Verb-Noun Collocation and Government Model Extraction from Large Corpora

Vladislav Tushkanov, Oksana Dereza

National Research University Higher School of Economics,
v.tushkanov@outlook.com, oksana.dereza@gmail.com

Abstract. Knowing the government model, or argument structure, of a verb is crucial for many NLP tasks. In this article, a method of automatic extraction of verbs from large annotated corpora is devised. This method allows to computationally efficiently extract government models and particular arguments for every verb using a simple window-based approach by iterating through each sentence with a window of fixed size and applying frequency filters to filter out noise.

Keywords: collocations, verb government models, argument structure, corpus methods, parsing