

Ejercicio 1: Clave de búsqueda: nombre.

Cantidad de claves de búsqueda:

$$4096 = N * \text{tamaño nombre} + (N+1) * 6 = N * 40 + (N + 1) * 6 = 46 N + 6$$

$4090 = 46 N$, $N = 88,91$. Luego $N = 88$ claves de búsqueda por nodo.

Altura del árbol B+

$$\begin{aligned}\text{Altura de árbol B+} &= h = \lceil \log_{10}(5000) / \log_{10}(44) \rceil \\ &= \lceil 3,698 / 1,643 \rceil = \lceil 2,25 \rceil = 3\end{aligned}$$

Operador de búsqueda: selección para índice primario con igualdad para no clave candidata.

Si un nombre no se repite mas de 15 veces, como el índice es primario en nombre, la tabla usuarios esta ordenada por nombre.

Una fila de usuarios contiene: $40(\text{nombre}) + 6(\text{uid}) + 6(\text{puntero}) = 52$

En un bloque de usuarios entran $= 4096/52 = 79$ tuplas.

Luego en el peor caso hay que acceder a dos bloques para encontrar todos los usuarios de un mismo nombre.

El costo se compone de: $(h + 2)$ transferencias de bloques.

Ejercicio 2:

Parte 1: para cada término: una tupla del índice invertido tiene identificador de documento (donde aparece el término) y frecuencia de documento o TF-IDF. Esta información me permite calcular la relevancia de cada documento que satisface la consulta y ordenar los resultados por relevancia.

Parte 2:

1. Buscar el vocabulario de la consulta.
2. Calcular el vector de la consulta.
3. Para cada término se busca la información de documentos en el índice invertido ($S'i$ para termino i).
4. Calcular los $S'i$ (a partir de $S'i$).
5. Evaluar la expresión de consulta booleana aplicando los operadores. Se obtiene así un conjunto de documentos.
6. Calcular similitud de vectores de cada documento arrojado por la búsqueda y el vector de la consulta.
7. Ordenar los documentos del resultado de la búsqueda por esas similitudes de vectores.

Ejercicio 3:

Análisis de los contextos:

- **Contexto local:** sonido: campanas de la iglesia, protestas globales: cambio climático, vivian en armonía: con la naturaleza, esa conexión: armonía con la naturaleza (oración anterior), esos tiempos: de nuestros ancestros.
- **Contexto global:** El tema general del texto es la desconexión moderna con la naturaleza y la necesidad de recuperar la armonía ambiental.
- **Contexto situacional:** La acción de tomar mate en la plaza implica un momento de tranquilidad y reflexión, lo que facilita la conexión con pensamientos más profundos y globales.
- **Contexto histórico:** La referencia a los ancestros y a cómo vivían en armonía con la naturaleza evoca un contexto histórico pasado, contrastándolo con la situación actual.
- **Contexto cultural:** La costumbre de tomar mate es una referencia clara a la cultura latinoamericana, y la mención de las protestas globales refleja un contexto cultural más amplio, relacionado con la conciencia ambiental y los movimientos sociales.

Ejercicio 4: ¿Hay alguna capa de un codificador de un transformer que haga reconocimiento de intenciones? ¿Alguna capa de decodificador hace la determinación del contenido? ¿Alguna capa del decodificador hace estructuración de la respuesta y planificación de frases? ¿Un transformer hace un post procesamiento de su output? Cuando no hay capas que hagan explícitamente una de estas tareas, indicar por qué otra tarea se lo reemplaza y qué capas colaboran en ella.

Los transformadores no hacen un reconocimiento de intenciones explícito antes de generar una respuesta. En lugar de eso, la intención está implícita en la información de contexto generada. Se tiene un proceso acumulativo donde cada capa contribuye a una representación más refinada del input. Las capas de auto-atención son clave aquí, ya que permiten al modelo captar las relaciones contextuales y semánticas entre los tokens, lo que indirectamente ayuda en el reconocimiento de intenciones.

Determinación del Contenido: Las capas de atención cruzada en el decodificador juegan un papel crucial al enfocarse en la representación contextual proporcionada por el codificador. Esto ayuda a determinar qué contenido es relevante para generar la respuesta. Los transformadores no determinan todo el contenido de antemano. En su lugar, generan el contenido de forma dinámica,

palabra por palabra, basándose en el input y el contexto previo. A cada paso, consideran el texto generado anteriormente y el contexto para decidir la palabra siguiente, ajustando el contenido sobre la marcha. Este proceso permite que las respuestas sean coherentes y relevantes, adaptándose continuamente al desarrollo del texto

Estructuración de la Respuesta: Aunque no estructuran toda la respuesta por adelantado de manera explícita, emplean una arquitectura que les permite comprender el contexto y generar cada palabra de manera informada.

Planificación de frases: En los transformadores, no hay una planificación explícita de frases completas. En lugar de eso, el modelo genera cada palabra secuencialmente, basándose en el contexto acumulado del texto anterior y el input. La capa lineal, junto con los mecanismos de atención, permite al modelo decidir la siguiente palabra a partir de la probabilidad de las posibles opciones en ese momento. La combinación de auto-atención y atención cruzada permite al modelo obtener una respuesta coherente y contextualmente apropiada.

Postprocesamiento del output: en los transformers usualmente no hay una fase explícita de postprocesamiento dentro del modelo mismo. Sin embargo, en implementaciones prácticas, se pueden aplicar algunas técnicas adicionales, como corrección gramatical, ajuste de formato, o filtros específicos para mejorar la calidad y coherencia de la respuesta generada.