Abstract

Over the past decades systems for automatic management of electronic documents have been one of the main fields of research. Text processing is a wide area that includes many important disciplines. In the processes of organizing unstructured text in order to implement a mining technique, preprocessing has to be applied. One of the most important preprocessing techniques is the removal of functional words which affects the performance of text mining tasks. In this paper, a statistical approach is presented to extract Arabic stop-words list. The extracted list was compared to a general list. The comparison yield an improvement in an ANN based classifier using the generated stop-words list over the general list.

References

Toward an ARABIC Stop-Words List Generation


Index Terms

Computer Science  Artificial Intelligence
Keywords
Arabic Text Processing  Stop-word List Generation