

ALGORITMOS PARA EL DESCUBRIMIENTO DE PROCESOS. ESTADO DEL ARTE

Ing. Asiel Díaz Vasallo¹, Dr.C Yasser Vázquez Alfonso², MSc. Orlenys López Pintado³& M.Sc. Velasteguí López Luis Efraín⁴

ABSTRACT

Process mining techniques allow the extraction of knowledge from various sources of information, the discovery of models is one of the stages that make up the process mining. At this stage conventional algorithms start from an event log and discover process models. In practical situations most cases contain several processes that collaborate and are mostly stored in relational databases. For these cases the researchers assume that a process can have different perspectives, each with its own space of states, in this way they have been able to discover process models in collaborative environments. There are several papers that present different methods focused on the processing of the data in relational databases. Some articles refer to a technique based on identifying an artifact-centric model, starting from a relational schema obtained from a database. The relationships existing in the scheme are extracted and organized with the use of graph algorithms..

Keywords

Process mining, process model, collaborative environments, artifacts scheme, relational schema.

CÓDIGO UNESCO: Informática 120317

RESUMEN

Las técnicas de la minería de procesos permiten la extracción de conocimientos de diversas fuentes de información, el descubrimiento de modelos es una de las etapas que componen la minería de procesos. En esta etapa los algoritmos convencionales parten inicialmente de un registro de eventos y descubren modelos de procesos. En situaciones prácticas la mayoría de los casos contienen varios procesos que colaboran y en su mayoría se encuentran almacenados en bases de datos relacionales. En este sentido, algunos investigadores asumen que un proceso puede tener diferentes perspectivas, cada una de ellas con su propio espacio de estados, de esta forma han podido descubrir modelos de procesos en entornos colaborativos. En varias investigaciones se plantean diferentes métodos enfocados en el procesamiento de los datos en bases de datos relacionales, existen artículos donde se hace referencia a una técnica basada en identificar un modelo centrado en artefactos, partiendo de un esquema relacional obtenido de una base de datos. Las relaciones existentes en el esquema son extraídas y organizadas con el uso de algoritmos de grafos.

¹ Universidad de la Agraria de la Habana, Habana, Cuba, asioldv@unah.edu.cu

² Universidad de la Habana. Habana, Cuba, yalfos01@gmail.com

³ Universidad de Tartu, Cuba orlenyslp@gmail.com

⁴ Universidad Técnica de Ambato, Ambato, Ecuador le.velastegui@uta.edu.ec

Palabras clave: minería de procesos, modelo de procesos, entornos colaborativos, esquema de artefactos, esquema relacional.

INTRODUCCIÓN

Los tiempos modernos con sus avances científicos y tecnológicos introducen nuevos retos; extraer conocimiento procesando grandes volúmenes de datos es uno de estos. Actualmente, las instituciones gestionan la ejecución de sus procesos de negocios mediante complejos sistemas computacionales; estos almacenan los datos relacionados con los procesos ejecutados en diversas fuentes de información, cuyo procesamiento manual se torna una tarea imposible de realizar.

En este sentido, la minería de procesos provee un conjunto de técnicas para examinar procesos reales en dominios de aplicaciones variados. Esta disciplina está compuesta fundamentalmente por tres etapas, descubrimiento del modelo, chequeo de conformidad y extensión del modelo, en la primera, las técnicas de descubrimiento parten de una información previa, sobre la que se efectúan análisis para la obtención de un modelo del proceso. En el chequeo de conformidad, se realiza la comparación de un modelo de procesos existente con el registro de eventos de un proceso (van der Aalst, Adriansyah, & Alves de Medeiros, 2011), para verificar cuán real es el comportamiento del modelo con respecto a los datos históricos obtenidos durante la ejecución del proceso.

La comparación de registros de eventos con modelos que especifiquen los requisitos de un proceso contribuye a la detección de casos de fraude. Por tal motivo, la comprobación de conformidad puede ser usada para descubrir, localizar y explicar inconsistencias en el proceso, por otro lado, la tercera etapa tiene como reto mejorar los modelos utilizando información acerca de los procesos actuales, almacenados en los registros de eventos.

Esta investigación abordará la etapa de descubrimiento de procesos, haciendo énfasis en los algoritmos existentes para descubrir modelos. En este sentido, se centrará en dos de los grandes retos que enfrenta la minería de procesos en la actualidad. La interacción de procesos en entornos colaborativos y la extracción de conocimiento de procesos almacenados en bases de datos relacionales.

1. Algoritmos para el descubrimiento de modelos

Múltiples algoritmos diseñados para descubrir modelos de procesos, se basan fundamentalmente en análisis estadísticos, métodos heurísticos y otras técnicas de la matemática aplicada. Entre los más conocidos se hallan: el *Alpha Miner* (van der Aalst, A. J.M.M., & Maruster, 2004), el *Heuristic Miner* (Weijters, van der Aalst, & Alves de Medeiros, 2006), el *Fuzzy Miner* (Gunter & van der Aalst, 2007) y el *Genetic Miner* (Alves de Medeiros, 2006). Todos estos se encuentran implementados en la herramienta ProM (B. F. van Dongen, 2005) desarrollada por investigadores de la Universidad Tecnológica de Eindhoven, en Holanda; centro donde se han obtenido varios de los principales resultados en el área de minería de procesos.

El *Alpha Miner* es un algoritmo pionero en esta rama, de fácil entendimiento y que se convierte en una referencia necesaria para todo joven investigador. A partir de un registro de eventos recibido como entrada, que contiene varias instancias de la ejecución de un proceso (van der Aalst, 2011); se ordenan las relaciones entre las tareas que son deducidas. Mediante el orden establecido se indica la precedencia entre las mismas, su ejecución en paralelo, entre otros aspectos que garantizan la reconstrucción de un flujo de procesos. El modelo descubierto se corresponde con una red de Petri (López, 2015), que representa un flujo de trabajo (van der Aalst, 2011). Este algoritmo no es robusto al ruido ni ante patrones estructurales complejos, por lo que no se aplica con frecuencia en situaciones prácticas y su uso es casi exclusivamente académico.

Los algoritmos *Fuzzy Miner*, *Heuristic Miner* y *Genetic Miner* son capaces de lidiar con registros de eventos ruidosos y con patrones estructurales complejos, por lo que tienen gran aplicabilidad en situaciones prácticas. En el caso del *Fuzzy Miner* se extraen modelos jerárquicos. Las actividades con baja frecuencia pero que se encuentran estrechamente relacionadas son agrupadas en subprocesos; según criterios y métricas previamente definidas, que determinan cuáles arcos deben ser incluidos en una matriz causal (Díaz, 2014). De este modo, se genera un modelo *fuzzy*, ideal para representar los llamados procesos espagueti (van der Aalst, 2011).

El algoritmo *Heuristic Miner* utiliza matrices causales (Alves de Medeiros, Weijters, & van der Aalst, 2007) para la representación del modelo que descubre. Este algoritmo tiene en cuenta la frecuencia de los eventos y secuencias al construir el modelo de procesos, estableciendo heurísticas que rigen la construcción del mismo. La idea principal está basada en que caminos poco frecuentes no deberían aparecer en el modelo de procesos resultante. Su tendencia de representar los modelos mediante matrices causales y el manejo de las frecuencias, hacen que este algoritmo sea uno de los más robustos.

El *Genetic Miner* utiliza un procedimiento iterativo para simular el proceso de evolución natural. A diferencia de los algoritmos *Alpha Miner*, *Fuzzy Miner* y *Heuristic Miner*, que construyen los modelos que generan mediante técnicas directas y deterministas locales, el algoritmo *Genetic Miner* utiliza una estrategia global para procesar los registros de eventos (Díaz Vasallo, 2014).

El empleo de estas técnicas ha permitido dar solución a una gran cantidad de problemas existentes. (Jagadeesh, Bose, & van der Aalst, 2012), exponen un mecanismo para alinear secuencias de ADN, facilitando la realización de diagnósticos. Los resultados obtenidos por (Mans, Schonenberg, Song, van der Aalst, & Bakker, 2009), se centran en casos reales de oncología ginecológica. (Peleg, Yeh, & Altman, 2002), descubren modelos que permiten la representación de procesos biológicos mediante redes de Petri. (Deevakar, 2015), muestra una variante para analizar los registros de salud electrónicos y comprender las desviaciones de los procesos que tienen lugar.

En todas estas investigaciones solo se contempla un único proceso, por lo que los enfoques clásicos de minería de procesos pueden ser aplicados. Sin embargo, existen escenarios colaborativos donde diversos procesos interactúan, en los cuales las técnicas convencionales existentes no brindan los resultados esperados. A esta perspectiva se le

denomina minería inter-organizacional y constituye uno de los grandes retos que enfrenta la minería de procesos (van der Aalst et al., 2011).

2. Minería de procesos inter-organizacional

En un inicio, la minería de procesos inter-organizacional pudiera evidenciarse desde dos escenarios diferentes. Un primer acercamiento sería la interacción de varios subprocesos, que trabajan en conjunto para manejar instancias de un proceso general (figura 1.a). En este caso, el análisis independiente de un subproceso sería insuficiente. El segundo escenario lo constituirá la ejecución del mismo proceso por parte de diferentes individuos, mientras comparten experiencias, conocimiento o una infraestructura común (figura 1.b).

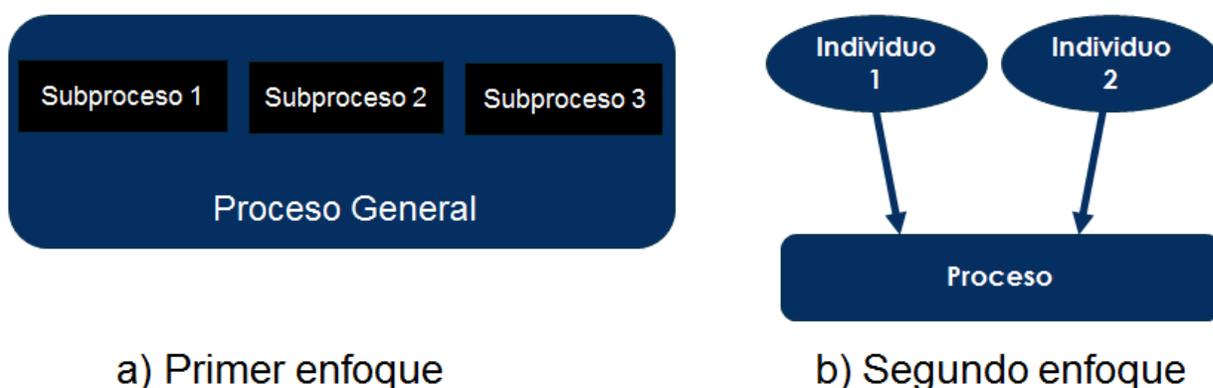


Figura 1: Enfoques de la minería de procesos inter-organizacional

Fuente: Elaboración propia.

De acuerdo con (van der Aalst, 2015), ante todo se deben considerar los ajustes colaborativos donde diferentes organizaciones trabajan juntas para manejar instancias de procesos. Lo que requiere que diversas partes estén disponibles para inter-operar, es decir, coordinar sus actividades. En la investigación realizada por (van der Aalst, 2000), se identifican cinco formas de interoperabilidad descritas a continuación:

- Capacidad de compartir. Esta forma de interoperabilidad asume un control centralizado, o sea, la dirección de los casos ocurre bajo el control de una sola organización. La ejecución de las tareas es distribuida, es decir, los recursos de diferentes organizaciones pueden ejecutar tareas.
- Ejecución en cadena. El proceso es separado en un número de subprocesos disjuntos, que son ejecutados por organizaciones en orden secuencial. Esta forma requiere que un compañero transfiera o inicialice el flujo para un caso después de completar todo el trabajo. El control de flujo es distribuido entre diferentes organizaciones.
- Subcontratación. Una organización subcontrata subprocesos a otras organizaciones.

- Transferencia de caso. Cada organización tiene una copia de la misma descripción del proceso, en otras palabras, las especificaciones de los procesos son replicadas. Sin embargo, en cualquier instante, cada caso (instancia del proceso) reside en exactamente una localización. Los casos pueden ser transferidos de una organización a otra, ya sea para balancear una carga de trabajo o porque las tareas no se han puesto en práctica en todas las organizaciones.
- Acoplado libremente. El proceso es separado en piezas que pueden activarse concurrentemente. Además, la definición de cada uno de los subprocesos es local, o sea, el ambiente no necesita conocer al proceso. Solo el protocolo que se emplea para efectuar la comunicación con las otras partes involucradas.

2.1 . Descubriendo procesos inter-organizacionales. Trabajos relacionados

(Chen & Hsu, 2001) citados por (Dayal, Hsu, & Ladin, 2001), presentan un *framework* de colaboración para ampliar la tecnología de administración de procesos centralizada. Se basa en el principio de que un proceso colaborativo involucra múltiples partes, donde cada una desempeña un rol dentro del proceso. Ellos plantean que existen dos aspectos a distinguir en el interior del modelo de procesos colaborativos respecto al modelo de procesos centralizado:

1. La definición de procesos está basada en un protocolo de interacción comercial comúnmente acordado, como el protocolo para la compra en línea o subasta.
2. La ejecución del proceso no es realizada por un mecanismo de flujo de trabajo centralizado, sino por múltiples mecanismos colaborativos.

(Di Ciccio et al., 2011), presentan una variante para extraer procesos interesantes de colecciones de mensajes de correo electrónico. La propuesta apunta al desafío de investigación abierto de transacciones con registros de eventos variados, es decir, detecta los rastros de múltiples procesos presentes en el sistema.

(Lu, Nagelkerke, van de Wiel, & Fahland, 2015), proponen un mecanismo de descubrimiento de servicios colaborativos haciendo uso de ontologías, las cuales pueden variar durante el proceso de descubrimiento. En este caso es necesario que ocurra la mediación de datos para resolver cuestiones de heterogeneidad.

La investigación realizada por (Zhao & Liu, 2006), se centra en darle seguimiento al flujo de trabajo a través de los límites organizacionales. En comparación con otras soluciones de rastreo de flujo de trabajo, el acercamiento propuesto no solo permite que una organización siga a otras por sus partes involucradas en procesos de negocios colaborativos, también posibilita que diferentes organizaciones persigan el mismo proceso de negocio de manera distinta. El *framework* desarrollado por los investigadores incluye tres representaciones matriciales y tres operaciones sobre matrices.

Los algoritmos emplean estas matrices y las operaciones para generar estructuras de seguimiento y se desarrolla la realización del rastreo del flujo de trabajo. Este *framework* permite a una organización generar su propia estructura de rastreo basada en su visibilidad respecto a otras organizaciones. Además, permite que una estructura de seguimiento pueda ser generada tiempo real, esta habilita la flexibilidad en la exploración del flujo de trabajo.

Basada en su propia estructura de seguimiento, una organización puede proactivamente trazar el progreso de ejecución de su parte involucrada en un proceso de negocio colaborativo.

Un aspecto importante es descubrir modelos de procesos apoyados en comportamientos capturados en datos de eventos. Estos modelos generalmente muestran las actividades que puedan ser ejecutadas durante el proceso y cómo ellas son ordenadas. Para el descubrimiento de procesos es de gran importancia deducir los estados de los procesos operacionales en el registro de eventos. Muchos algoritmos de minería solo tienen una noción de estado implícita, por lo que se centran en aprender el orden de las actividades. En este sentido, la noción de estado de los procesos puede presentarse explícitamente en sistemas de información. Tal es el caso de la información obtenida a partir del diagnóstico de un paciente con determinada afección de salud.

El trabajo propuesto por van Eck et al., (2016), se acerca a descubrir modelos para procesos donde diferentes facetas, o perspectivas, de los mismos puedan ser identificadas. En cambio, de centrarse en eventos o actividades que son ejecutadas en el contexto de un proceso particular; toman como punto de partida los estados de las diferentes perspectivas e intentan descubrir cómo están relacionadas. En este sentido asumen que un proceso puede tener diferentes perspectivas y cada una de ellas posee su propio espacio de estados. Cada estado del proceso de múltiples perspectivas bajo estudio se representa por un vector de estados de sus perspectivas.

Los autores modelan los procesos multi-perspectivas basados en estados con una máquina de estados finitos, a la que denominan “Máquina Compuesta por Estados”(van Eck et al., 2016a). Dado que el modelo puede tornarse complejo, proporcionan tres operaciones diferentes que puedan ser utilizadas para crear vistas simplificadas en máquinas de estados. Desarrollaron además una herramienta que permite visualizar las relaciones interesantes entre los estados y las transiciones que ocurren entre ellos(van Eck et al., 2016b).

Las investigaciones mencionadas anteriormente utilizan el XES como formato estándar asumido por los especialistas en el tema. Sin embargo, la mayoría de los sistemas de información emplean las bases de datos relacionales como fuente de almacenamiento y no este tipo de fichero. Esto ocasiona que para efectuar el minado de los datos se requiera del desarrollo de otras técnicas.

4. Extendiendo el descubrimiento a bases de datos relacionales

En esta temática se han realizado algunas investigaciones que proponen transformar las bases de datos en registros de eventos para posteriormente aplicarle las técnicas de minería de procesos. Los enfoques clásicos para convertir y extraer un registro de eventos desde fuentes de datos relacionales, se basan en identificar o definir una instancia del proceso, a partir de la cual se asocian y extraen los eventos(Lu et al., 2015). Finalmente, estos son ordenados por la instancia a la que pertenecen y el tiempo de ejecución. Este método permite solo extraer el registro de eventos de un proceso, asumiendo que este está aislado y no interactúa con otros.

Fahland et al., (2011), Dongen et al., (2012) y Lu, (2013), hacen uso de esquemas centrados en artefactos, que son entidades de negocio que interactúan y pueden ser descritas informacionalmente a través de estos esquemas, es decir, los artefactos pueden ser definidos como subprocesos que integran un proceso de negocio principal, los cuales pueden estar relacionados internamente.

Un método para descubrir de forma automática los esquemas de artefactos es presentado por (V. B. Dongen et al., 2012). Los autores plantean a partir de una fuente de datos estructurada, obtener un esquema que luego es particionado en varios esquemas de artefactos utilizando el algoritmo *k-means*. El parámetro *k* determina la cantidad de artefactos a descubrir y debe ser introducido por el usuario manualmente.

(Lu, 2013), propone una técnica basada en identificar un modelo centrado en artefactos, partiendo de un esquema relacional obtenido de una base de datos. Las relaciones existentes en el esquema son extraídas y organizadas con el uso de algoritmos de grafos. Por cada tipo de artefacto encontrado se obtiene un modelo independiente, lo que posibilitaría la realización de análisis para detectar posibles interacciones entre diferentes procesos, así como establecer comparaciones entre estos modelos. El enfoque aplicado sobre los esquemas centrados en artefactos que se menciona anteriormente constituye uno de los resultados más prometedores en este tema.

5. Conclusiones

Para desarrollar esta investigación se comprobó el estado del arte sobre las técnicas de descubrimiento de procesos mediante varios procedimientos y se detectaron dos desafíos vigentes para la aplicación de los algoritmos de esta etapa: el descubrimiento de modelos de procesos en entornos donde diversas organizaciones colaboran y la aplicación de las técnicas de minería de procesos sobre bases de datos relacionales.

Se han presentado algunos trabajos relacionados con estas temáticas, sin embargo, ninguno integra las dos problemáticas planteadas anteriormente. Esto trae como consecuencia que, a pesar de los avances obtenidos, la aplicación práctica de los algoritmos se ve limitada.

Como trabajos futuros se pretenden fusionar estas técnicas y conformar una estrategia que permita descubrir modelos de procesos capaces de representar subprocesos que interactúan entre sí. Al modelo descubierto se le podría aplicar posteriormente técnicas convencionales para analizar, predecir y proponer mejoras en el comportamiento del proceso global.

RECEIVED: JANUARY, 2017

REVISED: MARCH, 2017

Referencias Bibliográficas

- Alves de Medeiros, A. K. (2006). Genetic Process Mining. Eindhoven University of Technology.

- Alves de Medeiros, A. K., Weijters, A. J. M. M., & van der Aalst, W. M. P. (2007). Genetic Process Mining: An Experimental Evaluation. *DataMining and Knowledge Discovery*. Eindhoven University of Technology.
- Chen, Q., & Hsu, M. (2001). *Inter-Enterprise Collaborative Business Process Management. Proc. International Conference on Data Engineering (ICDE)*. Germany.
- Dayal, U., Hsu, M., & Ladin, R. (2001). Business Process Coordination: State of the Art, Trends, and Open Issues. In *VLDB* (Vol. 1, pp. 3–13). Retrieved from <http://www.vldb.org/conf/2001/P003.pdf>
- Deevakar, R. (2015). Process mining of medication revisions in electronic health records.
- Di Ciccio, C., Macella, M., Scannapieco, M., Zardetto, D., & Catarci, T. (2011). MailOfMine - Analyzing Mail Messages for Mining Artful Collaborative Processes. *1st International Symposium on Data-Driven Process Discovery and Analysis SIMPDA*. Italy.
- Díaz Vasallo, A. (2014). Algoritmo genético distribuido para el descubrimiento de procesos.
- Dongen, V. B., Fahland, D., & Nooijen, E. H. (2012). Automatic Discovery of Data-Centric and Artifact-Centric Processes.
- Fahland, D., Leoni, M. D., Dongen, V. B., & van der Aalst, W. M. P. (2011). Behavioral Conformance of Artifact-Centric Process Models.
- Gunter, C. W., & van der Aalst, W. M. P. (2007). Fuzzy Mining: Adaptive Process Simplification Based on Multi-Perspective Metrics.
- Jagadeesh, R. P., Bose, C., & van der Aalst, W. M. P. (2012). When process mining meets bioinformatics. Berlin.
- López Pintado, O. (2015). Refinamiento Iterativo de Matrices Causales.
- Lu, X. (2013). Artifact-Centric Log Extraction and Process Discovery.
- Lu, X., Nagelkerke, M., van de Wiel, D., & Fahland, D. (2015). Discovering interacting artifacts from ERP systems (extended version). Retrieved from <https://pdfs.semanticscholar.org/9f03/31ad41e9851c04730b60ba3bff3964634703.pdf>
- Mans, R. S., Schonenberg, M. H., Song, M., van der Aalst, W. M. P., & Bakker, P. J. M. (2009). Application of process mining in healthcare a case study in a dutch hospital. Berlin.

- Peleg, M., Yeh, I., & Altman, R. B. (2002). Modelling biological processes using work ow and petri net models.
- van der Aalst, W. M. P. (2000). Process-oriented Architectures for Electronic Commerce and Interorganizational Workflow. Information Systems.
- van der Aalst, W. M. P. (2011). *Process Mining. Discovery, Conformance and Enhancement of Business Processes*. Berlin: Springer.
- van der Aalst, W. M. P. (2015). Intra- and Inter-Organizational Process Mining: Discovering Processes Within and Between Organizations. Eindhoven University of Technology.
- van der Aalst, W. M. P., A. J.M.M., W., & Maruster, L. (2004). *Workflowmining: Discovering process models from event logs*. *IEEE Transactions on Knowledge and Data Engineering*.
- van der Aalst, W. M. P., Adriansyah, A., & Alves de Medeiros, A. K. (2011). Manifiesto sobre Minería de Procesos. IEEE Task Force on Process Mining. Retrieved from <http://www.win.tue.nl/ieeetfpm>
- van Dongen, B. F. (2005). The ProM Framework: A New Era in Process Mining Tool Support.
- van Eck, M. L., Sidorova, N., & van der Aalst, W. M. (2016a). Composite State Machine Miner: Discovering and Exploring Multi-perspective Processes. Retrieved from http://www.processmining.org/_media/blogs/pubs2016/compositestatemachineminer.pdf
- van Eck, M. L., Sidorova, N., & van der Aalst, W. M. (2016b). Discovering and Exploring State-Based Models for Multi-perspective Processes. In *International Conference on Business Process Management* (pp. 142–157). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-3-319-45348-4_9
- Weijters, A. J. M. M., van der Aalst, W. M. P., & Alves de Medeiros, A. K. (2006). Process Mining with Heuristics Miner Algorithm. BETA Working Paper Series. Eindhoven University of Technology.
- Zhao, X., & Liu, C. (2006). Tracking over Collaborative Business Processes. In *Bussiness Process Management: 4th International Conference, BPM* (pp. 30–48). Vienna: Springer. Retrieved from https://books.google.com/books?id=3Yh9bdxeW1UC&pg=PA47&lpg=PA47&dq=discovery+collaborative+business+process+management%2Balgorithms&source=bl&ots=UqwwXCSSK9&sig=jyF-9t0VlmwIqCM_XzBrcUEvuqk&hl=es-419&sa=X&sqi=2&ved=0ahUKEwiVxpjNhJPSAhVq6oMKHdbBD0wQ6AEIaDAJ#v=onepage&q=discovery%20collaborative%20business%20process%20management%2Balgorithms&f=false



El artículo que se publica es de exclusiva responsabilidad de los autores y no necesariamente reflejan el pensamiento de la Revista Ciencia Digital.

El artículo queda en propiedad de la revista y, por tanto, su publicación parcial y/o total en otro medio tiene que ser autorizado por el director de la Revista Ciencia

