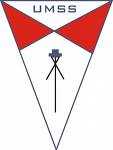
**UNIVERSIDAD MAYOR DE SAN SIMÓN**

**FACULTAD DE CIENCIAS Y TECNOLOGÍA**

**CARRERA DE LICIENCIATURA EN INGENIERÍA DE SISTEMAS**

**PERFIL DE TRABAJO DE GRADO**

**DISEÑO E IMPLEMENTACIÓN DE UN MOTOR DE BÚSQUEDA DE EDICTOS APLICANDO SPRING DATA Y ELASTICSEARCH**

**MODALIDAD: PROYECTO DE GRADO**

**ELABORADO POR: GILMER DANIEL FERNANDEZ PINTO**

**DOCENTE: ING. VILLARROEL TAPIA HENRY FRANK**

**PERIODO: II 2017**

COCHABAMBA – BOLIVIA

SEPTIEMBRE 2017

**ÍNDICE GENERAL**

[1. Introducción 3](#_Toc454867310)

[1.1. Antecedentes 3](#_Toc454867311)

[1.2. Definición del problema 3](#_Toc454867312)

[1.3. Objetivos 4](#_Toc454867313)

[1.3.1. Objetivo general 4](#_Toc454867314)

[1.3.2. Objetivos específicos 4](#_Toc454867315)

[1.4. Innovación tecnológica 4](#_Toc454867316)

[1.5. Justificación 4](#_Toc454867317)

[1.6. Metodología de desarrollo 5](#_Toc454867318)

[1.7. Alcance 5](#_Toc454867319)

1. Introducción

La integración de las tecnologías de información y las comunicaciones (TICs), ha contribuido a un mayor acceso a la información.

Hoy en día el acceso a la información se realiza por medio de la web. La naturaleza de la web es distribuida en relación a la información, la cantidad de la información va creciendo de manera progresiva día tras día sobre todo en la gestión de los documentos más específicamente en la publicación de edictos digitalizados. Esta situación ha generado la necesidad de crear un motor de búsqueda enfocado a la búsqueda de edictos digitalizados.

Los edictos son documentos que se publican en el periódico de manera cotidiana, ellos mismos representan una fuente de información importante que es manejada de una manera muy ineficiente dado que contienen datos importantes que no pueden ser encontrados directamente por el interesado dado su cantidad de contexto y el número de periódicos que hay en el país.

Dada esta problemática, este trabajo propone el desarrollo de un motor de búsqueda basada en bases de datos no relacionales aplicando Elasticsearch para la búsqueda de los edictos digitalizados proveyendo la información de manera más rápida y eficiente.

* 1. Antecedentes

La gestión de los documentos es un área muy importante dada la extensa cantidad de información que se maneja a través de ellos, información se va acumulando documentos desde que el hombre comenzó a escribirlos y a la necesidad de almacenarlos en bases de datos se vuelve imperiosa, esto teniendo en cuenta que los tipos y áreas de los documentos son variados una de esas áreas se encuentra en el aspecto legal más específicamente con la publicación de los edictos de prensa.

Edicto en el derecho moderno, es el mandato o decreto publicado con autoridad magistrado, juez o autoridad administrativa que dispone la observancia de ciertas reglas en algún asunto. También se determina así a los anuncios fijados en lugares públicos de las ciudades, villas o edificios gubernamentales sobre algún asunto para que sea notorio y de conocimiento general. ​En la práctica forense del derecho es un tipo de comunicación procesal. En nuestro caso nosotros trabajaremos con edictos de prensa digitalizados.

El crear un motor de búsqueda para realizar búsquedas en esa gran cantidad de información que contienen los edictos y las cuales son almacenadas en bases de datos y encontrar cualquier información de una manera más directa aprovechando la eficiencia de Elasticsearch combinado con Spring data para la implementación del motor de búsqueda.

* 1. Definición del problema

Los edictos son documentos publicados por un medio impreso un periódico de maneras distintas y con gran cantidad de información haciendo difícil encontrar la información de interés como ser los nombres, ci, direcciones otros, el interesado debe realizar una búsqueda manual y exhaustiva.

Se debe de leer una gran cantidad de información de manera exhaustiva dado la cantidad de edictos en un periódico y sin mencionar la calidad pésima del texto impreso que se tiene que se tiene que tratar de encontrar que varía de empresa a empresa además de tener que conseguir los periódicos de otros departamentos diferentes al que uno se encuentra haciendo bastante exhaustiva teniendo que perder mucho tiempo. En el peor de los casos no logrando encontrar la información y teniendo que sufrir las consecuencias del mismo.

* 1. Objetivo
     1. Objetivo general

Construir un motor de búsqueda de edictos digitalizados basado en Spring data y Elasticsearch para facilitar encontrar la información.

* + 1. Objetivos específicos
* Implementación del módulo del back end en Spring Boot con las dependencias de Spring Data y Elasticsearch.
* Implementar un algoritmo de búsqueda por texto en edictos digitalizados presentándolos de una manera más amigable.
* Implementación del módulo del cliente front end en Angular la interfaz de búsqueda por texto.
* Implementar un algoritmo de indexación de tablas relacionadas de base de datos para pasarla a colección de documentos en Elasticsearch.
* Implementar un algoritmo de búsqueda avanzada para realizar búsquedas más específicas.
* Implementar una interfaz de búsqueda intuitiva y amigable para el usuario del sistema.
* Pruebas de rendimiento al buscador por texto y al buscador avanzado.
  1. Innovación tecnológica

Se desarrolló un motor de búsqueda de edictos almacenados en una base de datos no relacional en Elasticsearch el cual estará implementado enteramente en la parte del back end en lenguaje orientado a objetos Java haciendo uso de framework Spring Boot añadiendo la dependencia Spring Data para poder manejar los Documentos persistidos en Elasticsearch y generar un algoritmo de búsqueda tanto por texto como para búsquedas avanzadas, generando un api Restful, para que sea consumido por una aplicación cliente basado en un framework JavaScript llamado Angular.

Elasticsearch es un motor de búsqueda orientado a documentos que nos permite indexar un gran volumen de datos para poder hacer consultas sobre ellos posteriormente.

Entre otras cosas nos permite hacer búsquedas de texto completo, búsquedas aproximadas, facetas y de resultado.

Las ventajas tecnológicas que nos provee Elasticsearch son:

* **Acceso en tiempo real**: Elasticsearch nos permite acceder sobre los datos que se están modificando en tiempo real.
* **Escalabilidad**: Gracias a su diseño nos permite escalar de forma horizontal e ir escalando nuestros servidores según nuestras necesidades.
* **Alta disponibilidad**: Los clúster de Elasticsearch son capaces de detectar qué nodos están fallando y reorganizarse para hacer que los datos sean siempre accesibles.
* **Multi-Tenant:** Nos permite operar sobre distintos índices al mismo tiempo y así potenciar nuestras búsquedas.
* **No utiliza esquemas:**Permite trabajar sin una estructura fija de base de datos.
* **Orientado a documentos:** Las entidades de Elasticsearch se almacenan como archivos JSON estructurados donde todos los campos son indexados y podemos incluir todos los índices en una misma consulta.
* **API:** Elasticsearch nos proporciona API Restful en JSON junto con API para diferentes lenguajes.
* **Búsquedas basadas en texto:**Elasticsearch está basado en Lucene, lo que incrementa las capacidades de búsqueda de texto, soportando geo localización, autocompletado.
* **Gestión de conflictos:** Previene la pérdida de datos al editar simultáneamente los registros.
  1. Justificación

Justificación técnica porque no hay un motor de búsqueda que este enfocados a la búsqueda de edictos, además del aprovechamiento de Elasticsearch que es una de las mejores tecnologías para realizar búsquedas.

Justificación económica porque al usuario que haga uso de la aplicación tendrá beneficios en el proceso de búsqueda de edictos de su interés minimizando el tiempo y logrando responder con los edictos que coinciden con la búsqueda.

Justificación social el proyecto puede tener una orientación de apoyo en la sociedad Boliviana, mejorando la calidad de las personas dando más tiempo a las personas que tengan la necesidad de buscar edictos en algún periódico a nivel nacional.

* 1. Metodología de desarrollo

El proyecto se desarrollará de manera incremental, en la cual cada módulo o subsistema se tomará como un incremento de la aplicación. Cada incremento seguirá este proceso, la primera es realizar la implementación de un algoritmo de indexación para poder llevar la información de la base de datos relacional a la base de datos de Elasticsearch, la segunda etapa la implementación de un algoritmo de búsqueda por texto, y después pasaría a la etapa de implementación de un algoritmo de búsqueda avanzada para realizar búsquedas más específicas, por último en la quinta etapa se realizara la aplicación cliente para realizar las búsquedas por el usuario y mostrar la información de los resultados de las búsquedas.

* + 1. Matriz de objetivos

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Objetivos | Actividades a Realizar | Resultados | Nro. Días Provistos | Fecha de Inicio | Fecha de Conclusión |
| Implementación del módulo del back end en Spring Boot con las dependencias de Spring Data y Elasticsearch. | * Creación de nuevo proyecto en Spring Boot. * Instalación de dependencias (Spring Data, Elasticsearch). * Configuración de propiedades para conexión con Elasticsearch | * Proyecto creado en Spring Boot, con dependencias instaladas y configuración para la conexión a Elasticsearch | 5 | 02/10/2017 | 06/10/2017 |
| Implementar un algoritmo de búsqueda por texto en edictos digitalizados presentándolos de una manera más amigable. | * Creación de la base de datos a través de los objetos persistentes Spring Data y JPA * Diseño de algoritmo de búsqueda. * Implementación del algoritmo en Spring Boot para búsquedas en Elasticsearch. | * Base de datos no relacional en Elasticsearch * Algoritmo de búsqueda diseñado e implementado en Spring Boot. | 20 | 09/10/2017 | 03/11/2017 |
| Implementación del módulo del cliente front end en Angular la interfaz de búsqueda por texto | * Implementación de arquitectura en Angular y instalación dependencias framework css (Bootstrap). * Implementación de interfaz de usuario para la búsqueda por texto | * Interfaz para la búsqueda por texto del lado del cliente. | 10 | 06/11/2017 | 17/11/2017 |
| Implementar un algoritmo de búsqueda avanzada para realizar búsquedas más específicas | * Diseñar el algoritmo de búsqueda con campos establecidos. * Implementación de algoritmo en Java Spring Boot tomando como criterios obligatorios los parámetros. | * Implementación de algoritmo para búsqueda avanzada con criterios | 30 | 20/11/2017 | 12/01/2018 |
| Implementar una interfaz de búsqueda avanzada para el usuario del sistema | * Implementación de interfaz de usuario y servicios en Angular para la ejecución de las búsquedas avanzadas | * Módulo de interfaz para realizar las búsquedas avanzadas par realizar búsquedas mas exactas | 20 | 15/01/2018 | 09/02/2018 |
| Pruebas de rendimiento al buscador por texto y al buscador avanzado. | * Implementación de pruebas de rendimiento para medir la mejora en las búsquedas | * Reporte de pruebas sobre el rendimiento en las búsquedas | 5 | 12/02/2018 | 16/02/2018 |

* + 1. Diagrama de Gantt

dasdasd

* 1. Alcance
* Se contempla únicamente el diseño la implementación del algoritmo de búsqueda.
* Para para la implementación se hará uso de Elasticsearch para realizar almacenar los datos.
* El sistema no contempla la gestión de los usuarios ni todo lo que respecta a los usuarios.
* El sistema no contempla la digitalización ni el registro de los edictos.
  1. Bibliografía