

Minería de Datos UTSOE

AGUIRRE-PUENTE, José*†, HUERTA-MASCOTTE, Eduardo, RODRIGUEZ-VARGAS, María y RAYA-PÉREZ, Jesús

Universidad Tecnológica del Suroeste de Guanajuato

Recibido Noviembre 18, 2015; Aceptado Enero 15, 2016

Resumen

El proyecto de minería de datos UTSOE se propuso para llevar a cabo el análisis de la información que se desplaza entre los diferentes departamentos internos de la universidad. En el presente trabajo se da a conocer el análisis de la información que se genera en las diferentes direcciones de carrera de la universidad para posteriormente conocer los procesos de información para la realización de la documentación que generan las mismas y las relaciones entre ellas y posteriormente el flujo de la información, de dónde se genera y su desplazamiento, así como el cliente final de dicho proceso-información. En la implementación del proyecto se realizaron prototipos para la extracción de la información y pruebas piloto, con ayuda de diagramas de proceso y actividades se plasma la técnica a seguir del usuario para realizar un documento y conocer su actividad durante el proceso. A través del análisis de los procesos y estudios de la información se desarrolla un concentrado de información y organización de la misma, para facilitar la toma de decisiones, la mejora continua y posible sistematización de los procesos minimizando errores humanos. En conjunto, la experiencia produce a una primera propuesta del modelo conceptual de la minería de datos pretendida. Esta determina la etapa de proyección para los siguientes momentos de diseño, planeación y modelado informático.

Minería, minería de datos, minería textual, interpretación, modelado

Abstract

The UTSOE data mining project proposed perform the analysis of information that is shared among the different internal departments of the university. The present paper discloses the analysis of the information generated in the different college career directions, to learn about the process of the documentation generated in this area and the relationships between them, the flow of information, from where it is generated to the end-customer of that process. In the implementation of the project, prototypes and pilot tests were created to extract information, with the help of process and activity diagrams to illustrate the technique and process that the user followed to generate a document as well as a guideline during the process. Through the analysis of processes and studies of the information it gets concentrated and organized to make easier decisions, continuous improvement and possible systematization of processes as well as minimizing human errors. In conjunction, the experience produces a first proposal of conceptual modeling of data mining that was expected. The projection phase is determined for future designs, plantation and informatic modeling.

Mining, data mining, text mining, interpretation, modeling

Citación: AGUIRRE-PUENTE, José, HUERTA-MASCOTTE, Eduardo, RODRIGUEZ-VARGAS, María y RAYA-PÉREZ, Jesús. Minería de Datos UTSOE. Revista de Análisis Cuantitativo y Estadístico. 2016. 3-6: 50-56

*Correspondencia al Autor (Correo Electrónico: jaaguirre@utsoe.edu.mx)

† Investigador contribuyendo como primer autor.

Introducción

En la actualidad, la minería es requerida en diversos contextos donde no se tiene un buen control del flujo de información como en el ámbito financiero, bancario, en el análisis de mercados y comercios, en el área de la salud tanto pública como privada, a nivel educativo, en procesos industriales, telecomunicaciones y diferentes áreas de trabajo.

La minería de datos es un proceso que consiste en encontrar información importante que no se encuentra plasmada directamente, pero que con base a ella se pueden obtener datos relevantes que sean de utilidad.

En la Universidad Tecnológica del Suroeste de Guanajuato (que en lo sucesivo se denominará UTSOE) se maneja un gran contenido de información que se dispersa a diferentes áreas y departamentos de la institución, en esta ocasión se hará énfasis en la que se genera en las direcciones de carrera y con base a minería de datos y minería textual, se pretende agilizar los procesos y disminuir el esfuerzo innecesario al igual que el error humano.

La UTSOE, está comprometida con la calidad académica de la oferta educativa mediante la evaluación externa de sus programas educativos y la acreditación por organismos. La Universidad trabaja con un Sistema de Gestión de Calidad y Ambiental (SGCA) que ayuda a la mejora continua de los procesos estratégicos de gestión de la institución, abarcando procesos educativos, educación continua, servicios tecnológicos y estadía así como al cuidado del medio ambiente cumpliendo con los requisitos de Normas ISO. El SGCA se tomó como base para el desarrollo del proyecto, para conocer quiénes son los dueños de los procesos y sus clientes.

Al concluir el proyecto se pretende tener como producto final una propuesta de procesos sistematizados de direcciones de carrera para su posterior utilización en la institución, integrando los resultados posteriores de los diferentes productos entregables de los siguientes proyectos de la minería de datos de los distintos departamentos de UTSOE para tener procesos de manera más eficientes.

Antecedentes

Desde los inicios en la UTSOE, se realizan procesos de manera intuitiva, manejando redundancia de información, incongruencia de datos y fechas, lo cual conlleva un posible error humano. Los empleados realizan varias actividades innecesarias para llevar a cabo un proceso de documentación académica, ya sea por falta de información, tiempo de espera para la validación de un proceso o por documentos que necesiten los mismos datos, pero reflejados de diferente manera o diferentes formatos, al igual que almacenamiento irrelevante (información histórica de carácter informativo con bajo impacto).

En la universidad se trabaja con el Sistema Institucional UTSOE (SINUT), que es un sistema que es usado para centralizar información referente a los educandos y sirve de base para realizar algunos de los procesos de las distintas área administrativas y directivas de la universidad, que también a su vez brinda servicios a los estudiantes tales como descargar boleta de calificaciones, documentos personales de los alumnos en formato digital (acta de nacimiento, CURP, certificado de preparatoria), calendario escolar, entre otras. Si bien se cuenta con el SINUT, tiene carencias en algunos módulos de aplicaciones para el acceso y filtración a la información de los usuarios, como son procesos de estadías, control de asesorías y tutorías, adeudos.

Marco Teórico

Metodología SCRUM.

Es un marco de trabajo para la gestión y desarrollo de software basada en un proceso iterativo e incremental utilizado comúnmente en entornos basados en el desarrollo ágil de software.

Un proceso en el que se aplican de manera regular un conjunto de buenas prácticas para trabajar colaborativamente en equipo, y obtener el mejor resultado posible de un proyecto.



Figura 1 Metodología SCRUM

Cada día de un sprint, se realiza la reunión sobre el estado de un proyecto, a esto se le llama daily standup. (Castro, 2015).

Minería de datos.

La minería de datos es un proceso de detectar la información procesable de grandes conjuntos de datos. Utiliza el análisis matemático para encontrar los patrones y tendencias que existen en los datos. En algunas ocasiones, estos patrones no se pueden detectar mediante la exploración tradicional de los datos, ya que las relaciones son demasiado complejas o porque hay demasiados datos.

El fin principal es encontrar patrones, perfiles y tendencias mediante el análisis de datos usando tecnologías que permitan reconocer patrones, redes neuronales, lógica difusa, algoritmos genéticos y muchas otras técnicas avanzadas del análisis de datos.

Procesos básicos de minería de datos:

1. Definir el problema.
2. Preparar los datos.
3. Explorar los datos.
4. Generar modelos.
5. Explorar y validar los modelos.
6. Implementar y actualizar los modelos.

PAD (Diagrama de arquitectura de procesos).

Los PAD están compuestos principalmente por un proceso. Un proceso puede ser definido como un conjunto de actividades enlazadas entre sí que, partiendo de uno o más inputs (entradas) los transforma, generando un output (resultado).

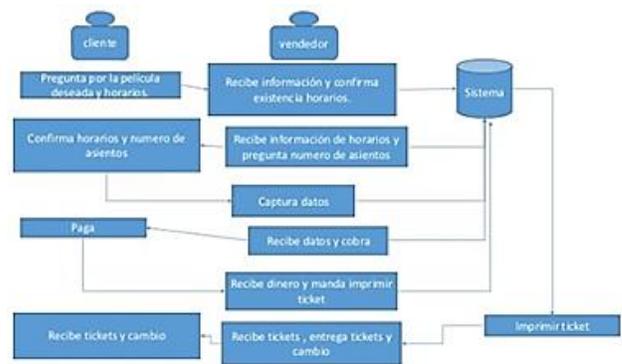


Figura 2 Diagrama PAD

Desarrollo

Implementación de SCRUM en el proyecto.

- Investigación Preliminar.
- Daily Standup.
- Levantamiento de Requerimientos.

- Daily Standup.
- Obtención de Información y mapeo.
- Daily Standup.
- Diseño de la Propuesta.
- Daily Standup.
- Resultados.
- Daily Standup.

Al finalizar cada etapa del proyecto, se realizaban reuniones periódicamente con el equipo de trabajo, obtener el mejor resultado posible de un proyecto. Durante el desarrollo del proyecto “Minería UTSOE” en el cual se llevaron a cabo las siguientes etapas con sus respectivas actividades las cuales son:

Investigación Preliminar.

Al iniciar el proyecto se realizó una investigación previa para indagar lo referente a ¿Qué es minería de datos?, cuáles son sus alcances y objetivos al igual las posibles herramientas que se utilizan para esta, sus fases del mismo y todo lo referente al estado del arte.

Levantamiento de Requerimientos.

Se realizó la entrevista inicial, en la cual se abordó la fase correspondiente al levantamiento de requerimientos con el Jefe del departamento de Planeación y Evaluación de la institución, en la cual se dio a conocer el proyecto de minería con sus alcances, objetivo general y el tiempo aproximado de duración de 2 años, ya que se pretende realizar en todas las áreas internas de la institución extendiendo la minería a los diversos departamentos y áreas que lo requieran dentro de la universidad. A través del análisis de los procesos y estudios de la información se desarrolla un concentrado de información y organización de la misma, para facilitar la toma de decisiones, la mejora continua y posible sistematización de los procesos minimizando errores humanos.

Debido al tiempo en el cual se desarrolla el proyecto se optó por realizar minería de datos en una sola área de la universidad para dar iniciativa al proyecto y dejar las demás áreas para futuros proyectos que den continuidad al que se aborda en esta ocasión.

Por lo cual se determinó como iniciativa de proyecto comenzar con direcciones de carrera, ya que es el área en que genera la información del cliente que funge como el educando, por ende, es esta quien genera más información que abastece a las diferentes áreas de la institución educativa.

Obtención de Información y Mapeo.

La institución trabaja actualmente con diez carreras de Técnico Superior Universitario y ocho de Ingeniería, cada una cuenta con su respectiva dirección de carrera, contando con cinco directores que a su cargo dirigen múltiples carreras.

Las cuales se agrupan de la siguiente manera:

- Desarrollo de Negocios (DEN), Contaduría (CONTA), Ingeniería en Negocios y Gestión de Empresarial (INGE) e Ingeniería Financiera y Fiscal (IFYF).
- Mecánica (MEC), Mantenimiento (MAN), Energías Renovables, Ingeniería en Mantenimiento Industrial (IMAN) e Ingeniería en Metal Mecánica (IMEM).
- Tecnologías de la Información y Comunicación (TIC) e Ingenierías en Tecnología de Información y Comunicación (ITIC).
- Procesos Alimentarios (PAL), Agricultura Sustentable y Protegida (ASIP), Ingeniería en Procesos Alimentarios (IPAL) e Ingeniería en Agricultura Sustentable y Protegida (IASIP).
- Diseño de Moda Industrial (DMI), Terapia Física () e Ingeniería en Diseño Textil y Moda (IDTM).

Se realizó una encuesta para la recopilación de información que se genera dentro de direcciones de carrera, que se muestra a continuación.

1. ¿Qué información se genera dentro de la dirección de carrera?
2. ¿De qué departamento viene la información?
3. ¿A qué áreas va dirigida la información?
4. ¿Quién consume la información?
5. ¿Quién la solicita?
6. ¿Con que frecuencia se solicita la información?

Al concluir con la entrevista, se programó una daily standup para evaluar los resultados de la entrevista de dirección de TIC, en la cual se propuso utilizar el documento obtenido adjuntando al cuestionario preestablecido y con este se llevó a cabo una prueba piloto en las restantes direcciones de carrera.

Al finalizar se convocó otra daily standup, para analizar la información obtenida de la prueba piloto, con base a ello se estableció que la siguiente actividad a realizar sería conocer el trazado de los documentos que se generan en las direcciones de carrera. Para la realización de dicha actividad, hubo necesidad de redirigirse al área de campo para así identificar el proceso de cada documento con ayuda de directores de carrera y secretarías se logró obtener el trayecto de cada documento.

En el transcurso de esto, se optó por conocer el SGCA para comparar los dueños del proceso y sus clientes.

Diseño de la propuesta.

Con base a lo anterior, se investigó con que herramienta sería más adecuada para modelar los resultados obtenidos. Los modelados que se adecuaban a la necesidad del proyecto son: Diagramas UML de Casos de Uso, Diagramas IDEF y Diagramas PAD, por lo que se eligió utilizar el modelado PAD, debido a que comprende los procesos de la institución y una representación (modelado) detallada de las actividades que se realizan en ellos.

Una vez determinado lo anterior, se decidió realizar el modelado de procesos de cada documento que se manejan en direcciones de carrera. Al obtener plasmado cada proceso en diagramas PAD por el equipo de trabajo, se sometieron al análisis cada diagrama buscando las relaciones entre la información y una manera de reducir actividades innecesarias de los procesos. Al tener la caracterización de la información se propuso entregar dos conjuntos de diagramas uno que contiene el proceso real de cada documento y en el segundo conjunto se incluye la propuesta de los procesos sistematizados.

Resultados

Los modelados de procesos se digitalizaron con ayuda de la herramienta de eDraw Max que permitió plasmar los diagramas PAD, facilitando la exportación a distintos formatos (Visual, HTML, WORD y PDF) realizando su entrega al asesor empresarial en formato PDF, adjuntando así la descripción y manejo de diagramas al igual que un glosario de términos.

De los 8 procesos analizados solo 2 calificaron para sistematización.

Los procesos de propuesta a sistematización son:

- Proceso y operación de estadías y titulación.
 - Convenio estadías.
 - Aprobación del reporte de estadías.
 - Acta de Exención.
- Control de asesorías y tutorías.
 - Canalización AIE y seguimiento.

Los cuales se muestran a continuación.

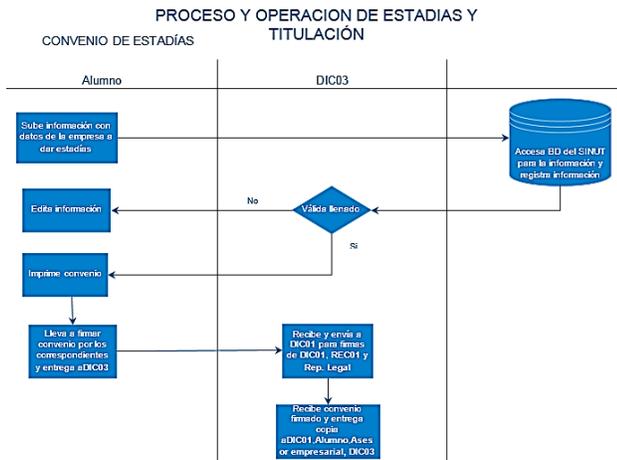


Figura 3 Proceso y operación de estadías y titulación

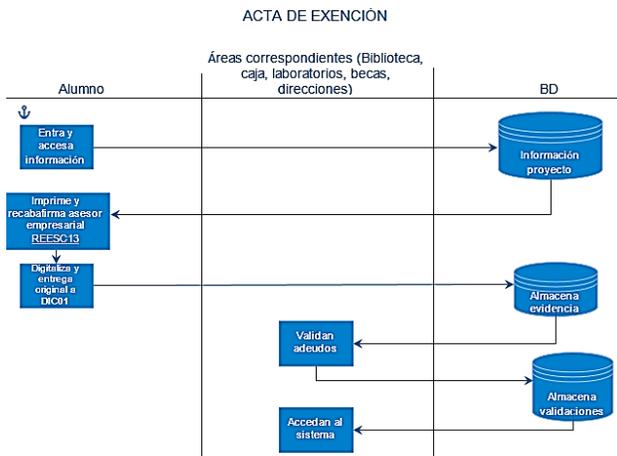


Figura 4 Proceso y operación de estadías y titulación

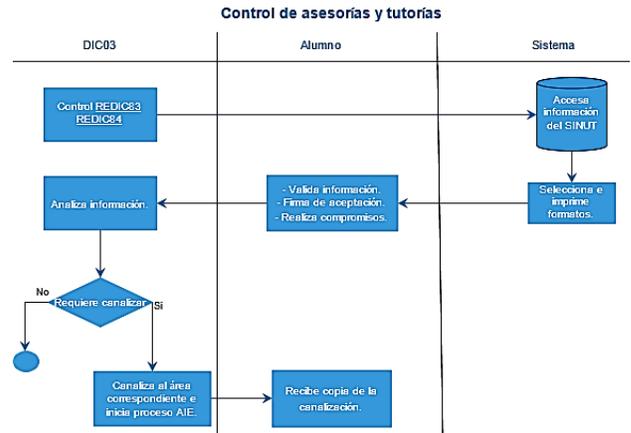


Figura 5 Control de asesorías y tutoría

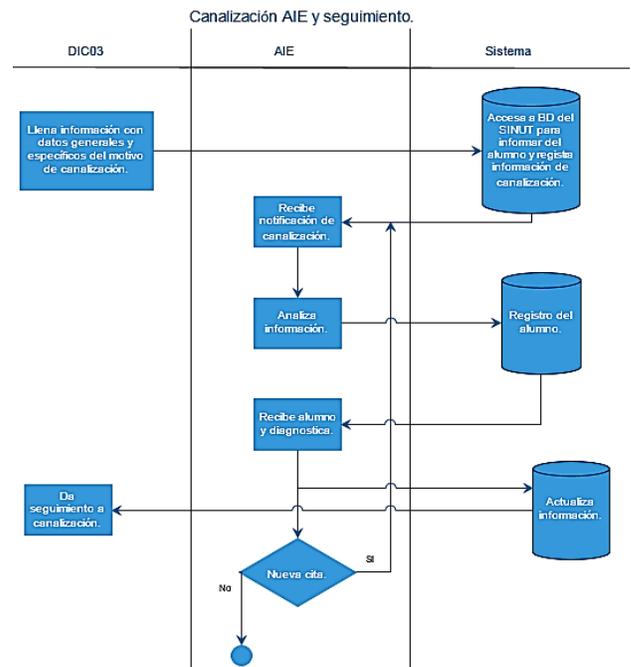


Figura 6 Canalización de AIE y seguimiento

Conclusiones

Con los resultados del proyecto de minería de datos presentado y después de haber analizado cada documento generado en las distintas direcciones de carrera y como se efectúan dichos procesos, se incluye modelado de los procesos en diagramas PAD para que los procesos calificados a sistematizar, explicando paso a paso las actividades de cómo se podría llevar el proceso con ayuda de una aplicación y así facilitando programador la manera de imaginarse el sistema y la manera de emplearlo, con la unión de éste proyecto y futuros trabajos en colaboración ayudar a mejorar el sistema con el que actualmente trabaja la universidad ya que no se cuenta con las herramientas necesarias para realizar de manera más dinámica los procesos, así mismo minimizando el error humano durante el transcurso de los procesos que se propone proseguir en este proyecto.

Es posible concluir que se obtuvo material de importancia la cual se adjunta en este reporte que será de utilidad a distintas áreas y departamentos de la institución educativa para realizar un análisis, con base a ello poder actualizar documentación para hacer más factible el funcionamiento interno de la UTSOE y posiblemente tener mejores resultados.

En conjunto, la experiencia produce a una primera propuesta del modelo conceptual de la minería de datos pretendida. Esta determina la etapa de proyección para los siguientes momentos de diseño, planeación y modelado informático, la propuesta de sistematización se basa en que los procesos son más realizados con más frecuencia invierten más horas de trabajo conllevando posibles errores en el transcurso de las actividades, de manera paralela en algunos procesos interviene el educando, siendo así los más factibles la sistematización para algunos procesos.

Agradecimiento

A la Universidad Tecnológica del Suroeste de Guanajuato, carreras de:

-Tecnologías de la Información y Comunicación área Sistemas Informáticos. Cuerpo Académico de Ambientes Inteligentes y Cómputo Suave.

Referencias

Ahumada, A. M. (7 de Abril de 2016). Minería de datos, de textos y de sentimientos. Obtenido de <http://www.gestiopolis.com/mineria-datos-textos-sentimientos-2/>

Booch, G. (12 de Febrero de 2012). En G. Booch, El Lenguaje Unificado de Modelado guía del usuario (pág. 527). Boston: Pearson. Obtenido de <https://prezi.com/uwmb4dpdae6a/modelado-de-procesos/>

Castro, J. V. (10 de Diciembre de 2015). Obtenido de <https://www.proyectosagiles.org/que-es-scrum> EdrawSoft. (2 de Junio de 2014). Obtenido de <https://www.edrawsoft.com/es/howitworks.php>

Nuño, R. R. (1 de Mayo de 2013). Textual Mining. Obtenido de <http://textmining.galeon.com/>

Rio, J. S. (27 de Febrero de 2015). Matriz Raci. Obtenido de <http://www.cantabriatic.com/que-es-una-matriz-raci/>