

Propuesta de Tesis de Maestría en Ingeniería de Software

a) Nombre y apellido del Alumno/Tesista.

Lic. Victor Martinez

b) Nombre y apellido del Director/es

Directores: Dra. Laura Lanzarini – Dr. Franco Ronchetti

c) Título del Tema de Tesis propuesto

Minería de Procesos y Minería de Datos. Análisis del proceso de distribución de piezas postales.

d) Objetivo

La minería de procesos es una herramienta sumamente útil para analizar procesos basados en registros de eventos. Para ello, hace uso de un conjunto de técnicas capaces de modelizar dichos procesos para luego extraer información útil relacionada con su funcionamiento que ayude a la toma de decisiones. La modelización del proceso no sólo permite comprender el funcionamiento real y contrastarlo con el diseñado originalmente, sino que facilita la detección de desvíos operativos y cuellos de botella; aspectos que redundan en una mejora del proceso agilizando su funcionamiento y reduciendo las congestiones.

El objetivo de esta tesis es construir un modelo que permita identificar patrones en el proceso de distribución postal. Dadas las características propias de la distribución, en ocasiones, debido a errores humanos u otras fallas, se producen retrasos o desvíos operativos que impactan en la calidad del servicio. Por tal motivo, se buscará detectar desvíos o retrasos en dicho proceso de distribución con el objetivo de sugerir posibles mejoras.

Se utilizarán las técnicas de minería de procesos para analizar los registros de eventos correspondientes a la distribución de piezas realizadas por la empresa Correo Argentino durante los años 2017 hasta 2020. Se aplicarán los conceptos de la ingeniería de software a la minería de procesos, analizando un proceso real que interactúa con un volumen importante de datos. Se modelizará el proceso actual con técnicas de Minería de Procesos y luego se comparará contra el ideal o el propuesto originalmente para encontrar diferencias, cuellos de botella y sugerir posibles mejoras.

Se trabajará con técnicas de agrupamiento para identificar recorridos similares de piezas. Del posterior análisis de los grupos formados se espera identificar motivos de retrasos que permitan sugerir mejoras al proceso de distribución.

Finalmente, en base a la media de los casos finalizados, se identificarán aquellos que estén en curso y puedan ser desvíos, enviando alertas antes de su finalización facilitando la solución y mejorando la calidad del servicio.

El aporte de esta tesis consistirá en el desarrollo de una herramienta de software que analice los datos de distribución, detecte desvíos y alerte sobre los mismos, oportunamente. El conocimiento adquirido será de ayuda tanto en el diagnóstico correcto del problema como en la formulación de estrategias de solución.

Los resultados alcanzados y los modelos obtenidos estarán dirigidos a personas que no estén familiarizadas con la minería de procesos y que no tengan un perfil técnico. Dichos resultados serán expresados de una forma clara que facilite su interpretación. Además, podrán servir de base para estudios posteriores que contribuyan al seguimiento del problema planteado y al análisis de resultados luego de la aplicación de las medidas que pueda adoptar la empresa.

e) Motivación /Estado del Arte del Tema

Según la definición de la IEEE, la ingeniería de software es la aplicación de un enfoque sistemático, disciplinado y cuantificable al desarrollo, operación y mantenimiento del software; es decir, la aplicación de la ingeniería al software.

La ingeniería de software se ha desarrollado desde hace aproximadamente 50 años buscando diseñar, desarrollar y mantener herramientas de software que se ajusten a las necesidades de las personas y de las organizaciones.

Generalmente se siguen una serie de pasos establecidos según la metodología utilizada y se realizan auditorías para verificar la calidad del proceso.

Por su parte, la minería de procesos tiene por objetivo central obtener conocimiento a través registros, o logs de eventos, con el fin de mejorar los procesos de las organizaciones. Por su intermedio, es posible tener detalles del funcionamiento del proceso en estudio con distintos niveles de abstracción.

Por tratarse de un proceso inductivo, basado en la generalización de logs, la calidad del modelo obtenido, asociado al proceso descubierto, dependerá de la cantidad y calidad de los datos disponibles.

Dado que el descubrimiento se realiza en base a los datos registrados, es de fundamental importancia relevar todos los aspectos centrales del proceso a analizar dado que aquellos que no se encuentre registrado no podrá formar parte del modelo y por lo tanto tampoco de los patrones a identificar y/o reconocer.

En sus comienzos la minería de procesos se encontró con una limitación significativa relacionada con la dificultad para contar con grandes volúmenes de datos. Este aspecto, gracias al avance de la tecnología, se ha visto resuelto. Actualmente una gran cantidad de herramientas de software dejan logs de las tareas realizadas ya sea en archivos de texto o en bases de datos. Estas herramientas pueden ser desarrollos propios o empresariales como software de gestión ERPs (SAP), de ventas CMRs, de procesos productivos MRPs, de almacenamiento WMSs, de mantenimiento (bug tracking), de control de calidad, o correos electrónicos. Incluso, con las nuevas tendencias, se generan grandes volúmenes de datos que se pueden explotar; tal es el caso de la información relacionada con la actividad de los usuarios en las redes sociales o el rastro dejado por internet de las cosas (IOT) en dispositivos del hogar o aparatos domésticos inteligentes.

Dada la gran variedad de logs disponibles, las técnicas de minería de procesos se pueden aplicar en un amplio espectro de actividades por ejemplo en la elaboración de productos, distribución, almacenamiento, logística, medicina, desarrollo de software, gestión de empresas u organizaciones públicas o privadas.

Las técnicas se adaptan prácticamente a cualquier proceso de negocio debido a que son genéricas y pueden aplicarse siempre que los datos posean las siguientes características:

1. Deben estar almacenados en uno o más registros de eventos.
2. En el registro debe haber datos de varias instancias.
3. Cada instancia debe tener sus eventos.
4. Cada evento debe tener el identificador de la instancia a la que pertenece, una descripción, y un timestamp de cuando se realizó.
5. Opcionalmente se puede agregar información acerca de quién realizó el evento, la fecha de finalización y cualquier dato relacionado al negocio.

Hoy uno de los desafíos es utilizar de forma eficiente los datos para realizar análisis de calidad. Dada la gran variedad y complejidad de los procesos operativos es importante vincular la ingeniería de software a la minería de procesos generando herramientas que obtengan resultados rápidos y faciliten el análisis para las personas que tienen que tomar decisiones. La amplia variedad de datos que hay actualmente almacenada en diversas fuentes requiere también un procesamiento previo de los mismos para ajustarlos a la forma requerida.

La minería de procesos utiliza técnicas de la minería de datos y se combina con esta para obtener información relevante de un proceso de negocio utilizando las trazas dejadas en los logs de eventos.

Se puede realizar la minería de procesos de tres formas diferentes:

1. Analizando un histórico de eventos se descubre el proceso real utilizando diferentes algoritmos (Alpha, Inductive Miner, Heuristics Miner) sin información previa.
2. Realizando una verificación de conformidad; para ello se contrasta un proceso (puede ser el descubierto o uno existente) con el log de eventos para comprobar cuántos casos se ajustan al proceso y cuáles son desvíos
3. Buscando mejoras al proceso; luego del análisis del log, se identifican cuellos de botella y se sugieren cambios en función de los resultados.

Basándose en estos datos se pueden aplicar las técnicas de esta metodología para encontrar y verificar el proceso real, no el que se diseñó originalmente sino el que se usa diariamente.

La minería de procesos se ubica como un nexo entre la gestión de procesos y la minería de datos, buscando generar información de fácil interpretación para personas que no tengan un gran conocimiento técnico, es decir no hace falta ser un científico de datos para interpretar los resultados obtenidos de un buen análisis basado en minería de procesos.

El proceso de distribución de piezas postales tiene una gran complejidad debido a las características de cada producto. Por ejemplo, algunos tienen diferentes estándares de entrega en los tiempos requeridos, o distinta cantidad de intentos de entrega o diferentes formas de ingreso. Todo esto representa un gran desafío para dar un servicio de calidad.

Las herramientas de análisis actuales se enfocan en los datos almacenados de las distribuciones finalizadas y si bien detectan varios problemas en la distribución lo hacen a posteriori, siempre sobre casos finalizados.

En esta tesis, se generará una herramienta que analice los datos, detecte los desvíos y alerte sobre los mismos, antes de su finalización, buscando una mejora en el servicio.

f) Temas de Investigación.

- Estudio y análisis de diferentes técnicas de representación de secuencias de eventos.
- Estudio de técnicas de Minería de Datos para determinar similitudes entre trazas (secuencias de eventos) utilizando técnicas de agrupamiento. Se trabajará con técnicas de clustering aplicadas sobre distintas representaciones.
- Estudio de técnicas de Minería de Procesos como herramienta para la identificación de patrones en un proceso de negocio.
- Estudio e implementación de distintas técnicas de visualización buscando destacar las partes del proceso que pueden ser mejoradas.

g) Desarrollos/Trabajo Experimental a Realizar

Las tareas experimentales y de desarrollo de esta tesis son las siguientes:

- Construcción de la vista minable a partir de la recopilación e integración de los datos de secuencia de eventos. Ajuste de los datos a formato adecuado para las técnicas de minería de datos y de procesos.
- Implementación de una técnica de agrupamiento que identifique similitudes entre trazas.
- Construcción del modelo del proceso a partir de los registros de eventos y teniendo en cuenta el resultado de las técnicas de agrupamiento.
- Visualización del modelo destacando las zonas más congestionadas.
- Análisis del modelo obtenido para establecer alarmas.

h) Esquema de Plan de Trabajo C/Actividades y Tiempos

ACTIVIDAD	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
Lectura bibliografía de Minería de procesos y de datos												
Búsqueda y análisis de investigaciones similares												
Recopilación e integración de datos												
Agrupamiento de secuencias de eventos												
Elección de técnicas y constr. de modelos												
Visualización del modelo del proceso												
Evaluación e interpretación de los patrones descubiertos												
Redacción de Tesis												

i) Posibilidades de Realización en el Ámbito del Tesista

El tesista se desempeña como analista de sistemas en el Correo Argentino desde hace más de 14 años, realizando nuevos proyectos de software y mantenimiento de software existente contando con acceso a datos para generar las muestras necesarias.

Dispone de la infraestructura adecuada para realizar la investigación y con conocimientos teóricos de la metodología utilizada en la minería de procesos.

Se espera que los resultados obtenidos de esta investigación puedan utilizarse para mejorar el proceso actual dentro de la empresa.

j) Bibliografía Básica Relacionada

- Aalst, W. v. (2016). *Process Mining - Data Science in Action Second edition*. London: Springer.
- Aalst, W. v. (2011). *Process Mining - Discovery, Conformance and Enhancement of Business Processes*. London: Springer.
- Aalst, W. v. (2012) *Process Mining Manifesto*. In: Daniel F., Barkaoui K., Dustdar S. (eds) *Business Process Management Workshops. BPM 2011. Lecture Notes in Business Information Processing*, vol 99. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-28108-2_19
- Aalst, W. v., W., Weijters, A., Maruster, L.: Workflow mining: discovering process models from event logs. *IEEE Trans. Knowl. Data Eng.* 16(9), 1128–1142 (2004)
- Cleiton dos Santos Garcia, Alex Meincheim, Elio Ribeiro Faria Junior, Marcelo Rosano Dallagassa, Denise Maria Vecino Sato, Deborah Ribeiro Carvalho, Eduardo Alves Portela Santos, Edson Emilio Scalabrin (2019). "Process mining techniques

and applications – A systematic mapping study”, *Expert Systems with Applications*, Volume 133, Pages 260-295, ISSN 0957-4174.

- M. Pegoraro and W. M. P. van der Aalst (2019), "Mining Uncertain Event Data in Process Mining," *2019 International Conference on Process Mining (ICPM)*, Aachen, Germany, pp. 89-96, doi: 10.1109/ICPM.2019.00023.
- P. Markowski and M. R. Przybyłek, "Process mining methods for post-delivery validation," *2017 Federated Conference on Computer Science and Information Systems (FedCSIS)*, Prague, 2017, pp. 1199-1202, doi: 10.15439/2017F372.