

Sistema de Administración de Farmacias

Prototipo 1 Lucene MySQL.

Versión 1.4

Historia de revisiones

Fecha	Versión	Descripción	Autor
27/08/2014	1.0	Redacción del documento	Mª José Rabaza
28/08/2014	1.1	Se expande objetivo y verificación	Guillermo López
28/08/2014	1.1	Revisión del SQA	Alfonso Methol
29/08/2014	1.2	Corrección del documento	Mª José Rabaza
30/08/2014	1.2	Verificación del prototipo	Juan Pablo Sierra
30/08/2014	1.3	Corrección del documento	Mª José Rabaza
31/08/2014	1.4	Verificación del nuevo prototipo	Juan Pablo Sierra

ÍNDICE

1. Información general	3
1.1 Requerimientos	3
1.2 Lucene	3
2. Objetivos del Prototipo	3
2.1 Importar un archivo CSV a MySQL	4
2.2 Indexar Base de Datos	4
2.3 Buscar y Mostrar	4
3. Resultados Obtenidos y conclusiones	5
4. Verificación.....	5
4.1 Planificación	5
4.2 Ejecución de la verificación	6
4.2.1 Preparación de la base de datos	6
4.2.2 Ejecución de las pruebas	6
4.3 Conclusión	7
4.5 Consideraciones adicionales	7
4.6 Entorno utilizado para las pruebas	7
5. Referencias.....	7
6. Descargas	8

1. Información general

1.1 Requerimientos

Se debe instalar MySQL en una versión 5.6.20 y, descargar el driver o conector a la base de datos de MySQL y la librería Lucene en su versión 4.9.0. Para este proyecto utilizaremos además Netbeans 8.0. Ver sección Descargas.

1.2 Lucene

Lucene es una librería open source (con licencia de Apache) que nos permite integrar funciones de indexación y búsquedas de información textual dentro de nuestros proyectos. Se puede usar Lucene para indexar y buscar datos almacenados en archivos: páginas web en servidores remotos, documentos almacenados en el sistema local de archivos, archivos de texto simple, documentos Microsoft Word, HTML, PDF, o cualquier otro formato del que se pueda extraer información textual.

Lucene se compone de dos procesos para realizar búsquedas efectivas:

1. Indexación: El proceso de indexación consiste en analizar y extraer de toda la información disponible, la que nos es relevante. Posteriormente, con esa información se crea el índice a partir del cual se realizarán las búsquedas. El índice es una estructura de datos que permite acceso rápido a la información (similar a lo que podría ser el índice de un libro).
2. Búsqueda: El proceso de búsqueda consiste en consultar el índice para obtener los documentos donde aparecen unas determinadas palabras o bien concuerdan con una determinada expresión de consulta.

2. Objetivos del Prototipo

El objetivo principal de este prototipo es desarrollar un bosquejo de módulo de indexación y búsqueda de productos, el cual realice consultas ágiles de forma de mejorar la experiencia del usuario.

Por parte del cliente se nos proporcionó una planilla Excel en formato .xls cargada con un stock de productos similar al que se utiliza en la empresa. La tabla que se nos brinda tiene la información de 20562 productos, la cual proporciona un volumen de datos adecuado para las pruebas a realizar sobre el módulo.

A partir de dicha fuente de datos se cargará una base de datos MySQL. Luego utilizando la librería Lucene se procederá a crear un índice de productos a partir de dicha base. Por último se realizarán distintas búsquedas sobre el índice creado mediante campos de texto a modo de prueba.

2.1 Importar un archivo CSV a MySQL

Primero tenemos que convertir el archivo de Excel de formato .xls al formato .csv. Para ello abrimos el documento y lo guardamos nuevamente pero con la extensión deseada. En caso de que haya varias opciones a elegir para el formato .csv se deberá elegir la que indica separación de los cambios por comas.

La importación del archivo CSV a una base de datos MySQL se deberá realizar por línea de comando.

Para ellos deberán seguir los siguientes pasos:

1. Conectar el sistema gestor de MySQL. Desde línea de comando ejecutaremos el comando `mysql` debiendo ingresar el host y nombre del usuario correspondientes (en este caso host: localhost y nombre de usuario: root). Cuando nos pida la contraseña se presionará Enter. Por más información ver referencia [1]
2. Crearemos una Base de Datos nueva, con el nombre Pruebas.
3. Se creará una tabla dentro de dicha Base de Datos llamada productos.
4. Usando el comando `LOAD DATA INFILE` se cargará la base de datos.

Para más información ver referencia [2].

2.2 Indexar Base de Datos

La indexación la haremos teniendo en cuenta sólo la descripción de producto (en los datos brindados por el cliente es el nombre del producto).

Se deberá realizar una consulta SQL en la Base de Datos ya creada y cargada para obtener la descripción de todas sus entradas y crear el índice de Lucene a partir de ella.

Cabe destacar dos cosas:

- Para obtener información sobre cómo conectar MySQL con Netbeans se proporcionó un tutorial realizado por los Especialistas Técnicos.
- Para conectar MySQL usando java ver referencia [4].

2.3 Buscar y Mostrar

Se deberá implementar una búsqueda en Lucene y mostrar el resultado obtenido. Para más información consultar la referencia [3].

3. Resultados Obtenidos y conclusiones

Se realizaron búsquedas por nombre de producto (en la tabla otorgada por el cliente el nombre está ubicado en el campo DESCRIPCIÓN). Se probó ingresar el nombre como una palabra completa y luego sólo la primer letra (para obtener una mayor cantidad de resultados). El objetivo de dicha búsqueda era ver el tiempo que se demoraba en devolver los resultados de la misma. En ambos casos el prototipo implementado demora **73ms**.

Concluimos que la búsqueda es rápida, por lo cual Lucene es una buena herramienta para realizarla. Sin embargo se decidió investigar y probar Solr en un próximo prototipo. Luego de realizar dicho prototipo se decidirá qué opción será utilizada para mitigar el riesgo de realizar búsquedas eficientes y ágiles para mejorar la experiencia del cliente de farmacia.

4. Verificación

El responsable de verificación deberá comprobar el correcto funcionamiento del módulo así como asegurar que los tiempos de búsqueda se adecuan a los requerimientos de performance solicitados. El cliente no nos supo dar una métrica para validar el requisito de búsquedas ágiles. Lo primordial es que la experiencia del cliente de farmacia sea la mejor, por lo cual la búsqueda debe ser rápida. Según los resultados obtenidos en este prototipo establecemos que la búsqueda deberá demorar a lo sumo **1 segundo**.

4.1 Planificación

Para la verificación de este prototipo es necesario considerar el ambiente en el que él mismo ha de ejecutarse y los resultados que se esperan de esa acción.

El ambiente que se necesita es:

- Un servidor de MySQL ejecutando en el equipo local.
- Una tabla de productos con descripciones de los mismos, que cuente con un número de registros comparable con el esperado en producción.
- Credenciales que permitan el acceso a la base de datos según lo asumido en el código del prototipo.

Por otra parte el procedimiento de la prueba es:

- Ejecutar el prototipo
- Evaluar el tiempo que toma su ejecución
- Determinar si el resultado observado es aceptable.

4.2 Ejecución de la verificación

Contando con un servidor de base de datos ejecutando, se procede a realizar las siguientes acciones.

4.2.1 Preparación de la base de datos

Se escribe en la terminal de MySQL el siguiente comando para crear la base de datos:

```
create database prueba_lucene
```

Se crea la tabla a utilizar:

```
create table productos (id integer not null primary key, DESCRIPCION varchar(300));
```

Importamos los datos desde el archivo CSV:

```
LOAD DATA INFILE 'C:\\Ruta\\al\\archivo.csv' INTO TABLE productos FIELDS TERMINATED BY ',' ENCLOSED BY '"' LINES TERMINATED BY '\n';
```

Query OK, 20561 rows affected

4.2.2 Ejecución de las pruebas

Para llevar a cabo las pruebas se optó por dos modalidades. Ejecutar el sistema desde el entorno integrado de desarrollo NetBeans y ejecutar la aplicación de forma standalone, de modo de observar posibles variaciones en el desempeño de la misma.

Desde NetBeans se selecciona la opción de ejecutar el proyecto, obteniéndose el siguiente resultado:

```
run:
420ms
BUILD SUCCESSFUL (total time: 6 seconds)
```

Similarmente, al ejecutar desde consola se obtiene lo siguiente:

```
=== java -jar LuceneDB.jar ===
357ms
```

Execution time: 6.258 s

4.3 Conclusión

Dado que el tiempo de ejecución se encuentra dentro del intervalo establecido, con una amplia holgura, tanto si el prototipo es ejecutado desde NetBeans como si se hace de forma independiente, **se aprueba** el mismo.

4.5 Consideraciones adicionales

Para realizar la evaluación se consideró un cronómetro incluido dentro del prototipo que mide el tiempo que es necesario para realizar la consulta y recibir los datos desde la base de datos. No se evalúa el tiempo de ejecución de todo el script dado que en el mismo se incluyen operaciones tales como la creación del índice, que deben llevarse a cabo una única vez, no afectando de esta manera a cada consulta.

Es necesario, de todas maneras, remarcar que si bien en los equipos en los que se ha llevado a cabo las pruebas del prototipo se han obtenido resultados con holguras, es posible que sea necesario evaluar el rendimiento en equipos aún más limitados o con una cierta carga al momento de las búsquedas.

4.6 Entorno utilizado para las pruebas

- mysql Ver 14.14 Distrib 5.6.20, for Win64 (x86_64)
- Windows 7 x64
- Java 7
- NetBeans 8.0
 - Java(TM) SE Runtime Environment 1.8.0_20-b26

5. Referencias

- [1]: <http://www.desarrolloweb.com/articulos/2408.php>
[2]: <http://elpoli.delphiaccess.com/mysql-importar-un-archivo-csv-con-load-data-infile/>
[3]: <http://www.lucenetutorial.com/lucene-in-5-minutes.html>
[4]: <http://www.chuidiang.com/java/mysql/EjemploJava.php>

6. Descargas

- Lucene 4.9.0
<http://download.nextag.com/apache/lucene/java/4.9.0/>
- MySQL Community Server 5.6.20
<http://dev.mysql.com/downloads/mysql/>
- Netbeans 8.0 (Versión de 210 MB)
<https://netbeans.org/downloads/>