



**VALIDACIÓN DE UN DOCUMENTO WORD BAJO LA NORMA NTC 1486 A PARTIR  
DE LA METADATA**

**JULIAN FELIPE PLATA ZUÑIGA**

**UNIVERSIDAD CATÓLICA DE COLOMBIA  
FACULTAD DE INGENIERÍA  
PROGRAMA DE INGENIERÍA DE SISTEMAS Y COMPUTACIÓN  
BOGOTÁ D.C  
2020**

**Validación de un documento Word bajo la norma NTC 1486 a partir de la metadata**

**Julián Felipe Plata Zúñiga**

**Trabajo de Grado para Optar al Título de  
Ingeniero de Sistemas**

**DIRECTOR  
HOLMAN DIEGO BOLÍVAR BARÓN Ph. D**

**UNIVERSIDAD CATÓLICA DE COLOMBIA  
FACULTAD DE INGENIERÍA  
PROGRAMA DE INGENIERÍA DE SISTEMAS Y COMPUTACIÓN  
BOGOTÁ D.C  
2020**



## Atribución 2.5 Colombia (cc BY 2.5 CO)

Este es un resumen legible por humanos (y no un sustituto) de la [licencia](#).

[Advertencia](#)



### Usted es libre para:



Compartir — copiar y redistribuir el material en cualquier medio o formato

Adaptar — remezclar, transformar y crear a partir del material

Para cualquier propósito, incluso comercialmente

El licenciante no puede revocar estas libertades en tanto usted siga los términos de la licencia

### Bajo los siguientes términos:



**Atribución** — You must give [appropriate credit](#), provide a link to the license, and [indicate if changes were made](#). You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

**No hay restricciones adicionales** — Usted no puede aplicar términos legales ni [medidas tecnológicas que restrinjan legalmente a otros hacer cualquier uso permitido por la licencia](#).

### Aviso:

[Usted no tiene que cumplir con la licencia para los materiales en el dominio público o cuando su uso esté permitido por una excepción o limitación aplicable.](#)

No se entregan garantías. La licencia podría no entregarle todos los permisos que necesita para el uso que tenga previsto. Por ejemplo, otros derechos como [relativos a publicidad, privacidad, o derechos morales](#) pueden limitar la forma en que utilice el material.

## NOTA DE ACEPTACIÓN

Aprobado por el comité de grado en cumplimiento de los requisitos exigidos por la Facultad de Ingeniería y la Universidad Católica de Colombia para optar al título de Ingenieros de Sistemas.

---

Jurado1

---

Jurado2

---

Holman Diego Bolívar  
Barón, Ph.D Asesor

Bogotá, mayo 22 de 2020

## **AGRADECIMIENTOS.**

Quiero agradecerle a Dios por darme las fuerzas y las capacidades suficientes para emprender mi camino como ingeniero, ya que sin su ayuda nada de esto hubiera sido posible.

Agradezco a mis padres por brindarme apoyo tanto económico como emocional en todos los retos de mi día a día, es por ellos que quiero alcanzar mis metas y sueños ya que estoy eternamente agradecido con ellos y quiero retribuirles todo lo que me han dado.

Agradezco a Shirley Sharay Menco Quintero, mi mejor amiga y compañera de muchos años atrás, por estar conmigo en las buenas y en las malas, por su compañía incondicional, además de aportar sus conocimientos y consejos para mi formación profesional y personal.

A Cristian González y Mauricio Rodríguez, por su apoyo y amistad incondicional en mi vida universitaria, ya que entre todos nos apoyamos para crecer como persona y como profesionales.

## CONTENIDO

	<b>pág.</b>
Glosario .....	13
RESUMEN .....	15
INTRODUCCIÓN .....	16
1. JUSTIFICACIÓN.....	17
2. PLANTEAMIENTO DEL PROBLEMA.....	20
3. OBJETIVO GENERAL.....	24
3.1. OBJETIVOS ESPECÍFICOS .....	24
4. MARCO TEÓRICO .....	25
4.1. NORMA TÉCNICA COLOMBIANA 1486 .....	25
4.2. WEB SEMÁNTICA .....	25
4.3. BIG DATA COMO BASE PARA EL ANÁLISIS Y LA EXTRACCIÓN DE DATOS.....	28
4.3.1. ¿Qué son los archivos binarios?.....	29
4.4. PROCESO DE ANALISIS DE DATOS .....	29
4.5. SOFTWARE PARA EXTRACCIÓN DE METADATOS.....	30
4.6. MODELO DE DATOS .....	33
4.6.1. Modelo conceptual.....	34
4.6.2. Modelo de datos lógico.....	35
4.6.3. Modelo de datos físico.....	35
4.6.5. Modelo de datos estructurado.....	36
4.6.6. Datos semiestructurados .....	37
4.6.7. Datos no estructurados .....	37
4.6.8. Base de datos Clave Valor .....	38
4.6.9. Base de datos Documentales .....	39
4.6.10. Bases de datos en grafo .....	40
5. MARCO CONCEPTUAL .....	42
5.1. INGENIERÍA DE SOFTWARE .....	42
5.1.1. Ingeniería inversa .....	43
5.2. DESARROLLO WEB.....	45
5.3. ARQUITECTURA MVC – MODELO VISTA CONTROLADOR .....	45
6. ESTADO DEL ARTE.....	47

7. METODOLOGIA.....	50
8. RESULTADOS .....	52
8.1. REQUERIMIENTOS DEL SISTEMA .....	52
8.1.1. Requerimientos funcionales. ....	52
8.2. CASOS DE USO.....	53
8.2.1. Caso de uso 1: Iniciar sesión.....	55
8.3. REQUERIMIENTOS NO FUNCIONALES.....	56
8.3.1. Requisito no funcional 1: Interfaz responsive. ....	56
8.3.2. Requisito no funcional 2: Permisos de acceso al sistema. ....	56
8.4. ATRIBUTOS DE CALIDAD .....	40
8.5. ARQUITECTURA DEL SISTEMA .....	57
8.5.1. Stakeholders.....	57
8.5.2. Diagrama de arquitectura del sistema .....	58
8.6. Manual de instalación del sistema .....	58
8.6.1. Requisitos mínimos del hardware.....	59
8.6.2. Instalación de componentes.....	59
8.7. MANUAL DE USUARIO.....	59
8.7.1. Inicio de sesión.....	59
8.7.2. Crear cuenta.....	60
8.8. PRUEBAS Funcionales DEL SISTEMA .....	61
8.8.1 Casos de prueba.....	61
9. CONCLUSIONES.....	64
10. TRABAJOS FUTUROS .....	65
11. Bibliografía.....	66
INTRODUCCIÓN.....	75
1. PROPÓSITO.....	76
1.1. ÁMBITO DEL SISTEMA.....	76
1.2. DEFINICIONES, ACRÓNIMOS Y ABREVIATURAS.....	77
1.3. VISIÓN GENERAL DEL DOCUMENTO.....	77
2. DESCRIPCIÓN DEL SISTEMA .....	78
2.1. PERSPECTIVA DEL PRODUCTO.....	78
2.2. REQUISITOS FUNCIONALES.....	78

Requerimiento Funcional 1: Iniciar Sesión.....	79
Requerimiento funcional 2: Crear cuenta.....	79
Requerimiento funcional 3: Recuperar contraseña.....	80
Requerimiento funcional 4: Editar cuentas de usuario.....	80
Requerimiento funcional 5: Editar administrador .....	82
Requerimiento funcional 6: Editar Usuario.....	83
Requerimiento funcional 7: Seleccionar usuario.....	84
Requerimiento funcional 8: Consultar Validaciones Realizadas.....	84
Requerimiento funcional 9: Descargar Validación.....	85
Requerimiento funcional 10: Cargar documento.....	85
Requerimiento funcional 11: Cargar información del trabajo de grado.....	86
Requerimiento funcional 12: Obtener información del documento.....	87
Obtener historial de documentos.....	88
Obtener historial de validaciones.....	89
2.1. REQUERIMIENTOS NO FUNCIONALES.....	90
Requisito no funcional 1: Integridad.....	90
Requerimiento no funcional 2: Escalabilidad .....	90
2.3. FUNCIONES DEL PRODUCTO.....	91
Detalle de Casos de Uso.....	1
2.4. CARACTERÍSTICAS DE LOS USAURIOS.....	17
2.5. RESTRICCIONES.....	18
2.6. SUPOSICIONES Y DEPENDENCIAS.....	18
3. REQUISITOS ESPECÍFICOS.....	19
3.1. INTERFACES EXTERNAS.....	19
3.1.1. Interfaces de usuario.....	19
Interfaces de software.....	19
1. INTRODUCCIÓN:.....	24
2. STAKEHOLDERS.....	25
3. REPRESENTACIÓN ARQUITECTÓNICA Y PUNTOS DE VISTA .....	26
3.1. VISTA FÍSICA .....	26
3.1.1. Arquitectura del sistema.....	26
3.1.2. Diagrama de Despliegue:.....	26



3.2. VISTA LÓGICA: .....	27
3.2.1. Diagrama de Componentes:.....	27
3.2.2. Diagrama de Clases .....	28
3.2.3. Modelo de datos .....	1
Introducción.....	5
1. REQUISITOS MINIMOS DEL HARDWARE: .....	6
1.1. SERVIDOR: .....	6
1.2. CLIENTE:.....	6
2. REQUISITOS DE SOFTWARE: .....	7
2.1. SERVIDOR: .....	7
2.2. CLIENTE:.....	7
3. INSTALACIÓN DE COMPONENTES:.....	8
3.1. INSTALACIÓN DE BASES DE DATOS MYSQL Y WORKBENCH:.....	8
3.2. INSTLACIÓN DE PYTHON:.....	10
3.3. INSTALACIÓN DE PIP:.....	11
3.4. INSTALACIÓN DE DJANGO: .....	11
3.5. INSTALACION DE PYMYSQL: .....	12
3.6. INSTALaCIÓN DE ZIPFILE:.....	12
3.7. INSTALACION DE PYTHON-LXML:.....	13
4. DESPLIEGUE DE LA APLICACIÓN.....	14
4.1. DESCARGAR NGINX: .....	14
4.2. INICIAR NGINX:.....	14
4.3. CONFIGURACIÓN POR DEFECTO DE NGINX PARA EL DESPLIEGUE DEL SITIO:.....	15
4.4. INSTALACIÓN DE GUNICORN:.....	16
Contenido .....	2
INTRODUCCION .....	4
1. FUNCIONALIDADES DEL SISTEMA. ....	5
1.1. INICIO DE SESIÓN.....	5
1.2. CREAR CUENTA.....	5
2. Gestión de documentos .....	7
2.1. Lista de archivos cargados.....	7
2.2. CARGAR DOCUMENTO. ....	8

2.3.	ANALISIS DE DOCUMENTOS. ....	8
2.4.	DESCARGAR RESULTADO DE VALIDACIÓN DE DOCUMENTOS. ....	9
3.	Administración de USUARIOS.....	11
3.1.	LISTADO Y SELECCIÓN DE USUARIOS .....	11
3.2.	RECUPERACIÓN DE CONTRASEÑA.....	11
3.3.	ADMINISTRACIÓN DEL SISTEMA.....	12
3.4.	Cambio de permisos de usuario.....	13
3.5.	Roles de usuario. ....	14
	Contenido.....	2
1.	INTRODUCCIÓN.....	5
2.	CASOS DE PRUEBA.....	6
2.	BIBLIOGRAFIA.....	17

## TABLA DE IMÁGENES

	pág.
Figura 1: Metadatos de un archivo formato .DOC .....	22
Figura 2: Metadatos de un archivo formato PPTX.....	23
Figura 3: Arquitectura de web semántica. ....	26
Figura 4: Mapa conceptual de la web semántica. ....	27
Figura 5: Formato XML.....	28
Figura 6: Fase de análisis de datos.....	30
Figura 7: Graph extrae datos de un grafo. ....	32
Figura 8: Características de los modelos de datos.....	<b>¡Error! Marcador no definido.</b>
Figura 9: Modelo de datos Conceptual.....	34
Figura 10: Modelo de datos lógico. ....	35
Figura 11: Modelo de datos físico. ....	36
Figura 12: Formato JSON. ....	<b>¡Error! Marcador no definido.</b>
Figura 13: Formato XML. ....	<b>¡Error! Marcador no definido.</b>
Figura 14: Base de datos Cadena Valor. ....	38
Figura 15: Modelo de datos conceptual. ....	39
Figura 16: Base de datos en grafo. ....	40
Figura 17: Proceso de Ingeniería. ....	44
Figura 18: Diagrama de casos de uso.....	54
Figura 19: Arquitectura del sistema ValidateDocx.....	58
Figura 20: Inicio de sesión.....	60
Figura 21: Crear cuenta. ....	60
Figura 22: Evidencia de inicio de sesión .....	62
Figura 23: Evidencia, registro de usuario. ....	63

## Lista de Tablas

	pág.
Tabla 1: Requerimiento Funcional Iniciar Sesión. ....	53
Tabla 2: Caso de Uso 1: Iniciar Sesión .....	55
Tabla 3: Requerimiento no Funcional 1: Interfaz Responsive .....	56
Tabla 4: Requerimiento no Funcional 2: Permisos de Acceso al sistema .....	56
Tabla 5: Requerimiento no funcional 3: Errores Informativos.....	56
Tabla 6: Stakeholders del sistema.....	57
Tabla 7: Caso de prueba: Inicio de sesión. ....	61
Tabla 8: Caso de prueba: Crear cuenta. ....	62

## Glosario

**Dataset:** “Una colección de conjuntos de información relacionados que se compone de elementos separados pero que una computadora puede manipular como una unidad”.<sup>1</sup>

**World Wide Web:** “Un sistema de información en Internet que permite que los documentos se conecten a otros documentos mediante enlaces de hipertexto, lo que permite al usuario buscar información moviéndose de un documento a otro”.<sup>2</sup>

**Web Semántica:** “La Web Semántica es una Web extendida, dotada de mayor significado en la que cualquier usuario en Internet podrá encontrar respuestas a sus preguntas de forma más rápida y sencilla gracias a una información mejor definida”.<sup>3</sup>

**XML:** “Se utiliza para el intercambio de datos estructurados. Más que un formato de archivos rígido, XML es un lenguaje que define los formatos aceptados que pueden utilizar los grupos para intercambiar información.”<sup>4</sup>

**Prototipo:** Un prototipo es un primer modelo que sirve como representación o simulación del producto final y que nos permite verificar el diseño y confirmar que cuenta con las características específicas planteadas.<sup>5</sup>

**Datos:** Se conoce que la palabra Datos proviene del latín “**Dtum**” cuyo significado es “**lo que se da**”. Los datos son la representación simbólica, bien sea mediante números o letras de una recopilación de información la cual puede ser cualitativa o cuantitativa, que facilitan la deducción de una investigación o un hecho.<sup>6</sup>

**Diccionario de datos:** Un conjunto de información que describe el contenido, el formato y la estructura de una base de datos y la relación entre sus elementos, que se utiliza para controlar el acceso y la manipulación de la base de datos.<sup>7</sup>

---

<sup>1</sup> LEXICO POERED BY OXFORD, UK Dictionary; [Consultado el 10 de 03 de 2020]. Disponible en: [https://www.lexico.com/definicion/data\\_set](https://www.lexico.com/definicion/data_set)

<sup>2</sup> LEXICO POERED BY OXFORD, UK Dictionary; [Consultado el 10 de 03 de 2020]. Disponible en: [https://www.lexico.com/definicion/world\\_wide\\_web](https://www.lexico.com/definicion/world_wide_web)

<sup>3</sup> W3C, Standards; [Consultado el 10 de 03 de 2020] Disponible en: <https://www.w3.org/standards/semanticweb/>

<sup>4</sup> FILMAKER PRO, XML format;[Consultado el 10 de 03 de 2020 ] Disponible en: [https://fmhelp.filemaker.com/help/18/fmp/es/index.html#page/FMP\\_Help/xml-format.html.p.8](https://fmhelp.filemaker.com/help/18/fmp/es/index.html#page/FMP_Help/xml-format.html.p.8).

<sup>5</sup> OXFORD LEARNER’S DICTIONARIES, definition english; [Consultado el 10 de 03 de 2020]. Disponible en: <https://www.oxfordlearnersdictionaries.com/definition/english/prototype>

<sup>6</sup> CONCEPTDEFINICION, Datos; [Consultado el 10 de 03 de 2020]. Disponible en: <https://conceptdefinicion.de/datos/>.

<sup>7</sup> OXFORD LEARNER’S DICTIONARIES, definition english; [Consultado el 10 de 03 de 2020]. Disponible en: [https://www.oxfordlearnersdictionaries.com/definicion/english/data\\_dictionary](https://www.oxfordlearnersdictionaries.com/definicion/english/data_dictionary)

**Metadatos:** “Información que describe otra información con el fin de ayudarle a entenderla o utilizarla”.<sup>8</sup>

---

<sup>8</sup> OXFORD LEARNER'S DICTIONARIES, Definition of metadata noun, 2020. Disponible en: Oxford Learner's Dictionaries, Definition of metadata noun

## RESUMEN

La investigación es uno de los pilares fundamentales para el desarrollo de la sociedad y la construcción de nuevo conocimiento, razón por la que se hace necesario que todas las personas que cuentan con una formación de carácter profesional participen en procesos investigativos, reconociendo las herramientas que dichas experiencias suponen para su desempeño profesional y su inmersión dentro del contexto laboral.

Estos procesos investigativos presentan unos requisitos mínimos relacionados con la calidad y pertinencia **de acuerdo con el objeto de estudio** y el campo donde se enmarquen, manifestados a través de las evidencias escritas y la construcción de documentos como tesis o monografías que funcionan como insumos para determinar nuevos temas de investigación surgidos a partir de las necesidades identificadas en cada campo. Por tanto, los documentos escritos representan un elemento importante dentro de la comunidad investigativa, y están sujetos a normas que regulan la calidad en términos de redacción forma y contenido. Estas normas son estandarizadas por diferentes entidades nacionales e internacionales, como lo es el caso de la norma NTC 1486 sexta edición, de las normas Icontec, que comprende una serie de sugerencias y requisitos para garantizar la calidad de un documento en aspectos de contenido y forma. Sin embargo, no existe una herramienta automatizada para identificar el cumplimiento de diferentes documentos respecto a esta norma, por lo cual la presente investigación se encuentra basada en la validación de un documento a partir del **metadato** de forma automatizada.

En el presente documento se encuentra contemplado el fundamento teórico de los tipos de datos y de los modelos de datos que se manejan en la actualidad, así como la web semántica y aspectos de ingeniería de software requeridos para la construcción de un prototipo de aplicación web que automatice el proceso de extracción de metadatos en un documento de Word, validando que la información contenida en este cumpla con la NTC 1486 en su totalidad, siendo una herramienta que facilite el proceso de calificación y caracterización de trabajos de grado y documentos de tipo investigativo. Esta investigación se basa en la metodología “Personal Software Process” o por sus siglas PSP, que brinda insumos para el desarrollo de acciones de programación en cualquier tipo de proyectos de desarrollo, dando lineamientos respecto a la planeación y ejecución del proceso.

Durante este proyecto de investigación, fue posible construir una herramienta llamada ValidateDocx, cuya función permitió el análisis de los metadatos de un archivo de Word de forma exitosa, arrojando información cualitativa y cuantitativa respecto a la forma y contenido de este. Esta investigación tiene un impacto dentro de la comunidad investigativa en la medida en que genera un prototipo de software que facilita el proceso de revisión de documentos investigativos en construcción.

Palabras Claves: Investigación, Metadatos, Trabajos de grado, NTC 1486.

## INTRODUCCIÓN

En la actualidad, la sociedad requiere de la creación de herramientas tecnológicas que vayan acorde con la evolución y la necesidad de las personas a partir del uso de recursos tecnológicos que faciliten las tareas que deben desempeñar, por lo cual, cada vez se busca una mayor articulación de los avances tecnológicos con el desarrollo personal y profesional de las personas en general.

Teniendo en cuenta la que el uso de herramientas tecnológicas facilita la creación y expansión de la información, la comunidad científica ha tenido transformaciones en la manera en que los procesos investigativos son publicados y divulgados. Debido a esto, la mayoría de las investigaciones en la actualidad obedecen a medios de comunicación como la internet en el momento de su exposición frente al resto de la comunidad, razón por la que los documentos escritos que se encuentran disponibles en la red requieren de un seguimiento en términos de calidad investigativa, con el fin de generar evidencias pertinentes para el campo y objeto de estudio a los que correspondan, estas evidencias abarcan el contenido y forma de las diferentes publicaciones de investigación, tomando como referente alguna de las normas existentes para la configuración de cada tipo de documento. Estas acciones requieren la supervisión y direccionamiento de expertos en el campo de investigación que se esté trabajando, sin embargo, en reiteradas ocasiones se puede identificar que las personas que realizan esta supervisión, pasan por alto detalles mínimos pero importantes en relación con las normas técnicas que regulan la escritura de este tipo de documentos, por lo cual se hace necesaria la automatización de este proceso de validación a través de la construcción de herramientas tecnológicas que promuevan la efectividad y excelencia en este tipo de procesos.

Teniendo en cuenta que actualmente los documentos son almacenados y publicados por medio de internet, el prototipo de software presentado en este proyecto de investigación busca la validación de un documento de Word a partir de los metadatos, tomando como referencia la NTC 1486, norma técnica que da lineamientos para la construcción de documentos de carácter investigativo en diferentes niveles de educación superior en Colombia.

El impacto que se puede generar con la creación de este prototipo a nivel de búsqueda de información es que se pueda facilitar el acceso a los metadatos de archivos editables, optimizando las búsquedas sobre información que puede ser útil para un usuario, además de aportar nuevos conocimientos que permitirán generar herramientas de búsqueda más eficaces permitiendo recolectar más cantidad de información fácilmente.



## 1. JUSTIFICACIÓN

En los últimos 20 años gran parte del conocimiento se ha publicado a través de medios tecnológicos como lo es el internet. El reciente desarrollo de la economía se ha centrado en automatizar y actualizar las herramientas con la que se permite y se busca la información ya sea en dispositivos móviles, computadoras, tabletas etc. Estos medios tecnológicos permiten que la información sea divulgada y así mismo pueda ser accedida de una forma mucho más sencilla y rápida para casi cualquier usuario<sup>9</sup>.

Teniendo en cuenta la importancia de los medios tecnológicos para la divulgación de la información, manejo y transformación de datos, son importantes en conceptos como Big data, ya que permite clasificar datos en estructuras organizadas con el propósito de hacer un análisis para inferir cualquier tipo de información que pueda ser valiosa para un usuario. Toda información es representada por un conjunto de datos (dataset). Según Steven Lynden, “Un conjunto de datos corresponde a los contenidos de una única tabla de base de datos o una única matriz de datos de estadística, donde cada columna de la tabla representa una variable en particular, y cada fila representa a un miembro determinado del conjunto de datos que se está tratando”<sup>10</sup>,

Todo dataset tiene información al respecto de sí mismo lo cual nos ayuda a describir cómo está organizada la información, a qué tipo de dato pertenece, estadísticamente como está representado. Este subconjunto de datos asociados a los dataset se conoce como metadatos, por lo que actualmente es almacenado de manera masiva gracias al Cloud Computing<sup>11</sup>, con el fin entre otros de ser analizado o utilizado para diferentes fines, esto también es llamado Big data, el cual ha convertido en un conjunto de herramientas de gran impacto social como lo son Google apps, que permite a los usuarios acceder más fácilmente a casi cualquier tipo de información desde diferentes dispositivos en aplicaciones diferentes como Google calendar, Google Docs, Dropbox entre otros, estas son herramientas que aportan a la adopción masiva del Cloud Computing.

A medida que pasan los años, los medios de difusión de la información como lo es el internet, avanzan rápidamente así como también las herramientas de búsqueda para acceder a ella, un claro ejemplo podría ser el navegador de Google, el cual busca facilitar el acceso a la información para los usuarios utilizando algoritmos de búsqueda avanzada y así mostrar la mayor cantidad posible información relacionada con lo que el usuario busca, es aquí donde la búsqueda binaria juega un papel importante “La búsqueda binaria

---

<sup>9</sup> ARRIOLA NAVARRETE, Oscar; BUTRÓN YÁÑEZ, Katya. Sistemas integrales para la automatización de bibliotecas basados en software libre. *Acimed*, 2008, vol. 18, no 6, p. 0-0.

<sup>10</sup> LYNDEN, Steven; TAVEEKARN, Waran. Semi-automated augmentation of pandas dataframes. En *International Conference on Data Mining and Big Data*. Springer, Singapore, 2019. p. 70-79.

<sup>11</sup> HAYES, Brian. Cloud computing. 2008.

es un algoritmo eficiente para encontrar un elemento en una lista ordenada de elementos. Funciona al dividir repetidamente a la mitad la porción de la lista que podría contener al elemento, hasta reducir las ubicaciones posibles a solo una<sup>12</sup> cómo lo dice la definición, para que la búsqueda binaria tenga éxito es importante que los elementos a buscar estén organizados, es por esta razón que los modelos de datos también juegan un papel importante en todo este proceso.

De igual manera, la construcción de conocimiento en la sociedad evoluciona, por lo que los gremios de ciencia e investigación hacen uso de herramientas tecnológicas tanto para el desarrollo de nuevas propuestas como para la regulación y validación de documentos científicos e investigativos. Estas herramientas tecnológicas tienen un papel importante en el proceso de investigación ya que permiten modernizar el paradigma investigativo y facilitan el intercambio de información verídica, el diálogo de saberes de diferentes disciplinas y la rigurosidad de acciones investigativas.

Teniendo en cuenta lo expuesto y con la ayuda de trabajos anteriores se plantea realizar un sistema que permita la extracción de datos y metadatos de archivos binarios, esta herramienta tiene como propósito validar un documento de Word a la luz de la norma bajo la cual debe estructurarse, en este caso la NTC 1486. Dicha extracción de metadatos permitirá identificar el nivel de cumplimiento que un documento tiene respecto a la norma, promoviendo la rigurosidad investigativa y aportando una herramienta de seguimiento para el desarrollo e implementación de propuestas investigativas en el país.

Por ejemplo, si fuese necesario validar el cumplimiento de criterios de estructura y contenido de varios archivos de Word de forma manual, tomaría un tiempo mayor que si se realiza a partir del metadato con un software que identifique no solamente el nivel de cumplimiento sino también los errores específicos en el documento y genere así una retroalimentación al autor que le permita mejorar la calidad investigativa de un documento en construcción.

Para fuentes específicas de este proyecto se ha decidido empezar con el uso de la suite de Microsoft Office por el gran uso que se le está dando a esta herramienta a nivel mundial<sup>13</sup>, siendo los sectores públicos y educativos aquellos que utilizan de forma constante cualquiera de los componentes principales de esta suite, herramientas como Word, Excel, PowerPoint entre otros, por lo que sí es posible realizar una extracción de datos y metadatos exitosa, puede lograrse a futuro una fácil identificación de cada uno de los elementos extraídos (datos y metadatos) para lograr centrar estos en un estudio analítico. Por lo que puede considerarse como una proyección no de volumen de datos, sino de variedad de estos.

---

<sup>12</sup> KHANACADEMY. Computer-science; [Consultado el 02 de 03 de 2020 ] Disponible en : <https://es.khanacademy.org/computing/computer-science/algorithms/binary-search/a/binary-search>

<sup>13</sup> MONTALBANO, Elizabeth. Forrester: Microsoft office in no danger from competitors. PC World, 2009.



## 2. PLANTEAMIENTO DEL PROBLEMA

La Norma Técnica Colombiana NTC 1486 es la encargada de estandarizar una serie de requerimientos de forma y contenido de un documento formal como por ejemplo trabajos de grado y monografías<sup>14</sup>. Dicha norma es aplicada por una gran variedad de investigadores colombianos, ya que presenta una guía para la redacción y esquematización de documentos investigativos, con el fin de dar una estructura generalizada a estos trabajos<sup>15</sup>, teniendo en cuenta el tipo y tamaño de letra, el interlineado, las márgenes, partes del texto y tipo de referenciación, entre otros.

Ahora bien, en muchas ocasiones algunos de sus aspectos más relevantes, e incluso algunas secciones obligatorias pueden no evidenciarse en los documentos publicados, ya que el análisis del cumplimiento de criterios relacionados con la norma NTC 1486 se realiza de forma manual por los asesores o investigadores que se encuentran inmersos en el proceso investigativo, por lo cual no hay un control de la calidad de los archivos, ni del cumplimiento que estos tienen en cuanto a los criterios establecidos por la NTC 1486<sup>16</sup>. Esto afecta a la comunidad investigativa, ya que disminuye las posibilidades de que un trabajo de grado sea referenciado en otras investigaciones si no cumple con los parámetros de calidad manifestados en la NTC 1486<sup>17</sup>. Sin embargo, no existe un estudio que mida el porcentaje de cumplimiento que un documento presenta frente a la norma, ni un programa que genere esta revisión de forma automática.

Teniendo en cuenta la cantidad de tiempo y recurso humano que conlleva realizar un análisis de la forma y contenido de un trabajo de grado o monografía, por lo cual se puede hacer uso de las herramientas tecnológicas, tales como la extracción de metadatos, para dar una solución a la necesidad de revisar estos documentos, automatizando el proceso de revisión y generando mayor confiabilidad y pertinencia del proceso<sup>18</sup>.

---

<sup>14</sup> INSTITUTO COLOMBIANO DE NORMAS TÉCNICAS Y CERTIFICACIÓN. Documentación, Presentación de tesis, trabajos de grado y otros trabajos de investigación, Norma Técnica Colombiana. NTC 1486. [En línea] Bogotá D.C.: el instituto. 41 p.

<sup>15</sup> UNIVERSIDAD CATÓLICA DE COLOMBIA. Dirección central de investigaciones [Sitio Web]. Bogotá; [Consultado 22 de febrero de 2020] Disponible en: <https://www.ucatolica.edu.co/portal/investigaciones/direccion-central-de-investigaciones/>

<sup>16</sup> CORNEJO, Marcela; SALAS, Natalia. Rigor y calidad metodológicos: un reto a la investigación social cualitativa. *Psicoperspectivas*, 2011, vol. 10, no 2, p. 12-34.

<sup>17</sup> MINISTERIO DE CIENCIA, TECNOLOGÍA E INNOVACIÓN. La ciencia en cifras [Sitio Web]. Bogotá; [Consultado el 23 de febrero de 2020] Disponible en: <https://minciencias.gov.co/la-ciencia-en-cifras>

<sup>18</sup> CASALI, A, et al. Asistente para el depósito de objetos en repositorios con extracción automática de metadatos. En *XV Simposio Internacional de Tecnologías de la Información y las Comunicaciones en la Educación SINTICE*. 2013. p. 133-136.

Es por esto que es necesario desarrollar una herramienta que permita la revisión detallada de los documentos ya publicados, para examinar el cumplimiento que tienen de acuerdo a la norma bajo la cual se rige su calidad, con miras a identificar debilidades en común y así proyectar un plan de mejoramiento para que los documentos a publicar se encuentren más cercanos a los estándares de calidad esperados por la comunidad investigativa<sup>19</sup>.

Hoy en día existen algunas herramientas de extracción de datos y metadatos para diferentes tipos de archivos binarios<sup>20</sup>, como por ejemplo PDF y algunos formatos de imágenes abiertos, pero no de formatos complejos como los de documentos editables. Estas herramientas de extracción que existen en la actualidad pueden extraer varias características de estos archivos permitiendo así separar todo tipo de información incorporada (como imágenes, texto, tablas, referencias y generación de estructura jerárquica del archivo)<sup>21</sup>. Cada una de estas herramientas sigue múltiples sentencias que permiten organizar y catalogar los datos de forma fácil y comprensible para que se pueda dar un uso futuro (un análisis de datos, por ejemplo), todo esto es posible dependiendo del tipo de formato del archivo, pero esto no se puede hacer actualmente de forma masiva.

Si se quisiera extraer datos de un archivo de Word o PowerPoint o cualquiera que sea de la suite de Microsoft, por ejemplo, se podría establecer características o palabras clave que permiten identificar los metadatos que desean extraer. Sin embargo, cada archivo tiene una estructura de datos interna completamente diferente a los demás<sup>22</sup>, lo que nos obliga a diseñar o a adquirir software específicamente creado para manipular un formato del archivo que se quiera abrir o extraer, en la figura 1 y figura 2 podemos ver la estructura interna que tiene un archivo de Word (.DOCX) y de PowerPoint (.PPTX):

---

<sup>19</sup> ARRIETA, Juan Gregorio, et al. Aplicación lean manufacturing en la industria colombiana. Revisión de literatura en tesis y proyectos de grado. Revista Virtual Pro, 2011, vol. 132, no 9.

<sup>20</sup> SHADISH, W. R., BRASIL, I. C. C., ILLINGWORTH, D. A., WHITE, K. D (2009). Using UnGraph to extract data from image files: Verification of reliability and validity. Behavior Research Methods, 41(1), 177–183. <https://doi.org/10.3758/BRM.41.1.177>

<sup>21</sup> BARRIGA MARIÑO, J. C., & RINCON YAÑEZ, D. A. (2017). Construcción De Una Herramienta Web Para La Extracción Y Almacenamiento De Datos De La Red Social Twitter. Universidad Católica de Colombia.

<sup>22</sup> PRIETO, J. (2016a). Lo que la extensión DOCX esconde. Retrieved from <https://www.gonduana.com/lo-que-la-extension-docx-esconde/>

Figura 1: Metadatos de un archivo formato .DOC

```
<?xml version="1.0" encoding="UTF-8" standalone="true"?>
- <Properties
xmlns:vt="http://schemas.openxmlformats.org/officeDocument/2006/docPropsVTypes"
xmlns="http://schemas.openxmlformats.org/officeDocument/2006/extended-
properties">
  <Template>Normal.dotm</Template>
  <TotalTime>3</TotalTime>
  <Pages>1</Pages>
  <Words>8</Words>
  <Characters>48</Characters>
  <Application>Microsoft Office Word</Application>
  <DocSecurity>0</DocSecurity>
  <Lines>1</Lines>
  <Paragraphs>1</Paragraphs>
  <ScaleCrop>false</ScaleCrop>
  <Company/>
  <LinksUpToDate>false</LinksUpToDate>
  <CharactersWithSpaces>55</CharactersWithSpaces>
  <SharedDoc>false</SharedDoc>
  <HyperlinksChanged>false</HyperlinksChanged>
  <AppVersion>14.0000</AppVersion>
</Properties>
```

Fuente: PRIETO, J. (2016b). Lo que la extensión DOCX esconde. Retrieved from <https://www.gonduana.com/lo-que-la-extension-docx-esconde/>

En la figura 1 se evidencia que el contenido del documento está dentro de etiquetas XML lo cual permite buscar información en específico entendiendo el significado de las etiquetas.

Figura 2: Metadatos de un archivo formato PPTX.

```
</a:p>
<a:p>
  <a:pPr indent="0" lvl="0" marL="2286000" marR="0" rtl="0" algn="l">
    <a:lnSpc>
      <a:spcPct val="100000" />
    </a:lnSpc>
    <a:spcBef>
      <a:spcPts val="0" />
    </a:spcBef>
    <a:spcAft>
      <a:spcPts val="0" />
    </a:spcAft>
    <a:buClr>
      <a:schemeClr val="dk1" />
    </a:buClr>
    <a:buSzPts val="1600" />
    <a:buFont typeface="Calibri" />
    <a:buNone />
  </a:pPr>
  <a:r>
    <a:rPr b="1" lang="en-US" sz="1600">
      <a:solidFill>
        <a:schemeClr val="dk1" />
      </a:solidFill>
    </a:rPr>
    <a:t>Código: 625448</a:t>
  </a:r>
  <a:endParaRPr b="1" i="0" sz="1600" u="none" cap="none" strike="noStrike">
    <a:solidFill>
      <a:schemeClr val="dk1" />
    </a:solidFill>
    <a:latin typeface="Arial" />
    <a:ea typeface="Arial" />
    <a:cs typeface="Arial" />
    <a:sym typeface="Arial" />
  </a:endParaRPr>
</a:p>
</p:txBody>
</p:sp>
<p:sp>
  <p:nvSpPr>
```

Fuente: El autor

En la figura 2 se evidencia que el contenido del documento está dentro de etiquetas XML, estas etiquetas son los metadatos de una presentación de PowerPoint, sin embargo, esta estructura es diferente a los metadatos de la figura 1 tanto en el orden de su estructura, como en su ubicación.

Teniendo en cuenta la funcionalidad de la extracción de metadatos para identificar la estructura y contenido de un documento, se ha decidido realizar este modelo respondiendo al formato DOCX de Word y manejando un proceso de ingeniería inversa<sup>23</sup> que permita extraer la información que contienen sus metadatos y representarlos para identificar qué tanto cumplen con los requisitos contemplados en la Norma bajo la cual son escritos.

Por todo lo anterior, la pregunta que da origen a la presente investigación es: ¿Cuáles serían los elementos necesarios para la automatización del proceso de validación de un documento en formato DOCX respecto a los requerimientos de estructura y contenido de la norma NTC 1486?

---

<sup>23</sup> MONROY, M. E.; ARCINIEGAS, J. L.; RODRÍGUEZ, J. C. Caracterización de herramientas de ingeniería inversa. *Información tecnológica*, 2012, vol. 23, no 6, p. 31-42.

### **3. OBJETIVO GENERAL**

Desarrollar un prototipo de software de análisis de metadatos para automatizar la validación de la estructura en documentos de formato DOCX de acuerdo con la norma NTC 1486 aplicando un proceso de desarrollo de acuerdo con los estándares de calidad propuestos por el Carnegie Mellon University.

#### **3.1. OBJETIVOS ESPECÍFICOS**

- Definir los requerimientos funcionales del prototipo a partir de la documentación de acciones y tareas que desarrollará con el fin de generar insumos para el diseño de la aplicación.
- Diseñar los elementos arquitecturales del sistema utilizando el lenguaje unificado de modelado (UML) con el fin de establecer una guía para la construcción del software.
- Desarrollar el prototipo de software con base en la definición de los requerimientos funcionales y el diseño de la arquitectura del sistema para realizar la validación automática de la norma NTC 1486 en la estructura y contenido de archivos Word.
- Validar la funcionalidad del prototipo a través de la realización de pruebas funcionales para verificar que el sistema ejecuta todas sus funcionalidades correctamente.



## 4. MARCO TEÓRICO

Para comprender mejor la naturaleza de esta investigación, será necesario identificar algunos conceptos que dan sentido al proyecto.

### 4.1. NORMA TÉCNICA COLOMBIANA 1486

La Norma Técnica Colombiana 1486 hace parte de las normas Icontec, es decir, el Instituto Colombiano de Normas Técnicas y Certificación, creado en el año de 1963 y que tiene especialidad en actividades de creación de normas para fomentar la producción de conocimiento de carácter investigativo. En consecuencia, la NTC 1486 surge con el objetivo de brindar herramientas de referencia para la presentación escrita de trabajos investigativos en diferentes niveles de educación superior, dando especificaciones a nivel de contenido y forma de este tipo de documentos.<sup>24</sup>

Esta norma da lineamientos para el desarrollo de documentos de carácter investigativos en cuanto al tipo y tamaño de papel, orden de las secciones del documento y aspectos de estructura tales como tamaño de letra, interlineado y espaciado, es decir, la disposición en general del documento. Dentro de la norma es posible evidenciar sugerencias en cuanto a la redacción del documento, lo cual permite tener una idea general de las características de redacción de un trabajo con una connotación científica<sup>25</sup>.

Algunas de las sugerencias propuestas por la NTC 1486 son: Tipo de papel: blanco opaco en tamaño carta, oficio u otro que designe la entidad solicitante del trabajo. Es posible realizar impresión a doble cara desde la página de contenido. Márgenes: Superior: 3 cm. (4 cm. título) Izquierdo: 4 cm. Derecho: 2 cm. Inferior: 3 cm. Número de página a 2 cm. y centrado. En caso de que se realice impresión a doble cara, las márgenes deben ser simétricas a 3 cm<sup>26</sup>.

### 4.2. WEB SEMÁNTICA

La Web Semántica, fomentada por Tim Berners-Lee corresponde a un enfoque en la ingeniería de sistemas, donde se pretende que las búsquedas realizadas en diferentes servidores tengan mayor eficacia, a través de la dotación de aspectos semánticos dentro de los contenidos de páginas web, con el fin de que la inteligencia de dichas páginas para arrojar resultados más certeros ante la exigencia del usuario sea mayor<sup>27</sup>. De acuerdo con esto, es importante reconfigurar las funciones de la web actual, con el fin de aumentar el contenido semántico de las páginas, yendo más allá de la lectura y visualización de datos en la red, logrando la interpretación de la información encontrada para así retornar

---

<sup>24</sup> ICONTEC, N. (2008). NTC 1486. *Elaboración de Informes y trabajos de grado, Sexta edición, Editada por el Icontec.*

<sup>25</sup> JOYA, R., et al. Guía para la elaboración de trabajos escritos. 2013.

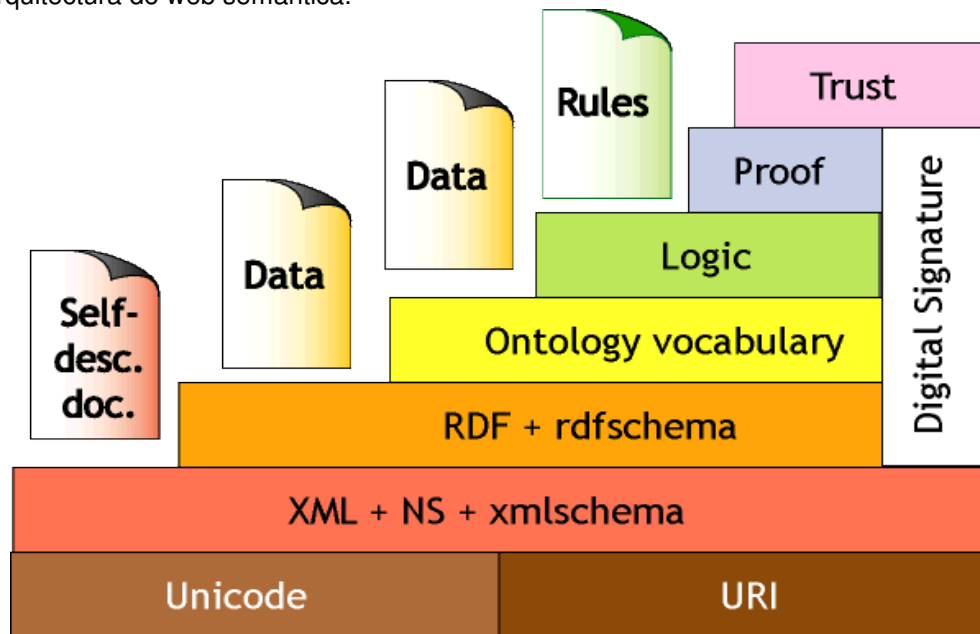
<sup>26</sup> Ibid, p.20.

<sup>27</sup> Tim Berners-Lee, James Hendler, Ora Lassila, The Semantic Web, Scientific American, May 2001

resultados que sean de mayor relevancia para la persona que realiza la búsqueda, sin que esto signifique que se reemplaza la web actual, sino más bien la adaptación de ciertos componentes para lograr así que se convierta en una Web integral.

La arquitectura de la web semántica es representada en la figura 3

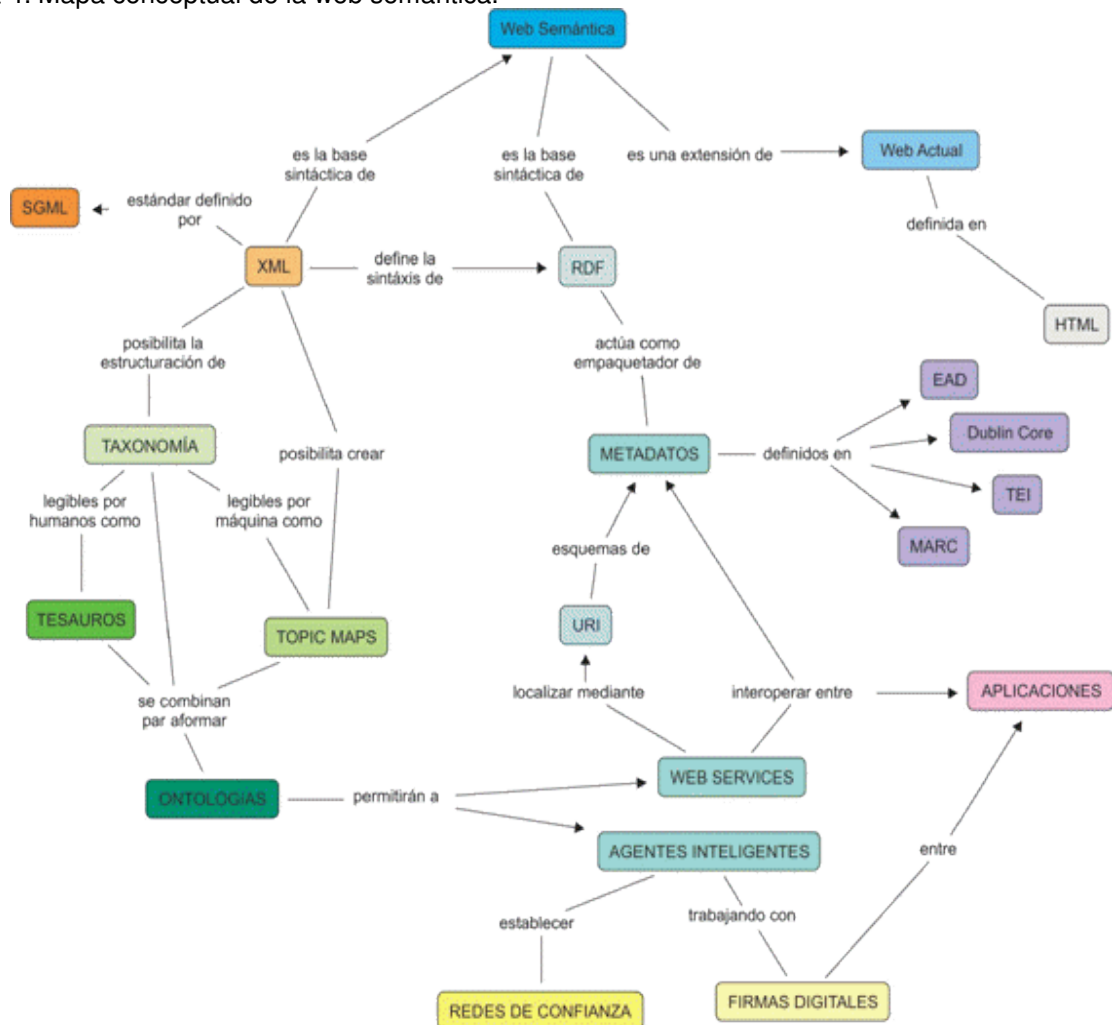
Figura 3: Arquitectura de web semántica.



Fuente: Tim Berners-Lee. *Semantic Web -XML2000. Architecture*  
<http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide11-0.html>

En la figura 3, se puede observar la importancia de los lenguajes semánticos para dar el mejor uso a la Web semántica teniendo en cuenta el uso de metadatos. Se evidencia uso de anotaciones RDF y RDF Schema donde se logran identificar conceptos y realizar asociaciones de una forma jerárquica. Dichos lenguajes requieren una estandarización con el fin de que su uso cumpla con criterios de universalidad, pueda ser reutilizable y compartido en toda la Web. Es importante usar un lenguaje particular que se base en el estilo web, y que exprese por medio de lo racional una representación semántica de las ontologías. De esta forma, el uso de lenguajes como OWL es parte consecutiva en el desarrollo de la Web Semántica, de acuerdo con lo manifestado en la figura 4.

Figura 4: Mapa conceptual de la web semántica.



Fuente: Keilyn Rodríguez Perojo y Rodrigo Ronda León. "Web Semántica: un nuevo enfoque para la organización y recuperación de información en la web". *Acimed*, vol. 13, núm. 6, November-December 2005. [http://bvs.sld.cu/revistas/aci/vol13\\_6\\_05/aci030605.htm](http://bvs.sld.cu/revistas/aci/vol13_6_05/aci030605.htm)

El lenguaje Extensible Markup Language (XML) (lenguaje extensible de marcas) es usado con el fin de intercambiar datos de tipo estructurado, definiendo los formatos aceptados que pueden ser utilizados por un grupo para intercambiar información. El uso de XML en empresas para la transferencia de información empresarial es bastante común<sup>28</sup>. Este formato maneja un lenguaje de etiquetas que representan la información, pero a diferencia del formato HTML el formato no tiene una estructura de etiquetas predefinidas por lo que se debe definir etiquetas propias, en la figura 5 se puede ver el ejemplo de cómo se ve un archivo en formato XML.

<sup>28</sup> FILMAKERPRO, Op.Cit. p.22.

Figura 5: Formato XML.

```
1 <?xml version="1.0" encoding="UTF-8" standalone="yes"?>
2 <businesscards xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
3   <card>
4     <name>Kelvin Kennedy</name>
5     <jobtitle>Manager</jobtitle>
6     <address>Excel to Indesign Street, 44</address>
7     <phone>111-111111</phone>
8     <mail>manager@mail.com</mail>
9   </card>
10  <card>
11    <name>Alfredo Hogan</name>
12    <jobtitle>Assistant</jobtitle>
13    <address>Excel to Indesign Street, 45</address>
14    <phone>222-222222</phone>
15    <mail>assistant@mail.com</mail>
16  </card>
```

Fuente: Pagination.com. Export Excel to XML Tutorial. Disponible en:  
<https://pagination.com/tutorials/export-xml-from-excel-spreadsheet/>

### 4.3. BIG DATA COMO BASE PARA EL ANÁLISIS Y LA EXTRACCIÓN DE DATOS

El proceso de Big Data es de suma importancia para el desarrollo de este proyecto, debido a que se busca manejar información de datos y metadatos para el análisis interno de un documento de Word, concentrándose en la variedad de información que se podría emplear al manejar metadatos teniendo en cuenta los diferentes tipos de archivos a manipular.

Como es necesario realizar un estudio de lo que se está ofreciendo actualmente, es necesario revisar cada uno de los componentes que se encuentra a disposición para la comunidad que permite realizar la tarea de extracción de metadatos de diferentes orígenes de datos que se encuentran contenidos en archivos binarios (para este caso en especial, los datos en bruto para un posterior análisis). Entiéndase componentes aquellos que son herramientas en línea (de uso libre o pagas) e incluso de instalación local que permite realizar el proceso de extracción, limpieza y organización de los datos para brindar al usuario un conjunto de datos y metadatos organizados para su posterior análisis.

El proceso de Big data se establece como un universo analítico donde se encarga de proveer información relevante de cualquier temática a tratar, teniendo en cuenta la veracidad de los datos, junto a su variabilidad y validez, se va a centrar el trabajo en aquella sección donde se encuentran los datos generales con el fin de establecer un eje central para poder realizar la temática de extracción de datos y metadatos.

Para brindar un enfoque específico, se va a tratar todo lo relacionado con datos y metadatos que van a tratar un conjunto cerrado de elementos como texto, imágenes, vídeos, gráficas y demás información que venga relacionada con cada objeto a tratar en el

proceso de extracción. Por lo tanto, se conoce que la forma más común de contener cada uno de estos es mediante archivos binarios<sup>29</sup>.

**4.3.1. ¿Qué son los archivos binarios?** Los archivos binarios hacen referencia a aquellos que requieren de un software especial para ser interpretados, entendiendo la secuencia de bits que se escriben en un archivo, decodificándolo y visualizándolo adecuadamente. Estos archivos son usados por imágenes, sonidos, programas ejecutables y datos comprimidos. El conjunto de bits que representa el archivo binario no puede ser interpretado a simple vista por un editor de texto plano, por lo que es una secuencia establecida para un software específico<sup>30</sup>.

#### **4.4. PROCESO DE ANALISIS DE DATOS**

Para la fase de análisis de datos es necesario poseer un glosario donde se pueda establecer cada una de las palabras clave a usar en el entorno de datos a analizar, esto es demasiado útil para el analítico de datos porque le permite identificar de forma fácil cada uno de los componentes tratados en el proceso de análisis de datos. Esta tarea solamente les corresponde a los encargados de establecer la extracción, limpieza y organización de los datos en bruto, conociendo este glosario en la rama de Big data como diccionario de datos (o Databook por su terminología en inglés).

Para realizar un análisis a cierto conjunto de datos y metadatos es necesario establecer de dónde van a surgir estos datos, o como son llamados normalmente, los datos en bruto. Una vez organizados ya deja de ser un material en bruto y se convierte en un material pre análisis este proceso se ilustra en la figura 6. Dicha organización ha de establecerse antes de dar inicio a la fase de análisis de datos. Estas fases son las siguientes<sup>31</sup>:

1. Requisitos de datos
2. Obtención de datos (Estructurados, no estructurados...)
3. Procesamientos de datos: Agrupación o filtrado.
4. Limpieza de Datos: Eliminar errores o datos duplicados.
5. Análisis exploratorio de datos: tendencias, histogramas, gráficas.
6. Modelado y Algoritmos: Datos Estadísticos (Medias, Modas, etc....)
7. Producto de datos: aplicaciones, documentos, servicios, Periscope, etc.
8. Comunicación/visualización de datos: Informes por audiencia.

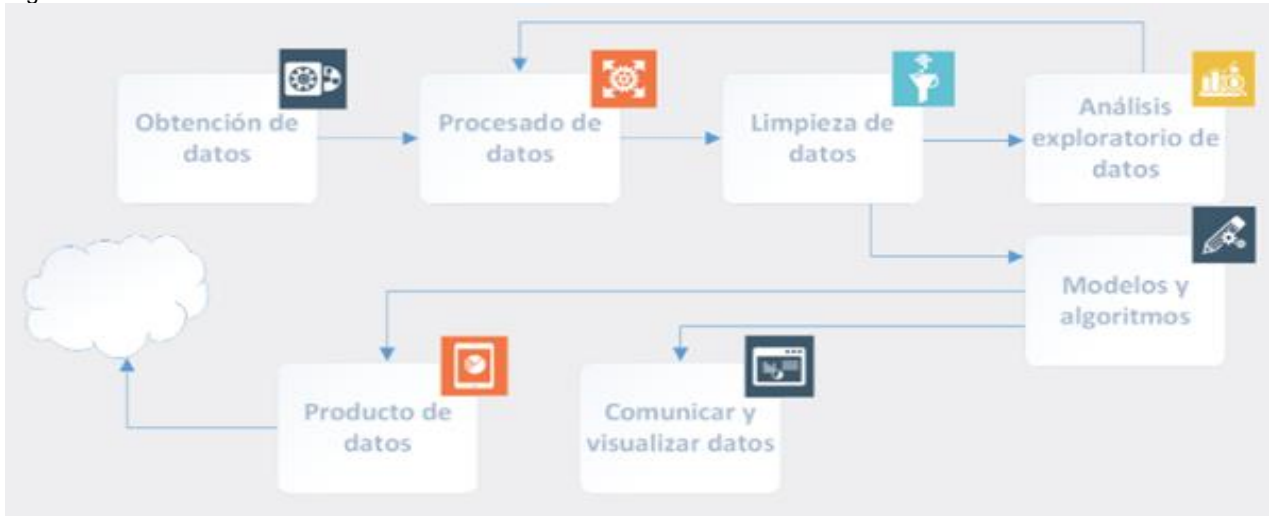
---

<sup>29</sup> POWER DATA, Big Data, Big Data: ¿En qué consiste? Su importancia, desafíos y gobernabilidad; [Consultado el 20 de 02 de 2020]. Disponible en: <https://www.powerdata.es/big-data>

<sup>30</sup> DEFINICION.DE, Definición de binario; [Consultado el 20 de 02 de 2020]. Disponible en: <https://definicion.de/binario/#ixzz2rE2HXYYr>

<sup>31</sup> MOLINA LOPEZ, Jose Manuel y GARCIA HERRERO, Jesus. Técnicas de análisis de datos: Aplicaciones prácticas utilizando Microsoft Excel y Weka.

Figura 6: Fase de análisis de datos.



Fuente: Computer Training. Análisis de Datos en Big Data: tipos y fases del análisis. Disponible en: <https://www.bit.es/knowledge-center/analisis-de-datos-en-big-data/>

#### 4.5. SOFTWARE PARA EXTRACCIÓN DE METADATOS

Existen hoy en día herramientas que permiten separar cada uno de los elementos contenidos en cualquier archivo binario (sea de extensión PDF, mp3, wmv), realizando un reconocimiento de las características que componen un archivo internamente, y que dan forma al archivo en general, que puede ser analizado de forma posterior si se llega a organizar cada uno de estos de forma correcta.

Tal es el caso de 3-Heights <sup>TM</sup> PDF Extract, este es un componente que sirve para leer contenidos junto con las propiedades de documentos PDF. Su función principal es extraer información rápida y eficientemente, sin importar el tamaño o las características que pueda tener el documento, la información y/o los documentos extraídos podrán ser almacenados en una base de datos o podrá ser utilizado para un análisis de datos según las características obtenidas.

PDF Extract también trae otros módulos en donde se clasifica el tipo de archivo al cual se buscará ser extraído, por ejemplo, de los documentos tipo .DOC, extrae características como: fuentes, anotaciones o gráficas, textos planos, token de textos, rangos de páginas, fragmentos de texto, codificación saliente. De los archivos como imágenes, extrae

características como, el tipo de imagen de salida, el rango de la página, y de otros tipos de objetos para extracción brinda características como fuentes, gráficas o anotaciones<sup>32</sup>.

En el caso de las imágenes, existe la herramienta UnGraph que permite extraer datos de archivos binarios tipo imagen, en el caso de estudio de la verificación de Fiabilidad y validez<sup>33</sup>, donde los autores afirman que la mayoría de los estudios se presenta de forma gráfica (diagramas de dispersión y pantallas de series de tiempo, por ejemplo). En ciertas ocasiones, los investigadores que leen los estudios desean acceder a los datos sin procesar de esos gráficos con el fin de realizar análisis adicionales no informados en el grafo original o para realizar un análisis. Cuando no es posible extraer los datos del gráfico y solamente se cuenta con la información que se plasma en la imagen, entonces es donde entra a jugar un papel muy importante la herramienta UnGraph.

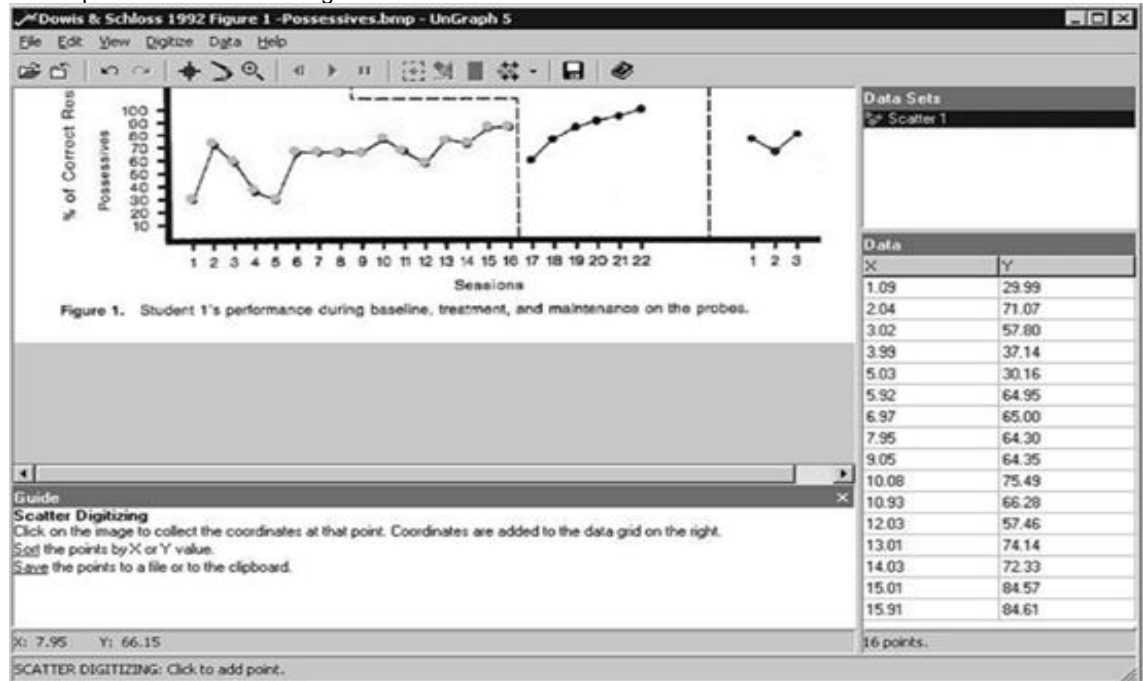
La recuperación de los datos es un proceso de dos pasos: identificar el sistema de coordenadas y extraer los datos. Para definir el sistema de coordenadas, el cursor se utiliza para seleccionar tres puntos a lo largo de los ejes del gráfico (uno que define el eje horizontal, uno en el origen y otro que define el eje vertical). Se ingresan las coordenadas asociadas con cada uno de esos puntos, formando los anclajes desde los cuales UnGraph interpola un sistema de coordenadas completo como se ve en la figura 7. La mayoría de las aplicaciones de UnGraph, incluida la presente, son para gráficos que tienen un sistema de coordenadas lineales.

---

<sup>32</sup> PDF-TOOLS. Automate your data extraction: [Consultado el 20 de noviembre de 2020]. Disponible en: <https://uao.libguides.com/c.php?g=529834&p=3623716#Autor>

<sup>33</sup> Shadish, William & Brasil, Isabel & Illingworth, David & White, Kristen & Galindo, Rodolfo & Nagler, Eden & Rindskopf, David. (2009). Using UnGraph to extract data from image files: Verification of reliability and validity. Behavior research methods. 41. 177-83. 10.3758/BRM.41.1.177.

Figura 7: Graph extrae datos de un grafo.



Fuente: Shadish, William R, Brasil, Using UnGraph to extract data from image files: Verification of reliability and validity, 2009. Disponible en: [https://www.researchgate.net/publication/23964018\\_Using\\_UnGraph\\_to\\_extract\\_data\\_from\\_image\\_files\\_Verification\\_of\\_reliability\\_and\\_validity](https://www.researchgate.net/publication/23964018_Using_UnGraph_to_extract_data_from_image_files_Verification_of_reliability_and_validity)

Así como existen herramientas que permiten extraer diferentes tipos de datos según las rutinas que se hallan dispuesto a la herramienta (sean imágenes, videos, texto, hipervínculos, referencias), también existen rutinas exclusivamente para extraer únicamente las tablas de los archivos PDF que la extracción de las tablas es una tarea que llamó mucho la atención de los investigadores ya que las tablas son uno de los elementos más utilizados para presentar y estructurar los datos y deben extraerse para su reutilización. Si bien los seres humanos pueden reconocer y comprender fácilmente las tablas, las cosas son diferentes para las computadoras, porque las tablas no tienen ninguna característica de identificación en común.

Además, pueden variar en términos de contener filas y / o columnas extensas. Otro punto que dificulta la extracción de tablas es que las tablas pueden contener diferentes tipos de contenido, como texto, figuras, fórmulas matemáticas, etc. <sup>34</sup>

Los autores desarrollaron varias herramientas para soportar el proceso de extracción. Una comparación les mostró que el resultado más útil para su propósito fue el software pdf2html desarrollada por Gueorgui Ovtcharov y Rainer Dorsch. Esta herramienta devuelve todos los elementos de texto (es decir, cadenas) en un archivo PDF con sus coordenadas absolutas en el archivo original. Usando esta herramienta, la tarea de los

<sup>34</sup> Yildiz, Burcu & Kaiser, Katharina & Miksch, Silvia. (2005). pdf2table: A Method to Extract Table Information from PDF Files. 1773-1785.



investigadores fue extraer información de tablas de archivos de texto semiestructurados utilizando sus coordenadas absolutas.

El trabajo al que se enfocaron se basó en los datos devueltos por pdftohtml. Por cada fragmento de texto en el archivo PDF, devuelve un elemento de texto en XML con los siguientes atributos:

- arriba = distancia vertical desde la parte superior de la página
- izquierda = distancia horizontal desde el borde izquierdo de la página
- ancho = ancho del fragmento de texto
- altura = altura del fragmento de texto
- fuente = este atributo describe el tamaño, la familia y el color del fragmento de texto

Los autores restringen su trabajo para utilizar solo estos cinco atributos para extraer información de tablas y no, por ejemplo, componentes gráficos como líneas. Después de aplicar la herramienta, fue necesario extraer información de tablas de un documento XML con elementos de texto que describen la posición absoluta de un trozo de texto en un archivo PDF<sup>35</sup>.

Basándose en lo anterior, es preciso hablar de la suite de Microsoft Office, exactamente Excel (XLS), Word (DOC) y PowerPoint (PPTX). Estos archivos binarios interpretados por cada programa de las herramientas ofimáticas permiten establecer el contenido en forma de bits, por lo que el programa interpreta esta organización a nivel jerárquico, distribuyendo cada elemento en carpetas donde el software con rutinas establecidas se le facilita interpretar la posición de cada elemento en el archivo a tratar.

#### 4.6. MODELO DE DATOS

El modelo de datos es la forma de estructurar y organizar los datos, permitiendo así dar un sentido y un contexto a la información, para esto existen diferentes niveles de modelado, contienen una lista de diversas características o atributos que representan esos datos modelados<sup>36</sup> en la figura 8 podremos detallar las características de cada nivel de modelado:

Tabla 1: Características de los modelos de datos.

Característica	Conceptual	Lógico	Físico
Nombres de entidad	✓	✓	
Relaciones de entidad	✓	✓	
Atributos		✓	
Claves primarias		✓	✓

<sup>35</sup> Ibid, p.28.

<sup>36</sup> LUCIDCHART. Que es un modelo de datos; [Consultado el: 20 de octubre de 2019]. Disponible en: <https://www.lucidchart.com/pages/es/que-es-un-modelo-de-base-de-datos>

Llaves extranjeras		✓	✓
Nombres de tabla			✓
Nombres de columna			✓
Tipos de datos de columna			✓

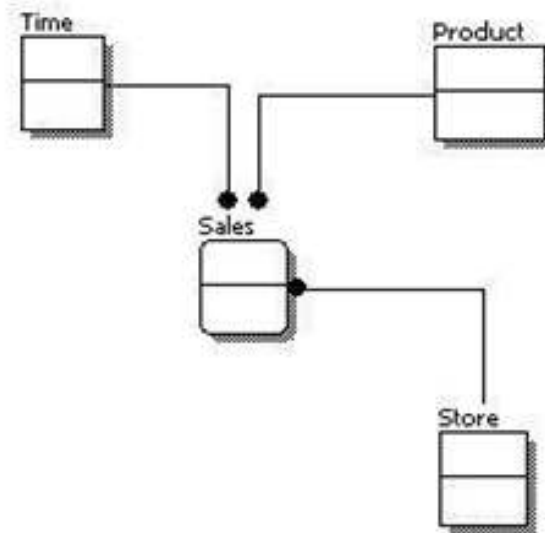
Fuente: 1keydata, Data Modeling 2020, Disponible en: <https://www.1keydata.com/datawarehousing/data-modeling-levels.html>

**4.6.1. Modelo conceptual.** Con el modelo de datos conceptual, se puede identificar las relaciones entre las entidades, las características principales de este modelo son<sup>37</sup>:

- Contiene la relación entre las entidades importantes del modelo.
- No hay definición de atributos.
- No tiene claves principales.

En la figura 9 se puede ver un ejemplo de cómo se ven los modelos de datos conceptuales:

Figura 8: Modelo de datos Conceptual.



Fuente: 1keydata, Data Modeling 2019. Disponible en: <https://www.1keydata.com/datawarehousing/conceptual-data-model.html>

---

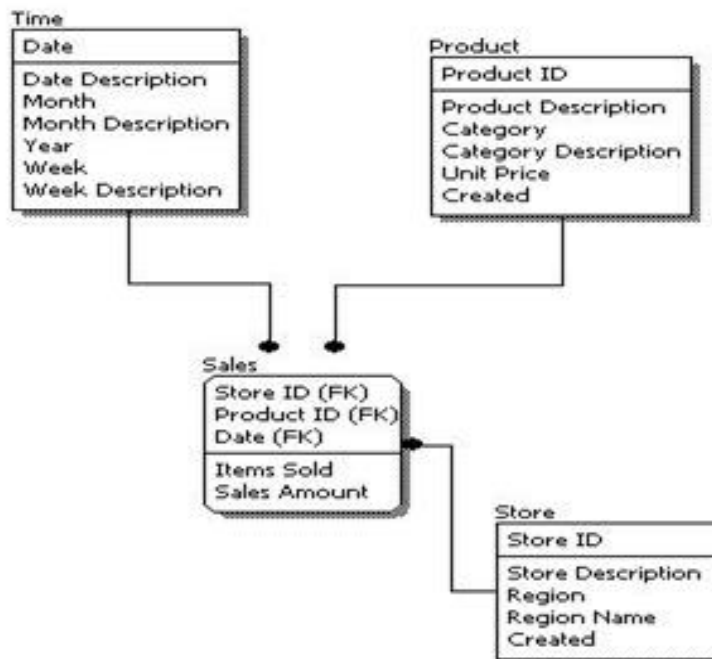
<sup>37</sup> 1KEYDATA. Conceptual Data Model; [Consultado el: 20 de octubre de 2019]. Disponible en: <https://www.1keydata.com/datawarehousing/conceptual-data-model.html>

**4.6.2. Modelo de datos lógico.** Este tipo de modelo de datos describe con el mayor detalle, el modelo de datos lógico tiene ciertas características que permiten identificar este tipo de modelos<sup>38</sup>:

- Se deben especificar los atributos de todas las entidades.
- Se especifican todas las relaciones entre las entidades.
- Se especifican las claves internas que se identifican en las entidades.

El modelo de datos lógico tiene representaciones de entidades, relaciones, subtipos, súper tipos e identificadores, este modelo de datos ayuda a hacer una transición hacia la creación de un modelo de datos físico, en la figura 10 se puede ver un ejemplo de la representación de un modelo de datos lógico:

Figura 9: Modelo de datos lógico.



Fuente: Tecnologías-información. Modelo de datos lógico. 2018 Disponible en:

<https://www.tecnologias-informacion.com/modelos-datos.html>

**4.6.3. Modelo de datos físico.** Este es un modelo específico que representa objetos de datos relacionales en una base de datos como por ejemplo tablas, columnas, claves principales, claves externas y sus relaciones, existen varios aspectos o características a tener en cuenta sobre un modelo de datos físico<sup>39</sup>:

<sup>38</sup> Tecnologías-Información, Modelo-datos; [Consultado el: 20 de octubre de 2019]. Disponible en: <https://www.tecnologias-informacion.com/modelos-datos.html>

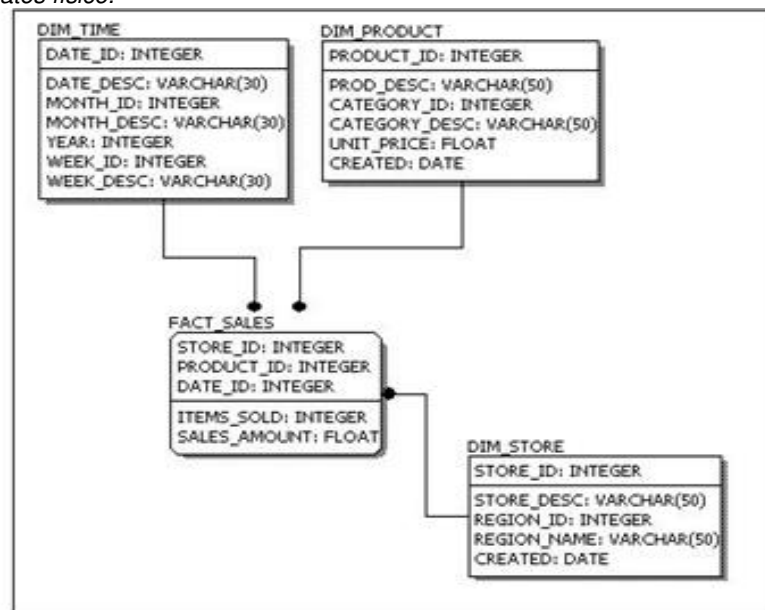
<sup>39</sup> 1KEYDATA. Physical Data Model; [Consultado el: 20 de octubre de 2019]. Disponible en: <https://www.1keydata.com/datawarehousing/physical-data-model.html>

- Se debe hacer especificación de todas las tablas y de todas las columnas.
- Se usan claves foráneas para identificar relaciones entre las tablas.
- El modelo de datos físico suele ser diferente del modelo de datos lógico.

Además, teniendo en cuenta estos aspectos mencionados, también se pueden definir una serie de pasos para la transformación de un modelo de datos lógico a un modelo de datos físico, estos pasos son:

- Las entidades se deben convertir en tablas.
- Las relaciones entre las tablas son claves foráneas.
- Los atributos se convierten en columnas.
- Modificar el modelo de datos físicos según las restricciones que se presenten en su entorno físico.

Figura 10: Modelo de datos físico.



Fuente: 1keydata, Physical Data Model 2020, Disponible en:

<https://www.1keydata.com/datawarehousing/physical-data-model.html>

Como se puede ver en la figura 11, este es el resultado del proceso de la transición del modelo de datos lógico a un modelo de datos físico siguiendo los lineamientos mencionados anteriormente.

Los modelos de datos también manejan estructuras distintas, todo esto depende del contexto en donde estos datos serán de utilidad, las estructuras más comunes son las siguientes:

**4.6.5. Modelo de datos estructurado.** Los modelos de datos estructurados son los que comúnmente se implementan en una base de datos relacional (RDBMS), en las columnas de este modelo de datos por lo general se representan, números de teléfono, códigos postales, direcciones y en general cadenas de texto con una longitud determinada, con este formato se facilita la manipulación de los datos, el lenguaje de consulta estructurado

(SQL) permite realizar consultas sobre este modelo de datos, dando facilidad a la hora de acceder a ellos. Este modelo de datos es generalmente usado para sistemas de reservas en línea, transacciones de ventas, controles de inventario, registros de usuarios etc.<sup>40</sup>

Este modelo también garantiza que siempre existirá una integridad de los datos para asegurar que se almacenarán de forma correcta y sin redundancias<sup>41</sup>.

**4.6.6. Datos semiestructurados.** Los datos semiestructurados son los datos que no tienen una estructura definida de modelo de datos que se asocian con bases de datos o de diferentes formas de tablas de datos, sin embargo, estos tienen información asociada como metadatos, los cuales describen atributos, objetos y sus relaciones. Existen diversos formatos de datos semiestructurados que son de gran utilidad en la actualidad, ya que son altamente usados en el mundo de las ciencias informáticas.

**4.6.7. Datos no estructurados.** Los datos no estructurados son aquellos datos que no están organizados bajo ningún patrón y tampoco están almacenados de forma relacional, estas características están dadas ya que no son un tipo de dato definido, a pesar de esto los datos deben poder ser organizados o clasificados de alguna manera. Este tipo de datos son de uso cotidiano, en situaciones específicas como el uso de correos electrónicos, envío de archivos de texto o mensajes desde diferentes dispositivos, así como la participación en las diversas redes sociales.<sup>42</sup>

Administrar este tipo de información no estructurada, no es una tarea simple sin embargo lo que se pretende con esta información, es poder acceder a los metadatos, los cuales describen características de los datos que componen un archivo o elemento informático con el fin de dar características específicas del mismo, como por ejemplo, la identificación de textos de acuerdo a información relacionada con el autor, el título o la editorial<sup>43</sup>

Las bases de datos no relacionales funcionan similares a los otros tipos de bases de datos mencionados en la sección 4.6 del documento, para ello existen diversas herramientas las cuales permiten manipular este tipo de datos a voluntad como por ejemplo las bases de datos NOSQL, estas bases de datos surgen a razón de los avances de la internet como por ejemplo con las redes sociales, ya que se facilitó considerablemente el acceso a cualquier usuario a la red ya sea para cargar o descargar información, pero a causa de esto el volumen de información masiva que se debió empezar a manejar generó dificultades para administrar o manipular estas grandes cantidades de información, se evidenció que

---

<sup>40</sup> Analítica Negocios. Diferencias entre datos estructurados y no estructurados. Disponible en: <https://www.analiticanegocios.com/guias/diferencias-datos-estructurados-no-estructurados/>

<sup>41</sup> MANUAL-TECNICO-BD-ORACLE, Bases de datos. Disponible en: <https://manual-tecnico-bd-oracle.readthedocs.io/es/latest/Modelo%20de%20datos.html#:~:text=En%20el%20caso%20del%20modelo%2Cregla%20de%20integridad%20del%20modelo>.

<sup>42</sup> Ibid. p.20.

<sup>43</sup> VARA, Juan M.; VELA, Belén; MARCOS, Esperanza. Diseño de bases de datos objeto-relacionales con UML. *Diseño de bases de datos objeto-relacionales con UML*, 2005, p. 0-0.

manejar la información con bases de datos relacionales ya no era una forma óptima de representar la información para los avances que ocurren en internet. Las bases de datos NOSQL fueron la solución para la estructuración que se le estaba dando a la información hasta el momento.

Las bases de datos NOSQL tienen ciertas ventajas sobre las bases de datos relacionales, como, por ejemplo:

- Se puede ejecutar en máquinas con pocos recursos.
- Permite manejar grandes cantidades de datos.
- No genera cuellos de botella

El problema con las bases de datos SQL es que se debe transcribir cada una de las sentencias para ejecutarse, así mismo, los procesos de máquina tienden a aumentar de complejidad y además esto demanda procesos de máquina cada vez más complejos por lo que si se analiza en un contexto mucho más grande, esto puede ralentizar el sistema considerablemente.

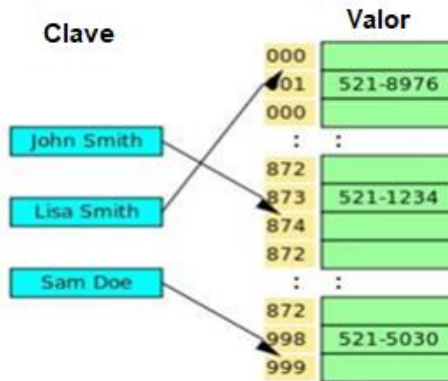
Existen algunas variaciones en la forma en que se puede almacenar la información en este tipo de base de datos NOSQL, algunas de las formas más utilizadas son:

**4.6.8. Base de datos Clave Valor.** Es uno de los modelos de bases de datos NOSQL más populares, este se basa en que cada elemento está identificado por una clave única, esto con el fin de poder recuperar información de forma óptima y eficaz, en la figura 14 se observa la representación gráfica de una base de datos clave valor, regularmente los datos se almacenan en un objeto binario y las herramientas que más se utilizan para su implementación son: Cassandra, BigTable o HBase<sup>44</sup>.

Figura 11: Base de datos Cadena Valor.

---

<sup>44</sup> CASTILLO, Jessica Nataly, et al. Base de Datos NoSQL: MongoDB vs. Cassandra en operaciones CRUD (Create, Read, Update, Delete). *Revista Publicando*, 2017, vol. 4, no 11 (1), p. 79-107.



Fuente: acenswhitepapers. *Bases de datos NoSQL. Qué son y tipos que nos podemos encontrar. 7.* Disponible en: <https://www.acens.com/wp-content/images/2014/02/bbdd-nosql-wp-acens.pdf>

**4.6.9. Base de datos Documentales.** Este tipo de base de datos generalmente se utiliza para almacenar información de tipo documento, por lo general esta información se representa en archivos de formato XML o de formato JSON como en la figura 7, en los cuales se utiliza una clave única para realizar consultas avanzadas sobre el contenido de un documento, otra característica importante es la versatilidad ya que este tipo de bases de datos también funciona sobre bases de datos relacionales, algunas herramientas como MongoDB o CouchDB son las más populares para este tipo de bases de datos<sup>45</sup>

Figura 12: Modelo de datos conceptual.

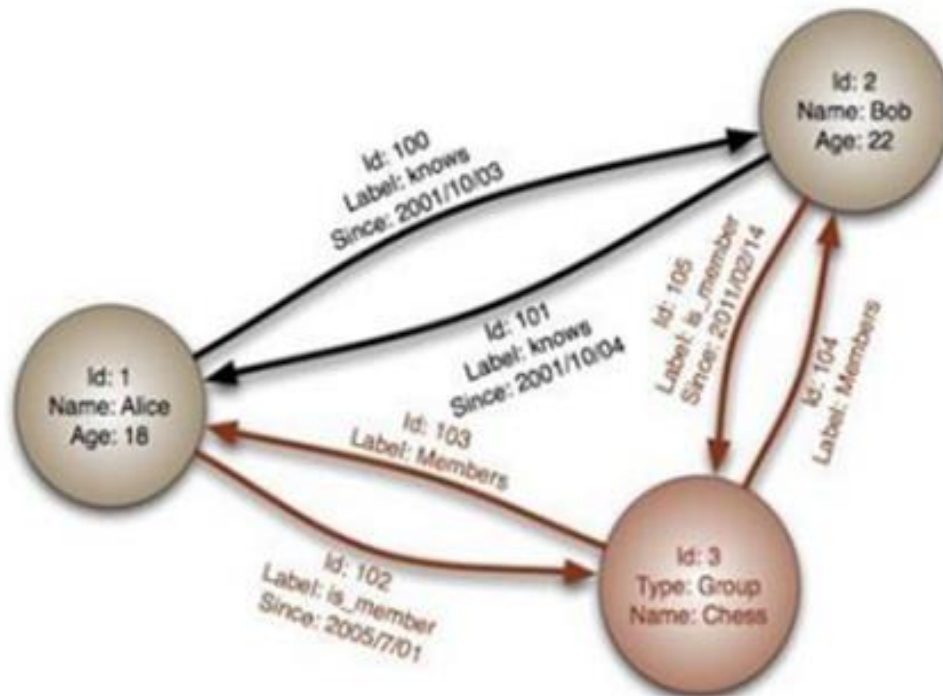


<sup>45</sup> DEL BUSTO, Hansel Gracia; ENRÍQUEZ, Osmel Yanes. Bases de datos NoSQL. *Revista Telemática*, 2013, vol. 11, no 3, p. 21-33.

Fuente: acenswhitepapers. *Bases de datos NoSQL. Qué son y tipos que nos podemos encontrar.* 7. Disponible en: <https://www.acens.com/wp-content/images/2014/02/bbdd-nosql-wp-acens.pdf>

**4.6.10. Bases de datos en grafo.** En este tipo de bases de datos la información es representada en nodos de un grafo y sus relaciones como en la figura 7 se representan con aristas, esto quiere decir que para recorrer su información se utiliza la teoría de grafos ya que esto permite aprovechar el máximo rendimiento que puede ofrecer. Existen ciertas reglas que esta estructura debe tener como, por ejemplo; la base de datos debe estar normalizada de tal manera que cada una de las tablas tengan una sola columna y cada relación tenga dos. Este tipo de bases de datos suelen ser más eficientes que una base de datos relacional<sup>46</sup>.

Figura 13: Base de datos en grafo.



Fuente: acenswhitepapers. *Bases de datos NoSQL. Qué son y tipos que nos podemos encontrar.* 7. Disponible en: <https://www.acens.com/wp-content/images/2014/02/bbdd-nosql-wp-acens.pdf>

## 4.7. ATRIBUTOS DE CALIDAD

---

<sup>46</sup> HERRERA, Harol Andrey; VALENZUELA, Carlos Rueda. NoSQL, la nueva tendencia en el manejo de datos. *Tecnología Investigación Y Academia*, 2016, vol. 4, no 1, p. 147-150.



Los atributos de calidad son los que describen las características deseables que se deben llevar a cabo en el sistema, ya que estos son factores determinantes en la calidad del resultado final del desarrollo, es por esta razón que para la construcción del software ValidateDocx se han considerado como relevantes los siguientes atributos<sup>47</sup>:

- **Usabilidad:** La usabilidad es la medida de calidad que busca brindarle la mejor experiencia de interacción al usuario con el sistema de tal modo que su uso sea lo más sencillo y eficaz posible.
- **Seguridad:** El sistema debe garantizar las siguientes medidas de seguridad:
  - Cada vez que se ingresa al sistema se debe comprobar la identidad del usuario a través de la verificación de sus credenciales de inicio de sesión.
  - Se debe garantizar que solo las personas que están autorizadas pueden ver ciertas porciones del sistema dependiendo del nivel de permisos de usuario.
  - Se debe garantizar que cualquier tipo de información solo puede ser modificada por los usuarios que estén específicamente autorizados para hacerlo.
- **Modificabilidad:** El sistema debe tener la capacidad de admitir cualquier cambio que sea necesario a nivel de requerimientos ya sea con el fin de modificar lo que ya se desarrolló o de añadir nuevas funcionalidades futuras que complementarán su funcionamiento actual.
- **Correctitud:** El sistema no debe retornar ningún error a nivel de funcionamiento una vez sea desplegado para su uso abierto al público.
- **Consistencia:** El sistema debe proporcionar coherencia entre todas las modificaciones y operaciones que el usuario realiza.
- **Completitud:** El sistema debe tener la capacidad de realizar y ejecutar todas las funciones que el usuario pueda llegar a requerir.

---

<sup>47</sup> UQBAR-FOUNDATION. Articles, Atributos de calidad; [Consultado el 15 de marzo de 2020] Disponible en: <http://wiki.uqbar.org/wiki/articles/atributos-de-calidad.html>

## 5. MARCO CONCEPTUAL

A continuación, se relacionan algunos conceptos clave para la comprensión de la investigación:

### 5.1. INGENIERÍA DE SOFTWARE

La ingeniería de software es una rama de la ingeniería de sistemas que se especializa en el uso de instrumentos y tecnologías para el desarrollo de programas informáticos. Esta rama está relacionada con los procesos de análisis de acuerdo con las problemáticas que se presenten en la sociedad y puedan ser solucionadas a partir del desarrollo de programas o aplicaciones que faciliten actividades dentro de un contexto determinado. Es importante mencionar que la ingeniería de software responde a una serie de pasos que se encuentra estrechamente relacionados con el ciclo de vida del software que consiste en siete fases<sup>48</sup>:

1. Fase de planificación. La fase de planificación se encuentra relacionada con el diseño y las características que sea necesario tener en cuenta para la implementación de un proyecto, logrando así establecer los pasos a seguir para realizar un desarrollo que tome en consideración todas las posibles eventualidades y permita el cumplimiento total del objetivo con que se realiza una aplicación.
2. Fase de análisis. En esta fase se evidencia un análisis de los requerimientos funcionales y no funcionales que debe contener la aplicación o programa a desarrollar con el fin de que se definan elementos para iniciar con la programación de este.
3. Fase de diseño. Durante la etapa de diseño, se realiza el planteamiento de posibles medidas a tomar en caso de que el programa no realice de manera adecuada su función o sea necesario realizar modificaciones, logrando así que el programa a desarrollar cuente con todas las garantías para la empresa.
4. Fase de desarrollo. Es el inicio de la vida del software como tal, donde se realiza la producción a nivel de software logrando que los desarrolladores realicen la programación de la aplicación o sistema que se busca desarrollar.
5. Fase de integración y periodo de pruebas. Durante esta fase es importante tener en cuenta las pruebas para validar el funcionamiento del software respecto a las expectativas que la empresa tenga con el sistema desarrollado. De ser necesario, se realizan modificaciones para optimizar el resultado.

---

<sup>48</sup> RAMOS, Daniel, et al. *Curso de Ingeniería de Software: 2ª Edición*. IT Campus Academy, 2017

6. Fase de implementación. Durante esta fase, se realiza la implementación del programa, desarrollo o sistema que se haya realizado, logrando que la empresa y los clientes de esta puedan ejecutarla de forma adecuada y completa.
7. Fase de mantenimiento. En esta etapa se evidencia la necesidad de realizar mantenimiento y actualizar la aplicación o desarrollo de acuerdo con las necesidades identificadas por la empresa.

**5.1.1. Ingeniería inversa.** “El proceso de analizar el código, documentación y comportamiento de un sistema para identificar sus componentes actuales y sus dependencias para extraer y crear una abstracción del sistema e información del diseño. El sistema en estudio no es alterado, sino que se produce conocimiento adicional acerca del sistema”<sup>49</sup>

Este proceso principalmente se utiliza para la recuperación del diseño de una aplicación a partir de su código, dependiendo del tipo de aplicación, esto se puede lograr mediante herramientas que extraen información. La ingeniería inversa se aplica principalmente a software con las siguientes características o necesidades:

- Software con documentación inexistente u obsoleta.
- La programación está en bloques de códigos muy grandes o sin estructurar.
- La aplicación puede estar sujeta a cambios constantes o frecuentes y por lo tanto se afecta su parte de diseño.
- Reducir la complejidad del sistema.
- Facilitar la reutilización.

El nivel de abstracción en el proceso de ingeniería inversa juega un rol bastante importante, entre más alto sea el nivel de abstracción, mayor será la cantidad de características que se podrán extraer y comprender del sistema del nivel de abstracción se puede derivar:

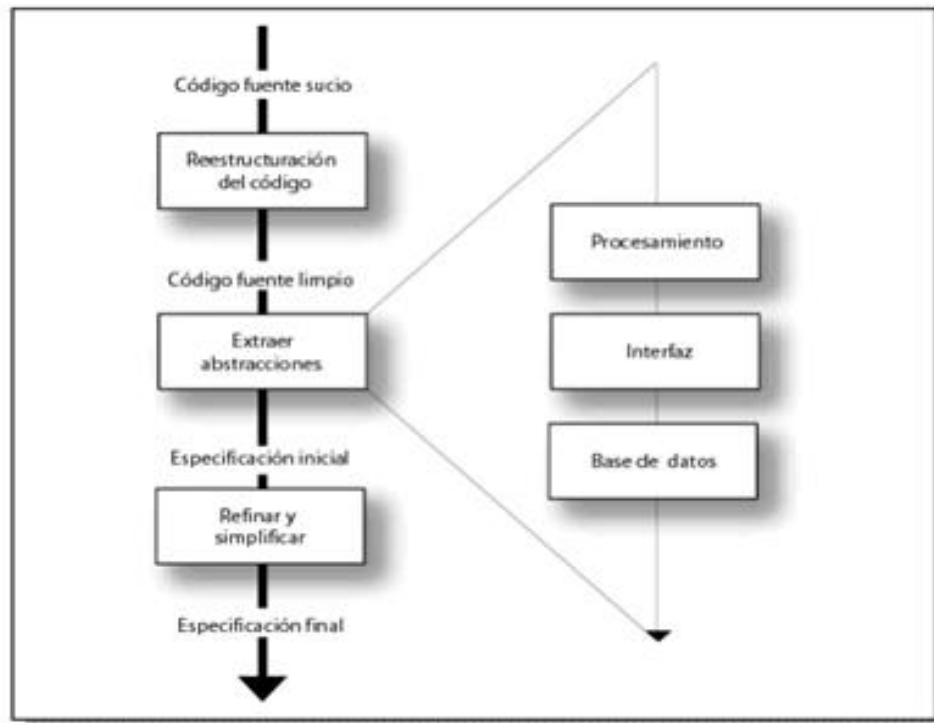
- Representaciones del diseño del procedimiento, esto se da si el nivel de abstracción es bajo.
- La información de las estructuras de datos y de programas, esto se da con un nivel de abstracción un poco más elevado.
- Modelos de flujos de datos, esto se da con un nivel de abstracción alto.
- Modelo de entidades y relaciones, esto se da con un nivel de abstracción elevado.

Antes de comenzar con el proceso de la ingeniería inversa es apropiado limpiar la mayor cantidad de código basura (código sobrante), esto con el fin de dejar un código más limpio y entendible como se muestra en la figura 17 ya que esto es la base para las actividades subsiguientes en este proceso.

---

<sup>49</sup> MONROY, Martín E.; ARCINIEGAS, José L.; RODRÍGUEZ, Julio C. Caracterización de herramientas de ingeniería inversa. *Información tecnológica*, 2012, vol. 23, no 6, p. 31-42.

Figura 14: Proceso de Ingeniería.



Fuente: catarina.udlap.mx, Ingeniería Inversa. 2018. Disponible en: [http://catarina.udlap.mx/u\\_dl\\_a/tales/documentos/lis/lopez\\_a\\_aa/capitulo4.pdf](http://catarina.udlap.mx/u_dl_a/tales/documentos/lis/lopez_a_aa/capitulo4.pdf)

En caso de que el proceso de la ingeniería inversa sea para un programa o sistema que no tiene documentación, el encargado de hacer este proceso debe especificar cuáles son las partes significativas en cuanto al procesamiento o estructuración de la información en el código original del sistema. En el caso de la investigación descrita en este documento, el proceso de abstracción se realiza para documentar y comprender el proceso y estructuración de los metadatos de un archivo de formato PPTX.

El primer paso que se debe realizar en esta práctica de la ingeniería inversa es comprender y extraer fragmentos del código fuente, para hacer una correcta comprensión del código, en este proceso es importante analizarlo en distintos niveles de abstracción: sistema, programa, componente, sentencia y configuración<sup>50</sup>.

Un aspecto importante para mencionar es la reestructuración del código el cual no modifica ningún diseño de la arquitectura del software, se pueden obtener los siguientes beneficios:

- Se aumenta la calidad del software y se mejora su documentación.
- Se reduce el esfuerzo de mantenimiento.
- Se hace más sencillo comprobar y depurar el software.

---

<sup>50</sup> AMENEYRO, López; GISWEB, A. A. *Reingeniería para la implementación de un Web FeatureService*. 2004. Tesis Doctoral. Tesis de Licenciatura en Sistemas Computacionales. Departamento de Ingeniería en Sistemas Computacionales, Escuela de Ingeniería, Universidad de las Américas, Puebla.

En muchos casos, el proceso de ingeniería inversa no sólo busca la reconstrucción del código de un programa, sino que también se hace con el fin de crear una réplica de este el cual aumenta el conocimiento sobre dicho software, si en la restauración llega a un nivel de especificación tan detallado a tal punto de que permite transformar el código fuente a un pseudocódigo, se considera que el nivel de abstracción realizado es de alto nivel.

## **5.2. DESARROLLO WEB**

Desarrollo web está relacionado con la construcción y mantenimiento de diferentes sitios; que favorece a la aplicación web en términos de apariencia, desempeño y funcionamiento, brindando la mejor experiencia al usuario<sup>51</sup>.

En términos prácticos, un desarrollador web, se encarga de realizar proyectos de desarrollo de aplicaciones web, tomando como referente vital el funcionamiento del software. Por otro lado, la persona que se encarga de la visualización (layout) de la página es el diseñador web, de igual manera, el proceso del webmaster es integrar las partes del desarrollo con la visualización(layout) de la información. En ocasiones el webmaster también se encarga de actualizar los contenidos de la página. Algunos de los lenguajes para programación que más se utilizan en desarrollo web son: ASP.NET, PHP y JSP, aunque aún hay quienes usan ASP, Macromedia ColdFusion y Perl<sup>52</sup>.

## **5.3. ARQUITECTURA MVC – MODELO VISTA CONTROLADOR**

Es un tipo de arquitectura en software que segmenta los datos en una aplicación, la interfaz de usuario, y la lógica de control en tres componentes como lo son el Modelo, la Vista y el Controlador. Es un modelo que se ha mantenido en el tiempo como una opción para el desarrollo de aplicaciones de diferentes tipos, con una gran variedad de lenguajes de desarrollo<sup>53</sup>.

El modelo es aquel que facilita el acceso a la capa de almacenamiento de datos. Esta parte de la arquitectura suele ser independiente al sistema de almacenamiento de la aplicación que se desarrolla. Permite definir la funcionalidad del sistema y la estructura de los datos que el software emplea. Así mismo, el modelo registra las vistas y controladores de la aplicación, favoreciendo la interacción con el controlador.

---

<sup>51</sup> OPENCLASSROOMS, Que es el desarrollo web?:[Consultado el 15 de 04 de 2020]. Disponible en: <https://blog.openclassrooms.com/es/2017/09/11/que-es-el-desarrollo-web/>

<sup>52</sup> Ibid.p.20.

<sup>53</sup> UNIVERSIDAD DE ALICANTE, Servicio de informatica ASP.NET MVC 3 Framework:[Consultado el: [20 de 03 de 2020]. Disponible en: <https://si.ua.es/es/documentacion/asp-net-mvc-3/1-dia/modelo-vista-controlador-mvc.html>

Por otro lado, el controlador establece una comunicación entre el modelo y la vista, determinando reglas para la gestión de diferentes eventos y así mismo la respuesta de acciones de acuerdo con lo que se le solicite desde el modelo.

Finalmente, la vista es el resultado que el usuario puede ver del sistema, en esta parte, se reciben los datos del modelo y se muestran al usuario<sup>54</sup>.

---

<sup>54</sup> Ibid.p.46.

## 6. ESTADO DEL ARTE

Dentro de las investigaciones que se han realizado en torno al tema de investigación que se está tratando en este proyecto, se pueden encontrar diferentes referentes de trabajo, desde una perspectiva de la extracción de metadatos de archivos para el análisis del contenido de los mismos, por ejemplo, el documento “Extracción automática de metadatos en un documento digital”, tesis del maestro Orlando Julián Piña, relata cómo se desarrolla un proceso automático de extracción de metadatos a partir de 4 heurísticas diferentes (Fecha de creación, Autor, Palabras clave del tema y Título del documento), fueron codificadas en Perl. Para las pruebas con los programas se utilizó un grupo de 45 documentos en español y 45 documentos en inglés; todos los documentos son de tipo investigación (tesis, artículos científicos, notas de congresos, etc.). La idea de utilizar documentos de diferentes fuentes es con la finalidad de utilizar una muestra variada de documentos para obtener una mejor confiabilidad en los resultados. Para observar si las heurísticas aplicadas a documentos en español y a documentos en inglés presentan resultados similares, se realizó una comparación de resultados para poder observar la efectividad y el comportamiento de los resultados para cada uno de los casos. Los resultados que se obtuvieron en la ejecución de los programas experimentales sobre los documentos en idioma español y sobre los documentos en idioma inglés son muy similares, lo que permite concluir que las búsquedas realizadas presentan resultados similares en los dos idiomas, teniendo cambios en el diccionario de contenido y no de forma<sup>55</sup>.

Otro de los textos que han sido un referente importante es el documento “Estudio de seguridad informática de los metadatos contenidos en archivos publicados en las web de las organizaciones”, con autoría de Jesús Antonio Durán Acevedo, desarrollado por la UNAD en el programa de seguridad informática, en donde se describe la metodología usada para un proyecto de seguridad informática basado en metadatos, mediante el uso de software que permita la extracción de los metadatos presentes en los objetos en medio magnético publicados por las empresas que fueron estudiadas, utilizando diferentes programas de software, con un total de 1797 documentos públicos, pertenecientes a 5 entidades diferentes y que respondían a diversos tipos. Dentro de las conclusiones de este trabajo, se encuentran que es mucha la información oculta que se obtuvo en los documentos analizados y que pueden ocasionar graves problemas de seguridad en la información de las entidades estudiadas<sup>56</sup>

En el documento “caracterización de la información mediante la extracción de metadatos utilizando recuperación de información sobre convocatorias” de los autores Andrey Mauricio Fernandez Franco Y Mauricio Muñoz Guzmán, de la universidad de San

---

<sup>55</sup> PIÑA GONZALEZ, Orlando Julián. Extracción Automática de Metadatos de un Documental Digital. Instituto tecnológico y de estudios superiores de monterrey. 2000

<sup>56</sup> DURÁN ACEVEDO, Jesús Antonio, et al. Estudio de Seguridad Informática de los metadatos contenidos en archivos publicados en los sitios web de las organizaciones: Alcaldía de Pamplona, Cámara de Comercio de Pamplona, Gobernación de Norte de Santander, Diario la Opinión y La Dian.

Buenaventura de Cali, en el programa de ingeniería de sistemas, se evidencia una arquitectura de software del sistema definida por el estilo de arquitectura de capas estratificado, donde el sistema se organiza a partir de diferentes capas que aumenta de complejidad conforme a los servicios que prestan para otras capas, logrando así que se realice un proceso de ejecución adecuada del aplicativo que se desarrolló. Este proceso es llamado arquitectura genérica, que se manifiesta al usar el framework Zathuracode como insumo para la programación del código base del sistema en el lenguaje Java Enterprise Edition. Es necesario señalar que este proyecto se basa en la creación de un sistema para la caracterización de las diferentes convocatorias y becas a partir de búsquedas de información basado en la extracción de metadatos con la herramienta AlchemyAPI<sup>57</sup>.

En el documento titulado “Extracción Automática de Metadatos de Objetos de Aprendizaje: un estudio comparativo” del autor Taihú Piré, se puede apreciar la metodología utilizada para la extracción de metadatos en diferentes objetos de aprendizaje de tipo modular, para el reconocimiento y análisis de la información contenida en dichos módulos de aprendizaje, facilitando el retorno de búsquedas de acuerdo con las necesidades educativas particulares de cada individuo. Uno de los resultados importantes de esta investigación se relaciona con la carencia de aspectos de vital importancia para realizar la caracterización de los objetos de aprendizaje, ya que se queda corto al realizar un análisis de ellos en cuanto a los metadatos<sup>58</sup>.

Otra investigación “Ponderación de metadatos de recursos educativos como forma de mejorar los resultados de una búsqueda” de autoría de Claudia Deco, muestra cómo los metadatos se han convertido en la actualidad en una herramienta importante para la búsqueda de información en un contexto educativo. En este trabajo se manifiesta la relevancia que tienen los metadatos en los objetos de aprendizaje como una herramienta para personalizar las recomendaciones que son dadas por un buscador. Es posible concluir que el análisis realizado en la extracción de metadatos deja a la luz una falta de información y estructura interna de los objetos de aprendizaje, lo cual influye en el proceso de descripción de dichos elementos<sup>59</sup>.

En el documento “Componente para la extracción automática de metadatos bibliográficos desde corpus textuales en formato PDF”, se evidencia que los metadatos son una herramienta para la gestión documental en bibliotecas virtuales, teniendo en cuenta los procesos de selección, tratamiento y explotación de recursos. En esta investigación se

---

<sup>57</sup> FERNANDEZ FRANCO, Andrey Mauricio, et al. Caracterización de la información mediante la extracción de metadatos utilizando recuperación de información sobre convocatorias. 2016.

<sup>58</sup> PIRE, Taihú, et al. Extracción automática de metadatos de objetos de aprendizaje: un estudio comparativo. En *VI Congreso de Tecnología en Educación y Educación en Tecnología*. 2011.

<sup>59</sup> DECO, Claudia; BENDER, Cristina; SAER, Jorge. Ponderación de metadatos de recursos educativos como forma de mejorar los resultados de una búsqueda. 2011.



muestra el desarrollo de una aplicación Web para la extracción de metadatos en archivos de tipo bibliográfico, lo cual se encuentra determinado a partir de ciertos procesos que dan forma a la aplicación, siguiendo una representación tubular de flujo de distintos datos. Se realiza un estudio de caso para corroborar que la extracción de metadatos sí presenta una reducción de tiempo frente al análisis de diferentes documentos<sup>60</sup>.

La investigación “Herramienta para el procesamiento en lote de documentos digitales para el sistema REPXOS 3.0.” llevada a cabo por José Javier Hernández Benítez, relata el diseño y aplicación de un sistema para la implantación de repositorios digitales en instituciones científicas y académicas, desarrollado en el centro de Informatización de la Gestión Documental (CIGED), teniendo como resultado la consolidación de una herramienta adecuada para la extracción de metadatos de diversos elementos digitales, para así permitir que dichos documentos fueran validados y compartidos a través de la plataforma REPXOS<sup>61</sup>

De igual manera, en la tesis “Servicio de recolección de metadatos genérico para documentos” de Julieta Paz Rodríguez, se detalla la implementación de una herramienta para la clasificación y recolección de metadatos para documentos, sin embargo, dicha herramienta es manual, lo cual demuestra que el proceso toma tiempo y requiere de la supervisión de una persona, quien puede cometer errores en el proceso<sup>62</sup>

---

<sup>60</sup> FLORES RIERA, Leduan, et al. Componente para la extracción automática de metadatos bibliográficos desde corpus textuales en formato PDF. *Revista Cubana de Ciencias Informáticas*, 2017, vol. 11, no 4, p. 85-98.

<sup>61</sup> BENÍTEZ, José Javier Hernández, et al. Herramienta para el procesamiento en lote de documentos digitales para el sistema REPXOS 3.0. *Revista Publicando*, 2018, vol. 5, no 14 (3), p. 571-582.

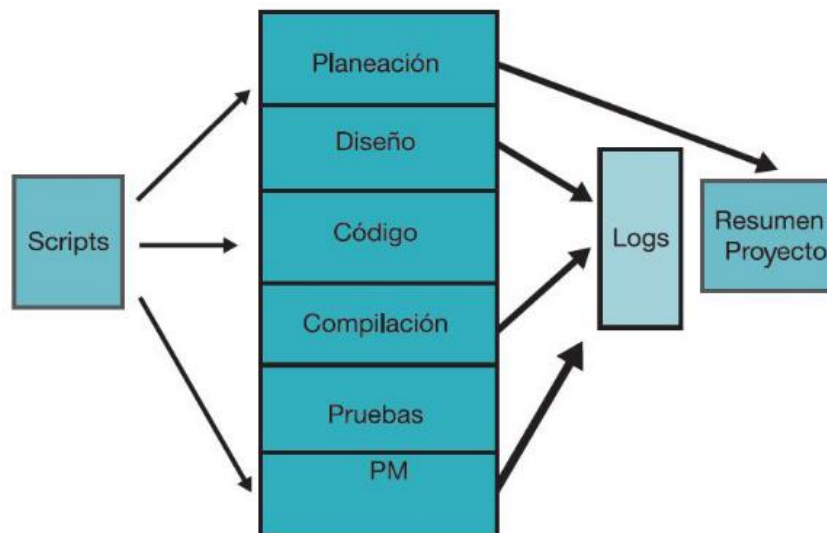
<sup>62</sup> JUAN, Rodríguez; PAZ, Julieta. *Servicio de recolección de metadatos genérico para documentos*. 2017. Tesis Doctoral. Universidad Nacional de La Plata.

## 7. METODOLOGIA

Para el desarrollo de esta investigación, se ha propuesto realizar una metodología Personal Software Process o por sus siglas en inglés PSP, que tiene como objetivo mejorar la productividad de los programadores o ingenieros de software a través de un seguimiento del desempeño que estos tengan frente a una planeación del desarrollo de sus tareas. El proceso que se realiza desde esta metodología permite que el desarrollador tenga una planeación y análisis de los requerimientos necesarios para la consolidación de un software.

De esta manera, en la figura 16 se ilustra que la metodología PSP tiene diferentes niveles, a saber<sup>63</sup>: Planeación, este nivel es donde se realiza la proyección, desarrollo y post mortem del proyecto. En dicho nivel se tienen en cuenta criterios relacionados con la disciplina del ingeniero encargado de desarrollar el software, permitiendo la estructuración elementos correspondientes a los requerimientos funcionales y no funcionales, así como el alcance que tendría el software. El siguiente nivel es código, donde el ingeniero realiza la construcción del sistema según los lineamientos definidos en el nivel de planeación y diseño. El siguiente nivel es compilación, en donde el ingeniero realiza el despliegue del desarrollo en el ambiente de producción lo cual el software disponible para la realización de las pruebas. El siguiente nivel pruebas busca realizar proceso de revisión y edición del software para garantizar que no haya errores frente al desarrollo y la programación de este.

Figura 15: Metodología PSP.



Fuente: sg.com.mx, Personal Software Process(PSP). Disponible en: <https://sg.com.mx/content/view/797>

---

<sup>63</sup> WATTS S. HUMPHREY, Introducción al PSP(Personal Software Process) 1997

La metodología PSP promueve un estilo de programación independiente al lenguaje, que permite a los ingenieros llevar un proceso de desarrollo analítico que favorece el diseño de software con un alcance y propósito definidos.<sup>64</sup> PSP facilita el diseño y desarrollo de productos de calidad, que cumplan con las expectativas de los ingenieros y clientes. Teniendo en cuenta esto, es posible plantear una serie de etapas a seguir dentro de la investigación actual, tomando en consideración los lineamientos propuestos por la metodología PSP:

**Etapas 1: Documentación de requerimientos funcionales y características del prototipo en general.** En esta etapa, se definen los requerimientos funcionales, a partir del planteamiento y especificación de tareas que se esperan sean realizadas por el prototipo. Es importante tener en cuenta que esta especificación se basa en diagramas, los cuales visualizan las diferentes acciones que puede desarrollar el sistema. Esta etapa es muy trascendental para el proyecto, a que en ella se resuelven aspectos que darán paso al diseño arquitectural del software, logrando que el desarrollo de este tenga precisión y responda por completo a los objetivos planteados durante dicha especificación.

**Etapas 2: Diseño arquitectural del sistema a partir del UML.** En este momento de la investigación, se realiza el diseño del sistema, lo cual permite tener claridad respecto al alcance del sistema y brinda un bosquejo general para la construcción del software, teniendo en cuenta la herramienta UML.

**Etapas 3: Desarrollo y prueba del software.** Teniendo los insumos recolectados en las etapas anteriores, se da paso al desarrollo del software como tal. En esta etapa es necesario identificar los aspectos que resultan relevantes para la validación de un documento de acuerdo con la norma NTC 1486, las especificaciones técnicas como lenguaje de desarrollo y la herramienta con la que se creó el software estará especificada en los anexos B y C del documento. De esta manera se logrará realizar un prototipo funcional que verifique el cumplimiento de documentos de Word bajo dicha norma. Finalmente, es necesario realizar pruebas para corroborar la funcionalidad del prototipo de acuerdo con las especificaciones de las tareas que se espera que desarrolle. Estas pruebas se realizarán mediante un piloto de validación de un documento de Word perteneciente al repositorio de la Universidad Católica de Colombia, verificando que arroja todos los metadatos necesarios para determinar el nivel de cumplimiento respecto a la norma NTC 1486.

---

<sup>64</sup> CARDONA TORRES, Sergio Augusto, et al. *Diseño de una estrategia de aprendizaje para implementar prácticas de PSP y TSP en cursos básicos de programación: caso programa de ingeniería de sistemas y computación Universidad del Quindío*. 2012. Tesis de Maestría. Universidad EAFIT.

## 8. RESULTADOS

Para dar cumplimiento a las etapas propuestas dentro de la metodología se ha realizado una descripción de las acciones ejecutadas en cada momento del proyecto de investigación, de esta manera, en los numerales 8.1; 8.2; 8.3 y 8.4 se brinda una breve explicación de la etapa del desarrollo de la metodología denominada “Documentación de requerimientos funcionales y características del prototipo en general”. Dicha información pertenece a un documento anexo en donde se detallan las características del prototipo, facilitando así el proceso de proyección del software, además, también está la descripción de forma cuantitativa y cualitativa la descripción de los requerimientos funcionales y los casos de uso. Este anexo tiene como título “ANEXO A” y se encuentra en el presente documento en la sección de anexos.

### 8.1. REQUERIMIENTOS DEL SISTEMA

Para exponer de forma clara y explícita las funcionalidades del software ValidateDocx y además el alcance de su desarrollo, se define la lista de requerimientos funcionales de este sistema, la definición de los casos de uso junto a su diagrama, requerimientos no funcionales, atributos de calidad entre otras especificaciones, teniendo en cuenta el comportamiento de la solución de este documento y la información que se debe manejar, ya que el objetivo principal es que el sistema sea de fácil acceso y manejo para los usuarios que requieran de su uso.

**8.1.1. Requerimientos funcionales.** La definición de requerimientos funcionales es muy importante porque genera bases sólidas para comenzar la elaboración de cualquier sistema, ya que con esta práctica además de evaluar si el resultado que plantea el desarrollo es correcto y cumple la solución que se espera obtener, también proporciona la información lo suficientemente detallada, para que un desarrollador que vaya a implementar estos requerimientos no tenga ningún problema al realizarlos.

La lista de requerimientos para el desarrollo de este proyecto que se muestra a continuación contiene cada uno de los nombres de los requerimientos funcionales definidos para la elaboración de este sistema:

- Requerimiento Funcional 1: Iniciar Sesión
- Requerimiento Funcional 2: Crear cuenta.
- Requerimiento funcional 3: Recuperar contraseña.
- Requerimiento funcional 4: Editar cuentas de usuario.
- Requerimiento funcional 5: Editar administrador
- Requerimiento funcional 6: Editar Usuario.
- Requerimiento funcional 7: Seleccionar usuario.
- Requerimiento funcional 8: Consultar Validaciones Realizadas.
- Requerimiento funcional 9: Descargar Validación.
- Requerimiento funcional 10: Cargar documento
- Requerimiento funcional 11: Cargar información del trabajo de grado.
- Requerimiento funcional 12: Obtener información del documento.

- Requerimiento funcional 13: Obtener historial de documentos.
- Requerimiento funcional 14: Obtener historial de validaciones

A continuación, se mostrará la definición del primer requerimiento funcional del sistema:

### 8.1.1.1. Requerimiento funcional 1: Iniciar sesión.

Tabla 2: Requerimiento Funcional Iniciar Sesión.

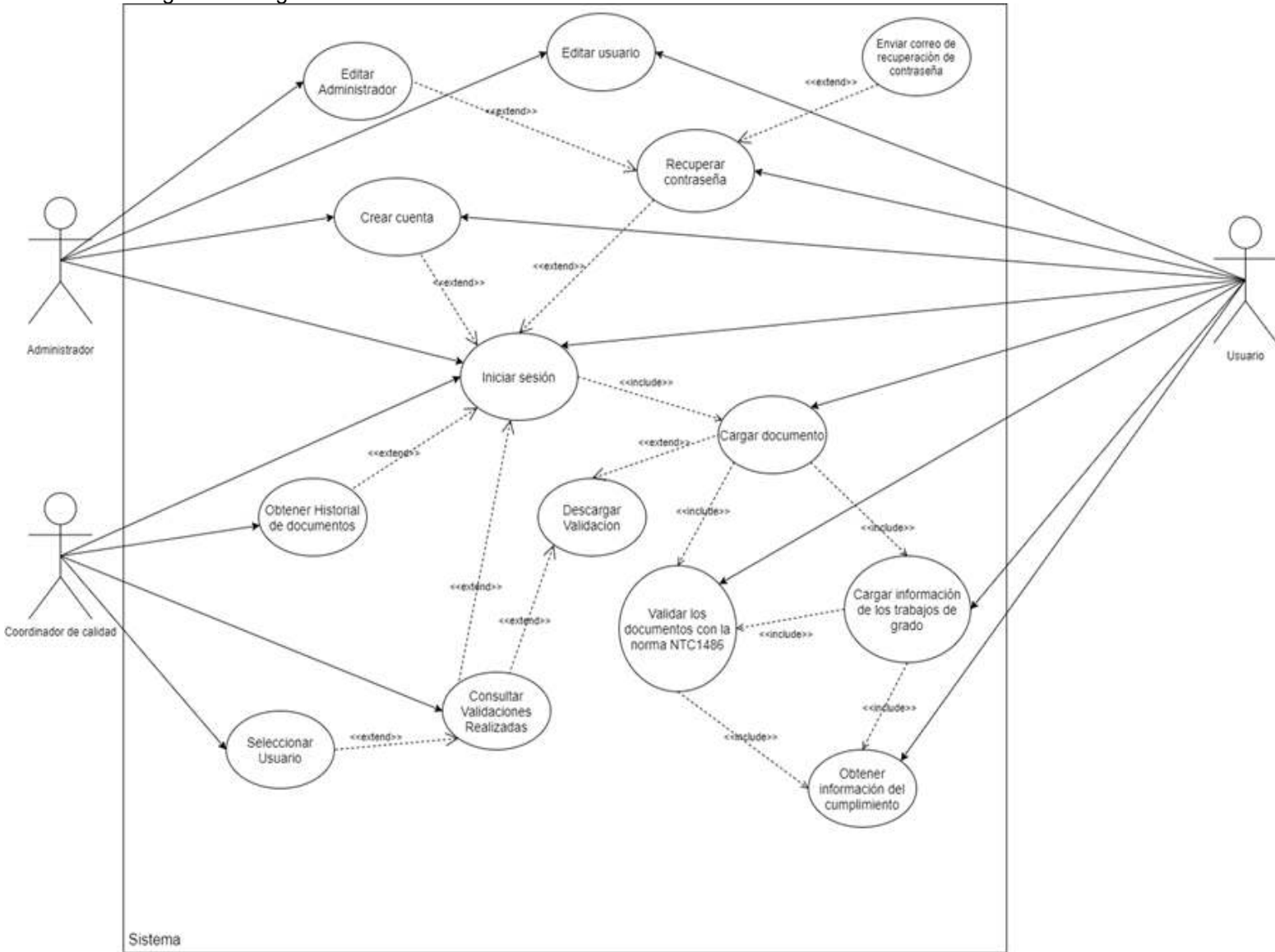
<b>Identificación del requerimiento</b>	RF01
<b>Nombre del requerimiento</b>	Iniciar Sesión
<b>Descripción del requerimiento</b>	El sistema validará que las credenciales ingresadas por el usuario existan en la base de datos, de lo contrario el usuario deberá crear una cuenta o validar que los datos sean correctos.
<b>Entradas</b>	<ul style="list-style-type: none"> <li>• usuario</li> <li>• password1</li> </ul>
<b>Salidas</b>	<ul style="list-style-type: none"> <li>• Usuario validado en B.D.</li> <li>• Datos incorrectos en validación de B.D.</li> <li>• <b>MENSAJE</b> "USUARIO INEXISTENTE, POR FAVOR VALIDAR O CREAR CUENTA".</li> <li>• <b>MENSAJE</b> "CONTRASEÑA INCORRECTA, POR FAVOR VALIDAR O RECUPERAR CONTRASEÑA".</li> </ul>

Fuente: Autor

## 8.2. CASOS DE USO

El sistema contará con 3 tipos diferentes de usuarios los cuales podrán tener diferentes tipos de interacciones con las funcionalidades del sistema, estas interacciones están definidas con detalle en la figura 18 del diagrama de casos de uso, el cual además de representar las interacciones del usuario con el sistema, también muestra la interacción de todas las funcionalidades del software entre sí, además también se definen los datos de entrada y de salida que tendrá cada caso de uso.

Figura 16:Diagrama de casos de uso.



Fuente: El autor

A continuación, se mostrará el primero de los 14 casos de uso definidos, en donde se detallará específicamente la descripción del caso de uso y también sus datos de entrada y salida.

### 8.2.1. Caso de uso 1: Iniciar sesión.

Tabla 3: Caso de Uso 1: Iniciar Sesión

<b>Identificación del caso de uso</b>	Caso de Uso-001
<b>Nombre del caso de uso</b>	Iniciar Sesión
<b>Descripción del caso de uso</b>	El sistema validará que las credenciales ingresadas por el usuario existan en la base de datos, de lo contrario el usuario deberá crear una cuenta o validar que los datos sean correctos.
<b>Entradas</b>	<ul style="list-style-type: none"> <li>• usuario</li> <li>• password</li> </ul>
<b>Salidas</b>	<ul style="list-style-type: none"> <li>• Usuario validado en B.D.</li> <li>• Datos incorrectos en validación de B.D.</li> <li>• <b>MENSAJE</b> “USUARIO INEXISTENTE, POR FAVOR VALIDAR O CREAR CUENTA”</li> <li>• <b>MENSAJE</b> “CONTRASEÑA INCORRECTA, POR FAVOR VALIDAR O RECUPERAR CONTRASEÑA”</li> </ul>
<b>Flujo Normal</b>	<ol style="list-style-type: none"> <li>1. SISTEMA: Despliega textbox para captura de usuario y contraseña</li> <li>2. USUARIO: Ingresa usuario e Ingresa password</li> <li>3. SISTEMA: Consulta la existencia de usuario.</li> <li>4. SISTEMA: Valida correspondencia de usuario y password en la Base de Datos.</li> <li>5. SISTEMA: Despliega página de Sesión de Usuario.</li> </ol>
<b>Flujos Alternos</b>	<ol style="list-style-type: none"> <li>3.1.1. SISTEMA: Valida la inexistencia de usuario</li> <li>3.1.2. SISTEMA: Despliega MENSAJE “USUARIO INEXISTENTE, POR FAVOR VALIDAR O CREAR CUENTA”</li> <li>4.1.1 SISTEMA: Valida la no correspondencia de usuario y password en la Base de Datos.</li> <li>4.1.2 SISTEMA: Despliega MENSAJE “CONTRASEÑA INCORRECTA, POR FAVOR VALIDAR O RECUPERAR CONTRASEÑA”</li> </ol>

Fuente: Autor

Para ver la lista de casos de uso completa y a detalle, puede remitirse al anexo: Especificación de Requerimientos Funcionales.

### 8.3. REQUERIMIENTOS NO FUNCIONALES

#### 8.3.1. Requisito no funcional 1: Interfaz responsive.

Tabla 4: Requerimiento no Funcional 1: Interfaz Responsive

<b>Identificación del requerimiento</b>	RNF01
<b>Nombre del requerimiento</b>	Interfaz Responsive
<b>Descripción del requerimiento</b>	El sistema debe procurar que la interfaz del usuario tenga tecnología responsive ya que esta tecnología permite que el sistema pueda ser visualizado correctamente desde cualquier dispositivo, sin importar el tamaño del a pantalla.

Fuente: Autor

#### 8.3.2. Requisito no funcional 2: Permisos de acceso al sistema.

Tabla 5: Requerimiento no Funcional 2: Permisos de Acceso al sistema

<b>Identificación del requerimiento</b>	RNF02
<b>Nombre del requerimiento</b>	Permisos de acceso al sistema
<b>Descripción del requerimiento</b>	Los permisos de acceso al sistema solo podrán ser modificados por un usuario administrador.

Fuente: Autor

#### 8.3.3. Requisito no funcional 3: Errores informativos.

Tabla 6: Requerimiento no funcional 3: Errores Informativos.

<b>Identificación del requerimiento</b>	RNF03
<b>Nombre del requerimiento</b>	Errores Informativos
<b>Descripción del requerimiento</b>	El sistema debe retornar errores informativos que el usuario final pueda entender.

Fuente: Autor

Para ver la lista de requerimientos completa y a detalle, puede remitirse al anexo A: Especificación de Requerimientos Funcionales.



## 8.4. ARQUITECTURA DEL SISTEMA

En la segunda etapa de este proyecto, denominada "Diseño arquitectural de software" se realizó un proceso de organización, planteamiento y construcción de los aspectos que componen el software, lo cual permitió un desarrollo ordenado siguiendo con la metodología PSP en sus diferentes estadios de acción. Esta fase se puede ver evidenciada en el anexo B, titulado diseño de arquitectura de software. A continuación, se presenta un resumen de este proceso en los numerales 8.5.1 y 8.5.2.

En esta parte del documento se describe de forma detallada y explícita, el diseño del software (ValidateDocx) para la validación de documentos de Word bajo la norma NTC1486 para documentos de tesis, trabajos de grado y trabajos de investigación en el territorio colombiano, esto permite proveer un conocimiento de toda la arquitectura del sistema, ya que este entregable es una pieza fundamental para su desarrollo.

ValidateDocx es una herramienta que implementa un proceso de ingeniería inversa con documentos de Word con el fin de extraer los metadatos que contienen estos de archivos. Este proceso de extracción le permite al sistema acceder a todas sus características y contenidos, y así poder validar a partir de esta información si el documento cumple con los estándares de la norma NTC1486, para llevar a cabo el desarrollo de este sistema, se pretende exponer a los usuarios, el diseño arquitectónico del software con notación UML como, por ejemplo: Diagrama de componentes, diagrama de despliegue, diagrama de clases. Estos diagramas nos permiten comprender de la mejor forma posible toda la composición del sistema.

**8.5.1. Stakeholders.** A continuación, se muestran las Stakeholders interesados en el sistema ValidateDocx.

Tabla 7: Stakeholders del sistema

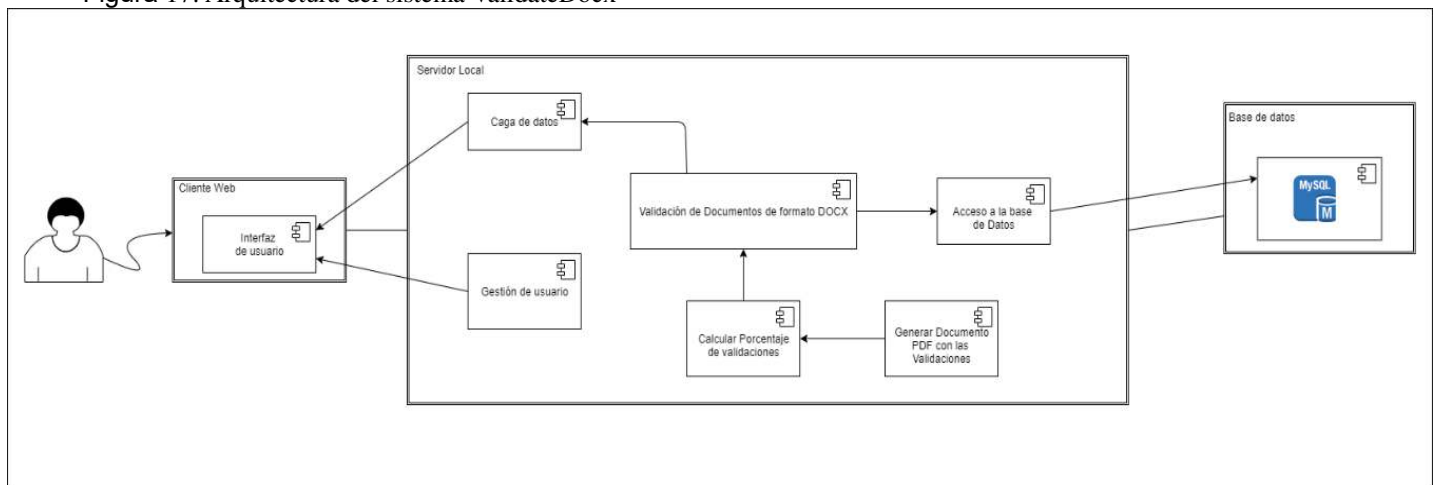
<b>STAKEHOLDERS</b>	<b>Descripción</b>
Administrador	Este actor es la persona encargada principalmente de administrar las cuentas de usuario de las personas que se registren en el sistema, esta persona es el único usuario que puede asignarle permisos de coordinador de calidad a un usuario.
Coordinador de calidad	Este actor es el encargado de supervisar y controlar las versiones de los documentos que se han cargado al sistema, junto con el resultado de las validaciones que se han realizado. Todo esto con el fin de tener un control detallado de la calidad con la que los documentos tienen en un principio.
Usuario	Este actor es el encargado de cargar los documentos al sistema e ingresar

la información relacionada con el documento de investigación, a diferencia del coordinador de calidad, este usuario solo puede ver el resultado de la validación de su propio documento.

Fuente: Autor

**8.5.2. Diagrama de arquitectura del sistema** En el diagrama de la figura 19, se muestra la arquitectura del sistema, se evidencia que el usuario accede a la aplicación desde un navegador el cual puede ser Google Chrome, Mozilla Firefox o Internet Explorer, los cuales pueden ser utilizados por la mayoría de dispositivos con conexión a internet ya sean móviles o computadores, esto con el fin de que el usuario pueda acceder a las funcionalidades del sistema y que pueda cargar sus archivos para que sean validados.

Figura 17: Arquitectura del sistema ValidateDocx



Fuente: El autor

Para ver esta información completa y a detalle, puede remitirse al anexo B: DOCUMENTO DE ARQUITECTURA DEL SOFTWARE.

## 8.5. MANUAL DE INSTALACIÓN DEL SISTEMA

Por último, se ejecutan actividades relacionadas con la etapa 3, desarrollo y prueba del software, dando como resultado la consolidación de los anexos C, D y E, donde se muestra el proceso de prueba del software con un documento de Word, haciendo uso del recurso construido. En esta fase se verifica la funcionalidad del software, de acuerdo con lo expuesto en los numerales 8.6; 8.7 y 8.8.

El sistema ValidateDocx se desarrolló con Django el cual es un framework de código abierto desarrollado con Python, por esta razón los componentes escogidos para el uso y despliegue del sistema se seleccionaron de acuerdo con los elementos que son compatibles con este framework.

### **8.6.1. Requisitos mínimos del hardware.**

**8.6.1.1. Servidor.** El servidor debe cumplir con los siguientes requisitos mínimos:

- Procesador Intel Core i5 o superior.
- Memoria RAM 8 GB o superior.
- Disco duro 256 GB o más.

**8.6.1.2. Cliente.** El dispositivo del cliente debe cumplir con los siguientes requisitos mínimos:

- Procesador Intel Core i3 o superior.
- Memoria RAM 2 GB o superior.
- El almacenamiento del disco duro no tiene un estándar mínimo, puede ser el de la preferencia del usuario.

### **8.6.2. Instalación de componentes.**

El proceso de instalación se debe hacer en el siguiente orden:

1. Servidor de bases de datos MYSQL5.7.
2. MYSQLWorkBench 8.0.
3. Python 3.8.2.
4. PIP.
5. Django 3.0.4.
6. PyMysql.
7. MysqlClient.
8. Zipfile.
9. Python-lxml.

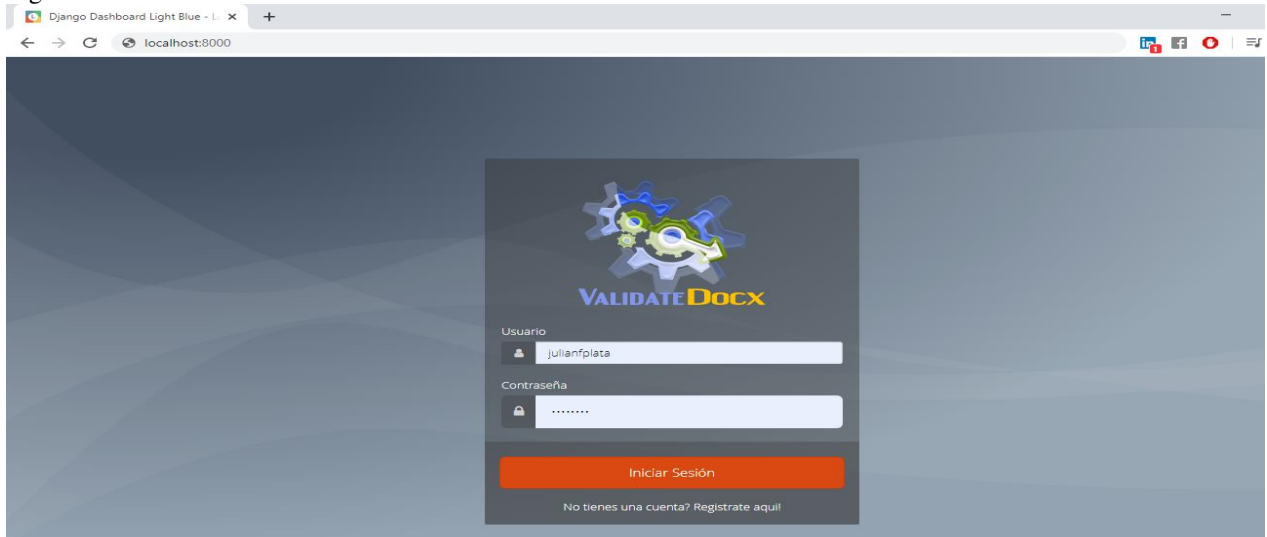
Para ver el manual de instalación completo y a detalle, puede remitirse al anexo C: Manual de Instalación del Sistema

## **8.6. MANUAL DE USUARIO**

A continuación, se describe con detalle las instrucciones para que cualquier usuario pueda utilizar al sistema ValidateDocx, accediendo desde un dispositivo móvil o pc por medio de un navegador web, este manual contiene la descripción y comportamiento de cada una de las funcionalidades del sistema, se puede comprobar también que las diferentes funcionalidades diseñadas en la arquitectura de software de este documento efectivamente fueron desarrolladas e implementadas.

**8.7.1. Inicio de sesión.** En la figura 20 se visualiza un recuadro el cual le pide que se ingrese una clave y una contraseña que previamente el usuario debió crear en el sistema con la funcionalidad de crear cuenta de usuario.

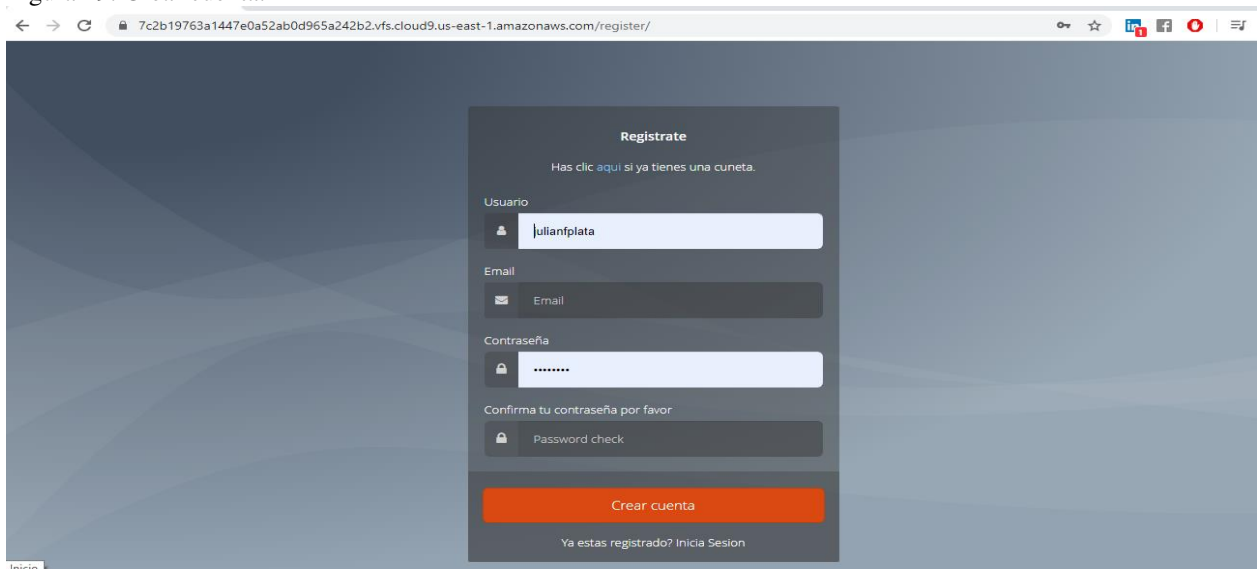
Figura 18: Inicio de sesión.



Fuente: El autor

**8.7.2. Crear cuenta.** En la figura 21 se visualiza un recuadro el cual le pide que se ingresen datos como; usuario, email, contraseña y otro recuadro para digitar la contraseña por segunda vez y verificar que el usuario esté totalmente seguro de lo que está escribiendo, con estos datos registrados en el sistema, el usuario podrá acceder en el momento en que lo prefiera para hacer uso de las funcionalidades.

Figura 19: Crear cuenta.



Fuente: El autor

Para ver la información de uso y funcionalidades completas y a detalle, puede remitirse al anexo D: Manual de usuario.

## 8.7. PRUEBAS FUNCIONALES DEL SISTEMA

Para el desarrollo de este proyecto, se documenta una serie de pruebas funcionales realizadas al sistema ValidateDocx, para verificar que el software desarrollado funciona de manera adecuada, todo esto basado en los anexos de arquitectura del sistema, de casos de uso y requerimientos funcionales, para asegurar que las pruebas son efectivas, se decide realizar pruebas funcionales del sistema por medio de la creación de casos de pruebas, los cuales contemplan la lista completa de requerimientos funcionales y el resultado esperado que se desea obtener al ejecutar cada función, esto garantiza que el sistema cumple con el objetivo inicial de automatizar el proceso de validación y verificación de un documento de Word con las normas NTC1486.

### 8.8.1 Casos de prueba

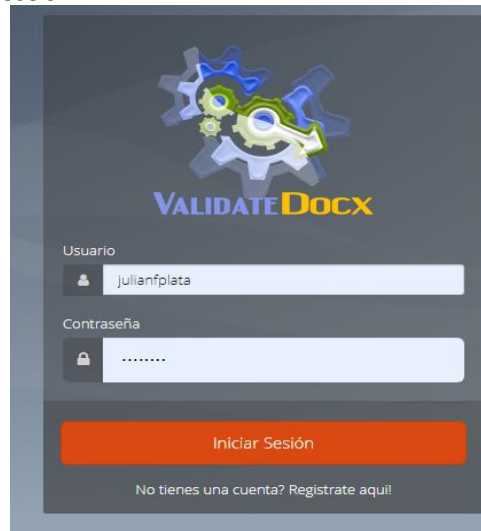
#### 8.7.1.1. Caso de prueba 1: Inicio de sesión.

Tabla 8: Caso de prueba: Inicio de sesión.

<b>Identificador</b>	CP - 01	<b>Versión</b>	1.0
<b>Nombre del caso de prueba</b>	Inicio de sesión		
<b>Módulo</b>	Gestión de usuarios		
<b>Descripción de la prueba</b>	Se validan las credenciales de ingreso del usuario para permitir o denegar el acceso al sistema.		
<b>Prerrequisitos</b>	El usuario debe registrarse previamente en el sistema con la funcionalidad de registrar usuario.		
<b>Resultados esperados</b>	Se almacenan los datos digitados por el usuario para que posteriormente pueda iniciar sesión el sistema.		
<b>Resultados reales</b>	El sistema almacena exitosamente los datos del usuario, para posteriormente permitirle iniciar sesión		

## Imagen

Figura 20: Evidencia de inicio de sesión



Fuente: El autor

Fuente: El autor

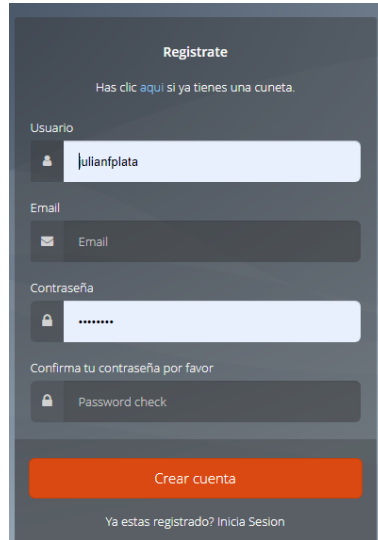
### 8.8.1.2. Caso de prueba 2: Crear Cuenta.

Tabla 9: Caso de prueba: Crear cuenta.

<b>Identificador</b>	CP - 02	<b>Versión</b>	1.0
<b>Nombre del caso de prueba</b>	Crear Cuenta		
<b>Módulo</b>	Gestión de usuarios		
<b>Descripción de la prueba</b>	El usuario ingresa sus datos personales, para que posteriormente el sistema los almacene en la base de datos y pueda iniciar sesión.		
<b>Prerrequisitos</b>	Ninguno, cualquier usuario puede acceder a esta funcionalidad		
<b>Resultados esperados</b>	Se almacenan los datos digitados por el usuario para que posteriormente pueda iniciar sesión el sistema.		
<b>Resultados reales</b>	El sistema almacena exitosamente los datos del usuario, para posteriormente permitirle iniciar sesión		

## Imagen

Figura 21: Evidencia, registro de usuario.



The image shows a registration form with the following fields and elements:

- Regístrate** (Title)
- Link: [Has clic aquí si ya tienes una cuenta.](#)
- Usuario** (User): Input field containing "julianplata".
- Email** (Email): Input field containing "Email".
- Contraseña** (Password): Input field containing "\*\*\*\*\*".
- Confirma tu contraseña por favor** (Confirm your password please): Input field containing "Password check".
- Crear cuenta** (Create account): Orange button.
- Link: [Ya estas registrado? Inicia Sesión](#)

Fuente: El autor

Fuente: El autor

Para ver la información de las pruebas funcionales y sus evidencias completas y a detalle, puede remitirse al anexo E: Pruebas Funcionales.

## 9. CONCLUSIONES

Dentro de esta investigación se realizaron actividades que aportaron grandes conocimientos desde un punto de vista diferente para el resultado final, logrando llegar a las siguientes conclusiones:

- La extracción de metadatos es una buena herramienta para realizar procesos de automatización en la validación de documentos, ya que permite identificar elementos más allá de lo que el usuario normalmente puede identificar en un documento de Word, lo cual permite un análisis más detallado y eficaz del cumplimiento de normas o requisitos básicos para la aceptación del documento dentro de un ambiente investigativo, aportando así a la comunidad científica y a la construcción de conocimiento en la sociedad.
- La herramienta generada en esta investigación aporta a la búsqueda de información y el uso de las tecnologías a partir de la metadata, proponiendo enfoques diferenciales para el manejo de información en la web y fortaleciendo la visión de la Web semántica como un facilitador para la navegación y el retorno de búsquedas más inteligentes y pertinentes de acuerdo con las necesidades de los usuarios.
- La documentación de la extracción de metadatos en archivos de Word es una característica novedosa, ya que dentro del estado del arte se puede evidenciar que los estudios realizados en torno a la extracción de metadatos se relacionan con documentos en formato PDF y otros. Esto permitió que el autor de esta investigación realizara un trabajo inicial respecto a la caracterización de los metadatos que contiene un archivo de Word, brindando espacios para realizar análisis de documentos, la estructura de las aplicaciones y su funcionamiento.
- La arquitectura de software MVC fue una metodología importante para el resultado positivo de esta investigación, ya que facilita la construcción de un software que cumpla con los requerimientos funcionales y no funcionales necesarios para garantizar su correcta ejecución de acuerdo a los objetivos trazados al momento de planear el avance del software, por lo cual esta arquitectura favorece el desarrollo de un proyecto investigativo con un sentido de aprendizaje, tal como lo es un proyecto de pregrado.



## 10. TRABAJOS FUTUROS

Esta investigación tiene un fuerte impacto en la construcción de conocimiento en el campo de la ingeniería ya que reconoce formas de investigación inexploradas, por lo cual se puede sugerir que se realicen investigaciones en torno al número de trabajos que cumplen con la NTC 1486 en la Universidad Católica de Colombia usando este prototipo, así como la añadidura de funciones más específicas respecto al software a fin de realizar una validación por jueces expertos que formalice su uso en diferentes contextos de educación superior del país.

De igual manera, se proyectan trabajos de investigación encaminados al diseño y construcción de plataformas automatizadas que permitan la divulgación únicamente de contenido que cumpla con la norma NTC 1486, sin embargo, es posible augurar el desarrollo de trabajos similares al presente con otras normas para la presentación de documentos investigativos como las APA.

Finalmente, se prevé que este trabajo de grado tiene implicaciones positivas y genera insumos para el desarrollo de programas en distintas universidades del país, que den herramientas de aprendizaje a los estudiantes para la construcción de sus documentos investigativos con antelación, favoreciendo los procesos de investigación en un entorno educativo.

## 11. Bibliografía

1KEYDATA. Conceptual Data Model; [Consultado el: 20 de octubre de 2019]. Disponible en:

<https://www.1keydata.com/datawarehousing/conceptual-data-model.html>

1KEYDATA. Physical Data Model; [Consultado el: 20 de octubre de 2019]. Disponible en: <https://www.1keydata.com/datawarehousing/physical-data-model.html>

ALONSO, Fernando, et al. *Introducción a la ingeniería del software*. Delta Publicaciones, 2005.

AMENEYRO, López; GISWEB, A. A. *Reingeniería para la implementación de un Web FeatureService*. 2004. Tesis Doctoral. Tesis de Licenciatura en Sistemas Computacionales. Departamento de Ingeniería en Sistemas Computacionales, Escuela de Ingeniería, Universidad de las Américas, Puebla.

ANALITICA NEGOCIOS. Guías; [Consultado el: 20 de octubre de 2019]. Disponible en: <https://www.analitanegocios.com/guías/diferencias-datos-estructurados-no-estructurados/>

ARRIETA, Juan Gregorio, et al. Aplicación lean manufacturing en la industria colombiana. Revisión de literatura en tesis y proyectos de grado. *Revista Virtual Pro*, 2011, vol. 132, no 9.

ARRIOLA NAVARRETE, Oscar; BUTRÓN YÁÑEZ, Katya. Sistemas integrales para la automatización de bibliotecas basados en software libre. *Acimed*, 2008, vol. 18, no 6, p. 0-0.

BARRIGA MARIÑO, J. C., & RINCON YAÑEZ, D. A. (2017). Construcción De Una Herramienta Web Para La Extracción Y Almacenamiento De Datos De La Red Social Twitter. Universidad Católica de Colombia.

BENÍTEZ, José Javier Hernández, et al. Herramienta para el procesamiento en lote de documentos digitales para el sistema REPXOS 3.0. *Revista Publicando*, 2018, vol. 5, no 14 (3), p. 571-582.

CALVO-MANZANO, José A., et al. Perfiles del ciclo de vida del software para pequeñas empresas: los informes técnicos ISO/IEC 29110. *REICIS. Revista Española de Innovación, Calidad e Ingeniería del Software*, 2008, vol. 4, no 2, p. 96-108.

CARDONA TORRES, Sergio Augusto, et al. *Diseño de una estrategia de aprendizaje para implementar prácticas de PSP y TSP en cursos básicos de programación: caso programa de ingeniería de sistemas y computación Universidad del Quindío*. 2012. Tesis de Maestría. Universidad EAFIT.

CASALI, A, et al. Asistente para el depósito de objetos en repositorios con extracción automática de metadatos. En *XV Simposio Internacional de Tecnologías de la Información y las Comunicaciones en la Educación SINTICE*. 2013. p. 133-136.

CASTILLO, Jessica Nataly, et al. Base de Datos NoSQL: MongoDB vs. Cassandra en operaciones CRUD (Create, Read, Update, Delete). *Revista Publicando*, 2017, vol. 4, no 11 (1), p. 79-107.

CONCEPTODEFINICION, Datos; [Consultado el 10 de 03 de 2020]. Disponible en: <https://conceptodefinicion.de/datos/>.

CORNEJO, Marcela; SALAS, Natalia. Rigor y calidad metodológicos: un reto a la investigación social cualitativa. *Psicoperspectivas*, 2011, vol. 10, no 2, p. 12-34.

DECO, Claudia; BENDER, Cristina; SAER, Jorge. Ponderación de metadatos de recursos educativos como forma de mejorar los resultados de una búsqueda. 2011.

DEFINICION.DE, Definición de binario; [Consultado el 20 de 02 de 2020]. Disponible en: <https://definicion.de/binario/#ixzz2rE2HXYYr>

DEL BUSTO, Hansel Gracia; ENRÍQUEZ, Osmel Yanes. Bases de datos NoSQL. *Revista Telemática*, 2013, vol. 11, no 3, p. 21-33.

DURÁN ACEVEDO, Jesús Antonio, et al. Estudio de Seguridad Informática de los metadatos contenidos en archivos publicados en los sitios web de las organizaciones: Alcaldía de Pamplona, Cámara de Comercio de Pamplona, Gobernación de Norte de Santander, Diario la Opinión y La Dian.

FERNANDEZ FRANCO, Andrey Mauricio, et al. Caracterización de la información mediante la extracción de metadatos utilizando recuperación de información sobre convocatorias. 2016.

FILMAKER PRO, XML format;[Consultado el 10 de 03 de 2020 ] Disponible en: [https://fmhelp.filemaker.com/help/18/fmp/es/index.html#page/FMP\\_Help/xml-format.html.p.8](https://fmhelp.filemaker.com/help/18/fmp/es/index.html#page/FMP_Help/xml-format.html.p.8)

FLORES RIERA, Leduan, et al. Componente para la extracción automática de metadatos bibliográficos desde corpus textuales en formato PDF. *Revista Cubana de Ciencias Informáticas*, 2017, vol. 11, no 4, p. 85-98.

GÓMEZ, Ricardo; MARTÍNEZ, Juliana. Internet: ¿para qué? 2001.

HAYES, Brian. Cloud computing. 2008.

HERRERA, Harol Andrey; VALENZUELA, Carlos Rueda. NoSQL, la nueva tendencia en el manejo de datos. *Tecnología Investigación Y Academia*, 2016, vol. 4, no 1, p. 147-150.

ICONTEC, N. (2008). NTC 1486. *Elaboración de Informes y trabajos de grado, Sexta edición, Editada por el Icontec*.

INSTITUTO COLOMBIANO DE NORMAS TÉCNICAS Y CERTIFICACIÓN. Documentación, Presentación de tesis, trabajos de grado y otros trabajos de investigación, Norma Técnica Colombiana. NTC 1486. [En línea] Bogotá D.C.: el instituto. 41 p.

KHANACADEMY. Computer-science; [Consultado el 02 de 03 de 2020 ] Disponible en : <https://es.khanacademy.org/computing/computer-science/algorithms/binary-search/a/binary-search>

LEXICO POERED BY OXFORD, UK Dictionary; [Consultado el 10 de 03 de 2020].  
Disponible en: [https://www.lexico.com/definition/data\\_set](https://www.lexico.com/definition/data_set)

LEXICO POERED BY OXFORD, UK Dictionary; [Consultado el 10 de 03 de 2020].  
Disponible en: [https://www.lexico.com/definition/world\\_wide\\_web](https://www.lexico.com/definition/world_wide_web)

LUCIDCHART. Que es un modelo de datos; [Consultado el: 20 de octubre de 2019].  
Disponible en: <https://www.lucidchart.com/pages/es/que-es-un-modelo-de-base-de-datos>

LYNDEN, Steven; TAVEEKARN, Waran. Semi-automated augmentation of pandas dataframes. En *International Conference on Data Mining and Big Data*. Springer, Singapore, 2019. p. 70-79.

MINISTERIO DE CIENCIA, TECNOLOGÍA E INNOVACIÓN. La ciencia en cifras [Sitio Web]. Bogotá; [Consultado el 23 de febrero de 2020] Disponible en: <https://minciencias.gov.co/la-ciencia-en-cifras>

MOLINA LOPEZ, Jose Manuel y GARCIA HERRERO, Jesus. Técnicas de análisis de datos: Aplicaciones prácticas utilizando Microsoft Excel y Weka.

MONROY, M. E.; ARCINIEGAS, J. L.; RODRÍGUEZ, J. C. Caracterización de herramientas de ingeniería inversa. *Información tecnológica*, 2012, vol. 23, no 6, p. 31-42.

MONROY, Martín E.; ARCINIEGAS, José L.; RODRÍGUEZ, Julio C. Caracterización de herramientas de ingeniería inversa. *Información tecnológica*, 2012, vol. 23, no 6, p. 31-42.

MONTALBANO, Elizabeth. Forrester: Microsoft office in no danger from competitors. PC World, 2009.

OXFORD LEARNER'S DICTIONARIES, definition english; [Consultado el 10 de 03 de 2020]. Disponible en: <https://www.oxfordlearnersdictionaries.com/definition/english/prototype>

OXFORD LEARNER'S DICTIONARIES, definition english; [Consultado el 10 de 03 de 2020]. Disponible en: [https://www.oxfordlearnersdictionaries.com/definition/english/data\\_dictionary](https://www.oxfordlearnersdictionaries.com/definition/english/data_dictionary)

OXFORD LEARNER'S DICTIONARIES, Definition of metadata noun, 2020. Disponible en: Oxford Learner's Dictionaries, Definition of metadata noun

PDF-TOOLS. Automate your data extraction: [Consultado el 20 de noviembre de 2020]. Disponible en: <https://uao.libguides.com/c.php?g=529834&p=3623716#Autor>

OPENCLASSROOMS, Que es el desarrollo web?:[Consultado el 15 de 04 de 2020]. Disponible en: <https://blog.openclassrooms.com/es/2017/09/11/que-es-el-desarrollo-web/>

PIÑA GONZALEZ, Orlando Julián. Extracción Automática de Metadatos de un Documental Digital. Instituto tecnológico y de estudios superiores de monterrey. 2000

PIRE, Taihú, et al. Extracción automática de metadatos de objetos de aprendizaje: un estudio comparativo. En *VI Congreso de Tecnología en Educación y Educación en Tecnología*. 2011.

PRIETO, J. (2016a). Lo que la extensión DOCX esconde. Retrieved from <https://www.gonduana.com/lo-que-la-extension-docx-esconde/>

POWER DATA, Big Data, Big Data: ¿En qué consiste? Su importancia, desafíos y gobernabilidad; [Consultado el 20 de 02 de 2020]. Disponible en: <https://www.powerdata.es/big-data>

RAMOS, Daniel, et al. *Curso de Ingeniería de Software: 2ª Edición*. IT Campus Academy, 2017

SCHWAB, Klaus. La cuarta revolución industrial. Debate, 2016.

SHADISH, W. R., BRASIL, I. C. C., ILLINGWORTH, D. A., WHITE, K. D (2009). Using UnGraph to extract data from image files: Verification of reliability and validity. *Behavior Research Methods*, 41(1), 177–183. <https://doi.org/10.3758/BRM.41.1.177>

Shadish, William & Brasil, Isabel & Illingworth, David & White, Kristen & Galindo, Rodolfo & Nagler, Eden & Rindskopf, David. (2009). Using UnGraph to extract data from image files: Verification of reliability and validity. *Behavior research methods*. 41. 177-83. [10.3758/BRM.41.1.177](https://doi.org/10.3758/BRM.41.1.177).

Tecnologías-Información, Modelo-datos; [Consultado el: 20 de octubre de 2019].

Disponible en: <https://www.tecnologias-informacion.com/modelos-datos.html>

Tim Berners-Lee, James Hendler, Ora Lassila, The Semantic Web, *Scientific American*, May 2001

UNIVERSIDAD CATÓLICA DE COLOMBIA. Dirección central de investigaciones [Sitio Web]. Bogotá; [Consultado 22 de febrero de 2020] Disponible en:

<https://www.ucatolica.edu.co/portal/investigaciones/direccion-central-de-investigaciones/>

UNIVERSIDAD DE ALICANTE, Servicio de informatica ASP.NET MVC 3

Framework; [Consultado el: [20 de 03 de 2020]. Disponible en:

<https://si.ua.es/es/documentacion/asp-net-mvc-3/1-dia/modelo-vista-controlador-mvc.html>

UQBAR-FOUNDATION. Articles, Atributos de calidad; [Consultado el 15 de marzo de 2020] Disponible en: <http://wiki.uqbar.org/wiki/articles/atributos-de-calidad.html>

VARA, Juan M.; VELA, Belén; MARCOS, Esperanza. Diseño de bases de datos objeto-relacionales con UML. *Diseño de bases de datos objeto-relacionales con UML*, 2005, p. 0-0.

JUAN, Rodríguez; PAZ, Julieta. *Servicio de recolección de metadatos genérico para documentos*. 2017. Tesis Doctoral. Universidad Nacional de La Plata.

W3C, Standards; [Consultado el 10 de 03 de 2020] Disponible en:

<https://www.w3.org/standards/semanticweb/>

Yildiz, Burcu & Kaiser, Katharina & Miksch, Silvia. (2005). pdf2table: A Method to Extract Table Information from PDF Files. 1773-1785.



**UNIVERSIDAD CATÓLICA**  
de Colombia

## **ANEXO A ESPECIFICACIÓN DE REQUERIMIENTOS DE SOFTWARE**

**JULIAN FELIPE PLATA ZUÑIGA**

**UNIVERSIDAD CATÓLICA DE COLOMBIA  
FACULTAD DE INGENIERÍA  
PROGRAMA DE INGENIERÍA DE SISTEMAS Y COMPUTACIÓN  
BOGOTÁ  
2020**

## Contenido

1. Introducción .....	75
1.1. Propósito .....	76
1.2. Ámbito del sistema .....	76
1.3. Definiciones, acrónimos y abreviaturas .....	77
1.4. Visión general del documento .....	<b>¡Error! Marcador no definido.</b>
2. Descripción general .....	<b>¡Error! Marcador no definido.</b>
2.1. Perspectiva del producto .....	<b>¡Error! Marcador no definido.</b>
2.2. Requisitos Funcionales .....	<b>¡Error! Marcador no definido.</b>
2.1.1. Requerimientos Funcional 1: Iniciar Sesión .....	79
2.1.2. Requerimiento funcional 2: Crear cuenta. ....	79
2.1.3. Requerimiento funcional 3: Recuperar contraseña.....	80
2.1.4. Requerimiento funcional 4: Editar administrador. ....	80
2.1.5. Requerimiento funcional 5: Enviar correo de recuperación de contraseña.	82
2.1.6. Requerimiento funcional 6: Editar Usuario. ....	83
2.1.7. Requerimiento funcional 7: Seleccionar usuario.....	84
2.1.8. Requerimiento funcional 8: Consultar Validaciones Realizadas.....	84
2.1.9. Requerimiento funcional 9: Descargar Validación. ....	84
2.1.10. Requerimiento funcional 10: Cargar documento .....	85
2.1.11. Requerimiento funcional 11: Cargar información del trabajo de grado.	86
2.1.12. Requerimiento funcional 12: Obtener información del documento.	87
2.1.13. Obtener historial de documentos.....	88
2.1.14. Obtener historial de validaciones. ....	89
2.2. Requerimientos no funcionales .....	<b>¡Error! Marcador no definido.</b>
2.2.1. Requisito no funcional 1: Integridad.....	90
2.2.2. Requisito no funcional 2: Modificabilidad. <b>¡Error! Marcador no definido.</b>	
2.2.3. Requisito no funcional 3: Verificabilidad. <b>¡Error! Marcador no definido.</b>	
2.2.4. Requisito no funcional 4: Disponibilidad. <b>¡Error! Marcador no definido.</b>	
2.3. Funciones del producto .....	<b>¡Error! Marcador no definido.</b>
2.3.1. Detalle de Casos de Uso. ....	1
2.3. Características de los usuarios .....	<b>¡Error! Marcador no definido.</b>
2.4. Restricciones.....	<b>¡Error! Marcador no definido.</b>
2.5. Suposiciones y dependencias .....	<b>¡Error! Marcador no definido.</b>
3. Requisitos Específicos.....	<b>¡Error! Marcador no definido.</b>
3.1. Interfaces Externas .....	<b>¡Error! Marcador no definido.</b>
3.1.1. Interfaces de usuario. ....	19
3.1.2. Interfaces de software. ....	19



## Lista de tablas.

Tabla 1: Definiciones, acrónimos y abreviaturas. ....	77
Tabla 2: Requerimiento funcional: Iniciar Sesión .....	79
Tabla 3: Requerimiento funcional: Crear cuenta. ....	79
Tabla 4: Requerimiento funcional: Recuperar contraseña. ....	80
Tabla 5:Requerimiento funcional: Editar cuentas de usuario. ....	80
Tabla 6: Requerimiento funcional: Editar Administrador. ....	82
Tabla 7: Requerimiento funcional: Editar Usuario .....	83
Tabla 8: Requerimiento funcional: Seleccionar Usuario.....	84
Tabla 9: Requerimiento funcional: Consultar Validaciones Realizadas.....	84
Tabla 10: Requerimiento funcional: Descargar Validación.....	85
Tabla 11: Requerimiento funcional: Cargar Documento.....	85
Tabla 12: Requerimiento funcional: Cargar información del trabajo de grado.....	86
Tabla 13: Requerimiento funcional: Obtener información del documento. ....	87
Tabla 14: Requerimiento funcional: Obtener historial de documentos. ....	88
Tabla 15: Requerimiento funcional: Obtener historial de validaciones. ....	89
Tabla 16: Requerimiento no funcional: Interfaz responsive.....	90
Tabla 17: Requerimiento no funcional: Escalabilidad.....	90
Tabla 18: Caso de uso: Iniciar sesión. ....	1
Tabla 19: : Caso de uso: Crear cuenta.....	2
Tabla 20: Caso de uso: Recuperar contraseña. ....	3
Tabla 21: Caso de uso: Editar cuentas de usuario.....	5
Tabla 22: Caso de uso: Editar administrador. ....	6
Tabla 23: Caso de uso: Editar usuario. ....	8
Tabla 24: Caso de uso: Seleccionar usuario. ....	9
Tabla 25: Caso de uso: Consultar validaciones realizadas.....	10
Tabla 26: : Caso de uso: Descargar validación. ....	11
Tabla 27: Caso de uso: Cargar documento.....	12
Tabla 28: Caso de uso: Cargar información del trabajo de grado. ....	13
Tabla 29: Caso de uso: Obtener información del documento.....	14
Tabla 30: Caso de uso: Obtener historial de documentos.....	15
Tabla 31: Caso de uso: Obtener historial de validaciones. ....	16
Tabla 32: Tipo de usuario: Administrador.....	17
Tabla 33:Tipo de usuario: Coordinador de calidad.....	17
Tabla 34:Tipo de usuario: Usuario .....	18

**Lista de imágenes.**

Figura 1: Diagrama de casos de uso.....¡Error! Marcador no definido.

## **INTRODUCCIÓN.**

Este documento pretende realizar de forma detallada la especificación de requerimientos de software(ERS) ValidateDocx, para extraer los metadatos de un documento de Word con el fin de verificar si el contenido de este archivo sigue los estándares de la norma NTC1486 ya que el cumplimiento de esta norma es un requisito para realizar documentos de tesis, trabajos de grado y trabajos de investigación en el territorio colombiano.

El usuario tendrá una visión general de la estructura del sistema y sus funcionalidades. También se define el comportamiento del sistema en las diferentes situaciones que puedan ocurrir.

ValidateDocx tiene una funcionalidad específica para generar porcentajes que indican el nivel de cumplimiento que contiene uno o varios documentos de Word respecto a la norma NTC1486.

El documento se estructura a partir de las normas dadas por el estándar IEEE830 ya que esta es una práctica recomendada para la especificación de los requisitos de software.

## 1. PROPÓSITO.

Se pretende desarrollar un software que sirva como herramienta que permita automatizar el proceso de verificación del cumplimiento de un documento de Word con la norma NTC1486. El proceso que se debe efectuar para que la verificación de un documento siga los estándares de esta norma, se debe realizar de forma manual y es por esta razón que este tipo de procesos puede llegar a ser dispendioso ya que el usuario debe repetir el mismo proceso con varios documentos.

Es por esta razón que se busca automatizar este proceso, ya que permite optimizar tanto el tiempo, como los resultados que se obtienen al revisar un documento que deba cumplir con esta norma, por otro lado, con la extracción de estos metadatos pueden llegar a ser de mucha utilidad para el mundo investigativo ya que los metadatos contienen información vital no solo de las características del documento sino de todo su contenido, lo cual abre un mundo de posibilidades para que este sistema sea integrado a las herramientas tecnológicas que actualmente son de mucha utilidad para la investigación.

### 1.1. ÁMBITO DEL SISTEMA.

El software que se describe en este documento se llamará “ValidateDocx”, inicialmente se desempeñará dentro de la institución de la Universidad Católica de Colombia, esto con el fin de demostrar que el software realmente es una herramienta que le aporta un valor agregado al proceso de calificación de los trabajos de grado, pero debido a que la norma NTC1486 es la norma que se utiliza a nivel nacional para tesis, trabajos de grado y trabajos de investigación, ValidateDocx puede ser utilizado en cualquier institución u organización que incurra en el mundo de la investigación en el país.

Para poder automatizar el proceso de verificación del nivel de cumplimiento de la norma NTC1486 de un documento de Word, el sistema tiene la capacidad de extraer los metadatos de un archivo de Word de formato DOCX los cuales contienen la información completa del contenido del documento y sus características, estas características son las que proveen la información que se definirá como el conjunto de criterios que permiten validar si realmente el documento está cumpliendo la norma exigida, el conjunto de criterios que se validan son los siguientes:

- Márgenes del documento.
- Tamaño de la fuente de letra.
- Interlineado.
- Citas bibliográficas.
- Tabla de contenido.
- Tabla de imágenes.
- Tabla de tablas.
- Tabla de anexos.
- Tabla de ecuaciones.
- Jerarquía de títulos.
- Referencias.

Existen otros aspectos que están definidos oficialmente en la norma NTC1486 como por ejemplo el tipo de redacción que se debe aplicar en los párrafos, el contenido de la portada o cualquier análisis semántico del documento, pero debido a la complejidad que conllevaría

hacer este tipo de análisis y validaciones, estos aspectos mencionados no se tendrán en cuenta.

El conjunto de validaciones de estas características da como resultado un porcentaje del nivel de cumplimiento que cada documento tiene, los usuarios podrán visualizar los porcentajes individuales de cada documento, o si bien lo prefiere, el porcentaje de cumplimiento de todos los documentos que el usuario ha validado.

En cuanto al diseño del sistema, deberá desarrollarse utilizando tecnología web, también deberá usar diseño responsive debido a que el entorno web y la tecnología responsive facilita el acceso a los usuarios desde cualquier tipo de dispositivo facilitando su visualización.

## 1.2. DEFINICIONES, ACRÓNIMOS Y ABREVIATURAS.

Tabla 10: Definiciones, acrónimos y abreviaturas.

Nombre	Descripción
Usuario	Persona que usará el sistema de revisión de documentos.
ERS	Especificación de Requisitos Software.
RF	Requerimiento Funcional.
RNF	Requerimiento No Funcional

Fuente: El autor

## 1.3. VISIÓN GENERAL DEL DOCUMENTO

Este documento está compuesto por 3 secciones. En la primera sección se realiza una introducción y descripción a este documento, además se describe la especificación de recursos del sistema.

En la segunda sección de este documento se realiza una descripción general del sistema, con la finalidad de conocer las funciones que el software debe realizar, también se identifican características como factores, restricciones y supuestos que afectan el desarrollo del sistema sin mucho detalle.

En la tercera sección se definen los requisitos que el sistema debe satisfacer.

## **2. DESCRIPCIÓN DEL SISTEMA**

En esta sección se realizará una descripción de todos los procesos y funcionamiento del sistema, los cuales serán representados por medio de diagramas UML.

### **2.1. PERSPECTIVA DEL PRODUCTO**

El sistema ValidateDocx es un producto desarrollado en la modalidad de trabajo de grado de la línea software inteligente y convergencia tecnológica de la Universidad Católica de Colombia.

Este sistema será un prototipo que tendrá como objetivo inicial, validar documentos de trabajos de grado desarrollados en la universidad ya que estos documentos están basados en la NTC1486 dado que este es el estándar que inicialmente validará el sistema.

Además, la NTC1486 es una norma es utilizada para la presentación de documentos como:

- Trabajo de introducción a la investigación.
- Trabajos de grado.
- Trabajo de investigación profesional.
- Ensayos.
- Monografías.
- Tesis
- Informes científicos.

Como esta norma es indispensable para este tipo de documentos, ValidateDocx no solo está diseñado para que su uso se implemente en instituciones educativas, sino también en lugares como minciencias o semilleros de investigación los cuales son los encargados de aportar conocimiento a la sociedad en que vivimos en día, por esta razón es que se le da importancia esta norma.

Es importante resaltar que además de validar los documentos, el sistema también manejará una administración para usuarios, ya que el sistema puede manejar 3 tipos de usuario con roles diferentes, esto con el fin de que el sistema sea amigable y compatible con los procesos que las instituciones emplean para la calificación de los trabajos de grado.

### **2.2. REQUISITOS FUNCIONALES.**

En esta sección se hará la identificación y especificación de cada uno de los requerimientos funcionales del sistema ValidateDocx.

## Requerimiento Funcional 1: Iniciar Sesión

Tabla 11: Requerimiento funcional: Iniciar Sesión

<b>Identificación del requerimiento</b>	RF01
<b>Nombre del requerimiento</b>	Iniciar Sesión
<b>Descripción del requerimiento</b>	El sistema validará que las credenciales ingresadas por el usuario existan en la base de datos, de lo contrario el usuario deberá crear una cuenta o validar que los datos sean correctos.
<b>Entradas</b>	<ul style="list-style-type: none"><li>• usuario</li><li>• password1</li></ul>
<b>Salidas</b>	<ul style="list-style-type: none"><li>• Usuario validado en B.D.</li><li>• Datos incorrectos en validación de B.D.</li><li>• <b>MENSAJE</b> "USUARIO INEXISTENTE, POR FAVOR VALIDAR O CREAR CUENTA".</li><li>• <b>MENSAJE</b> "CONTRASEÑA INCORRECTA, POR FAVOR VALIDAR O RECUPERAR CONTRASEÑA".</li></ul>

Fuente: El autor.

## Requerimiento funcional 2: Crear cuenta.

Tabla 12: Requerimiento funcional: Crear cuenta.

<b>Identificación del requerimiento</b>	RF02
<b>Nombre del requerimiento</b>	Crear Cuenta
<b>Descripción del requerimiento</b>	El usuario ingresa sus datos personales, para que posteriormente el sistema los almacene en la base de datos y pueda iniciar sesión.
<b>Entradas</b>	<ul style="list-style-type: none"><li>• primerNombre</li><li>• segundoNombre</li><li>• primerApellido</li><li>• segundoApellido</li><li>• email</li><li>• usuario</li><li>• password1</li><li>• password2</li></ul>
<b>Salidas</b>	<ul style="list-style-type: none"><li>• Usuario creado con éxito.</li><li>• <b>Mensaje:</b> "Campo faltante, por favor valide que los datos ingresados estén completos".</li><li>• <b>Mensaje:</b> "Los campos no coinciden por favor verifique la contraseña ingresada".</li></ul>

Fuente: El autor.

### Requerimiento funcional 3: Recuperar contraseña.

Tabla 13: Requerimiento funcional: Recuperar contraseña.

<b>Identificación del requerimiento</b>	RF03
<b>Nombre del requerimiento</b>	Recuperar contraseña
<b>Descripción del requerimiento</b>	El sistema verifica que exista un usuario y un correo electrónico registrados en la base de datos, para reestablecer la contraseña.
<b>Entradas</b>	<ul style="list-style-type: none"> <li>● usuario</li> <li>● email</li> </ul>
<b>salidas</b>	<ul style="list-style-type: none"> <li>● <b>Mensaje:</b> “Se ha enviado un mensaje a su correo electrónico, por favor revise su bandeja de entrada”</li> <li>● <b>Mensaje:</b> “Su contraseña ha sido actualizada satisfactoriamente, por favor regrese al inicio”.</li> <li>● <b>Mensaje:</b> “El usuario no está registrado en la base de datos, por favor verifique de nuevo”.</li> <li>● <b>Mensaje:</b> “El correo electrónico no está registrado en la base de datos, por favor verifique de nuevo.”</li> <li>● <b>Mensaje:</b> “Los Campos de password no coinciden, por favor verifique”</li> </ul>

Fuente: El autor.

### Requerimiento funcional 4: Editar cuentas de usuario.

Tabla 14:Requerimiento funcional: Editar cuentas de usuario.

<b>Identificación del requerimiento</b>	RF04
<b>Nombre del requerimiento</b>	Editar cuentas de usuario
<b>Descripción del requerimiento</b>	El sistema verifica que el tipo de usuario sea administrador, para poder editar la información de cualquier usuario.
<b>Entradas</b>	<ul style="list-style-type: none"> <li>● primerNombre</li> <li>● segundoNombre</li> <li>● primerApellido</li> <li>● segundoApellido</li> <li>● email</li> <li>● usuario</li> <li>● password1</li> <li>● password2</li> <li>● tipoUsuario</li> </ul>



<p><b>Salidas</b></p>	<ul style="list-style-type: none"> <li>● Lista de todos los usuarios que se han registrado en el sistema.</li> <li>● Acceso a todos los campos de información que pertenecen a los usuarios registrados en el sistema.</li> <li>● Asignación de roles a todos los usuarios del sistema.</li> <li>● Actualización de los datos de los usuarios.</li> <li>● <b>Mensaje:</b> “Su usuario no está autorizado para acceder a esta funcionalidad, por favor contacte a un administrador”.</li> <li>● <b>Mensaje:</b> “No hay usuarios registrados en el sistema por el momento”.</li> <li>● <b>Mensaje:</b> “Los datos del usuario han sido modificados con éxito”.</li> </ul>
-----------------------	--

Fuente: El autor.

## Requerimiento funcional 5: Editar administrador

Tabla 15: Requerimiento funcional: Editar Administrador.

<b>Identificación del requerimiento</b>	RF05
<b>Nombre del requerimiento</b>	Editar Administrador
<b>Descripción del requerimiento</b>	El sistema verifica que el tipo de usuario sea administrador y que el perfil que va a editar sea el propio para modificar sus datos.
<b>Entradas</b>	<ul style="list-style-type: none"><li>● primerNombre</li><li>● segundoNombre</li><li>● primerApellido</li><li>● segundoApellido</li><li>● email</li><li>● usuario</li><li>● password1</li><li>● tipoUsuario</li></ul>
<b>Salidas</b>	<ul style="list-style-type: none"><li>● primerNombre</li><li>● segundoNombre</li><li>● primerApellido</li><li>● segundoApellido</li><li>● email</li><li>● usuario</li><li>● password1</li><li>● tipoUsuario</li> <li>● <b>Mensaje:</b> “Su modificación ha sido realizada exitosamente”</li> <li>● <b>Mensaje:</b> “Campos vacíos o incorrectos, por favor verifica.”</li></ul>

## Requerimiento funcional 6: Editar Usuario.

Tabla 16: Requerimiento funcional: Editar Usuario

<b>Identificación del requerimiento</b>	RF06
<b>Nombre del requerimiento</b>	Editar Usuario
<b>Descripción del requerimiento</b>	El sistema verifica que el usuario esté registrado en la base de datos para que pueda acceder a sus datos personales, y le permita modificarlos.
<b>Entradas</b>	<ul style="list-style-type: none"><li>● primerNombre</li><li>● segundoNombre</li><li>● primerApellido</li><li>● segundoApellido</li><li>● email</li><li>● usuario</li><li>● password1</li></ul>
<b>Salidas</b>	<ul style="list-style-type: none"><li>● primerNombre</li><li>● segundoNombre</li><li>● primerApellido</li><li>● segundoApellido</li><li>● email</li><li>● usuario</li><li>● password1</li><li>● <b>Mensaje:</b> "Usuario modificado exitosamente."</li><li>● <b>Mensaje:</b> "Su cuenta de usuario no está autorizada para editar estos datos".</li><li>● <b>Mensaje:</b> "Campos vacíos o incorrectos, por favor verifica."</li></ul>

Fuente: El autor.

## Requerimiento funcional 7: Seleccionar usuario.

Tabla 17: Requerimiento funcional: Seleccionar Usuario.

<b>Identificación del requerimiento</b>	RF07
<b>Nombre del requerimiento</b>	Seleccionar Usuario
<b>Descripción del requerimiento</b>	El sistema verifica que el usuario que ingrese a esta funcionalidad sea coordinador de calidad o administrador para buscar y seleccionar a uno de los usuarios que han utilizado la herramienta.
<b>Entradas</b>	<ul style="list-style-type: none"><li>• usuario</li><li>• rolUsuario</li></ul>
<b>Salidas</b>	<ul style="list-style-type: none"><li>• Listado de usuarios.</li><li>• <b>Mensaje:</b> “Su cuenta de usuario no está autorizada para consultar estos datos”.</li><li>• <b>Mensaje:</b> “No hay usuarios registrados en el sistema en este momento”</li></ul>

Fuente: El autor.

## Requerimiento funcional 8: Consultar Validaciones Realizadas.

Tabla 18: Requerimiento funcional: Consultar Validaciones Realizadas.

<b>Identificación del requerimiento</b>	RF08
<b>Nombre del requerimiento</b>	Consultar Validaciones Realizadas
<b>Descripción del requerimiento</b>	El sistema verifica que el usuario logueado sea el autor de el listado de las validaciones que el sistema mostrará.
<b>Entradas</b>	<ul style="list-style-type: none"><li>• usuario</li></ul>
<b>Salidas</b>	<ul style="list-style-type: none"><li>• Listado de documentos validados y el resultado de la validación.</li><li>• <b>Mensaje</b> “Su cuenta de usuario no está autorizada para editar estos datos”</li></ul>

Fuente: El autor.

## Requerimiento funcional 9: Descargar Validación.

Tabla 19: Requerimiento funcional: Descargar Validación.

<b>Identificación del requerimiento</b>	RF09
<b>Nombre del requerimiento</b>	Descargar Validación.
<b>Descripción del requerimiento</b>	El sistema verifica que el usuario logueado sea el coordinador de calidad o un administrador para descargar el archivo que contiene el resumen de las validaciones realizadas por el sistema a un documento.
<b>Entradas</b>	<ul style="list-style-type: none"><li>• usuario</li><li>• rolUsuario</li><li>• IDvalidación</li></ul>
<b>Salidas</b>	<ul style="list-style-type: none"><li>• Documento PDF con la información del resultado de la validación.</li><li>• <b>Mensaje:</b> "EI USUARIO NO ESTA AUTORIZADO PARA ACCEDER A ESTA FUNCION, POR FAVOR VALIDAR O CREAR CUENTA"</li><li>• <b>Mensaje:</b> "NO HAY ARCHIVOS PARA DESCARGAR EN ESTE MOMENTO"</li></ul>

Fuente: El autor.

## Requerimiento funcional 10: Cargar documento

Tabla 20: Requerimiento funcional: Cargar Documento.

<b>Identificación del requerimiento</b>	RF010
<b>Nombre del requerimiento</b>	Cargar documento
<b>Descripción del requerimiento</b>	El sistema tendrá la opción para que el usuario pueda adjuntar uno o más documentos de formato DOC.
<b>Entradas</b>	<ul style="list-style-type: none"><li>• usuario</li><li>• nombreDocumento</li></ul>
<b>Salidas</b>	<ul style="list-style-type: none"><li>• Documento cargado con éxito.</li><li>• <b>Mensaje:</b> " Este formato de archivo no es permitido para cargar en el sistema, por verifíquelo".</li></ul>

Fuente: El autor.

## Requerimiento funcional 11: Cargar información del trabajo de grado.

Tabla 21: Requerimiento funcional: Cargar información del trabajo de grado.

<b>Identificación del requerimiento</b>	RF011
<b>Nombre del requerimiento</b>	Cargar información del trabajo de grado
<b>Descripción del requerimiento</b>	El sistema almacenará la información relacionada con el documento que el usuario ingresa. para realizar las validaciones pertinentes con el fin de validar los estándares NTC1486 correctamente.
<b>Entradas</b>	<ul style="list-style-type: none"><li>• Título del documento.</li><li>• Autor(es) del documento.</li><li>• Asesor del trabajo de grado.</li><li>• Jurado del trabajo de grado.</li><li>• Nombre del archivo formato DOCX</li></ul>
<b>Salidas</b>	<ul style="list-style-type: none"><li>• Información del documento cargada con éxito.</li><li>• <b>Mensaje:</b> “Su información ha sido almacenada correctamente”.</li><li>• <b>Mensaje:</b> “Por favor valide nuevamente los datos del documento de trabajo de grado.”</li></ul>

Fuente: El autor.

## Requerimiento funcional 12: Obtener información del documento.

Tabla 22: Requerimiento funcional: Obtener información del documento.

<b>Identificación del requerimiento</b>	RF012
<b>Nombre del requerimiento</b>	Obtener información del documento.
<b>Descripción del requerimiento</b>	El sistema realizará una serie de validaciones sobre el archivo de formato DOCX las cuales permiten concluir si el documento cumple o no con la norma NTC1486, además se retorna la lista de características del documento que cumplen y las que no.
<b>Entradas</b>	<ul style="list-style-type: none"><li>● ID del documento.</li></ul>
<b>Salidas</b>	<ul style="list-style-type: none"><li>● Porcentaje de cumplimiento.</li><li>● Items de la norma NTC1486 por corregir.</li><li>● Nombre del usuario que lo cargó.</li><li>● <b>Mensaje:</b> "Su documento cumple con el n% del total de normas establecidas por la NTC1486".</li><li>● <b>MENSAJE:</b> "El documento que quiere validar no existe, por favor contacte al administrador".</li></ul>

Fuente: El autor.

## Obtener historial de documentos.

Tabla 23: Requerimiento funcional: Obtener historial de documentos.

<b>Identificación del requerimiento</b>	RF013
<b>Nombre del requerimiento</b>	Obtener Historial de documentos
<b>Descripción del requerimiento</b>	El coordinador de calidad y el administrador podrán visualizar todos y cada uno de los documentos que se han cargado al sistema.
<b>Entradas</b>	<ul style="list-style-type: none"><li>• usuario</li><li>• rolUsuario</li></ul>
<b>Salidas</b>	<ul style="list-style-type: none"><li>• Listado de todos los documentos que se han cargado.</li><li>• Listado de todos los usuarios que han cargado los documentos.</li><li>• <b>Mensaje:</b> "EI USUARIO NO ESTÄ AUTORIZADO PARA ACCEDER A ESTA FUNCIONALIDAD, POR FAVOR CONTACTE AL ADMINISTRADOR."</li><li>• <b>Mensaje:</b> "Esta lista está vacía o hubo un problema al cargar la información, por favor contacte al administrador"</li></ul>

Fuente: El autor.



## Obtener historial de validaciones.

Tabla 24: Requerimiento funcional: Obtener historial de validaciones.

<b>Identificación del requerimiento</b>	RF014
<b>Nombre del requerimiento</b>	Obtener historial de validaciones
<b>Descripción del requerimiento</b>	El coordinador de calidad y el administrador podrán visualizar todas y cada una de las validaciones realizadas con todos los documentos que se han cargado anteriormente con todos los usuarios que pertenecen al sistema.
<b>Entradas</b>	<ul style="list-style-type: none"><li>• usuario</li><li>• rolUsuario</li></ul>
<b>Salidas</b>	<ul style="list-style-type: none"><li>• Listado de todos porcentajes de los documentos que se han validado con la herramienta.</li><li>• Porcentaje general del nivel de cumplimiento del todos los porcentajes de las validaciones de los documentos que se han cargado al sistema.</li><li>• <b>Mensaje:</b> "EI USUARIO NO ESTÄ AUTORIZADO PARA ACCEDER A ESTA FUNCIONALIDAD, POR FAVOR CONTACTE AL ADMINISTRADOR."</li></ul>

Fuente: El autor.

## 2.1. REQUERIMIENTOS NO FUNCIONALES

### Requisito no funcional 1: Integridad.

Tabla 25: Requerimiento no funcional: Interfaz responsive.

<b>Identificación del requerimiento</b>	RNF01
<b>Nombre del requerimiento</b>	Interfaz Responsive
<b>Descripción del requerimiento</b>	El sistema debe procurar que la interfaz que el usuario utilizará deba tener tecnología responsive ya que esta tecnología permite que el sistema pueda ser visualizado correctamente desde cualquier dispositivo, sin importar el tamaño del a pantalla.

Fuente: El autor.

### Requerimiento no funcional 2: Escalabilidad

Tabla 26: Requerimiento no funcional: Escalabilidad.

<b>Identificación del requerimiento</b>	RNF02
<b>Nombre del requerimiento</b>	Escalabilidad
<b>Descripción del requerimiento</b>	Este sistema deber tener escalabilidad horizontal debido a que este software tiene el potencial de extender sus funcionalidades y demanda por los usuarios basados en el proceso de extracción de metadatos ya que los usos que se le pueden dar son muy diversos.

Fuente: El autor.

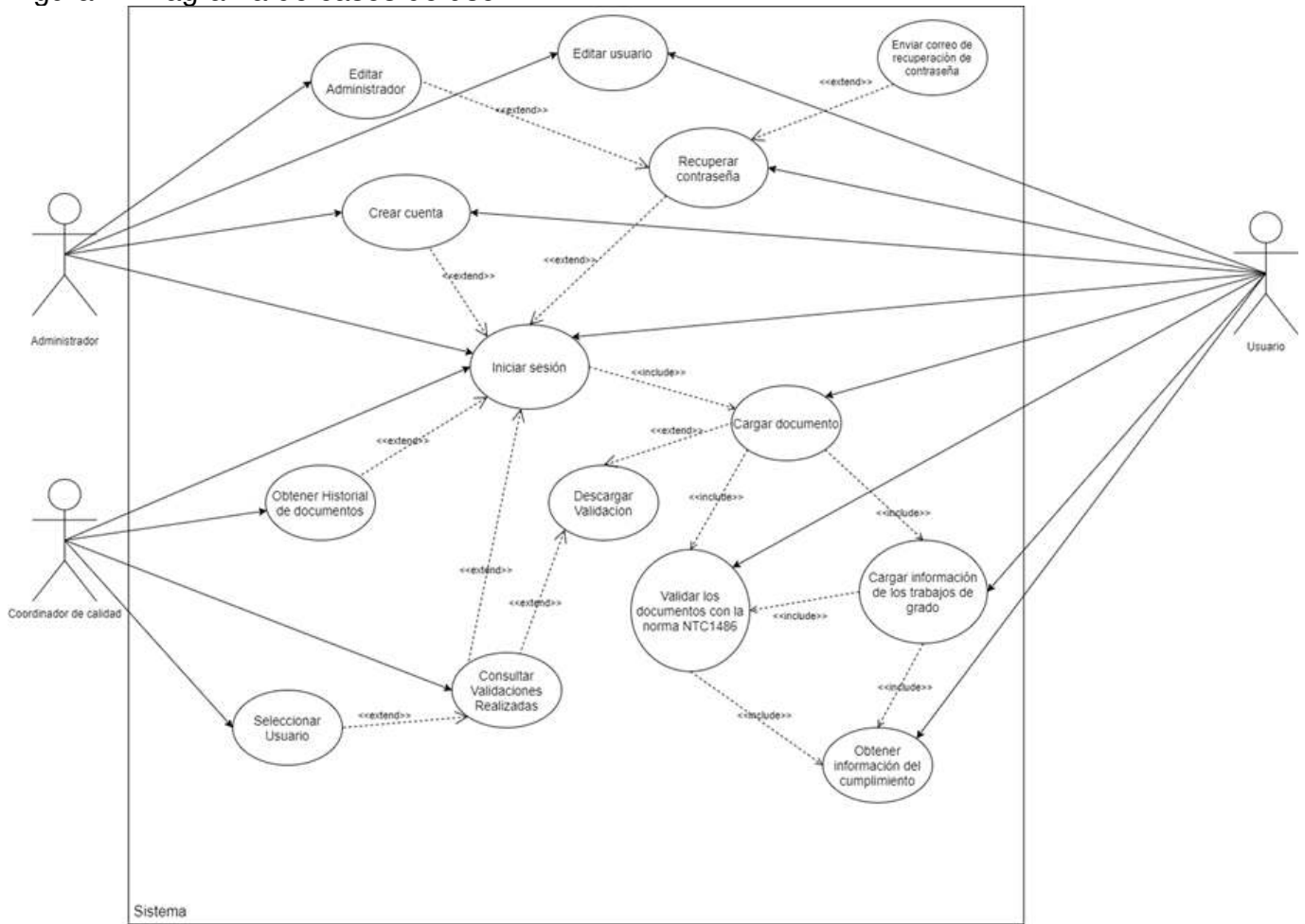
### **2.3. FUNCIONES DEL PRODUCTO.**

El sistema ValidateDocx como primer requisito solicitará credenciales de acceso, si el usuario del sistema no cuenta con este recurso, el sistema ofrecerá la opción de crear una cuenta de acceso, el usuario administrador también tiene la opción de crear cuentas para nuevos usuarios y asignar las respectivas contraseñas, si el usuario por diversas razones olvida su contraseña, el sistema también tendrá una funcionalidad para recuperarla mediante un correo electrónico.

Una vez se haya iniciado sesión, el usuario tiene varias opciones para ejecutar, la primera es cargar uno o varios documentos para utilizar la función de validación la cual verificará que el documento cumpla con los estándares de la NTC1486 o puede acceder al historial de documentos que hayan sido cargados con anterioridad y visualizar el porcentaje de cumplimiento que tuvo cada uno.

A continuación, se muestra el diagrama de casos de uso que corresponde al sistema descrito anteriormente:

Figura 1: Diagrama de casos de uso.



Fuente: el autor.

## Detalle de Casos de Uso.

### 2.3.1.1. Caso de uso 1: Iniciar Sesión

Tabla 27: Caso de uso: Iniciar sesión.

<b>Identificación del caso de uso</b>	Caso de Uso-001
<b>Nombre del caso de uso</b>	Iniciar Sesión
<b>Descripción del caso de uso</b>	El sistema validará que las credenciales ingresadas por el usuario existan en la base de datos, de lo contrario el usuario deberá crear una cuenta o validar que los datos sean correctos.
<b>Entradas</b>	<ul style="list-style-type: none"> <li>● usuario</li> <li>● password</li> </ul>
<b>Salidas</b>	<ul style="list-style-type: none"> <li>● Usuario validado en B.D.</li> <li>● Datos incorrectos en validación de B.D.</li> <li>● <b>MENSAJE</b> “USUARIO INEXISTENTE, POR FAVOR VALIDAR O CREAR CUENTA”</li> <li>● <b>MENSAJE</b> “CONTRASEÑA INCORRECTA, POR FAVOR VALIDAR O RECUPERAR CONTRASEÑA”</li> </ul>
<b>Flujo Normal</b>	<ol style="list-style-type: none"> <li>1. SISTEMA: Despliega textbox para captura de usuario y contraseña</li> <li>2. USUARIO: Ingresa usuario e Ingresa password</li> <li>3. SISTEMA: Consulta la existencia de usuario.</li> <li>4. SISTEMA: Valida correspondencia de usuario y password en la Base de Datos.</li> <li>5. SISTEMA: Despliega página de Sesión de Usuario.</li> </ol>
<b>Flujos Alternos</b>	<ol style="list-style-type: none"> <li>3.1.1. SISTEMA: Valida la inexistencia de usuario</li> <li>3.1.2. SISTEMA: Despliega MENSAJE “USUARIO INEXISTENTE, POR FAVOR VALIDAR O CREAR CUENTA”</li> <li>4.1.1 SISTEMA: Valida la no correspondencia de nombreUsuario y Password en la Base de Datos.</li> <li>4.1.2 SISTEMA: Despliega MENSAJE “CONTRASEÑA INCORRECTA, POR FAVOR VALIDAR O RECUPERAR CONTRASEÑA”</li> </ol>

Fuente: El autor.

### 2.3.1.2. Caso de uso 2: Crear Cuenta

Tabla 28: Caso de uso: Crear cuenta.

<b>Identificación del caso de uso</b>	Caso de Uso-002
<b>Nombre del caso de uso</b>	Crear Cuenta
<b>Descripción del caso de uso</b>	El usuario ingresa sus datos personales, para que posteriormente el sistema los almacene en la base de datos y pueda iniciar sesión.
<b>Entradas</b>	<ul style="list-style-type: none"> <li>● primerNombre</li> <li>● segundoNombre</li> <li>● primerApellido</li> <li>● segundoApellido</li> <li>● email</li> <li>● usuario</li> <li>● password1</li> <li>● password2</li> </ul>
<b>Salidas</b>	<ul style="list-style-type: none"> <li>● Usuario creado con éxito</li> <li>● <b>Mensaje:</b> “Campo faltante, por favor valide que los datos ingresados estén completos”</li> <li>● <b>Mensaje:</b> “Los campos no coinciden por favor verifique la contraseña ingresada”</li> </ul>
<b>Flujo Normal</b>	<ol style="list-style-type: none"> <li>1. SISTEMA: Despliega textbox para captura de: <ul style="list-style-type: none"> <li>● primerNombre</li> <li>● segundoNombre</li> <li>● primerApellido</li> <li>● segundoApellido</li> <li>● email</li> <li>● usuario</li> <li>● password1</li> <li>● password2</li> </ul> </li> <li>2. USUARIO: Ingresa primerNombre, segundoNombre, primerApellido, SegundoApellido, email, usuario, password1, password2.</li> <li>3. SISTEMA: Verifica que los campos password1 y password2, tengan la misma información.</li> <li>4. SISTEMA: Almacena los datos ingresados por el usuario en la base de datos y redirige al usuario al inicio de sesión.</li> </ol>

<b>Flujos Alternos</b>	<p>3.1.1. SISTEMA: Verifica que los campos password1 y password2 no coinciden con la misma información.</p> <p>3.1.2. SISTEMA: Despliega MENSAJE “Los campos no coinciden por favor verifique la contraseña ingresada”.</p> <p>4.1.1. SISTEMA: Valida la inexistencia de primerNombre, primerApellido, email, usuario y password solicitados al usuario.</p> <p>4.1.2. SISTEMA: Despliega MENSAJE “Campo faltante, por favor valide que los datos ingresados estén completos”</p>
------------------------	---

Fuente: El autor.

### 2.3.1.3. Caso de uso 3: Recuperar contraseña

Tabla 29: Caso de uso: Recuperar contraseña.

<b>Identificación del caso de uso</b>	Caso de Uso-003
<b>Nombre del caso de uso</b>	Recuperar contraseña
<b>Descripción del caso de uso</b>	El sistema verifica que exista un usuario y un correo electrónico registrados en la base de datos, para reestablecer la contraseña.
<b>Entradas</b>	<ul style="list-style-type: none"> <li>● usuario</li> <li>● email</li> </ul>
<b>Salidas</b>	<ul style="list-style-type: none"> <li>● <b>Mensaje:</b> “Se ha enviado un mensaje a su correo electrónico, por favor revise su bandeja de entrada”</li> <li>● <b>Mensaje:</b> “Su contraseña ha sido actualizada satisfactoriamente, por favor regrese al inicio”.</li> <li>● <b>Mensaje:</b> “El usuario no está registrado en la base de datos, por favor verifique de nuevo”.</li> <li>● <b>Mensaje:</b> “El correo electrónico no está registrado en la base de datos, por favor verifique de nuevo.”</li> <li>● <b>Mensaje:</b> “Los Campos de password no coinciden, por favor verifique”</li> </ul>
<b>Flujo Normal</b>	<ol style="list-style-type: none"> <li>1. SISTEMA: Despliega textbox para captura de correo email.</li> <li>2. USUARIO: Ingresa usuario.</li> <li>3. USUARIO: Ingresa email.</li> </ol>

	<p>4. SISTEMA: Valida la correspondencia del usuario en la base de datos.</p> <p>5. SISTEMA: Valida correspondencia de email en la Base de Datos.</p> <p>6. SISTEMA: El sistema ejecuta una función que utiliza un protocolo SMTP para enviar un mensaje al correo electrónico del usuario que contiene un link que le permite ingresar una nueva contraseña.</p> <p>7. SISTEMA: Despliega MENSAJE: “Se ha enviado un mensaje a su correo electrónico, por favor revise su bandeja de entrada”</p> <p>8. SISTEMA: Despliega textbox para captura de password1 y password2.</p> <p>9. SISTEMA: Modifica el registro de la contraseña anterior, con la nueva contraseña que el usuario digita.</p> <p>10. SISTEMA: Despliega Mensaje: “Su contraseña ha sido actualizada satisfactoriamente, por favor regrese al inicio”.</p>
<p><b>Flujos Alternos</b></p>	<p>4.1.1. SISTEMA: Valida la inexistencia de usuario en la base de datos.</p> <p>4.1.2. SISTEMA: Despliega MENSAJE “El usuario no está registrado en la base de datos, por favor verifique de nuevo.”</p> <p>5.1.1. SISTEMA: Valida la inexistencia de email en la base de datos.</p> <p>5.1.2. SISTEMA: Despliega MENSAJE “El correo electrónico no está registrado en la base de datos, por favor verifique de nuevo.”</p> <p>8.1.1. SISTEMA: Valida que los campos password1 y password2 no coinciden.</p> <p>8.1.1 SISTEMA: Despliega Mensaje “Los Campos de password no coinciden, por favor verifique”</p>

Fuente: El autor.



### 2.3.1.4. Caso de uso 4: Editar cuentas de usuario

Tabla 30: Caso de uso: Editar cuentas de usuario.

<b>Identificación del caso de uso</b>	Caso de Uso-004
<b>Nombre del caso de uso</b>	Editar cuentas de usuario
<b>Descripción del caso de uso</b>	El sistema verifica que el tipo de usuario sea administrador para poder editar la información de cualquier usuario.
<b>Entradas</b>	<ul style="list-style-type: none"> <li>● primerNombre</li> <li>● segundoNombre</li> <li>● primerApellido</li> <li>● segundoApellido</li> <li>● email</li> <li>● usuario</li> <li>● password1</li> <li>● tipoUsuario</li> </ul>
<b>Salidas</b>	<ul style="list-style-type: none"> <li>● Lista de todos los usuarios que se han registrado en el sistema.</li> <li>● Acceso a todos los campos de información que pertenecen a los usuarios registrados en el sistema.</li> <li>● Asignación de roles a todos los usuarios del sistema.</li> <li>● Actualización de los datos de los usuarios.</li> <li>● <b>Mensaje:</b> “Su usuario no está autorizado para acceder a esta funcionalidad, por favor contacte a un administrador”.</li> <li>● <b>Mensaje:</b> “No hay usuarios registrados en el sistema por el momento”.</li> <li>● <b>Mensaje:</b> “Los datos del usuario han sido modificados con éxito”.</li> </ul>
<b>Flujo Normal</b>	<ol style="list-style-type: none"> <li>1. SISTEMA: Despliega textbox para captura de: <ul style="list-style-type: none"> <li>● primerNombre</li> <li>● segundoNombre</li> <li>● primerApellido</li> <li>● segundoApellido</li> <li>● email</li> <li>● usuario</li> <li>● password1</li> <li>● password2</li> <li>● tipoUsuario</li> </ul> </li> </ol>

	<p>2. USUARIO: Ingresa primerNombre, segundoNombre, primerApellido1, SegundoApellido2, email, usuario, password1, password2, tipoUsuario.</p> <p>3. SISTEMA: Valida que el perfil del usuario sea correspondiente al tipo administrador</p> <p>4. SISTEMA: Despliega página de editar los datos de los usuarios.</p>
<b>Flujos Alternos</b>	<p>3.1.1. SISTEMA: Valida la no correspondencia del tipo de perfil del usuario de tipo administrador.</p> <p>3.1.2 SISTEMA: Despliega MENSAJE “Su usuario no está autorizado para acceder a esta funcionalidad, por favor contacte a un administrador”.</p> <p>4.1.1. SISTEMA: Valida la inexistencia de usuarios registrados en el sistema.</p> <p>4.1.1 SISTEMA: Despliega Mensaje “No hay usuarios registrados en el sistema por el momento”.</p>

Fuente: El autor.

### 2.3.1.5. Caso de uso 5: Editar Administrador.

Tabla 31: Caso de uso: Editar administrador.

<b>Identificación del caso de uso</b>	Caso de Uso-005
<b>Nombre del caso de uso</b>	Editar Administrador
<b>Descripción del caso de uso</b>	El sistema verifica que el tipo de usuario sea administrador y que el perfil que va a editar sea el propio para modificar sus datos.
<b>Entradas</b>	<ul style="list-style-type: none"> <li>● primerNombre</li> <li>● segundoNombre</li> <li>● primerApellido</li> <li>● segundoApellido</li> <li>● email</li> <li>● usuario</li> <li>● password1</li> </ul>
<b>Salidas</b>	<ul style="list-style-type: none"> <li>● primerNombre</li> <li>● segundoNombre</li> <li>● primerApellido</li> <li>● segundoApellido</li> </ul>

	<ul style="list-style-type: none"> <li>● email</li> <li>● usuario</li> <li>● password1</li> <li>● <b>Mensaje:</b> “Su modificación ha sido realizada exitosamente”</li> <li>● <b>Mensaje:</b> “Campos vacíos o incorrectos, por favor verifica.”</li> </ul>
<b>Flujo Normal</b>	<ol style="list-style-type: none"> <li>1. SISTEMA: Valida que el usuario sea el administrador del sistema y que la información a editar sea la de su propio perfil.</li> <li>2. SISTEMA: Despliega textbox para captura de primerNombre, segundoNombre, primerApellido, segundoApellido, email, usuario, password1.</li> <li>3. USUARIO: Puede modificar uno o todos los siguientes campos: primerNombre, segundoNombre, primerApellido, segundoApellido, email, usuario, password1</li> <li>4. SISTEMA: Modifica todos los campos en la Base de datos.</li> <li>5. SISTEMA: Despliega Mensaje: “Su modificación ha sido realizada exitosamente”</li> </ol>
<b>Flujos Alternos</b>	<ol style="list-style-type: none"> <li>1.1.1. SISTEMA: Valida que el usuario no es administrador y el perfil a editar no es el propio.</li> <li>1.1.2. SISTEMA: Despliega Mensaje: “Su usuario no está autorizado para acceder a esta funcionalidad, por favor contacte a un administrador”</li> <li>3.1.1. SISTEMA: Valida que existen campos que están vacíos o incorrectos.</li> <li>3.1.2. SISTEMA: Despliega MENSAJE “Campos vacíos o incorrectos, por favor verifica.”</li> </ol>

Fuente: El autor.

### 2.3.1.6. Caso de uso 6: Editar Usuario.

Tabla 32: Caso de uso: Editar usuario.

<b>Identificación del caso de uso</b>	Caso de Uso-006
<b>Nombre del caso de uso</b>	Editar Usuario
<b>Descripción del caso de uso</b>	El sistema verifica que el usuario esté registrado en la base de datos para acceder a sus datos personales y le permite modificarlos.
<b>Entradas</b>	<ul style="list-style-type: none"> <li>● primerNombre</li> <li>● segundoNombre</li> <li>● primerApellido</li> <li>● segundoApellido</li> <li>● email</li> <li>● usuario</li> <li>● password1</li> </ul>
<b>Salidas</b>	<ul style="list-style-type: none"> <li>● primerNombre</li> <li>● segundoNombre</li> <li>● primerApellido</li> <li>● segundoApellido</li> <li>● email</li> <li>● usuario</li> <li>● password1</li> <li>● <b>Mensaje:</b> "Usuario modificado exitosamente."</li> <li>● <b>Mensaje:</b> "Su cuenta de usuario no está autorizada para editar estos datos"</li> <li>● <b>Mensaje:</b> "Campos vacíos o incorrectos, por favor verifica."</li> </ul>
<b>Flujo Normal</b>	<ol style="list-style-type: none"> <li>1. SISTEMA: Valida que el usuario sea el propietario de los datos que va a modificar.</li> <li>2. SISTEMA: Despliega textbox para captura de primerNombre, segundoNombre, primerApellido, segundoApellido, email, usuario, password1.</li> <li>3. USUARIO: Puede modificar uno o todos los siguientes campos primerNombre, segundoNombre, primerApellido, segundoApellido, email, usuario, password1.</li> <li>4. SISTEMA: Modifica todos los campos en la Base de datos.</li> <li>5. SISTEMA: Despliega Mensaje: "Usuario modificado exitosamente."</li> </ol>

<b>Flujos Alternos</b>	<p>1.1.1. SISTEMA: Valida que el usuario no está autorizado a modificar los datos.</p> <p>1.1.2. SISTEMA: Despliega Mensaje:” Su cuenta de usuario no está autorizada para editar estos datos”.</p> <p>3.1.1. SISTEMA: Valida que existen campos que están vacíos o incorrectos.</p> <p>3.1.2. SISTEMA: Despliega MENSAJE “Campos vacíos o incorrectos, por favor verifica.”</p>
------------------------	--

Fuente: El autor.

### 2.3.1.7. Caso de uso 7: Seleccionar Usuario.

Tabla 33: Caso de uso: Seleccionar usuario.

<b>Identificación caso de uso</b>	Caso de Uso-007
<b>Nombre del caso de uso</b>	Seleccionar Usuario
<b>Descripción del caso de uso</b>	El sistema verifica que el usuario que ingrese a esta funcionalidad sea coordinador de calidad o administrador para buscar y seleccionar a uno de los usuarios que han utilizado la herramienta.
<b>Entradas</b>	<ul style="list-style-type: none"> <li>● usuario</li> <li>● rolUsuario</li> </ul>
<b>Salidas</b>	<ul style="list-style-type: none"> <li>● Listado de usuarios.</li> <li>● <b>Mensaje:</b> “Su cuenta de usuario no está autorizada para consultar estos datos”.</li> <li>● <b>Mensaje:</b> “No hay usuarios registrados en el sistema en este momento”</li> </ul>
<b>Flujo Normal</b>	<ol style="list-style-type: none"> <li>1. SISTEMA: Valida que el usuario sea el coordinador de calidad o administrador del sistema.</li> <li>2. SISTEMA: Despliega página de listado de usuarios.</li> </ol>
<b>Flujos Alternos</b>	<p>1.1.1. SISTEMA: Valida que el usuario no está autorizado a modificar los datos.</p>

	<p>1.1.2. SISTEMA: Despliega Mensaje:” Su cuenta de usuario no está autorizada para editar estos datos”.</p> <p>2.1.1. SISTEMA: Despliega Mensaje” No hay usuarios registrados en el sistema en este momento”</p>
--	---

Fuente: El autor.

### 2.3.1.8. Caso de uso 8: Consultar Validaciones Realizadas.

Tabla 34: Caso de uso: Consultar validaciones realizadas.

<b>Identificación del caso de uso</b>	Caso de Uso-008
<b>Nombre del caso de uso</b>	Consultar Validaciones Realizadas
<b>Descripción del caso de uso</b>	El sistema verifica que el usuario logueado sea el autor de el listado de las validaciones que el sistema mostrará.
<b>Entradas</b>	<ul style="list-style-type: none"> <li>• usuario</li> </ul>
<b>Salidas</b>	<ul style="list-style-type: none"> <li>• Listado de documentos validados y el resultado de la validación.</li> <li>• <b>Mensaje</b> “Su cuenta de usuario no está autorizada para editar estos datos”</li> </ul>
<b>Flujo Normal</b>	<ol style="list-style-type: none"> <li>1. SISTEMA: Despliega textbox para captura de usuario.</li> <li>2. SISTEMA: Valida correspondencia de usuario en la Base de Datos.</li> <li>3. SISTEMA: Despliega página del listado de documentos validados, con el nombre de los usuarios que los cargó.</li> </ol>
<b>Flujos Alternos</b>	<p>2.1.1. SISTEMA: Valida la inexistencia de usuario.</p> <p>2.1.2. SISTEMA: Despliega MENSAJE “El USUARIO INEXISTENTE, POR FAVOR VALIDAR O CREAR CUENTA”.</p> <p>3.1.1. SISTEMA: Despliega MENSAJE “No hay historial de validaciones realizadas”</p>

Fuente: El autor

### 2.3.1.9. Caso de uso 9: Descargar Validación.

Tabla 35: Caso de uso: Descargar validación.

<b>Identificación del caso de uso</b>	Caso de Uso-009
<b>Nombre del caso de uso</b>	Descargar Validación.
<b>Descripción del caso de uso</b>	El sistema verifica que el usuario logueado sea el coordinador de calidad o un administrador para descargar el archivo que contiene el resumen de las validaciones realizadas por el sistema a un documento.
<b>Entradas</b>	<ul style="list-style-type: none"> <li>• usuario</li> <li>• rolUsuario</li> <li>• IDvalidación</li> </ul>
<b>Salidas</b>	<ul style="list-style-type: none"> <li>• Documento PDF con la información del resultado de la validación.</li> <li>• <b>Mensaje:</b> “EI USUARIO NO ESTA AUTORIZADO PARA ACCEDER A ESTA FUNCION, POR FAVOR VALIDAR O CREAR CUENTA”</li> <li>• <b>Mensaje:</b> “NO HAY ARCHIVOS PARA DESCARGAR EN ESTE MOMENTO”</li> </ul>
<b>Flujo Normal</b>	<ol style="list-style-type: none"> <li>1. SISTEMA: Consulta que el tipo del usuario sea coordinador de calidad o administrador del sistema.</li> <li>2. SISTEMA: Despliega página del listado de documentos validados, con el nombre de los usuarios que los cargó con la opción de descarga.</li> <li>3. SISTEMA: Exporta archivo PDF con la información de los resultados de la validación.</li> </ol>
<b>Flujos Alternos</b>	<p>1.1.1. SISTEMA: Consulta que el tipo de usuario no es el administrador o el coordinador de calidad.</p> <p>1.1.2. SISTEMA: Despliega MENSAJE “EI USUARIO NO ESTA AUTORIZADO PARA ACCEDER A ESTA FUNCION, POR FAVOR VALIDAR O CREAR CUENTA”.</p> <p>2.1.1. SISTEMA: Despliega Mensaje “No hay archivos para descargar en este momento”</p>

Fuente: El autor.

### 2.3.1.10. Caso de uso 10: Cargar documento.

Tabla 36: Caso de uso: Cargar documento.

<b>Identificación del caso de uso</b>	Caso de Uso-010
<b>Nombre del caso de uso</b>	Cargar documento
<b>Descripción del caso de uso</b>	El sistema tendrá la opción para que el usuario pueda adjuntar uno o más documentos de formato DOC.
<b>Entradas</b>	<ul style="list-style-type: none"> <li>● usuario</li> <li>● nombreDocumento</li> </ul>
<b>Salidas</b>	<ul style="list-style-type: none"> <li>● Documento cargado con éxito.</li> <li>● <b>Mensaje:</b> " Este formato de archivo no es permitido para cargar en el sistema, por verifíquelo".</li> </ul>
<b>Flujo Normal</b>	<ol style="list-style-type: none"> <li>1. USUARIO: Adjunta el archivo de Word en el sistema.</li> <li>2. SISTEMA: Valida que el formato del archivo es .DOCX.</li> <li>3. SISTEMA: Almacena usuario y nombreDocumento en la Base de Datos</li> <li>4. SISTEMA: Carga el archivo de formato DOCX y asocia el documento al usuario, al nombreDocumento y le asigna un ID al archivo.</li> </ol>
<b>Flujos Alternos</b>	<ol style="list-style-type: none"> <li>2.1.1. SISTEMA: Valida que el formato del archivo no es .DOCX</li> <li>2.1.2. SISTEMA: Despliega Mensaje" Este formato de archivo no es permitido para cargar en el sistema, por verifíquelo".</li> </ol>

Fuente: El autor.



### 2.3.1.11. Caso de uso 11: Cargar información del trabajo de grado.

Tabla 37: Caso de uso: Cargar información del trabajo de grado.

<b>Identificación del caso de uso</b>	Caso de Uso-011
<b>Nombre del caso de uso</b>	Cargar información del trabajo de grado
<b>Descripción del caso de uso</b>	El sistema almacenará la información relacionada con el documento que el usuario carga al sistema, para realizar las validaciones pertinentes con el fin de verificar los estándares NTC1486 correctamente.
<b>Entradas</b>	<ul style="list-style-type: none"> <li>● tituloDocumento.</li> <li>● autor1</li> <li>● autor2.</li> <li>● nombreArchivo.</li> </ul>
<b>Salidas</b>	<ul style="list-style-type: none"> <li>● <b>Mensaje:</b> “LOS DATOS DE SU DOCUMENTO HAN SIDO CARGADOS CON EXITO”.</li> <li>● <b>Mensaje:</b> “Por favor valide nuevamente los datos del documento de trabajo de grado.”</li> </ul>
<b>Flujo Normal</b>	<ol style="list-style-type: none"> <li>1. USUARIO: Carga la información que el sistema solicita para asociarla a un documento que se cargará al servidor</li> <li>2. SISTEMA: Almacena, tituloDocumento, autor1, autor2, nombreArchivo en la base de datos.</li> <li>3. SISTEMA: Despliega Mensaje: “EL DOCUMENTO HA SIDO CARGADO CON ÉXITO”.</li> </ol>
<b>Flujos Alternos</b>	<ol style="list-style-type: none"> <li>1.1.1. SISTEMA: Valida que todos los datos no están correctamente diligenciados.</li> <li>1.1.2. SISTEMA: Despliega MENSAJE “Por favor valide nuevamente los datos del documento de trabajo de grado.”</li> </ol>

Fuente: El autor.

### 2.3.1.12. Caso de uso 12: Obtener información del documento.

Tabla 38: Caso de uso: Obtener información del documento.

<b>Identificación del caso de uso</b>	Caso de Uso-012
<b>Nombre del caso de uso</b>	Obtener información del documento.
<b>Descripción del caso de uso</b>	El sistema realizará una serie de validaciones sobre el archivo de formato DOCX las cuales permiten concluir si el documento cumple o no con la norma NTC1486, además se retorna la lista de características del documento que cumplen y las que no.
<b>Entradas</b>	<ul style="list-style-type: none"> <li>● ID del documento.</li> </ul>
<b>Salidas</b>	<ul style="list-style-type: none"> <li>● Porcentaje de cumplimiento.</li> <li>● Ítems de la norma NTC1486 por corregir.</li> <li>● Nombre del usuario que lo cargó.</li> <li>● <b>Mensaje:</b> "Su documento cumple con el n% del total de normas establecidas por la NTC1486".</li> <li>● <b>MENSAJE:</b> "El documento que quiere validar no existe, por favor contacte al administrador".</li> </ul>
<b>Flujo Normal</b>	<ol style="list-style-type: none"> <li>1. SISTEMA: Identifica el ID del documento que se encuentra almacenado en la base de datos, el cual está asociado a un documento previamente cargado.</li> <li>2. SISTEMA: Realiza un proceso de ingeniería inversa el cual permite extraer los metadatos del archivo de trabajo de grado.</li> <li>3. SISTEMA: Accede a los metadatos extraídos del documento y comienza el proceso de validación de los metadatos que contienen la información que ayuda a validar si el documento realmente está cumpliendo con las normas NTC1486.</li> <li>4. SISTEMA: Calcula un porcentaje con el número de ítems validados en el documento versus el número de ítems que faltan por cumplir.</li> <li>5. SISTEMA: Almacena en la base de datos la información resultante de la validación realizada al documento.</li> </ol>

	6. SISTEMA despliega MENSAJE: “Su documento cumple con el n% del total de normas establecidas por la NTC1486”.
<b>Flujos Alternos</b>	1.1.1. SISTEMA: Valida la inexistencia del ID del documento a validar.  1.1.2. SISTEMA despliega MENSAJE: “El documento que quiere validar no existe, por favor contacte al administrador”.

Fuente: El autor.

### 2.3.1.13. Caso de uso 13: Obtener historial de documentos.

Tabla 39: Caso de uso: Obtener historial de documentos.

<b>Identificación del caso de uso</b>	Caso de Uso-013
<b>Nombre del caso de uso</b>	Obtener Historial de documentos
<b>Descripción del caso de uso</b>	El coordinador de calidad y el administrador podrán visualizar todos y cada uno de los documentos que se han cargado al sistema.
<b>Entradas</b>	<ul style="list-style-type: none"> <li>● usuario</li> <li>● rolUsuario</li> </ul>
<b>Salidas</b>	<ul style="list-style-type: none"> <li>● Listado de todos los documentos que se han cargado.</li> <li>● Listado de todos los usuarios que han cargado los documentos.</li> <li>● <b>Mensaje:</b> “EI USUARIO NO ESTÄ AUTORIZADO PARA ACCEDER A ESTA FUNCIONALIDAD, POR FAVOR CONTACTE AL ADMINISTRADOR.”</li> <li>● <b>Mensaje:</b> ”Esta lista está vacía o hubo un problema al cargar la información, por favor contacte al administrador”</li> </ul>
<b>Flujo Normal</b>	<ol style="list-style-type: none"> <li>1. SISTEMA: Valida que el rol del usuario sea de un administrador o un coordinador de calidad.</li> <li>2. SISTEMA: despliega la lista de documentos cargados por el usuario.</li> </ol>
<b>Flujos Alternos</b>	1.1.1. SISTEMA: Valida que el rol del usuario no está autorizado para acceder a esa funcionalidad.

	<p>1.1.2. SISTEMA: Despliega MENSAJE “EI USUARIO NO ESTÄ AUTORIZADO PARA ACCEDER A ESTA FUNCIONALIDAD, POR FAVOR CONTACTE AL ADMINISTRADOR.”</p> <p>5.1.1. SISTEMA: Valida que la lista de documentos no existe o no se puede mostrar.</p> <p>5.1.2. Sistema : Despliega MENSAJE: ”Esta lista está vacía o hubo un problema al cargar la información, por favor contacte al administrador”</p>
--	--

Fuente: El autor.

### 2.3.1.14. Caso de uso 14: Obtener historial de validaciones.

Tabla 40: Caso de uso: Obtener historial de validaciones.

<b>Identificación del caso de uso</b>	Caso de Uso-014
<b>Nombre del caso de uso</b>	Obtener historial de validaciones
<b>Descripción del caso de uso</b>	El coordinador de calidad y el administrador podrán visualizar todas y cada una de las validaciones realizadas con todos los documentos que se han cargado anteriormente con todos los usuarios que pertenecen al sistema.
<b>Entradas</b>	<ul style="list-style-type: none"> <li>• usuario</li> <li>• rolUsuario</li> </ul>
<b>Salidas</b>	<ul style="list-style-type: none"> <li>• Listado de todos porcentajes de los documentos que se han validado con la herramienta.</li> <li>• Porcentaje general del nivel de cumplimiento del todos los porcentajes de las validaciones de los documentos que se han cargado al sistema.</li> <li>• <b>Mensaje:</b> “EI USUARIO NO ESTÄ AUTORIZADO PARA ACCEDER A ESTA FUNCIONALIDAD, POR FAVOR CONTACTE AL ADMINISTRADOR.”</li> </ul>
<b>Flujo Normal</b>	<ol style="list-style-type: none"> <li>1. SISTEMA: Valida que el rol del usuario sea de un administrador o un coordinador de calidad.</li> <li>2. SISTEMA: despliega la lista de los porcentajes de las validaciones.</li> </ol>
<b>Flujos Alternos</b>	1.1.1. SISTEMA: Valida que el rol del usuario no está autorizado para acceder a esa funcionalidad.

	<p>1.1.2. SISTEMA: Despliega MENSAJE “EI USUARIO NO ESTÁ AUTORIZADO PARA ACCEDER A ESTA FUNCIONALIDAD, POR FAVOR CONTACTE AL ADMINISTRADOR.”</p> <p>5.1.1. SISTEMA: Valida que la lista de validaciones está vacía.</p> <p>5.1.2. Sistema : Despliega MENSAJE: ”Esta lista está vacía.</p>
--	--

Fuente: El autor.

## 2.4. CARACTERÍSTICAS DE LOS USAURIOS.

El sistema cuenta con tres usuarios el cual puede ser un docente o también un estudiante, el coordinador de trabajos de grado y un administrador, ya que para el caso del docente la herramienta le servirá para optimizar tanto el tiempo que se gasta haciendo una revisión como en la calidad del contenido que se está revisando, en el caso del estudiante, esta herramienta le es muy útil en el momento de verificar que todo el trabajo que ha realizado cumple con un listado de mínimas condiciones para que el documento pueda ser aceptado, al coordinador de trabajos de grado le servirá para llevar un control más exhaustivo del proceso que llevan los documentos para que sean aceptados y por último, el administrador puede controlar todo el sistema por si se llega a dar alguna contingencia con algún usuario, el tendrá las facultades para solucionar lo que surja.

Tabla 41: Tipo de usuario: Administrador.

<b>Tipo de Usuario</b>	<b>Funciones</b>
Administrador	El administrador tiene acceso a todas las funcionalidades del sistema, además puede crear usuarios para más personas que requieran credenciales para utilizar el software o cambiar sus respectivas contraseñas.

Fuente: El autor.

Tabla 42: Tipo de usuario: Coordinador de calidad.

<b>Tipo de Usuario</b>	<b>Funciones</b>
Coordinador de calidad	Es el usuario que le hace seguimiento a todos los documentos de Trabajos de Grado.

Fuente: El autor.

Tabla 43: Tipo de usuario: Usuario

<b>Tipo de Usuario</b>	<b>Funciones</b>
Usuario	Este perfil podrá ser utilizado por cualquier persona que necesite acceder tanto a la herramienta de validación, como a los resultados relacionados con los documentos

Fuente: El autor.

## 2.5. RESTRICCIONES.

- Para el desarrollo de este proyecto únicamente se utilizarán los documentos que están en el repositorio de la Universidad Católica de Colombia.
- Para probar el funcionamiento del software, solo se utilizará en un computador local.

## 2.6. SUPOSICIONES Y DEPENDENCIAS.

- Se asume que las características utilizadas para validar el porcentaje de cumplimiento de un documento de Word con las normas NT1486 no cambian.

### 3. REQUISITOS ESPECÍFICOS

#### 3.1. INTERFACES EXTERNAS.

**3.1.1. Interfaces de usuario.** La interfaz con el usuario debe permitir realizar la carga de los archivos que necesite verificar con el sistema, además de que debe permitir la visualización de las validaciones que se den como resultado posterior al análisis realizado.

#### **Interfaces de software.**

- Sistema Operativo: Windows 7 o superior.
- Explorador: Google Chrome, Mozilla, Internet Explorer.



**UNIVERSIDAD CATÓLICA**  
de Colombia

**ANEXO B**  
**DOCUMENTO DE ARQUITECTURA DEL SOFTWARE**

**JULIAN FELIPE PLATA ZUÑIGA**

**UNIVERSIDAD CATÓLICA DE COLOMBIA**  
**FACULTAD DE INGENIERÍA**  
**PROGRAMA DE INGENIERÍA DE SISTEMAS Y COMPUTACIÓN**  
**BOGOTÁ**  
**2020**



## Contenido

1.	INTRODUCCIÓN: .....	24
2.	STAKEHOLDERS .....	25
3.	REPRESENTACIÓN ARQUITECTÓNICA Y PUNTOS DE VISTA .....	26
3.1.	VISTA FÍSICA .....	26
3.1.1.	Arquitectura del sistema. ....	26
3.1.2.	Diagrama de Despliegue: .....	26
3.2.	VISTA LÓGICA: .....	27
3.2.1.	Diagrama de Componentes:.....	27
3.2.2.	Diagrama de Clases .....	28
3.2.3.	Modelo de datos .....	1

### **Tabla de imágenes.**

Figura 1: Arquitectura del sistema ValidateDocx.....	58
Figura 2: Diagrama de despliegue del sistema .....	27
Figura 3: Diagrama de componentes del sistema. ....	28
Figura 4: Diagrama de clases.....	29
Figura 5: Modelo Entidad Relación del sistema. ....	1

## Lista de tablas

Tabla 1: Stakeholdes del sistema.....	25
---------------------------------------	----

## 1. INTRODUCCIÓN:

Este documento pretende mostrar de forma detallada y explícita, el diseño del software(ValidateDocx) para la validación de documentos de Word bajo la norma NTC1486 para documentos de tesis, trabajos de grado y trabajos de investigación en el territorio colombiano, esto permite proveer un conocimiento de toda la arquitectura del sistema, ya que este entregable es una pieza fundamental para su desarrollo.

ValidateDocx es una herramienta que implementa un proceso de ingeniería inversa con documentos de Word con el fin de extraer los metadatos que contienen estos de archivos. Este proceso de extracción le permite al sistema acceder a todas sus características y contenidos, y así poder validar a partir de esta información si el documento cumple con los estándares de la norma NTC1486, para llevar a cabo el desarrollo de este sistema, se pretende exponer a los usuarios, el diseño arquitectónico del software con notación UML como, por ejemplo: Diagrama de componentes, diagrama de despliegue, diagrama de clases. Estos diagramas nos permiten comprender de la mejor forma posible toda la composición del sistema.

## 2. STAKEHOLDERS

A continuación, se muestran las personas o Stakeholders interesados en el sistema ValidateDocx.

Tabla 44: Stakeholders del sistema.

<b>STAKEHOLDERS</b>	<b>Descripción</b>
Administrador	Este actor es la persona encargada principalmente de administrar las cuentas de usuario de las personas que se registren en el sistema, esta persona es el único usuario que puede asignarle permisos de coordinador de calidad a un usuario.
Coordinador de calidad	Este actor es el encargado de supervisar y controlar las versiones de los documentos que se han cargado al sistema, junto con el resultado de las validaciones que se han realizado. Todo esto con el fin de tener un control detallado de la calidad con la que los documentos tienen en un principio.
Usuario	Este actor es el encargado de cargar los documentos al sistema e ingresar la información relacionada con el documento de investigación, a diferencia del coordinador de calidad, este usuario solo puede ver el resultado de la validación de su propio documento.

### 3. REPRESENTACIÓN ARQUITECTÓNICA Y PUNTOS DE VISTA

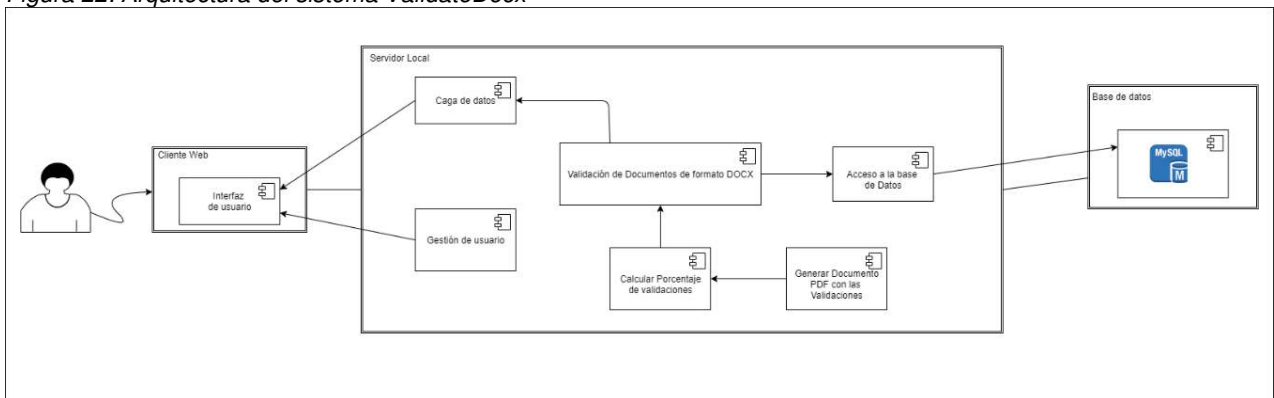
El presente documento mostrará una visión arquitectural del software ValidateDocx a través de 2 vistas diferentes: Vista Física y Vista Lógica del sistema.

#### 3.1. VISTA FÍSICA

En la vista física se especifica y se ilustra la configuración del sistema en el momento de su ejecución.

**3.1.1. Arquitectura del sistema.** En el diagrama de la figura 1, se muestra la arquitectura del sistema, se evidencia que el usuario accede a la aplicación desde un navegador el cual puede ser google Chrome, Mozilla Firefox o Internet Explorer, los cuales pueden ser utilizados por la mayoría de dispositivos con conexión a internet ya sean móviles o computadores, esto con el fin de que el usuario pueda acceder a las funcionalidades del sistema y que pueda cargar sus archivos para que sean validados.

Figura 22: Arquitectura del sistema ValidateDocx



Fuente: El autor

**3.1.2. Diagrama de Despliegue:** En el diagrama de despliegue que se muestra a continuación en la figura 2, se especifican los detalles del software requerido para que el Sistema se pueda ejecutar.

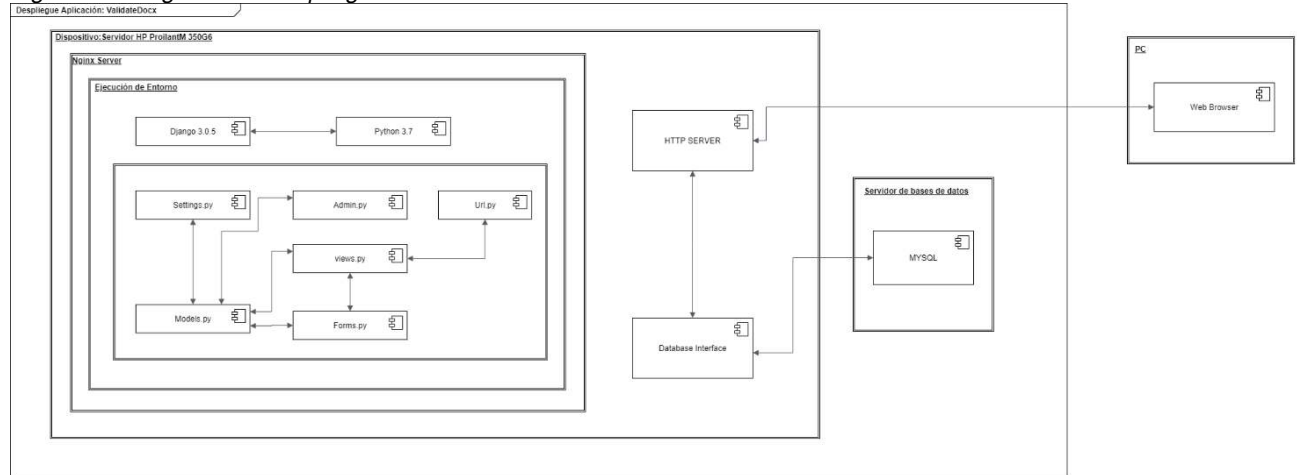
En el diagrama se especifica que se puede acceder al sistema utilizando un pc o cualquier dispositivo electrónico que soporte un navegador web ya sea: Google Chrome, Mozilla Firefox o Internet Explorer.

Este sistema está desarrollado con un framework de aplicaciones web gratuito y de código abierto llamado Django el cual está escrito en Python, para poder desplegar este

framework se utiliza Nginx Server el cual permite realizar conexiones HTTP para hacer intercambios de información entre el servidor y los clientes web.

La base de datos será de tipo relacional, estará desplegada en el mismo servidor y utilizará MYSQL como motor principal para permitir la conexión con Django.

Figura 23: Diagrama de despliegue del sistema



Fuente: El autor

### 3.2. VISTA LÓGICA:

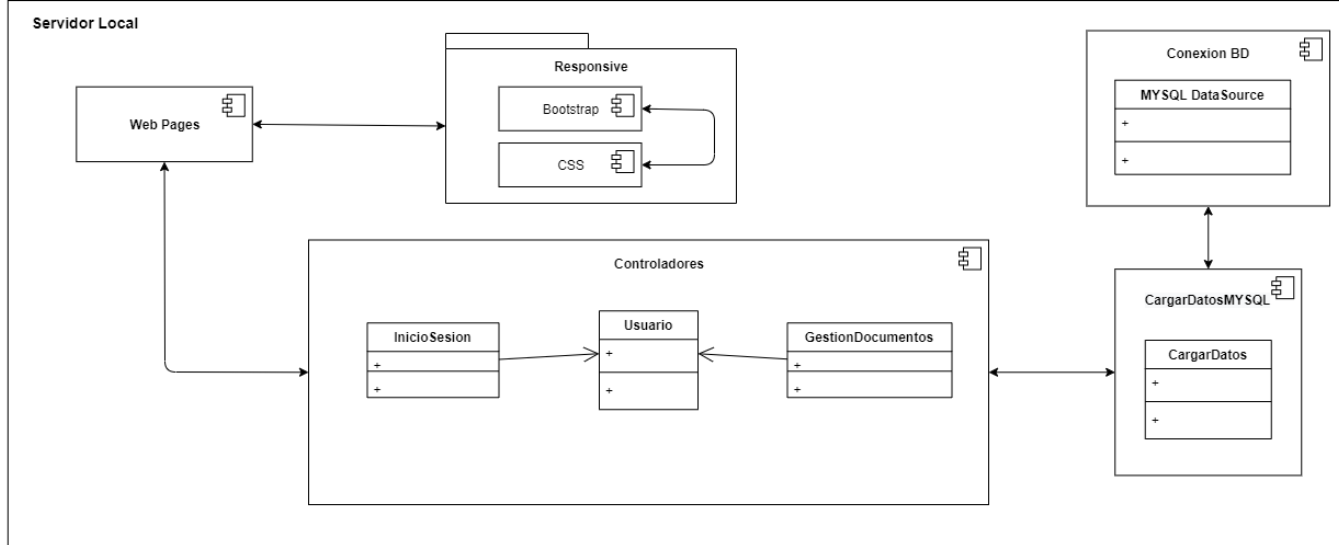
La vista lógica representa la estructura de la aplicación desde la parte de las funcionalidades del sistema.

**3.2.1. Diagrama de Componentes:** El objetivo de este diagrama de componentes es presentar las funcionalidades esenciales del sistema que se pueden ver en la figura 3.

La funcionalidad principal de este software es extraer los metadatos que tiene un archivo de Word con la finalidad de poder verificar si el documento cumple con los estándares de la NTC1486, para esta verificación es necesario utilizar el componente de carga de datos, ya que esta es la manera de que se pueden cargar tanto los datos del documento como los datos del usuario al sistema.

Con la información que retorne la validación y el componente de calcular porcentaje de validaciones, el componente de generar Documento PDF creará un reporte completo con los detalles las características con las que el documento cumple y no cumple para que el usuario lo pueda descargar.

Figura 24: Diagrama de componentes del sistema.

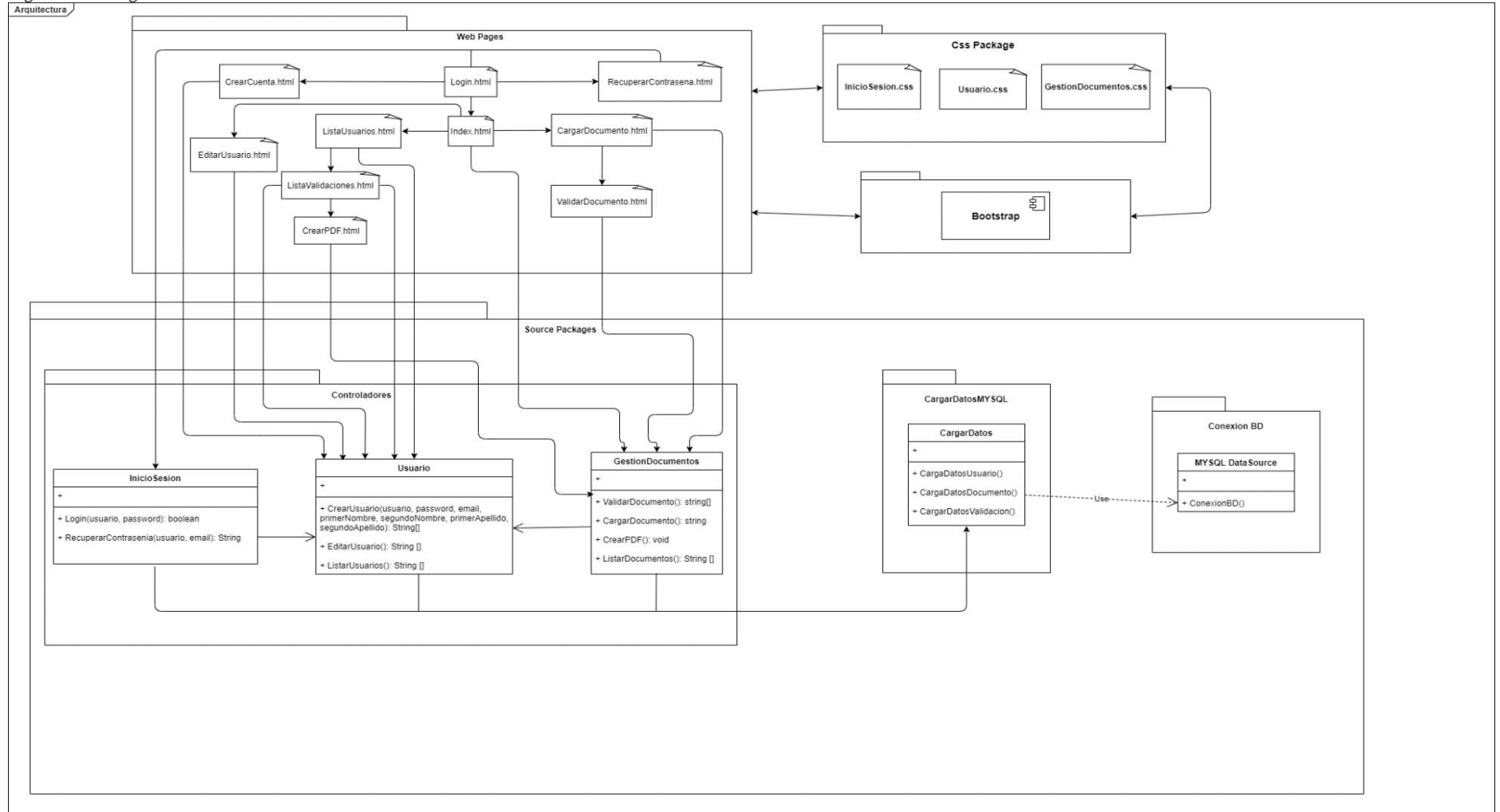


Fuente: El autor.

**3.2.2. Diagrama de Clases** En la figura 4 se representa el diagrama de clases el cual esta implementado en una arquitectura MVC para poder acceder a los datos que se encuentran alojados en la base de datos relacional, esta arquitectura se puede ver reflejada en los WebPages que vendría el paquete de las vistas, que a su vez se conectan con el paquete de controladores los cuales están distribuidos entre las clases.



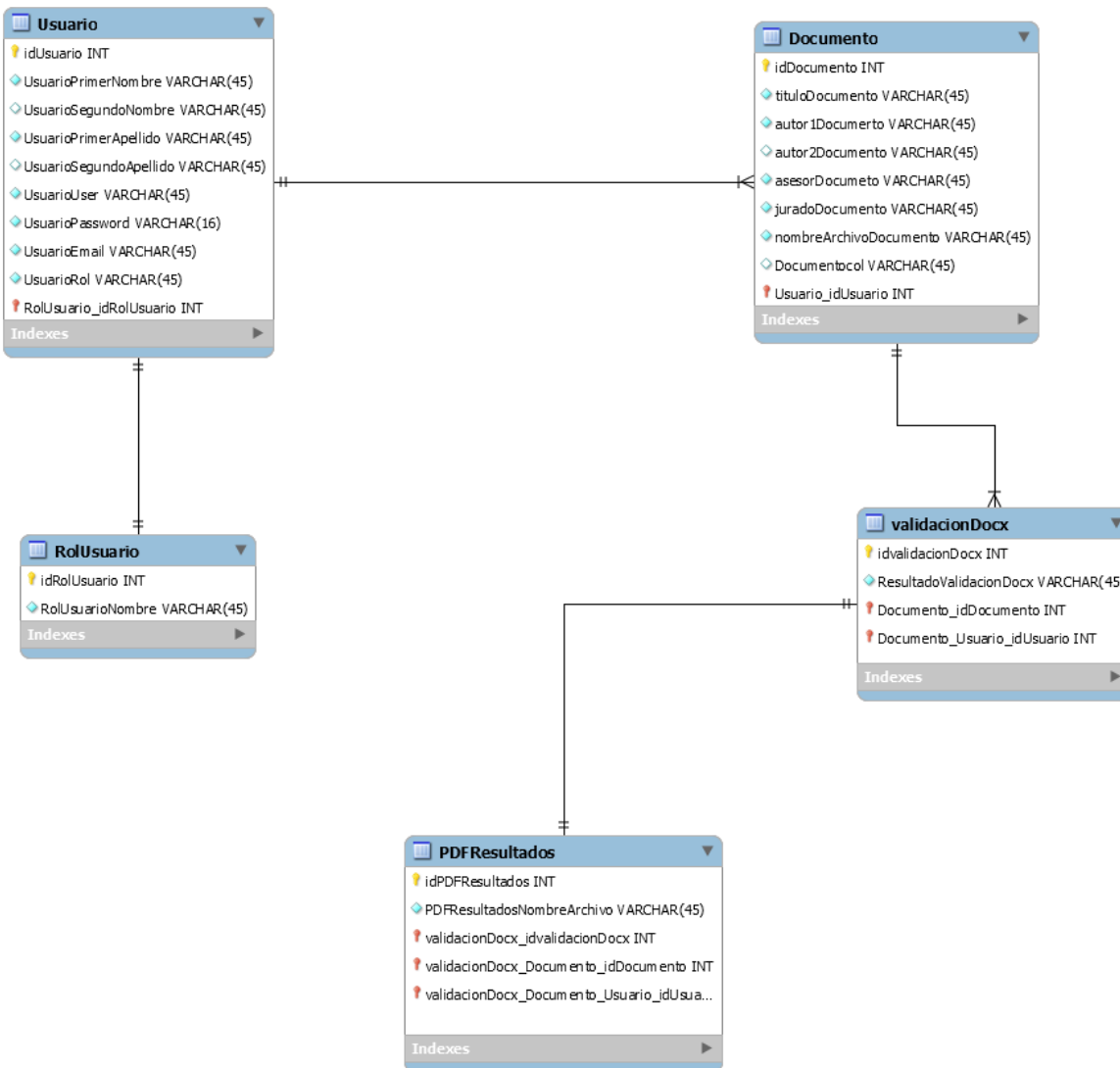
Figura 25: Diagrama de clases.



Fuente: El auto

**3.2.3. Modelo de datos** A continuación, en la figura 5 se presenta el modelo de datos del sistema ValidateDocx en donde se puede evidenciar la estructura lógica que tendrán los datos de modo relacional, en donde un usuario puede tener un rol asignado y puede cargar uno o varios documentos, a su vez, el documento puede ser validado una o varias veces y se puede generar un documento pdf por cada validación que se haya realizado

Figura 26: Modelo Entidad Relación del sistema.



Fuente: El autor



**UNIVERSIDAD CATÓLICA**  
de Colombia

**ANEXO C**  
**MANUAL DE INSTALACIÓN DEL SISTEMA**

**JULIAN FELIPE PLATA ZUÑIGA**

**UNIVERSIDAD CATÓLICA DE COLOMBIA**  
**FACULTAD DE INGENIERÍA**  
**PROGRAMA DE INGENIERÍA DE SISTEMAS Y COMPUTACIÓN**  
**BOGOTÁ**  
**2020**

## Contenido

Introducción.....	5
1. REQUISITOS MINIMOS DEL HARDWARE:.....	6
1.1. SERVIDOR: .....	6
1.2. CLIENTE: .....	6
2. REQUISITOS DE SOFTWARE:.....	7
2.1. SERVIDOR: .....	7
2.2. CLIENTE: .....	7
3. INSTALACIÓN DE COMPONENTES: .....	8
3.1. INSTALACIÓN DE BASES DE DATOS MYSQL Y WORKBENCH:.....	8
3.2. INSTALACIÓN DE PYTHON:.....	10
3.3. INSTALACIÓN DE PIP:.....	11
3.4. INSTALACIÓN DE DJANGO: .....	11
3.5. INSTALACION DE PYMYSQL: .....	12
3.6. INSTALACIÓN DE ZIPFILE: .....	12
3.7. INSTALACION DE PYTHON-LXML: .....	13
4. DESPLIEGUE DE LA APLICACIÓN. ....	14
4.1. DESCARGAR NGINX: .....	14
4.2. INICIAR NGINX:.....	14
4.3. CONFIGURACIÓN POR DEFECTO DE NGINX PARA EL DESPLIEGUE DEL SITIO: 15	
4.4. INSTALACIÓN DE GUNICORN:.....	16

## Tabla de Figuras.

Figura 1: Instalador de MYSQL .....	8
Figura 2: Configuración de acceso a MYSQL .....	9
Figura 3: Ventana de Comandos MYSQL .....	9
Figura 4: MYSQL Workbench.....	10
Figura 5: Instalación de Python .....	10
Figura 6: Consola de Comandos Python.....	11
Figura 7: Consola de comandos: instalación del PIP .....	11
Figura 8: Consola de comandos: Instalación de Django. ....	11
Figura 9: Consola de comandos: Instalación de PYMYSQL. ....	12
Figura 10: Consola de comandos: Instalación de la librería ZipFile36 .....	12
Figura 11: Consola de comandos: Instalación de la librería lxml.....	13
Figura 12: Página de descarga de nginx. ....	14
Figura 13: Inicio del servicio nginx .....	14
Figura 14: Proceso activo de nginx. ....	15
Figura 15: Configuración de nginx.....	16

## **Introducción**

Este documento se realiza con el fin de mencionar y describir con amplio detalle, la lista de componentes necesarios para la instalación del ambiente y el despliegue del sistema ValidateDocx. Este tipo de despliegues se realizan para que cualquier usuario pueda acceder al sistema desde su ordenador o dispositivo móvil, con un navegador web.

El sistema ValidateDocx se desarrolló con Django el cual es un framework de código abierto desarrollado con Python, por esta razón los componentes escogidos para el uso y despliegue del sistema, se seleccionaron de acuerdo a los elementos que son compatibles con este framework.

## **1. REQUISITOS MINIMOS DEL HARDWARE:**

### **1.1. SERVIDOR:**

- Procesador Intel Core i5 o superior.
- Memoria RAM 8 GB o superior.
- Disco duro 256 GB o más.

### **1.2. CLIENTE:**

- Procesador Intel Core i3 o superior.
- Memoria RAM 2 GB o superior.
- El almacenamiento del disco duro no tiene un estándar mínimo, puede ser el de la preferencia del usuario.

## **2. REQUISITOS DE SOFTWARE:**

### **2.1. SERVIDOR:**

- Instalador de MYSQL 6.0.
- Instalador de MYSQL WorkBench 8.1.
- Instalador Python 3.8.2.
- Instalador PIP.
- Django 3.0.4
- Pymysql
- Zipfile
- Python-lxml

### **2.2. CLIENTE:**

- Google Chrome.
- Word Office.



### 3. INSTALACIÓN DE COMPONENTES:

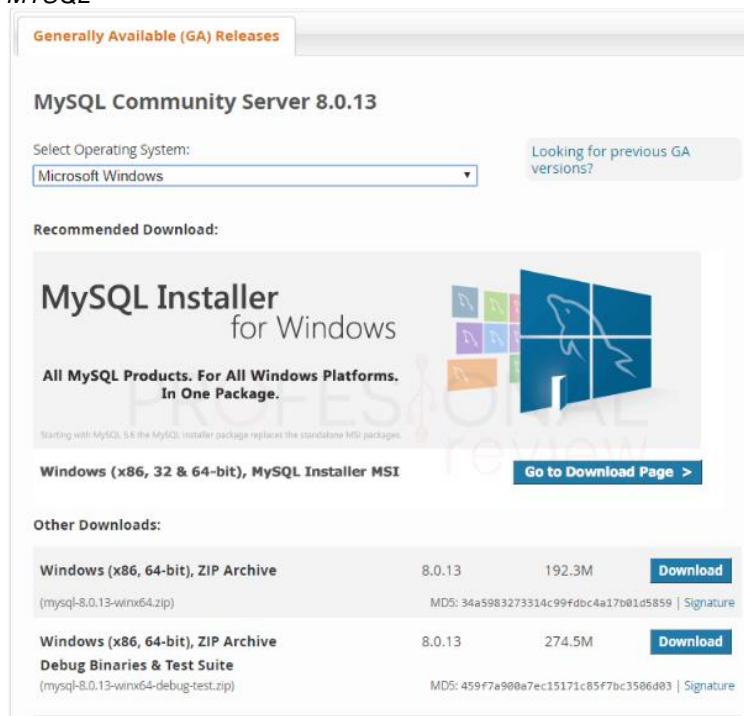
El proceso de instalación se debe hacer en el siguiente orden:

1. Servidor de bases de datos MYSQL5.7.
2. MYSQLWorkBench 8.0.
3. Python 3.8.2.
4. PIP.
5. Django 3.0.4.
6. PyMysql.
7. MysqlClient.
8. Zipfile.
9. Python-lxml.

#### 3.1. INSTALACIÓN DE BASES DE DATOS MYSQL Y WORKBENCH:

- Descargar MYSQL installer:

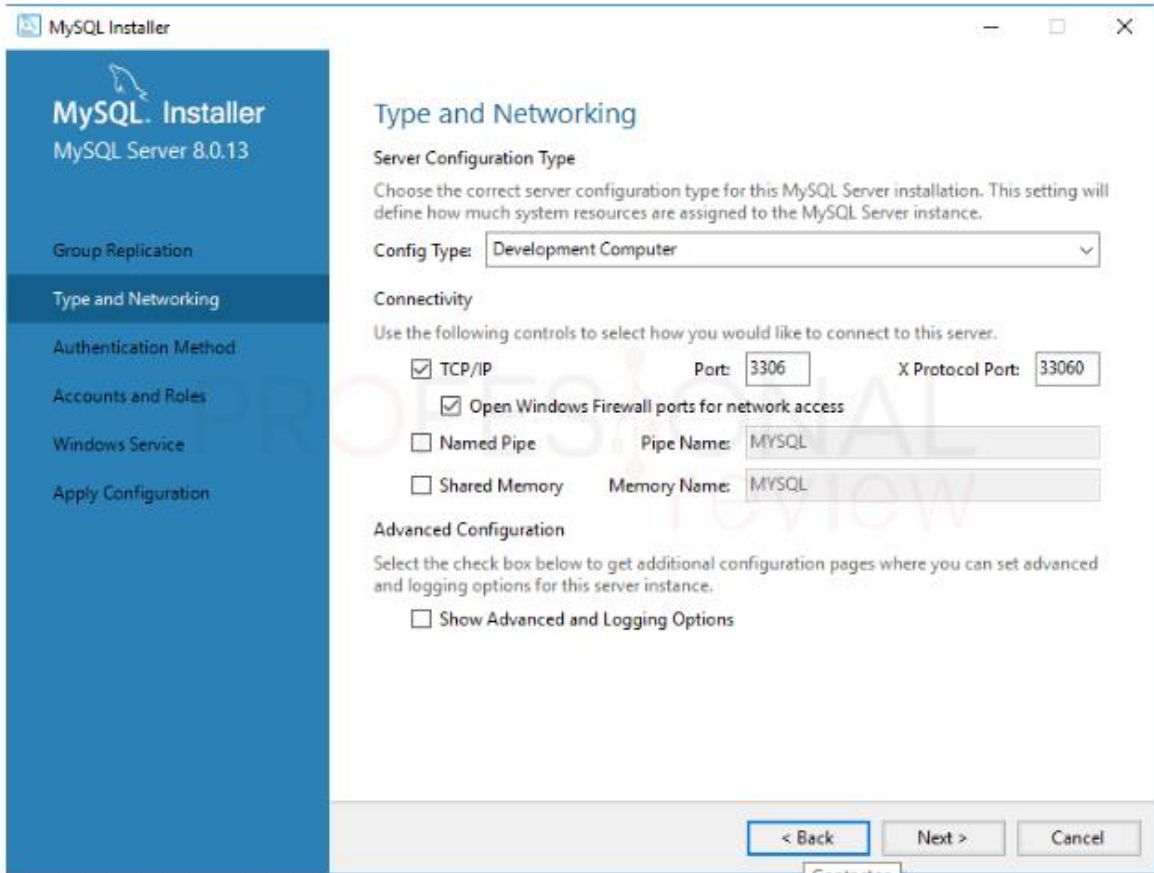
Figura 27: Instalador de MYSQL



Fuente: El autor

- Instalar Servidor MYSQL:

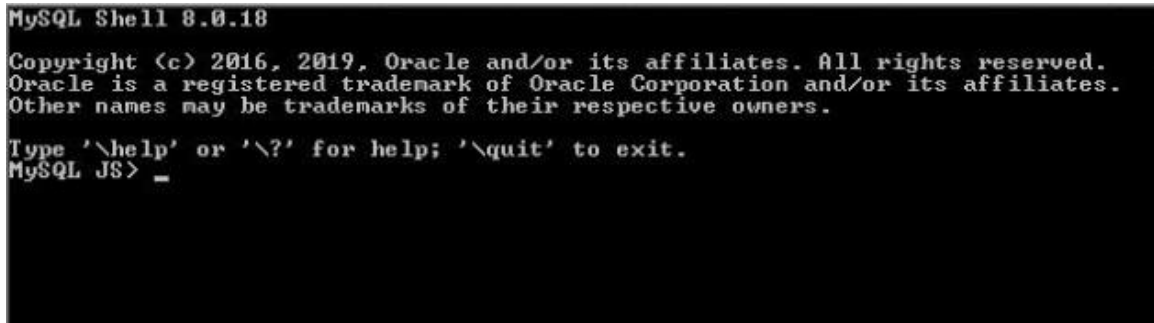
Figura 28: Configuración de acceso a MYSQL



Fuente: El autor.

- Se verifica que servidor de MYSQL esté ejecutándose correctamente:

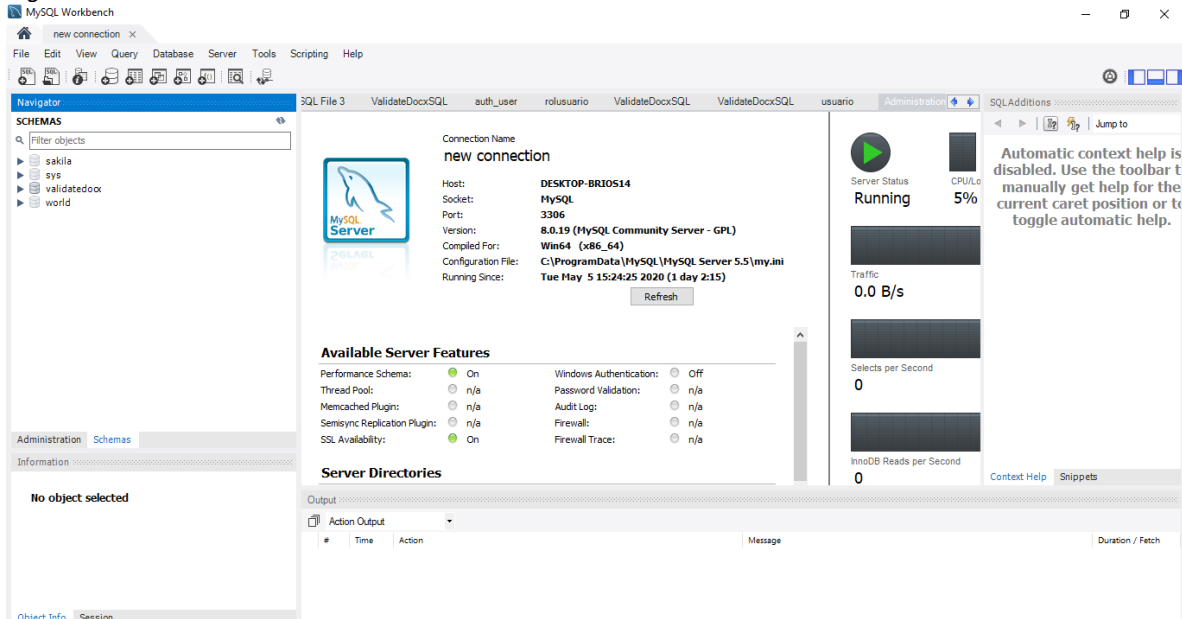
Figura 29: Ventana de Comandos MYSQL



Fuente: El autor.

- Se verifica que el WorkBench esté funcionando correctamente.

Figura 30: MYSQL Workbench

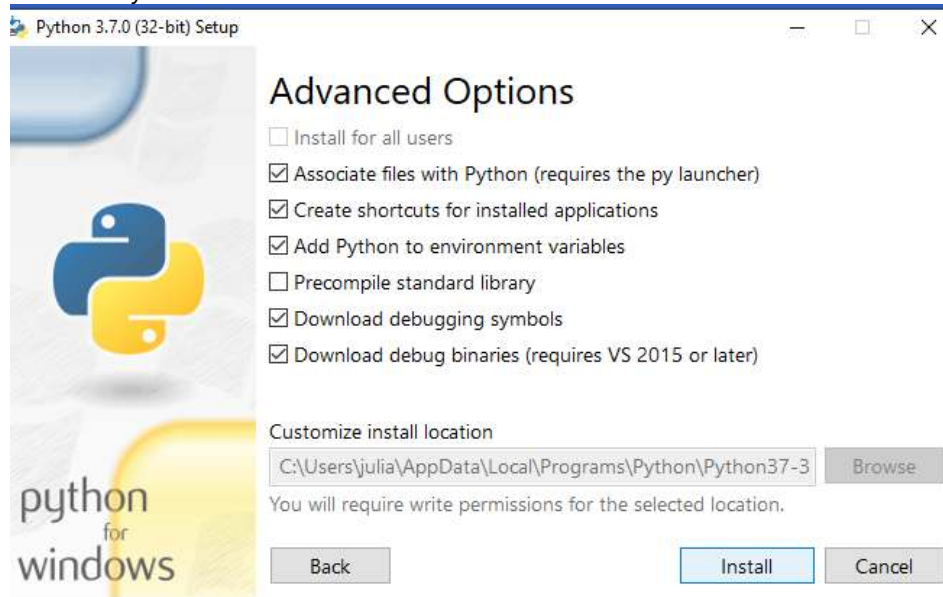


Fuente: El autor.

### 3.2. INSTALACIÓN DE PYTHON:

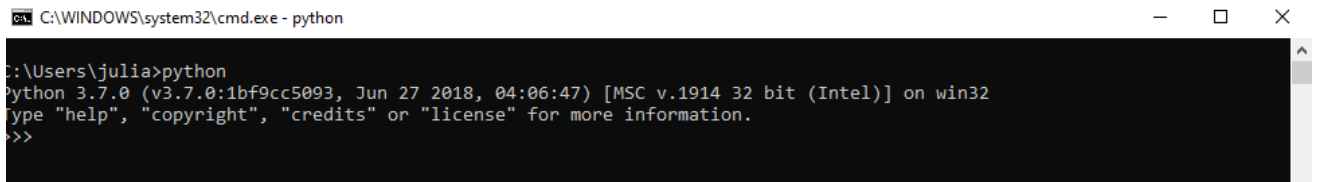
- Se ejecuta el instalador:

Figura 31: Instalación de Python



Fuente: El autor

Figura 32: Consola de Comandos Python.



```
C:\WINDOWS\system32\cmd.exe - python
C:\Users\julia>python
Python 3.7.0 (v3.7.0:1bf9cc5093, Jun 27 2018, 04:06:47) [MSC v.1914 32 bit (Intel)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>
```

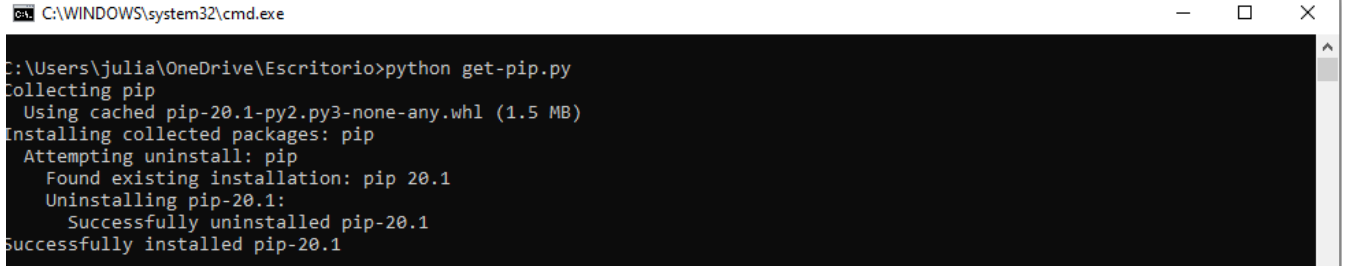
Fuente: El autor

### 3.3. INSTALACIÓN DE PIP:

El PIP es una herramienta de Python que permite descargar las librerías requeridas por el programador para realizar su desarrollo dependiendo de los requerimientos que el sistema a realizar requiera.

Después de descargar el archivo get-pip.py se ejecuta en la consola CMD para instalarlo en el entorno de desarrollo.

Figura 33: Consola de comandos: instalación del PIP



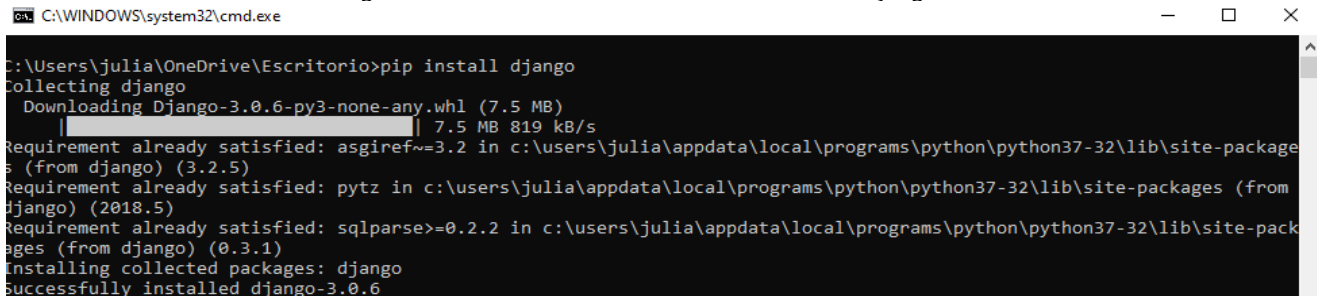
```
C:\WINDOWS\system32\cmd.exe
C:\Users\julia\OneDrive\Escritorio>python get-pip.py
Collecting pip
  Using cached pip-20.1-py2.py3-none-any.whl (1.5 MB)
Installing collected packages: pip
  Attempting uninstall: pip
    Found existing installation: pip 20.1
    Uninstalling pip-20.1:
      Successfully uninstalled pip-20.1
Successfully installed pip-20.1
```

Fuente: El autor

### 3.4. INSTALACIÓN DE DJANGO:

Utilizando el PIP que previamente se instaló en el sistema se instala el framework Django con el comando "pip install Django en el CMD del sistema.

Figura 34: Consola de comandos: Instalación de Django.



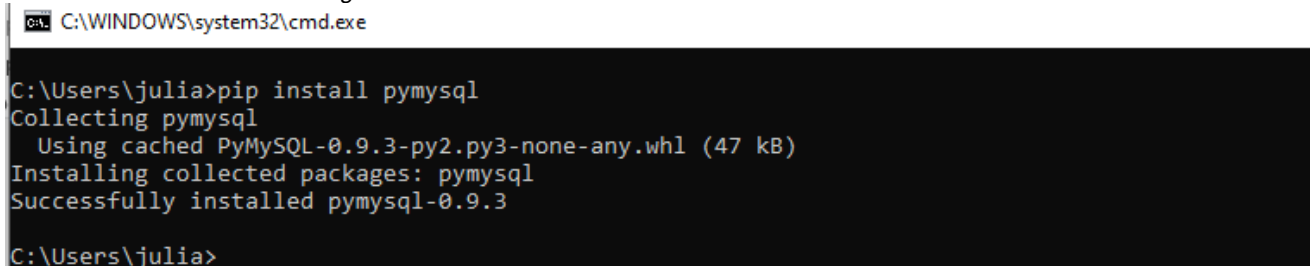
```
C:\WINDOWS\system32\cmd.exe
C:\Users\julia\OneDrive\Escritorio>pip install django
Collecting django
  Downloading Django-3.0.6-py3-none-any.whl (7.5 MB)
    |-----| 7.5 MB 819 kB/s
Requirement already satisfied: asgiref<=3.2 in c:\users\julia\appdata\local\programs\python\python37-32\lib\site-packages (from django) (3.2.5)
Requirement already satisfied: pytz in c:\users\julia\appdata\local\programs\python\python37-32\lib\site-packages (from django) (2018.5)
Requirement already satisfied: sqlparse>=0.2.2 in c:\users\julia\appdata\local\programs\python\python37-32\lib\site-packages (from django) (0.3.1)
Installing collected packages: django
Successfully installed django-3.0.6
```

Fuente: El autor

### 3.5. INSTALACION DE PYMYSQL:

El componente de PYMYSQL es esencial ya que este componente permite realizar conexiones entre el framework Django y la base de datos MYSQ, como se mostrará la en la siguiente figura:

*Figura 35: Consola de comandos: Instalación de PYMYSQL.*



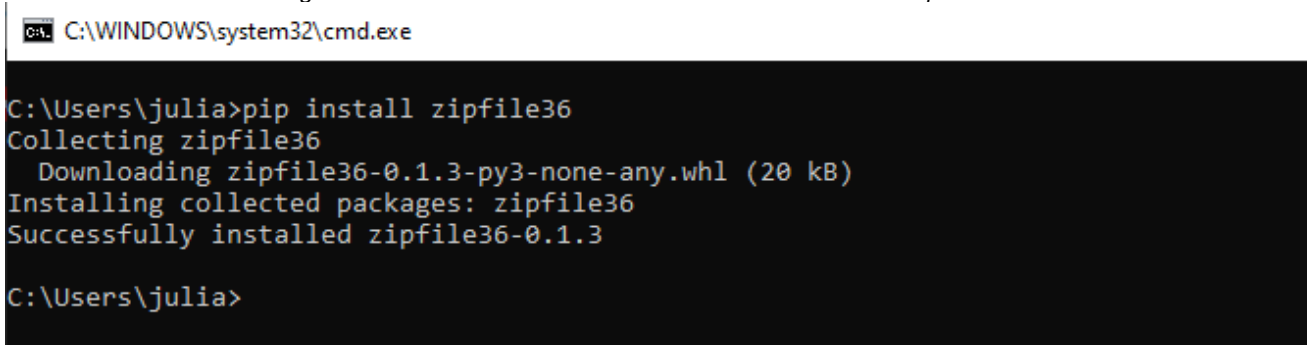
```
C:\WINDOWS\system32\cmd.exe
C:\Users\julia>pip install pymysql
Collecting pymysql
  Using cached PyMySQL-0.9.3-py2.py3-none-any.whl (47 kB)
Installing collected packages: pymysql
Successfully installed pymysql-0.9.3
C:\Users\julia>
```

Fuente: El autor

### 3.6. INSTALACIÓN DE ZIPFILE:

Este componente es una librería que permite abrir archivos de formato Rar o Zip, debido a que el sistema se dedica a la extracción de metadatos, y estos formatos son los que contienen los archivos con los metadatos que se necesitan leer para tomar las decisiones pertinentes que el sistema necesita.

*Figura 36: Consola de comandos: Instalación de la librería ZipFile36*



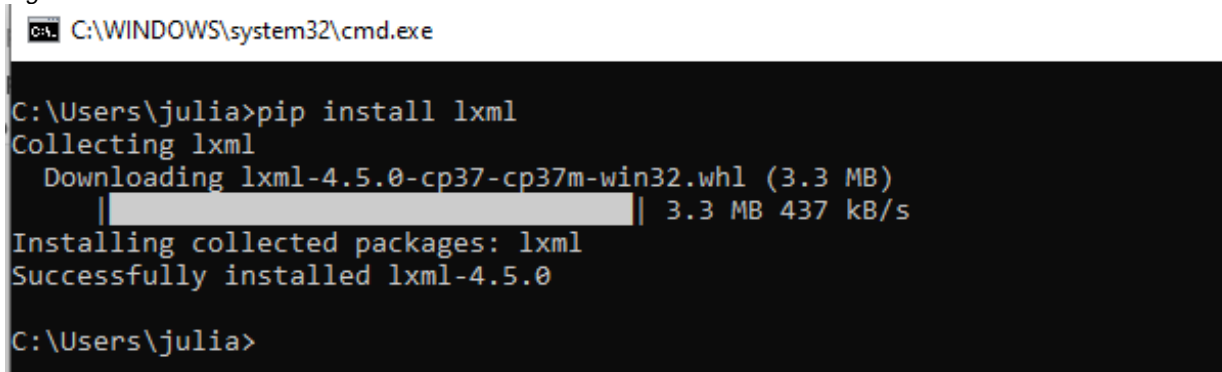
```
C:\WINDOWS\system32\cmd.exe
C:\Users\julia>pip install zipfile36
Collecting zipfile36
  Downloading zipfile36-0.1.3-py3-none-any.whl (20 kB)
Installing collected packages: zipfile36
Successfully installed zipfile36-0.1.3
C:\Users\julia>
```

Fuente: El autor.

### 3.7. INSTALACION DE PYTHON-LXML:

Este componente es una librería que permite leer y extraer información de archivos de formato XML, esto es indispensable ya que los metadatos necesarios para esta investigación, se encuentran alojados en archivos de este formato, en la siguiente imagen se puede ver la forma en que se instala en el servidor donde se va a desplegar.

Figura 37: Consola de comandos: Instalación de la librería lxml



```
C:\WINDOWS\system32\cmd.exe

C:\Users\julia>pip install lxml
Collecting lxml
  Downloading lxml-4.5.0-cp37-cp37m-win32.whl (3.3 MB)
    |-----| 3.3 MB 437 kB/s
Installing collected packages: lxml
Successfully installed lxml-4.5.0

C:\Users\julia>
```

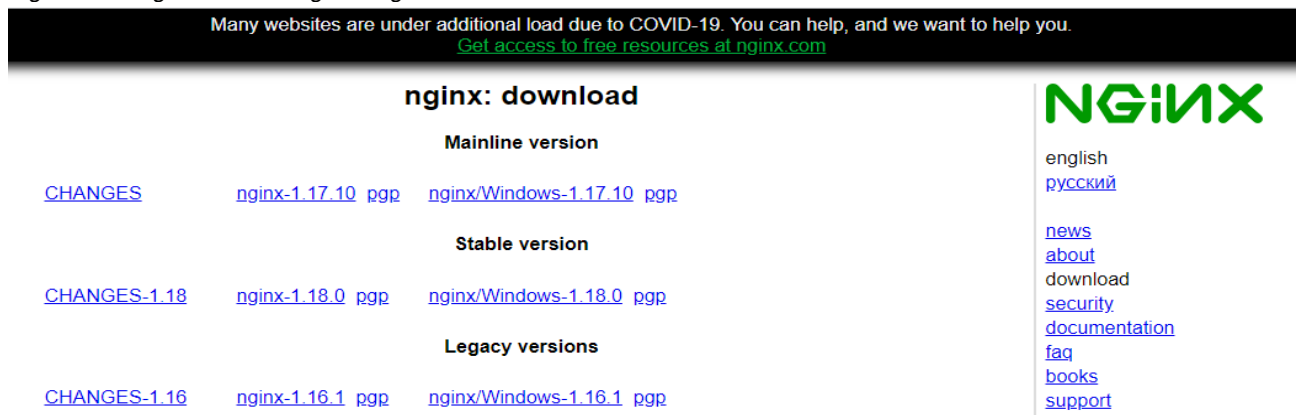
Fuente: El autor.

## 4. DESPLIEGUE DE LA APLICACIÓN.

Para desplegar la aplicación ValidateDocx en el servidor, es necesario descargar nginx ya que este es un servidor web/proxy de alto rendimiento, y este permite que la aplicación quede disponible para cualquier usuario que tenga una conexión a internet y busque acceder a la aplicación, para hacer este despliegue se deben hacer los siguientes pasos:

### 4.1. DESCARGAR NGINX:

Figura 38: Página de descarga de nginx.



Fuente: <http://nginx.org/en/download.html>

### 4.2. INICIAR NGINX:

Para iniciar el nginx se debe escribir el comando “start nginx” en la ventana de comandos de Windows como se ve en la figura 13.

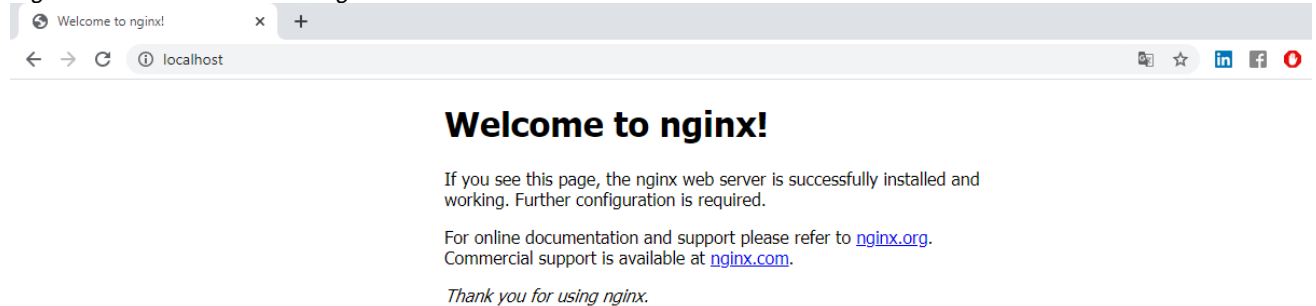
Figura 39: Inicio del servicio nginx

```
C:\WINDOWS\system32\cmd.exe
C:\Users\julia\OneDrive\Escritorio\nxinx-1.17.10>start nginx.exe
C:\Users\julia\OneDrive\Escritorio\nxinx-1.17.10>_
```

Fuente: El autor

Si el componente ha iniciado sin problemas, la consola de comandos no mostrará ningún mensaje, además una vez iniciado el servicio de nginx, se debe verificar que el proceso esté corriendo correctamente, en la figura 14 se muestra el mensaje de bienvenida a nginx el cual indica que el proceso ha sido activado con éxito.

Figura 40: Proceso activo de nginx.



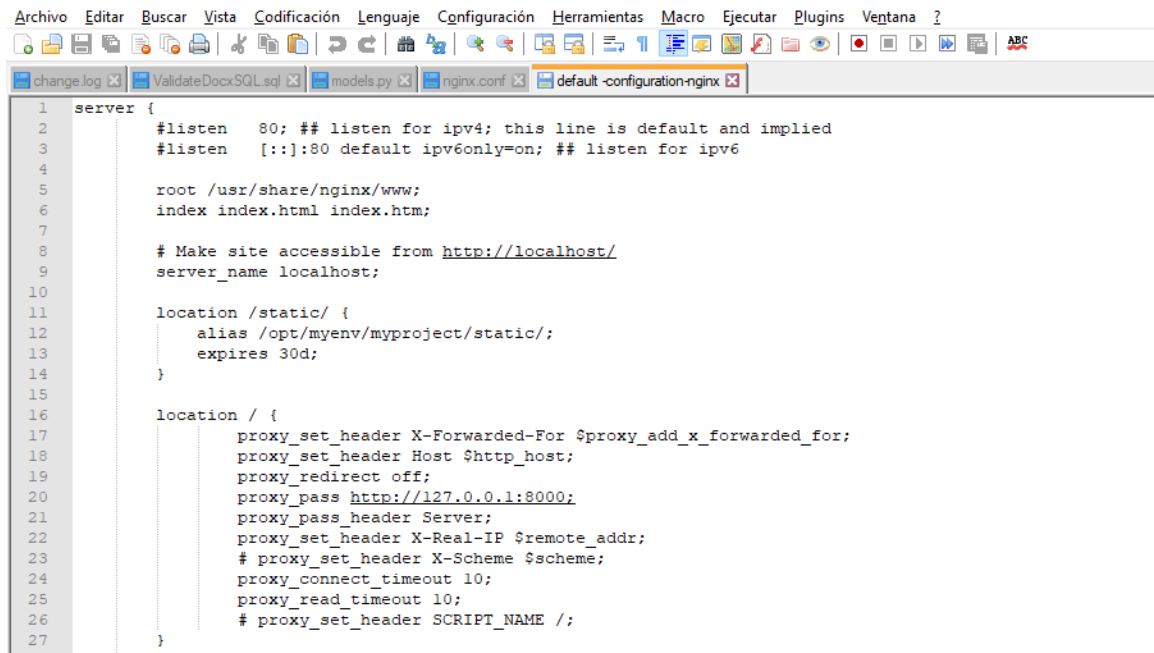
Fuente: El autor

### 4.3. CONFIGURACIÓN POR DEFECTO DE NGINX PARA EL DESPLIEGUE DEL SITIO:

Nginx debe configurarse como se ilustra en la figura 15 para que la aplicación de ValidateDocx pueda ser accedida a través de cualquier navegador que tenga conexión al servidor, además se debe especificar la ubicación de los archivos estáticos de la aplicación para su correcto funcionamiento.



Figura 41: Configuración de nginx.



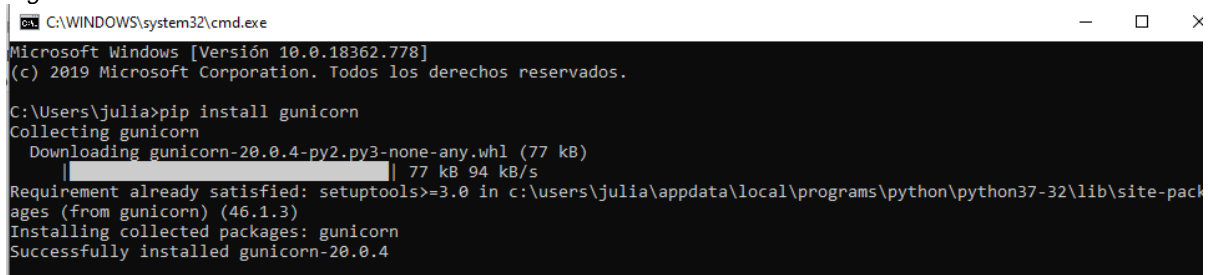
```
1 server {
2     #listen 80; ## listen for ipv4; this line is default and implied
3     #listen [::]:80 default ipv6only=on; ## listen for ipv6
4
5     root /usr/share/nginx/www;
6     index index.html index.htm;
7
8     # Make site accessible from http://localhost/
9     server_name localhost;
10
11     location /static/ {
12         alias /opt/myenv/myproject/static/;
13         expires 30d;
14     }
15
16     location / {
17         proxy_set_header X-Forwarded-For $proxy_add_x_forwarded_for;
18         proxy_set_header Host $http_host;
19         proxy_redirect off;
20         proxy_pass http://127.0.0.1:8000;
21         proxy_pass_header Server;
22         proxy_set_header X-Real-IP $remote_addr;
23         # proxy_set_header X-Scheme $scheme;
24         proxy_connect_timeout 10;
25         proxy_read_timeout 10;
26         # proxy_set_header SCRIPT_NAME /;
27     }
28 }
```

Fuente: El autor

#### 4.4. INSTALACIÓN DE GUNICORN:

Este componente nos permitirá activar el proceso de la aplicación ValidateDocx constantemente para permitir el acceso de clientes web a la aplicación, para instalar este componente es necesario utilizar el comando “pip install” en la consola de comandos como se ve en la figura 16:

Figura 42: Instalación de Gunicorn.



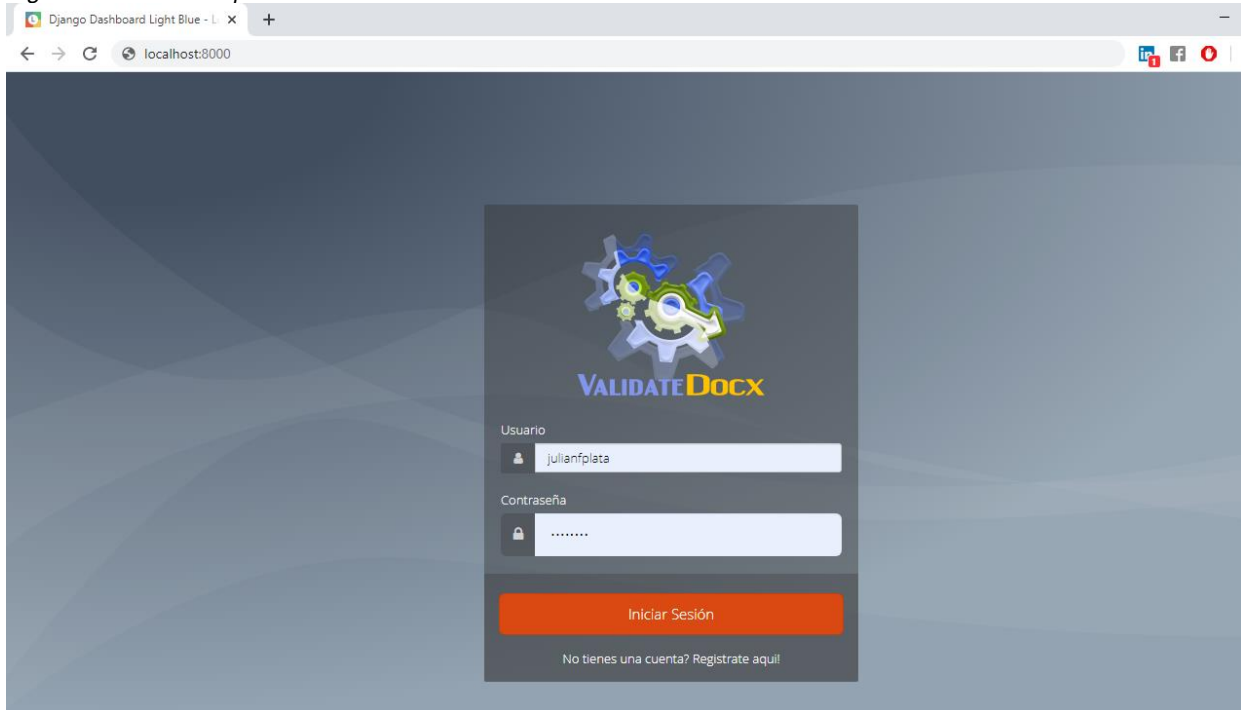
```
C:\WINDOWS\system32\cmd.exe
Microsoft Windows [Versión 10.0.18362.778]
(c) 2019 Microsoft Corporation. Todos los derechos reservados.

C:\Users\julia>pip install gunicorn
Collecting gunicorn
  Downloading gunicorn-20.0.4-py2.py3-none-any.whl (77 kB)
    |-----| 77 kB 94 kB/s
Requirement already satisfied: setuptools>=3.0 in c:\users\julia\appdata\local\programs\python\python37-32\lib\site-packages (from gunicorn) (46.1.3)
Installing collected packages: gunicorn
Successfully installed gunicorn-20.0.4
```

Fuente: El autor

Una vez que se ha instalado Gunicorn y se ha iniciado el componente de nginx, se procede a verificar que el sistema de ValidateDocx, en la figura 17 se puede apreciar que la aplicación ha iniciado correctamente.

Figura 43: Inicio de Aplicación ValidateDocx.



Fuente: El autor



**UNIVERSIDAD CATÓLICA**  
de Colombia

**ANEXO D**  
**MANUAL DE USUARIO**

**JULIAN FELIPE PLATA ZUÑIGA**

**UNIVERSIDAD CATÓLICA DE COLOMBIA**  
**FACULTAD DE INGENIERÍA**  
**PROGRAMA DE INGENIERÍA DE SISTEMAS Y COMPUTACIÓN**  
**BOGOTÁ**  
**2020**

## Contenido

INTRODUCCION .....	4
1. FUNCIONALIDADES DEL SISTEMA. ....	5
1.1. INICIO DE SESIÓN.....	5
1.2. CREAR CUENTA.....	5
2. Gestión de documentos .....	7
2.1. Lista de archivos cargados.....	7
2.2. CARGAR DOCUMENTO. ....	8
2.3. ANALISIS DE DOCUMENTOS. ....	8
2.4. DESCARGAR RESULTADO DE VALIDACIÓN DE DOCUMENTOS. ....	9
3. Administración de USUARIOS.....	11
3.1. LISTADO Y SELECCIÓN DE USUARIOS .....	11
3.2. RECUPERACIÓN DE CONTRASEÑA.....	11
3.3. ADMINISTRACIÓN DEL SITEMA.....	12
3.4. Cambio de permisos de usuario.....	13
3.5. Roles de usuario. ....	14

### **Tabla de imágenes.**

Figura 1: Inicio de sesión. ....	60
Figura 2: Crear cuenta. ....	60
Figura 3: Lista de archivos Cargados. ....	7
Figura 4: Carga de información y archivo de documento. ....	8
Figura 5: Resultado del análisis de un documento. ....	8
Figura 6: Descargar resultado de la validación ....	9
Figura 7: PDF generado, con los resultados de la validación de un documento. ...	10
Figura 8: Lista de Usuarios. ....	11
Figura 9: Recuperación de contraseña. ....	11
Figura 10: Autenticación para el administrador del sistema. ....	12
Figura 11: Menú de administración del sistema. ....	13
Figura 12: Modificación de un usuario. ....	14

## **INTRODUCCION**

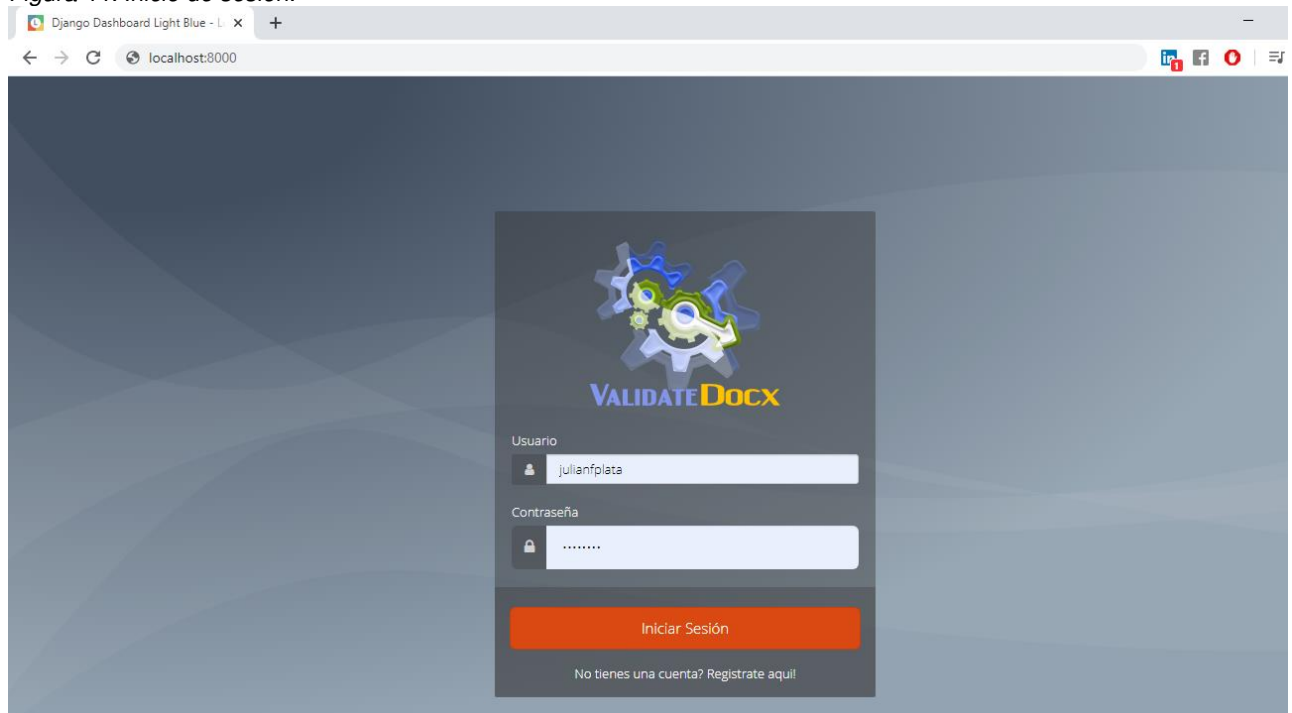
En este documento se explicará con gran detalle las instrucciones para manejar el sistema ValidateDocx el cual está construido como una aplicación web con el fin de brindarle una facilidad de acceso a los usuarios que tengan dispositivos con conexión a internet. La información plasmada en este anexo proviene de las evidencias tomadas al final de la construcción del sistema ValidateDocx, la cual permite ilustrar muy fácilmente, los pasos que se deben seguir para que cualquier usuario pueda utilizar este sistema.

# 1. FUNCIONALIDADES DEL SISTEMA.

## 1.1. INICIO DE SESIÓN

En la figura 1 se visualiza un recuadro el cual le pide que se ingrese una clave y una contraseña que previamente el usuario debió crear en el sistema con la funcionalidad de crear cuenta de usuario.

Figura 44: Inicio de sesión.



Fuente: El autor

## 1.2. CREAR CUENTA.

En la figura 2 se visualiza un recuadro el cual le pide que se ingresen datos como; usuario, email, contraseña y otro recuadro para digitar la contraseña por segunda vez y verificar que el usuario esté totalmente seguro de lo que está escribiendo, con estos datos registrados en el sistema, el usuario podrá acceder en el momento en que lo prefiera para hacer uso de las funcionalidades.

Figura 45: Crear cuenta.

The image shows a web browser window with the URL `7c2b19763a1447e0a52ab0d965a242b2.vfs.cloud9.us-east-1.amazonaws.com/register/`. The page features a registration form with the following elements:

- Header:** "Regístrate" and a link "Has clic aqui si ya tienes una cuneta."
- Usuario:** Input field containing "julianfplata".
- Email:** Input field containing "Email".
- Contraseña:** Input field with masked characters "\*\*\*\*\*".
- Confirmación:** Input field labeled "Confirma tu contraseña por favor" with "Password check" text.
- Botón:** "Crear cuenta" (orange button).
- Footer:** "Ya estas registrado? Inicia Sesion".

A small "Inicio" link is visible in the bottom left corner of the page.

Fuente: El autor

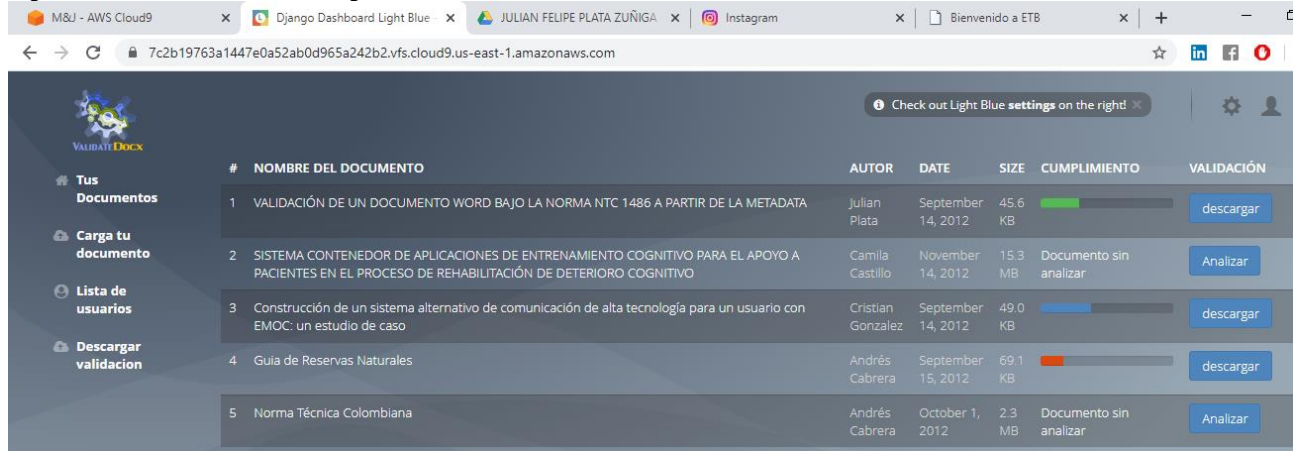


## 2. Gestión de documentos

### 2.1. LISTA DE ARCHIVOS CARGADOS

Esta funcionalidad es la que el usuario visualizará inmediatamente después de que se haya logueado con sus credenciales en el sistema, si el rol del usuario del sistema es el de un usuario común, no podrá visualizar los documentos cargados de los demás usuarios sino solamente los documentos propios, por el contrario, si el usuario logueado en el sistema es un coordinador de calidad o un administrador, podrá visualizar todos los documentos que todos los usuarios del sistema han cargado recientemente junto con un porcentaje representado en una gráfica de colores, que indicará cual es el nivel de cumplimiento de los archivos que se han cargado y validado con el sistema esta funcionalidad se puede visualizar en la figura 3.

Figura 46: Lista de archivos Cargados.



#	NOMBRE DEL DOCUMENTO	AUTOR	DATE	SIZE	CUMPLIMIENTO	VALIDACIÓN
1	VALIDACIÓN DE UN DOCUMENTO WORD BAJO LA NORMA NTC 1486 A PARTIR DE LA METADATA	Julian Plata	September 14, 2012	45.6 KB	<div style="width: 100%; background-color: green;"></div>	<a href="#">descargar</a>
2	SISTEMA CONTENEDOR DE APLICACIONES DE ENTRENAMIENTO COGNITIVO PARA EL APOYO A PACIENTES EN EL PROCESO DE REHABILITACIÓN DE DETERIORO COGNITIVO	Camila Castillo	November 14, 2012	15.3 MB	Documento sin analizar	<a href="#">Analizar</a>
3	Construcción de un sistema alternativo de comunicación de alta tecnología para un usuario con EMOC: un estudio de caso	Cristian Gonzalez	September 14, 2012	49.0 KB	<div style="width: 100%; background-color: blue;"></div>	<a href="#">descargar</a>
4	Guia de Reservas Naturales	Andrés Cabrera	September 15, 2012	69.1 KB	<div style="width: 100%; background-color: orange;"></div>	<a href="#">descargar</a>
5	Norma Técnica Colombiana	Andrés Cabrera	October 1, 2012	2.3 MB	Documento sin analizar	<a href="#">Analizar</a>

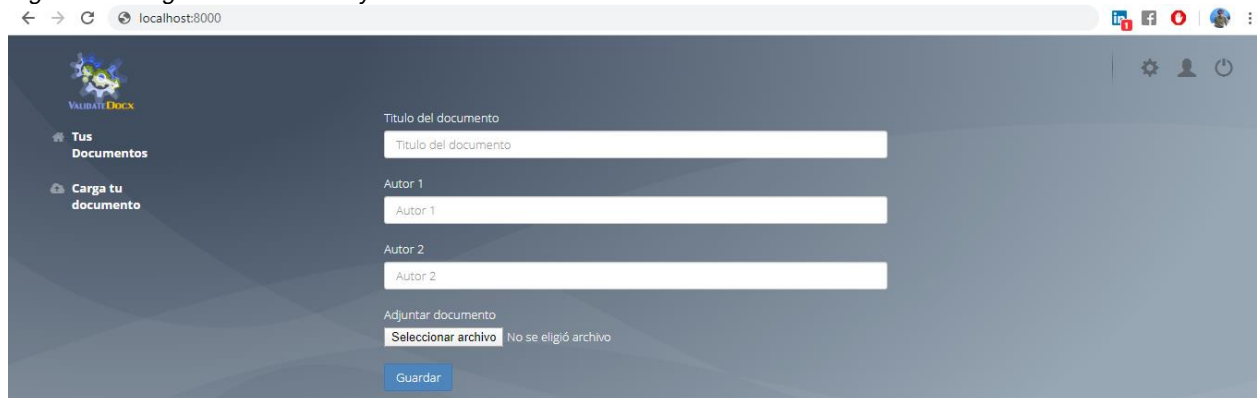
Fuente: El autor.

por el contrario, si el documento cargado aún no ha sido analizado por el sistema, aparecerá un botón que redirigirá al usuario a la funcionalidad de análisis de documentos el cual informará y almacenará las características del documento para verificar si cumple o no con los estándares de la norma NTC1486.

## 2.2. CARGAR DOCUMENTO.

Esta funcionalidad le permitirá al usuario ingresar los datos relacionados con el documento que desea analizar, junto con un archivo en formato Word el cual contiene la información del documento investigativo, este paso es totalmente obligatorio antes de iniciar el análisis del documento, en la figura 4 se pueden visualizar los campos que la funcionalidad solicita.

Figura 47: Carga de información y archivo de documento.



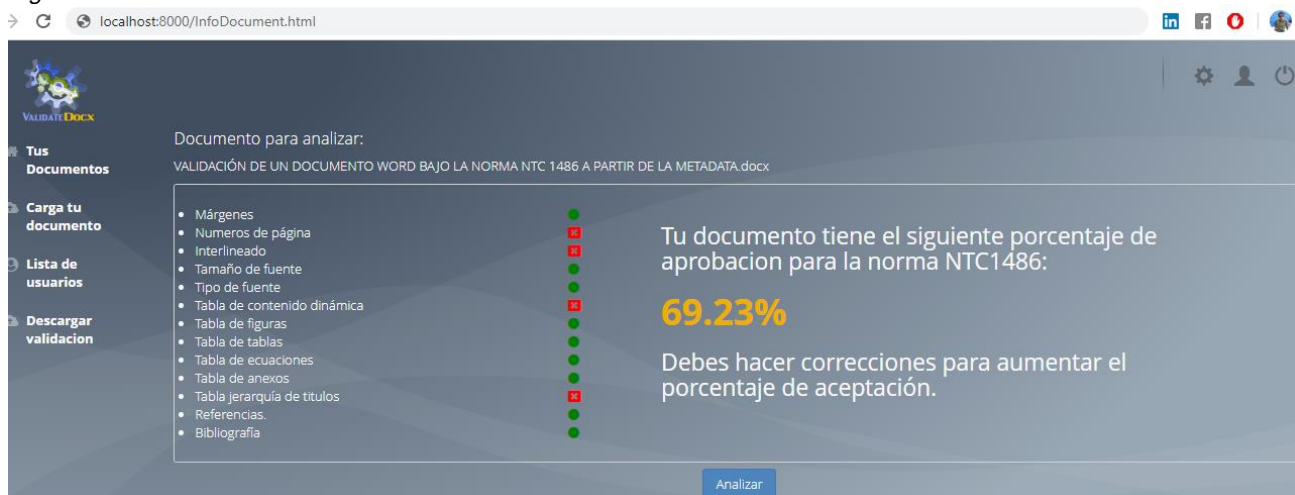
The screenshot shows a web browser window with the URL 'localhost:8000'. The application interface includes a sidebar with 'VALIDATE DOCX' logo and navigation options: 'Tus Documentos', 'Carga tu documento', 'Lista de usuarios', and 'Descargar validacion'. The main content area contains a form with the following fields: 'Titulo del documento' (input field with placeholder 'Titulo del documento'), 'Autor 1' (input field with placeholder 'Autor 1'), 'Autor 2' (input field with placeholder 'Autor 2'), and 'Adjuntar documento' (file selection button labeled 'Seleccionar archivo' with the text 'No se eligió archivo'). A 'Guardar' button is located at the bottom of the form.

Fuente: El autor.

## 2.3. ANALISIS DE DOCUMENTOS.

Esta funcionalidad del sistema le mostrará al usuario la lista de ítems que ValidateDocx tiene en cuenta para retornar la respuesta del análisis del documento, la lista se mostrará junto a varios iconos de color verde o rojo, los cuales indican si el documento tiene o no esa característica que va de acuerdo a las reglas de la norma NTC1486, en la figura 5 se puede visualizar una prueba de ejemplo que se realizó con un documento aleatorio.

Figura 48: Resultado del análisis de un documento.



Fuente: El autor.

## 2.4. DESCARGAR RESULTADO DE VALIDACIÓN DE DOCUMENTOS.

Esta funcionalidad solo está disponible tanto para el usuario de control de calidad, como par el administrador del sistema, y les permite descargar un archivo PDF con el resultado de la validación que el sistema generó, ya que podría darse la necesidad de que el usuario necesite guardar esa información en otro lugar, para descargar el documento se da clic en el botón descargar, como se ve en la figura

Figura 49: Descargar resultado de la validación



DATE	SIZE	CUMPLIMIENTO	VALIDACIÓN
September 14, 2012	45.6 KB	<div style="width: 100%; height: 10px; background-color: green;"></div>	descargar
November 14, 2012	15.3 MB	Documento sin analizar	Analizar
September 14, 2012	49.0 KB	<div style="width: 50%; height: 10px; background-color: blue;"></div>	descargar
September 15, 2012	69.1 KB	<div style="width: 20%; height: 10px; background-color: orange;"></div>	descargar
October 1, 2012	2.3 MB	Documento sin analizar	Analizar

Mostrar todo X

Fuente: El autor

Seguido de presionar el botón descargar, el sistema genera el archivo PDF con los resultados de la validación, en la figura 7 se puede ver un ejemplo del contenido del PDF que genera.

Figura 50: PDF generado, con los resultados de la validación de un documento.

### Resultados ValidateDocx

**Documento:** Construcción de un sistema alternativo de comunicación de alta tecnología para un usuario con EMOC: un estudio de caso.

Aspectos evaluados:

- Márgenes:
  - Superior: 3 cm. (4 cm. título)
  - Izquierdo: 4 cm.
  - Derecho: 2 cm.
  - Inferior: 3 cm.
  - Número de página a 2 cm. y centrado
- Números de página
- Interlineado
- Tamaño de fuente
- Tipo de fuente
- Tabla de contenido dinámica
- Tabla de figuras
- Tabla de tablas
- Tabla de ecuaciones
- Tabla de anexos
- Tabla jerarquía de títulos
- Referencias.
- Bibliografía

Teniendo en cuenta los aspectos anteriormente mencionados, el documento; "Construcción de un sistema alternativo de comunicación de alta tecnología para un usuario con EMOC: un estudio de caso", tiene un porcentaje del 69.23%, lo que indica que el documento aún debe corregir alguno de los aspectos anteriormente mencionados para que cumpla con los estándares NTC1486

Fuente: El autor

### 3. Administración de USUARIOS.

#### 3.1. LISTADO Y SELECCIÓN DE USUARIOS

Esta sección del sistema solo estará disponible para los roles de usuario de administrador y el de control de calidad, el cual permite visualizar la lista completa de los usuarios que están registrados en el sistema el cual se puede visualizar en la figura 6, además también permite acceder a la información de los documentos que se han cargado.

Figura 51: Lista de Usuarios.

#	NOMBRE DEL USUARIO	EMAIL	CANTIDAD DE DOCUMENTOS CARGADOS
1	Julian Plata	jfplata48@ucatonlica.edu.co	1
2	Camila Castillo	ccastilloulloa@ucatonlica.edu.co	2
3	Cristian Gonzalez	ccgonzalez82@ucatonlica.edu.co	2
4	Andrés Cabrera	afcabrera26@ucatonlica.edu.co	1

Fuente: El autor

#### 3.2. RECUPERACIÓN DE CONTRASEÑA.

Si el usuario olvida su contraseña tiene la posibilidad de recuperarla utilizando la opción de recuperar contraseña que se encuentra al final de la ventana donde el usuario se loguea con sus credenciales para iniciar sesión, el sistema le solicitará a la persona, el nombre de usuario y el correo electrónico que utilizó al momento de registrarse como se ve en la figura 9, posteriormente el sistema enviará un correo con un link que le permitirá digitar una nueva contraseña para poder ingresar.

Figura 52: Recuperación de contraseña.

Valibate Docx

Check out Light Blue settings on the right!

Nombre de usuario

Email


Enviar

Fuente: El autor.

### 3.3. ADMINISTRACIÓN DEL SISTEMA.

ValidateDocx tiene otra interfaz adicional diferente a la que se visualiza en las figuras anteriores, ya que cuenta con un administrador tanto de usuarios como de contenido del sistema, una gran característica es, que esta parte del sistema solo podrá ser utilizada por un usuario de tipo administrador, ya que el administrador tiene acceso a todas las funcionalidades del sistema sin ninguna restricción, además de que es el encargado de aumentarle o de quitarle permisos a cualquier usuario registrado en el sistema.

Figura 53: Autenticación para el administrador del sistema.



The image shows a login form for the 'ValidateDocx' system. The form is centered on a light gray background. It features a dark blue header bar with the text 'ValidateDocx' in white. Below the header, there are two input fields. The first is labeled 'Nombre de usuario:' and contains the text 'julianfplata'. The second is labeled 'Contraseña:' and contains a masked password represented by seven asterisks. Below the password field is a blue button with the text 'Iniciar sesión' in white.

Fuente: El autor

El administrador podrá iniciar sesión en esta sección como se puede ver en la figura10, y luego de esto, en la figura 11 el usuario podrá ver un administrador de contenido el cual tiene acceso a todos los datos que el sistema ha almacenado, tanto de los usuarios del sistema como de los documentos cargados.

Figura 54: Menú de administración del sistema.

ValidateDocx BIENVENIDO/A, JULIANFPLA

Sitio administrativo

**APP**

Documentos	+ Añadir	Modificar
Pdfresultadoss	+ Añadir	Modificar
Rolusuarios	+ Añadir	Modificar
Usuarios	+ Añadir	Modificar
Validaciondocxs	+ Añadir	Modificar

**AUTENTICACIÓN Y AUTORIZACIÓN**

Grupos	+ Añadir	Modificar
Usuarios	+ Añadir	Modificar

**Acciones recientes**

**Mis acciones:**

- + PDF Documento
- x PDF5 Documento
- x PDF2 Documento
- x PDF1 Documento
- x PDF Documento
- + PDF5 Documento

Fuente: El autor.

### 3.4. CAMBIO DE PERMISOS DE USUARIO

Como se ha mencionado anteriormente en este documento, el administrador es el único que puede acceder a la funcionalidad de cambio de permisos del sistema para cualquier usuario, en la figura 7 se puede visualizar los datos de un usuario registrado en el sistema el cual tiene un rol de usuario normal, pero puede ser modificado a administrador o a control de calidad.

Figura 55: Modificación de un usuario.

Inicio > App > Usuarios > Julian

### Modificar usuario

Usuarioprimernombre:

Usuariosegundonombre:




Usuarioprimerapellido:

Usuariosegundoapellido:

Usuariouser:

Usuariopassword:

Usuarioemail:

Rolusuario idrolusuario:    

- 
- Usuario
- Administrador
- Control de calidad

Fuente: El autor.

### 3.5. ROLES DE USUARIO.

El sistema puede asignar o quitar roles de usuario dependiendo de las preferencias del administrador, en la figura 11 se puede ver a lista de roles que el sistema maneja.

ValidadeDocx BIENVENIDO/A, JULIANFPLATA VER EL SITIO / CAMBIAR CONTRASEÑA /

Inicio > App > Rolusuarios

Escoja rolusuario a modificar

Acción:   seleccionados 0 de 3

- RolUSUARIO
- Usuario
- Administrador
- Control de calidad

3 rolusuarios





**UNIVERSIDAD CATÓLICA**  
de Colombia

**ANEXO E  
PRUEBAS FUNCIONALES DEL SISTEMA.**

**JULIAN FELIPE PLATA ZUÑIGA**

**UNIVERSIDAD CATÓLICA DE COLOMBIA  
FACULTAD DE INGENIERÍA  
PROGRAMA DE INGENIERÍA DE SISTEMAS Y COMPUTACIÓN  
BOGOTÁ  
2020**

## Contenido

INTRODUCCIÓN .....	5
1. CASOS DE PRUEBA.....	61

## Lista de Figuras.

Figura 1: Evidencia de inicio de sesión .....	62
Figura 2: Evidencia, registro de usuario. ....	63
Figura 3: Evidencia: Edición de datos de los usuarios. ....	8
Figura 4: Evidencia: Ventana de modificación de datos del administrador. ....	9
Figura 5: Evidencia :Modificación de datos de usuario. ....	10
Figura 6: Evidencia: Lista de usuarios del sistema.....	11
Figura 7: Evidencia: Listado de documentos cargados al sistema. ....	12
Figura 8: Evidencia: Botón de descargas de evidencias. ....	14
Figura 9:Evidencia: PDF generado por ValidateDocx con los resultados de las validaciones.....	14
Figura 10:Evidencia: Ventana de carga de documentos de ValidateDocx. ....	15
Figura 11: Evidencia: Información detallada del documento después del análisis con el sistema. ....	16

## Lista de tablas

Tabla 1: Caso de prueba: Inicio de sesión. ....	61
Tabla 2: Caso de prueba: Crear cuenta. ....	62
Tabla 3: Caso de prueba: Editar cuentas de usuario. ....	8
Tabla 4: Caso de prueba: Editar administrador. ....	9
Tabla 5: Caso de prueba: Editar usuario. ....	10
Tabla 6: Caso de prueba: Seleccionar usuario. ....	11
Tabla 7 : Caso de prueba: Consultar validaciones realizadas y documentos cargados. .....	12
Tabla 8: Caso de prueba: Descargar Validación. ....	13
Tabla 9: Caso de prueba: Cargar documento. ....	15
Tabla 10: Caso de prueba: Obtener información del documento. ....	16

## 1. INTRODUCCIÓN

Este anexo se construye con el fin de documentar una serie de pruebas funcionales realizadas al sistema ValidateDocx, para verificar que el software desarrollado funciona de manera adecuada, todo esto basado en los anexos de arquitectura del sistema, de casos de uso y requerimientos funcionales.

Para asegurar que las pruebas son efectivas, se decide realizar pruebas funcionales del sistema por medio del diseño de casos de pruebas los cuales contemplan la lista completa de requerimientos funcionales y el resultado esperado que se desea obtener al ejecutar cada función, esto garantiza que el sistema cumple con el objetivo inicial de automatizar el proceso de validación y verificación de un documento de Word con las normas NTC1486, además de brindar diversas facilidades a los usuarios para administrar la información que se carga al sistema.

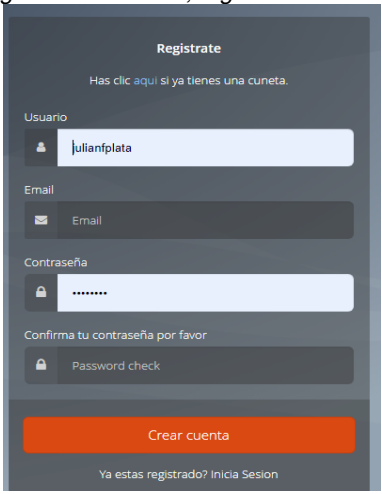
Las tablas de este documento que representan los casos de prueba realizados al sistema ValidateDocx, están basadas en las tablas del documento de plan de pruebas de la alcaldía mayor de Bogotá<sup>65</sup>, debido a que la información que se debe plasmar, representa los resultados que se desean mostrar para la ejecución de pruebas del sistema desarrollado en este proyecto.

---

<sup>65</sup> ALCALDIA MAYOR DE BOGOTÁ, FORMATOS; [Consultado el 30 de abril del 2020]. Disponible en: <http://www.saludcapital.gov.co/Formatos%20Gua%20Desarrollo%20SDS/Gu%C3%ADa%20para%20el%20documento%20de%20plan%20de%20pruebas%20de%20software.pdf>

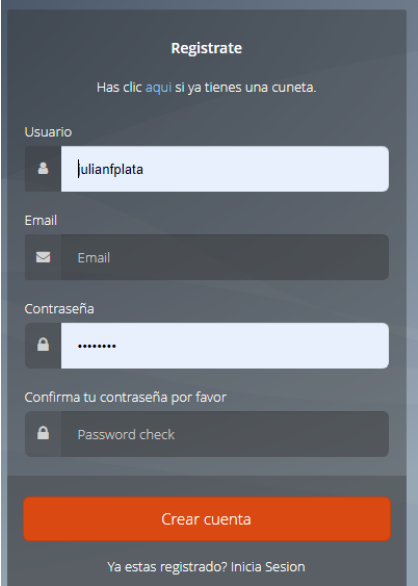
## 2. CASOS DE PRUEBA.

Tabla 45: Caso de prueba: Inicio de sesión.

<b>Identificador</b>	CP - 01	<b>Versión</b>	1.0
<b>Nombre del caso de prueba</b>	Inicio de sesión		
<b>Módulo</b>	Gestión de usuarios		
<b>Descripción de la prueba</b>			
Se validan las credenciales de ingreso del usuario para permitir o denegar el acceso al sistema.			
<b>Prerrequisitos</b>			
El usuario debe registrarse previamente en el sistema con la funcionalidad de registrar usuario.			
<b>Resultados esperados</b>			
Se almacenan los datos digitados por el usuario para que posteriormente pueda iniciar sesión el sistema.			
<b>Resultados reales</b>			
El sistema almacena exitosamente los datos del usuario, para posteriormente permitirle iniciar sesión			
<b>Imagen</b>			
<p><i>Figura 1: Evidencia, registro de usuario.</i></p> 			
Fuente: El autor			

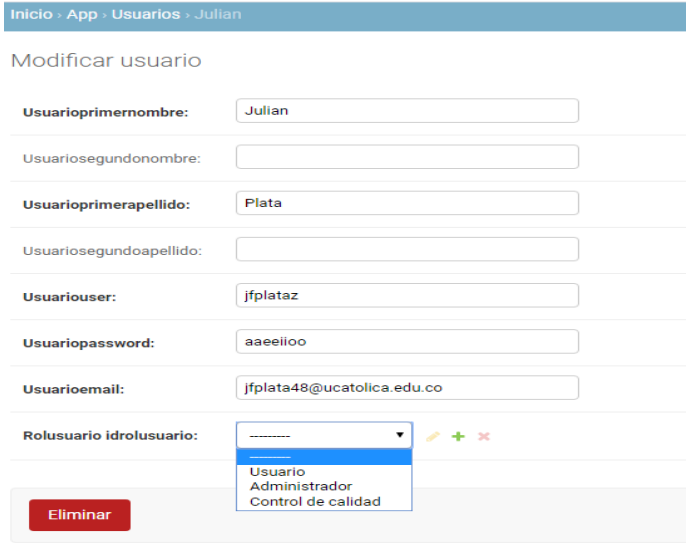
Fuente: El autor

Tabla 46: Caso de prueba: Crear cuenta.

<b>Identificador</b>	CP - 02	<b>Versión</b>	1.0
<b>Nombre del caso de prueba</b>	Crear Cuenta		
<b>Módulo</b>	Gestión de usuarios		
<b>Descripción de la prueba</b>			
El usuario ingresa sus datos personales, para que posteriormente el sistema los almacene en la base de datos y pueda iniciar sesión.			
<b>Prerrequisitos</b>			
Ninguno, cualquier usuario puede acceder a esta funcionalidad			
<b>Resultados esperados</b>			
Se almacenan los datos digitados por el usuario para que posteriormente pueda iniciar sesión el sistema.			
<b>Resultados reales</b>			
El sistema almacena exitosamente los datos del usuario, para posteriormente permitirle iniciar sesión			
<b>Imagen</b>			
<p style="text-align: center;"><i>Figura 56: Evidencia, registro de usuario.</i></p>  <p style="text-align: center;"><b>Fuente: El autor</b></p>			

Fuente: El autor

Tabla 47: Caso de prueba: Editar cuentas de usuario.

<b>Identificador</b>	CP - 03	<b>Versión</b>	1.0
<b>Nombre del caso de prueba</b>	Editar cuentas de usuario		
<b>Módulo</b>	Gestión de usuarios		
<b>Descripción de la prueba</b>			
El sistema verifica que el tipo de usuario sea administrador, para poder editar la información de cualquier usuario.			
<b>Prerrequisitos</b>			
El usuario debe iniciar sesión y ser administrador del sistema para poder acceder a los datos de cualquier usuario y editarlos.			
<b>Resultados esperados</b>			
El administrador del sistema puede modificar cualquiera de los datos de los usuarios cuando lo requiera sin ninguna restricción.			
<b>Resultados reales</b>			
El administrador puede modificar los datos de cualquier usuario a voluntad.			
<b>Imagen</b>			
<p><i>Figura 57: Evidencia: Edición de datos de los usuarios.</i></p> 			
Fuente: El autor			

Fuente: El autor



Tabla 48: Caso de prueba: Editar administrador.

<b>Identificador</b>	CP - 04	<b>Versión</b>	1.0
<b>Nombre del caso de prueba</b>	Editar administrador		
<b>Módulo</b>	Gestión de usuarios		
<b>Descripción de la prueba</b>			
El sistema verifica que el tipo de usuario sea administrador y que el perfil que va a editar sea el propio para modificar sus datos.			
<b>Prerrequisitos</b>			
Como primera condición, el usuario debe ser administrador del sistema para acceder a esta función, para identificarse como administrador, debe iniciar sesión en el sistema.			
<b>Resultados esperados</b>			
El usuario administrador puede modificar cualquiera de sus datos cuando lo requiera sin ninguna restricción.			
<b>Resultados reales</b>			
El usuario administrador, puede modificar tanto el módulo de administración del sitio, como los datos propios, además de los permisos del sistema.			
<b>Imagen</b>			
<p><i>Figura 58: Evidencia: Ventana de modificación de datos del administrador.</i></p>			
Fuente: El autor			

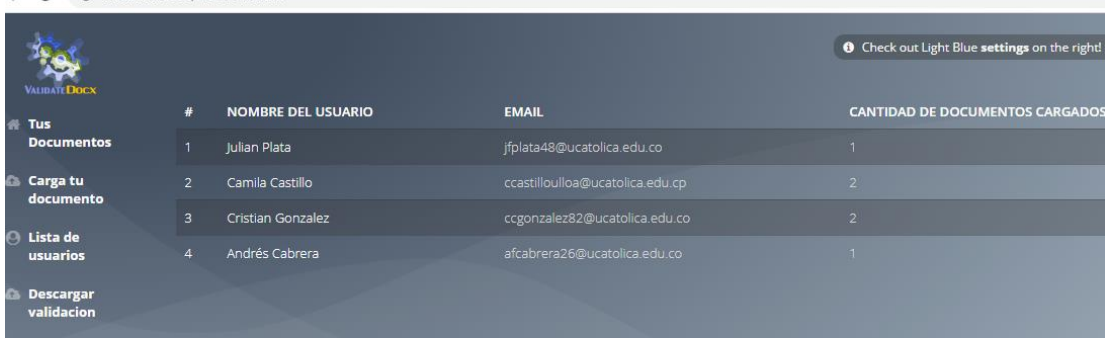
Fuente: El autor

Tabla 49: Caso de prueba: Editar usuario.

<b>Identificador</b>	CP – 05	<b>Versión</b>	1.0
<b>Nombre del caso de prueba</b>	Editar Usuario		
<b>Módulo</b>	Gestión de usuarios		
<b>Descripción de la prueba</b>			
El sistema verifica que el usuario esté registrado en la base de datos para que pueda acceder a sus datos personales, y le permita modificarlos.			
<b>Prerrequisitos</b>			
El usuario debe iniciar sesión previamente en el sistema para poder acceder a sus datos y editarlos.			
<b>Resultados esperados</b>			
El usuario puede modificar cualquiera de sus datos cuando lo requiera sin ninguna restricción.			
<b>Resultados reales</b>			
El usuario puede modificar sus propios datos cuando lo desee.			
<b>Imagen</b>			
<p><i>Figura 59: Evidencia :Modificación de datos de usuario.</i></p> <p>Modificar usuario</p> <p>Nombre de usuario: <input type="text" value="julianfplata"/>  <small>Requerido. 150 caracteres como máximo. Únicamente letras, dígitos y @./+/_</small></p> <p>Contraseña: <input type="password" value="algoritmo: pbkdf2_sha256 iteraciones: 180000 sal: WPB0nm***** funci"/>  <small>Las contraseñas no se almacenan en bruto, así que no hay manera de ver la contraseñ</small></p> <p><b>Información personal</b></p> <p>Nombre: <input type="text"/></p> <p>Apellidos: <input type="text"/></p> <p>Dirección de correo electrónico: <input type="text" value="jfplata48@ucatolica.edu.co"/></p> <p><b>Permisos</b></p> <p>Fuente: El autor</p>			

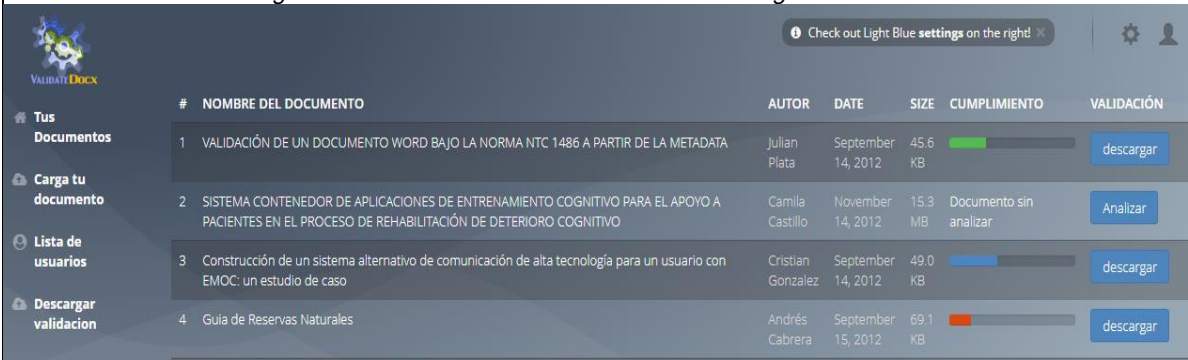
Fuente: El autor

Tabla 50: Caso de prueba: Seleccionar usuario.

<b>Identificador</b>	CP - 06	<b>Versión</b>	1.0
<b>Nombre del caso de prueba</b>	Seleccionar Usuario		
<b>Módulo</b>	Gestión usuarios		
<b>Descripción de la prueba</b>			
El sistema verifica que el usuario que ingrese a esta funcionalidad sea coordinador de calidad o administrador para buscar y seleccionar a uno de los usuarios que han utilizado la herramienta.			
<b>Prerrequisitos</b>			
El usuario debe iniciar sesión previamente en el sistema además de tener el rol de administrador o coordinador de calidad para poder acceder a la lista de los usuarios y seleccionarlos.			
<b>Resultados esperados</b>			
El usuario puede consultar la lista de los usuarios del sistema y ver sus actividades realizadas recientemente con los documentos cargados cuando lo requiera sin ninguna restricción.			
<b>Resultados reales</b>			
El sistema muestra correctamente, la lista de usuarios registrados y activos en el sistema.			
<b>Imagen</b>			
<p><i>Figura 60: Evidencia: Lista de usuarios del sistema.</i></p> 			
Fuente: El autor			

Fuente: El autor

Tabla 51 : Caso de prueba: Consultar validaciones realizadas y documentos cargados.

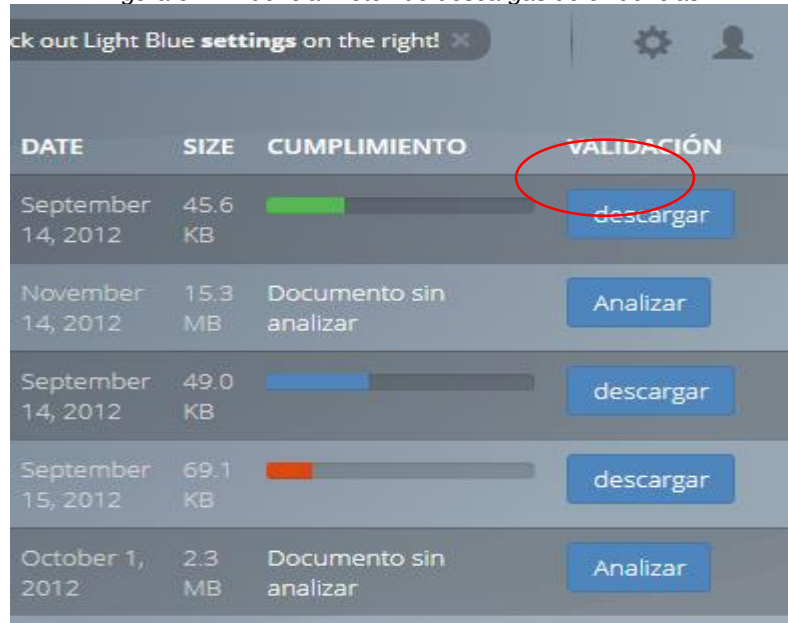
<b>Identificador</b>	CP - 07	<b>Versión</b>	1.0
<b>Nombre del caso de prueba</b>	Consultar Validaciones Realizadas y documentos cargados		
<b>Módulo</b>	Gestor Documental		
<b>Descripción de la prueba</b>			
El sistema verifica que el usuario logueado sea el autor de el listado de las validaciones que el sistema mostrará.			
<b>Prerrequisitos</b>			
El usuario debe iniciar sesión previamente en el sistema además de tener el rol de administrador o coordinador de calidad para poder acceder a la lista de todas las validaciones que se han realizado en el sistema.			
<b>Resultados esperados</b>			
El usuario puede consultar la lista de las validaciones realizadas por los usuarios con los documentos cargados cuando lo requiera sin ninguna restricción.			
<b>Resultados reales</b>			
El sistema retorna la lista de documentos y sus respectivos usuarios, que han cargado los documentos al sistema,			
<b>Imagen</b>			
<p><i>Figura 61: Evidencia: Listado de documentos cargados al sistema.</i></p> 			
<p>Fuente: El autor</p>			

Fuente: El autor

Tabla 52: Caso de prueba: Descargar Validación.

<b>Identificador</b>	CP - 08	<b>Versión</b>	1.0
<b>Nombre del caso de prueba</b>	Descargar Validación.		
<b>Módulo</b>	Gestor Documental		
<b>Descripción de la prueba</b>			
<p>El sistema verifica que el usuario logueado sea el coordinador de calidad o un administrador para descargar el archivo que contiene el resumen de las validaciones realizadas por el sistema a un documento.</p>			
<b>Prerrequisitos</b>			
<p>El usuario debe iniciar sesión previamente en el sistema además de tener el rol de administrador o coordinador de calidad para poder acceder a la lista de todas las validaciones que se han realizado en el sistema y descargar las validaciones que requiera.</p>			
<b>Resultados esperados</b>			
<p>El usuario puede consultar la lista de las validaciones realizadas por los usuarios con los documentos cargados cuando lo requiera sin ninguna restricción y descargarlas en un documento PDF.</p>			
<b>Resultados reales</b>			
<p>El sistema generó un PDF con las conclusiones de la validación de los documentos.</p>			
<b>Imagen</b>			

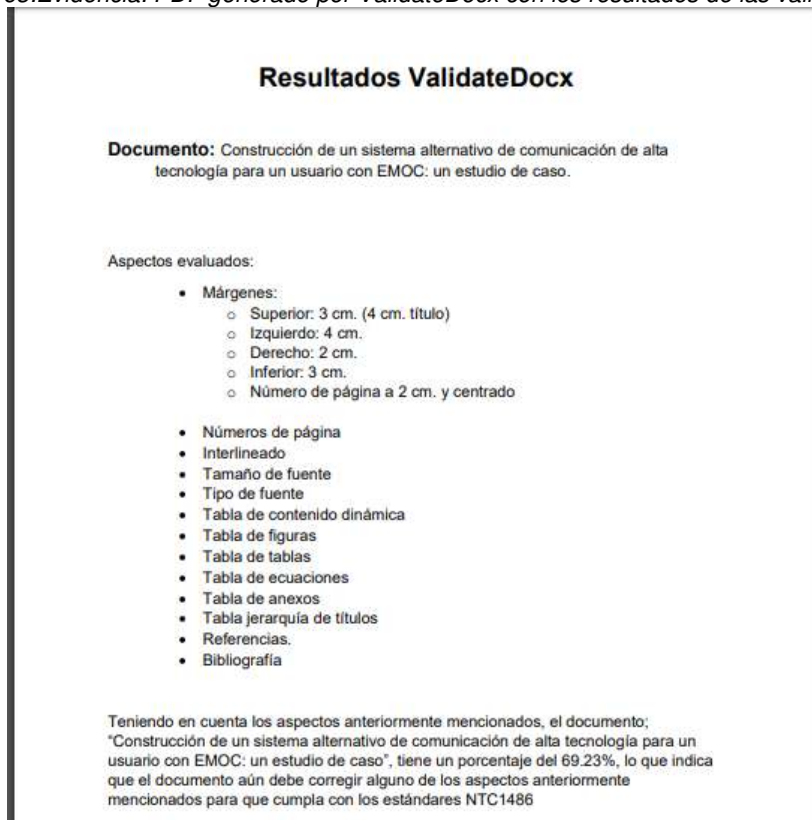
Figura 62: Evidencia: Botón de descargas de evidencias.



DATE	SIZE	CUMPLIMIENTO	VALIDACIÓN
September 14, 2012	45.6 KB	<div style="width: 50%; background-color: green;"></div>	descargar
November 14, 2012	15.3 MB	Documento sin analizar	Analizar
September 14, 2012	49.0 KB	<div style="width: 75%; background-color: blue;"></div>	descargar
September 15, 2012	69.1 KB	<div style="width: 30%; background-color: orange;"></div>	descargar
October 1, 2012	2.3 MB	Documento sin analizar	Analizar

Fuente: el autor

Figura 63: Evidencia: PDF generado por ValidateDocx con los resultados de las validaciones



**Resultados ValidateDocx**

**Documento:** Construcción de un sistema alternativo de comunicación de alta tecnología para un usuario con EMOC: un estudio de caso.

Aspectos evaluados:

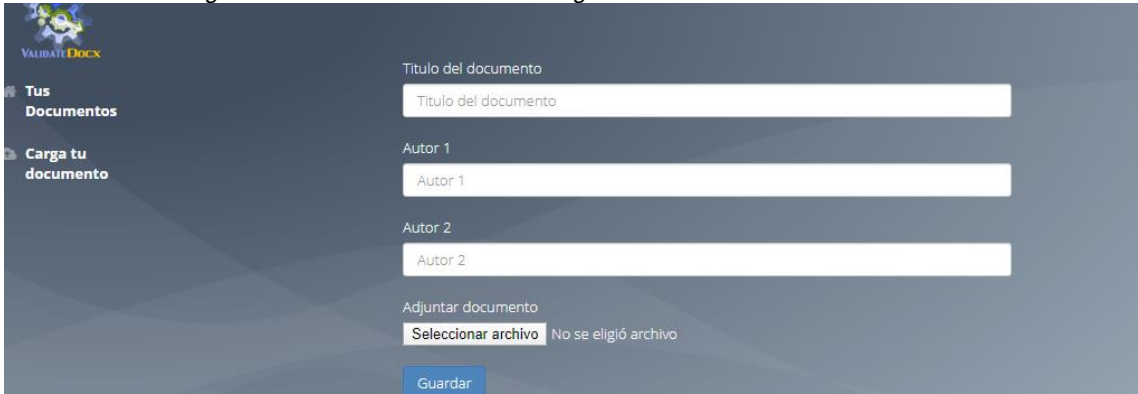
- Márgenes:
  - Superior: 3 cm. (4 cm. título)
  - Izquierdo: 4 cm.
  - Derecho: 2 cm.
  - Inferior: 3 cm.
  - Número de página a 2 cm. y centrado
- Números de página
- Interlineado
- Tamaño de fuente
- Tipo de fuente
- Tabla de contenido dinámica
- Tabla de figuras
- Tabla de tablas
- Tabla de ecuaciones
- Tabla de anexos
- Tabla jerarquía de títulos
- Referencias.
- Bibliografía

Teniendo en cuenta los aspectos anteriormente mencionados, el documento; "Construcción de un sistema alternativo de comunicación de alta tecnología para un usuario con EMOC: un estudio de caso", tiene un porcentaje del 69.23%, lo que indica que el documento aún debe corregir alguno de los aspectos anteriormente mencionados para que cumpla con los estándares NTC1486

Fuente: El autor

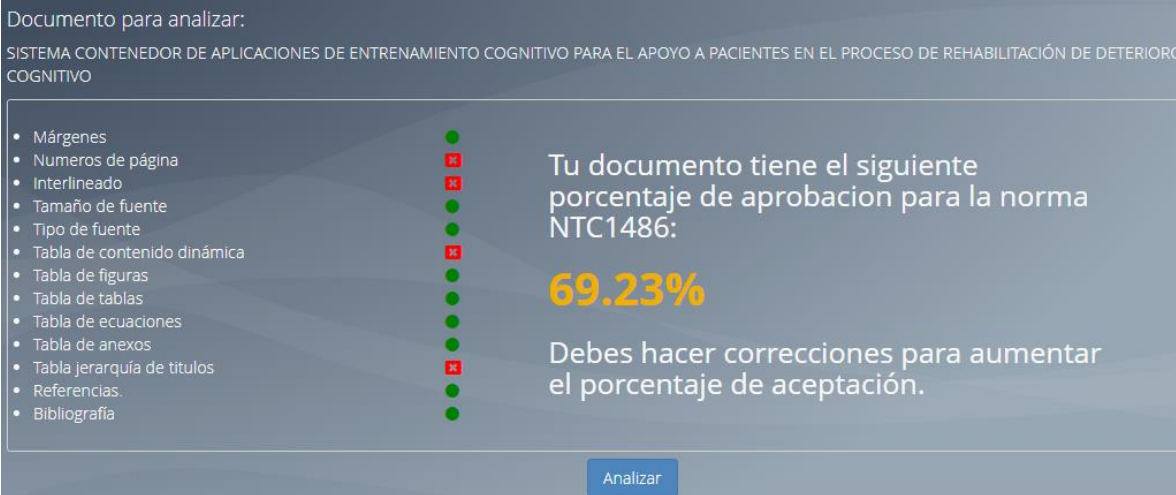
Fuente: El autor

Tabla 53: Caso de prueba: Cargar documento.

<b>Identificador</b>	CP - 09	<b>Versión</b>	1.0
<b>Nombre del caso de prueba</b>	Cargar documento		
<b>Módulo</b>	Gestor Documental		
<b>Descripción de la prueba</b>			
El sistema tendrá la opción para que el usuario pueda adjuntar uno o más documentos de formato DOC.			
<b>Prerrequisitos</b>			
El usuario debe iniciar sesión previamente en el sistema para poder cargar los documentos que requiera.			
<b>Resultados esperados</b>			
El usuario puede cargar tanto los documentos como la información relacionada con los documentos cuando lo requiera.			
<b>Resultados reales</b>			
El sistema almacena correctamente tanto el documento, como los datos relacionados con el documento.			
<b>Imagen</b>			
<p><i>Figura 64:Evidencia: Ventana de carga de documentos de ValidateDocx.</i></p> 			
<p>Fuente: El autor.</p>			

Fuente: El autor

Tabla 54: Caso de prueba: Obtener información del documento.

<b>Identificador</b>	CP - 10	<b>Versión</b>	1.0
<b>Nombre del caso de prueba</b>	Obtener información del documento.		
<b>Módulo</b>	Gestor Documental		
<b>Descripción de la prueba</b>			
El sistema realizará una serie de validaciones sobre el archivo de formato DOCX las cuales permiten concluir si el documento cumple o no con la norma NTC1486, además se retorna la lista de características del documento que cumplen y las que no.			
<b>Prerrequisitos</b>			
El usuario debe iniciar sesión previamente en el sistema para poder ejecutar la función del sistema que permite retornar las características del documento que cumplen y no cumplen con la norma NTC1486.			
<b>Resultados esperados</b>			
El usuario puede visualizar el porcentaje del cumplimiento que tiene el documento, además de un listado de características que cumplen y no cumplen con la norma NTC1486.			
<b>Resultados reales</b>			
El sistema retorna una lista completa de todos los ítems que se evalúan a la hora de analizar los documentos, junto con un porcentaje en general que le indica al usuario cual es el nivel de cumplimiento que tiene su documento, respecto a las normas de la NTC1486.			
<b>Imagen</b>			
<p><i>Figura 65: Evidencia: Información detallada del documento después del análisis con el sistema.</i></p>  <p>Documento para analizar: SISTEMA CONTENEDOR DE APLICACIONES DE ENTRENAMIENTO COGNITIVO PARA EL APOYO A PACIENTES EN EL PROCESO DE REHABILITACIÓN DE DETERIORO COGNITIVO</p> <ul style="list-style-type: none"> <li>• Márgenes</li> <li>• Numeros de página</li> <li>• Interlineado</li> <li>• Tamaño de fuente</li> <li>• Tipo de fuente</li> <li>• Tabla de contenido dinámica</li> <li>• Tabla de figuras</li> <li>• Tabla de tablas</li> <li>• Tabla de ecuaciones</li> <li>• Tabla de anexos</li> <li>• Tabla jerarquía de títulos</li> <li>• Referencias.</li> <li>• Bibliografía</li> </ul> <p>Tu documento tiene el siguiente porcentaje de aprobación para la norma NTC1486: <b>69.23%</b></p> <p>Debes hacer correcciones para aumentar el porcentaje de aceptación.</p> <p>Analizar</p>			
Fuente: El autor			

Fuente: El auto



## 2. BIBLIOGRAFIA

ALCALDIA MAYOR DE BOGOTÁ, FORMATOS; [Consultado el 30 de abril del 2020].

Disponible en:

<http://www.saludcapital.gov.co/Formatos%20Gua%20Desarrollo%20SDS/Gu%C3%ADa%20para%20el%20documento%20de%20plan%20de%20pruebas%20de%20software.pdf>