

## 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](mailto:g7946192@itcolima.edu.mx)

**Nicandro Farías Mendoza**

Instituto Tecnológico de Colima, México

[nmendoza@ucol.mx](mailto:nmendoza@ucol.mx)

**Jesús Alberto Verduzco Ramírez**

Instituto Tecnológico de Colima, México

[averduzco@itcolima.edu.mx](mailto:averduzco@itcolima.edu.mx)

Número 12. Enero – Junio 2016

### 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.

**Palabras Clave:** 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.

## Resumo

Hoje, as ferramentas para a extração de informações estão a melhorar o processo para as empresas e agências podem obter informações a partir de grandes volumes de dados. Sistemas de extração de informações é tradicionalmente aplicado como uma seqüência de módulos de extração para fins especiais torna-se, como uma classe particular de peças de informação relevantes que são usados por agências ou empresas a tomar decisões que melhoram a processos funcionalidade. Este documento fornece uma visão geral do Sistema Web para cursos de programação em CECATI (SWPCC) é feita. Em particular, a nossa investigação centra-se no desenvolvimento de um módulo para extração de informações a partir da análise de dados históricos, agendar cursos CECATI no Estado de Colima, durante o ano escolar 2013-2014 por ferramentas de Data Warehouse e Data Mining. A análise dos dados históricos produz informações sobre os cursos regulares, mal planejado, e outras áreas de oportunidade e de outros aspectos como o tempo ea duração dos cursos que podem influenciar a demanda por cursos ministrados em CECATI , o que nos permite tomar decisões para cursos de planejamento eficazes. A metodologia utilizada para a nossa investigação, consiste nas seguintes três fases:

Na primeira fase, a contextualização desta pesquisa é apresentada, descrevendo os mecanismos e técnicas de extração de informações, a exposição dos bancos de dados históricos, System Web para cursos de programação em CECATI (SWPCC) e módulo extracção de informação.

Em uma segunda fase apresenta em detalhe cada uma das etapas que foram executadas para obter um conjunto de dados limpo que poderiam ser analisados por uma ferramenta chamada Weka (Waikato Ambiente para a análise do conhecimento) (Sudhir, Kodge, 2013) a partir de um origem conjunto de dados de CECATI no Estado de Colima o ano letivo

2013-2014, a conversão do formato de arquivo original, processamento de dados, análise de dados, filtragem de trabalho e discretização os dados.

Numa terceira fase detalhada cada um dos resultados obtidos na análise de dados utilizando diferentes algoritmos que tem Weka, buscando uma apresentação de forma simples e clara a estes resultados, de modo que eles podem fornecer o novas informações interessado para a tomada de decisão. A metodologia que é detalhado neste documento pode servir como uma base para pesquisas futuras com outros ciclos escolares, com o objectivo de obter novos conhecimentos.

**Palavras-chave:** Mineração de Dados, Data Warehouse, dados históricos, o processo de KDD.

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

---