

Linguistic discourse refers to the meaning of larger text segments, and could be very useful for guiding attempts at text mining such as document selection or summarization. The aim of this project is to apply discourse information to Knowledge Discovery in Databases. As far as we know, this is the first attempt at combining these two very different fields, so the goal is to create a basis for this type of knowledge extraction. We approach the problem by extracting discourse relations using unsupervised methods, and then model the data using pattern structures in Formal Concept Analysis. Our method is applied to a corpus of medical articles compiled from PubMed. This medical data can be further enhanced with concepts from the UMLS MetaThesaurus, which are combined with the UMLS Semantic Network to apply as an ontology in the pattern structures. The results show that despite having a large amount of noise, the method is promising and could be applied to domains other than the medical domain.