BÚSQUEDA HÍBRIDA

Dos buscadores, una sola lista

Una consulta puede llevar un código exacto y una idea con otras palabras. Cada buscador cubre una parte.

> **anular inscripción MAT-17**

Por palabras	**encuentra el código MAT-17**
Por significado	**anular ≈ cancelar**
Juntos	**una lista ordenada**

EL CASO

Una consulta y tres textos

Las letras A, B y C sirven para reconocer cada texto cuando aparezca en una lista.

Consulta de la persona

> anular inscripción **MAT-17**

A **MAT-17**: antes de tramitar la baja, revisa los recibos pendientes con administración.

B Para **cancelar tu matrícula** y pedir la devolución, envía una solicitud a secretaría.

C **MAT-17**: consulta el estado de esta incidencia en el portal de soporte.

MAT-17 es un código que figura en la consulta y en A y C, no un identificador del programa. B no lo tiene, pero habla de cancelar una matrícula.

DOS LISTAS

Cada buscador devuelve sus candidatos

Dos como máximo, de mejor a peor: 1 es la primera posición. Listas ilustrativas, no medidas con un modelo.

Por palabras

Coinciden las palabras, como MAT-17.

1. A recibos pendientes
2. C incidencia en el portal

Por significado

Se parece el significado: anular ≈ cancelar.

1. B cancelar matrícula
2. A recibos pendientes

A está en las dos listas; C solo en la de palabras; B solo en la de significado.

Reunidas, sin repetir: A, B y C

RECIPROCAL RANK FUSION

Cada posición vale unos puntos

RRF da puntos según la posición del texto en cada lista. Así no sumamos un recuento de palabras con una similitud entre vectores.

$$\text{puntos} = 1 / (\text{constante} + \text{posición})$$

Elegimos constante = 1 para poder seguir las cuentas.

Primero en la lista — **0.5**

1 / (1 + 1)

Segundo en la lista — **0.3333**

1 / (1 + 2) ≈ 0.3333

No aparece en la lista — **0**

esa lista no aporta nada

LAS CUENTAS

Sumamos lo que aporta cada lista

Cada texto suma sus puntos de las dos listas y se ordena por el total.

texto	palabras	significado	total
A	0.5 1.º	0.3333 2.º	0.8333
B	0 no está	0.5 1.º	0.5
C	0.3333 2.º	0 no está	0.3333

A: 0.5 + 0.3333 = 0.8333, el total más alto. Orden final: A, B, C.

EN PYTHON

El programa acumula puntos en un diccionario

La función recorre cada lista con sus posiciones y apunta el total de cada texto en un diccionario, scores.

1. Lista por palabras

A es primero: +0.5. C es segundo: +0.3333.

```
{A: 0.5, C: 0.3333}
```

2. Lista por significado

B es primero: +0.5. A es segundo: +0.3333, que se suma a su 0.5.

```
{A: 0.8333, C: 0.3333, B: 0.5}
```

3. Ordenar de mayor a menor

Devuelve solo los identificadores:

```
["A", "B", "C"]
```

EL RESULTADO

Conservamos A y B y recuperamos sus textos

Ordenados por su total (a la derecha), nos quedamos con los dos primeros y recuperamos su texto original.

A	MAT-17: antes de tramitar la baja, revisa los recibos pendientes con administración.	0.8333
B	Para cancelar tu matrícula y pedir la devolución, envía una solicitud a secretaría.	0.5
C	Queda fuera del corte de dos.	0.3333

Recuperamos la condición (recibos) y el procedimiento general.

Son fuentes para consultar: nadie ha comprobado recibos ni ejecutado una cancelación.

LA IDEA

Combinar posiciones, no puntuaciones

La búsqueda híbrida reúne candidatos de dos buscadores y los ordena por posiciones.

1 **Dos señales**
Palabras para MAT-17; significado para «cancelar matrícula».

2 **Puntos por posición**
RRF suma 1 / (constante + posición).

3 **Fusionar no es reevaluar**
Un reranker opcional lee consulta y texto; RRF solo usa posiciones.

4 **Hay límites**
Un texto que nadie propone no se recupera, y el total no es una probabilidad.