2019 CIINATIC
CONGRESO INTERNACIONAL
EN INNOVACIÓN Y APROPIACÓN
DE LAS TECNOLOGÍAS DE LA INFORMACIÓN
Y LAS COMUNICACIONES

1

Hacia una Revisión Sistemática de la Literatura: Caso aplicado en Análisis de Sentimientos en política Latinoamericana

Laura F. García[1]
Johana A. Manrique[1]
Liliana Calderón-Benavides[2]

[1] Universidad Popular del Cesar, Seccional Aguachica, {lfernandagarcia, jandreahernandez}@unicesar.edu.co
[2] Universidad Autónoma de Bucaramanga, mcalderon@unab.edu.co

RESUMEN

La investigación es una actividad orientada a la generación de nuevo conocimiento útil para la solución de necesidades en los diferentes campos del conocimiento. Para iniciar esta actividad, es necesario identificar qué se conoce sobre el tema a tratar y qué aspectos aún permanecen desconocidos, abordándolos desde un punto de vista holístico. La Revisión Sistemática de la Literatura es una metodología práctica que permite identificar, evaluar y sintetizar información proveniente de fuentes primarias para la caracterización y el reconocimiento de brechas en cualquier campo de investigación. Este artículo presenta una guía que estructura el proceso de desarrollo de una Revisión Sistemática de la Literatura ejemplificando su uso al campo de investigación en Análisis de Sentimientos en política Latinoamericana.

***Palabras Clave* - Análisis de Datos, Análisis de sentimientos, Investigación, Revisión Sistemática de la Literatura.**

ABSTRACT

Research is an activity aimed at generating new and useful knowledge for the solution of needs in different fields of knowledge. As a starting point of this activity, it is necessary to identify what is known about the desired topic as well as the aspects remain unknown, addressing them from a holistic point of view. Systematic Literature Review (SLR) is a methodology that allows to identify, evaluate and synthesize information from primary sources aiming to determine and characterize gaps in any field of research. This paper presents a guide to developing a SLR, exemplifying its use in the research on Sentiment Analysis in Latin American Politics.

***Key Words* - Data Analytics, Research, Sentiment Analysis, Systematic Literature Review (SLR).**

I. INTRODUCCIÓN

La investigación científica es el más importante instrumento con que cuenta el ser humano para conocer, explicar, interpretar y transformar la realidad. Su desarrollo desde las diferentes disciplinas científicas es indispensable para la búsqueda de soluciones a los principales problemas que afronta la sociedad y para la generación de nuevos conocimientos que expliquen y orienten su transformación [1]. La investigación es un proceso sistemático porque implica pasos ordenados; empírico, por la recolección y análisis de datos; y crítico, por la evaluación y mejora constante, aplicada al estudio de un fenómeno [2].

Una de las primeras etapas para el desarrollo de investigación, independientemente de su objetivo (básica o aplicada), su profundidad (exploratoria, explicativa, descriptiva), o su enfoque (cualitativo, cuantitativo o mixto) es la revisión del estado del arte. Con la revisión del estado del arte se establece una aproximación al conocimiento de un tema, y se puede decir que es la primera etapa del proceso de investigación ya que permite identificar qué se conoce y qué se desconoce del tema a abordar [3]. Esta información es necesaria para proponer nuevo conocimiento o para realizar modificaciones al conocimiento existente; por ejemplo, si se trata de una investigación aplicada, conocer qué existe sobre una técnica particular es básico, de cara a la generación de tecnologías susceptibles de patentamiento que puedan convertirse en innovaciones y que produzcan regalías para sus desarrolladores, o por el contrario, frustraciones por la generación de tecnologías ya existentes en otros lugares del mundo.

[1]Primer Autor: lfernandagarcia@unicesar.edu.co, estudiante de Ingeniería de sistemas, Universidad Popular del Cesar, Seccional Aguachica.
Segundo Autor: jandreahernandez@unicesar.edu.co, Profesor, Programa Ingeniería de Sistemas, Universidad Popular del Cesar, Seccional Aguachica.
Tercer Autor: mcalderon@unab.edu.co, Profesor, Programa Ingeniería de Sistemas, Universidad Autónoma de Bucaramanga.

2019 CIINATIC
CONGRESO INTERNACIONAL
EN INNOVACIÓN Y APROPIACÓN
DE LAS TECNOLOGÍAS DE LA INFORMACIÓN
Y LAS COMUNICACIONES

Para desarrollar un estado del arte con estudios primarios publicados por expertos, se realiza una revisión y análisis de la literatura en fuentes bibliográficas científicas. Para esto, es necesario seguir un enfoque metodológico detallado, que, a diferencia de modelos narrativos, informales y subjetivos, presente de manera explícita el proceso seguido para buscar, organizar y evaluar la información que ha conducido al planteamiento de conclusiones [4]. Un método muy difundido en los últimos años es la Revisión Sistemática de la Literatura (RSL) [5], [6].

Este artículo se enfoca en presentar de manera resumida, pero didáctica, en qué consiste una RSL, así como la forma de realizarla, como se describe en la sección II; ejemplificando la forma de conducir la revisión con un tema de actualidad como el análisis de sentimientos en política Latinoamericana, descrito en la sección III. Finalmente, se presenta el trabajo futuro y las conclusiones.

II. REVISIÓN SISTEMÁTICA DE LA LITERATURA

La Revisión Sistemática de la Literatura – RSL (*en inglés, Systematic Literature Review – SLR)*, es un método sistemático, explícito comprensible y reproducible que sirve para identificar, evaluar y sintetizar estudios relevantes de las diferentes áreas del conocimiento, los cuales han sido publicados por académicos y profesionales en fuentes bibliográficas de carácter científico e investigativo [5], [7], [8]. Los estudios empíricos sobre un tema específico que contribuyen a una revisión de la literatura se denominan estudios primarios; Una RSL es una forma de estudio secundario porque sintetiza evidencia sobre el estado actual de una investigación [9]. Este tipo de estudio tiene como objetivos [3], [5], [10]: (i) identificar brechas actuales que requieren más atención de los investigadores; (ii) proporcionar una base teórica para organizar adecuadamente los nuevos parámetros de investigación; (iii) direccionar nuevos enfoques para investigaciones posteriores.

Cabe resaltar que, aunque una RSL es una metodología rigurosa y estandarizada, tiene un protocolo con características cualitativas y cuantitativas que son abordadas por los investigadores de acuerdo al área del conocimiento en el cual es aplicada. Okoli y Schabram [10, p. 12] presentan una comparación de protocolos para la RSL que fueron publicados entre 1976 – 2008. Si bien es cierto, no son los únicos, éstos han sido muy referenciados y han servido como base para la mayoría de las RSL que actualmente se encuentran publicadas. Sin embargo, aunque las metodologías de la RSL planteadas tengan más o menos fases de implementación, tienen en común actividades enmarcadas en: un propósito y protocolo de búsqueda, datos de extracción, la síntesis y el reporte de los resultados obtenidos [11]–[14].

Kitchenham et. al [15] presentan un marco metodológico riguroso que proporciona un conjunto de fases direccionadas de acuerdo a un protocolo de actividades predefinidas. Este marco especializado en RSL para el campo de la ingeniería del software ha sido publicado como desarrollo del proyecto (EP/CS51839/X) en el Reino Unido [5] desde el año 2007. En la Fig. 1 se presenta un resumen de las tres fases principales de la RSL.

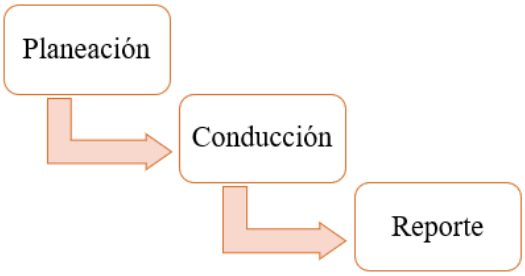

Fig. 1. Fases de una RSL
Fuente: Las autoras

La metodología presenta un estilo en cascada en donde los parámetros de salida de una fase, sirven como parámetros de entrada para la siguiente. Las fases deben responder a una lista de verificación [16] como se observa en la Tabla 1, para garantizarle a los investigadores que: (i) es necesaria la realización de una revisión sistemática; (ii) no existe una con los mismos objetivos planteados y; (iii) ésta cumple con requisitos de calidad y pertinencia.

Tabla 1. Lista de verificación de una RSL

Parámetros	Descripción
Planeación	
¿Cuál es el objetivo de la revisión?	
¿Cuáles son las palabras clave para realizar la revisión?	
¿Qué fuentes se buscaron para identificar estudios primarios?	
¿Cuáles fueron los criterios de inclusión/exclusión y cómo se aplicaron?	
¿Qué criterios se utilizaron para evaluar la calidad de los estudios primarios?	
Conducción	
¿Cómo se extrajeron los datos de los estudios primarios?	
¿Cómo se aplicaron los criterios de calidad?	
¿Cómo se sintetizaron los datos?	
¿Cómo se investigaron las diferencias entre los estudios?	

2019 CIINATIC
CONGRESO INTERNACIONAL EN INNOVACIÓN Y APROPIACÓN DE LAS TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES

¿Cómo se combinaron los datos?	
Reporte	
¿Fue razonable combinar los estudios?	
¿Las conclusiones fluyen de la evidencia?	

Fuente: Las autoras

A. Planeación

En esta fase se establece el protocolo de búsqueda. Este protocolo incluye una serie de pasos con actividades que direccionan el éxito de la RSL (ver Fig. 2). El primer paso es definir el objetivo de la búsqueda el cual puede ser amplio, y guiará el alcance de la RSL [17].

El segundo paso es establecer las preguntas de revisión, las cuales dirigen toda la metodología de la RSL, de tal forma que en el proceso de búsqueda se puedan identificar estudios primarios que den respuesta a las preguntas establecidas. Este proceso que debe ser reflexivo, debatido y reformulado por los involucrados en el desarrollo de la RSL. Es recomendable que la RSL sea realizada por un grupo interdisciplinario de investigadores que conozcan de la temática a tratar, de tal forma que los resultados no sean solo juicios de valor. Además, la mayoría de las RSL, debido a su amplio alcance, no pueden ser realizadas por un solo investigador [7].

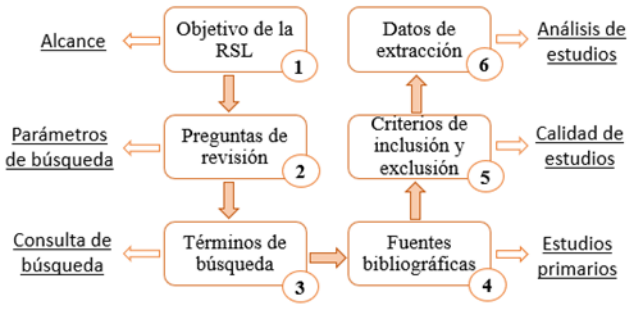

Fig. 2. Actividades de la Fase Planeación de una RSL
Fuente: Las autoras

En el tercer paso se determinan los términos *(keywords)* para construir la consulta de búsqueda. Para establecer cuáles son los términos de búsqueda adecuados, es necesario tener en cuenta el área en el cual está enmarcada la investigación. Un tesauro (diccionario) es una herramienta que agrupa términos según la similitud del significado y la relación que existe entre ellos [18].

Con esta herramienta, los investigadores pueden establecer de forma práctica los términos clave para la consulta. Es importante tener en cuenta que entre más términos similares (sinónimos) tenga un conjunto de palabras, se pueden obtener más resultados. Sin embargo, entre más conjuntos de palabras se establecen para una consulta, se reducen el número de resultados de búsqueda. Se recomienda tener un número adecuado de términos para no sesgar los resultados de la RSL.

En cuarto lugar, se seleccionan las fuentes bibliográficas como las bibliotecas digitales, revistas científicas, congresos, en las cuales se realiza la consulta de búsqueda que previamente se formó con los términos clave para extraer los estudios primarios que den respuesta a las preguntas de revisión.

En quinto lugar, se establecen los criterios de inclusión y exclusión de los estudios, con base a las preguntas de revisión. Los investigadores deben garantizar que estos criterios puedan interpretarse de manera confiable para clasificar los estudios correctamente. En este paso es importante determinar el rango de fecha en el que se delimitará la búsqueda para que la RSL pueda ser replicable.

Por último, se establece el esquema de extracción de datos el cual debe tener los elementos necesarios para abordar las preguntas de revisión. Este esquema puede tener cambios a medida que se aplica la evaluación de calidad en los estudios primarios preseleccionados.

B. Conducción

En esta fase se identifican, seleccionan y evalúan los estudios primarios de acuerdo con el protocolo previamente establecido. El primer paso es ejecutar la consulta de búsqueda en las fuentes bibliográficas seleccionadas y extraer los metadatos de los resultados obtenidos. Luego, los investigadores realizan una revisión inicial de los estudios teniendo en cuenta el título, resumen y palabras clave que abordan el objetivo y las preguntas de revisión. Con esta evaluación inicial, se descarta aproximadamente el 70% de los estudios debido a que la mayoría no cumplen los criterios de inclusión y exclusión.

En segunda instancia, se descargan en formato *.pdf* los estudios potencialmente elegibles, y se examinan más detenidamente para evaluar su calidad. Esta evaluación de calidad sirve para dos propósitos, a saber: (i) facilita la eliminación de estudios primarios que no cumplen con los requisitos estándar establecido por los investigadores; (ii) asigna una calificación de la calidad metodológica usada para los estudios incluidos en la RSL, ya que la calidad de la revisión final depende de la calidad de los estudios primarios seleccionados. Aunque el propósito del puntaje del estudio (calificar los estudios según los criterios de calidad) es distinto al de la selección metodológica (eliminar estudios por deficiencias en la calidad), los dos pasos están relacionados. Por lo tanto, en términos prácticos, tiene sentido clasificar la calidad de los estudios utilizando los mismos criterios de inclusión y exclusión establecidos en la fase de planeación [10], [15]. Toda decisión que tomen los investigadores con respecto a mantener

o no estudios dentro del proceso de revisión, debe ser informada de forma explícita.

Finalmente, se realiza la extracción de datos de los estudios que incluirán la información final para la fase de reporte de la RSL. En este punto, la información se tomará sistemáticamente de cada estudio para que sirva como materia prima en esta fase. Se toma como base el esquema de extracción de datos definido en el protocolo, el cual debe ser revisado por los investigadores una vez más, ya sea para eliminar o incluir nuevos parámetros de extracción.

Después de haber seguido los pasos anteriores, los investigadores deberían desarrollar un formulario de extracción limpio que se utilizará para almacenar la información pertinente sobre cada estudio. Se recomienda que este esquema incluya espacios para almacenar los detalles requeridos por las preguntas de revisión, así como una sección de comentarios generales. La información de este formulario se combinará con el protocolo de búsqueda y la evaluación de la calidad de los estudios para construir el reporte de la RSL.

C. Reporte

El reporte es la etapa final de la RSL en donde se agrega información extraída de los estudios primarios relevantes que atienden las preguntas de búsqueda, y finalmente, se presentan las conclusiones de la revisión. Este reporte debe tener características [8], [10], [15]como: (i) escritura clara y técnica, usando términos entendibles para el público objetivo; (ii) descripción de forma detallada del protocolo de revisión junto a cada una de las actividades realizadas, de modo que la RSL sea replicable; (iii) proceso del análisis de datos, sintetizando los datos de manera que las preguntas de revisión puedan ser respondidas; (iv) detalle del análisis de los metadatos y de los resultados cuantitativos a través de gráficas; (v) análisis y descripción de los hallazgos principales obtenidos en el proceso de revisión de modo que sean útiles para futuras investigaciones en esa área; (vi) discusión de las fortalezas y debilidades de la evidencia incluida en la revisión para permitirle a los lectores evaluar la confianza de los estudios recopilados; (vii) relación de la RSL obtenida con otras revisiones, considerando cualquier diferencia y similitud en calidad y resultados; (viii) establecimiento de una dirección para futuras RSL tomando como base los hallazgos obtenidos.

III. CASO APLICADO: ANÁLISIS DE SENTIMIENTOS EN POLÍTICA LATINOAMERICANA

Para ejemplificar la metodología descrita en la sección II, las autoras de este artículo están realizando como caso aplicado una RSL en el campo de investigación de análisis de sentimientos en política Latinoamericana.

Dentro de la fase de planeación se estableció el objetivo de la RSL, el cual aborda la identificación, selección y análisis de manera sistemática de estudios relacionados con análisis de sentimientos en política. No se especificó explícitamente en Latinoamérica puesto que se sesga parte de la revisión, ya que esa palabra no es ampliamente usada dentro de los metadatos de los estudios a seleccionar. Para lograr el objetivo planteado, se propusieron las siguientes preguntas de búsqueda:

RQ1: ¿Qué tipos técnicas son usadas para el procesamiento de datos en análisis de sentimientos en política?
RQ2: ¿Qué tipos de herramientas son usadas para el procesamiento de datos en análisis de sentimientos en política?
RQ3: ¿Cuál es el enfoque sobre los estudios de análisis de sentimientos en política?
RQ4: ¿Cuáles son los países latinoamericanos (Mapeo) que están trabajando en análisis de sentimientos en política?

Para dar respuesta a estas preguntas de revisión, se definió la cadena de búsqueda, tomando como base términos clave para formar grupos con características semejantes (sinónimos).

TITLE-ABS (("Sentiment Analysis" OR "Opinion Mining") AND (politic OR political OR government OR governmental))

La cadena de búsqueda fue consultada en fuentes bibliográficas digitales como Scopus, ScientDirect, ACM y Springer las cuales abordan estudios primarios concernientes al análisis de sentimientos.

Para reducir progresivamente la cantidad de estudios encontrados en la búsqueda a un número con alta calidad relevantes para atender las preguntas planteadas, se establecieron unos criterios de selección: inclusión (CI) y exclusión (CE), los cuales se describen a continuación.

CI1: El estudio está escrito en inglés o español.
CI2: El estudio es de tipo artículo, libro, capítulo de libro o artículo de conferencia.
CI3: El título y resumen del estudio coinciden con los criterios de búsqueda.
CI4: El estudio describe un desarrollo/aplicación basada en análisis de sentimientos en política.

CE1: No se tiene acceso a todo el documento *(full text).*
CE2: El estudio está relacionado con análisis de sentimientos, pero no propone o discute herramientas, técnicas o dominios de aplicación en política.

CE3: El estudio no es primario ya que expone un estado del arte sobre análisis de sentimientos en aspectos políticos.

CE4: El estudio se repite en dos o más fuentes bibliográficas consultadas.

Para garantizar la calidad de los estudios recolectados, se estableció un esquema de extracción de datos el cual responde a las preguntas de revisión, como se observa en la Tabla 2. Como se explicó en la sección anterior, este esquema puede tener cambios a medida que se evalúan los estudios primarios preseleccionados.

Tabla 2. Lista de verificación de una RSL

Pregunta	Campo	Subcampo	Descripción
RQ1	Técnicas para análisis de sentimientos	Documento	
		Sentencia	
		Entidad	
		Aspecto	
	Modelo	Manual	
		Semiautomático	
		Automático	
RQ2	Hardware	--	
	Software	Herramienta	
		Lenguaje de programación	
	Diccionario	Nombre	
		Cantidad de términos	
		Idioma	
RQ3	Enfoque	Elecciones	
		Leyes	
		Campañas	
		Actores políticos	
		Ideología	
		Debate	
RQ4	Países Latinoamericanos	Institución	
		País	
		Ciudad	

Fuente: Las autoras

Una vez establecido el protocolo, se ejecutó la consulta en las fuentes bibliográficas mencionadas anteriormente y se obtuvieron las siguientes cantidades de estudios: Scopus 506, Springer 825, ScientDirect 54 y ACM 67, para un total de 1452 estudios.

Este es un trabajo en desarrollo, por esto no se reportan las siguientes fases. Sin embargo, a la fecha de entrega de este documento se está llevando a cabo la segunda parte de la fase conducción haciendo lectura del título y resumen para obtener los estudios potencialmente elegibles, los cuales serán examinados más detenidamente para evaluar su calidad. Finalmente, se generará el reporte de la RSL usando como base los datos de extracción que permitan responder cada una de las preguntas de búsqueda planteadas.

IV. CONCLUSIONES

La RSL es un estudio secundario aplicable a cualquier área del conocimiento con un enfoque metodológico práctico que permite sintetizar la información científica disponible, dar a conocer los resultados obtenidos en estudios primarios e identificar áreas de incertidumbre para la realización de nuevos proyectos de investigación.

Este artículo describe una guía sobre los pasos necesarios para realizar una RSL, útil para cualquier investigador. Se tomó como base estudios científicos encontrados en la literatura, y se aplicaron a un caso práctico en el campo de análisis de sentimientos en política Latinoamericana. Como trabajo futuro, se espera finalizar la RSL para dar a conocer los resultados obtenidos, de manera que la comunidad interesada en esta área de investigación disponga de un recurso para el desarrollo de nuevas investigaciones.

REFERENCIAS

[1] C. Monje, "Metodología de la investigación cuantitativa y cualitativa: Guia didáctica," 2011.

[2] C. F. Collado, R. H. Sampieri, and P. B. Lucio, *Metodología de la investigación*, Quinta edi. McGraw-Hill Interamericana, 2010.

[3] R. Guevara Patiño, "El estado del arte en la investigación: análisis de los conocimientos acumulados o indagación por nuevos sentidos?," *Folios*, no. 44, pp. 165–179, 2016.

[4] G. Goris and S. J. Adolf, "Utilidad y tipos de revisión de literatura," *Ene*, vol. 9, no. 2, p. 0, 2015.

[5] B. Kitchenham, S. Charters, D. Budgen, P. Brereton, M. Turner, and S. Linkman, "Guidelines for performing systematic literature reviews in software engineering," 2007.

[6] J. P. Gisbert and X. Bonfill, "¿Cómo realizar, evaluar y utilizar revisiones sistemáticas y metaanálisis?," *Gastroenterol. Hepatol.*, vol. 27, no. 3, pp. 129–149, 2004.

[7] A. Fink, *Conducting research literature reviews: From the internet to paper*, Second edi. California: Sage publications, 2005.

[8] L. A. Machi and B. T. McEvoy, *The literature review: Six steps to success*, Third edit. Corwin Press, 2016.

[9] S. Jalali and C. Wohlin, "Systematic literature studies: database searches vs. backward snowballing," in *Proceedings of the 2012 ACM-IEEE international symposium on empirical software engineering and measurement*, 2012, pp. 29–38.

[10] C. Okoli and K. Schabram, "A guide to conducting a systematic literature review of information systems research," *Work. Pap. Inf. Syst.*, p. 51, 2010.

[11] G. M. Marakas, M. Y. Yi, and R. D. Johnson, "The multilevel and multifaceted character of computer self-efficacy: Toward clarification of the construct and an integrative framework for research," *Inf. Syst. Res.*, vol. 9, no. 2, pp. 126–163, 1998.

[12] L. Dubé and G. Paré, "Rigor in information systems positivist case research: current practices, trends, and recommendations," *MIS Q.*, pp. 597–636, 2003.

[13] D. E. Leidner and T. Kayworth, "A review of culture in information systems research: Toward a theory of information technology culture conflict," *MIS Q.*, vol. 30, no. 2, pp. 357–399, 2006.

[14] D. Robey, G. Im, and J. D. Wareham, "Theoretical foundations of empirical research on interorganizational systems: assessing past contributions and guiding future directions," *J. Assoc. Inf. Syst.*, vol. 9, no. 9, p. 4, 2008.

[15] B. A. Kitchenham, O. P. Brereton, D. Budgen, and others, "Protocol for Extending an existing Tertiary study of Systematic Literature Reviews in Software Engineering," 2017.

[16] K. S. Khan, G. Ter Riet, J. Glanville, A. J. Sowden, J. Kleijnen, and others, *Undertaking systematic reviews of research on effectiveness: CRD's guidance for carrying out or commissioning reviews*, no. 4 (2n. NHS Centre for Reviews and Dissemination, 2001.

[17] M. D. J. Peters, C. M. Godfrey, H. Khalil, P. McInerney, D. Parker, and C. B. Soares, "Guidance for conducting systematic scoping reviews," *Int. J. Evid. Based. Healthc.*, vol. 13, no. 3, pp. 141–146, 2015.

[18] J. Aitchison, D. Bawden, and A. Gilchrist, *Thesaurus construction and use: a practical manual*, Fourth edi. Routledge, 2003.