Efficient text indexing and search

Mgr. Adam Kövári

Search engines need to be able to efficiently perform great amounts of queries over huge indexes of text. Increasing operational costs of search engines are driving research in this area. This presentation focuses on some modern ways to make search more efficient.

This presentation will explain the basic theory and concepts of search data structures, matching problem, and its typical solutions, such as inverted indexes. Additionally, traditional inverted indexes will be compared with modern approaches, such as BitFunnel [1]. To set a baseline for comparison of BitFunnel approach, it is important to study improvements made to inverted indexes, in this case partitioned Elias-Fano indexes [2]. Elias-Fano monotone sequences are explained, their use in inverted indexes and their further improvements, specifically partitioning is considered. Finally, bit-string signature-based indexes and bit-sliced signature-based indexes are explained and are compared to partitioned Elias-Fano indexes.

References

[1] BitFunnel: Revisiting Signatures for Search: https://danluu.com/bitfunnel-sigir.pdf

[2] Partitioned Elias-Fano Indexes: http://www.di.unipi.it/~ottavian/files/elias_fano_sigir14.pdf