

Goal of this work was to design and test methods for creating and parsing input data from file or stream in a way they are optimized for following compression algorithms. We used two runs through data, first run collected analytical data and second run was for parsing. First part of this document is focused on methods TD2 and TD3 [4] for trie compression and their extension for unicode alphabets. In second part there are presented methods for creating dictionary without considering partitioning on words or syllables. All methods are considered over unicode alphabet and implemented as part of modular program XBW.