

Bienvenido/a al curso **Aprende motores de búsqueda con Redisearch**. Su objetivo es describir el concepto de motor de búsqueda, particularizando en **Redisearch**. Al finalizarlo, conocerá este tipo de bases de datos, su terminología y, además, cómo trabajar con **Redisearch**.

La lección comienza introduciendo el concepto de motor de búsqueda. A continuación, se introduce el concepto de base de datos de búsqueda, sus principales características y usos. Cerramos la lección introduciendo nuestro motor en estudio y el plan del curso.

Al finalizar la lección, sabrá:

- Qué es y cuáles son las principales funciones de un motor de de búsqueda.
- Cuáles son las principales características de los motores de búsqueda.
- Qué es **Redisearch**.
- Cuál es el programa del curso.

## Introducción

---

Un **motor de búsqueda** (*search engine*) es un sistema de gestión de bases de datos desarrollado específicamente para encontrar términos o frases en grandes cantidades de datos de manera rápida y sencilla. Entre otros motores, encontramos **Elasticsearch**, **Solr**, **Sphinx** y **Redisearch**, siendo este último el objeto del curso.

Las principales funciones de un motor de búsqueda son:

- Administrar adecuadamente el almacenamiento de los datos en los dispositivos de almacenamiento.
- Administrar adecuadamente la memoria utilizada por el SGBD para que no robe toda la memoria del sistema e impida la ejecución de otros programas en la misma máquina.
- Controlar o regular el acceso a los datos según el usuario que se encuentre detrás. No todos los usuarios tiene por qué poder acceder a todos los datos.

Se pueden desarrollar como bibliotecas integrables, las cuales se pueden incorporar fácilmente a nuestras propias aplicaciones, ejecutándose dentro del proceso de la aplicación. Siendo **Lucene** un ejemplo de este tipo de motor. O bien, se pueden desarrollar como sistemas cliente/servidor, donde encontramos la mayoría de productos como **Elasticsearch**, **Solr** y **Redisearch**.

Entre las organizaciones que usan motores de búsqueda, encontramos **Amazon**, **AOL**, **Apple**, **BBC**, **Bloomberg**, **Cisco**, **Dell**, **Disney**, **eBay**, **Facebook**, **Fender**, **GitHub**, **Groupon**, **IBM**, **Instagram**, **Microsoft**, **Mozilla**, **MTV**, **NASA**, **Netflix**, **Sears**, **Telefónica**, **The Guardian**, **Ticketmaster**, **Twitter** y **Wikimedia**.

## Bases de datos de búsqueda

---

Como ya sabemos, una **base de datos** (*database*) es una colección estructurada de datos y/u objetos, organizada atendiendo a la especificación de un determinado modelo de base de datos. La idea de una base de datos de este tipo es organizar la información de tal manera que le resulte al motor fácil y muy rápido el acceso a aquellos registros que cumplen un determinado criterio de búsqueda.

Tengamos en cuenta que generalmente hay que consultar muchos registros con mucho texto. Por lo que los sistemas de gestión de bases de datos **SQL** y los nuevos **NoSQL** no suelen ser buenas opciones para resolver este tipo de situaciones. Necesitamos algo especializado y diseñado para este fin.

Los motores de búsqueda proporcionan varios conceptos. Por un lado, suelen utilizar el término **documento** (*document*), como unidad que describe una entidad de datos como, por ejemplo, un producto, una organización, una persona o una página web. Es importante no asociar el término al

concepto de archivo como un PDF o un Word. Podría ser así, pero no tiene por qué serlo siempre. Podemos tener una descripción detallada de un producto, sin necesidad de estar asociada a ningún archivo, representado mediante un documento en la base de datos.

Por otra parte, tenemos los **índices** (*indexes* o *indices*), los contenedores de la base de datos donde se almacena los documentos, de manera organizada. Cuando añadimos un documento a un índice, lo que hace el motor en primer lugar es analizarlo. El objetivo del **análisis** (*analysis*) es extraer los términos que aparecen en el documento, sus posiciones, el número de veces y otra información. Así cuando busquemos por un término o palabra, la búsqueda será muy rápida. El índice tiene registrados, en una estructura interna, los términos de todos sus documentos y, como no, en qué documentos aparecen. Cada vez que se realiza una búsqueda no se recorrerá todos los documentos, sino la estructura interna de términos mantenida por el índice. La cual es mucho más pequeña y fácil de recorrer.

Para poder realizar un buen análisis, es necesario indicar el idioma del documento. No es lo mismo analizar un documento en español, en italiano, en inglés o en chino. Cada idioma tiene sus cosas y debe ser por tanto analizado por un componente específico.

También relacionado con el análisis, tenemos el concepto de **refinamiento** (*stemming*). Esta operación consiste en reducir o extraer las raíces de las palabras que aparecen en el documento. Así mediante el refinamiento, las conjugaciones de verbos se reducirán al verbo infinitivo, los plurales a singular, etc. Esto ayudará al motor a realizar búsquedas independientemente del género, número y la conjugación de las palabras que aparecen en el documento. Por ejemplo, si el usuario realiza una búsqueda de la palabra programador, el motor podrá devolver aquellos documentos donde aparece programador, programadora, programadores y programadoras.

### Características de los motores de búsqueda

Las principales características de un motor de búsqueda son:

- Las consultas de búsqueda realizadas por los usuarios deben responderse rápidamente.
- Las consultas de búsqueda deben atender situaciones donde no aparezca la palabra exacta pero sí algún sinónimo o variación de la palabra.
- Se debe poder acceder al motor de búsqueda mediante distintos *drivers*.
- Debe permitir el escalado de bases de datos de búsqueda grandes mediante algún tipo de partición como, por ejemplo, *sharding*.
- Para asegurar la alta disponibilidad y la recuperación ante fallos de sistema, debe permitir la replicación de datos.
- Debe ser capaz de analizar documentos escritos en varios idiomas.
- Debe realizar la indexación de manera rápida, eficiente y eficaz.

### Principales usos de los motores de búsqueda

Generalmente se usa motores de búsqueda para:

- Búsqueda de términos o frases en grandes cantidades de datos como, por ejemplo, en catálogos de productos.
- Autocompleción. Es muy habitual utilizar motores de búsqueda para encontrar términos que comienzan por un determinado texto o prefijo habitualmente usados por muchas aplicaciones webs y móviles.

Por ejemplo, en la investigación de los **Papeles de Panamá** se utilizó un motor de búsqueda para realizar búsquedas en los **2.6 terabytes** de datos disponibles.

### RediSearch

**RediSearch**, el objeto del curso, es un motor de búsqueda implementado como un módulo de **Redis**, creado y desarrollado por **Redis Labs**. Su principal ventaja, con respecto a otros motores, es que no se implementa como un sistema de gestión de bases de datos, sino que se integra con **Redis**. Esto reduce considerablemente la curva de aprendizaje, pues no hay que aprender un nuevo motor de bases de

datos, con sus peculiaridades y aspectos de administración y monitorización. Sólo hay que tener claro los conceptos de los motores de búsqueda, cómo integrarlo en **Redis**, algo que es extraordinariamente fácil, y con qué comandos interactuar con el motor de búsqueda.

Recordemos que **Redis** es un motor de bases de datos clave-valor en memoria. Lo que lo hace extraordinariamente rápido. En algunos estudios de rendimiento, se observa mejoras entre el 100 y el 300% con respecto a **Elasticsearch** y **Solr**, los otros motores más utilizados hoy en día. Pero cuidado, no vale para todo tipo de consultas, pero sí para cubrir el 80% de los casos más habituales.

Antes de elegir un motor de búsqueda es importante tener claro qué consultas se realizarán y si el motor puede satisfacer nuestras necesidades, al menos las principales.

Su página oficial es [redisearch.io](http://redisearch.io).

## Principales características

Sus principales características son:

- Al integrarse en **Redis**, es un sistema de bases de datos en memoria con soporte de persistencia.
- Al integrarse en **Redis**, no hay que aprender un motor de bases de datos adicional.
- Al integrarse en **Redis**, es ligero, de pequeño tamaño y muy pero que muy rápido.
- Tiene muy buen rendimiento tanto en las lecturas como en las escrituras.
- Al integrarse en **Redis**, se puede utilizar su soporte de alta disponibilidad y escalabilidad.
- Se puede realizar consultas complejas mediante **Lua**.
- Soporta índices de búsqueda para realizar búsquedas de términos en grandes cantidades de documentos.
- Puede analizar documentos en los principales idiomas como, por ejemplo, alemán, español, francés, inglés e italiano.
- Soporta diccionarios de sugerencias con los que las aplicaciones pueden implementar las autocompleciones.

## Información del curso

Este curso está dedicado al motor de búsqueda **RedisSearch**, implementado como un módulo independiente fácilmente integrable en **Redis**.

Tiene como objetivo introducir al estudiante en el mundo de las bases de datos de búsqueda, mediante una descripción detallada de este módulo por su buen soporte del concepto de motor de búsqueda.

Al finalizarlo, sabrá:

- Qué es un motor de búsqueda.
- Por qué y para qué utilizar un motor de búsqueda.
- Cómo compilar, instalar y cargar el módulo **RedisSearch** en **Redis**.
- Qué es y para qué se usa un índice de búsqueda.
- Cómo crear y suprimir índices de búsqueda.
- Cómo añadir, actualizar y suprimir documentos de los índices de búsqueda.
- Cómo realizar consultas de búsqueda contra índices.
- Qué es y para qué se usa un diccionario de sugerencias.
- Cómo añadir, actualizar y suprimir sugerencias a un diccionario.
- Cómo consultar un diccionario de sugerencias.

## Conocimientos previos

Cómo no puede ser de otra manera, debe tener conocimiento de **Redis**.

## Plan del curso

El curso tiene una duración aproximada de **2 horas**. Y se divide en **6 lecciones**.

A continuación, se enumera las distintas lecciones y el tiempo estimado para su estudio:

Lección	Teoría	Descripción
1 <b>Introducción</b>	15min	Esta lección.
2 <b>Instalación de Redisearch</b>	5min	Cómo descargar, compilar y cargar <b>Redisearch</b> .
3 <b>Índices de búsqueda</b>	10min	Qué es un índice de búsqueda y cómo crearlos y suprimirlos.
4 <b>Documentos</b>	15min	Qué es un documento y cómo insertarlos, actualizarlos y suprimirlos.
5 <b>Consultas de búsqueda</b>	15min	Cómo realizar consultas de búsqueda que devuelven qué documentos presentan uno o más términos dados.
6 <b>Diccionarios de sugerencias</b>	10min	Qué es un diccionario de sugerencias, cómo actualizar su contenido y cómo consultarlo.

## Información de publicación

Título **Aprende motores de búsqueda con Redisearch**

Autor **Raúl G. González - raulggonzalez@nodemy.com**

Primera edición **Abril de 2017**

Versión actual **1.0.0**

Versión de **Redisearch** **0.12**

Contacto **hola@nodemy.com**