Title: Tree-based Indexing Methods for Similarity Search in Metric and Nonmetric Spaces

Author: Mgr. Jakub Lokoč
Department: Department of Software Engineering
Faculty of Mathematics and Physics
Charles University in Prague
Supervisor: Doc. RNDr. Tomáš Skopal, Ph.D.
Author's e-mail address: lokoc@ksi.mff.cuni.cz
Supervisor's e-mail address: skopal@ksi.mff.cuni.cz

Abstract: The M-tree is a well-known indexing method enabling efficient similarity search in metric spaces. Although the M-tree is an aging method nowadays, we believe it still offers an undiscovered potential. We present several approaches and directions that show how the original M-tree algorithms and structure can be improved. To allow more efficient query processing by the M-tree, we propose several new methods of (parallel) M-tree construction that achieve more compact M-tree hierarchies and preserve acceptable construction cost. We also demonstrate that the M-tree can be simply extended to a new indexing method – the NM-tree, which allows efficient nonmetric similarity search by use of the TriGen algorithm. All these experimentally verified improvements show that the M-tree can still be regarded as an important dynamic metric access method suitable for management of large collections of unstructured data. Moreover, all the improvements can be further adopted by M-tree descendants (e.g. the PM-tree), so that the results presented in this thesis open the door for future research in this area.

Keywords: similarity search, metric access methods, indexing, M-tree, TriGen