Abstract
Here, we address the task of assigning relevant terms to thematically and semantically related sub-corpora and achieve superior results compared to the baseline performance. Our results suggest that more reliable sets of keyphrases can be assigned to the semantically and thematically related subsets of some corpora if the automatically determined sets of keyphrases for the individual documents of an entire corpus are identified first. The sets of keyphrases assigned by our proposed method for the workshops present in the ACL Anthology Corpus over a 6-year period were considered better in more than 60% of the test cases compared to our baseline system when evaluated against an aggregation of different human judgements.

Keywords
Multi-document keyphrase extraction, knowledge management, information retrieval.