

## Autocorrección

### Implementación Básica del Modelo Vectorial

---

#### Corrección de la práctica

Dado que los datos coinciden con los del ejercicio de la página Excel, el alumno puede autoevaluarse tan solo comparando la salida de esta práctica con la de la hoja de cálculo

#### Autocorrección

##### ¿Se ha incluido alguno de los siguientes elementos?

- **Índice en base de datos.** En el enunciado de la práctica el índice inverso se ha almacenado mediante diccionarios en memoria RAM por sencillez. Sin embargo, los índices realmente se guardan en bases de datos de diversos tipos. Para la práctica del motor de búsqueda se implementará el índice inverso con una base de datos MySQL por ejemplo.
- **Tokenizer.** En este caso el tokenizer puede ser muy básico. Sin embargo, podría mejorarse tanto para limpiar las etiquetas HTML como para separar los términos. Para la práctica del motor de búsqueda deberán incorporarse estas mejoras.
- **Normalización de Términos.** Los **términos** se podrían normalizar, por ejemplo pasándolos a minúsculas y reduciéndolos a la raíz, para que casos como *Teclado* y *teclados* se consideren el mismo término. Para la práctica del motor de búsqueda se deberán implementar este tipo de mejoras.
- **Separar el cálculo de pesos de la función de similitud.** Tanto la función de similitud del producto escalar como la del **coseno** están definidas independientemente de los pesos que se den para cada término. En vez de implementar cada modelo con una asignación de pesos concreta (e.g. ProductoEscalarTF, ProductoEscalarTFIDF, CosenoTF y CosenoTFIDF), se podría separar en una clase para la función de similitud (e.g. ProductoEscalar y Coseno) y otra clase para el cálculo de pesos (e.g. TF y TFIDF).

