

Sort all the report documents of the topic in chronological order and process the documents according to the time sequence



Process the documents with word segmentation, NER, POS tagging and word frequency statistics, get the characteristic word set and form the content vector and time vector of the document



Calculate the similarity between the document and the existing event. If the largest similarity is greater than the threshold θ , join the document into the event set which the largest similarity is corresponded to. After that, update the vector space of the event set



If all the similarities are smaller than the threshold θ , create a new event, join the document into the new event and update the vector space of the event set



After processing all the report documents, the algorithm ends