

Análisis bibliométrico de la producción científica sobre Minería de Procesos registrada en Scopus en el período 2012-2016

Bibliometric analysis of scientific production about Process Mining registered in Scopus in the period 2012-2016

Sady Carina Fuentes Reyes
Waldo Pérez García

RESUMEN

La minería de procesos es una poderosa técnica para la gestión y la inteligencia empresarial que en los últimos años ha despertado interés en la comunidad científica, notándose un incremento de las investigaciones en esta área del conocimiento. En el presente trabajo la metodología que se aplica abarca un conjunto de indicadores bibliométricos que permitió cuantificar, visualizar y evaluar los resultados de la producción científica sobre minería de procesos. A partir del análisis realizado en la base de datos Scopus en el período 2012 - 2016 se pudo evidenciar que el sector académico representado por las universidades es el más destacado, el continente europeo es el que representa el núcleo de investigaciones científicas, siendo al respecto, Holanda el país con mayores resultados de investigación en el tópico examinado. Las ponencias o conferencias son las que más sobresalen respecto a otras salidas de investigación y el año 2015 se destaca como el más productivo. En América, Estados Unidos es el país líder. Este estudio nos permite concluir que en los países altamente industrializados existe una sostenible producción científica sobre el tema en cuestión y que en América Latina y el Caribe comienza emerger una tendencia a incluir como estrategia de gestión de procesos de negocio la Minería de Procesos.

Palabras clave: minería de procesos; producción científica; bibliometría; indicadores bibliométricos

ABSTRACT

Process mining is a powerful technique for management and business intelligence that in recent years has aroused interest in the scientific community noticing an increase in research in this knowledge area. In the present work, the applied methodology includes bibliometric indicators set that allowed to quantify, visualize and evaluate the scientific production results on process mining. From the analysis carried out in the Scopus database in the period 2012 - 2016 it was possible to show that the academic sector represented by the universities is the most outstanding, the European continent is the one that represents the scientific research core, being in this respect, Holland the country with the highest research results in the examined topic. The papers or conferences are determined as most outstanding related to other research outputs and the year 2015 is outstanding as the most productive. In America, the United States is the leading country. This study allows us to conclude that in highly industrialized countries there is a sustainable scientific production on the subject in question and that in Latin America and the Caribbean begin to emerge a tendency to include the process mining as a strategy for business Process Management.

Keywords: process mining; scientific production; bibliometrics; bibliometric indicators

Introducción

Debido al crecimiento exponencial de la información y la necesidad de reaccionar rápidamente dentro de un entorno competitivo en la toma de decisiones, las empresas requieren de estrategias que permitan la mejora continua de los servicios dispuestos a partir de las tecnologías de la información. En este sentido son variadas las herramientas y técnicas

informáticas que utilizan los decisores en el entorno empresarial que contribuyen en gran medida al proceso de toma de decisión y la correcta gestión de los procesos de negocio. Entre las técnicas más novedosas se destacan las de minería: minería de datos, texto, web y de procesos. (Gregorio, 2008).

En la actualidad los procesos de negocio de las organizaciones están soportados por sistemas de información que registran datos

valiosos con respecto a ejecutores, actividades, eventos, tiempos y variables asociadas a la ejecución de los procesos.

La Minería de Procesos (MP) es una herramienta importante para las organizaciones modernas que necesitan gestionar procesos operacionales no triviales. Por un lado, hay un increíble crecimiento en la cantidad de datos de eventos y por otro, los procesos y la información necesitan estar perfectamente alineados para cumplir requerimientos relacionados con el cumplimiento de normas, eficiencia y servicio al cliente (Van der Aalst et al., 2012).

Van Der Aalst (2011) define la MP como una disciplina que tiene el objetivo de descubrir, monitorear y mejorar procesos de negocio a través de la extracción de conocimiento del registro de eventos de los sistemas de información.

Entre las principales ventajas de la minería de procesos para las organizaciones se pueden citar:

- Descubrir el modelo de ejecución real del proceso.
- Determinar si el proceso cumple con la reglamentación y procedimientos documentados.
- Analizar la interacción del personal que ejecuta el proceso.
- Descubrir cuellos de botella.
- Monitorear la productividad del personal.
- Predecir el tiempo de ciclo de un caso.
- Determinar la relación entre las variables de un caso.

Incrementar el entendimiento de la MP y su utilización por parte de no expertos es uno de los once retos identificados en el Manifiesto de la Minería de Procesos (Van der Aalst et al., 2012) escrito por los miembros y personas que respaldan la Fuerza de Trabajo sobre Minería de Procesos (IEEE-TFPM, Institute of Electrical and Electronics Engineers - Task Force on Process Mining, *por sus siglas en inglés*).

En Cuba la minería de procesos es un área del conocimiento en exploración. Siendo la Universidad de Ciencias Informática la líder en las investigaciones en esta temática. Entre las líneas de investigaciones desarrolladas se destacan:

- Análisis y transformación de trazas
- Algoritmos para la detección de subprocesos
- Modelos para detectar variabilidad en la ejecución de procesos

Estas investigaciones se han aplicado fundamentalmente en Sistemas hospitalarios, Sistema de Negocios y el Sistema Único de Identificación Nacional.

Atendiendo a la importancia que tiene la MP como herramienta que apoya la toma de decisiones y la gestión empresarial el presente estudio se propone realizar un análisis bibliométrico que permita conocer el comportamiento de la producción científica en esta área del conocimiento.

Materiales y Métodos

Para el análisis de la producción científica sobre minería de procesos, a partir de un marco teórico conceptual sobre los aspectos metodológicos

y las relaciones interdisciplinarias de los estudios métricos de la Información (Spinak, 1996), se analizaron las salidas de investigación representadas en: artículos publicados en revistas científicas, ponencias o conferencias publicados en memorias de eventos y capítulos de libros.

Atendiendo al amplio cubrimiento temático se utilizó como fuente de información la base de datos bibliográfica Scopus, lo que permitió tener una visión de la producción global sobre el tema en cuestión. Se realizó una descarga directa, utilizando como estrategia de búsqueda la frase exacta «process mining». La recuperación contempló un periodo de 5 años, desde el 2012 hasta el 2016. El fichero recuperado fue importado hacia el gestor de referencias bibliográficas EndNote X7.

Se obtuvieron un total de 910 registros. Posteriormente se procedió a la normalización, con vistas a homogenizar el nombre de los autores, la afiliación de los artículos, el país de las instituciones colaboradoras y la clasificación temática de los artículos.

Se aplicó la Ley de Lotka (Lotka, 1926; Urbizagástegui Alvarado, 1999) para agrupar a los autores por nivel de productividad, y para el análisis de productividad por sectores se tuvo en cuenta la filiación autoral y se relacionaron las instituciones con sectores específicos.

Para graficar los resultados se utilizaron el Microsoft Excel 2016, Tableau 10.3, Bibexcel Versión 2009, VosViewer 1.6.5, Ucinet 6 y NetDraw 2.38.

Indicadores analizados

- Producción total
- Producción por años
- Producción por autores
- Colaboración autoral
- Producción por entidad y sector
- Producción por publicaciones
- Producción por continentes y países
- Palabras clave más representativas
- Co-ocurrencia de términos en el resumen y en el título

Resultados y Discusión

Productividad por años

Se analizó el período comprendido entre 2012 - 2016. En la figura 1 muestra el comportamiento de este indicador.

Es evidente el aumento sostenido en los últimos tres años, siendo el año 2015 el más productivo con un total de 240 contribuciones.

A partir del año 2012 se produce un número significativo de contribuciones, se infiere que este incremento puede estar dado porque en el año 2011 aparece la versión original del Manifiesto sobre Minería de Procesos en los BPM (*Business Process Mining, por sus siglas en inglés*) *Workshops proceedings, Lecture Notes in Business Information Processing* y Springer-Verlag, es un manifiesto escrito por los miembros y personas que respaldan la IEEE-TFPM. El objetivo de esta fuerza de trabajo es promover la investigación, desarrollo, educación, implementación, evolución y

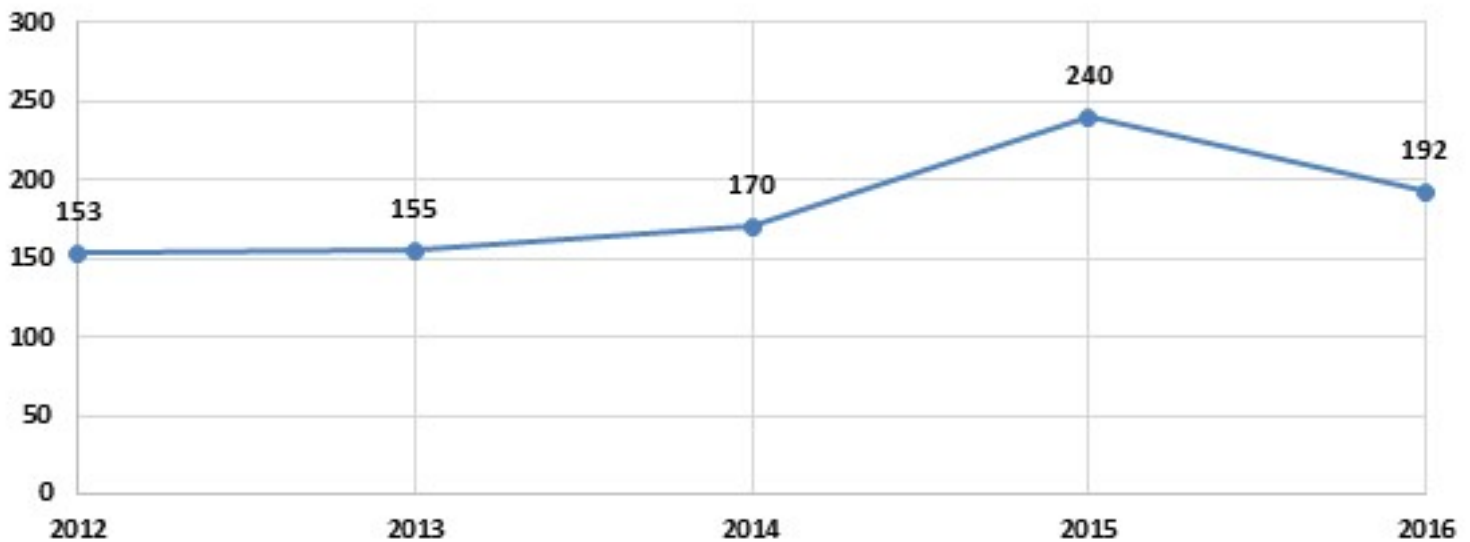


Figura 1. Cantidad de contribuciones por año

Fuente: elaboración propia

entendimiento acerca de la minería de procesos. (Van der Aalst et al., 2012).

Autores más destacados

Un total de 1627 autores fueron responsables de la producción científica sobre minería de procesos durante el período analizado. Existe una marcada tendencia a la autoría múltiple con el 95 % de las contribuciones bajo esta categoría. El mayor por ciento de las investigaciones se publicó bajo la autoría de investigadores de diferentes países, lo que denota una alta colaboración entre científicos de diferentes países.

El autor más productivo con un total de 126 contribuciones es el holandés Wil van der Aalst de la Universidad Tecnológica de Eindhoven, seguido con 33 contribuciones por Fabrizio Maggi de la Universidad de Tartu de Estonia.

Wil van der Aalst es profesor titular del Departamento de Matemáticas Informáticas de la Technische Universiteit Eindhoven, donde preside el grupo de Arquitectura de Sistemas de Información y es director científico del Data Science Center.

Entre las principales líneas de investigación de Wil van der Aalst se destacan: la gestión de flujo de trabajo, minería de procesos, redes de Petri, gestión de procesos de negocio, modelado de procesos y análisis de procesos. Sobre estas temáticas el autor ha publicado 200 artículos de revistas, 20 libros (como autor o redactor), 450 publicaciones de conferencias y 65 capítulos de libros.

La tabla 1 muestra los autores que originaron 20 o más contribuciones. En ella se ratifica la tendencia a la autoría múltiple y a la colaboración entre autores de diferentes entidades y países.

Wil Van der Aalst es el autor que tiene mayor cantidad de contribuciones como autor principal y mayor cantidad como autor individual.

De los ocho autores más productivos tres pertenecen a la Universidad Tecnológica de Eindhoven de Holanda, entidad que más ha investigado sobre minería de procesos.

La aplicación de la Ley de Lotka, permitió agrupar los autores por niveles de productividad. En la tabla 2 se muestran los niveles de productividad. (Lotka, 1926).

Tabla 1. Autores más productivos

Fuente: elaboración propia

Autores	Total de contribuciones	Primer autor	Autoría individual
Wil van der Aalst	126	23	13
Fabrizio Maggi	34	11	1
Wichian Premchaiswadi	28	2	
Jan Mendling	26		
Massimiliano de Leoni	24	10	
Boudewijn van Dongen	24	2	
Jan Vanthienen	21		
Josep Carmona	20	2	

Tabla 2. . Niveles de productividad

Fuente: elaboración propia

Niveles	Cant. de autores	%
Grandes Productores	31	2%
Medianos Productores	238	15%
Pequeños Productores	1358	83%
Total	1627	100%

Del total de 1627 autores, todos aquellos que tienen 10 o más contribuciones, se clasifican como grandes productores. En el caso del holandés Wil van der Aalst tiene una producción científica muy superior a los autores ubicados en la categoría de Grandes Productores.

Colaboración autorial

En la figura 2 se muestra, una red de colaboración autorial que representa los autores con más de cinco contribuciones (grandes y medianos productores).

En la red se puede observar que existe una alta colaboración entre autores y se visualiza, además, las colaboraciones más fuertes a través del grosor de las líneas que une a los diferentes nodos. En el período analizado destacan los holandeses Van Der Aalst y Van Dongen con 20 contribuciones realizadas en conjunto, ambos autores pertenecen a la Universidad Tecnológica de Eindhoven.

La figura 3 muestra con mayor claridad la relación de colaboración de los autores más productivos (con más de 10 contribuciones).

La gráfica muestra la colaboración entre los autores más productivos,

destacándose en el centro de la red a Massimiliano de Leoni profesor de Sistema de Información en la Universidad Tecnológica de Eindhoven como el autor más colaborativo. Massimiliano de Leoni tiene una fuerte colaboración con Hajo A. Reijers (6 contribuciones) profesor de Ciencia de la Computación de la Universidad de Amsterdam.

Sobresale la colaboración autorial entre Josep Carmona de la Universidad Politécnica de Cataluña y Jorge Muñoz-Gama de la Pontificia Universidad Católica de Chile con 9 contribuciones. Wichian Premchaiswadi ocupa el tercer lugar de autores más productivo con 28 contribuciones bajo autoría múltiple, sin embargo, no tiene colaboración con los grandes productores.

Países más productivos

En relación a la distribución geográfica de la producción científica la tabla 3 muestra los países más productivos, con veinte o más contribuciones, 66 países son responsables de la producción científica en esta área del conocimiento en el periodo analizado.

El continente europeo produce más de la mitad de las contribuciones (58 %) sobre minería de procesos, seguido de Asia (22%) y América (11 %) y en menor medida se sitúan Australia (6%) y África (1%). El Medio Oriente representado por Irán, Israel, Arabia Saudita, Emiratos Árabes Unidos y Turquía cuenta con 23 contribuciones y Oceanía con 4 de Nueva Zelanda.

El país más productivo es Holanda con 164 contribuciones, seguido de Italia con 111 y China y Alemania con 81. Entre los países más productivos 10 pertenecen al continente europeo, lo que identifica a este como el continente más prolífero. En América se destaca


Figura 2. Red de colaboración autorial

Fuente: elaboración propia

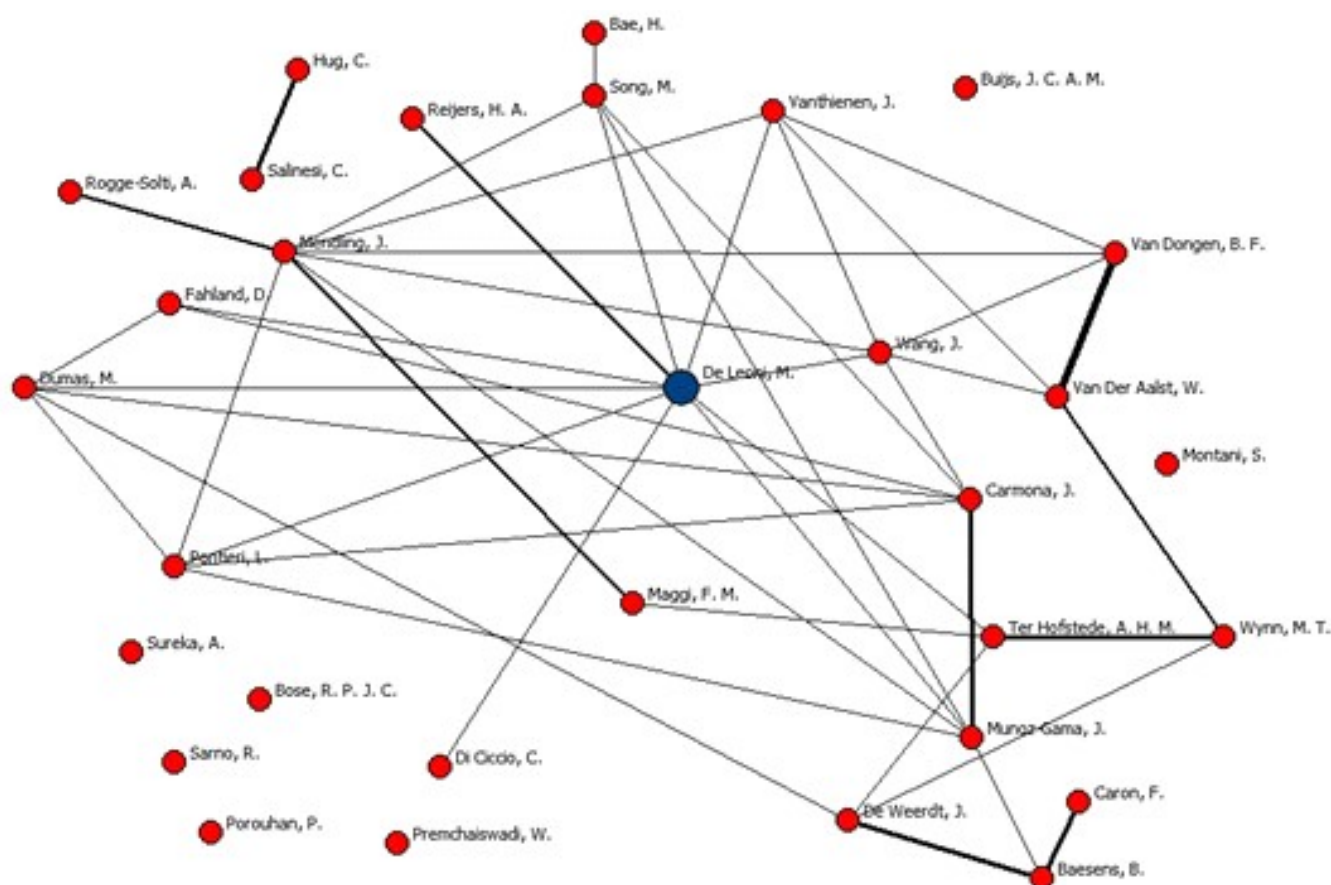


Figura 5. Red de colaboración entre autores más productivos

Fuente: elaboración propia

Tabla 3. Países más productivos

Fuente: elaboración propia

Países	Cant. de contribuciones
Holanda	164
Italia	111
China	81
Alemania	81
Austria	59
Australia	58
Estados Unidos	51
España	49
Bélgica	46
Francia	44
Estonia	40
Corea de Sur	35
Reino Unido	35
Tailandia	31
India	30
Brasil	26
Rusia	20

Estados Unidos como el país más representativo en el área con 51 contribuciones seguido de Brasil con 26.

En el continente americano nueve países son los responsables de la producción científica en esta área del conocimiento, Cuba ocupa

quinto lugar por encima de Argentina, Colombia, Ecuador y México. De las 6 contribuciones de Cuba cinco pertenecen a la Universidad de Ciencias Informáticas (UCI), una de ellas en colaboración con el Instituto Nacional de Astrofísica, Óptica y Electrónica de Puebla, México. El Instituto Superior Politécnico José Antonio Echevarría cuenta con una publicación en colaboración con la Universidad de Sao Paulo de Brasil.

Aunque España ocupa el octavo lugar de los países más productivos, en Iberoamérica tiene el liderazgo con 49 contribuciones que corresponden a 19 entidades, de ellas, trece universidades, destacándose la Universidad Politécnica de Cataluña con 21 aportes.

Entidades y sectores más productivos

Sobre Minería de procesos en el periodo analizado publicaron más de 600 entidades sobresaliendo la Universidad Tecnológica de Eindhoven de Holanda. La tabla 4 muestra las siete entidades más productivas con más de veinte contribuciones.

En 2017, la World University Rankings colocó a la Universidad Tecnológica Eindhoven en el puesto 177 internacionalmente y en el lugar 64 en el mundo de Ingeniería y Tecnologías de la Información. Esta entidad tiene entre sus principales líneas de investigación los sistemas automotrices, tecnologías de telecomunicaciones de banda ancha, sistemas de información y comunicación, nanoingeniería, plasmas y polímeros entre otros. Las siete entidades más productivas son universidades, cinco pertenecen al continente europeo y una al asiático.

Tabla 4. Entidades más productivas

Fuente: elaboración propia

Entidades	País	Cant. de contribuciones
Universidad Tecnológica de Eindhoven	Holanda	155
Universidad de Tartu	Estonia	40
Universidad de Queensland	Australia	32
Universidad de Siam	Tailandia	28
Universidad de Leuven	Bélgica	27
Universidad Politécnica de Cataluña	España	21
Universidad de Viena	Austria	21

El mayor por ciento de las investigaciones corresponden al sector Educación (80%) que encierra las investigaciones que se realizan desde instituciones relacionadas directamente con este sector y el cual está representado significativamente por universidades. Sectores como: el Empresarial, Salud, Ciencia y Técnica entre otros tienen escasa productividad que en su conjunto solo representan un 20%.

Publicaciones más productivas

El núcleo Bradford está integrado por tres publicaciones alemanas: *Lecture Notes in Business Information Processing* seguido de

Lecture Notes in Computer Science ambas editadas por Springer Verlag. El tercer lugar lo ocupa la revista *Information Systems* del Reino Unido. Un por ciento significativo de contribuciones se centra en estas publicaciones, lo que infiere que ocupan la preferencia de los autores para divulgar los resultados de investigación sobre MP.

La figura 4 muestra las publicaciones que en el período analizado tienen diez o más contribuciones.

Lecture Notes in Business Information Processing informa de los resultados más avanzados en áreas relacionadas con los sistemas de información empresarial y el desarrollo de software de aplicación industrial. Tiene un Índice H de 27 y publica acta de eventos, informes de proyectos, tutoriales, conferencias, seminarios, tesis premiadas entre otros materiales. Según *Scimago Journal Report* en el periodo comprendido entre 2009 y 2016 esta publicación tiene un ranking de 0.22, la cantidad de citas por documentos es de 0.52 y el total de citas es de 992.

Lecture Notes in Computer Science tiene un Índice H de 251 y publica resultados de todas las áreas de investigación, desarrollo y educación en informática y tecnología de la información. Según *Scimago Journal Report* en el periodo comprendido entre 2009 y 2016 esta publicación tiene un ranking de 0.32, la cantidad de citas por documentos es de 0.67 y el total de citas es de 44840.

Information Systems es una revista del cuartil 1 con un Índice H de 68 y publica artículos sobre diseño e implementación de lenguajes, modelos de datos, algoritmos, software y hardware para sistemas de información.

Según *Scimago Journal Report* en el periodo comprendido entre 2009 y 2016 esta publicación tiene un ranking de 0.91, la cantidad de citas por documentos es de 3.57 y el total de citas es de 911.

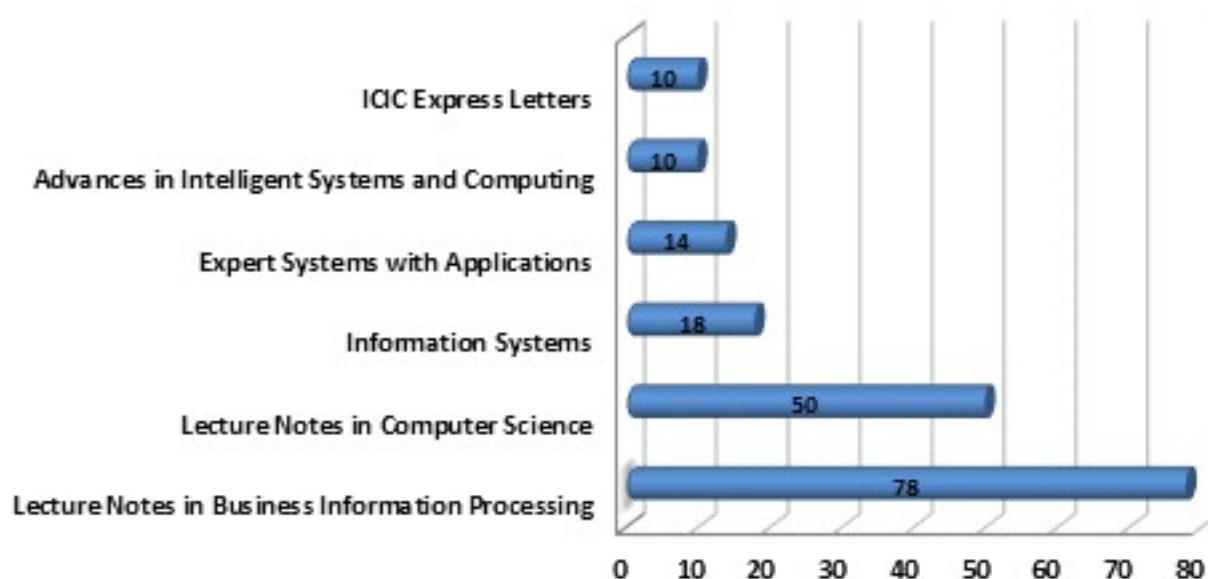


Figura 4. Publicaciones más productivas

Fuente: elaboración propia

Frecuencia de aparición de palabras clave

Del análisis de las palabras clave se obtuvieron un total de 1746 palabras, de ellos, 1368 términos aparecen una sola vez y los que más se destacan con más de 20 repeticiones son: *Process Mining* (590), *Process Discovery* (57), *Business Process Management* (44), *Conformance Checking* (38), *Data Mining* (36), *Petri Nets* (28) y *Event Log*, *Clustering*, *Event Logs* (21). En criterio de los autores, las palabras claves asignadas en ocasiones son poco abarcadoras y no describen con exactitud el contenido de las investigaciones, lo cual limita el análisis y recuperación de información.

La significativa representatividad del término *Process Mining* en el campo palabras claves (keywords) nos indica la efectividad de la recuperación partiendo que en la estrategia de búsqueda se utilizó la frase exacta («process mining»).

Co-ocurrencia de términos en el resumen y en el título

Al realizar el conteo de las palabras más frecuentes en el resumen y el título, se puede constatar que los términos más frecuentes en los resúmenes fueron: «process», «event log», «model», «process model» y «data». Estos términos también pueden ser considerados tópicos consolidados o más abordados en este período de tiempo. De manera similar en los títulos se destacaron los términos: «process model», «event log», «business process».

A través del análisis de los términos del resumen y del título se aprecia coincidencia en los que más se destacan. La diferencia entre ellos son los frentes de investigación sobre minería de procesos, los cuales, de cierta manera, coinciden con los planteados por la IEEE-TFPM en el Manifiesto sobre Minería de Procesos (Van der Aalst, W., et al., 2012) y sobresaliendo:

- Estudios relacionados con encontrar, fusionar y limpiar datos de eventos.
- Mejorar el sesgo representacional utilizado para el descubrimiento de procesos.
- Proporcionar soporte operacional
- Combinar la minería de procesos con otros tipos de análisis
- Mejorar la usabilidad y entendimiento para los no expertos

Conclusiones

En el año 2011 apareció la versión original del Manifiesto sobre Minería de Procesos en los BPM 2011 Workshops proceedings, Lecture Notes in Business Information Processing y Springer-Verlag, que tiene entre sus principales objetivos promover la investigación, desarrollo, educación, implementación, evolución y entendimiento acerca de la minería de procesos, y es precisamente a partir de este año donde se visualizó un incremento sostenido en la producción científica sobre esta área del conocimiento.

Se identificó un marcado comportamiento a la autoría múltiple y una fuerte colaboración científica internacional en temas de investigación relacionada con la minería de procesos, liderada por el sector académico, lo que garantiza se propague con inmediatez el uso de esta herramienta.

Las entidades más productivas corresponden al sector académico, siendo la Universidad Tecnológica de Eindhoven la que más se destaca en salidas de investigaciones sobre el tema en cuestión, por lo que se infiere que la minería de procesos es un área de consolidación con un crecimiento exponencial en sus contribuciones. Atendiendo a las ventajas que tiene esta técnica en el contexto organizacional sería beneficioso que otros sectores de la economía implementen su uso.

En los países industrializados existe una sostenida producción científica sobre minería de procesos liderado por Holanda. En América, Estados Unidos es el país más productivo, pero comienza a visualizarse un aumento en las investigaciones sobre este tema en otros países del área.

Referencias

- Gregorio Chaviano, O. (2008). Aplicaciones y perspectivas de los estudios métricos de la información (EMI) en la gestión de información y el conocimiento en las organizaciones. AIBDA, 29(1-2), 1-27.
- Lotka, A. J. (1926). The frequency distribution of scientific productivity. Journal of the Washington Academy of Sciences, 16(12), 317-323.
- Spinak, E. (1996). Diccionario enciclopédico de Bibliometría, Cienciometría e Informetría. Caracas: UNESCO, 245.
- Urbizagástegui alvarado, R. (1999). La Ley de Lotka y la literatura de bibliometría. Investigación Bibliotecológica, 3(27), 17.
- Van der Aalst, W. M. P. (2011). Process mining: discovery, conformance and enhancement of business processes. Springer Science & Business Media.
- Van der Aalst, W. M. P., Adriansyah, A., Alves de Medeiros, A. K., & et, al. (2012). Process mining manifesto. Paper presented at the 9th International Conference on Business Process Management, BPM 2011P, Clermont-Ferrand.

Recibido: 11 de septiembre de 2018
Aprobado en su forma definitiva:
1 de enero de 2019

Sady Carina Fuentes Reyes
Instituto de Información Científica y
Tecnológica (IDICT), La Habana, Cuba.
Correo-e.: sady@idict.cu

Waldo Pérez García
Universidad Central «Marta Abreu» de Las
Villas (UCLV), La Habana, Cuba.
Correo-e.: waldop@uclv.edu.cu
