

TOWARDS CLUSTERING FROM CORPORA BY USING PATTERN ORIENTED APPROACHES

DANA AVRAM LUPȘA

ABSTRACT. Clustering from untagged corpora is very important especially for languages that have no such hierarchies as WordNet. The idea is to combine automatic noun clustering from unannotated corpora with some supervised learning methods. This paper presents a study on automatic noun clustering from texts selected from corpora by using a pattern-oriented filter. The patterns used are oriented to Romanian language but can be extended also to other languages. A comparison between results obtained by not applying any pattern and by applying different patterns (as filters) is also presented.

BABES BOLYAI UNIVERSITY
COMPUTER SCIENCE DEPARTMENT
KOGALNICEANU 1, 400084 CLUJ-NAPOCA, ROMANIA
E-mail address: `davram@cs.ubbcluj.ro`