

The goal of this thesis is an effective implementation of the methods of multiword expression extraction from text, so that designed program would be capable of processing large textual corpora containing up to billions of words. Additional function of the program is context tracing of extracted N-grams. For thesis purposes the program implementation is specially adjusted for collocation extraction from The Prague Dependency Treebank, but the program is designed in such manner that allows an easy future extensibility.