



Wil van der Aalst  
 Universidad Técnica de Eindhoven (Países Bajos)

<w.m.p.v.d.aalst@tue.nl>

© 2013 ACM, Inc. Este artículo fue publicado previamente en inglés con el título "Process Mining" en Communications of the ACM CACM, Volume 55, Issue 8, August 2012, páginas 76-83, doi>10.1145/2240236.2240257. Se publica con los permisos de reproducción correspondientes.

## 1. Espectro de la minería de procesos

La minería de procesos pretende descubrir, seguir y mejorar procesos reales mediante la extracción de conocimiento de los registros (logs) de eventos disponibles en los sistemas de información actuales [1][2].

Aunque los datos de eventos son omnipresentes, las organizaciones carecen de una buena comprensión de sus procesos reales. Las decisiones de dirección tienden a basarse en diagramas de Powerpoint, en políticas locales o en paneles de gestión en lugar de en un cuidadoso análisis de datos de sucesos. El conocimiento oculto en los registros de eventos no se puede convertir en información procesable. Los avances en minería de datos hacen posible encontrar valiosos patrones en grandes conjuntos de datos y apoyar decisiones complejas basadas en tales datos. Sin embargo, cuestiones clásicas de la minería de datos como la clasificación, el agrupamiento, la regresión, el aprendizaje de reglas de asociación, y la minería de secuencia/episodio *no* se centran en procesos.

Por lo tanto, los enfoques de la gestión de procesos de negocio (*Business Process Management*, BPM) tienden a recurrir a modelos hechos a mano. La investigación sobre la minería de procesos persigue cerrar la brecha que existe entre la minería de datos y BPM. Metafóricamente, la minería de procesos se puede ver como la obtención de radiografías a través de rayos-x para diagnosticar/predecir problemas y recomendar tratamientos.

Un importante vehículo para la minería de procesos es el increíble crecimiento de los datos de eventos [4][6]. Estos datos están en todas partes – en cada sector, en cada economía, en cada organización, y en cada hogar donde uno pueda encontrar sistemas que registren eventos. Por menos de 600 dólares cualquiera puede comprar una unidad de disco duro con capacidad para almacenar toda la música del mundo. [6]. Un reciente estudio publicado en Science muestra que el espacio de almacenamiento creció de 2,6 exabytes óptimamente comprimidos (2.6 x 10

# Minería de procesos: Obtenga una radiografía de sus procesos de negocio

**Traducción:** Josep Moya Pérez (Grupo de Trabajo de Lengua e Informática de ATI)

**Resumen:** Grandes avances recientes en la investigación de la minería de procesos hacen posible descubrir, analizar y mejorar procesos de negocio basados en datos de eventos. Las actividades realizadas por las personas, por las máquinas y el software dejan rastros en los así llamados event logs o registros de sucesos. Eventos como introducir un pedido de un cliente en un sistema SAP, registrarse para un vuelo, cambiar la dosis para un paciente y rechazar un permiso de obras tienen en común que todos ellos son registrados por sistemas de información. Durante la última década ha habido un espectacular crecimiento de datos. Además, el universo digital y el universo físico se están alineando cada vez más. Por lo tanto, los procesos de negocio se deberían gestionar, soportar y mejorar en base a datos de eventos en lugar de hacerlo con opiniones subjetivas y experiencias obsoletas. La aplicación de la minería de procesos en cientos de organizaciones ha mostrado que tanto gestores como usuarios tienden a sobreestimar su conocimiento de los procesos en los que están involucrados. Por consiguiente, los resultados de la minería de procesos pueden contemplarse como una radiografía que muestra lo que realmente sucede dentro de los procesos. Como las radiografías, pueden ser utilizadas para diagnosticar problemas y proponer el tratamiento adecuado. La importancia práctica de la misma y los interesantes retos científicos hacen de la minería de procesos una de las cuestiones de más actualidad en la gestión de procesos de negocio (Business Process Management, BPM). Este artículo proporciona una introducción a la minería de procesos explicando los conceptos principales y tratando de varias aplicaciones de esta tecnología emergente.

**Palabras clave:** Administración, dirección de proceso de negocio, inteligencia de negocio, medición, minería de datos, minería de procesos, rendimiento.

## Autor

**Wil van der Aalst** es profesor en la Universidad Técnica de Eindhoven y cuenta con un índice h de más de 90 puntos entre los científicos más citados en Europa. Se le conoce sobre todo por sus trabajos en modelos de *Workflow* y se le considera sobradamente como el "padrino" de la minería de procesos. Su web personal se encuentra en <www.vdaalst.com>.

<sup>12</sup> bytes) en 1986, a 295 exabytes comprimidos en 2007. En ese año, el 94 por ciento de toda la capacidad de almacenamiento de información sobre la Tierra era digital. El otro 6 por ciento residía en los libros, en revistas y en otros formatos no digitales. Esto es en marcado contraste con 1986, cuando sólo el 0,8 por ciento de toda la capacidad de almacenamiento de información era digital. Estas cifras ilustran el crecimiento exponencial de los datos.

La mayor adopción de tecnologías como la identificación por radiofrecuencia (RFID, *Radio Frequency Identification*), los servicios basados en la localización, computación en la nube (*cloud computing*) y redes de sensores, acelerarán todavía más el crecimiento de datos de eventos. Sin embargo, las organizaciones tienen problemas para usar eficazmente esas grandes cantidades de datos de eventos. De hecho, la mayoría de corporaciones todavía diagnostican problemas basándose en ficciones (diapositivas de Powerpoint, diagramas de Visio, etc) en lugar de en hechos

(datos de eventos). Esto es ilustrativo de la mala calidad de los modelos de procesos en la práctica; por ejemplo, más del 20% de los diagramas de procesos 604 en el modelo de referencia de SAP tienen errores clarísimos y su relación con los procesos de negocio actuales soportados por SAP es confusa [7]. Por lo tanto, es vital convertir ingentes cantidades de datos de eventos en conocimiento relevante y percepciones fiables. Aquí es donde puede ayudar la minería de procesos.

La creciente madurez de la minería de procesos queda ilustrada en el Manifiesto de Minería de Procesos (*Process Mining Manifesto*) [5], publicado recientemente por la *IEEE Task Force on Process Mining*. Este manifiesto está apoyado por 53 organizaciones y 77 expertos en minería de procesos han contribuido en él. La participación activa de usuarios finales, vendedores de herramientas, consultores, analistas e investigadores ilustra la importancia de la minería de procesos como puente entre el *data mining* minería de datos y el modelado de procesos de negocio.

“ La verificación de conformidad se puede usar para comprobar si la realidad, tal y como se grabó en el registro, es conforme con el modelo y viceversa ”

El punto de partida para la minería de procesos es un *registro (log) de eventos*. Cada suceso en este registro se refiere a una *actividad* (por ejemplo, un paso bien definido en algún proceso) y está relacionado con un *caso particular* (por ejemplo, una *instancia de proceso*). Los eventos pertenecientes a un caso están ordenados y se pueden ver como una ejecución de ese proceso. Los registros de eventos pueden almacenar información adicional sobre estos sucesos. De hecho, siempre que es posible, las técnicas de minería de procesos utilizan información extra como el *recurso* (por ejemplo, una persona o un dispositivo) que ejecuta o inicia la actividad, la *marca de tiempo* del evento, o *elementos de datos* registrados con el mismo (por ejemplo, el tamaño de una orden).

Los registros de eventos se pueden usar para realizar tres tipos de minería de procesos, tal y como se muestra en la **figura 1** [1]. El primer tipo de minería de procesos es el *descubrimiento*. Una técnica de descubrimiento toma un registro de sucesos y crea un modelo sin usar a priori ninguna información. El descubrimiento de procesos es la técnica más importante en la minería de procesos. En muchas organizaciones es sorprendente ver que las técnicas existentes son realmente capaces de descubrir auténticos procesos basados simplemente en ejemplos de comportamientos grabados en los registros de eventos.

El segundo tipo de minería de procesos es la *conformidad*. Aquí, un modelo de proceso existente es comparado con un registro de eventos del mismo proceso. La verificación de conformidad se puede usar para comprobar si la realidad, tal y como se grabó en el registro, es conforme con el modelo y viceversa.

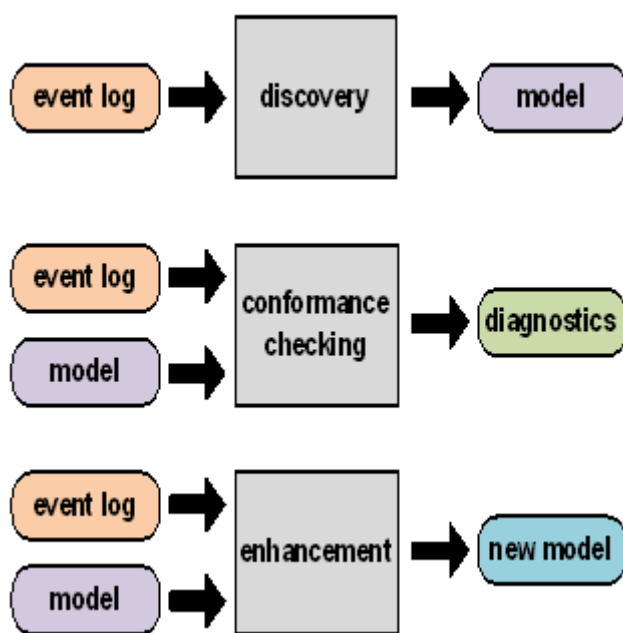
El tercer tipo es la *mejora*. Aquí la idea es ampliar o mejorar un modelo de proceso existente utilizando información sobre el proceso real registrado en el mismo fichero de eventos. Considerando que la comprobación de la conformidad mide el alineamiento entre modelo y realidad, este tercer tipo de minería de procesos pretende cambiar o extender el modelo a priori. Por ejemplo, usando marcas de tiempo en el registro de eventos se puede extender el modelo para que muestre cuellos de botella, niveles de servicio, tiempos de rendimiento y frecuencias.

**2. Descubrimiento de procesos**

Tal y como se muestra en la **figura 1**, el objetivo del descubrimiento de procesos es descubrir un modelo basado en un registro de eventos. Estos pueden tener toda clase de atributos (marcas de tiempo, información transaccional, uso de recursos, etc.). Todos se pueden usar en el descubrimiento de procesos. Sin embargo, por simplicidad, solemos representar los eventos sólo con nombres de actividades.

De esta manera, un caso (una instancia de proceso, por ejemplo) se puede representar por una traza que describe una secuencia de acciones. Considérese como ejemplo el registro de eventos mostrado en la **figura 1** (ejemplo tomado de [1]). Este registro contiene 1.391 casos, es decir, instancias de algún proceso de reembolso. Hay 455 instancias de proceso que siguen el rastro **acdeh**. Las actividades se representan con un único carácter: **a** = *petición de registro*, **b** = *examinar minuciosamente*, **c** = *examinar casualmente*, **d** = *comprobar resguardo*, **e** = *decidir*, **f** = *reiniciar solicitud*, **g** = *pago de indemnización* y **h** = *rechazar petición*. Por lo tanto, la traza **acdeh** modela una petición de reembolso que fue rechazada después de los pasos de registro, examen, comprobación y decisión. 455 casos siguen esta secuencia consistente en cinco etapas, es decir, la primera línea en la tabla corresponde a  $455 \times 5 = 2.275$  eventos. El registro entero comprende 7.539 sucesos.

Las técnicas de descubrimiento de procesos producen modelos de procesos basados en registros de eventos como el mostrado en la **figura 2**. Por ejemplo, el algoritmo clásico  $\alpha$  produce un modelo  $M_i$  para este registro. Este modelo se representa como una *red de Petri*. Una red de Petri consiste en *posiciones* y *transiciones*. El estado de una red de Petri, también conocido como *marcado*, se define por la distribución de *tokens* sobre posiciones. Una transición se activa si cada una de sus posiciones de entrada contiene un *token*. Por ejemplo,  $\alpha$  se *activa* en la marca inicial de  $M_i$  debido a que la única posición de entrada de  $\alpha$  contiene un *token* (un punto negro). La transición de **e** en  $M_i$  solamente se activa si ambas posiciones contienen un *token*. Una transición activada puede de este modo lanzar el consumo de un *token* desde cada una de sus posiciones de entrada y producir un *token* por cada una de sus posiciones de salida. Lanzar  $\alpha$  en la marca inicial se corresponde con eliminar un *token* del inicio y con producir dos *tokens* (uno por cada posición de salida). Después de lanzar  $\alpha$  se activan tres transiciones: **b**, **c** y **d**. El lanzamiento de **b** desactivará **c** porque el *token* está borrado de la posición de entrada compartida (y viceversa). La transición **d** es concurrente con **b** y **c**, es decir, se puede lanzar sin desactivar otra transición. La transición **e** se vuelve activa después de que **d** y **b** ó **c** hayan sucedido. Después de ejecutarse **e** tres transiciones se activan: **f**, **g** y **h**. Estas transiciones compiten por el mismo *token* modelando así una elección. Cuando **g** ó **h** se han lanzado, el proceso termina con un *token*



**Figura 1.** Los tres tipos básicos de minería de procesos explicados en términos de entrada y salida.

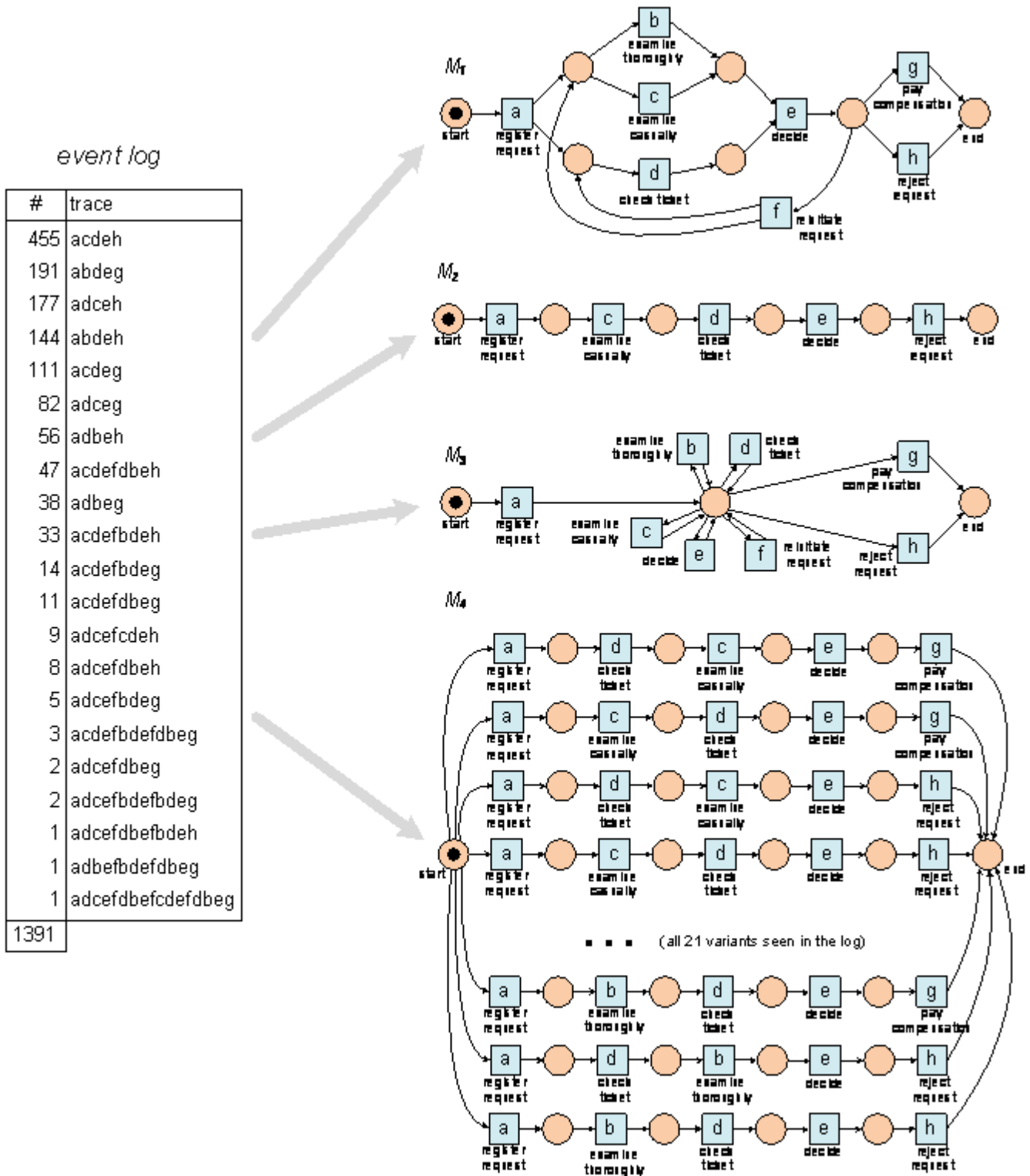


Figura 2. Un registro de eventos y cuatro modelos potenciales de procesos ( $M_1$ ,  $M_2$ ,  $M_3$  y  $M_4$ ) intentando describir el comportamiento observado.

en la posición *end*. Si **f** se lanza, el proceso regresa al estado justo después de ejecutar **a**.

Obsérvese que la transición **d** es concurrente con **b** y **c**. Las técnicas de minería de procesos necesitan poder descubrir patrones de proceso más avanzados y no se deberían restringir a simples procesos secuenciales.

Es fácil comprobar que todas las trazas en el registro de eventos pueden reproducirse por  $M_1$ . Esto no es válido para el segundo modelo de proceso de la figura 2.  $M_2$  sólo puede reproducir la traza más frecuente **acdch**. El modelo no *encaja* bien en el registro porque los rastros observados como **abdcg** no son posibles de acuerdo a  $M_2$ .

El tercer modelo puede reproducir el registro de eventos completo, pero  $M_2$  también lo permite para trazas como **ah** y **addddddg**. Por lo tanto, consideramos a  $M_2$  como poco adecuado; se permite demasiado funcionamiento porque  $M_2$  sobregeneraliza claramente el comportamiento observado. El modelo  $M_4$  puede también reproducir el registro de

Una red de Petri consiste en *posiciones* y *transiciones*. El estado de una red de Petri, también conocido como *marcado*, se define por la distribución de *tokens* sobre posiciones ””

$$\gamma_1 = \begin{array}{|c|c|c|c|c|} \hline a & d & c & e & g \\ \hline a & d & c & e & g \\ \hline \end{array} \quad \text{and} \quad \gamma_2 = \begin{array}{|c|c|c|c|c|} \hline a & b & \gg & e & f & d & \gg & e & g \\ \hline a & b & d & e & f & d & b & e & g \\ \hline \end{array} \quad \text{and} \quad \gamma_3 = \begin{array}{|c|c|c|c|c|} \hline a & b & e & f & d & e & g \\ \hline a & b & \gg & \gg & d & e & g \\ \hline \end{array}$$

Tabla 1. Alineaciones de ejemplo entre las trazas y el modelo  $M_1$ .

eventos. Sin embargo, el modelo únicamente codifica las trazas de ejemplo en el registro. Tal modelo lo denominamos "sobreadecuado" al no generalizar el comportamiento más allá de los ejemplos observados.

En años recientes se han desarrollado poderosas técnicas de minería que pueden construir automáticamente un modelo personalizado dado un registro de eventos. El objetivo de esas técnicas es crear un modelo simple que sea capaz de explicar la mayor parte de los comportamientos observados sin subvalorar o sobrevalorar el registro.

**3. Comprobación de conformidad**

La minería de procesos no se limita al descubrimiento de procesos. De hecho, el proceso descubierto es tan sólo el punto de partida de un análisis más profundo. Como se muestra en la figura 1, la comprobación de conformidad y la mejora relacionan modelo y registro. El modelo puede haber sido creado a mano o hallado por medio de descubrimiento de procesos. Para la comprobación de la conformidad, el comportamiento modelado y el comportamiento observado (es decir, el registro de eventos), se comparan. Cuando se comprueba la conformidad de  $M_2$  con respecto al registro mostrado en la figura 2 es fácil observar que únicamente los 455 casos que siguieron la traza **acdeh** pueden ser repetidos de principio a fin. Si intentamos repetir la traza **acdeg** nos quedaremos parados después de la ejecución de **acde** porque **g** no está activado. Si intentamos repetir el rastro **adceh** nos atoraremos después de ejecutarse el primer paso, porque **d** no está (todavía) activado.

Existen varios enfoques para diagnosticar y cuantificar la conformidad. Un enfoque es encontrar una alineación óptima entre cada traza en el registro y el comportamiento más parecido en el modelo. Considérese por ejemplo el modelo de proceso  $M_1$ , una traza adecuada  $\sigma_1 = \text{adceg}$  una traza no adecuada  $\sigma_2 = \text{abefdeg}$ , y las tres alineaciones mostradas en la tabla 1.

$\gamma_1$  muestra una alineación perfecta entre  $\sigma_1$

y  $M_1$ : todos los movimientos de la traza en el registro de eventos (parte superior de la alineación) pueden ser seguidos por movimientos del modelo (parte inferior de la alineación).  $\gamma_2$  muestra una alineación óptima para la traza  $\sigma_2$  en el registro de eventos, y el modelo  $M_1$ . Los dos primeros movimientos de la traza en el registro pueden ser seguidos por el modelo. Sin embargo, *e* no está activada después de ejecutar sólo *a* y *b*. En la tercera posición de la alineación  $\gamma_2$ , vemos un movimiento *d* del modelo que no está sincronizado con un movimiento en el registro de eventos.

Un movimiento sólo en el modelo se denota como ( $\gg$ , *d*). En los tres movimientos siguientes el modelo y el registro concuerdan. En la séptima posición del alineamiento  $\gamma_2$  sólo hay un paso del modelo y ninguno en el registro: ( $\gg$ , *b*).  $\gamma_3$  muestra otra alineación óptima para la traza  $\sigma_2$ . Aquí se dan dos situaciones donde el registro y el modelo no se mueven juntos: (*e*,  $\gg$ ) y (*f*,  $\gg$ ). Los alineamientos  $\gamma_2$  and  $\gamma_3$  son ambos óptimos si las penalizaciones por moverse en el registro de eventos y por moverse en el modelo son las mismas. En ambas alineaciones hay dos  $\gg$  pasos y no existen alineamientos con menos de dos  $\gg$  pasos.

La conformidad puede verse desde dos ángulos: (a) el modelo no captura el comportamiento real ("el modelo está equivocado") y (b) la realidad se desvía del modelo deseado ("el registro de sucesos es erróneo"). El primer punto de vista se toma cuando se supone que es *descriptivo*, es decir, que captura o predice la realidad. El segundo punto de vista se toma cuando el modelo es *normativo*, es decir, usado para influenciar o controlar la realidad.

Existen varios tipos de conformidad y el crear un alineamiento entre el registro y el modelo sólo es el punto de partida para la comproba-

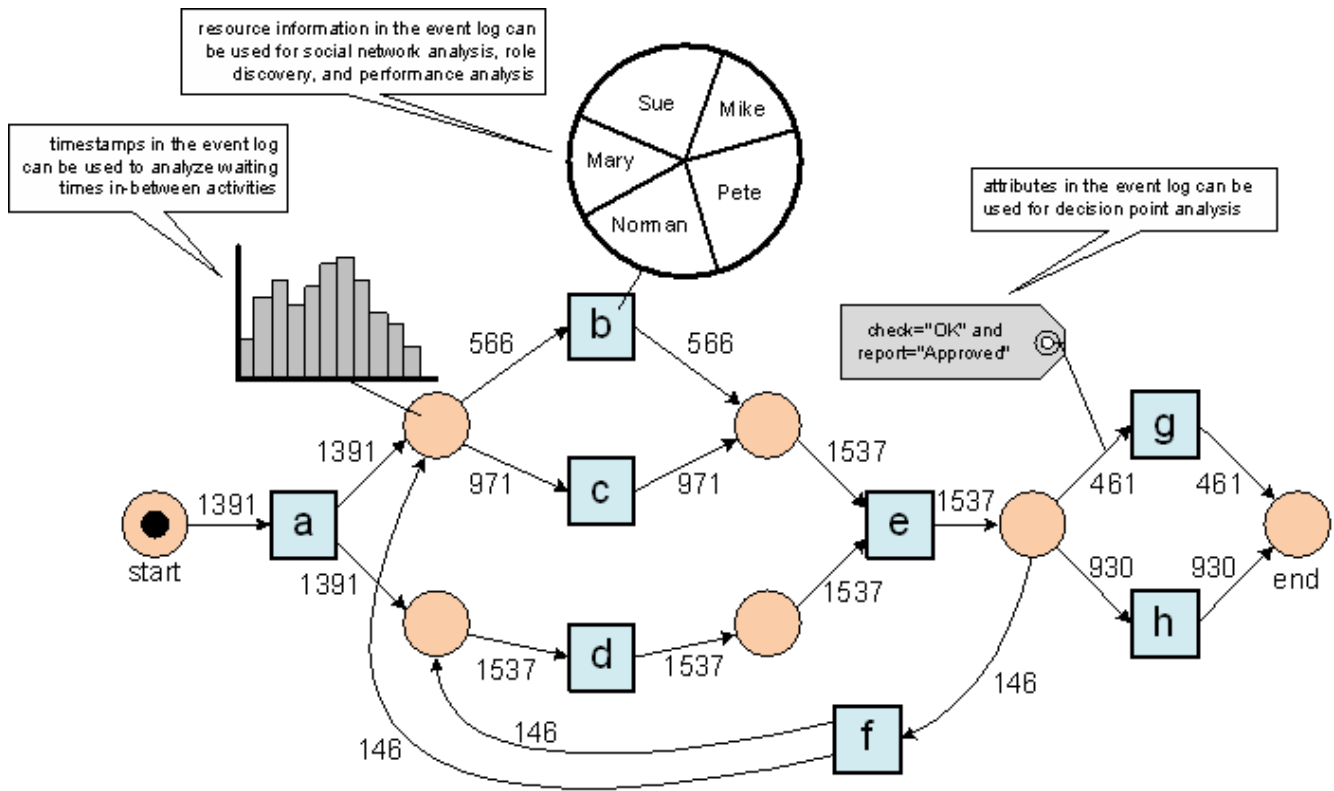
ción de la conformidad [1]. Por ejemplo, hay varias métricas de estado físico (la habilidad para repetir). Un modelo tiene el estado 1 si todas las trazas pueden repetirse de principio a fin. Otro modelo tiene estado 0 si dicho modelo y el registro no concuerdan en ningún evento. Los modelos de procesos  $M_1$ ,  $M_2$  y  $M_3$  tienen un estado de 1 (o sea, en perfecto estado) con respecto al registro de sucesos mostrado en la figura 2. El modelo  $M_2$  tiene un estado de 0,8 para el registro de eventos compuesto por 1.391 casos. Intuitivamente, esto significa que el 80% de los sucesos en el registro pueden ser explicados por el modelo. El estado físico es sólo una de varias métricas de conformidad.

Las experiencias con comprobaciones de conformidad en docenas de organizaciones revelan que los procesos en la vida real suelen desviarse de las representaciones simplificadas en Visio y PowerPoint que suelen usar los analistas de procesos.

**4. Mejora de modelo**

Es también posible extender o mejorar un modelo de proceso existente usando el alineamiento entre registro de eventos y modelo. Un modelo de proceso no adecuado se puede corregir utilizando los diagnósticos proporcionados por la alineación. Si el alineamiento contiene muchos (*e*,  $\gg$ ) pasos, entonces podría tener sentido el salto de la actividad *e* en el modelo. Además, los registros de eventos pueden contener información sobre recursos, marcas de fecha y datos del caso. Por ejemplo, un evento refiriéndose a la acción "petición de registro" y con el caso "992564" puede tener también atributos que describan a la persona que registró la solicitud (es decir, "John"), la fecha del evento (por ejemplo, "30-11-2011:14:55"), la edad del cliente (como "45"), y el importe reclamado (por ejemplo, "650 euros"). Después del alineamiento del modelo y del registro es posible repetir el registro de sucesos sobre el modelo. Mientras se repite se pueden analizar estos atributos adicionales.

Por ejemplo, como muestra la figura 3, es posible analizar tiempos de espera entre actividades. Midiendo simplemente la diferencia



**Figura 3.** El modelo de proceso se puede ampliar usando atributos de eventos como las marcas de fecha, información de recursos y datos del caso. El modelo muestra también frecuencias, como por ejemplo, una decisión que se tomó 1.537 veces y en 930 casos se rechazó.

de tiempo entre eventos causalmente relacionados y calculando estadísticas básicas tales como medias, varianzas e intervalos de confianza. De esta forma es posible identificar los principales cuellos de botella.

La información sobre recursos se puede utilizar para hallar roles, como por ejemplo grupos de personas ejecutando frecuentemente actividades relacionadas. Aquí se pueden usar las técnicas de agrupamiento convencionales. También es posible construir redes sociales basadas en el flujo de trabajo y analizar las prestaciones de los recursos (como la relación entre la carga de trabajo y los tiempos de servicio).

Las técnicas de clasificación estándar pueden usarse para analizar los puntos de decisión en el modelo de proceso. Por ejemplo, la actividad 'e' ("decidir") tiene tres posibles resultados: ("pagar", "rechazar" y "volver a hacer"). Utilizando los datos conocidos sobre el caso antes de la decisión, podemos construir un árbol de decisión explicando el comportamiento observado.

La **figura 3** muestra que la minería de procesos no se limita a hallar controles de flujo. Además, la minería de procesos no se restringe al análisis fuera de línea y se puede usar para predicciones y recomendaciones en tiempo de ejecución. Por ejemplo, el tiempo de terminación de un pedido de cliente parcial-

mente realizado puede ser predicho mediante un modelo de proceso hallado con información de tiempo.

**5. La minería de procesos crea valor de varias formas**

Después de introducirnos en las tres clases de minería de procesos utilizando un pequeño ejemplo, nos centraremos ahora en el valor práctico de esta disciplina. Como dijimos anteriormente, la minería de procesos se mueve por el crecimiento exponencial de los datos de eventos. Por ejemplo, de acuerdo con MGI, las empresas almacenaron más de 7 exabytes de nuevos datos en unidades de disco en el 2010 mientras que los consumidores guardaron más de 6 exabytes de nueva información en dispositivos como PC's y notebooks [6].

En lo que resta, mostraremos que la minería de procesos puede aportar valor de varias maneras. Para ilustrar esto nos referiremos a casos de estudio donde hemos usado nuestro paquete de software de código abierto ProM [1]. ProM ha sido creado y es mantenido por el grupo de minería de procesos de la Eindhoven University of Technology. No obstante, han contribuido a él grupos de investigación de todo el mundo, como la Universidad de Padua, la *Universitat Politècnica de Catalunya*, la Universidad de Calabria, la Universidad Humboldt de Berlín, la Queensland University of Technology, la Universidad Técnica de

Lisboa, la Universidad de Economía y Empresas de Viena, el Instituto Nacional de Ciencia y Tecnología de Ulsan, el K.U. Leuven, la Universidad de Tsinghua, y la Universidad de Innsbruck. Aparte de ProM existen alrededor de una decena de suministradores de software comercial que proporcionan software de minería de procesos (suelen estar embebidos en grandes herramientas), como Pallas Athena, Software AG, Futura Process Intelligence, Fluxicon, Businesscape, Iontas/Verint, Fujitsu, y Stereologic.

**5.1. Proporcionar percepciones**

Durante la última década hemos aplicado nuestro software de minería de procesos ProM en más de 100 organizaciones. Algunos ejemplos son municipios (sobre 20 en total, como Alkmaar, Heusden y Harderwijk), agencias del gobierno (*Rijkswaterstaat*, *Centraal Justitiele Incasso Bureau*, y el departamento de Justicia holandés), agencias relacionadas con los seguros (como UWV), bancos (como el ING Bank), hospitales (por ejemplo, el hospital AMC y el hospital Catharina), multinacionales (como por ejemplo DSM y Deloitte), fabricantes de sistemas de alta tecnología y sus clientes (Philips Healthcare, ASML, Ricoh, y Thales), y compañías de medios de comunicación (como Winkwaves). En cada una de estas corporaciones hemos encontrado algunos de sus procesos basados en los datos de eventos que proporcionaron. En cada proceso descubierto habían partes que

“ Es también posible extender o mejorar un modelo de proceso existente usando el alineamiento entre registro de eventos y modelo ”

sorprendieron a algunos de los interesados. La variabilidad de los procesos es típicamente mucho mayor de la esperada. Tales percepciones representan un tremendo valor debido a que las sorprendentes diferencias suelen apuntar a menudo al derroche y a la mala gestión.

**5.2. Mejorar el rendimiento**

Como hemos explicado anteriormente, es posible reproducir los registros de eventos en modelos de proceso encontrados o hechos manualmente. Esto se puede usar para la comprobación de conformidad y en la mejora de modelos. Puesto que la mayoría de registros de sucesos contienen marcas de fecha, la repetición se puede utilizar para ampliar el modelo con información del rendimiento.

La **figura 4** ilustra algunos de los diagnósticos relacionados con el rendimiento que se pueden obtener por medio de la minería de procesos. El modelo mostrado fue hallado

basándose en 745 objeciones contra la valoración denominada WOZ ("Waardering Onroerende Zaken") en una población holandesa. Los municipios holandeses necesitan estimar el valor de las casas y apartamentos. El valor WOZ se usa como base para determinar el impuesto sobre la propiedad inmobiliaria. Cuanto más alto sea el valor WOZ, más impuesto debe pagar el propietario. Por lo tanto, muchos ciudadanos apelan contra la valoración WOZ afirmando que es demasiado alto.

Cada una de las 745 objeciones corresponden a una instancia de proceso. Juntas, estas instancias generaron 9.583 eventos teniendo todos marcas de tiempo. La **figura 4** muestra la frecuencia de las diferentes trayectorias en el modelo. Por otra parte, las diferentes etapas del modelo están coloreadas para mostrar dónde, en promedio, se emplea más tiempo. Las etapas púrpura del proceso nece-

sitan más tiempo mientras que las etapas azules usan menos tiempo. También es posible seleccionar dos actividades y medir el tiempo que transcurre entre ellas.

Como se ve en la **figura 4**, pasan de media 202,73 días entre la terminación de la actividad "OZ02 Voorverieden" (preparación) y la conclusión de "OZ16 Uitspraak" (decisión final). Esto es más largo que el promedio de tiempo de flujo total que es aproximadamente de 178 días. Alrededor de 416 de estas objeciones (aproximadamente el 56%) siguen este camino; los demás casos siguen el ramal "OZ15 Zelf uitspraak", que por término medio usa menos tiempo.

Como se puede observar en la **figura 4**, los diagnósticos se pueden utilizar para mejorar los procesos eliminando los cuellos de botella y desviando los casos. Dado que el modelo está conectado a datos de eventos, es posible

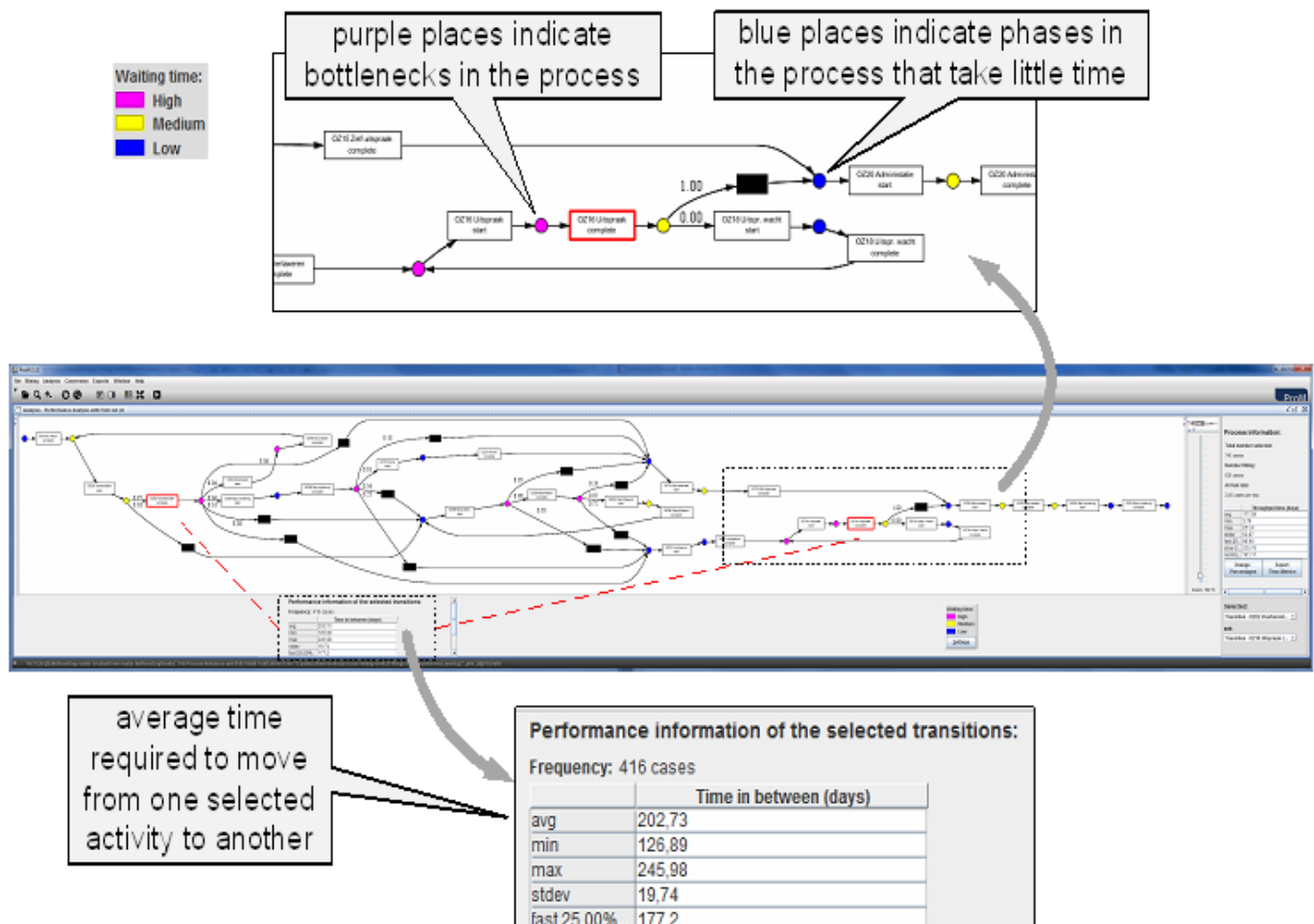


Figura 4. Análisis de rendimiento basado en 745 apelaciones contra la valoración WOZ.

“ A menudo, una "realidad PowerPoint" tiene poco en común con el proceso real, que tiene mucha más variabilidad. Sin embargo, para mejorar la conformidad y el rendimiento no debería dejar de abstraerse esta variabilidad ”

profundizar de inmediato e investigar grupos de casos que usan más tiempo que otros [1].

**5.3. Asegurar la conformidad**

La repetición también se puede utilizar para comprobar la conformidad tal y como se muestra en la **figura 5**. En base a las 475 apelaciones en contra de la valoración WOZ, hemos comparado también el modelo normativo y el comportamiento observado: 628 de los 475 casos se pueden repetir sin encontrar ningún problema. La adecuación del modelo y del registro es de 0,98876214 indicando que casi todos los eventos registrados se pueden explicar por el modelo. A pesar de su buena adecuación, ProM muestra claramente todas las desviaciones. Por ejemplo, "OZ12 Hertaxeren" (reevaluar propiedad) sucedió 23 veces cuando esto no estaba permitido según el modelo normativo (indicado por el "-23" en la **figura 5**). Es fácil de nuevo profundizar y ver qué tienen estos casos en común.

La conformidad del proceso de apelación que acabamos de describir es muy alta (alrededor del 99% de los eventos son posibles según el

modelo). También hemos encontrado muchos procesos con una conformidad muy baja; no es infrecuente hallar procesos donde sólo el 40% de los sucesos son posibles de acuerdo con el modelo. Por ejemplo, la minería de procesos reveló que el proceso de modelado de prueba ASML se desvió fuertemente del proceso real [9].

La creciente importancia del gobierno corporativo, riesgo y gestión de cumplimiento normativo y legislación como los de la Ley Sarbanes-Oxley (SOX), y el acuerdo Basel II, ilustran la importancia práctica de la comprobación de conformidad. La minería de procesos puede ayudar a los auditores a comprobar si los procesos se ejecutan dentro de ciertos límites establecidos por administradores, gobiernos y otras partes interesadas [3].

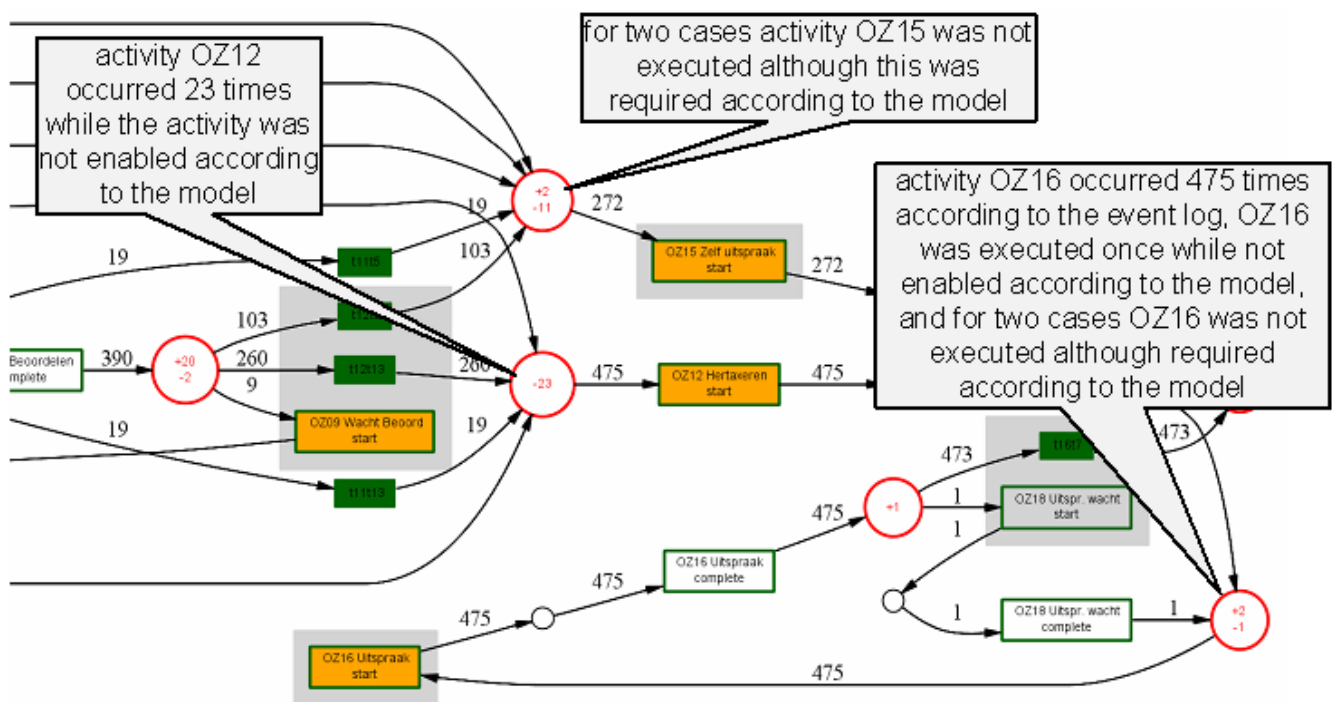
Las violaciones encontradas por medio de la minería pueden indicar fraudes, negligencias, riesgos e ineficiencias. Por ejemplo, en el municipio donde analizamos el proceso de apelación WOZ, hemos descubierto errores de configuración de su sistema de gestión de flujo de trabajo eIStream. Las personas tam-

bién evadieron el funcionamiento del sistema. Esto fue posible porque los administradores del sistema podrían haber cambiado manualmente el estado de los casos [8].

**5.4. Mostrar variabilidad**

Los modelos de proceso artesanales tienden a proporcionar una visión idealizada en el proceso de negocios que se modela. A menudo, una "realidad PowerPoint" tiene poco en común con el proceso real, que tiene mucha más variabilidad. Sin embargo, para mejorar la conformidad y el rendimiento no debería dejar de abstraerse esta variabilidad.

En el contexto de la minería de procesos solemos ver modelos tipo "spaghetti" como se puede ver en la **figura 6**. El modelo se encontró en base a un registro que contenía 24.331 eventos referidos a 376 actividades diferentes. El fichero de sucesos describe la diagnosis y el tratamiento de 627 pacientes de oncología ginecológica en el hospital AMC de Amsterdam. Las estructuras parecidas a *spaghettis* no están causadas por el algoritmo de descubrimiento, sino por la auténtica variabilidad del proceso.



**Figura 5.** Análisis de la conformidad mostrando desviaciones entre el registro de eventos y el modelo de proceso.



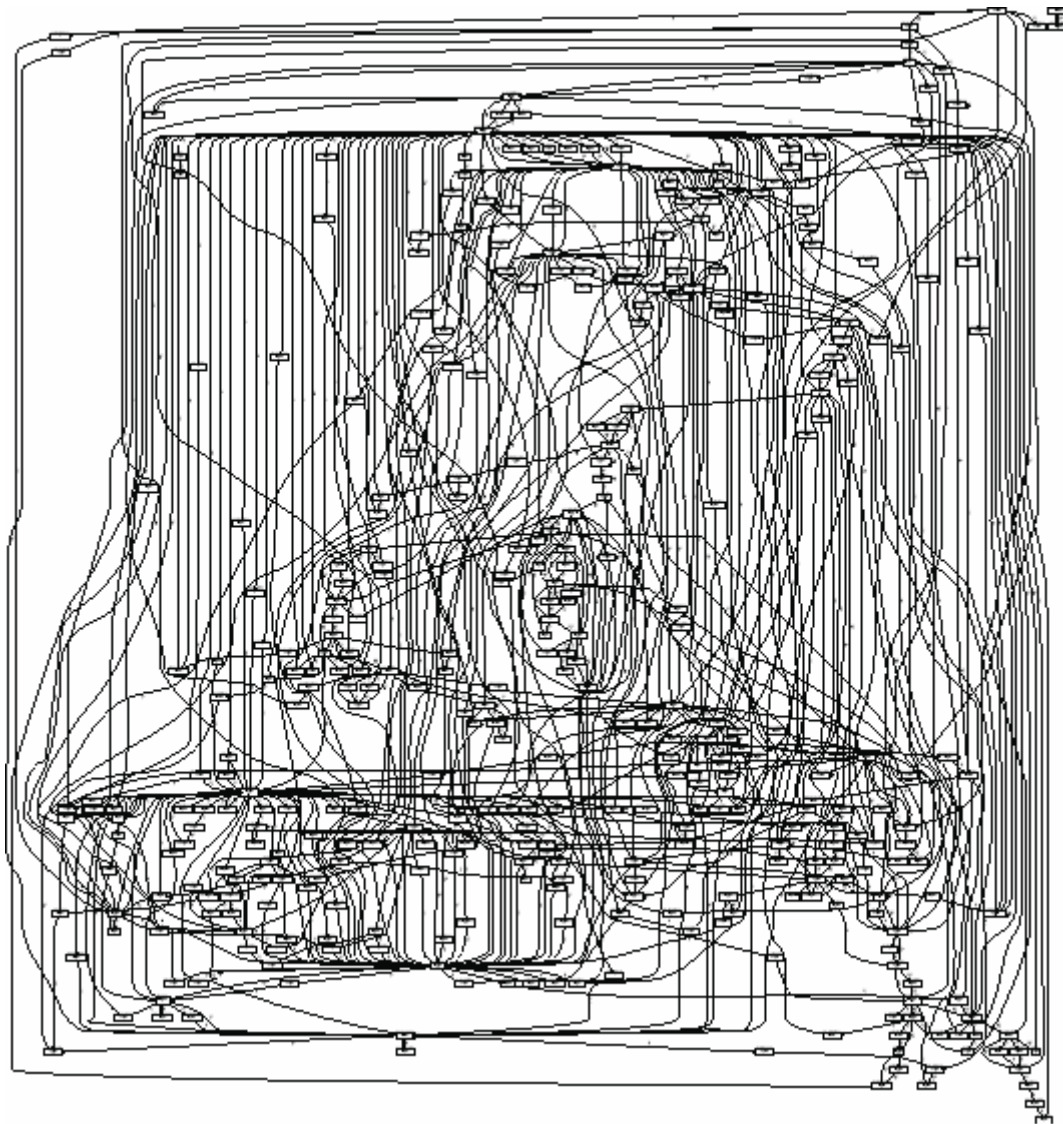


Figura 6. Modelo de proceso descubierto en un grupo de 627 pacientes de oncología ginecológica.

Aunque es importante enfrentar a las partes interesadas con la realidad, como se muestra en la figura 6, podemos también simplificar perfectamente estos modelos similares a *spaghetti*. Como cuando usamos mapas electrónicos, es también perfectamente posible ampliar y reducir [1]. Mientras reducimos, las cosas insignificantes se omiten o quedan dinámicamente agrupadas en capas añadidas, como las calles y los suburbios se amalgaman en ciudades en el Google Maps. El nivel de importancia de una actividad o conexión se puede basar en la frecuencia, en el costo, o en el tiempo.

### 5.5. Mejora de la fiabilidad

La minería de procesos se puede usar también para mejorar la fiabilidad de sistemas y procesos. Por ejemplo, desde 2007 hemos estado involucrados en un esfuerzo continuado para analizar los registros de eventos de los equipos de rayos-X de Philips Healthcare utilizando minería de procesos [1]. Estas máquinas registran cantidades

ingentes de eventos. En los equipamientos médicos es esencial probar que los sistemas han sido testeados bajo circunstancias realistas. Por lo tanto, el descubrimiento de procesos se usó para construir perfiles de prueba realistas. Philips Healthcare usó también minería de procesos para diagnosticar fallos. Aprendiendo de problemas anteriores, es posible hallar la causa de nuevos problemas que puedan surgir. Por ejemplo, al utilizar ProM hemos analizado bajo qué circunstancias se reemplazan ciertos componentes. Esto dio como resultado un conjunto de registros de comportamiento. Cuando un equipo de rayos-X con mal funcionamiento muestra el registro de un comportamiento particular, el ingeniero de mantenimiento sabe qué componente ha de cambiar.

### 5.6. Permitir predicciones

La combinación de datos históricos de eventos con otros datos en tiempo real puede ser útil para predecir problemas. Por ejemplo,

Philips Healthcare puede anticipar que un tubo de rayos-X en servicio está a punto de fallar por descubrirse patrones en el registro de eventos. Por lo tanto, se puede reemplazar el tubo antes de que la máquina empiece a fallar.

Hoy, muchas fuentes de datos se actualizan (casi) en tiempo real y existe suficiente potencia de cálculo para analizar eventos en cuanto suceden. Por lo tanto, la minería de procesos no se restringe tan sólo al análisis fuera de línea y se puede usar también para el soporte operacional *online*. Para una instancia de proceso en ejecución es posible hacer predicciones, como por ejemplo el tiempo de flujo esperado restante [1].

### 6. Conclusión

Las técnicas de minería de procesos permiten a las organizaciones radiografiar sus procesos de negocio, diagnosticar problemas y obtener sugerencias para solucionarlos. El descubrimiento de procesos proporciona a

menudo nuevas y sorprendentes ideas. Se pueden usar para rediseñar procesos o mejorar su gestión. La comprobación de conformidad puede utilizarse para observar dónde se desvían los procesos. Esto es muy importante, ya que las organizaciones necesitan poner más énfasis en el gobierno corporativo, en los riesgos y en el cumplimiento. Las técnicas de minería de procesos ofrecen un medio para comprobar más rigurosamente la conformidad mientras se mejora el rendimiento.

Este artículo ha introducido los conceptos básicos y mostrado que la minería de procesos puede aportar valor de varias maneras. Al lector interesado en la minería de procesos se le remite al primer libro sobre ésta [1], y al manifiesto ya mencionado [5], que está disponible en 12 idiomas. También le invitamos a visitar <[www.processmining.org](http://www.processmining.org)> para acceder a registros de eventos de ejemplo, videos, diapositivas, artículos y software.

El autor desea dar las gracias a los miembros del Grupo de Trabajo sobre Minería de Procesos del IEEE y a todos los que han contribuido al Manifiesto sobre Minería de Procesos y al *framework* ProM.

## Referencias

- [1] **W. van der Aalst.** *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Springer-Verlag, Berlin, 2011. ISBN 978-3-642-19345-3.
- [2] **W. van der Aalst.** Using Process Mining to Bridge the Gap between BI and BPM. *IEEE Computer* 44, 12, pp. 77–80, 2011.
- [3] **W. van der Aalst, K. van Hee, J.M. van Werf, M. Verdonk.** Auditing 2.0: Using Process Mining to Support Tomorrow's Auditor. *IEEE Computer* 43, 3, pp. 90–93, 2010.
- [4] **M. Hilbert, P. Lopez.** The World's Technological Capacity to Store, Communicate, and Compute Information. *Science* 332, 6025, pp. 60–65, 2011.
- [5] **TFPM Task Force on Process Mining.** Process Mining Manifesto. *Business Process Management Workshops*, F. Daniel, K. Barkaoui, and S. Dustdar, Eds. Lecture Notes in Business Information Processing Series, vol. 99. Springer-Verlag, Berlin, pp. 169–194, 2012.
- [6] **J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, A. Byers.** Big Data: The Next Frontier for Innovation, Competition, and Productivity. McKinsey Global Institute, 2011. <[http://www.mckinsey.com/insights/business\\_technology/big\\_data\\_the\\_next\\_frontier\\_for\\_innovation](http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation)>.
- [7] **J. Mendling, G. Neumann, W. van der Aalst.** Understanding the Occurrence of Errors in Process Models Based on Metrics. Proceedings of the OTM Conference on Cooperative Information Systems (CoopIS 2007). En F. Curbera, F. Leymann, and M. Weske, Eds. *Lecture Notes in Computer Science Series, vol. 4803*. Springer-Verlag, Berlin, pp. 113–130, 2007.
- [8] **A. Rozinat, W. van der Aalst.** Conformance Checking of Processes Based on Monitoring Real Behavior. *Information Systems* 33, 1, pp. 64–95, 2008.
- [8] **A. Rozinat, I. de Jong, C. Günther, W. van der Aalst.** Process Mining Applied to the Test Process of Wafer Scanners in ASML. *IEEE Transactions on Systems, Man and Cybernetics, Part C* 39, 4, pp. 474–479, 2009.



## ACTUALIZACIÓN DATOS SOCIO ATI

¿Has cambiado de domicilio, de empresa, y lo has comunicado a la Secretaría General?

¿Recibes el correo postal de la asociación?

¿Te llegan los correos electrónicos enviados por las Secretarías de ATI?

Si has contestado que NO a todas estas preguntas, te agradeceríamos que enviaras un mensaje a [secregen@ati.es](mailto:secregen@ati.es) con tus nuevos datos con el fin de tener actualizada tu ficha de socio y, de este modo, nos ayudes a mejorar la comunicación entre la asociación y sus miembros.

\* Del mismo modo, si sabes de algún compañero tuyo, miembro de ATI, que no recibirá esta información, te agradeceríamos que se la hagas llegar para que se pueda poner en contacto con nosotros.

---

ATI Secretaría General | Vía Laietana 46, ppal. 1a. | 08003 Barcelona | 93 412 52 33 | [secregen@ati.es](mailto:secregen@ati.es) | [www.ati.es](http://www.ati.es)