

Conceptually related lexicon clustering based on word context association mining.

ABSTRACT

Automatic lexicon generation is a useful task in learning text fragment patterns. In our previous work we have focused on text fragment pattern learning through the fuzzy grammar method which inputs include a predefined lexicon and text fragments that represents the expression of the grammar class to be learned. However, the bottleneck of the success of the fuzzy grammar creation and in common with other text learner often lies in the knowledge acquisition phase; due to the labour intensive text annotation which also demands skills and background knowledge of the text. For this reason, a semi-automated technique called automatic Terminal Grammar Recommender (TGR) is devised to identify conceptually related lexicons in the texts and their related to create terminal grammars by mining associations of words contexts. The approach recognizes that there is a degree of local structure within such text and the technique exploits the local structure without the large computational overhead of deeper analysis. Result from the comparison of the associative words detected by TGR with the definition of a content category tool called General Inquirer on the data from European Central Bank data is reported. Our findings show that our proposed method has managed to reduce the manual effort of identifying conceptually similar lexicons to form terminal grammars. The average of matched generated terminal grammar clusters compared to General Inquirer is 54.85% which indicates that at least half the expensive effort to construct conceptually related lexicon is saved. This hint the potential of word context association mining in automated conceptual lexicon generation.

Keyword: Conceptual lexicon relation; Word context mining; Terminal grammar; Fuzzy grammar.