
**BÚSQUEDA APROXIMADA EN ESPACIOS MÉTRICOS USANDO
AGRUPAMIENTOS BASADOS EN DISTANCIAS CUANTIZADAS**

**MAXIMILIANO ADOLFO VERDUGO WALSEN
INGENIERÍA CIVIL EN COMPUTACIÓN**

RESUMEN

La evolución de la cantidad y calidad de la información que es manipulada va de la mano con el rápido crecimiento que ha presentado el área de la computación. Junto con esto, las bases de datos han tenido que adaptarse para poder dar abasto a todas las necesidades, las cuales no contemplan únicamente el almacenamiento de la información, sino que también deben proveer formas de obtener información según se requiera. Dentro de las búsquedas que se solicitan a una base de datos comúnmente se encuentran las búsquedas exactas, es decir, dado un identificador único se quiere conseguir información vinculada con éste. Por ejemplo, al realizar una búsqueda de la información de morosidad asociada a una persona a través de su RUT. Por otro lado dada la gran cantidad de características que existen aparece otra necesidad, la cual consiste en obtener los elementos que son los más parecidos a otro. Siguiendo con el ejemplo, buscar el conjunto de personas que tengan las mismas características respecto al pago de deudas como: tiempo de morosidad del pago de cuentas, cantidad de créditos obtenidos, montos adeudados y muchas otras características que son agregadas según la cantidad de información capturada. A este tipo de búsquedas se les llama búsquedas de proximidad, en donde se tiene un elemento de consulta y una función de distancia que indica que tan similares son dos elementos, para obtener el conjunto de elementos más parecidos a la consulta. Actualmente existen métodos para indizar datos no convencionales que entregan una respuesta exacta. Sin embargo el gran problema que estos sufren es la cantidad de memoria RAM que requieren. También existen índices de búsqueda que carecen de este problema, en los que no se asegura que el resultado obtenido sea el correcto, pero si es muy aproximado a éste. El propósito de ésta memoria consiste en el diseño e implementación de un índice nuevo que provecha las características de índices probabilísticos existentes y que busca reducir el esfuerzo de la búsqueda a una porción de la base de datos. Los resultados obtenidos indican que la exactitud del índice propuesto es menor respecto al índice probabilístico del que éste se

basa. Pese a esto, los tiempos de reacción obtenidos se mantienen constantes y predecibles, superando en algunos casos a su antecesor.