

September 2011

MADALGO seminars by Djamel Belazzougui, Université Paris Diderot

Applications of Minimal Perfect Hashing in Compressed Full-Text Indexing

Abstract:

Given a set S of n numbers from the range $[1..U]$, a monotone minimal perfect hashing [SODA 2009] is a function which associates to each element x of S its rank among all elements of S . The function is allowed to return an arbitrary answer when evaluated on an element outside of S . In [SODA 2009] it is shown that a monotone minimal perfect hash function can be encoded using $O(n \log \log(U/n))$ bits of space (which is sublinear when compared with the $O(n \log \log(U/n))$ bits needed to encode the set S) such that the evaluation takes $O(1)$.

Given a text T , a full-text index on T is an index which supports queries which can efficiently search for patterns in T . The full-text index is said to be compressed if it occupies a space close to the space occupied by the compressed text.

In this talk, I will show how monotone minimal perfect hash functions can be used to speed-up some known compressed text indexes.

Joint work with Gonzalo Navarro