

Deciphering event logs in SharePoint Server: A methodology based on Process Mining

Michael Arias Chaves
Universidad de Costa Rica
Pontificia Universidad Católica de Chile
michael.arias@uc.cl

Eric Rojas Córdoba
Universidad de Costa Rica
Pontificia Universidad Católica de Chile
eurojas@uc.cl

Abstract— Nowadays, the information systems are an indispensable resource for the organizations. The processes that are managed through these systems most of the time are hard to understand, maintain and improve. The data associated to the process becomes the main source and input to do all types of analysis. Process Mining allows the extraction of useful knowledge from the generated information of the corporate systems.

This work suggests a methodology based on process mining to execute process support analysis of a corporate intranet implemented in SharePoint Server. With the extracted information it is possible to do analysis from several perspectives. The obtained results allow administrators of this type of technology platforms to evaluate the techniques used and generate benefits. The methodology was applied in a case study for a Retail client, besides doing an exploratory analysis of the data for two additional clients in the Industrial Safety industry in Chile.

Key words: *Business Processes; Process Mining; BPMN; SharePoint Server; Intranet; PAIS; Methodology; Process Analysis; Event Logs.*

I. INTRODUCCIÓN

Las tecnologías de información se han convertido en un recurso indispensable en las operaciones diarias de una organización. Con ello, se tiene el desafío de contar con aplicaciones de software que funcionen de manera integrada y en donde se pueda almacenar información relacionada con los procesos organizacionales que se ejecutan.

Sin embargo, la información que se registra sobre la ejecución de los procesos no siempre se analiza más a fondo o se utiliza para generar oportunidades de mejora que impacten en la calidad de sus procesos.

El poder conocer mejor los procesos de negocio se convierte en una oportunidad que pueden aprovechar las empresas para trabajar mejor sus procesos de acuerdo a las necesidades reales que se tienen. Para diseñar y trabajar con procesos de negocio, las organizaciones pueden recurrir a lo que se conoce como *Business Process Management* (BPM). BPM es una metodología en donde se integran un conjunto de herramientas de software, técnicas, métodos y disciplinas de gestión que se encargan de administrar y mejorar los procesos de negocio [1]. Adoptar la implementación de esta metodología corporativa en

una organización ha traído grandes beneficios en lo que se refiere a la gestión de procesos según lo reportado en [2].

La minería de procesos (*process mining*) ha emergido como una nueva disciplina que integra la inteligencia de máquina, la minería de datos, el modelamiento y el análisis de procesos. Debido a ello, es posible llegar a descubrir, supervisar y mejorar procesos de negocios de una organización mediante la extracción de conocimiento desde registros de eventos (conocidos como *logs* de eventos) que se encuentran en los sistemas de información corporativos [3]. Mediante los *logs* de eventos se puede realizar el descubrimiento de modelos y además analizar cuál es la cercanía que tienen con respecto a cómo se ejecuta un proceso en la realidad versus lo que se tiene definido en los modelos preexistentes. Existen algunas técnicas para realizar minería en los procesos de negocios, tales como las presentadas en [4]. Estas técnicas están relacionadas con algoritmos genéticos y el análisis de tipo clúster, las que consideran el flujo de trabajo para detectar imperfecciones en el diseño del proceso en [5], o las técnicas relacionadas con la extracción de modelos a partir de *logs* de flujos de trabajo y representarlos mediante las redes de Petri [6].

La minería de procesos ofrece una serie de perspectivas que facultan la posibilidad de realizar un análisis amplio dentro del contexto de una organización. Esta posibilidad permite no solo realizar el descubrimiento de un proceso, sino que también proporciona la oportunidad de ejecutar un análisis del tipo organizacional, temporal, y de caso [7].

De acuerdo al contexto del negocio, es posible a través de BPM llegar a modelar un proceso determinado por medio de la identificación de actividades del proceso y roles presentes; no obstante, con el uso de técnicas de minería de procesos se puede descubrir cómo se lleva a cabo el proceso realmente, debido a que la información que se almacena en un *log* de eventos contiene cada una de las actividades del proceso.

Con la existencia de sistemas de información atentos al proceso, llamados PAIS (*Process-Aware Information Systems*) [8], es posible extraer *logs* de eventos y luego aplicar técnicas de minería de procesos enfocadas a descubrir un modelo de proceso, realizar un análisis de conformidad para comparar el modelo real versus un modelo ideal, o bien un análisis social desde la perspectiva organizacional. El tener esta posibilidad

de investigación, permite a los encargados de los procesos de negocio tener a disposición una serie de herramientas para poder revisar detalladamente el comportamiento de los procesos y así poder proponer mejoras a estos. Es importante también indicar que además de los PAIS, también se han realizado casos de estudio con sistemas legados que no registran *logs* de eventos, permitiendo de esta forma también la aplicación de minería de procesos [9].

Microsoft SharePoint Server es una herramienta colaborativa creada por Microsoft, la cual brinda un entorno de colaboración empresarial que utiliza navegadores web y puede ser utilizado por organizaciones que desean mejorar la eficacia de sus procesos organizacionales. SharePoint tiene un conjunto de productos y aplicaciones de software que involucran distintas tecnologías Microsoft (ej: herramientas de Office, herramientas comunicación como Lync, y también herramientas en el campo de inteligencia de negocios y desarrollo de aplicaciones. SharePoint Server tiene la facilidad de generar *logs* de eventos que contienen información importante para estudiar procesos organizacionales y realizar un análisis más profundo sobre los datos y el comportamiento del proceso. Estos *logs* no solo son informativos, también contienen datos relacionados con procesos computacionales, eventos como las advertencias, los problemas del sistema operativo y/o de aplicaciones. Este tipo de *logs* han sido utilizados en ambientes de espacios colaborativos y redes sociales, cuyo enfoque principal está en analizar solamente el comportamiento de los usuarios [10].

Por medio de esta tecnología Microsoft, es posible tener información oportuna sobre procesos en particular y sus recursos. Por ejemplo, la información del flujo de trabajo de una intranet es almacenada en los servidores de la plataforma. A nivel de negocio, dicha información podría no estar siendo analizada del todo por parte del administrador de la herramienta, por lo que existe la oportunidad de poder analizar los datos que se están registrando en la intranet y poder aplicar distintas perspectivas de minería de procesos.

En este trabajo se crea una metodología que utiliza distintas técnicas y herramientas de minería de procesos para analizar información acerca de procesos computacionales que fue extraída de servidores de una intranet corporativa. Para probar la metodología, se realiza un caso de estudio en el cual se tiene acceso a la información del proceso de soporte de un cliente de tipo *Retail*. Esta intranet se implementó en la plataforma de colaboración de SharePoint Server, y forma parte de un proyecto de consultoría que brinda una empresa consultora de tecnología que ofrece este tipo de servicios en Chile y Costa Rica. Además, se realiza la exploración de la información relacionada con dos clientes ubicados en el área de seguridad laboral a nivel de prevención de riesgos y enfermedades en el trabajo, en donde se tiene implementada también un intranet.

El artículo está organizado de la siguiente forma. En la sección 2 se introducen los antecedentes con conceptos de

administración de procesos de negocio y de minería de procesos. La sección 3 se enfoca en mostrar las principales herramientas disponibles para realizar minería de procesos. En la sección 4 se establece la propuesta de investigación definida por los autores, mientras que la sección 5 muestra la metodología de trabajo PMSLA (*Process Mining SharePoint Log Analysis*) definida en este estudio. La sección 6 muestra la discusión de los resultados generados en el caso de estudio propuesto y además se hace una comparación de los datos del log de eventos asociado al cliente *Retail* junto con los dos clientes en el área de seguridad laboral. La sección 7 muestra las conclusiones obtenidas y la sección 8 el trabajo futuro.

II. ANTECEDENTES

Actualmente, los sistemas de información pueden apoyar los procesos de negocio de cualquier organización, considerando información relacionada con el proceso y los recursos que participan en su ejecución. Existe un área de investigación que ha emergido para la gestión de los procesos de negocio y que aporta técnicas y herramientas para el análisis y la generación de conocimiento a partir de registros de eventos. Esta área se le conoce como minería de procesos.

A continuación se introducen dos conceptos importantes que están presentes en el desarrollo de este artículo. El primero de ellos es la administración de procesos de negocio y el segundo relacionado con la disciplina de minería de procesos.

A. Administración de procesos de negocio

Según Weske, los procesos de negocio consisten en una serie de actividades que se realizan de manera coordinada para completar un objetivo de negocio [1]. Para alcanzar su efectiva realización, estas deben estar incorporadas en un marco organizacional y tecnológico que permita su ejecución y administración, de manera que no solo faculte la ejecución de los procesos productivos de la organización, sino que brinde la posibilidad de administrar los recursos participantes en cada una de las etapas requeridas.

En su libro de Administración de los Procesos de Negocio [1], Weske indica que este proceso de vida es cíclico e incluye conceptos, métodos y técnicas para soportar las etapas del proceso de administración de los procesos. Estas fases se pueden observar en la figura 1.

Según Harvey, este ciclo de vida brinda una serie de ventajas competitivas a una organización o empresa que realice una correcta administración de los procesos de negocio. Dentro de estas, se encuentran las siguientes [11]:

- Identifica mejoras puntuales en los procesos.
- Formaliza procesos existentes.
- La adopción de la administración de los procesos de negocio potencia a las organizaciones a pensar y formalizar su entendimiento de los procesos propios de negocio.
- Facilita el flujo eficiente y automatizado de los procesos de negocio.

- Aumenta la productividad, disminuyendo la cantidad de recursos.
- Agiliza los procesos de cumplimiento de regulaciones.

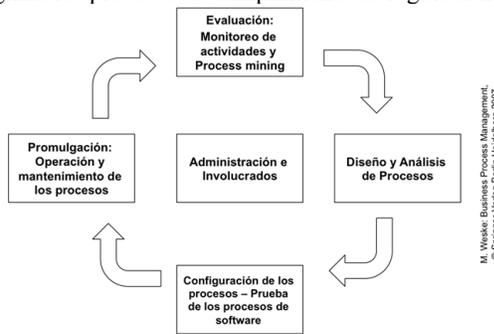


Figura 1. Ciclo de vida de los procesos de negocio

Durante las últimas décadas se han creado una serie de distintos estándares que incluyen notaciones simbólicas para la definición de los procesos de negocio, permitiendo el correcto entendimiento y claridad de las actividades y los recursos involucrados en un proceso de negocio. Actualmente, los más utilizados corresponden a: *Business Process Management Notation* (BPMN) [12], *Yet Another Workflow Language* (YAWL) [13], Redes de Petri [14], *Workflow Language* [15], entre otros.

B. Minería de procesos

Ubicada como un área de investigación relativamente nueva, la minería de procesos es conocida con una disciplina que busca extraer información que se encuentra almacenada en los registros de eventos de los sistemas de información actuales [3].

Esta disciplina permite realizar descubrimiento de procesos, ejecutar un monitoreo de estos, así como exponer propuestas de mejora fundamentadas en el análisis que se realiza al extraer conocimiento a partir de la información que se encuentra almacenada en los registros o *logs* de eventos disponibles en los sistemas de información actuales [16]. Se considera que dentro de esta información es posible guardar datos tales como: cada evento referido con una actividad, cada evento referido a un caso o instancia del proceso, información sobre la persona que ejecuta o inicia una actividad del proceso, así como un registro de tiempo o *timestamp* [3].

La minería de procesos se enfoca en realizar su análisis fundamentado en tres tipos principales: (1) el descubrimiento, (2) la verificación de conformidad, y (3) el mejoramiento de procesos. En el descubrimiento, a partir de un registro de eventos se puede obtener como resultado un modelo del proceso. En el caso del segundo tipo, con la verificación de conformidad se puede comparar un modelo de proceso existente contra un registro de eventos del mismo proceso. Al hacer esto, se puede chequear si la realidad a como está guardada en el registro de eventos, es equivalente al modelo y viceversa, mostrando las diferencias y elementos en común entre un modelo y un registro de eventos. El tercer tipo de

minería es el mejoramiento. En este, la idea es extender o mejorar un modelo de proceso existente usando la información real acerca del proceso que está almacenada en algún registro de eventos [17].

Complementariamente, es posible implementar el análisis de minería de procesos de acuerdo a cuatro perspectivas, a saber: (a) la perspectiva de control de flujo, (b) la perspectiva organizacional, (c) la perspectiva de caso y (d) la perspectiva temporal [13]. En el caso de la perspectiva de control de flujo, esta se enfoca en determinar orden en el que se ejecutan las actividades. La perspectiva organizacional describe quién o quiénes están involucrados en el desempeño de las actividades de un proceso y cómo están relacionados entre sí (roles, redes sociales, etc.). La perspectiva de caso permite mostrar las características que identifican a cada caso del proceso y de esta manera poder determinar similitudes y diferencias entre cada uno de ellos. Finalmente, en la perspectiva temporal se busca analizar cuándo ocurren las actividades dentro del proceso, medir la frecuencia con que suceden y poder determinar cuellos de botella, entre otros.

Existen una serie de algoritmos que han sido creados para realizar el análisis por medio de la minería de procesos. Algoritmos como el α (Alpha) [5], el *Heuristic miner* [18], el *Genetic* [19], o el *Fuzzy Mining* [20], son ejemplos de rutinas que permiten realizar el descubrimiento de procesos. Otros algoritmos facilitan el análisis desde la perspectiva organizacional por medio de los métricas conocidas como: *handover of work*, *subcontracting*, *working together* y *doing similar tasks*. Estas métricas permiten realizar una clasificación de personas en términos de roles o información organizacional, además de la interacción que existe entre las personas que llevan a cabo las actividades en un proceso [21] [22].

Por medio de herramientas enfocadas a apoyar las labores de la minería de procesos, es posible construir *logs* de eventos, hacer un estudio detallado de ellos, así como desplegar resultados producto de ese análisis. Como ejemplos de estas herramientas se pueden mencionar, la herramienta de código abierto llamada ProM [23], la aplicación comercial DISCO [24], y una nueva iniciativa de desarrollo para análisis de procesos basada en Python llamada PMLab, la cual está en desarrollo. En la siguiente sección se muestra una reseña de las principales herramientas.

La aplicación de minería de procesos utilizando información obtenida a partir de SharePoint no es un dominio que haya sido analizado ampliamente. Se registra un trabajo previo [25], donde se utilizó la herramienta llamada Nitro para manejar archivos en formato basado en XES y aplicar minería de procesos. Sin embargo, este trabajo no revela la creación de ninguna metodología como la presentada en esta propuesta. La oportunidad que surge hoy día para implementar análisis mediante minería de procesos se ve alimentada por la gran cantidad de datos que se encuentran almacenados en los

sistemas de información en las compañías. Una realidad es que no todas las organizaciones aprovechan su información *post-mortem* para impulsar mejoras a sus procesos actuales.

III. HERRAMIENTAS DISPONIBLES

Existe un grupo de herramientas mediante las cuales se puede realizar la exploración y la minería de procesos como tal. La selección de una herramienta en particular dependerá mucho de las necesidades que una persona o empresa tenga con respecto al análisis de sus procesos y las preguntas que desee llegar a responder.

Entre las herramientas más conocidas están:

- CELONIS¹: es una herramienta comercial con funcionalidades enfocadas en el monitoreo y la auditoría continua de procesos. Brinda conexión directa a bases de datos de escogencia del usuario. Permite el análisis de información a distintos niveles y posee capacidades de inteligencia de negocios, uso de indicadores claves de desempeño, análisis estadísticos y distintas opciones para desplegar resultados por medio de “dashboards”.
- DISCO²: es considerada como una solución bastante completa, fácil de usar y muy amigable para el usuario. Para quienes inician en minería de procesos, DISCO se convierte en una de las mejores alternativas de aprendizaje. Ofrece una serie de algoritmos que se consideran eficientes y rápidos, junto con tener buenas capacidades para administrar, realizar filtros en los *logs* de eventos y brindar buenas características para visualizar las salidas producto del análisis. Tiene una modalidad de licencia académica, adicional a la de tipo comercial.
- ProM³: en sus inicios fue considerado como un *framework* para realizar minería de procesos, y hoy día se valora como una herramienta con amplias capacidades para realizar análisis de procesos de negocio [26]. ProM ofrece una arquitectura escalable, la cual permite la incorporación de *plug-ins* desarrollados por los creadores de la herramienta o usuarios de la misma, como un beneficio de ser una aplicación de código abierto. Según un estudio realizado sobre las técnicas y herramientas de minería de procesos [27], ProM es catalogada como la herramienta de mayor uso por parte de usuarios que practican la profesión. La versión actual es la 6.3.
- QPR / Process Analyzer⁴: es una herramienta comercial que permite la integración con MS Excel (parte del *front-end* para visualización) y con MS-SQL (almacenamiento de los datos). Es una solución que combina el análisis de procesos de negocio, la arquitectura empresarial y la administración del desempeño.
- XESame: es una aplicación que permite realizar la extracción de un log de eventos a partir de fuentes de datos

¹ Para más información referirse a: www.celonis.de/en/

² Para más información referirse a <https://fluxicon.com/>

³ Para más información referirse a: <http://www.promtools.org/prom6/>

⁴ Para más información referirse a: www.qpr.com

diversas (ej: archivos de texto, tablas de bases de datos, etc.), donde permite especificar cómo el *log* de eventos debe ser extraído y convertido a formatos compatibles para minería de procesos como lo son el XES o MXML. La versión 1.2 de XESame se encuentra incluida a partir de la versión 6.2 de ProM.

Con la existencia de herramientas como las anteriores, es posible implementar análisis por medio de la minería de procesos. Cabe mencionar que indistintamente de la o las herramientas seleccionadas, los pasos para llevar a cabo la extracción y preparación de los datos representan un claro desafío para cualquier proyecto que se quiera efectuar.

IV. PROPUESTA DE INVESTIGACIÓN

La propuesta de esta investigación busca considerar el comportamiento de distintos *logs* de eventos de actividades de soporte de una intranet creada mediante tecnología de Microsoft SharePoint Server 2010 Enterprise Edition SP2.

Se pretende que este trabajo incorpore un caso de estudio que permita aplicar la metodología PMSLA (explicada en la siguiente sección) en un cliente de tipo *Retail* para descubrir y entender el proceso soportado por la intranet y poder identificar las actividades principales que se ejecutan. Para ello, se propone utilizar al menos dos algoritmos de descubrimiento de minería de procesos con la intención de analizar sus resultados y determinar cuál de ellos puede ser el más adecuado de acuerdo al contexto de negocio que se está analizando. Además, otro de los objetivos planteados está relacionado con la aplicación de cuatro métricas relacionadas con la perspectiva organizacional, con el fin de descubrir información acerca del comportamiento de los recursos de SharePoint Server en la intranet y las relaciones que existen entre ellos.

Adicionalmente, se propone utilizar también la metodología PMSLA para el análisis de la intranet de dos tipos de clientes adicionales. Se desea llegar a analizar al menos las características principales de estos dos registros de eventos que corresponden a otras intranets implementadas bajo la misma tecnología y que se encuentran en uso actualmente.

V. METODOLOGÍA PMSLA

Para llevar a cabo esta investigación se consideró crear la metodología PMSLA (*Process Mining SharePoint Log Analysis*), para la cual se establecen una serie de tareas por realizar. Primero, se realiza la recolección de datos de las fuentes, es decir, obtener por medio de los sistemas de información los *logs* registrados de las actividades de los usuarios. Posterior a esto, se construye el *log* de eventos y se realiza una inspección del mismo, con el objetivo de clarificar y entender las características básicas del proceso. Luego de inspeccionar el log se procede a la aplicación de las técnicas, herramientas y algoritmos de minería de procesos, que generarán los resultados finales. La figura 2 muestra un diagrama que resume la metodología PMSLA.

Los pasos iniciales de la metodología PMSLA consisten en la extracción de los datos y la generación del log para el caso de

estudio seleccionado. Las tareas realizadas se detallan a continuación.

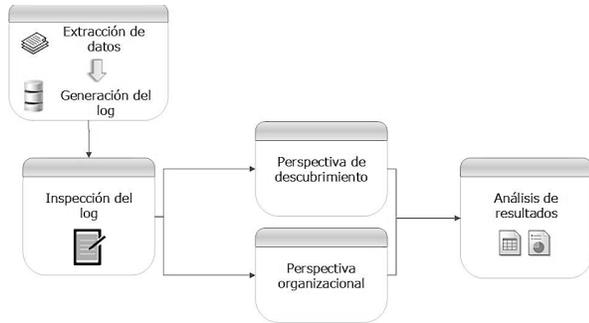


Figura 2. Diagrama de la metodología PMSLA

1. Extraer los datos de la intranet de SharePoint Server del cliente de tipo *Retail* de una consultora de tecnologías chilena – costarricense.
2. Guardar los datos en formato CSV y abrirlos en la herramienta MS Excel.
3. Se identifican las columnas asociadas al *log* y se les asocia un nombre distintivo a lo que representan. Seguidamente, se indica cada una de las columnas y su significado asociado:
 - a. *Timestamp*: almacena el dato del tiempo exacto en que se incluyó el evento en el *log*.
 - b. *Process*: corresponde al proceso que se está ejecutando. Cada proceso puede ejecutarse en distintos hilos o casos.
 - c. *TID*: es el *Thread ID* o identificador de cada hilo o caso incluido en el *log*.
 - d. *Area*: contiene al área específica que ejecuta el hilo o caso. Por ejemplo, se tienen el *SharePoint Server* o el *SharePoint Translation Services*.
 - e. *Category*: se almacena el dato que corresponde al tipo de hilo o caso que se ejecuta. Por ejemplo, se presentan el *Database* o el *Monitoring*.
 - f. *EventID*: es el identificador del evento asociado a la categoría.
 - g. *Level*: corresponde a la criticidad del hilo o caso ejecutado.
 - h. *Message*: almacena la descripción del evento ejecutado.
4. Antes de construir el *log* de eventos, se realizó la eliminación del símbolo asterisco (*) que se encontraba ubicado en varias de las fechas (este asterisco provenía en el archivo original de extracción). Para eliminarlo se utilizó la aplicación NOTEPAD ++, en donde se reemplazó el asterisco que se encontraba al final de la información de fecha por un espacio vacío. Este cambio era necesario de aplicar, debido a que impactaba en el formato de la columna del *Timestamp*.
5. Una vez realizados los pasos anteriores, se procedió a generar el *log* de eventos. Se definió como actividad única el *TID*, que indica cada caso o hilo ejecutado y almacenado en el *log*. Nótese que no se eliminan las

columnas individuales, pues estas brindan información necesaria para el análisis.

Un fragmento del *log* de eventos final utilizado para el proceso de minería de proceso se muestra en la figura 3.

Timestamp	Process	TID	Area
03/04/2014 11:24:21.69	mssearch.exe (0x0894)	0x1CE0	SharePoint Server
03/04/2014 11:24:21.69	mssearch.exe (0x0894)	0x1CE0	SharePoint Server
03/04/2014 11:24:21.69	mssearch.exe (0x0894)	0x1CE0	SharePoint Server
03/04/2014 11:24:21.69	mssearch.exe (0x0894)	0x1CE0	SharePoint Server
03/04/2014 11:24:21.69	mssearch.exe (0x0894)	0x1CE0	SharePoint Server Search
03/04/2014 11:24:21.69	mssearch.exe (0x0894)	0x1CE0	SharePoint Server Search
03/04/2014 11:24:21.69	mssearch.exe (0x0894)	0x1CE0	SharePoint Server Search
03/04/2014 11:24:21.69	mssearch.exe (0x0894)	0x1CE0	SharePoint Server Search
03/04/2014 11:24:21.70	mssearch.exe (0x0894)	0x1CE0	SharePoint Server
03/04/2014 11:24:21.70	mssearch.exe (0x0894)	0x1CE0	SharePoint Server

Figura 3. Fragmento del *log* de eventos final utilizado para el análisis y minería

6. Este *log* incluye una muestra de las transacciones del *log* de la intranet de SharePoint Server de un periodo de tiempo de casi dos horas (desde 11:24 am a las 13:01 pm del 4 de marzo del 2014). El total de datos recopilados corresponden a 100511 líneas de registro.
7. Como nota aclaratoria, hay que indicar que para el formato de fecha (*Timestamp*) se utilizó Mm/dd/yyyy hh:mm:ss.SS, que es el que resulta correcto de aplicar de acuerdo con la información contenida en el *log*.

Una vez ejecutados los pasos anteriores, se cuenta con el *log* de eventos listo para continuar aplicando la metodología. Con el *log* ya generado, se realiza una inspección o diagnóstico inicial del proceso por medio de la herramienta DISCO, incluyendo un análisis cualitativo y cuantitativo de los datos asociados al proceso (actividades, tiempos de ejecución, recursos y casos ejecutados).

Comprendidas las características básicas del proceso, se pasa a ejecutar la perspectiva de descubrimiento del modelo del proceso y la perspectiva organizacional, utilizando los algoritmos y técnicas incluidas en la herramienta ProM 5.2 y 6.3. Con la información resultante de estos análisis, se muestra un resumen del proceso, el modelo descubierto y sus principales características, un análisis organizacional y posibles mejoras que se derivan de los resultados.

Con respecto a la metodología PMSLA es importante indicar que algunos de sus pasos se podrían automatizar, buscando hacer más eficiente el proceso. Los principales pasos de la metodología que se pueden automatizar corresponden a la extracción de los datos, la generación del *log* de eventos y la aplicación de las técnicas de minería de procesos.

VI. ANÁLISIS DE RESULTADOS

Como se mencionó, se tuvo acceso a un *log* de eventos de soporte de una intranet implementada en SharePoint Server. La información del *log* corresponde al monitoreo que se lleva del correcto funcionamiento de la intranet y al momento que surja cualquier actividad del sistema o se produzca un error, se genera un registro del evento con atributos como la fecha y hora del evento, la descripción del proceso, el ID del evento o *TID* (*Thread ID*), el área donde se ejecuta, entre otros.

Análisis con DISCO

Posterior a la obtención del log de eventos se procede al paso de inspección del log de eventos, el cual se realiza con la ayuda de la aplicación DISCO. Se analizó el *log* correspondiente al registro de eventos del cliente *Retail*. Aplicando los pasos de la metodología PMSLA, se logra obtener la información que se muestra en la tabla I.

TABLA I. DESCRIPCIÓN INICIAL DEL LOG DE RETAIL

Característica	Log de Cliente <i>Retail</i>
Eventos	100,511
Variaciones	217
Casos totales	983
Actividades	62
Recursos participantes	14
Tiempo promedio de ejecución por caso	45.6

El modelo inicial completo que se genera del proceso mediante la herramienta DISCO es el siguiente:

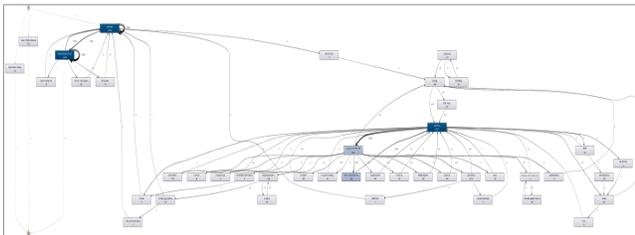


Figura 4. Modelo inicial del proceso del cliente Retail

A pesar de ser un modelo amplio, es importante destacar las tres actividades que se ejecutan con mayor frecuencia en el *log* (en color azul). A la izquierda de la figura están: *Database* (24.74%) y *Crawler: Gatherer Plugin* (23.21%), mientras que en el centro de la figura está *Monitoring* (21.58%).

Con la herramienta DISCO, fue posible conocer el comportamiento de las actividades del *log* con respecto al nivel de prioridad de estas. Los niveles que más se presentaron en el periodo del análisis fueron el nivel medio con un 54.87% y el nivel alto con un 44.99%. El que menos se presentó fue el nivel crítico con 0%.

Asimismo, se consideró la característica de los recursos involucrados en la ejecución de los distintos casos. Como resultado, el *SharePoint Foundation* es a nivel de la intranet el recurso que más se utiliza (37.34%), seguido del *SharePoint Server Search* (25.55%), y el *SharePoint Server* (23.61%). En tanto, los recursos que menos participación en el rango de tiempo analizado fueron el *InfoPath Forms Services*, el *Access Services* y *Access Services 2010*, los tres con 0%.

Seguidamente, mediante la herramienta *ProM* en sus versiones 5.2 y 6.3 se amplía el análisis del modelo obtenido desde la perspectiva de descubrimiento y la perspectiva organizacional.

Descubrimiento con *ProM*

Cuando ya se posee una mayor claridad de las características del *log* de eventos se procede al paso de ejecutar la perspectiva de descubrimiento, el cual se realiza con las técnicas y algoritmos implementados en la herramienta *ProM*.

A continuación, se explica el detalle del análisis realizado en la herramienta *ProM*. Primero se muestra el descubrimiento realizado mediante el algoritmo *Heuristic Mining* y seguidamente, se muestra lo descubierto al aplicar el algoritmo *Genetic Mining*.

Heuristic Mining

Debido a la complejidad y tamaño al descubrir el proceso completo (ver figura 4), se utilizó la técnica de descomposición para dividir en dos subprocesos (mediante el filtro de actividades en *ProM* y DISCO). El primer subproceso inicia desde que se gatilla con cualquier evento que hace referencia a la base de datos de la intranet hasta que inicia el monitoreo. El segundo subproceso abarca el monitoreo y las actividades que se generan a partir de su ejecución.

En el primer subproceso están involucradas las siguientes actividades:

- *Database, Crawler: Gatherer Plugin, Monitoring, Topology, Search Platform Services, FsPlugin indexing tasks, Micro Trace, Feed Cache, Microfeeds, Report Server Catalog, Search Component, Configuration, Web Controls.*

En el segundo subproceso se encuentran las siguientes actividades:

- *Start, Logging Correlation Data, Personal Site Instantiation, Taxonomy, User Profiles, Site Management, Administration, Exchange Task Synchronization, Work Management Service, App Deployment, Publishing, Translation, End, Analytics, Linguistic Processing, General, Usage Infrastructure, Unified Logging Service, Timer Job, Tracing Controller Service, Discovery, Query, Timer, Analysis Engine, Social Data, App Marketplace, DistributedCache, Health, State Service, Crawler: Content Plugin, FastServer log, Authentication Authorization, Claims Authentication, Files, Cobalt, Upgrade, Object Cache, Template Cache, Web Parts, Runtime, Asp Runtime, Performance, Data Source Control, WOPI, Publishing Cache.*

El hacerlo de esta manera permite mostrar de una forma más clara el proceso respectivo. El primer bloque incorpora un componente importante de clasificación del tipo de evento ocurrido en la intranet y en el segundo bloque, los componentes más importantes del monitoreo y ejecución de tareas asociadas a las características respectivas del evento.

Para ejecutar el *Heuristic Mining* se consideró la siguiente configuración utilizando *ProM 5.2*.

TABLA II. CONFIGURACIÓN *HEURISTIC MINING PROM 5.2*

Característica	Valor
<i>Relative-to-best threshold</i>	0.05
<i>Positive observations</i>	10
<i>Dependency threshold</i>	0.9
<i>Lenght-one-loops threshold</i>	0.9
<i>Lenght-two-loops threshold</i>	0.9
<i>Long distance threshold</i>	0.9
<i>Dependency divisor</i>	1
<i>AND threshold</i>	0.1

Modelo del primer subproceso

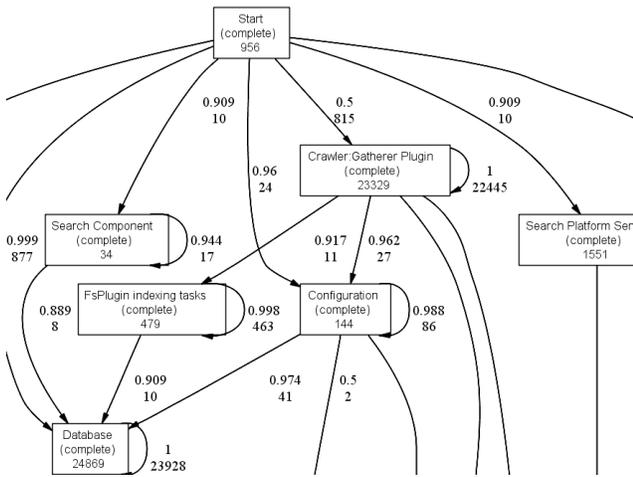


Figura 5. Porción del modelo del proceso descubierto con Heuristic Miner

Las actividades inician y se dirigen principalmente hacia a la actividad de *DataBase*. Llegan a esta de diferente manera: de forma directa, provenientes del *Search Component* y por medio del *Crawler: Gatherer Plugin* a través de dos tareas de *FS Plugin Indexing* y de *Configuration*.

Posterior a la actividad de *DataBase*, se llega a la actividad de *Monitoring*, que es una de las actividades más ejecutadas en este proceso. A ella se llega posterior al *DataBase*, o de forma directa desde el comienzo del proceso. *Monitoring*, al igual que *DataBase*, corresponden a actividades finales de este subproceso 1. Aparte, existen otras actividades que finalizan este subproceso como lo son: *Report Server Catalog*, *Search Plataforma Services*, *Crawler: Gatherer Plugin*.

Lo más relevante de esta partición es que la mayor parte de los casos antes de terminar su participación en este subproceso incluyen las actividades de *DataBase*, *Monitoring* o *Crawler: Gatherer Plugin*.

Modelo del segundo subproceso

Debido a la complejidad del segundo modelo y al tamaño del mismo, se generan dos sub-modelos (porciones) donde sólo se muestran las actividades que poseen mayor participación. La figura 6 muestra la porción 1 del modelo del segundo subproceso:

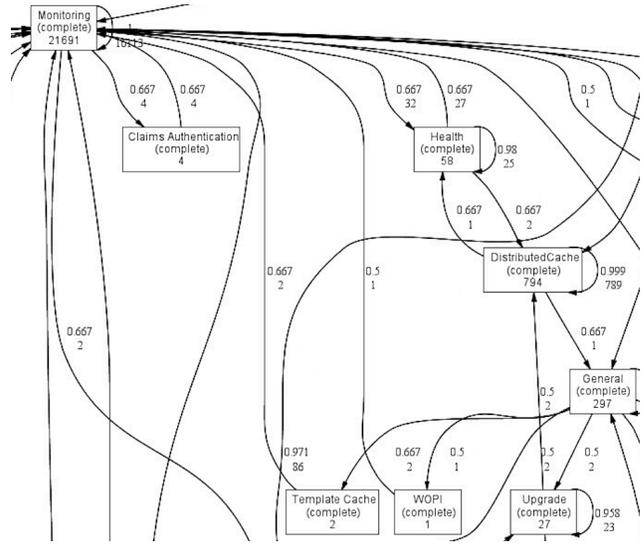


Figura 6. Porción 1 del modelo del segundo subproceso descubierto con Heuristic Miner

- Se puede apreciar que la actividad de *Monitoring* es un componente principal del proceso (se ejecuta en 21691 casos). Esto se debe a que se encarga de vincular la primera parte del proceso con la segunda y además tiene el rol de hacer el monitoreo de los servicios que llaman a los procesos y además de velar por el estado de la plataforma.
- La tarea de *Monitoring* analiza las actividades monitoreadas y procede a realizar las acciones correspondientes. Se relaciona con la tarea *Logging Correlation Data* (que se ejecuta un total de 9847 veces), cuya funcionalidad se centra en manejar la información de ID's de cada elemento en el *log* para un análisis posterior. El detalle de esta actividad se muestra en la porción 2 del modelo del segundo subproceso en la figura 7.

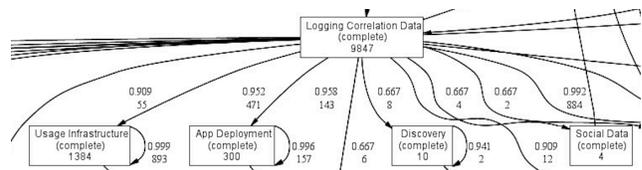


Figura 7. Porción 2 del modelo del subproceso descubierto con Heuristic Miner

En esta parte de la porción 2 del segundo subproceso se puede apreciar que:

- El *Logging Correlation Data* es la segunda actividad más importante del subproceso, y se encarga de recibir datos provenientes principalmente de la actividad del *Monitoring*

y dispara actividades asociadas a las características del monitoreo según corresponda.

- Por ejemplo, algunas de las actividades que se disparan después de realizar el *Logging Correlation Data* son el *Translation*, el *App Marketplace* y el *Discovery*.
- Una vez que se lleva a cabo las actividades de monitoreo, el subproceso se termina, y por ende el proceso en general.

El algoritmo *Heuristic Mining* muestra que una vez que se ejecuta el monitoreo de actividades se pasa al evento *Logging*, el cual se encarga de almacenar por medio de actividades específicas los datos requeridos en el *log*.

Ventajas de utilizar Heuristic Mining

- Menor tiempo de procesamiento.
- Modelo generado es menos complejo de evaluar.
- Modelo es más entendible y representa el proceso de una forma más completa y real con respecto a un procedimiento habitual de registro de eventos.
- El modelo descubierto permite de forma simple, analizar sus actividades, además de las relaciones entre ellas.

Desventajas de utilizar Heuristic Mining

- No brinda información del ajuste (medida que indica cuantos casos del log de eventos permite ejecutar el modelo) de forma directa.
- De igual forma, genera un modelo grande y complejo de evaluar en primera instancia.
- Puede ser necesario recurrir a otros algoritmos u herramientas (DISCO o ProM), para poder filtrar el log en dos subprocesos que permitieran descomponer el proceso global en dos partes para realizar un mejor análisis.

Genetic Mining

Tal y como se hizo con el algoritmo *Heuristic Mining*, debido a la complejidad de descubrir el proceso completo, también se utilizó en el *Genetic Mining* la misma descomposición de dos subprocesos antes explicada.

Para ejecutar el *Genetic Mining* se consideró la configuración de la tabla III utilizando ProM 5.2. La figura 8 muestra el modelo resultante del primer subproceso.

TABLA III. CONFIGURACIÓN *GENETIC MINING* *PROM* 5.2

Característica	Valor	Característica	Valor
Population size	25	Show Advanced Fitness Parameters	
Initial population type	Possible Duplicates	Use genetic operators	✓
Max number generations	10	Selection method type	Tournament 5
Seed	1	Crossover type	Enhanced
Power value	1	Crossover rate	0.8
Elitism rate	0.02	Mutation type	Enhanced
Fitness type	ExtraBehaviorPunishment	Mutation rate	0.2

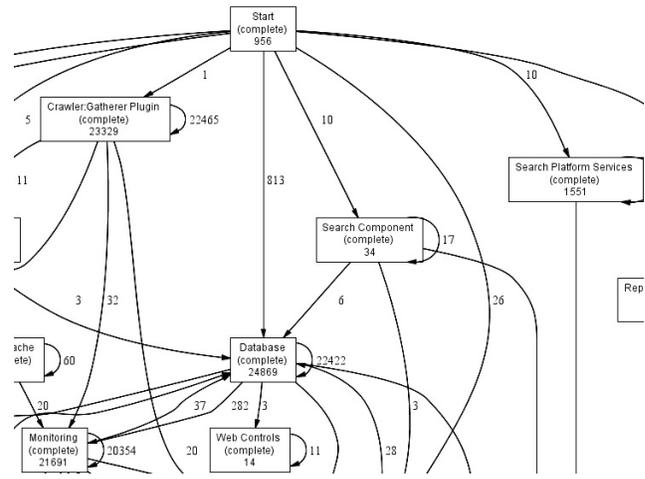


Figura 8. Porción del modelo del primer subproceso descubierto con *Genetic Miner*

A partir de esto, se pueden analizar algunas características que el modelo del primer subproceso brinda:

- Existen tres actividades principales que se ejecutan en la mayoría de los casos del subproceso 1, estas actividades corresponden al *Database*, al *Crawler: Gatherer Plugin* y al *Monitoring*. Estas corresponden a procesos de acceso a la base de datos, a la indexación de contenidos y generación de palabras claves, así como el monitoreo. Es importante recalcar que la actividad de monitoreo es la que da inicio al segundo subproceso.
- Otra actividad que se ejecuta en 1551 casos es la de *Search Platform Services*, que es un proceso del servicio principal de búsqueda.
- Este modelo no deja ciclos ni actividades aisladas, por lo tanto, se puede considerar que corresponde a un buen modelo para representar el primer subproceso que se centra en clasificar las actividades monitoreadas según su tipo, para realizar tareas específicas en el segundo subproceso.

Después de analizar el primer modelo, en el segundo subproceso (figura 9) se aprecia lo siguiente:

- Debido a la complejidad del segundo modelo y al tamaño del mismo, sólo se muestran las actividades que poseen mayor participación.
- Lo más relevante es que se inicia con la tarea de monitoreo que también se incluyó en el primer subproceso, lo que permite ver que el componente principal que vincula ambos modelos es esta tarea.
- La tarea de *Monitoring* se ejecuta en 21691 casos, siendo esta en el subproceso 2 la tarea más ejecutada. *Monitoring* se encarga de analizar las actividades monitoreadas y proceder a realizar las acciones correspondientes. Está vinculada con la segunda tarea más ejecutada del segundo subproceso, *Logging Correlation Data* (9874 veces), cuya

funcionalidad es almacenar los datos correspondientes a los ID's de cada elemento en el *log* para futuro análisis.

- Otras tareas importantes que se ejecutan en el subproceso 2 corresponden a *Personal Site Instantiation*, ejecutada 7002 veces y *User Profiles* ejecutada 3073.
- El resto de las actividades complementan el proceso de *logging* iniciado en el subproceso 1 y corresponden a tareas asociadas a *Administration*, *Site Management*, *Timer Jobs*, *Analytics*, *Work Management Services*, entre otras.
- Posterior a que se ejecutan las actividades para registrar en el *log* la información deseada, se termina el subproceso.

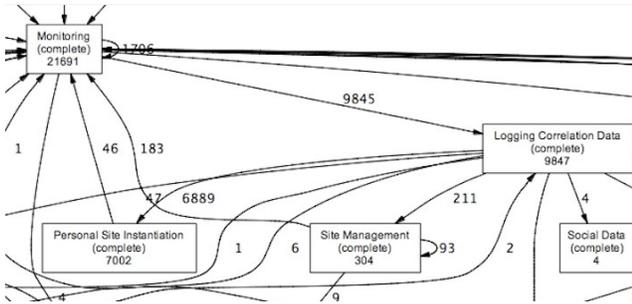


Figura 9. Porción 2 del modelo del segundo sub proceso descubierto con Genetic Miner

Ventajas de utilizar Genetic Mining

- Proporciona el ajuste asociado a cada modelo.
- Obtiene buenos modelos pero a un alto precio de procesamiento.

Desventajas de utilizar Genetic Mining

- Requiere alto procesamiento para obtener buenos modelos.
- Requiere definición de parámetros asociados a algoritmos genéticos (población inicial, generaciones, mutación, tipo de mutación), para obtener distintos modelos.
- El experimento se realizó con parámetros que incluyen una población inicial baja y pocas generaciones, lo cual, a pesar de tardar un tiempo significativo, no brinda modelos con un alto ajuste.

Análisis organizacional con ProM

Posterior, o en paralelo al paso de la ejecución de la perspectiva de descubrimiento, se procede a ejecutar la perspectiva organizacional. Para realizar el análisis organizacional en este caso, se utilizó el *log* completo de la intranet. Se consideraron para este análisis las siguientes métricas: *Handover of work*, *Working together*, *Doing similar task* y *Subcontracting*.

Handover of work

Esta métrica permite obtener la relación en la cual un recurso le pasa trabajo a otro recurso. Para el proceso de soporte de la intranet, se obtuvo la figura 10 utilizando ProM 6.3.

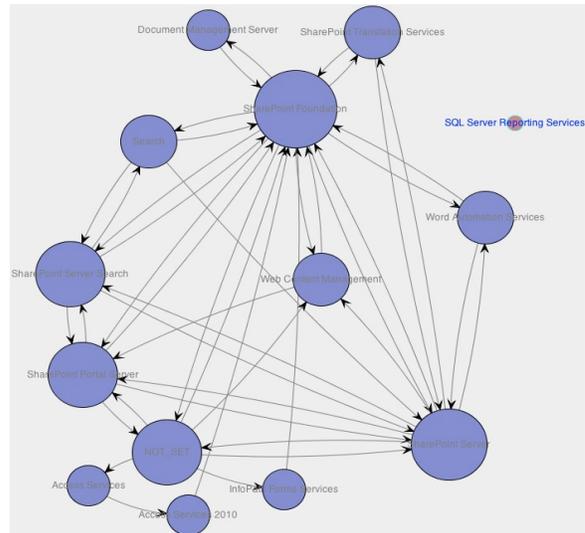


Figura 10. Diagrama de la métrica *Handover of work*

Es posible identificar que en el proceso de soporte los recursos que más reciben trabajo de otros recursos son el *SharePoint Server* y el *SharePoint Foundation*. Los recursos relacionados con *Access Services* e *InfoPath* tienen poca interacción con el resto de los recursos, mientras que el *SQL Server Reporting Services* no le pasa ni recibe trabajo de algún otro recurso.

Working together

La métrica del *Working together* es aquella que permite identificar qué tan frecuentemente dos individuos están trabajando en el mismo caso. Al aplicar esta métrica se genera la figura 11 utilizando ProM 6.3.

Se puede ver que hay una alta relación de recursos que trabajan juntos cuando se ejecuta un caso, pues prácticamente todos se involucran. El único recurso que no participa es el *SQL Server Reporting Services*, el cual no interactúa con ningún otro recurso debido a posiblemente no se encontraba activado en la intranet durante el momento de tomar la muestra de datos.

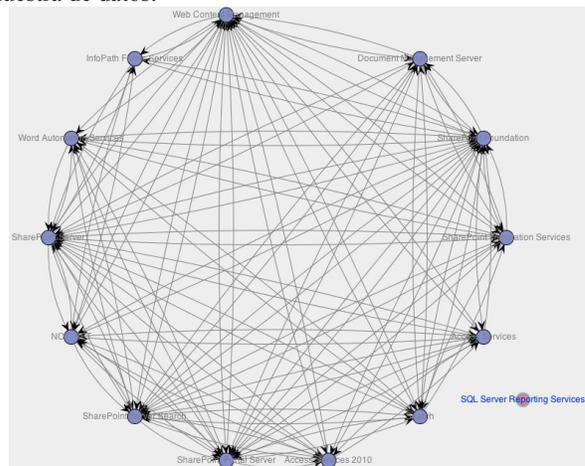


Figura 11. Diagrama de la métrica *Working together*

Doing similar tasks

La métrica del *Doing similar tasks* se enfoca en las actividades que hacen los recursos y permite determinar cuáles recursos realizan tareas similares, facilitando la identificación de roles. Al aplicar esta métrica se genera la figura 12 utilizando ProM 6.3. Se logran identificar 3 roles principales. El primero de ellos involucra al *SharePoint Server Search*, el *Access Services* y el *Access Services 2010*, que corresponden a las búsquedas y los servicios brindados por el Access Services.

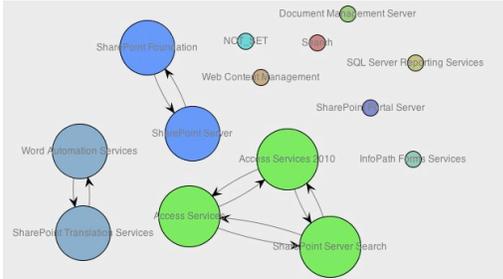


Figura 12. Diagrama de la métrica *Doing Similar Tasks*

El segundo rol (el más relevante) lo tienen el *SharePoint Foundation* y el *SharePoint Server*, que llevan a cabo actividades propias del *SharePoint*, como la administración, la configuración, el mantenimiento y uso de la herramienta. El tercer rol lo tienen los recursos de *Word Automation Services* junto con *SharePoint Translation Services* que se encargan de las actividades que tienen que ver con gestión de los archivos de MS Word que se utilizan en la intranet. Adicionalmente, quedan identificados 7 roles más que son desempeñados por cada uno de los recursos involucrados, donde es posible indicar que las actividades que estos recursos desempeñan solo pueden ser realizadas por ellos mismos. Por ejemplo, está el *Web Content Management*, que se encarga de administrar el contenido web de la intranet, y el *SQL Server Reporting Services* que tiene el rol de ayudar a crear, implementar y administrar informes para ser mostrados en la intranet corporativa.

Subcontracting

Esta métrica permite establecer la relación cuando un recurso le solicita a otro recurso la ejecución de alguna actividad presente en el caso. Por medio de la herramienta ProM 6.3 se obtiene la siguiente figura de subcontratación en el proceso.

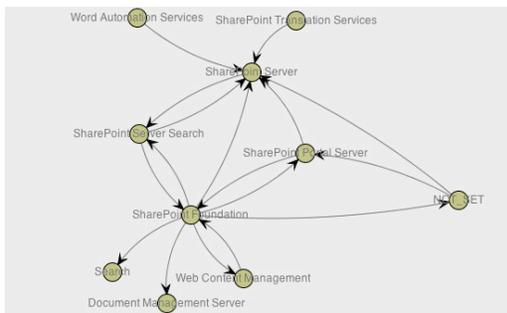


Figura 13. Diagrama de la métrica *Subcontracting*

A partir del diagrama generado, se identificaron ciertas relaciones de subcontratación. La primera de ellas se puede considerar a nivel de los recursos de *Word Automation Services* y *SharePoint Translation Services*, que subcontratan tareas monitoreadas por el *SharePoint Server*. A su vez, el *SharePoint Server* subcontrata tareas del *SharePoint Server Search* y recibe solicitudes de este mismo servicio de búsqueda junto con el *SharePoint Portal Server*, el *SharePoint Foundation* y del *Not Set* (conjunto de recursos no establecidos). El *SharePoint Foundation* tiene una relación en dos sentidos con las tareas de los recursos *SharePoint Server Search*, *SharePoint Portal Server* y del *Web Content Management*. Mientras tanto, solo tiene una relación de un sentido (solicitud de tareas) con el *Search* y al *Document Management Server*. Existen algunos recursos que no tienen ningún tipo de subcontratación de tareas y no se muestran en el diagrama. Ellos son: *Access Services*, *Access Services 2010*, *InfoPath Forms Services* y el *SQL Server Reporting Services*.

Con base en los diagramas y modelos generados en cada uno de los pasos que ejecutan las perspectivas de descubrimiento y organizacional, se realizó un análisis de los resultados, los cuales brindan información relevante al negocio acerca de su proceso y cómo sus recursos interactúan. Esto corresponde al paso final de la metodología, y en este artículo se incluyó el análisis en la misma sección en cada paso.

Extensión de la metodología PMSLA en logs de eventos de otros clientes

Una vez probada la metodología PMSLA con el caso de estudio del cliente *Retail*, se desea extender la aplicación de PMSLA a *logs* de distintos clientes que también poseen intranets implementadas en *SharePoint*. Por ejemplo, a partir de registros de eventos de dos clientes adicionales, fue posible aplicar los primeros pasos de la metodología relacionados con la extracción de los datos, generación del log, y la inspección del mismo. Se presenta en la tabla IV la información general obtenida por medio de la herramienta DISCO, correspondiente al cliente *Retail* y los dos clientes adicionales.

TABLA IV. INFORMACIÓN DE LOGS DE EVENTOS DE TRES CLIENTES QUE POSEEN INTRANET EN SHAREPOINT

Característica	Log de Cliente 1 Retail	Log de Cliente 2 Seguridad Laboral: Prevención de Riesgos	Log de Cliente 3 Seguridad Laboral: Accidentes Laborales
Eventos	100,511	412,027	132,170
Variaciones	217	509	104
Casos totales	983	1849	257
Actividades	62	62	16
Recursos participantes	14	12	5
Tiempo promedio de ejecución por caso	45.6	252	102,5

Como información general, es posible observar que el cliente de tipo *Retail* posee mayor similitud con el cliente 2 si se comparan las actividades y los recursos involucrados; sin embargo, con respecto a la cantidad de eventos, variaciones y tiempo promedio, el cliente de *Retail* posee menor cantidad, lo cual está directamente asociado al tamaño del *log* de eventos almacenado. El cliente 2 presenta variaciones del proceso y el tiempo promedio de ejecución, en donde ambos son significativamente mayores. El cliente 3 presenta un *log* que posee un tamaño menor considerando el número de eventos, variaciones, casos, actividades y recursos. No obstante, estos casos a pesar de ser menor cantidad, tardan mucho más en completarse. Es relevante poder llegar a encontrar hallazgos significativos que expliquen esta situación. Se desea extender el análisis utilizando la metodología PMSLA propuesta en este artículo para descubrir posibles diferencias con respecto al proceso y la perspectiva organizacional. Extender el análisis y mostrar los resultados obtenidos quedan fuera del alcance de la presente investigación; sin embargo, serán considerados para una versión posterior del trabajo, donde no solamente se puedan analizar estos clientes, sino también considerar la información de otros clientes adicionales.

VII. CONCLUSIONES Y TRABAJO FUTURO

Como objetivo de esta investigación se tenía poder definir una metodología para trabajar con *logs* de eventos extraídos de una intranet. Lo novedoso de esta metodología llamada *Process Mining SharePoint Log Analysis* es que utiliza herramientas de minería de procesos en conjunto con información de procesos que están soportados en la tecnología *SharePoint Server*. Esta metodología busca generar resultados que permitan una correcta y eficiente administración de los procesos de negocios en una organización. Esto se lograría al poder tener un mayor entendimiento de los procesos de negocio y generar oportunidades de mejora. Para probar la metodología planteada, se realizó un caso de estudio que involucró el análisis de procesos computacionales relacionados con el soporte de una intranet para un cliente *Retail* de una consultora de tecnologías de información en Chile. A partir de la construcción de un *log* de eventos y por medio de la utilización de técnicas y algoritmos, se ha descubierto el proceso que se ejecuta y algunas de sus características principales. A pesar de enfocarnos en procesos computacionales, con la herramienta *SharePoint* es posible generar procesos de negocio que cumplan con estándares como BPMN y YAWL, por lo que se podría extender el uso de la metodología PMSLA para analizar otros procesos.

Fue posible mostrar los resultados obtenidos al cliente para obtener su retroalimentación. Según su valoración, el aporte que le brinda este trabajo es posible enmarcarlo en dos contextos. Primero, está el contexto relacionado con el ámbito de negocio, y en segundo lugar, están los hallazgos asociados a la metodología aplicada como tal, y al desempeño obtenido al utilizar las herramientas de minería de procesos en la implementación del caso de estudio.

Desde el contexto del negocio, la metodología PMSLA le brinda al cliente una serie de ventajas que pueden ser aprovechadas por recursos que desempeñen el rol de administradores de plataformas *SharePoint Server* dentro de una consultora de negocio, entre ellas:

- El aporte de la metodología puede ayudar a definir requerimientos de hardware y software. A nivel de hardware, puede establecer las necesidades de dispositivos que se requiere activar o no dentro de la plataforma. Si se tiene poco hardware es posible que no se puedan activar todos los servicios pues no se cuentan con los recursos adecuados. Con respecto al software, ayuda a determinar cuáles servicios se necesitan activar.
- Generar *insights* que se pueden utilizar para abaratar costos a nivel de hardware en el sentido de saber si está o no sobredimensionada la plataforma de hardware (más recursos de los que se necesitan).
- Ayudar a establecer cuáles son los servicios y procesos que no deben ser revisados por los antivirus de la compañía, debido a que los antivirus empresariales buscan procesos repetitivos y los bloquean.
- Colaborar en la definición de las necesidades de conexión entre los diferentes servidores que componen la granja existente. De esta manera, se pueden establecer los diversos niveles de conexión que necesitan entre ellos.
- Permite un mejor análisis y control del seguimiento que se le debe dar a la plataforma para asegurar su estabilidad, al mantener los *logs* dentro de la plataforma.

Desde el punto de vista de la metodología, la misma involucraba el uso de algoritmos y técnicas de minería de procesos. Para el cliente fue relevante notar que al aplicar la metodología fue posible aprovechar la información almacenada sobre procesos computacionales y realizar la construcción del *log* de eventos para hacer el análisis de estos. Esto es importante debido a que no se había utilizado este tipo de información de *SharePoint* para realizar tareas de minería sobre ella. Con respecto al desempeño obtenido al utilizar la metodología, los autores coincidimos que es importante considerar factores como el tamaño del proceso descubierto y la aplicación o no de técnicas de *clustering* para mejorar el descubrimiento del proceso. Además, hay que tomar en cuenta cuáles algoritmos de descubrimiento son más óptimos de aplicar (según recursos, tiempo y capacidad de procesamiento disponible) y también cuáles métricas de la perspectiva organizacional ejecutar dentro del análisis.

En el caso de estudio ejecutado, los resultados en cuanto a la perspectiva de descubrimiento muestran que el algoritmo *Heuristic Mininig* descubre mejores modelos con menor uso de recursos y en menor tiempo, mientras que el *Genetic Mining* puede generar modelos similares pero invirtiendo mayor cantidad de recursos. Estas consideraciones tienen impacto en cuanto a poder realizar un análisis más efectivo según los resultados que se buscan obtener. Asimismo, es importante incorporar la opinión de los dueños de los procesos

por parte del cliente para determinar cuál enfoque les genera mayor valor al implementar minería de procesos.

El conocimiento del contexto del negocio, mezclado con el dominio de las técnicas, herramientas y algoritmos de minería de procesos, brinda la oportunidad de descubrir cómo se están ejecutando los procesos organizacionales en todo nivel de detalle, cómo se están administrando los recursos y cómo se pueden identificar oportunidades de mejora. Analizar procesos de corte más computacional permite generar *insights* relevantes sobre el comportamiento de este tipo de procesos dentro de un ámbito de negocio que está siendo soportado por una infraestructura tecnológica como SharePoint Server.

Como trabajo futuro, se puede apuntar a incluir más algoritmos de descubrimiento y hacer un análisis exhaustivo de todas sus diferencias en cada uno de los procesos. Es importante ampliar el rango de tiempo de la muestra de los datos para analizar mayor cantidad de casos y verificar los resultados obtenidos actualmente. El incluir más tiempo supone un desafío en cuanto a las capacidades de procesamiento de los datos y el periodo de tiempo que se necesita para realizar el análisis.

Adicionalmente, se pretende analizar el registro de eventos de los otros dos clientes, incluyendo todos los pasos de la metodología PMSLA para obtener un análisis más completo de los mismos.

VIII. AGRADECIMIENTOS

Para este trabajo se utilizaron datos de clientes de una consultora que brinda servicios de Microsoft SharePoint Server en Chile y Costa Rica, a la cual se le agradece la disposición a colaborar en esta investigación.

REFERENCIAS

- [1] M. Weske, *Business process management: concepts, languages, architectures*. Springer, 2012.
- [2] Learned from Successful BPM Organizations. *Business Rules Journal* 12.10 (2011).
- [3] W.M.P. Van der Aalst, et al., "Business process mining: An industrial application." *Information Systems*, 2007. 32(5): pp. 713-732.
- [4] A. Tiwari, C. J. Turner, and B. Majeed. "A review of business process mining: state-of-the-art and future trends." *Business Process Management Journal* 14.1. 2008, pp 5-22.
- [5] W. M. P. Van der Aalst, B. F. van Dongen, J. Herbst, L. Maruster, G. Schimm, and A. J. M. M. Weijters. *Workflow mining: A survey of issues and approaches*. Data and Knowledge Engineering. 2003.
- [6] W. Van der Aalst, T. Weijters, and L. Maruster, "Workflow mining: Discovering process models from event logs." *Knowledge and Data Engineering, IEEE Transactions On*, 16(9), 2004, pp 1128-1142
- [7] W. M. Van der Aalst, *Process mining: discovery, conformance and enhancement of business processes*. Springer, 2011.
- [8] W. M. Van der Aalst, "Process-aware information systems: Lessons to be learned from process mining. In *Transactions on Petri Nets and Other Models of Concurrency II*", Springer Berlin Heidelberg, 2009, pp 1-26.
- [9] Perez-Castillo, Ricardo, et al. "Generating event logs from non-process-aware systems enabling business process mining." *Enterprise Information Systems* 5.3, 2011, pp 301-335.
- [10] P. Nasirifard, V. Peristeras, C. Hayes and S. Decker, "Extracting and Utilizing Social Networks from Log Files of Shared Workspaces" Book Section: *Leveraging Knowledge for Innovation in Collaborative Networks*. IFIP Advances in Information and Communication Technology. Springer, Berlin Heidelberg, 2009, pp 643-650.
- [11] M. Havey, *Essential business process modeling*. O'Reilly, 2009.
- [12] S. A. White, "Introduction to BPMN." IBM Cooperation. 2004, pp 2008-029.
- [13] W. M. Van Der Aalst and A. H. T. Hofstede, "YAWL: yet another workflow language." *Information systems* 30.4 . 2005, pp 245-275.
- [14] T. Murata, "Petri nets: Properties, analysis and applications." *Proceedings of the IEEE* 77.4. 1989, pp 541-580.
- [15] L. Fischer, ed. *The Workflow Handbook*, 2002. Future Strategies Inc., 2002.
- [16] W. M. P. Van der Aalst, "Process Mining: Overview and Opportunities." *ACM Trans. Manage. Inf. Syst.* 3, 2. 2012.
- [17] W. M. P. Van der Aalst, et al, "Manifiesto de Minería de Procesos". IEEE Task Force on Process Mining. 2011. Recuperado el 5 de agosto de 2013, de www.win.tue.nl/ieeetfpm/lib/exe/fetch.php?media=shared:pmm-spanish-v1.pdf
- [18] A. J. M. M. Weijters and W. M. P. Van der Aalst, "Rediscovering workflow models from event-based data using little thumb." 2003.
- [19] A. K. A. de Medeiros, *Genetic Process Mining* (PhD Thesis). Eindhoven University of Technology, Eindhoven. 2006.
- [20] C. W. Günther and W. M. Van der Aalst, "Fuzzy Mining: Adaptive Process Simplification Based on Multi-Perspective Metrics." In G. Alonso, P. Dadam, and M. Rosemann, editors, *International Conference on Business Process Management (BPM 2007)*, volume 4714 of *Lecture Notes in Computer Science*. Springer, Berlin, 2007, pp 328-343
- [21] W. M. Van der Aalst, H. Reijers and M. Song, "Discovering Social Networks from Event Logs." *Computer Supported Cooperative Work (CSCW)*, 14(6). 2005, pp 549-593.
- [22] S. Minseok, W. M. Van der Aalst, "Towards comprehensive support for organizational mining." *Decision Support Systems*, Volume 46, Issue 1. 2008 pp 300-317.
- [23] W. M. P. Van der Aalst, et al. *ProM: The Process Mining Toolkit*, 2009.
- [24] N. Gehrke and M. Werner, *Process Mining*, 2013. p. 10.
- [25] Naderipour, Farhad. "Mining of Ad-hoc Business Processes using Microsoft Sharepoint, Nitro & Prom 6.0-An Industrial Practice." *ICEIS* (4). 2011.
- [26] B. F. van Dongen, A. K. A. de Medeiros, H. M. W. Verbeek, A. J. M. M. Weijters, and W. M. P. van der Aalst. 2005. "The prom framework: a new era in process mining tool support." In *Proceedings of the 26th international conference on Applications and Theory of Petri Nets (ICATPN'05)*, Gianfranco Ciardo and Philippe Darondeau (Eds.). Springer-Verlag, Berlin, Heidelberg, 444-454.
- [27] J. Claes, G. Poels, "Process Mining and the ProM Framework: An Exploratory Survey." *BPM 2012 Workshops*. 2012.