

Instituto Tecnológico de Costa Rica

Campus Tecnológico Central Cartago

Escuela de Ingeniería en Computación

Bases de Datos II

IC-4302

Resumen 9

Fecha de entrega: 30/05/2023

I-Semestre 2023

Profesor:

Nereo Campos Araya.

Elaborado por:

Diana Sanabria Calvo 2021436548.

Inverted indexes: Types and techniques

Los sistemas de recuperación de información utilizan índices invertidos como estructuras de datos para agilizar los procesos de búsqueda.

Los índices invertidos se analizan en los párrafos iniciales del artículo junto con su importancia para mejorar la eficacia de la búsqueda. Se demuestra cómo los índices invertidos se usan con frecuencia en bases de datos y motores de búsqueda para facilitar búsquedas de palabras clave rápidas y efectivas.

Hay varios tipos de índices invertidos. El índice posicional invertido, que almacena tanto los términos como sus ubicaciones dentro de los documentos indexados, se menciona como el tipo más frecuente. Como resultado, se mejoran la coincidencia de resultados y la clasificación de relevancia. Los índices de dos palabras, los índices invertidos comprimidos y los índices invertidos distribuidos son variedades adicionales de índices invertidos. Luego se trata en detalle la creación de índices invertidos. Se analiza cómo tokenizar documentos de texto, lo que los separa en términos o tokens separados. Los autores enfatizan el valor de los métodos de lematización y derivación para disminuir las variaciones de palabras y aumentar la precisión de la búsqueda.

A lo largo del artículo se mencionan los métodos de compresión de índices invertidos. Mencionan métodos como la codificación de bytes variables, la codificación gamma y la compresión basada en diccionarios que reducen la cantidad de almacenamiento necesario para los índices invertidos y al mismo tiempo, mantienen un procesamiento de consultas eficaz. También cubre las técnicas de indexación incremental y manejo de actualizaciones. Se explica cómo actualizar el índice invertido de manera efectiva sin tener que volver a indexar toda la colección utilizando técnicas como la indexación basada en combinación y la indexación delta.

Por último se concluye con un resumen de los índices invertidos, enfatizando sus diferentes tipos y métodos. Cubre varias aplicaciones, estrategias comerciales, tokenización, ponderación de términos, compresión y varios tipos de índices invertidos.