

Tipo de artículo: Artículo original
Temática: Minería de Procesos
Recibido: 10/09/16 | Aceptado: 10/10/16 | Publicado: 21/11/16

Componente para la transformación y extracción de trazas del Sistema de Gestión de Decisiones para aplicar técnicas de Minería de Procesos

Component for the transformation and extraction of log the System of Decision Management to apply techniques of Process Mining

Manuel Alvarez Alonso ^{1*}, Abraham Calás Torres ², Pavel Eduardo Barzaga De la Paz ³

¹Centro de Gobierno Electrónico, Universidad de las Ciencias Informáticas, La Habana, Cuba, malvareza@uci.cu.

²Centro de Informatización de Entidades, Universidad de las Ciencias Informáticas, La Habana, Cuba, acalas@uci.cu.

³Centro de Informatización de Entidades, Universidad de las Ciencias Informáticas, La Habana, Cuba, pebarzaga@uci.cu.

* Autor para correspondencia: malvareza@uci.cu.

Resumen

La Fiscalía General de la República de Cuba se ha insertado en el proceso de informatización que vive la sociedad actual potenciando el empleo de las Tecnologías de la Información y las Comunicaciones en sus áreas. Ejemplo de ello, es el Sistema para la Gestión de Decisiones (SGD), encaminado al proceso de “controlar y evaluar el cumplimiento de las decisiones, para garantizar la veracidad y oportunidad de las informaciones que se emiten”. La Minería de Procesos (MP) es una disciplina informática dirigida al desarrollo de técnicas y herramientas para analizar los registros de eventos generados durante la explotación de otros sistemas, extraer información de ellos y presentar explícitamente el conocimiento que contienen; lo que, aplicadas a los procesos del SGD, aumentarían la calidad y eficacia de los mismos. El punto de partida para la aplicación de técnicas de MP son los eventos registrados y, la calidad de los mismos está directamente relacionada con la del resultado que arrojan las técnicas de MP aplicadas

sobre dichos registros. La presente investigación aborda la solución desarrollada para la creación de registros de eventos a partir de las trazas y datos de negocio del SGD, garantizando que cumplan con los estándares y formatos internacionalmente definidos para la utilización en la MP. La propuesta se valida evaluando los principales riesgos de calidad que existen en este tipo de soluciones y analizando cómo se mitigaron estos riesgos. Además, se muestran los resultados obtenidos al aplicar técnicas de MP utilizando los registros generados por la solución.

Palabras clave: Minería de Procesos, registro de evento, calidad

Abstract

The Attorney General's Office of the Republic of Cuba has been inserted in the process of computerization that the current society is experiencing, enhancing the use of Information and Communication Technologies in its areas. An example of this is the Decision Management System (DGS), which is aimed at "monitoring and evaluating compliance with decisions, in order to guarantee the truthfulness and timeliness of the information that is issued." Process Mining (MP) is a computer discipline aimed at developing techniques and tools to analyze the records of events generated during the exploitation of other systems, extract information from them and explicitly present the knowledge they contain; which, applied to the processes of the DGS, would increase the quality and effectiveness of that system. The starting point for the application of PM techniques are the recorded events and their quality is directly related to the quality of the result that the MP techniques applied on those registers. This research deals with the solution developed for the creation of event records based on the business data and records of the SGD, ensuring that they comply with internationally defined standards and formats for use in the PM. The proposal is validated by evaluating the main quality risks that exist in this type of solutions and analyzing how these risks were mitigated. In addition, the results obtained by applying MP techniques using the logs generated by the solution are shown.

Keywords: Process Mining, event log, quality

Introducción

El desarrollo de la sociedad actual está estrechamente relacionado con el avance de las TIC, marcando un cambio significativo en el mundo en cuanto a informática se refiere, ésta se ha expandido en varios niveles, surgiendo así nuevas tendencias como Gobierno Electrónico. Las nuevas tecnologías además de brindar la posibilidad de recolectar,

almacenar y procesar la información, también permiten que las organizaciones gubernamentales puedan ofrecer sus servicios y con ello establecer una comunicación más estrecha con los ciudadanos (Acosta, 2016).

La Fiscalía General de la República de Cuba (FGR) se ha insertado en esta tendencia y ha potenciado la informatización de sus distintas áreas (Machado, 2016). La Secretaría de la FGR, según contempla la resolución 120/13, tiene como uno de sus procesos principales, el de “controlar y evaluar el cumplimiento de las decisiones, para garantizar la veracidad y oportunidad de las informaciones que se emitan” (Cura, 2013), en el año 2016 puso en explotación un sistema para la gestión de sus decisiones (SGD). Las decisiones en la FGR se definen como las tareas, orientaciones, posibles actividades, etc. designadas por los jefes y que deben realizar los subordinados en un término específico. Para la FGR, el SGD constituye una fuente de información con un impacto muy alto en la toma de decisiones de este órgano, encargado de velar por el cumplimiento de la legalidad, además, agiliza el procesamiento de los datos, su envío y recepción en las instancias superiores y deja abierta la puerta para desde los datos que este procesa mejorar la calidad del trabajo de La Secretaria de la FGR (Noa, y otros, 2016).

Las organizaciones y empresas poseen y generan diariamente un gran cúmulo de datos imposibles de analizar a simple vista. Mucho esfuerzo es dedicado por las mismas en el análisis y la mejora de procesos para optimizar su desempeño organizacional y operacional. Analizar procesos consume mucho tiempo, involucra a muchas personas y es costoso (Heijden, 2012). La Minería de Procesos (MP) es una disciplina dirigida al desarrollo de técnicas y herramientas para analizar los registros de eventos, extraer información a partir de ellos y presentar de forma explícita el conocimiento que contienen (Aalst, 2011), su aplicación aprovecha los datos de eventos en una forma significativa, por ejemplo, para proveer un mejor entendimiento, identificar cuellos de botella, anticipar problemas, registrar violaciones de políticas, recomendar contramedidas, y simplificar procesos (Aalst, 2011). Estas mejoras aplicadas a los procesos que gestiona el SGD facilitarían un aumento en la calidad y eficacia de los mismos lo que convierte a la MP en una disciplina que está en consonancia con los intereses de la FGR.

El punto de partida para cualquier actividad de Minería de Procesos son los eventos registrados (Aalst, 2011), en el caso SGD almacena registros de cada una de las actividades que se ejecutan en el sistema que se dividen en dos grupos las concerniente a las acciones y las que registran el acceso a los datos, lo que provoca que estos registros puedan no estar relacionados con procesos que son el objetivo de los futuros análisis y la información concerniente a cada uno de los procesos, se encuentra dispersa y no recoge todos los datos relacionados con la ejecución de los mismos.

Teniendo en cuenta lo anterior, además de que la calidad del resultado de Minería de Proceso en gran medida depende de la entrada (Aalst, 2011) y que las técnicas de esta disciplina son menos efectivas cuando deben manejar

información incompleta (Bose, 2013), se identificó la pertinencia de desarrollar de una solución informática que permita la extracción y transformación de las trazas registradas por el SGD para realizar análisis de los procesos ejecutados utilizando técnicas de Minería de Procesos.

Materiales y métodos o Metodología computacional

Registros de eventos

Los registros de eventos están compuestos por instancias de proceso y cada instancia contiene información relacionada con la ejecución de una acción u operación específica que se ha ejecutado en un sistema y su marca de tiempo correspondiente. Estos registros de elementos también pueden contener atributos. Como el registro de elementos solo se crea una vez, el impacto de incluir muchos atributos en el registro es mínimo (Buijs, 2010). Los siguientes atributos son frecuentemente tomados en consideración para su adición en el registro y brindan un conjunto de metadatos útiles para el futuro análisis de los registros (García, y otros, 2015):

- Nombre de Proceso: El nombre del proceso al cual el registro le graba su ejecución.
- Fuente de Datos Una descripción del Sistema de Información del cual se extrae el Registro de eventos.
- Organización Fuente: El nombre de la organización que provee los datos.
- Descripción: Una breve descripción del contenido del Registro de eventos.
- Versión: Un identificador para diferenciar versiones de Registros de eventos.
- Autor: Nombre y detalles de contacto del que definió la conversión.
- Proyecto de Minería de proceso: Una referencia del Proyecto de Minería de proceso o el propósito del Registro de eventos.

Para formalizar la estructura de los registros de eventos a utilizar en la Minería de Procesos se han definido dos estándares: Mining eXtensible Markup Language (MXML) y Extensible Event Stream (XES). En el 2010 XES reemplaza a MXML como el nuevo formato para Minería de Procesos independiente de la herramienta (Aalst, 2011). Es un estándar basado en experiencias prácticas de MXML, menos restrictivo y verdaderamente extensible. Su principal propósito es ofrecer un formato de intercambio de registros de eventos entre herramientas y dominios de aplicaciones (Günther, 2009).

Calidad de los registros

La combinación de los tres aspectos de calidad se refleja en los niveles de madurez definidos por el Institute of Electrical and Electronics Engineers (IEEE) para los registros de eventos (Aalst, 2011).

Nivel-1: Los eventos se registran manualmente, los eventos registrados podrían no corresponder a la realidad y podrían faltar eventos.

Nivel-2: Los eventos se registran automáticamente, no se sigue un enfoque sistemático para decidir qué eventos se registran, podrían faltar eventos o estos podrían no registrarse correctamente.

Nivel-3: Los eventos se registran automáticamente, pero no se sigue un enfoque sistemático para hacerlo. Hay un nivel de garantía de que los eventos registrados se corresponden con la realidad (el registro de eventos es confiable pero no necesariamente completo).

Nivel-4: Los eventos se registran automáticamente y de manera sistemática y confiable, a diferencia del nivel 3 da soporte de manera explícita a nociones tales como instancia de proceso (caso) y actividad.

Nivel-5: El registro de eventos es de excelente calidad (confiable y completa) y los eventos están bien definidos. Los eventos se registran de manera automática, sistemática y segura. Se toman en cuenta adecuadamente consideraciones acerca de la privacidad y la seguridad.

El conocimiento del nivel en que se encuentran los registros con los que se cuenta permite entender las limitaciones potenciales de los mismos y traza un camino para la mejora constante de los mismos y con ello, la potencial mejora de los resultados obtenidos en el análisis de los mismos.

Extracción de registros XES para el Sistema de Gestión de Decisiones

En la actualidad el SGD almacena trazas en una estructura de datos donde se agrupa toda la información concerniente a la misma, sin relaciones directas con el resto de las estructuras del SGD. Atendiendo a que los eventos se registran automáticamente y hay un nivel alto de fiabilidad en los registros recogidos, pero no se sigue un enfoque sistemático para registrar los eventos podemos situar las trazas que hoy genera el SGD en la categoría 3 de 5 posibles lo que deja una brecha de mejora que influye en los resultados que pudieran brindar futuros trabajos de Minería de Procesos sobre estos registros.

Para que el registro sea válido para algoritmos o técnicas de Minería de Procesos, debe referenciar las extensiones del estándar, contener una serie de atributos globales y definir clasificadores para las trazas y/o eventos. En el caso del registro a generar las extensiones utilizadas son: Concept, Time, Lifecycle, Organizational y Micro. Los atributos globales son elementos comunes en todas las trazas o los eventos; para la solución se definen los siguientes atributos a nivel de eventos: nombre, marca de tiempo, usuario (que ejecutó el evento), el rol del usuario y la fiscalía a la que

pertenece. Los clasificadores asignan un identificador a los eventos para realizar comparaciones entre ellos (XES Standard Definition, 2014), atendiendo a esto en la solución los eventos se clasifican por su nombre.

Como parte de la solución desarrollada se definió una estructura de datos compuesta por los campos mostrados en la figura 1 y que satisfacen las necesidades de almacenamiento enunciadas anteriormente. Con esto se persiguen dos objetivos, primero separar el procesamiento de estos datos para la obtención de los registros XES del SGD y de esta forma mitigar la posible competencia por los mismos recursos de procesamiento o posible concurrencia sobre un determinados datos; y como segunda meta, separar los datos que son relevantes para futuros análisis del total generado por el SGD y de esta manera disminuir la cantidad de registros a procesar y con esto los tiempos de procesamiento.



El diagrama muestra una estructura de datos para registros, representada como una tabla con un fondo naranja. El título de la tabla es "Registros". La tabla contiene las siguientes columnas y filas:


Registros	
 Identificado	numeric(19, 0)
 Instancia de proceso	varchar(255)
 Evento	varchar(255)
 Recurso que ejecuta la actividad	varchar(255)
 Marca de tiempo	timestamp
 Instancia de fiscalía	integer(10)

Figura 1 Estructura de datos para salvar los nuevos registros

Un trigger o disparador es un objeto procedimental asociado a una tabla que entre tus aplicaciones tiene generar automáticamente valores de columnas derivadas en base a un valor proporcionado por una sentencia INSERT o UPDATE (Informáticos, 2012). Aprovechando estas potencialidades el llenado de la nueva estructura de datos se efectúa a través de trigger dispuestos en la tabla de registros del sistema de gestión. En la función utilizada por el trigger, se garantiza que el registro a procesar sea relevante para el análisis, se seleccionan los valores relevantes del registro en cuestión y a partir de estos, se buscan los que no están implícitos en el mismo de manera que se puedan completar los datos necesarios en la estructura auxiliar creada. Para la obtención de los datos que permiten la creación de los registros se utilizan los registros del sistema y las relaciones indirectas de estos con los datos de sistema de la forma que se describe a continuación.

La instancia del proceso y el recurso que ejecuta la actividad son obtenidos directamente del registro que hoy genera el SGD. El evento y la fiscalía en la que se está manejando el proceso, son obtenidos mediante la asociación con datos del negocio en la base de datos; en el caso del primero, es indispensable para la conformación y posterior utilización de cualquier archivo XES, y la no existencia del segundo, limitaría los futuros análisis de la información y cerraría la puerta a la Minería Inter-Organizacional uno de los desafíos de la minería de procesos (Aalst, 2011). Unido a esto, como no están registrados en las tablas de trazas propiamente y en las que se encuentran es susceptible a actualizaciones, podría introducirse ruido en la información que manejamos de no salvarse inmediatamente después de que se genera el registro. Para el caso de la fecha y hora en la que se sucede el evento al que hace referencia el registro, elementos que nos facilitarían la marca de tiempo indispensable en un registro XES, se fusionan datos: los datos propios del registro.

A partir de los datos salvados en esta estructura se desarrolló la funcionalidad exportar registro, capaz de crear los registros XES teniendo como punto de partida los registros salvados y que le da la posibilidad al usuario de seleccionar el intervalo de fecha que tendría en cuenta el recurso involucrado en los procesos, la posibilidad de

abarcar todas las actividades correspondientes a los procesos desarrollados en el intervalo de tiempo seleccionado, aunque las mismas no estén en el rango de tiempo seleccionado, y la o las instancias de la fiscalía a la que pertenecen.

Resultados y discusión

Problemas de calidad en archivos XES

Existe un conjunto de problemas independientes del enfoque, que están presentes en los registros de eventos extraídos de sistemas reales y están asociados a la herramienta empleada para conformar el registro, el algoritmo utilizado o las condiciones del negocio: Granularidad de eventos, Heterogeneidad de los casos, Datos voluminosos, Problemas relacionados con las marcas de tiempo, Pérdida de datos, Ambigüedad entre eventos, Flexibilidad de los procesos, Ruido, Procesos agrupados y Alcance (Bose, 2013). Estos problemas fueron considerados a la hora de crear el registro de eventos para que la información obtenida sea lo más completa posible.

Atendiendo a las categorías descritas anteriormente y los distintos componentes de un registro de eventos, se han detectado 27 problemas de calidad que se encuentran frecuentemente en los registros. En la tabla 1 Bose 2013 señala la relación entre estos dos elementos y los enumeran para un mejor análisis de los mismos.

	Caso	Evento	Asociación	Atributo de caso	Posición	Nombre de actividad	Marca de tiempo	Recurso	Atributo de evento
Ausente	P1	P2	P3	P4	P5	P6	P7	P8	P9
Incorrecto	P10	P11	P12	P13	P14	P15	P16	P17	P18
Impreciso			P19	P20	P21	P22	P23	P24	P25
Irrelevante	P26	P27							

Tabla 1 Relación entre las distintas categorías y problemas en registros de eventos

En el desarrollo de la solución se han ejecutado un conjunto de acciones y se han aprovechado un grupo de premisas o características con las que se contó inicialmente para mitigar estas posibles problemáticas y de esta forma contribuir a

la calidad de los registros obtenidos con la solución y con ello mejorar la certeza de los resultados. A continuación, se listan estos elementos y se denotan con una **S** seguido de un número consecutivo:

S1. Datos presentes en los registros del SGD.

S2. Datos obtenidos mediante asociación entre los datos de los registros del SGD y la información que, referente al proceso, se encuentra en otras tablas en la base de datos del SGD.

S3. Selección de las tuplas a incluir en el registro de eventos.

S4. Información brindada por el usuario al seleccionar los valores de entrada para la creación del registro.

S5. Atendiendo a las características de los procesos que atiende el SGD no existen registros con características que provoquen la ocurrencia de este error.

En las tabla 2 y 3 se señalan cuáles son las medidas o premisas que impactaron en la mitigación de cada uno de los problemas enunciados anteriormente.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14
S1	X	X	X		X	X	X			X	X	X	X	X
S2				X				X	X					
S3														
S4														
S5														

Tabla 2 Medidas o premisas que impactan en los problemas del 1 al 14

	P15	P16	P17	P18	P19	P20	P21	P22	P23	P24	P25	P26	P27
S1	X	X			X	X	X	X	X				
S2			X	X						X	X		
S3												X	X
S4												X	X
S5									X				

Tabla 3 Medidas o premisas que impactan en los problemas del 15 al 27

Conclusiones

Se detectó la ausencia de la información mínima necesaria para aplicar MP en las trazas del SGD por lo que se desarrolló una solución que ubicara e incorporara los datos necesarios en los datos de negocio del sistema.

Atendiendo a las necesidades mínimas de la mayoría de las técnicas de MP y las posibles aplicaciones que estas tengan en dominio al que pertenece el SGD se definieron como atributos a salvar en la estructura de datos creada: instancia de proceso, proceso, marca de tiempo, recurso e instancia de la fiscalía.

El análisis de los problemas de calidad más comunes en este tipo de soluciones permitió identificar soluciones para mitigar su ocurrencia.

La solución implementada permite a la dirección de la FGR analizar sus procesos y utilizar esta información en la toma de decisiones.

Referencias

Aalst, Wil M.P. Van der. 2011. Process Mining: Discovery, Conformance and Enhancement of Business Processes. [Online] 2011.

https://books.google.com/cu/books?hl=en&lr=&id=I1KOAfiqfxYC&oi=fnd&pg=PR4&dq=Discovery,+Conformance+and+Enhancement+of+Business+Processes&ots=K-kS69vVBL&sig=WLc3SuDDHQYJbX4dZpraXFzxsxGI&redir_esc=y#v=onepage&q=Discovery%20Conformance%20and%20Enhancement

Aalst, Wil van der et al. 2011. IEEE TASK FORCE ON PROCESS MINING, 2011. Manifiesto sobre Minería de proceso. [Online] 2011. <http://www.win.tue.nl/ieeetfpm>.

ACOSTA, JOSÉ CARLOS PUPO. 2016. Foro de Empresarios y Líderes en Tecnologías de la Información. [Online] 2016.

<http://www.felti.org/sites/default/files/archivos/postulacion/latinatec2016sistemadegestionfiscal.pdf>.

BUIJS, J. C. A. M. 2010. Mapping data sources to xes in a generic way. [Online] 2010. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.653.8165&rep=rep1&type=pdf>.

Cura, Darío Delgado. 2013. *Resolución 120/13 del Fiscal General de la República de Cuba*. La Habana : s.n., 2013.

García, Arturo Orellana, Armenteros, Osvaldo Larrea and Alfonso, Damián Pérez. 2015. Generador de Registros de Eventos para el análisis de procesos en el Sistema de Información Hospitalaria xavia HIS. *Convención Salud 2015*. [Online] 2015.

<http://www.convencionsalud2015.sld.cu/index.php/convencionsalud/2015/paper/viewPaper/380>.

Günther, Christian. 2009. Process mining in flexible environments. [Online] 2009.

<http://repository.tue.nl/dc50e490-e9ce-4c20-bcfe-5c9e992daa54>.

HEIJDEN, VAN DER. 2012. Process Mining Project Methodology Developing a General Approach to Apply Process Mining in Practice. [Online] 2012.

http://alexandria.tue.nl/extra2/afstversl/tm/Van_der_Heijden_2012.pdf.

Informáticos, Departamento de Lenguajes y Sistemas. Universidad de Sevilla. [Online] <http://www.lsi.us.es/docencia/get.php?id=4570>.

Machado, Marlén Fernández. 2016. *Sitio Web de la Fiscalía General de la Republica de Cuba*. [Online] 2016. <http://www.fgr.cu/es/por-el-bien-del-ciudadano>.

Noa, Arnel Abad and Morales, Heidi Infante. 2016. Sistema para la gestión de las decisiones de la Fiscalía General de la República. *Congreso Internacion de Informacion INFO 2016*. [Online] 2016. <http://www.congreso-info.cu/index.php/info/2016/paper/viewFile/230/186>.

Wanna Improve Process Mining Results? It's High Time We Consider Data Quality Issues Seriously.

BOSE, RP Jagadeesh Chandra, MANS, Ronny S. and VAN DER AALST, Wil MP. 2013. s.l. : Computational Intelligence and Data Mining (CIDM) IEEE Symposium, 2013.

Wanna Improve Process Mining Results? It's High Time We Consider Data Quality Issues Seriously.

Chandra Bose, R.P. Jagadeesh, Mans, Ronny S. and van der Aalst, Wil M.P. 2013. 2013.

XES Standard Definition. **Günther, Christian and Verbeek, Eric. 2014.** 2014.