

# Sumário

Um método de acesso multidimensional (MAM) é uma estrutura de indexação voltada ao suporte de objetos espaciais, especialmente de retângulos. O principal objetivo de um MAM é propiciar uma rápida obtenção dos objetos espaciais que satisfazem um certo relacionamento topológico, métrico ou direcional. Neste sentido, o espaço indexado é organizado de tal forma que, por exemplo, a recuperação dos retângulos de dados contidos em uma área particular requiera apenas o acesso aos retângulos próximos a esta área, em oposição à análise do conjunto completo de retângulos armazenados em memória secundária. Um MAM, portanto, é projetado como um caminho otimizado aos dados espaciais e o seu uso melhora significativamente o desempenho de sistemas gerenciadores de banco de dados espaciais no processamento de consultas. Nesta tese, nós investigamos o desempenho de um conjunto de MAM, a maioria dos quais tem sido identificado na literatura como um MAM muito eficiente no suporte a consultas espaciais de seleção. Este grupo consiste dos seguintes métodos de acesso: **R-tree**, **R-tree Greene**, **R\*-tree**, **Hilbert R-tree**, **SR-tree** e três variantes da **R\*-tree** chamadas de **R\*-tree CR** (i.e., *close reinsert*), de **R\*-tree FR** (isto é, *far reinsert*) e de **R\*-tree WR** (isto é, *without reinsertion*). A comparação do desempenho destes MAM foi realizada visando-se analisar prioritariamente a influência do fator distribuição espacial dos dados. Neste sentido, nós propusemos uma metodologia de avaliação de desempenho que permite a geração de um conjunto de tipos de distribuição espacial com diferentes características, as quais tornam possível que a influência do fator distribuição espacial dos dados seja analisada sob diferentes perspectivas, desde uma fraca até uma forte influência. Por meio de diversos testes de desempenho, nós observamos de que forma a distribuição espacial dos dados afetou os custos de inserção e de armazenamento de novas entradas no índice espacial, além do custo de *point queries*, *intersection range queries*, *enclosure range queries* e *containment range queries*. Com relação a estas consultas espaciais de seleção, os resultados de desempenho mostraram que a **R\*-tree** foi a melhor estrutura de indexação espacial para *point queries* e *enclosure range queries*, ao passo que as variantes da **R\*-tree** produziram os melhores resultados de desempenho para *intersection* e *containment range queries*. Por outro lado, os métodos **Hilbert R-tree** e **SR-tree** geraram um baixo desempenho para as quatro consultas espaciais investigadas. No entanto, em testes de desempenho adicionais, os quais modificaram tanto o tamanho quanto o formato dos retângulos de dados, os métodos de acesso **Hilbert R-tree** e **SR-tree** geraram resultados competitivos, particularmente para *intersection* e *containment range queries*.

**palavras-chave:** métodos de acesso multidimensionais, estruturas de indexação espacial, árvores-R, distribuição espacial dos dados, análise de desempenho, técnica experimental de *benchmark* e banco de dados espaciais.

# Abstract

A multidimensional access method (MAM) is an index structure designed to the support of spatial data objects, mainly data rectangles. The main goal of a MAM is to provide a fast retrieving of spatial data objects that satisfy a given topologic, metric, or directional relationship. In this sense, the indexed space is organized in such a way that allows the retrieval of data objects that are located close to each other. For instance, the execution of spatial queries about data objects contained in a certain area just requests the access to data objects located close to this area, as opposed to the analysis of the whole set of spatial data objects stored in the secondary memory. A MAM, therefore, is designed for providing an optimized access path to spatial data objects and its utilization improves considerably the performance of spatial database management systems in the query processing. In this thesis, we have investigated the performance variation of a certain group of MAM, most of which has been identified in the literature as a very efficient MAM in the support of spatial selection queries. This group consists of the following access methods: **the R-tree**, **the R-tree Greene**, **the R-tree**, **the Hilbert R-tree**, **the SR-tree** and three variants of **the R\*-tree**, called as **the R\*-tree CR** (that is, Close Reinsert technique), **the R\*-tree FR** (that is, Far Reinsert technique) and **the R\*-tree WR** (that is, Without Reinsertion). The performance comparison of these MAM was accomplished aiming at analyzing mainly the influence of the spatial data distribution. In this sense, we have proposed a performance evaluation methodology that allows the generation of a set of spatial data distribution types, which have different characteristics to each other. This set effectively enabled us to analyze the influence of the spatial data distribution from distinct perspectives, since a weak until a strong influence. Through several experimental performance tests, we have observed the way that the spatial data distribution has affected the storage and the insertion costs for new entries into the index structures. Besides, we have analyzed query costs for point queries, intersection range queries, enclosure range queries, and containment range queries. In particular, for spatial selection queries, the performance results have showed that **the R-tree** was the best spatial index structure for the point queries and the enclosure range queries, whereas **the R\*-tree** variants have produced the best outcomes for the intersection and the containment range queries. On the other hand, **the Hilbert R-tree** and **the SR-tree** have generated a low performance for any investigated spatial selection query. However, additional performance tests have showed that a variation in both the size and the shape properties of the data rectangles provided some competitive results for **the Hilbert R-tree** and **the SR-tree**, especially for the intersection and the containment range queries.

**keywords:** multidimensional access methods, spatial index structures, **R-trees**, spatial data distribution, performance evaluation, database benchmark technique, spatial databases.