LLAMAINDEX · RECUPERAR CONTEXTO

La regla sin su excepción

Prospecto o política de devoluciones: la excepción va justo debajo de la regla.

Prospecto de un medicamento

Regla	Tome un comprimido cada ocho horas.
Excepción	En menores de doce años, consulte a su médico antes de usarlo.

Política de devoluciones

Regla	Dispone de treinta días para devolver el producto.
Excepción	Salvo en artículos personalizados.

Un asistente que busca en estos textos y solo recupera la regla responde algo incompleto.

EL CASO

Un manual con dos apartados y tres párrafos

Normas ficticias de room_rules.md. Llamamos P1, P2 y P3 a cada párrafo.

	Apartado «Cancelaciones»
P1	Una reserva puede cancelarse hasta veinticuatro horas antes.
P2	Si faltan menos de veinticuatro horas, debe solicitarse la cancelación en recepción.
	Apartado «Horario»
P3	El centro abre de lunes a viernes, de nueve a veinte horas.

> **Consulta de la persona**
>
> cancelar reserva

EL PROBLEMA

Si solo se devuelve P1, falta la excepción

Se localiza la regla, no lo que la completa.

El buscador elige P1 y es lo único que recibe el asistente

> Una reserva puede cancelarse hasta veinticuatro horas antes.

Queda fuera P2

> Si faltan menos de veinticuatro horas, debe solicitarse la cancelación en recepción.

P1 es cierto, pero incompleto. Con la reserva dentro de diez horas, el asistente diría que ya no puede cancelar; según P2, debe pedirlo en recepción.

Se busca en el párrafo porque un apartado largo mezcla temas y su parte útil pesa menos al compararlo; un párrafo se parece más a una pregunta concreta. Y se devuelve el apartado entero.

LA ORGANIZACIÓN · 1

Nodos: texto, ID y metadatos

LlamaIndex, una biblioteca de Python para conectar documentos con aplicaciones de IA, guarda el manual en nodos.

Padre · un nodo por apartado · clase TextNode

ID	cancellations Lo damos nosotros, como el nombre de una variable.
Texto	P1 + P2, el apartado completo
Metadatos	room_rules.md · Cancelaciones

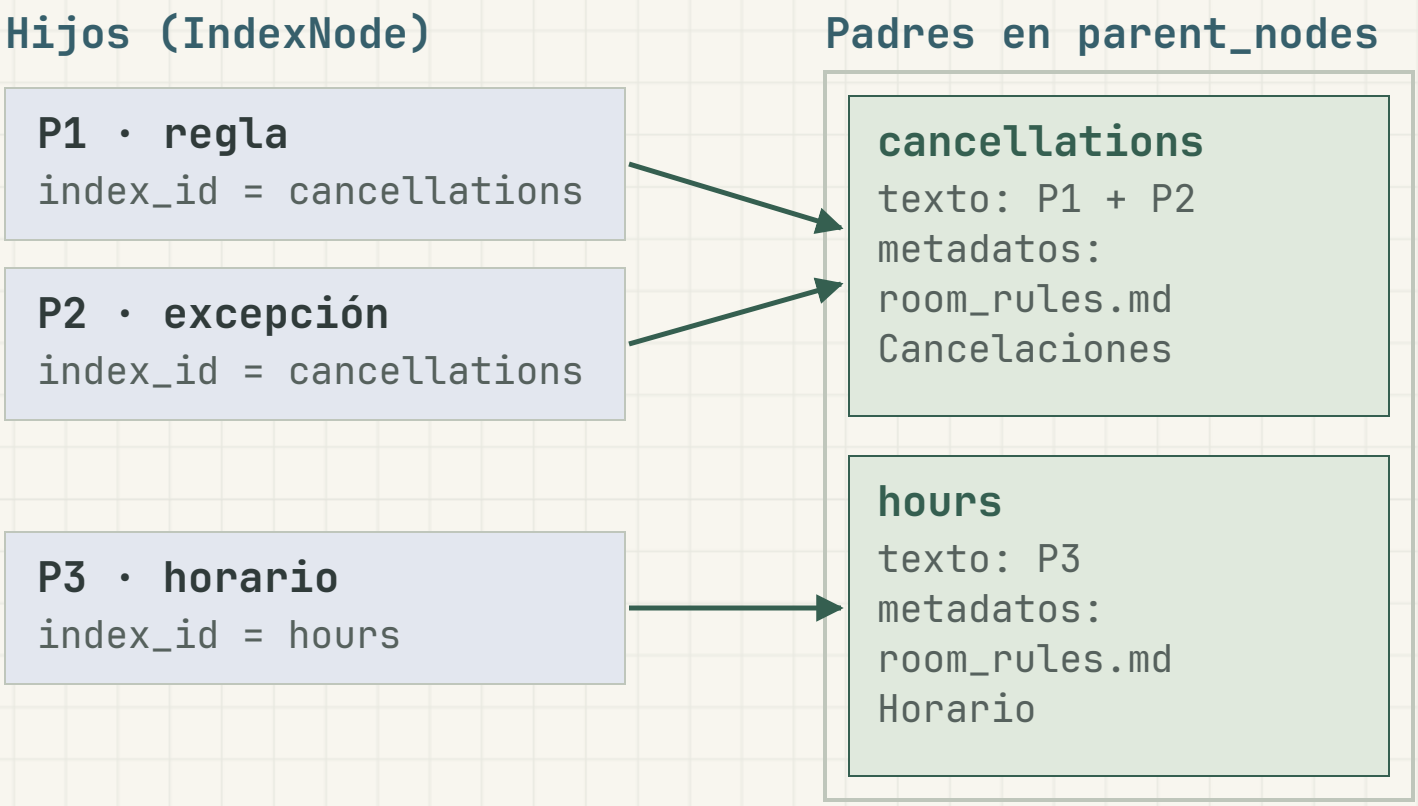
Hijo · un nodo por párrafo · IndexNode (un TextNode con index_id)

Texto	P1
ID	Lo genera LlamaIndex
index_id	cancellations El ID del padre al que apunta.

LA ORGANIZACIÓN · 2

Cada hijo apunta a su padre con index_id

Los padres se guardan aparte, en el diccionario parent_nodes.



Cada hijo guarda el ID de su padre. El ID solo sirve para localizar el apartado; no contiene sus normas.

LA BÚSQUEDA

El recuperador BM25 puntúa los hijos

Un recuperador recibe una consulta y devuelve nodos. BM25 puntúa P1, P2 y P3 y devuelve el mejor.

Palabras de la consulta que aparecen en cada hijo

P1 · regla — **2/2**

contiene **cancelarse** y **reserva**

P2 · excepción — **1/2**

contiene **cancelación**, pero no «reserva»

P3 · horario — **0/2**

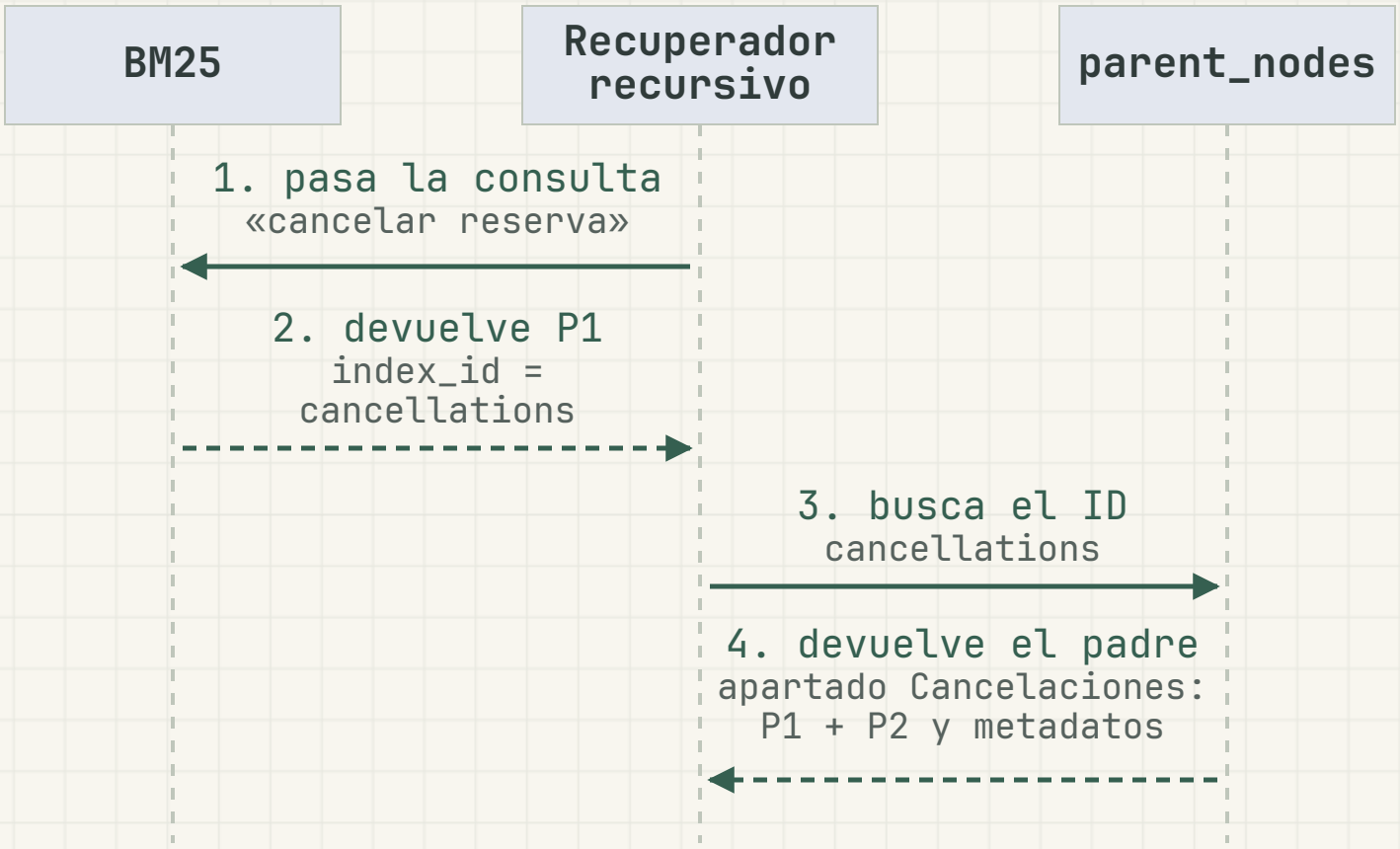
ninguna palabra de la consulta

«cancelar», «cancelarse» y «cancelación» comparten la raíz «cancel». BM25 calcula una puntuación; aquí contamos coincidencias.

EL RECORRIDO

Del hijo encontrado al padre

El recuperador recursivo envuelve a BM25: recibe la consulta, se la pasa y, con el index_id de P1, busca en parent_nodes.



La aplicación muestra el texto de P1 y P2 con su archivo y su título.

DÓNDE COMPENSA

Buscar en pequeño, devolver en grande

Compensa con apartados largos; con apartados diminutos no hace falta.

1 **Condiciones de venta**
Un plazo con su excepción.

2 **Normativa y convenios**
Un artículo con su regla y sus supuestos.

3 **Documentación técnica**
Se busca una línea; se devuelve la sección.

4 **Manuales de procedimiento**
Cada paso con su advertencia.

Tampoco hace falta con fichas o preguntas de una línea.

Si BM25 no encuentra ninguna palabra, su candidato puntúa 0, pero el recuperador recursivo entrega el padre con 1.0: descarta antes ese hijo.