

Lab 8 – Elasticsearch on Wikipedia – Java Guide

CC5212-1 – May 7, 2025

Esta es una guía opcional para ayudarte a codificar tu solución en Java. Hay tres partes en este ejercicio. La primera parte consiste en completar `BuildWikiIndexBulk.java`, que indexa documentos en Elasticsearch. Específicamente, debes (1) indexar adicionalmente el título, el abstract (si está disponible) y los archivos modificados. La segunda parte consiste en completar `SearchWikiIndex.java`, que realiza una búsqueda de palabras clave sobre un índice. Específicamente, debes (2.1) considerar adicionalmente el campo de resumen para las búsquedas, y (2.2) imprimir los detalles de los resultados en la consola.

Paso 1: BuildWikiIndexBulk

Completa la clase con los siguientes snippets (los cuales pueden aparecer desordenados):

Put values: Añade valores para diferentes fields (title/abstract/modified) al documento:

```
json.put(FieldNames.TITLE.name(), tabs[1]);  
  
json.put(FieldNames.ABSTRACT.name(), tabs[2]);  
  
json.put(FieldNames.MODIFIED.name(), System.currentTimeMillis());
```

Verify array length: Asegúrate de que tengamos al menos tres valores. Uno o más fragmentos adicionales deben ir dentro de este bloque.

```
if(tabs.length>2){  
    ?  
}
```

Pasos 2.1 & 2.2: SearchWikiIndex

Completa la clase con los siguientes snippets (que pueden ser presentados fuera de orden). Los snippets para los pasos 2.1 y 2.2 están mezclados aquí.

Include abstract: Incluye el campo abstract al query builder. Nota que BOOSTS contiene pesos para distinguir los resultados en diferentes campos, donde con la definición actual, una coincidencia de palabras clave en el título tendrá un valor cinco veces mayor que una coincidencia de las mismas palabras clave en el abstract.

```
float absBoost = BOOSTS.get(FieldNames.ABSTRACT.name());  
multiMatchQueryBuilder.field(FieldNames.ABSTRACT.name(),absBoost);
```

Print result: Imprime algunos detalles de un resultado en la consola.

```
System.out.println((hit.getIndex()) + "\t" + url + "\t" + title + "\t" + abs + "\t" + hit.getScore());
```

Extract values: Obtén los valores de un resultado.

```
String url = (String) json.get(FieldNames.URL.name());  
String abs = (String) json.get(FieldNames.ABSTRACT.name());
```