Sokolov, A. M.
Vector representations for efficient comparison and search for similar strings. (English. Russian original) Zbl 1148.68017

Summary: A method is proposed for approximation of the classic edit distance between strings. The method is based on a mapping of strings into vectors belonging to a space with an easily calculable metric. The method preserves the closeness of strings and makes it possible to accelerate the computation of edit distances. The developed q-gram method of approximation of edit distances and its two randomized versions improves the approximation quality in comparison with well-known results.

MSC:
68P10 Searching and sorting

Keywords:
approximating edit distance; metric embedding; approximate nearest-neighbor search; neural-network-based information technologies

Full Text: DOI

References:


This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.