Validation of Clustering Algorithms in Weka

Student: Tanasie Virgil - Alexandru (alexnadru_tanasie@yahoo.com)
Coordinator: Cristian Mihăescu, PhD (mihaescu@software.ucv.ro)

Project Goal

The goal of this applicatin is : compute metrics to compare the results of two clustering algorithms, create templates and improve new or existing algorithms by benchmark analysis using new or existing clustering quality metrics.

Short Description

With this tool we first plan to test it on the MTree which is already in Weka and also SKM (Simple K-Means) so that we can start to compare them based on the clustering quality metrics (CQM) and to see which one is better and why.

Our final product will be an extension of the applet already in Weka which will show scores of both clustering algorithms based on some CQM offering a basis for comparison between the two so that the user knows which algorithm is suited for some particular problems.

We hope to extend the applet so that it can compare not only SKM and MTree but other clustering algorithms out there because there is a need to know which algorithm is better on clustering data and why thus improving the efficiency of every clustering algorithms or as many as possible.

Involved Technologies

XML	Is used to store data about students obtained with k-means and for some server configuration files.
JAXB	Is used to work with XML.
JAPPLET	Is used for application client side.
DBCP	Is used to get the connection to the database.
iTEXT	Is used to save students data in a PDF document.
hartDirector	Is used to create the graph with clusters of students.
JavaMail	Is used to send emails to students.
Weka	Is used for applying the K-Means clustering algorithm and Mtree clustering algorithm.