Abstract

Nowadays, the news is updated very frequently, especially in the Middle East region where the Arabic language is the primary language of all its countries. The people in this region are interested in following up these updates through the available online news platforms. In order to automate the work in the news agencies, there is an urgent need for an automated system that is able to detect the topic of the news once it has arrived at the agency. In this paper, an efficient system is presented for classifying the online Arabic news into its proper topic. The proposed system uses various natural language processing techniques along with different classification methods. The experimental results show that utilizing the Information Gain, as a feature selection technique, with the Naïve Bayes algorithm, achieves the best accuracy in order to solve the topic detection problem for the online Arabic news.

References

Speech and Natural Language (HLT ’91; Vol. 91. pp. 312–318). Association for Computational Linguistics, Stroudsburg, PA.


17. RapidMiner Studio® Tool for Data Mining Tasks and Implementations: https://rapidminer.com/


Index Terms

Computer Science  Security

Keywords

Machine Learning, Text Mining, Arabic Online News, Topic Detection, Feature Selection.