Análisis de los Datos Históricos de la Programación de Cursos en los CECATI del Estado de Colima

Historical Data Analysis for Scheduling of Cecati's Courses in Colima State

Manuel Espinosa Ortega
Instituto Tecnológico de Colima, México
g7946192@itcolima.edu.mx

Nicandro Farías Mendoza
Instituto Tecnológico de Colima, México
nmendoza@ucol.mx

Jesús Alberto Verduzco Ramírez
Instituto Tecnológico de Colima, México
averduzco@itcolima.edu.mx

Resumen

Hoy en día las herramientas para la extracción de información están mejorando el proceso para que las empresas y dependencias puedan obtener información a partir de grandes volúmenes de datos. Los sistemas de extracción de información se aplican tradicionalmente como una secuencia de módulos de propósito especial, la extracción se convierte, como una clase particular de piezas relevantes de información, que son utilizados por las dependencias o empresas con el fin de tomar decisiones que mejoran la funcionalidad de sus procesos. En este documento se hace una descripción general del Sistema Web para la Programación de Cursos en los CECATI (SWPCC). En particular, nuestra investigación se enfoca a desarrollar un módulo para la extracción de información, a partir del análisis de datos históricos, de la programación de cursos en los CECATI del Estado de Colima, durante el ciclo escolar 2013-2014 mediante herramientas de Data Warehouse y Minería de Datos. El análisis de los datos históricos arroja información sobre los cursos más

programados, los escasamente programados, así como otras áreas de oportunidad y otros aspectos como los horarios y duración de los cursos que pueden influir en la demanda de los cursos que se imparten en los CECATI, lo que nos permite tomar las decisiones para lograr una planeación efectiva de los cursos. La metodología empleada para nuestra investigación, consiste en las siguientes tres fases:

En la primera fase se presenta la contextualización de este trabajo de investigación, describiendo los mecanismos o técnicas de extracción de información, la exposición de las bases de datos históricas, el Sistema Web para la Programación de Cursos en los CECATI (SWPCC) y el módulo de extracción de información.

En una segunda fase presenta en forma detallada cada una de las etapas que se realizaron para obtener un set de datos limpios que se pudieran analizar mediante una herramienta llamada Weka (Waikato Environment for Knowledge Analysis) (Sudhir, Kodge, 2013) a partir de un conjunto de datos de origen de los CECATI en el Estado de Colima del ciclo escolar 2013-2014, la conversión del formato de este archivo original, el procesado de los datos, el análisis de los datos, el trabajo de filtrado, y la discretización de los datos.

En una tercera fase se detallan cada uno de los resultados que se obtienen en el análisis de los datos con el uso de los diferentes algoritmos que posee Weka, procurando una presentación en forma sencilla y clara de estos resultados, de manera que puedan brindar a los interesados nueva información para la toma de decisiones. La metodología de trabajo que se detalla en este documento puede servir de base para futuras investigaciones con otros ciclos escolares, con propósitos de obtener nuevos conocimientos.

Key words: Minería de datos, Data Warehouse, datos históricos, proceso KDD.

Abstract

Today the tools for information extraction are improving the process for companies and dependencies can obtain information from large volumes of data. The information extraction systems are traditionally implemented as a sequence of special-purpose modules, extraction is converted, as a particular class of relevant pieces of information, which they are used by dependencies or companies in order to make decisions that enhance the functionality of its processes. This document provides an overview of the Web System for

Programming of Courses in the CECATI (SWPCC). In particular, our research focuses on developing a module for extracting information, from the analysis of historical data, of programming of courses in the CECATI of state of Colima, during the 2013-2014 school year, using the tools of Data Warehouse and Data Mining. The analysis of historical data yields information about the more scheduled courses, the less programmed, as well as other areas of opportunity and other aspects such as the time and duration of the courses that can influence the demand for courses taught in CECATI, allowing us to make decisions for effective planning of courses. The methodology used for our research, consists of the following three phases:

In the first phase the contextualization of this research is presented, describing the mechanisms and techniques of information extraction, the exposure of historical databases, Web System for Programming of Courses in CECATI (SWPCC) and module of information extraction.

In a second phase presents in detail each of the steps that were performed to obtain a set of clean data that could be analyzed by a tool called Weka (Waikato Environment for Knowledge Analysis) (Sudhir, Kodge, 2013) from a dataset origin of CECATI in the State of Colima during the 2013-2014 school year, the conversion of the original file format, the data processing, the data analysis, the data filtering and discretization the data.

In a third phase detailed each of the results obtained in the analysis of data using different algorithms that has Weka, endeavoring a presentation in simple and clear way of these results, so that can provide to interested parties new information for decision-making. The methodology that is detailed in this document can serve as a basis for future research with other school years, with the purpose of obtaining new knowledge.

Key words: Data Mining, Data Warehouse, historical data, KDD process.

Fecha Recepción: Julio 2015 **Fecha Aceptación:** Enero 2016

Introduction

Many companies do not give it the importance it is due to the use of technology so it is important for the development of the same as it is today requires companies to compete in the electronic market and have systems adequate information their needs (Talo, 2015). Today companies need to have (SI) information systems that enable them to obtain reliable information and to help them decision making (Rojas, 2010), in fact there is the claim that whoever controls the information will dominate the world (Johnson, 2011) and it is clear that currently there are huge amounts of information on that ground to control these huge volumes of information required use of tools to be able to extract, sort, sorting, cleaning, structure, process, etc. . so that this information becomes useful for decision-making. For this need arose mechanisms such as data warehouse and data mining that allow these tasks on behalf of the institutions.

Data Warehouse

After the difficulties of traditional systems to meet informational needs, the Data Warehouse concept emerges as a solution to global informational needs of the company (Boza, 2004 et al). Data Warehouse The term was first coined by Inmon (Pablos, Albarran, Castilla, 1998), translates literally as Data Warehouse. However, if the data warehouse was only a data store, problems would remain the same as in the Information Centers[1]. The main advantage of these systems is based on the fundamental concept, the information structure. This concept means the storage of homogeneous and reliable information in a structure based on consultation and treatment nested in the same, and in a differentiated environment of operational systems. As defined Inmon (2000), the Data Warehouse is characterized by:

- Integrated: data stored in the data warehouse must be integrated into a consistent structure, so that existing inconsistencies between the various operational systems

[1]They are specialized centers, created with the purpose of collecting data, produce information and make it available to all those institutions, universities, unions and business associations, as well as for international cooperation. Rosanna Silva "Information and Documentation Centers (México) www.monografias.com

must be eliminated. The information is often also structured in different levels of detail to suit different user needs.

• Theme: only the data necessary for the process of generating business knowledge are integrated from the operational environment. The data are organized by topic for easy access and understanding by end users. For example, all customer data can be consolidated into a single table of the Data Warehouse. In this way, requests for customer information will be easier to answer because all information resides in the same place.

• History: time is an implicit part of the information contained in a Data Warehouse. In operating systems, data always reflect the state of business activity at present. On the contrary, the information stored in the Data Warehouse serves, among other things, for trend analysis. Therefore, the data warehouse is loaded with different values of a variable in time to allow comparisons.

• Non-volatile: store information in a data warehouse exists to be read, not modified. The information is therefore permanent, meaning the update the data warehouse incorporating the latest values taking the variables contained in it without any action on what already existed.

Data mining

Silberschatz, Abraham, Korth, Henry F. y Sudarshan, S. (2002) define Data Mining (DM) as an automatic or semi-automatic process that seeks to uncover hidden patterns in a dataset and also potentially useful for users of the database (BD). In the MD contemplated various strategies to identify different types of patterns, such as classification trees, neural networks, Bayesian networks, association techniques, among others (Olmos Gonzalez, 2007). The goal in all PI process is to obtain patterns of interest to the end user. To achieve this, it is necessary to properly prepare the data for processing, choose a suitable method to extract the desired patterns and finally determine how to evaluate the patterns found.

Extraction of knowledge is mainly associated with the discovery process known as Knowledge Discovery in Databases (KDD). It is a process that extracts quality information

that can be used to draw conclusions based on relationships or patterns within the data (webmining, 2011). The following figure illustrates the steps of the KDD process (Figure 1):

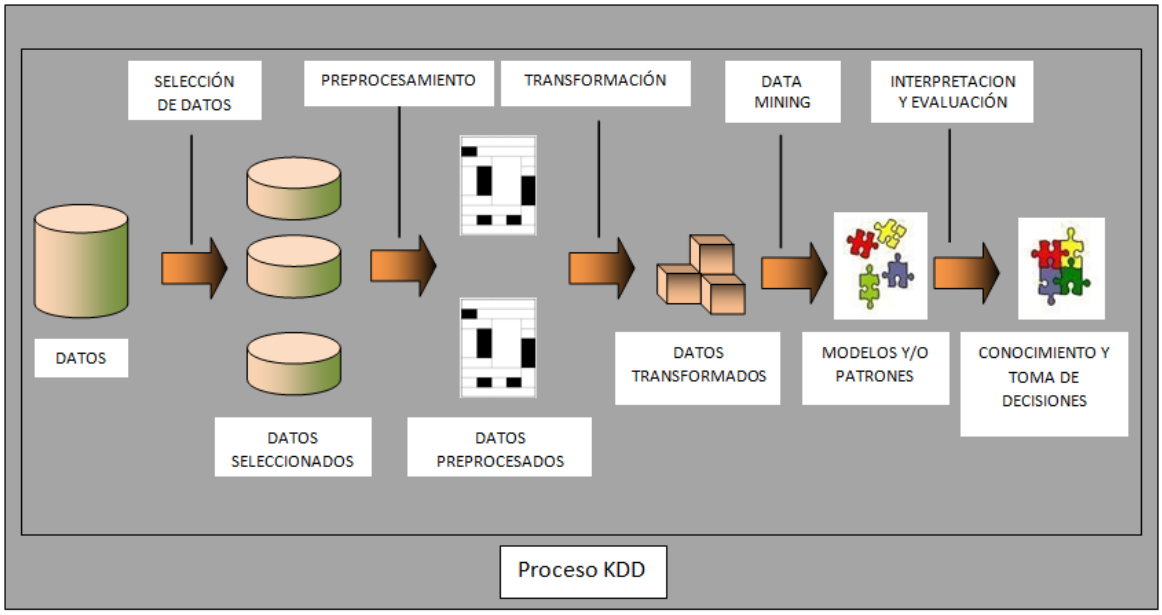

Figure 1. Process Stepso KDD

As shown in the figure above, the steps of the KDD process is divided into five phases are:

Data Selection

At this stage data sources and types of information to be used are determined. It is the stage where the relevant data for analysis are extracted from the data sources or.

I preprocessing

This step is the preparation and cleaning of data extracted from different data sources in a manageable form, necessary to the later stages. At this stage various strategies are used to handle missing data or blank, or inconsistent data that are out of range, resulting in the end a data structure suitable for further processing.

Transformation

It consists of the preliminary processing of data, processing and generation of new variables from existing with appropriate data structure. Here aggregation or normalization operations are performed, consolidating data from a form required for the next phase.

Data Mining

It is the phase of modeling itself, where intelligent methods are applied with the aim of extracting previously unknown, valid, novel, potentially useful and understandable patterns which are contained or "hidden" in the data.

Performance and Evaluation

The patterns obtained are identified and are really interesting, based on some measurements and evaluation of the results is performed.

Historical databases

Training Centres for Industrial Work (CECATI) have a system called "Web System Programming Course in CECATI" (SWPCC) .With this system (Figure 2) is performed automatically scheduling of courses CECATI taught in every school year.

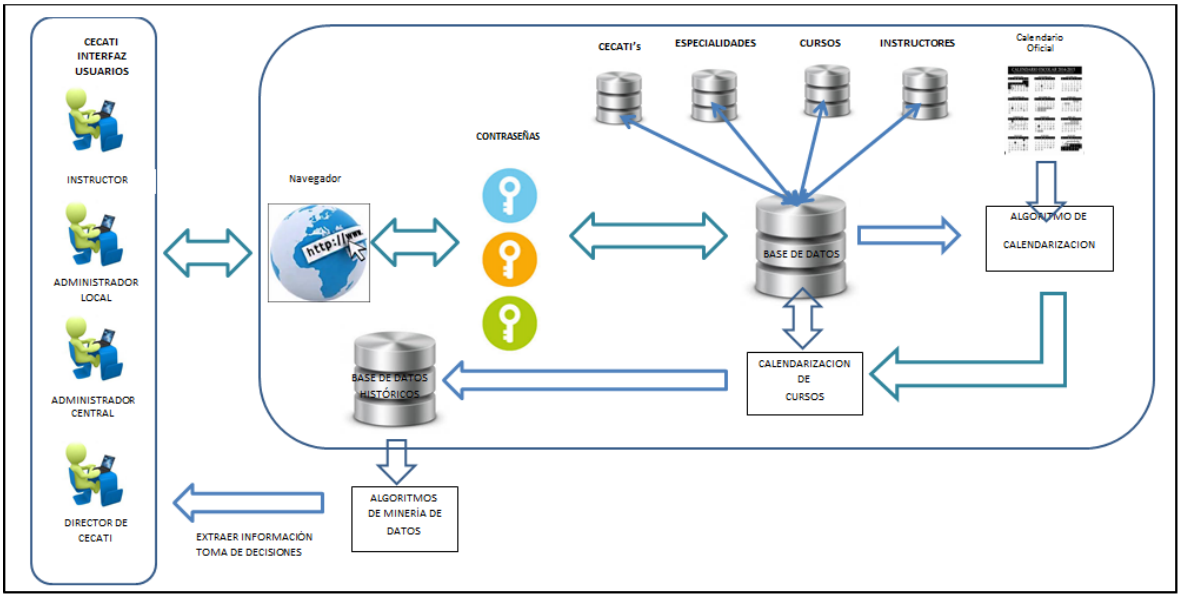

Figure 2. Web System for Programming Courses in CECATI

The system also has an information extraction module comprises the following components: First, the historical database of course plannings. This database is generated with the power of programming courses, which are offered nationwide by year-on-year in each CECATI. Second, a process of applying algorithms to extract information and third, an interface for interpretation and display of the extracted information (Figure 3).

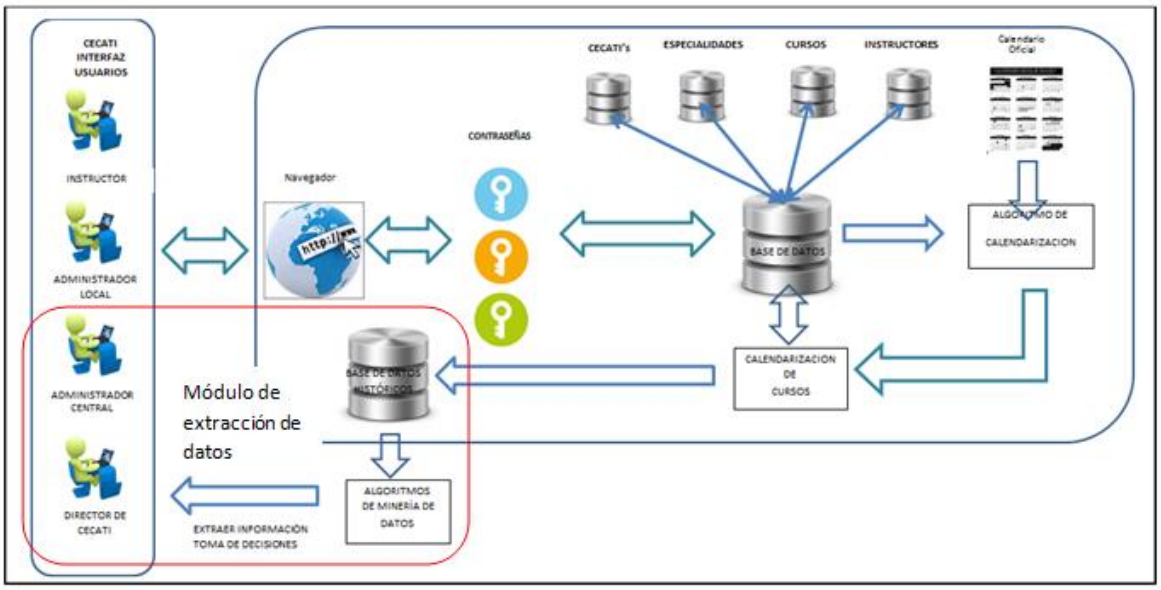

Figure 3. Information extraction module

This historical basis to analyze the information using tools Data Warehouse and Data Mining for purposes of obtaining patterns, trends or statistical projections to help decision-making (Molina, García, 2006). In this research Weka was chosen to be a free license application, very reliable results and because it has most of the algorithms used (classification trees, neural networks, Bayesian networks, association techniques, etc.) to the analysis of data. To perform the analysis of the data mentioned here, they are taken into account only those incurred in the State of Colima CECATI for ease of access to such information.

Getting the Facts

The historical database of SWPCC system is the source for obtaining the source data to analyze by Weka. The format "xls" is for files that can be opened with Excel, the source data are ".sql" format, you must import the Excel program to change the original format ".sql" format ".xls" . The reason for this conversion is that Excel makes cleaning the data, leaving only the data that are useful for the analysis to be performed usandoWeka.

Cleaning the Data

Through Excel tool replacement name fields for all attributes that have data written to them differently unified. In addition to unifying the data content, attributes that are not

necessary for the analysis of the data set, for example eliminated: the attribute "address", "colony", "postal_code", "telephone", "NOMBRE_JEFE_AREA" its contents are unnecessary for the information sought to be obtained; the "specialty" attribute is removed because you can guess from the name of the scheduled courses; also "non-working days" also removed because its content is not relevant.

Data discretization

The set of data to be analyzed contains some fields that are numeric, eg "DURACION_EN_HORAS"; for the analysis process with Weka are required to be of a symbolic nature is therefore necessary discretization of this type of data, before this action should first convert the clean data set to the type of CSV comma delimited format, then copy the attributes and data file type .arff (Diego Garcia) the following structure (Figure 4):

```
% comentarios

@relation NOMBRE_RELACION

@attribute r1 real
@attribute r2 real ...
...
@attribute i1 integer
@attribute i2 integer
…
@attribute s1 {v1_s1, v2_s1,…vn_s1}
@attribute s2 {v1_s1, v2_s1,…vn_s1}
…
@data
```

Figure 4. Structure Weka for data analysis

In the top comments relevant preceded by the% sign is placed. Then the relation name preceded by "@ relationship" is written. Subsequently they listed in column form each of the attributes preceded by "attribute" followed by the data type: real, integer, or string, as appropriate. This structure is recognized by WEKA and must be scrupulous in the syntax of each concept as any change in itself does not allow you to open the file to analyze. Underdata go all comma delimited data, as seen in the following example (Figure 5):

```
%CECATI_No,NOMBRE_CURSO,TIPO_CURSO,DURACION,HORARIO_CURSO


@relationprogramacion_cursos_cecati_colima

@attribute CECATI_No {CECATI_34,CECATI_126,CECATI_145,CECATI_183}
@attribute NOMBRE_CURSO String
@attribute TIPO_CURSO {REGULAR,CAE,EXTENSION,ACCION_MOVIL}
@attribute DURACION Real
@attribute HORARIO_CURSO String
@data


CECATI_34,"INSTALACION DEL SISTEMA ELECTRICO RESIDENCIAL",REGULAR,240,"07:30-10:46"
CECATI_34,"MANTENIMIENTO DE APARATOS DOMESTICOS",REGULAR,200,"10:46-13:18"
CECATI_34,"MANTENIMIENTO DE APARATOS DOMESTICOS",REGULAR,200,"16:00-19:00"
CECATI_34,"INSTALACION DEL SISTEMA ELECTRICO INDUSTRIAL",REGULAR,240,"18:24-21:00"
CECATI_34,"INSTALACION DEL SISTEMA ELECTRICO RESIDENCIAL",REGULAR,240,"19:00-22:00"
CECATI_34,"BOBINADO DE MOTORES ELECTRICOS",CAE,57,"10:00-13:00"
CECATI_34,"INSTALACION DEL SISTEMA ELECTRICO RESIDENCIAL",REGULAR,240,"16:00-19:00"
CECATI_34,"INSTALACION DEL SISTEMA ELECTRICO INDUSTRIAL",REGULAR,240,"19:00-22:00"
CECATI_34,"MANTENIMIENTO DE SISTEMA DE A/A Y REFRIGERACION",ACCION_MOVIL,200,"10:00-16:53"
CECATI_34,"REPARACION DE REFRIGERADORES DOMESTICOS SIN ESCARCHA",EXTENSION,112,"07:00-10:00"
CECATI_126,"MANTENIMIENTO DE EQUIPOS RECEPTORES DE TELEVISION",REGULAR,360,"08:00-14:00"
CECATI_126,"REPARACIONES BASICAS DE UN AUTOESTEREO",EXTENSION,168,"08:00-14:00"
CECATI_126,"REPARACION DE MOTORES A GASOLINA",REGULAR,450,"14:00-20:00"
CECATI_126,"REPARACION DEL SISTEMA DE FRENOS BASICOS",REGULAR,280,"14:00-17:00"
CECATI_126,"REPARACION DEL SISTEMA DE FRENOS BASICOS",REGULAR,280,"17:00-20:00"
CECATI_126,"REPARACION DEL SISTEMA DE TRANSMISION MANUAL",REGULAR,234,"14:00-20:00"
CECATI_126,"CONFECCION DE PRENDAS PARA DAMA Y NIÑA",REGULAR,350,"08:00-11:00"
CECATI_126,"CONFECCION DE PRENDAS PARA DAMA Y NIÑA",REGULAR,350,"11:00-14:00"
CECATI_126,"CONFECCION DE PRENDAS PARA CABALLERO Y NIÑO",REGULAR,273,"08:00-11:00"
CECATI_126,"CONFECCION DE PRENDAS PARA CABALLERO Y NIÑO",REGULAR,273,"11:00-14:00"
CECATI_126,"ALTA COSTURA",REGULAR,350,"16:30-19:30"
CECATI_126,"ELABORACION DE BLANCOS",EXTENSION,150,"16:30-19:30"
CECATI_126,"CONFECCION DE PRENDAS PARA CABALLERO Y NIÑO",REGULAR,300,"08:00-11:00"
CECATI_126,"CONFECCION DE PRENDAS PARA CABALLERO Y NIÑO",REGULAR,300,"11:00-14:00"
CECATI_126,"ALTA COSTURA",REGULAR,350,"08:00-12:00"
CECATI_126,"DECORACION DE PRENDAS DE VESTIR",EXTENSION,176,"12:00-14:00"
CECATI_126,"DECORACION DE PRENDAS DE VESTIR",EXTENSION,120,"08:00-14:00"
CECATI_126,"INGLES COMUNICATIVO BASICO INICIAL",ACCION_MOVIL,180,"08:30-14:30"
CECATI_126,"INGLES COMUNICATIVO BASICO INICIAL",ACCION_MOVIL,180,"08:30-14:30"
CECATI_126,"VERBOS REGULARES E IRREGULARES",EXTENSION,78,"08:30-14:30"
CECATI_126,"VERBOS REGULARES E IRREGULARES",EXTENSION,66,"08:30-14:30"
CECATI_145,"MANTENIMIENTO DE AIRE ACONDICIONADO MINISPLIT",EXTENSION,111,"15:00-18:00"
CECATI_145,"MANTENIMIENTO DE AIRE ACONDICIONADO MINISPLIT",EXTENSION,111,"18:00-21:00"
```

```
CECATI_145,"INSTALACION DEL SISTEMA ELECTRICO RESIDENCIAL",REGULAR,240,"07:00-10:00"
CECATI_145,"INSTALACION DEL SISTEMA ELECTRICO RESIDENCIAL",REGULAR,240,"10:00-13:00"
CECATI_145,"INSTALACION DEL SISTEMA ELECTRICO RESIDENCIAL",REGULAR,240,"07:00-10:00"
CECATI_145,"INSTALACION DEL SISTEMA ELECTRICO RESIDENCIAL",REGULAR,240,"10:00-13:00"
CECATI_145,"INSTALACION Y REPARACION DE SISTEMAS DE COMUNICACIÓN",REGULAR,192,"07:00-11:00"
CECATI_183,"PREPARACION DE ALIMENTOS",REGULAR,360,"08:00-11:00"
CECATI_183,"PREPARACION DE BEBIDAS",REGULAR,200,"11:00-14:00"
CECATI_183,"SERVICIO A COMENSALES",REGULAR,216,"08:00-11:00"
CECATI_183,"ELABORACION DE PASTELES Y PRODUCTOS DE REPOSTERIA",REGULAR,350,"11:00-14:00"
CECATI_183,"BOCADILLOS Y PANADERIA CASERA",EXTENSION,60,"11:00-14:00"
CECATI_183,"PREPARACION DE ALIMENTOS",REGULAR,360,"16:00-20:00"
CECATI_183,"ELABORACION DE PASTELES Y PRODUCTOS DE REPOSTERIA",REGULAR,350,"16:00-20:00"
CECATI_183,"GELATINAS ARTISTICAS",EXTENSION,92,"16:00-20:00"
CECATI_183,"CORTE Y PEINADO DEL CABELLO",REGULAR,200,"08:00-11:00"
CECATI_183,"CORTE Y PEINADO DEL CABELLO",REGULAR,200,"11:00-14:00"
CECATI_183,"COLOR Y TRANSFORMACION EN EL CABELLO",REGULAR,220,"08:00-11:00"
CECATI_183,"COLOR Y TRANSFORMACION EN EL CABELLO",REGULAR,220,"11:00-14:00"
…
```

Figure 5. Structure analysis of historical data CECATI

Caring this structure and the correct syntax, the program Weka (which is very scrupulous) receives the information without restriction allowing use all the functions available to analyze the data set.

Data analysis

The data set Programming Course in CECATI in the State of Colima in this document will be analyzed by the Weka tool it contains the following attributes: CECATI_No, NOMBRE_CURSO, TIPO_CURSO, duration, HORARIO_CURSO, with which it is sought relevant information on the courses scheduled in CECATI in the State of Colima during the 2013-2014 school year. For analysis of the data set some guidelines are followed: The attributes whose content is a list of options and also quite large requiring containing more than one word, they should be of type String as the case of "NOMBRE_CURSO" shall be removed other attributes and leave only the attribute to be analyzed, then select the StringToNominal Weka can filter to display the list of instances that contains the "NOMBRE_CURSO" attribute. Such is the case of the attribute "HORARIO_CURSO" which is of type String, so for analysis is equally applicable. Then

the figures of each of these attributes and the results obtained by applying the filter shown StringToNominal:

The resulting figure (Figure 6) shows a total of 111 different courses scheduled during the 2013-2014 school year in CECATI in the State of Colima, of which specialty beauty services courses Cutting and styling hair, and COLOR MAKEUP FACE AND TRANSFORMATION IN THE HAIR, are the that is scheduled more often in relation to other courses of other specialties. After this field, specializing in refrigeration and A / C with system maintenance A / C AND REFRIGERATION course, shows second frequency program this course, the specialty of Electricity is in third place with the course called INSTALLATION RESIDENTIAL ELECTRICAL. Among the regular courses they were programmed only once in the entire state of Colima in these training centers include: ELECTRIC MOTORS MAINTENANCE, REPAIR SYSTEM CLUTCH CONTROL SYSTEM REPAIR emission pollutants, Brake Repair ABS, MAINTAINING A LOCAL AREA NETWORKS (LAN), beverage preparation and service to guests.

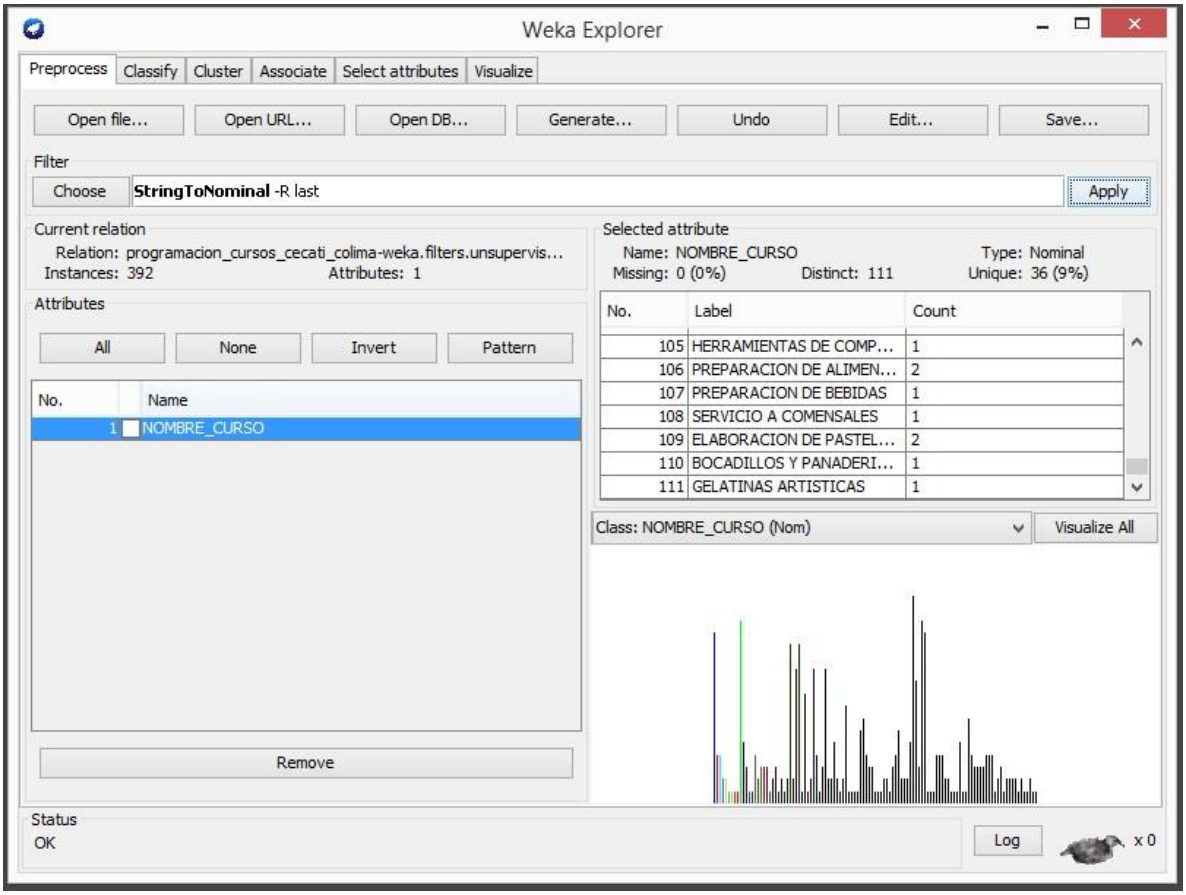

Figure 6. Análisis del atributo "NOMBRE_CURSO" con el filtro StringToNominal

Following the order of the attributes of type String we have the case of "HORARIO_CURSO" which Figure 7, we see that the following is obtained: The five most used for programming courses in order from highest to lowest frequency schedules they are: 07: 00-10: 00 with a number of frequency of eighteen, 11: 00-14: 00 with a number of frequency seventeen, 18: 00-20: 00 with a number of every ten six, 07: 00-09: 00 and 08: 00-11: 00 with a frequency number fifteen, finally times 15: 00-18: 00, 18: 00-21: 00 and 11:00 -13: 00 had a rate of thirteen. The times were programmed only once in the entire state of Colima are: 13:18-15:33, 18:00-22:00, 09:00-11:30, 15:10-18:30, 19:00-21:00, 13:00-15:24 y 15:00-20:00.

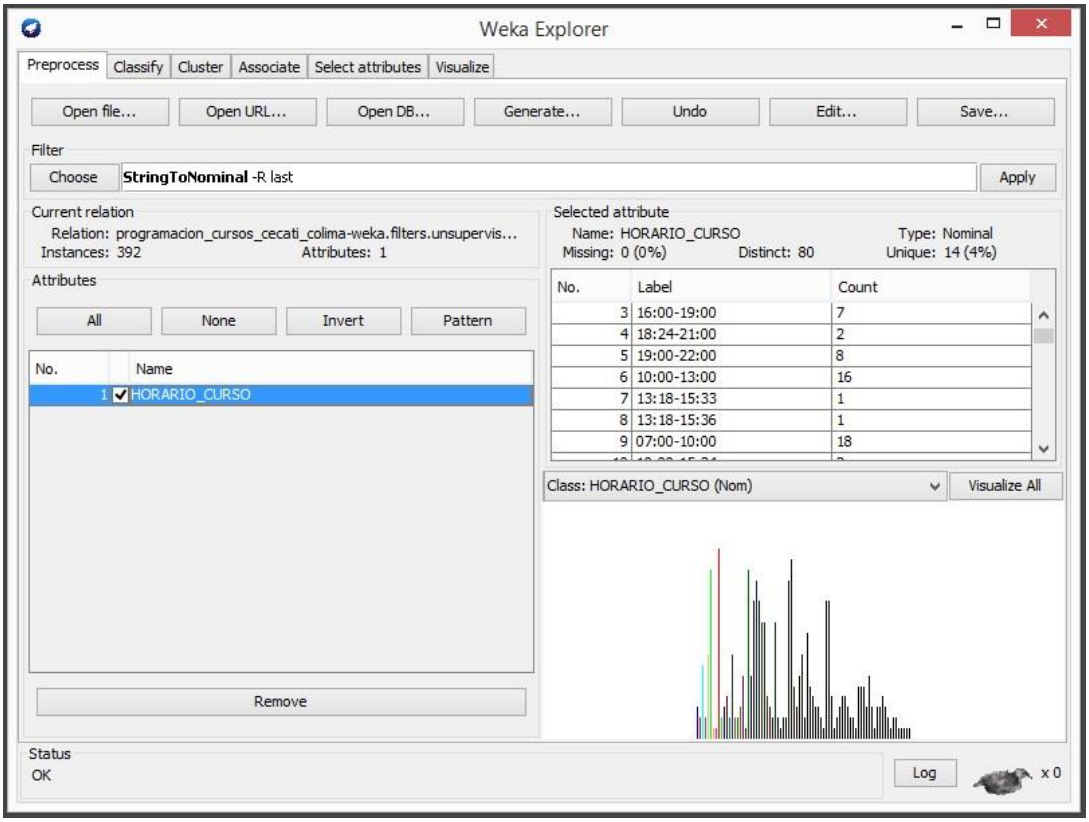

Figure 7. Análisis del atributo "HORARIO_CURSO" con el filtro StringToNominal

The attribute "CECATI No" to have few instances, their analysis is easier and does not require selecting a filter or remove other attributes, just make sure the box is checked to "CECATI_No." And shows the list of instances corresponding, as shown in Figure 8. This Figure 8 shows that the 145 was the CECATI more courses scheduled during the school year being analyzed a total of 120 courses representing a 30.61% of the total scheduled on throughout the State of Colima in that school year; secondly the CECATI 34 scheduled a total of 118 courses this school year, which represented 30.10% of all scheduled courses throughout the State of Colima; once the scheduled 97 courses CECATI 183 representing 24.74% and finally the CECATI 126 with a total of 57 courses and 14.54% of the total. It should be mentioned that these figures do not correspond to the courses described in Figure 6, because that course different names are mentioned, whereas here the names of the courses may be repeated.

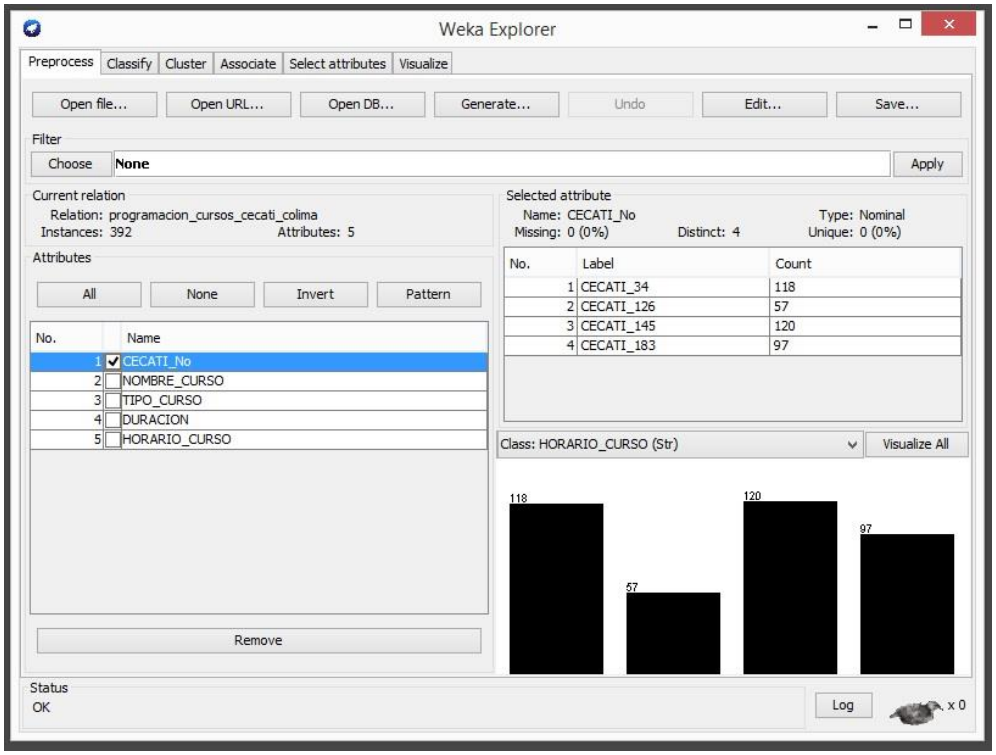

Figure 8. Attribute analysis "CECATI_No."

he attribute analysis "TIPO_CURSO" is similar to the above, also has few instances, as shown in Figure 9 has four instances, namely courses

"REGULAR" for a total of 272 scheduled courses of this type, accounting for 69.39% of all scheduled courses in the State of Colima during the 2013-2014 school year; courses "STRETCH" for a total of 97 courses representing 24.74% of the total; courses "CAE" with a total of 19 scheduled courses represent a 4.85% relative to the total; finally scheduled courses as "MOBILE ACTION" making a total of four courses, representing 1.02% of courses scheduled in the state of Colima.

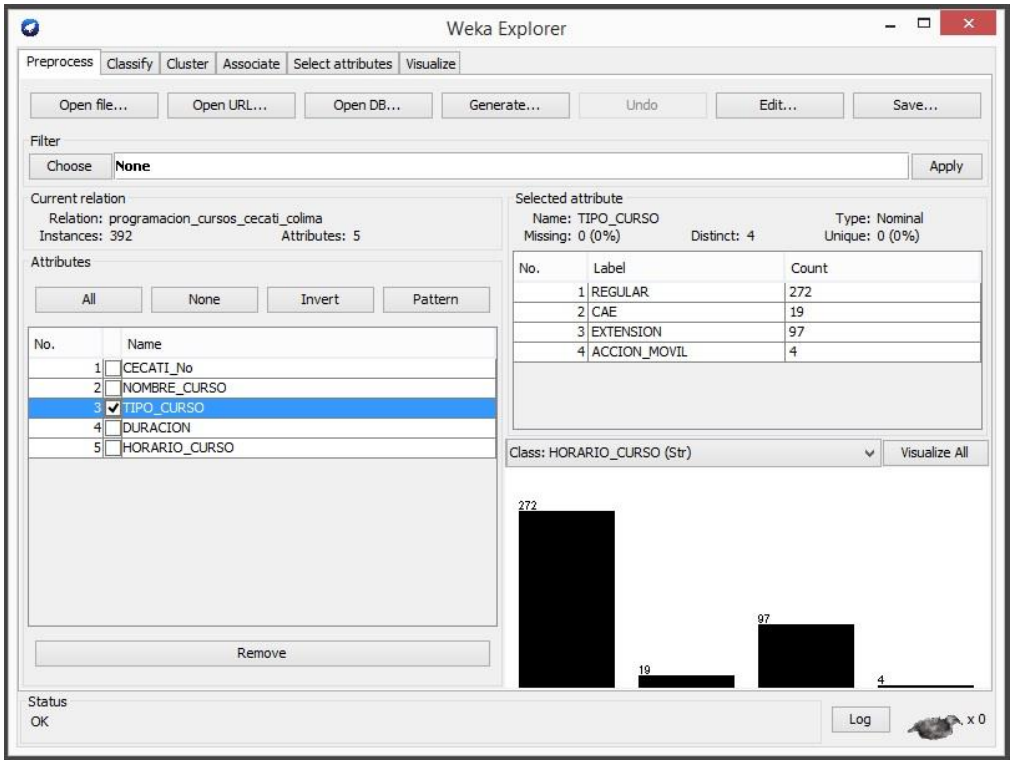

Figura 9. Análisis del atributo "TIPO_CURSO"

In the case of "duration" attribute (Figure 10) is convenient to use the filter "discretize" to prevent instances are a huge list that can not be analyzed, therefore the amounts shown in ranges and with longer courses in hours are firstly those found in the range of 100 to 144 years, secondly the range 56-100 hours, and thirdly the range from 187 to 231 hours, the fourth range 144 187 hours are fifth courses lasting less than 57 hours, the last place the range of 362-406 hours.

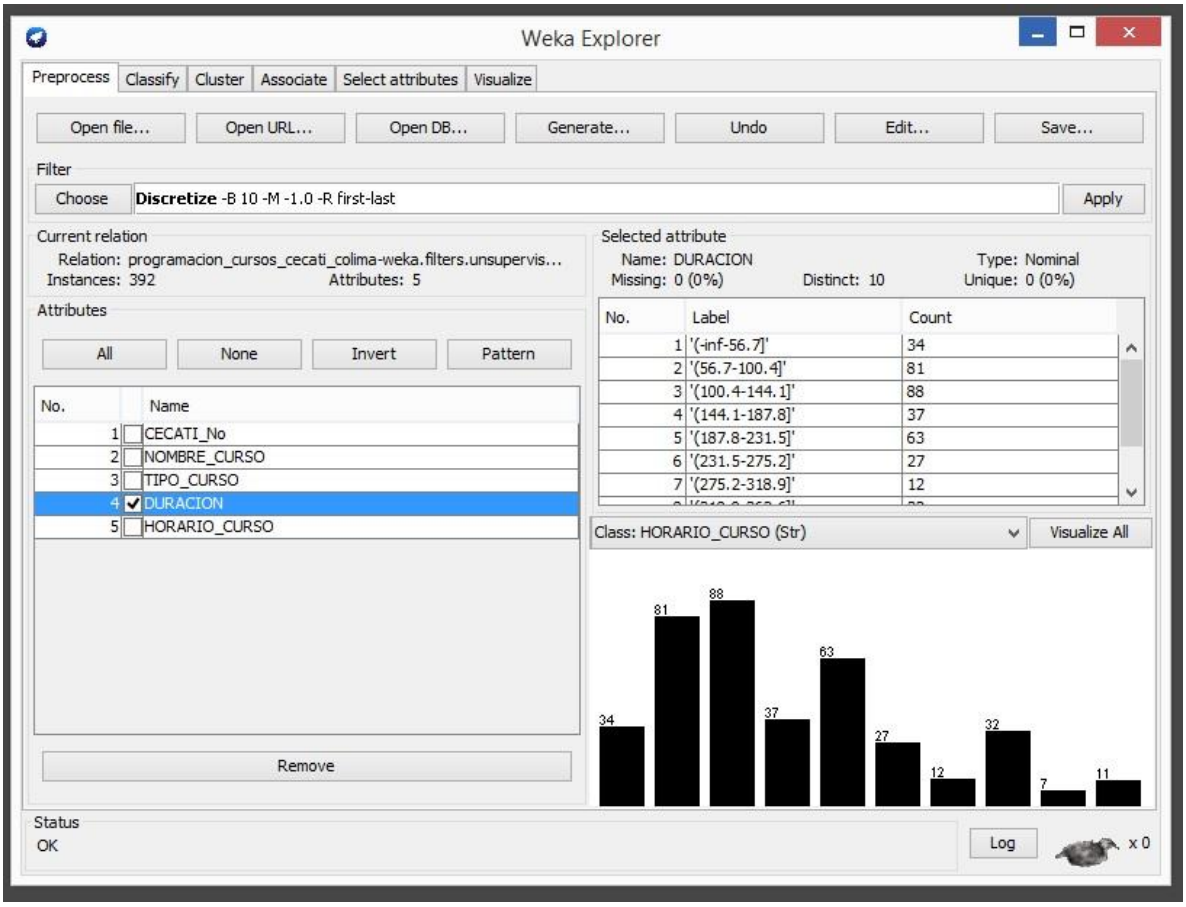

Figure 10. Attribute analysis "DURACION"

Association attributes

To search for some kind of association between the attributes apply the Associate Weka tab and select the filter A priori and we click on Start, if there is any association between the instances corresponding to the five attributes, the list of instances that have some association unfolds in this case to be independent each attribute not exhibit any association as shown in Figure 11.

Figure 11. Attributes Interface Association

Results

The results presented in the section on data analysis represent reliable data are available for interpretation and decision-making by stakeholders. These data have an objective basis and scientific foundation and are shown as algorithms of Weka apply presented without bias the results, nor maquillarlos to cause trends, stakeholders will be responsible for interpreting them properly for a decision correct.

Recommendations

We recommend using the methodology used in this paper for further analysis for other statewide school years historical data. We also recommend expanding the historical database for analysis of historical data in a national context taking into account that the CECATI operating throughout the Mexican Republic.

Bibliography

Dr. Sudhir B. Jagtap y Dr. Kodge B.G. “Census Data Mining and Data Analysis using Weka” (ICETSTM) International Conference in “Emerging Trends in Science, Tecnology and Management” 2013, Singapore.Recuperado de: http://arxiv.org/ftp/arxiv/papers/1310/1310.4647.pdf

Gonzalo Talo 25 de ene. De 2015. Copy of Tecnología y empresa. Recuperado de: https://prezi.com/6uhj2zx1izl3/copy-of-tecnologia-y-empresa/

Gonzalez Rojas, H.D.: "Importancia de la tecnología en las empresas" en Contribuciones a la Economía, febrero 2010. Recuperado de http://www.eumed.net/ce/2010a/

Dick Johnson 2011 “El Control de la Información: 6 empresas para gobernar a todos”. Recuperado de: http://elespiritudeltiempo.org/blog/el-control-de-la-informacion-6-empresas-para-gobernarlos-a-todos/

D. Andrés Boza García, Dr. Angel Ortiz Bas, Dr. Eduardo VicénsSalort, Dña. Llanos Cuenca Gonzalez, “Data Warehouse para la gestión por procesos en el sistema productivo”. Second World Conference on POM and 15tn Annual POM Conference, Cancun, Mexico, April 30 – May 3, 2004. Recuperado de: http://www.pomsmeetings.org/ConfProceedings/002/POMS_CD/Browse%20This%20CD/PAPERS/002-0278.pdf

Carmen de Pablos Heredero, Irene Albarrán Lozano, Guillermo Castilla Alcalá, “El Proceso de Implantación del Data Warehouse en la Organización: Análisis de un caso”. Investigaciones Europeas de Dirección y Economía de la Empresa Vol. 4, 03,1998, pp. 73-92. Recuperado de: http://www.aedem-irtual.com/articulos/iedee/v04/043073.pdf

W.H. Inmon “Building the Data Warehouse: Getting Started”. Recuperado de: http://www.academia.edu/3081161/Building_the_data_warehouse

Silberschatz, Abraham, Korth, Henry F. y Sudarshan, S. “Fundamentos de Bases de Datos”. McGrawHill, 4a Ed., 2002.

I.Olmos Pineda y J.A. González Bernal “Casos de éxito de Minería de Datos”. Recuperado de: http://es.scribd.com/doc/93421745/Caso-de-Exito-Mineria-de-Datos#scribd

Webmining Consultores “KDD: Proceso de Extracción de conocimiento”. 10 de Enero de 2011 • En la Categoría Business Intelligence & Analytics, Data Mining. Recuperado de: http://www.webmining.cl/2011/01/proceso-de-extraccion-de-conocimiento/

José Manuel Molina López y Jesús García Herrera “Técnicas de Analisis de Datos Aplicaciones Prácticas utilizando Microsoft Excel y Weka” 2006. Recuperado de: http://www.giaa.inf.uc3m.es/docencia/II/ADatos/apuntesAD.pdf

Diego García Morate “Manual de Weka” licencia CreativeCommons Reconocimiento-NoComercial-SinObraDerivada2.0 .Recuperado de: http://www.metamotion.com/diego.garcia.morate/download/weka.pdf