Abstract

In this study we have implemented the Kmeans and Kmediods algorithms in order to make a practical comparison between them. The system was tested using a manual set of clusters that consists from 242 predefined clustering documents. The results showed a good indication about using them especially for Kmediods. The average precision and recall for Kmeans compared with Kmediods are 0.56, 0.52, 0.69 and 0.60 respectively. We have also extract feature set of keywords in order to improve the performance, the result illustrates that two algorithms can be applied to Arabic text, a sufficient number of examples for each category, the selection of the feature space, the training data set used and the value of K can enormously affect the accuracy of clustering.

References

- D. Mladenic and M. Grobelnik, "Word Sequenc as Features in
Automatic Arabic Text Clustering using K-means and K-mediods

- Ghosh , Joydeep, 2003. &quot;Scalable Clustering&quot;,The Handbook of Data Mining, Nong Ye(Ed), Lawrence Erlbam Assoc. chapter 10,pp. 247-278.
- Jiaiwei Han and Micheline Kamber, 2006, Data Mining: Concept and Techniques chapter 7, Department of Computer science ,University of Illinois at Urbana-chapaign: www. cs. uiuc. edu/~hanj.

Index Terms
Computer Science Pattern Recognition

Keywords
Arabic Text Clustering Data mining Kmeans Kmediods