



Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Ingeniería en Ciencias y Sistemas

**PLATAFORMA WEB PARA LA INDEXACIÓN Y BÚSQUEDA DE INFORMACIÓN EN EL
INSTITUTO DE PROBLEMAS NACIONALES DE LA UNIVERSIDAD DE SAN CARLOS DE
GUATEMALA (IPNUSAC)**

Gustavo Adolfo Siguantay Chiché

Asesorado por el Ing. José Samuel Bulux Calderón

Guatemala, septiembre de 2017

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE INGENIERÍA

**PLATAFORMA WEB PARA LA INDEXACIÓN Y BÚSQUEDA DE INFORMACIÓN EN EL
INSTITUTO DE PROBLEMAS NACIONALES DE LA UNIVERSIDAD DE SAN CARLOS DE
GUATEMALA (IPNUSAC)**

TRABAJO DE GRADUACIÓN

PRESENTADO A LA JUNTA DIRECTIVA DE LA
FACULTAD DE INGENIERÍA
POR

GUSTAVO ADOLFO SIGUANTAY CHICHÉ
ASESORADO POR EL ING. JOSÉ SAMUEL BULUX CALDERÓN

AL CONFERÍRSELE EL TÍTULO DE

INGENIERO EN CIENCIAS Y SISTEMAS

GUATEMALA, SEPTIEMBRE DE 2017

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
FACULTAD DE INGENIERÍA



NÓMINA DE JUNTA DIRECTIVA

DECANO	Ing. Pedro Antonio Aguilar Polanco
VOCAL I	Ing. Angel Roberto Sic García
VOCAL II	Ing. Pablo Christian de León Rodríguez
VOCAL III	Ing. José Milton de León Bran
VOCAL IV	Br. Jurgen Andoni Ramírez Ramírez
VOCAL V	Br. Oscar Humberto Galicia Nuñez
SECRETARIA	Inga. Lesbia Magalí Herrera López

TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO

DECANO	Ing. Pedro Antonio Aguilar Polanco
EXAMINADOR	Ing. Marlon Antonio Pérez Türk
EXAMINADOR	Ing. Sergio Leonel Gómez Bravo
EXAMINADORA	Inga. Floriza Felipa Ávila Pesquera
SECRETARIA	Inga. Lesbia Magalí Herrera López

HONORABLE TRIBUNAL EXAMINADOR

En cumplimiento con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

**PLATAFORMA WEB PARA LA INDEXACIÓN Y BÚSQUEDA DE INFORMACIÓN EN EL
INSTITUTO DE PROBLEMAS NACIONALES DE LA UNIVERSIDAD DE SAN CARLOS DE
GUATEMALA (IPNUSAC)**

Tema que me fuera asignado por la Dirección de la Escuela de Ingeniería en Ciencias y Sistemas, con fecha 5 de marzo de 2016.



Gustavo Adolfo Siguantay Chiché

Guatemala, 3 de agosto de 2017

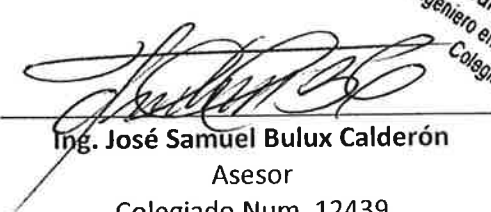
Inga. Christa del Rosario Classon de Pinto
Director de la Unidad de EPS
Facultad de Ingeniería
Universidad de San Carlos de Guatemala

Estimada Ingeniera Christa del Rosario Classon de Pinto:

Por este medio atentamente le comunico que el INFORME FINAL de la Práctica del Ejercicio Profesional Supervisado (E.P.S.) del estudiante universitario **GUSTAVO ADOLFO SIGUANTAY CHICHÉ** quien se identifica con registro académico **199919582** y CUI **1942125480101**, titulado **“PLATAFORMA WEB PARA LA INDEXACIÓN Y BÚSQUEDA DE INFORMACIÓN EN EL INSTITUTO DE PROBLEMAS NACIONALES DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA (IPNUSAC)”**, fue revisado y aprobado, solicitándole darles el trámite respectivo.

Sin otro particular, me es grato suscribirme.

Atentamente,


Ing. José Samuel Bulux Calderón
Asesor
Colegiado Num. 12439

*José Samuel Bulux Calderón
Ingeniero en Ciencias y Sistemas
Colegiado 12439*



Guatemala, 08 de agosto de 2017.
REF.EPS.DOC.518.08.2017.

Inga. Christa Classon de Pinto
Directora Unidad de EPS
Facultad de Ingeniería
Presente

Estimada Ingeniera Classon de Pinto:

Por este medio atentamente le informo que como Supervisora de la Práctica del Ejercicio Profesional Supervisado, (E.P.S) del estudiante universitario de la Carrera de Ingeniería en Ciencias y Sistemas, **Gustavo Adolfo Siguantay Chiché, Registro Académico 199919582 y CUI 1942 12548 0101** procedí a revisar el informe final, cuyo título es **PLATAFORMA WEB PARA LA INDEXACIÓN Y BÚSQUEDA DE INFORMACIÓN EN EL INSTITUTO DE PROBLEMAS NACIONALES DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA (IPNUSAC)**.

En tal virtud, **LO DOY POR APROBADO**, solicitándole darle el trámite respectivo.

Sin otro particular, me es grato suscribirme.

Atentamente,

"Id y Enseñad a Todos"

Inga. Floriza Felipa Ávila Pesquera de Medina
Supervisora de EPS
Área de Ingeniería en Ciencias y Sistemas



FFAPdM/RA



Guatemala, 08 de agosto de 2017.
REF.EPS.D.245.08.2017.

Ing. Marlon Antonio Pérez Turk
Director Escuela de Ingeniería Ciencias y Sistemas
Facultad de Ingeniería
Presente

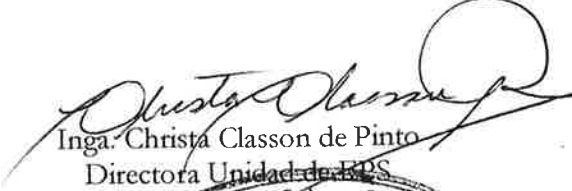
Estimado Ingeniero Perez Turk:

Por este medio atentamente le envío el informe final correspondiente a la práctica del Ejercicio Profesional Supervisado, (E.P.S) titulado **PLATAFORMA WEB PARA LA INDEXACIÓN Y BÚSQUEDA DE INFORMACIÓN EN EL INSTITUTO DE PROBLEMAS NACIONALES DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA (IPNUSAC)**, que fue desarrollado por el estudiante universitario **Gustavo Adolfo Siguantay Chiché, Registro Académico 199919582 y CUI 1942 12548 0101** quien fue debidamente asesorado por el Ing. José Samuel Bulux Calderón y supervisado por la Inga. Floriza Felipa Ávila Pesquera de Medinilla.

Por lo que habiendo cumplido con los objetivos y requisitos de ley del referido trabajo y existiendo la aprobación del mismo por parte de la Asesora y la Supervisora de EPS, en mi calidad de Director apruebo su contenido solicitándole darle el trámite respectivo.

Sin otro particular, me es grato suscribirme.

Atentamente,
"Id y Enseñad a Todos"


Inga. Christa Classon de Pinto
Directora Unidad de EPS

CCsP/ra





Universidad San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Ingeniería en Ciencias y Sistemas

Guatemala, 9 de Agosto de 2017


Ingeniero
Marlon Antonio Pérez Türk
Director de la Escuela de Ingeniería
En Ciencias y Sistemas

Respetable Ingeniero Pérez:

Por este medio hago de su conocimiento que he revisado el trabajo de graduación-EPS del estudiante **GUSTAVO ADOLFO SIGUANTAY CHICHÉ** carné **199919582** y CUI **1942 12548 0101**, titulado: **"PLATAFORMA WEB PARA LA INDEXACIÓN Y BÚSQUEDA DE INFORMACIÓN EN EL INSTITUTO DE PROBLEMAS NACIONALES DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA (IPNUSAC)"** y a mi criterio el mismo cumple con los objetivos propuestos para su desarrollo, según el protocolo.

Al agradecer su atención a la presente, aprovecho la oportunidad para suscribirme,

Atentamente,


Ing. Carlos Alfredo Azurdia
Coordinador de Privados
y Revisión de Trabajos de Graduación



UNIVERSIDAD DE SAN CARLOS
DE GUATEMALA



FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA EN
CIENCIAS Y SISTEMAS
TEL: 24767644

*El Director de la Escuela de Ingeniería en Ciencias y Sistemas de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer el dictamen del asesor con el visto bueno del revisor y del Licenciado en Letras, del trabajo de graduación **“PLATAFORMA WEB PARA LA INDEXACIÓN Y BÚSQUEDA DE INFORMACIÓN EN EL INSTITUTO DE PROBLEMAS NACIONALES DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA (IPNUSAC)”**, realizado por el estudiante GUSTAVO ADOLFO SIGUANTAY CHICHÉ aprueba el presente trabajo y solicita la autorización del mismo.*

“ID Y ENSEÑAD A TODOS”

Ing.  Pérez Türk
Director

Escuela de Ingeniería en Ciencias y Sistemas



Guatemala, 01 de septiembre de 2017

Universidad de San Carlos
de Guatemala



Facultad de Ingeniería
Decanato

Ref.DTG.D.387.2017

El Decano de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Ingeniería en Ciencias y Sistemas, al trabajo de graduación titulado: **PLATAFORMA WEB PARA LA INDEXACIÓN Y BÚSQUEDA DE INFORMACIÓN EN EL INSTITUTO DE PROBLEMAS NACIONALES DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA (IPNUSAC)**, presentado por el estudiante universitario: **Gustavo Adolfo Siguantay Chiché**, y después de haber culminado las revisiones previas bajo la responsabilidad de las instancias correspondientes, se autoriza la impresión del mismo.

IMPRÍMASE.


Ing. Pedro Antonio Aguilar Polanco
Decano



Guatemala, septiembre de 2017

/cc

ACTO QUE DEDICO A:

Dios	Por ser su voluntad permitirme alcanzar esta meta en mi carrera.
Mis padres	Carmen e Ismael, por todo el esfuerzo y apoyo incondicional.
Mis hermanas	Ivón y Jessica por ser un gran apoyo e inspiración.
Mi novia	Por sus palabras de aliento, por su amor y cariño.

AGRADECIMIENTOS A:

Universidad de San Carlos de Guatemala	Por ser mi casa de formación académica profesional.
Facultad de Ingeniería	Porque en sus aulas adquirí y desarrollé los conocimientos para lograr ser un profesional.
Mis amigos de la Facultad	Por compartir sus conocimientos.
IPNUSAC	Institución que me permitió realizar mi trabajo de graduación y a la persona responsable que me brindó el apoyo en todo el proceso.
Las instituciones en donde he laborado	Por brindarme el apoyo para finalizar mis estudios.

ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES.....	V
LISTA DE SÍMBOLOS	VII
GLOSARIO	IX
RESUMEN.....	XI
OBJETIVOS.....	XIII
INTRODUCCIÓN	XV
1. FASE DE INVESTIGACIÓN	1
1.1. Antecedentes de la institución	1
1.1.1. Reseña histórica	1
1.1.2. Misión	2
1.1.3. Visión.....	2
1.1.4. Servicios que realiza.....	2
1.2. Descripción de las necesidades	3
1.2.1. Categorizar las fuentes de información	4
1.2.1.1. Instituciones.....	4
1.2.1.2. Blogs.....	4
1.2.1.3. Base de datos.....	4
1.2.1.4. Revistas	4
1.2.1.5. Estudios.....	5
1.2.1.6. IPNUSAC.....	5
1.2.2. Definición de repositorio de documentos	5
1.2.2.1. Repositorio de documentos	5
1.2.2.2. Repositorio de índices	5
1.2.3. Búsqueda de la información	6

1.2.4.	Administración y configuración.....	6
1.3.	Priorización de necesidades	6
2.	FASE TÉCNICO-PROFESIONAL.....	7
2.1.	Descripción del proyecto.....	7
2.2.	Investigación preliminar para la solución del proyecto	8
2.2.1.	Metodologías de desarrollo de software.....	8
2.2.1.1.	Metodología en cascada	9
2.2.1.2.	Metodología de prototipos	9
2.2.1.3.	Metodología incremental o iterativo.....	10
2.2.1.4.	Metodología SCRUM	10
2.2.1.5.	Metodología Kanban	11
2.2.1.6.	Metodología XP	11
2.2.1.7.	Metodología de proceso unificado ágil.....	11
2.2.2.	Patrón de diseño MVC	12
2.2.3.	<i>Frameworks</i> para desarrollo web	12
2.2.3.1.	Laravel.....	13
2.2.3.2.	CodeIgniter	13
2.2.3.3.	Yii	13
2.2.3.4.	Struts 2.....	14
2.2.3.5.	Spring MVC.....	14
2.2.4.	<i>Frameworks</i> de diseño web.....	15
2.2.4.1.	Bootstrap.....	15
2.2.4.2.	jQuery UI	15
2.2.5.	Motor de indexación y búsqueda.....	16
2.2.5.1.	LUCENE.....	17
2.2.5.2.	XAPIAN	17
2.2.5.3.	SWISH-E	18

2.3.	Presentación de la solución del proyecto	18
2.3.1.	Tecnologías para el desarrollo.....	19
2.3.2.	Infraestructura.....	20
2.3.3.	Arquitectura	21
2.3.4.	Ciclo de vida proceso unificado ágil.....	22
2.3.4.1.	Iniciación.....	22
2.3.4.2.	Elaboración.....	23
2.3.4.3.	Construcción.....	23
2.3.4.4.	Transición	23
2.3.5.	Solución presentada	24
2.3.5.1.	Inicio	24
2.3.5.2.	Búsqueda.....	25
2.3.5.2.1.	General.....	25
2.3.5.2.2.	Directa	25
2.3.5.2.3.	Resultados.....	26
2.3.5.3.	Documentos.....	27
2.3.5.3.1.	Carga masiva	27
2.3.5.3.2.	Bandeja	28
2.3.5.4.	Indexación	29
2.3.5.4.1.	Plantilla.....	30
2.3.5.5.	Seguridad	31
2.3.5.5.1.	Usuarios	31
2.3.5.6.	Parametrización.....	32
2.3.5.6.1.	Catálogos	32
2.3.5.7.	Lucene.....	35
2.3.5.7.1.	Configuración	35
2.4.	Costos del proyecto.....	37
2.5.	Beneficios del proyecto.....	37

3.	FASE DE ENSEÑANZA-APRENDIZAJE	39
3.1.	Capacitación	39
3.1.1.	Objetivo	39
3.1.2.	Acciones y medios	39
3.1.2.1.	Guías de usuario	40
3.1.2.2.	Capacitación presencial	40
3.2.	Material elaborado.....	40
	CONCLUSIONES.....	43
	RECOMENDACIONES	45
	BIBLIOGRAFÍA.....	47

ÍNDICE DE ILUSTRACIONES

FIGURAS

1.	Infraestructura	20
2.	Arquitectura MVC	22
3.	Ingreso de usuario.....	24
4.	Búsqueda general	25
5.	Búsqueda directa	26
6.	Resultado de la búsqueda.....	26
7.	Carga masiva	27
8.	Bandeja de documentos.....	28
9.	Formulario de indexación	29
10.	Lista de plantillas.....	30
11.	Plantilla de indexación.....	30
12.	Lista de usuarios	31
13.	Formulario datos de usuario.....	32
14.	Lista de catálogos	33
15.	Formulario de catálogo.....	33
16.	Lista detalle de catálogo.....	34
17.	Formulario detalle de catálogo	34
18.	Lista de parámetros de Lucene	35
19.	Formulario de parámetros	36

TABLAS

I.	Costos del proyecto.....	37
----	--------------------------	----

LISTA DE SÍMBOLOS

Símbolo	Significado
Q	Moneda quetzal.

GLOSARIO

Arquitectura	Es el diseño o conjunto de relaciones entre las partes que constituyen un sistema.
Carga	Proceso por el cual se realiza el ingreso de un documento.
Catálogo	Lista de opciones dentro del sistema.
Credenciales	Valores asociados a un usuario, rol o perfil, nombre de usuario y contraseña.
Icono	Imagen dentro del sistema que refleja una funcionalidad.
IndBusInf	Nombre de la plataforma, indexación y búsqueda de información.
Indexación	Ordenación de los documentos con base en una serie de datos para facilitar su consulta y análisis.
IPNUSAC	Instituto de Problemas Nacionales de la Universidad de San Carlos de Guatemala.
Módulo	Parte del sistema que engloba una funcionalidad específica.

Plantilla

Definición de una estructura o forma que identifica un tipo de documento.

Repositorio

Ubicación física dentro de un disco duro donde se almacenan o guardan archivos.

RESUMEN

Este proyecto se realiza para contribuir con el Instituto de Problemas Nacionales de la Universidad de San Carlos de Guatemala (IPNUSAC), cuyo objetivo y responsabilidad como parte de la USAC es cooperar con el estudio y solución de problemas nacionales, expresando opiniones y propuestas particulares de la problemática que se trate.

El proyecto consiste en el desarrollo e implementación de una plataforma web para la indexación y búsqueda de información, realizado a la medida de las necesidades actuales y futuras.

Para conocer la situación actual y flujos de procesos que siguen los investigadores para realizar estudios y proponer soluciones, se realizó un diagnóstico, con el objetivo de definir los requerimientos de una solución para gestión documental que cumpla y satisfaga las necesidades de los investigadores.

Se desarrolla e implementa una plataforma web con la funcionalidad principal de recuperar documentos, mediante el uso de un motor de recuperación basado en texto. Se complementa con módulos de gestión administrativa y de configuración.

OBJETIVOS

General

Implementar una plataforma web para mantener organizadas y actualizadas las fuentes de estudio de las diferentes áreas de trabajo del IPNUSAC y proveer de recuperación de la información mediante una estructura de indexación de los tipos de referencia.

Específicos

1. Realizar un análisis de los procesos que se realizan en la búsqueda de información de las investigaciones por parte de los profesionales.
2. Determinar los actores y sus funciones.
3. Definir y dividir las áreas de referencia de fuentes de información.
4. Establecer los tipos de recursos por manejar e indexar.
5. Diseñar una arquitectura de biblioteca de los recursos y una arquitectura de indexación.
6. Diseñar e implementar un sistema para la recuperación de los recursos almacenados mediante el uso de un motor de recuperación basado en texto.

7. Permitir búsquedas relativamente complejas sobre los datos.
8. Diseñar e implementar el sistema basado en patrones básicos de uso.
9. Aplicar el sistema para las ocho áreas de trabajo del IPNUSAC.
10. Diseñar e implementar un repositorio de las fuentes de información.
11. Facilitar la búsqueda y recuperación de la información.

INTRODUCCIÓN

El Instituto de Análisis e Investigación de los Problemas Nacionales de la Universidad de San Carlos de Guatemala (IPNUSAC) tiene como objetivo y responsabilidad de parte de la universidad, cooperar con el estudio y solución de los problemas nacionales, expresando opiniones y propuestas particulares de la problemática que se trate.

Para realizar estudios, proponer iniciativas, formular propuestas, desarrollar estudios de caso y emitir opiniones sobre temas de trascendencia e impacto nacional, los profesionales especializados realizan las investigaciones en fuentes de referencia compuestas por páginas web, blogs científicos, libros y revistas propias, bases de datos locales e internacionales. En la mayoría de casos, estas fuentes están desactualizadas, no categorizadas y, además, desordenadas.

Se necesita que las referencias y sus distintos tipos estén ordenadas y sean de fácil acceso para cada una de las ocho áreas de trabajo del IPNUSAC.

Se propone como solución, una plataforma web para la indexación y búsqueda de la información. La función principal del sistema es la recuperación de documentos de texto y hacer uso de las ventajas ofrecidas por los motores de recuperación basados en texto.

Se implementará una estructura de indexación como solución para la recuperación de la información, con el objetivo de encontrar de manera sencilla,

fácil y automática la información, utilizando consultas en forma de lenguaje natural.

La indexación ayudará a representar de una forma apropiada los documentos y referencias web, seleccionando los términos que identifiquen a los recursos y así facilitar la recuperación de la información.

Para su realización e implementación se utilizará un motor para la recuperación, basado en texto.

La plataforma está seccionada por áreas de trabajo y subdividida en los tipos de áreas de referencia, las cuales serán alimentadas por los usuarios para la búsqueda de la información, definidas en módulos de administración y de configuración.

1. FASE DE INVESTIGACIÓN

1.1. Antecedentes de la institución

El IPNUSAC es un centro académico de carácter multidisciplinario que asesora al Consejo Superior Universitario (CSU) y al rector en el tratamiento de los asuntos relevantes de la vida democrática del país.

1.1.1. Reseña histórica

Durante 2007 se realizó el diseño de una propuesta, crear una entidad dedicada a la reflexión y propuesta de solución de los problemas nacionales, incluyendo la utilización del derecho que tiene la USAC de presentar ante el Congreso iniciativas de ley.

Durante el segundo semestre de 2008, coincidiendo con un requerimiento del Consejo Superior Universitario, que solicitó realizar gestiones para la creación de una unidad ágil con capacidad de análisis, propuesta oportuna y pertinente en el tiempo y el contexto social guatemalteco. Se presenta formalmente ante ese cuerpo colegiado, el documento que plantea la creación del Instituto.

En septiembre de 2008, el Consejo Superior Universitario de la USAC, según consta en el punto sexto, inciso 6.3, del Acta No. 23-2008 de su sesión ordinaria celebrada el 24 del mes referido, creó el Instituto de Análisis e Investigación de los Problemas Nacionales, quedando aprobados: su base legal, el marco referencial y sociocontextual al que responde, así como la

justificación de la necesidad de su implementación, su estructura organizacional y los recursos necesarios para su funcionamiento.

1.1.2. Misión

Contribuir a que la Universidad de San Carlos (USAC) cumpla con el mandato constitucional de cooperar en el estudio y la solución de los problemas nacionales, a través de investigaciones científicas, propuestas institucionales y debates públicos.

1.1.3. Visión

Es la unidad académica que contribuye de manera eficaz y eficiente a que la USAC cumpla con el mandato constitucional de cooperar en el estudio y la solución de los problemas nacionales, promoviendo un Estado social, democrático y de derecho, en un clima de libertad, respeto y pluralidad.

1.1.4. Servicios que realiza

Contribuir a la solución de la problemática nacional y al desarrollo del país, en áreas estratégicas tales como la educación, salud, seguridad social, vivienda, empleo, ambiente, desarrollo rural, seguridad ciudadana, fortalecimiento del estado de derecho, entre otras.

Es responsable de realizar estudios de coyuntura, proponer iniciativas de ley, formular propuestas de políticas públicas, desarrollar estudios de caso, y emitir dictámenes u opiniones colegiadas, con el propósito de contribuir a plantear opciones de solución a problemas de trascendencia e impacto nacional, para que la USAC esté en mejores condiciones de posicionarse

propositivamente frente a los problemas nacionales de orden coyuntural y estructural, con una visión que integre el corto, mediano y largo plazo.

Asimismo, tiene como una de sus finalidades formular un plan general de desarrollo integral del país, a largo plazo, con base en las investigaciones científicas que se hagan en la USAC y en otras instituciones.

Igualmente, es responsable de promover el diálogo y la deliberación en la forma de foros, conferencias, seminarios y talleres sobre problemas y temas diversos a nivel nacional, regional e internacional. Contribuir a articular y cohesionar los distintos esfuerzos de la USAC, que tengan como eje la incidencia en temas de interés nacional, con el fin de lograr el posicionamiento institucional de forma coordinada y oportuna.

1.2. Descripción de las necesidades

Para realizar estudios y proponer iniciativas, los profesionales especializados realizan las investigaciones en fuentes de referencia formadas por páginas web, blogs científicos, libros, revistas y otro tipo de documentos. Las fuentes están desactualizadas, desordenadas y no categorizadas.

Se detectó la necesidad de que las referencias y sus distintos tipos estén ordenadas y sean de fácil acceso para cada una de las ocho áreas de trabajo del IPNUSAC.

Se identifican y clasifican las necesidades proporcionadas por los investigadores, las cuales se exponen a continuación.

1.2.1. Categorizar las fuentes de información

Las fuentes de información consultadas se categorizan por tipo de documento, asociado a la información que manejan, con el objetivo de realizar búsquedas por tema en un tipo de documento específico y agilizar las búsquedas.

Se definen seis categorías:

1.2.1.1. Instituciones

Documentos realizados por instituciones gubernamentales, dirección de páginas web asociadas y contactos de las instituciones.

1.2.1.2. Blogs

Registro de blogs científicos, marcando el tema y la dirección web.

1.2.1.3. Base de datos

Documentos en formato de hojas de cálculo que representan estadísticas y datos cuantificables.

1.2.1.4. Revistas

Documentos tipo revistas de investigación que son realizados por profesionales de distintas áreas.

1.2.1.5. Estudios

Documentos de investigación realizados por profesionales o estudiantes, que no necesariamente sean documentos formales u oficiales.

1.2.1.6. IPNUSAC

Documentos asociados a la revista y artículos realizados por la institución.

1.2.2. Definición de repositorio de documentos

Se identifica la necesidad de guardar los documentos en un servidor centralizado y de alta disponibilidad.

1.2.2.1. Repositorio de documentos

Se identifica la necesidad de que los documentos o fuentes de información digitales se almacenen en un único sitio centralizado y protegido. Se define una estructura de repositorio eficiente de acuerdo con la tipificación de los documentos y áreas de estudio de la institución.

1.2.2.2. Repositorio de índices

En el proceso de indexación de los documentos se creará un repositorio para el control de índices. Se define una estructura eficiente de acuerdo con el repositorio de documentos.

1.2.3. Búsqueda de la información

En la búsqueda de la información, se analizan e identifican dos tipos de búsqueda:

- General: búsqueda dentro del contenido de los documentos, mediante consultas en forma de lenguaje natural.
- Específica: búsqueda directa por los datos que identifican a un tipo de documento.

1.2.4. Administración y configuración

Para el mantenimiento y funcionamiento correcto de la plataforma, se necesita de funcionalidades específicas para la gestión de usuarios y administración del sistema.

1.3. Priorización de necesidades

- Definición y tipificación de los documentos.
- Definición de la estructura de repositorios.
- Instalación y configuración de servidor virtual para alojar la herramienta y los repositorios.

2. FASE TÉCNICO-PROFESIONAL

2.1. Descripción del proyecto

Se desarrolla e implementa una solución software vista como una plataforma web para la indexación y búsqueda de información, en el Instituto de Problemas Nacionales de la Universidad de San Carlos de Guatemala (IPNUSAC).

Se implementa, además, una estructura de indexación como solución para la recuperación de la información, con el objetivo de encontrar de manera sencilla, fácil y automática la información, utilizando consultas en forma de lenguaje natural.

La indexación ayudará a representar de una forma apropiada los documentos y referencias web, seleccionando los términos que mejor representen a los recursos digitales y así facilitar la recuperación de la información. Para la realización e implementación se utiliza un motor para la recuperación basados en texto.

Se creará un repositorio de documentos, con el objetivo de integrar y centralizar en un único sitio los documentos de referencia, para ser utilizado por los profesionales especializados de la institución en las investigaciones y estudios.

Se crean funcionalidades necesarias para la gestión de la seguridad, administración y configuración del sistema.

2.2. Investigación preliminar para la solución del proyecto

El diseño de un sistema informático requiere de la definición de una serie de pasos para lograr un objetivo final, es decir, crear un sistema para satisfacer una necesidad.

El diseño comprende la definición de la arquitectura de hardware y software, componentes, módulos y datos del sistema, con el objetivo de satisfacer ciertos requerimientos. Al definir plataformas se establecen los tipos de arquitectura, sistema operativo, lenguaje de programación o interfaz de usuario compatibles.

Una plataforma web de indexación y búsqueda de información, es una aplicación diseñada para que los usuarios puedan ingresar y obtener información de forma ordenada y eficiente desde una fuente de datos centralizada.

Para enmarcar todo este proceso para el desarrollo de la plataforma, se realiza la investigación y estudio de metodologías, herramientas y tecnología para el desarrollo del software. En cuanto a la tecnología por utilizar se analiza con base en las restricciones o delimitaciones definidas por el Departamento Técnico de la institución.

2.2.1. Metodologías de desarrollo de software

Una metodología de desarrollo de software, consiste, principalmente, en hacer uso de diversas herramientas, técnicas, métodos y modelos para el desarrollo de un sistema.

Existe una gran cantidad de metodologías que, con el paso del tiempo, han ido evolucionando debido a que no todos los sistemas de la información, son compatibles con ellas, pues el ciclo de vida del software puede ser variable. Por esta razón, se analizarán algunas metodologías y se identificará la idónea para el diseño de la plataforma.

2.2.1.1. Metodología en cascada

El estilo en modelo cascada, es que no se puede avanzar a la siguiente fase, si la anterior no se encuentra totalmente finalizada, pues no se permite regresar a la etapa anterior. Define que cada una de las fases elaboradas se encuentra documentada perfectamente.

Las fases de la metodología son, análisis de requerimientos, diseño, codificación, pruebas, verificación y mantenimiento.

2.2.1.2. Metodología de prototipos

Esta metodología consiste en que, con base a los requerimientos y necesidades del cliente, se realiza de forma rápida un prototipo, que no estará completo ni terminado. Es el desarrollo de versiones no terminadas del producto que se le entregan al usuario final.

Las fases de la metodología son, planeación, modelado, elaboración del prototipo, desarrollo, entrega y retroalimentación, comunicación con el cliente y entrega del producto final.

2.2.1.3. Metodología incremental o iterativo

Esta metodología consiste, en la combinación del modelo lineal e iterativo. Consiste en completar varias iteraciones del modelo de cascada, pero sin completar ninguna. Haciendo iteraciones se logra crear una evolución en el producto, permitiendo que se agreguen nuevas especificaciones, funcionalidades, opciones, funciones y lo que el usuario requiera después de cada iteración.

Las fases de la metodología son, inicialización, periodos de iteración y lista de control.

2.2.1.4. Metodología SCRUM

Esta metodología ágil se destaca porque es bastante amigable y fomenta el trabajo en equipo en todo momento, con la finalidad de conseguir los objetivos de una forma rápida.

Se realizan entregas parciales y regulares del producto final, priorizadas por el beneficio que se aporta al receptor del proyecto. Está especialmente indicada para proyecto en entornos complejos, donde se necesita obtener resultados pronto, con requisitos cambiantes o poco definidos, en donde la innovación, la competitividad, la flexibilidad y la productividad son fundamentales.

Las fases de la metodología son, planificación de la iteración, ejecución de la iteración, inspección y adaptación.

2.2.1.5. Metodología Kanban

Metodología ágil japonesa, consiste en ir etiquetando con tarjetas cada uno de los procesos que se deben llevar a cabo. Se le visualiza como un sistema de producción de alta efectividad y productividad. La palabra Kanban, en japonés significa “tarjetas visuales”. Los principios básicos de la metodología son, garantía de la calidad, desperdicios, mejora continua, flexibilidad.

2.2.1.6. Metodología XP

Metodología ágil que define la idea de que no se trate de mantener ciertos requisitos desde que se está elaborando el proyecto, sino que durante el proceso, estos vayan cambiando o evolucionando gradualmente sin complicaciones. Es adaptarse en el proceso a los requisitos que vayan apareciendo, y no iniciar con requisitos y desarrollar en función de ellos.

Sus características son desarrollo iterativo e incremental, pruebas unitarias, trabajo en equipo, alguien del equipo trabaja con el cliente, corrección de errores, reestructuración del código, el código es de todos y el código simple es la clave.

2.2.1.7. Metodología de proceso unificado ágil

Es una versión simplificada del proceso unificado de racional (RUP). Describe de una manera simple y fácil de entender la forma de desarrollar aplicaciones de software de negocio usando técnicas ágiles y conceptos que aun se mantienen válidos en RUP. El AUP aplica técnicas ágiles incluyendo desarrollo dirigido por pruebas, modelado ágil, gestión de cambios ágil, y refactorización de base de datos para mejorar la productividad.

En general, el proceso unificado ágil supone un enfoque intermedio entre XP y el proceso unificado racional, y tiene la ventaja de ser un proceso ágil que incluye explícitamente actividades y artefactos a los que la mayoría de desarrolladores ya están, de alguna manera, acostumbrados.

El proceso unificado ágil consta de cuatro fases que el proyecto atraviesa de forma secuencial: iniciación, elaboración, construcción y transición. A lo largo de estas cuatro fases, se desarrollan actividades relativas a siete disciplinas de manera iterativa: modelado, implementación, pruebas, despliegue, gestión de la configuración, gestión del proyecto y entorno.

2.2.2. Patrón de diseño MVC

Una arquitectura orientada a objetos bien estructurada está llena de patrones. La calidad de un sistema orientado a objetos se mide por la atención que los diseñadores han prestado a las colaboraciones entre sus objetos. Los patrones conducen a arquitecturas más pequeñas, más simples y más comprensibles.

Define un diseño que desacople la vista del modelo, con la finalidad de mejorar la reusabilidad, mantenimiento y escalabilidad del código. De esta forma las modificaciones en las vistas impactan en menor medida en la lógica de negocio. Los elementos del patrón MVC son: modelo, vista y controlador.

2.2.3. Frameworks para desarrollo web

Para el diseño y desarrollo de un sistema de software se hará uso de un framework, que se refiere a una estructura de software integrada de componentes personalizables e intercambiables para desarrollo. Su objetivo

principal es acelerar el proceso de desarrollo, reutilizar código ya existente y promover buenas prácticas de desarrollo como el uso de patrones.

Un *framework* web, por tanto, se puede definir como un conjunto de componentes (por ejemplo clases en java y descriptores y archivos de configuración en XML) que componen un diseño reutilizable que facilita y agiliza el desarrollo de sistemas web.

2.2.3.1. Laravel

Es uno de los mejores *frameworks* en PHP, gracias a su diseño extremadamente fácil de aprender y de utilizar. Además de ser súper rápido para el desarrollo de aplicaciones web, cuenta con potentes herramientas necesarias para construir grandes aplicaciones, un contenedor de controles de inversión, sistema de migración expresiva y el apoyo de las pruebas unitarias estrechamente integrada.

2.2.3.2. CodeIgniter

CodeIgniter es el segundo *framework* más popular entre los desarrolladores de PHP. Es muy potente y destaca por su velocidad, además de su sencillez y facilidad para ser aprendido, por lo que resulta ideal para iniciar cualquier proyecto.

2.2.3.3. Yii

Yii es un *framework* de desarrollo de aplicaciones libres para la web, de código abierto, que promueve el diseño limpio y motiva el desarrollo rápido. Ayuda a garantizar un producto final eficiente, extensible y mantenible.

Cuenta con características como el desarrollo rápido, el almacenamiento en memoria temporal, autenticación y control de acceso basado en roles, andamios y pruebas.

2.2.3.4. Struts 2

Framework de Java utilizados por gran parte de los ingenieros de software más modernos es el sucesor de Apache Struts 1, su versión número dos. Este *framework* es todo un hallazgo para los desarrolladores que construyen nuevas aplicaciones web con Java EE.

Apache ha entregado a estos un conjunto enorme de herramientas para crear aplicaciones web orientadas a empresas, optimizando el proceso de desarrollo desde el inicio hasta el final y el mantenimiento posterior a la entrada en funcionamiento. Struts 2 es la elección correcta si la web va a tener una carga alta de datos, como puede ser una web de broadcast.

2.2.3.5. Spring MVC

A pesar de ser un *framework* veterano (hace ya más de diez años de su primera versión), sigue siendo utilizado. Después de su expansión al MVC (modelo vista controlador), Spring siguió evolucionando convirtiéndose en un *framework* de gran escala de Java para las aplicaciones de Internet, ofreciendo a los ingenieros una poderosa herramienta para el desarrollo de aplicaciones web, así como para proyectos de seguridad.

El *framework* tiene un conjunto de interfaces que después se implementan para proporcionar la funcionalidad correspondiente. Estas abstracciones ofrecidas por las interfaces son potentes, pues permiten configurar un conjunto

de interfaces juntas y ofrecen un conjunto en el top del Api de Servlet y una división limpia entre controladores, modelos y vistas.

2.2.4. Frameworks de diseño web

Conjunto de herramientas para el diseño de sitios y aplicaciones web, basados en html, css y extensiones de javascript. Integra componentes que facilitan el desarrollo de aplicaciones como el soporte de programa, bibliotecas y plantillas.

2.2.4.1. Bootstrap

Framework o conjunto de herramientas de código abierto para diseño de sitios y aplicaciones web. Contiene plantillas de diseño con tipografía, formularios, botones, cuadros, menús de navegación y otros elementos de diseño basado en HTML y CSS, así como, extensiones de JavaScript opcionales adicionales.

Proporciona una vía rápida para el diseño web adaptado a cualquier tipo de dispositivo y es compatible con la mayoría de los navegadores web.

2.2.4.2. jQuery UI

Es una biblioteca de componentes para el *framework* jQuery que le añaden un conjunto de *plug-ins*, *widgets* y efectos visuales para la creación de aplicaciones web.

2.2.5. Motor de indexación y búsqueda

Según la mayoría de estudios que se han estado realizando en los últimos años, la recuperación y organización de la información es uno de los aspectos que han cobrado mayor relevancia. En la actualidad estos estudios resaltan la vital importancia que ha cobrado ese campo. Esto se debe en gran medida a que los buscadores de internet están situados como el primer método utilizado para obtener cualquier tipo de información, sea para el uso que sea.

La indexación de documentos es la operación más significativa del análisis documental referido al proceso de descripción y representación del contenido de un documento, mediante un número limitado de conceptos extraídos del texto de los documentos (palabras clave) o de vocabularios controlados (descriptores, términos o encabezados).

Para la indexación de cualquier tipo de datos se generan metadatos o etiquetas, altamente estructurados y que contienen información relevante de cada documento, facilitando, por tanto, su recuperación.

La recuperación de información se puede definir como el problema de la selección de documentos en respuesta a consultas o demandas de información por parte de un usuario. Los sistemas de recuperación de información utilizan bases de datos compuestas por documentos y procesan las consultas de los usuarios permitiéndoles acceder a la información relevante en un intervalo de tiempo apropiado.

Los motores de búsqueda son programas encargados de realizar las búsquedas dentro de las bases de datos de documentos web. Actualmente, se clasifican en tres categorías principales: motores de búsqueda temática,

también conocidos como directorios o catálogos; motores de búsqueda por palabras claves y sistemas basados en el contenido.

A continuación se listan algunos sistemas de recuperación de información:

2.2.5.1. LUCENE

Apache Lucene es una API de código abierto para recuperación de información, originalmente implementada en Java por Doug Cutting. Este sistema está apoyado por el Apache Software Foundation y se distribuye bajo la Apache Software License.

En el centro de la arquitectura lógica de Lucene se encuentra el concepto de documento que contiene campos de texto. Esta flexibilidad permite a Lucene ser independiente del formato del fichero. Textos que se encuentran en PDF, páginas HTML, documentos de Microsoft Word, así como muchos otros, pueden ser indexados mientras que se pueda extraer información de ellos.

Lucene ofrece características potentes a través de una API sencilla: algoritmos de búsquedas eficientes, precisos y potentes, permite realizar búsquedas clasificadas, distintos tipos de peticiones, búsqueda por campos o fecha, clasificación con base en cualquier campo, combinación de resultados de búsquedas multiíndice, permite actualización y búsqueda simultaneas.

2.2.5.2. XAPIAN

Xapian es una biblioteca de código abierto de recuperación probabilística de información, publicada con la Licencia Pública General (GPL/GNU), es decir, es una biblioteca completa que sirve como motor de búsqueda de texto.

El diseño de Xapian está pensado para ser un conjunto de herramientas altamente adaptables que permiten agregar fácilmente indexación y búsqueda avanzada en sus propias aplicaciones.

2.2.5.3. SWISH-E

Swish-e es un pequeño programa de indización y búsqueda de documentos no estructurados, este programa está basado en el sistema de indización SWISH que fue escrito originalmente por Kevin Hughes.

Desarrollado como una solución que sirve de base para el desarrollo de programas más avanzados, necesita un nivel intermedio de conocimientos en informática y de algún lenguaje de programación.

Puede indexar texto plano, e-mail, pdf, html, xml, documentos word y casi cualquier archivo que pueda convertirse a xml o html. También se usa como suplemento de base de datos para búsqueda rápida de texto.

2.3. Presentación de la solución del proyecto

Para desarrollar la solución del proyecto se hizo un análisis de las metodologías, herramientas y tecnologías posibles por utilizar, tomando en consideración las restricciones definidas por el departamento técnico del IPNUSAC.

Se selecciona una metodología de desarrollo ágil y liviano, proceso unificado ágil (AUP). Define disciplinas que se llevan de manera sistemática a lo largo de las cuatro fases de la metodología de manera iterativa. Y su principal

prioridad es satisfacer al cliente mediante la entrega temprana y continua de software con valor.

2.3.1. Tecnologías para el desarrollo

Luego de identificar las fases del proceso de desarrollo mediante una metodología, se definen y eligen las herramientas utilizadas en el desarrollo e implementación del sistema.

- Sistema operativo Debian, un sistema operativo GNU basado en software libre.
- Lenguaje de programación java, implementando Spring MVC.
- Spring Security, utilizado para el manejo de la seguridad y autenticación.
- Apache Tomcat, es un contenedor de *servlets* que se utiliza como un servidor web con soporte de *servlets* y JSPs.
- Apache Lucene, es un API de código abierto que funciona como el motor de indexación y recuperación de información.
- Bootstrap, implementado para el diseño de las vistas de cada módulo.
- JQuery, biblioteca de JavaScript utilizado para interactuar con las vistas de Bootstrap e implementación de AJAX.
- PostgreSQL, sistema de gestión de bases de datos relacional.

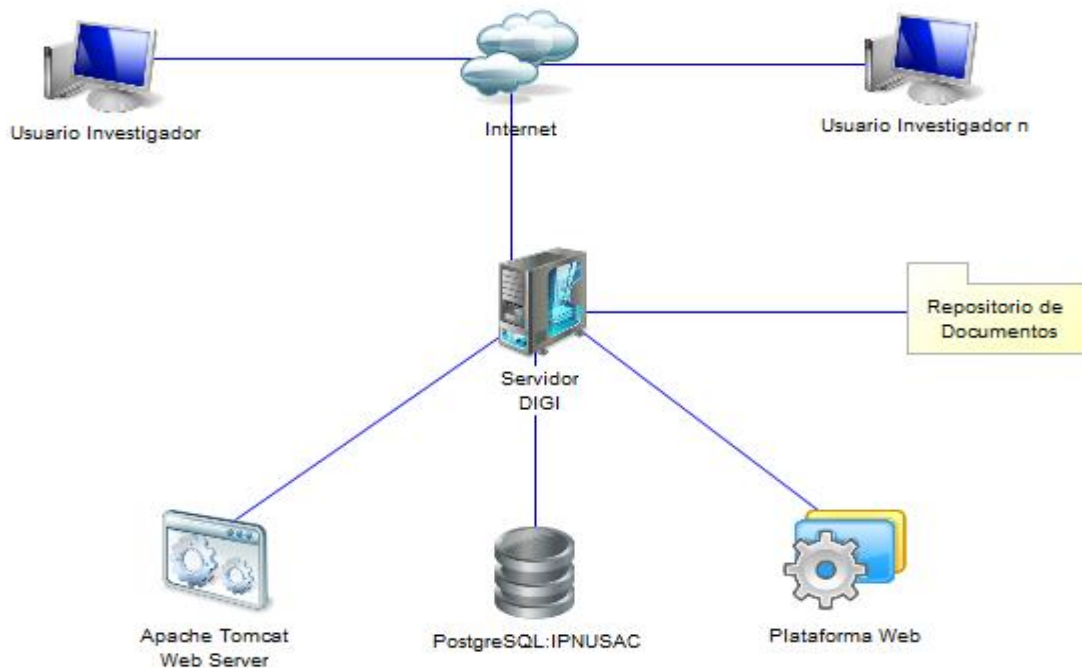
- Librerías o paquetes para conexión a base de datos y para conversión de formatos de documentos.

2.3.2. Infraestructura

La infraestructura se refiere a todos los recursos informáticos y de comunicaciones requeridos para el despliegue y utilización de la plataforma. Define una importancia estratégica ya que puede limitar o potenciar el crecimiento del sistema.

La publicación de la plataforma se realizará en los servidores brindados por la Dirección General de Investigación de la Universidad de San Carlos de Guatemala (DIGI).

Figura 1. Infraestructura



Fuente: elaboración propia.

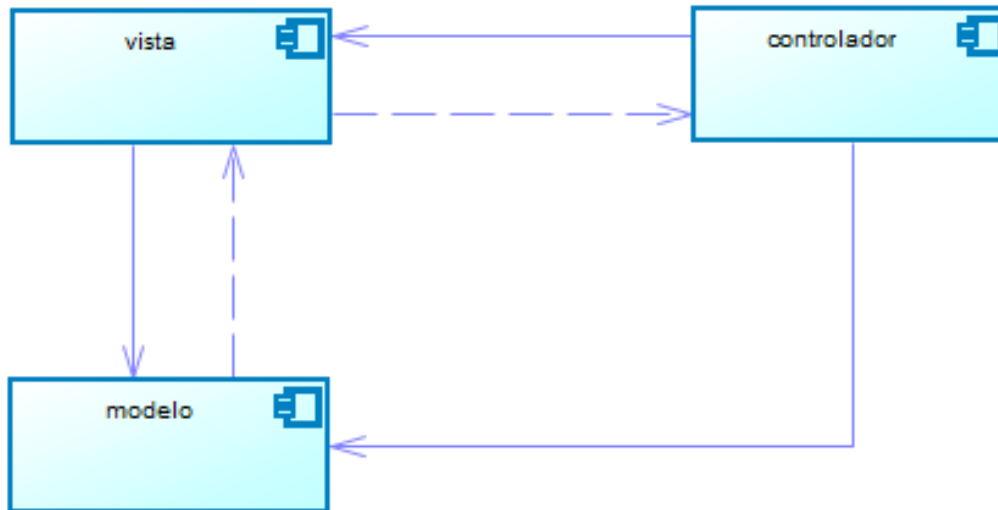
2.3.3. Arquitectura

La definición de la arquitectura toma como base el patrón MVC, este patrón ha cobrado gran popularidad en los últimos años, considerado como un estándar en aplicaciones web. Se identifican tres componentes que cumplen diferentes roles dentro de una aplicación web. Se realiza la implementación de Spring MVC.

- Modelo, es la representación de la información utilizada por el sistema para llevar a cabo sus operaciones.
- Vista, es la presentación del modelo en un formato que permita la interacción con los usuarios del sistema, generalmente la interfaz de usuario.
- Controlador, es el encargado de responder a los eventos de la vista producidos por los usuarios del sistema para invocar cambios en el modelo y consecuentemente en la vista.

En el diagrama siguiente se muestra la relación entre el modelo, la vista y el controlador, las líneas sólidas indican asociación directa: invocación a método; las líneas punteadas, asociación indirecta: evento.

Figura 2. **Arquitectura MVC**



Fuente: elaboración propia.

2.3.4. Ciclo de vida proceso unificado ágil

La metodología de desarrollo utilizada, proceso unificado ágil consta de cuatros fases que el proyecto atraviesa de forma secuencial. Para cada una se desarrollan actividades iterativas, modelado, implementación, pruebas, despliegue, gestión de la configuración, gestión del proyecto y entorno.

2.3.4.1. Iniciación

En esta fase se realizan reuniones y entrevistas con los actores del sistema, investigadores y personal técnico. Se definió el alcance inicial del proyecto: definir y entender la lógica del negocio e identificar una solución viable para los requerimientos del sistema.

Se delimitan y priorizan los requerimientos con base a los tiempos definidos para la realización del proyecto. También se revisan los diseños de

modelos que representan los procesos y los cuales son transformados en funcionalidades del sistema mediante código ejecutable.

2.3.4.2. Elaboración

Con base a los requerimientos definidos y entendimiento de la lógica del negocio, en esta fase se realiza el análisis para identificar y validar la arquitectura del sistema y tecnologías por utilizar. Se analizan los recursos disponibles, correspondientes a limitaciones y restricciones de hardware y software, que permitan implementar de una manera eficiente el proyecto.

2.3.4.3. Construcción

En esta fase se realizan la construcción o desarrollo del sistema de software, desde un punto de vista incremental. Se empiezan a desarrollar los módulos o funcionalidades de acuerdo con las prioridades definidas y acordadas.

Se transforman los modelos en código funcional, se realizan pruebas básicas y unitarias. Se evalúa el cumplimiento de los requisitos para asegurar la calidad y que el sistema funcione tal y como fue diseñado.

2.3.4.4. Transición

En esta fase se valida y despliega el sistema en entorno de producción. Se planean las entregas del sistema y se logra que el sistema quede disponible para los usuarios finales.

Asimismo, se procede a gestionar la configuración, proveer el acceso a los artefactos del proyecto, a las versiones de código fuente, documentación técnica y funcional.

2.3.5. Solución presentada

La solución implementada por el proyecto está dividida en módulos funcionales y de administración del sistema. Se denomina al sistema como IndBusInf, simplificado de la funcionalidad del sistema, indexación y búsqueda de información.

2.3.5.1. Inicio

La página de inicio diseñada bajo un patrón de usabilidad simple, muestra el nombre del sistema y de la institución. Además, muestra un formulario para el ingreso de nombre y contraseña del usuario.

Figura 3. **Ingreso de usuario**



The image shows a user login interface. At the top, there is a header with a folder icon, the text 'IndBusInf', and a horizontal line. Below the header, the text 'IPNUSAC' is displayed. Underneath, there are two yellow input fields. The first field contains the text 'Rural'. The second field contains four dots, indicating a password field. Below the password field, there is a button with the text 'Entrar'.

Fuente: elaboración propia, empleando Bootstrap.

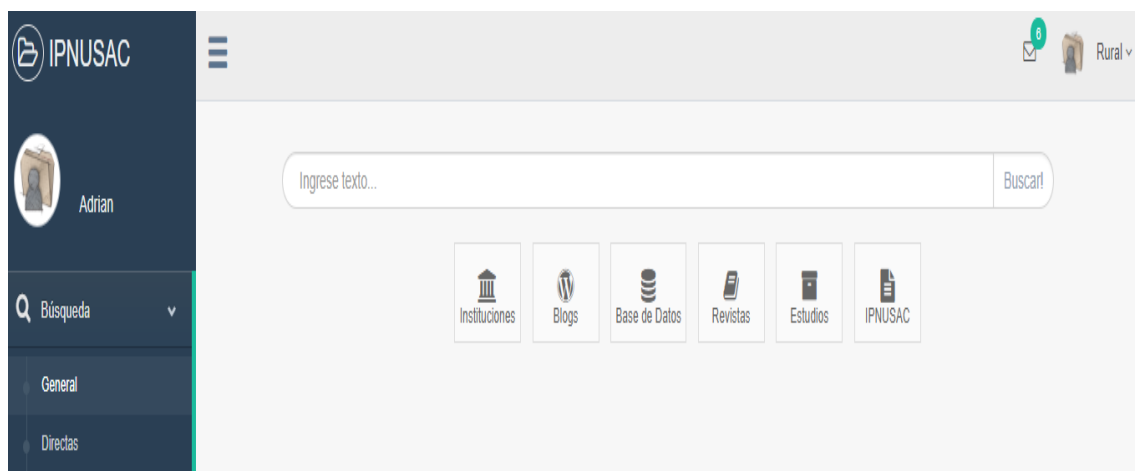
2.3.5.2. Búsqueda

Este módulo es el principal del sistema, es el encargado de realizar la recuperación de la información mediante consultas directas o específicas. Se divide en dos módulos: general y directa.

2.3.5.2.1. General

Este tipo de búsqueda realiza la consulta sobre el contenido de los documentos. Se puede realizar una búsqueda sobre todos los documentos o sobre un tipo específico.

Figura 4. Búsqueda general

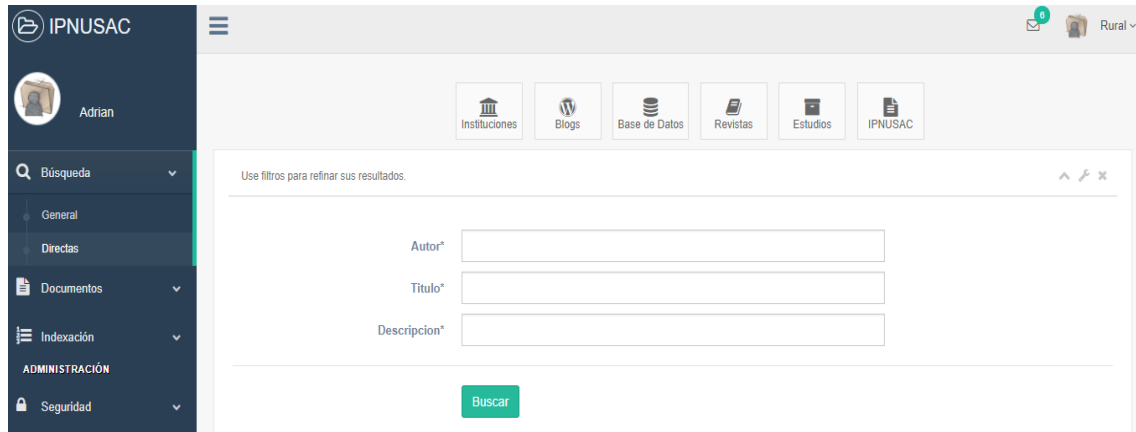


Fuente: elaboración propia, empleando Bootstrap.

2.3.5.2.2. Directa

Este tipo de búsqueda realiza una consulta específica con base en parámetros o filtros definidos para un tipo de documento específico.

Figura 5. **Búsqueda directa**

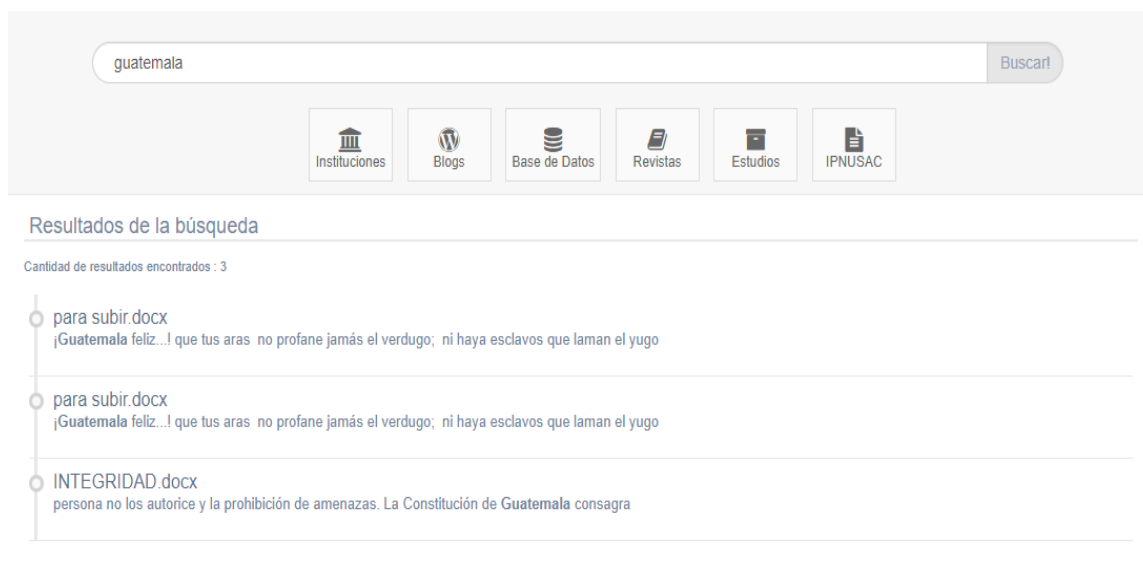


Fuente: elaboración propia, empleando Bootstrap.

2.3.5.2.3. Resultados

Los resultados de una búsqueda para cualquiera de los dos tipos de búsqueda, se muestra como una lista de los documentos que cumplen con las condiciones de consulta ingresados, permitiendo la descarga del documento.

Figura 6. **Resultado de la búsqueda**



Fuente: elaboración propia, empleando Bootstrap.

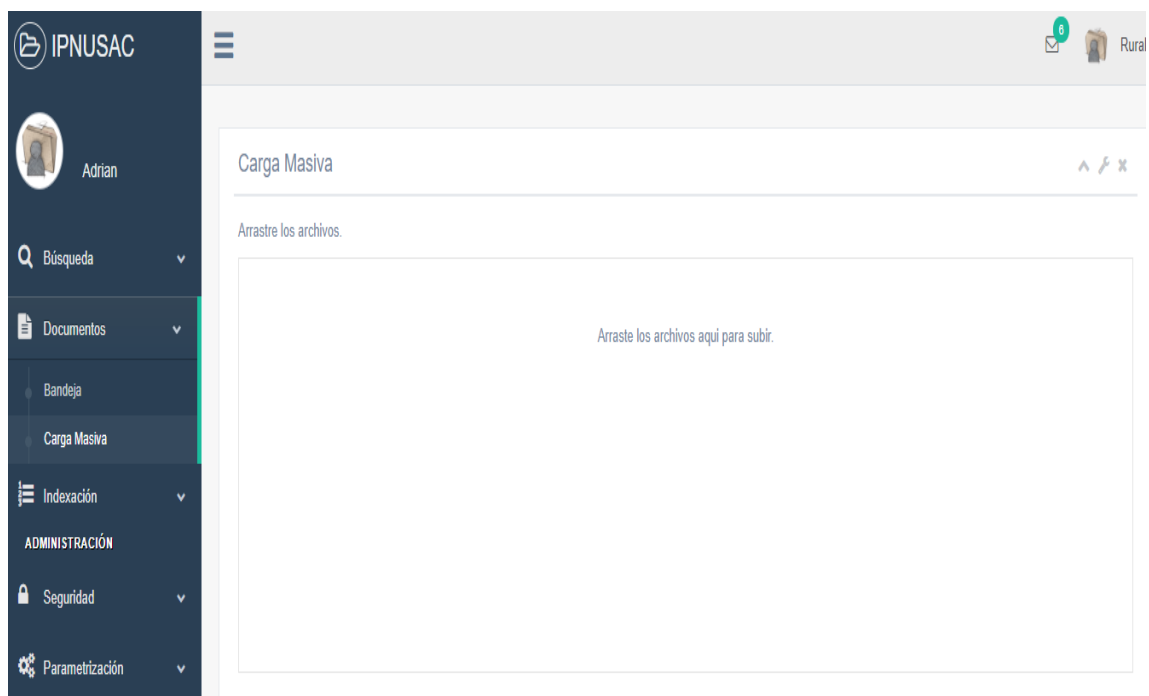
2.3.5.3. Documentos

Módulo encargado para la gestión de los documentos, ingreso de los documentos al repositorio del sistema y una bandeja de documentos para realizar la tipificación e indexación de los documentos. Se divide en bandeja y carga masiva de documentos.

2.3.5.3.1. Carga masiva

Módulo donde se realiza el ingreso de los documentos al repositorio. Permite la selección o arrastrar los documentos por subir.

Figura 7. Carga masiva

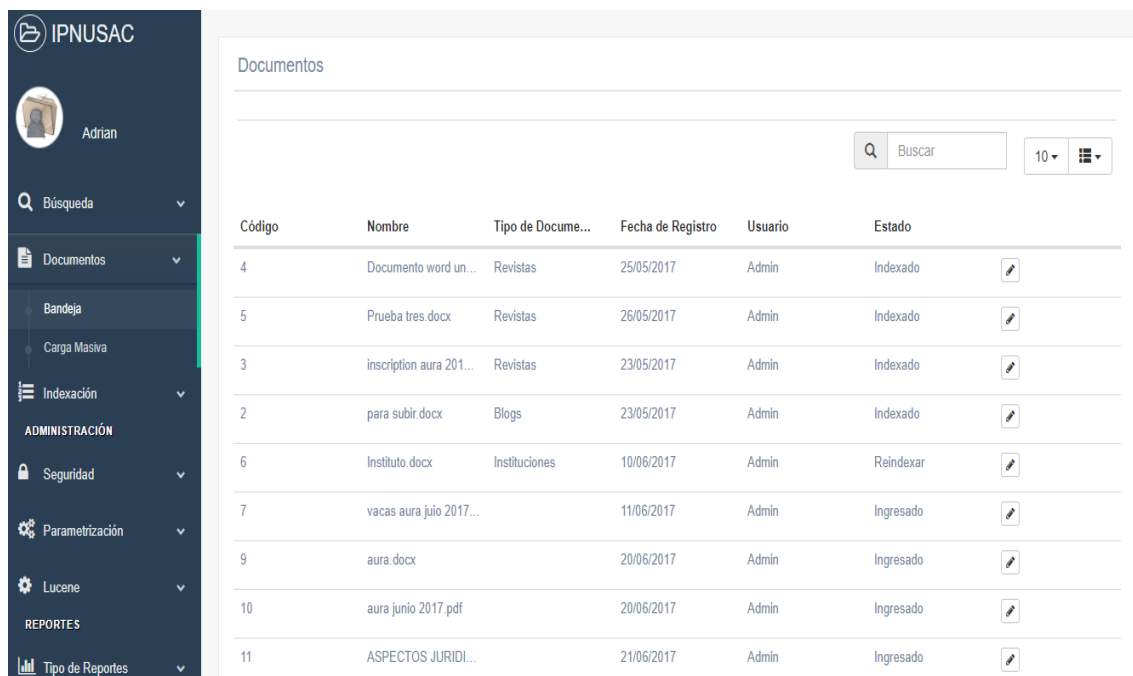


Fuente: elaboración propia, empleando Bootstrap.

2.3.5.3.2. Bandeja

Módulo que se encarga de indexar los documentos, presenta una lista de los documentos subidos al sistema mostrando su estado. Estado ingresado, indica que el documento se subió al sistema pero no está indexado; estado indexado, indica que el documento está indexado y disponible para las consultas; estado reindexar, indica que se actualizó el documento y es necesario actualizar sus índices.

Figura 8. Bandeja de documentos



Código	Nombre	Tipo de Docume...	Fecha de Registro	Usuario	Estado	
4	Documento word un...	Revistas	25/05/2017	Admin	Indexado	
5	Prueba tres.docx	Revistas	26/05/2017	Admin	Indexado	
3	inscription aura 201...	Revistas	23/05/2017	Admin	Indexado	
2	para subir.docx	Blogs	23/05/2017	Admin	Indexado	
6	Instituto.docx	Instituciones	10/06/2017	Admin	Reindexar	
7	vacas aura julio 2017 ...		11/06/2017	Admin	Ingresado	
9	aura.docx		20/06/2017	Admin	Ingresado	
10	aura junio 2017.pdf		20/06/2017	Admin	Ingresado	
11	ASPECTOS JURIDI...		21/06/2017	Admin	Ingresado	

Fuente: elaboración propia, empleando Bootstrap.

El proceso de indexación consiste, en seleccionar un documento de la bandeja, en el icono en forma de lápiz el cual mostrará un formulario para seleccionar el tipo de documento e ingresar la información para la indexación.

Figura 9. **Formulario de indexación**

Indexacion

Nombre *

Tipo *

Autor*

Titulo*

Descripcion*

Fuente: elaboración propia, empleando Bootstrap.

2.3.5.4. **Indexación**

El sistema soporta varios tipos de documentos, cada uno tiene información asociada que lo identifica. Los datos que componen esta información pueden ser desde el nombre, descripción, autor o título del documento.

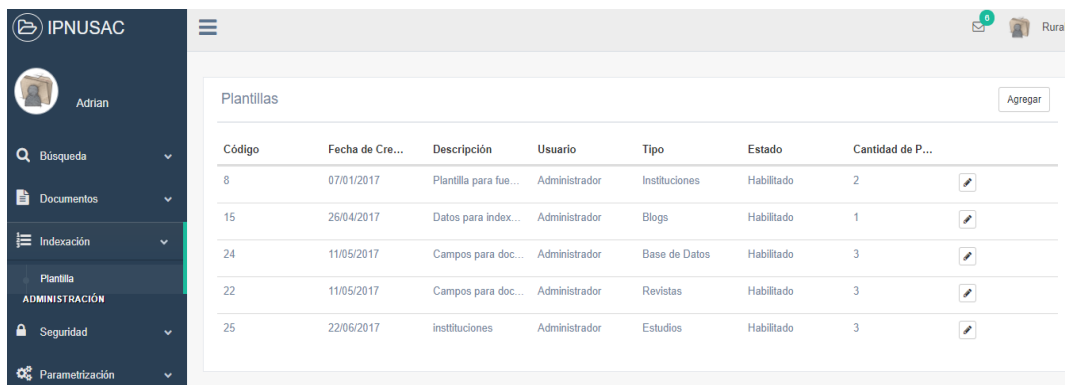
Asocia datos que identifican a un documento en una plantilla de índices con base a un tipo de documento. El sistema tiene definidos tipos de documentos y estos, a su vez, le definen qué tipos de datos componen o distinguen a cada tipo.

Se crean las plantillas de índices con los datos definidos para cada tipo de documento; estas plantillas serán utilizadas para indicarle al motor de indexación los valores por almacenar e identificar al documento indexado.

2.3.5.4.1. Plantilla

Se muestra una tabla donde se listan las plantillas creadas. El módulo permite agregar y modificar.

Figura 10. Lista de plantillas

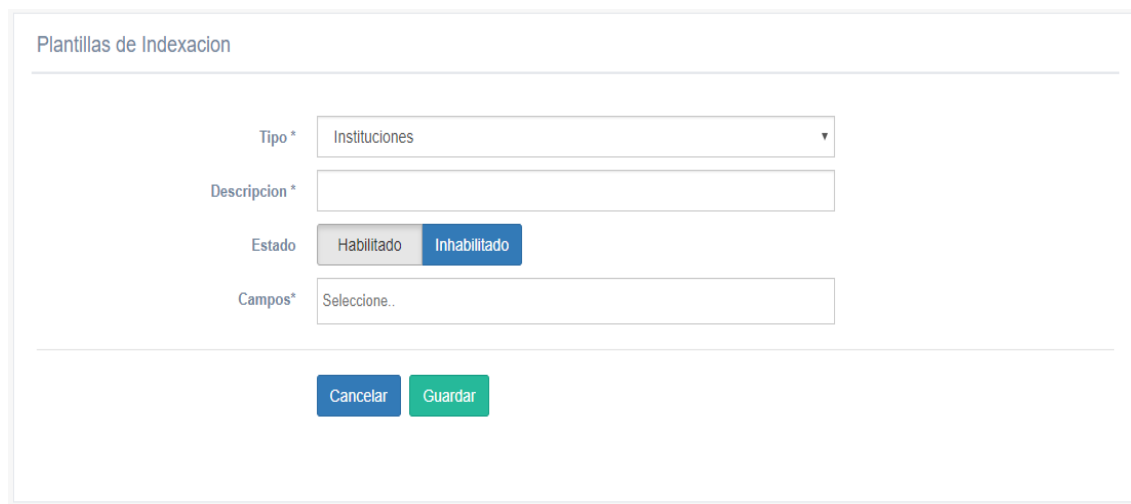


Código	Fecha de Cre...	Descripción	Usuario	Tipo	Estado	Cantidad de P...
8	07/01/2017	Plantilla para fue...	Administrador	Instituciones	Habilitado	2
15	26/04/2017	Datos para index...	Administrador	Blogs	Habilitado	1
24	11/05/2017	Campos para doc...	Administrador	Base de Datos	Habilitado	3
22	11/05/2017	Campos para doc...	Administrador	Revistas	Habilitado	3
25	22/06/2017	instituciones	Administrador	Estudios	Habilitado	3

Fuente: elaboración propia, empleando Bootstrap.

En el proceso de agregar o modificar una plantilla se muestra un formulario para configurar los datos que componen a la plantilla de indexación.

Figura 11. Plantilla de indexación



Plantillas de Indexación

Tipo *

Descripcion *

Estado Habilitado Inhabilitado

Campos*

Fuente: elaboración propia, empleando Bootstrap.

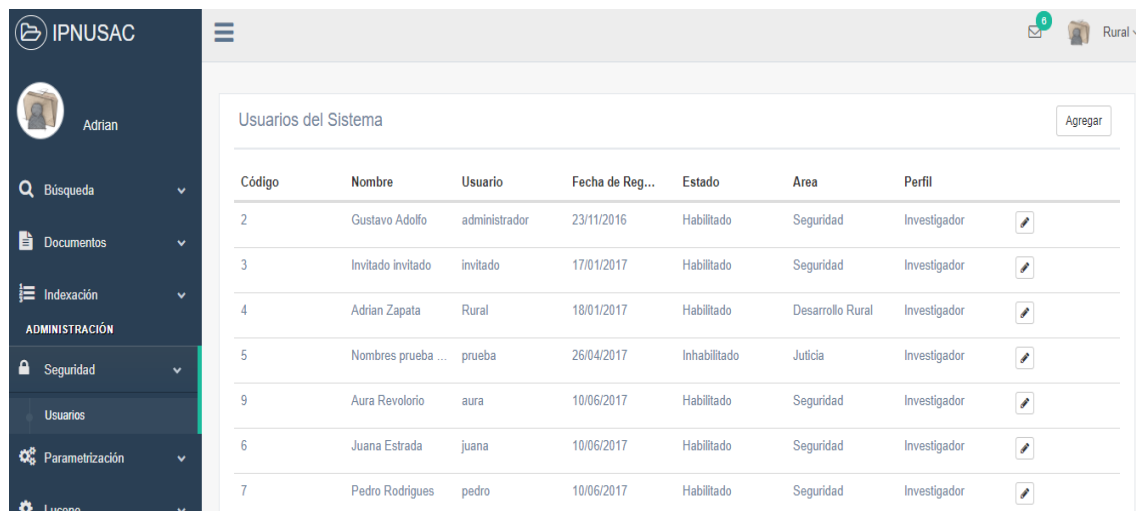
2.3.5.5. Seguridad

Para la autenticación de los usuarios del sistema es necesaria la creación, configuración y modificación del módulo de usuarios, el cual permite la creación de un usuario dentro del sistema. Este módulo asigna un rol, un nombre de usuario y contraseña para que se pueda hacer uso del sistema de una forma segura.

2.3.5.5.1. Usuarios

Módulo para la gestión de usuarios. Se muestra una lista de los usuarios del sistema.

Figura 12. Lista de usuarios



Código	Nombre	Usuario	Fecha de Reg...	Estado	Area	Perfil	
2	Gustavo Adolfo	administrador	23/11/2016	Habilitado	Seguridad	Investigador	
3	Invitado invitado	invitado	17/01/2017	Habilitado	Seguridad	Investigador	
4	Adrian Zapata	Rural	18/01/2017	Habilitado	Desarrollo Rural	Investigador	
5	Nombres prueba ...	prueba	26/04/2017	Inhabilitado	Juticia	Investigador	
9	Aura Revolorio	aura	10/06/2017	Habilitado	Seguridad	Investigador	
6	Juana Estrada	juana	10/06/2017	Habilitado	Seguridad	Investigador	
7	Pedro Rodrigues	pedro	10/06/2017	Habilitado	Seguridad	Investigador	

Fuente: elaboración propia, empleando Bootstrap.

El módulo permite la creación y modificación de los usuarios, muestra un formulario para el ingreso de los datos necesarios y rol del usuario dentro del sistema.

Figura 13. **Formulario datos de usuario**

Datos de Usuario

Nombres Apellidos

Email Telefono

Perfil * Investigador

Area * Seguridad

Estado Habilitado Inhabilitado

Usuario *

Contraseña *

Repetir contraseña *

Cancelar Guardar

Fuente: elaboración propia, empleando Bootstrap.

2.3.5.6. Parametrización

El sistema está compuesto por catálogos en algunos módulos; se utiliza para seleccionar un tipo de dato o valor. Debido a que estos valores pueden variar en nombre, o descripción, es necesario agregar una nueva opción o tipo. La gestión y configuración se realiza en este módulo.

2.3.5.6.1. Catálogos

Módulo para la configuración de catálogos. Muestra una lista de los catálogos del sistema.

Figura 14. Lista de catálogos

Código	Nombre Catálogo	Descripción
6	prueba_uno	catalogo de prueba
8	cat_prueba	tabla con estrucutra
9	cat_area	Áreas en que se divide la institucion
10	cat_estado_documento	Determina el estado de un documento d...
11	cat_tipo_dato	Representa los tipos de datos

Fuente: elaboración propia, empleando Bootstrap.

Para la creación o modificación de un catálogo se muestra un formulario para el ingreso del catálogo.

Figura 15. Formulario de catálogo

Datos de Catalogo

Nombre*

Descripcion*

Fuente: elaboración propia, empleando Bootstrap.

Para la gestión de los datos de un catálogo, se selecciona el catálogo y muestra la lista del detalle.

Figura 16. Lista detalle de catálogo

Código	Nombre	Descripción	Estado
1	Ingresado	El documento esta ingresado d...	true
3	Reindexar	Volver a indexar un documento ...	true
2	Indexado	El documento esta indexado en...	true

Fuente: elaboración propia, empleando Bootstrap.

Para la gestión del detalle, se agregan o modifican los valores en el formulario de detalle de catálogo.

Figura 17. Formulario detalle de catálogo

Datos de Detalle de Catalogo

Nombre*

Descripción*

Estado Habilitado Inhabilitado

Fuente: elaboración propia, empleando Bootstrap.

2.3.5.7. Lucene

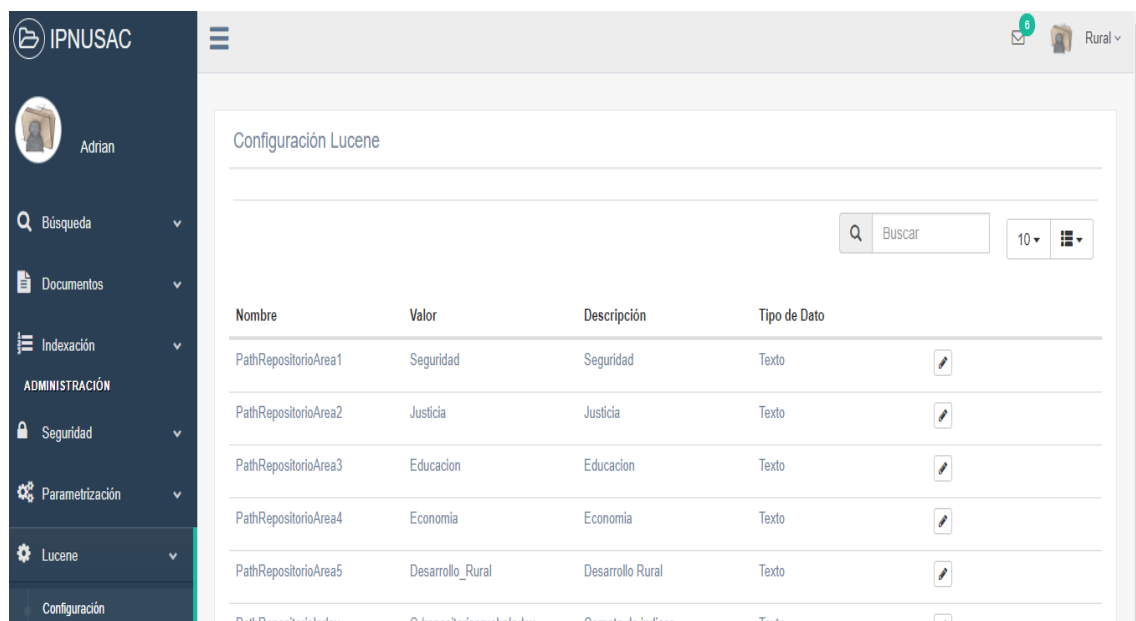
Lucene es el motor de indexación y búsqueda implementado dentro de la plataforma.

Para su funcionamiento se definieron parámetros que pueden ser configurables o modificados por el administrador del sistema, se recomienda que luego de la instalación del sistema no se cambien estos valores.

2.3.5.7.1. Configuración

Módulo que gestiona los valores de los parámetros creados por defecto en la configuración de Lucene. Se listan los parámetros creados en el sistema.

Figura 18. Lista de parámetros de Lucene

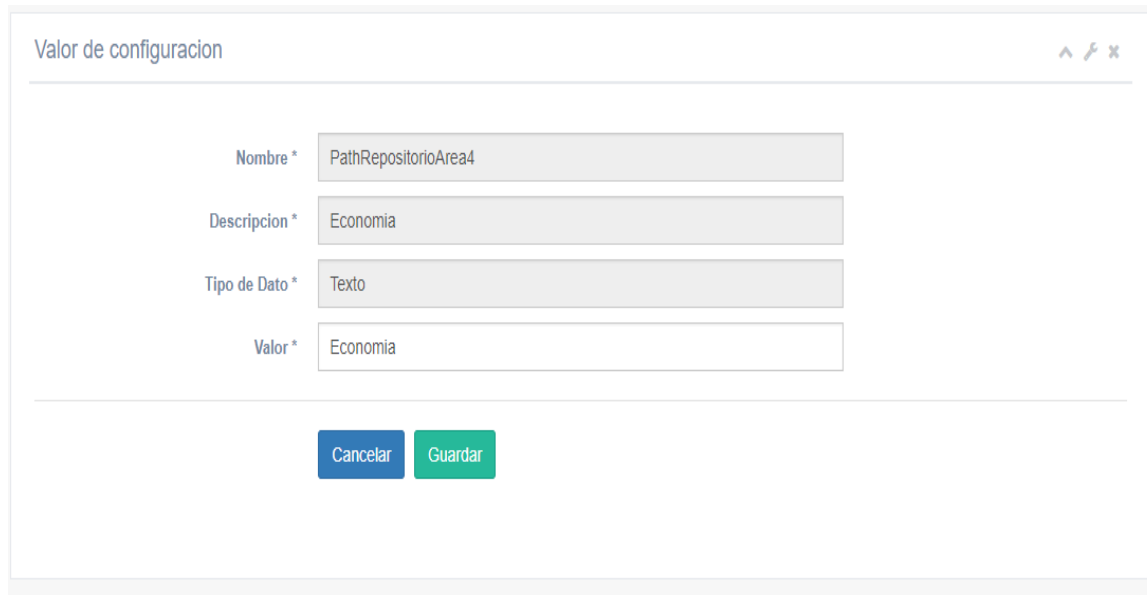


Nombre	Valor	Descripción	Tipo de Dato
PathRepositorioArea1	Seguridad	Seguridad	Texto
PathRepositorioArea2	Justicia	Justicia	Texto
PathRepositorioArea3	Educacion	Educacion	Texto
PathRepositorioArea4	Economia	Economia	Texto
PathRepositorioArea5	Desarrollo_Rural	Desarrollo Rural	Texto
PathRepositorioIndex	C:\repositorio\nuebalIndex	Carteta de índices	Texto

Fuente: elaboración propia, empleando Bootstrap.

Permite la modificación del valor de los parámetros, al seccionar “edición del parámetro” aparece un formulario para la modificación del valor.

Figura 19. **Formulario de parámetros**



The image shows a web form titled "Valor de configuracion" with a close button in the top right corner. The form contains four input fields, each with a label and an asterisk indicating it is required:

- Nombre *: PathRepositorioArea4
- Descripcion *: Economia
- Tipo de Dato *: Texto
- Valor *: Economia

At the bottom of the form, there are two buttons: "Cancelar" (blue) and "Guardar" (green).

Fuente: elaboración propia, empleando Bootstrap.

2.4. Costos del proyecto

Como en todo proyecto, es pertinente conocer los costos que requería su implementación en caso de que llegue a aprobarse.

Tabla I. Costos del proyecto

Recursos	Costo Unitario (Q)	Subtotal (Q)
Sueldos		
Desarrollador	9 000,00 / mes	54 000,00
Asesor informática	2 000,00 / mes	12 000,00
Asesor de área	1 000,00 / mes	6 000,00
Asesor técnico de área	1 000,00 / mes	6 000,00
Equipo y mobiliario		
Pc de desarrollo	5 000,00	5 000,00
Servidor	8 000,00	8 000,00
Consumibles de cómputo	800,00	800,00
Servicios		
Telefónico y energía eléctrica	500,00	500,00
Internet	600,00	600,00
Viáticos		
Gasolina	1 200,00	1 200,00
Otros gastos	500,00	500,00
Total		94 600,00

Fuente: elaboración propia.

2.5. Beneficios del proyecto

- Desarrollo e implementación de un sistema a la medida definida como una plataforma web para la indexación y búsqueda de información.

- Mayor productividad, eficacia y eficiencia en la búsqueda y recuperación de documentos y referencias.
- Acceso rápido y seguro a la información reduciendo los tiempos de búsqueda de información y de respuesta.
- Ahorro de tiempo, al tener toda la documentación en un solo repositorio y con reglas de descripción, creación, clasificación y tipificación.
- Homologación, mediante la unificación de los procedimientos para realizar todas las tareas de investigación, estableciendo un mismo modelo para toda institución.
- Mejora en los procesos de investigación y control de los documentos que se consultan y que se generan.
- Sistema para la gestión y administración de los artículos que conforman la revista digital del instituto.

3. FASE DE ENSEÑANZA-APRENDIZAJE

Capacitación de los investigadores del Instituto de Problemas Nacionales de la Universidad de San Carlos de Guatemala (IPNUSAC), para el uso adecuado de la plataforma. Capacitación a personal del área técnica sobre la arquitectura y componentes del sistema.

3.1. Capacitación

El proceso de capacitación de los usuarios se realiza mediante presentaciones presenciales y documentos de guías de usuario en formato digital, para brindar la preparación y el conocimiento necesario para trabajar con el sistema.

3.1.1. Objetivo

Dar a conocer la funcionalidad del sistema y lograr la adaptación de los usuarios en la utilización de la plataforma para realizar sus tareas de investigación y análisis, y lograr un mejor desempeño.

3.1.2. Acciones y medios

Una vez planteado el objetivo de la capacitación para los usuarios, se procede a ejecutar las acciones en donde se deben llevar a cabo una serie de tareas, como las que se mencionan a continuación.

3.1.2.1. Guías de usuario

Documentos con el objetivo de dar asistencia a los usuarios del sistema. Describen una guía sobre el proceso o flujos de los módulos, de manera escrita, acompañadas de imágenes asociadas claramente, detalladas y sencillas.

3.1.2.2. Capacitación presencial

Reunión presencial con los usuarios abordando los siguientes temas:

- Presentación del proyecto, exponiendo justificación, características, resultados, beneficios y costos.
- Capacitación del sistema mostrando la funcionalidad y explicando los procesos, flujos y datos que maneja cada módulo.
- Capacitación sobre las características técnicas, arquitectura, configuración, tecnologías utilizadas y despliegue del sistema en ambiente de producción.

3.2. Material elaborado

Se elaboran modelos de documentación con el propósito de transmitir información de una manera persistente. Documentación que representa la funcionalidad de la solución.

Guías de usuario por módulo, representan el flujo y funcionalidad de cada pantalla que integra un módulo, dirigida a los usuarios del sistema.

Documentación técnica por módulo, se representa por el diseño de modelos de base de datos, diccionario de datos y diagrama de composición de estructura de los módulos.

Documento de instalación y configuración del servidor y despliegue del sistema.

CONCLUSIONES

1. La plataforma web para la indexación y búsqueda de información, implementada en el Instituto de Problemas Nacionales de la Universidad de San Carlos de Guatemala contribuye al control y administración de documentos que utilizan los investigadores.
2. Se logra desarrollar e implementar la plataforma gracias al apoyo y participación de las autoridades y personal de la institución, quienes influyeron en la eficiente y efectiva implementación.
3. Se implementa un sistema en línea para la indexación, consulta y búsqueda de información, mediante un repositorio centralizado de documentos.
4. Con la constante alimentación de documentos al sistema, se logrará optimizar los tiempos de respuesta en las actividades de estudio e investigación.

RECOMENDACIONES

1. Alimentar e indexar constantemente el repositorio de documentos para brindar más opciones de consulta.
2. Establecer un usuario con rol administrador que sea el encargado de las gestiones de configuración y administración del sistema.
3. Implementar políticas de seguridad y planes de respaldo para la base de datos y de los repositorios de documentos e índices.
4. Proporcionar capacitación constante a los usuarios nuevos.
5. Documentarse sobre el funcionamiento del sistema mediante el estudio de los documentos o guías de usuario.

BIBLIOGRAFÍA

1. *Apache Lucene*. [en línea]. <<https://lucene.apache.org/core/>> [Consulta: 1 de junio de 2017].
2. BAEZA-YATES, Ricardo. y RIBEIRO-NETO, Berthier. *Modern Information Retrieval*. [en línea]. <<http://people.ischool.berkeley.edu/~hearts/irbook/print/chap10.pdf>>. [Consulta: 1 de junio de 2017].
3. *Diseño de framework web para el desarrollo dinámico de aplicaciones*. [en línea]. <<http://revistas.utp.edu.co/index.php/revistaciencia/article/viewFile/1817/1127>>. [Consulta: 1 de junio de 2017].
4. *Modelos de recuperación de información*. [en línea]. <<http://modelosrecuperacion.tripod.com/>>. [Consulta: 1 de junio de 2017].
5. *Motores de búsqueda en internet*. [en línea]. <<http://www.unlu.edu.ar/~tyr/tyr/TYR-motor/stark-motor.pdf>>. [Consulta: 1 de junio de 2017].
6. *Proceso unificado ágil*. [en línea]. <<http://nosolopau.com/2012/06/07/mas-sobre-el-proceso-unificado-agil-fases-y-disciplinas/>>. [Consulta: 1 de junio de 2017].
7. *Spring Framework*. [en línea]. <<http://www.uv.es/grimo/teaching/SpringMVCv3PasoAPaso/>>. [Consulta: 1 de junio de 2017].

8. *Swish-e*. [en línea]. <<http://eprints.rclis.org/13252/1/Swish-e.pdf>>. [Consulta: 1 de junio de 2017].

9. *Xapian*. [en línea]. <<http://xapian.org>>. [Consulta: 1 de junio de 2017].