CONCEPTO · BÚSQUEDA SEMÁNTICA

Buscar por significado

Encontrar una misma idea expresada con otras palabras.

> **Consulta que escribe una persona**
>
> «Quiero dar de baja mi inscripción»

Tres textos guardados en el buscador

T1 Puedes cancelar tu matrícula escribiendo a secretaría.

T2 La biblioteca abre de lunes a viernes, de nueve a veinte horas.

T3 Las aulas disponen de proyector y conexión wifi.

Compara la consulta con T1, T2 y T3.
En este ejemplo devuelve T1.

01 · ANTES DE USARLO

Qué es un vector

Un vector numérico reúne varios números en un orden fijo. Se puede escribir entre corchetes.

[2, 1]

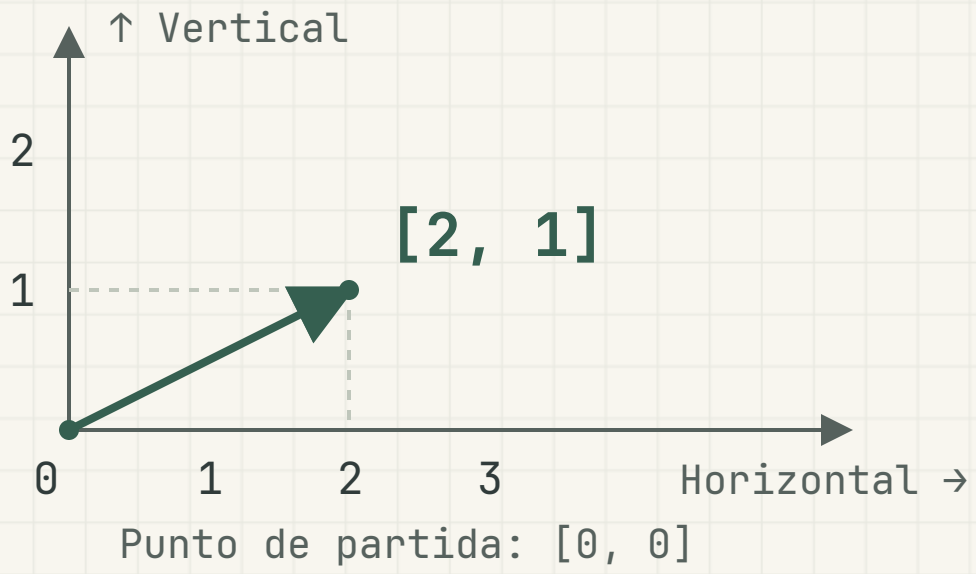
Primera componente	El número 2.
Segunda componente	El número 1.
Dos componentes	Dos números en la lista.

El orden importa: [2, 1] y [1, 2] son vectores distintos. Este es un ejemplo sencillo, no el vector de un texto.

02 · VER LO QUE SIGNIFICAN DOS COORDENADAS

Cómo se dibuja [2, 1]

Desde [0, 0], avanzamos 2 a la derecha y 1 hacia arriba.



2: posición horizontal.

1: posición vertical.

Cada número es una coordenada. Dos componentes: dos dimensiones.

03 · APLICARLO A UNA BÚSQUEDA

Del texto al vector

Un embedding es un vector que representa un texto. Lo calcula un modelo de IA.

> **Consulta de la persona**
>
> «Quiero dar de baja mi inscripción»

↓ Modelo de IA

```
[0.0331, 0.0061, 0.0044,
 -0.0750, …]
```

Vemos 4 de 384 números. Los puntos ocultan los otros 380.

Consulta, T1, T2 y T3 → cuatro listas de 384 números. El plano anterior solo mostraba dos componentes.

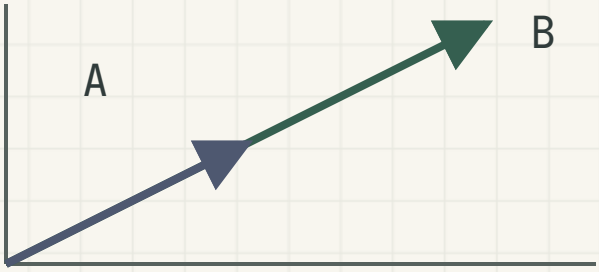
El modelo es un programa entrenado para calcular la lista; los números no se eligen a mano.

04 · COMPARAR LOS VECTORES

Comparar direcciones

El coseno compara direcciones. Más alineación significa un valor mayor, hasta 1.

Ejemplo con 2 componentes: dos flechas alineadas.



A = [2, 1]
B = [4, 2]

Coseno = 1

Con los vectores de los textos de la búsqueda

Consulta con T1 · Cancelación	0.878
Consulta con T2 · Horario	0.810
Consulta con T3 · Equipamiento	0.797

La búsqueda compara los 384 números de cada vector.

05 · ELEGIR Y RECUPERAR

Se elige T1 y se devuelve su texto

0.878 es la mayor de las tres similitudes. Corresponde al texto de cancelación.

> **Consulta de la persona**
> «Quiero dar de baja mi inscripción»

↓ El mejor candidato es T1

«Puedes cancelar tu matrícula escribiendo a secretaría».

«Dar de baja una inscripción» y «cancelar una matrícula» se refieren a la misma gestión.

El programa recupera la frase que tenía guardada. No está generando una respuesta nueva.

06 · QUÉ SIGNIFICA EL RESULTADO

Encontrar un texto no garantiza una respuesta

La búsqueda ha elegido un candidato. Todavía hay que comprobar si sirve.

0.878 no es un 87,8 % de acierto

Es una medida de similitud de este modelo, no una probabilidad de respuesta correcta.

Siempre existe un valor máximo

Si ninguno de los tres textos responde, el programa sigue pudiendo elegir uno.

Devolver T1 no tramita una baja

El buscador muestra una instrucción. No escribe a secretaría ni cancela la matrícula.

LA IDEA DEL CONCEPTO

Buscar textos por su significado

Un vector es la lista de números que permite realizar la comparación.

01 **Representar**
La consulta y cada texto se convierten en una lista de 384 números.

02 **Comparar**
Vector de consulta frente a los vectores de T1, T2 y T3.

03 **Recuperar**
La mayor puntuación señala T1; se devuelve su texto original.

Comparamos números para recuperar un texto.