cvlac_url==url) or (User.rgate_url==url)).get()
12            except:
13                return False
14
15        def update(self, user, names, sName, lastName, sLastName, nNationality, gen):
16            user.name=names
17            user.second_name=sName
18            user.last_name=lastName
19            user.second_last_name=sLastName
20            user.nationality=nNationality
21            user.gender=gen
22            user.save()
23
24        def getAwaitCvlac(self):
25            try:
26                query = User.select(User.id, User.cvlac_url).where(User.cvlac_url % "http%")
27                return query
28            except:
29                return False

```

Figura 21. Paquete DAO

Modificamos el archivo de Pipeline.py para que use el paquete DAO y pueda asentar los ítems en la base de datos de MySQL.

```

4     from Crawler.Data_Access import Utilidades as util
5     from Crawler.Data_Access.Dao import *
6     from scrapy.exceptions import DropItem
7
8     class CvCvlacPipeline(object):
9
10        authorCode      = 30 # Autor de la obra principal (obj_role)
11        articleCode     = 2  # Código de artículo (obj_type)
12        magazineCode    = 16 # Código de revista (obj_type)
13        bookCode        = 1  # Código de Libro (obj_type)
14        chapterBookCode = 3  # Código de Capitulo de libro (obj_type)
15        statusCode      = "p" # Publico (Status_field obj_bibliographic)
16
17        def open_spider(self, spider):...
18
19        def spider_closed(self, spider):...
20
21        def process_item(self, item, spider):
22
23            # if (item["tipo"] == "general"): Suspendido por reformas en Cvlac
24            if (item["tipo"] == "estudio"): self.proc_Study(item)
25            if (item["tipo"] == "experiencia"): self.proc_Experience(item)
26            if (item["tipo"] == "idioma"): self.proc_language(item)
27            if (item["tipo"] == "reconocimiento"): self.proc_distinction(item)
28            if (item["tipo"] == "proyecto"): self.proc_project(item)
29            if (item["tipo"] == "artículo"): self.proc_article(item)
30            if (item["tipo"] == "libros"): self.proc_book(item)
31            if (item["tipo"] == "capitulos"): self.proc_chpBook(item)
32
33            return item

```

Figura 22. Pipeline.py

Algunas veces es necesario realizar una validación antes de insertar los datos, como es el caso de productos bibliográficos o colaboraciones de ítems bibliográficos ya existentes.

```

def proc_chpBook(self, item):
208     chpBookItem = ObjBibliographic()
209     chpBookUtil = objBibliographicDao()
210     chpBook = chpBookUtil.findByIsbn(item["isbn"].replace("|", ""))
211     chpBookNew = util.parseChpBook(item["bibliografia"])
212     chpBookNew.parse_complemento(item["complemento"])
213     if(chpBook):
214         # El libro ya existe
215         chapter = chpBookUtil.findChapter(chpBook.get_id(), chpBookNew.pages, self.chapterBookCode)
216         if(chapter):
217             ## El artículo ya existe
218             #print("Autor: ->", chpBookNew.author)
219             self.set_author(chpBookNew.author, chapter.get_id(), self.authorCodeCode, item["userid"])
220         else:
221             ## El artículo no existe aun
222             chpBookItem.father_obj_bibliographic = chpBook.get_id()
223             chpBookItem.title = chpBookNew.title
224             chpBookItem.pages_num = chpBookNew.pages
225             chpBookItem.type = self.chapterBookCode
226             chpBookItem.status_field = self.statusCode
227             try:
228                 chpBookItem.save()
229                 self.set_author(chpBookNew.author, chpBookItem.get_id(), self.authorCode, item["userid"])
230             except:
231                 print ("ERROR INSERT CHAPTER")
232     else:
233         # El libro no existe aun

```

**Figura 23.** Validación de ítem bibliográficos

### 9.3 Desarrollo de la Interfaz de Presentación de Resultados.

Este proyecto se integra al Proyecto Diseño e implementación de un portal web para soportar la enseñanza de la filosofía desarrollado por Galeano & Salamanca (2016), por tal razón por tal razón el componente desarrollado se comunica e integra con los formularios de ingreso y de registro como se puede ver en la **Figura 24**.



Registro

Tipo Usuario: Estudiante

Usuario: Usuario

Contraseña: Contraseña

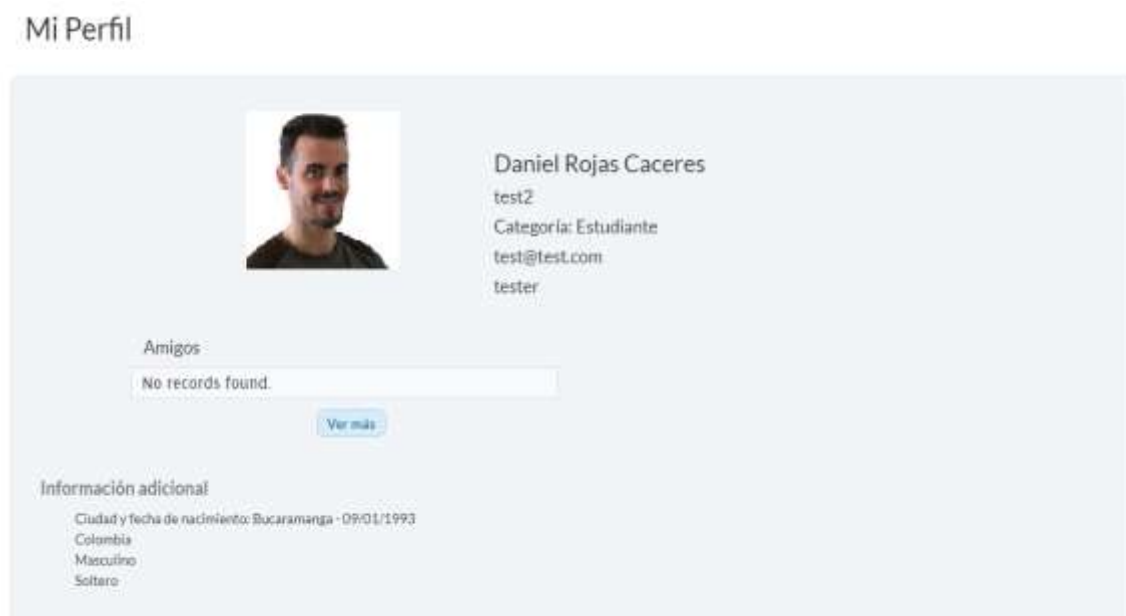
Repetir Contraseña: Contraseña

Correo: Correo electrónico


Crear

**Figura 24.** Formulario Registro

Si el usuario ya posee una cuenta registrada en la plataforma FiloEn, debe dirigirse al menú desplegable, al menú mi perfil, ahí encontrara una vista resumen, de sus datos personales, como se puede ver en las **Figura 25**, **Figura 26** y **Figura 27**.



Mi Perfil



Daniel Rojas Caceres  
test2  
Categoría: Estudiante  
test@test.com  
tester

Amigos:  
No records found.  
[Ver más](#)

Información adicional  
Ciudad y fecha de nacimiento: Bucaramanga - 09/01/1993  
Colombia  
Masculino  
Soltero

**Figura 25.** Mi Perfil

En la Opción de Información personal muestra inicialmente la cantidad de idiomas que maneja el usuario y su respectivo nivel de dominio

Habilidades	
N/N	
Idiomas (3)	
<b>Español</b>	Nivel de escucha: Alto Nivel de escritura: Medio Nivel de habla: Alto Nivel de lectura: Alto
<b>Árabe</b>	Nivel de escucha: Alto Nivel de escritura: Bajo Nivel de habla: Alto Nivel de lectura: Medio
<b>Bengali</b>	Nivel de escucha: Bajo Nivel de escritura: Bajo Nivel de habla: Bajo Nivel de lectura: Bajo

**Figura 26.** Información personal de idiomas

También presenta los logros obtenidos por el usuario y condecoraciones

Condecoraciones (4)	
<b>Cum Laude</b>	Institución: Universidad de los Andes Tipo de distinción: Reconocimiento Fecha: 2005-07-01
<b>Segundo lugar en la promoción de Economistas 2005</b>	Institución: Universidad de los Andes Tipo de distinción: Reconocimiento Fecha: 2005-07-01
<b>Segundo lugar en el Concurso Nacional de Ensayos "Alberto Adriani"</b>	Institución: Consejo Nacional de Estudiantes de Economía Tipo de distinción: Reconocimiento Fecha: 2004-01-01
<b>Excelencia Docente</b>	Institución: Univ. Católica Andrés Bello Tipo de distinción: Reconocimiento Fecha: 2014-06-01

**Figura 27.** Información Personal de Condecoraciones

Al finalizar la información, encontrará el botón para gestionar la información lo cual lo conducirá a una pantalla como lo muestra la **Figura 28**.



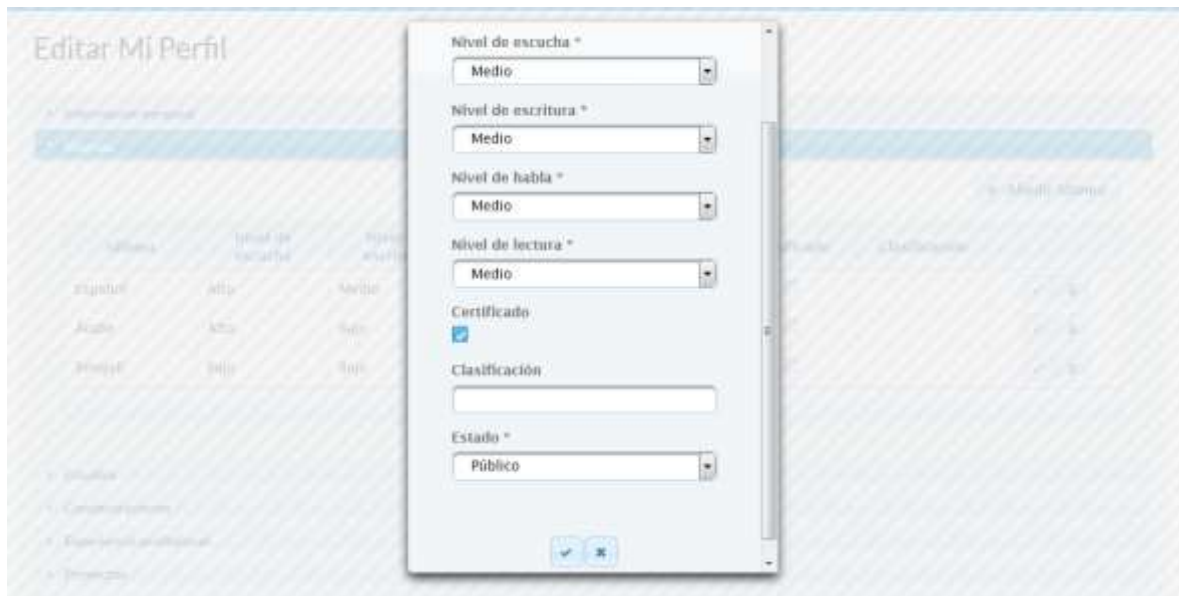
**Figura 28.** Gestión de Perfil

En la opción de Idiomas, se despliega un menú el cual permite gestionar los idiomas que el usuario maneja, asimismo agregar, eliminar y editar el nivel de estudio de cada uno.



**Figura 29.** Menú de Idiomas

En la opción de Añadir idioma, se presentan los distintos dominios del idioma, agregar también proporciona al usuario la alternativa de decidir si desea compartir la información con los demás usuarios



**Figura 30.** Añadir Idioma

En la Opción de eliminar idioma el usuario autenticado puede eliminar el idioma si así lo desea.



**Figura 31.** Eliminar Idioma

En el menú de estudios permite al usuario gestionar los estudios realizados (añadir, editar, eliminar estudio), de igual manera le presenta los estudios realizados detalladamente.

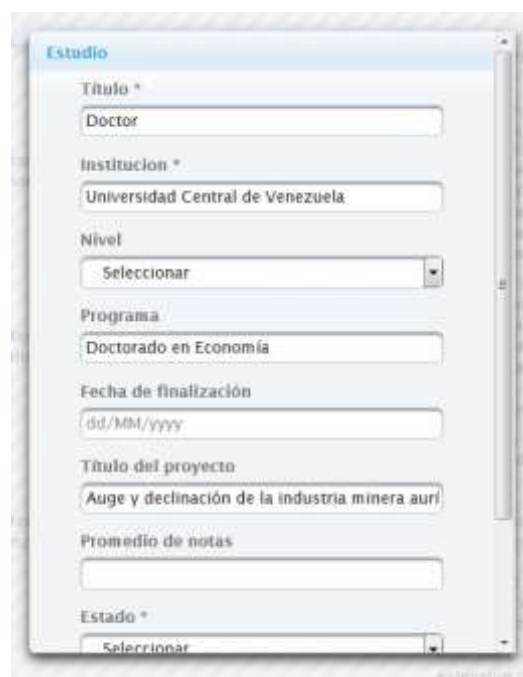


The screenshot shows a web interface with a blue header bar containing the text 'Estudios'. In the top right corner of the interface, there is a button labeled 'Añadir estudio'. Below the header is a table with the following columns: Título, Institución, Programa, Nivel, Fecha de grado, Título de proyecto, Promedio de notas, and an empty column for actions. The table contains three rows of data:

Título	Institución	Programa	Nivel	Fecha de grado	Título de proyecto	Promedio de notas	
Doctor	Universidad Central de Venezuela	Doctorado en Economía			Auge y declinación de la industria minera aurífera en Venezuela (1866-1920): Una mirada a los vínculos internacionales		 
Magister	Instituto de Estudios Superiores de Administración	Master en Administración		01/02/2010	Incidencia de un desempeño pobre de la acción sobre la probabilidad de fusiones y adquisiciones.		 
Magister	Instituto de Estudios Superiores de Administración	Master en Finanzas		01/09/2007	Evaluación de proyectos de inversión pública mediante el uso de opciones reales bayesianas: Una aplicación al sector eléctrico		 

Figura 32. Menú estudio

En la opción de añadir estudio, permite al usuario añadir un estudio a su hoja de vida en donde tendrá que insertar los datos de la institución y el título obtenido.



The screenshot shows a form titled 'Estudio' with the following fields and controls:

- Título \*: Doctor
- Institución \*: Universidad Central de Venezuela
- Nivel: Selecccionar (dropdown menu)
- Programa: Doctorado en Economía
- Fecha de finalización: dd/MM/yyyy
- Título del proyecto: Auge y declinación de la industria minera auri
- Promedio de notas: (empty text field)
- Estado \*: Selecccionar (dropdown menu)

Figura 33. Añadir Estudio

En la sección de condecoraciones el usuario puede añadir nuevas condecoraciones, editar una condecoración existente o eliminar una condecoración haciéndole clic a la respectiva opción.

Distinción	Institución	Fecha	Motivo	
Cum Laude	Universidad de los Andes	01/07/2005	0	
Segundo lugar en la promoción de Economistas 2005	Universidad de los Andes	01/07/2005	0	
Segundo lugar en el Concurso Nacional de Ensayos "Alberto Adriani"	Consejo Nacional de Estudiantes de Economía	01/01/2004	0	
Excelencia Docente	Univ. Católica Andrés Bello	01/06/2014	0	

(1 of 1) 10

**Figura 34.** Menú Condecoraciones

En la opción de proyectos el usuario gestiona los proyectos en los que ha participado el cual debe asignar la fecha en que se realizó o si aún está en desarrollo deberá informar la fecha en que lo inició; permitiendo también eliminar un proyecto, haciendo clic en la opción preferida.

Proyecto	Institución	Etapa	Fecha	Problema	Objetivo general	Objetivos específicos	
Indicador de vulnerabilidad social: Perspectivas para el caso colombiano	Universidad Industrial de Santander - UIS	Culminado	01/01/2015	0	0	0	
Estimación de la volatilidad del precio en la onza troy de oro para el periodo 2001-2014	Universidad Industrial de Santander - UIS	Cancelado	01/01/2015	0	0	0	
El sistema de salud en Colombia: condicionantes frente al traslado de afiliados entre entidades promotoras de salud (E.P.S.)	Universidad Santo Tomás Seccional Bucaramanga	Cancelado	01/01/2017	0	0	0	
Impacto financiero de la							

**Figura 35.** Gestión Proyectos

En la opción de experiencia profesional, el usuario puede agregar las empresas con su respectivo cargo, salario, los rangos de tiempo en el que laboraba en la empresa y los motivos por los cuales ya no se encuentra en la empresa, así mismo editar, y eliminar esta información.




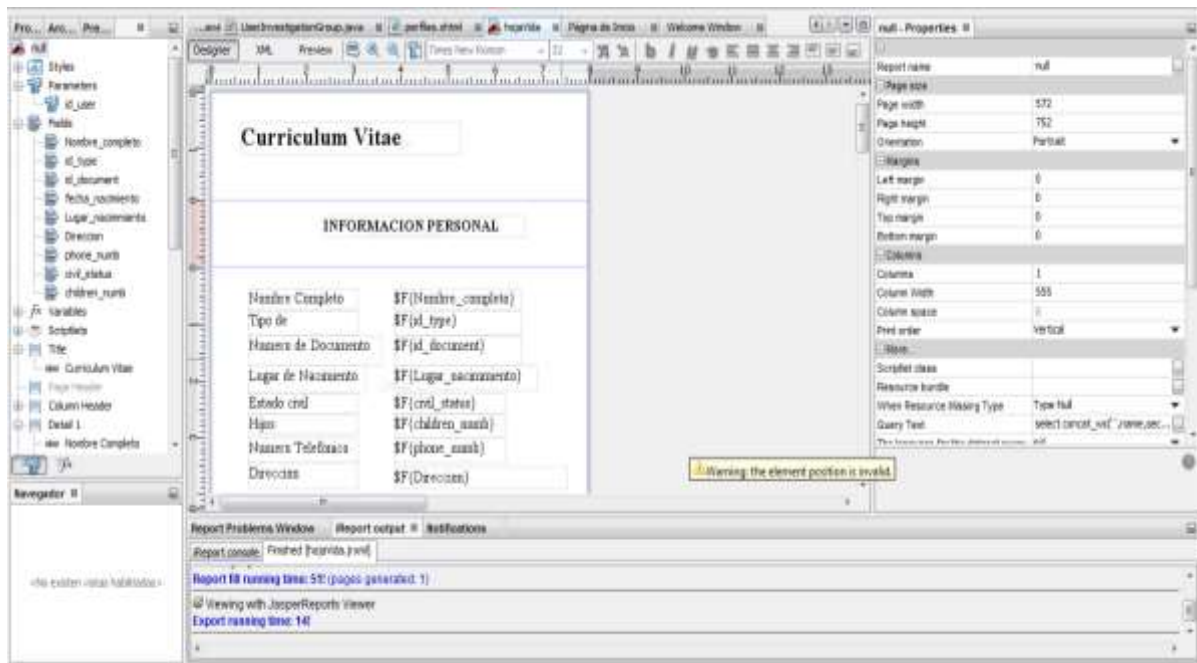
Empresa	Cargo	Fecha de ingreso	Fecha de retiro	Salario	Motivo de retiro
Universidad Pontificia Bolivariana Seccional Bucaramanga	Actividades de docencia				  
Universidad Santo Tomás Seccional Bucaramanga	Actividades de docencia				  
Universidad Industrial de Santander - UIS	Actividades de docencia				  
Universidad Central De Venezuela	Actividades de docencia				  
Banco Central de Venezuela	Actividades de administración				  

Figura 36. Experiencia profesional

#### 9.4 Componente de generación de PDF

Para la generación de la información en formato PDF se utilizó la biblioteca más utilizada para la generación de informes en java, iReport, el cual cuenta con un editor, gráfico de informes facilitando el diseño y la implementación de reportes dinámicos.

Para la configuración del reporte, se necesita configurar previamente los campos que deseamos mostrar de las entidades trayendo dichos campos bajo un parámetro de criterios, en este caso el parámetro elegido fue id\_user.



**Figura 37.** Reporte fase de diseño

Una vez configurados los campos resta aplicar un tema o plantilla para las hojas, vale la pena resaltar el reporte admite la integración de sub-reportes. En tal caso se igualan los parámetros de su reporte con el de sub-reportes.

## 10. Pruebas

Con el fin de evaluar la eficacia de la herramienta web se realizaron pruebas empíricas y técnicas que evidencian el correcto funcionamiento de cada uno de los componentes que integra dicha herramienta.

Las pruebas se realizaron de manera local un equipo de características:

**Tabla 51.**  
*Especificaciones Equipo Pruebas*

<b>Sistema operativo</b>	<b>Windows 7</b>
<b>Versión</b>	Ultimate – 64 bits
<b>Memoria Instalada (RAM)</b>	8,00 GB
<b>Procesador</b>	Intel core i5 M460 @ 2.53 GHz

### Pruebas componente de extracción de datos (Scrapy).

**Tabla 52.**

*Pruebas de Extracción de datos Scrapy*

Caso de Prueba	Resultado esperado	Resultado obtenido
Extracción de la información con enlaces correctos	Ok	Ok
Extracción de la información con enlaces incorrectos	Error	Error
Validación de información inexistente contra la base de datos	Ok	Ok
Validación de información existente contra la base de datos	Ok	Ok
Inserción de información en la base de datos	Ok	Ok

Para el módulo Scrapy se utilizó el método de caja blanca en el cual se busca determinar la funcionalidad interna de los módulos de spiders, pipelines y data Access (los DAO y paquetes de utilidades).

Luego en la integración se utilizó el método de caja negra, el cual presentó resultados satisfactorios.

### Pruebas Componentes Interfaz de Usuario.

**Tabla 53.**

*Prueba Componente Interfaz de Usuario*

Caso de Prueba	Resultado obtenido
Registro de usuario	Exitoso
Login de usuario	Exitoso
Edición de perfil	Exitoso
Generación de pdf	Exitoso
Obtención de datos	Exitoso

Para las pruebas de interfaz de usuario se utilizó el método de caja negra en la cual se insertaron valores validos e inválidos con el fin de establecer la correcta funcionalidad de los módulos evaluados en donde arrojaron resultados exitosos.

## **11. Conclusiones**

Se analizaron los requerimientos necesarios para el desarrollo de la herramienta software cumpliendo las necesidades planteadas para la administración de hojas de vida de investigadores.

Se logró establecer una estructura de datos adecuada en el desarrollo de la herramienta, de manera tal que brinda flexibilidad y escalabilidad, dejando la puerta abierta para la ampliación de su alcance.

Se diseñó una herramienta software que permite la administración adecuada de información de hojas de vida de investigadores, a través de la gestión de perfiles de usuario.

Con el desarrollo de esta herramienta web, se evidencia una plataforma que mejora la gestión de hojas de vida, ahorrando tiempo al usuario y centralizando la información, y facilitando llevar a cabo procesos de valoración de la producción bibliográfica de la comunidad UIS.

Se validó el uso y funcionalidad de cada uno de los componentes correspondientes a la herramienta software que permite la gestión de información a través de perfiles de usuario

## **12. Recomendaciones**

Expandir el alcance de algunas funciones ya implementadas como la extracción de datos, mediante la implementación de algoritmos para el raspado de información de más paginas relevantes.

Se recomienda la implementación de mecanismos que permitan la valoración la cantidad de producción bibliográfica generada por los investigadores.

### Referencias bibliográficas

- Aguilar, Cancino & Gonzalez. 2018, Expociencias Veracruz: *Sistema de Reutilización de Aguas Grises*. Universidad Tecnológica del Centro de Veracruz. Pg.3. Recuperado de: [https://www.academia.edu/37254424/Protocolo\\_de\\_Investigacion\\_SISRAG.docx](https://www.academia.edu/37254424/Protocolo_de_Investigacion_SISRAG.docx)
- Alegsa, L. (2018). Diccionario de informática y tecnología. Santa Fe, Argentina. Recuperado de: <http://www.alegsa.com.ar/Dic/python.php>
- Arbeláez Gómez, M. C. (2014). Las tecnologías de la información y la comunicación (TIC) un instrumento para la investigación. *Investigaciones Andina*, 16(29). Recuperado de: [http://www.scielo.org.co/scielo.php?script=sci\\_arttext&pid=S0124-81462014000200001](http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0124-81462014000200001)
- Arias, Á. (2014). Bases de Datos con MySQL: 2ª Edición. IT Campus Academy. Recuperado de: [https://books.google.es/books?hl=es&lr=lang\\_es&id=EojJCgAAQBAJ&oi=fnd&pg=PA13&dq=que+es+mysql+&ots=FEe3-LnR-i&sig=nNxfeZxBZOdFTpgCM4kM7bXvgBQ#v=onepage&q=que%20es%20mysql&f=false](https://books.google.es/books?hl=es&lr=lang_es&id=EojJCgAAQBAJ&oi=fnd&pg=PA13&dq=que+es+mysql+&ots=FEe3-LnR-i&sig=nNxfeZxBZOdFTpgCM4kM7bXvgBQ#v=onepage&q=que%20es%20mysql&f=false)
- Bell, D & Parr, M. (2003) *Java Para Estudiantes*. 3a. ed. México: Pearson Educación. Print. Recuperado de: [https://books.google.es/books?hl=es&lr=lang\\_es&id=TRUdyfwdaSoC&oi=fnd&pg=PR19&dq=que+es+java&ots=DqHFp7cbvD&sig=epfBC7e-qn66CxPjSm3Q8YLTW4U#v=onepage&q=que%20es%20java&f=false](https://books.google.es/books?hl=es&lr=lang_es&id=TRUdyfwdaSoC&oi=fnd&pg=PR19&dq=que+es+java&ots=DqHFp7cbvD&sig=epfBC7e-qn66CxPjSm3Q8YLTW4U#v=onepage&q=que%20es%20java&f=false)
- Bernal Bermúdez, J., & Fernández Muñoz, L. (2012). Programación Orientada a objetos con Java (POOJ). Universidad Politécnica de Madrid. Recuperado de: [https://www.etsisi.upm.es/sites/default/files/curso\\_2013\\_14/MASTER/MIW.JEE.PO OJ.pdf4](https://www.etsisi.upm.es/sites/default/files/curso_2013_14/MASTER/MIW.JEE.PO OJ.pdf4)
- Blog Gidahatari (2018). *Extrae información de páginas web con Scrapy*. Recuperado de: <http://gidahatari.com/ih-es/extrae-informacion-de-paginas-web-con-scrapy>
- Cabral, Vargas, B. (2016). *La sociedad del conocimiento, la sociedad de la información*. Disponible en: <http://www.bibliotecarios.concepto28.com/programa2016/la-sociedad-del-conocimiento.pdf>
- Carreño, C. H. & Gamboa, S. C. (2014). Fundamentos para la gestión de conocimiento en grupos de investigación. *I+ D Revista de investigaciones*, 3(1), 40-50. Recuperado de: <http://www.udi.edu.co/revistainvestigaciones/index.php/ID/article/view/53/50>
- Carvajal, M., & Saab, J. (2010). Fundamentos conceptuales de las Directrices de Usabilidad de Gobierno en línea. *Ministerio de Tecnologías de la Información y*

- Comunicaciones*, Bogotá, Colombia. Recuperado de: <http://www.mariocarvajal.com/Lineamientos-de-estructura-Manual-de-Usabilidad.pdf>
- Chaparro, E., & Salazar, J., (2014). Sistema de información para generación de hojas de vida de investigadores (tesis de pregrado). *Universidad Industrial de Santander*. Colombia. Recuperado de: <http://tangara.uis.edu.co/biblioweb/>
- Çivici, Ç. (2014). PrimeFaces User Guide 5.0. *PrimeTek Informatics*. Recuperado de: [https://www.primefaces.org/docs/guide/primefaces\\_user\\_guide\\_5\\_0.pdf](https://www.primefaces.org/docs/guide/primefaces_user_guide_5_0.pdf)
- Cobo, A., Gómez, P., Pérez, D., & Rocha, R., (2005). PHP y MySQL Tecnologías para el desarrollo de aplicaciones web. Ediciones Díaz Santos. Recuperado de: [https://books.google.es/books?hl=es&lr=lang\\_es&id=zMK3GOMOpQ4C&oi=fnd&pg=PR17&dq=que+es+mysql+&ots=Fgls1-Ienm&sig=X53wp0aVG1jBv8MeEsnVx3DWoWQ#v=onepage&q=que%20es%20mysql&f=false](https://books.google.es/books?hl=es&lr=lang_es&id=zMK3GOMOpQ4C&oi=fnd&pg=PR17&dq=que+es+mysql+&ots=Fgls1-Ienm&sig=X53wp0aVG1jBv8MeEsnVx3DWoWQ#v=onepage&q=que%20es%20mysql&f=false)
- Colciencias (2016), Manual del aplicativo CvLac Currículum Vitae para Latinoamérica y el Caribe. Bogotá D.C: Departamento Administrativo de Ciencia, Tecnología e Innovación- Colciencias. Recuperado de: [https://colciencias.gov.co/sites/default/files/ckeditor\\_files/manual-de-usuario-cvlac.pdf](https://colciencias.gov.co/sites/default/files/ckeditor_files/manual-de-usuario-cvlac.pdf)
- Colciencias (s.f). *Glosario*. Recuperado de: <http://www.colciencias.gov.co/sites/default/files/upload/glosario-colciencias.pdf>
- Flórez, H. A. (2012). *Programación orientada a objetos usando java*. Ecoe Ediciones. Recuperado de: <https://ebookcentral.proquest.com/lib/ucooperativasp/reader.action?docID=3203026#>
- Galeano, J., & Salamanca, R.,(2016). Diseño e implementación de un portal web para soportar la enseñanza de la filosofía (tesis de pregrado). *Universidad Industrial de Santander*, Colombia. Recuperado de: <http://tangara.uis.edu.co/biblioweb/>
- Groussard, T. (2010). Java Enterprise Edition: desarrollo de aplicaciones web con JEE 6. Ediciones Eni. Recuperado de: [https://books.google.com.co/books?hl=es&lr=&id=MdoMHI4QiJ0C&oi=fnd&pg=PP21&dq=java%20jee&ots=wOCFgJA3CH&sig=1sboT0zGcMONBBqFEcAakotPrwQ&redir\\_esc=y&fbclid=IwAR22qCbKcpYShXHkGJGC5Aha7eM9fBLPZjTIU7W1Y CzIp3feSMnlQU3zSwo#v=onepage&q=java%20jee&f=false](https://books.google.com.co/books?hl=es&lr=&id=MdoMHI4QiJ0C&oi=fnd&pg=PP21&dq=java%20jee&ots=wOCFgJA3CH&sig=1sboT0zGcMONBBqFEcAakotPrwQ&redir_esc=y&fbclid=IwAR22qCbKcpYShXHkGJGC5Aha7eM9fBLPZjTIU7W1Y CzIp3feSMnlQU3zSwo#v=onepage&q=java%20jee&f=false)
- Java (2018). *¿Qué es la tecnología Java y para qué la necesito?* Recuperado de: [https://www.java.com/es/download/faq/whatis\\_java.xml](https://www.java.com/es/download/faq/whatis_java.xml)
- Java (2018). *Java Platform, Enterprise Edition*. Recuperado de: [https://es.wikipedia.org/wiki/Java\\_EE](https://es.wikipedia.org/wiki/Java_EE)

- Martínez, L.; Parra, P. (2016). *Tecnologías de la información y la comunicación (TICS) en el sector salud*. Recuperado de: <http://www.bdigital.unal.edu.co/11172/1/laurayanethavellamartinez.2013.pdf>
- Páez, D., Techio, E., Marques, J., & Beristain, C. M. *Psicología Social*. 2007. McGraw Hill. Recuperado de: [https://www.researchgate.net/profile/Dario\\_Paez/publication/286457442\\_Memoria\\_social\\_y\\_Colectiva\\_Representaciones\\_sociales\\_de\\_la\\_historia/links/566b023f08ae62b05f046941/Memoria-social-y-Colectiva-Representaciones-sociales-de-la-historia.pdf](https://www.researchgate.net/profile/Dario_Paez/publication/286457442_Memoria_social_y_Colectiva_Representaciones_sociales_de_la_historia/links/566b023f08ae62b05f046941/Memoria-social-y-Colectiva-Representaciones-sociales-de-la-historia.pdf)
- Patiño, P. (2013). “Evolución de la investigación y el desarrollo tecnológico en los dos últimos siglos”. En: Ensayos Pensamiento Universitario. Medellín: Universidad de Antioquia. Recuperado de: <http://www.udea.edu.co/wps/wcm/connect/udea/bbc7300d-4ffd-40f6-aea4-e92def2f04e3/evolucion-investigacion-desarrollo-tecnologico-ciencia.pdf?MOD=AJPERES&CVID=kQRgqK6>
- Pech, F., Gómez, M. A., & de la Cruz Diaz, L. A. (2012). Desarrollo de Aplicaciones web con JPA, EJB, JSF y PrimeFaces. Instituto Tecnológico Superior de los Ríos. Recuperado de: <https://www.tamps.cinvestav.mx/~fpech/sd/files/paper001.pdf>
- PrimeFaces (2019), Why PrimeFaces? PrimeFaces. PrimeTek Informatics. Recuperado de: <https://www.primefaces.org/showcase/>
- Python (2018). *Python de pila completa*. Recuperado de: <https://www.fullstackpython.com/peewee.html>
- Python (2019). *The Python Tutorial*, <https://www.python.org/doc/> Recuperado de: <https://docs.python.org/3/tutorial/index.html>
- Ríos, J. I. (2005). Programación orientada a objetos en lenguajes no orientados a objetos: c, una experiencia. *Scientia et technica*, 3(29). Recuperado de: <http://www.redalyc.org/pdf/849/84911948018.pdf>
- Scrapy (2016). *Descripción de la arquitectura*. Recuperado de: <https://doc.scrapy.org/en/latest/topics/architecture.html>
- Scrapy (2016). *Documentos y procesamientos Scrapy*. Recuperado de: <https://doc.scrapy.org/en/latest/intro/overview.html>
- Scrapy (2018). *Docs: Scrapy at a glance: Read the Docs*. Retrieved from: <https://doc.scrapy.org/en/latest/intro/overview.html>.
- Sourceforge (2018). *iText, A Java PDF Library*. Recuperado de: <https://sourceforge.net/projects/itext/>

- Thibaud, C. (2006). *MySQL 5: instalación, implementación, administración, programación*. Ediciones ENI. Recuperado de: <https://books.google.com.co/books?id=wY0bHPmW-NUC&pg=PA30&dq=que+es+mysql&hl=es&sa=X&ved=0ahUKEwjOpsu69abjAhXouFkKHeRFBboQ6AEIKDAA#v=onepage&q=que%20es%20mysql&f=false>
- Universidad de Alicante (2018). *Modelo vista controlador (MVC)*. Disponible en: <https://si.ua.es/es/documentacion/asp-net-mvc-3/1-dia/modelo-vista-controlador-mvc.html>
- Universidad del Valle (2019). ¿Dónde estudiar? Estas son las mejores universidades de Colombia. Cali, Colombia: Agencia de noticias Univalle. Recuperado de: <https://www.univalle.edu.co/mejoresuniversidadesdecolombia>
- Universidad industrial de Santander (2018). *Principios e información general*. Bucaramanga, Colombia. Recuperado de: <https://www.uis.edu.co/webUIS/es/acercaUis/principios.html>
- Wipo (2004), ¿Qué es la propiedad intelectual?: Organización Mundial de la propiedad intelectual. recuperado de: <https://www.wipo.int/publications/es/details.jsp?id=99&plang=ES>