

**UNIVERSIDAD NACIONAL DE INGENIERÍA**  
**FACULTAD DE INGENIERÍA INDUSTRIAL Y DE SISTEMAS**



**EVALUACIÓN DE LA PRECISIÓN DEL DESCUBRIMIENTO DE  
MODELOS DE PROCESOS APLICANDO TÉCNICAS DE  
MINERÍA DE PROCESOS**

**TESIS**

**Para optar por el Título Profesional de  
Ingeniera de Sistemas**

**Mg. Bach. Aldave Castillo, Cecilia Valentina.**

**Lima, Perú.**

**2014**

**Digitalizado por:**

**Consortio Digital del  
Conocimiento MebLatam,  
Hemisferio y Dalse**

## **DEDICATORIA**

El presente trabajo lo dedico con mucho amor a mi familia y a Dios por haberme brindado su apoyo a lo largo de todo el camino de mis estudios profesionales.

A Dios por estar siempre junto a mí dándome fuerzas en los momentos más difíciles de mi vida y ayudándome a hacer todo de manera correcta.

A mi padre y a mi madre, les agradezco infinitamente por sus sabios consejos, por todo el amor que me han brindado y por haberme apoyado para avanzar más cada día.

A mi hermana por haberme acompañado y ayudado a resolver todo problema que apareció en el camino hacia el logro de mis objetivos y a mi hermano por su soporte constante cada vez que lo he necesitado.

No existen más palabras que agradecerles muchísimo.

Los quiero mucho, muchas gracias!

## **AGRADECIMIENTOS**

Quisiera agradecer al Dr. Emilio Alberto Un Jan Liao Hing por haber sido mi asesor de tesis brindándome su disponibilidad, paciencia y sus recomendaciones a lo largo del desarrollo del presente trabajo. De manera similar quiero agradecer al Ing. Luis Zuloaga Rotta y al Dr. Ricardo Rodríguez Ulloa por su apoyo en la realización del presente trabajo.

También quiero hacer un agradecimiento especial a mi profesor de la maestría "MIAGE – Universidad Pantheón Sorbonne" (Paris, Francia), el Dr. Camile Sallinesi por haberme orientado en la formulación de mi tema de tesis.

Muchas gracias a todos mis profesores, así como a la plana docente de la FIIS.

A fin de evitar el omitir agradecimientos sin querer, quisiera también agradecer a todos aquellos que me han ayudado de alguna manera en el avance del presente tema ya sea de cerca o de lejos.

# ÍNDICE

RESUMEN.....	6
ABSTRACT.....	10
DESCRIPTORES TEMATICOS.....	14
INTRODUCCIÓN.....	15
CAPITULO I: PENSAMIENTO ESTRATÉGICO.....	16
I.1 DIAGNÓSTICO FUNCIONAL.....	16
I.2 DIAGNÓSTICO ESTRATÉGICO.....	16
CAPITULO II: MARCO TEÓRICO Y METODOLÓGICO.....	18
II.1 RETRO INGENIERÍA.....	18
II.2 MINERÍA DE PROCESOS.....	20
II.2.1 FUERZA DE TRABAJO IEEE.....	20
II.2.2 DEFINICIÓN Y BENEFICIOS DE LA MINERÍA DE PROCESOS.....	20
II.2.3 HERRAMIENTAS DE MINERÍA DE PROCESOS.....	23
II.3 ELEMENTOS DE LA MINERÍA DE PROCESOS.....	24
II.4 DESCUBRIMIENTO DEL MODELO DE PROCESOS A PARTIR DE UN REGISTRO DE EVENTOS.....	26
II.4.1 PATRONES DE CONTROL.....	28
II.4.2 ALGORITMO PROM-ALPHA PARA LA MINERÍA DE PROCESOS.....	32
II.5 CLARIDAD DE LOS MODELOS GENERADOS.....	33
CAPITULO III: PROCESO DE TOMA DE DECISIONES.....	36
III.1 IDENTIFICACIÓN DE PROBLEMA.....	36
III.2 OBJETIVOS DE LA INVESTIGACIÓN.....	36
III.4 MATRIZ DE CONSISTENCIA.....	37
III.5 PLANTEAMIENTO DE LA SOLUCIÓN.....	39
III.5.1 ETAPAS DE LA INVESTIGACIÓN.....	39
III.5.2 ESCENARIOS ESTABLECIDOS.....	39
III.6 PLANES DE ACCIÓN PARA DESARROLLAR LA SOLUCIÓN PLANTEADA.....	40
III.6.1 CRITERIOS DE SELECCIÓN DE LAS: “ENTRADAS/SALIDAS”.....	40
III.6.2 POBLACIÓN.....	41
III.6.3 MUESTRA.....	46
III.6.4 ESQUEMA: “POBLACIÓN – MUESTRA”.....	47
III.6.5 VARIABLES.....	49

III.7 ANÁLISIS COSTO BENEFICIO .....	55
CAPITULO VI. ANÁLISIS DE RESULTADOS .....	56
IV.1 RESULTADOS DE LA ETAPA 1: .....	56
IV.2 RESULTADOS DE LA ETAPA 2: CUESTIONARIO DE CLARIDAD .....	68
IV.3 ANÁLISIS DE LA HERRAMIENTA .....	74
IV.4 VALIDACIÓN DE LA HIPÓTESIS.....	83
IV.5 ANÁLISIS GLOBAL DE LOS RESULTADOS.....	94
CONCLUSIONES Y RECOMENDACIONES.....	97
CONCLUSIONES.....	97
RECOMENDACIONES.....	102
INDICE DE FIGURAS.....	104
INDICE DE ECUACIONES: .....	106
INDICE DE TABLAS .....	107
BIBLIOGRAFIA .....	108
1. FUENTES BIBLIOGRAFICAS.....	108
2. FUENTES SITIOS WEB DE INTERNET.....	111
ANEXOS .....	113
1. COMPOSICIÓN DE LA SECCIÓN DE EVENTOS DEL MICRO REGISTRO .....	113
2. EJEMPLO DEL ANÁLISIS EMPLEADO EN LA MINERÍA DE PROCESOS .....	115
3. RESULTADO DETALLADO DE LAS EVALUACIONES .....	122
4. ENCUESTA PARA LA EVALUACIÓN DE LA CLARIDAD DE LOS MODELOS .....	130

## RESUMEN

En los siguientes años los sistemas aumentarán aún más, generando millones de eventos que reflejen el verdadero flujo de procesos en la empresa. Por ello la tendencia actual en la retro ingeniería está relacionada a la aplicación de la técnica de minería de procesos que nos permite obtener información sobre los procesos a partir de registros de eventos. La minería de procesos puede ser aplicada a diferentes dominios como la banca, educación, salud, gobierno entre otros, así también dentro de una sola organización o entre organizaciones según como se encuentren los procesos distribuidos.

Ya en los 90's diferentes grupos trabajaban para crear diferentes técnicas a ser implementadas en las herramientas informáticas. Técnicas que nos permitieran obtener los modelos de procesos del negocio a partir de los eventos observados dentro de los registros de eventos. Pero los primeros trabajos formales aparecen en el 95.

En el año 2009 aparece la Fuerza de trabajo de la IEEE compuesta por: HP, IBM, Fujitsu, las consultoras internacionales Gartner y Deloitte, así como más de 20 universidades entre ellas la Universidad de Tecnología de Eindhoven en Países Bajos (representada por su jefe del departamento de tecnología y editor asociado a la IEEE el profesor y Doctor Wil Van Der Aalst quién ha publicado diversos artículos sobre el tema), con la finalidad de promover el uso y crear la documentación sobre la minería de procesos.

Hoy en día existen cada vez más proveedores de herramientas agregando funcionalidades de minería de procesos en sus productos. De esta manera los investigadores de la Universidad de Tecnología de Eindhoven han creado y actualizado constantemente la herramienta libre de minería de procesos ProM que ahora integra otras herramientas similares como son EMIT, Thumb, Mison, entre otras, con la finalidad de ser hacerla mucho más completa. Se ha creado una página de difusión sobre dicha herramienta y se

ha hecho un llamado a todos aquellos investigadores y personas que deseen probarla y a contribuir con sus aportes para su mejora continua.

En cuanto a esto, si bien ProM provee un soporte a la minería de procesos, aún nos preguntamos:

- Si es posible asegurar que los modelos de procesos generados por esta herramienta serán verdaderamente precisos y claros. Es decir, si al momento de aplicar la minería de procesos a partir de los registros de eventos se obtendrán modelos que representarán la realidad fiel de dichos registros o si los modelos obtenidos serán incompletos o poco comprensibles.

- Si verdaderamente sería recomendable emplear estas herramientas de manera habitual, como soporte para el descubrimiento y la posterior mejora de procesos de la empresa y sino: ¿Qué es lo que les faltaría desarrollar? ¿Que está faltando para que éstas funcionen bien?

Con la finalidad de responder dichas inquietudes es que la presente investigación pretende proponer un marco de evaluación que será usado para evaluar la herramienta de minería de procesos ProM con el propósito de tener una opinión y recomendación de esta herramienta representativa.

Para esto, se comenzará por el desarrollo de un estado del arte en el que se hablará acerca de las justificaciones y conceptos a usar acerca de la aplicación de la minería de procesos, los diagramas de procesos, los registros de eventos y los patrones de flujo de trabajo. Se detallará una equivalencia entre la forma de leer un registro de eventos y como esta lectura se representará a través de los patrones de flujo de trabajo mediante el empleo del algoritmo Alpha por ProM.

La comparación entre el modelo manual generado a partir de la observación del registro de eventos fuente y aquel modelo que obtendremos como fruto de la aplicación de la minería de procesos, nos permitirá saber si se está o no generando bien el modelo y llegar al porqué en caso éste no se esté generando bien.

La investigación de la generación de los modelos por medio de la minería de procesos a partir de un registro de eventos será muy detallada. Las instancias de modelos así como el registro de eventos a evaluar contendrán eventos de lo más variados. Es decir que el número de eventos se hallará en diferentes cantidades por cada instancia de proceso y estos contendrán una diversidad de patrones de flujo de trabajo como por ejemplo: La Sucesión Directa, el Paralelismo, la Elección, la Causalidad entre otros. De modo que los resultados sean lo más fiables a la realidad.

El análisis se hará mediante la comparación de cada patrón de flujo de trabajo del modelo original manual que representa al registro de eventos que será usado en la minería de procesos y los patrones dentro del modelo generado luego de la aplicación de la minería de procesos. De igual manera, durante el estudio de esta herramienta representativa podremos avizorar cuales son las funcionalidades de las herramientas actuales de minería de procesos que se encuentran mejor desarrolladas y cuáles no.

La hipótesis girará en torno a si el uso de la herramienta de minería de procesos ProM ha generado aquello que realmente debía generar o no, por lo que adicionalmente al método de observación se planteará contar el total de elementos generados y mediante el método T Student se hará la evaluación de estas hipótesis. Esta validación nos permitirá contrastar el resultado de la prueba con lo observado y enseguida llegar a las conclusiones del estudio y a las recomendaciones.

Finalmente, solo queda por mencionar que la presente investigación ha sido inspirada en los cursos llevados en mi maestría de Métodos Informáticos Aplicados a la Gestión de Empresas en la Universidad Pantheon Sorbonne y Paris Descartes (Paris, Francia 2010-2012). Así como, en el trabajo de investigación de mi memoria de maestría [2] en donde: Evalué la precisión de una herramienta en la generación de los modelos de clases aplicando la retro ingeniería de programas que habían sido previamente generados por ingeniería directa. Dicho trabajo lo realicé junto a mi profesor y asesor de Memoria de Máster: Camille Salinesi (Profesor de la Universidad Paris 1

Panthéon Sorbonne en París, Francia, director del Centro de Investigación en Informática y director del Máster 2 “Sistemas de Información y de Decisión”), quién también me facilitó parte de la información sobre el tema de la presente tesis. Así como el Dr. W.M.P. Van Der Aalst quien me ayudó en el acceso a un real registro de eventos de un hospital nacional en Países Bajos. De igual manera, esta tesis también se encuentra inspirada en el trabajo de retro modelización que realicé en el área de la Dirección de Sistemas de Información del banco BNP Paribas en París.

## **ABSTRACT**

In the next years the number of systems will increase, and it will become in millions of events that reflect the true flux of process in the Enterprise. Because of this, the current trend in retro-engineering is related to the application of the process mining which permit us obtain information about the process from the events logs. The process mining can be applied to different domains as the bank, education, health, government between others, also in one organization or between organizations according to the distribution of the process.

In the 90's different groups are working to create different techniques to be implemented in the computer tools. Techniques that will allow us to obtain the models of the business process from the observed events inside of the events logs. But the first formal works appear in 1995.

In 2009 starts the work force of the IEEE composed by: HP, IBM, Fujitsu, the international cabinet of consulting Gartner y Deloitte, as well as more than 20 universities among them the University of Technology of Eindhoven in Netherlands (represented by the head of technology department and the editor associated to the IEEE the professor and Doctor Wil Van Der Aalst who has published different articles about the subject), with the finality of promote the use and create the documentation about the process mining.

Nowadays there are more and more tools suppliers that they are adding functionalities of process mining in their products. In this way, the scientists of the University of Technology of Eindhoven have created and continually updated the free tool of Process Mining ProM that now integrates others similar tools as EMIT, Thumb, Mison, between others, in order to make it more complete. They have created a web page to make this tool popular and there is a call to all the scientists and people who wish to try it and contribute with feedback for the continuous improvement of ProM.

In relation to this, although ProM provides a support to the process mining, we still ask us:

- If it is possible to assure that the models of process generated by this tool will be really precise and clear. In other words, if in the moment that we applied the process mining from the events logs we will obtain models that represent the exact reality of these events logs or if these models will be incompletes or difficult of understand.
- If indeed, it would be advisable to employ these tools regularly, in order to support for the discovering and in consequence the improvement of process in the Enterprise and if not: ¿What else is needed to implement ProM? ¿What is it necessary for this tool to function well?.

With the objective to answer these questions, this research tries to propose an evaluation framework. This evaluation framework will be used to evaluate the process mining tool ProM with the purpose of having an opinion and a recommendation about this representative tool.

Because of this, we will begin developing a state of art which will speak about the the theories and concepts to use about the application of the process mining, the process diagrams, the event logs and the workflow patterns. We will detail an equivalence between the way of reading an event log and the way this reading will be represented through the workflow patterns by employing the algorithm Alpha by ProM.

The comparison between the manual model generated by the observation of the source event log and the model that will be obtained as a result of the application of the process mining, will allow us to know if there are or not a good generation of the model and in case of failure, get to know the reasons why the generation the model was not so good.

The reaserch of the generation of the models through process mining from an event log will be very detailed. The instances of models and the event log to

be evaluated will have a variety of events. In other words, the number of events will be in different amounts for each process instance and these process instances will have a diversity of workflow patterns such as the following: The directly succession, the parallelism, the election, the causality between others. So the results will be more reliable to the reality.

The analysis will be done by the comparison of each workflow pattern of the manual original model, which represents the event log will be used in the process mining, and the patterns inside the model generated after the application of the process mining. Also, during the study of this representative tool, we can know what are the functionalities of the current process mining tools which are best developed and which are not.

The hypothesis will focus on whether the use of the process mining tool ProM has generated what really should generate or not, for this in addition to the observation method we will count the total of elements generated and employ the T student method to evaluate this hypothesis. This validation will allow us contrast the result of the test with what is observed and then come to the conclusions and recommendations of the research.

Finally, I would like to mention that this reseach has been inspired in the courses that I took in my studies of the master degree "Informatics Methods Applied to the Management of Enterprises" in the University Pantheón Sorbonne and Paris Descartes (Paris, France 2010-2012). Also, in my work of investigation for my master memory [2] where: I perform the evaluation of the accuracy of a tool in the generation of the class models by applying the retro engineering of programs which have been previously generated by direct engineering. This work was made with my professor and master memory advisor: Camille Salinesi (Professor of the University Paris 1 Panthéon Sorbonne in Paris, France, Director of the research Center in Informatics and Director of the Master 2 "Systems of Information and of Decision"), who also provided me some information about the subject of this thesis. Similarly, the Dr. W.M.P. Van Der Aalst who helped me with the

access to a real event log of a national hospital in Netherlands. As well, this thesis was also inspired by the work of retro modeling that I did in the area of Management Information Systems at the bank BNP Paribas in Paris

## DESCRIPTORES TEMATICOS

- Ingeniería Directa.
- Sistemas de Flujo de Trabajo.
- Retro Ingeniería.
- Minería de Procesos.
- Herramientas de Minería de Procesos.
- Registro de Eventos.
- Patrones de Flujo de Trabajo.
- Modelo de Procesos.
- ProM.
- Algoritmo Alpha.
- Redes Petri.
- Claridad de Procesos.

## INTRODUCCIÓN

Según la encuesta de BPTrends [1] en la actualidad, se han incrementado el número de empresas que poseen sistemas de gestión de flujo de trabajo. Dichos sistemas contienen incorporados en sí mismo el manejo de los diferentes procesos de la empresa mientras esta ha ido evolucionando a lo largo de los años.

Muchas veces la forma de trabajar de las personas que se apoyan en el uso de las herramientas de gestión de procesos no se ve reflejada en el flujo de trabajo formal que la empresa presenta públicamente a sus empleados. Generalmente el conocimiento sobre el verdadero flujo de trabajo puesto en práctica se concentra en pocas personas dentro de la empresa, lo que hace que al pasar los años éste se pierde luego de que las personas portadoras del conocimiento ya se han ido de la empresa. Es por ello que en algún momento surge la inmensa necesidad de recuperar toda esta información perdida en la mente de las personas pero incluida dentro de la herramienta de gestión de flujo de trabajo. Su recuperación nos permitirá conocer mejor las actividades reales de los procesos de gestión, verificar si éstas son conformes a la realidad de la empresa y a la vez a mejorar el diseño de estos procesos actuales mediante una reingeniería tal, que los procesos apoyen más a la estrategia de la empresa.

Esto se apoya en la retro ingeniería, que según Elliot Chikofsky y James Cross II es la recuperación de los modelos en un alto nivel de abstracción y legibilidad y que contienen toda la información que ha estado perdida en diferentes medios. Medios como los registros de eventos en cuyo caso se le denomina a la técnica de recuperación: "Minería de Procesos".

En torno al tema de la retro ingeniería y la minería de procesos los principales temas de investigación tienen relación con la evaluación de nuevas herramientas informáticas que permiten la obtención de modelos y que han incorporado en sí diferentes e innovadoras técnicas para ello.

## **CAPITULO I: PENSAMIENTO ESTRATÉGICO**

### **I.1 DIAGNÓSTICO FUNCIONAL**

#### **ORGANIZACIÓN**

Se trata de un hospital de Los Países Bajos de naturaleza anónima, en el cual se ofrecen diferentes tratamientos para los pacientes. Cuenta con diferentes áreas, entre las cuales se estudiarán los procesos de atención en el área de ginecología. Ginecología cuenta con los servicios de diagnóstico y los servicios de rehabilitación.

#### **CLIENTES**

Son todos los pacientes que siguen un tratamiento en el hospital y que han sido diagnosticados con algún tipo de cáncer (de cérvix, de vulva, de útero, de ovario, de vagina) en algún nivel de gravedad.

### **I.2 DIAGNÓSTICO ESTRATÉGICO**

#### **ANALISIS EXTERNO**

La nueva ley "Sarbanes Oxley (SOX)" está forzando a las organizaciones a seguir sus actividades de negocios más de cerca. Por lo que los procesos deben estar más monitoreados, lo que a su vez puede aumentar la performance y la eficiencia.

Según la ley de Moore, los sistemas de información van a expandirse en los años siguientes, lo que traerá como consecuencia el aumento de los registros de eventos (como por ejemplo el retiro de dinero de una caja automática) que contienen las informaciones más detalladas sobre los procesos. Nosotros debemos aprovechar toda esta información para comprender mejor aquello que pasa y los posibles cuellos de botella existentes en los sistemas.

## CAPITULO II: MARCO TEÓRICO Y METODOLÓGICO

### II.1 RETRO INGENIERÍA

Según Elliot Chikofsky and James Cross [3]: *“La retro ingeniería es el proceso de analizar un sistema sujeto para identificar sus componentes y sus relaciones y crear representaciones del sistema en otra forma o un nivel de abstracción más alto”*. Así se puede a partir del análisis de cualquier tipo de información generada (como código java, registros de eventos) obtener información de más alto nivel como son los modelos de producto, de procesos, entre otros.

La retro ingeniería es útil cuando no se cuenta con información documentada sobre cómo se manejan los procesos y actividades que se llevan a cabo en la empresa y cuya lógica se encuentra implementada en los diferentes sistemas. Según Elliot Chikofsky y James Cross [3] la retro ingeniería: *“Surge con el objetivo de ayudar a la re documentación, la generación de vistas gráficas alternativas, la recuperación de las informaciones perdidas, detectar los efectos secundarios debido a las sucesivas modificaciones no documentadas y a la facilitación para la reutilización del sistema”*.

Si nosotros conocemos la realidad acerca de cómo funciona el sistema, entonces será más fácil realizar cualquier modificación sobre éste y de esta manera podríamos ahorrar en costos de tiempo de comprensión por parte de los desarrolladores de sistemas.

El proceso de retro ingeniería está compuesto según Ati Jain, Swapnil soner, Anand Gadwal [4] en cuatro etapas: *“El análisis de contexto, el análisis de componentes, la recuperación y la reconstrucción de diseños”*.

Hoy en día los principales artículos científicos detallan diferentes evaluaciones sobre las distintas técnicas implementadas en las herramientas de retro ingeniería. Jawaria Sadiq y Tabinda Waheed [5] presentan los criterios de evaluación propuestos por ellos y las técnicas que evaluaron como son: El análisis estático, el análisis estructural, la analogía, el análisis de datos, entre otras que han sido ya implementadas en las diferentes herramientas de retro ingeniería. De igual manera, Chih-Wei Lu, William C. Chu, Chih – Hung Chang, Yeh – Ching Chung, Xiaodong Liu y Hongji Yang [6] evaluaron las herramientas de retro ingeniería VIFOR, Logiscope, Seela, Rigi, Mansart, Simba en relación a las principales tecnologías de investigación existentes en el campo como son: El entendimiento de programas, la recuperación de requerimientos, la recuperación de modelos, la recuperación de la arquitectura, etc.

## II.2 MINERÍA DE PROCESOS

### II.2.1 FUERZA DE TRABAJO IEEE

Ya en los 90's existían diversos grupos trabajando en el desarrollo de las diferentes técnicas de minería de procesos que se basan en la observación de los eventos dentro de los registros de eventos, pero los primeros trabajos recién se hacen públicos en el 95.

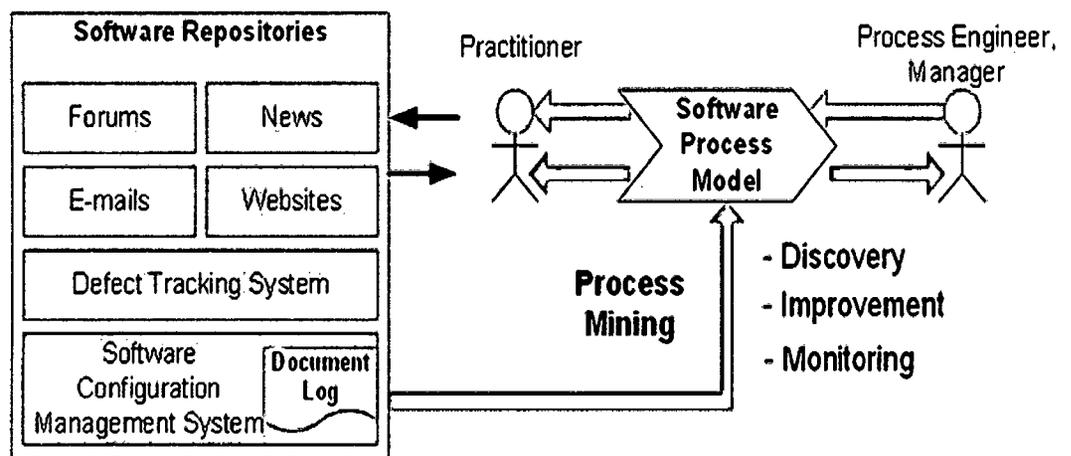
En el año 2009 oficialmente se funda la fuerza de trabajo de la IEEE compuesta por los representantes de los más grandes vendedores comerciales, así como por 20 universidades de diferentes países, entre ellos: [7] HP, BPMLATAM, Universidad Cranfield, Fluxicon, Fujitsu, IBM, Universidad de Haifa, Universidad Hasselt, Universidad Técnica de Lisboa, Pontificia Universidad Católica de Chile, Universidad de Tecnología Eindhoven, Universidad Politécnica de Catalunya, Deloitte, Excellentia BPM, Fluxicon, con la finalidad de promover la investigación, el desarrollo, la educación y el entendimiento en la minería de procesos.

### II.2.2 DEFINICIÓN Y BENEFICIOS DE LA MINERÍA DE PROCESOS

Wil Van Der Aalst [8] define: *“La minería de procesos como una disciplina entre la inteligencia de computación y la minería de datos y el modelado y análisis de procesos. Toda técnica de minería de procesos asume que se puede recorrer un registro de eventos de manera secuencial. Cada evento es una actividad y está relacionado a una instancia de proceso”*. La minería de procesos es una nueva tecnología que puede ser aplicada a diferentes dominios como la banca, educación, salud, gobierno e inclusive según Wil M. P. y Van Der Aalst [9] dentro de una sola organización y en diferentes organizaciones según como estén distribuidos los procesos.

Según Vladimir Rubin, Christian W. Günther, Wil M.P. Van Der Aalst, Ekkart Kindler, Boudewijn F. Van Dongen, y Wilhelm Schafer [10]: *“La arquitectura*

tradicional de los ambientes de ingeniería de software centrados en procesos contaba con repositorios de software donde el producto del software se guardaba en repositorios. Así el ingeniero de procesos diseñaba el proceso original que debía ser instanciado por los empleados en su trabajo cotidiano, el cual a veces era instanciado de manera incorrecta, dejando de reflejar el modelo inicial planteado. Por ello, hoy en día se le sugiere al ingeniero de procesos aplicar la minería de procesos y tomar el registro de eventos que corresponde a las instancias de procesos y mejorar el modelo de procesos de manera continua e incremental”.



**Figura 1: Arq. Tradicional vs Incremental según Vladimir Rubin, Christian W. Günther, Wil M.P. van der Aalst, Ekkart Kindler, Boudewijn F. van Dongen, and Wilhelm Schafer [8].**

La minería de procesos surge entonces con la finalidad de analizar los registros de eventos que contienen la descripción de los procesos instanciados por los participantes de la empresa para obtener información sobre las diferentes perspectivas:

- La perspectiva de procesos del negocio: Conocida como el ¿Cómo? y cuyo objetivo es la buena caracterización de todos los posibles caminos que se pueden dar expresados en términos de una red PETRI, UML, BPMN, etc.

- La perspectiva organizacional: Conocida como el ¿Quién? y se centra en los recursos ocultos en el registro de eventos, tales como los actores con la finalidad de estructurar la organización en términos de roles y unidades organizacionales o como una red social.
- La perspectiva del tiempo, la cual tiene relación con el tiempo y con la frecuencia de los eventos y que permite el monitoreo del uso de los recursos, la predicción del tiempo que resta, la generación de indicadores fiables y la calidad total.

La Fuerza de Trabajo en minería de procesos de la IEEE [11] nos menciona en uno de sus comunicados publicados acerca de la minería de procesos, que existen 3 tipos de minería de procesos y que son los que se mencionan a continuación:

- El descubrimiento: Donde se toma un registro de eventos y se produce un modelo de procesos (expresado en una red PETRI, BPMN, Diagrama de Actividad UML) sin tener información previa sobre el modelo.
- La conformidad: Se tiene un modelo de procesos inicial y se compara con un registro de eventos del mismo proceso y se verifica si ambos corresponden.
- La mejora: Se mejora el proceso existente con la información del registro de eventos. Por ejemplo, al leer la información sobre la hora en la que se llevo a cabo el evento en el registro de eventos, podemos verificar la existencia de cuellos de botella en el nivel de servicio, frecuencias, etc.

Muchas tendencias de gestión relacionadas a la mejora de procesos (Six Sigma, TQM, CPI, CPM) y conformidad (SOX, BAM, etc.) podrían beneficiarse de la minería de procesos.

### II.2.3 HERRAMIENTAS DE MINERÍA DE PROCESOS

Muchos sistemas de información guardan actualmente, los eventos más importantes de manera estructurada en algún tipo de registro de eventos, tal como es el caso de los sistemas de administración de flujo de trabajo (o BPMS, business process management systems), los ERP (SAP) ó los CRM.

Ante esto, Wil M. P. y Van Der Aalst [10] nos indican que existen cada vez más vendedores que han comenzado a agregar funcionalidades de minería de procesos en sus herramientas. Así las herramientas: ARIS, Process Performance Manager, Enterprise Visualization Suite, Interstage BPME, OKT Process Mining Suite, Process Discovery Focus, Process Analyser, ProM, Rbminer/Dbminer, Reflect|one, Reflect ya incluyen dichas funcionalidades. Sin embargo, el comunicado de la Fuerza de Trabajo en minería de procesos de la IEEE nos indica que como la minería de procesos es un paradigma nuevo y estas herramientas no son aún muy maduras.

W.M.P. Van Der Aalst, H.A. Reijers, A.J.M.M. Weijters, B.F. Van Dongen, A.K. Alves de Medeiros, M. Song, H.M.W. Verbeek [12] nos indican que a pesar de que existen muchas técnicas desarrolladas para la minería de procesos, la mayoría tienen asunciones y dificultades con las excepciones: Así algunas solo se limitan a procesos con una estructura particular, por lo que es necesario cotejar las herramientas y técnicas existentes con los registros de eventos de la vida real.

En los últimos 5 años se han creado varias herramientas de minería de procesos por la Universidad de Tecnología en Eindhoven. Por ejemplo: EMIT, Thumb y Mison, las cuales actualmente se han integrado para formar el framework ProM que contiene más de 90 programas siendo los más interesantes los de análisis y minería de procesos.

“ProM Framework” y “ProM Import Framework” son dos herramientas desarrolladas en la “Universidad de Tecnología de Eindhoven” por Wil Van Der Aalst y otros colaboradores. Estas herramientas son gratuitas.

ProM proporciona una variedad de algoritmos en una sola interface. Además, también soporta la minería de procesos en el descubrimiento de procesos por medio del algoritmo Alpha, la identificación de cuellos de botella, análisis de redes sociales, verificación de redes de negocio, etc. También permite extraer la información desde diferentes fuentes como son los sistemas de gestión de configuración de software "SCM".

Generalmente los registros contienen muchos detalles por lo que se necesita una herramienta de abstracción. PROM contiene un conjunto de filtros para resolver este problema.

### **II.3 ELEMENTOS DE LA MINERÍA DE PROCESOS**

Los elementos que intervienen en la minería de procesos son los siguientes:

- **Un proceso:** Es la realización de un conjunto de actividades que son ejecutadas por individuos, máquinas y/o sistemas de información con la finalidad de cumplir un objetivo específico y cuyo resultado produce un valor agregado. El concepto de proceso podría tener un equivalente con el concepto de clase.
- **Un modelo de procesos de negocios:** Es la descripción por medio de un gráfico que representa las actividades y estados que componen el proceso de negocios.
- **Una instancia de proceso de negocios:** Está compuesta por el camino a través de las actividades que serán recorridas dentro de un modelo de procesos. Las actividades pueden ser: Una orden de un cliente, una aplicación a un trabajo, una queja al seguro, etc. Una instancia podría tener un equivalente con el concepto de objeto. Las instancias se crean, existen por un tiempo y desaparecen (llegan a un estado final). Durante su vida una instancia solo puede estar en un estado en cada instante del tiempo.

- **Un evento:** Es una actividad de una instancia de proceso que debe ser rastreada. Un registro de un evento puede contener información adicional sobre el evento, como por ejemplo sobre: El recurso, el actor (sea una persona o un dispositivo) que realiza el evento, el tiempo en el cuál se realizó el evento, los elementos de datos que se han guardado con el evento (como el tamaño y el orden). Los eventos son guardados para proveer una historia detallada acerca de los procesos.
- **Un registro o log de eventos:** Es una grabación que lista todos los eventos que se ejecutaron en una instancia de proceso determinada, junto con la información tal como cuando fueron ejecutados y quién los ejecuto. Un registro de eventos puede también contener información sobre la persona que ejecutó el evento. A partir de septiembre del año 2010 el formato adoptado por la Fuerza de Trabajo en minería de procesos de la IEEE es el formato XES (eXtensible Event Stream) que es un estándar basado en XML para los registros de eventos.
- **Patrones de flujo de trabajo:** Son patrones de diseño que sirven para documentar una solución para un problema de diseño en un campo. Así los requerimientos para los lenguajes de flujo de trabajo son indicados por medio de estos patrones, que a su vez sirven como base para realizar comparaciones entre los diferentes sistemas de gestión de flujo de trabajo existentes. Riehle y Zullighoven [13] indican que: “Un patrón es la abstracción de una forma concreta que se mantiene recurrente en contextos específicos no arbitrarios”.

E. Gamma, R. Helm, R. Johnson y J. Vlissides [14], son los primeros en catalogar los patrones en 20 tipos diferentes, permitiendo visualizar las interacciones entre los sistemas orientados a objetos. Pero se debe saber que no todos los sistemas de gestión de flujo de trabajo soportan todos los patrones definidos. Wil Van Der Aalst, Arthur H.M. Hofstede, Bartek Kiepuszewski, y Alistair P. Barros [15] formulan una clasificación de patrones de flujo de trabajo a partir del estudio de diferentes herramientas.

Ellos crean una página de [16]: "Patrones de Flujo de Trabajo", donde se alienta a los investigadores, desarrolladores y usuarios a generar una retroalimentación sobre los patrones establecidos bajo investigación. El resultado es que muchos investigadores han usado los patrones propuestos para evaluar la utilidad de diferentes sistemas de gestión de flujos de trabajo y también nuevos lenguajes de diseño de flujo de trabajo.

Así A. Hirnschall en [17] señala que se han usado dichos patrones para evaluar el sistema de gestión de flujo de control CONDIS y H. Lavana en [18] que ha evaluado el ambiente Omniflow usando 10 de los patrones formulados.

#### **II.4 DESCUBRIMIENTO DEL MODELO DE PROCESOS A PARTIR DE UN REGISTRO DE EVENTOS**

Según Wil Van Der Aalst [19]: *"Un proceso consiste de un conjunto de casos. Un caso consiste de un grupo de eventos. Los eventos dentro de un caso (instancia de proceso) se encuentran ordenados. Los eventos pueden tener atributos (actividad, tiempo, costo, recurso, entre otros)".*

A continuación se muestra un ejemplo del registro de eventos y la manera en que se generan las posibles secuencias de actividades para el modelo de procesos

Registro de eventos	
Instancia de Proceso 1	Tarea A
Instancia de Proceso 2	Tarea A
Instancia de Proceso 3	Tarea A
Instancia de Proceso 3	Tarea B
Instancia de Proceso 1	Tarea B
Instancia de Proceso 1	Tarea C
Instancia de Proceso 2	Tarea C
Instancia de Proceso 4	Tarea A
Instancia de Proceso 2	Tarea B
Instancia de Proceso 2	Tarea D
Instancia de Proceso 5	Tarea E
Instancia de Proceso 4	Tarea C
Instancia de Proceso 1	Tarea D
Instancia de Proceso 3	Tarea C
Instancia de Proceso 3	Tarea D
Instancia de Proceso 4	Tarea B
Instancia de Proceso 5	Tarea F
Instancia de Proceso 4	Tarea D

La Instancia de proceso 1:  
 $A > B, B > C, C > D$   
 La Instancia de proceso 2:  
 $A > C, C > B, B > D$   
 La Instancia de proceso 3:  
 $A > B, B > C, C > D$   
 La Instancia de proceso 4:  
 $A > C, C > B, B > D$   
 La Instancia de proceso 5:  
 $E > F$



Las posibles  
 secuencias son:  
 ABCD, ACBD, EF

**Figura 2: Desde un registro de eventos hacia un modelo de procesos.**  
*Elaboración propia.*

Cada secuencia en un registro de eventos toma un patrón diferente según sus características, así para representar estas secuencias en un modelo de procesos, tomaremos la clasificación de los 20 patrones de flujo de trabajo establecidos por Wil Van Der Aalst, Arthur H.M. Hofstede, Bartek Kiepuszewski y Alistair P. Barros en [15]. Estos son los siguientes:

## II.4.1 PATRONES DE CONTROL

### ***Los Patrones de Control Básicos:***

- 1.*Secuencia:* Este patrón es utilizado para modelar dependencia entre tareas, es decir, una tarea no puede empezar hasta que otra del mismo proceso no haya terminado [15].
- 2.*División Paralela:* Una división paralela es un punto del proceso donde un camino es dividido en dos o más ramas paralelas las cuales son ejecutadas al mismo tiempo [15].
- 3.*Sincronización,* La sincronización es un punto en el proceso donde dos o más ramas del proceso se unen en una sola. Se llama sincronización porque se espera a que todas las ramas entrantes se completen antes de continuar con la siguiente actividad [15].
- 4.*Elección exclusiva,* La decisión exclusiva representa un punto en el proceso donde se debe escoger un solo camino de varios disponibles dependiendo de una decisión o de datos del proceso [15].
- 5.*Unión Simple,* La unión simple es un punto del proceso donde dos o más caminos alternativos convergen en uno solo. Es un supuesto de este patrón que los caminos alternativos no son ejecutados en paralelo [15].

### ***Ramificaciones Avanzadas - Patrones de Sincronización:***

- 1.*Elección Múltiple:* El patrón de elección múltiple es utilizado para modelar puntos en el flujo de proceso donde varios caminos son escogidos dependiendo de datos del proceso y/o puntos de decisión [15].
- 2.*Estructura de Unión Sincronizada:* Es un punto en el proceso donde múltiples caminos que fueron activados antes en el proceso convergen en

una sola rama, el proceso continua hasta que todas las actividades de las ramas entrantes hayan terminado [15].

3. *Unión Múltiple*: Se usa para hacer converger dos o más ramas en un solo camino. Cada activación de una de las ramas entrantes resulta en la activación de la siguiente actividad en el proceso [15].

4. *Discriminador Estructurado*: Este patrón describe un punto en el proceso donde se espera a que una de las ramas entrantes haya sido completada antes de continuar con la siguiente actividad, las otras ramas son omitidas después de ser completadas. Una vez todas las ramas entrantes han sido completadas, el discriminador se reinicia [15].

#### ***Patrones Estructurales:***

1. *Ciclo Arbitrario*

2. *Terminación Implícita*

#### ***Patrones Relacionados a Múltiples Instancias:***

1. *Múltiples Instancias Sin Sincronización*: Es utilizado para modelar actividades que han sido instanciadas muchas veces en el proceso y no necesitan ser sincronizadas para que el flujo continúe. Este patrón también es conocido como Multi Caminos Sin Sincronización [15].

2. *Múltiples Instancias Con Conocimiento a Priori del Número de Ejecuciones en la Etapa de Diseño*: Permite instanciar una actividad varias veces en el proceso. El número de instancias es conocido y se ejecutan al mismo tiempo. Es necesario terminar todas las actividades antes de continuar con el flujo del proceso [15].

3. *Múltiples Instancias Con Conocimiento en el Tiempo de Ejecución*: Permite instanciar una actividad muchas veces durante el proceso. La

cantidad de instancias es conocida hasta el momento de la ejecución del proceso. Las actividades son ejecutadas de forma paralela y es necesario terminar todas las instancias abiertas antes de continuar con la siguiente actividad [15].

#### 4. *Múltiples Instancias Sin Conocimiento en el Tiempo de Ejecución.*

##### **Patrones Basados en Estados:**

1. *Elección Diferida:* Este patrón describe un punto en el proceso donde una rama es escogida de entre varias disponibles. La decisión es tomada de acuerdo a los datos del proceso. Cuando una rama es escogida, las demás se deben deshabilitar [15].
2. *Ejecución Intercalada:* Este patrón describe que cada actividad de un conjunto de actividades debe ser ejecutada una vez. Las tareas pueden ser realizadas en cualquier orden, sin embargo, no es posible realizar dos tareas al mismo tiempo [15].
3. *Milesone:* Este patrón define que una actividad es habilitada solamente hasta cuando el proceso alcance un estado específico (Hito). Si el proceso ha avanzado más allá de ese estado, la tarea no puede volver a ser habilitada [15].

##### **Patrones de Cancelación:**

1. *Cancelación de tareas:* Permite cancelar o suprimir una tarea que está siendo ejecutada [15].
2. *Cancelación del Caso:* Este patrón describe la eliminación de una instancia completa del proceso. Esto incluye tareas que están siendo ejecutadas y las tareas que todavía no lo han sido [15].

A continuación un gráfico explicativo de cada patrón anteriormente mencionados:

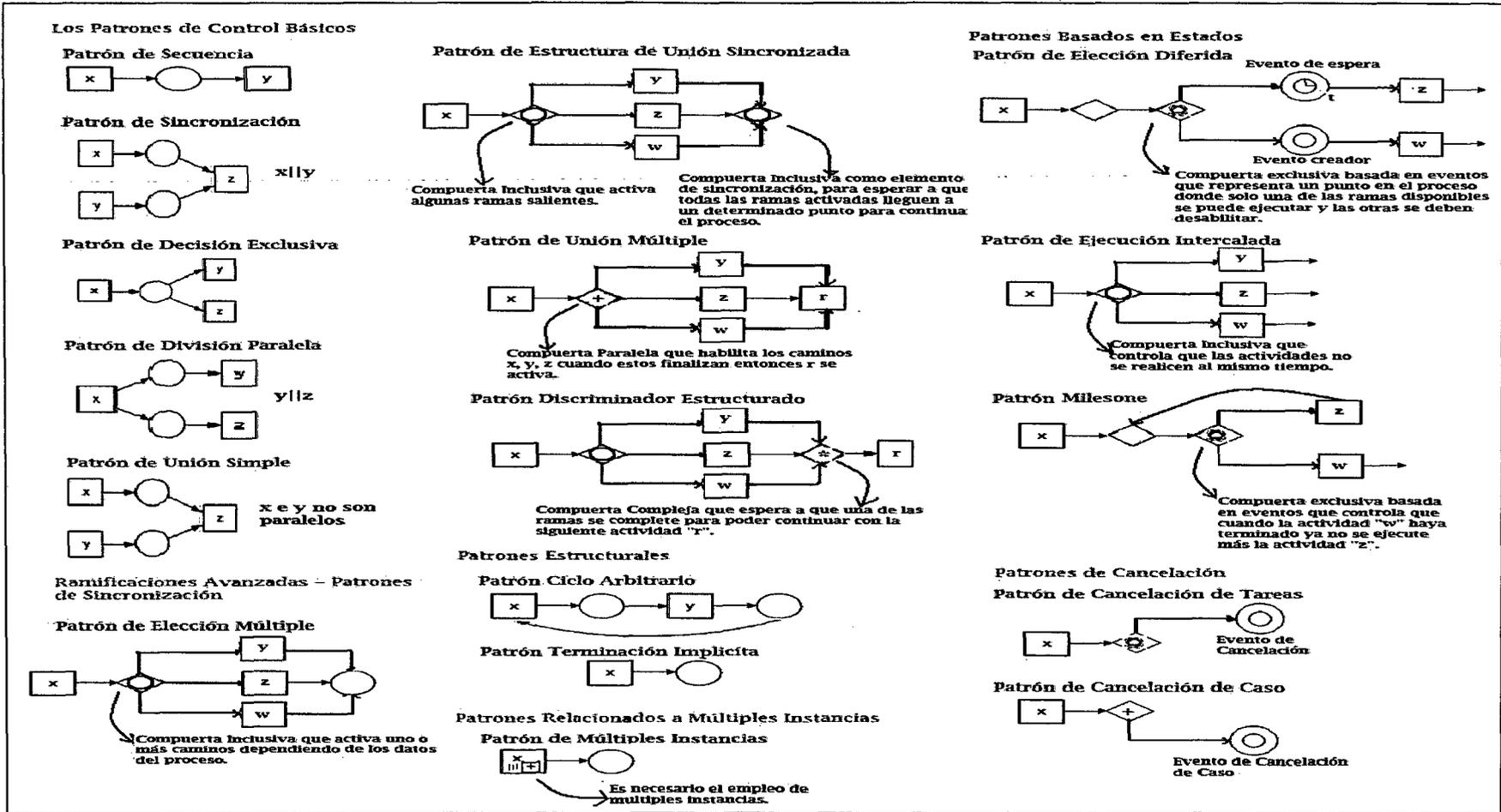


Figura 3: Patrones de Control. Elaboración Propia

## II.4.2 ALGORITMO PROM-ALPHA PARA LA MINERÍA DE PROCESOS

Según el estudio presentado en [20]: *“El algoritmo Alpha para la minería de procesos es el segundo algoritmo más usado en la última versión de ProM (6.1) por ser intuitivo, fácil de entender, confiable y rápido”*. La lista la lidera un simple programa que solo permite adicionar los eventos artificiales de “inicio”, “fin”, “y” y “o” para cada conjunto de eventos. Por lo que tomáremos como algoritmo representativo de la herramienta ProM el algoritmo “Alpha” para la obtención de modelos de procesos en forma de Petri Nets.

El algoritmo Alpha maneja la siguiente lógica para la obtención del modelo global por minería de procesos:

1. Obtención de todas las relaciones de secuencia existentes entre las actividades por cada instancia del registro de eventos. Como se asume que los eventos se encuentran ordenados en cada instancia de procesos: El primer y el último evento de la instancia de procesos serán la primera y última actividad del diagrama de flujo de procesos respectivamente.
2. A partir de estas relaciones de secuencia se verifica si surgen otros nuevos patrones.
3. Se elabora el modelo.

Entonces a partir del registro de eventos anterior tendríamos que:

1. Las relaciones de secuencia son:

a>b, entonces a→b

a>c, entonces a→c

b>c, entonces b→c

c>b, entonces c→b

c>d, entonces c→d

b>d, entonces b→d

$e \rightarrow f$ , entonces  $e \rightarrow f$

2. Los nuevos patrones que surgirían a partir de las relaciones de secuencia anteriores son:

- A partir de a, tenemos que:  $a \rightarrow b$ ,  $a \rightarrow c$ ,  $b \rightarrow c$ ,  $c \rightarrow b \Rightarrow$  De a saldrá que  $b \parallel c$ .
- Como ya tenemos la relación entre b y c pasamos a la siguiente. Enseguida a partir de c tendremos que:  $c \rightarrow d$  y  $b \rightarrow d \Rightarrow$  c y b sincronizarán en d.
- A partir de e:  $e \rightarrow f \Rightarrow$  e será seguido por f.
- Finalmente como a y e no las antecede ninguna otra actividad en su respectiva instancia de proceso, éstas serán actividades de origen de flujo. Mientras que como f y d no son precedidas por ninguna actividad en su respectiva instancia de proceso, éstas serán actividades de fin de flujo.

Enseguida se obtiene el modelo que será comprobado con uno original que ha sido generado manualmente a fin de encontrar las incoherencias o poder afirmar que ambos son idénticos.

Se pueden obtener una diversidad de modelos originales correctos, es decir que contengan a todos los caminos de los eventos contenidos en el registro de eventos. Por ello se usará uno de los modelos que ha sido creado correctamente por minería de procesos, para realizar la evaluación de la claridad, a fin de que la comparación entre ambos sea lo más precisa.

## **II.5 CLARIDAD DE LOS MODELOS GENERADOS**

La claridad de los modelos es un tema importante que no se debe dejar del lado, cuando analizamos la precisión con la que una herramienta crea un modelo. Para crear un modelo claro, primeramente debemos investigar que significa la claridad y luego evaluar que tan claros son nuestros modelos.

Daniela Oelke, David Spretke, Andreas Stoffel y Daniel A. Keim en [21] indican que: *“Para tener un texto claro, se deben evitar usar palabras desconocidas, con una estructura gramatical difícil de comprender, con faltas de ortografía. Existen aspectos básicos en el problema de la claridad: Primero están los problemas de lingüística y dificultades de contenido y enseguida esta el contexto de la coherencia y la coexistencia como por ejemplo: El diseño de impresión de una página”*.

En cuanto a la forma de medir la claridad, los autores antes mencionados indican que: *“La mayoría de mediciones se hace en base a propiedades estadísticas como el largo de las palabras, el número de palabras en una oración, la complejidad en el vocabulario y la complejidad de la estructura de una oración (la oración es compleja si tiene oraciones incluidas entre paréntesis ya que el cerebro está obligado a recordar partes incompletas de la oración)”*.

Akhilesh Bajaj, Stephen Rockwell en [22] consideran que la claridad que obtenemos al usar un método de modelización nos indicará esencialmente la facilidad con la que podemos leer un modelo y aún más reconstruir el dominio detrás del mismo. Ellos mencionan que existen variables independientes a considerar durante la investigación de la claridad: *“El nivel de familiaridad de las personas con el tema y la fragmentación de la información en el modelo”*. La claridad se encontrará vinculada a 3 aspectos fundamentales: La eficacia, la eficiencia y la facilidad de aprendizaje.

Yvon Laframboise en [23] expone que para: *“La mayoría de investigadores la claridad se encuentra estrechamente vinculada a la presentación material del texto escrito (claridad en lo impreso), es decir en el tipo de caracteres tipográficos, la disposición de los gráficos en la página, la calidad del papel, la composición de las imágenes”*.

Los científicos han deducido fórmulas sobre diferentes factores en busca de medir la claridad, tales como el número de frases, el número de palabras que pertenecen a un vocabulario básico, así como el nivel de abstracción de

los textos. En [22] también se indica que la percepción de la claridad puede variar en función de la edad y nivel de estudios.

## **CAPITULO III: PROCESO DE TOMA DE DECISIONES**

### **III.1 IDENTIFICACIÓN DE PROBLEMA**

La creación de diferentes herramientas de minería de procesos que están en proceso de desarrollo genera que muchas veces todos los patrones estándar de flujos de proceso dentro de los modelos implementados no son generados de manera adecuada, generando modelos incorrectos y poco claros.

### **III.2 OBJETIVOS DE LA INVESTIGACIÓN**

La investigación tiene como objetivo:

1. Evaluar enfocándonos principalmente en la precisión y adicionalmente en la claridad de los modelos producidos por la herramienta representativa (ProM) y su algoritmo Alpha que permite la obtención de los modelos de procesos del negocio por medio de la aplicación de la minería de procesos a partir de los registros de eventos.

### **III.3 HIPÓTESIS:**

H<sub>0</sub> etapa 1: La herramienta crea el modelo correctamente luego de la minería de procesos.

$$\Leftrightarrow \mu_{\text{ProM01}} = \mu_{01}$$

H<sub>1</sub> etapa 1: La herramienta no crea el modelo correctamente luego de la minería de procesos.

$$\Rightarrow \mu_{\text{ProM01}} \neq \mu_{01}$$

### III.4 MATRIZ DE CONSISTENCIA

DISEÑO TEORICO		
Problema	Objetivo	Hipótesis
<p>1. La creación de diferentes herramientas de minería de procesos que están en proceso de desarrollo genera que muchas veces todos los patrones estándar de flujos de proceso dentro de los modelos implementados no son generados de manera adecuada, generando modelos incorrectos.</p>	<p>1. Evaluar enfocándonos principalmente en la precisión y adicionalmente en la claridad de los modelos producidos por la herramienta representativa (ProM) y su algoritmo Alpha que permite la obtención de los modelos de procesos del negocio por medio de la aplicación de la minería de procesos a partir de los registros de eventos.</p> <p>Con Este objetivo se busca identificar las principales fallas de la</p>	<p>H<sub>0</sub> etapa 1: La herramienta crea el modelo correctamente luego de la minería de procesos.</p> <p><math>\Rightarrow \mu_{\text{ProM01}} = \mu_{01}</math></p> <p>H<sub>1</sub> etapa 1: La herramienta no crea el modelo correctamente luego de la minería de procesos.</p>

	minería de procesos, por medio de la herramienta ProM, para finalmente proponer algunas recomendaciones de mejora.	
--	--	--

### III.5 PLANTEAMIENTO DE LA SOLUCIÓN

La investigación tendrá la siguiente metodología de solución de la hipótesis:

#### III.5.1 ETAPAS DE LA INVESTIGACIÓN

Tiene 2 etapas:

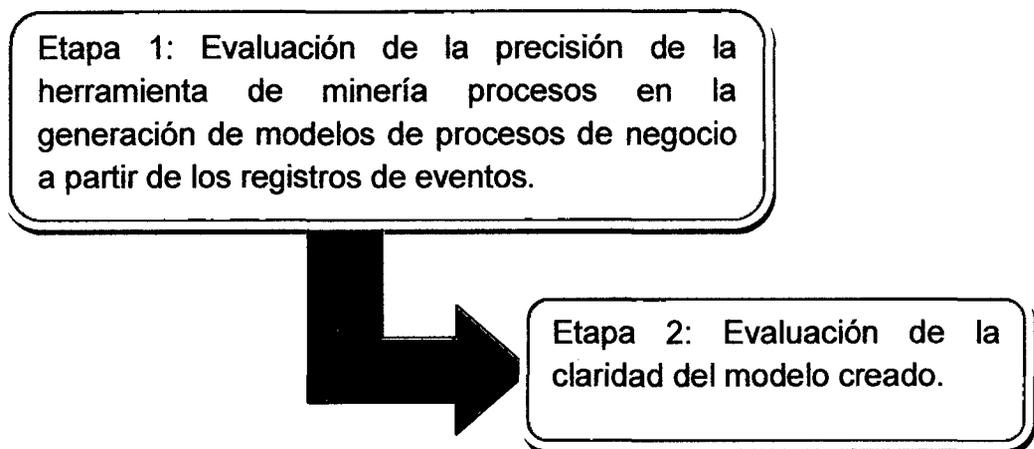


Figura 4: Etapas de la investigación. *Elaboración propia.*

#### III.5.2 ESCENARIOS ESTABLECIDOS

**Etapa 1:** “Evaluación de la precisión de la herramienta de minería de procesos en la generación de los modelos de procesos de negocio a partir de los registros de eventos”.

- **Necesidad:** Es necesario tener inicialmente, para llevar a cabo la investigación, la pareja (modelo de proceso de negocio, registro de eventos del modelo de proceso de negocio).
- **Entradas:** Registro de eventos del proceso a estudiar en el formato estándar XES XML.
- **Salidas:** Modelo de procesos de negocio en forma de redes PETRI.

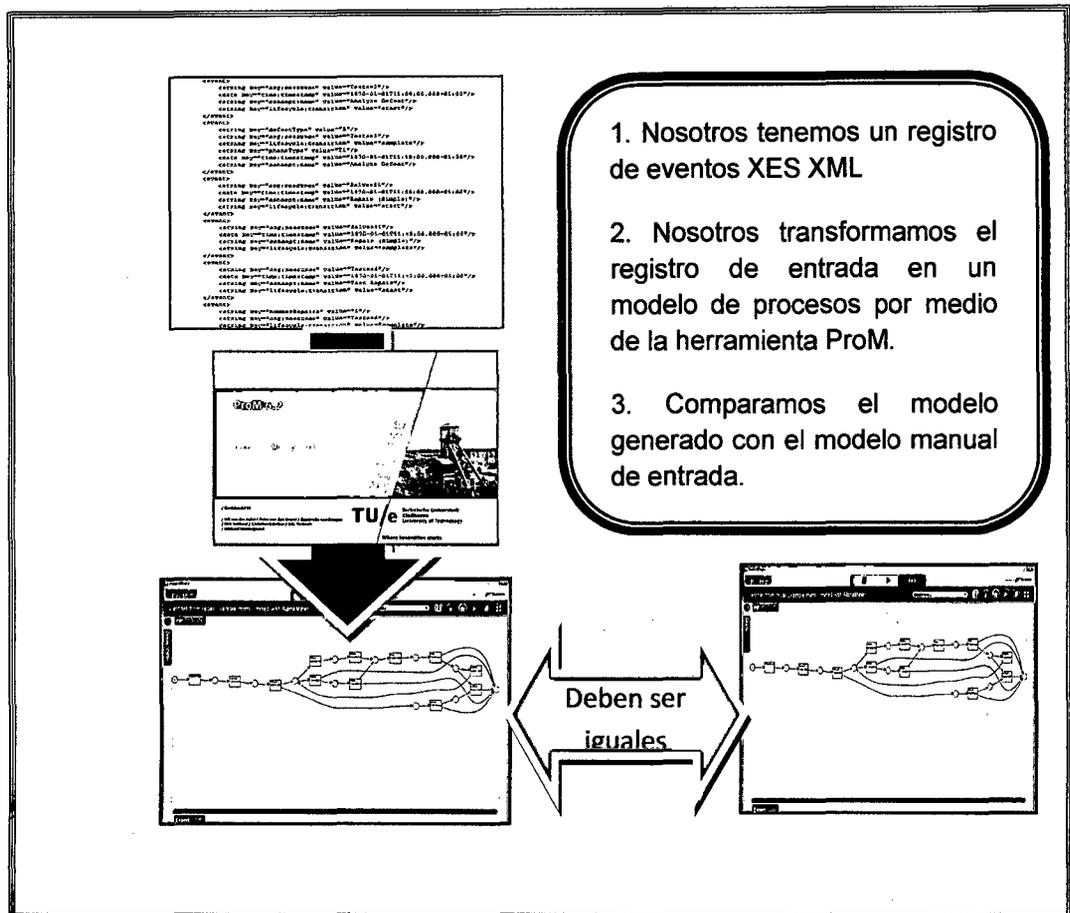


Figura 5: Escenario de investigación. *Elaboración propia.*

### III.6 PLANES DE ACCIÓN PARA DESARROLLAR LA SOLUCIÓN PLANTEADA

#### III.6.1 CRITERIOS DE SELECCIÓN DE LAS: “ENTRADAS/SALIDAS”

**Entradas:** ¿Por qué un registro de eventos? ¿Por qué el formato XES XML?

- La expansión de los sistemas de información traerá como consecuencia el aumento de registros de eventos con muchos datos que contienen abundante información.
- Los registros de eventos contienen abundante información detallada sobre los procesos instanciados de la empresa, los recursos y los tiempos.

- El formato XES es elegido por haber sido anunciado como estándar para el almacén de eventos y por ser soportable por la mayor cantidad de herramientas de generación de procesos por retro ingeniería.

**Salidas: ¿Por qué un modelo de procesos? ¿Por qué en formato PETRI?**

- Los procesos ya existen dentro de una empresa, lo importante es identificarlos gestionarlos. Una correcta gestión permitirá alcanzar de manera rápida efectiva los objetivos de la empresa.
- Debemos tener un modelo de procesos actualizado ya que a veces el modelo planteado es mal instanciado por los trabajadores, por lo que el ingeniero de procesos necesita mejorarlo constantemente.
- Las redes PETRI son fáciles de entender gracias a su naturaleza gráfica y precisa.

### **III.6.2 POBLACIÓN**

La población estará compuesta por un registro de eventos de la vida real proveniente de un hospital académico de Los Países Bajos, por lo que el idioma en el que se encuentra es el neerlandés. El registro contiene unos 150 000 eventos y más de 1 100 casos.

El registro de eventos fue proporcionado por el Doctor W.M.P van der Aalst luego de que se le realizará la consulta sobre la existencia de data pública (eventos) disponible para la presente investigación. Este registro de eventos se encuentra disponible en [28] en donde también encontramos información acerca de que el mismo registro ha sido también empleado en investigaciones realizadas para el “Business Process Intelligence Contest (BPIC 2011)”.

Cada caso es un paciente de un departamento de Ginecología. El registro de eventos contiene información acerca de cómo y cuándo se llevaron a cabo ciertas actividades y qué grupo llevo a cabo la actividad. Muchos atributos relevantes para los procesos han sido registrados, algunos atributos se repiten más de una vez para un paciente, lo que indica que este paciente pasó por diferentes fases (tal vez se superponen). Una fase consiste en la combinación de un diagnóstico y el tratamiento.

El registro de alguna manera es anónimo, ya que proviene de los sistemas de un hospital real, por lo que contiene datos sensibles reales. Este registro de eventos fue usado en el primer Congreso de Inteligencia de Procesos de Negocio (BPIC 2011) y ha sido proporcionado por la Universidad de Tecnología de Eindhoven.

Dado que la herramienta ProM tiene limitaciones con el manejo de registro de eventos de talla muy grande (como en nuestro caso que tenemos 1684937 líneas), *la población ideal del estudio no estará compuesta por el total de eventos del registro completo sino que estará compuesto por micro-registros de eventos de igual tamaño cada uno de los cuales permitirá generar un modelo de procesos.*

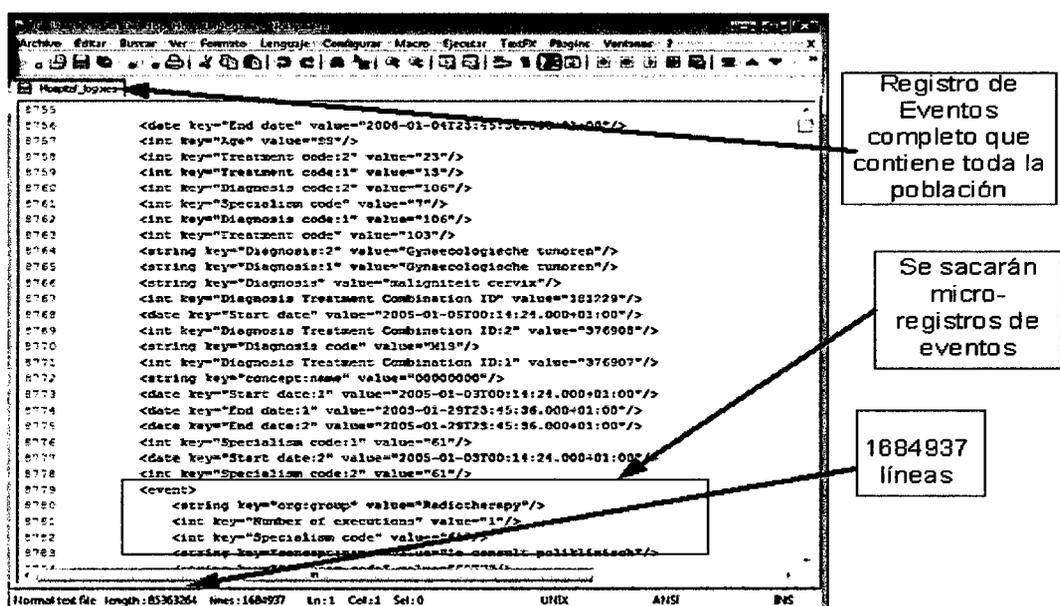


Figura 6: De registros a micro – registros de eventos. Elaboración Propia

Propia

En la definición de dichos micro- registros se debe tener en consideración dos puntos importantes:

1. El tamaño del micro- registro debe representar el comportamiento real del registro completo. La única manera de que cada micro- registro contenga el comportamiento representativo del registro completo es usando un método de muestreo que nos permita calcular el valor común sugerido del tamaño de cada muestra o micro-registro.
2. Coherencia en cada micro- registro, la secuencia en la que aparecen los eventos en los registros permite identificar con coherencia como se llevan a cabo las actividades en los procesos. Si tomamos eventos del registro original de manera aleatoria para formar cada micro-registro se perderá esta coherencia y se generarán modelos incoherentes lo cual no es correcto. Por lo que se mantendrá la secuencia de los eventos entre el primer y el último evento escogidos para el micro –registro.

The image shows a screenshot of an XML editor window titled 'Monitor\_log.xml'. The XML content consists of three event blocks, each starting with `<event>` and ending with `</event>`. Each event block contains several key-value pairs, such as `<string key="org:group" value="General Lab Clinical Chemistry"/>`, `<int key="Number of executions" value="1"/>`, `<int key="Specialism code" value="86"/>`, `<string key="concept:name" value="aaname laboratorioonderzoek"/>`, `<string key="Producer code" value="CRJA"/>`, `<string key="Section" value="Section 4"/>`, `<int key="Activity code" value="370000"/>`, and `<date key="time:timestamp" value="2008-03-20T00:00:00.000+01:00"/>`. The first and third event blocks end with `<string key="lifecycle:transition" value="complete"/>`. Annotations on the right side of the image use arrows to point to specific parts of the XML. An arrow points to the first event block labeled 'X<sup>0</sup> evento'. Another arrow points to the second event block labeled 'X+1<sup>o</sup> evento'. A third arrow points to the third event block labeled 'X+2<sup>o</sup> evento'. To the right of these annotations, there is a vertical line. Text next to the line says: 'Es correcto, se debe mantener la coherencia, primero evento x<sup>0</sup> luego evento x+1<sup>o</sup>'. Below this, another text block says: 'No, porque al saltarnos el evento X+1<sup>o</sup>, se pierde la coherencia.'

Figura 7: Micro – registros y la coherencia de eventos. Elaboración Propia

Tomando en lo anterior, se ha decidido usar el tipo de muestreo “No Probabilístico – Intencional u opinático”.

Muestreo No Probabilístico: Según Scharager, J y Armijo, I. [24], en este tipo de muestras también llamadas muestras dirigidas o intencionales, la elección de los elementos no depende de la probabilidad sino de las condiciones que permiten hacer el muestreo (acceso o disponibilidad, **conveniencia**, etc.). Como en nuestro caso que nos conviene tomar los eventos de una manera determinada o continua y no al azar.

Estas muestras son seleccionadas con mecanismos informales. En el Muestreo intencional u opinático es el investigador quien selecciona la muestra e intenta que sea representativa, por lo tanto, la representatividad depende de su “intención” u “opinión”.

Como conocemos que el número de eventos del registro de eventos es menor que 300000 eventos, entonces el método ideal para hallar el tamaño de la muestra representativa será el de “Tamaño de una muestra con población finita” (Método independiente del tipo de muestreo probabilístico o no probabilístico).

Tamaño de la muestra con población finita: Según [25], el tamaño de la muestra se puede determinar mediante la siguiente fórmula:

$$n = \frac{N}{1 + \frac{e^2(N - 1)}{z^2 pq}} \dots \dots \dots (\alpha)$$

**Ecuación 1: Fórmula del tamaño de una muestra con población finita.**

Donde:

n = tamaño de la muestra que deseamos conocer,

N = tamaño conocido de la población,

pq (o  $\sigma^2$ ) = Varianza de la población

El suponer que  $p = q$  quiere decir que para escoger la muestra nos ponemos en la hipótesis de que en la población hay la máxima diversidad posible: un 50% va a decir que sí y otro 50% va a decir que no, de esta manera, y por lo que respecta a la varianza de la población, no corremos riesgos de quedarnos cortos en el número de sujetos. Este valor de pq (= .25) es válido (válido para calcular el tamaño de la muestra) aún cuando las preguntas no sean dicotómicas.

e = Error de la muestra

El error de la muestra significa error o desviación posible cuando extrapolamos los resultados. Es el margen de error que aceptamos. El margen de error no debe ser superior a .05 (5%) para que los resultados sean realmente informativos y útiles.

z = Valor de z correspondiente al nivel de confianza

El valor de z (o distancia en errores típicos con respecto a la media) lo decidimos libremente y así establecemos los límites o intervalos de confianza de la verdadera media (la media de la población, no la de la muestra). Usualmente, escogemos un nivel de confianza del 95%  $\Rightarrow \alpha = .05$  (lo que significa que nos equivocáramos 5 veces de cada 100), al que corresponde  $z = 1.96$ .

Luego en el Caso, para el total de eventos del registro original con los que se cuenta (150000) se tendrá que:

$$n = \frac{150000}{1 + \frac{0.05^2(150000 - 1)}{1.96^2 \times 0.25}}$$

n: Tamaño recomendado de la muestra es igual a 384 eventos por micro registro (lo que no necesariamente es igual a 384 actividades diferentes ya que algunas actividades y caminos se pueden repetir más de una vez dentro de un micro registro).

Luego el tamaño de la población (o el número de micro - registros) se calculará:

$$\text{Número de Micro registros} = \frac{\text{Número de eventos del registro Original}}{\text{Número de eventos que conforman un micro - registro (n)}}$$

**Ecuación 2: Fórmula para calcular el número de micro registro.**

En nuestro caso la población será igual a:

$$\text{Número de Micro registro} = \frac{150000}{384} = 390.625 \text{ Micro registros}$$

### III.6.3 MUESTRA

Como nuestra población ideal está conformada por 390.625 Micro registros usaremos la misma fórmula usado antes (explicada en  $\alpha$ ) para obtener el tamaño de una muestra. De aquí se obtiene que la nueva muestra esté conformada por 194 micro- registros de eventos que generarán a su vez 194 modelos de procesos.

Estos micro- registros serán elegidos por muestreo probabilístico o aleatorio **“es decir como cada uno de ellos es independiente del otro, serán seleccionados al azar”**.

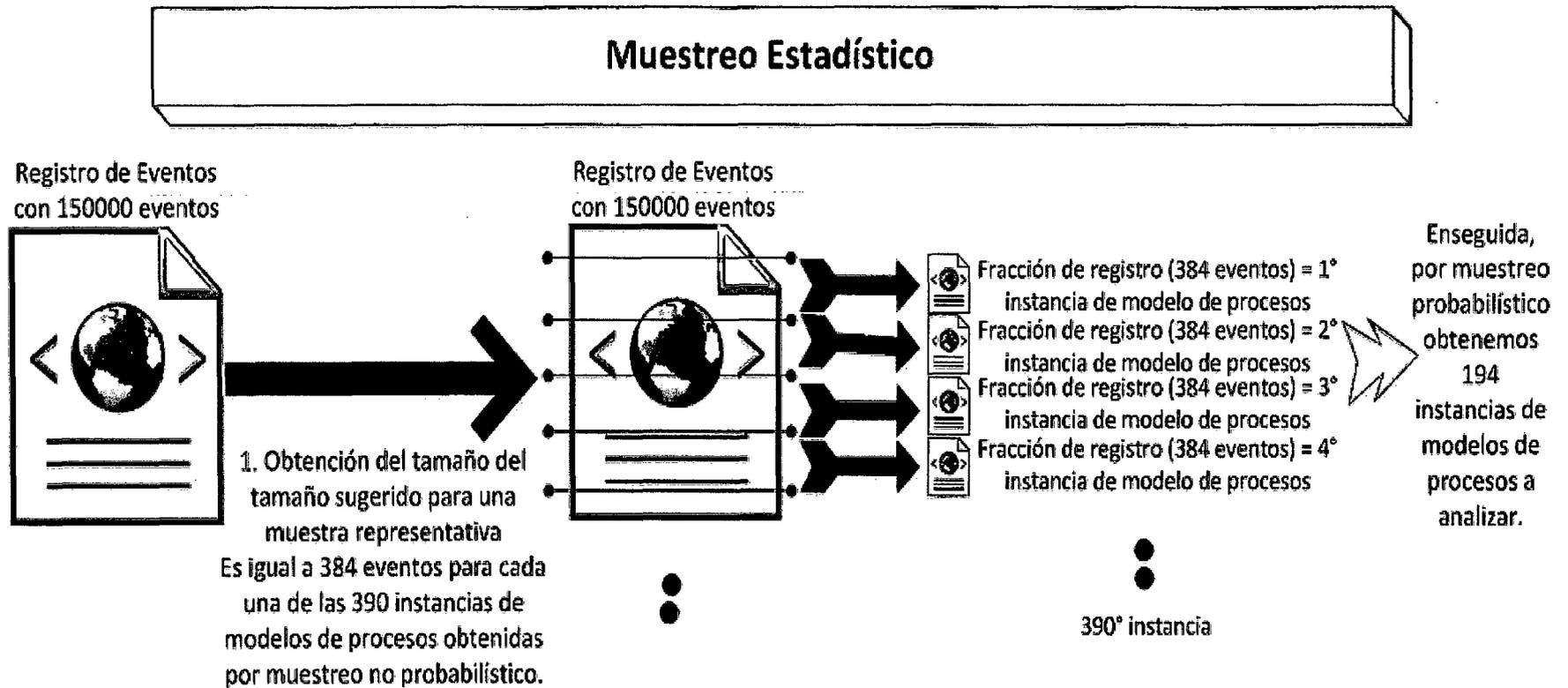
**Muestras probabilísticas o aleatorias:** Una muestra aleatoria o probabilística es aquella en la que todos los sujetos de la población han tenido la misma probabilidad de ser escogidos. En una muestra aleatoria tenemos más

seguridad de que se encuentran representadas las características importantes de la población en la proporción que les corresponde.

Finalmente, Para mantener la coherencia de las fracciones de registro cuando usemos la herramienta ProM para crear los modelos de procesos, se tomará en cuenta el filtrar aquellos eventos que se encuentran incompletos, es decir en los que no se puede generar todo el proceso desde el inicio hasta el final, ya que solo contiene información parcial del proceso. Por esto solo se tomarán en cuenta las instancias de procesos que contengan toda la secuencia de eventos desde el inicio hasta el final.

#### **III.6.4 ESQUEMA: “POBLACIÓN – MUESTRA”**

El esquema de obtención de la población y de la muestra para el estudio que ya ha sido antes explicado, se puede observar gráficamente a continuación



**Figura 8: Obtención de la muestra a analizar. Elaboración propia.**

### III.6.5 VARIABLES

#### III.6.5.1 VARIABLES INDEPENDIENTES

- El uso o no de la herramienta de minería de procesos ProM (Para la etapa 1).
- La edad de las personas (Para la etapa 2): Para [25;35] y >35 años.
- El nivel académico (Para la etapa 2), para esta variable se tienen los rangos de evaluación:
  - Universitario.
  - Maestría.
  - Doctorado.

#### III.6.5.2 VARIABLES DEPENDIENTES

**Para la etapa 1:** El número de elementos visibles en el modelo de procesos "V1" se define como:

**Número de elementos en el modelo de procesos =**

**Número de actividades + Número de patrones ( $\Sigma$  Patrones Secuencia + División Paralela + Sincronización + Elección Exclusiva + Unión Simple + Elección Múltiple + Estructura de Unión Sincronizada + Unión Múltiple + Discriminador Estructurado + Ciclo Arbitrario + Terminación Implícita + Múltiples Instancias Sin Sincronización + Múltiples Instancias Con Conocimiento a Priori del Número de Ejecuciones en la Etapa de Diseño + Múltiples Instancias Con Conocimiento en el Tiempo de Ejecución + Múltiples Instancias Sin Conocimiento en el Tiempo de Ejecución + Elección Diferida + Ejecución Intercalada + Milestone + Cancelación de Tareas + Cancelación del Caso)**

**Ecuación 3: Fórmula del número de elementos dentro de un modelo de procesos.**

Según apliquen.

**Para la etapa 2:** La claridad de un modelo de procesos generado por la herramienta de minería de procesos “V2” ProM, se puede medir según los artículos [21,22 ,23], por medio de los siguientes componentes:

- Complejidad:

- El número de nodos, patrones de diseño.
- La complejidad en la estructura de una proposición (proposiciones subordinadas entre paréntesis).
- El número de palabras fáciles y difíciles (más de 3 sílabas).
- El número de palabras desconocidas.
- El nivel de abstracción de los textos.

- Largo de los nombres (de las actividades):

- El largo de una palabra.
- El número de palabras en una oración.

- Presentación (de las actividades, patrones de diseño):

- El color, posición, talla.
- Las faltas de ortografía.
- El tipo de caracteres tipográficos.
- La composición o disposición del modelo en la página.

- Precisión o completitud del contenido de los modelos:

- El número de elementos en el modelo.
- La información que falta.
- La información ambigua.

La metodología de evaluación será la empleada en [2].

### III.6.5.3 INSTRUMENTOS DE MEDICIÓN

#### III.6.5.3.1 VARIABLES INDEPENDIENTES

**Para la etapa 1:** Nosotros escogemos cuando usar o no la herramienta de minería de procesos.

**Para la etapa 2:** La edad y el grado académico son medidos por medio de una pregunta en la encuesta.

#### III.6.5.3.2 VARIABLES DEPENDIENTES

**Para la etapa 1:** El número de elementos en el modelo se mide por medio de una inspección manual y los valores obtenidos son escritos en una ficha de resumen.

**Para la etapa 2:** La complejidad, el largo de los nombres, la presentación y la precisión completitud serán medidos por medio de un cuestionario.

#### III.6.4.3.3 DISEÑO DEL EXPERIMENTO

Aquí presentamos las combinaciones de las variables dependientes con las variables independientes:

Etapa 1:

	Etapa 1: Las entradas son los registros de eventos					
	Sin uso de la herramienta ProM			Con uso de la herramienta ProM		
Sujeto	Modelo de Procesos 1	Modelo de Procesos 2	...	Modelo de Procesos 1	Modelo de Procesos 2	...
V1						

Tabla 1: Tabla de la 1º etapa del diseño del experimento.

Etapa 2:

La evaluación de la claridad del modelo generado se hará por medio de una encuesta, disponible en:

[https://docs.google.com/forms/d/1Jy9RAoKPtn37YPkHcOLMioJ\\_8anJXx9slc82asPrnA/viewform](https://docs.google.com/forms/d/1Jy9RAoKPtn37YPkHcOLMioJ_8anJXx9slc82asPrnA/viewform)

En ella, se ha elegido para la evaluación el siguiente par de modelos: “Modelo creado manualmente o modelo A, modelo creado por ProM o modeloB” que provienen del mismo registro de eventos:

Para evitar los problemas de dificultad con la familiaridad con el tema, la encuesta se realizará a las personas que desarrollen una actividad en relación con la ingeniería de sistemas.

\*\*Las actividades en la encuesta, se tradujeron al español ya que anteriormente se hallaban en su idioma original: El holandés.

Modelo A: Es el modelo creado manualmente.

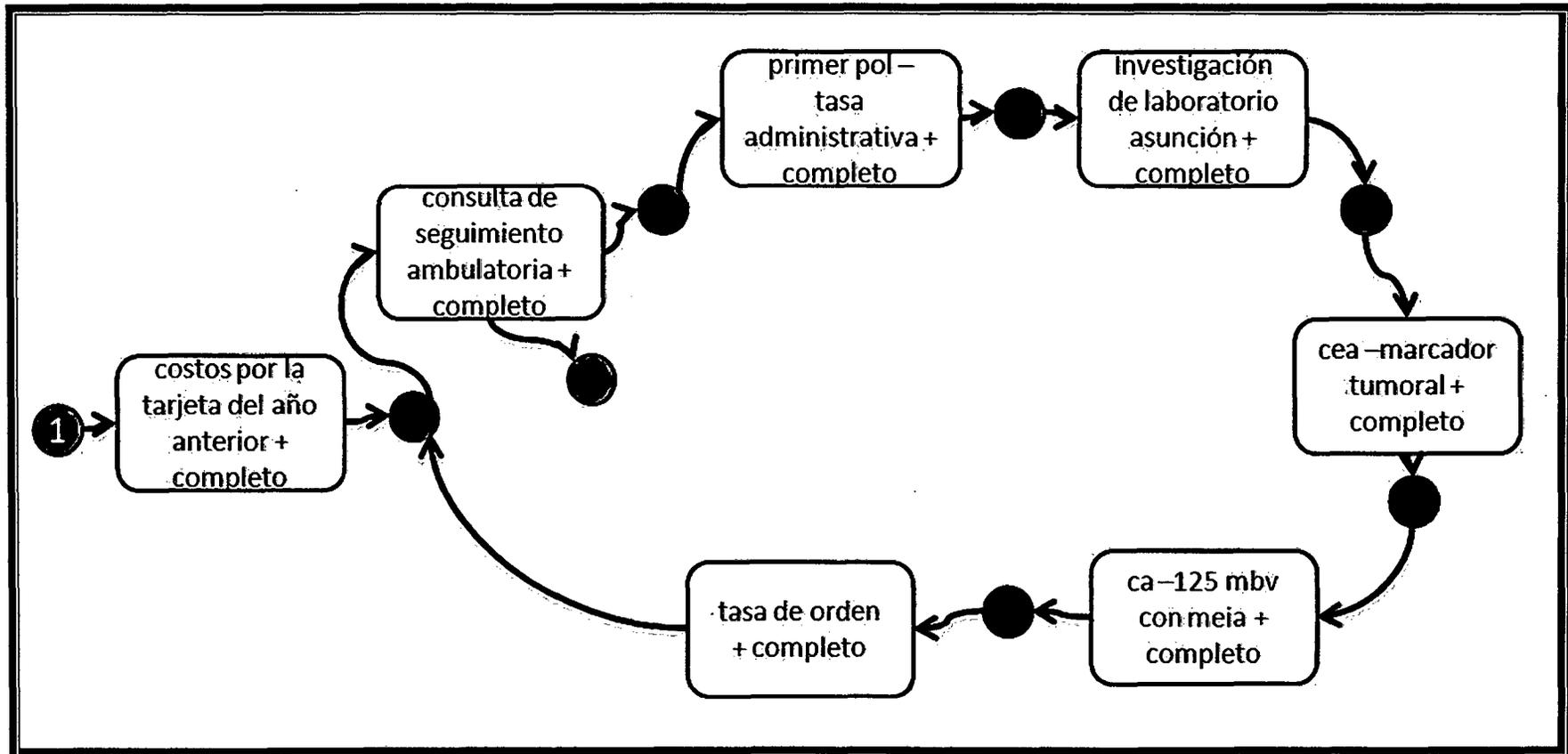
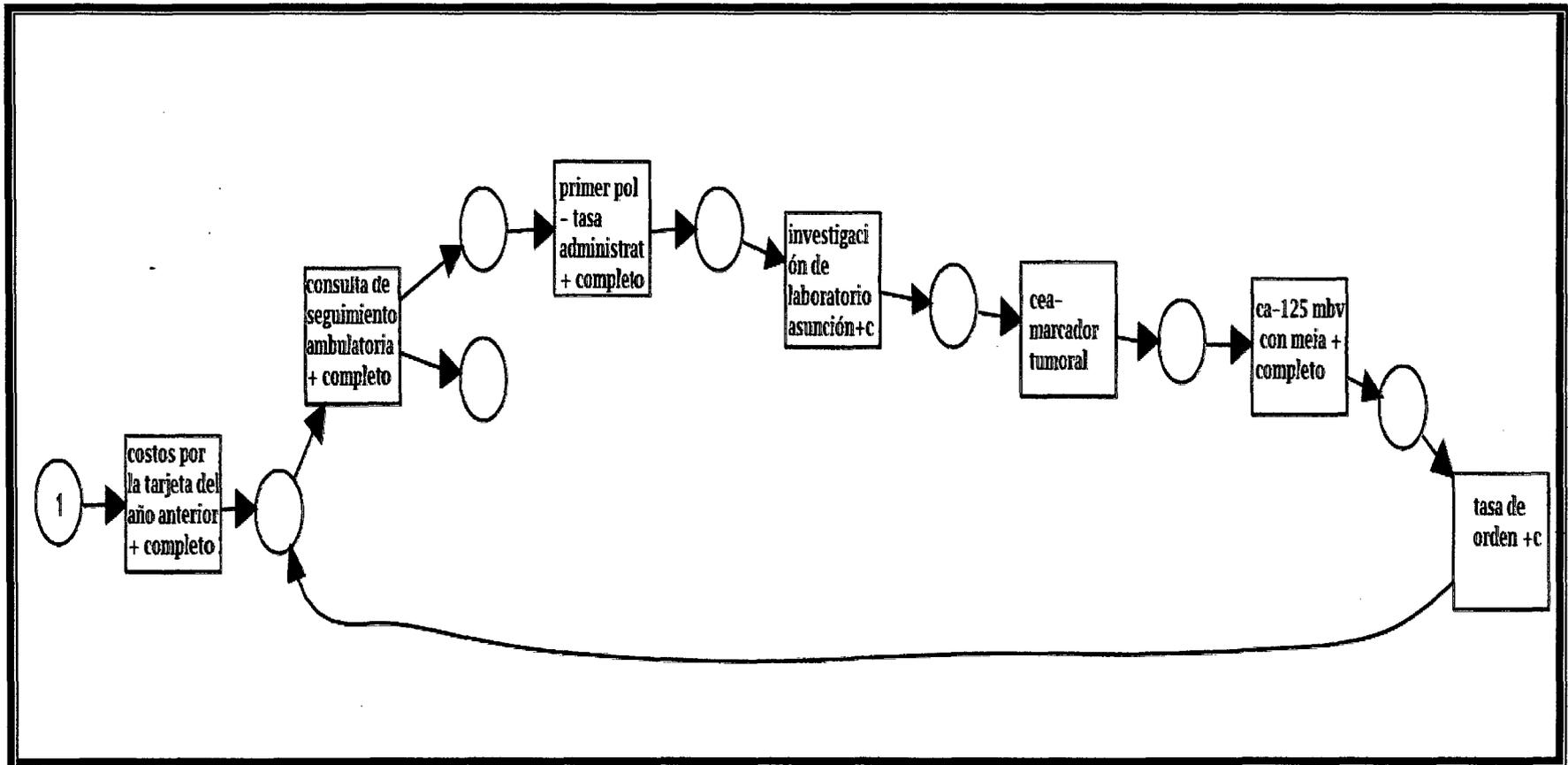


Figura 9: Modelo creado manualmente, empleado para la evaluación de la claridad. *Elaboración propia.*

**Modelo B: Es el modelo creado correctamente por ProM.**



**Figura 10: Modelo creado con ProM, empleado para la evaluación de la claridad. *Elaboración propia***

### **III.7 ANALISIS COSTO BENEFICIO**

La realización del presente trabajo trae innumerables beneficios entre los principales:

- La identificación de las principales deficiencias en las herramientas de generación de modelos por minería de procesos como ProM conllevará a la corrección de las mismas.
- La identificación de los elementos de la claridad de los modelos de minería de procesos permitirá su aplicación y la posterior obtención de modelos más claros.

En relación a costos incurridos para la realización del presente trabajo, no se tuvieron costos dado que en la investigación no se realiza ninguna implementación. Nuestra investigación se basa en el estudio de una herramienta gratuita, los datos son anónimos y publicados de manera formal en internet, los libros y artículos usados en el estudio fueron libres y obtenidos de fuentes formales.

## **CAPITULO VI. ANÁLISIS DE RESULTADOS**

### **IV.1 RESULTADOS DE LA ETAPA 1:**

Los valores indicados en amarillo indican las coincidencias en los modelos comparados.

Etapa 1: Las entradas son los registros de eventos.

Etapa 1: Las entradas son los registros de eventos.																		
Sin uso de la herramienta ProM										Con uso de la herramienta ProM								
Suje to:	MP1	MP2	MP3	MP4	MP5	MP6	MP7	MP8	MP9	MP1	MP2	MP3	MP4	MP5	MP6	MP7	MP8	MP9
V1	185	46	40	24	24	10	16	45	19	119	23	25	24	15	10	16	45	19
	MP10	MP11	MP12	MP13	MP14	MP15	MP16	MP10	MP11	MP12	MP13	MP14	MP15	MP16				
	12	15	41	21	26	16	12	12	15	33	14	26	13	12				
	MP17	MP18	MP19	MP20	MP21	MP22	MP23	MP17	MP18	MP19	MP20	MP21	MP22	MP23				
	23	13	25	14	9	43	15	23	13	13	13	9	25	6				
	MP24	MP25	MP26	MP27	MP28	MP29	MP30	MP24	MP25	MP26	MP27	MP28	MP29	MP30				
	10	15	17	7	9	14	15	10	6	17	7	9	6	15				
	MP31	MP32	MP33	MP34	MP35	MP36	MP37	MP31	MP32	MP33	MP34	MP35	MP36	MP37				
	17	20	27	10	42	29	40	9	20	17	10	21	29	27				
	MP38	MP39	MP40	MP41	MP42	MP43	MP44	MP38	MP39	MP40	MP41	MP42	MP43	MP44				
	25	12	17	11	27	12	16	25	12	9	5	27	12	13				
	MP45	MP46	MP47	MP48	MP49	MP50	MP51	MP45	MP46	MP47	MP48	MP49	MP50	MP51				
	33	17	13	12	14	9	10	24	9	6	12	14	9	10				
	MP52	MP53	MP54	MP55	MP56	MP57	MP58	MP52	MP53	MP54	MP55	MP56	MP57	MP58				
	7	13	10	28	18	32	15	7	13	10	28	7	13	6				
MP59	MP60	MP61	MP62	MP63	MP64	MP65	MP59	MP60	MP61	MP62	MP63	MP64	MP65					
29	14	19	9	15	15	25	29	6	15	9	6	6	20					

MP66	MP67	MP68	MP69	MP70	MP71	MP72	MP66	MP67	MP68	MP69	MP70	MP71	MP72
21	24	9	9	15	22	44	21	20	9	9	6	22	33
MP73	MP74	MP75	MP76	MP77	MP78	MP79	MP73	MP74	MP75	MP76	MP77	MP78	MP79
31	14	9	14	22	13	14	20	6	9	6	14	7	6
MP80	MP81	MP82	MP83	MP84	MP85	MP86	MP80	MP81	MP82	MP83	MP84	MP85	MP86
9	16	14	11	31	9	16	9	17	14	6	26	9	6
MP87	MP88	MP89	MP90	MP91	MP92	MP93	MP87	MP88	MP89	MP90	MP91	MP92	MP93
32	11	12	11	12	21	21	29	6	12	6	12	14	21
MP94	MP95	MP96	MP97	MP98	MP99	MP100	MP94	MP95	MP96	MP97	MP98	MP99	MP100
12	12	32	16	9	11	9	12	12	26	13	9	6	9
MP101	MP102	MP103	MP104	MP105	MP106	MP107	MP101	MP102	MP103	MP104	MP105	MP106	MP107
14	32	15	10	17	14	33	6	24	6	10	13	6	34
MP108	MP109	MP110	MP111	MP112	MP113	MP114	MP108	MP109	MP110	MP111	MP112	MP113	MP114
14	15	31	52	11	36	19	6	15	31	40	6	20	19
MP115	MP116	MP117	MP118	MP119	MP120	MP121	MP115	MP116	MP117	MP118	MP119	MP120	MP121
14	38	30	27	19	11	10	6	38	30	19	14	6	10
MP122	MP123	MP124	MP125	MP126	MP127	MP128	MP122	MP123	MP124	MP125	MP126	MP127	MP128
14	14	49	14	40	18	17	6	6	39	6	33	18	11
MP129	MP130	MP131	MP132	MP133	MP134	MP135	MP129	MP130	MP131	MP132	MP133	MP134	MP135
9	61	23	11	14	14	26	9	64	20	6	6	6	26

MP136	MP137	MP138	MP139	MP140	MP141	MP142	MP136	MP137	MP138	MP139	MP140	MP141	MP142
14	34	58	33	31	12	14	6	27	58	27	31	12	6
MP143	MP144	MP145	MP146	MP147	MP148	MP149	MP143	MP144	MP145	MP146	MP147	MP148	MP149
12	24	14	16	11	30	25	12	26	14	9	5	24	25
MP150	MP151	MP152	MP153	MP154	MP155	MP156	MP150	MP151	MP152	MP153	MP154	MP155	MP156
19	14	36	9	12	44	20	19	6	24	9	12	35	14
MP157	MP158	MP159	MP160	MP161	MP162	MP163	MP157	MP158	MP159	MP160	MP161	MP162	MP163
16	23	35	12	14	40	43	6	7	35	12	6	21	43
MP164	MP165	MP166	MP167	MP168	MP169	MP170	MP164	MP165	MP166	MP167	MP168	MP169	MP170
11	18	39	9	45	9	22	5	11	18	9	45	9	22
MP171	MP172	MP173	MP174	MP175	MP176	MP177	MP171	MP172	MP173	MP174	MP175	MP176	MP177
24	19	15	47	40	33	32	15	19	8	27	35	33	16
MP178	MP179	MP180	MP181	MP182	MP183	MP184	MP178	MP179	MP180	MP181	MP182	MP183	MP184
19	11	26	18	15	24	34	15	5	29	21	8	13	21
MP185	MP186	MP187	MP188	MP189	MP190	MP191	MP185	MP186	MP187	MP188	MP189	MP190	MP191
23	38	38	48	37	37	39	12	38	25	38	20	27	39
MP192	MP193	MP194					MP192	MP193	MP194				
37	13	24					37	13	24				

Tabla 2: Tabla de la 1º etapa con la recopilación de resultados del experimento.

En relación a los casos en que el modelo obtenido difiere del modelo original se tienen las siguientes observaciones:

**Patrones de Ciclo Arbitrario:**

- Se observa que para los eventos que deberían generar dos o más actividades consecutivas, donde cada una de ellas genere un respectivo ciclo arbitrario, al aplicar la minería de procesos se anularán los patrones sobre y entre las actividades, generándose actividades aisladas. Así por ejemplo, los eventos “Bacteriologisch onderzoek met kweek-nie” y “Resistentie bepalingen-5 bepalingen”, deberían haberse creado luego de la minería de procesos como se muestra a la izquierda, pero se crearon tal como se muestra a la derecha:

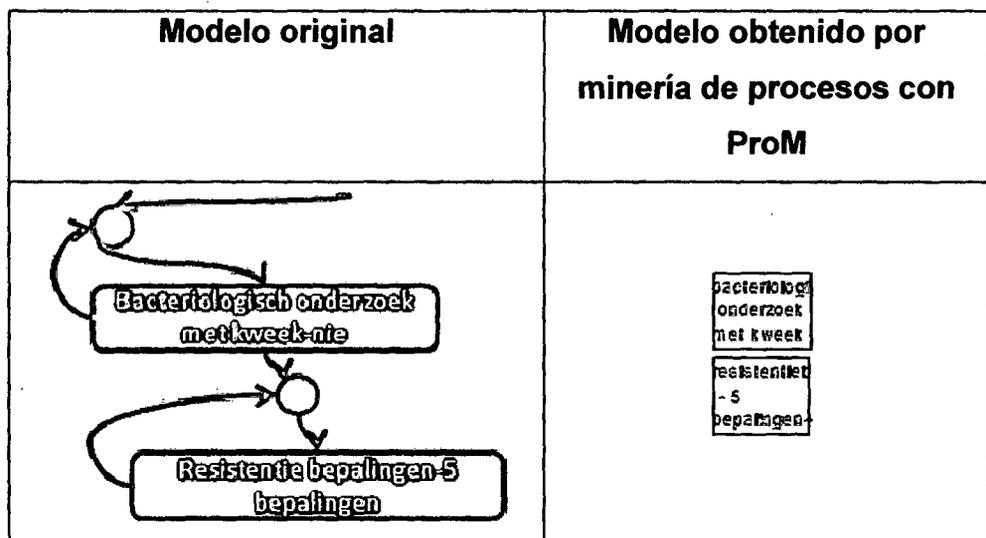


Figura 11: Consecuencia de los Ciclos Arbitrarios continuos.

*Elaboración propia.*

- Así, si debiera existir una actividad entre las dos actividades que tienen sus respectivos ciclos arbitrarios, al crearse las dos actividades de manera aislada, la actividad intermedia también se generará como una actividad aislada:

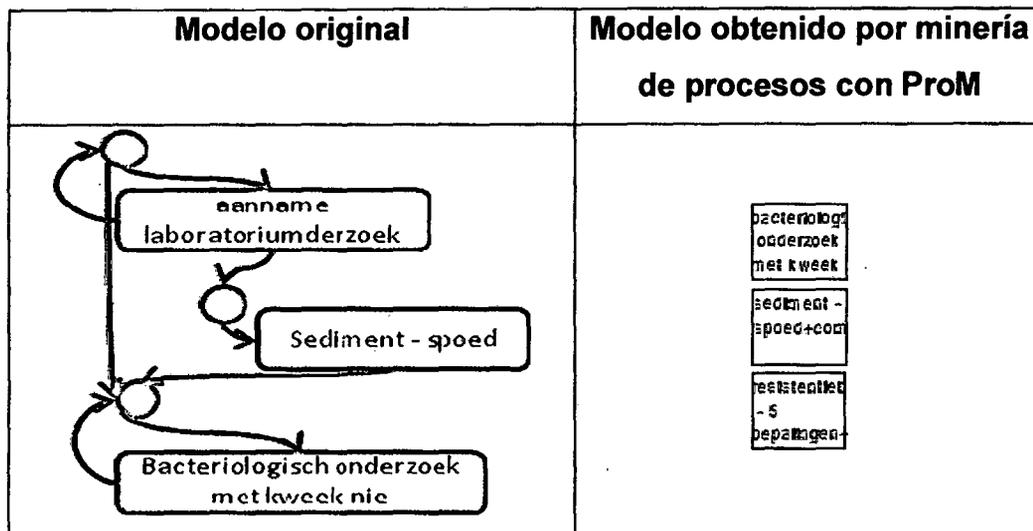


Figura 12: El efecto de los Ciclos Arbitrarios continuos en actividades intermedias. *Elaboración propia.*

Esto conlleva a que los: **“Patrones de Secuencia, de Elección y de Unión”** entre las actividades desaparezcan lo que origina un modelo creado de manera incompleta por minería de procesos.

⊗ Los ciclos arbitrarios mal generados son capaces de dividir un flujo de actividades por la mitad, tal como se observa:

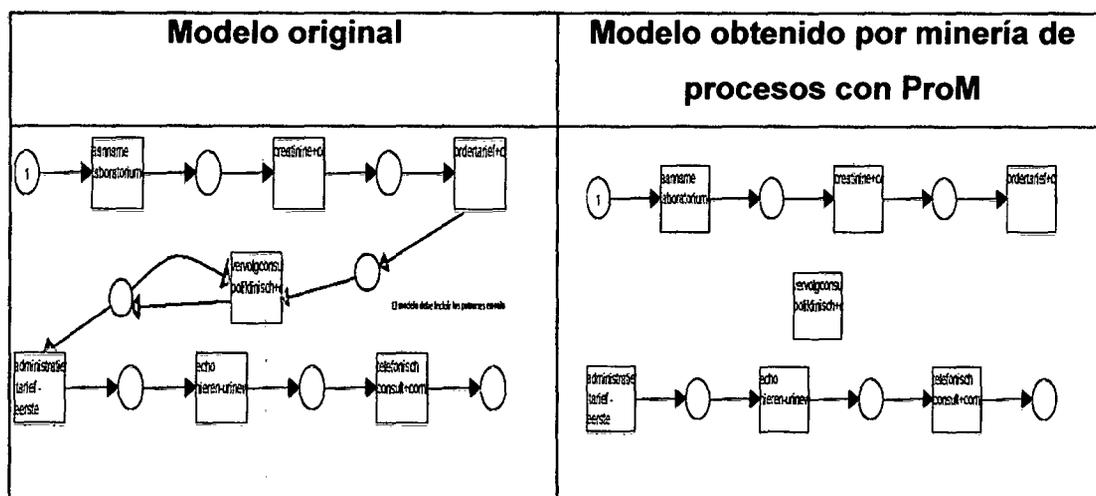


Figura 13: El efecto de los Ciclos Arbitrarios mal generados en el proceso global. *Elaboración propia.*

⊗ De tratarse una sola actividad con un ciclo arbitrario, los patrones alrededor de dicha actividad no se crearán completos. Así los bucles encima de una actividad que genera su propio bucle (como en la figura) no se crearán.

→ La lógica seguida por los nodos alrededor de los cuales se forma el ciclo arbitrario es que: si hay una única flecha entrante al nodo, está se mantendrá y también el ciclo arbitrario, pero si por el contrario hay alguna flecha saliente al nodo, está se perderá al igual que el ciclo arbitrario (a menos que al mismo tiempo exista una flecha entrante sobre el nodo).

→ Cualquier otro patrón de flujo que se debería crear sobre alguna actividad sobre la cual se crea un ciclo arbitrario, no se creará.

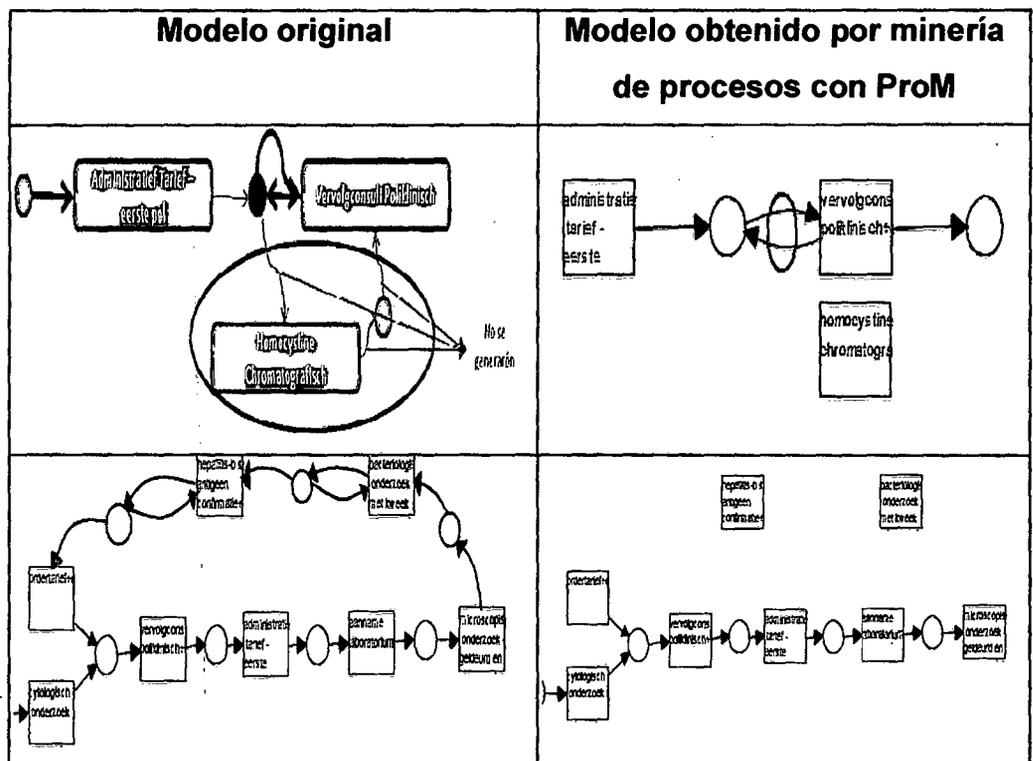
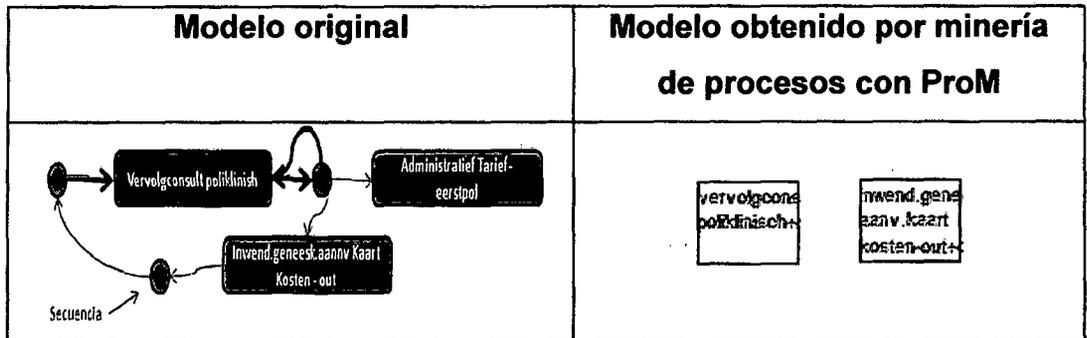


Figura 14: El efecto de los Ciclos Arbitrarios en cualquier otro patrón de flujo de trabajo. *Elaboración propia.*

**Patrones de Secuencia:**

- La aparición de ciclos arbitrarios cerca a Patrones de Secuencia conlleva a que estos no se generen adecuadamente:

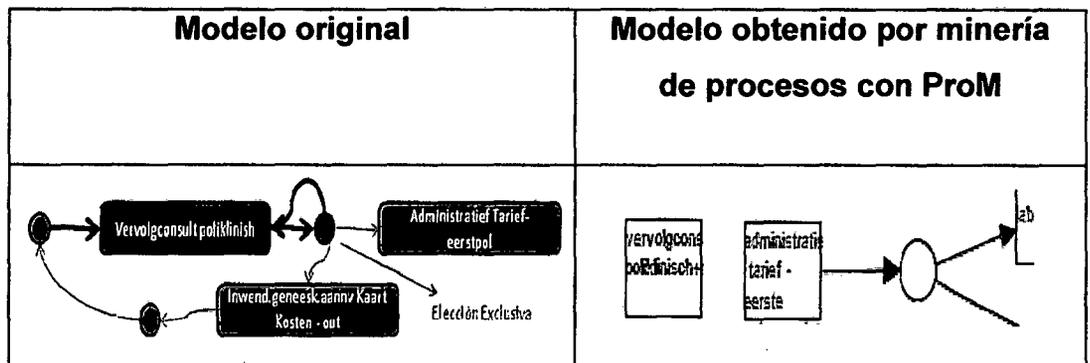


**Figura 15: Ciclos Arbitrarios y Patrones de Secuencia.**

*Elaboración propia.*

**Patrones de Elección:**

- La aparición de los ciclos arbitrarios alrededor de los Patrones de Elección los altera conllevando a que estos no se generen adecuadamente, como podemos observar en el siguiente ejemplo:



**Figura 16: Los Ciclos Arbitrarios y su efecto en los Patrones de Elección. *Elaboración propia.***

- Algunas veces se crean Patrones de Elección redundantes como se observa a continuación:

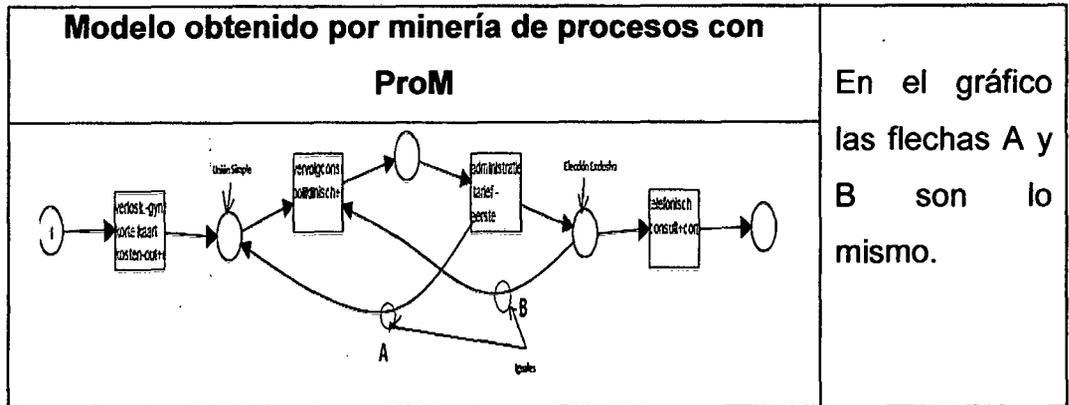


Figura 17: Patrones de Elección redundantes. *Elaboración propia.*

**Patrones de Unión Simple:**

- Los Patrones de Unión (tal como se observa en el gráfico), ciertas veces no se generan correctamente. En el gráfico la flecha que debería salir de dicha unión no se ha creado por estar posicionada sobre la actividad generadora del ciclo arbitrario.

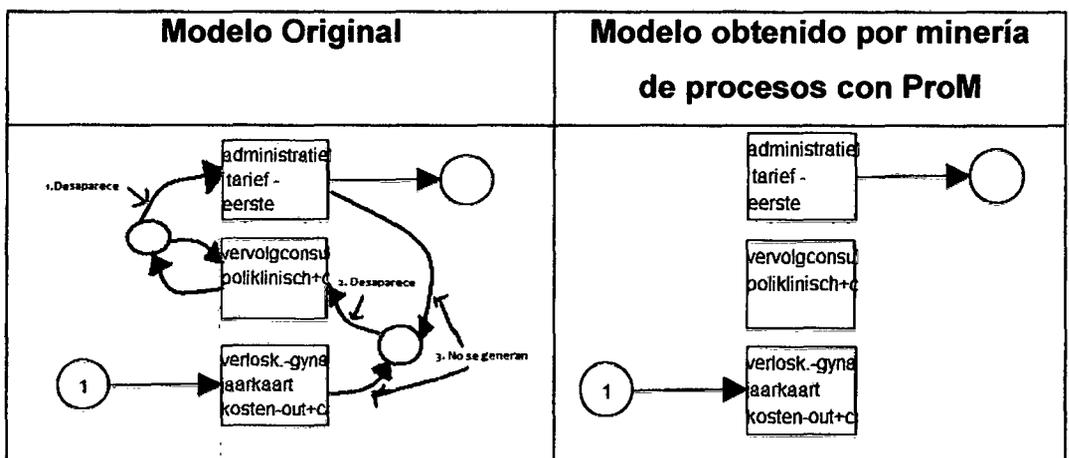


Figura 18: Ciclos Arbitrarios y Patrones de Unión Simple. *Elaboración propia.*

- ⊗ Algunas veces los Patrones de Unión Simple no son bien generados. En el ejemplo que se muestra a continuación se debió crear la flecha roja:

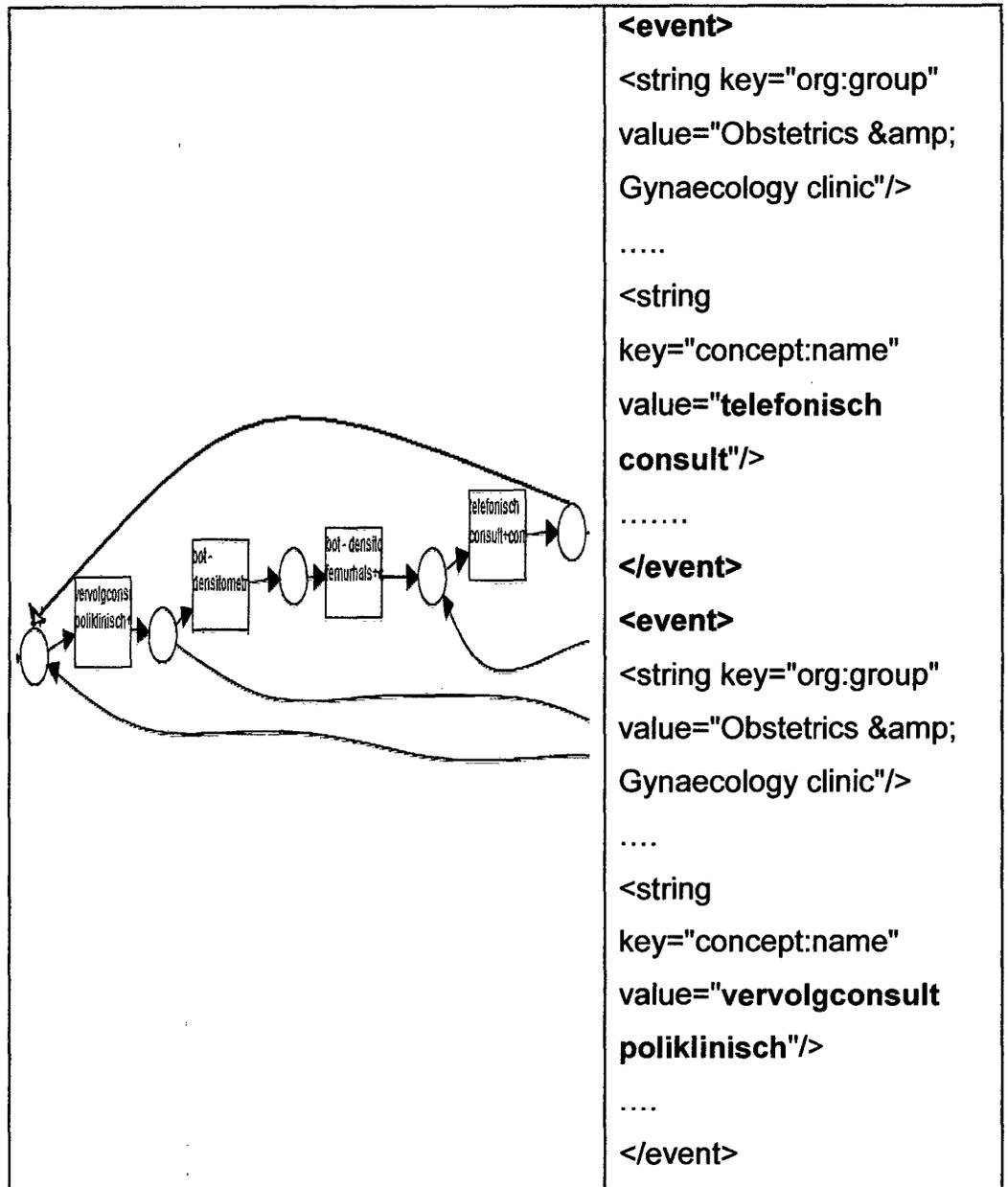


Figura 19: Patrones de Unión Simple incompletos. *Elaboración propia.*

- ⊗ Ocasionalmente, se crean Patrones de Unión redundantes cuando se crean flechas duplicadas:

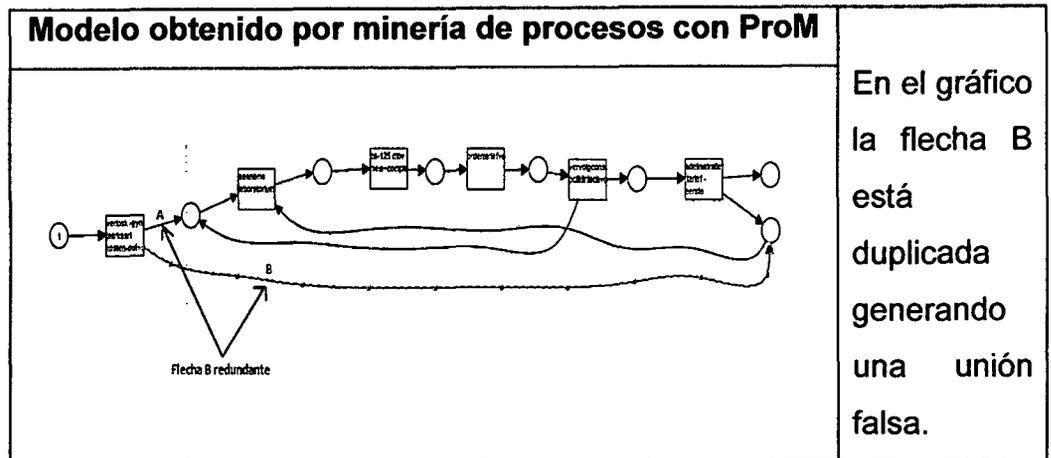


Figura 20: Patrones de Unión redundantes. *Elaboración propia.*

- La aparición de un bucle mal generado crea algunas veces Uniones Simples o Decisiones Exclusivas erróneas. A la izquierda se creó una flecha falsa lo que originó una Unión Simple irreal y a la derecha se muestra como se debieron haber generado correctamente los patrones de diseño:

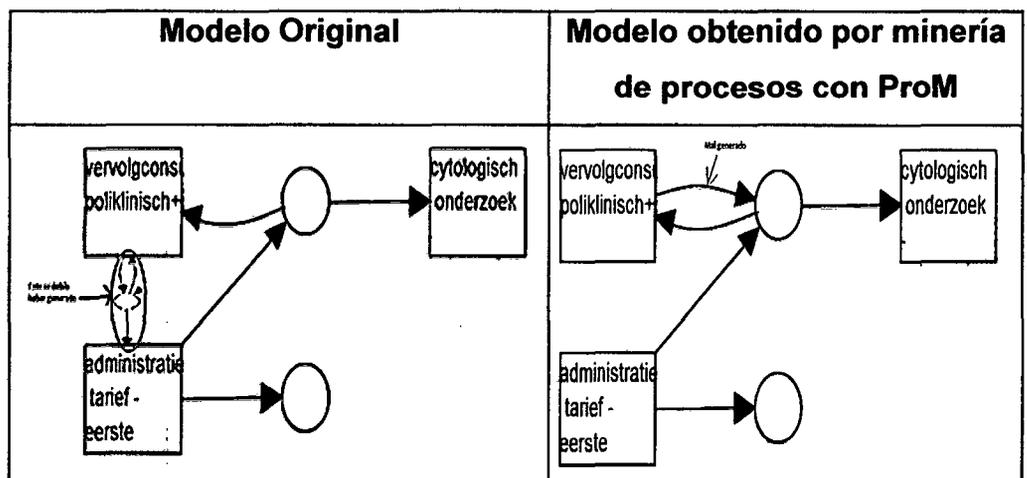


Figura 21: Uniones Simples irreales. *Elaboración propia.*

**Patrones de División Paralela y Sincronización:**

⊗ ProM crea con poca frecuencia, Patrones de Sincronización redundantes.

Por ejemplo:

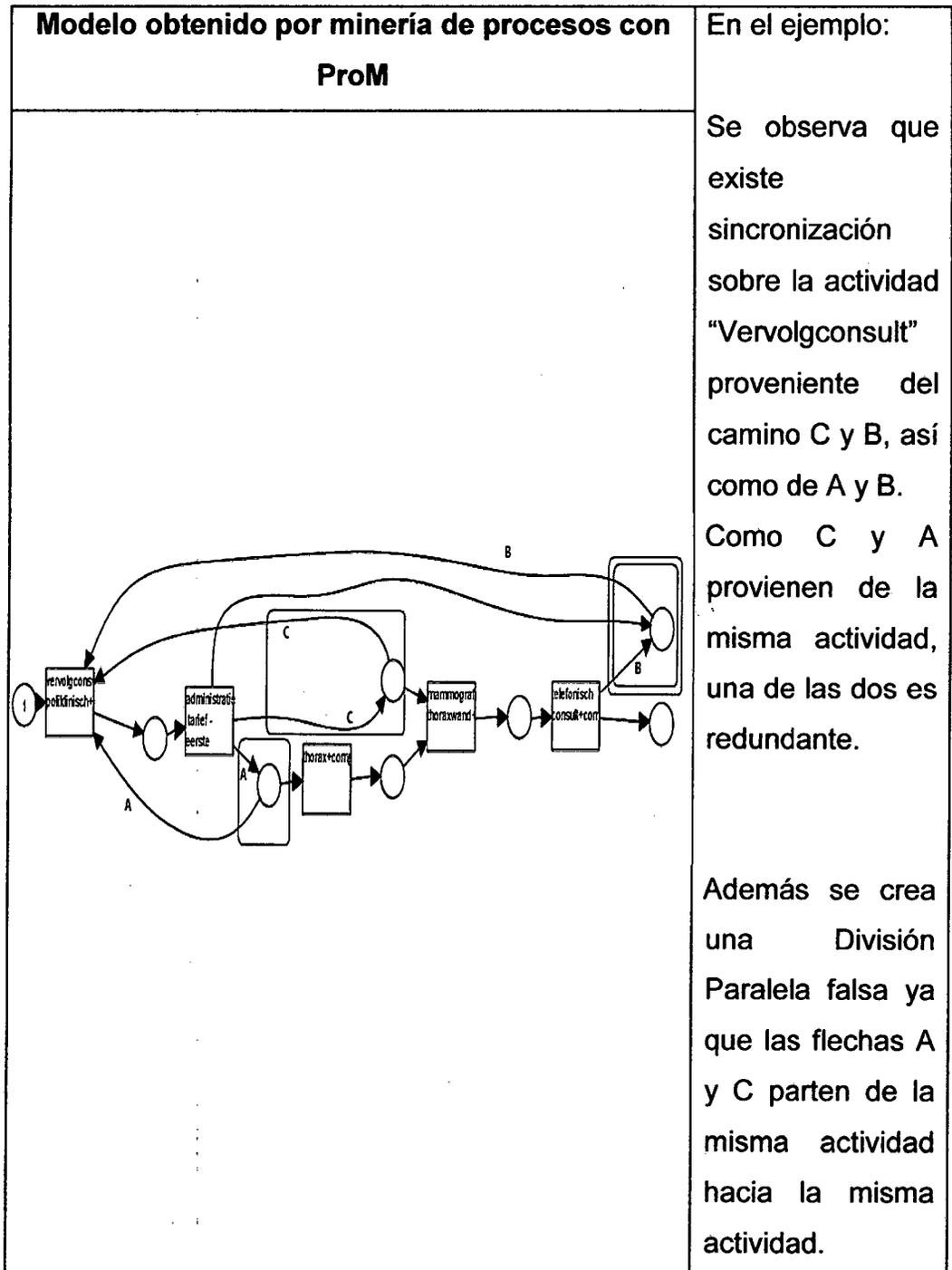
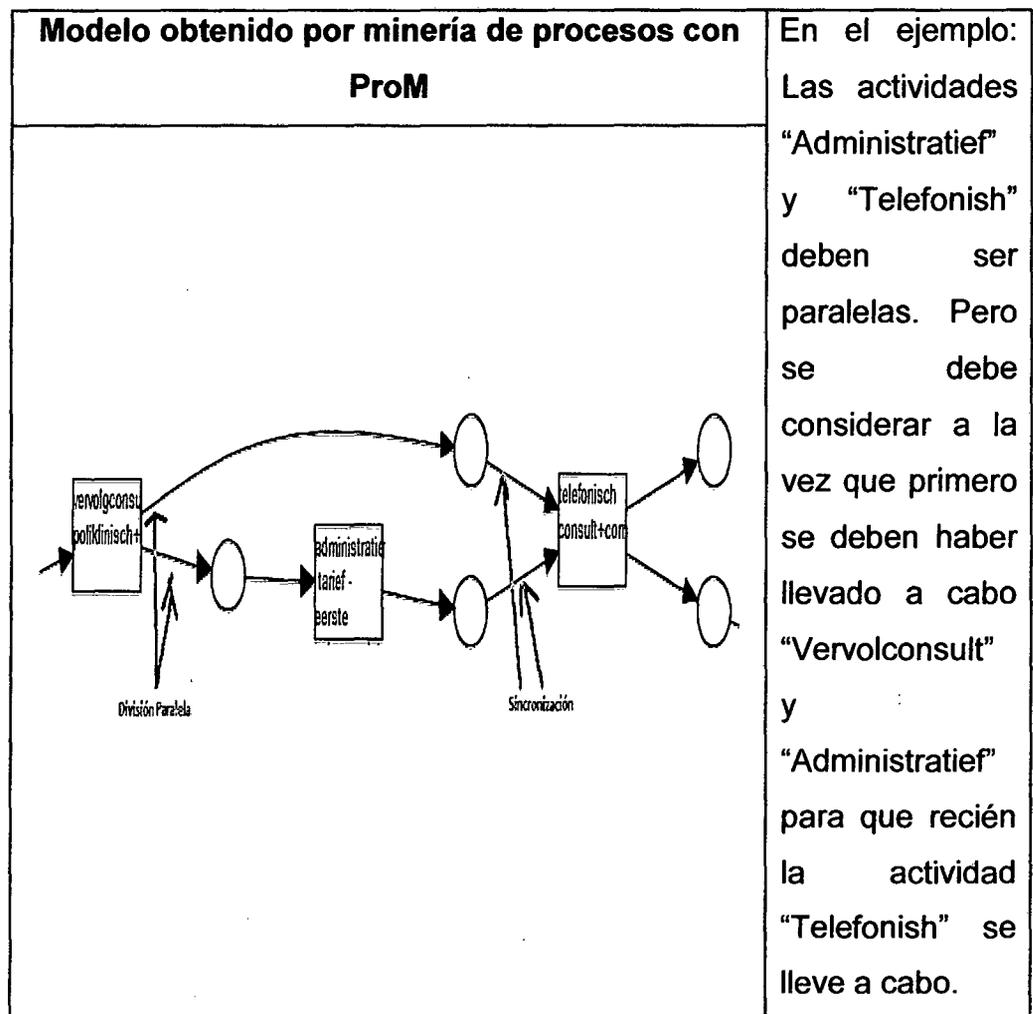


Figura 22: Patrones de Sincronización redundantes. *Elaboración propia.*

- ⊗ Algunas veces la herramienta crea los Patrones de División Paralela y Sincronización, pero no se pregunta si lo que se está creando es lógico o no. Por ejemplo:



**Figura 23: División Paralela y Sincronización incoherentes.**  
*Elaboración propia.*

#### IV.2 RESULTADOS DE LA ETAPA 2: CUESTIONARIO DE CLARIDAD

El cuestionario ha sido realizado a 29 personas que siguen la profesión de ingeniería de sistemas para evitar de esta manera una posible influencia de la variable independiente "tipo de profesión". Los resultados para cada

pregunta del cuestionario y el análisis respectivo se muestran a continuación.

**Pregunta: Complejidad de los modelos**

La complejidad se mide en relación a la complejidad en la estructura de las oraciones, el número de palabras simples y complejas (más de 3 sílabas), existencia de palabras desconocidas y el nivel de abstracción de los textos en el modelo.

PCMA, PCMB: Modelo A poco complejo, modelo B poco complejo.

MCMA, MCMB: Modelo A medianamente complejo, modelo B medianamente complejo.

TCMA, TCMB: Modelo A muy complejo, modelo B muy complejo.

**Complejidad del Modelo A y B: "Entre 25 - 35 años".**

	PCMA	MCMA	TCMA	PCMB	MCMB	TCMB
<b>Universitario</b>	8	10	4	6	14	2
<b>Maestría</b>	2	1	0	1	2	0
<b>Doctorado</b>	0	1	0	0	1	0

Tabla 3: Tabla de la 2º etapa con la recopilación de resultados de la evaluación de la complejidad, edad 25-35 años.

**Complejidad del Modelo A y B: "Mayor que 35 años".**

	MCMA	TCMA	MCMB	TCMB
<b>Universitario</b>	1	0	1	
<b>Maestría</b>	1	1	0	2

Tabla 4: Tabla de la 2º etapa con la recopilación de resultados de la evaluación de la complejidad, edad >35 años.

Se observa que:

- **Entre 25 y 35 años:** La mayoría de universitarios indican que si bien ambos modelos son medianamente complejos, el modelo B es más complejo que el modelo A. Para la mayoría de los magísteres el modelo A es poco complejo mientras que el modelo B es medianamente complejo. El único doctor encuestado señala que ambos modelos son medianamente complejos.
- **Más de 35 años:** El único entrevistado universitario indica que el modelo A y B son ambos medianamente complejos, mientras que los magísteres indican en consenso que el modelo B es muy complejo.

De lo anterior se puede deducir que si la variable grado académico se mantiene constante a [25; 35] para la mayoría de encuestados, el modelo B será más complejo que el modelo A (siendo para los universitarios y el doctor ambos modelos medianamente complejos, para los magísteres el modelo A poco complejo y el modelo B medianamente complejo). Mientras que para la mayoría de encuestados > 35 años el modelo B mucho más complejo que el modelo A.

#### **Pregunta: Número de elementos de los modelos**

PCMA, PCMB: Modelo A poco comprensible, modelo B poco comprensible.

MCMA, MCMB: Modelo A medianamente comprensible, modelo B medianamente comprensible.

TCMA, TCMB: Modelo A muy comprensible, modelo B muy comprensible.

**Comprensión del Modelo A y B: "Entre 25 - 35 años".**

	PGMA	MGMA	TCMA	PGMB	MGMB	TCMB
<b>Universitario</b>	4	15	3	7	13	2
<b>Maestría</b>	1	1	1	1	1	1
<b>Doctorado</b>		1			1	

**Tabla 5:** Tabla de la 2º etapa con la recopilación de resultados de la evaluación de la comprensión del número de elementos del modelo, edad 25-35 años.

**Comprensión del Modelo A y B: "Mayor que 35 años".**

	PGMA	MGMA	PGMB	TCMB
<b>Universitario</b>		1		1
<b>Maestría</b>	1	1	2	

**Tabla 6:** Tabla de la 2º etapa con la recopilación de resultados de la evaluación de la comprensión del número de elementos del modelo, edad >35 años.

Se observa que:

- **Entre 25 y 35 años:** La mayoría de universitarios indican que si bien ambos modelos son medianamente comprensibles, el modelo A es más comprensible que el modelo B. Para el único doctor encuestado ambos modelos resultan medianamente comprensibles.
- **Más de 35 años:** El único universitario encuestado indica que el modelo A es menos comprensible que el modelo B, mientras que los magísteres indican en consenso que el modelo B es poco comprensible.

De lo anterior se puede deducir que para la mayoría de entrevistados entre 25-35 años el modelo A será más comprensible que el modelo B. Mientras

que para los magísteres mayores de 35 el modelo B es poco comprensible, siendo A muy comprensible para algunos de entre ellos.

**Pregunta: Presentación de modelos**

PPMA, PPMB: Modelo A poco presentable, modelo B poco presentable.

MPMA, MPMB: Modelo A medianamente presentable, modelo B medianamente presentable.

TPMA, TPMB: Modelo A muy presentable, modelo B muy presentable.

**Presentación del Modelo A y B: "Entre 25 - 35 años".**

	PPMA	MPMA	TPMA	PPMB	MPMB	TPMB
<b>Universitario</b>	2	12	8	14	6	2
<b>Maestría</b>		1	2	2	1	
<b>Doctorado</b>			1		1	

Tabla 7: Tabla de la 2º etapa con la recopilación de resultados de la evaluación de la presentación del modelo, edad 25-35 años.

**Presentación del Modelo A y B: "Mayor que 35 años".**

	MPMA	TPMA	PPMB
<b>Universitario</b>	1		1
<b>Maestría</b>	1	1	2

Tabla 8: Tabla de la 2º etapa con la recopilación de resultados de la evaluación de la presentación del modelo, edad >35 años

Se observa que:

- **Entre 25 y 35 años:** La mayoría de encuestados indican que el modelo A es más presentable que el modelo B. Siendo para los universitarios el

modelo A medianamente presentable y para los magísteres y el doctor un modelo muy presentable.

- **Más de 35 años:** El único universitario encuestado indica que el modelo A (considerado por él como medianamente presentable) es más presentable que el modelo B (considerado por él como poco presentable). Mientras que los magísteres indican en consenso que el modelo B es poco presentable, llegando a ser para ellos el modelo A algunas veces muy presentable.

De lo anterior se puede deducir que para la mayoría de encuestados entre 25-35 y > 35 años el modelo A será más presentable que el modelo B.

#### **Pregunta: Contenido de los modelos**

PCMA, PCMB: Modelo A poco comprensible, modelo B poco comprensible.

MCMA, MCMB: Modelo A medianamente comprensible, modelo B medianamente comprensible.

TCMA, TCMB: Modelo A muy comprensible, modelo B muy comprensible.

#### **Contenido del Modelo A y B: "Entre 25 - 35 años".**

	PCMA	MCMA	TCMA	PCMB	MCMB	TCMB
<b>Universitario</b>	10	8	4	12	8	2
<b>Maestría</b>		3		1	2	
<b>Doctorado</b>		1			1	

**Tabla 9: Tabla de la 2º etapa con la recopilación de resultados de la evaluación del contenido del modelo, edad 25-35 años**

#### **Contenido del Modelo A y B: "Mayor que 35 años".**

	PCMA	MCMA	PCMB	MCMB
Universitario		1		1
Maestría	1	1	2	

Tabla 10: Tabla de la 2º etapa con la recopilación de resultados de la evaluación del contenido del modelo, edad >35 años

Se observa que:

- **Entre 25 y 35 años:** La mayoría de universitarios encuestados indican que ambos modelos son poco comprensibles, siendo A más comprensible que B. Para la mayoría de los magisteres y el doctor ambos modelos son medianamente comprensibles.
- **Más de 35 años:** Para el único universitario encuestado, ambos modelos son medianamente comprensibles. Los magisteres señalan en consenso que el modelo B es poco comprensible, llegando algunas veces a ser el modelo A para ellos muy comprensible.

De lo anterior se puede deducir que para los universitarios encuestados entre 25-35 el modelo A es más comprensible que B siendo ya ambos de por sí poco comprensibles. La comprensión de los modelos se incrementa a medida que el grado académico aumenta. Para los magisteres mayores de 35 años el modelo A será más comprensible que el modelo B que es considerado en consenso como poco comprensible.

#### IV.3 ANÁLISIS DE LA HERRAMIENTA

- En relación a la fenomenología hermenéutica de los procesos con la minería de procesos, si bien [8] un log de eventos representa el

comportamiento de un proceso en un intervalo de tiempo, sabemos que el comportamiento de un proceso no es único y varía con el tiempo, por lo que un log de eventos contiene solo una fracción de todos los posibles comportamientos. Es por ello que para que un modelo de procesos que ha sido generado por medio de una herramienta de minería de procesos sea fiable es necesario que se cumplan previamente a la generación del modelo ciertas principios:

1. Los datos de los eventos deben ser de calidad.
2. La extracción de los eventos debe ser hecha como consecuencia a ciertas preguntas bien definidas.
3. El lenguaje de modelización de procesos elegido debe soportar como mínimo los conceptos básicos de control de flujo.
4. Se deben eliminar los eventos ambiguos.
5. Se debe tener en claro que no existe un solo modelo (representación de la realidad) de procesos se pueden obtener modelos en diferentes perspectivas, granularidades y precisión.
6. La minería de procesos es un proceso continuo, como los procesos cambian mientras son estudiados, por lo que no se debe ver la minería de procesos como una actividad puntual para obtener un modelo fijo sino que se debe usar de manera constante.

☉ Ahora, dado que la minería de procesos proveerá diferentes versiones o instancias de procesos según la traza de eventos tomada proveniente de un intervalo de tiempo dado, la mejor visión consenso del proceso será aquella que haya sido obtenida bajo los principios anteriormente mencionados en [11] y será aquel proceso que [11]: Luego de ser descubierto por minería de procesos se le haya dado conformidad (es decir que a partir del proceso ya obtenido se compare el mismo con el registro de eventos que original) y mejora (a partir del log de eventos obtenido se mejora el modelo de procesos inicial) a manera de obtener el

modelo más similar a la realidad. Esto dado que [8] no existe una clara relación entre el tamaño de un modelo y su comportamiento global.

Ahora, algunas limitaciones en el diseño de ProM (algoritmo Alpha) y la Minería de Procesos en forma de redes Petri son:

- Las redes Petri que son creadas por el algoritmo Alpha solo soportan: La Sucesión Directa (Patrón de Secuencia), la Causalidad (Patrón de Decisión Exclusiva), el Paralelismo (Patrón de División Paralela), la Elección (Patrón de Sincronización) [19]. Por lo que se analizaron a detalle estos patrones, además del Patrón de Ciclo Arbitrarios y la Terminación Implícita.

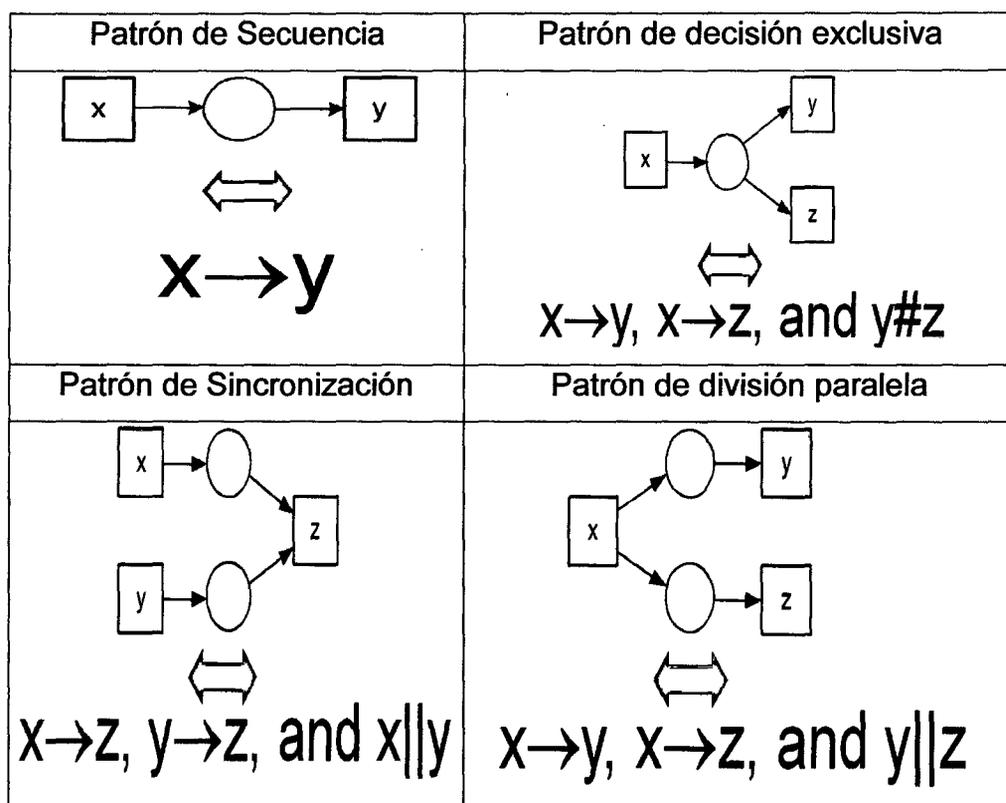


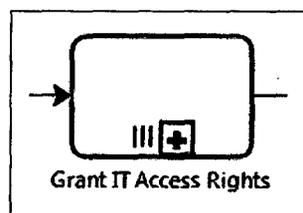
Figura 24: Patrones soportados por el algoritmo Alpha - ProM usado en minería de procesos. *Elaborado por Wil Van Der Aalst [19].*

- En cuanto a los patrones de flujo de trabajo que no son soportados en los modelos de procesos creados por minería de procesos con ProM:

Patrones Relacionados a Múltiples Instancias, los siguientes:

- Múltiples Instancias Sin Sincronización.
- Múltiples Instancias Con Conocimiento a Priori del Número de Ejecuciones en la Etapa de Diseño.
- Múltiples Instancias Con Conocimiento en el Tiempo de Ejecución.
- Múltiples Instancias Sin Conocimiento en el Tiempo de Ejecución.

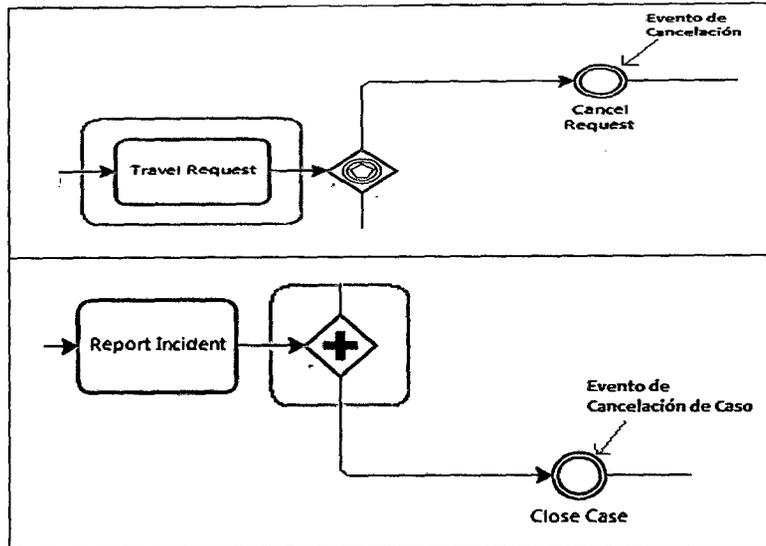
Es probable que estos no se encuentren implementados, ya que la implementación de los mismos necesitaría de un artefacto de diseño que indique que una actividad ha sido instanciada múltiples veces. Tal como en la herramienta de flujo de trabajo BizAgi:



Patrones de Cancelación:

- Cancelación de Tareas.
- Cancelación del Caso.

Probablemente no se encuentran debido a que para su implementación se necesita de un artefacto de diseño en forma de evento que nos indique que se está cancelando una actividad, tal como sucede en la herramienta de flujo de trabajo BizAgi:

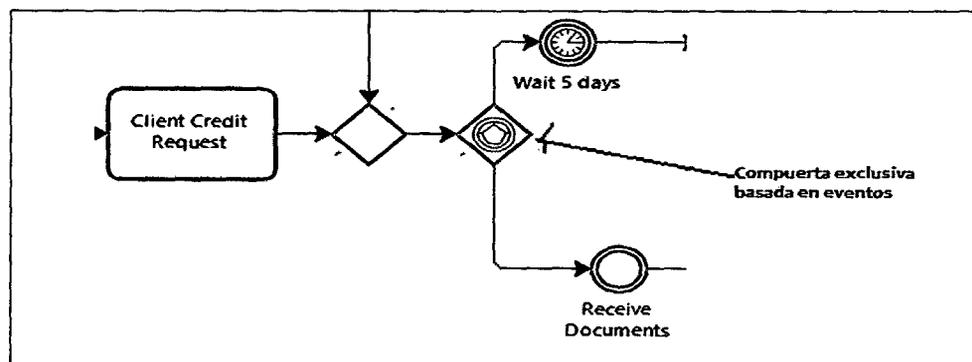


**Figura 25: Artefactos de diseño: Eventos de Cancelación en Bizagi.**  
*Elaborado por Bizagi [26].*

Patrones Basados en Estados:

- Elección Diferida.

Para su implementación es necesario un artefacto de diseño en forma de una compuerta exclusiva basada en eventos que permita que solo una de las ramas “la elegida” se ejecute a la vez ó que una de las ramas se deje de usar cuando la otra rama llega al valor hito. Tal como en Bizagi:



**Figura 26: Artefactos de diseño: Compuertas basadas en eventos en Bizagi.** *Elaborado por Bizagi [26].*

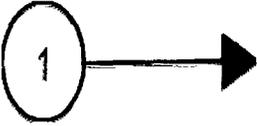
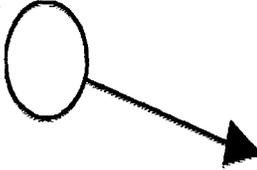
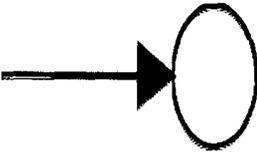
- Ejecución Intercalada.

Es probable que este no se encuentre implementado, ya que la implementación del mismo necesitaría de un artefacto de diseño como una compuerta inclusiva que habilite la actividad seleccionada de un conjunto de actividades para que se ejecute, ejecutándose una a la vez.

- ☉ Se observa que existe una diferencia entre en enfoque PETRI y BPMN, esto puede ser porque [30] no es posible construir una Red de Petri en forma precisa e inequívoca sobre la base de Diagramas BPMN, ya que algunas operaciones se pueden realizar de forma automática y pueden simplificar las actividades posteriores.

Además se sabe que [31] las Redes PETRI proporcionan un método formal para describir el comportamiento. Una red de PETRI se compone de lugares y transiciones, conectados por arcos dirigidos. Existen las denominadas redes clásicas de PETRI, donde estos son los únicos elementos permitidos. Sin embargo, las Redes de PETRI se pueden extender para ser capaces de expresar un comportamiento más.

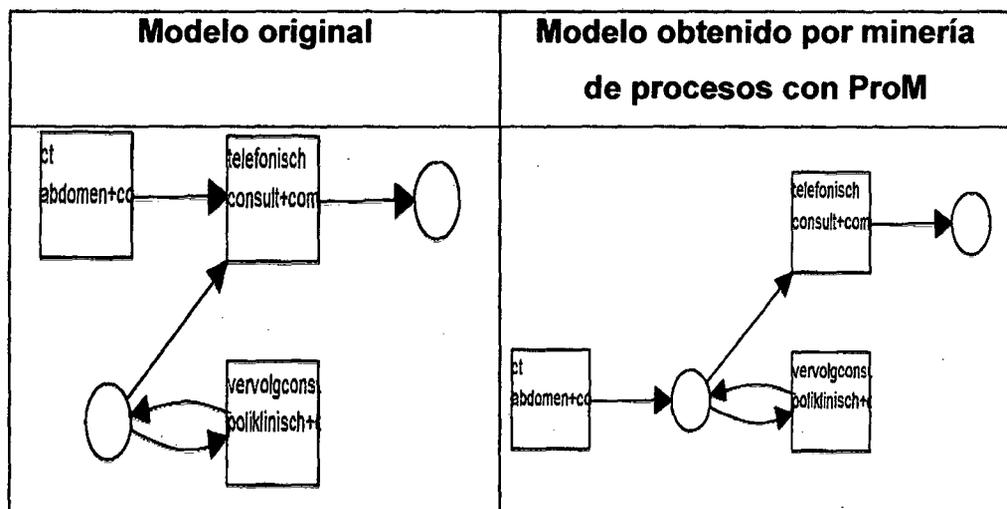
- ☉ Si bien el nodo de inicio del flujo de actividades en el diagrama de procesos es diferente a los demás nodos por lo que es fácil de diferenciar, el nodo que indica el final del flujo de actividades no presenta un formato diferente al de los otros nodos, generando confusión en la lectura del diagrama generado por minería de procesos:

Nodo Inicial	Nodo dentro del Flujo de Actividades	Nodo Final
		

**Figura 27: Sugerencias sobre el diseño del nodo Final. *Elaboración propia.***

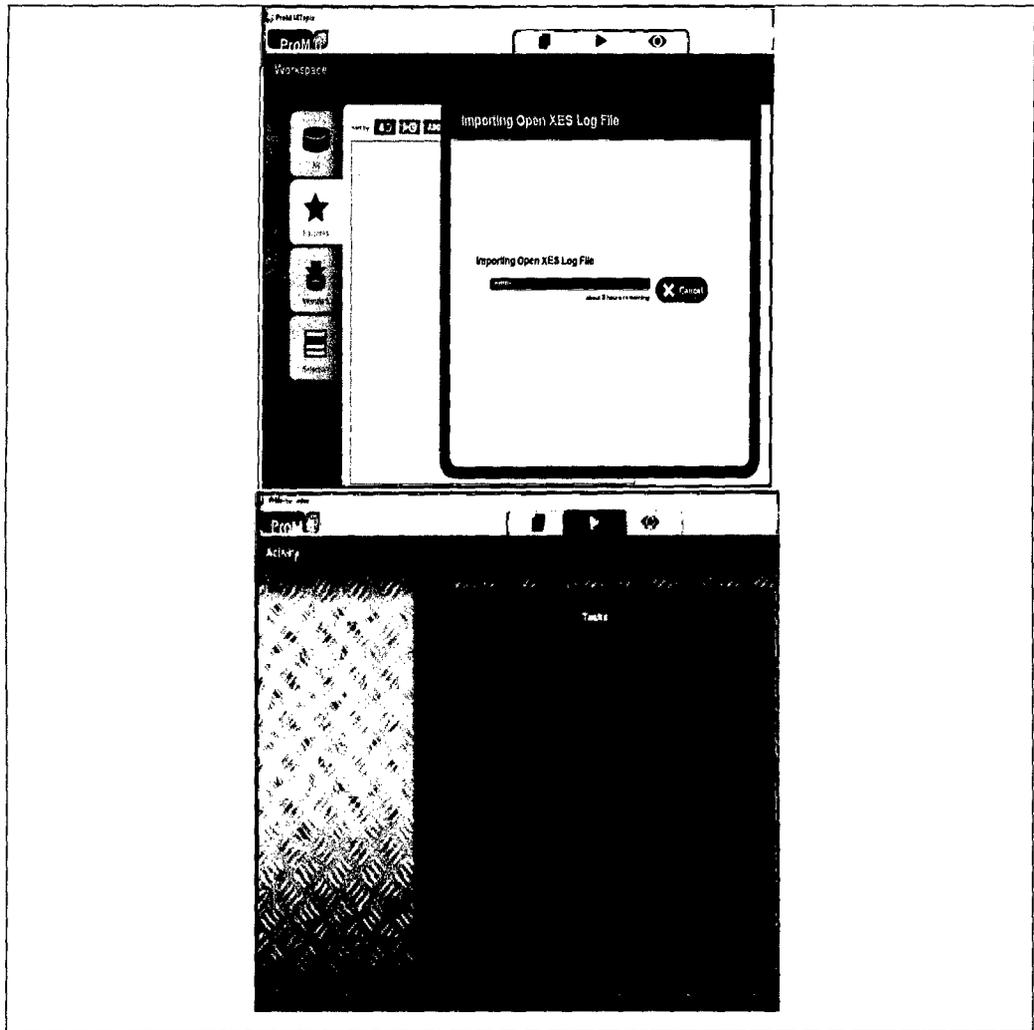
- ❶ La herramienta algunas veces genera modelos que si bien representan el registro de eventos, adicionalmente contienen nuevos flujos de actividades no incluidos.

En el ejemplo, según el registro de eventos la actividad “Ct-abdomen” debe ser seguida por la actividad “Telefonish consult”, tal como se observa en el modelo generado por ProM. Sin embargo ProM no ha escogido la mejor opción al colocar la actividad “Ct-abdomen” junto al nodo de la actividad “Vervolgconsult” dado que esto nos indica que luego de pasar por la actividad “Ct abdomen” se pueda seguir con la actividad “Vervolgconsult” lo que no se indica en el registro de eventos.



**Figura 28: Flujos de actividades adicionales en el modelo generado por ProM. *Elaboración propia.***

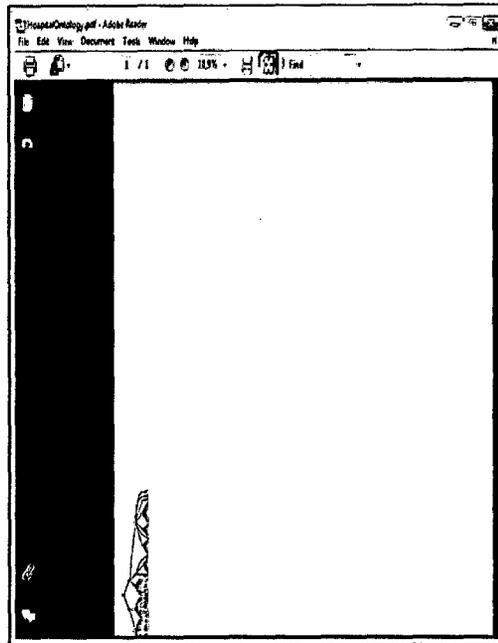
- ❷ La herramienta ProM tiene dificultades en la realización de la minería de procesos de un registro de eventos de gran tamaño (más de 150 000 eventos). La importación de dicho registro de eventos llevará varios minutos y aún más la creación del modelo haría que ProM se quede colgado durante muchas horas, incluso hasta días.



**Figura 29: Dificultades de ProM con registros de eventos de inmenso tamaño. *Elaboración propia.***

Se debe hacer necesariamente empleo de filtros automáticos o manuales previamente. Los filtros de ProM se limitan a seleccionar eventos por atributo (número de caso, área de tratamiento, etc.) pero no se pueden aplicar filtros más avanzados como a los patrones de flujo de trabajo.

☉ Si el tamaño del registro es muy grande se generará un modelo generalmente incomprensible ó incompleto. Como se muestra a continuación:



**Figura 30: Modelos incompletos provenientes de registros de eventos de gran tamaño. *Elaboración propia.***

Por otro lado, si el registro es muy pequeño se creará un gráfico incompleto ya que solo mostrará una pequeña parte de lo que pasa en la realidad.

- ProM hace la asunción de que los registros recibidos en formato XES se encuentran ordenados de manera fiel a la realidad, pero: ¿Esto realmente es siempre cierto?. Supongamos que si por algún problema técnico, el que debió ser el evento inicial no se pudo guardar sino hasta mucho después, entonces esto podría haber generado una incoherencia que sería difícil de identificar entre una multitud de eventos dentro del registro de eventos.
- Si el registro de eventos recibido se toma completo sin filtrar las actividades incompletas, entonces no se reflejará la realidad en el modelo obtenido por minería de procesos. Por ejemplo: Si luego del corte, el último evento que quedo capturado dentro del registro de eventos era en

realidad la primera actividad del proceso, no será cierto que ese último evento sea la última actividad a realizar tal como lo mostrará el flujo de actividades creado por ProM.

#### IV.4 VALIDACIÓN DE LA HIPÓTESIS

Para validar las hipótesis, primeramente buscamos conocer las principales características de la muestra con la finalidad de poder elegir el tipo de test a usar.

Las características la muestra en el experimento son las siguientes:

- ☉ Como cada individuo tiene un valor en la primera muestra (el modelo elaborado manualmente) y otro en la segunda (el modelo obtenido con ProM), las muestras se denominarán: "Apareadas o dependientes".
- ☉ Los datos son cuantitativos.
- ☉ Las muestras tienen un tamaño (n), donde  $n > 30$ .
- ☉ Para obtener el valor de las varianzas, usamos la herramienta estadística gratuita R (la podemos descargar de su página web en [27]):

```
>MuestraModelosSinProM=c(135,46,40,24,24,10,16,45,19,12,15,41,21,26,16,12,28,13,25,14,9,43,15,10,15,17,7,9,14,15,17,20,27,10,42,29,40,25,12,17,11,27,12,16,33,17,13,12,14,9,10,7,13,10,28,18,32,15,29,14,19,9,15,15,26,21,24,9,9,15,22,44,31,14,9,14,22,18,14,9,16,14,11,31,9,16,32,11,12,11,12,21,21,12,12,32,16,9,11,9,14,32,15,10,17,14,33,14,15,31,52,11,36,19,14,38,30,27,19,11,10,14,14,49,14,40,18,17,9,61,28,11,14,14,26,14,34,58,33,31,12,14,12,24,14,16,11,30,25,19,14,36,9,12,44,20,16,23,35,12,14,40,43,11,18,39,9,45,9,22,24,19,15,47,40,33,32,19,11,26,18,15,24,34,23,38,38,48,37,37,39,37,13,24)
```

```
>MuestraModelosConProM=c(119,28,25,24,15,10,16,45,19,12,15,3
8,14,26,13,12,23,13,18,13,9,25,6,10,6,17,7,9,6,15,9,20,17,10,21,29,2
7,25,12,9,5,27,12,13,24,9,6,12,14,9,10,7,13,10,28,7,16,6,29,6,15,9,6,
6,20,21,21,9,9,6,22,33,20,6,9,6,14,7,6,9,17,14,5,26,9,6,29,5,12,5,12,
14,21,12,12,23,13,9,5,9,6,24,6,10,13,6,34,6,15,31,40,5,20,19,6,38,30
,19,14,5,10,6,6,39,6,33,18,11,9,54,20,5,6,6,26,6,27,58,27,31,12,6,12,
26,14,9,5,24,25,19,6,24,9,12,35,14,6,7,35,12,6,21,43,5,11,18,9,45,9,
22,15,19,8,27,35,33,16,15,5,29,21,8,13,21,12,38,25,38,20,27,30,37,1
3,24)
```

```
> var(MuestraModelosSinProM)
```

```
[1] 200.1193
```

```
> var(MuestraModelosConProM)
```

```
[1] 166.5936
```

**Tabla 11: Tabla de la 1º etapa con la evaluación de varianzas de las muestras del experimento.**

⇒ Se observa que las varianzas son diferentes.

- ☉ Para conocer el tipo de distribución de cada muestra, usamos el test de normalidad Shapiro-Wilk porque se encuentra ya disponible en R. Este test verifica si una muestra sigue una ley Normal o no sobre todo para muestras que no son de gran tamaño.

En R, el test de Shapiro Wilk indica que si P-value < 0.01, entonces la muestra sigue la ley Normal. Para nuestras muestras el resultado será:

```
> shapiro.test(MuestraModelosSinProM)
```

```
Shapiro-Wilk normality test
```

```
data: MuestraModelosSinProM  
W = 0.7575, p-value < 2.2e-16  
  
> shapiro.test(MuestraModelosConProM)  
Shapiro-Wilk normality test  
data: MuestraModelosConProM  
W = 0.7633, p-value < 2.2e-16 ≈ 0.00000000.....22 < 0.01 ⇒
```

**Tabla 12: Tabla de la 1º etapa con la evaluación de la normalidad de las muestras mediante el “test de normalidad Shapiro Wilk”**

Como para ambas muestras se cumple que:

**p-value < 2.2e-16 ≈ 0.00000000.....22 < 0.01 ⇒ p-value < 0.01**  
  
**⇒ Las muestras siguen la ley Normal.**

Con las características encontradas, el test más recomendable para la validación de la hipótesis es el test “T-Student”, el cuál explicaremos a continuación:

**Test T Student**

Como los promedios de las muestras siguen la ley Normal, entonces:

**El Test indica que:**

**Con i: 1 y a: ProM**

- El promedio de elementos con uso de la herramienta ProM tiende a seguir la ley Normal:

$$X_{a0i} \sim N(\mu_{a0i}, \sigma_{a0i}^2/n_{a0i})$$

- El promedio de elementos sin usar la herramienta ProM tiende a seguir la ley Normal:

$$X_{0i} \sim N(\mu_{0i}, \sigma_{0i}^2/n_{0i})$$

- En consecuencia, la diferencia de promedios tenderá también a seguir la ley Normal:

$$X_{a0i} - X_{0i} \sim N(\mu_{a0i} - \mu_{0i}, (\sigma_{a0i}^2 + \sigma_{0i}^2)/n_{0i})$$

- Adecuando las hipótesis a las que será aplicado el test, éstas serán:

$H_{0\text{étape1}}$  : La herramienta ProM crea el modelo correctamente:

$$\mu_{a0i} - \mu_{0i} = 0$$

$H_{1\text{étape1}}$  : La herramienta ProM crea un modelo diferente:

$$\mu_{a0i} - \mu_{0i} \neq 0$$

- La variable estadística ( $T_{\text{calculado}}$ ) que emplearemos en el test se calcula de la siguiente manera:

$$T_{\text{calculado}} = \frac{\bar{d} - \mu_0}{\frac{\sigma_d}{\sqrt{N}}}$$

Donde:

—

$\bar{d}$  : Es el valor promedio o media aritmética de las diferencias en los tiempos sin uso de ProM y con uso de ProM.

$\sigma_d$ : Es la desviación estándar de las diferencias entre los tiempos sin uso de ProM y con uso de ProM.

$N$  : Es el tamaño de la muestra.

El promedio aritmético se obtiene de la siguiente manera:

—

$$\bar{d} = \frac{\sum d}{N}$$

Y la desviación estándar de las diferencias:

$$\sigma_d = \sqrt{\frac{\sum (d - \bar{d})^2}{N - 1}}$$

$T_{\text{calculado}}$  tendrá una distribución T Student con  $k=N-1$  grados de libertad.

#### Ecuación 4: Ecuación para la obtención del $T_{\text{calculado}}$ en el test T-Student.

Finalmente, consideraremos que la hipótesis  $H_0$  será aceptada cuando esta encuentre su valor dentro de la zona de aceptación (la zona dentro de un tipo de distribución bilateral o con dos colas, con una probabilidad alfa  $0,05/2=0,025$  y  $k$  grados de libertad) cuando  $T_{\text{calculado}}$  se encuentre entre los valores críticos  $[-T_{\text{escrito en la tabla}}; + T_{\text{escrito en la tabla}}]$ .

Ahora pasaremos a evaluar la hipótesis de nuestro experimento. Se tiene que:

Hipótesis H <sub>0</sub> etapa 1							
Posició n Modelo s	Sin ProM	Con ProM	d	$\bar{d}-d$	$(\bar{d}-d)^2$	$\sigma d = \sqrt{\frac{\sum(d - \bar{d})^2}{N-1}}$	$T_{calc} = \frac{\bar{d}}{\sigma d} \sqrt{N}$
1	135	119	16	11.1	123.21	$\sigma d = \sqrt{\frac{5607.54}{193}}$ $\sigma d = 5.39$	$T_{calc} = \frac{4.9}{5.39} \sqrt{194}$ $T_{calc} = 12.66$
2	46	28	18	13.1	171.61		
3	40	25	15	10.1	102.01		
4	24	24	0	-4.9	24.01		
5	24	15	9	4.1	16.81		
6	10	10	0	-4.9	24.01		
7	16	16	0	-4.9	24.01		
8	45	45	0	-4.9	24.01		
9	19	19	0	-4.9	24.01		
10	12	12	0	-4.9	24.01		
11	15	15	0	-4.9	24.01		
12	41	38	3	-1.9	3.61		
13	21	14	7	2.1	4.41		
14	26	26	0	-4.9	24.01		
15	16	13	3	-1.9	3.61		
16	12	12	0	-4.9	24.01		
17	28	23	5	0.1	0.01		
18	13	13	0	-4.9	24.01		
19	25	18	7	2.1	4.41		
20	14	13	1	-3.9	15.21		
21	9	9	0	-4.9	24.01		
22	43	25	18	13.1	171.61		
23	15	6	9	4.1	16.81		
24	10	10	0	-4.9	24.01		
25	15	6	9	4.1	16.81		
26	17	17	0	-4.9	24.01		
27	7	7	0	-4.9	24.01		
28	9	9	0	-4.9	24.01		
29	14	6	8	3.1	9.61		
30	15	15	0	-4.9	24.01		
31	17	9	8	3.1	9.61		

32	20	20	0	-4.9	24.01		
33	27	17	10	5.1	26.01		
34	10	10	0	-4.9	24.01		
35	42	21	21	16.1	259.21		
36	29	29	0	-4.9	24.01		
37	40	27	13	8.1	65.61		
38	25	25	0	-4.9	24.01		
39	12	12	0	-4.9	24.01		
40	17	9	8	3.1	9.61		
41	11	5	6	1.1	1.21		
42	27	27	0	-4.9	24.01		
43	12	12	0	-4.9	24.01		
44	16	13	3	-1.9	3.61		
45	33	24	9	4.1	16.81		
46	17	9	8	3.1	9.61		
47	13	6	7	2.1	4.41		
48	12	12	0	-4.9	24.01		
49	14	14	0	-4.9	24.01		
50	9	9	0	-4.9	24.01		
51	10	10	0	-4.9	24.01		
52	7	7	0	-4.9	24.01		
53	13	13	0	-4.9	24.01		
54	10	10	0	-4.9	24.01		
55	28	28	0	-4.9	24.01		
56	18	7	11	6.1	37.21		
57	32	16	16	11.1	123.21		
58	15	6	9	4.1	16.81		
59	29	29	0	-4.9	24.01		
60	14	6	8	3.1	9.61		
61	19	15	4	-0.9	0.81		
62	9	9	0	-4.9	24.01		
63	15	6	9	4.1	16.81		
64	15	6	9	4.1	16.81		
65	26	20	6	1.1	1.21		
66	21	21	0	-4.9	24.01		

67	24	21	3	-1.9	3.61		
68	9	9	0	-4.9	24.01		
69	9	9	0	-4.9	24.01		
70	15	6	9	4.1	16.81		
71	22	22	0	-4.9	24.01		
72	44	33	11	6.1	37.21		
73	31	20	11	6.1	37.21		
74	14	6	8	3.1	9.61		
75	9	9	0	-4.9	24.01		
76	14	6	8	3.1	9.61		
77	22	14	8	3.1	9.61		
78	18	7	11	6.1	37.21		
79	14	6	8	3.1	9.61		
80	9	9	0	-4.9	24.01		
81	16	17	1	-3.9	15.21		
82	14	14	0	-4.9	24.01		
83	11	5	6	1.1	1.21		
84	31	26	5	0.1	0.01		
85	9	9	0	-4.9	24.01		
86	16	6	10	5.1	26.01		
87	32	29	3	-1.9	3.61		
88	11	5	6	1.1	1.21		
89	12	12	0	-4.9	24.01		
90	11	5	6	1.1	1.21		
91	12	12	0	-4.9	24.01		
92	21	14	7	2.1	4.41		
93	21	21	0	-4.9	24.01		
94	12	12	0	-4.9	24.01		
95	12	12	0	-4.9	24.01		
96	32	23	9	4.1	16.81		
97	16	13	3	-1.9	3.61		
98	9	9	0	-4.9	24.01		
99	11	5	6	1.1	1.21		
100	9	9	0	-4.9	24.01		
101	14	6	8	3.1	9.61		

102	32	24	8	3.1	9.61
103	15	6	9	4.1	16.81
104	10	10	0	-4.9	24.01
105	17	13	4	-0.9	0.81
106	14	6	8	3.1	9.61
107	33	34	-1	-5.9	34.81
108	14	6	8	3.1	9.61
109	15	15	0	-4.9	24.01
110	31	31	0	-4.9	24.01
111	52	40	12	7.1	50.41
112	11	5	6	1.1	1.21
113	36	20	16	11.1	123.21
114	19	19	0	-4.9	24.01
115	14	6	8	3.1	9.61
116	38	38	0	-4.9	24.01
117	30	30	0	-4.9	24.01
118	27	19	8	3.1	9.61
119	19	14	5	0.1	0.01
120	11	5	6	1.1	1.21
121	10	10	0	-4.9	24.01
122	14	6	8	3.1	9.61
123	14	6	8	3.1	9.61
124	49	39	10	5.1	26.01
125	14	6	8	3.1	9.61
126	40	33	13	8.1	65.61
127	18	18	0	-4.9	24.01
128	17	11	6	1.1	1.21
129	9	9	0	-4.9	24.01
130	61	54	7	2.1	4.41
131	28	20	8	3.1	9.61
132	11	5	6	1.1	1.21
133	14	6	8	3.1	9.61
134	14	6	8	3.1	9.61
135	26	26	0	-4.9	24.01
136	14	6	8	3.1	9.61

137	34	27	7	2.1	4.41		
138	58	58	0	-4.9	24.01		
139	33	27	6	1.1	1.21		
140	31	31	0	-4.9	24.01		
141	12	12	0	-4.9	24.01		
142	14	6	8	3.1	9.61		
143	12	12	0	-4.9	24.01		
144	24	26	-2	-6.9	47.61		
145	14	14	0	-4.9	24.01		
146	16	9	7	2.1	4.41		
147	11	5	6	1.1	1.21		
148	30	24	6	1.1	1.21		
149	25	25	0	-4.9	24.01		
150	19	19	0	-4.9	24.01		
151	14	6	8	3.1	9.61		
152	36	24	12	7.1	50.41		
153	9	9	0	-4.9	24.01		
154	12	12	0	-4.9	24.01		
155	44	35	9	4.1	16.81		
156	20	14	6	1.1	1.21		
157	16	6	10	5.1	26.01		
158	23	7	16	11.1	123.21		
159	35	35	0	-4.9	24.01		
160	12	12	0	-4.9	24.01		
161	14	6	8	3.1	9.61		
162	40	21	19	14.1	198.81		
163	43	43	0	-4.9	24.01		
164	11	5	6	1.1	1.21		
165	18	11	7	2.1	4.41		
166	39	18	21	16.1	259.21		
167	9	9	0	-4.9	24.01		
168	45	45	0	-4.9	24.01		
169	9	9	0	-4.9	24.01		
170	22	22	0	-4.9	24.01		
171	24	15	9	4.1	16.81		

172	19	19	0	-4.9	24.01
173	15	8	7	2.1	4.41
174	47	27	20	15.1	228.01
175	40	35	5	0.1	0.01
176	33	33	0	-4.9	24.01
177	32	16	16	11.1	123.21
178	19	15	4	-0.9	0.81
179	11	5	6	1.1	1.21
180	26	29	-3	-7.9	62.41
181	18	21	-3	-7.9	62.41
182	15	8	7	2.1	4.41
183	24	13	11	6.1	37.21
184	34	21	13	8.1	65.61
185	23	12	11	6.1	37.21
186	38	38	0	-4.9	24.01
187	38	25	13	8.1	65.61
188	48	38	10	5.1	26.01
189	37	20	17	12.1	146.41
190	37	27	10	5.1	26.01
191	39	39	0	-4.9	24.01
192	37	37	0	-4.9	24.01
193	13	13	0	-4.9	24.01
194	24	24	0	-4.9	24.01
			$\sum d=948$		$\sum (d-d)^2$
			-		=5607.54
			$\Rightarrow d=$		
			948/194		
			-		
			$\Rightarrow d=4,9$		

Tabla 13: Tabla de la 1º etapa con los resultados de la evaluación de la hipótesis usando el “test T-Student”.

Para esta etapa tenemos que

$$T_{\text{escrito en la tabla}} = T_{(0,025, \text{gl}=193)} = 1,96$$

Entonces para aceptar la  $H_{0\text{étape1}}$ , el valor de  $T_{\text{calculado}}$  debe encontrarse incluido en el intervalo  $[-T_{\text{escrito en la tabla}}, T_{\text{escrito en la tabla}}] : [-1,96; 1.96]$

Se observa que  $T_{\text{calculado}} = 12.66 \notin [-1,96; 1.96]$ , por lo que la hipótesis " $H_{0\text{étape1}}$ " no es aceptada y si rechazada. Ahora, podemos afirmar que la herramienta ProM no crea el modelo correctamente y que toda variación en el resultado del modelo no se debe a la suerte.

#### IV.5 ANÁLISIS GLOBAL DE LOS RESULTADOS

A partir del análisis de la herramienta de minería de procesos ProM durante la **primera etapa** podemos deducir que: "Aún no es del todo recomendable apoyarnos completa y únicamente en el empleo de dicha herramienta para la obtención de modelos de procesos de negocio ya que aún no se encuentra del todo madura".

Actualmente ProM aún tiende a generar modelos de procesos de negocio con "Patrones de Flujo de Trabajo" faltantes o mal generados. Mediante la investigación se pudo descubrir que ProM tiene las siguientes fallas:

- ⊗ Existe un fuerte problema con los "Patrones de Ciclo Arbitrario" y las actividades que los generan:
  - ⊗ Las actividades generadoras de dichos patrones se crean como aisladas.
  - ⊗ Las demás actividades existentes entre dichas actividades generadoras también se ven afectadas creándose también como aisladas.
  - ⊗ Los "Patrones de Secuencia, Elección y Unión Simple" que están en contacto con las actividades generadoras de los ciclos arbitrarios, desaparecen o se generan incorrectamente. Lo mencionado trae como consecuencia la creación de modelos de procesos divididos.

- ⊕ Si el “Patrón de Ciclo Arbitrario” se creó mal posicionado, éste puede generar “Uniones Simples o Elecciones Exclusivas” erróneas.
  
- ⊕ Algunas veces se crean: “Patrones de Elección, Unión Simple, División Paralela o de Sincronización redundantes”.
  
- ⊕ Puede suceder que falte que se cree alguna de las flechas que conforman un: “Patrón de Unión Simple”.
  
- ⊕ El algoritmo Alpha de ProM empleado para la minería de procesos tiene limitaciones ya que no permite el diseño de los siguientes patrones:
  - ⊕ Ramificaciones Avanzadas o Patrones de Sincronización.
  - ⊕ Relacionados a Múltiples Instancias.
  - ⊕ Basados en Estados.
  - ⊕ De Cancelación.
  
- ⊕ En cuanto a las funcionalidades de la herramienta ProM:
  - ⊕ No crea nodos finales diferentes a los nodos intermedios.
  - ⊕ Escoge donde colocar la actividad según su criterio, al encontrarse en algunas ocasiones entre diversas opciones. Pero en realidad no siempre considera la mejor opción ya que algunas veces sitúa la actividad de tal manera que crea nuevos caminos inexistentes dentro del registro de eventos trabajado.
  - ⊕ Tiene problemas en el manejo de registro de eventos de gran tamaño. La importación de los mismos se llevará a cabo durante varios minutos y la generación del modelo inclusive puede que hasta en días. Siendo el resultado final muchas veces un modelo de procesos incomprensible o incompleto.
  - ⊕ Si bien proporciona filtros para reducir el tamaño de los registros de eventos a analizar, éstos son de tipo genérico.

- ⊕ Realiza la asunción de que los registros de eventos se encuentran ordenados de manera fiel a la realidad, lo que algunas veces puede ser erróneo.
- ⊕ ProM no se pregunta si el modelo de procesos de negocios que ha creado es lógico.

Del análisis de la **segunda etapa** podemos deducir que para la mayoría de entrevistados:

**Universitarios entre [25; 35] años:**

*“Ambos modelos son medianamente complejos, medianamente comprensibles en relación al número de elementos. El modelo A es el más presentable y más comprensible en relación al contenido”.*

**Magísteres entre [25; 35] años:**

*“El modelo A es menos complejo, ambos modelos son igualmente de comprensibles en relación al número de elementos. El modelo A es mucho más presentable y más comprensible en relación al contenido”.*

**Para la mayoría de magísteres entrevistados mayores que 35 años:**

*“El modelo A es menos complejo. El modelo A es más comprensible en relación al número de elementos, en relación al contenido y también es el más presentable”*

## **CONCLUSIONES Y RECOMENDACIONES**

### **CONCLUSIONES**

Siendo la problemática de investigación del presente trabajo: “La evaluación de la precisión del descubrimiento de modelos de procesos aplicando técnicas de minería de procesos”, siendo la utilidad de la minería de procesos el ser la disciplina que se encarga de analizar los registros de eventos para obtener información sobre la perspectiva de procesos de negocio, la perspectiva organizacional y la perspectiva del tiempo, se formuló la hipótesis nula de no variación y la hipótesis alternativa de variación del modelo obtenido por minería de procesos.

Con la finalidad de responder a nuestra pregunta de investigación y de validar nuestras hipótesis, hemos usado una metodología de investigación que implica además de la formulación de las hipótesis, la identificación de etapas, objetivos de investigación, la identificación de criterios de selección de entradas y salidas, la población, la muestra, las variables, los instrumentos de medición, los escenarios, el diseño del estudio. Así como la recolección de datos y el análisis de resultados. De dicho análisis, se ha logrado lo siguiente:

- 1. Se logró validar la hipótesis con la investigación realizada, teniendo como resultado su rechazo. Lo que demuestra que ProM es una herramienta que aún no se encuentra muy madura, por lo que no es muy recomendable apoyarse únicamente en dicha herramienta para la obtención de modelos de procesos de negocio.**
- 2. Se logró encontrar el porqué del rechazo de la hipótesis nula, es decir los principales puntos débiles de ProM mediante la aplicación de su algoritmo Alpha:**

*“El principal punto débil de ProM está en el manejo de los Patrones de Ciclo Arbitrario que se generan incorrectamente y tienen influencia en la errónea generación de los demás patrones”.*

Siendo otros problemas no menos importantes:

- La redundancia y la falta de creación de algunos patrones durante la generación del modelo.
  - La generación de caminos inexistentes, debido a que la herramienta no analiza donde le conviene más colocar la actividad cuando se halla entre múltiples opciones. ProM no se pregunta si el modelo creado es lógico.
3. Se logró identificar los patrones que son soportados por ProM y su algoritmo Alpha y aquellos que no. ProM y su algoritmo Alpha solo soportan los patrones de Control Básico y los patrones Estructurales.
4. En torno a la evaluación de ProM también se logró identificar otras debilidades en torno al diseño y funcionalidad de la herramienta:
- Gran dificultad en el manejo directo de registros de eventos de gran tamaño. Los que son comunes en los sistemas actuales.
  - Filtros genéricos que no permiten realizar muchas funciones.
  - Asunción en el trabajo sobre que los registros se encuentran ordenados. Lo que no siempre es cierto.
  - No se maneja diferentes niveles de abstracción. ProM crea los modelos de procesos al mismo nivel de detalle que los registros de eventos recibidos.

5. También luego de aplicar la minería de procesos, podemos concluir que es posible indagar acerca de los procesos cognitivos de los actores relacionados al proceso a partir de los patrones de los procesos. Para ello es necesaria la observación y el registro de los diferentes eventos realizados por dichos actores y que representan los patrones de procesos. Los resultados mostrarán la manera en que los actores han aprendido sobre el proceso y como utilizan ese aprendizaje para llevarlo a cabo según sus necesidades.

Además al tener la precisión relación directa con la claridad del modelo creado se puso en práctica una encuesta comparativa entre un modelo manual y aquel obtenido por ProM con el empleo del algoritmo Alpha.

6. De aquello, si bien se sabe que la claridad depende de la edad y el grado académico, se logró conocer que para la mayoría de los entrevistados existe un consenso acerca de que los modelos generados por ProM:

- Son poco presentables ya que se presentan como gráficos con solo dos colores blanco y negro, caracteres tipográficos pequeños y estándar, con actividades cada vez más pequeñas y con una disposición del modelo en la página cada vez más concentrada a medida que el registro de eventos se hace más grande.
- Son poco comprensibles en relación al contenido ya que ProM cuando necesita de espacio porque está trabajando con un registro de eventos muy grande tiende a cortar los nombres dentro de las actividades según como necesite espacio, haciendo algunas veces el modelo totalmente incomprensible.

7. En cuanto a la comprensión en relación al largo de palabras existe una discrepancia ya que las personas entre [25; 35] años indican que el modelo generado por ProM es comprensible mientras que los mayores de

35 señalan lo contrario al preferir modelos más detallados que incluyan más palabras.

8. En cuanto a la complejidad, la opinión de los entrevistados es que el modelo generado por la herramienta ProM puede ser complejo en el vocabulario y en la estructura de las oraciones. Esto se debe a que ProM solo genera el modelo de procesos a partir del registro de eventos que recibe sin presentar opciones de idioma, sinónimos para las palabras a tratar. Tampoco da aviso acerca de si el nombre del evento a tratar contiene paréntesis ó otros símbolos (como el símbolo + para el ejemplo), lo que conlleva a que algunas veces se generen más de 3 sílabas en las oraciones haciéndolas difíciles de comprender.

En relación a la teoría existente, se puede afirmar que no existe un cambio en las teorías existentes, al contrario los resultados de este trabajo reafirman y agregan nuevas y más detalladas informaciones a aquello que la teoría expone sobre las herramientas de minería de procesos, así como sobre la definición de la claridad.

La respuesta obtenida como fruto de la investigación es muy importante para los futuros trabajos ya que muestra en detalle, mediante un caso práctico, la realidad actual de las herramientas de minería de procesos. Brindando aún más, una definición para la claridad de los modelos de procesos generados por la herramienta ProM y su algoritmo Alpha.

Pero, debemos tomar en cuenta que nuestra investigación ha sido limitada a una sola herramienta representativa en el mercado, por lo que, lo que quedaría por hacer es tomar más herramientas y evaluarlas para poder tener una visión más precisa acerca del funcionamiento global de las herramientas de minería de procesos. De igual manera, para las investigaciones futuras, se espera la evolución y la incorporación de las recomendaciones del presente trabajo en las herramientas de minería de procesos, así como la consideración de las definiciones de la claridad halladas para la mejora de los modelos de procesos obtenidos por minería de procesos.

Finalmente la pregunta de investigación que se puede explorar más tarde y que se deriva del presente trabajo será:

“¿Cuáles son las características de un proyecto informático que harán que la aplicación de la minería de procesos sea indispensable?”

## **RECOMENDACIONES**

Las recomendaciones son las siguientes:

- 1. Perfeccionar la funcionalidad de generación de los Patrones de Ciclo Arbitrario y su influencia en los demás patrones, a fin de que estos se generen adecuadamente y se dejen de crear actividades aisladas y modelos mal creados o divididos.**
- 2. Reparar el problema de redundancia en los Patrones de Elección, Unión Simple, División Paralela o de Sincronización. Y lo necesario para que no falte la creación de ninguna flecha que conforme algún Patrón de Unión Simple.**
- 3. Incorporar la lógica necesaria que permitirá a la herramienta tener la capacidad para poder optar por colocar la actividad en la ubicación que sea más conveniente, consiguiendo así evitar la generación de caminos inexistentes en el registro de eventos.**
- 4. Agregar los artefactos y el diseño necesario para que en el modelo sea posible soportar todos los patrones existentes incluyendo: Los patrones de Ramificaciones Avanzadas o patrones de Sincronización, los patrones Relacionados a Múltiples Instancias, los patrones Basados en Estados, los patrones de Cancelación. Además de los patrones Básicos y los patrones Estructurales que ya son soportados.**
- 5. Implementar nuevas funcionalidades para manejo de filtros más personalizados, para el manejo de registros de gran tamaño pero proporcionando siempre opciones al usuario para obtener sus modelos de**

procesos de negocio a diferentes niveles de abstracción. Evitando así, modelos incompletos o poco comprensibles.

**6.** Añadir las opciones necesarias para obtener una presentación variada en nuestros modelos de procesos de negocio generados. Opciones tales como:

- Diferentes Colores, tipo y tamaño de carácter.
- Que se pueda estirar el cuadro de cada actividad según se desee.
- Que el nodo que indica el fin del flujo de actividades tenga un diseño que lo diferencie de los nodos intermedios.
- Que se cuente con opciones acerca de si se desea un gráfico con las actividades coloreadas según la especialidad a la que pertenecen, pues ello ayudaría a una mejor comprensión del proceso.

**7.** Subsanan la generación de actividades con nombres cortados por falta de espacio, en lugar de ello hacer que la actividad se amolde al tamaño del texto.

**8.** Enmendar la creación de modelos altamente concentrados, distribuyendo mejor las actividades a lo largo de la hoja.

## INDICE DE FIGURAS

Figura 1: Arq. Tradicional vs Incremental según Vladimir Rubin, Christian W. Günther, Wil M.P. van der Aalst, Ekkart Kindler, Boudewijn F. van Dongen, and Wilhelm Schafer [8].	21
Figura 2: Desde un registro de eventos hacia un modelo de procesos. <i>Elaboración propia.</i>	27
Figura 3: Patrones de Control. <i>Elaboración Propia</i>	31
Figura 4: Etapas de la investigación. <i>Elaboración propia.</i>	39
Figura 5: Escenario de investigación. <i>Elaboración propia.</i>	40
Figura 6: De registros a micro – registros de eventos. <i>Elaboración Propia</i>	42
Figura 7: Micro – registros y la coherencia de eventos. <i>Elaboración Propia.</i>	43
Figura 8: Obtención de la muestra a analizar. <i>Elaboración propia.</i>	48
Figura 9: Modelo creado manualmente, empleado para la evaluación de la claridad. <i>Elaboración propia.</i>	53
Figura 10: Modelo creado con ProM, empleado para la evaluación de la claridad. <i>Elaboración propia.</i>	54
Figura 11: Consecuencia de los Ciclos Arbitrarios continuos. <i>Elaboración propia.</i>	60
Figura 12: El efecto de los Ciclos Arbitrarios continuos en actividades intermedias. <i>Elaboración propia.</i>	61
Figura 13: El efecto de los Ciclos Arbitrarios mal generados en el proceso global. <i>Elaboración propia.</i>	61
Figura 14: El efecto de los Ciclos Arbitrarios en cualquier otro patrón de flujo de trabajo. <i>Elaboración propia.</i>	62
Figura 15: Ciclos Arbitrarios y Patrones de Secuencia. <i>Elaboración propia.</i>	63
Figura 16: Los Ciclos Arbitrarios y su efecto en los Patrones de Elección. <i>Elaboración propia.</i>	63
Figura 17: Patrones de Elección redundantes. <i>Elaboración propia.</i>	64
Figura 18: Ciclos Arbitrarios y Patrones de Unión Simple. <i>Elaboración propia.</i>	64
Figura 19: Patrones de Unión Simple incompletos. <i>Elaboración propia.</i>	65
Figura 20: Patrones de Unión redundantes. <i>Elaboración propia.</i>	66
Figura 21: Uniones Simples irreales. <i>Elaboración propia.</i>	66
Figura 22: Patrones de Sincronización redundantes. <i>Elaboración propia.</i>	67
Figura 23: División Paralela y Sincronización incoherentes. <i>Elaboración propia.</i>	68
Figura 24: Patrones soportados por el algoritmo Alpha - ProM usado en minería de procesos. <i>Elaborado por Wil Van Der Aalst [19].</i>	76
Figura 25: Artefactos de diseño: Eventos de Cancelación en Bizagi. <i>Elaborado por Bizagi [26].</i>	78

Figura 26: Artefactos de diseño: Compuertas basadas en eventos en Bizagi. <i>Elaborado por Bizagi [26].</i> .....	78
Figura 27: Sugerencias sobre el diseño del nodo Final. <i>Elaboración propia.</i> 79	
Figura 28: Flujos de actividades adicionales en el modelo generado por ProM. <i>Elaboración propia.</i> .....	80
Figura 29: Dificultades de ProM con registros de eventos de inmenso tamaño. <i>Elaboración propia.</i> .....	81
Figura 30: Modelos incompletos provenientes de registros de eventos de gran tamaño. <i>Elaboración propia.</i> .....	82
Figura 31: Composición de una sección de eventos del micro registro. <i>Publicado por Universidad de Tecnología de Eindhoven [28].</i> .....	113
Figura 32: Especialidades del Hospital bajo análisis. <i>Proporcionado por [29]</i> .....	114
Figura 33: Ejemplo de un Modelo de Procesos Inicial. <i>Elaboración Propia</i> 115	
Figura 34: Ejemplo de una Parte del log de eventos del Ejemplo. <i>Elaboración Propia</i> .....	119
Figura 35: Ejemplo del modelo de procesos generado por Minería de Procesos con PROM. <i>Elaboración Propia</i> .....	120
Figura 36: Ejemplo de la forma en que debió ser generado el Modelo de Procesos. <i>Elaboración Propia</i> .....	121
Figura 37: Modelo de Encuesta empleada en la 2da etapa de la investigación. <i>Elaboración Propia</i> .....	130

## INDICE DE ECUACIONES:

Ecuación 1: Fórmula del tamaño de una muestra con población finita. ....	44
Ecuación 2: Fórmula para calcular el número de micro registro. ....	46
Ecuación 3: Fórmula del número de elementos dentro de un modelo de procesos.....	49
Ecuación 4: Ecuación para la obtención del $T_{\text{calculado}}$ en el test T-Student. ...	87

## INDICE DE TABLAS

Tabla 1: Tabla de la 1º etapa del diseño del experimento.....	51
Tabla 2: Tabla de la 1º etapa con la recopilación de resultados del experimento. ....	59
Tabla 3: Tabla de la 2º etapa con la recopilación de resultados de la evaluación de la complejidad, edad 25-35 años. ....	69
Tabla 4: Tabla de la 2º etapa con la recopilación de resultados de la evaluación de la complejidad, edad >35 años.....	69
Tabla 5: Tabla de la 2º etapa con la recopilación de resultados de la evaluación de la comprensión del número de elementos del modelo, edad 25-35 años. ....	71
Tabla 6: Tabla de la 2º etapa con la recopilación de resultados de la evaluación de la comprensión del número de elementos del modelo, edad >35 años. ....	71
Tabla 7: Tabla de la 2º etapa con la recopilación de resultados de la evaluación de la presentación del modelo, edad 25-35 años.....	72
Tabla 8: Tabla de la 2º etapa con la recopilación de resultados de la evaluación de la presentación del modelo, edad >35 años.....	72
Tabla 9: Tabla de la 2º etapa con la recopilación de resultados de la evaluación del contenido del modelo, edad 25-35 años.....	73
Tabla 10: Tabla de la 2º etapa con la recopilación de resultados de la evaluación del contenido del modelo, edad >35 años.....	74
Tabla 11: Tabla de la 1º etapa con la evaluación de varianzas de las muestras del experimento. ....	84
Tabla 12: Tabla de la 1º etapa con la evaluación de la normalidad de las muestras mediante el “test de normalidad Shapiro Wilk” .....	85
Tabla 13: Tabla de la 1º etapa con los resultados de la evaluación de la hipótesis usando el “test T-Student”.....	93
Tabla 14: Tabla de la 1ra Comparación MPInicial vs. MPFinal .....	123
Tabla 15: Tabla de la 2da Comparación MPInicial vs. MPFinal .....	125
Tabla 16: Tabla de la 3ra Comparación MPInicial vs. MPFinal .....	126
Tabla 17: Tabla de la 4ta Comparación MPInicial vs. MPFinal .....	128
Tabla 18: Tabla de la 5ta Comparación MPInicial vs. MPFinal .....	129

## **BIBLIOGRAFIA**

### **1. FUENTES BIBLIOGRAFICAS**

- [1]** Nathaniel Palmer, A survey of business process initiatives – BPTrends, pp: 6.
  
- [2]** Aldave Cecilia, Problemática de la retro ingeniería “Evaluación del Proceso de Ida/Retorno usando dos herramientas evaluación de la claridad de los modelos”, Memoria profesional de Máster, Universidad Paris 1 Pantheón Sorbonne, Paris, Francia, 2012.
  
- [3]** Elliot Chikofsky and James Cross II, “Reverse Engineering and Design Recovery: A Taxonomy”, IEEE Software, vol. 7, no. 1, January 1990, pp. 13-17.
  
- [4]** Ati Jain, Swapnil Soner, Anand Gadwal, “Reverse Engineering: Journey from Code to Design”, Electronics Computer Technology (ICECT), 2011 3rd International Conference on, April 2011, pp. 102-106, 8-10.
  
- [5]** Jawaria Sadiq, Tabinda Waheed, “Reverse Engineering & Design Recovery: An Evaluation of Design Recovery Techniques”, Computer Networks and Information Technology (ICCNIT), 2011 International Conference on, 2011, pp. 325 – 332.
  
- [6]** Chih-Wei Lu, William C. Chu, Chih-Hung Chang, Yeh-Ching Chung, Xiaodong Liu and Hongji Yang, “Reverse engineering”, Department Of Computer Science, Leicester England, 01/2002.
  
- [8]** Wil Van der Aalst, “Process mining Knowledge Discovery Process Centric”, ACM SIGKDD Explorations Newsletter, Volume 13 Issue 2, December 2011, pp. 45-49.

- [9] Wil M. P. van der Aalst, "Inter-Organizational Process Mining Discovering Processes Within and Between Organizations", In proceeding of: The Practice of Enterprise Modeling - 4th IFIP WG 8.1 Working Conference, PoEM 2011 Oslo, Norway, November 2-3, 2011 Proceedings.
- [10] Vladimir Rubin, Christian W. Günther, Wil M.P. van der Aalst, Ekkart Kindler, Boudewijn F. van Dongen, and Wilhelm Schafer, "Process Mining Framework for Software Processes", Eindhoven University of Technology, Eindhoven, The Netherlands and University of Paderborn, Paderborn, Germany.
- [11] IEEE Task Force on Process Mining, "Manifiesto sobre Minería de Procesos", In BPM Workshops, volume 99 of Lecture Notes in Business Information Processing. Springer-Verlag, Berlin, 2011, pp. 169-194.
- [12] W.M.P. van der Aalst, H.A. Reijers, A.J.M.M. Weijters, B.F. van Dongen, A.K. Alves de Medeiros, M. Song, b, H.M.W. Verbeek, "Business Process mining", Information Systems, Volume 32, Issue 5, July 2007, pp. 713–732.
- [13] D. Riehle and H. Zullighoven, "Understanding and Using Patterns in software development", Theory and Practice of Object Systems, 1996, 2(1): pp. 3-13.
- [14] E. Gamma, R. Helm, R.Johnson, and J. Vlissides, "Design Patterns: Elements of reusable Object-Oriented Software", Addison – Wesley, Reading, Massachusetts, 1995.
- [15] Wil van Der Aalst, Arthur H.M. Hofstede, Bartek Kiepuszewski, and Alistair P. Barros, "Workflow Patterns", Distributed and Parallel Databases 14 (1): pp. 5—51.

- [17] A. Hirsenschall, "Patterns discusión in the Context of Co-flow, CONDIS Workflow Management System", Technical Report, University of Linz, Austria, 2001.
- [18] H. Lavana, "A Universally Configurable Architecture for Taskflow-Oriented Design of a Distributed Collaborative Computing Environment" PhD Tesis, Departamento de Ciencias de la computación, Universidad del estado de Carolina del Norte, Raleigh, NC, USA, 2000.
- [19] Wil van der Aalst, "Process Mining: Discovery, Conformance and Enhancement of Business Processes", 2011, Chapter 04, 05.
- [20] Jan Claes y Geert Poels, "Process Mining and the ProM Framework: An Exploratory Survey", BPM 2012 International Workshops, Tallinn, Estonia, September 3, 2012, pp 9.
- [21] Daniela Oelke, David Spretke, Andreas Stoffel, Daniel A.Keim, "Visual Readability Analysis: How to Make Your Writings Easier to Read", IEEE, , 2010, pp123-130.
- [22] Akhilesh Bajaj, Stephen Rockwell, "The Effect Of Fragmentation On the Readability of Schemas: An Empirical Study With Multiple Model Methods", The University of Tulsa.
- [23] Yvon Laframboise, "La lisibilité: Qu'est – ce que la lisibilité? Quels éléments rendent un texte lisible et un autre pas?", Les Publications Québec français Numéro 32, décembre 1978, pp.27-29.  
(<http://id.erudit.org/iderudit/56560ac>).
- [24] Scharager, J y Armijo, I., "Metodología de la Investigación para las Ciencias Sociales" [CD-ROM]: Versión 1.0 Santiago: Escuela de Psicología, SECICO Pontificia Universidad Católica de Chile. Programa computacional, 2001.

- [25] “Estadística aplicada a las Ciencias Sociales”, Universidad Pontificia Comillas, Facultad de Humanidades, Madrid, Diciembre, 2012.
- [26] Bizagi, “Patrones de modelado de Procesos”, pp: 34, 39-40.  
<http://www.bizagi.com/docs/Workflow Patterns using BizAgi Process Modeler Esp.pdf>
- [30] Koniewski R., Dzielinski A., Amborski K., “Use of Petri Nets and Business Processes Management Notation in Modeling and Simulation of Multimodal Logistics Chains”, Proceedings 20th European Conference on Modeling and Simulation, Institute of Control and Industrial Electronics, Warsaw University of Technology, Koszykowa 75, 00-662 Warszawa, Poland, 2006, Pp 2
- [31] Raedts, I., Petkovic, M., Usenko, Y., van der Werf, J., Groote, J., Somers, L.: “Transformation of BPMN models for behaviour analysis” In: Proceedings of the 5th International Workshop on Modelling, Simulation, Verification and Validation of Enterprise Information Systems, INSTICC Press (2007) 126-137, pp 3

## **2. FUENTES SITIOS WEB DE INTERNET**

*Página de la Fuerza de Trabajo en Minería de Procesos de la IEEE, consultada el 08/11/2012:*

[7] <http://www.win.tue.nl/ieeetfpm/doku.php?id=shared:org:start>

*Página del grupo de la Iniciativa de Patrones de Flujo de Trabajo, consultada el 10/12/2012:*

[16] <http://www.workflowpatterns.com/>

*Página de la herramienta R para la computación estadística, consultada el 17/07/2012:*

[27] <http://www.r-project.org/>

*Página donde se encuentra formalmente el log de eventos de un hospital de naturaleza anónima de los Países Bajos publicado por la Universidad de Tecnología de Eindhoven, consultada el 10/12/2012:*

**[28]** <http://data.3tu.nl/repository/uuid:d9769f3d-0ab0-4fb8-803b-0d1120ffc54>

*Página del Congreso Internacional en Inteligencia de Procesos de Negocio-BPI'11, Clermont-Ferrand, France, Agosto 28th - Septiembre 2nd 2011, Se encuentra el detalle de las especialidades del Hospital bajo análisis, consultada el 10/12/2012:*

**[29]** <http://www.win.tue.nl/bpi/2011/challenge>

## ANEXOS

### 1. COMPOSICIÓN DE LA SECCIÓN DE EVENTOS DEL MICRO REGISTRO

El micro- registro de eventos contiene como sección más importante “la sección de eventos” ya que esta permitirá formar los modelos de procesos a partir de la minería de procesos.

```
<event>
<string key="org:group" value="Pathology"/>
<int key="Number of executions" value="1"/>
<int key="Specialism code" value="88"/>
<string key="concept:name" value="cytologisch onderzoek -
vagina -"/>
<string key="Section" value="Section 4"/>
<string key="Producer code" value="LVPT"/>
<int key="Activity code" value="355427"/>
<date key="time:timestamp" value="2005-01-18T18:00:00-
05:00"/>
<string key="lifecycle:transition" value="complete"/>
</event>
```

**Figura 31: Composición de una sección de eventos del micro registro. Publicado por Universidad de Tecnología de Eindhoven [28]**

Esta sección tiene las siguientes partes:

1. **<string key="org:group" value="Pathology"/>**, permite definir el grupo ó especialidad en el hospital donde se llevo a cabo el evento. El hospital cuenta con las siguientes especialidades [19]:

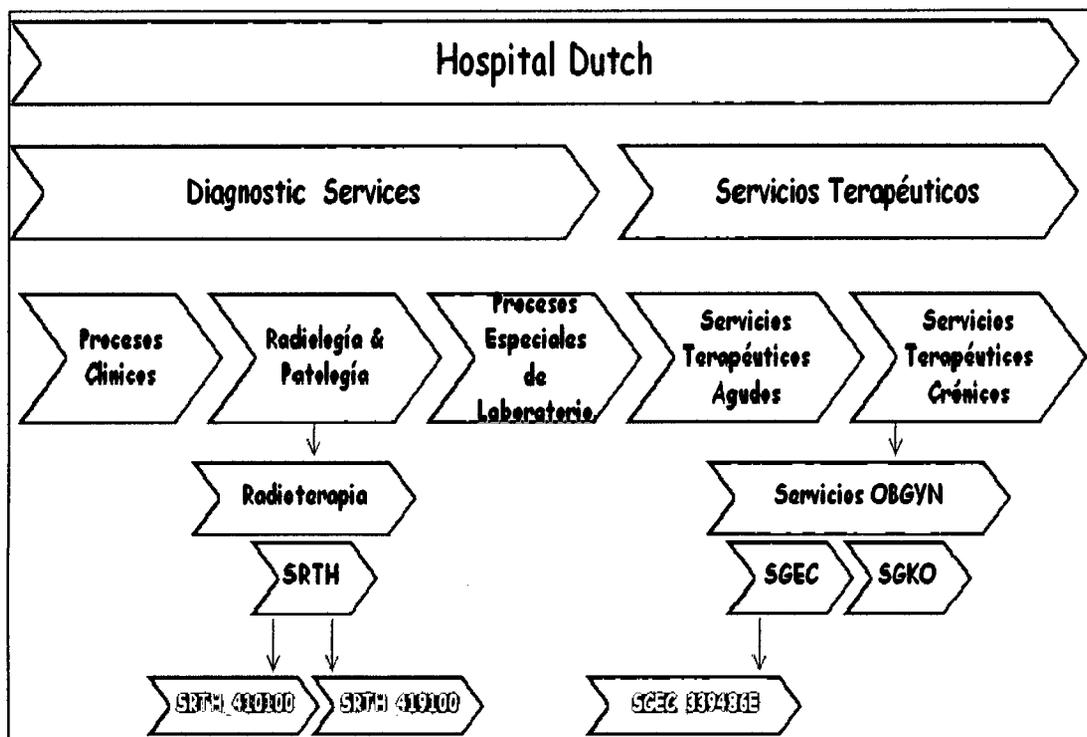


Figura 32: Especialidades del Hospital bajo análisis. *Proporcionado por [29]*

2. `<int key="Number of executions" value="1"/>`, nos indica el número de veces que se registro el evento. Para todos los eventos dentro de cada micro- registro este valor siempre es uno.
3. `<int key="Specialism code" value="88"/>`, define el código de la especialidad del hospital relacionada al evento.
4. `<string key="concept:name" value="cytologisch onderzoek - vagina -"/>`, este es el punto más importante en la sección de eventos ya que indica el nombre de la actividad (evento) que se llevo a cabo.
5. `<string key="Section" value="Section 4"/>`, es la sección a la que pertenece la actividad.
6. `<string key="Producer code" value="LVPT"/>`, indica el código del que produjo el evento.

7. `<int key="Activity code" value="355427"/>`, es el código de la actividad o evento realizado.
8. `<date key="time:timestamp" value="2005-01-18T18:00:00-05:00"/>`, indica la hora a la que se completó la actividad.
9. `<string key="lifecycle:transition" value="complete"/>`, informa que la actividad ya se completo.

## 2. EJEMPLO DEL ANÁLISIS EMPLEADO EN LA MINERÍA DE PROCESOS

A continuación se muestra un ejemplo del análisis empleado para la comparación entre cada instancia de procesos fuente y la instancia de procesos que ha sido generada a partir del registro de eventos.

Inicialmente tenemos como entrada el modelo de procesos inicial instanciado que ha sido creado de manera manual a partir del micro - registro de eventos:

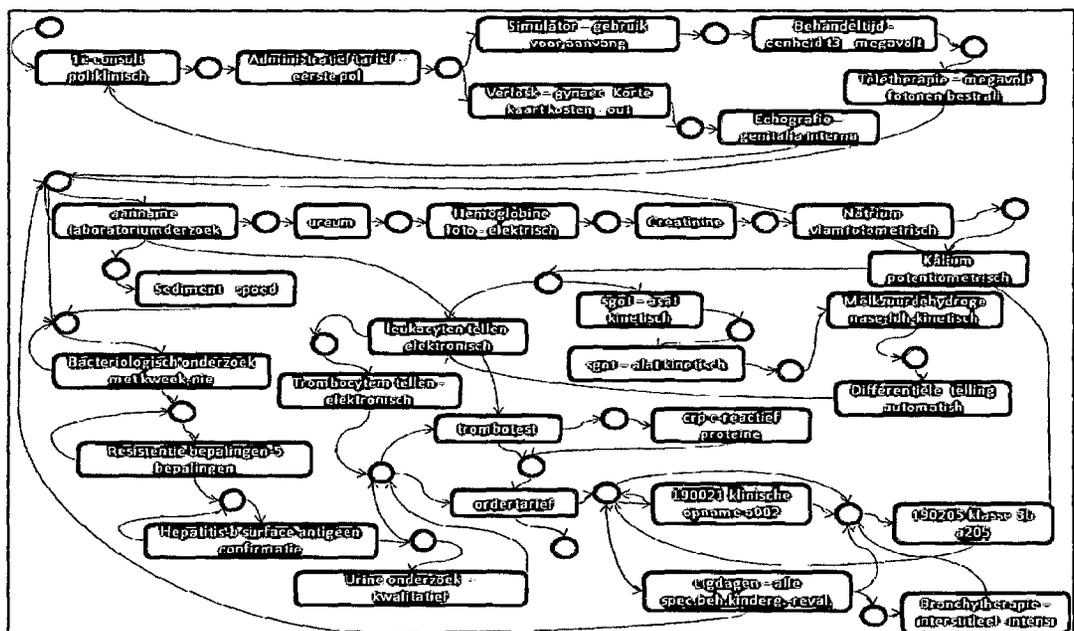


Figura 33: Ejemplo de un Modelo de Procesos Inicial. *Elaboración Propia*

También tenemos como entrada y a continuación, una parte del micro registro de eventos, creado bajo las condiciones establecidas y que se empleo para crear el modelo anterior. En este se puede observar las actividades (en negrita) que permitirán la creación del modelo de procesos instanciado luego de la aplicación del algoritmo Alpha para la minería de procesos:

*Observación: Solo se muestra una parte porque la composición del micro-registro completo es muy grande.*

```
.  
.
<event>
  <string key="org:group" value="Radiotherapy"/>
  <int key="Number of executions" value="1"/>
  <int key="Specialism code" value="61"/>
  <string key="concept:name" value="1e consult poliklinisch"/>
  <string key="Section" value="Section 5"/>
  <string key="Producer code" value="SRTH"/>
  <int key="Activity code" value="410100"/>
  <date key="time:timestamp" value="2005-01-02T18:00:00-05:00"/>
  <string key="lifecycle:transition" value="complete"/>
</event>
<event>
  <string key="org:group" value="Radiotherapy"/>
  <int key="Number of executions" value="1"/>
  <int key="Specialism code" value="61"/>
  <string key="concept:name" value="administratief tarief - eerste
pol"/>
  <string key="Section" value="Section 5"/>
  <string key="Producer code" value="SRTH"/>
  <int key="Activity code" value="419100"/>
  <date key="time:timestamp" value="2005-01-02T18:00:00-05:00"/>
```

```

    <string key="lifecycle:transition" value="complete"/>
</event>
<event>
    <string key="org:group" value="Nursing ward"/>
    <int key="Number of executions" value="1"/>
    <int key="Specialism code" value="7"/>
    <string key="concept:name" value="verlosk.-gynaec. korte kaart
kosten-out"/>
    <string key="Section" value="Section 2"/>
    <string key="Producer code" value="SGEH"/>
    <int key="Activity code" value="10107"/>
    <date key="time:timestamp" value="2005-01-04T18:00:00-05:00"/>
    <string key="lifecycle:transition" value="complete"/>
</event>
<event>
    <string key="org:group" value="Obstetrics & Gynaecology clinic"/>
    <int key="Number of executions" value="1"/>
    <int key="Specialism code" value="7"/>
    <string key="concept:name" value="echografie - genitalia interna"/>
    <string key="Section" value="Section 2"/>
    <string key="Producer code" value="SGEC"/>
    <string key="Activity code" value="339486E"/>
    <date key="time:timestamp" value="2005-01-04T18:00:00-05:00"/>
    <string key="lifecycle:transition" value="complete"/>
</event>
<event>
    <string key="org:group" value="Nursing ward"/>
    <int key="Number of executions" value="1"/>
    <int key="Specialism code" value="7"/>
    <string key="concept:name" value="1e consult poliklinisch"/>
    <string key="Section" value="Section 2"/>
    <string key="Producer code" value="SGEH"/>

```

```

<int key="Activity code" value="410100"/>
<date key="time:timestamp" value="2005-01-04T18:00:00-05:00"/>
<string key="lifecycle:transition" value="complete"/>
</event>
<event>
  <string key="org:group" value="Nursing ward"/>
  <int key="Number of executions" value="1"/>
  <int key="Specialism code" value="7"/>
  <string key="concept:name" value="administratief tarief- eerste pol"/>
  <string key="Section" value="Section 2"/>
  <string key="Producer code" value="SGEH"/>
  <int key="Activity code" value="419100"/>
  <date key="time:timestamp" value="2005-01-04T18:00:00-05:00"/>
  <string key="lifecycle:transition" value="complete"/>
</event>
<event>
  <string key="org:group" value="Radiotherapy"/>
  <int key="Number of executions" value="1"/>
  <int key="Specialism code" value="61"/>
  <string key="concept:name" value="simulator - gebruik voor aanvang
egavol"/>
  <string key="Section" value="Section 5"/>
  <string key="Producer code" value="RATH"/>
  <int key="Activity code" value="390520"/>
  <date key="time:timestamp" value="2005-01-23T18:00:00-05:00"/>
  <string key="lifecycle:transition" value="complete"/>
</event>
<event>
  <string key="org:group" value="Radiotherapy"/>
  <int key="Number of executions" value="1"/>
  <int key="Specialism code" value="61"/>
  <string key="concept:name" value="behandeltijd - eenheid t3 - megavolt"/>
  <string key="Section" value="Section 5"/>

```

```

<string key="Producer code" value="RATH"/>
<int key="Activity code" value="390003"/>
<date key="time:timestamp" value="2005-01-30T18:00:00-05:00"/>
<string key="lifecycle:transition" value="complete"/>
</event>
<event>
<string key="org:group" value="Radiotherapy"/>
<int key="Number of executions" value="1"/>
<int key="Specialism code" value="61"/>
<string key="concept:name" value="teletherapie - megavolt fotonen
bestrali"/>
<string key="Section" value="Section 5"/>
<string key="Producer code" value="RATH"/>
<int key="Activity code" value="390550"/>
<date key="time:timestamp" value="2005-01-30T18:00:00-05:00"/>
<string key="lifecycle:transition" value="complete"/>
</event>
<event>
<string key="org:group" value="General Lab Clinical Chemistry"/>
<int key="Number of executions" value="1"/>
<int key="Specialism code" value="86"/>
<string key="concept:name" value="aaname laboratoriumonderzoek"/>
<string key="Section" value="Section 4"/>
<string key="Producer code" value="CRLA"/>
<int key="Activity code" value="370000"/>
<date key="time:timestamp" value="2005-02-14T18:00:00-05:00"/>
<string key="lifecycle:transition" value="complete"/>
</event>

```

**Figura 34: Ejemplo de una Parte del log de eventos del Ejemplo.  
Elaboración Propia**

Enseguida, al aplicar la minería de procesos al micro – registro de eventos con la herramienta ProM se obtendrá la siguiente instancia de modelo de procesos:

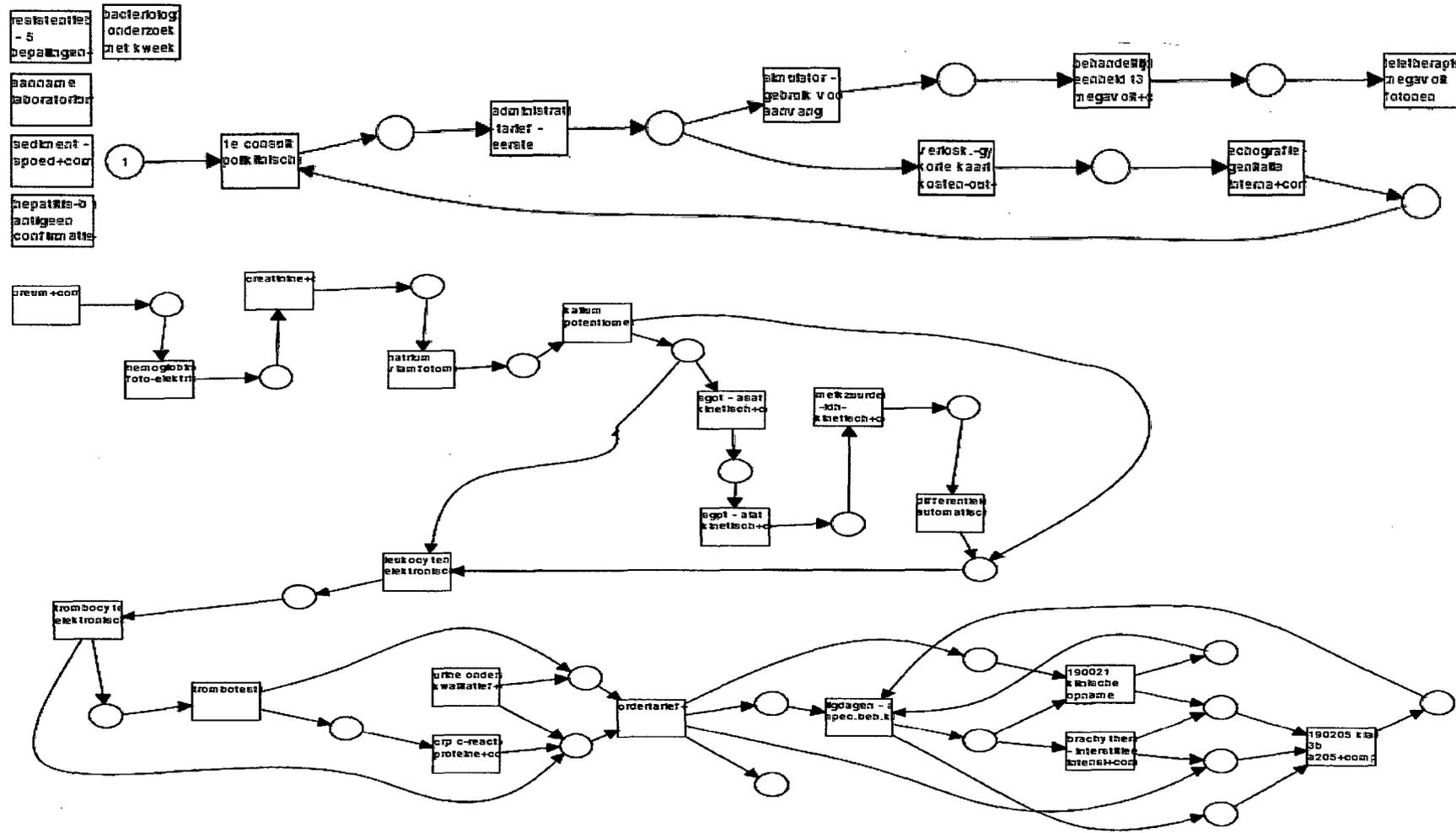


Figura 35: Ejemplo del modelo de procesos generado por Minería de Procesos con PROM. *Elaboración Propia.*

De la comparación entre ambos modelos se observa que existen diferencias en los patrones de flujo de trabajo, precisamente estas diferencias se observan en la parte azul del gráfico:

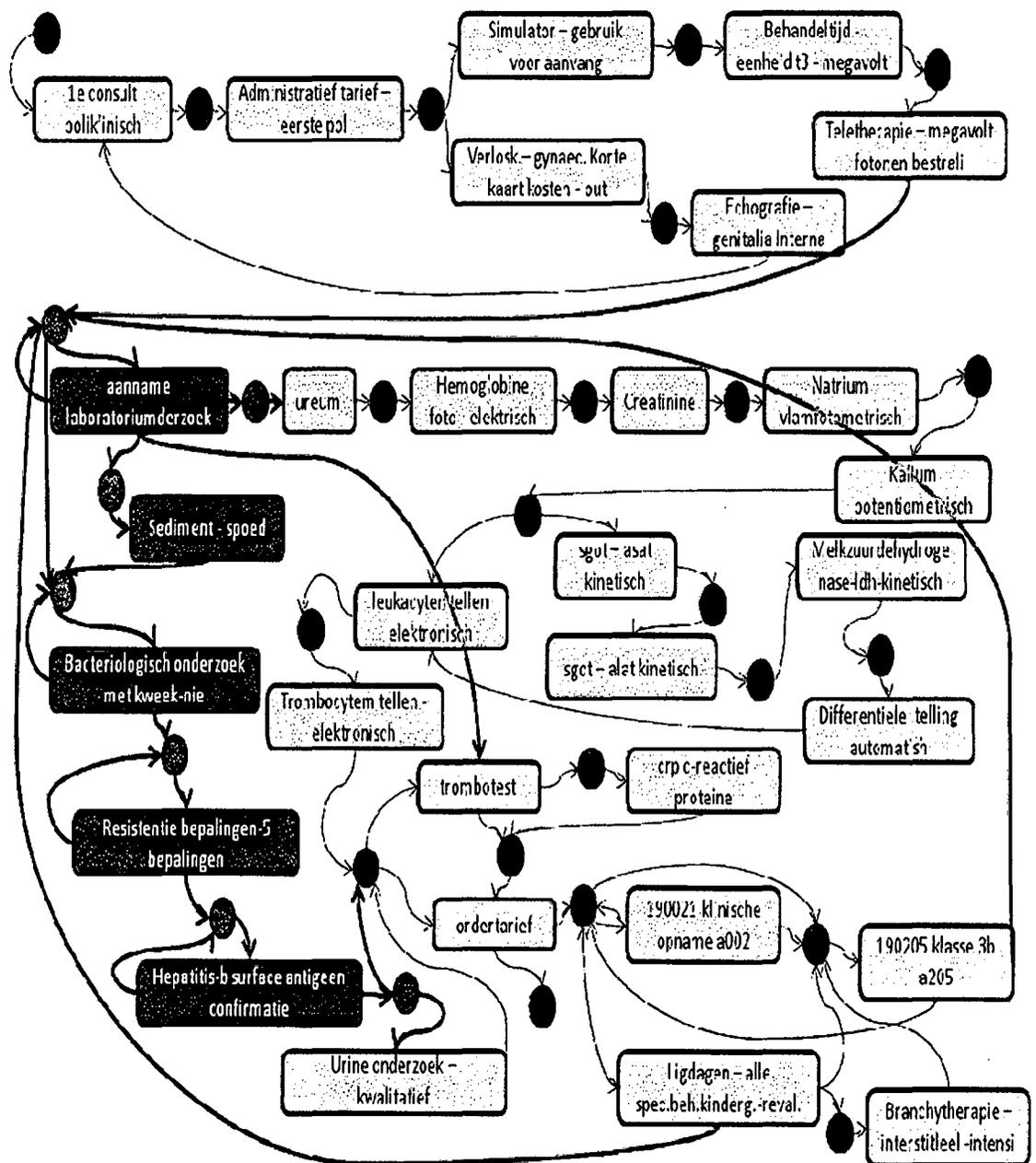


Figura 36: Ejemplo de la forma en que debió ser generado el Modelo de Procesos. *Elaboración Propia*

De esta comparación entre el modelo de procesos inicial vs modelo de procesos final es que deduciremos los valores numéricos para los cuadros resumen siguientes.

### 3. RESULTADO DETALLADO DE LAS EVALUACIONES

Solo se muestra el detalle del 1er modelo de proceso al 5to de los 194 que se evaluaron, para mayor detalle en el archivo adjunto a la tesis:

Entradas: Modelo de procesos de la instancia N°1 (MPI1), micro-registro de eventos N°1.		
Salidas: Modelo de procesos N°1 obtenido por minería de procesos con ProM. (MPF1)		
Comparación MPI1 vs. MPF1		
	Modelo original	Modelo obtenido por minería de procesos con ProM
<b>Número de actividades</b>	<b>31</b>	<b>31</b>
<b><i>Patrones de Control Básico</i></b>		
<b>Secuencia</b>	77	67
<b>División Paralela</b>	3	3
<b>Sincronización</b>	1	1
<b>Elección exclusiva</b>	8	9
<b>Unión Simple</b>	10	7
<b><i>Ramificaciones Avanzadas Patrones de Sincronización</i></b>		
<b>Elección Múltiple</b>	--	--
<b>Estructura de Unión Sincronizada</b>	--	--
<b>Unión múltiple</b>	--	
<b>Discriminador Estructurado</b>	--	
<b><i>Patrones Estructurales</i></b>		
<b>Ciclo Arbitrario</b>	4	0
<b>Terminación implícita</b>	1	1
<b><i>Patrones Relacionados a Múltiples Instancias</i></b>		
<b>Múltiples Instancias sin</b>	--	--

<b>Sincronización</b>		
<b>Múltiples instancias con conocimiento a priori del número de ejecuciones en la etapa de diseño</b>	--	--
<b>Múltiples instancias con conocimiento en el tiempo de ejecución</b>	--	--
<b>Múltiples instancias sin conocimiento en el tiempo de ejecución</b>	--	--
<b><i>Patrones Basados en Estados</i></b>		
<b>Elección Diferida</b>	--	--
<b>Ejecución Intercalada</b>	--	--
<b>Milesone</b>	--	--
<b><i>Patrones de Cancelación</i></b>		
<b>Cancelación de tareas</b>	--	--
<b>Cancelación del Caso</b>	--	--
<b><i>Número de Elementos en el Modelo de Procesos</i></b>	<b>135</b>	<b>119</b>

**Tabla 14: Tabla de la 1ra Comparación MPI inicial vs. MP Final**

<b>Entradas: Modelo de procesos de la instancia N°2 (MPI2), micro-registro de eventos N°2.</b>		
<b>Salidas: Modelo de procesos N°2 obtenido por minería de procesos con ProM. (MPF2)</b>		
<b>Comparación MPI2 vs. MPF2</b>		
	<b>Modelo original</b>	<b>Modelo obtenido por minería de procesos con ProM</b>
<b>Número de actividades</b>	<b>11</b>	<b>11</b>

<b>Patrones de Control Básico</b>		
<b>Secuencia</b>	28	14
<b>División Paralela</b>	--	--
<b>Sincronización</b>	--	--
<b>Elección exclusiva</b>	3	1
<b>Unión Simple</b>	2	1
<b>Ramificaciones Avanzadas Patrones de Sincronización</b>		
<b>Elección Múltiple</b>	--	--
<b>Estructura de Unión Sincronizada</b>	--	--
<b>Unión múltiple</b>	--	--
<b>Discriminador Estructurado</b>	--	--
<b>Patrones Estructurales</b>		
<b>Ciclo Arbitrario</b>	1	0
<b>Terminación implícita</b>	1	1
<b>Patrones Relacionados a Múltiples Instancias</b>		
<b>Múltiples Instancias sin Sincronización</b>	--	--
<b>Múltiples instancias con conocimiento a priori del número de ejecuciones en la etapa de diseño</b>	--	--
<b>Múltiples instancias con conocimiento en el tiempo de ejecución</b>	--	--
<b>Múltiples instancias sin conocimiento en el tiempo de ejecución</b>	--	--
<b>Patrones basados en Estados</b>		
<b>Elección Diferida</b>	--	--

<b>Ejecución Intercalada</b>	--	--
<b>Milesone</b>	--	--
<b><i>Patrones de Cancelación</i></b>		
<b>Cancelación de tareas</b>	--	--
<b>Cancelación del Caso</b>	--	--
<b><i>Número de Elementos en el Modelo de Procesos</i></b>	46	28

**Tabla 15: Tabla de la 2da Comparación MPIncial vs. MPFinal**

<b>Entradas: Modelo de procesos de la instancia N°3 (MPI3), micro-registro de eventos N°3.</b>		
<b>Salidas: Modelo de procesos N°3 obtenido por minería de procesos con ProM. (MPF3)</b>		
<b>Comparación MPI3 vs. MPF3</b>		
	<b>Modelo original</b>	<b>Modelo obtenido por minería de procesos con ProM</b>
<b>Número de actividades</b>	<b>9</b>	<b>9</b>
<b><i>Patrones de Control Básico</i></b>		
<b>Secuencia</b>	<b>25</b>	<b>12</b>
<b>División Paralela</b>	--	--
<b>Sincronización</b>	--	--
<b>Elección exclusiva</b>	<b>3</b>	<b>1</b>
<b>Unión Simple</b>	<b>1</b>	<b>1</b>
<b><i>Ramificaciones Avanzadas Patrones de Sincronización</i></b>		
<b>Elección Múltiple</b>	--	--
<b>Estructura de Unión Sincronizada</b>	--	--
<b>Unión múltiple</b>	--	--
<b>Discriminador Estructurado</b>	--	--

<b>Patrones Estructurales</b>		
<b>Ciclo Arbitrario</b>	1	1
<b>Terminación implícita</b>	1	1
<b>Patrones Relacionados a Múltiples Instancias</b>		
<b>Múltiples Instancias sin Sincronización</b>	--	--
<b>Múltiples instancias con conocimiento a priori del número de ejecuciones en la etapa de diseño</b>	--	--
<b>Múltiples instancias con conocimiento en el tiempo de ejecución</b>	--	--
<b>Múltiples instancias sin conocimiento en el tiempo de ejecución</b>	--	--
<b>Patrones Basados en Estados</b>		
<b>Elección Diferida</b>	--	--
<b>Ejecución Intercalada</b>	--	--
<b>Milesone</b>	--	--
<b>Patrones de Cancelación</b>		
<b>Cancelación de tareas</b>	--	--
<b>Cancelación del Caso</b>	--	--
<b>Número de Elementos en el Modelo de Procesos</b>	40	25

**Tabla 16: Tabla de la 3ra Comparación MPInicial vs. MPFinal**

**Entradas: Modelo de procesos de la instancia N°4 (MPI4), micro-registro de eventos N°4.**

**Salidas: Modelo de procesos N°4 obtenido por minería de procesos con ProM. (MPF4)**

<b>Comparación MPI4 vs. MPF4</b>		
	<b>Modelo original</b>	<b>Modelo obtenido por minería de procesos con ProM</b>
<b>Número de actividades</b>	<b>7</b>	<b>7</b>
<b><i>Patrones de Control Básico</i></b>		
<b>Secuencia</b>	<b>15</b>	<b>15</b>
<b>División Paralela</b>	--	--
<b>Sincronización</b>	--	--
<b>Elección exclusiva</b>	<b>0</b>	<b>0</b>
<b>Unión Simple</b>	<b>1</b>	<b>1</b>
<b><i>Ramificaciones Avanzadas Patrones de Sincronización</i></b>		
<b>Elección Múltiple</b>	--	--
<b>Estructura de Unión Sincronizada</b>	--	--
<b>Unión múltiple</b>	--	--
<b>Discriminador Estructurado</b>	--	--
<b><i>Patrones Estructurales</i></b>		
<b>Ciclo Arbitrario</b>	<b>0</b>	<b>0</b>
<b>Terminación implícita</b>	<b>1</b>	<b>1</b>
<b><i>Patrones Relacionados a Múltiples Instancias</i></b>		
<b>Múltiples Instancias sin Sincronización</b>	--	--
<b>Múltiples instancias con conocimiento a priori del número de ejecuciones en la etapa de diseño</b>	--	--
<b>Múltiples instancias con conocimiento en el tiempo de ejecución</b>	--	--

<b>Múltiples instancias sin conocimiento en el tiempo de ejecución</b>	--	--
<b><i>Patrones Basados en Estados</i></b>		
<b>Elección Diferida</b>	--	--
<b>Ejecución Intercalada</b>	--	--
<b>Milesone</b>	--	--
<b><i>Patrones de Cancelación</i></b>		
<b>Cancelación de tareas</b>	--	--
<b>Cancelación del Caso</b>	--	--
<b><i>Número de Elementos en el Modelo de Procesos</i></b>	24	24

**Tabla 17: Tabla de la 4ta Comparación MPIInicial vs. MPFinal**

<b>Entradas: Modelo de procesos de la instancia N°5(MPI5), micro-registro de eventos N°5.</b>		
<b>Salidas: Modelo de procesos N°5 obtenido por minería de procesos con ProM. (MPF5)</b>		
<b>Comparación MPI5 vs. MPF5</b>		
	<b>Modelo original</b>	<b>Modelo obtenido por minería de procesos con ProM</b>
<b>Número de actividades</b>	<b>5</b>	<b>5</b>
<b><i>Patrones de Control Básico</i></b>		
<b>Secuencia</b>	15	6
<b>División Paralela</b>	--	--
<b>Sincronización</b>	--	--
<b>Elección exclusiva</b>	1	1
<b>Unión Simple</b>	1	1
<b><i>Ramificaciones Avanzadas Patrones de Sincronización</i></b>		
<b>Elección Múltiple</b>	--	--

<b>Estructura de Unión Sincronizada</b>	--	--
<b>Unión múltiple</b>	--	
<b>Discriminador Estructurado</b>	--	
<b><i>Patrones Estructurales</i></b>		
<b>Ciclo Arbitrario</b>	1	1
<b>Terminación implícita</b>	1	1
<b><i>Patrones Relacionados a Múltiples Instancias</i></b>		
<b>Múltiples Instancias sin Sincronización</b>	--	--
<b>Múltiples instancias con conocimiento a priori del número de ejecuciones en la etapa de diseño</b>	--	--
<b>Múltiples instancias con conocimiento en el tiempo de ejecución</b>	--	--
<b>Múltiples instancias sin conocimiento en el tiempo de ejecución</b>	--	--
<b><i>Patrones Basados en Estados</i></b>		
<b>Elección Diferida</b>	--	--
<b>Ejecución Intercalada</b>	--	--
<b>Milesone</b>	--	--
<b><i>Patrones de Cancelación</i></b>		
<b>Cancelación de tareas</b>	--	--
<b>Cancelación del Caso</b>	--	--
<b><i>Número de Elementos en el Modelo de Procesos</i></b>	24	15

Tabla 18: Tabla de la 5ta Comparación MPInicial vs. MPFinal

## 4. ENCUESTA PARA LA EVALUACIÓN DE LA CLARIDAD DE LOS MODELOS

### Evaluación de Modelos

Evaluemos la claridad de los modelos A y B en función a diferentes características.  
\*Obligatoire

**Modelo A**

**Modelo B**

Su edad es: \*  
[ ] [ ] [ ] [ ]

Su nivel académico es: \*  
[ ] [ ] [ ] [ ]

**Evalua los modelos en relación a la complejidad \***  
Complejidad en relación a la complejidad en la estructura de las oraciones, al número de palabras simples y complejas (más de 3 sílabas), existencia de palabras desconocidas y al nivel de abstracción de los textos.

	Poco Complejo	Medianamente Complejo	Muy Complejo
Modelo A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Modelo B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Evalua los modelos en función al largo de las palabras \***  
Largo de las palabras, número de palabras en una oración.

	Poco Comprensible	Medianamente Comprensible	Muy Comprensible
Modelo A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Modelo B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Evalua los modelos en función a su presentación \***  
Colores, posición y tamaño de los elementos del modelo, faltas de ortografía, tipo de caracteres tipográficos, la composición o disposición del modelo en la página.

	Poco Presentable	Medianamente Presentable	Muy Presentable
Modelo A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Modelo B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Evalua los modelos en función a su exhaustividad de contenido \***  
Número de elementos en el modelo, información que falta, información ambigua.

	Poco Comprensible	Medianamente Comprensible	Muy Comprensible
Modelo A	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Modelo B	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Interpretación del Modelo \***  
Explique en sus propias palabras, lo que representa el modelo de procesos.

N'envoyez jamais de mots de passe via Formulaires Google.

Ce contenu n'est ni rédigé, ni cautionné par Google

Fourni par Google Drive

Signaler un cas d'utilisation abusive - Conditions d'utilisation - Clauses additionnelles

Figura 37: Modelo de Encuesta empleada en la 2da etapa de la investigación. *Elaboración Propia*