ONTOLOGY-BASED SIMILARITY CALCULATION METHOD FOR EMAIL CLUSTERING

Dharmadasa K.N.N.[1*] and Kumara B.T.G.S.[1]

[1]Department of Computing and Information Systems, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Sri Lanka

*knndharmadasa@std.appsc.sab.ac.lk

Communication is very important thing around the world. Because of the current situation and development of the technology as well there are lots of methods were used by people. As well as communication methods are increased day by day. Among those communication methods, email is an important and famous method. When it comes to email, basically there are three tabs in the inbox of email account. The primary tab is a one of them. When it comes to the primary tab lots of types of emails are received to this tab. Some of them are academic purpose emails, google forms, security alerts and conference invitations. Considering those types of emails, in this research, researcher has been tried to categorize those emails under few categories. For doing this task researcher has used "clustering" emails according to similarity values of each email contents which has been got by comparing contents of each email messages. In this research the agglomerative hierarchical clustering algorithm is used for clustering email contents because uses of this clustering algorithm is very rare comparing with other algorithms. Although there are few methods for calculating similarity values, co-sign similarity and ontology based similarity methods were used separately for this research. The main objective of the research is developing an algorithm for email clustering using ontology based similarity calculation, an addition to main objective researcher has compared both of similarity calculation methods and accuracy of both similarity value calculation methods are measured as specific objective of the research. When it comes to end of the research, categorizing emails will be very useful for email users for their day to day activities.

Keywords: *Clustering, Cosine similarity, Ontology based similarity, Clustering Algorithm*