

A Web Search Engine Model Based on Index-Query Bit-Level Compression

Hussein Al-Bahadili¹, Saif Al-Saab¹, Reyadh Naoum¹ Shakir M. Hussain²

¹The Arab Academy for Banking & Financial Sciences
P.O. Box 13190
Amman 11942, Jordan
00962-6-5552900
Hbahadili@aabfs.org, saif.alsaab@gmail.com,
Rnaoum@aabfs.org

²Petra University
P.O. Box 961343
Amman 11196, Jordan
00962-6-5799555
shussain@uop.edu.jo

ABSTRACT

In this paper, we propose a new web search engine model based index-query bit-level compression. The model incorporates two bit-level compression layers both implemented at the backend processor (server) side, one layer resides after the indexer acting as a second compression layer to generate a double compressed index, and the second layer be located after the query parser for query compression to enable bit-level compressed index-query search. This contributes to reducing the size of the index file as well as reducing disk I/O overheads, and consequently yielding higher retrieval rate and performance. The data compression scheme used in this model is the adaptive character wordlength (ACW(n,s)) scheme, which is an asymmetric, lossless, bit-level scheme that permits compressed index-query search. Results investigating the performance of the ACW(n,s) scheme is presented and discussed.

Keywords

Web search engine; index files; query optimization; full-text compressed self-index; bit-level data compression; ACW(n,s) scheme.