

### 2.3. Criterios de evaluación de la calidad de las preguntas de investigación

Como parte del proceso de formulación de las preguntas de investigación, se llevó a cabo la evaluación de su calidad, para esto se aplicó el método FINER [7], un enfoque ampliamente reconocido para formular y evaluar la calidad de preguntas de investigación en diversos campos de estudio. Este método se basa en cinco criterios específicos: Factibilidad, Interés, Novedad, Ética y Relevancia [7], que proporcionan un marco completo para garantizar que las preguntas formuladas sean apropiadas, factibles, originales, relevantes y éticas. Para la evaluación de cada criterio, se asignó una puntuación de 1 punto si la pregunta cumplía totalmente con el criterio, 0.5 puntos si lo cumplía parcialmente y 0 puntos si no lo cumplía, se consideró que una pregunta era aceptable si su promedio de puntuación total se encontraba entre 3.5 y 5 puntos, lo que indica que cumplía con los estándares de calidad establecidos, aquellas preguntas que no alcanzaron este umbral fueron revisadas y, en caso necesario, modificadas para garantizar su relevancia y pertinencia para la RSL en cuestión, a continuación, se relaciona a cada criterio la pregunta correspondiente a ser evaluada:

- **Factibilidad (F):** ¿Es posible abordar esta pregunta con los recursos disponibles, como acceso a la literatura relevante y capacidad de análisis?
- **Interesante (I):** ¿La pregunta aborda un tema de interés y relevancia para la comunidad académica e industrial?
- **Novedoso (N):** ¿La pregunta explora un aspecto único o innovador que no ha sido ampliamente abordado en la literatura existente?
- **Ético (E):** ¿La investigación derivada de esta pregunta se conducirá de manera ética y transparente, respetando los principios de integridad académica y profesional?
- **Relevante (R):** ¿La pregunta es relevante para el campo de estudio y tiene implicaciones significativas para la práctica y la investigación en el desarrollo de software?

Para llevar a cabo el método FINER se contó con la participación de 5 evaluadores expertos en el área de Ingeniería de Software, especialmente en el área de mejora y calidad de procesos, los cuales ayudaron a obtener el nivel de relevancia y pertinencia de las preguntas de investigación para la RSL, esta evaluación se llevó a cabo haciendo uso de una encuesta en Google Forms, a través del cual se obtuvo un conjunto de resultados para posteriormente llevar a cabo el análisis de los resultados obtenidos, los cuales se resumen en la Tabla 3.

Tabla 3. Criterios de evaluación de la calidad para las preguntas de investigación.

Id	Pregunta	E	Puntuación por persona a cada uno de los criterios de calidad					Total	P
			Factible	Interesante	Novedoso	Ético	Relevante		
PI1	¿Cuáles son los tipos de soluciones propuestas en la literatura que permitan analizar y/o gestionar la deuda de procesos en el desarrollo de software?	1	1	1	0.5	0.5	1	4	3.7
		2	0.5	1	1	1	1	4.5	
		3	0.5	0.5	0	0	0.5	1.5	
		4	1	1	0.5	1	1	4.5	
		5	0.5	1	0.5	1	1	4	
PI2	¿Cuáles son los tipos de investigación	1	1	1	1	1	1	5	4.1
		2	1	1	0.5	0.5	1	4	
		3	0.5	1	0.5	0	0.5	2.5	

	identificados en la literatura?	4	1	1	1	0.5	1	4.5	
		5	1	1	0.5	1	1	4.5	
PI3	¿Cuál es el nivel de madurez y/o desarrollo actual en la literatura con respecto a los conceptos y enfoques de gestión de la deuda de procesos en el desarrollo de software?	1	1	1	1	1	1	5	4.5
		2	1	1	1	1	1	5	
		3	1	1	1	0	0.5	3.5	
		4	1	1	1	0.5	1	4.5	
		5	1	1	0.5	1	1	4.5	
PI4	¿Cuáles son los beneficios, desafíos, causas, impacto, efectos y/o consecuencias asociadas con la investigación sobre la deuda de procesos en el desarrollo de software?	1	1	1	1	1	1	5	3.9
		2	0.5	0.5	0.5	0.5	0.5	2.5	
		3	1	1	0.5	0	0.5	3	
		4	1	1	1	0.5	1	4.5	
		5	1	1	0.5	1	1	4.5	

Acrónimos utilizados: E: evaluador, P: Promedio.

Basado en las respuestas obtenidas utilizando el método FINER, los evaluadores consideran que las preguntas de investigación cumplen con los criterios de calidad establecidos por el método, y aunque se puede observar que el puntaje de las preguntas no alcanzó el máximo esperado, estas lograron superar el límite inferior del umbral establecido anteriormente.

## 2.6. Evaluación de la relevancia y pertinencia de los estudios primarios

Evaluar la relevancia y pertinencia para la investigación de los artículos primarios permite minimizar los sesgos y maximizar la validez interna y externa de una investigación [5]. Siguiendo esta premisa y con el propósito de mantener una rigurosidad científica en la investigación, se evaluó la relevancia y pertinencia de los artículos primarios seleccionados a partir de una adaptación del instrumento propuesto por Kitchenham y Charters [5] Kitchenham *et al.* [8]. La evaluación de la relevancia y pertinencia se basó en tres factores principales: **(i) validez interna**, que se refiere a la capacidad del estudio para responder adecuadamente a la pregunta de investigación planteada; **(ii) validez externa**, que según Dybá *et al.* [9], permite analizar la generalización de los resultados para la población de interés y la aplicabilidad de los hallazgos; **(iii) sesgo**, definido por Dybá *et al.* [9] como un error sistemático o una desviación de la verdad en los resultados o inferencias, lo que refleja una tendencia a generar resultados que se alejan sistemáticamente de la realidad. Teniendo en cuenta lo anterior, para esta evaluación se diseñó un instrumento que incluye una lista de verificación que abarca cuatro categorías: claridad, rigor, credibilidad y relevancia de cada estudio. La Tabla 5 presenta las preguntas utilizadas para medir el nivel de relevancia y pertinencia a partir de las cuatro categorías establecidas. A continuación, se describen dichas categorías:

- **Claridad:** Esta categoría hace énfasis en la importancia de presentar los resultados de la RSL de manera clara y concisa para que puedan ser comprendidos por otros investigadores y profesionales en el campo de la ingeniería de software [10].

- **Rigor:** Esta categoría destaca la necesidad de seguir una metodología rigurosa y confiable durante todo el proceso de la RSL para garantizar la validez de los resultados obtenidos [6].
- **Credibilidad:** Esta categoría permite conocer el nivel en el cual los resultados de la investigación son significativos, válidos y efectivos mediante la evaluación de los métodos científicos utilizados, la discusión de las limitaciones del proceso de investigación y el análisis de los resultados obtenidos [11].
- **Relevancia:** Esta categoría permite determinar si los resultados del estudio son válidos, útiles, bien presentados y significativos. Además, identifica los métodos científicos utilizados y el análisis de los hallazgos encontrados [9].

Tabla 5. Relevancia y pertinencia de los artículos primarios.

Id	Relevancia y Pertinencia de los Artículos Primarios	Puntuación para las respuestas			Categoría
		+1	0	-1	
1	¿El artículo se enfoca claramente en la deuda de procesos en el desarrollo de software?	Si	Parcialmente	No	Claridad
2	¿El artículo describe claramente el problema y los objetivos de la investigación?	Si	Parcialmente	No	
3	¿El artículo proporciona una definición clara y operativa de la deuda de procesos?	Si	Parcialmente	No	
4	¿El artículo sigue una metodología de investigación formal y estructurada?	Si	Parcialmente	No	Rigurosidad
5	¿Los métodos de recolección y análisis de datos son apropiados y bien descritos?	Si	Parcialmente	No	
6	¿El artículo proporciona suficiente detalle para permitir la replicación del estudio?	Si	Parcialmente	No	
7	¿El artículo tiene implicaciones prácticas claras para la industria del software?	Si	Parcialmente	No	Relevancia
8	¿El artículo contribuye significativamente al conocimiento académico sobre la deuda de procesos?	Si	Parcialmente	No	
9	¿El artículo propone trabajos futuros o áreas de investigación adicionales?	Si	Parcialmente	No	
10	¿El artículo utiliza métodos científicos adecuados para la investigación?	Si	Parcialmente	No	Credibilidad
11	¿El artículo discute las limitaciones y sesgos del proceso de investigación?	Si	Parcialmente	No	
12	¿Los resultados del artículo son significativos y válidos?	Si	Parcialmente	No	

## 2.7. Estrategia de extracción de datos y métodos de síntesis

La Tabla 6 presenta el diseño de la ficha resumen utilizada para facilitar el proceso de clasificación de la información y aspectos más importantes que se consideraron para cada estudio primario, esta permitió que la extracción de datos y la clasificación de información fuera uniforme. Asimismo, se estableció la relación de los estudios primarios respecto a las PI definidas (ver Tabla 11).

Tabla 6. Ficha resumen de estudios.

Identificación		
Título:		
DOI:		Número de citaciones:
Fecha de publicación:	Tipo de publicación:	
Autores:		
Nombre	País	Universidad
Resumen		

Descripción		
Tipo de investigación (Clasificación de [6]):		
Validación industrial <input type="checkbox"/> Investigación de validación <input type="checkbox"/> Investigación de evaluación <input type="checkbox"/> Propuesta de solución <input type="checkbox"/> Metodología de investigación <input type="checkbox"/> Documentos de experiencia <input type="checkbox"/>		
Tipo de solución(es) ofrecida(s):		
Definición conceptual <input type="checkbox"/>	Causas, efectos, impactos y limitaciones <input type="checkbox"/>	Método o técnica de evaluación <input type="checkbox"/>
Implicaciones prácticas <input type="checkbox"/>	Estudio de caso <input type="checkbox"/>	Otros <input type="checkbox"/>
Propuesta:		
Evaluación de la propuesta:		

La Tabla 8 presenta un resumen del proceso de clasificación y selección de estudios primarios, los identificadores numéricos de cada fuente de datos, el nombre de la base de datos científica utilizada, el número total de estudios encontrados inicialmente en cada fuente, el número de estudios considerados relevantes para la RSL después de una evaluación inicial, el número de estudios que se identificaron como duplicados tras la comparación semiautomática y el número de estudios únicos seleccionados como primarios para la RSL. La evaluación de los resultados obtenidos de las distintas bases de datos muestra que Google Scholar destacó por su amplia cobertura, ya que, de los 288 estudios encontrados, 233 fueron considerados relevantes y ninguno fue identificado como duplicado. Este hecho refuerza la utilidad de Google Scholar como la fuente principal de datos, ya que permitió recopilar eficientemente estudios relevantes, cubriendo hallazgos de otras bases de datos especializadas como Scopus, Springer Link, Science Direct, IEEE Xplore, Web of Science y ACM, los cuales resultaron ser duplicados. La identificación semiautomática de duplicados resultó ser clave para asegurar la precisión del proceso, reduciendo el esfuerzo manual y evitando sobrevaloraciones de estudios repetidos, lo que permitió una mayor concentración en los estudios únicos y verdaderamente relevantes. Además, el uso del método de bola de nieve con búsqueda hacia atrás (conocido en inglés como backward snowballing) [13] permitió identificar 239 estudios que cumplieran con los CI y CE, sumando así un total de 20 estudios primarios para la investigación, tal como se ilustra en la Tabla 9.

**Tabla 8.** Esquema de clasificación.

Id.	Fuentes de Datos	Estudios Encontrados	Estudios Relevantes	Estudios Duplicados	Primarios Seleccionados
1	Google Scholar	288	233	0	20
2	Scopus	25	25	25	0
3	Springer Link	13	9	9	0
4	IEEE Xplore	21	18	18	0
5	Web of Science	189	114	114	0
6	Science Direct	11	9	9	0
7	ACM	18	14	14	0
<b>Total</b>		<b>565</b>	<b>422</b>	<b>183</b>	<b>20</b>

**Tabla 9.** Artículos primarios.

Id	Artículos Primarios	Año	Ref.
A1	Process Debt: a First Exploration	2020	[1]
A2	Process Debt: Definition, Risks and Management.	2024	[2]
A3	The Pandora's box of social, process, and people debts in software engineering.	2022	[3]
A4	Technical-, social-and process debt in large-scale agile: an exploratory case-study	2019	[4]

A5	A Multivocal Literature Review on Non-Technical Debt in Software Development: An Insight into Process, Social, People, Organizational, and Culture Debt	2024	[24]
A6	A study of factors that lead development teams to incur technical debt in software projects	2018	[15]
A7	A systematic mapping study on technical debt definition	2015	[25]
A8	A systems perspective on technical debt	2021	[26]
A9	A taxonomy of assets for the development of software-intensive products and services	2023	[27]
A10	Actions and impediments for technical debt prevention: results from a global family of industrial surveys	2020	[14]
A11	Technical debt: towards a crisper definition report on the 4th international workshop on managing technical debt	2013	[22]
A12	Beyond Technical Debt Unravelling Organisational Debt Concept	2024	[28]
A13	Exploring Process Debt in Large-Scale Agile Software Development For Secure Telecom Solutions	2024	[16]
A14	Identification and management of technical debt: A systematic mapping study	2016	[29]
A15	Multivocal Literature Review on Non-Technical Debt in Software Development: An Exploratory Study.	2023	[17]
A16	Society 5.0 and Soft Skills in Agile Global Software Development	2022	[18]
A17	Technical debt is not only about code and we need to be aware about it	2021	[19]
A18	Technical Debt Management: Definition of a Technical Debt Reduction Software Engineering Methodology for SMEs	2019	[20]
A19	Towards a taxonomy of code review smells	2022	[21]
A20	Towards an ontology of terms on technical debt	2014	[23]

Acónimos utilizados: Ref.: referencia.

### 3.1 Evaluación de la relevancia y pertinencia

La evaluación de la relevancia y pertinencia de los estudios primarios, detallada en la Tabla 10, no solo refleja el fundamento teórico de los artículos seleccionados, sino que ofrece una comprensión profunda sobre las áreas donde la investigación en deuda de procesos aún puede avanzar. Cada estudio fue evaluado con base en doce criterios que permiten determinar la importancia de su contribución a la comprensión y gestión de la deuda de procesos. Los puntajes obtenidos, que van desde -12 hasta 12 puntos, reflejan tanto el valor como las posibles limitaciones de cada artículo, haciendo visible la diversidad de enfoques en esta área emergente.

La Figura 1 proporciona una agrupación visual clara de los estudios evaluados, y su relevancia radica en destacar cómo los diferentes enfoques han sido calificados en términos de pertinencia para la comunidad académica. Es notable que solo un 5% de los estudios [2] haya alcanzado la máxima puntuación de 12 puntos, lo que podría indicar que existen enfoques sobresalientes que sirven como referentes clave en la gestión de la deuda de procesos. Este hecho subraya la necesidad de observar con detalle cuáles fueron los criterios y estrategias que permitieron a este artículo destacarse tan significativamente por encima del resto. Por otro lado, el 5% de los estudios [1] alcanzó una puntuación de 11 puntos, mientras que un 15% [3], [16], [21] logró 10 puntos, lo que sugiere que hay un grupo relativamente pequeño de investigaciones que, aunque no alcanzan el puntaje más alto, ofrecen importantes aportes a la gestión de la deuda de procesos. Estos artículos destacan como aportes significativos que animan a los lectores interesados a profundizar en su exploración.

La distribución de los estudios con puntajes intermedios, como aquellos que obtuvieron entre 6 y 9 puntos. Un 10% de los estudios [24], [19] alcanzó 9 puntos, mientras que otro 10% [14], [29] obtuvo 7 puntos, mostrando enfoques que, si bien son efectivos, podrían beneficiarse de una integración más profunda con los contextos reales de desarrollo de software. No obstante, estos estudios constituyen una base valiosa sobre la cual seguir construyendo, y sería recomendable examinarlos con mayor detalle para identificar posibles mejoras o áreas de aplicación en entornos empresariales.

La Figura 1 también resalta la presencia de artículos con calificaciones más bajas, como aquellos con 5 puntos o menos, que representan el 40% de los estudios evaluados. Este grupo incluye artículos con puntajes negativos, como el estudio [22] (-1) y [26] (-6), lo cual abre un espacio de reflexión crítica sobre la pertinencia de sus contribuciones. Estos estudios no solo indican una falta de alineación con los criterios evaluados, sino que también invitan a una lectura más profunda para comprender si la baja puntuación se debe a limitaciones metodológicas, contextuales o de enfoque.

Dada la diferencia en las calificaciones, es necesario revisar con cuidado tanto la Tabla 10 como la Figura 1, pues ambas ofrecen una representación visual que facilita identificar patrones de relevancia en los estudios. Estas diferencias de puntuación deben ser vistas como una oportunidad para evaluar el estado actual de la investigación sobre la deuda de procesos. En este sentido, aquí es donde la comunidad académica puede focalizar futuros esfuerzos, por ejemplo: fomentar la colaboración interdisciplinaria y desarrollar teorías que mejoren las prácticas de gestión actuales [2]. De igual manera, se podrían realizar estudios de caso en empresas, impulsar programas de capacitación y fortalecer la colaboración con la industria [3]. Además, es esencial desarrollar marcos conceptuales, adoptar enfoques holísticos que integren las diferentes formas de deuda y estudio de caso en entornos reales [17].

Tabla 10. Evaluación de la relevancia y pertinencia de los artículos primarios.

Id	Relevancia y pertinencia de los artículos primarios												Total
	1	2	3	4	5	6	7	8	9	10	11	12	
A1	1	1	1	1	1	0	1	1	1	1	1	1	11
A2	1	1	1	1	1	1	1	1	1	1	1	1	12
A3	1	1	1	0	0	1	1	1	1	1	1	1	10
A4	0	1	0	1	1	0	0	0	1	1	0	1	6
A5	1	1	1	1	0	0	1	1	1	1	0	1	9
A6	-1	1	-1	1	1	0	1	0	1	1	0	1	5
A7	0	1	0	1	1	0	0	0	1	1	0	0	5
A8	-1	0	-1	0	-1	-1	1	-1	0	-1	-1	0	-6
A9	0	1	-1	1	0	0	0	0	1	1	1	0	4
A10	0	1	0	1	1	0	1	0	1	1	0	1	7
A11	-1	1	-1	0	0	-1	1	-1	1	0	0	0	-1
A12	1	1	1	0	0	-1	1	1	1	0	0	1	6
A13	1	1	1	1	1	0	1	1	1	1	0	1	10
A14	0	1	0	1	1	0	1	0	1	1	1	0	7
A15	1	1	1	1	0	0	1	1	1	0	0	1	8
A16	-1	1	-1	1	1	0	0	-1	1	1	0	0	2
A17	1	1	0	1	1	0	1	1	1	1	0	1	9
A18	0	1	0	1	0	0	1	0	1	0	1	0	5

A19	1	1	1	1	1	0	1	1	1	1	0	1	10
A20	0	1	0	1	0	-1	0	0	1	0	-1	0	1

Acrónimos utilizados: Id: identificación del artículo (ver Tabla 9).

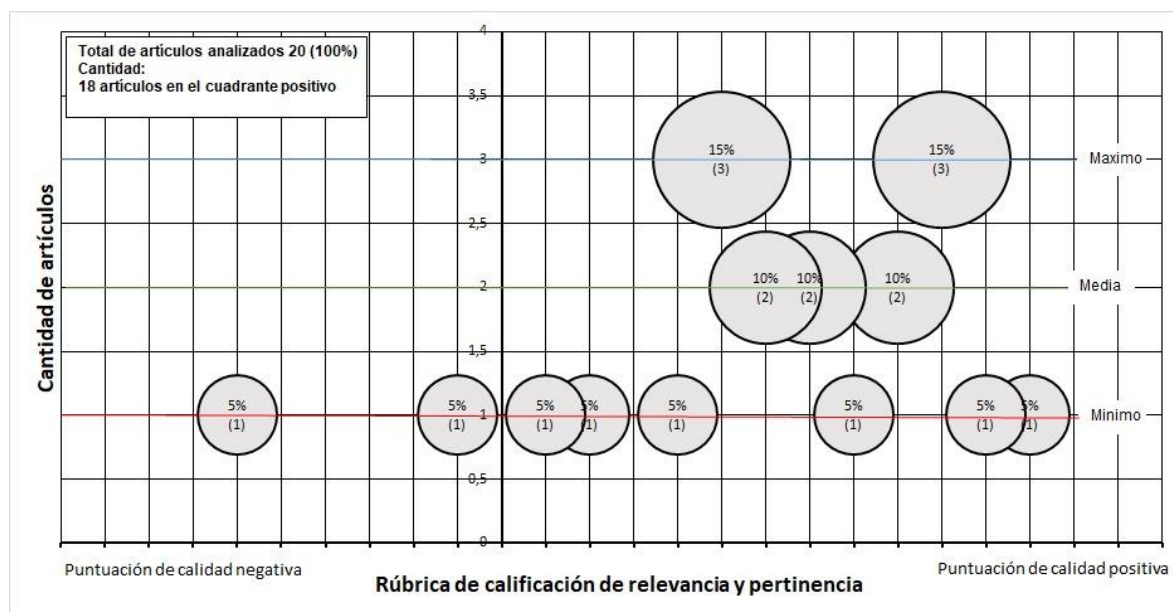
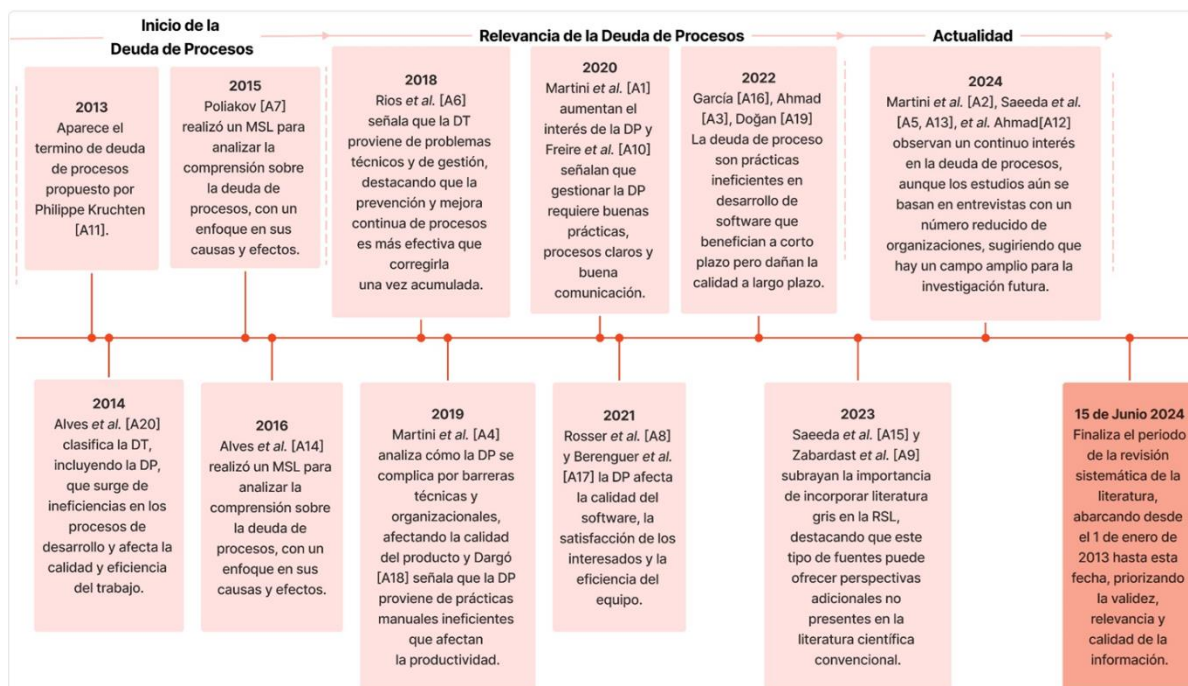


Figura 1. Visualización de la relevancia y pertinencia de los artículos primarios en una gráfica de burbuja.

### 3.2 Frecuencia y tipos de publicaciones realizadas sobre la deuda de procesos

Como muestra la Figura 2, aunque ha aumentado la frecuencia de las publicaciones sobre la deuda de procesos, lo que refleja un interés creciente por el tema, el número de estudios sigue siendo limitado en comparación con la deuda técnica y otros tipos de deuda. Este mayor interés se manifiesta, como mencionan Martini *et al.* [2], en estudios empíricos, documentos de conferencias y revisiones conceptuales; sin embargo, los estudios actuales, que a menudo se basan en entrevistas con un número reducido de organizaciones, sugieren que los resultados aún no son exhaustivos y que existe un amplio margen para futuras investigaciones [1]. Además, se destaca que, antes de 2019, los estudios pertinentes sobre la deuda de procesos eran escasos.



Acrónimos utilizados: DP: deuda de procesos, DT: deuda técnica, MSL: mapeo sistemático de la literatura, RSL: revisión sistemática de la literatura

**Figura 2.** Línea de tiempo de la evolución y segmentación de artículos según su distribución temporal.

## 5.2. Limitaciones de la revisión

- **Sesgos en las preguntas de investigación.** Durante la formulación de las preguntas de investigación, se evidenció un avance significativo en la identificación de desafíos derivados de la limitada cantidad de información disponible sobre el tema de la deuda de procesos. Esta escasez de datos hizo necesario que el enfoque de cada pregunta fuera específico al período de tiempo seleccionado (2013-2024). Aunque estos ajustes fueron necesarios para abordar las limitaciones del tema, también introducen un sesgo al restringir el análisis a estudios recientes y muy específicos.
- **Sesgos en la cadena de búsqueda.** La cadena de búsqueda inicial utilizada en esta revisión fue bastante extensa, lo que resultó en la exclusión de algunos documentos relevantes. Para abordar esta limitación, se decidió utilizar una cadena de búsqueda más corta y general, lo que permitió abarcar un mayor número de estudios y obtener una visión más completa del tema. Se realizaron búsquedas en diferentes motores de búsqueda, incluyendo Google Scholar, Scopus, Science Direct, IEEE Xplore, Web of Science, ACM y Springer Link. Finalmente, se seleccionó Google Scholar como motor de búsqueda principal debido a su capacidad para cubrir de manera integral los estudios relevantes que también se encontraban en otras bases de datos. Sin embargo, este enfoque en un solo motor de búsqueda principal pudo haber introducido un sesgo al depender en gran medida de su algoritmo.



- **Sesgos en la selección de estudios.** La selección de estudios se basó en criterios de inclusión y exclusión predefinidos, lo que podría haber llevado a la omisión de investigaciones relevantes que no cumplieran con todos los criterios establecidos. Este proceso de selección está sujeto a sesgos subjetivos, donde la interpretación de los criterios pudo haber influido en la decisión de incluir o excluir ciertos estudios. La accesibilidad de los documentos también fue un factor relevante en la selección, ya que se incluyeron documentos de tesis de pregrado, maestría y doctorado, así como tipos de publicaciones que abarcan conferencias, artículos científicos y literatura gris. Esta diversidad en los tipos de documentos permitió una visión más amplia del tema, aunque también pudo introducir un sesgo al depender de la disponibilidad y accesibilidad de estos materiales.
- **Sesgos en la extracción de datos.** Para minimizar los sesgos en la extracción de datos, se siguió un protocolo estructurado para seleccionar los artículos primarios. Este protocolo incluyó una evaluación cuidadosa del tema tratado, la introducción de cada artículo y las palabras clave, así como los criterios de inclusión y exclusión definidos. A pesar de que este enfoque sistemático ayudó a garantizar la relevancia y calidad de los estudios seleccionados, la interpretación subjetiva de estos elementos pudo haber introducido sesgos. Es posible que se haya dado mayor peso a ciertos estudios en función de su alineación con los objetivos de la revisión, lo que podría haber influido en la forma en que los datos fueron extraídos y presentados.
- **Sesgos en la interpretación de resultados.** La interpretación de los resultados pudo haber estado influenciada por expectativas previas y por la limitada cantidad de estudios disponibles que abordan directamente la deuda de procesos. Este fenómeno puede haber llevado a una sobreinterpretación de ciertos hallazgos o a la subestimación de otros, dependiendo de cómo los resultados coincidieran con las hipótesis iniciales. Esto limita la generalización de los resultados a un contexto más amplio, lo que es un aspecto a tener en cuenta en futuras investigaciones.