

Consultas eficientes sobre registros de eventos basadas en análisis de conformidad*

Jacobo Casas Ramos, Manuel Lama, Manuel Mucientes

Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS)
Universidade de Santiago de Compostela

`jacobo.casas.ramos@rai.usc.es`, `{manuel.lama,manuel.mucientes}@usc.es`

Abstract. En este artículo se presenta una aproximación basada en técnicas de conformidad de procesos para la resolución de consultas orientadas a definición de modelos de proceso y de condiciones temporales sobre actividades y sobre indicadores clave de negocio.

Keywords: Minería de procesos · Análisis de la conformidad · Consultas.

1 Introducción

Las técnicas de conformidad de procesos (*conformance checking*) comprueban si las trazas de ejecución de un proceso son consistentes con el modelo que describe su comportamiento [1]. Sin embargo, en ocasiones los usuarios están interesados en consultar si un submodelo se verifica y/o si ciertas condiciones temporales o restricciones sobre indicadores clave de negocio han tenido lugar.

Para resolver este problema se han propuesto trabajos basados en una aproximación declarativa [2] y/o en chequeo de modelos [3]. El principal problema de estas aproximaciones es su falta de flexibilidad a la hora de, por una parte, definir submodelos de proceso que contienen cualquier estructura de control, y, por otra parte, añadir a dichos modelos condiciones temporales complejas sobre actividades y condiciones sobre indicadores clave de negocio. En este artículo se abordará este problema con una aproximación basada en técnicas de conformidad de procesos y en tecnologías de análisis masivo de datos (*Big Data*).

2 Descripción del sistema

El algoritmo está basado en alineamientos, usando una implementación de *token replay* en la que se comprueba si el modelo se ejecuta en alguna parte de la traza, o lo que es lo mismo, si la traza se adapta al modelo que constituye la

* Este trabajo ha sido financiado por el Ministerio de Economía y Competitividad a través del proyecto TIN2017-84796-C2-1-R; por la Consellería de Cultura, Educación e Ordenación Universitaria de Galicia (acreditación 2016-2019, ED431G/08) y por el European Regional Development Fund (ERDF).

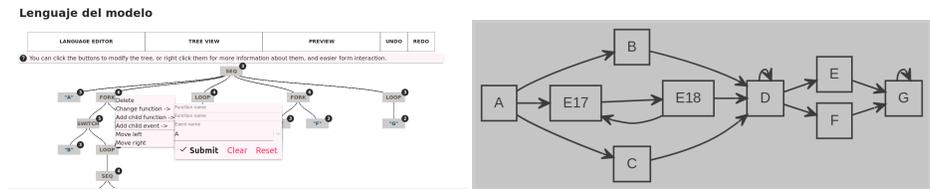
consulta del usuario. Además, existen dos modos de consulta: *estricto*, en el que la traza debe adaptarse completamente al modelo, es decir, no deben existir movimientos ni en el modelo ni en la traza, y *relajado*, en el que se permiten movimientos que únicamente tienen lugar en la traza. Este modo permite realizar consultas con condiciones sobre determinadas actividades, independientemente de si se han registrado otras actividades entre ellas. En estos casos, el registro de eventos contiene más información que la que desea consultar el usuario.

Para realizar las consultas sobre los modelos se ha hecho uso de dos lenguajes: uno que permite especificar la *estructura del modelo*, que está basado en el modelado de procesos a través de árboles (*process trees*); y otro que permite definir las condiciones que restringen la exploración del modelo a la hora de realizar consultas. Más específicamente:

- *Lenguaje sobre la estructura del modelo*. Contiene las estructuras de control básicas de los árboles de procesos y que, típicamente, son secuencias (*SEQ*), selecciones (*SWITCH*), paralelos (*FORK*) y bucles de actividades. En la Figura 1 se muestra un ejemplo de consulta de un modelo que contiene tres bucles de actividades y dos paralelos.

Fig. 1. Ejemplo de un modelo, especificado como un árbol de proceso, que constituye la consulta que se desea realizar sobre el registro de eventos

Consulta	Red Causal (<i>Causal Net</i>)
SEQ("A", LOOP(SEQ(E17, LOOP("D"), LOOP("G"))), FORK(SWITCH("B", E18))), "C"), FORK("E", "F")	$I(F_6)=[5,1]$ $O(F_6)=[]$ $I(E_5)=[3,4]$ $O(E_5)=[6]$ $I(D_4)=[2]$ $O(D_4)=[5]$ $I(C_3)=[2]$ $O(C_3)=[5]$ $I(B_2)=[0]$ $O(B_2)=[3,4]$ $I(A_1)=[0]$ $O(A_1)=[6]$ $I(I_0)=[]$ $O(I_0)=[1,2]$



- *Lenguaje sobre condiciones*. Permite especificar, por una parte, condiciones que deben cumplir los indicadores clave de negocio asociados a los eventos que componen una traza y, por otra parte, restricciones temporales sobre las actividades que forman parte del modelo a consultar. Por ejemplo, con la condición $((A["tiempo"] > B["tiempo"]) \text{AND} (C["users"] < 2000))$ las trazas devueltas deben cumplir que el tiempo de ocurrencia de la última actividad *A* que se ha ejecutado sea posterior al tiempo de ocurrencia de la actividad *B* y que el número de usuarios sea inferior a 2000 cuando se haya ejecutado la siguiente actividad *C*.

Una consulta puede estar formada por un modelo y/o por una o varias condiciones sobre indicadores clave de negocio y/o sobre relaciones temporales entre actividades, de modo que como resultado se devolverán el número de trazas en las que se verifica dicha consulta y el número de veces que se ha cumplido (en una traza la consulta puede cumplirse en más de una ocasión). Opcionalmente se pueden devolver las trazas para poder extraer una serie de estadísticas que pueden ser de interés para el usuario como, por ejemplo, la frecuencia y el tiempo medio de ejecución de los arcos y de las actividades del modelo.

3 Arquitectura Software

En la Figura 2 se muestra la arquitectura software desarrollada para dar soporte a la resolución de las consultas sobre registros de eventos. La principal característica de esta arquitectura es que está basada en tecnologías de análisis de masivo de datos; concretamente, hace uso de un clúster Apache Spark, puesto que la resolución de las consultas se puede paralelizar aprovechando que la comprobación de la verificación del modelo sobre cada traza es independiente de las demás. De este modo, el registro de eventos se divide en diferentes partes sobre las que se aplicará el algoritmo de resolución de consultas de forma paralela (*fase map*). El resultado final consistirá en agregar los resultados de dicha consultas (*fase reduce*).

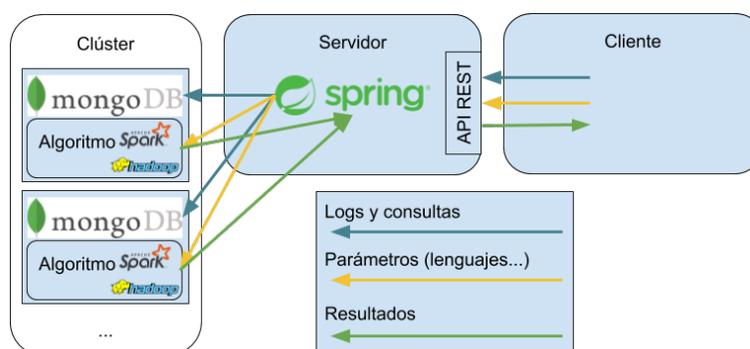


Fig. 2. Arquitectura software del sistema

References

1. W.P.M. van der Aalst, Process Mining: Discovery, Conformance and Enhancement of Business Processes, 1st edition, Springer, 2011.
2. S. Schönig, C. Di Ciccio, F. M. Maggi, J. Mendling, Discovery of Multi-perspective Declarative Process Models, Springer, 2016, pp.87–103.
3. S. Hernández, Pedro Álvarez, Javier Fabra, Joaquin Ezpeleta, Analysis of Users' Behavior in Structured e-Commerce Websites, IEEE Access, 5 (2017) 11941–11958.