

Novática, revista fundada en 1975 y decana de la prensa informática española, es el órgano oficial de expresión y formación continua de **ATI** (Asociación de Técnicos de Informática), organización que edita también la revista **REICIS** (Revista Española de Innovación, Calidad e Ingeniería del Software).

<<http://www.ati.es/novatica/>>
<<http://www.ati.es/reicis/>>

ATI es miembro fundador de **CEPIS** (Council of European Professional Informatics Societies) y es representante de España en **IFIP** (International Federation for Information Processing); tiene un acuerdo de colaboración con **ACM** (Association for Computing Machinery), así como acuerdos de vinculación o colaboración con **AdaSpain**, **AI2**, **ASTIC**, **RITSI** e **HispaLinux**, junto a la que participa en **ProInnova**.

Consejo Editorial

Ignacio Aguiló Sousa, Guillem Alsina González, María José Escalona Cuarema, Rafael Fernández Calvo (presidente del Consejo), Jaime Fernández Martínez, Luis Fernández Sanz, Didac Lopez Viñas, Celestino Martín Alonso, José Onofre Montes Andrés, Francesc Noguera Puig, Ignacio Pérez Martínez, Andrés Pérez Payeras, Viktu Pons i Colomer, Juan Carlos Vigo López

Coordinación Editorial

Llorenç Pagés Casas <lpages@ati.es>

Composición y autoedición

Jorge Llácer Gil de Ramales

Traducciones

Grupo de Lengua e Informática de ATI <<http://www.ati.es/gt/lengua-informatica/>>

Administración

Tomás Brunete, María José Fernández, Enric Camarero

Secciones Técnicas - Coordinadores

Acceso y recuperación de la información

José María Gómez Hidalgo (Optenet), <jmgomez@yahoo.es>

Manuel J. María López (Universidad de Huelva), <manuel.maria@diehsia.uhu.es>

Administración Pública electrónica

Francisco López Crespo (MAE), <flc@ati.es>

Sebastià Justicia Pérez (Diputación de Barcelona), <sjusticia@ati.es>

Arquitecturas

Enrique F. Torres Moreno (Universidad de Zaragoza), <enrique.torres@unizar.es>

José Filich Cardo (Universidad Politécnica de Valencia), <jfilich@disca.upv.es>

Auditoría SITIC

Marina Tourino Irolifio, <marinatourino@marinatourino.com>

Sergio Gómez-Landero Pérez (Endesa), <sergio.gomezlandero@endesa.es>

Derecho y tecnologías

Isabel Hernández Collazos (Fac. Derecho de Donostia, UPV), <isabel.hernandez@ehu.es>

Elena Davara Fernández de Marcos (Davara & Davara), <edavara@davara.com>

Enseñanza Universitaria de la Informática

Cristóbal Parja Flores (DSIP-UCM), <cparja@sip.ucm.es>

J. Angel Velázquez Iturbide (DLSI, URJC), <angel.velazquez@urjc.es>

Entorno digital personal

Andrés Marín López (Univ. Carlos III), <amarin@it.uc3m.es>

Diego Gachet Pérez (Universidad Europea de Madrid), <dgachet@uem.es>

Estandares Web

Encarna Quesada Ruiz (Virati), <encarna.quesada@virati.com>

José Carlos del Arco Prieto (TCP, Sistemas e Ingeniería), <jcarco@gmail.com>

Gestión del Conocimiento

Juan Baiget Solé (Cap Gemini Ernst & Young), <juan.baiget@ati.es>

Gobierno Cooperativo de las TI

Manuel Palao García-Suñito (ATI), <manuel@palao.com>

Miguel García-Moneda (ITI), <mgarciamoneda@ititrendsinstitute.org>

Informática y Filosofía

José Ángel Olivás Varela (Escuela Superior de Informática, UCLM), <joseangel.olivas@uclm.es>

Roberto Feltrero Oreja (UNED), <rfeltrero@gmail.com>

Informática Gráfica

Miguel Chover Sellés (Universitat Jaume I de Castellón), <mchover@lsi.uji.es>

Roberto Vivó Hernando (Eurographics, sección española), <rvido@dsic.upv.es>

Ingeniería del Software

Javier Dolado Cosin (DLSI-UPV), <adolado@lsi.ehu.es>

Daniel Rodríguez García (Universidad de Alcalá), <daniel.rodriguez@uah.es>

Inteligencia Artificial

Vicente Boti Navarro, Vicente Julián Inglada (DSIC-UPV), <[vbotti.vinglada@dsic.upv.es](mailto:(vbotti.vinglada)@dsic.upv.es)>

Interacción Persona-Computador

Pedro M. Latorre Andrés (Universidad de Zaragoza, AIPO), <platorre@unizar.es>

Francisco L. Gutierrez Vela (Universidad de Granada, AIPO), <fgutierrez@ugr.es>

Lenguaje e Informática

M. del Carmen Ugarte García (ATI), <cugarte@ati.es>

Lenguajes Informáticos

Oscar Belmonte Fernández (Univ. Jaime I de Castellón), <obelmonte@lsi.uji.es>

Inmaculada Coma Taty (Univ. de Valencia), <inmaculada.coma@uv.es>

Lingüística computacional

Xavier Gómez Guzmán (Univ. de Vigo), <xgg@uvigo.es>

Manuel Palomar (Univ. de Alicante), <mpalomar@lsi.ua.es>

Mundo estudiantil y jóvenes profesionales

Federico G. Mon Trotti (RITSI), <gmon.trotti@gmail.com>

Mikel Sáizar Peña (Asoc. de Jóvenes Profesionales, Junta de ATI Madrid), <mikelbo_uni@yahoo.es>

Profesión Informática

Rafael Fernández Calvo (ATI), <rfcalvo@ati.es>

Miguel Sarrías Gilardi (ATI), <miguel@sarrias.net>

Redes y servicios telemáticos

José Luis Marzo Lázaro (Univ. de Girona), <joseluis.marzo@udg.es>

Juan Carlos López López (UCLM), <juancarlos.lopez@uclm.es>

Robótica

José Cortés Arenas (Sopra Group), <joscorare@gmail.com>

Juan González Gómez (Universidad CARLOS III), <juan@iearobotics.com>

Seguridad

Javier Areñio Bertollin (Univ. de Deusto), <jareñio@deusto.es>

Javier López Muñoz (ETSI Informática-UMA), <jlmm@lcc.uma.es>

Sistemas de Tiempo Real

Alejandro Alonso Muñoz, Juan Antonio de la Puente Alfaro (DIT-UPM), <caalmon@puente@dit.upm.es>

Software Libre

Jesús M. González Barahona (GSYC - URJC), <jgb@gsyc.es>

Israel Herráiz Tabernero (Universidad Politécnica de Madrid), <isra@herraz.org>

Tecnología de Objetos

Jesús García Molina (DIS-UM), <jmolina@um.es>

Gustavo Rossi (LIFIA-UNLP Argentina), <gustavo@sol.info.unlp.edu.ar>

Tecnologías para la Educación

Juan Manuel Dodero Beardo (UC3M), <jdodero@inf.uc3m.es>

César Pablo Córcoles Briogio (UOC), <ccorcoles@uoc.edu>

Tecnologías y Empresa

Didac Lopez Viñas (Universitat de Girona), <didac.lopez@ati.es>

Alonso Álvarez García (TID), <aa@tid.es>

Tendencias tecnológicas

Gabriel Martí Fuentes (Interbits), <gabi@atinet.es>

Juan Carlos Vigo (ATI), <juancarlosvigo@atinet.es>

TIC y Turismo

Andrés Aguayo Maldonado, Antonio Guevara Plaza (Univ. de Málaga), <[aguayo.guevara@lcc.uma.es](mailto:(aguayo.guevara)@lcc.uma.es)>

Las opiniones expresadas por los autores son responsabilidad exclusiva de los mismos.

Novática permite la reproducción, sin ánimo de lucro, de todos los artículos, a menos que lo impida la modalidad de © o copyright elegida por el autor, debiéndose en todo caso citar su procedencia y enviar a **Novática** un ejemplar de la publicación.

Coordinación Editorial, Redacción Central y Redacción ATI Madrid

Plaza de España 6, 2ª planta, 28008 Madrid

Tfno. 91 4029391; fax. 91 3093685 <novatica@ati.es>

Composición, Edición y Redacción ATI Valencia

Av. del Reino de Valencia 29, 46005 Valencia

Tfno. 963740173 <novatica_prod@ati.es>

Administración y Redacción ATI Cataluña

Via Laietana 46, ppal. 1º, 08003 Barcelona

Tfno. 934125235; fax. 934127713 <secretgen@ati.es>

Redacción ATI Andalucía

Av. de la Constitución 10, 41013 Sevilla

Tfno. 952011111 <secretand@ati.es>

Redacción ATI Galicia

Av. de la Constitución 10, 15101 Santiago de Compostela

Tfno. 981011111 <secretgal@ati.es>

Suscripción y Ventas

<novatica.subscriptions@atinet.es>

Publicidad

Plaza de España 6, 2ª planta, 28008 Madrid

Tfno. 91 4029391; fax. 91 3093685 <novatica@ati.es>

Imprenta: Derra S.A.

Juan de Austria 86, 08005 Barcelona

Depósito legal: B 15.154-1975

ISSN: 0211-2124; CODEN NOVACB

Portada: "Mineral, vegetal, animal"

Concha Arias Pérez / © ATI

Diseño: Fernando Agresta / © ATI 2003

editorial

El "caso Snowden" y la seguridad de las redes de telecomunicación en resumen > 02

Soporte al negocio y práctica profesional: El sueño del buen editor > 03

Llorenç Pagés Casas

noticias de IFIP

IFIP TC6 Latin American Tutorials in Networking (LATIN 2013) > 03

Ramon Puigjaner Trepap

monografía

Minería de procesos

Editores invitados: Antonio Valle Salas y Anne Rozinat

Presentación. Una perspectiva sobre la minería de procesos > 05

Antonio Valle Salas, Anne Rozinat

Minería de procesos: La objetivación de la intuición en los procesos de toma de decisiones en los negocios, más transparentes gracias al análisis de los datos > 07

Anne Rozinat, Wil van der Aalst

Minería de procesos: Obtenga una radiografía de sus procesos de negocio > 11

Wil van der Aalst

El viaje del descubrimiento de procesos > 20

Josep Carmona Vargas

Posibilidades de uso de la minería de procesos en ITSM > 24

Antonio Valle Salas

Optimización dirigida por minería de procesos de un proceso de aprobación de préstamos al consumo > 31

Arjel Bautista, Lalit Wangikar, S.M. Kumail Akbar

Mejoramiento de procesos con técnicas de minería de procesos, simulación y optimización: Caso de estudio > 41

Santiago Aguirre Mayorga, Carlos Alberto Parra Rodríguez

Detección de cambios temporales en los procesos de negocio mediante el uso de técnicas de segmentación > 49

Daniela Lorena Luengo Mundaca, Marcos Sepúlveda Fernández

secciones técnicas

Referencias autorizadas

visiones sobre Lenguajes de Programación

Cómo la metáfora de objetos llegó a ser el principal paradigma de programación > 62

Jesús J. García Molina

Elección de lenguajes de programación para la enseñanza universitaria > 67

Baltasar García Pérez-Schofield

La importancia de la labor del programador. ¿Qué se espera? ¿Cómo se prepara? > 70

Análisis desde los lenguajes de programación

Óscar Belmonte Fernández, Carlos Granell Canut

Para pensar > 79

Rafael Martínez Martínez

Programando caminos y resolviendo necesidades > 81

Alejandro Fuentes Penna

sociedad de la información

Programar es crear

El problema del CUIT (corrección del publicado en el número anterior) (Competencia UTN-FRC 2012, problema D, enunciado) > 82

Julio Javier Castillo, Diego Javier Serrano, Marina Elizabeth Cárdenas

Asuntos Interiores

Coordinación editorial / Programación de Novática / Socios Institucionales > 83

Tema del próximo número:

"Eficiencia energética en centros de proceso de datos"

Josep Carmona Vargas
Departament de Llenguatges i Sistemes
Informàtics, Universitat Politècnica de
Catalunya

<jcarmona@lsi.upc.edu>

El viaje del descubrimiento de procesos

Traducción: María del Carmen Ugarte García (Grupo de Trabajo de Lengua e Informàtica de ATI)

1. Introducción

La velocidad a la que crecen los datos en los sistemas de Tecnologías de la Información (TI) [1] hace crucial su automatización para permitir a empresas e instituciones gestionar sus procesos. Las técnicas automáticas abren la puerta a manejar grandes cantidades de datos, algo imposible para la capacidad humana. En este artículo hablamos sobre una de esas técnicas: modelos de descubrimiento de procesos. Para ilustrar la tarea principal detrás del descubrimiento de procesos nos serviremos de un ejemplo que esperamos les resulte simpático.

2. Un ejemplo divertido: la visita del marciano

Imaginemos que un marciano nos visita (ver figura 1) y, no importa por qué medios, quiere comunicar sus planes en lo que respecta a su visita a la Tierra. Por razones obvias no entendemos los mensajes del marciano, que tienen la apariencia mostrada en la figura 2.

A pesar de desconocer el significado de cada una de las letras en el mensaje anterior, se pueden detectar en él algunos patrones, por ejemplo una repetición de la secuencia I A C D M E (primera y últimas seis letras en la secuencia). Por lo tanto nos preguntamos: ¿cómo podemos representar el compor-



Figura 1. Imagen de nuestro marciano imaginario.

Resumen: Los modelos de proceso son un elemento de valor incalculable dentro de un sistema de Tecnologías de la Información (TI): sirven para analizar, monitorear o mejorar los procesos reales que proporcionan funcionalidad al sistema. La tecnología ha permitido a los sistemas TI almacenar en ficheros de log de eventos las marcas de las ejecuciones de los procesos, que pueden usarse para derivar los modelos de proceso que se corresponden con los procesos reales, mediante una disciplina llamada descubrimiento de procesos. En este artículo proporcionamos una visión general de esta disciplina con algunas de las alternativas existentes al día de hoy.

Palabras clave: Descubrimiento de procesos, Ingeniería del Software, métodos formales.

Autor

Josep Carmona Vargas obtuvo una Ingeniería y un Doctorado en Informática por la Universitat Politècnica de Catalunya en 1999 y en 2004, respectivamente. Es profesor asociado del Departament de Llenguatges i Sistemes Informàtics de dicha universidad. Sus intereses investigadores incluyen métodos formales, sistemas concurrentes y minería de procesos y datos. Ha sido autor o coautor de más de 50 artículos en congresos y revistas.

tamiento de los marcianos sin conocer exactamente el significado de cada unidad de información?

El descubrimiento de procesos puede ser una buena solución en este caso: el algoritmo de descubrimiento de procesos tratará de producir un modelo (formal) del comportamiento que subyace en un conjunto de secuencias. Por ejemplo, el modelo formal en notación para el Modelado de Procesos de Negocio (BPMN, Business Process Modeling Notation) [2] mostrado en la figura 3 representa muy acertadamente el comportamiento expresado en las secuencias del marciano. Para aquellos que no están familiarizados con la notación BPMN, el modelo anterior describe el siguiente proceso: cuando I ocurre, entonces (operador 'x') o el proceso B seguido de X ocurre, o bien el proceso A seguido de C y D en paralelo (operador '+') ocurre, seguidos a su vez de M. Ambos procesos activan E, que a su vez reactiva I. Es claro, que incluso sin conocer nada sobre las acciones que ha tomado el marciano, la estructura general de estas actividades resulta patente con una simple inspección del modelo BPMN.

Ahora imaginemos que en un determinado momento se descifra el significado de las letras: evalúa el total de energía en la Tierra

(I, nivel de energía alto (B), invade la Tierra (X), nivel de energía bajo (A), reúne algunos ejemplares humanos (C), infórmate sobre el sistema de reproducción humana (D), enseña a los humanos cómo incrementar sus recursos energéticos (M), informa a los marcianos del platillo volante más próximo (E). Ante esta nueva información, el valor del modelo obtenido se incrementa significativamente (a pesar de que no podamos relajarnos al descubrir la situación global que el modelo pone en evidencia).

3. Anatomía de un algoritmo simple de descubrimiento de procesos

El ejemplo anterior ilustra uno de las principales tareas de un algoritmo de descubrimiento de procesos: dado un conjunto de trazas (llamado log de eventos) correspondiente a un comportamiento particular sometido a estudio, derivar un modelo formal que represente fielmente el proceso que ha producido esas trazas. En su forma más sencilla, los algoritmos de descubrimiento de procesos ponen el foco sobre la perspectiva de control de flujo, por ejemplo se calcula la ordenación entre actividades para realizar las ejecuciones del proceso. El ejemplo anterior ha considerado esta perspectiva.

Un log de eventos debe contener suficiente información para extraer la secuencia de las

I A C D M E I B X E I A D C M E I B X E I A C D M E

Figura 2. Mensajes enviados por el marciano.

“ El núcleo de un algoritmo de descubrimiento de procesos es la capacidad de extracción de la información necesaria para llegar mediante aprendizaje a un modelo que represente el proceso ”

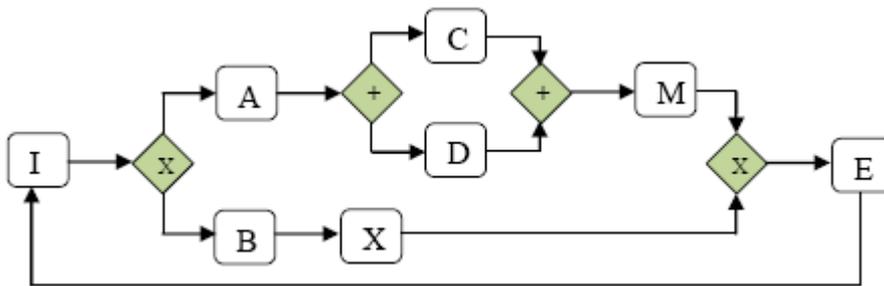


Figura 3. Modelo formal del comportamiento de las secuencias del marciano en BPMN.

actividades que están siendo monitorizadas. Típicamente se requiere un identificador de trazas, un nombre de actividad e información temporal para permitir la secuenciación correspondiente (mediante el sello de tiempo) para las actividades pertenecientes a una traza dada (determinada por un identificador de trazas). Otra información puede ser necesaria si el algoritmo de descubrimiento debe tener en cuenta información adicional como recursos (*qué cantidad se compró*), causante de la actividad (*quién llevó a cabo esa actividad?*), duración de la actividad (*cuánto duró la actividad*), entre otras. Un ejemplo del algoritmo de descubrimiento que tiene en cuenta otras dimensiones es el algoritmo para descubrir redes sociales [3], que deriva la red a través de un colaborador que realiza un proceso dado.

El núcleo de un algoritmo de descubrimiento de procesos es la capacidad de extracción de la información necesaria para llegar mediante aprendizaje a un modelo que represente el proceso.

El descubrimiento de procesos es a menudo una *tarea de aprendizaje no supervisada*, ya que el algoritmo está normalmente expuesto solo a ejemplos positivos, por ejemplo ejecuciones exitosas en el proceso en estudio: en el ejemplo de la introducción, estamos solo expuestos a lo que el marciano planea hacer, pero no sabemos lo que el marciano no piensa hacer. Esto complica la tarea de aprendizaje, ya que se espera que los algoritmos del descubrimiento del proceso produzcan modelos que sean a la vez *precisos* (el modelo resultante no debería desviarse mucho del comportamiento observado) y *generales* (el modelo debería ser capaz de generalizar los modelos observados en el *log* de eventos) [4]. Obviamente la presencia de ejemplos negati-

vos ayudaría al algoritmo de descubrimiento a mejorar estas dos métricas de calidad, pero la información negativa no está normalmente en los registros TI.

¿Cómo llegar un modelo de proceso a partir de un conjunto de trazas? Existen hoy día varios algoritmos para distintos modelos (ver **sección 4**). Sin embargo, usemos el ejemplo del marciano para razonar sobre el descubrimiento del modelo BPMN anterior. Si nos centramos en la primera letra de la secuencia (I), vemos que a veces le sigue A y a veces le sigue B, y siempre (excepto en la primera ocurrencia) E los precede. Estas observaciones pueden verse expresadas de modo gráfico en la **figura 4**.

En la notación BPMN, la relación *disjunción exclusiva* entre A o B después de I se modela usando el operador "x". La precedencia entre E e I se modela mediante una línea que conecta ambas letras. De forma simétrica, o bien M o bien X preceden a E. Igualmente, tanto C como D pueden seguir a A en cualquier orden. El bien conocido algoritmo *alfa* [5] puede encontrar la mayoría de las relaciones existentes entre pares en el *log* de eventos, y después pueden utilizarse para fabricar el modelo BPMN tal como ilustramos en la **tabla 1**.

Esta tabla puede leerse de la siguiente forma: si en el *log* de eventos A precede a B siempre, pero B es único (no hay otra letra que vaya detrás de A), entonces se crea un arco directo entre A y B. Si por el contrario hay siempre más de una letra que siga a A, entonces el operador "+" se inserta entre A y las letras que le siguen. Cuando la relación ocurre solo algunas veces la lectura es similar.

Por lo tanto se puede escanear el *log* de eventos para extraer estas relaciones (en el caso peor complejidad cuadrática respecto al *log* de eventos) y usar la tabla para crear el modelo BPMN. Sin embargo, esta es una forma de descubrimiento muy restrictiva ya que otras relaciones en la notación BPMN pueden estar escondidas en el *log* de eventos, como la relación de disjunción inclusiva, pero el algoritmo no las considera.

Los algoritmos de descubrimiento de procesos están siempre balanceando la complejidad de un algoritmo y su capacidad de modelado: el algoritmo propuesto en esta sección podría extenderse para considerar operadores o-inclusivos, pero ello lo complicaría significativamente. Más abajo analizaremos someramente este y otros problemas.

4. Algoritmos y modelos

Hay distintos modelos que pueden obtenerse a través de distintos algoritmos de descubrimiento de procesos: *redes de Petri, cadenas de eventos de proceso, redes causales, redes heurísticas, mapas de procesos de negocio*, entre otros. Hay que destacar que la mayoría de estos modelos disponen de una semántica que permite simular el modelo para así certificar si se adecúa al *log* de eventos.

La descripción de cada uno de ellos cae fuera de los límites de este artículo, pero comentaré brevemente sobre las redes de Petri, que es un modelo que a menudo producen los

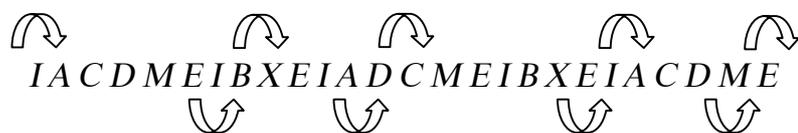


Figura 4. Patrones observados en los mensajes del marciano.

“ Decidir cuál es la mejor notación de modelado para un registro de eventos es un problema arduo para el que la investigación debe proporcionar técnicas en la próxima década ”

	Always	Sometimes
A precedes B	B Unique: 	B Unique:
	General: 	General:

Tabla 1. Construcción de un modelo BPMN a partir de las relaciones descubiertas.

algoritmos de descubrimiento, debido a su semántica formal y la capacidad para representar la concurrencia. Para el modelo de nuestro ejemplo, la correspondiente red de Petri que sería puesta de manifiesto por la mayoría de los algoritmos de descubrimiento basados en Petri sería la mostrada en la figura 5.

Aquellos lectores familiarizados con las redes de Petri encontrarán un perfecto encaje entre el comportamiento subyacente en la red de Petri y la traza del marciano. Adviertan que mientras en el modelo BPMN, además de las unidades de información (en este caso las letras del alfabeto) hay otros componentes del modelo (operadores) cuya semántica define la forma en la que el modelo representa las trazas en el log de eventos.

Lo mismo ocurre con la red de Petri anterior, en la que los círculos se corresponden con el comportamiento global del modelo, que se distribuye entre la red (hay marcados algunos

círculos). Mientras el algoritmo de descubrimiento por BPMN necesita encontrar tanto las conexiones como los operadores, el algoritmo análogo de las redes de Petri debe computar tanto los círculos como las conexiones.

Hoy en día existen algunas técnicas para llevar a cabo el descubrimiento mediante redes de Petri; estas oscilan desde las relaciones extraídas de la ordenación del log de eventos mediante el algoritmo *alfa*, hasta estructuras complejas basadas en grafos que se computan sobre un autómata que representa las trazas del log de eventos.

¿Qué notación de algoritmo/modelado de descubrimiento de proceso deberemos elegir? Buena pregunta que solo tiene una respuesta parcial: no hay un modelo mejor que el resto, pero a cambio sí hay modelos mejores para un determinado tipo de comportamiento. Realmente decidir cuál es la mejor notación de modelado para un log de eventos es un pro-

blema arduo para el que la investigación debe proporcionar técnicas en la próxima década (un problema llamado *selección de sesgo representativo*). Desde un punto de vista pragmático, se deben seleccionar las notaciones de modelado de procesos con las que se esté más familiarizado, y esperar que los algoritmos de descubrimiento para esa notación sean adecuados para nuestras necesidades.

Como se ha dicho con anterioridad pueden considerarse otras perspectivas diferentes al flujo de control para los algoritmos de descubrimiento del proceso: tiempo, recursos, organización, etc.

Puede consultarse el manual de referencia [6] para profundizar en estos otros algoritmos de descubrimiento de procesos.

5. Herramientas

El descubrimiento de procesos es una disciplina bastante nueva si la comparamos con otras áreas ajenas tales como la minería de

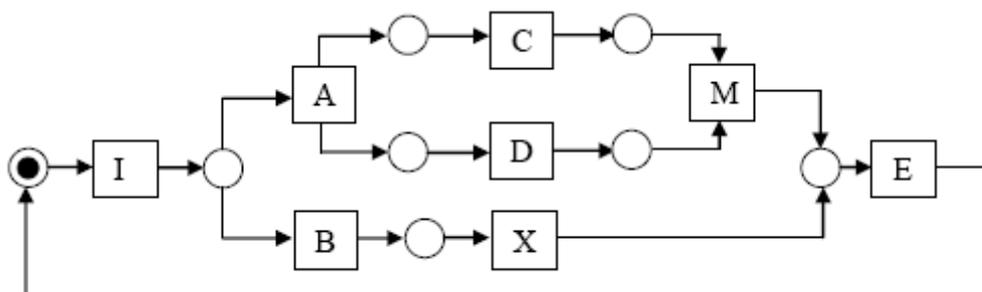


Figura 5. Red de Petri para el modelo de nuestro ejemplo.

datos o el aprendizaje automático. A pesar de ello, se pueden encontrar herramientas de minería de procesos tanto en el mundo académico (la mayor parte) como en la industria. La siguiente clasificación no es en absoluto exhaustiva, pero recoge algunas de las herramientas más relevantes que se pueden utilizar como herramientas de descubrimiento de procesos:

- **Mundo académico:** El ProM Framework, de la Universidad Técnica de Eindhoven (TU/e) es hoy día la herramienta de referencia. Es el resultado de una gran colaboración entre distintas universidades del mundo para reunir soporte algorítmico para la minería de procesos (por ejemplo no solo descubrimiento de procesos). Adicionalmente, distintos grupos han desarrollado algunas herramientas autónomas en el ámbito universitario que incorporan modernos algoritmos de descubrimiento de procesos.

- **Industria:** Algunas importantes empresas han invertido recursos en construir herramientas para el descubrimiento de procesos, por ejemplo Fujitsu (APD), pero también empresas de mediano tamaño y empresas emergentes que están más enfocadas hacia las prácticas de minería de procesos, por ejemplo Pallas Athena (ReflectOne), Fluxicon (Disco), Perspective Software (BPMOne, Futura Reflect), Software AG (ARIS Process Performance Manager), entre otras.

6. Retos

La tarea de descubrimiento de procesos puede dificultarse si aparecen alguno de los siguientes aspectos:

- **Registros de eventos incompletos:** El log de eventos contiene solo una fracción del total del comportamiento representativo del proceso. Por lo tanto, el algoritmo de descubrimiento de procesos necesita suponer parte del comportamiento que no está presente en el log de eventos, algo que en general es difícil.

- **Ruido:** El comportamiento registrado puede presentar a veces raras excepciones que no forman parte del proceso. Por lo tanto, los algoritmos de descubrimiento del proceso podrían obstaculizarse cuando aparece el ruido, p. ejp. en el descubrimiento del control de flujo algunas relaciones entre las actividades pueden resultar contradictorias. Separar el ruido de la información válida en un log de eventos es actualmente una línea de investigación.

- **Complejidad:** Debido a la magnitud de los actuales registros de eventos TI, es a menudo difícil utilizar algoritmos complejos que pueden requerir cargar el log de eventos en memoria para derivar el modelo de proceso, o aplicar técnicas cuya complejidad es no lineal al tamaño del log de eventos. En estos casos, estrategias de alto nivel (p. ejp., *divide y vencerás*) son la única posibilidad de derivar el modelo de proceso.

- **Visualización:** Incluso si el algoritmo de descubrimiento de procesos hace su trabajo y sabe cómo derivar un modelo de proceso, podría ser difícil para un humano entenderlo si hay más de un centenar de elementos (nodos, arcos). En estos casos, una descripción jerárquica, similar a la aplicación *Google Maps* donde se pueda ampliar o reducir parte del modelo, facilitará la comprensión de un modelo complejo de proceso.

Agradecimientos

Agradezco a David Antón la creación del dibujo del marciano de este artículo.

Referencias

[1] S. Rogers. Data is Scaling BI and Analytics-Data Growth is About to Accelerate Exponentially - Get Ready. *Information and Management - Brookfield*, 21(5):p. 14, 2011.

[2] D. Miers, S.A. White. BPMN Modeling and Reference Guide: Understanding and Using BPMN. *Future Strategies Inc.*, 2008. ISBN-10: 0977752720.

[3] W. M. P. van der Aalst, H. Reijers, M. Song. *Discovering Social Networks from Event Logs. Computer Supported Cooperative Work*, 14(6):pp. 549-593, 2005.

[4] A. Rozinat, W. M. P. van der Aalst. Conformance Checking of Processes Based on Monitoring Real Behavior. *Information Systems*, 33(1):pp. 64-95, 2008.

[5] W.M.P. van der Aalst, A. Weijters, L. Maruster. Workflow Mining: Discovering Process Models from Event Logs. *IEEE Transactions on Knowledge and Data Engineering*, 16 (9):pp. 1128-1142, 2004.

[6] W.M.P. van der Aalst. *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Springer, 2011. ISBN-10: 3642193447.