

PROTOCOLO

Mapeo Sistemático de la Literatura

Título: “Estado de la investigación en el uso de Domain-Driven Design para el desarrollo de sistemas con Arquitectura de Microservicios: Un mapeo sistemático de la literatura”.

Josué Sangabriel Alarcón

Proyecto Guiado | Lic. en Ingeniería de Software - FEI

Índice

1.	Introducción	3
2.	Preguntas de investigación	3
3.	Estrategia de búsqueda	4
3.1	Proceso de búsqueda de estudios primarios	4
3.2	Términos de búsqueda	6
3.3	Cadenas de búsqueda.....	6
3.4	Selección de motores de búsqueda	8
4.	Selección de los estudios primarios	9
4.1	Criterios de selección de estudios primarios.....	9
4.2	Procedimiento de selección de estudios primarios.....	9
5.	Extracción de los datos.....	11
6.	Estrategia para la síntesis de datos	13
7.	Limitaciones	14
8.	Informe.....	15
1.	Gestión de la revisión	16
2.	Referencias.....	17

1. Introducción

El objetivo de este mapeo sistemático de la literatura (MSL) es identificar el estado de la investigación respecto al uso de Domain-Driven Design (DDD) para el desarrollo de sistemas con Arquitectura de Microservicios (MSA) a través del análisis y síntesis de los estudios primarios recolectados con la finalidad de proponer áreas de oportunidad en el uso de DDD para el desarrollo de dichos sistemas.

En relación con el diagrama de actividad de la Figura 1, se ha decidido realizar un mapeo sistemático de la literatura titulada: “Estado de la investigación en el uso de Domain-Driven Design para el desarrollo de sistemas con arquitectura de Microservicios”.

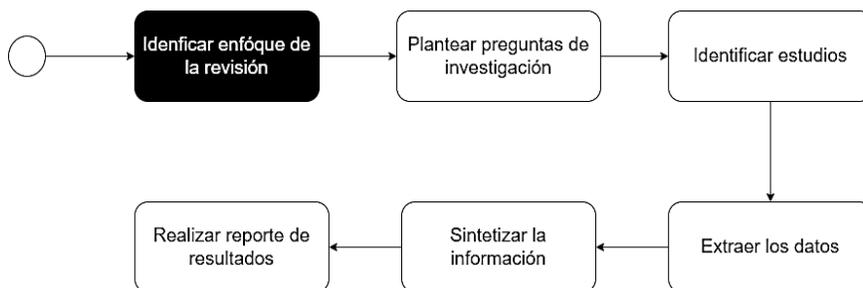


Figura 1: Actividades del proceso de mapeo sistemático (Popay et al., 2006)

2. Preguntas de investigación

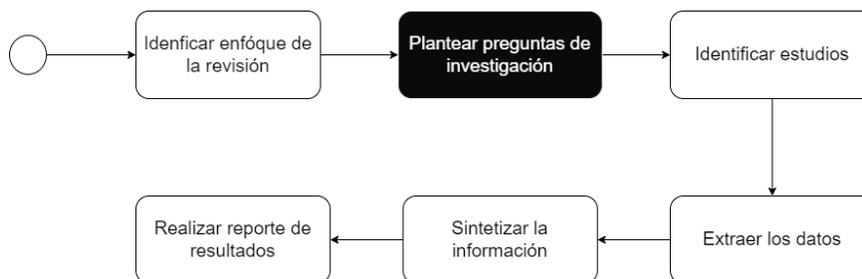


Figura 2: Actividades del proceso de mapeo sistemático (Popay et al., 2006)

Continuando con el planteamiento de las preguntas de investigación de han definido las siguientes:

- PI-1.** ¿Cuáles son los propósitos del uso de Domain-Driven Design en el desarrollo de sistemas con arquitectura de microservicios?
- PI-2.** ¿Qué evidencia del uso de Domain-Driven Design en el desarrollo de sistemas con arquitectura de microservicios?
- PI-3.** ¿Cuáles son los patrones de Domain-Driven Design utilizados en el desarrollo sistemas con arquitectura de microservicios?
- PI-4.** ¿Qué tecnologías son utilizadas con Domain-Driven Design en el desarrollo de sistemas con arquitectura de microservicios?
- PI-5.** ¿Cuáles son las técnicas utilizadas en el desarrollo de sistemas con arquitectura de microservicios diseñados con DDD?

- PI-6.** ¿Cuáles son las limitantes encontradas en el uso de Domain-Driven Design para el desarrollo de sistemas con arquitectura de microservicios?
- PI-7.** ¿Qué propuestas existen para el desarrollo de sistemas con arquitectura de microservicios utilizando Domain-Driven Design?

3. Estrategia de búsqueda

La estrategia de búsqueda comprende procesos de localización de investigaciones que toman lugar en la etapa de identificación de los estudios relevantes al tema de investigación, en otras palabras, los estudios que responden al menos una pregunta de investigación planteada, misma etapa referenciada en el diagrama de actividad de la Figura 3 mostrada a continuación:

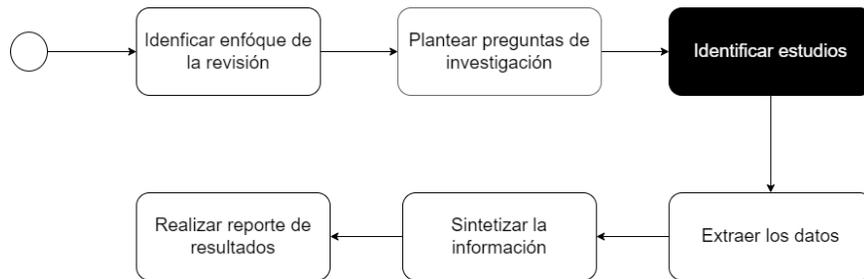


Figura 3: Actividades del proceso de mapeo sistemático (Popay et al., 2006)

3.1 Proceso de búsqueda de estudios primarios

Para comprender el proceso de búsqueda de estudios primarios, usaremos como referencia las actividades del proceso completo de selección de estudios únicamente haciendo énfasis en las actividades de búsqueda, el proceso completo es mostrado en el diagrama de actividades de la Figura 4, mientras que las actividades del proceso de búsqueda únicamente se encuentran en la Tabla 1.

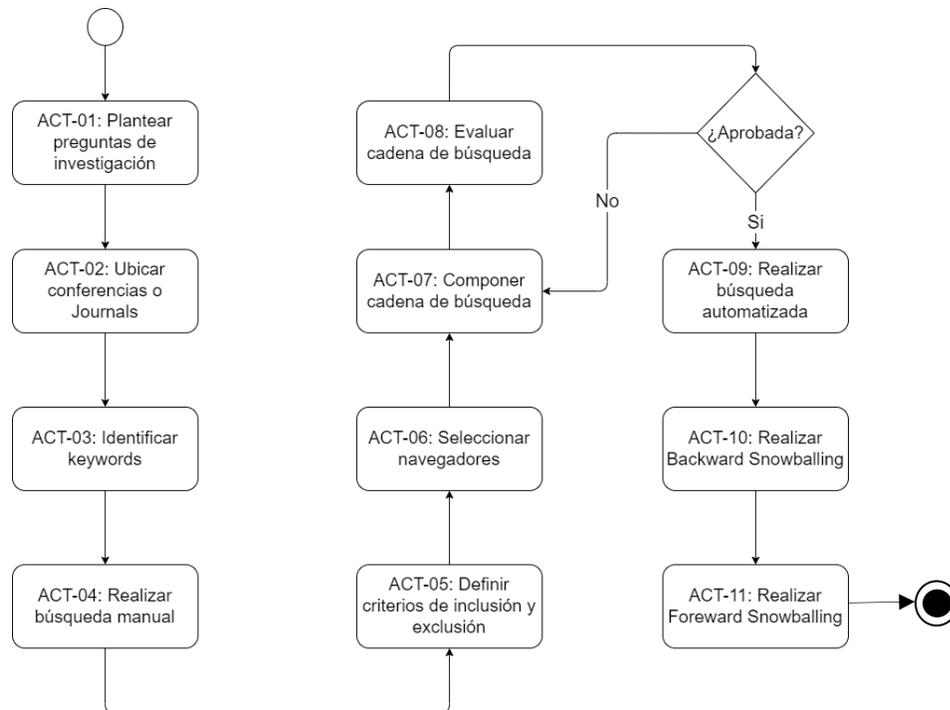


Figura 4: Proceso de búsqueda y selección de estudios primarios

A continuación, se presenta el detalle de las actividades llevadas a cabo en el proceso de búsqueda de los estudios primarios, teniendo como base la actividad ACT-01 que comprende el establecimiento de las preguntas de investigación para llevar a cabo la búsqueda y selección e los estudios primarios:

Tabla 1: Proceso de búsqueda de los estudios primarios

Fase	Actividades
Búsqueda manual	<ul style="list-style-type: none"> • ACT-01: Planear preguntas de investigación. <ul style="list-style-type: none"> ○ Acordar preguntas de investigación con base a consenso de los investigadores. • ACT-02: Ubicar conferencias o journals con estudios publicados relevantes al tema de investigación (Zhang et al., 2011) <ul style="list-style-type: none"> ○ Realizar búsqueda en Google Scholar. ○ Identificar estudios relacionados al tema de investigación. ○ Identificar las conferencias, journals o workshops que publicaron los estudios identificados. • ACT-03: Identificar términos de búsqueda relevantes y sus sinónimos (Zhang et al., 2011) <ul style="list-style-type: none"> ○ Leer los estudios relacionados al tema de investigación completamente. ○ Localizar términos relacionados a las preguntas de investigación y al tema. ○ Eliminar términos duplicados e identificar sinónimos. • ACT-04: Realizar búsqueda manual de estudios relevantes con base en término identificados (Kitchenham et al., 2015; Zhang et al., 2011) <ul style="list-style-type: none"> ○ Preparar cadenas de búsqueda con sinónimos para captura el mayor número de estudios relevantes. ○ Lanzar cadenas de búsqueda en navegadores seleccionados para obtener estudios relevantes. ○ Refinar términos de búsqueda en base a los estudios encontrados. ○ Identificación de autores y sus publicaciones relevantes al tema de investigación. ○ Iterar la búsqueda manual con términos refinados. • ACT-05: Definir criterios de inclusión y exclusión para la búsqueda automatizada. <ul style="list-style-type: none"> ○ Identificar las características deseables de los estudios primarios en base a los estudios identificados en la búsqueda manual.
Búsqueda automatizada	<ul style="list-style-type: none"> • ACT-06: Seleccionar los motores de búsqueda (engines) para llevar a cabo la búsqueda automatizada. <ul style="list-style-type: none"> ○ Localizar los motores de búsqueda más utilizados en investigaciones basadas en evidencia de la ingeniería de software. ○ Identificar los motores de búsqueda donde se encuentran los estudios identificados en la búsqueda manual. • ACT-07: Componer cadena de búsqueda para búsqueda automatizada <ul style="list-style-type: none"> ○ Combinar los términos localizados en la búsqueda manual en base al dominio del tema de investigación por los investigadores. (Kitchenham et al., 2015; Zhang et al., 2011)

Snowballing

- ACT-08: Evaluar cadena de búsqueda en un motor de búsqueda específico (se seleccionó IEEE Explorer) (Zhang et al., 2011)
 - Seleccionar motor de búsqueda donde será evaluada la cadena de búsqueda.
 - Calcular el Recall con base a (Zhang et al., 2011).
 - Calcular el Precision con base a (Zhang et al., 2011).
 - Comprobar que la sensibilidad de la cadena (**Recall**) sea superior al 80%.
- ACT-09: Realizar búsqueda automatizada
 - Lanzar la cadena de búsqueda en los motores de búsqueda seleccionados para obtener estudios relevantes.
 - Filtrar estudios con base a los criterios de selección establecidos previamente.
- ACT-10: Realizar Backward Snowballing (Wohlin, 2014)
 - Identificar estudios citados en los estudios primarios que den indicios de responder alguna pregunta de investigación.
 - Obtener estudio desde la fuente donde fue publicado.
- ACT-11: Realizar Foreward Snowballing (Wohlin, 2014)
 - Identificar estudios que citen a los estudios primarios y que den indicios de responder alguna pregunta de investigación.
 - Buscar el estudio en varios motores de búsqueda.

3.2 Términos de búsqueda

Los términos seleccionados a partir de la búsqueda manual y lectura de los estudios relevantes al tema de investigación son los siguientes:

Palabra clave	Sinónimos
Domain-Driven Design	<ul style="list-style-type: none">• Domain driven design• Domain driven design pattern• Domain driven model design• DDD• DDD pattern
Microservice	<ul style="list-style-type: none">• Microservices• Micro-service• Remote API• API Rest• API
Decomposition	<ul style="list-style-type: none">• Decompose• Microservice identification• Split

3.3 Cadenas de búsqueda

La cadena de búsqueda fue conformada por la terminología encontrada tanto en las conferencias

seleccionadas como de los estudios relevantes seleccionados a través del proceso de búsqueda manual como parte de la propuesta de (Zhang et al., 2011) . Así mismo tenemos el conocimiento del dominio de la investigación, conforme a las directrices de (Kitchenham et al., 2015). De este modo tenemos la conformación de las cadenas de búsqueda a partir de los siguientes criterios:

- Terminología encontrada en los estudios seleccionados relevantes al tema, que corresponde a los estudios bien conocidos de acuerdo con (Zhang et al., 2011)
- Estudio de terminología referente al dominio como parte del conocimiento del investigador de acuerdo con (Kitchenham et al., 2015).

Para la evaluación de la cadena de búsqueda se ha seleccionado la técnica de evaluación de (Zhang et al., 2011) basado en el cálculo de 2 métricas para una evaluación objetiva:

- **Recall (Sensitivity):** $\left(\frac{\text{Numero de estudios relevantes obtenidos}}{\text{Total de estudios relevantes}}\right) \times 100\%$
- **Effort (Precision):** $\left(\frac{\text{Numero de estudios rlevantes obtenidos}}{\text{Total de estudios obtenidos}}\right) \times 100\%$

La siguiente tabla muestra la terminología utilizada en la especificación de las cadenas de búsqueda en la Tabla 2.

Término	Significado
TE	Total de estudios obtenidos por la cadena de búsqueda
TER	Total de estudios relevantes obtenidos
ER	Número de estudios relevantes obtenidos (seleccionados)
CB	Cadena de búsqueda
ID	Identificados de cadena de búsqueda
EF	Effort (Precision) en porcentaje %
RC	Recall (Sensitivity) en porcentaje &

Tabla 2: Cadenas de búsqueda

ID	CB	ER	TER	TE	Recall	Effort
CB-1	("Domain-driven design" OR "domain driven design" OR "DDD" OR "bounded context" OR "decompos*") AND ("microservice*" or "micro-service" OR "API" OR "REST API")	13	25	64	No válido	39.06%
CB-2	("Domain-driven design" OR "domain driven design" OR "domain driven model design" OR ("DDD" AND "software") OR "DDD patern") AND ("microservice*" OR "micro-service*" OR "API") AND ("decompos*" OR "microservice identification" OR "split*")	10	10	12	No válido	No válido
CB-3	("Domain-driven design" OR "domain driven design" OR "domain driven model design" OR ("DDD" AND "software") OR "DDD patern") AND ("microservice*" OR "micro-service" OR "API") AND ("decompos*" OR "using" OR "use" OR "identification" OR "appl*" OR "split*" OR	8	15	30	53.33%	50%

	"modeling")					
CB-4	("Domain-driven design" OR "domain driven design" OR "domain driven model design" OR "Domain driven design pattern" OR ("DDD" AND "software") OR "DDD patern") AND ("microservice*" OR "micro-service" OR "API" OR "decompos*" OR "microservice identification" OR "split")	12	16	34	75%	75.05%
CB-5	("Domain-driven design" OR "domain driven design" OR "DDD pattern") AND ("microservice" OR "microservices" OR "micro-service" OR "decompose" OR "decomposition")	14	17	33	82.35%	47.05%
CB-6	("Domain-driven design" OR "domain driven design" OR "(DDD)" OR "DDD pattern") AND ("microservice" OR "microservices" OR "micro-service" OR "decompose" OR "decomposition")	14	22	31	63.63%	70.96%

Respecto a los resultados obtenidos tras la evaluación de cada cadena de búsqueda, cuyos resultados son mostrados en la Tabla 7, podemos concluir lo siguiente:

- Tanto la cadena 1 como la cadena 2 obtienen estudios que no alcanzan el número de estudios seleccionados (14 estudios). Dentro del conjunto de resultados faltaban aquellos que habían sido seleccionados y que respondían al menos 1 pregunta de investigación planteada, por lo que quedan descartadas y no son válidas para efectuar la búsqueda automatizada.
- La cadena 4 obtiene un porcentaje de Recall y Effort bajo, además, el número de estudios no alcanza los seleccionados, sino que se pierden 2 estudios, mismo caso que ocurre con la cadena 3.
- Las cadenas 5 y 6 obtienen el número de estudios bien conocidos, sin embargo, la cadena 6 presenta un Recall inválido, a diferencia de la cadena 5 que cumple con las recomendaciones de Zhang.

A partir de las conclusiones anteriores, la cadena de búsqueda 5 fue seleccionada para efectuar la búsqueda automatizada.

CB-5: ("Domain-driven design" OR "domain driven design" OR "DDD pattern") AND ("microservice" OR "microservices" OR "micro-service" OR "decompose" OR "decomposition")

3.4 Selección de motores de búsqueda

Los motores de búsqueda han sido seleccionados a partir del raking propuesto por (Zhang et al., 2011), así como resultado de la ubicación de conferencias o Journals con investigaciones relevantes de acuerdo con las preguntas de investigación, seleccionando 4 buscadores donde se localizaron los estudios relevantes en la búsqueda manual y automatizada:

1. IEEE Explorer
2. ACM Digital Library
3. ScienceDirect
4. SpringerLink

Donde la fuente IEEE Explorer ha sido considerada como principal motor de búsqueda para evaluación de la cadena de búsqueda seleccionada debido a que fue el motor de búsqueda donde se localizó el mayor número de estudios relevantes. Como principales categorías de los estudios en los buscadores se ha considerado aquellos que provienen de alguna conferencia, workshop o journal únicamente, excluyendo artículos de revistas u otras categorías.

4. Selección de los estudios primarios

La selección de los estudios primarios toma lugar en la etapa de identificación de los estudios relevantes al tema de investigación, al igual que la estrategia de búsqueda, solo que la selección de estudios comprende la recolección de los estudios relevantes a través la lectura y aplicación de criterios de inclusión y exclusión:

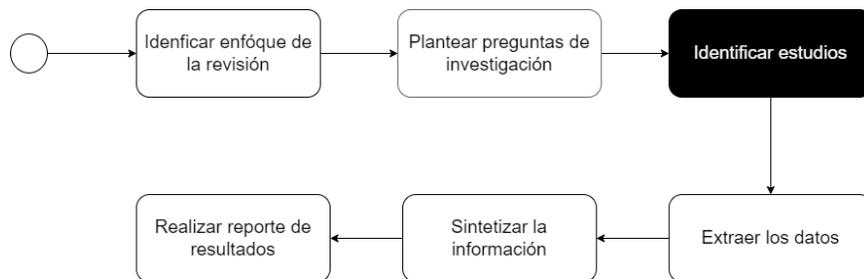


Figura 5: Actividades del proceso de mapeo sistemático (Popay et al., 2006)

4.1 Criterios de selección de estudios primarios

Se plantearon 5 criterios de inclusión y 4 criterios de exclusión para la selección de los estudios primarios a través de un proceso de selección anteriormente definido:

Criterios de inclusión:

- CI-1. Estudios publicados entre 2014 y 2022.
- CI-2. El estudio fue publicado en idioma inglés.
- CI-3. El título presenta indicios de contestar al menos 1 pregunta de investigación.
- CI-4. El resumen da indicios de contestar al menos 1 pregunta de investigación.
- CI-5. El texto del estudio contesta al menos 1 pregunta de investigación.

Criterios de exclusión

- CE-1. El estudio no forma parte de alguna conferencia o journal, workshop.
- CE-2. No se tiene acceso al estudio completo.
- CE-3. Se hace uso de Domain-Driven Design en un contexto distinto a microservicios.
- CE-4. El estudio es duplicado.

4.2 Procedimiento de selección de estudios primarios

El procedimiento seguido para la selección de los estudios primarios se basa en 5 etapas, donde cada una de estas etapas es conformada por la aplicación de un conjunto de criterios de selección de estudios. Este procedimiento se muestra a continuación en la Figura 6:

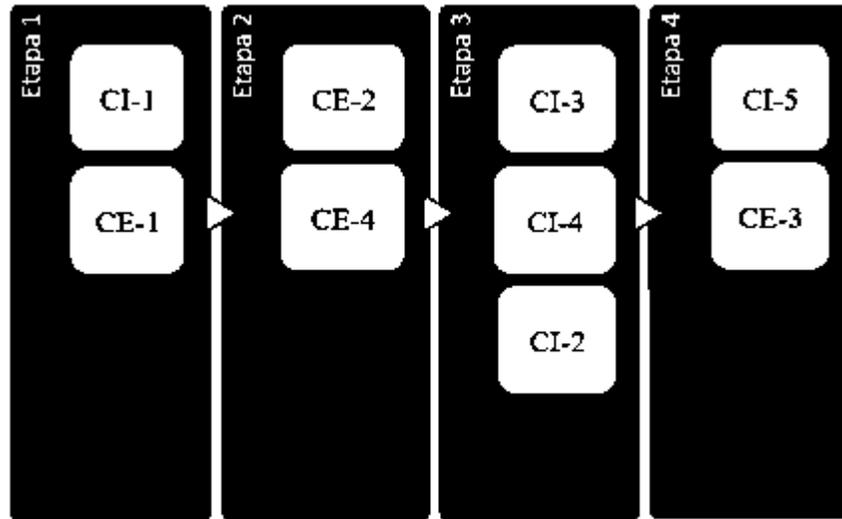


Figura 6: Proceso de selección de estudios primarios

Etapa	Criterio aplicado	Motivación
Etapa 1	CI-1, CI-2	<ul style="list-style-type: none"> Filtración de los estudios a través de las herramientas de filtro de los motores de búsqueda seleccionados.
Etapa 2	CE-2, CE-3	<ul style="list-style-type: none"> Exclusión directa de los estudios no accesibles o que no correspondan a Journals, Conferencias o Workshops sin contemplar criterios de inclusión.
Etapa 3	CI-3, CI-4	<ul style="list-style-type: none"> Lectura de título y abstract como elementos con contenido breve que permite descartar o seleccionar el estudio sin leerlo completo
Etapa 4	CI-5, CE-3, CE-4	<ul style="list-style-type: none"> Proceso de refinamiento de estudios seleccionados a partir de la lectura completa del texto

5. Extracción de los datos

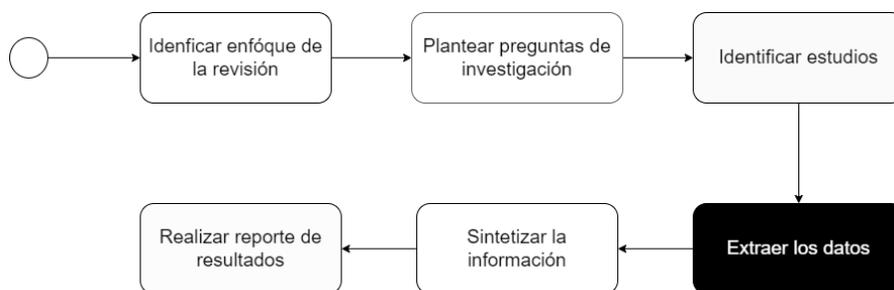


Figura 7: Actividades del proceso de mapeo sistemático (Popay et al., 2006)

Continuando con la extracción de los datos, representada por el diagrama de actividad de la figura 7, tenemos la siguiente plantilla para el informe de los datos de los estudios seleccionados:

DATOS GENERALES	
ID	Identificador único del estudio generado con la nomenclatura EP-#, que corresponde a Estudio Primario (EP) - #número.
DOI	Identificador único del estudio publicado.
Título	Título del estudio.
Autores	Responsables de la publicación del <u>estudio</u> .
Año de publicación	Año de publicación que debe estar entre 2015 y 2022 alineándose a los criterios de inclusión.
Fuente	URL donde se encuentra el estudio.
Base de datos	Motor de búsqueda donde fue encontrado el estudio [IEEE Explorer, ACM Digital library, Science Direct o SpringerLink].
Tipo de publicación	Publicación en [Conferencia, Journal o Workshop].
Venue	Nombre del congreso o journal donde fue presentado el estudio.
Palabras clave	Palabras clave establecidas en el estudio.
<i>Abstract</i> o resumen	Resumen establecido en el estudio.
Pregunta/s de investigación relacionada/s	Preguntas de investigación a las cuales el contenido del estudio responde.

PI-1. ¿Cuáles son los propósitos del uso de Domain-Driven Design en el desarrollo de sistemas con arquitectura de microservicios?	Capturar la finalidad reportada por los autores de los estudios primarios respecto al uso de DDD en los sistemas de microservicios.
PI-2. ¿Qué evidencia del uso de Domain-Driven Design en el desarrollo de sistemas con arquitectura de microservicios?	Capturar los artefactos en los que se presenta el uso de los patrones de DDD en el desarrollo de los sistemas de microservicios.
PI-3. ¿Cuáles son los patrones de Domain-Driven Design utilizados en el desarrollo sistemas con arquitectura de microservicios?	Capturar los patrones de DDD y la finalidad del uso de estos en el desarrollo de sistemas de microservicios.
PI-4. ¿Qué tecnologías son utilizadas con Domain-Driven Design en el desarrollo de sistemas con arquitectura de microservicios?	Capturar las tecnologías que son utilizadas con DDD o junto a DDD para el desarrollo de sistemas de microservicios.
PI-5. ¿Cuáles son las técnicas utilizadas en el desarrollo de sistemas con arquitectura de microservicios diseñados con DDD?	Capturar las técnicas que utilizan DDD o complementan el uso de DDD en el desarrollo de sistemas de microservicios.
PI-6. ¿Cuáles son las limitantes encontradas en el uso de Domain-Driven Design para el desarrollo de sistemas con arquitectura de microservicios?	Capturar las limitantes que reportan los autores en el uso de DDD para el desarrollo de sistemas de microservicios.
PI-7. ¿Qué propuestas existen para el desarrollo de sistemas con arquitectura de microservicios utilizando Domain-Driven Design?	Capturar lo que los autores proponen para desarrollar sistemas de microservicios con el uso de DDD.

6. Estrategia para la síntesis de datos

Como parte del mapeo sistemático de la literatura, tenemos como actividad clave, llevar a cabo la extracción de los datos y su sinterización, pues es esto lo que nos permite obtener respuestas fundamentadas (hallazgos) para nuestras preguntas de investigación. Como parte de las actividades del proceso de revisión sistemática mostrado en la Figura 8, tenemos la selección de una síntesis narrativa para efectos de la realización de una síntesis de hallazgos en los estudios primarios.

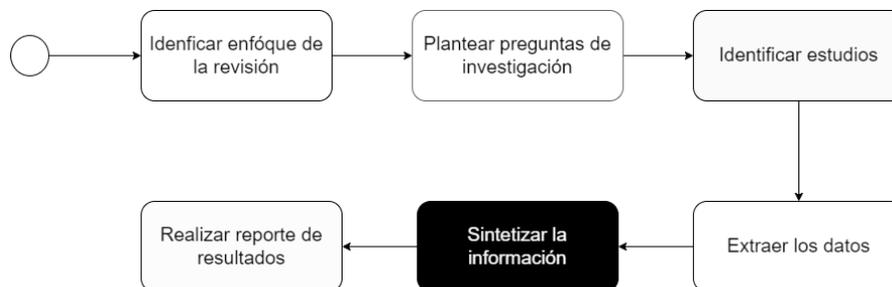


Figura 8: Actividades del proceso de mapeo sistemático (Popay et al., 2006)

La síntesis narrativa refiere a un enfoque para la revisión sistematizada y síntesis de hallazgos de múltiples estudios que se basa principalmente en el uso de las palabras en el texto para resumir y explicar los hallazgos de la síntesis (Popay et al., 2006). De acuerdo con (Popay et al., 2006) el uso de una síntesis narrativa refiere al proceso de síntesis que puede ser usado en enfoques de revisión sistemática para una gran gama de preguntas. La síntesis narrativa comprende 4 elementos identificados para llevar a cabo el proceso, basándonos en la propuesta de (Popay et al., 2006):

Elemento de la síntesis a utilizar	Propósito	Estrategias
Explorar relaciones en los datos	<ul style="list-style-type: none"> Relacionar hallazgos y agruparlos en “clusters”. 	<p>Uso de herramienta gráficas: Se usarán graficas para explorar las relaciones entre los estudios como una técnica propuesta por (Popay et al., 2006).</p> <p>Redacción de texto donde se resuman los hallazgos de forma ordenada y concisa (Popay et al., 2006).</p>

7. Limitaciones

Como limitaciones de la revisión sistemática de la literatura tenemos las siguientes:

- LI-1. El enfoque del mapeo sistemático comprende la recolección de casos de aplicación de Domain-Driven Design para Microservicios en investigaciones y la sinterización de los estudios sin proponer buenas o malas prácticas.
- LI-2. Los casos del uso de Domain-Driven Design se limitan al uso del estilo arquitectónico de Microservicios, excluyendo otros estilos orientados a los servicios como SOA.
- LI-3. El enfoque de la revisión sistemática no está basado en la evaluación de la calidad de implementación de Microservicios ni de Domain-Driven Design.
- LI-4. Se dispone de un tiempo aproximado de 2 meses para realizar el proceso de revisión sistemática de la literatura.

8. Informe

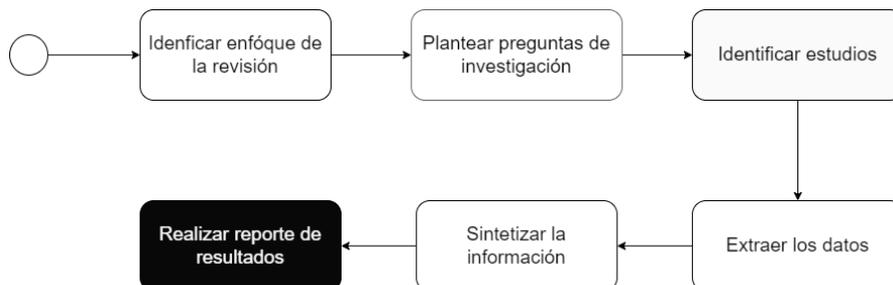


Figura 9: Actividades del proceso de mapeo sistemático (Popay et al., 2006)

Como índice propuesto para el informe de la revisión sistemática de la literatura donde se plasmarán los resultados a través de un reporte como lo muestra el diagrama de actividad de la Figura 9, se ha definido la siguiente estructura de índices con base a las directrices de (Kitchenham et al., 2015).

- Título
- Resumen
 - Contexto
 - Objetivos
 - Métodos
 - Resultados
 - Conclusiones
- Palabras clave
- Contexto o introducción
 - Preguntas de investigación
- Método de revisión sistemática
 - Criterios de inclusión y exclusión
 - Proceso de búsqueda
 - Evaluación de la calidad
 - Extracción y síntesis de datos
- Resultados
 - Respuesta a preguntas de investigación
- Discusión
- Amenazas a la validez
- Conclusiones
- Referencias

1. Gestión de la revisión

Para llevar un control de la revisión sistemática como proceso de gestión, se ha elaborado un cronograma de actividades en el cual aparecen las actividades principales para efectuar la revisión, así como las fechas planeadas para la revisión del progreso.

Actividades	Septiembre				Octubre				Noviembre				Diciembre				Enero			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Desarrollo de primera versión el protocolo	■	■	■	■	■	■	■													
Primera presentación de avances					■															
Proceso de búsqueda						■	■	■	■	■										
Selección de estudios primarios						■	■	■	■	■	■									
Obtención de datos de los estudios							■	■	■	■	■	■								
Síntesis de la información								■	■	■	■	■	■	■						
Segunda presentación de avances										■										
Redacción del reporte														■	■	■	■	■		
Tercera presentación																			■	

2. Referencias

- Kitchenham, B., Budgen, D., & Brereton, P. (2015). *Evidence-based software engineering and systematic reviews* (1st ed., Vol. 4).
- Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., & Britten, N. (2006). Guidance on the conduct of narrative synthesis in systematic Reviews. A Product from the ESRC Methods Programme. Version 1. *Undefined*. <https://doi.org/10.13140/2.1.1018.4643>
- Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering. *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering - EASE '14*, 1–10. <https://doi.org/10.1145/2601248.2601268>
- Zhang, H., Babar, M. A., & Tell, P. (2011). Identifying relevant studies in software engineering. *Information and Software Technology*, 53(6), 625–637. <https://doi.org/10.1016/j.infsof.2010.12.010>