

October 2009

MADALGO seminar by Rasmus Pagh, IT University of Copenhagen

Storing a Compressed Function with Constant Time Access

Abstract:

We consider the problem of representing, in a space-efficient way, a function $f: S \rightarrow \Sigma$ such that any function value can be computed in constant time on a RAM. Specifically, our aim is to achieve space usage close to the 0th order entropy of the sequence of function values. Our technique works for any set S of machine words, without storing S , which is crucial for applications.

Our contribution consists of two new techniques, of independent interest, that we use in combination with an existing result of Dietzfelbinger and Pagh (ICALP 2008). First of all, we introduce a way to support more space efficient approximate membership queries (Bloom filter functionality) with arbitrary false positive rate. Second, we present a variation of Huffman coding using approximate membership, providing an alternative that improves the classical bounds of Gallager (IEEE Trans. Information Theory, 1978) in some cases. The end result is an entropy-compressed function supporting constant time random access to values associated with a given set S . This improves both space and time compared to a recent result by Talbot and Talbot (ANALCO 2008).

Joint work with:

Johannes Hreinsson
Morten Krøyer