Recommendation system for students based on clustering

Student: Marius Ştefan Chiriţoiu (chiritoiu.mariusstefan@yahoo.com)
Coordinator: S.L. Dr. Ing. Cristian Mihăescu (mihaescu@software.ucv.ro)

Project Goal

This application has two main goals: **efficient and intuitive visualization of the activities performed by students** and **easy interaction with them**.

Short Description

This software tool is mainly designed for the use of professors in order of assist them in a better monitoring of students activity from an online educational environment. The tool presents in an intuitive and graphical way the activities performed by students. After viewing the activity of students, the teacher can interact with sets of students (that are easily selected) by sending an email to them (by which he can give them a recommendation) or he can save students data in a PDF document.

The application server-side has the role of manipulate the data: it connects to a database, read the necessary data about students from which creates a ARFF file, then, from this ARFF file it creates clusters (groups) of students based on students specific attributes ; Finally, clusters of students built in memory is saved in XML files. Also, the server responds to client requests.

On the client side we have a java applet that runs in an Internet browser. It connects to server, takes necessary data and displays the students grouped according to certain features chosen by the professor. The interface of the client application allows specifying data for a new student and viewing its position on the chart (in the cluster to which it belongs).

The tool integrates K-Means clustering algorithm for grouping students and for facilitating the customization of parameters and number of clusters that are displayed. The web server administration interface (*index.html*) allows building a number of clusters and viewing them. Also, Principal Component Analysis (PCA) is used to obtain composed features by combining two simple features : for example, the *MessagingActivity* feature is computed as a combination between the *NumberOfMessages* feature and *AvgNrOfCharacters* feature.

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.
ChartDirector	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 PCA algorithm.