
**DISEÑO Y DESARROLLO DE UN ÍNDICE BASADO EN
HIPERPLANOS PARA BÚSQUEDA EN ESPACIOS MÉTRICOS**

**JOSÉ MARÍA ANDRADE ANDRADE
INGENIERO CIVIL EN COMPUTACIÓN**

RESUMEN

La constante evolución de los datos en los últimos tiempos ha provocado que organizar la información con los métodos tradicionales se vuelva muy difícil. Es por esto que la investigación en esta área ha crecido bastante. Uno de los problemas más importantes al momento de administrar bases de datos con los métodos tradicionales, es que existen algunos conjuntos de datos donde la búsqueda exacta no tiene mayor sentido. En estos casos es necesario considerar la búsqueda por proximidad o similitud. Para enfrentar este nuevo problema los datos se pueden modelar como un espacio métrico (EM), que se define como el par $(X; d)$, en que X denota el universo de objetos válidos y d una función de distancia (no negativa) entre cualquier par de objetos, que generalmente es costosa de calcular. Es por esta última razón que los EM se indexan con estructuras que permiten reducir el número de cálculos de distancia al momento de realizar búsquedas por similitud.

Existen índices que permiten resolver consultas por similitud de forma exacta.

Sin embargo, hay varios espacios métricos de interés práctico donde el rendimiento de estos índices decae. En estos casos existen alternativas no exactas, como por ejemplo el índice basado en permutantes (IBP). Este índice permite cambiar el espacio del problema a uno en donde es más barato calcular la distancia y de esta forma definir un orden de revisión promisorio al momento de resolver las búsquedas.

En esta memoria se proponen dos índices no exactos que utilizan una estrategia basada en el particionado por hiperplanos. En ellos cada elemento se representa con una firma binaria que es una secuencia de bits que se calcula a partir de los elementos que constituyen los hiperplanos. Similar al caso del IBP, al representar los elementos con firmas binarias se puede definir un orden promisorio de revisión. Una característica importante de este trabajo es la utilización de poco espacio de memoria.

Para evaluar los índices propuestos, se realiza una comparación de rendimiento a mismo espacio de memoria utilizada con respecto al IBP. Los resultados obtenidos para el primer índice muestran un mejor comportamiento al momento de realizar búsquedas. Por su parte, aún cuando el segundo índice no logró superar a los permutantes es importante continuar con su investigación, ya que podría ser útil para categorizar a los elementos.