

On the Construction of Minimal Perfect Hashing and Compressed Datastructures using the CONSENSUS Technique

Praxis der Forschung 2025

Betreuer: Stefan Walzer, Stefan Hermann

Abstract

Space efficiency is an essential metric for the performance of compressed data structures like perfect hash functions and retrieval data structures. Perfect hash functions assign a unique name to each key of a given input set. Retrieval structures store a mapping from keys to respective values. If the values underlie a known probability distribution, there is further potential for compression. The potential for compression is not yet fully exhausted in state-of-the-art techniques. Recently, the development of a generalized compression technique, called CONSENSUS, has created promising research opportunities in the field of compressed data structures.

The project investigates how CONSENSUS can be applied to perfect hashing and retrieval structures. It is conceivable that an overarching approach can be established that solves perfect hashing and simultaneously retrieval structures in a particularly space efficient manner. The aim of this project is to obtain a theoretical understanding of such an overarching approach and to beat the state-of-the-art in terms of the trade-off between construction throughput, query throughput and space efficiency.