Abstract

In information retrieval system, inverted index is used to evaluate query. More documents to be store can causes larger inverted index to be create and more queries that must be processed in search system. So, needed an optimization query, one of which is inverted index compression.

Inverted index is expected to reduce storage space requirements and increase the usage of memory cache, thus avoiding the full access to the disk during query evaluation. One of the inverted index compression method is the Gamma code. Gamma code is one of the compression technique that turns an integer into a binary codeword. Compressed data is document ID and term frequency.

The testing is an implementation of inverted index compression in information retrieval system with a small document collection and a large document collection. From the analysis of test results, we conclude that Gamma code can increase performance in size of inverted index and query processing time on a large document collection. In inverted index size, the terms of large document collection are distributed in many documents, so it result shorter encoding of Gamma code. In query processing, average query processing time ratio of large document collection is lower than small document collection.

Keywords*: compression, inverted index, gamma code, integer*