



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

ENFOQUE OLAP PARA EL DESCUBRIMIENTO INTERACTIVO EN MINERÍA DE PROCESOS

GUSTAVO EDUARDO PIZARRO ALVAREZ

Tesis para optar al grado de
Magister en Ciencias de la Ingeniería

Profesor Supervisor:
MARCOS SEPÚLVEDA

Santiago de Chile, Diciembre, 2013

© 2013, Gustavo Pizarro



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

ENFOQUE OLAP PARA EL DESCUBRIMIENTO INTERACTIVO EN MINERÍA DE PROCESOS

GUSTAVO EDUARDO PIZARRO ÁLVAREZ

Tesis presentada a la Comisión integrada por los profesores:

MARCOS SEPULVEDA FERNÁNDEZ

JAIME CAICEO DUQUE

DIEGO MOYA OLAVE

GONZALO CORTAZAR SANZ

Para completar las exigencias del grado de
Magister en Ciencias de la Ingeniería

Santiago de Chile, Diciembre, 2013

A mi familia y mis amigos,

AGRADECIMIENTOS

Quisiera agradecer a todos aquellos que estuvieron siempre presentes cuando apareció alguna inquietud básica o compleja. Al profesor guía Marcos Sepúlveda, a Mauricio Arriagada, estudiante de postgrado, a Sebastián Vicencio, a Nicolás Pizarro y más de alguno que probablemente se me haya olvidado mencionar.

También quiero agradecer a las 38 personas que permitieron realizar el caso de estudio y se dieron el tiempo de poder colaborar con esta investigación. En particular al grupo de minería de procesos por colaborar también en esta labor.

Finalmente quisiera agradecer a todo aquel que tuvo una palabra de aliento cuando esta tesis parecía no terminar, gracias por su apoyo probablemente sin él esta tesis aún estaría en elaboración.

ÍNDICE GENERAL

Pág.

DEDICATORIA	ii
AGRADECIMIENTOS	iii
ÍNDICE DE TABLAS	vi
ÍNDICE DE FIGURAS	vii
RESUMEN.....	viii
ABSTRACT	ix
1. INTRODUCCIÓN	1
2. ESTADO DEL ARTE	4
2.1. El analista y su tarea.....	4
2.2. Medición de calidad	7
2.3. Descubrimiento de procesos.....	11
3. PROPUESTA.....	14
4. DESCRIPCIÓN DE OLAPDISCOVERY.....	15
4.1. Definición de descubrimiento para usuarios de negocio.....	15
4.1.1. Propósito	15
4.1.2. Usuario	16
4.1.3. Brechas con el descubrimiento actual.....	16
4.1.4. Justificación OLAP	17
4.2. Propuesta inicial	18
4.2.1. Diseño propuesta inicial.....	18
4.2.2. Funcionalidades adicionales	20
4.2.3. Implementación inicial	21
4.2.4. Prueba exploratoria	22
4.2.5. Versión final OLAPDiscovery.....	23
5. CASO DE ESTUDIO	25

5.1. Descripción de los usuarios.....	25
5.2. Descripción de las tareas a realizar por los usuarios.....	25
5.3. Descripción de la evaluación.....	27
5.4. Métricas propuestas.....	28
5.5. Resultados	31
5.6. Limitaciones de la evaluación	40
6. Conclusiones.....	41
7. Trabajo futuro	43
Referencias	44
A N E X O S.....	47
Anexo A: FUNCIONALIDADES OLAP-DISCOVERY.....	48
Anexo B : FUNCIONALIDADES PROM	51
Anexo C : FUNCIONALIDADES DISCO.....	52
Anexo D: MEDIDAS DE DISPERSIÓN.....	54
Anexo E: ESTADÍSTICO DE PRUEBA.....	57
Anexo F: RESULTADOS ESTADÍSTICOS DE PRUEBA.....	58
Anexo G: PREFERENCIAS POR TIPO DE USUARIO	61

ÍNDICE DE TABLAS

	Pág.
Tabla 1 - Métricas objetivas.....	28
Tabla 2 - Métricas subjetivas	29
Tabla 3 - Medias de dispersión de tiempo unidimensional.....	32
Tabla 4 - Medidas de dispersión de tiempo multidimensional	32
Tabla 5 - Distribución de respuestas subjetivas entre ProM y OLAP-D	36
Tabla 6 - Distribución de respuestas subjetivas entre Disco y OLAP-D	38
Tabla 7 - Estadísticos de prueba para cada afirmación entre herramientas	58
Tabla 8 - Estadísticos de prueba por tipo de usuario entre OLAP-D y Disco	59
Tabla 9 - Estadísticos de prueba por uso de herramienta previa entre OLAP-D y Disco	60
Tabla 10 - Preferencias Usuarios Básicos.....	61
Tabla 11 - Preferencias Usuarios Expertos	62

ÍNDICE DE FIGURAS

	Pág.
Figura 1 - Tipos de minería de procesos, presentado por Van der Aalst et al. (2012)	2
Figura 2 - Herramientas para distintos tipos de usuario, obtenido del libro de Eckerson (2013)	6
Figura 3 - Metodología ISO9126, obtenida de International Organization for Standardization (2001)	8
Figura 4 - Atributos internos y externos de medición ISO9126-2 e ISO9126-3, obtenido de International Organization for Standardization (2001)	9
Figura 5 - Atributos de medición en ISO9126-4, obtenida de International Organization for Standardization (2001)	9
Figura 6 - Manejo de Versiones en OLAPDiscovery	20
Figura 7 - OLAPDiscovery Final	24
Figura 8 - Selección de atributos a medir de ISO9126-4	27
Figura 9 - Atributos seleccionados de ISO9126-3	28
Figura 10 – Promedio tiempo unidimensional (minutos) según intento	31
Figura 11 – Promedio tiempo multidimensional (minutos) según intento	32
Figura 12 – Logro en tareas unidimensionales	34
Figura 13 – Logro en tareas unidimensionales	34
Figura 14 – Logro en tareas unidimensionales	35
Figura 15 - Productividad efectiva multidimensional	35

RESUMEN

Un analista de procesos de negocio debe enfrentarse a la labor de analizar, monitorear y promover mejoras para distintas áreas del negocio. En los últimos años, ha aumentado su potencial de acción gracias al análisis automático de los registros de eventos de los sistemas de información utilizando técnicas de minería de procesos. Estas herramientas permiten descubrir modelos de procesos considerando distintos objetivos, distintas perspectivas (control de flujo, organizacional, temporal, entre otras) y creando distintas representaciones. Sin embargo, actualmente carecen de la posibilidad de explorar de manera conjunta e interactiva las distintas perspectivas, lo cual influye en la comprensión de lo que está sucediendo en la organización. En este artículo se propone una nueva forma de descubrimiento interactiva para analistas de procesos, orientada a explorar distintas perspectivas y considerando distintos niveles de detalle, desde lo general a lo particular. El presente documento plantea el desarrollo de este concepto considerando un enfoque de interacción OLAP, combinando un barrido multidimensional y acciones de navegación como drill down, roll up, slices and dice y pivot. Lo anterior se implementó a través de un plug-in de ProM, y se puso a prueba en un caso de estudio con usuarios reales. Las pruebas realizadas indican que hay mejoras sustantivas en relación al uso de ProM y ligeras mejoras al contraponerlo con Disco, dos de las principales herramientas disponibles actualmente para minería de procesos. La ventaja más relevante es la efectividad alcanzada en cuanto a tiempo al realizar tareas de análisis tanto unidimensionales como multidimensionales. Además, los usuarios expertos le dan una evaluación notoriamente mejor en productividad, efectividad y “es operativo” en relación a las otras dos herramientas. Esta investigación da indicios de que el usuario podría mejorar el análisis de procesos si cuenta con herramientas que le permitan un descubrimiento exploratorio, multidimensional y de navegación interactiva, permitiendo iterar dinámicamente en el análisis del proceso.

Palabras clave: Process mining, business process discovery, OLAP.

ABSTRACT

OLAP approach for interactive discovery in process mining

Business process analysts must face the task of analyzing, monitoring and promoting **improvements to different business processes**. In recent years, an **emerging discipline** called process mining has emerged as a useful tool for helping them, allowing an automatic analysis of event logs that are registered by corporate information systems. These tools allow discovering process models considering different goals, perspectives (control-flow, organizational, time, among others) and representations. However, currently they lack the ability to explore jointly and interactively the different perspectives, which influences the understanding of what is happening in the organization. This article proposes a novel approach for interactive discovery aimed at providing process analysts with a tool that allow them to explore diverse perspectives and considering different levels of detail. This paper explains the development of this concept, based on the OLAP interactive approach, combining multidimensional navigation actions, such as drill down, roll up, slices and dice, and pivot. This approach was implemented as a ProM plug-in and it was tested in a case of study with real users. Results indicate that there are substantial improvements in relation to the use of ProM and some advantages compare to the use of Disco, two of the main tools currently available for process mining. The most important advantage is the effectiveness achieved in time when performing both one-dimensional and multidimensional analysis. Additionally, expert users assess it considerably better in productivity, effectiveness and in “being operative” in comparison to the other two tools. This research shows that process analysts could improve their analysis when counting with tools that allow an exploratory, multidimensional and interactive navigation of the process.

Keywords: Process mining, business process discovery, OLAP.

1. INTRODUCCIÓN

Durante las últimas décadas, se ha adoptado paulatinamente el concepto de gestión por procesos de negocio en las organizaciones. En muchas de éstas se han creado áreas, integradas en parte por analistas que modelan, monitorean y auditan dichos procesos. Sin embargo, en los últimos años el crecimiento del universo digital ha hecho posible que estas labores tengan un alcance aún mayor gracias a que los sistemas de información crean registros de eventos (Van der Aalst et al., 2012), que almacenan cuándo los ejecutores realizan las actividades en que se descompone el proceso. Estos registros permiten que hoy sea factible efectuar un análisis complejo y robusto sobre lo que realmente está sucediendo en la organización a través de la minería de procesos. Esta disciplina utiliza la información almacenada en estos registros de eventos utilizando distintas técnicas de descubrimiento, conformidad y mejora, varias de ellas descritas por Van der Aalst et al. (2012) y ordenadas por frecuencia de uso a nivel de usuarios en un estudio de Claes, J., & Poels, G (2013). La primera categoría produce un modelo a partir del log sin utilizar ninguna información a-priori. La segunda compara el comportamiento observado en el registro de eventos con un modelo de proceso existente. La tercera tiene como objetivo modificar o ampliar el modelo teniendo en cuenta cuellos de botella, niveles de servicio, tiempos, rendimientos y frecuencias.

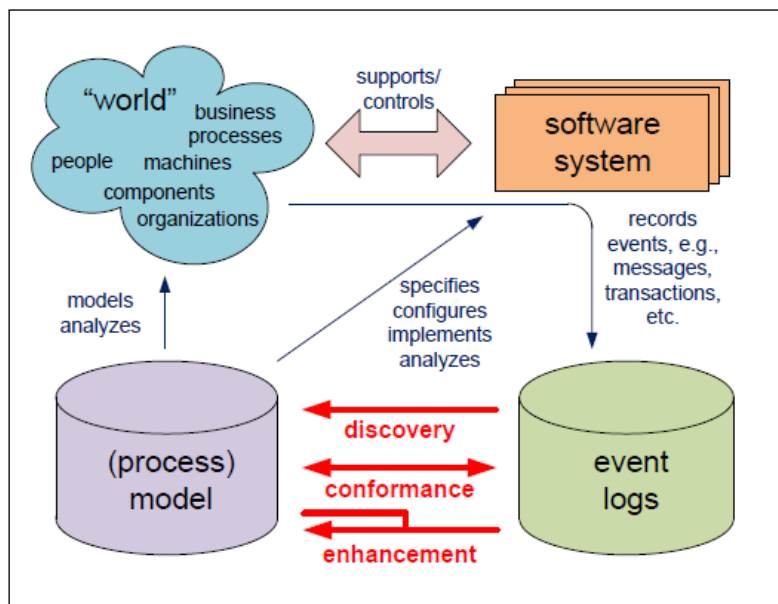


Figura 1 - Tipos de minería de procesos, presentado por Van der Aalst **et al.** (2012)

Las herramientas actuales de descubrimiento proponen un diagrama de control de flujo o de interacción de recursos, pero no permiten explorar de manera interactiva las distintas perspectivas del proceso. Según İsmail Bayraktar (2011), la comprensión del proceso completo requiere un esfuerzo que está fuera del alcance de los analistas de procesos tradicionales, ya que cada análisis de control de flujo, social, de variantes o de tiempo, debe hacerse de manera separada y luego procesarla en algún algoritmo de descubrimiento que permita analizarla en profundidad. Esto sumado a la gran cantidad de herramientas que se puede elegir, el sesgo representacional que tiene cada una de ellas y los parámetros que se deben configurar, hacen que sea difícil lograr un descubrimiento analítico eficiente, productivo y satisfactorio para un usuario de negocio utilizando las herramientas actuales.

En este artículo se propone una nueva forma de descubrimiento interactiva para analistas de procesos, orientada a explorar mediante filtrado instantáneo distintas perspectivas de un proceso, considerando distintos niveles de detalle, expresando todo lo anterior en distintos diagramas y escenarios que permitirán un análisis comparativo en las perspectivas de control de flujo, organizacional y temporal. Considerando conceptos de

usuario, tarea, usabilidad y calidad en el uso se concluyó que un concepto de descubrimiento de este tipo debería ser implementado a través de un paradigma de uso tipo OLAP. Para esto, incorpora acciones de navegación como *drill down*, *roll up* y *slice and dice*, que permiten ir desde grupos que ejecutan tareas similares o grupos de trabajo en conjunto, hasta ejecutores con el detalle de sus actividades y variantes agrupadas en clusters. Sumado a esto, se puede explorar en la dimensión temporal filtrando por duraciones mínimas y máximas. Todo lo anterior se ve reflejado en ilustraciones que muestran diagramas de control de flujo o pivotearlas para observar las relaciones de interacción entre los ejecutores. Todo lo anterior se puso a prueba en un caso de estudio a través de un plug-in de ProM, con usuarios reales.

2. ESTADO DEL ARTE

En este capítulo se exponen los distintos temas relevantes para la elaboración de la tesis y su contribución al desarrollo de la investigación.

2.1. El analista y su tarea

Dentro del área de procesos existen distintos roles a considerar. Los más importantes son: gerente de procesos, dueño del proceso, arquitecto de procesos y analista de procesos (Doebeli, Fisher, Gapp y Sanzogni, 2011). Para cada uno de ellos se requiere distintas habilidades y aptitudes, ya que tienen distintas responsabilidades.

En particular, el rol de analista de procesos surgió hace ya bastante tiempo, como se puede ver en el artículo de Hammer (1999). Según Bandara et al. (2010), este rol se ha encargado de apoyar una visión integral de los procesos de negocio y ofrecer la capacidad de transformar la organización rápidamente. También en la investigación de Mathiesen et al. (2011) se señala que debe apoyar el ciclo de vida BPM, que entre otras actividades, abarca el análisis, diseño, monitoreo y control de los procesos de negocio. Para ello se requiere demostrar conocimiento para auditar procesos, realizar análisis de deficiencias y apoyar transformaciones de cambios en ambientes de negocio.

En minería de procesos existen herramientas, pero no está catalogado qué tipo de herramienta es más adecuada a cada tipo de usuario. Es por ello que consideraremos una clasificación genérica que se propone en el ámbito de Inteligencia de Negocios, área que incluye a la minería de procesos. Existen variadas herramientas las cuales han sido documentadas por tipo de usuario, según Eckerson (2010), existen dos grandes categorías de usuario: el consumidor de información y el productor de información. Dentro de los primeros están los

analistas de negocio y los usuarios casuales. Los usuarios ocasionales incluyen ejecutivos, gerentes, empleados y usuarios externos; ven regularmente informes, pero no calculan números o realizan análisis de tendencias detallados sobre una base diaria. Los analistas de negocio en cambio, pueden además hacer análisis más sofisticados. El consumidor de información, en general, ocupa cuadros de mando, que abarca las herramientas de consulta y presentación de informes, herramientas OLAP, informes de hojas de cálculo, informes estándar y la salida de los modelos estadísticos. Dentro de los productores de información están los autores de reportes y estadísticos que utilizan las herramientas de minería de datos, además de algunos analistas de negocio. Ellos utilizan herramientas de diseño de reportes o herramientas de programación para generar reportes personalizados. Los autores de informes pueden ser los desarrolladores de TI o usuarios empoderados y capacitados por el departamento de TI para generar dichos reportes. Algunos usuarios empoderados pueden ser productores de información y consumidores de información. Un ejemplo de ese usuario mixto es el analista de negocio. Son los típicos que usan hojas de cálculo, tablas dinámicas, consultas simples a bases de datos y crean reportes a la medida. La Figura que se presenta a continuación indica qué herramienta se aconseja para los distintos tipos de usuario:

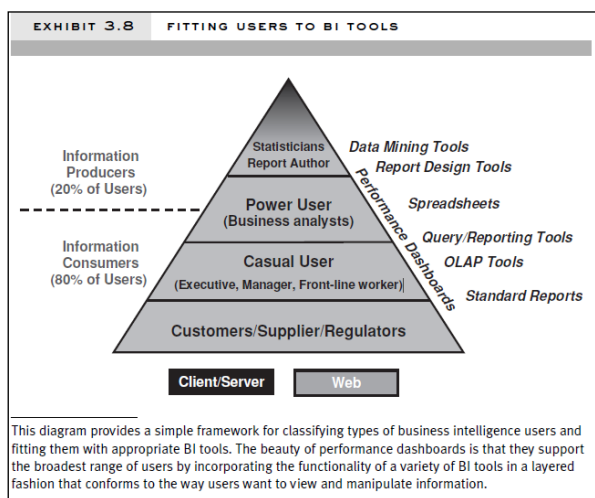


Figura 2 - Herramientas para distintos tipos de usuario, obtenido del libro de Eckerson (2013)

En la clasificación anterior se propone más de una herramienta para un analista de negocio por ser usuarios frontera entre los productores y consumidores. Sin embargo, se proponen tablas dinámicas, hojas de cálculo, OLAP, herramientas de consultas y reportes, todas ellas muy familiarizadas con el filtrado dinámico de información.

Dentro de las mencionadas, OLAP (Online Analytical Processing) es una solución cuyo objetivo es agilizar la consulta a grandes cantidades de datos. Para ello se utilizan estructuras multidimensionales. Algunos aspectos característicos a considerar de este enfoque son la vista conceptual multidimensional, la manipulación de datos intuitiva y dimensiones y niveles de agregación ilimitados (Codd, Codd y Salley, 1993).

En este artículo se considera como tipo de usuario al analista de procesos y se incorporan los aspectos característicos de OLAP en el diseño de la solución propuesta.

2.2. Medición de calidad

Dentro de las mediciones de calidad, no hay una forma 100% aceptada y establecida para la evaluación de productos de software (Behkamal, Kahani, Akbari, 2009). Sin embargo, existen distintos modelos como McCall (Cavano y McCall, 1977), Boehm (Brown, Kaspar, Lipow, MacLeod y Merrit, 1978), FURPS (Grady, 1994), Dromey (Dromey, 1995), Star (Khosravi y Guéhéneuc, 2004) e ISO9126 (2001) . El modelo ISO ha sido utilizado en varias investigaciones para medir la calidad de distintas herramientas; ejemplos se pueden encontrar en investigaciones de Chua y Dyson (2004), Stefani y Xenos (2008), Ribeiro (2013).

ISO 9126 es un proceso de evaluación de software con distintas etapas. La primera etapa introduce el modelo de calidad, especificando que en la etapa dos se debe medir calidad interna; en la etapa tres, calidad externa; y en la etapa cuatro, calidad de uso. Cada etapa puede obtener mejor resultados si se considera el feedback de la etapa anterior. En específico, la segunda etapa mide la calidad interna del software y se realiza durante la elaboración del mismo; y describe las medidas que pueden ser utilizadas para especificar o evaluar el comportamiento del software cuando es utilizado por el usuario. La tercera etapa describe las medidas que se pueden utilizar para crear los requisitos que describen las propiedades estáticas de la interfaz, y puede ser evaluado mediante inspección sin hacer funcionar el software. En esta parte evalúan principalmente los desarrolladores.

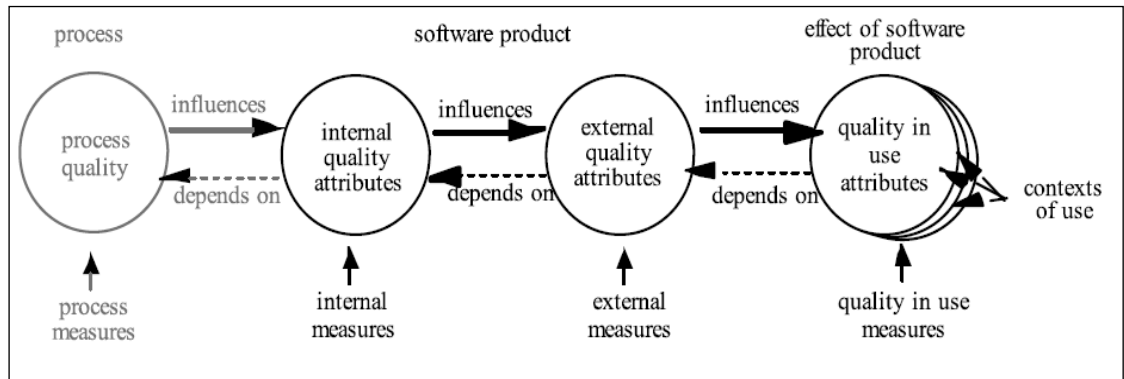


Figura 3 - Metodología ISO9126, obtenida de International Organization for Standardization (2001)

Para medir ambas etapas se utilizan seis características: funcionalidad, fiabilidad, facilidad de uso, eficiencia, mantención y portabilidad. Cada una de éstas tiene sus propias características que se pueden apreciar en la siguiente figura.

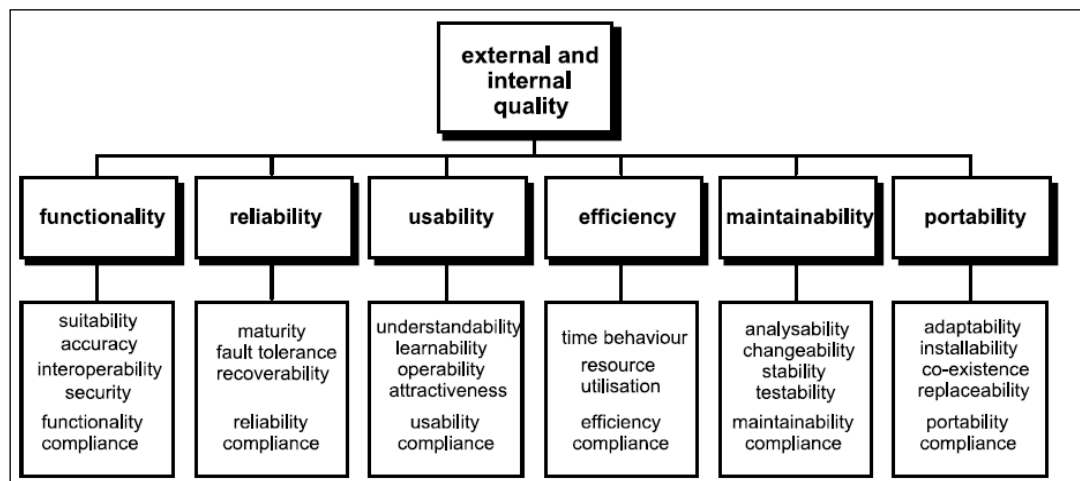


Figura 4 - Atributos internos y externos de medición ISO9126-2 e ISO9126-3, obtenido de International Organization for Standardization (2001)

La cuarta y última etapa de medición de calidad en el uso, se ve influenciada por la mejora de la anterior. Esta calidad es medida en cuanto a satisfacción, productividad, efectividad y seguridad.

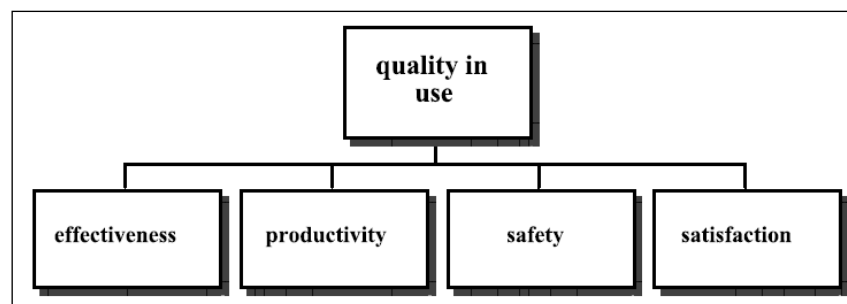


Figura 5 - Atributos de medición en ISO9126-4, obtenida de International Organization for Standardization (2001)

Dentro de minería de procesos ha sido el único enfoque que se ha utilizado y se puede encontrar en el trabajo de Ribeiro (2013). Si bien el modelo presenta una

estructura completa y robusta, existen críticas por ser un modelo que con factores demasiado amplios y generales. Es por eso que en distintas investigaciones se han ocupado versiones modificadas del mismo modelo. Algunos ejemplos se pueden encontrar en un estudio de evaluación de aplicaciones BB2 efectuado por Behkamal, Kahani, Akbari (2009) o en otros, donde se sugieren modificaciones al concepto de satisfacción, como en la evaluación de aplicaciones e-learning elaborado por Chua y Dyson (2004).

La mayoría de estas modificaciones han sido en relación al concepto de usabilidad utilizado. “La usabilidad se define como la capacidad de un software de ser comprendido, aprendido, usado y ser atractivo para el usuario, en condiciones específicas de uso” (International Organization for Standardization, 2001, p.17).

Los atributos de este concepto según Nielsen (1994) son:

1. Facilidad de aprendizaje: Indica qué tan fácil es aprender la funcionalidad básica del sistema.
2. Eficiencia: Se determina por el número de transacciones por unidad de tiempo que el usuario puede realizar usando el sistema. Lo que se busca es la máxima velocidad de realización de tareas del usuario.
3. Manejo de errores: Este atributo se refiere a la manera en que el sistema puede manejar los errores cometidos por el usuario mientras está realizando una tarea. Los errores reducen la eficiencia y satisfacción del usuario.
4. Presentación propia: El concepto de sistema se materializa al realizar el diseño de la parte visual de la interacción, es decir, la “interfaz gráfica del usuario”.

5. Satisfacción: Es el atributo más subjetivo. Muestra la impresión subjetiva que el usuario desarrolla sobre el sistema.

Los conceptos anteriores contribuyen al modelo modificado que se utilizará en esta investigación. Si bien está claro que no hay una norma 100% aceptada, ésta nos dará un marco de referencia y se podrán usar las consideraciones de pruebas realizadas anteriormente.

2.3. Descubrimiento de procesos

En los últimos años ha existido una evolución en los métodos para descubrir procesos a través de log de eventos en minería de procesos, pero un escaso avance en esta tarea considerando un usuario de negocio, según el manifiesto de minería de procesos (Van der Aalst et al., 2012).

Desde el año 2005 se han propuesto distintos algoritmos de descubrimiento en control de flujo unidimensional, tales como AlphaMiner propuesto por Van der Aalst, Weijters y Maruster (2004), HeuristicMiner propuesto por Weijters, Van der Aalst, De Medeiros (2006), Regiones propuesto por Carmona, Cortadella, Kishinevsky (2008) y otros. Estas herramientas poseen diferentes enfoques a nivel algorítmico y contribuyen en distintos escenarios, sin embargo son la base, y no necesariamente presentan un enfoque orientado a la tarea, ni vistas que contribuyan a una comprensión del proceso multidimensional por parte del usuario.

El año 2007 Günther y Van Der Aalst proponen la herramienta FuzzyMinning, primera herramienta orientada a una tarea específica más compleja que solo conseguir un flujo. En ella se presenta un enfoque de cartografía que captura dimensiones de análisis tales como: agregación, para agrupar actividades en clusters; abstracción, para eliminar datos que son poco significativos; énfasis, para

destacar lo más importante; y personalización, para destacar el propósito. Esta propuesta, a pesar de ser innovadora y enfocada en una tarea particular, no es una evolución en como analizar procesos estructurados.

Durante el año 2011 se propone la herramienta BMPNAnalysist (İsmail Bayraktar, 2011) cuya primera motivación es disminuir la complejidad de la tarea de enfrentarse a cientos de algoritmos en ProM y a un lenguaje poco cercano al negocio. Contribuye con un descubrimiento multidimensional combinando las distintas herramientas anteriores, agregando una representación de negocio y presentación de datos holístico, donde se visualiza la dimensión de tiempo, recursos y variantes. Si bien, es un avance en el descubrimiento orientado a usuarios de negocio, no está cubierta la tarea desde un punto de vista de análisis y tampoco prueba que realmente la propuesta sea mejor para el usuario de negocio.

Finalmente, durante 2013 se han presentado propuestas multidimensionales que cubren el proceso a través de distintas dimensiones e incorporan el concepto de OLAP a nivel de algoritmo. Es el caso de Multidimensional (Ribeiro, 2013), que explora qué herramientas son extensibles a nivel multidimensional, desarrollando por ejemplo, descubrimientos en HeuristicMiner, considerando flujos con roles o tiempo. Si bien este trabajo es desarrollado desde el punto de vista algorítmico, aporta con la valorización de lo multidimensional, la incorporación de métricas de conformance al análisis multidimensional, y la evaluación de su herramienta a través de ISO9126.

Durante el mismo año se desarrolló la herramienta ProcessCube (Mamaliaga, 2013) a nivel comparativo de procesos. Este trabajo incorpora la idea de comparar el mismo proceso para distintos segmentos, o a nivel de distintas organizaciones, incorpora una formalidad al diseño multidimensional del log de eventos. Estos últimos dos trabajos no se enfocan en la tarea de análisis del usuario propiamente

tal en su uso cotidiano, como si lo hace BPMNAnalysist y la presente investigación.

De la revisión bibliográfica realizada se observa entonces que, para un descubrimiento de procesos para un analista de procesos, hace falta un método que sea diseñado para facilitar la tarea de análisis multidimensional. Esto puede ser logrado a través de un enfoque que considere la tarea más allá de la solución algorítmica, ocupando las herramientas existentes, considerando una visión holística de proceso y teniendo en cuenta conceptos de usabilidad.

3. PROPUESTA

Este artículo desarrolla un concepto de descubrimiento interactivo de procesos para usuarios de negocio en minería de procesos, combinando exploración, filtros dinámicos, navegación y actualización automática aplicando operaciones de OLAP. Lo anterior se propuso para mejorar algunos aspectos de la usabilidad con respecto a las herramientas actuales, mantener la funcionalidad y mejorar el logro en la comprensión de lo que sucede en la organización. Para diseñar la herramienta se enfatizó en el tipo de usuario y la tarea que debe cumplir. En la elaboración se considera el uso de herramientas de descubrimientos ya disponibles, a nivel algorítmico.

El enfoque OLAP permite al usuario interactuar con el proceso de manera similar a una tabla dinámica de hojas de cálculo, explorando el proceso a través de sus distintas dimensiones, filtrando, manteniendo constantes algunas dimensiones, buscando desde lo general a lo particular, permitiendo vistas de flujo e interacción de recursos y comparación de versiones. La herramienta propuesta la hemos denominada OLAPDiscovery, y está integrada dentro del framework ProM. Las dimensiones que se utilizaron son: redes sociales a través de SocialMiner (Van der Aalst, Reijers y Song, 2005), variantes y agrupación de clusters a través de TraceAlignment (Bose y Van der Aalst, 2010) y tiempo a través de los registros temporales de cada caso. Para visualización de resultados se utilizaron los diagramas resultantes de HeuristicMiner y SocialMiner.

Finalmente, se presenta un caso de estudio donde se buscó poner a prueba la herramienta con usuarios reales, y medir lo propuesto.

4. DESCRIPCIÓN DE OLAPDISCOVERY

4.1. Definición de descubrimiento para usuarios de negocio

4.1.1. Propósito

El propósito que buscamos cubrir en esta investigación es “analizar un proceso de negocio a través de herramientas de minería de procesos” por un analista de procesos. Analizar es distinguir y separar las partes de un todo hasta llegar a conocer sus principios y elementos constituyentes (Real Academia Española, 2001). En este caso, el todo es el proceso, un conjunto de actividades lógicamente relacionada que usa los recursos de la organización, para proveer resultados definidos, con el fin de alcanzar los objetivos del negocio. En la definición anterior se mencionan al menos cuatro componentes importantes:

Actividades lógicamente relacionadas: Relevante para entender el orden en el cual se ejecuta el proceso. En minería de procesos se utilizan distintas representaciones para describir dichas relaciones.

Recursos de la organización: Son quienes ejecutan las actividades e interactúan entre sí en el proceso trabajando juntos, delegando trabajo, ejecutando actividades similares, entre otras cosas. También existen diagramas específicos para representar las relaciones que existen entre ellos.

Proveer resultados definidos: Si bien los resultados esperados de un proceso pueden ser variados, una métrica muy importante es el tiempo. Ya sea en la duración, tiempo de espera u otros.

Alcanzar objetivos del negocio: Cada proceso puede tener más de un resultado esperado y objetivos distintos. Generalmente cada variación del proceso se puede deber a un objetivo distinto o variaciones de un objetivo; pero también se puede producir por anomalías.

4.1.2. Usuario

Un proceso de negocio pertenece a la organización, pero no cualquier perfil es encargado de auditarlo. El analista de procesos debe preocuparse de hacer seguimiento y control a los procesos para promover mejoras e informes de gestión. Este tipo de usuarios está familiarizado con la Gestión Orientada a Procesos y tiene conocimiento del negocio en que está su organización, pero no necesariamente tiene dominio experto en minería de datos o minería de procesos. Complementando esa información con la clasificación de usuarios según el área de Inteligencia de Negocios, se puede describir que en algunos casos estos usuarios son consumidores de información y en otros productores de información, y en el caso de ser muy capacitados utilizan a lo más hojas de cálculo para distintos tipos de informes. Para este tipo de usuario se recomiendan herramientas query/reporting tools, standard reports, OLAP y hojas de cálculo.

4.1.3. Brechas con el descubrimiento actual

Actualmente el descubrimiento de procesos considera resultados unidimensionales de flujo de actividades u organizacionales, como en ProM (Process Mining Group,

2013), o en algunos casos bidimensionales informativos como en Disco (Fluxicon Process Laboratories, 2013), donde no es clara la posibilidad de interactuar con otras dimensiones o de explorar dinámicamente el registro de eventos. Esta última opción existe solo filtrando previamente una y otra vez para cada aspecto que se quiere observar de los registros. Un analista de procesos requiere hacer y deshacer cambios dentro de su registro para observar como varía el resultado según lo que se requiera mantener y sacar. Además, hoy necesita internalizar muchos parámetros y el hecho de saberlos no necesariamente se traduce en grandes cambios en la información obtenida. En el año 2011, 20 investigadores propusieron los desafíos que existían en el área, y uno de los diez desafíos hacía referencia a mejorar la usabilidad para usuarios no expertos (Van der Aalst et al., 2012). Además, a comienzos de este año apareció un estudio que indicaba que los mayores inconvenientes de la herramienta era que es poco intuitiva y difícil de entender (Claes y Poels, 2013). Este problema puede tener más de una causa. En esta investigación se ve desde el punto de vista de la definición de la tarea, y desde ahí se analiza la brecha entre lo actual y lo deseable.

El descubrimiento deseable debe poder variar las distintas variables que influyen en el proceso, y ver dinámicamente como esa variación determina el comportamiento de flujos organizacionales y de actividades.

4.1.4. Justificación OLAP

Dentro del concepto de descubrimiento interactivo es donde aparece la herramienta OLAP. OLAP es un tipo de aplicación que pretende facilitar el análisis multidimensional de la información (datos que han sido agregados en varias categorías, o “dimensiones”). Facilita al usuario sintetizar información de la organización a través de vistas comparativas, personalizadas; y analizar información tanto histórica (pasado) como estimada (futura).

Dadas las definiciones del propósito que se quiere abordar y del usuario objetivo, es razonable escoger un estilo OLAP para implementar un uso interactivo, por las siguientes razones:

El tipo de usuario no es experto en análisis avanzado y está alejado de conceptos técnicos detallados de minería de procesos.

La información sobre la ejecución de un proceso de negocio contenida en un log de eventos se ajusta perfectamente para implementarlo en un cubo con al menos tres dimensiones: recursos, tiempo y variantes. Lo cual hace que la acción de analizar se pueda hacer sobre él de manera natural y no forzada.

Es una herramienta que sus acciones permiten llevar a cabo la acción de analizar. Las acciones slices and dices, pivot, drill down y bottom up, facilitan la comprensión del proceso, explorando la información poco a poco, de manera interactiva, profundizando en cada paso en una perspectiva a la vez.

4.2. Propuesta inicial

4.2.1. Diseño propuesta inicial

a) Dimensiones

i) Variante

Una variante es una forma de ejecutar el proceso, una cierta secuencia de actividades, que ha sido registrada en el log de eventos. Puede repetirse en distintas instancias a través de casos que tienen ejecutores y duraciones que pueden ser distintos. La dimensión tiene como objetivo agrupar variantes que tengan características en común, y poder navegar dentro de estas agrupaciones de tal manera que los registros de eventos se pueden explorar fácilmente. Para representar esta dimensión se ocupó el plug-in

TraceAlignment (Bose y Van der Aalst, 2010). TraceAlignment agrupa variantes en clusters según distintos parámetros. Puede ser utilizada para explorar el proceso en las primeras etapas de análisis, y para responder a las preguntas específicas en etapas posteriores. Por tanto, complementa las técnicas de minería de proceso existentes que se centran en el descubrimiento y comprobación de conformidad. Permite conocer la columna vertebral del proceso, los puntos críticos que comparten las variantes, visualizar brechas entre lo que sucede y lo que debería suceder, detectar patrones y descubrir variaciones con pequeñas variantes a las deseadas, entre otras cosas.

En la herramienta será posible elegir el número de agrupaciones o clusters con el cual partir el análisis.

ii) Recurso

Se ocupará la herramienta SocialMiner, pero esta vez se utilizarán las agrupaciones obtenidas para permitir que el usuario pueda hacer drill down y roll up con grupos, según el algoritmo elegido. Además, es posible cambiar el valor umbral para decidir cuándo dos o más recursos están agrupados.

iii) Tiempo

Para el tiempo se ha definido extraer la duración de cada caso, y permitir el filtrado a través de navegar entre la duración mínima y máxima. Además, será posible visualizar la media y desviación estándar del tiempo de ejecución del proceso.

b) Formas de ver el proceso

Para la visualización, se ha propuesto ver el resultado del filtrado del proceso en un flujo de actividades o de recursos a través de la función pivote. Para el flujo

de actividades se escogió HeuristicMiner, ya que puede hacer frente al ruido, y se puede utilizar para expresar el comportamiento principal (es decir, no todos los detalles y excepciones) del log de eventos (Weijters, Van der Aalst y De Medeiros, 2006). Para el flujo de recursos se ocupó SocialMiner. Esta última permite descubrir interacciones de los ejecutores del proceso de manera gráfica. Se ocupará Working Together, Similar Task, Handover Of Work, por ser los que más información entregan a un usuario de negocio.

4.2.2. Funcionalidades adicionales

i) Vistas

Se puede utilizar una vista de análisis individual o una vista comparativa donde aparece el diagrama base y el diagrama que se está filtrando. En la dimensión de versión, es posible cargar análisis realizados anteriormente. El usuario puede decidir si cargar el análisis en la parte superior como base de análisis o en la parte inferior para seguir filtrando la versión. La imagen a continuación señala cual sería la parte inferior, la parte superior y el cuadro para cargar las versiones guardadas.

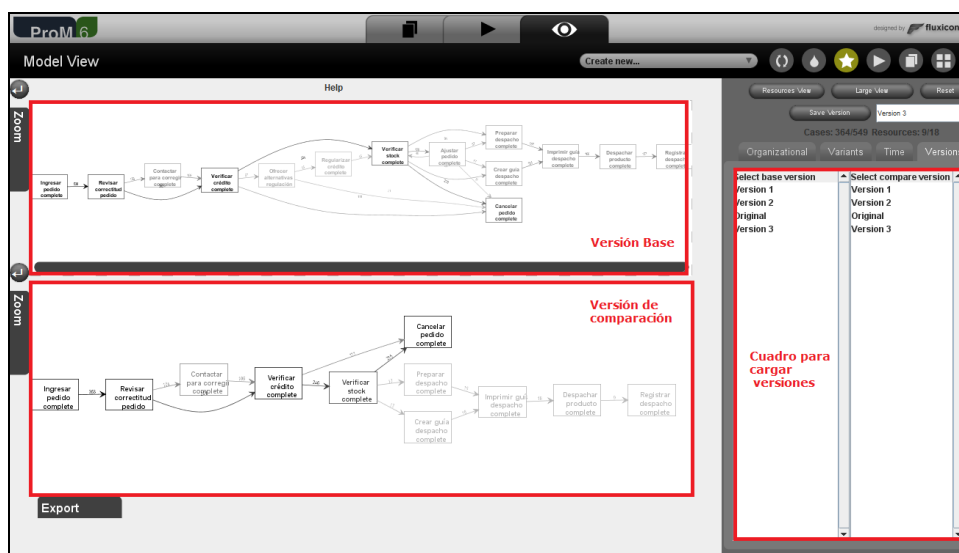


Figura 6 - Manejo de Versiones en OLAPDiscovery

ii) Reset

Botón que permite comenzar de cero todo el análisis.

iii) Tooltip

Se utilizan tooltips (textos emergentes) en recursos para mostrar qué actividades ejecuta cada recurso. Además, se utilizan en el filtro de tiempo para indicar la selección actual y en variantes para saber cuántas actividades hay por variante.

4.2.3. Implementación inicial

Inicialmente se construyó una herramienta que consideraba las dimensiones de variantes, el tiempo y recursos con las consideraciones antes mencionadas.

Un punto a considerar en esta herramienta es que cada vez que se cambia de dimensión, el plug-in vuelve a calcular los miembros a desplegar en cada dimensión, ya que si se sacan recursos pueden cambiar variantes o si se sacan variantes pueden quedar recursos fuera del log de trabajo. Además, recalculando los miembros se podrían haber obtenido nuevos clusters o grupos de interacción social, y el recalcular podría haber sido valioso. Debido a estas dos razones, cada cambio de dimensión corresponde a comenzar a trabajar con un sub-cubo del original y con un nuevo despliegue de miembros de estos dos algoritmos, lo cual no lo hace ser un OLAP 100% y lo llevo a ser llamado inicialmente StackOLAP.

4.2.4. Prueba exploratoria

Inicialmente se probó la versión inicial de la herramienta para evaluar qué se podía mejorar, utilizando como guía los atributos de ISO9126-2.

i) Usuarios

Se realizó una prueba exploratoria a veinticuatro usuarios en dos grupos. El primer grupo era de diecinueve personas y estaba compuesto por estudiantes que durante todo el semestre realizaron un curso de modelos de procesos, donde habían aprendido todos los conceptos relacionados al enfoque BPM. Además, conocían la herramienta ProM a nivel básico. Es decir, conocían algoritmos de descubrimiento de procesos organizacionales y de actividades. Incluso ya habían utilizado filtros y herramientas de verificación de conformidad. Sin embargo, no habían estudiado en detalle los parámetros, limitaciones e investigaciones relacionados con las herramientas. El segundo grupo estaba conformado por personas que habían finalizado el curso de minería de procesos y conocían al menos el 90% de las técnicas más populares del área (Claes y Poels, 2013). Además, habían sido enfrentados a resolución de problemas dado un log de eventos en al menos 7 ocasiones.

ii) Descripción de prueba

Para ejecutar las pruebas se le entregó a cada usuario un documento con tres secciones, cada una se describe a continuación:

- Descripción del contexto del rol que tenían ellos en la prueba. Básicamente consistía en ser un analista de negocio con un log de eventos y la herramienta para analizarlo.
- Pequeño tutorial con las funcionalidades en caso de ser necesario.

- Sección de preguntas que no era necesario responder, pero se proponía como útil para evaluar la herramienta.

Además, dos observadores apoyaron respondiendo dudas sobre la herramienta y observaron el desarrollo de la actividad.

iii) Evaluación

Al finalizar la prueba exploratoria, los usuarios debían dar feedback sobre qué mejorarían en la herramienta, y su apreciación subjetiva sobre usabilidad y funcionalidad.

iv) Resumen de resultados

Los resultados obtenidos muestran mejoras en la percepción de la herramienta con respecto a ProM. Sin embargo, los usuarios manifestaron su opinión contraria al re cálculo por dimensión, y además agregaron que sería útil un manejo de versiones.

4.2.5. Versión final OLAPDiscovery

i) Mejoras

Debido a la prueba exploratoria se decidió trabajar con una lógica 100% OLAP sin re cálculo de dimensión, y con la opción de guardar versiones para el análisis. Además, se agregó la característica de texto emergente para que el usuario pudiera ver cuáles eran las actividades de un ejecutor o cuál era el tiempo actual seleccionado. La versión del programa final con sus componentes se muestra en la figura a continuación.

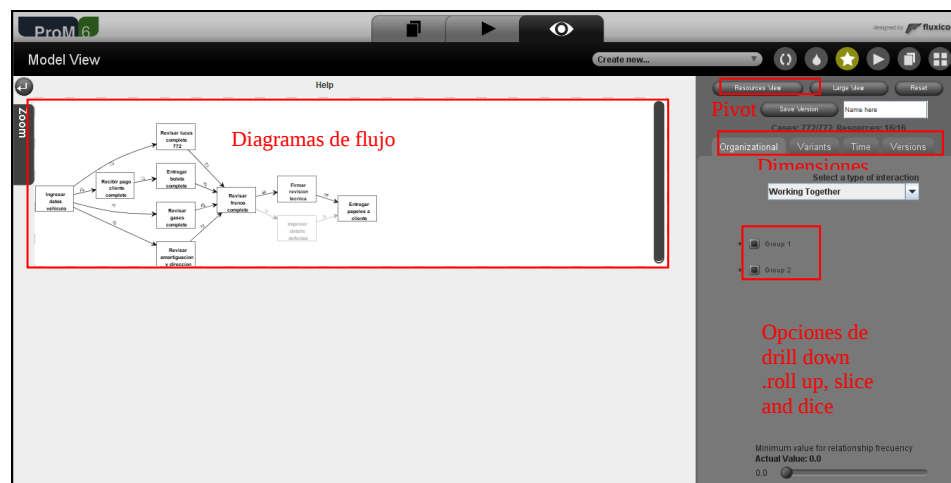


Figura 7 - OLAPDiscovery Final

Como se puede ver, la herramienta tiene una vista inicial con el diagrama de flujo, botones en la parte superior derecha que permiten cambiar a vista de recursos, botón de vista de página completa, resetear el análisis y botón de guardado de versiones. Además, en la parte media de ese sector se puede ver una pestaña por dimensión. Haciendo click en cada una se puede filtrar el registro. Para mayor detalle en los anexos hay vistas de cada funcionalidad particular.

5. CASO DE ESTUDIO

En esta sección se describe el caso de estudio, en el que se puso a prueba las herramientas OLAPDiscovery, Disco y ProM con distintos usuarios, para evaluar el uso de estos programas y analizar si existe una mejor percepción entre este grupo de usuarios de la herramienta diseñada en comparación a las otras.

5.1. Descripción de los usuarios

El perfil de los usuarios es ser profesionales del área de tecnologías de información o estar en los últimos dos años de carreras afines. Además, deben haber cursado algún curso de modelación de procesos y/o minería de procesos. En el primer curso se aprenden conceptos de BPM y se utiliza de manera básica la herramienta ProM; en el segundo curso se aprende exhaustivamente minería de procesos en base al libro de van der Aalst. En estas pruebas se consideró que las personas que habían cursado solo el primer curso serían usuarios básicos y las personas que habían cursado el segundo curso se considerarían expertos. La composición de la muestra fue de 38 personas, de los cuales 16 fueron expertos; y 22, básicos. Cada prueba fue tomada en distintas ocasiones, en grupos de 1 a 6 usuarios.

5.2. Descripción de las tareas a realizar por los usuarios

La tarea a realizar por los usuarios está compuesta por dos secciones de cuatro preguntas cada una. La primera es de análisis de temas unidimensionales, una por cada dimensión del proceso considerada. Para las afirmaciones se debía determinar si eran verdaderas o falsas, y para las preguntas se debía marcar una alternativa. Las afirmaciones y preguntas eran:

- Observe el diagrama y determine si de la actividad X se desprenden 2 o más caminos.

- Observe el diagrama social y determine si las personas X, Y, Z trabajan juntas.
- ¿Cuántas variantes tiene su proceso?
- ¿Cuánto dura el caso más largo y el caso más cortó?

La segunda sección se compuso de tareas que combinaban diferentes dimensiones. Para cada aseveración se debía contestar si era verdadera o falsa; las afirmaciones se presentan a continuación:

- Para el grupo X del recurso Y, para todos los casos de duración mínima posible no participa Y.
- En el filtro anterior no se encuentra la actividad Z.
- En todos los casos de duración máxima del grupo de Y solo hay 1 variante.
- El grupo X ejecuta más de 10 variantes que el grupo W.

Para completar cada tarea el usuario no tenía un tiempo máximo asociado, pero sí existía una duración aconsejable que no excedía los 12 minutos para tareas multidimensionales y 5 minutos para tareas unidimensionales. Si el usuario consideraba pertinente podía continuar, o si no, podía determinar que no lo logró por tiempo.

Previo a las pruebas, cada usuario recibía una capacitación exacta de lo que debía hacer, además en cada pregunta se explicaba paso a paso cómo lograr realizar el análisis. Esto se determinó porque se hicieron pruebas exploratorias previamente y era muy complejo que los usuarios se acordarán de todos los pasos o los descubrieran en un tiempo aceptable. Además, si para las tres herramientas se realizaba la misma introducción y se entregaba una explicación de cómo llevar a cabo cada tarea, esto no debería influir en los resultados obtenidos.

Finalmente, cada usuario debía contestar tres test de ocho preguntas, en base a tres registros de eventos distintos, sin límite de tiempo, pero con la posibilidad de terminar la tarea en caso que haya excedido el tiempo aconsejable. Además, debía anotar el tiempo utilizado en cada tarea, y al finalizar el test debía contestar que tan de acuerdo o en desacuerdo estaban con las afirmaciones planteadas. Cada herramienta se evaluó en distinto orden por cada usuario, para tener presente qué tanto influye usar una después de la otra, o usarla primero en relación a usarla en último lugar.

5.3. Descripción de la evaluación

Para la evaluación se diseñó un método que combina técnicas de ISO9126-3 e ISO9126-4. De ISO9126-4 se midió efectividad, productividad y satisfacción; se dejó de lado seguridad, por considerarse que no es el foco de esta investigación. La Figura a continuación muestra en azul el subconjunto de atributos que se desean medir de esta metodología:

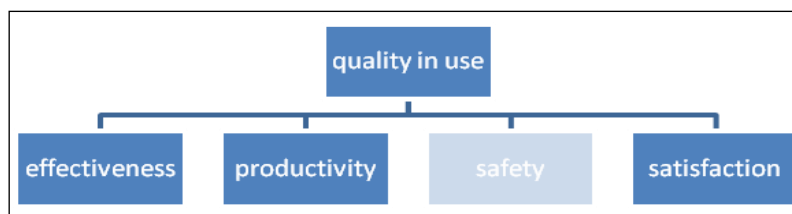


Figura 8 - Selección de atributos a medir de ISO9126-4

Además, se agregó atributos de ISO9126-3 como idoneidad, o es operativo, comprensión y aprendizaje. Los demás atributos no correspondía medirlos, porque no tienen relación al enfoque del diseño de la herramienta. La Figura siguiente muestra el subconjunto de propiedades que se desean medir de ISO9126-3.

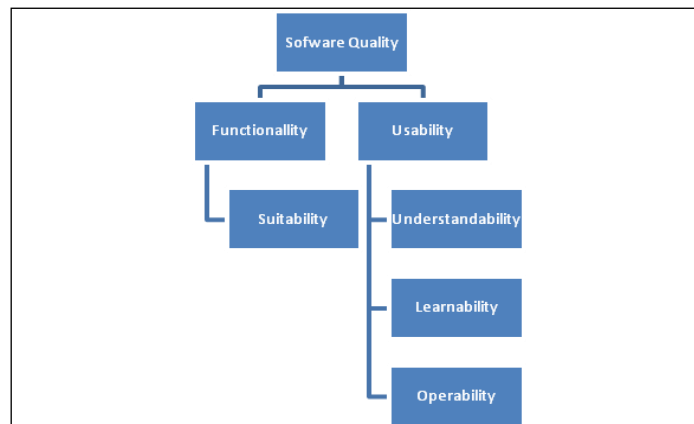


Figura 9 - Atributos seleccionados de ISO9126-3

5.4. Métricas propuestas

Para la medición se contempló medidas subjetivas y objetivas. Las primeras corresponden a la percepción del usuario y las segundas corresponden al desempeño en la tarea propuesta en cuanto a tiempo, logro, errores. Para los atributos de productividad, efectividad y métricas objetivas, se proponen las siguientes métricas:

Tabla 1 - Métricas objetivas

Dimensión	Métrica
Tareas completas	Tareas terminadas/Tareas Totales
Efectividad	Tareas logradas correctamente/Tareas Totales
Tiempo unidimensional	Tiempo total de ejecución de tareas unidimensionales
Tiempo multidimensional	Tiempo total de ejecución de tareas multidimensionales
Productividad efectiva	Efectividad/Tiempo total de ejecución de las tareas

Para los demás atributos se realizó una lista de al menos 100 afirmaciones. Cada una se categorizó en una de las características que se querían medir, es decir, idoneidad, satisfacción, efectividad, productividad, “es comprensible”, “es operativo” y aprendizaje. Luego, con el juicio experto de personas del área, se seleccionaron las más representativas. La idea era que no fuera un cuestionario extenso, ya que el usuario debía contestarlo para cada herramienta. El resumen se muestra a continuación.

Tabla 2 - Métricas subjetivas

Dimensión	Métrica
Idoneidad	Frecuencia de percepción sobre las afirmaciones: -Los parámetros disponibles son adecuados -La herramienta es útil para el análisis de mi procesos -La organización de los menús es bastante lógica
Satisfacción	Frecuencia de percepción sobre las afirmaciones: -Recomendaría este programa a mis colegas -Volvería a usar este programa para analizar procesos
Efectividad	Frecuencia de percepción sobre las afirmaciones: -Cree que puede entender un proceso completo. *Completo: acabado, con todas las particularidades que lo distinguen -Cree que puedo entender con precisión un proceso.

	*Precisión: fijo, exacto, que se distingue con claridad
Productividad	Frecuencia de percepción sobre las afirmaciones: -Los pasos necesarios para hacer una tarea de análisis son los adecuados -El tiempo utilizado para hacer el análisis es razonable
Es comprensible	Frecuencia de percepción sobre las afirmaciones: -Las funcionalidades de esta herramienta son fáciles de entender -Puedo entender la información
Es operativo	Frecuencia de percepción sobre las afirmaciones: -Es fácil actuar en base a la información proporcionada -Es fácil combinar un análisis de todas las dimensiones de un proceso -Es fácil pasar de una parte de una funcionalidad a otra
Aprendizaje	Frecuencia de percepción sobre las afirmaciones: -Es fácil recordar cómo hacer las cosas -La herramienta es sencilla de usar

5.5. Resultados

A continuación se presentan los resultados de la evaluación. En el punto 5.5.1 se pueden ver los tiempos de uso, en el punto 5.5.2 se puede observar logro, efectividad y productividad. Finalmente, en el punto 5.5.3 se pueden ver los resultados para las métricas subjetivas.

5.5.1. Tiempos de uso

Para llegar a conclusiones respaldadas estadísticamente se utilizó el estadístico de prueba de Mann-Whitney (Paul Newbold, 2008) dado que a priori no se tenían distribuciones normales. Además, este método es muy usado para estudio de variables ordinales como los resultados de afirmaciones de escala Likert.

Los tiempos muestran resultados favorables para OLAPDiscovery (en adelante en los gráficos aparecerá como OLAP-D) tanto en las tareas multidimensionales como las tareas unidimensionales. El mejor intento de Disco no supera al peor intento de OLAPDiscovery. Además, el mejor intento de ProM casi duplica en duración al peor intento de OLAPDiscovery, y casi triplica al mejor.

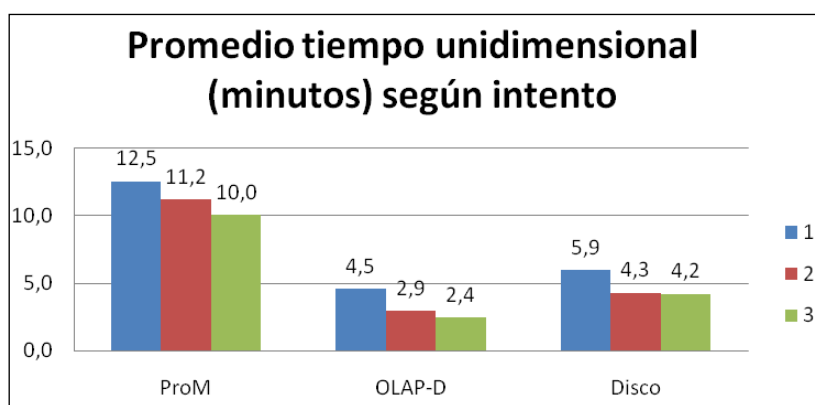


Figura 10 – Promedio tiempo unidimensional (minutos) según intento

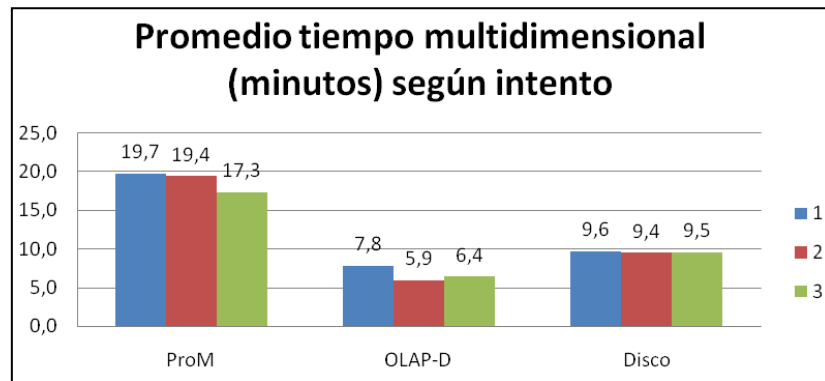


Figura 11 – Promedio tiempo multidimensional (minutos) según intento

Las siguientes tablas muestran otras medidas de dispersión para ambos casos. Las definiciones de Kurtosis, coeficiente de variación y asimetría y la interpretación de estos indicadores se puede encontrar en el anexo 10.4.

Tabla 3 - Medias de dispersión de tiempo unidimensional

Herramienta	Media	Dev.estándar	Coef.Variación	Kurtosis	Asimetría
Prom	10,8	4	0,4	2,2	1,4
Disco	4,9	2,5	0,5	0,4	1,1
OLAP-D	3,3	1,8	0,5	5	1,9

Tabla 4 - Medidas de dispersión de tiempo multidimensional

Herramienta	Media	Dev.estándar	Coef.Variación	Kurtosis	Asimetría
Prom	19,3	7,2	0,4	0,7	1
Disco	9,3	2,5	0,3	-0,5	0,2
OLAP-D	6,5	2,5	0,4	0,1	0,6

Para los tiempos unidimensionales se ve que ProM y OLAP-D tienen un indicador leptocúrtico, es decir, tienen una mayor concentración de valores, en contraposición con Disco, que tiene una distribución con concentraciones normales de valores. Para

las tres herramientas los valores tienen una asimetría positiva por lo que los resultados tienden a ser menores que la media. Además, los coeficientes de variación muestran que todas las desviaciones estándar están 40% por sobre la media, salvo Disco (30%).

Para los tiempos multidimensionales solo ProM cae en la categoría de kurtosis leptocurtiva, es decir con concentraciones muy cerca de la media, mientras que las otras herramientas muestran un resultado mesocúrtico, es decir los datos tienen una concentración normal. Para la asimetría se puede decir que ProM y OLAP-D son asimétricos positivos, es decir sus valores tienden a ser menores que la media y Disco es simétrico.

5.5.2. Tarea y productividad

Para la correcta ejecución de tareas se observa tres datos relevantes. En primer lugar, ProM presenta menores índices en capacidad de completar las tareas multidimensionales. En un 29% de los casos, los usuarios no pudieron terminas las tareas, siendo en su mayoría usuarios básicos. Para ese mismo ítem, sólo un usuario tuvo tareas incompletas con OLAP; y dos, con Disco. A continuación se muestra un resumen de resultados sólo para respuestas multidimensionales. Si bien ProM tiene menos porcentaje de error, entre las respuestas no completadas y las incorrectas tiene un mayor porcentaje de no logradas, mientras que OLAP y Disco tuvieron resultados muy similares.

Para las tareas unidimensionales se puede observar un mayor porcentaje de error en Disco. Esto se dio por el análisis de flujo de recursos, ya al tener que realizarse en registros separados (por limitaciones de la herramienta), el usuario se confundía y respondía en base a un registro que no correspondía.

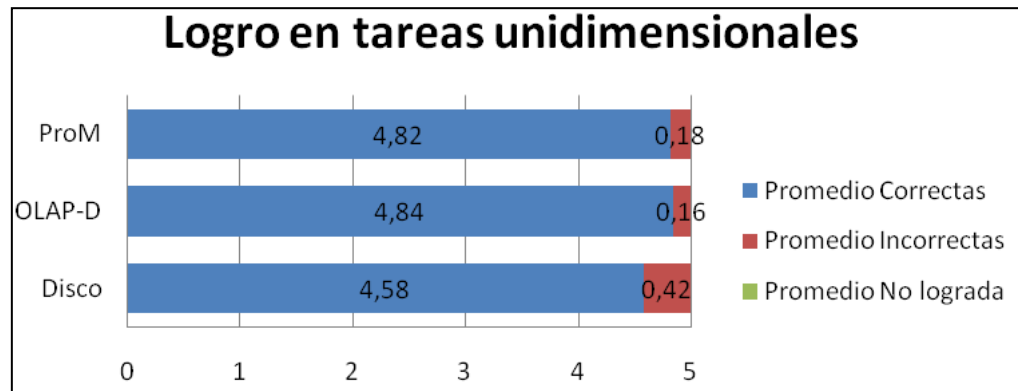


Figura 12 – Logro en tareas unidimensionales

Por otro lado para las tareas multidimensionales, ocurre lo que ya se mencionó. Se ve que ProM tiene un elevado índice de no logrados, para OLAP-D y Disco hay una tendencia similar.

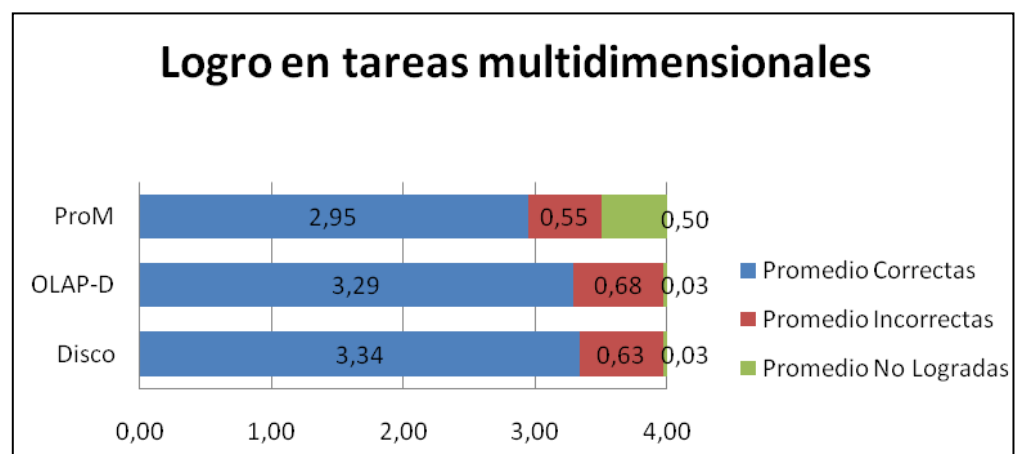


Figura 13 – Logro en tareas unidimensionales

Continuando se puede ver que el índice de respuestas correctas en relación al tiempo utilizado para las tareas unidimensionales beneficia a OLAP-D en todos los intentos.

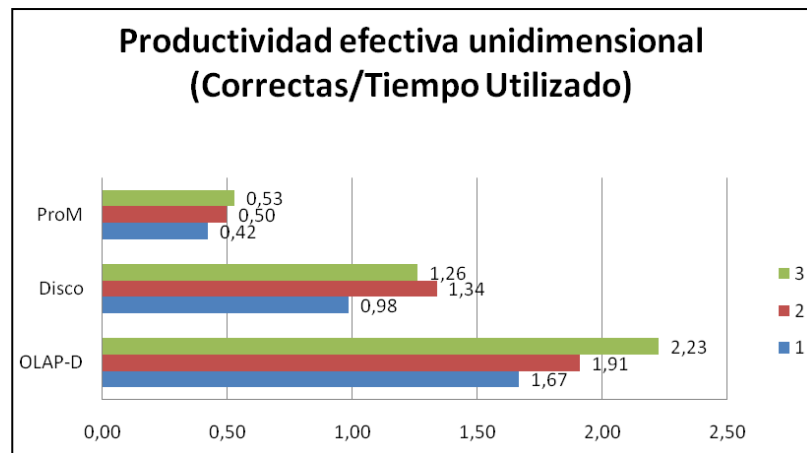


Figura 14 – Logro en tareas unidimensionales

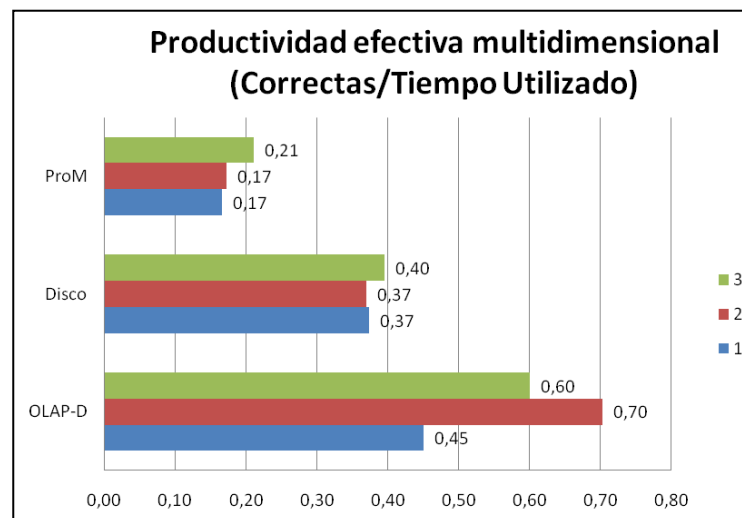


Figura 15 - Productividad efectiva multidimensional

5.5.3. Apreciaciones subjetivas

La siguiente tabla muestra la distribución de resultados según la frecuencia en que respondieron las preguntas. En primer lugar, se presenta el contraste de OLAP-D con ProM. El número 1 indica totalmente en desacuerdo, el 2 en desacuerdo, el 3 ni de acuerdo ni en desacuerdo, el 4 de acuerdo y el 5 totalmente de acuerdo.

Tabla 5 - Distribución de respuestas subjetivas entre ProM y OLAP-D

Afirmación	Dimensión	Totalmente en desacuerdo		En desacuerdo		Ni de acuerdo, ni en desacuerdo		De acuerdo		Totalmente de acuerdo	
		ProM	OLAP-D	ProM	OLAP-D	ProM	OLAP-D	ProM	OLAP-D	ProM	OLAP-D
Es fácil recordar cómo hacer las cosas	Aprendizaje	8%	0%	32%	5%	26%	0%	29%	63%	5%	32%
La herramienta es sencilla de usar	Aprendizaje	24%	0%	42%	0%	18%	13%	13%	53%	3%	34%
Las funcionalidades de esta herramienta son fáciles de entender	Es comprendido	13%	0%	47%	0%	26%	24%	11%	55%	3%	21%
Puedo entender la información	Es comprendido	5%	0%	21%	0%	21%	11%	53%	58%	0%	32%
Es fácil actuar en base a la información proporcionada.	Es operativo	8%	0%	16%	3%	34%	5%	37%	53%	5%	39%
Es fácil combinar un análisis de todas las dimensiones de un proceso	Es operativo	18%	0%	47%	11%	16%	8%	18%	42%	0%	39%
Es fácil pasar de una parte de una funcionalidad a otra.	Es operativo	13%	0%	29%	3%	24%	5%	21%	50%	13%	42%
Los parámetros disponibles son adecuados.	Idoneidad	3%	0%	16%	3%	29%	24%	47%	50%	5%	24%
La herramienta es útil para el análisis de mi procesos	Idoneidad	0%	0%	16%	0%	26%	13%	34%	45%	24%	42%
La organización de los menús es bastante lógica.	Idoneidad	11%	0%	32%	11%	32%	16%	16%	47%	11%	26%
Los pasos necesarios para hacer una tarea de análisis son los adecuados	Productividad	11%	0%	34%	5%	32%	11%	21%	55%	3%	29%
El tiempo utilizado para hacer el análisis es razonable	Productividad	21%	0%	29%	0%	26%	11%	24%	32%	0%	58%
Recomendaría este programa a mis colegas.	Satisfacción	13%	0%	21%	0%	32%	16%	29%	29%	5%	55%
Volvería a usar este programa para analizar procesos	Satisfacción	18%	0%	8%	0%	29%	11%	39%	45%	5%	45%
Cree que puede entender un proceso completo. *Completo: acabado, con todas las particularidades que lo distinguen	Efectividad	8%	0%	26%	3%	21%	16%	39%	53%	5%	29%
Cree que puedo entender con precisión un proceso. *Precisión: fijo, exacto, que se distingue con claridad	Efectividad	5%	0%	21%	5%	29%	29%	37%	39%	8%	26%

En general, la diferencia en cada afirmación es bastante amplia. OLAP tiene su moda entre el “totalmente de acuerdo” y “de acuerdo” para todas sus afirmaciones, mientras que ProM, solo en siete casos tiene modas correspondientes a la apreciación “de acuerdo” y en ninguna afirmación supera a OLAP-D en frecuencia. En el anexo 10.5 se pueden observar las pruebas no paramétricas de Mann – Whitney aplicadas a la distribución de respuestas de ambas muestras. Estas últimas arrojan que en todas las métricas es posible decir con valores menores a 0,05 de error que la apreciación de OLAP-D es distinta en estas afirmaciones con respecto a ProM. Eso combinado a la tabla de frecuencia anterior nos permite señalar que esa diferencia es favorable a OLAP-D.

Con respecto a Disco, se puede observar la siguiente tabla de porcentajes en la distribución de las respuestas:

Tabla 6 - Distribución de respuestas subjetivas entre Disco y OLAP-D

Afirmación	Dimensión	Totalmente en desacuerdo		En desacuerdo		Ni de acuerdo, ni en desacuerdo		De acuerdo		Totalmente de acuerdo	
		Disco	OLAP-D	Disco	OLAP-D	Disco	OLAP-D	Disco	OLAP-D	Disco	OLAP-D
Es fácil recordar cómo hacer las cosas	Aprendizaje	0%	0%	3%	5%	24%	0%	53%	63%	21%	32%
La herramienta es sencilla de usar	Aprendizaje	0%	0%	0%	0%	18%	13%	55%	53%	26%	34%
Las funcionalidades de esta herramienta son fáciles de entender	Es comprendido	0%	0%	3%	0%	18%	24%	45%	55%	34%	21%
Puedo entender la información	Es comprendido	0%	0%	0%	0%	8%	11%	53%	58%	39%	32%
Es fácil actuar en base a la información proporcionada.	Es operativo	0%	0%	3%	3%	24%	5%	53%	53%	21%	39%
Es fácil combinar un análisis de todas las dimensiones de un proceso	Es operativo	5%	0%	24%	11%	21%	8%	39%	42%	11%	39%
Es fácil pasar de una parte de una funcionalidad a otra.	Es operativo	0%	0%	8%	3%	24%	5%	45%	50%	24%	42%
Los parámetros disponibles son adecuados.	Idoneidad	0%	0%	0%	3%	16%	24%	66%	50%	18%	24%
La herramienta es útil para el análisis de mi procesos	Idoneidad	0%	0%	0%	0%	21%	13%	37%	45%	42%	42%
La organización de los menús es bastante lógica.	Idoneidad	0%	0%	13%	11%	16%	16%	45%	47%	26%	26%
Los pasos necesarios para hacer una tarea de análisis son los adecuados	Productividad	0%	0%	3%	5%	18%	11%	61%	55%	18%	29%
El tiempo utilizado para hacer el análisis es razonable	Productividad	0%	0%	0%	0%	18%	11%	53%	32%	29%	58%
Recomendaría este programa a mis colegas.	Satisfacción	0%	0%	0%	0%	8%	16%	55%	29%	37%	55%
Volvería a usar este programa para analizar procesos	Satisfacción	0%	0%	0%	0%	8%	11%	63%	45%	29%	45%
Cree que puede entender un proceso completo. *Completo: acabado, con todas las particularidades que lo distinguen	Efectividad	5%	0%	8%	3%	18%	16%	45%	53%	24%	29%
Cree que puedo entender con precisión un proceso. *Precisión: fijo, exacto, que se distingue con claridad	Efectividad	3%	0%	8%	5%	26%	29%	39%	39%	24%	26%

En relación a los datos de la tabla se puede apreciar indicadores bastante similares entre ambas herramientas. En aprendizaje y “es operativo”, OLAP-D tiene mayor proporción de apreciaciones nivel “totalmente de acuerdo” y “de acuerdo”. Para productividad y satisfacción, supera a Disco en “totalmente de acuerdo”. La segunda lectura que debe hacerse a las apreciaciones anteriores es lo que nos dice el test de Mann-Whitney para cada afirmación entre OLAP y Disco. En general, el test arroja errores mayores a 0,05 lo que no permite señalar que un atributo es evaluado de manera distinta para ambas herramientas. Salvo para la afirmaciones de “es operativo” y para “el tiempo utilizado para hacer el análisis es razonable” de la dimensión productividad, en que el test demuestra que el atributo es evaluado de manera distinta para ambas herramientas y al observar se ve resultados favorables a OLAP-D.

Para la frecuencia por tipo de usuario se puede ver el anexo 10.6. En general, lo más llamativo es que los usuarios expertos valoran más ProM, pero su valoración no llega a un punto que ProM supere las otras herramientas. Para los usuarios expertos se observa una tendencia similar al estudio general, donde se valora más en OLAP-D el atributo “es operativo” y su productividad. Los estadísticos de prueba del anexo 10.7 muestran que las afirmaciones son favorables para los mismos casos que en el estudio general para OLAP-D, salvo para la afirmación de “es operativo” que señala “Es fácil actuar en base a la información proporcionada”. Para los usuarios básicos, no hay diferencias significativas entre OLAP-D y Disco. Esto se puede explicar, porque si bien OLAP-D ofrece una funcionalidad más avanzada, Disco es más amigable para los usuarios básicos.

Con respecto al uso previo de una herramienta antes que la otra, no se encontraron diferencias significativas al usar Disco antes que OLAP-D y OLAP-D antes que Disco. Cuando se ocupaba ProM antes que las dos anteriores se repetía la tendencia de encontrar más productivo a OLAP-D y con un error menor a 0,05. Por otro lado, cuando no se usaba ninguna antes, se presentaron datos favorables a OLAP-D en “Es

fácil actuar en base a la información proporcionada” correspondiente a la dimensión “es operativo”.

En la medida que se fue haciendo un análisis más específico de la información, fue más complejo encontrar diferencias significativas entre las evaluaciones de las herramientas. Sin embargo, es relevante señalar que OLAP-D superó a ProM en todas las dimensiones, en la mayoría empató según estadísticos de prueba y en “es operativo” y un componente de productividad salió favorecido. Los mismos indicadores se mantuvieron al separar la muestra encontrando que para usuarios básicos no había diferencia entre Disco y OLAP-D. Para usuarios expertos se mantuvieron casi las mismas observaciones positivas que en general (salvo una de “es operativo”) y para intentos previos se observaron observaciones positivas con respecto a la productividad cuando se usaba ProM antes y con cuando no se usaba ninguna se valoraba bastante el hecho de poder operar con la información dada.

5.6. Limitaciones de la evaluación

Las limitaciones del caso de estudio son:

- La prueba fue diseñada para medir análisis multidimensional pero de manera acotada, lo cual no permite sacarle provecho 100% a ProM y a Disco.
- Los resultados individuales están sujetos a la comprensión lectora de cada persona. Varias personas requirieron asistencia, pese a que las instrucciones eran claras respecto a las tareas que debían realizar.
- La poca cantidad de usuarios en relación al universo no permite inferencias relevantes con validez externa. Sin embargo, es un ejercicio exploratorio con una cantidad de usuarios relevante para el nivel de adopción que estas herramientas tienen en esta disciplina.

6. CONCLUSIONES

Del análisis de las métricas objetivas se desprende que el descubrimiento dinámico e interactivo provisto por OLAPDiscovery influye directamente en la efectividad en cuanto a tiempo tanto en tareas unidimensionales como en tareas multidimensionales. La clave aquí es la interactividad dinámica que ofrece al usuario, y la capacidad de filtrar y automáticamente obtener resultados tanto en la perspectiva de control de flujo como en la perspectiva organizacional. Las otras herramientas separan las tareas de filtrar de la tarea de aplicar algoritmos de descubrimiento, tareas que están integradas en OLAPDiscovery. Sin embargo, este ahorro de tiempo no es sinónimo de menos errores, obteniéndose un nivel de logro (calidad de respuestas) similar entre OLAPDiscovery y Disco. No así en relación a ProM, en que con OLAPDiscovery se logra un porcentaje de respuestas mayor.

En el análisis de las métricas subjetivas se concluye que para todos los usuarios esta forma de descubrimiento les parece más operativa y más productiva; en contraste, para las demás métricas no hay información concluyente. Probablemente con un test más amplio podría haberse analizado más perspectivas de cada dimensión, sin embargo una condicionante al experimento era que no se podía ocupar más de una hora para no aburrir al usuario y que respondiera a consciencia.

Es muy interesante que para los usuarios no expertos OLAP-D no representa una real ventaja en relación a Disco. Esto se puede deber a la poca experiencia que han tenido estos usuarios trabajando con ejercicios de minería de procesos, sumada a la valoración de la visualización y otras bondades que tiene la herramienta Disco. Sin embargo, es muy positivo que se obtengan resultados similares y no peores. En el caso de los usuarios expertos se ve reflejada una parte del sentido de esta investigación. Mejorar la productividad y el concepto de “es operativo” es un real acierto con respecto a la tarea del analista. Para el concepto de “es comprensible”, satisfacción, idoneidad y

aprendizaje no se muestran grandes diferencias en relación a Disco, pero sí a ProM. Lo más rescatable en estos casos es que se mantiene la calidad de lo que ya existe.

En todas las métricas medidas se obtuvieron resultados iguales o mejores que Disco y siempre mejores que ProM. Objetivamente demora notoriamente menos en la tarea multidimensional manteniendo la misma tasa de error que los otros descubrimientos. El uso previo de ProM dio como resultado que la valoración en es productivo para OLAP-D era mayor que para Disco. Usar ninguna antes que Disco u OLAP-D mostró resultados favorables a este últimos en “es operativo”. No se encontraron otras tendencias significativas importantes con respecto al uso previo

Finalmente se puede decir que el descubrimiento mejora la usabilidad del descubrimiento desde el punto de vista de efectividad en tiempo y no de la satisfacción, ni aprendizaje.

7. TRABAJO FUTURO

A futuro visualizamos distintos tipos de líneas de investigación:

- A nivel de uso se debe evaluar incorporar este concepto de descubrimiento en las herramientas de minería de procesos. Además, se debe potenciar su uso combinado con otros algoritmos de descubrimiento como FuzzyMining y herramientas de verificación de conformidad.
- Es relevante también hacer experimentos exploratorios de evaluación de calidad, usabilidad y funcionalidad en las herramientas, ya que no solo importa el algoritmo, sino que también su uso satisfactorio por parte del usuario.
- También es importante considerar posibles mejoras en el descubrimiento a través de las métricas de uso donde este trabajo no puso énfasis, como lo son la percepción de aprendizaje y comprensión de la herramienta, la idoneidad y la satisfacción.
- Otra oportunidad es considerar distintos enfoques en la mejora del uso de las herramientas de minería de procesos, en este caso se propuso la tarea de análisis, en otros se considera la tarea de procesos poco estructurados (Günther y Van Der Aalst, 2007) y en otra se plantea la conceptualización del ciclo de vida BPM (İsmail Bayraktar, 2011).
- Extender la mecánica OLAP a otros atributos de datos contenidos en el log, como por ejemplo zonas geográficas, tipos de productos y otros posibles elementos.

REFERENCIAS

- Abran, A., Al-Qutaish, R. E., & Desharnais, J. M. (2005). Harmonization issues in the updating of the ISO standards on software product quality. *Metrics News*, 10(2), 35-44.
- Bandara, W., Chand, D. R., Chircu, A. M., Hintringer, S., Karagiannis, D., Recker, J. C., ... & Welke, R. J. (2010). Business process management education in academia: Status, challenges, and recommendations. *Communications of the Association for Information Systems*, 27, 743-776.
- Behkamal, B., Kahani, M., & Akbari, M. K. (2009). Customizing ISO 9126 quality model for evaluation of B2B applications. *Information and software technology*, 51(3), 599-609.
- Bose, R. J. C., & van der Aalst, W. (2010). Trace alignment in process mining: opportunities for process diagnostics. In *Business Process Management* (pp. 227-242). Springer Berlin Heidelberg.
- Brown, J. R., Kaspar, H., Lipow, M., MacLeod, G. J., & Merrit, M. J. (1978). *Characteristics of software quality* (Vol. 1). North-Holland Publishing Company.
- Carmona, J., Cortadella, J., & Kishinevsky, M. (2008). A region-based algorithm for discovering Petri nets from event logs. In *Business Process Management* (pp. 358-373). Springer Berlin Heidelberg.
- Cavano, J. P., & McCall, J. A. (1978, January). A framework for the measurement of software quality. In *ACM SIGMETRICS Performance Evaluation Review* (Vol. 7, No. 3-4, pp. 133-139). ACM.
- Chua, B. B., & Dyson, L. E. (2004, December). Applying the ISO 9126 model to the evaluation of an elearning system. In *Proc. of ASCILITE* (pp. 5-8).
- Claes, J., & Poels, G. (2013, January). Process mining and the ProM framework: an exploratory survey. In *Business Process Management Workshops* (pp. 187-198). Springer Berlin Heidelberg.
- Codd, E. F., Codd, S. B., & Salley, C. T. (1993). Providing OLAP (on-line analytical processing) to user-analysts: An IT mandate. *Codd and Date*, 32.
- Doebeli, G., Fisher, R., Gapp, R., & Sanzogni, L. (2011). Using BPM governance to align systems and practice. *Business Process Management Journal*, 17(2), 184-202.
- Dromey, R. G. (1995). A model for software product quality. *Software Engineering, IEEE Transactions on*, 21(2), 146-162.

Eckerson, W. W. (2010). *Performance dashboards: measuring, monitoring, and managing your business*. Wiley. com.

Fluxicon Process Laboratories, [Download]: Disco .Versión 1.5. Eindhoven, Inc. Programa computacional.

Grady, R. B. (1994). Successfully applying software metrics. *Computer*, 27(9), 18-25..

Günther, C. W., & Van Der Aalst, W. M. (2007). Fuzzy mining–adaptive process simplification based on multi-perspective metrics. In *Business Process Management* (pp. 328-343). Springer Berlin Heidelberg.

Hammer, M., & Stanton, S. (1999). How process enterprises really work. *Harvard business review*, 77, 108-120.

İsmail Bayraktar (2011). The Business Value of Process Mining Bringing It All Together, Eindhoven university of technology, Eindhoven University of Technology, Eindhoven.

International Organization for Standardization (2001). ISO 9126: Software Engineering – Product quality, Geneva, Switzerland.

Joel Tiago Soares Ribeiro (2013), Multidimensional process discovery (), Eindhoven University of Technology, Eindhoven.

Khosravi, K., & Guéhéneuc, Y. G. (2004). On issues with software quality models. In *The Proceedings of the 11th Working Conference on Reverse Engineering. GEODES* (pp. 172-181).

Mamaliga, T. (2013). *Realizing a Process Cube Allowing for the Comparison of Event Data* (Doctoral dissertation, Master's thesis, Eindhoven University of Technology, Eindhoven).

Mathiesen, P., Bandara, W., Delavari, H., Harmon, P., & Brennan, K. (2011). A comparative analysis of business analysis (BA) and business process management (BPM) capabilities. *ECIS 2011 Proceedings*..

Nielsen, J. (1994). *Usability engineering*. Access Online via Elsevier.

Paul Newbold, B. T. (2008). *Estadísticas para la administración y economía*, VI edición. Editorial Pearson.

Process Mining Group, Eindhoven Technical University, [Download]: Prom .Versión 6.3. Eindhoven, Inc. Programa computacional.

- Real Academia Española. (2001). Diccionario de la lengua española [Dictionary of the Spanish Language] (22nd ed.). Madrid, Spain: Author.
- Stefani, A., & Xenos, M. (2008). E-commerce system quality assessment using a model based on ISO 9126 and Belief Networks. *Software Quality Journal*, 16(1), 107-129.
- Van Der Aalst, W. M., Reijers, H. A., & Song, M. (2005). Discovering social networks from event logs. *Computer Supported Cooperative Work (CSCW)*, 14(6), 549-593.
- Van der Aalst, W., Adriansyah, A., de Medeiros, A. K. A., Arcieri, F., Baier, T., Blickle, T., ... & Pontieri, L. (2012, January). Process mining manifesto. In *Business process management workshops* (pp. 169-194). Springer Berlin Heidelberg.
- Van der Aalst, W., Weijters, T., & Maruster, L. (2004). Workflow mining: Discovering process models from event logs. *Knowledge and Data Engineering, IEEE Transactions on*, 16(9), 1128-1142.
- Weijters, A. J. M. M., van der Aalst, W. M., & De Medeiros, A. A. (2006). Process mining with the heuristics miner-algorithm. *Technische Universiteit Eindhoven, Tech. Rep. WP*, 166.

ANEXOS

ANEXO A: FUNCIONALIDADES OLAP-DISCOVERY

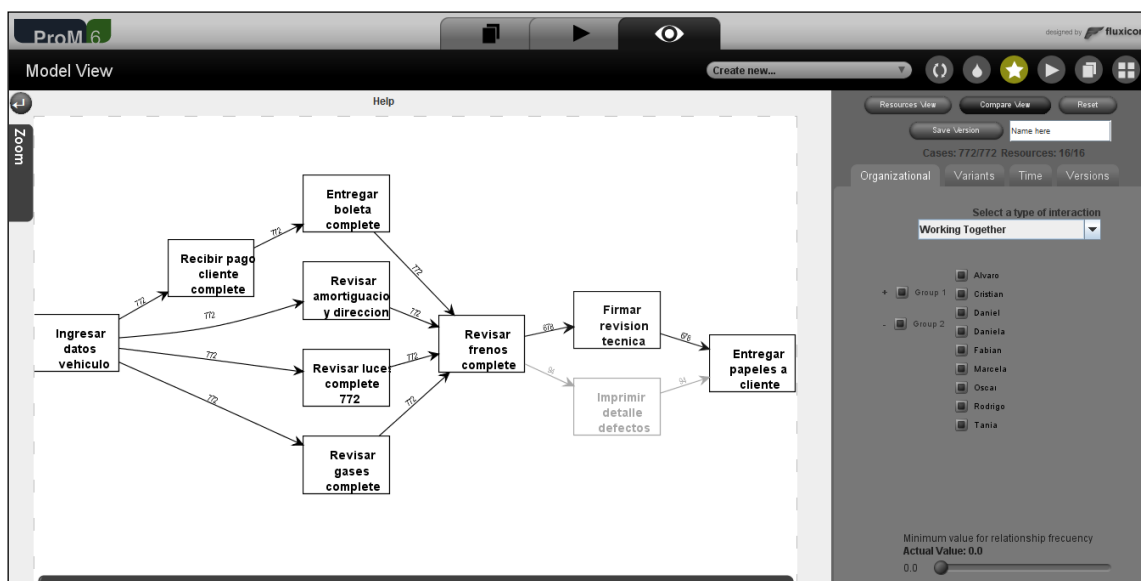


Figura 16 - Vista Completa OLAPDiscovery

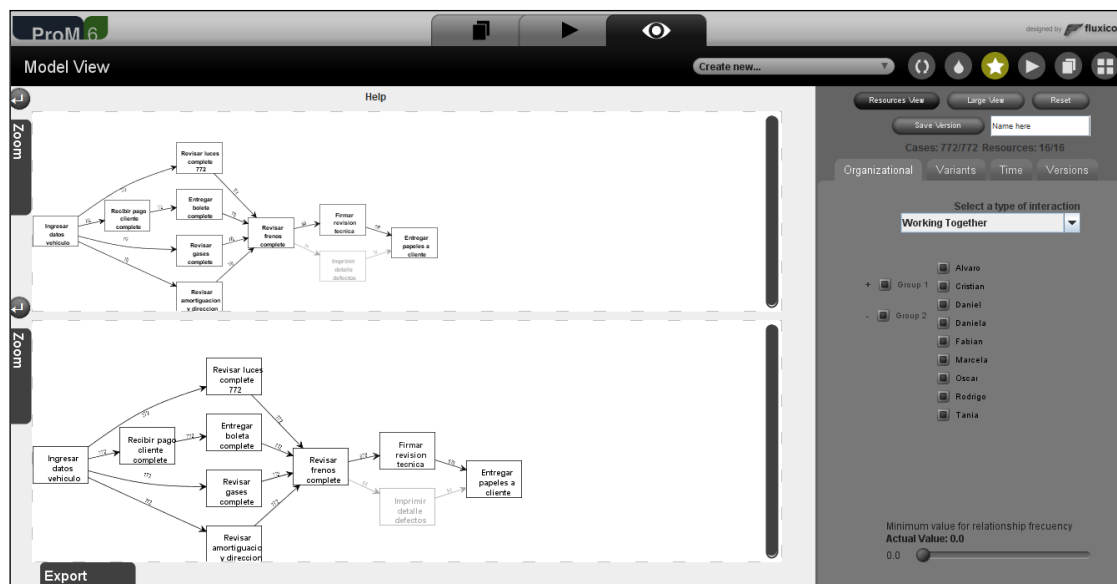


Figura 17 - Vista de recursos OLAPDiscovery

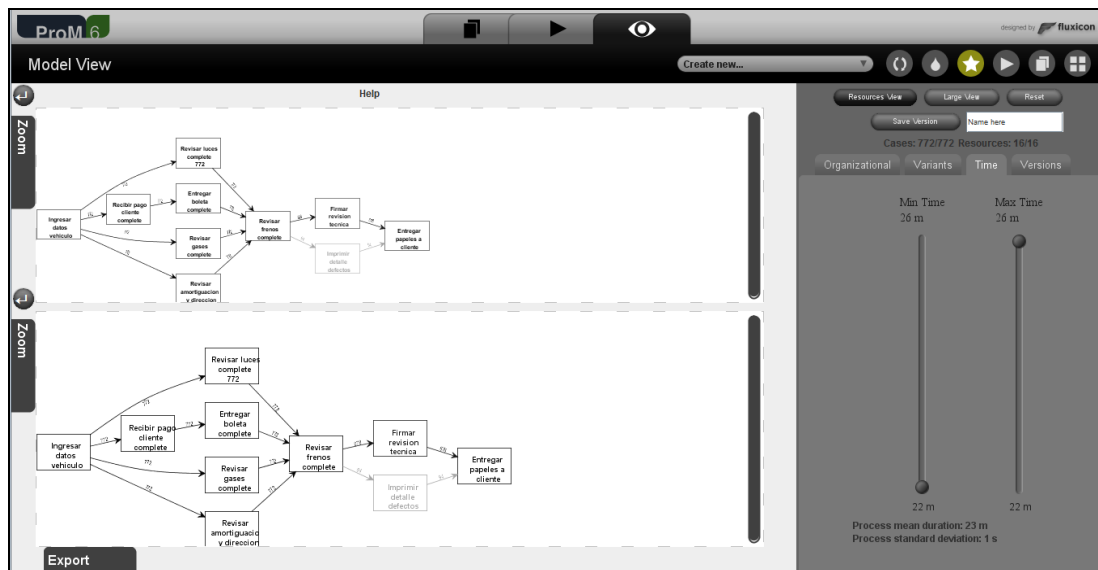


Figura 18 - Vista de tiempo OLAPDiscovery

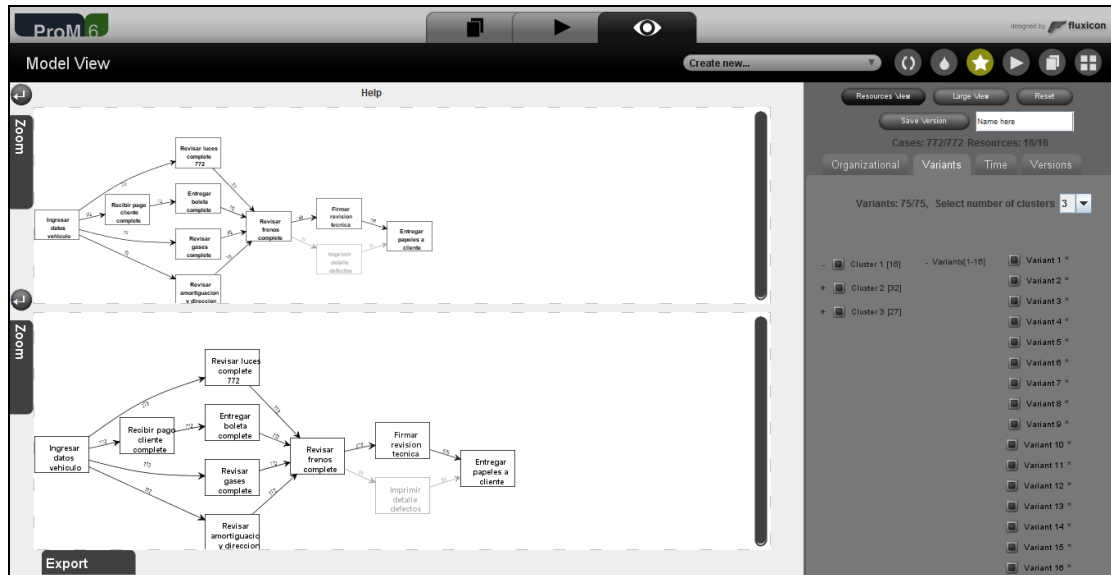


Figura 19 - Vista de Variantes

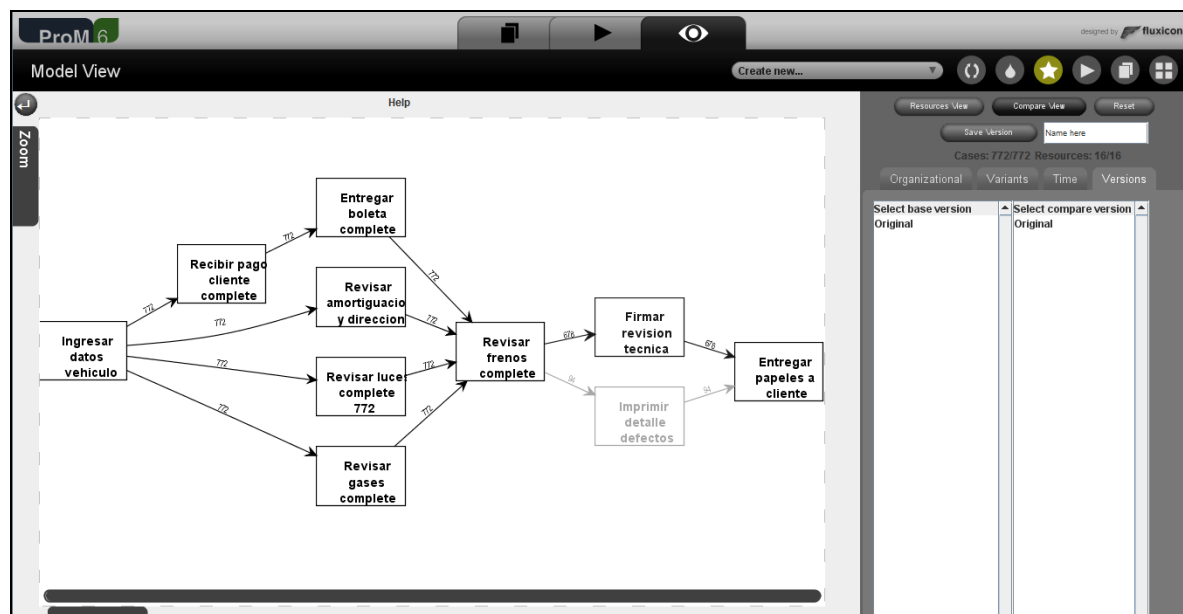


Figura 20 - Vista de Versiones OLAPDiscovery

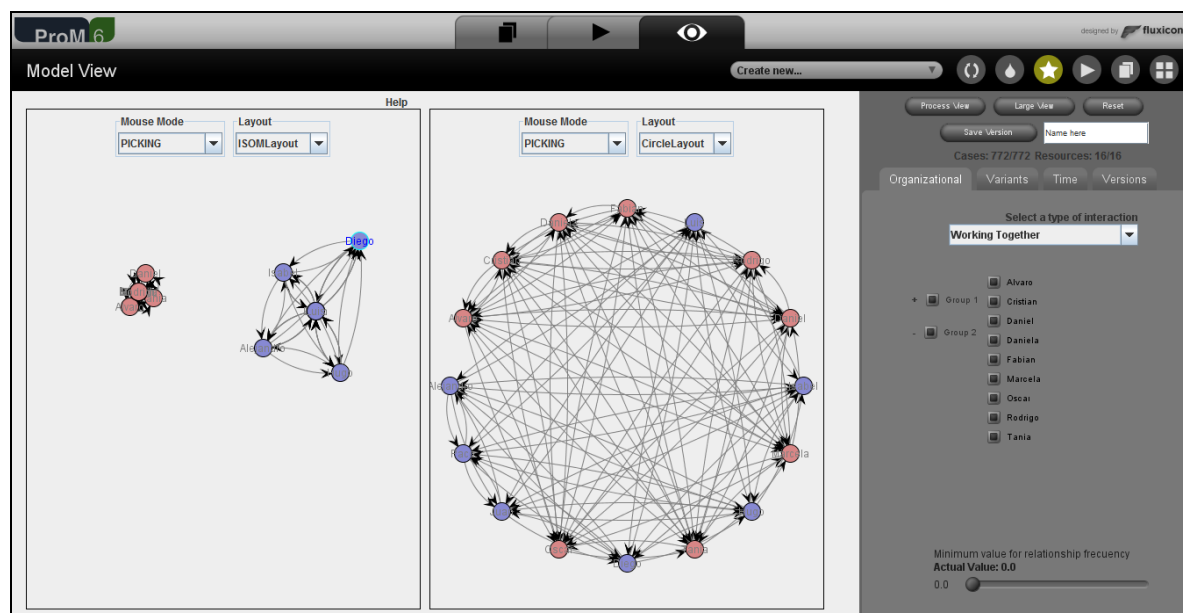


Figura 21 - Vista de flujo de recursos OLAPDiscovery

ANEXO B : FUNCIONALIDADES PROM

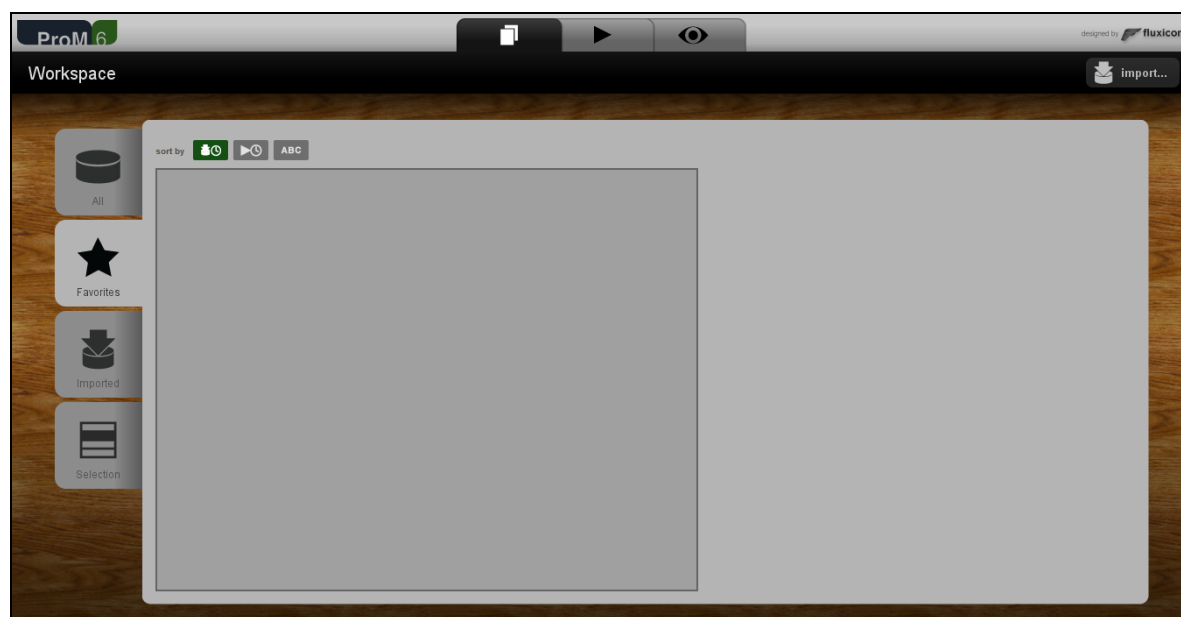


Figura 22 - Vista inicial ProM

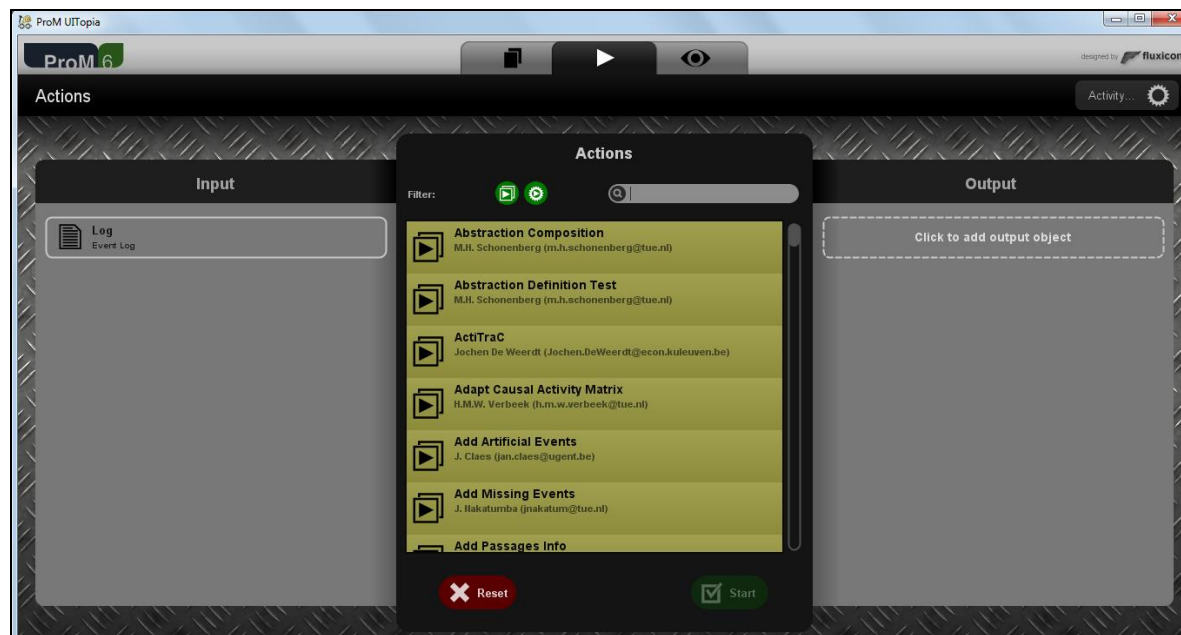


Figura 23 - Vista de carga de plug-ins

ANEXO C : FUNCIONALIDADES DISCO

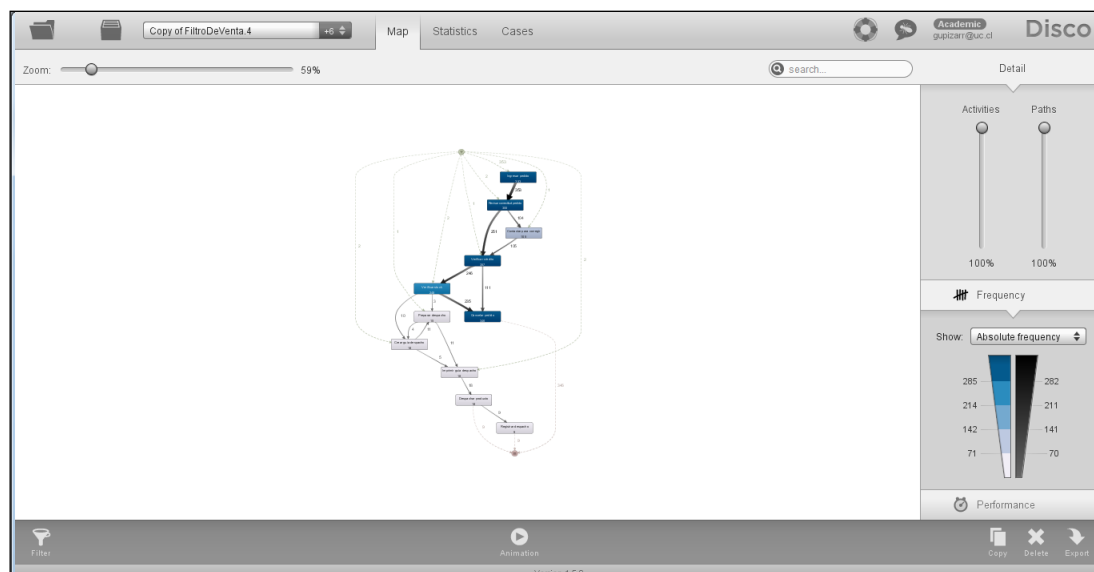


Figura 24 - Funcionalidad de vista de flujo Disco

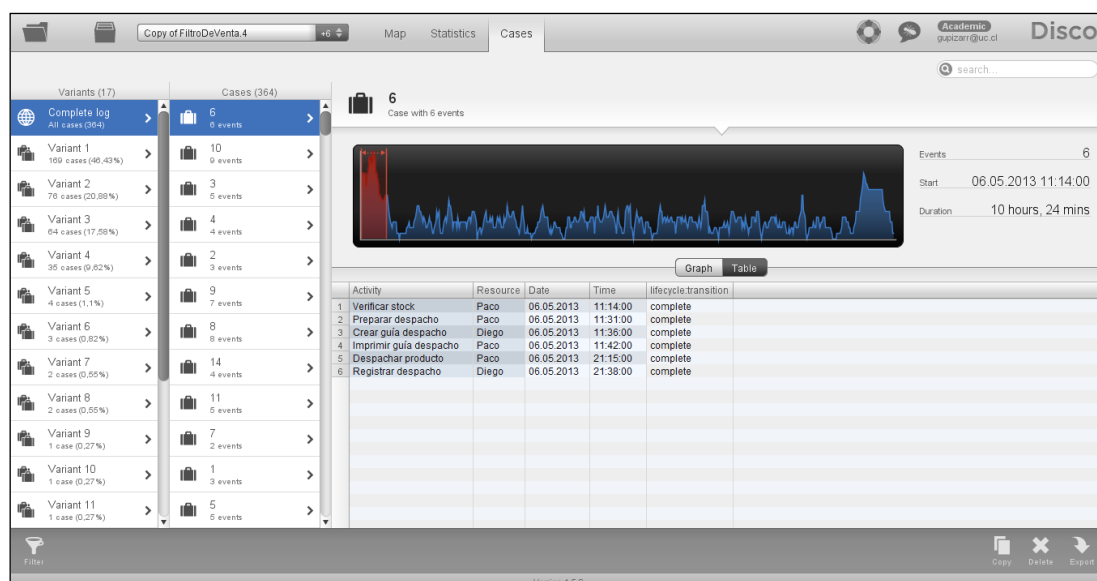


Figura 25 - Funcionalidad de casos Disco

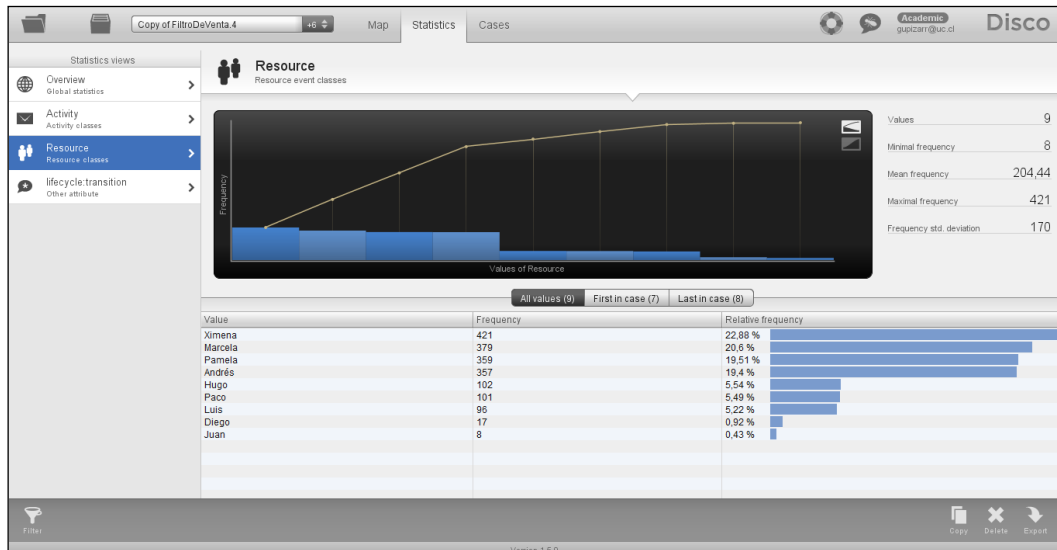


Figura 26 - Funcionalidad de estadísticas Disco

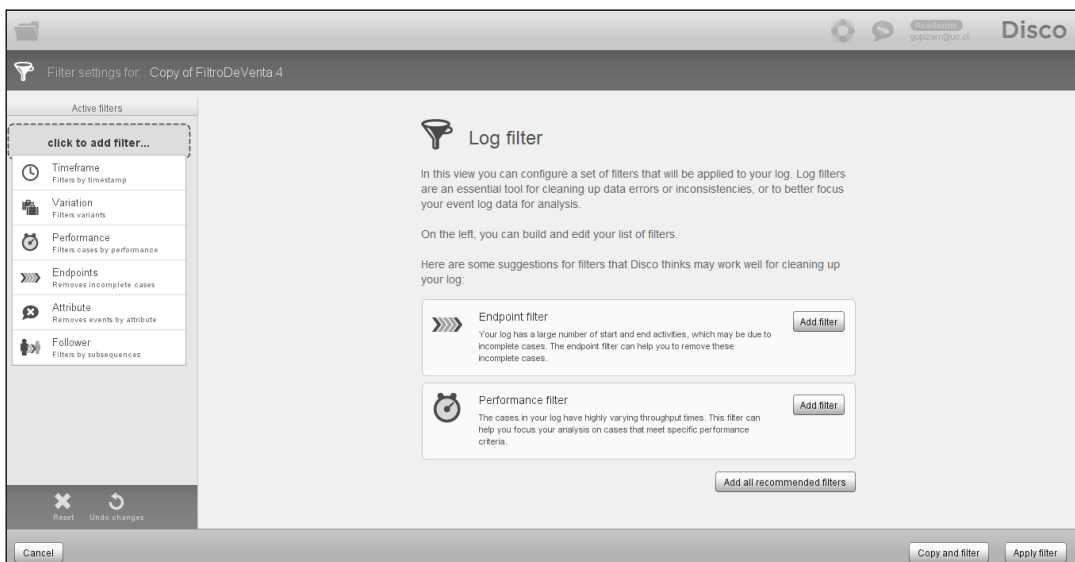


Figura 27 - Funcionalidad de filtro Disco

ANEXO D: MEDIDAS DE DISPERSIÓN

a) Asimetría¹

Esta medida nos permite identificar si los datos se distribuyen de forma uniforme alrededor del punto central (Media aritmética). La asimetría presenta tres estados diferentes que se pueden ver en la Figura 27, cada uno de los cuales define de forma concisa como están distribuidos los datos respecto al eje de asimetría. Se dice que la *asimetría es positiva* cuando la mayoría de los datos se encuentran por encima del valor de la media aritmética, la curva es *Simétrica* cuando se distribuyen aproximadamente la misma cantidad de valores en ambos lados de la media y se conoce como *asimetría negativa* cuando la mayor cantidad de datos se aglomeran en los valores menores que la media.

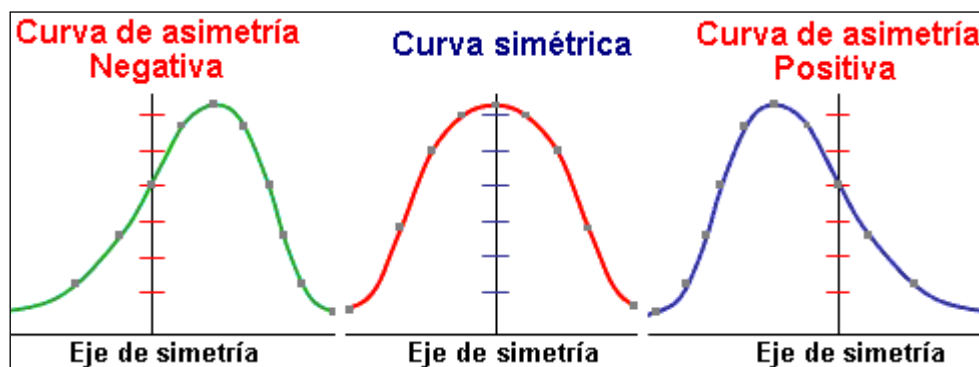


Figura 28 - Ejes de simetría

El *Coefficiente de asimetría*, se representa mediante la ecuación matemática,

¹ <http://www.spssfree.com/spss/analisis3.html>

$$g_1 = \frac{\frac{1}{n} \sum (X_i - \bar{X})^3 * n_i}{\left(\frac{1}{n} \sum (X_i - \bar{X})^2 * n_i \right)^{\frac{3}{2}}}$$

Figura 29 - Coeficiente de asimetría

Donde (g_1) representa el coeficiente de asimetría de Fisher, (X_i) cada uno de los valores, ($\bar{X}$) la media de la muestra y (n_i) la frecuencia de cada valor. Los resultados de esta ecuación se interpretan:

- ($g_1 = 0$): Se acepta que la distribución es Simétrica, es decir, existe aproximadamente la misma cantidad de valores a los dos lados de la media. Este valor es difícil de conseguir por lo que se tiende a tomar los valores que son cercanos ya sean positivos o negativos (± 0.5).
- ($g_1 > 0$): La curva es asimétricamente positiva por lo que los valores se tienden a reunir más en la parte izquierda que en la derecha de la media.
- ($g_1 < 0$): La curva es asimétricamente negativa por lo que los valores se tienden a reunir más en la parte derecha de la media.

Desde luego entre mayor sea el número (Positivo o Negativo), mayor será la distancia que separa la aglomeración de los valores con respecto a la media.

b) Kurtosis

Esta medida determina el grado de concentración que presentan los valores en la región central de la distribución. Por medio del *Coeficiente de Curtosis*, podemos identificar si existe una gran concentración de valores (*Leptocúrtica*), una concentración normal (*Mesocúrtica*) ó una baja concentración (*Platicúrtica*).

$$g_2 = \frac{\frac{1}{n} \sum (X_i - \bar{X})^4 * n_i}{\left(\frac{1}{n} \sum (X_i - \bar{X})^2 * n_i \right)^2} - 3$$

Figura 30 - Kurtosis

Donde (g_2) representa el coeficiente de Curtosis, (X_i) cada uno de los valores, ($\bar{X}$) la media de la muestra y (n_i) la frecuencia de cada valor. Los resultados de esta fórmula se interpretan:

- ($g_2 = 0$) *la distribución es Mesocúrtica*: Al igual que en la asimetría es bastante difícil encontrar un coeficiente de Curtosis de cero (0), por lo que se suelen aceptar los valores cercanos (± 0.5 aprox.).
- ($g_2 > 0$) *la distribución es Leptocúrtica*
- ($g_2 < 0$) *la distribución es Platicúrtica*

Cuando la distribución de los datos cuenta con un coeficiente de asimetría ($g_1 = \pm 0.5$) y un coeficiente de Curtosis de ($g_2 = \pm 0.5$), se le denomina Curva Normal. Este criterio es de suma importancia ya que para la mayoría de los procedimientos de la estadística de inferencia se requiere que los datos se distribuyan normalmente.

c) Coeficiente de variación

El coeficiente de variación de Pearson se define como la división de la desviación típica entre la media de la distribución. Indica el número de veces que la desviación típica está contenida en la media, y permite la comparación entre diferentes distribuciones.

La varianza y la desviación típica son medidas de carácter absoluto, en el sentido de que van referidas en las mismas unidades de medida que los datos de la distribución. Esto trae consigo un problema: no se pueden comparar dos distribuciones con medidas diferentes.

ANEXO E: ESTADÍSTICO DE PRUEBA

La distribución del estadístico de Mann-Whitney² se utiliza como método no paramétrico para cuando no se puede garantizar distribución normal. También sirve para hacer test de hipótesis a datos ordinales. Un requisito es aplicarlo cuando se toman muestras aleatorias independientes de las dos poblaciones.

Para contrastar la hipótesis nula de que la posición central de las dos distribuciones poblacionales es igual, suponemos que, aparte de la existencia de cualquier posible diferencia entre las posiciones centrales, las dos distribuciones poblacionales son idénticas.

$$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

Se supone que las dos distribuciones poblacionales son idénticas, aparte de las diferencias que puedan existir entre sus posiciones centrales. Para contrastar la hipótesis nula de que las dos distribuciones poblacionales tienen la misma posición central, las reglas de decisión para un nivel de significación dado en el caso particular de esta investigación es:

Si la hipótesis alternativa es la hipótesis bilateral, la regla de decisión es:

Rechazar la hipótesis nula si

$$\frac{U - \mu_U}{\sigma_U} < -z_{\alpha/2} \quad \text{o} \quad \frac{U - \mu_U}{\sigma_U} > z_{\alpha/2}$$

² Paul Newbold, B. T. (2008). *Estadísticas para la administración y economía*, VI edición. Editorial Pearson.

ANEXO F: RESULTADOS ESTADÍSTICOS DE PRUEBA

Tabla 7 - Estadísticos de prueba para cada afirmación entre herramientas

Afirmación	Dimensión	OLAP-D vs Disco		OLAP-D vs ProM	
		Sig. Asintótica (bilateral)	U de Mann-Whitney	Sig. Asintótica (bilateral)	U de Mann-Whitney
Es fácil recordar cómo hacer las cosas	Aprendizaje	0,050	555,000	0,000	250,000
La herramienta es sencilla de usar	Aprendizaje	0,390	647,500	0,000	124,00
Las funcionalidades de esta herramienta son fáciles de entender	Es comprendido	0,348	639,000	0,000	157,000
Puedo entender la información	Es comprendido	0,453	658,000	0,000	316,000
Es fácil actuar en base a la información proporcionada.	Es operativo	0,022	522,500	0,000	272,000
Es fácil combinar un análisis de todas las dimensiones de un proceso	Es operativo	0,001	409,000	0,000	174,000
Es fácil pasar de una parte de una funcionalidad a otra.	Es operativo	0,013	502,000	0,000	273,500
Los parámetros disponibles son adecuados.	Idoneidad	0,700	689,000	0,007	481,500
La herramienta es útil para el análisis de mi procesos	Idoneidad	0,710	689,000	0,006	470,500
La organización de los menús es bastante lógica.	Idoneidad	0,851	705,000	0,000	354,000
Los pasos necesarios para hacer una tarea de análisis son los adecuados	Productividad	0,315	636,000	0,000	225,500
El tiempo utilizado para hacer el análisis es razonable	Productividad	0,017	511,000	0,000	110,000
Recomendaría este programa a mis colegas.	Satisfacción	0,323	635,500	0,000	217,500
Volvería a usar este programa para analizar procesos	Satisfacción	0,286	630,500	0,000	268,500
Cree que puede entender un proceso completo. *Completo: acabado, con todas las particularidades que lo distinguen	Efectividad	0,214	611,000	0,000	357,000
Cree que puedo entender con precisión un proceso. *Precisión: fijo, exacto, que se distingue con claridad	Efectividad	0,666	682,500	0,0008	476,500

Tabla 8 - Estadísticos de prueba por tipo de usuario entre OLAP-D y Disco

Afirmación	Dimensión	Usuario Experto		Usuario Básico	
		Sig. Asintótica (bilateral)	U de Mann-Whitney	Sig. Asintótica (bilateral)	U de Mann-Whitney
Es fácil recordar cómo hacer las cosas	Aprendizaje	0,016	75,500	0,487	214,500
La herramienta es sencilla de usar	Aprendizaje	0,079	88,000	0,379	208,000
Las funcionalidades de esta herramienta son fáciles de entender	Es comprendido	0,787	121,500	0,318	203,500
Puedo entender la información	Es comprendido	0,567	115,000	0,204	192,000
Es fácil actuar en base a la información proporcionada.	Es operativo	0,247	100,000	0,369	207,000
Es fácil combinar un análisis de todas las dimensiones de un proceso	Es operativo	0,001	44,500	0,619	223,000
Es fácil pasar de una parte de una funcionalidad a otra.	Es operativo	0,001	46,000	0,791	231,500
Los parámetros disponibles son adecuados.	Idoneidad	0,850	123,500	0,893	236,500
La herramienta es útil para el análisis de mi procesos	Idoneidad	0,772	121,000	0,467	214,000
La organización de los menús es bastante lógica.	Idoneidad	0,246	98,500	0,818	233,000
Los pasos necesarios para hacer una tarea de análisis son los adecuados	Productividad	0,004	61,500	0,166	187,500
El tiempo utilizado para hacer el análisis es razonable	Productividad	0,021	73,500	0,166	189,000
Recomendaría este programa a mis colegas.	Satisfacción	0,660	117,500	0,053	168,000
Volvería a usar este programa para analizar procesos	Satisfacción	0,254	101,500	0,119	178,500
Cree que puede entender un proceso completo. *Completo: acabado, con todas las particularidades que lo distinguen	Efectividad	0,111	89,000	0,711	227,500
Cree que puedo entender con precisión un proceso. *Precisión: fijo, exacto, que se distingue con claridad	Efectividad	0,364	106,000	0,641	224,500

Tabla 9 - Estadísticos de prueba por uso de herramienta previa entre OLAP-D y Disco

Afirmación	Dimensión	Ninguna herramienta previa		Uso previo ProM	
		Sig. Asintótica (bilateral)	U de Mann-Whitney	Sig. Asintótica (bilateral)	U de Mann-Whitney
Es fácil recordar cómo hacer las cosas	Aprendizaje	0,268	54,500	0,924	75,000
La herramienta es sencilla de usar	Aprendizaje	0,748	66,500	0,521	66,000
Las funcionalidades de esta herramienta son fáciles de entender	Es comprendido	0,515	61,000	0,786	72,000
Puedo entender la información	Es comprendido	0,361	58,500	0,456	64,000
Es fácil actuar en base a la información proporcionada.	Es operativo	0,041	42,000	0,953	75,500
Es fácil combinar un análisis de todas las dimensiones de un proceso	Es operativo	0,060	40,500	0,053	42,500
Es fácil pasar de una parte de una funcionalidad a otra.	Es operativo	0,129	47,000	0,305	59,000
Los parámetros disponibles son adecuados.	Idoneidad	0,348	56,500	0,536	66,500
La herramienta es útil para el análisis de mi procesos	Idoneidad	0,077	43,500	0,221	55,500
La organización de los menús es bastante lógica.	Idoneidad	0,128	46,500	0,204	55,000
Los pasos necesarios para hacer una tarea de análisis son los adecuados	Productividad	0,763	67,000	0,523	66,000
El tiempo utilizado para hacer el análisis es razonable	Productividad	0,427	59,000	0,012	34,000
Recomendaría este programa a mis colegas.	Satisfacción	0,512	61,000	0,440	64,500
Volvería a usar este programa para analizar procesos	Satisfacción	0,534	62,000	0,246	57,500
Cree que puede entender un proceso completo. *Completo: acabado, con todas las particularidades que lo distinguen	Efectividad	0,732	66,000	0,070	45,000
Cree que puedo entender con precisión un proceso. *Precisión: fijo, exacto, que se distingue con claridad	Efectividad	0,853	68,500	0,409	62,000

ANEXO G: PREFERENCIAS POR TIPO DE USUARIO

Tabla 10 - Preferencias Usuarios Básicos

Afirmación	Dimensión	Totalmente en desacuerdo		En desacuerdo		Ni de acuerdo, ni en desacuerdo		De acuerdo		Totalmente de acuerdo	
		Disco	OLAP-D	Disco	OLAP-D	Disco	OLAP-D	Disco	OLAP-D	Disco	OLAP-D
Es fácil recordar cómo hacer las cosas	Aprendizaje	0%	0%	0%	4,50%	18,20%	0%	45,50%	54,50%	36,30%	41,00%
La herramienta es sencilla de usar	Aprendizaje	0%	0%	0%	0%	18,20%	18,20%	45,50%	50,00%	36,40%	31,80%
Las funcionalidades de esta herramienta son fáciles de entender	Es comprendido	0%	0%	4,50%	0%	9,10%	27,30%	45,50%	50,00%	40,90%	22,70%
Puedo entender la información	Es comprendido	0%	0%	0%	0%	4,50%	13,70%	45,50%	54,50%	50,00%	31,80%
Es fácil actuar en base a la información proporcionada.	Es operativo	0%	0%	4,50%	0%	18,20%	0%	50,00%	54,50%	27,30%	45,50%
Es fácil combinar un análisis de todas las dimensiones de un proceso	Es operativo	4,50%	0%	18,20%	13,70%	22,70%	13,60%	40,90%	40,90%	13,70%	31,80%
Es fácil pasar de una parte de una funcionalidad a otra.	Es operativo	0%	0%	9,10%	4,50%	18,20%	9,10%	40,90%	59,10%	31,80%	27,30%
Los parámetros disponibles son adecuados.	Idoneidad	0%	0%	0%	0%	13,60%	18,20%	59,10%	59,10%	27,30%	22,70%
La herramienta es útil para el análisis de mi procesos	Idoneidad	0%	0%	0%	0%	27,30%	13,60%	31,80%	40,90%	40,90%	45,50%
La organización de los menús es bastante lógica.	Idoneidad	0%	0%	9,10%	13,70%	9,10%	13,60%	54,50%	54,50%	27,30%	18,20%
Los pasos necesarios para hacer una tarea de análisis son los adecuados	Productividad	0%	0%	0%	9,10%	18,20%	13,60%	50,00%	59,10%	31,80%	18,20%
El tiempo utilizado para hacer el análisis es razonable	Productividad	0%	0%	0%	0%	22,70%	13,60%	50,00%	40,90%	27,30%	45,50%
Recomendaría este programa a mis colegas.	Satisfacción	0%	0%	0%	0%	9,10%	18,20%	59,10%	27,30%	31,80%	54,50%
Volvería a usar este programa para analizar procesos	Satisfacción	0%	0%	0%	0%	13,60%	13,60%	59,10%	50,00%	27,30%	36,40%
Cree que puede entender un proceso completo. *Completo: acabado, con todas las particularidades que lo distinguen	Efectividad	4,50%	0%	4,50%	4,50%	22,70%	18,30%	40,90%	54,50%	27,40%	22,70%
Cree que puedo entender con precisión un proceso. *Precisión: fijo, exacto, que se distingue con claridad	Efectividad	4,50%	0%	4,50%	9,10%	31,80%	31,80%	27,40%	31,80%	31,80%	27,30%

Tabla 11 - Preferencias Usuarios Expertos

Afirmación	Dimensión	Totalmente en desacuerdo		En desacuerdo		Ni de acuerdo, ni en desacuerdo		De acuerdo		Totalmente de acuerdo	
		Disco	OLAP-D	Disco	OLAP-D	Disco	OLAP-D	Disco	OLAP-D	Disco	OLAP-D
Es fácil recordar cómo hacer las cosas	Aprendizaje	0%	0%	6,20%	6,20%	31,30%	0%	62,50%	75%	0%	18,80%
La herramienta es sencilla de usar	Aprendizaje	0%	0%	0%	0%	18,70%	6,30%	68,80%	56,20%	12,50%	37,50%
Las funcionalidades de esta herramienta son fáciles de entender	Es comprendido	0%	0%	0%	0%	31,20%	18,80%	43,80%	62,50%	25,00%	18,70%
Puedo entender la información	Es comprendido	0%	0%	0%	0%	12,50%	6,20%	62,50%	62,50%	25,00%	31,30%
Es fácil actuar en base a la información proporcionada.	Es operativo	0%	0%	0%	6,20%	31,20%	12,50%	56,30%	50%	12,50%	31,30%
Es fácil combinar un análisis de todas las dimensiones de un proceso	Es operativo	6,20%	0%	31,30%	6,20%	18,80%	0%	37,50%	43,80%	6,20%	50,00%
Es fácil pasar de una parte de una funcionalidad a otra.	Es operativo	0%	0%	6,20%	0%	31,30%	0%	50,00%	37,50%	12,50%	62,50%
Los parámetros disponibles son adecuados.	Idoneidad	0%	0%	0%	6,20%	18,80%	31,30%	75,00%	37,50%	6,20%	25,00%
La herramienta es útil para el análisis de mi procesos	Idoneidad	0%	0%	0%	0%	12,50%	12,50%	43,70%	50,00%	43,80%	37,50%
La organización de los menús es bastante lógica.	Idoneidad	0%	0%	18,80%	6,20%	25,00%	18,80%	31,20%	37,50%	25,00%	37,50%
Los pasos necesarios para hacer una tarea de análisis son los adecuados	Productividad	0%	0%	6,20%	0%	18,80%	6,20%	75,00%	50,00%	0%	43,80%
El tiempo utilizado para hacer el análisis es razonable	Productividad	0%	0%	0%	0%	12,50%	6,20%	56,30%	18,80%	31,20%	75,00%
Recomendaría este programa a mis colegas.	Satisfacción	0%	0%	0%	0%	6,20%	12,50%	50,00%	31,30%	43,80%	56,20%
Volvería a usar este programa para analizar procesos	Satisfacción	0%	0%	0%	0%	0%	6,20%	68,80%	37,50%	31,20%	56,30%
Cree que puede entender un proceso completo. *Completo: acabado, con todas las particularidades que lo distinguen	Efectividad	6,20%	0%	12,50%	0%	12,50%	12,50%	50,00%	50,00%	18,80%	37,50%
Cree que puedo entender con precisión un proceso. *Precisión: fijo, exacto, que se distingue con claridad	Efectividad	0%	0%	12,50%	0%	18,80%	25,00%	56,20%	50,00%	12,50%	25,00%