Abstract

DF-ICF is an algorithm designed by modifying the well known TF-IDF, for the purpose of improving the performance and reliability. The work mainly presents the validation of this new algorithm. The algorithm has been implemented with Hadoop using Cloudera, VMware and WampServer in order to conduct experiments. It also presents the results of an experiment conducted on the algorithm. Finally, the performance of the algorithm is predicted based on assumptions by comparing it with that of the TF-IDF. Overall it was found out that DF-ICF is actually better than TF-IDF.

References

9. Stamatis Karnouskos-“Big data analytics for Smart Grid Cities”. Eurescom mess@ge 1-2013.
10. LiThomas H Davenport, Jill Dyche- “Big Data in Big Companies” International Institute for Analytics, may 2013

Index Terms

Computer Science Algorithms

Keywords

TF-IDF, DF-ICF, Cosine Similarity, Document, Term, Corpus