Thesis Title Text Compression by Sorting Transformation

Name Somphong Lerwongrat

Degree Master of Science (Computer Science)

Thesis Supervisory Committee

Damras Wongsawang, Ph.D.

Supachai Tangwongsan, Ph.D.

Date of Graduation 15 May B.E. 2540 (1997)

ABSTRACT

The traditional compression paradigm; "Modeling and Coding" may be alternatively replaced by a new paradigm; "Transformation, Modeling and Coding". The recent works described the Block-sorting Transformation and lossless compression algorithms that give the compression ratio comparable to those algorithms currently in use. The transformation processes the raw input into a sequence of less disorder form to obtain more compressible form. It is already known that this is a context-based compressor of unbounded order. The originating of new paradigm starts to restructure contexts by sorting phase then processes the permuted text with the Move-to-front and finally the statistical compressor is applied. This technique not only gives good speed but also an excellent compression ratio.

This research studied this new technique in detail. The Block-sorting transformation was explored and studied in many aspects, the effects of those changed context and their entropy, several different orders context modeling. Finally various types of lossless compression were applied to improve the compression performance eventually yielding a compressor which was the best of this type. It was shown that the Block-sorting technique is comparable to other compressors in terms of compression ratio and speed.