

# Caracterización de la producción científica en el área disciplinar de la minería de proceso

**Sady Carina Fuentes Reyes\***

**Adonis Domínguez Castro\***

**Waldo García Pérez\*\***

**Patricia Romero Lazcano\***

**Liliet Leyva Pérez\***

*Artículo recibido:*

*3 de marzo de 2018*

*Artículo aceptado:*

*29 de octubre de 2018*

*Artículo de investigación*

## RESUMEN

La minería de proceso 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 proceso. A partir del análisis realizado en la base de datos Scopus en el periodo 2002-2017 se pudo evidenciar que el sector educativo

\* Instituto de Información Científica y Tecnológica (IDICT), Cuba  
sady@idict.cu  
adonis@idict.cu  
patricia.romero@idict.cu  
liliet.leyva@idict.cu

\*\* Universidad Central "Marta Abreu" (UCLV), Cuba  
waldop@uclv.edu.cu

representado por las universidades es el más destacado, y el continente europeo es el que representa el núcleo de investigaciones científicas, con Holanda como el país con mayores resultados de investigación en el tópico examinado. En relación a la tipología documental se determina que fue la publicación de artículos científicos la que sobresale 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 a emerger una tendencia a incluir como estrategia de gestión de procesos de negocio la minería de proceso.

**Palabras clave:** Minería de Proceso; Producción Científica; Bibliometría; Indicadores Bibliométricos

### **Characterization of scientific production in the disciplinary area of Process Mining**

*Sady Carina Fuentes Reyes, Adonis Domínguez Castro, Waldo García Pérez, Patricia Romero Lazcano and Liliet Leyva Pérez*

#### **ABSTRACT**

Process mining is a powerful technique for management and business intelligence that in recent years has aroused interest in the scientific community, where there has been an increase in research in this knowledge area. In the present work, the applied methodology includes a set of bibliometric indicators that allowed to quantify, visualize and evaluate the scientific production results on process mining. Thanks to the analysis realized in the Scopus database during the period 2002-2017, it was possible to show that the education sector represented by the universities is the most outstanding, and that the European continent is the one that incarnates the scientific research core, being in this respect Holland the country with the highest research in the examined topic. Regarding the documentary typology, it was the publication of scientific articles published which stood out with respect to other research outputs, and the year 2015 excels 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 begins to emerge a tendency which includes 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 a la necesidad de reaccionar rápidamente dentro de un entorno competitivo en la toma de decisiones (Carneiro, 2009; Ochoa Reyes, Orellana García y Lizama Mué, 2014), las empresas requieren de estrategias que permitan la mejora continua de los servicios dispuestos a partir de la gestión de las tecnologías de la información (Shrestha *et al.*, 2014). 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 (Mora *et al.*, 2013). Entre las técnicas más novedosas destacan las de minería: minería de datos, texto, web y de procesos.

La minería de proceso destaca por su capacidad de reflejar a través de modelos la realidad empresarial, ya sea con carácter predictivo o auditable. A través de la extracción de datos utilizando esta técnica, donde los *log* son una pieza fundamental de información, se pueden identificar las rutas críticas de los procesos dentro de la entidad y de esta manera apoyar a los directivos en su labor de gestión (Van der Aalst *et al.*, 2012). Es la más joven de las técnicas mencionadas y es la que tiene una visualización fuerte en los procesos de la empresa (Orellana García, Larrea Armenteros y Pérez Alfonso, 2014). Se ubica entre la inteligencia computacional y la minería de datos, por una parte, y la modelación y análisis de procesos, por otra (Van der Aalst, 2011). Surge con el objetivo de analizar los registros de eventos destinados a recopilar las trazas de los procesos que ocurren en los sistemas de información actuales en busca de errores, inconsistencias y vulnerabilidades.

Van der Aalst (2011) define la minería de proceso como una disciplina que tiene como objetivo descubrir, monitorear y mejorar procesos de negocio

a través de la extracción de conocimiento del registro de eventos de los sistemas de información. Entre sus principales ventajas para las organizaciones se pueden citar las siguientes:

- 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.

El descubrimiento de proceso, la verificación de conformidad y el mejoramiento de modelos son los tres tipos fundamentales de minería de proceso identificados por Van der Aalst (2011). La *Figura 1* muestra el posicionamiento de estas técnicas.

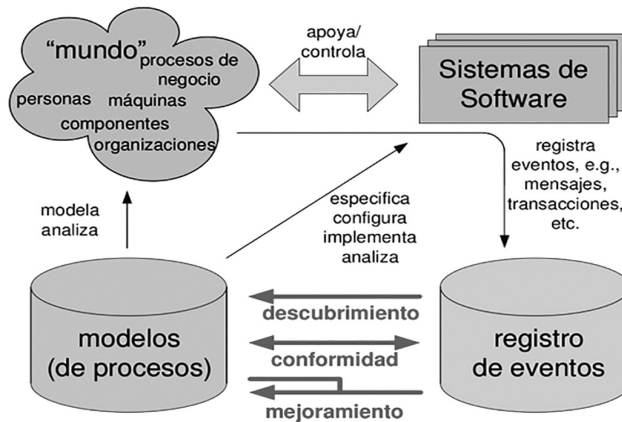


Figura 1. Tipos básicos de minería de proceso (Van der Aalst *et al.*, 2012)

*Descubrimiento.* Se usa un registro de eventos para producir un modelo sin usar a *priori*. El descubrimiento de procesos es la técnica de minería de proceso más destacada. Para muchas organizaciones es sorprendente ver que las técnicas existentes son realmente capaces de descubrir los procesos reales meramente basados en las muestras de ejecución de los registros de eventos.

Entre los principales algoritmos para el descubrimiento se encuentran:

- Fuzzy Miner
- Alpha Miner
- Heuristic Miner
- Genetic Miner

*Verificación de conformidad.* Necesitan un registro de eventos y un modelo como entrada. La salida consiste en información de diagnóstico que muestra las diferencias y elementos en común entre el modelo y el registro de eventos. Aquí se compara un proceso existente con un registro de eventos del mismo proceso, para verificar si la realidad, según el registro, es equivalente al modelo y viceversa.

*Mejoramiento.* Se busca extender o mejorar un modelo de proceso existente con la información del proceso real almacenado en un registro de eventos. También necesitan un registro de eventos y un modelo como entrada. La salida es un modelo mejorado o extendido.

En Cuba, la minería de proceso es un área del conocimiento en exploración, en la cual la Universidad de Ciencias Informática (UCI) es la líder en las investigaciones. Las líneas de investigaciones desarrolladas en UCI sobre minería de proceso se centran en tres aspectos:

1. Análisis y transformación de trazas.
2. Algoritmos para la detección de subprocesos.
3. Modelos para detectar variabilidad en la ejecución de procesos.

Estas investigaciones se han aplicado fundamentalmente en sistemas hospitalarios (Mans, Van der Aalst y Vanwersch, 2015; Orellana García *et al.*, 2016; Ochoa Reyes, Orellana García y Lizama Mué, 2014), sistemas de negocios (Yzquierdo Herrera, 2013; Giraldo Mejía, 2016) y el Sistema Único de Identificación Nacional.

Los estudios bibliométricos pertenecen a un campo de investigación interdisciplinario que tiene el potencial de extenderse a casi todos los campos científicos. En el análisis documental realizado no se encontró evidencias de estudios bibliométricos sobre el tema en cuestión; atendiendo a la importancia que tiene la minería de proceso como herramienta que apoya la toma de decisiones y la gestión empresarial, el presente trabajo se propone realizar un estudio bibliométrico sobre el comportamiento de la producción científica que permita conocer el estado del arte de esta área del conocimiento.

## METODOLOGÍA

Para el análisis de la producción científica sobre minería de proceso, a partir de un marco teórico conceptual sobre los aspectos metodológicos y las relaciones interdisciplinarias de los estudios métricos de la información, se estudiaron las salidas de investigación representadas en artículos en revistas científicas, conferencias en eventos y capítulos de libros. El método bibliométrico es un método documental que ha alcanzado un importante desarrollo durante las últimas tres décadas (Ayala Picazo, 2016) y constituye en la actualidad la herramienta esencial para el estudio de la actividad investigadora (Bojo Canales *et al.*, 2004).

Atendiendo a la amplia cobertura temática 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 16 años, desde 2002 hasta marzo de 2017, por lo que se asume que la información referente a este año está incompleta. El fichero recuperado fue importado hacia el gestor de referencias bibliográficas EndNote v17.0.1.7212.

En el estudio se utilizó la metodología para la aplicación de indicadores bibliométricos en la actividad científica (Rodríguez Sánchez, 2012) que cuenta con cuatro fases:

- Fase I. Compilación de la producción científica
- Fase II. Construcción de la base de datos bibliométrica
- Fase III. Procesamiento de la producción científica
- Fase IV. Aplicación de los módulos bibliométricos

Se obtuvieron un total de 1 618 registros y posteriormente se procedió a la normalización. Comúnmente ésta se centra en el campo de adscripción (Hernández García y Navarro Contreras, 2017), en nuestro caso se estandarizaron 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 (Urbizagástegui Alvarado, 1999; Lotka, 1926) para agrupar a los autores por nivel de productividad, mientras que la ley de Bradford (Bradford, 1948) permitió identificar las publicaciones núcleos. La aplicación práctica de la ley de Bradford proporcionó los mecanismos para seleccionar las publicaciones periódicas no sólo más productivas sino también más relevantes para cubrir el área del conocimiento estudiada (Urbizagástegui Alvarado, 2016).

Para graficar los resultados se utilizaron Microsoft Excel 2016, Tableau 10.3, AntCont 3.4.4w y VosViewer 1.6.5. Microsoft Excel 2016, del paquete de Microsoft Office, es ampliamente conocido e ideal para la elaboración de tablas, gráficos, entre muchas otras funciones; Tableau es un programa creado en el Departamento de Ciencias de la Computación de la Universidad Stanford entre los años 1997 y 2002 por el profesor Pat Hanrahan y el estudiante de doctorado Chris Stolte, quien se especializaba en técnicas de visualización, exploración y análisis de bases de datos; lograron obtener una herramienta que de manera sencilla permite visualizar e interactuar con sus datos en un tiempo relativamente corto (Sood *et al.*, 2017).

AntCont es un programa creado por Anthony (2014), muy útil para realizar análisis lingüísticos de conjuntos de textos (corpus lingüístico) en formato .txt. Está compuesto de herramientas (Concordance-Clusters-Collocates-Wordlist) dentro de las cuales hay una serie de instrumentos de análisis y de funciones que permiten, entre otras acciones, elaborar listados de palabras monoléxicas, poliléxicas o polilexemáticas, de agrupamientos léxicos (clústeres), bien de todo el conjunto de textos o bien de una palabra base, y de palabras clave (*keywords*) (Navarro Colorado, 2015).

Finalmente, VOSviewer es una herramienta gratuita que se descarga de la página web de la Universidad de Leiden (Centre for Science and Technology Studies, 2017), con una curva de aprendizaje moderada, una interfaz amigable y de fácil entendimiento; no requiere mayores ajustes para obtener grafos claros y clústeres bien agrupados (Pichuante Escalda, 2016). Sus creadores afirman haber probado las fórmulas predefinidas que modelan la herramienta y le dan la mayor importancia a la posibilidad de representación de grandes mapas bibliométricos de fácil interpretación (Van Eck y Waltman, 2010).

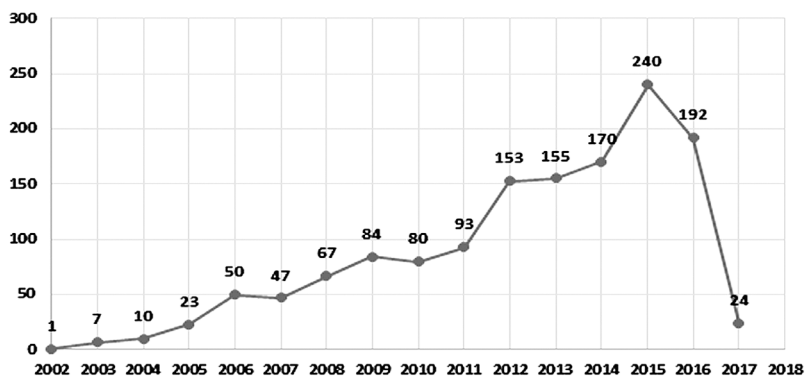
Batería de indicadores utilizados:

- Producción total: total de contribuciones de la muestra.
- Productividad por autores: cantidad de registros por cada autor.
- Producción por años: cantidad de registros por cada año comprendido en el estudio.
- Producción por entidades y sectores.
- Producción por tipología documental.
- Producción por continentes y países: cantidad de registros por continentes y países.
- Palabras clave más representativas: frecuencia de palabras clave.
- Co-ocurrencia de términos en el resumen y en el título.

## RESULTADOS Y DISCUSIÓN

*Productividad por años*

Se analizó el periodo comprendido entre 2002 y marzo de 2017. La *Figura 2* muestra el comportamiento de este indicador.



*Figura 2.* Cantidad de contribuciones por año

Es evidente el aumento sostenido en los últimos años a partir de 2011, siendo 2015 el más productivo con un total de 240 contribuciones. El decrecimiento que se aprecia en 2017 es debido a que las publicaciones recuperadas corresponden a los tres primeros meses del año.

A partir de 2011 se producen 1 027 contribuciones que representan 63 % del total de las contribuciones, se infiere que este incremento puede deberse a que en 2011 aparece la versión original del *Process Mining Manifesto* (Van der Aalst *et al.*, 2012). Este manifiesto fue escrito por los miembros y personas que respaldaban la IEEE Task Force on Process Mining (Fuerza de Trabajo de la IEEE sobre Minería de Proceso), cuyo objetivo es promover la investigación, desarrollo, educación, implementación, evolución y entendimiento acerca de la minería de proceso.

*Autores más destacados*

Un total de 2 426 autores fueron responsables de la producción científica sobre minería de proceso durante el periodo analizado. La mayor parte de los trabajos se publicó bajo la autoría de investigadores de diferentes países; se



observó alta colaboración entre científicos de diferentes países, lo que denota una marcada tendencia a la autoría múltiple con más de 90 % de las contribuciones bajo esta categoría.

El autor más productivo con un total de 286 contribuciones es el holandés Wil Van der Aalst,<sup>1</sup> de la Universidad Tecnológica de Eindhoven, seguido con 53 publicaciones por Boudewijn van Dongen, profesor asistente de Ciencias de la Computación de la misma institución.

Autores	Total de contribuciones	Primer autor	Autoría individual	Últimos años (2012-marzo 2017)
Wil van der Aalst	286	77	39	129
Boudewijn van Dongen	53	12		24
Fabrizio Maggi	37	12	2	33
Josep Carmona	36	7	1	20
Jan Mendling	33			27
Jianmin Wang	33	6		19
Minseok Song	29	4		11
Wichian Premchaiswadi	28	2		28
Jan Vanthienen	28			21
Eric Verbeek	26	11	2	11

Tabla 1. Autores más productivos

En la *Tabla 1* 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 de autoría individual. Todas las contribuciones de Jan Mendling y Jan Vanthienen las realizaron en colaboración con otros autores y en el periodo analizado, en ninguna de ella son autores principales. De los 10 autores más productivos, tres pertenecen a la Universidad Tecnológica de Eindhoven de Holanda, entidad que más ha investigado sobre minería de proceso.

La aplicación de la ley de Lotka permitió agrupar a los autores por niveles de productividad (*Tabla 2*).

1 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 (AIS) y es director científico del Data Science Center. Entre sus principales líneas de investigación se destacan la gestión de flujo de trabajo, minería de proceso, 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.

Número de Artículos o Contribuciones	No. de Autores recuperados en la investigación	No. de Autores esperados según la ley de Lotka	Dif. entre autores esperados y recuperados	Proporción Obtenida	Proporción Esperada	Dif. de las proporciones esperadas y obtenidas
1	1598	1598.00	0.00	65.87%	62.39%	-3.48%
2	404	399.50	-4.50	16.65%	15.60%	-1.05%
3	133	177.56	44.56	5.48%	6.93%	1.45%
4	94	99.88	5.88	3.87%	3.90%	0.02%
5	54	63.92	9.92	2.23%	2.50%	0.27%
6	45	44.39	-0.61	1.85%	1.73%	-0.12%
7	15	32.61	17.61	0.62%	1.27%	0.66%
8	13	24.97	11.97	0.54%	0.97%	0.44%
9	11	19.73	8.73	0.45%	0.77%	0.32%
10	10	15.98	5.98	0.41%	0.62%	0.21%
11	5	13.21	8.21	0.21%	0.52%	0.31%
12	5	11.10	6.10	0.21%	0.43%	0.23%
13	8	9.46	1.46	0.33%	0.37%	0.04%
14	1	8.35	5.15	0.12%	0.32%	0.19%
15	4	7.10	3.10	0.16%	0.28%	0.11%
16	2	6.24	4.24	0.08%	0.24%	0.16%
18	1	4.91	1.91	0.12%	0.19%	0.07%
19	1	4.43	3.43	0.04%	0.17%	0.13%
20	1	4.00	1.00	0.12%	0.16%	0.03%
24	2	2.77	0.77	0.08%	0.11%	0.03%
25	1	2.56	0.44	0.12%	0.10%	-0.02%
26	1	2.36	1.36	0.04%	0.09%	0.05%
28	2	2.04	0.04	0.08%	0.08%	0.00%
29	1	1.90	0.90	0.04%	0.07%	0.03%
33	2	1.47	-0.53	0.08%	0.06%	-0.03%
36	1	1.23	0.23	0.04%	0.05%	0.01%
37	1	1.17	0.17	0.04%	0.05%	0.00%
53	1	0.57	-0.43	0.04%	0.02%	-0.02%
286	1	0.02	-0.98	0.04%	0.00%	-0.04%
<b>TOTALES</b>	<b>2426</b>	<b>2561.23</b>	<b>135.23</b>	<b>100.00%</b>	<b>100.00%</b>	<b>0.00%</b>

Tabla 2. Análisis de la ley de Lotka o ley de crecimiento exponencial

Se pudo comprobar que los autores recuperados en la investigación constituyen 95 % de los esperados según la ley de Lotka, mientras que 5 % restante se ubica en las posiciones de los menos productivos. Se puede afirmar además que hay una distribución desigual de productividad en los autores y que, al menos en la disciplina de minería de proceso, la mayoría publica el menor número de trabajos, mientras que unos pocos autores publican la mayor parte de la bibliografía relevante y forman el grupo más prolífico. A pesar de que la ley de crecimiento exponencial es criticada basándose en que la misma calcula como valor de calidad la cantidad y no necesariamente los autores que más artículos publican son los que más impacto tienen en su campo, en el caso de la minería de proceso los autores que más publican son los de mayor impacto, lo que concuerda con la investigación, con esta ley bibliométrica, enunciada por Alfred Lotka en 1926.

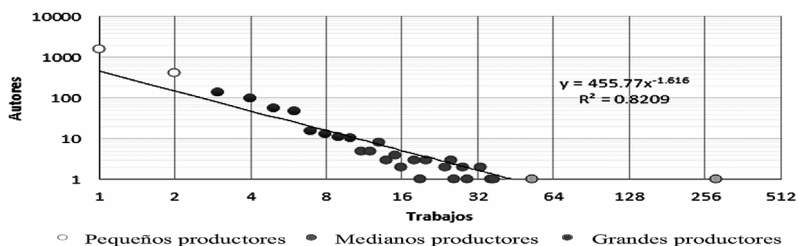


Figura 3. Nivel de productividad de autores



El país más productivo es Holanda, con 352 contribuciones, seguido de China con 178 y los que tienen más entidades responsables de la producción científica en esta área del conocimiento son China (84), Estados Unidos (81) y Alemania (72). Entre los países más productivos ocho pertenecen al continente europeo, lo que identifica a éste como el continente más prolífero, destacándose Holanda y Italia en los últimos años con más de 100 contribuciones.

Como se aprecia en la *Tabla 3*, de los ocho países productivos del continente americano Cuba ocupa el quinto lugar, por encima de Argentina, Colombia y México. De las 10 contribuciones de Cuba siete pertenecen a la 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 tres publicaciones, una de ellas en colaboración con la Universidad de São Paulo, de Brasil.

		Países	Contribuciones	Últimos Años (2012- marzo 2017)	%
Países de Iberoamérica	Continente americano	Estados Unidos	96	53	55.2
		Canadá	23	14	60.8
		Brasil	41	27	65.8
		Chile	20	17	85
		Cuba	10	7	70
		Argentina	4	2	50
		Colombia	2	2	100
		México	3	1	33.3
		Ecuador	1	1	100
				<b>Total</b>	<b>200</b>
		<b>Total</b>	<b>196</b>	<b>125</b>	<b>63.7</b>

*Tabla 3.* Producción científica de los países del continente americano y de Iberoamérica

Aunque España ocupa el octavo lugar de los países más productivos en Iberoamérica tiene el liderazgo con 82 contribuciones que corresponden a 23 entidades, de ellas 14 universidades, destacándose la Universidad Politécnica de Cataluña con 38 aportes. En los últimos años (2012-marzo 2017) generó 49 aportes, que representan 59.7 % del total de las contribuciones de este país.

**Entidades y sectores más productivos**

Sobre minería de proceso en el periodo analizado publicaron 876 entidades, sobresaliendo la Universidad Tecnológica de Eindhoven, de Holanda. La *Tabla 4* muestra las entidades más productivas con más de 20 contribuciones.

Entidades	País	Contribuciones
Universidad Tecnológica de Eindhoven	Holanda	328
Universidad de Queensland	Australia	55
Universidad de Tartu	Estonia	45
Universidad de Tsinghua	China	39
Universidad Politécnica de Cataluña	España	38
Universidad de Leuven	Bélgica	32
Universidad de Siam	Tailandia	28
Universidad de Ciencia y Tecnología	Corea del Sur	27
Universidad de Viena	Austria	25
Escuela de Altos Estudios Económicos	Rusia	24
Instituto de Cálculos y Redes de Alto Rendimiento del Consejo Nacional de Investigación (ICAR, CNR)	Italia	24
Universidad de Calabria	Italia	21
Universidad de Economía y Administración de Negocios	Austria	21

*Tabla 4.* Entidades más productivas

La Universidad Tecnológica de Eindhoven ocupó, en 2017, el puesto 177 internacionalmente y el lugar 64 en el mundo de Ingeniería y Tecnologías de la Información del World University Rankings. 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 investigaciones en esta área del conocimiento se centran en las universidades (1 439). Las 13 entidades más productivas pertenecen al sector de la educación, nueve pertenecen al continente europeo y tres a Asia, continentes más productivos. El mayor porcentaje de las investigaciones corresponden al sector educación (77 %), que agrupa las investigaciones que se realizan desde instituciones relacionadas directamente con este sector y el cual está representado significativamente por universidades. Los sectores empresarial, de salud, ciencia y técnica tienen escasa productividad, pues en conjunto sólo representan 23 %.

### *Tipología documental*

Los resultados de las investigaciones en esta área del conocimiento, en el periodo analizado, se presentaron en artículos de revista, ponencias de eventos y capítulos de libros. El artículo de revista es la tipología documental predominante (55 %), superando a las memorias de eventos (34 %) y a los capítulos de libros (11 %).

En la actualidad una de las principales fuentes para comunicar el conocimiento son las revistas científicas, especializadas en cada rama de la ciencia; son periódicas y el riguroso sistema de revisión constituye una forma de validar el conocimiento. Las revistas son reconocidas como medios para obtener información, esto puede ser también una razón por la cual los autores las prefieren, pues el usuario o consumidor logra acceder a ellas fácilmente, están diseñadas para ser atractivas y su salida regular garantiza actualidad en los textos que las incluyen.

### *Productividad de publicaciones*

El núcleo Bradford acumula una porción sustancial de contribuciones (más de 100) producidas, 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 *CEUR Workshop Proceedings*. La mayor cantidad de contribuciones se centra en estas publicaciones, de lo que se infiere que ocupan la preferencia de los autores para divulgar los resultados de investigación sobre minería de proceso. La *Figura 5* muestra las publicaciones que en el periodo analizado tienen más de 10 contribuciones sobre minería de proceso.

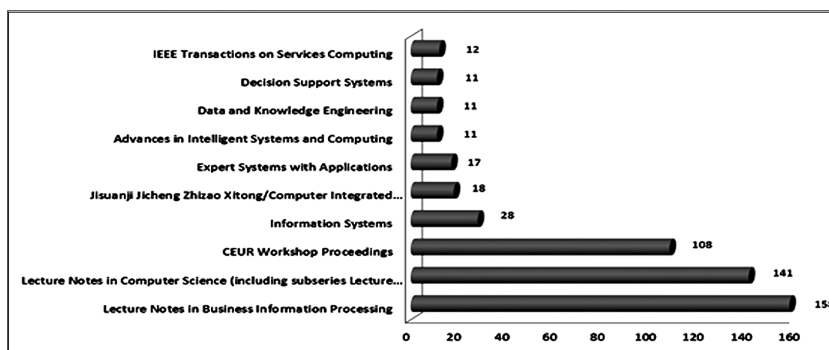


Figura 5. Publicaciones más productivas

*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 actas de eventos, informes de proyectos, tutoriales, conferencias, seminarios y tesis premiadas, entre otros materiales. Según Scimago Journal Report, en el periodo comprendido entre 2009 y 2016 esta publicación tuvo 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 tuvo un *ranking* de 0.32, la cantidad de citas por documentos es de 0.67 y el total de citas es de 44840.

*CEUR Workshop Proceedings* es una serie monográfica publicada en Alemania, de libre acceso, de Sun SITE Europa Central, operado con el auspicio de RWTH Aachen University.

En relación a los eventos o congresos existen 397 contribuciones, entre las que destacan *International Conference on ICT and Knowledge Engineering* con 27 contribuciones, seguida de *Proceedings of the ACM Symposium on Applied Computing* con 18 y *Proceedings - International Conference on Research Challenges in Information Science* con 10.

### ***Frecuencia de aparición de palabras clave***

Del análisis de las palabras clave de los registros se obtuvieron un total de 2 638 palabras clave. De la muestra, 76 % sólo aparecen una vez, el término PROCESS MINING es el que más se destaca (960), seguido de BUSINESS PROCESS MANAGEMENT (87) y PROCESS DISCOVERY (84). De manera general las palabras clave asignadas en ocasiones abarcan poco y no describen con exactitud el contenido de las investigaciones, lo cual limita los análisis de información.

La significativa representatividad del término PROCESS MINING en el campo de las palabras clave (*keywords*) nos indica la efectividad de la recuperación partiendo que en la estrategia de búsqueda se utilizó la frase exacta (“process mining”). Utilizando la herramienta AntConc 3.4.4 para analizar el corpus de todos los resúmenes, se obtuvo, luego de filtrar todas las Stop-Words en inglés que, de los 121 709 términos, PROCESS y MINING siguen ocupando el primer y segundo lugar del *ranking*, con frecuencias de aparición de 6 405 y 2 950 respectivamente. Se destacan otros términos como EVENT, DATA, MODEL, BUSINESS. La *Figura 6* muestra el análisis de los términos en el resumen.

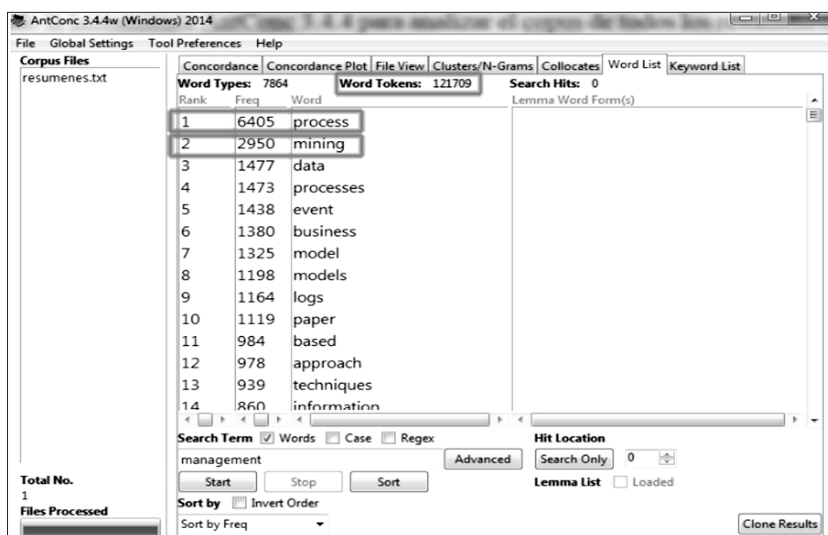


Figura 6. Análisis del corpus lingüístico del resumen con AntConc

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

El análisis de co-palabras estudia la co-ocurrencia de palabras en un texto, permite detectar clústeres, programas o líneas de investigación presentes en el objeto de estudio (Restrepo Arango y Urbizagástegui Alvarado, 2017). El abordaje de este tipo de análisis ha estado siempre sometido a debate en relación a cuáles son los métodos más adecuados para representar la constitución del dominio disciplinar desde su estructuración temática (Liberatore y Herrero Solana, 2013), centrándose la discusión en cuanto a los niveles de subjetividad.

Al realizar el conteo de las palabras más frecuentes del resumen y el título y visualizarlas, se puede observar los términos más frecuentes en las publicaciones sobre minería de proceso. Estos términos identificados pueden caracterizarse por conceptos de proximidad y distancia y, a su vez, representarse gráficamente a través de mapas de ciencia obteniéndose como resultado la estructura temporal del campo del conocimiento sobre esta temática (Restrepo Arango y Urbizagástegui Alvarado, 2017).

En la *Figura 7* se muestran las dos variantes de visualización de la densidad de co-ocurrencia de términos. A la izquierda se muestran grupos de términos asociados por tonalidades de gris, la cantidad de gris, utilizados nos da el número de clúster de términos presentes en los resúmenes. A la derecha, se observa la densidad de co-ocurrencia de cada término. La cercanía al gris más





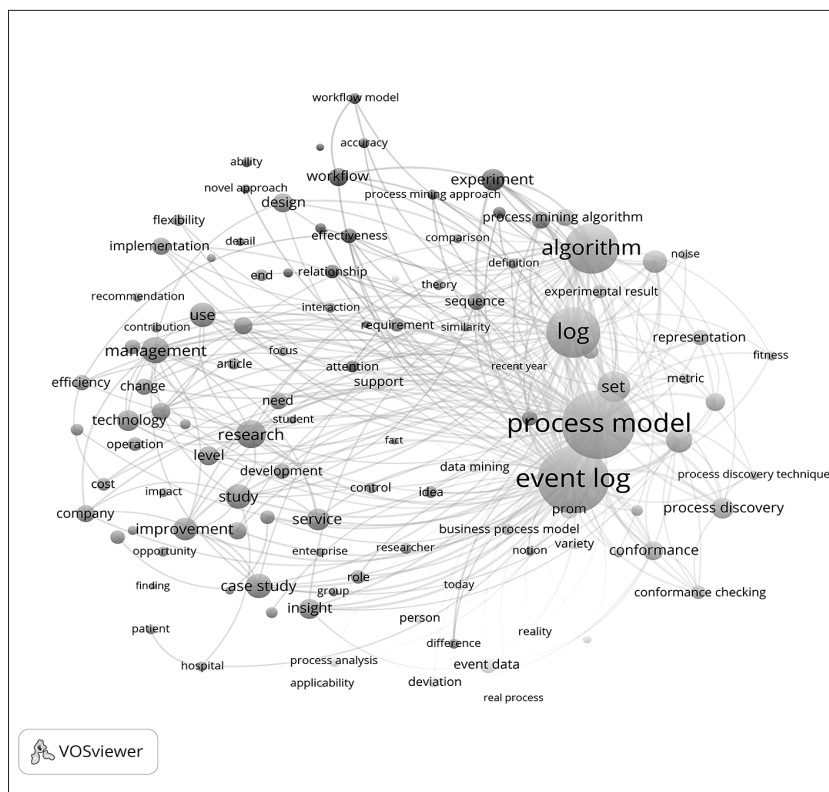


Figura 8. Visualización de la red de términos en los resúmenes

La red de términos (Figura 8) nos muestra el grado de asociación de los términos entre sí con la utilización de un mismo tono de gris, las líneas que los unen son más fuertes a partir del aumento de grosor de las mismas y el peso del término está en correspondencia con el tamaño del nodo y de sus textos. Como se aprecia, los *log* en el modelo del proceso ocupa un papel fundamental y están relacionados con el criterio de la comunidad científica, que los datos de eventos deben ser tratados como ciudadanos de primera clase (Van der Aalst *et al.*, 2012).

Los términos más frecuentes en los resúmenes, como se aprecia en las Figuras 7 y 8, son “process model”, “log”, “event log”, “algorithm” y “management”. Estos términos también pueden ser considerados tópicos consolidados o calientes, es decir, los temas más abordados en este periodo de tiempo; los términos más alejados de la zona gris más intensa son emergentes o que en algún momento fueron muy tratados, pero han perdido actualidad.

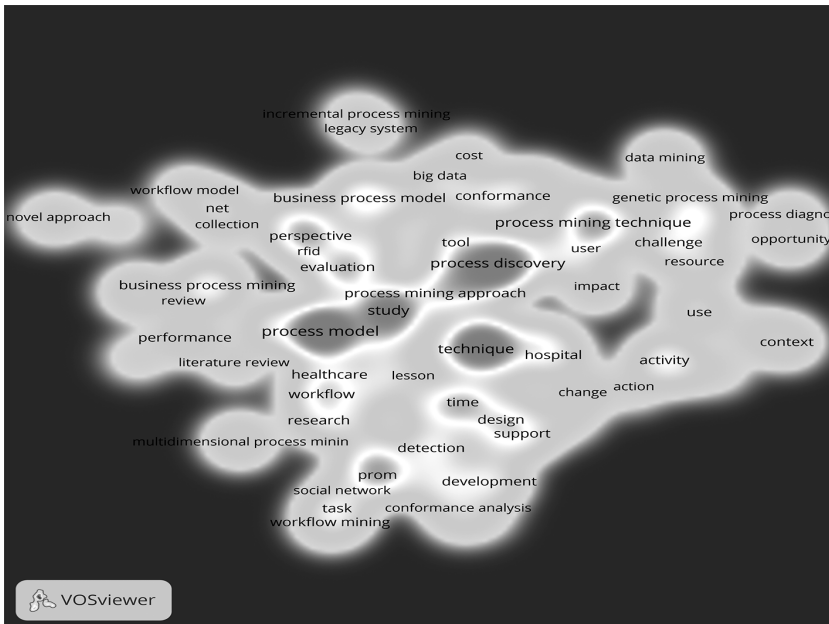
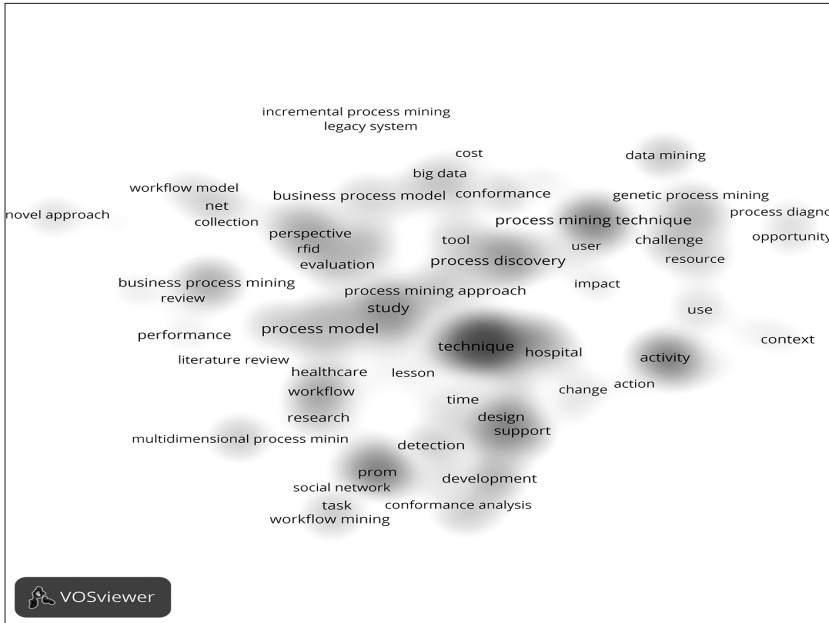


Figura 9. Co-ocurrencia de términos en los títulos (arriba-densidad de los clústeres o grupos y abajo-densidad de co-ocurrencia de términos)



## CONSIDERACIONES FINALES

El análisis de la producción científica sobre minería de proceso a partir de herramientas métricas permitió la caracterización de esta área del conocimiento desde una mirada tanto cuantitativa como cualitativa. La identificación de los principales frentes, tendencias, líderes de investigación y publicaciones núcleo constituye una valiosa herramienta para la vigilancia tecnológica.

En el año 2011 aparece la versión original del *Process Mining Manifesto*, cuyo principal objetivo es promover la investigación, desarrollo, educación, implementación, evolución y entendimiento acerca de la minería de proceso. En ese año se visualizó un incremento sostenido en la producción científica sobre esta área del conocimiento. Se demostró la efectividad de la estrategia trazada por la Fuerza de Trabajo de la IEEE sobre minería de proceso, la cual pudiera ser extendible a otras áreas del conocimiento en desarrollo.

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

Las entidades más productivas corresponden al sector educativo, en el cual destaca la Universidad Tecnológica de Eindhoven en salidas de investigaciones sobre el tema en cuestión, por lo que se infiere que la minería de proceso 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.

A pesar de que estadísticamente las ponencias o conferencias publicadas en memorias de eventos en ocasiones superan en cantidad a los artículos científicos, en el caso de la minería de proceso se demostró que los artículos en revistas científicas son las principales fuentes para comunicar el conocimiento. Este comportamiento resulta lógico para una ciencia en evolución y crecimiento, donde de manera general los investigadores prefieren esta vía para lograr una mayor visibilidad y reconocimiento en la comunidad científica.

Se determinaron las principales líneas de investigación y se denota que el empleo de esta técnica puede tener aplicabilidad en diferentes áreas del conocimiento, lo que evidencia la importancia de este tipo de análisis para destacar los diversos frentes de investigación existentes y nuevas tendencias investigativas.

En los países industrializados existe una sostenida producción científica sobre minería de proceso liderada 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. Se evidenció la brecha

tecnológica existente en las diferentes áreas geográficas e inclusive dentro de países de un mismo continente. La aplicación de la ley de Lotka en este periodo demostró una distribución desigual de las contribuciones al estar concentrada en una pequeña proporción de autores altamente productivos.

El núcleo de Bradford permitió identificar las principales fuentes donde se publican investigaciones sobre esta área del conocimiento, lo cual deriva en una herramienta académica para los interesados en conocer o investigar sobre minería de proceso. Contribuye, además, a establecer políticas de gestión de colecciones en entidades que trabajan o investigan sobre este tema ya que permite determinar los títulos de publicaciones que deben adquirirse.

#### Agradecimientos

A la doctora Yaniris Rodríguez Sánchez, experta en estudios bibliométricos, por la revisión que hizo a esta investigación.

#### REFERENCIAS

- Anthony, Laurence. 2014. AntConc (Versión 3.4.4.). Tokyo, Japan: Waseda University, Faculty of Science and Engineering.
- Ayala Picazo, Micaela. 2016. “Estudio bibliométrico de la revista *Anales Médicos* (1955-2015). Parte 1, análisis de la producción científica”. *Anales Médicos* 61 (4): 246-250.
- Bojo Canales, C., D. Carabantes Alarcón, J. Veiga de Cabo y D. Martínez Hernández. 2004. “Análisis bibliométrico de la *Revista Española de Quimioterapia* (1996-2000)”. *Revista Española de Quimioterapia* 17 (2): 161-168.
- Bradford, S. C. 1948. *Documentation*. Londres: Crosby Lockwood. Reimpresión, Washington, DC: Public Affair Press, 1950.
- Carneiro, A. 2009. *Auditoria e Controlo de Sistemas de Informação*. Lisboa: FCA-Editora de Informática.
- Centre for Science and Technology Studies. 2017. “VOSviewer: Visualizing Scientific Landscapes”. Leiden University. Fecha de consulta: marzo de 2017. <http://www.vosviewer.com/>
- Gálvez, Carmen. 2016. “Visualización de las principales líneas de investigación en salud pública: un análisis basado en mapas bibliométricos aplicados a la *Revista Española de Salud Pública* (2006-2015)”. *Revista Española de Salud Pública* 90:1-10.
- Giraldo Mejía, Juan Camilo. 2016. “Modelo de integración de BPM y Minería de Procesos con un Enfoque Dimensional, para la Optimización de Indicadores KPI”. Tesis (doctorado), Departamento de Ciencias de la Computación y de la Decisión, Facultad de Minas, Universidad Nacional de Colombia.
- Hernández García, Yoscelina I. y Hugo Navarro Contreras. 2017. “Estudio bibliométrico del sistema de investigación local de San Luis Potosí (1920-2015)”. *Investigación bibliotecológica. Número especial de Bibliometría*: 249-270.

- Liberatore, Gustavo y Victor Herrero Solana. 2013. "Caracterización temática de la investigación en Ciencia de la Información en Brasil en el período 2000-2009". *TransInformação* 25 (3): 225-235.
- Lotka, A. J. 1926. "The frequency distribution of scientific productivity". *Journal of the Washington Academy of Sciences* 16 (12): 317-323.
- Mans, Ronny S., Wil M. P Van der Aalst y Rob J. B. Vanwersch. 2015. "Process Mining in Healthcare Evaluating and Exploiting Operational Healthcare". *Springer briefs in business process management*. doi: 10.1007/978-3-319-16071-9
- Mora, Manuel, Jorge Marx Gómez, Leonardo Garrido y Francisco Cervantes Pérez. 2013. *Engineering and Management of IT-based Service Systems: An Intelligent Decision-Making Support Systems Approach*, vol. 55. Springer.
- Navarro Colorado, Borja. 2015. *Guía rápida de análisis de corpus (con AntConc)*. Universidad de Alicante.
- Ochoa Reyes, Alexeis Joel, Arturo Orellana García y Yadira Lizama Mué. 2014. "System for Processing and Analysis of Information Using Clustering Technique". *IEEE Latin America Transactions* 12 (2).
- Orellana García, Arturo, Osvaldo Ulises Larrea Armenteros y Damián Pérez Alfonso. 2014. "Generador de registros de eventos para el análisis de procesos en un sistema de información hospitalaria."
- Orellana García, Arturo, Osvaldo Ulises Larrea Armenteros, Yosbani Enrique Pérez Ramírez y Damián Pérez Alfonso. 2016. "Inductive Visual Miner Plugin Customization for the Detection of Eventualities in the Processes of a Hospital Information System". *IEEE Latin America Transaction* 14 (4): 1930-1936.
- Pichuante Escaida, Charlotte. 2016. "Visualización de grafos de co-autoría y de conocimiento basado en publicaciones científicas, implementada en VOSviewer". Magister en Procesamiento y Gestión de la Información, Universidad Católica de Chile.
- Restrepo Arango, Cristina y Rubén Urbizagástegui Alvarado. 2017. "Red de co-palabras en la bibliometría mexicana". *Investigación bibliotecológica* 31 (73): 17-45.
- Rodríguez Sánchez, Yaniris. 2012. "Metodología bibliométrica para la evaluación de la actividad científica". Tesis (doctorado), Instituto Superior Politécnico José Antonio Echeverría.
- Shrestha, A., Aileen Cater Steel, Mark Toleman y Wui Gee Tan. 2014. "A Method to Select IT Service Management Processes for Improvement". *JITTA. Journal of Information Technology Theory and Application* 15 (3): 31-56.
- Sood, Archit, Neha Sinha, Shashank Dewjee y Wei Zhao. 2017. "Tableau Tutorial. User Documentation". <http://www.tableausoftware.com/products/desktop>
- Urbizagástegui Alvarado, Rubén. 1999. "La Ley de Lotka y la literatura de bibliometría". *Investigación Bibliotecológica* 3 (27):17.
- Urbizagástegui Alvarado, Rubén. 2016. "El crecimiento de la literatura sobre la ley de Bradford". *Investigación bibliotecológica* 30 (68): 51-72.
- Van der Aalst, Wil M. P. 2011. *Process mining: discovery, conformance and enhancement of business processes*, vol. 2. Springer.
- Van der Aalst, Wil M. P, Arya Adriansyah, Ana Karla Alves de Medeiros, F. Arcieri, T. Baier, F. Daniel, K. Barkaoui *et al.* 2012. "Process mining manifesto". 9th International Conference on Business Process Management, BPM 2011P, Clermont-Ferrand.

- Van Eck, Nees Jan y Ludo Waltman. 2010. "Software survey: VOSviewer, a computer program for bibliometric mapping". *Scientometrics* 84 :523-538. doi: 10.1007/s11192-009-0146-3
- Yzquierdo Herrera, Raykenler. 2013. "Minería de proceso como herramienta para la auditoría". *Ciencias de la Información* 44 (2): 25-32.

*Para citar este texto:*

Fuentes Reyes, Sady Carina, Adonis Domínguez Castro, Waldo García Pérez, Patricia Romero Lazcano y Liliet Leyva Pérez. 2019. "Caracterización de la producción científica en el área disciplinar de la minería de proceso". *Investigación Bibliotecológica: archivonomía, bibliotecología e información* 33 (78): 193-216.  
<http://dx.doi.org/10.22201/iibi.24488321xe.2019.78.57925>