UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

COLEGIO DE CIENCIAS E INGENIERÍAS

Métodos empíricos del estudio del estado del arte – caso de estudio: *A Systematic Mapping Study on Security Compliance for Agile Software Development*

Pamela Elizabeth Almeida Salazar

Ingeniería en Sistemas

Trabajo de titulación presentado como requisito

para la obtención del título de

Ingeniera en Sistemas

Quito, 3 de diciembre de 2019

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ
COLEGIO DE CIENCIAS E INGENIERÍAS

HOJA DE CALIFICACIÓN
DE TRABAJO DE TITULACIÓN

Métodos empíricos del estudio del estado del arte – caso de estudio: *A Systematic Mapping Study on Security Compliance for Agile Software Development*

Pamela Elizabeth Almeida Salazar

Calificación:	
Nombre del profesor, Título académico	Daniel Riofrío, Ph.D. in Computer Science
Firma del profesor	

Quito, 3 de diciembre de 2019

Derechos de Autor

Por medio del presente documento certifico que he leído todas las Políticas y Manuales de la Universidad San Francisco de Quito USFQ, incluyendo la Política de Propiedad Intelectual USFQ, y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo quedan sujetos a lo dispuesto en esas Políticas.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo en el repositorio virtual, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

Firma del estudiante: _____

Nombres y apellidos: Pamela Elizabeth Almeida Salazar

Código: 00127708

Cédula de Identidad: 172197364-0

Lugar y fecha: Quito, 3 de diciembre de 2019

RESUMEN

El desarrollo de estudios sistemáticos del estado del arte es un tipo de investigación documental altamente utilizada en la academia que provee bases para proponer, evaluar, validar y experimentar permitiendo el desarrollo de conocimiento en las diferentes áreas. Se pueden utilizar diferentes metodologías de acuerdo con el nivel de rigurosidad y calidad deseado. Estas características son establecidas por los investigadores y dependen de la naturaleza del estudio, los resultados esperados y disponibilidad del equipo de trabajo. La desventaja de este proceso es que el tiempo invertido en el desarrollo de los estudios sistemáticos del estado del arte es ineficiente dada la extensa documentación y aumenta con un nivel de rigurosidad más alto. Una solución propuesta es el uso de herramientas que ayuden a reducir el tiempo manteniendo la rigurosidad. Este documento de titulación incluye y explica diferentes metodologías para el desarrollo del estudio sistemático del estado del arte y herramientas disponibles para las diferentes fases. Adicionalmente, se discute, brevemente, un caso de estudio sobre la importancia de incorporar actividades de seguridad basadas en estándares en las metodologías ágiles para industrias críticas, en el que se detalla fase a fase la metodología utilizada, resultados cuantitativos junto una descripción de las herramientas desarrolladas y utilizadas en el proceso. Finalmente, como trabajo futuro se plantea el desarrollo de una metodología que incorpore herramientas, la mejora de las herramientas desarrolladas y para el caso de estudio, el desarrollo de un marco conceptual que incorpore actividades de seguridad basadas en estándares en las metodologías ágiles.

Palabras clave: *estudio sistemático del estado del arte, metodologías ágiles, seguridad, estándares, herramientas.*

ABSTRACT

The development of systematic mapping studies is a type of documental research widely used in academia that provides the basis for proposing, evaluating, validating and experimenting allowing the development of knowledge in different areas. Different methodologies can be used according to the level of rigor and quality desired. These characteristics are established by the researchers and depend on the nature of the study, the expected results and availability of the research group. The disadvantage of this process is that the time invested in the development of systematic mapping studies is inefficient given the extensive documentation and increases with the level of rigor. A proposed solution is the use of tools that help reduce the time while maintaining rigor. This document includes and explains different methodologies for the development of the systematic mapping and the tools available for the different phases. Additionally, this document discuses a case study about the importance of incorporating security compliance activities in agile methodologies for critical industries, including the methodology used, quantitative results and a description of the tools developed and used in the process. Finally, as future work, this work proposes the development of a methodology that incorporates tools, the improvement of tools already developed and for the case study; the development of a framework that incorporates security compliance activities in agile methodologies.

Keywords: *systematic mapping study, agile methodologies, security, compliance, tools*

TABLA DE CONTENIDOS

ÍNDICE DE FIGURAS

INTRODUCCIÓN

Importancia de los Estudios Empíricos

Los avances en las diferentes diciplinas dependen tanto de los estudios empíricos como de la formación de teorías y el desarrollo de herramientas. Los estudios empíricos se enfocan en crear modelos para expresar conocimiento en diferentes dominios de interés. Este conocimiento debe ser analizado y sintetizado para un posterior desarrollo. Debido a que el desarrollo científico es realizado por humanos, es necesario estudiar sus actividades para entender y mejorar el trabajo desarrollado. El proceso de aprendizaje es continuo y evolutivo. Por lo tanto, una buena relación entre la industria y la academia es necesaria (Basili & Zelkowitz, 2007).

En las introducciones de los diferentes estudios científicos, los autores suelen incluir un breve contexto basados en publicaciones seleccionadas por el autor. Esto hace que el estudio no sea confiable debido a que incluyen un sesgo. Es posible eliminar este sesgo mediante un estudio del arte sistemático, ya que en este se incluyen datos considerando a todos los estudios que se han realizado en un área determinada en un tiempo determinado. Además, como lo dice su definición debe ser “Un método sistemático, explícito y reproducible de identificar, evaluar y sintetizar un conjunto de trabajo completado y documentado por investigadores, escolares y practicantes” (Okoli & Schabram, 2010). Estos factores dan al estudio más credibilidad y por lo tanto sirven como base para futuras investigaciones (Cooper, 2016). Cuando un área de estudio madura, por lo general, hay un gran incremento en el número de resultados y reportes disponibles, por lo cual es importante sintetizarlos. Los estados de arte metodológicos son utilizados satisfactoriamente para identificar tendencias y mejorar las prácticas en las diversas áreas de investigación (J. Randolph, Julnes, Sutinen, & Lehman, 2008).

Un estudio del arte sistemático es también recomendado cuando se estudia un área con pocos estudios primarios. Una publicación de un estudio del arte generalmente contiene una breve introducción, el contexto, la metodología utilizada, los resultados, una discusión y sus respectivas conclusiones. Adicionalmente, se suelen incluir gráficos y tablas cualitativas y cuantitativas para facilitar la comprensión del estudio (Petersen, Feldt, Mujtaba, & Mattsson, 2008).

La metodología de los estudios del arte sistemáticos depende del área de investigación y su aplicabilidad. Generalmente se componen de una definición de preguntas de investigación, revisión del espectro, revisión del panorama, conducción de la búsqueda, obtención de publicaciones relevantes, clasificación, extracción de datos y finalmente el mapeo sistemático (Petersen et al., 2008).

Organización del Documento

El siguiente documento incluye una descripción de las diferentes metodologías para realizar estudios del arte sistemáticos. Posteriormente, se adiciona una breve descripción de la metodología utilizada en el caso de estudio: A Systematic Mapping Study on Security Compliance for Agile Software Development. El reporte del caso de estudio realizado será incluído en una publicación científica. Se continúa con una introducción a los procesos que pueden ser automatizados en los estudios del arte. Además, se incluye un ejemplo experimental de la inclusión de scripts en el proceso. Se prosigue con las herramientas utilizadas y las áreas de mejora. Adicionalmente, se incluyen los valores macro del caso de estudio. Finalmente, se incluyen conclusiones y trabajo futuro en esta área.

METODOLOGÍA

Metodologías para el estudio del arte sistemático

Existen diferentes niveles de rigurosidad para realizar un estudio del arte sistemático. Un estudio puede ser tan sencillo como una selección anotada de bibliografía, así como una síntesis científica de una búsqueda primaria. A continuación, se incluyen 4 procedimientos rigurosos de como realizar un estudio del arte sistemático. En particular, se describen las fases que componen cada metodología. Los tres primeros corresponden a guías mientras que el cuarto corresponde a la metodología utilizada para el caso de estudio, por lo cual se incluye un mayor detalle.

Metodología Utilizada en el Caso de Estudio

Para el estudio, *A Systematic Mapping Study on Security Compliance for Agile Software Development*, se utiliza la guía propuesta por Kurhmann, Méndez y Daneva. La cual se basa en su experiencia y fue desarrollada para estudios de ingeniería de software. En esta guía se incluyen alternativas para complementar el mapeo; sin embargo, aquí se detallan de forma específica aquellas fases que se ocuparon para el caso de estudio.

Preparación

En la fase de preparación se incluyen las preguntas de investigación, junto con el conjunto de criterios de inclusión y exclusión. Este proceso se encuentra estrechamente relacionado con la planificación previa que realiza el equipo de trabajo.

Preguntas de Investigación

Debido a que cada estudio debe tener un objetivo bien definido, es necesario que las preguntas se encuentren relacionadas con el área que cubre. Estas preguntas por lo general tratan de clasificar a las publicaciones de acuerdo con el lugar en donde fueron publicados, los

autores y sus contribuciones al área de estudio. Es importante que los autores detallen no solo las preguntas sino también la perspectiva de cada una. Las preguntas por lo general se realizan al inicio del estudio, pero es posible modificarlas mientras se avanza con el proceso.

Cadenas de Búsqueda

Para cada una de las fuentes bibliográficas, se genera un conjunto de cadenas de búsqueda. Estas cadenas no solo incluyen el formato de cada fuente junto con los debidos conectores, pero también incluyen un conjunto de palabras claves relevantes al área de estudio. Las palabras claves deben ser seleccionadas por los miembros del equipo luego de los procesos de prueba de error y de bola de nieve; y en caso de que se tengan estudios conocidos como relevantes en el área, se pueden incluir las palabras claves detalladas en éstos. Además, es recomendable buscar sinónimos que se utilizan en la academia para las palabras clave.

Bola de Nieve

Se utilizan publicaciones conocidas previamente, se sugiere utilizar una nube de palabras que considere las frecuencias de todas las palabras clave de las distintas publicaciones. La nube de palabras permite visualizar las palabras con más frecuencia con letra de mayor tamaño por lo cual simplifica el proceso de selección para las cadenas de búsqueda.

Búsqueda de Prueba y Error

Se puede utilizar la búsqueda por medio de meta indexadores como Scopus para verificar brevemente si los resultados con esas palabras claves seleccionadas tienen relación con el área de estudio. Es importante también realizar un análisis cuantitativo, este puede ser en porcentaje, por ejemplo: si se consideran las primeras 100 publicaciones, de éstas cuántas se consideran como potenciales relevantes al estudio. Si se cuenta con estudios de los cuales se conoce previamente, se debe considerar si entre los resultados de las cadenas de prueba, se puede encontrar a los mismos, o a su vez, publicaciones por los mismos autores.

Criterios de Inclusión y Exclusión

Los criterios de inclusión son aquellos que permiten que una publicación se incluya en el conjunto resultante. Por lo general se encuentran relacionados con las preguntas de investigación y el área de estudio. Los criterios de exclusión corresponden a características que deben tener las publicaciones. Como ejemplos podemos listar: el que las publicaciones tengan el título o un resumen relacionado con el tema de estudio, que el texto completo de la publicación se encuentre disponible o que ya se lo haya considerado en el estudio. Esto quiere decir que hubo duplicación de resultados con diferentes cadenas de búsqueda o con las librerías.

La fase de planificación se puede dar por concluida cuando todos los miembros del equipo hayan llegado a un consenso sobre las preguntas, palabras clave y cadenas de búsqueda. Debido a que mediante se vaya avanzando con las diferentes fases, se pueden encontrar nuevas palabras clave o características que no se habían considerado previamente, este proceso se vuelve iterativo. La documentación de cada uno de los cambios es indispensable para permitir la repetitividad.

Recolección de Datos

Fuentes de Datos

Las fuentes más usadas son las librerías digitales. De acuerdo con Kurhmann, Méndez y Daneva, se sugiere el uso de IEEE, ACM y SpringerLink. Se deben generar cadenas de búsquedas para cada una de las librerías seleccionadas.

Revisión del Conjunto de Resultados y Búsqueda primaria

Para cada una de las cadenas, se deben verificar los resultados tanto cualitativamente como cuantitativamente. En esta sección también se utiliza la búsqueda de las publicaciones antes conocidas. Se sugiere incluir tablas con las diferentes versiones de las cadenas de

búsqueda y los resultados. Hay que considerar que pequeñas variaciones en los conectores, el uso de sinónimos o los paréntesis pueden variar los resultados obtenidos considerablemente. La documentación permite que el equipo realice un análisis en conjunto sobre las versiones finales para cada librería. A partir de las cadenas de búsquedas seleccionadas, se procede a validar la búsqueda.

Prácticas de Exportación

Se procede a la extracción de atributos para cada publicación obtenida en el conjunto de resultados. Es importante mantener una coherencia de los atributos a extraer en todas las librerías. Usualmente se obtiene el título, resumen, año de publicación, dónde fue publicado, fuente y cadena de búsqueda relacionada.

Limpieza de los Datos

Unificación y Reducción del Conjunto de Datos

Para el proceso de unificación se sugiere utilizar los títulos como identificador. Debido a los diferentes formatos manejados por las librerías se sugiere el usar mayúsculas. Dependiendo de la librería, es posible que algunas publicaciones nada más consten de un título pero que no brinden más información por lo cual se eliminan.

Completar los Datos Faltantes Sabiamente

Debido a los diferentes formatos, es posible que algunos atributos no se encuentren disponibles y si se encuentran, no sean compatibles. Por ejemplo, algunas librerías incluyen solo el año de publicación, otras la fecha, por lo cual se debe seleccionar un formato en común. Para los atributos no disponibles, se suele encontrar “N/a” o “nan”. En caso de que se haya incluido como criterios de exclusión la falta de atributos completos, se deben descartar estas publicaciones.

Estructura de los Datos

Para que haya uniformidad en los atributos, se sugiere realizar una tabla en donde se enlisten los campos, la cardinalidad y una pequeña descripción del atributo extraído.

Selección de Estudios

Definir la Selección de Estudios y el Enfoque

Se sugiere una reunión del equipo para determinar los criterios de votación. Es importante definir la relación entre los criterios para formar conjuntos y subconjuntos. Usualmente el criterio más específico es aquel que se va a encontrar más relacionado con el área de estudio y las preguntas de investigación. Se sugiere el uso de un identificador y de un gráfico que se pueda tener al alcance cuando se realice el proceso de votación para evitar confusiones.

Procedimiento de Votación

Para este proceso se sugiere el uso de una herramienta de votación. La votación depende del número de críticos en el equipo y del nivel de experiencia de cada uno.

Votación por Mayoría

Para que este proceso sea imparcial y a la vez práctico. Cada uno de los miembros del equipo recibe el mismo conjunto de datos, no pueden ver las respuestas del otro participante, se mantiene un registro de los criterios que corresponden a cada publicación, puede incluir un comentario y la acepta o rechaza. En un equipo de tres personas, en caso de que una publicación sea aceptada por dos miembros del equipo, se acepta. Si fue negada por dos miembros, se rechaza. Si hay una que acepta y una que rechaza, el tercer miembro del equipo vota sobre esa publicación y se decide por mayoría simple.

Integrar y Finalizar el Conjunto y Reportar

En esta fase se integran las publicaciones que fueron aceptadas en la primera fase de votación con las que fueron aceptadas en la segunda fase. Se recomienda el realizar una tabla que permita visualizar todo el proceso realizado previamente de forma cuantitativa por librería hasta la fase de votación. Obteniendo, así, el conjunto final.

Análisis de Datos

Para el análisis de datos se utilizan gráficos y tablas que permitan resumir la información antes obtenida relacionada con las preguntas de investigación. Es importante incluir una descripción analítica de los datos más importantes en cada gráfico (Kuhrmann, Fernández, & Daneva, 2016).

Otras Metodologías

Metodología propuesta por Okoli y Schabram

1. **Planificación**
    - **a.** Propósito de la revisión de la literatura: identificar los propósitos de la revisión, es necesario que sea explícito para los lectores.
    - **b.** Protocolo y entrenamiento: para las visiones que cuentan con más de un crítico, es necesario llegar a un acuerdo con respecto a los criterios.
2. **Selección**
    - a. Búsqueda de la Literatura: Se debe describir en detalle la búsqueda de la literatura y las justificaciones que llevan a la comprensión.
    - b. Visualización práctica: un crítico debe establecer el conjunto de publicaciones a incluir y los motivos por los cuales no se consideran a los otros estudios.

3. **Extracción**
    a. Evaluación de Calidad: dependiendo de la calidad deseada por los críticos, estos van a establecer criterios particulares para que una publicación sea considerada más adelante.
    b. Extracción de Datos: los críticos extraen la información relevante al estudio de cada publicación.
4. **Ejecución**
    a. Síntesis de los Estudios: se utilizan diversas técnicas para combinar los datos extraídos en la fase anterior.
    b. Escribir la reseña: introducir detalles que permitan la reproducción del proceso (Okoli & Schabram, 2010).

Metodología propuesta por Petersen

1. **Definición de las Preguntas de Investigación**: trazar las frecuencias de las publicaciones con respecto al tiempo.
2. **Revisión del Alcance**: mediante el uso de cadenas de búsqueda en las bibliotecas digitales.
3. **Conducir la Búsqueda**: se utilizan los criterios de inclusión y exclusión para no incluir publicaciones que no son relevantes con respecto a las preguntas de investigación.
4. **Extracción de Palabras Claves Utilizando los Resúmenes y Esquema de Clasificación**: se utiliza para simplificar el trabajo de esquematización mediante la segmentación de tópicos relacionados.

5. **Extracción de Datos y Proceso de Mapeo**: este proceso permite visualizar un resumen de los resultados obtenidos en las búsquedas (Petersen et al., 2008).

Metodología propuesta por Ieradi, Orihuela y Jurado

1. **Planificación**
    a. Preguntas de Investigación: deben ser coherentes con el objetivo de la investigación y la delimitación del problema
    b. Bases de Datos: es necesario seleccionar las librerías digitales apropiadas para el estudio de forma que cubran el contenido más relevante y los campos de búsqueda deseados.
    c. Criterios de Inclusión y Exclusión: el propósito de los criterios es obtener las publicaciones que capturen la mayoría de las características para el estudio.
    d. Pregunta Binaria: se encuentra relacionada con las palabras claves del área a investigar para verificar que las librerías digitales seleccionadas sean apropiadas.
2. **Conducción**
    a. Selección de Estudios Primarios: en esta etapa hay una lectura parcial de los documentos por diferentes personas y por último una lectura total de los documentos antes de su selección final.
    b. Extracción y Síntesis de Información: depende de los investigadores y sus técnicas el definir como sintetizar y organizar la información. Es común el uso de tablas y gráficos conceptuales.

3. Reporte

a. Documentación de los datos extraídos: incluir una crítica objetiva y razonada de la literatura incluida (Ierardi, Orihuela, & Jurado, 2018).

HERRAMIENTAS

Estudio Automático

El proceso de desarrollar estados del arte sistemáticos se ha vuelto una metodología muy utilizada. Esta metodología tiene el problema de que consume mucho tiempo debido a que comprende bastante trabajo manual. El uso de herramientas para producir evidencia científica se ha incrementado, ya que permiten aprovechar el avance tecnológico, resultando en publicaciones de propuestas de soluciones. A pesar de su disponibilidad, las herramientas aún se encuentran en las primeras etapas de desarrollo y su uso es en ambientes controlados. Este factor, adicionado al hecho de que las evaluaciones de las herramientas rara vez son realizadas por una entidad externa, demuestra que el área aún no cuenta con una madurez. Aún así, se evidencia que la mayoría de las herramientas documentadas se encuentran relacionadas con la minería de texto para la fase de selección de estudios y síntesis de datos (Marshall & Brereton, 2013).

De acuerdo con L. Fornari, I. Pinho, et. Al, 2019, las áreas que han reportado más uso de las herramientas para el desarrollo de estados del arte son aquellas de las ciencias de la salud, ciencias sociales e ingeniería y tecnología en último lugar con solo una publicación. A partir de un análisis de la calidad metodológica de lo estudios del arte con el uso de herramientas digitales, se puede afirmar una mejora en la recolección, exploración, clasificación de los datos, análisis e interpretaciones de los investigadores. Actualmente, el uso de las herramientas no comprende todo el proceso de desarrollo de un estudio del arte sistemático. Pero presenta un gran potencial debido a la transparencia y credibilidad de los resultados en cada búsqueda.

Herramientas de Automatización

En el capítulo anterior, metodología, se presentaron un grupo de distintas metodologías para realizar un estudio sistemático del estado del arte. A pesar de que ellas difieren en el número de fases, todas tratan los siguientes aspectos: definición del objeto de estudio, selección de fuentes de información, análisis crítico de esas fuentes y desarrollo del documento de resumen del estudio. De estos aspectos la selección de fuentes de información y parte del análisis crítico de esas fuentes puede ser automatizado o semiautomatizado. En particular, durante el desarrollo de este trabajo se detectaron múltiples procesos en cada fase de la metodología utilizada en las cuáles se pueden construir herramientas autónomas o semiautónomas que permitan reducir el tiempo del estudio. En particular, la figura 1 muestra en color rosa los procesos que pueden ser automatizados o semiautomatizados.

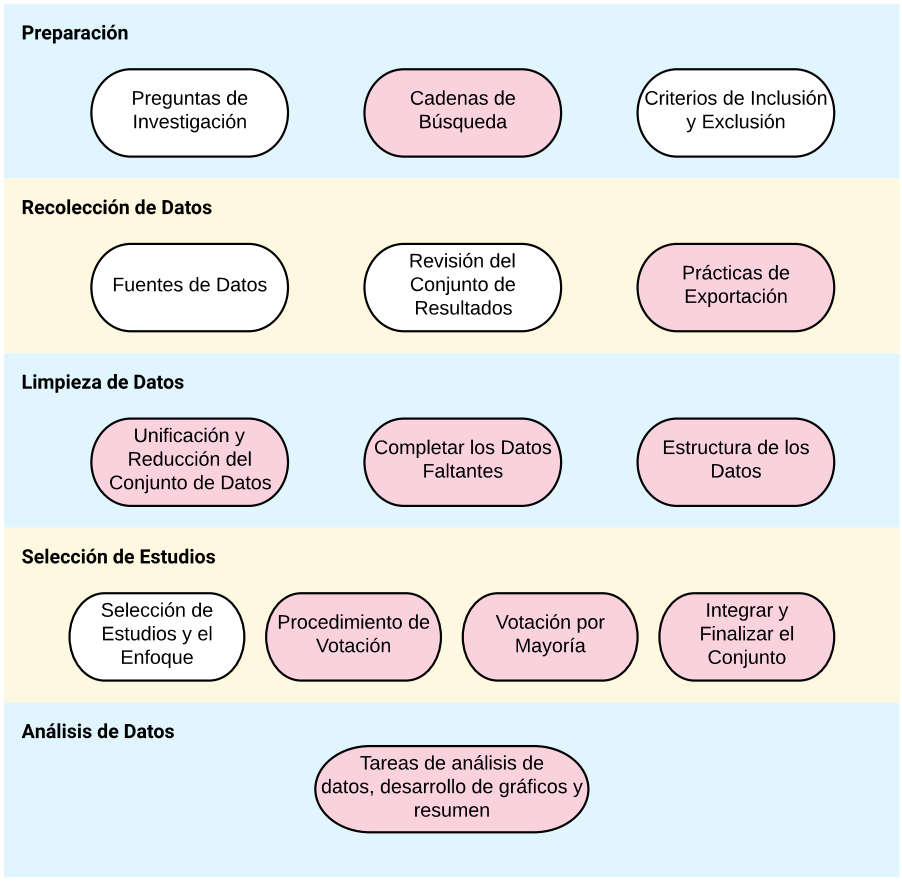

Figura 1: Fases de la Metodología Utilizada

A continuación, se detallan 3 herramientas de automatización que han sido desarrolladas para optimizar el proceso de los estudios sistemáticos. Y, se documentan las herramientas utilizadas durante el caso de estudio que incluyen desarrollos propios (*Crawler*, *Integrator* y *NLP Text Miner*) y la herramienta web para el proceso de votación (*SMS Tool*) provista por los colaboradores externos de este trabajo.

Herramientas Disponibles

Clustering

Este método de automatización utiliza minería de texto con el propósito de optimizar el proceso de clasificación en los estudios sistemáticos. Es posible utilizarla cuando se tiene al menos 30 estudios como conjunto de entrenamiento. Otra técnica consiste en identificar grupos de documentos con combinaciones de palabras similares. Este proceso se considera de mayor ayuda en las fases de revisión de resultados, cuando los investigadores se encuentran con un gran conjunto de datos. Una herramienta disponible es Lingo 3D, la cual realiza agrupaciones con identificadores. Estas agrupaciones son revisadas por los investigadores. Este proceso tiene la debilidad con palabras que tienen múltiples significados en el área de investigación y que los identificadores no siempre son adecuados para la agrupación. Con respecto al proceso manual es mucho más rápido, permite descubrir temas no aparentes, pero depende mucho de la claridad del lenguaje de los estudios incorporados (Stansfield, Thomas, & Kavanagh, 2013).

Thoth – Una herramienta basada en la web para las revisiones sistemáticas

Debido a la necesidad del trabajo en equipo, para facilitar el trabajo entre investigadores, se elige una herramienta basada en la web. Esta herramienta considera las características consideradas como más deseadas y las que se consideran como obligatorias. *Thoth* permite mejorar las cadenas de búsqueda, manejo de divergencia en los criterios de inclusión y exclusión, importar archivos de formato .csv o código .bib. Esta herramienta

permite ahorrar tiempo en todo el proceso del estudio sistemático, pero requiere acceso a Internet y no cuenta con la administración de documentos. La versión detallada anteriormente corresponde a un prototipo, el cual se encontraba en su primera versión, pero aun así mostró resultados satisfactorios (Marchezan, Bolfe, Rodrigues, Bernardino, & Basso, 2019).

CADIMA

Es una herramienta de libre acceso que fue desarrollada usando Scrum para facilitar el proceso de un estado del arte sistemático e incrementar la rigurosidad del estudio. Es una aplicación cliente-servidor que cuenta con almacenamiento hasta por 6 meses con minería de texto integrada. Esta herramienta mantiene un registro de todo el proceso y facilita la comunicación entre los miembros del equipo. El uso de la herramienta se encuentra disponible para cualquier persona que se registre y permite el manejo de roles. Esta herramienta permite revisar el protocolo, facilita la asociación de las cadenas de búsqueda con las máquinas de búsqueda, extracción de datos fuera de línea y síntesis de datos. Esta herramienta a pesar de ser útil en la mayoría de las etapas no cuenta con un soporte en la síntesis cuantitativa ni permite el cambio de roles durante el proceso y no cuenta con remoción automática de los duplicados. Se esperan mejoras de estas deficiencias junto con modificaciones adicionales en la siguiente versión de la herramienta (Kohl et al., 2018).

Herramientas Utilizadas

Las herramientas utilizadas en el caso de estudio semiautomatizan las fases de recolección de datos, limpieza de datos y selección de estudios. Las herramientas *Crawler, Integrator* y *NLP Text Miner* fueron desarrolladas utilizando Python 3.6 y probadas en sistemas operativos Microsoft, macOS y Ubuntu. En el caso de *SMS Tool* es una herramienta desarrollada en PHP 7.3.10 sobre Apache Server 2.4.29

Crawler

Es un *script* desarrollado en *Python* que lee un conjunto de cadenas de búsqueda diseñadas tanto para los repositorios digitales de *IEEE*, *SpringerLink* y *ACM*. Estas cadenas de búsqueda son construidas manualmente revisando la especificación de operadores en cada librería digital y son ejecutados a través de la librería *Selenium* para Python 3 sobre un buscador web que accede a cada librería digital: realiza la consulta especificada en la cadenas de búsqueda, abre uno a uno los resultados y copia la información relevante de cada registro: nombre de autores, año de publicación, título de la publicación y resumen de la publicación. Esta información es recolectada en una hoja de cálculo para ser filtrada por el *Integrator*. La figura 2 resume el proceso de ejecución del *crawler*.

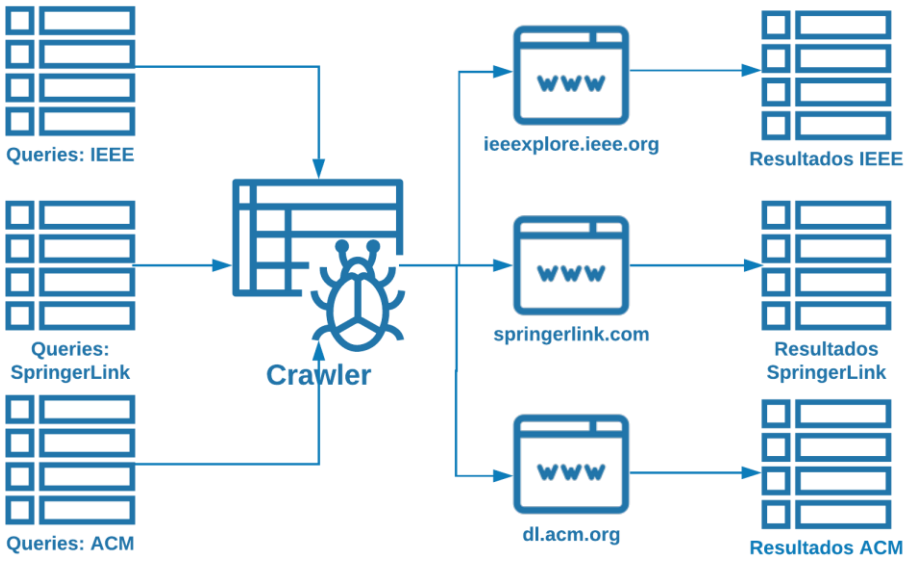

Figura 2: Esquema de funcionamiento del Crawler.

Integrator

Es un *script* desarrollado en *Python* que permite la lectura de los resultados de las diferentes librerías digitales y las consolida en un único repositorio de resultados. Al momento de consolidar, el *script* se encarga de completar información faltante, eliminar duplicados y

adjuntar las cadenas de búsqueda que produjeron ese resultado. La figura 3 muestra el esquema de funcionamiento de esta herramienta.

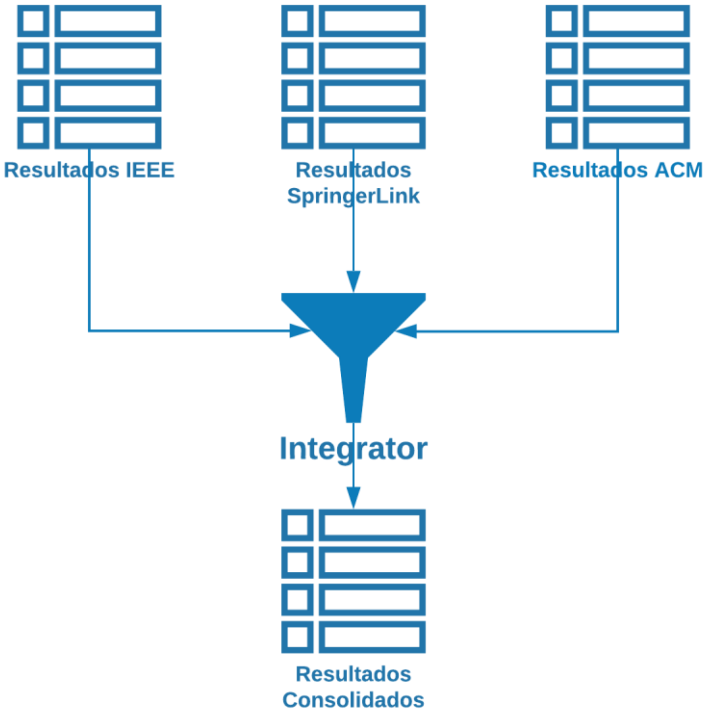

Figura 3: Esquema de funcionamiento del Integrator.

NLP Text Miner

Esta herramienta es una primera versión en busca de automatizar el proceso de filtrado/priorización de los resultados consolidados. En particular, esta herramienta utiliza procesamiento de lenguaje natural (*NLP*) para priorizar a los documentos científicos más parecidos a una base de conocimiento previo el cual es considerado como el conjunto de entrenamiento. En esta primera versión *NLP Text Miner* permite definir un grupo de documentos científicos relevantes iniciales y priorizar los resultados consolidados en función de la cercanía a éstos. Transforma los documentos científicos a arreglos multidimensionales de números en función de la frecuencia absoluta de palabras más relevantes del documento y mide la distancia euclidiana con respecto al centro de los documentos científicos relevantes. Esta estrategia permite producir un grupo de documentos priorizados, sin embargo, en la práctica

esta priorización solo muestra un sesgo ante los estudios parecidos a la base de documentos inicial. La figura 4 muestra el esquema de funcionamiento de esta herramienta.

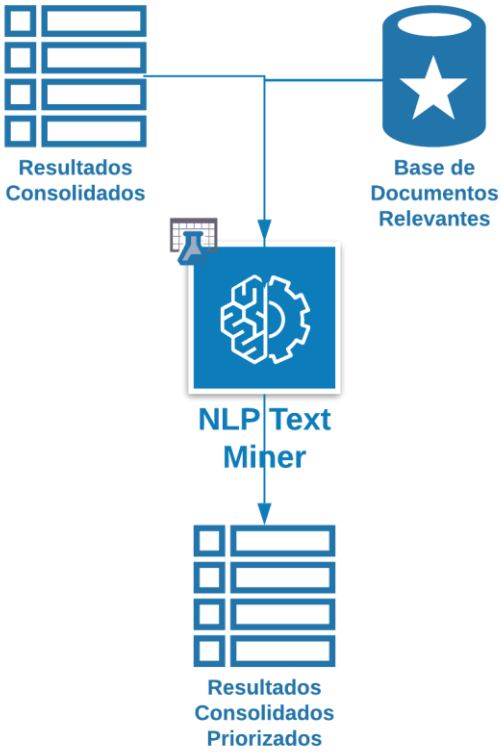

Figura 4: Esquema de funcionamiento de NLP Text Miner.

SMS Tool

El *SMS Tool* es una aplicación web que toma los resultados consolidados y los integra en una base de datos *SQLite* para el proceso de votación. Esta aplicación puede ser configurada para cualquier número de personas que pertenezcan al grupo de investigación para acceso remoto. Permite agregar los criterios de votación y muestra un formulario para calificar si el documento científico es relevante o no a la investigación según cada criterio de votación. Adicionalmente, la aplicación muestra de forma aleatoria cada uno de los documentos científicos (título, resumen, autores y año de publicación) a cada investigador. La herramienta marca los artículos con votación de mayoría, aquellos rechazados y los que tienen votación inconclusa. Finalmente, permite la exportación de la base de datos. Esta aplicación fue

desarrollada y proporcionada por Severin Kacianka colaborador externo del caso de estudio. La figura 5 permite visualizar el esquema de funcionamiento de esta herramienta.

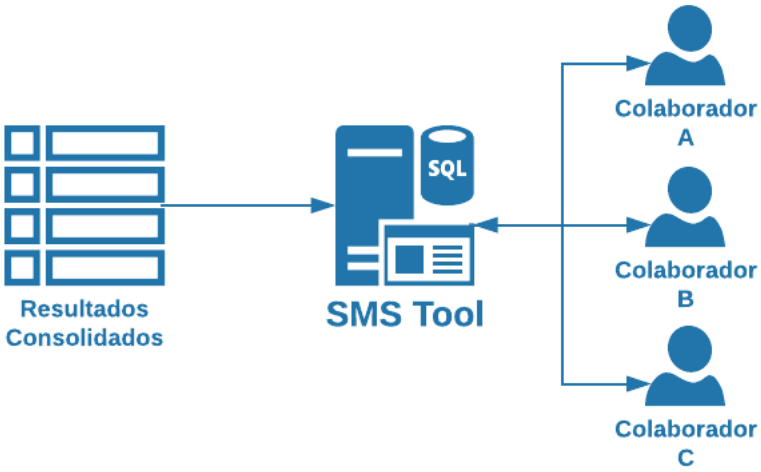

Figura 5: Esquema de funcionamiento de SMS Tool.

CASO DE ESTUDIO

Introducción

El incremento del desarrollo tecnológico ha aumentado exponencialmente en este siglo. Junto con este incremento, las vulneraciones a la seguridad también han ido en aumento. El 97% de las empresas utilizan metodologías ágiles actualmente para el desarrollo de herramientas tecnológicas (Hoda, Salleh, & Grundy, 2018). Estas metodologías no solo se caracterizan por permitir la disminución de tiempo en el desarrollo de productos, pero también son criticadas por la falta de inclusión de seguridad dentro de su ciclo de vida. Este problema se vuelve crítico cuando se trata de empresas que se encuentran relacionadas con las industrias críticas tales como la financiera, el sector público o salud. Actualmente hay algunas soluciones propuestas y evaluadas con respecto a cómo implementar seguridad en las metodologías de desarrollo. Pero el problema de las industrias altamente reguladas es más complejo ya que se encuentran obligadas a cumplir estándares tanto nacionales como internacionales (Moyon, Beckers, Klepper, Lachberger, & Bruegge, 2018). Con el objetivo de determinar el estado de madurez de los métodos ágiles que incorporen seguridad y cumplimiento con regulaciones, se realiza un estudio sistemático del estado del arte siguiendo la metodología propuesta por Kurhmann, Méndez y Daneva, 2016, tal como se explicó en la sección de metodología.

Detalles del Estudio

Los detalles del estudio serán incluidos en una publicación realizada en colaboración con Siemens y la Universidad Técnica de Múnich. A continuación, se incluye un cuadro cuantitativo del proceso de recolección de datos, limpieza de datos y selección de archivos (ver

figura 6). En total se obtuvieron 11 publicaciones relevantes al cumplimiento de inclusión de actividades de seguridad basadas en estándares en las metodologías ágiles.

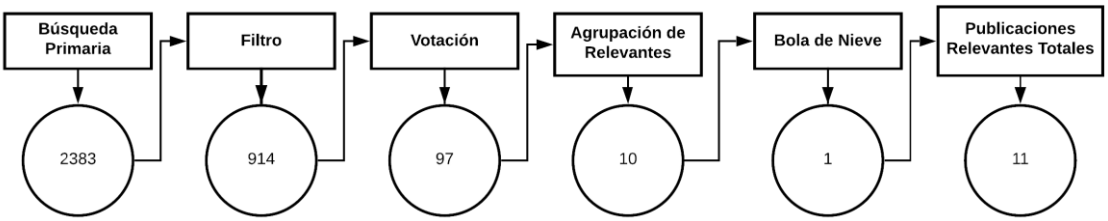

Figure 6: Resultados Cuantitativos del Proceso de Filtrado

Se incluye, además, en la figura 7 el detalle de las cadenas de búsqueda utilizadas, el procedimiento de filtrado junto con descripción cuantitativa por librería en cada etapa. Como se observa en la figura, los procesos de lectura del texto completo de los documentos agrupados para el cumplimiento de estándares y seguridad en el agilismo se realizó en paralelo con la lectura del título y resumen de los documentos agrupados para seguridad en el agilismo.

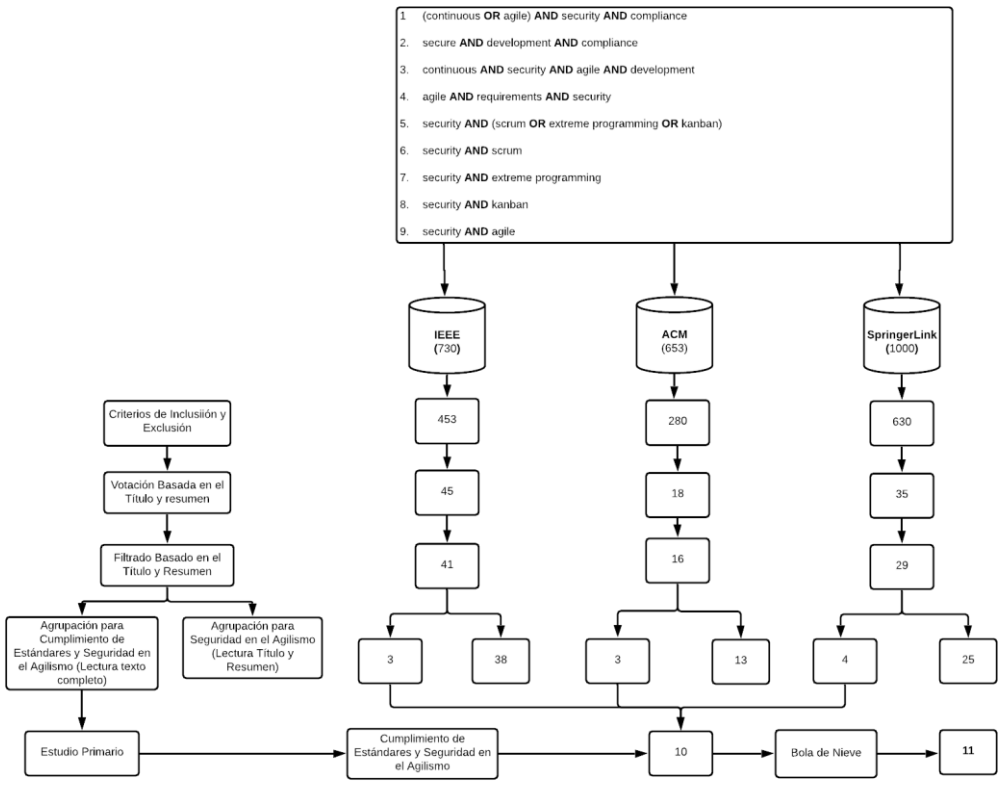

Figure 7: Representación Cuantitativa Detallada del Proceso de Filtrado

CONCLUSIONES Y TRABAJO FUTURO

Conclusiones

El desarrollo tecnológico se encuentra estrechamente relacionado con el desarrollo de estudios científicos. El aporte de la academia permite clarificar el estado de madurez de las diferentes áreas de estudio. Con los estudios científicos como punto de partida, es posible proponer soluciones a los problemas actuales, innovar, comparar y generar sinergias que permitan la generación de conocimiento.

El caso de estudio fue desarrollado y documentado de acuerdo a la metodología propuesta por Kurhmann, Méndez y Daneva, 2016. Este proceso se considera como riguroso ya que en cada fase hubo validaciones internas por miembros del equipo y externas por colaboradores externos. La rigurosidad y confiabilidad del proceso aumentan debido a los diferentes niveles y experiencia del equipo tanto interno como externo. Debido a la rigurosidad del proceso y el tamaño del equipo de trabajo este proceso tuvo una gran duración. El tiempo invertido se justifica por la calidad y el aporte del resultado final.

Existen diferentes metodologías propuestas para realizar un estudio sistemático del estado del arte. Se recomienda que cada equipo de trabajo escoja o cree una metodología que vaya de acuerdo con las características tanto del equipo de trabajo como del nivel de rigurosidad y tiempo disponible para el desarrollo del estudio científico.

Durante este trabajo de titulación se desarrollaron algunas herramientas con la finalidad de reducir el trabajo manual e ineficiente que involucra un estudio sistemático del estado del arte. Particularmente, las herramientas desarrolladas permiten la realización de consultas automáticas en las bibliotecas digitales escogidas para el estudio, consolidan la información obtenida, la repara en caso de registros redundantes y sirven como fuente de alimentación para la aplicación de votación para que los miembros del equipo de trabajo

agreguen su criterio considerando los criterios de votación, las preguntas de investigación en conjunto con la clasificación de los documentos científicos determinados como relevantes para el estudio.

Las metodologías ágiles se han vuelto un estándar en el desarrollo de herramientas tecnológicas. Debido a su importancia es importante no solo el desarrollo continuo, pero también la mejora en el área de seguridad y de los estándares, permitiendo su uso en la industria crítica de cada país. Se encontraron algunas publicaciones académicas relacionadas con el cumplimiento de seguridad. Pero, se considera que el área de cumplimiento de estándares y seguridad para metodologías ágiles aún es un área en desarrollo. La propuesta del desarrollo de nuevos estándares basados en las metodologías ágiles o a su vez el desarrollo de guías o marcos conceptuales de cómo automatizar y/o incluir actividades de seguridad y el cumplimiento de estándares parecerían solucionar este problema y expandir el conocimiento en esta área.

Trabajo Futuro

Las metodologías documentadas incluyen a detalle las diferentes fases para realizar un estudio sistemático del estado del arte generalmente no incluyen herramientas automáticas o semiautomáticas dentro del proceso. Se considera la creación de una metodología que no solo considere a los procesos manuales realizados por los miembros del equipo, pero a su vez incorpore herramientas tecnológicas para realizar un proceso de alta calidad, riguroso y eficiente. Adicionalmente, como trabajo futuro es necesaria la validación de esta metodología por investigadores externos al equipo en diferentes proyectos.

Considerando los resultados obtenidos en el caso de estudio y la amplitud del área de investigación que aún no se encuentra con madurez, como trabajo futuro, se propone la realización de un marco conceptual que permita la implementación eficiente de los estándares

y actividades de seguridad relacionadas a la metodología ágil que es utilizada por las empresas.

Si bien en este trabajo se implementó una herramienta para la priorización de documentos científicos en base a una base de conocimiento previa relevante al tema de investigación (*NLP Text Miner*), cabe mencionar que sus resultados fueron excluidos del caso de estudio. La estrategia inicial de esta herramienta se basa en que todos los artículos científicos deben parecerse a la base de conocimiento, pero cuando este no es el caso, la herramienta induce un sesgo notorio en la priorización lo que no ayuda a disminuir el trabajo al momento de realizar la fase de votación. Es por esto que, como trabajo futuro, se plantea profundizar las estrategias de minería de texto, al igual de aquellas de aprendizaje de máquina avanzado como aprendizaje profundo (*Deep Learning*) para mejorar la transformación de texto a vectores numéricos que puedan caracterizar mejor el contenido de los documentos científicos antes de la priorización. Se plantea, además, el incluir comparaciones con respecto a las herramientas desarrolladas, con herramientas disponibles y con el trabajo manual con el fin de obtener comparaciones cuantitativas y cuantitativas de la efectividad de los procesos.

BIBLIOGRAFÍA

Basili, V. R., & Zelkowitz, M. V. (2007). Empirical studies to build a science of computer science. *Communications of the ACM*, *50*(11), 33. https://doi.org/10.1145/1297797.1297819

Cooper, I. D. (2016). What is a "mapping study?" *Journal of the Medical Library Association : JMLA*, *104*(1), 76–78. https://doi.org/10.3163/1536-5050.104.1.013

Fornari, L. F., Pinho, I., de Almeida, C. A., & Costa, A. P. (2019). Systematic Literature Review with Support of Digital Tools. *2019 IEEE 9th Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, 182–187. https://doi.org/10.1109/ISCAIE.2019.8743787

Hoda, R., Salleh, N., & Grundy, J. (2018). The Rise and Evolution of Agile Software Development. *IEEE Software*, *35*(5), 58–63. https://doi.org/10.1109/MS.2018.290111318

Ierardi, C., Orihuela, L., & Jurado, I. (2018). Guidelines for a systematic review in systems and automatic engineering. Case study: Distributed estimation techniques for cyber-physical systems. *2018 European Control Conference (ECC)*, 2230–2235. https://doi.org/10.23919/ECC.2018.8550218

J. Randolph, J., Julnes, G., Sutinen, E., & Lehman, S. (2008). A Methodological Review of Computer Science Education Research. *Journal of Information Technology Education: Research*, *7*, 135–162. https://doi.org/10.28945/183

Kohl, C., McIntosh, E. J., Unger, S., Haddaway, N. R., Kecke, S., Schiemann, J., & Wilhelm, R. (2018). Online tools supporting the conduct and reporting of systematic reviews and systematic maps: A case study on CADIMA and review of existing tools. *Environmental Evidence*, *7*(1), 8. https://doi.org/10.1186/s13750-018-0115-5

Kuhrmann, M., Fernández, D. M., & Daneva, M. (2016). On the Pragmatic Design of Literature Studies in Software Engineering: An Experience-based Guideline. *ArXiv:1612.03583 [Cs]*. Retrieved from http://arxiv.org/abs/1612.03583

Marchezan, L., Bolfe, G., Rodrigues, E., Bernardino, M., & Basso, F. P. (2019). Thoth: A Web-based Tool to Support Systematic Reviews. *2019 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, 1–6. https://doi.org/10.1109/ESEM.2019.8870160

Marshall, C., & Brereton, P. (2013). Tools to Support Systematic Literature Reviews in Software Engineering: A Mapping Study. *2013 ACM / IEEE International Symposium on Empirical Software Engineering and Measurement*, 296–299. https://doi.org/10.1109/ESEM.2013.32

Moyon, F., Beckers, K., Klepper, S., Lachberger, P., & Bruegge, B. (2018). Towards continuous security compliance in agile software development at scale. *Proceedings of the 4th International Workshop on Rapid Continuous Software Engineering - RCoSE '18*, 31–34. https://doi.org/10.1145/3194760.3194767

Okoli, C., & Schabram, K. (2010). A Guide to Conducting a Systematic Literature Review of Information Systems Research. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1954824

Petersen, K., Feldt, R., Mujtaba, S., & Mattsson, M. (2008, June 1). *Systematic Mapping Studies in Software Engineering*. Presented at the 12th International Conference on Evaluation and Assessment in Software Engineering (EASE). https://doi.org/10.14236/ewic/EASE2008.8

Stansfield, C., Thomas, J., & Kavanagh, J. (2013). 'Clustering' documents automatically to support scoping reviews of research: A case study: "Clustering" to support scoping reviews. *Research Synthesis Methods*, n/a-n/a. https://doi.org/10.1002/jrsm.1082.