Abstract

Entity resolution is the problem of recognizing which entry in database refers to same cluster. in this we have to run the ER in order to reduce the running time and to obtain good results. This paper investigates how we can reduce the running of ER with minimum amount of work using k-means clustering algorithm. In this, clustering can be done according to the matching of entries. We introduce a concept of technique called as k-means clustering to maximize the matching of entries identified using a limited amount of work. We illustrate the potential gains of this entity resolution approach using k-means.

References


Index Terms

Computer Science
Algorithms

Keywords
Data cleaning
Entity resolution-means Clustering Algorithm