Desktop Application For Data Management Using Open Source Technologies

Student: Popescu Paul Ştefan (popescu_paul_stefan@yahoo.com)
Coordinator: Cristian Mihăescu, PhD., (mihaescu@software.ucv.ro)
Presentation date: July, 2012

Project Goal

This application has two main goals: **data conversion** and **intelligent data analysis**.

For the **data conversion** I developed the capability of transfering database records into XML files in a generic way. For each table from the database will be created one XML file, which contains the corresponding records and all the details necessary for the reverse operation: the restoration of the database using the entries from the XML file. Based on the XML files already created the application is able to create PDF,SVG,HTML and XLS files.

The application also has an **intelligent data analysis** module, realized using WEKA platform and the **J48 classification algorithm**, which is able to classify students depending on their marks, time spent on learning and their presence at courses. The data needed for this module is gathered from the database. Using this data, the system generates the five attributes needed for classifying the students: the users' ID, the number of hours spent to learn, the average mark, the presence at the course and the class they belong to. This information is used for training the algorithm. This module also analyzes the problem of choosing the optimum number of instances needed to train, by generating a chart. There is also generated a chart for visualizing the distribution of the instances and the decision tree obtaind from the training set based on which the classification will be made.

Involved Tools and Technologies

XML	Data representation: records from database are transfered to XML files using a DOM parser, and the database is restored from the XML files using a SAX parser.
Reflection API	Is used to maintain the generic nature of the application.
DBCP	Is used to establish the database connection and make queries.
XSLT	Is used for creating PDF, SVG and HTML files from the existing XML files
JXL	Is used to create the XLS files based on the existing XML files
ANT	Is used for generating a build of the application
Weka	Is used for applying the J48 clasification algorithm on the data from the .arff file.

Project Output

Data Conversion Tool	Converts the records from the database into XML files and restores the database using data from generated XML files. Based on the XML files the HTML,SVG,PDF and XLS files ar created. There is also generated a JTree to view the professors, sections where they teach,courses, chapters,etc.
Data Analysis Tool	Is used to obtain a decision tree for classifying the users depending on the five attributes mentioned above and drawing a chart in order to determine the optimum number of instances.