



EDUCACIÓN

SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO
NACIONAL DE MÉXICO

Tecnológico Nacional de México

Centro Nacional de Investigación
y Desarrollo Tecnológico

Tesis de Maestría

Mapeo sistemático del reconocimiento del habla,
proceso del lenguaje natural y uso de ontologías para
identificar el dominio del problema y los
requerimientos de solución

presentada por

Ing. Santiago Hernández Guzmán

como requisito para la obtención del grado de
Maestro en Ciencias de la Computación

Director de tesis

Dr. Moisés González García

Cuernavaca, Morelos, México. Septiembre de 2019.



"2019, Año del Caudillo del Sur, Emiliano Zapata"

Cuernavaca, Mor.,
Oficio No.
Asunto:

17/septiembre/2019
DCC/083/2019
Aceptación de documento de tesis

DR. GERARDO VICENTE GUERRERO RAMÍREZ
SUBDIRECTOR ACADÉMICO
PRESENTE

Por este conducto, los integrantes de Comité Tutorial del Ing. Santiago Hernández Guzmán, con número de control M17CE092, de la Maestría en Ciencias de la Computación, le informamos que hemos revisado el trabajo de tesis profesional titulado "Mapeo sistemático del reconocimiento del habla, proceso de lenguaje natural y uso de ontologías para identificar el dominio del problema y los requerimientos de solución" y hemos encontrado que se han realizado todas las correcciones y observaciones que se le indicaron, por lo que hemos acordado aceptar el documento de tesis y le solicitamos la autorización de impresión definitiva.

DIRECTOR DE TESIS

Dr. Moisés González García
Doctor en Ciencias en la
Especialidad de Ingeniería
Eléctrica
7501724

REVISOR 1

Dr. René Santaolaya Salgado
Doctor en Ciencias de la
Computación
4454821

REVISOR 2

Dr. Juan Carlos Rojas Pérez
Doctor en Ciencias en Ciencias de
la Computación
6099372

C.p. M.E. Guadalupe Garrido Rivera - Jefa del Departamento de Servicios Escolares.
Estudiante
Expediente

NACS/lmz



"2019, Año del Caudillo del Sur, Emiliano Zapata"

Cuernavaca, Mor.,
No. de Oficio:
Asunto:

18/septiembre/2019
SAC/266/2019
Autorización de
impresión de tesis

ING. SANTIAGO HERNÁNDEZ GUZMÁN
CANDIDATO AL GRADO DE MAESTRO EN CIENCIAS
DE LA COMPUTACIÓN
PRESENTE

Por este conducto, tengo el agrado de comunicarle que el Comité Tutorial asignado a su trabajo de tesis titulado "Mapeo sistemático del reconocimiento del habla, proceso de lenguaje natural y uso de ontologías para identificar el dominio del problema y los requerimientos de solución", ha informado a esta Subdirección Académica, que están de acuerdo con el trabajo presentado. Por lo anterior, se le autoriza a que proceda con la impresión definitiva de su trabajo de tesis.

Esperando que el logro del mismo sea acorde con sus aspiraciones profesionales, reciba un

ATENTAMENTE

Excelencia en Educación Tecnológica
"Conocimiento y tecnología al servicio de México"

DR. GERARDO VICENTE GUERRERO RAMÍREZ
SUBDIRECTOR ACADÉMICO



SEP TecNM
CENTRO NACIONAL
DE INVESTIGACIÓN
Y DESARROLLO
TECNOLÓGICO
SUBDIRECCIÓN
ACADÉMICA

C.p. Mtra. Guadalupe Garrido Rivera.- Jefa del Departamento de Servicios Escolares.
Expediente

GVGR/mcr

Dedicatoria

A mi abuelita Tina, que en paz descansa, no le dedicaré solamente este trabajo de posgrado. A ella le quiero dedicar cada uno de mis logros, tanto académicos como personales. Gracias a ella pude emprender este camino y realizar este sueño; gracias a todo su amor, a todo su esfuerzo y su empeño por hacerme una buena persona. Le estaré eternamente agradecido por sus años de dedicación, sus noches de desvelo, y su cariño siempre incondicional. Estás conmigo siempre, en cada momento y en cada instante. Te amo y te amaré eternamente.

A mi mamá Celia y mi hermana Guadalupe, por enseñarme a enfrentar las adversidades de la vida, a continuar luchando, aunque todo se complique a nuestro alrededor, por darme el ejemplo de nunca darme por vencido.

A mi novia Martha Patricia, por estar junto a mí y apoyarme en estos años de trabajo y dedicación. Eres mi tesoro.

A mi suegra Martha, por creer que podía lograrlo y alentarme a perseguir esta estrella fugaz.

Agradecimientos

Al Consejo Nacional de Ciencia y Tecnología (CONACYT), por proporcionarme a mí y a muchos otros estudiantes el apoyo necesario para potencializar nuestros sueños y habilidades.

Al Tecnológico Nacional de México / Centro Nacional de Investigación y Desarrollo Tecnológico (CENIDET), por su gran servicio como Institución Académica, empeñada en formar e impulsar el conocimiento científico.

Especialmente le quiero agradecer a mi director de tesis, el Dr. Moisés González García, por el tiempo, esfuerzo, paciencia y valiosos aportes invertidos en este proyecto de tesis.

A mis revisores: Dr. René Santaoyala Salgado y Dr. Juan Carlos Rojas Pérez, por su invaluable intervención y sugerencias para realizar un trabajo de calidad.

A mis compañeros y amigos de generación: Alex, Heidi, Iván, Xico, Juan Carlos, Violeta y Gudiel, por acompañarme en esta aventura de dos años.

Agradezco a mis amigos Raúl, Jaziel y a sus familias. Siempre valoraré el apoyo que me han brindado a lo largo de todos estos años, gracias por siempre estar ahí cuando los necesité, gracias por estos años juntos.

A mis amigos: Vicky y Héctor, por su apoyo incondicional en los primeros meses de esta travesía.

Resumen

El proceso de desarrollo de software consta de varias etapas, entre las cuales varios autores señalan que una de las más importantes es la etapa de análisis, y una de sus principales actividades es la elicitación de requerimientos. Esto es el proceso de entender el dominio del problema e interactuar con los clientes para identificar sus necesidades y luego capturar los requerimientos de una solución propuesta. Sin embargo, esta actividad se realiza manualmente por el analista, utilizando lenguaje natural en la interacción con el cliente, por lo que se presentan problemas de ambigüedad, inconsistencia y captura incompleta de requerimientos. En los últimos años se han publicado trabajos de aplicación de Procesamiento de Lenguaje Natural y uso de ontologías para atacar los problemas antes mencionados. En este documento se presentan los resultados de un estudio de mapeo sistemático, realizado para reunir la evidencia publicada de las distintas técnicas de reconocimiento del habla, procesamiento de lenguaje natural y uso de ontologías en la elicitación de requerimientos, dentro del periodo 2010-2018. A partir de una búsqueda inicial, donde las bases de datos arrojaron 9783 artículos, posteriormente se aplicaron ciertos criterios de exclusión e inclusión y de esta manera se obtuvo una selección final de 64 artículos de investigación, que se analizaron para proveer un resumen y representaciones visuales de la evidencia existente en la comunidad científica, donde se identificaron técnicas y enfoques automatizados, semi-automatizados y manuales, que apoyan dicha actividad y se pueden implementar en la industria.

Abstract

The software development process consists of several phases, among which several authors point out that one of the most important is the analysis phase, and one of its main activities is the requirements elicitation. This is the process of understanding the problem domain and interacting with stakeholders to identify their needs and then capture the requirements of a proposed solution. However, this activity is done manually by the analyst, using natural language in the interaction with stakeholders, so there are problems of ambiguity, inconsistency and incomplete capture of requirements. In recent years, applications of Natural Language Processing and the use of ontologies to tackle the aforementioned problems have been published. This document presents the results of a systematic mapping study, conducted to gather the published evidence of the different techniques of speech recognition, natural language processing and use of ontologies in the requirements elicitation, in the period 2010-2018. From an initial search, where the databases returned 9783 articles, then exclusion and inclusion criteria were applied and in this way a final selection of 64 research articles was obtained, which were analyzed to provide a summary and visual representations of the evidence existing in the scientific community, where automated, semi-automated and manual techniques and approaches were identified that support this activity and can be implemented in the industry.

Índice general

Lista de figuras	V
Lista de tablas	VI
Acrónimos	VIII
Glosario de términos	IX
Presentación	X
Capítulo 1. Introducción	1
1.1 Contexto del problema	2
1.2 Descripción del problema	3
1.3 Objetivos	3
1.3.1 Objetivo General	3
1.3.2 Objetivos Específicos.....	3
1.4 Justificación.....	3
1.5 Alcances y Limitaciones	4
1.5.1 Alcances	4
1.5.2 Limitaciones.....	4
1.6 Resumen del Capítulo	4
Capítulo 2. Marco de Referencia	5
2.1 Antecedentes	6
2.1.1 Generación de Modelos Conceptuales a partir de Modelos de Requisitos: un Enfoque Basado en Tratamiento Automático del Lenguaje Natural.....	6
2.1.2 Traductor de Consultas en Lenguaje Natural a SPARQL para Realizar Búsquedas sobre Ontologías.....	6
2.1.3 Generación de ontologías en OWL a partir de diagramas de clases de UML mediante MDA (Arquitectura Dirigida por Modelos)	7
2.1.4 Almacenamiento y manipulación de ontologías que describen patrones de análisis y su semántica	7
2.1.5 Revisión del estado del arte de los algoritmos K-means y sus Mejoras	8
2.1.6 Discusión de antecedentes	9
2.2 Trabajos relacionados.....	12

2.2.1 The state of the art in automated requirements elicitation	12
2.2.2 Formal Rule Representation and Verification from Natural Language Requirements Using an Ontology	12
2.2.3 Software Requirement Elicitation Using Natural Language Processing	13
2.2.4 Survey of Works that Transform Requirements into UML Diagrams.....	14
2.2.5 Use of Ontologies in Embedded Systems: A Systematic Mapping	14
2.2.6 The Applications of Natural Language Processing (NLP) for Software Requirement Engineering - A Systematic Literature Review	15
2.2.7 A Framework using NLP to automatically convert User-Stories into Use Cases in Software Projects.....	15
2.2.8 Software patterns and requirements engineering activities in real-world settings: A systematic mapping study	16
2.2.9 Discusión de trabajos relacionados	17
2.3 Resumen del Capítulo	20
Capítulo 3. Marco Conceptual	21
3.1 Dominio del problema.....	22
3.2 Dominio de solución	22
3.3 Ingeniería de requerimientos.....	23
3.4 Reconocimiento del habla.....	23
3.5 Procesamiento de Lenguaje Natural.....	24
3.6 Ontología.....	24
3.7 Artículo encuesta o sondeo	25
3.8 Estudio sistemático	26
3.8.1 Tipos de estudios sistemáticos	26
3.9 Resumen del Capítulo	30
Capítulo 4. Metodología de búsqueda	31
4.1 Definición de preguntas de investigación	32
4.1.1 Espacio de publicación	33
4.1.2 Espacio de investigación.....	33
4.2 Búsqueda.....	36
4.2.1 Selección de base de datos científica	36
4.2.2 Creación de cadena de búsqueda	37

4.2.3 Ejecución de cadena de búsqueda.....	38
4.3 Revisión de artículos.....	39
4.4 Identificación de palabras clave.....	40
4.5 Análisis de estos artículos y creación del mapa resultante.....	42
4.6 Resumen del Capitulo.....	42
Capítulo 5. Resultados del estudio.....	44
5.1 Preguntas de publicación.....	45
5.1.1 PQ1.- ¿Qué tipo de publicación es? (Conferencia/Revista).....	45
5.1.2 PQ2.- ¿Cuál es la cantidad de artículos o conferencias publicados por año?.....	46
5.1.3 PQ3.- ¿Cuál es la cantidad de artículos o conferencias por país de procedencia? (De acuerdo a la institución de procedencia del autor).....	48
5.1.4 PQ4.- ¿Cuál es el ámbito del autor? (Académico/Industrial).....	49
5.1.5 PQ5.- ¿Cuáles son los congresos o revistas donde se publicaron los artículos?..	50
5.2 Preguntas de investigación.....	50
5.2.1 RQ1.- ¿Qué tema se trata en la elicitación de requerimientos y/o dominio del problema?.....	53
5.2.2 RQ2.- ¿Qué aporte se presenta en el artículo/conferencia? (Contribución).....	55
5.2.3 RQ3.- ¿Qué tipo de entrada se procesa en el trabajo reportado?.....	56
5.2.4 RQ4.- ¿Qué tipo de salidas o resultados presenta el trabajo reportado?.....	58
5.2.5 RQ5.- ¿Qué tipo de validación se utilizó en el trabajo reportado?.....	60
5.2.6 RQ6.- ¿Qué tipo de investigación se publicó?.....	62
5.2.7 RQ7.- ¿Cuál es el nivel de automatización en el trabajo reportado?.....	64
5.2.8 RQ8.-¿Qué técnicas de reconocimiento del habla, procesamiento de lenguaje natural y/o ontologías se presentan?.....	65
5.3 Resumen del capítulo.....	72
Capítulo 6. Principales hallazgos.....	73
6.1 Aspectos importantes.....	74
6.2 Técnicas de procesamiento de lenguaje natural.....	76
6.3 Resumen del capítulo.....	94
Capítulo 7. Conclusiones.....	95
8.1 Conclusiones.....	96
8.2 Aportaciones.....	97

8.3 Trabajos futuros	97
Referencias	98
Anexos.....	107

Lista de figuras

Figura 3.1. Proceso de revisión sistemática de literatura.	27
Figura 3.2. Proceso del estudio de mapeo sistemático.	29
Figura 4.1. Proceso de selección de artículos de investigación.....	40
Figura 5.1. Porcentaje de artículos de investigación de acuerdo a su tipo.	45
Figura 5.2. Artículos por año.....	46
Figura 5.3. Países con más aportes de trabajos relacionados.	48
Figura 5.4. Esquema de clasificación dado el análisis de la literatura.	52
Figura 5.5. Distribución de artículos respecto a al tipo de enfoque.	54
Figura 5.6. Tipos de artículos por año.....	54
Figura 5.7. Tipos de contribuciones en los artículos.....	55
Figura 5.8. Tipos de entrada.....	57
Figura 5.9. Distribución de artículos con base en su tipo de salida.....	59
Figura 5.10. Distribución porcentual de artículos sobre sus tipos de evaluación.....	61
Figura 5.11. Distribución de artículos de acuerdo al tipo de investigación.....	63
Figura 5.12. Distribución porcentual de artículos de acuerdo al grado de automatización..	64
Figura 5.13. Distribución de artículos de acuerdo la clasificación de técnicas de procesamiento de lenguaje natural.	66
Figura 5.14. Fases de la ingeniería de requerimientos soportados por ontologías.	69
Figura 5.15. Distribución de artículos de acuerdo el tipo de ontología.....	71
Figura 6.1. Flujo de trabajo del NLP.....	77
Figura 6.2. Clasificación de técnicas de procesamiento de lenguaje natural.	78

Lista de tablas

Tabla 2-1. Comparativa de antecedentes.....	10
Tabla 2-2. Comparativa de trabajos relacionados.	17
Tabla 4-1. Descripción de preguntas de investigación.	33
Tabla 4-2. Resultados de las diferentes bases de datos científicas al ejecutar la cadena de búsqueda.	39
Tabla 4-3. Plantilla para captura clasificación de artículos.	40
Tabla 4-4. Artículo con datos de preguntas de publicación.	42
Tabla 4-5. Artículos con datos de preguntas de investigación	42
Tabla 5-1. Referencia de artículos de acuerdo al tipo de publicación.	46
Tabla 5-2. Referencia de artículos por año.	47
Tabla 5-3. Referencia de artículos con mayor contribución por país.	48
Tabla 5-4. Referencia de artículos de acuerdo al ámbito del autor.	49
Tabla 5-5. Congresos con más aportes de artículos.	50
Tabla 5-6. Referencia de artículos de acuerdo al tipo de enfoque.	54
Tabla 5-7. Referencia de artículos de acuerdo al tipo de contribución.	56
Tabla 5-8. Clasificación de tipos de entrada identificados.	56
Tabla 5-9. Referencia de artículos de acuerdo al tipo de entrada que procesa el enfoque. ..	58
Tabla 5-10. Clasificación de tipos de salida.	59
Tabla 5-11. Referencia de artículos con base al tipo de salida o finalidad del artículo.	60
Tabla 5-12. Descripción de tipos de evaluación.	60
Tabla 5-13. Referencia de artículos de acuerdo tipo de evaluación empleado.	62
Tabla 5-14. Referencia de artículos respecto al tipo de investigación.	63
Tabla 5-15. Referencia de artículos de acuerdo al grado de automatización.	65
Tabla 5-16. Artículos de acuerdo al tipo de enfoque.	66
Tabla 5-17. Referencia de artículos de acuerdo al flujo de trabajo del NLP.	67
Tabla 5-18. Descripción de las fases de la ingeniería de requerimientos.	68
Tabla 5-19. Referencia de artículos de acuerdo a la fase de la RE que apoya el enfoque ontológico.	69
Tabla 5-20. Descripción de los tipos de ontología de acuerdo a(Castañeda et al., 2010). ...	70

Tabla 5-21. Referencia de artículos de acuerdo al tipo de ontología implementado.....	71
Tabla 6-1. Descripción de técnicas de Procesamiento de Lenguaje Natural.....	79
Tabla A-1. Referencia de artículos seleccionados para este estudio.	107
Tabla A-2. Clasificación de artículo de acuerdo a las preguntas de publicación.	113
Tabla A-3. Clasificación de artículo de acuerdo a las preguntas de publicación	120

Acrónimos

API:	<i>Application Programming Interface</i> (Interfaz de Programación de Aplicaciones)
ATL:	<i>Atlas Transformation Language</i> (Lenguaje de transformación Atlas).
CENIDET:	Centro Nacional de Investigación y Desarrollo Tecnológico.
CNL:	<i>Controlled Natural Language</i> (Lenguaje Natural Controlado).
MDA:	<i>Model Driven Architecture</i> (Arquitectura Dirigida por Modelos).
NCBO:	<i>National Center for Biomedical Ontology</i> (Centro Nacional de Ontologías Biomédicas).
NL:	<i>Natural Language</i> (Lenguaje Natural).
NLTK:	<i>Natural Language Tool Kit</i> (Kit de Herramientas de Lenguaje Natural).
NLP:	<i>Natural Language Processing</i> (Procesamiento de Lenguaje Natural).
ODM:	<i>Ontology Definition Metamodel</i> (Metamodelo para la Definición de Ontologías).
OMG:	<i>Object Management Group</i> (Grupo de Administración de Objetos).
OWL:	<i>Web Ontology Language</i> (Lenguaje Ontologías Web).
POS:	<i>Parts Of Speech</i> (Partes del habla).
PQ:	<i>Publication Question</i> (Pregunta de Publicación).
RE:	<i>Requirement Engineering</i> (Ingeniería de Requerimientos).
RQ:	<i>Research Question</i> (Pregunta de Investigación).
SBVR:	<i>Semantic Business Vocabulary and Rules</i> (Vocabulario y Reglas Semánticas de Negocios).
SLR:	<i>Systematic Literature Review</i> (Revisión Sistemática de Literatura).
SMS:	<i>Systematic Mapping Study</i> (Estudio de Mapeo Sistemático).
SPARQL:	<i>SPARQL Protocol and RDF Query Language</i> (Protocolo SPARQL y lenguaje de consulta RDF).
UML:	<i>Unified Modeling Language</i> (Lenguaje Unificado de Modelado).
W3C:	<i>World Wide Web Consortium</i> (Consortio de la Mundial de la Red).

Glosario de términos

Dominio del problema: Conocimiento sobre un dominio que se requiere saber para resolver un problema. En ingeniería de software, el dominio del problema es el conjunto de conceptos interrelacionados que son necesarios conocer para comprender el proceso de negocio y, de esta forma, comunicarse eficazmente con clientes. El objetivo es entender las necesidades de los clientes y, de esta manera, proponer una solución adecuada.

Elicitar: Según el *Diccionario panhispánico de dudas*, publicado en 2005, es una adaptación innecesaria del verbo inglés *to elicit*, que aparece a veces en textos de psicología y otras áreas como la computación, con el sentido que corresponde a los verbos españoles *provocar*, *suscitar* u *obtener*. En este documento se considera sinónimo de obtener.

Elicitación: Consiste en extraer información de componentes que generalmente no son tan expresivos o la información no es tan evidente, para esto se aplican distintas técnicas para obtener los requerimientos (de los *stakeholders* o involucrados). En este documento es sinónimo de *proceso de obtener los requerimientos que desean los involucrados*.

Estudio de Mapeo Sistemático: Un estudio de mapeo sistemático es un tipo de estudio secundario (estudio basado en el análisis de investigaciones previas). Su principal objetivo es determinar el alcance de la investigación sobre un tema de investigación específico y clasificar dicho conocimiento. Un SMS tiene un método definido para construir clasificaciones y conducir el análisis temático de manera que se obtenga un mapa visual del conocimiento existente dentro de un tema amplio. El análisis de los resultados se realiza mediante la clasificación de los hallazgos y contando la frecuencia de publicaciones dentro de cada categoría para determinar la cobertura de las distintas áreas de un tema de investigación específico. La información generada se puede combinar para responder a preguntas de investigación más específicas, con el objetivo de ahorrar tiempo y esfuerzo de investigación. Para que esto sea posible, los estudios de mapeos sistemáticos deben ser de calidad en términos de completitud y rigurosidad (Petersen, Feldt, Mujtaba, & Mattsson, 2008).

Estudio primario: Es un estudio empírico que investiga un tema en específico (Kitchenham & Charters, 2007).

Estudio secundario: Es un estudio que revisa todos los estudios primarios relacionados con un tema de investigación en específico con el objetivo de integrar y sintetizar dicha evidencia (Kitchenham & Charters, 2007).

Presentación

Actualmente, el software juega un papel importante en la sociedad. Prácticamente todos los hogares tienen al menos una computadora o un dispositivo móvil. Las empresas utilizan el software para controlar sus operaciones, los centros de salud para llevar el control de pacientes, las escuelas para el control de calificaciones, entre otras aplicaciones. Por lo tanto, se puede decir que el software es importante en las actividades diarias de las personas.

El desarrollo de software es un proceso que consta de varias etapas: análisis, diseño, desarrollo, pruebas e implementación. En la primera etapa, el analista se pone en contacto con los involucrados, ya sean usuarios y/o interesados, para saber todas las actividades que se realizan y sus restricciones (dominio del problema) e identificar los requerimientos para una solución posible (dominio de la solución). La recopilación de la descripción de las actividades, sus restricciones, así como de los requerimientos se realizan principalmente en lenguaje natural y se produce una representación del problema, así como un documento de especificación de requerimientos, usando una combinación de modelos visuales y de texto en lenguaje natural, en un alto nivel de abstracción. Así, en la etapa de análisis se trabaja en el dominio del problema (modelo del negocio) y en el dominio de solución (especificación de requerimientos). Debido a que estas actividades se realizan manualmente por el analista utilizando lenguaje natural, se presentan problemas de ambigüedad, inconsistencia y representación incompleta, además de tomar demasiado tiempo en su elaboración.

Dentro de la etapa de análisis, uno de sus objetivos es el establecimiento de los requerimientos de la solución, que generalmente, se especifican mediante expresión hablada y se transcriben a un documento redactado en lenguaje natural. Hoy en día, existen técnicas computacionales que hacen uso del Procesamiento de Lenguaje Natural (*Natural Language Processing*) que toman como entrada estos requerimientos (expresados en lenguaje informal) para tratarlos y generar una salida (modelos formales), como los trabajos de (Azzazi, 2017) y (Mohanani & Philip, 2016). Sin embargo, su utilización requiere que se conozcan los avances existentes, el estado del arte correspondiente y las áreas donde se requiere investigación. Uno de los avances es usar ontologías para la validación de los requerimientos, por ejemplo, validar que sean consistentes, completos y no ambiguos (Farfeleder, Moser, Krall, Stålhane, et al., 2011). O se pueden usar ontologías para transformación de modelos (Harmain & Gaizauskas, 2000), por nombrar algunos ejemplos.

Un SMS es un método definido para proporcionar un panorama general de un área de investigación a través de clasificación y conteo de contribuciones en relación a las categorías de dicha clasificación (Kitchenham & Charters, 2007). Esto implica buscar en la literatura qué temas se han cubierto y dónde se ha publicado la literatura (Petersen et al., 2008). Los estudios de mapeo sistemáticos se ocupan principalmente de estructurar un área de investigación. Un SMS tiene un proceso definido que consta de cinco pasos, que son los siguientes:

- 1) Definición de preguntas de investigación: Mediante las preguntas de investigación se define el panorama general de un área de investigación, se identifica la cantidad y tipos de investigación disponibles en el estado del arte actual.
- 2) Búsqueda de artículos: Se crea una cadena de búsqueda, derivado de las preguntas de investigación, con términos relevantes y operadores booleanos para ser ejecutada en las distintas bases de datos científicas y, de esta manera, obtener artículos respecto al área de investigación.
- 3) Selección de artículos relevantes: Se aplican criterios de inclusión y exclusión a los artículos arrojados por las bases de datos científicas para excluir esos artículos de investigación que no son relevantes para contestar las preguntas de investigación.
- 4) Análisis y extracción de palabras clave (Esquema de clasificación): Por medio del análisis de lectura de los artículos e identificación de palabras clave que permite desarrollar el esquema de clasificación.
- 5) Extracción de datos: Una vez definido el esquema de clasificación, los artículos son categorizados de acuerdo a dicho esquema, aquí es donde la extracción de datos toma lugar. Dichos datos son capturados generalmente en una plantilla o formato (por ejemplo, una tabla en Excel) para el proceso de extracción de información. El formato contiene las categorías del esquema de clasificación. Una vez capturados, las frecuencias de publicación en cada categoría pueden ser calculado. De esta manera, el analista de resultados se enfoca sólo en presentar las frecuencias de publicación en cada categoría en un documento (mapa) que muestra un panorama visual del área de investigación. Esto permite ver qué categorías se han enfatizado en investigaciones pasadas y así identificar brechas y posibilidades para futuras investigaciones.

El resultado de un SMS es un reporte visual (mapa) que brinda un panorama general de la evidencia existente de un área de investigación particular. En este trabajo de investigación se realizó un estudio de mapeo sistemático (SMS), bajo las directrices de (Petersen et al., 2008), acerca del uso del procesamiento de lenguaje natural (tanto voz como texto) y las ontologías en la elicitación de requerimientos y/o descripción del dominio del problema. Este documento presenta el reporte de los resultados del SMS, en el cual, a partir de una búsqueda inicial que resultó en 9783 artículos, posteriormente se aplicaron criterios de inclusión y exclusión para obtener un resultado final de 64 artículos en el periodo 2010-2018. A este conjunto de artículos se procedió a realizar un análisis para generar un esquema Resultando la identificación de distintas técnicas, métricas y modelos para el procesamiento de requerimientos en lenguaje natural, que apoyan la automatización de la elicitación de requerimientos y/o la descripción del dominio del problema. Del mismo modo, se identificaron tres tipos de ontologías que brindan soporte a las cuatro fases de la ingeniería

de requerimientos: elicitation de requerimientos; análisis y negociación de requerimientos; documentación y validación de requerimientos; y la administración de requerimientos.

Este documento de tesis se divide en seis capítulos:

Capítulo 1. Introducción: Se presenta el contexto del problema, descripción del problema, el objetivo general y los objetivos específicos, la justificación, alcances y limitaciones.

Capítulo 2. Marco de Referencia: Se muestran información respecto a antecedentes y trabajos relacionados que sirven como punto de referencia del trabajo propuesto.

Capítulo 3. Marco Conceptual: Se describen los conceptos fundamentales que determinan el contexto de este trabajo de investigación.

Capítulo 4. Metodología de búsqueda: Se presentan las actividades realizadas para llevar a cabo la revisión de contribuciones bibliográficas de acuerdo a las directrices de (Petersen et al., 2008).

Capítulo 5. Resultados del estudio: Se presentan el resumen visual del análisis de los artículos de investigación recopilados mediante el proceso de estudio de mapeo sistemático.

Capítulo 6. Principales hallazgos: Se presentan los principales hallazgos que se identificaron en resultados obtenidos en este estudio de mapeo sistemático.

Capítulo 7. Conclusiones: Se incluyen las conclusiones obtenidas del análisis de evidencia existente y los trabajos futuros que se derivan de los resultados de este estudio de mapeo sistemático.

Capítulo 1.

Introducción

En el presente capítulo se expone el problema que dio origen al proyecto de investigación descrito en esta tesis; se definen los objetivos, la justificación los alcances y limitaciones de la investigación.

1.1 Contexto del problema

Hoy en día, el software forma parte importante en la vida de las personas, es el responsable de soportar las tecnologías de las que depende la sociedad, para operar procesos de negocios; salud; energía; educación y ciencia. Sin embargo, las demandas de software innovador y rentable, está aumentando exponencialmente, en todos los dominios.

El reporte “*Future of Software Engineering Research (FoSER)*” (Networking and Information Technology Research and Development, 2011), es el resultado del análisis realizado por un comité del gobierno del Estados Unidos, considerando los trabajos presentados en un taller internacional, donde se reunieron a los mejores investigadores del área. Que identifica problemas de investigación y oportunidades para el futuro de la ingeniería de software, en el cual se identificaron cinco temas, basándose en los hilos temáticos comunes:

1. Ayudar a la gente a producir y usar sistemas de software;
2. Diseñar sistemas complejos para el futuro;
3. Crear sistemas de software confiables;
4. Mejorar la toma de decisiones, evolución y economía;
5. Avanzar nuestra disciplina y metodología de investigación.

Dentro del tema tres, se identificaron cinco subtemas, descritos a continuación:

- a. Habilitar programación amigable al usuario.
- b. Automatizar la evolución del software.
- c. Diseñar para confiabilidad.
- d. Emplear análisis de programas diferenciales e interactivos.
- e. Mejorar la usabilidad.

Enfocándose en el subtema “Habilitar programación Amigable-al-Usuario”, se identifican las siguientes necesidades:

- i. Habilitar a la gente a programar computadoras expresando el comportamiento computacional deseado, usando lenguajes informales y otras maneras naturales de expresión, en lugar de lenguajes de programación formales y esotéricos.
- ii. Desarrollar nuevos enfoques, asistidas por computadora, para refinamiento iterativo incremental de especificaciones en lenguaje natural informal, hacia programas de computadora con semántica restringida. De manera que el refinamiento iterativo produzca código ejecutable, especificaciones precisas, conjuntos de pruebas y demás.

- iii. Desarrollar los fundamentos para procesar lenguaje natural y otras maneras de expresión humana, ligadas a la síntesis automatizada de software que amplíe la participación del usuario en el cómputo, mientras se mejora el diseño de software y la productividad.

1.2 Descripción del problema

Se identificó que no se cuenta con un estado del arte actual sobre el uso de técnicas de reconocimiento del habla, procesamiento del lenguaje natural y enfoques de ontologías para apoyar el modelado del dominio del problema e identificación de los requerimientos de solución (primeras actividades de la etapa de análisis del desarrollo de software). Debido a lo anterior, se presentan brechas de investigación en esta área, lo que conduce a que todavía se realice la elicitación de requerimientos por técnicas tradicionales llevadas a cabo por un analista (proceso manual). Esta intervención humana en la elicitación de requerimientos tiende a generar problemas de entendimiento del dominio del problema, comunicación ineficiente con los clientes y captura incompleta de requerimientos que afectan el éxito del desarrollo de un proyecto de software.

1.3 Objetivos

Considerando la descripción del problema, en esta sección se presenta el objetivo general y los objetivos específicos de este trabajo de investigación.

1.3.1 Objetivo General

Realizar un estudio bibliográfico (estado del arte) que brinde un panorama general acerca del reconocimiento del habla, procesamiento del lenguaje natural y uso de ontologías en apoyo al modelado del dominio del problema y la elicitación de requerimientos.

1.3.2 Objetivos Específicos

1. Identificar aquellas contribuciones bibliográficas que presenten estudios acerca del reconocimiento del habla, procesamiento de lenguaje natural y uso ontologías en el modelado del dominio del problema y elicitación de requerimientos.
2. Realizar un esquema de clasificación mediante el análisis temático de las contribuciones bibliográficas con el propósito de brindar un panorama general (estructuración del área de investigación) del estado del arte actual.

1.4 Justificación

Durante el transcurso de los años, numerosos estudios se han publicado con respecto al uso del procesamiento de lenguaje natural y ontologías, en el ámbito de actividades relacionadas

con el modelado del dominio del problema y la elicitación de los requerimientos de software. A medida que un área de investigación madura, se incrementa la cantidad de informes y resultados disponibles, por lo que es importante resumir y proporcionar una visión general (Petersen et al., 2008).

En este contexto, la identificación de evidencia bibliográfica, mediante un estado del arte, permite brindar un panorama general de los estudios disponibles e identificar la estructura de las líneas de investigación. Esto proporcionará elementos de decisión para establecer nuevas áreas de investigación en el ámbito de la ingeniería de software.

1.5 Alcances y Limitaciones

En este apartado se presentan los alcances y limitaciones de este trabajo de investigación.

1.5.1 Alcances

1. Se consideraron como objeto de estudio aquellos artículos que tratan acerca del uso del procesamiento de lenguaje natural y uso de ontologías en el modelado del negocio y elicitación de requerimientos. Específicamente en la etapa de análisis, sólo se tomaron como evidencia las investigaciones relacionados al dominio del problema.
2. Se elaboró un reporte de investigación con los resultados presentados como resumen visual.

1.5.2 Limitaciones

1. Se recopiló material de las fuentes bibliográficas a las que tiene acceso el CENIDET, así mismo, se buscó acceso a otras bases de datos científicas.
2. Sólo se consideraron artículos disponibles en texto completo.
3. El esquema de clasificación se realizó con base al análisis temático de los artículos seleccionados para este estudio.

1.6 Resumen del Capítulo

En este capítulo se presentó la motivación para el desarrollo del tema de investigación, el planteamiento del problema identificado y por último el enfoque de solución propuesto. En el siguiente capítulo se abordan algunos trabajos de investigación que fueron la base para este trabajo de tesis.

Capítulo 2.

Marco de
Referencia

En este capítulo se presentan los cinco trabajos antecedentes (investigaciones realizadas en CENIDET) y los ocho trabajos relacionados (investigaciones realizadas fuera de CENIDET), explicitando las similitudes, diferencias y elementos que sirvieron de base para este trabajo de investigación.

2.1 Antecedentes

2.1.1 Generación de Modelos Conceptuales a partir de Modelos de Requisitos: un Enfoque Basado en Tratamiento Automático del Lenguaje Natural

En esta tesis de maestría, (Sánchez Pacheco, 2008) presentó una herramienta, llamada “Enoch”, para la generación automática de un modelo conceptual orientado a objetos que parte de las especificaciones textuales de los casos de uso de un sistema en idioma español, basado en reglas heurísticas y detección de patrones sintácticos preposicionales en el texto. La herramienta no depende de conocimiento extralingüístico en el dominio del problema. Dicha herramienta hizo uso de técnicas de procesamiento de lenguaje natural y grafos conceptuales con enfoque en el análisis lingüístico para identificar los elementos del modelado. Además, la herramienta resuelve algunos fenómenos lingüísticos como son: oraciones complejas, composición de nombres y perífrasis verbal.

En dicha investigación, la herramienta generó modelos conceptuales de casos de prueba: los resultados generados se compararon con aquellos obtenidos por un analista para medir su desempeño con lo que se concluyó que entre más rica y expresiva sea la descripción del sistema, más completo será el modelo conceptual resultante. Con el resultado de esta tesis se observa el procesamiento del lenguaje natural de texto de casos de uso para la generación automática de modelos conceptuales.

2.1.2 Traductor de Consultas en Lenguaje Natural a SPARQL para Realizar Búsquedas sobre Ontologías

En esta tesis de maestría, (Vázquez Vázquez, 2010) menciona que las ontologías junto con las interfaces de lenguaje natural crearon las bases para los sistemas de pregunta-respuesta sobre ontologías. El principal objetivo es localizar información almacenada en ontologías mediante una consulta expresada en lenguaje natural. Estos tipos de sistemas reducen la brecha entre una consulta en lenguaje natural y el lenguaje SPARQL, que es un lenguaje de consultas para acceder a ontologías.

En esta investigación se desarrolló una herramienta que recibe, analiza y transforma una consulta en lenguaje natural a una sintaxis del lenguaje de consulta SPARQL a través del reconocimiento de nombres y acciones. Esta herramienta genera tripletas y una consulta en

lenguaje SPARQL a partir de una consulta expresada en lenguaje natural, sin importar el dominio de la base de conocimientos. La entrada de lenguaje natural se analiza mediante expresiones regulares. Como resultado se detectan las entidades(nombres), acciones(verbos) y el predicado, posteriormente se genera la tripleta y se construye la consulta en lenguaje SPARQL.

2.1.3 Generación de ontologías en OWL a partir de diagramas de clases de UML mediante MDA (Arquitectura Dirigida por Modelos)

En esta tesis de maestría, (Galindo García, 2011) propone un método de generación de ontologías a partir de diagramas de clases UML para representar modelos de conocimiento en OWL mediante transformaciones de meta-modelos de UML, meta-modelos ODM y lenguaje de transformaciones entre modelos ATL. ATL es un lenguaje de transformación de modelos que permite especificar cómo uno o más modelos objetivo pueden producirse desde un conjunto de modelos fuente. Es un lenguaje especificado como un metamodelo y que cuenta con una sintaxis concreta de texto. El principal objetivo es proveer una notación visual en el inicio del desarrollo de ontologías con el enfoque MDA que permita crear ontologías de tal manera que ayude al usuario en su formulación. Esta tesis concluyó que la herramienta generada escala perfectamente a un número creciente de clases y relaciones en el modelo UML.

Las ontologías representan conocimiento acerca de un dominio específico con el propósito de obtener una representación formal de tal conocimiento mediante conceptos y sus relaciones. Una de sus aplicaciones ha sido como herramienta de soporte a la elicitación de requerimientos de software, mediante la creación de modelos conceptuales, es decir, ontologías, lo cual resulta relevante en esta propuesta.

2.1.4 Almacenamiento y manipulación de ontologías que describen patrones de análisis y su semántica

En esta tesis de maestría, (Candia Zavaleta, 2017) menciona que los patrones de análisis son muy poco utilizados y que esto se debe a la falta de conceptos de estabilidad y la semántica en su descripción cuando se aplican en dominios diferentes. La propuesta para resolver el problema mencionado fue una estructura para almacenar descripciones de patrones de análisis utilizando una jerarquía de ontologías dividida en tres niveles de abstracción donde incluye los conceptos de estabilidad y semántica. Esta estructura se encuentra alojada en un repositorio disponible en la web que se utiliza mediante la aplicación NCBO BioPortal, de la Universidad de Standford. Utilizando NCBO Bioportal, se logró obtener los requerimientos funcionales especificados en la investigación. De esta manera, usando la aplicación NCBO BioPortal se obtuvo un repositorio de descripciones de patrones de análisis.

El resultado de esta tesis fue el almacenamiento permanente de la estructura de información mediante una aplicación web. Permitiendo que el experto en el dominio aporte conocimiento (descripciones de patrones de análisis) y que, además, el experto en ontologías genere las ontologías de los patrones de análisis y las agregue al repositorio. De esta manera, se concluyó que disponer de un repositorio de ontologías de patrones de análisis en la web ayuda a compartir conocimiento, a que personas inexpertas comprendan los procesos de domino correspondiente y ayuda a que proliferen los patrones de análisis.

El uso de una aplicación web para el almacenamiento de patrones de análisis en repositorios de ontologías indica un avance importante para el desarrollo de software. Esto justifica la búsqueda de trabajos relacionados con ontologías.

2.1.5 Revisión del estado del arte de los algoritmos K-means y sus Mejoras

En esta tesis de maestría, (Vega Villalobos, 2017) realizó una taxonomía, de las publicaciones más relevantes, de acuerdo a las etapas del algoritmo K-means, las cuales son: inicialización, clasificación, cálculo de centroides y convergencia. Dicha tesis llevo el proceso de investigación basado en los principios de revisión sistemática. El resultado de aplicar la metodología de revisiones literarias fue que se detectaron 1125 documentos, los cuales se aplicaron criterios de inclusión y exclusión, restando categóricamente 79 estudios, de los cuales 38 pertenecen a la fase de inicialización; 33 para una combinación entre las fases de clasificación y cálculo de centroides; 4 publicaciones para la etapa de convergencia; 2 para la combinación de convergencia-inicialización; y 2 para convergencia-clasificación. Esta investigación concluyó que fue factible realizar la taxonomía de acuerdo a las etapas de K-means sobre las publicaciones más relevantes de mejoras del algoritmo.

Se menciona que el algoritmo K-means es uno de los mejores algoritmos para resolver problemas de agrupamiento en distintos dominios debido a la fácil interpretación de sus patrones y la sencilla implementación computacional. Durante el transcurso de los años se han publicado un número de artículos aportando mejoras y variantes, es decir, la literatura acerca del algoritmo K-means es muy extensa lo que hace difícil la selección de artículos específicos en alguna de las etapas del algoritmo.

El aporte de esta tesis fue el proceso de investigación, que usa los principios de revisión sistemática, con el propósito de obtener las publicaciones más relevantes acerca del dominio, puesto que la información que se encuentra es muy extensa. Estos procesos sistemáticos permiten definir un área de investigación y establecer una ruta de acción en dicha área. Además, se identificaron las mejores prácticas del algoritmo K-means en sus distintas etapas.

2.1.6 Discusión de antecedentes

En la Tabla 2-1, se muestra un cuadro comparativo de los trabajos de antecedente realizados en CENIDET. Como se observa, existen investigaciones referentes al procesamiento del lenguaje natural y al uso de ontologías que soportan la etapa de análisis del proceso de desarrollo de software.

La investigación (Sánchez Pacheco, 2008) aporta el uso del procesamiento del lenguaje natural en el texto de los escenarios de los casos de uso para generar modelos conceptuales. La investigación (Vázquez Vázquez, 2010) resalta el procesamiento del lenguaje natural para generar consultas en SPARQL, que es un lenguaje de consultas de ontologías, que almacenan conocimiento.

La investigación (Galindo García, 2011) es un avance en la creación de ontologías mediante diagramas de clases (UML) para proveer de una notación visual. Es decir, esa transformación de diagrama de clases a ontologías resulta relevante en la ingeniería de requerimientos. El aporte de (Candia Zavaleta, 2017) es una aplicación web que aloja y manipula ontologías de patrones de análisis, resulta relevante puesto que la utilización de ontologías, en la etapa de análisis, permite compartir el conocimiento de patrones de análisis.

Por último la investigación de (Vega Villalobos, 2017) aporta con el uso de proceso de búsqueda (revisión sistemática de literatura) de publicaciones relevantes sobre el uso del algoritmo K-means para generar una taxonomía, dado que no se conocían las mejores prácticas en cada una de las etapas de K-means. Con este proceso pudo establecer una línea de investigación y una estrategia para abordarla.

El trabajo de (Vega Villalobos, 2017), que reporta el procesamiento para llevar a cabo la revisión del estado del arte, resultó relevante debido a que el objetivo es similar al de este trabajo de investigación.

En el presente trabajo de tesis se considera de suma importancia, el uso del reconocimiento del habla, procesamiento del lenguaje natural y ontologías en la etapa de análisis, enfocándose en el dominio del problema y dominio de solución. Sin embargo, se tiene muy poco conocimiento con respecto su estado y avances. Por lo tanto, se considera necesaria una búsqueda sistemática de contribuciones publicadas acerca de esta área. El aporte del presente trabajo de tesis es brindar un panorama general (estado del arte) actualizado respecto a esta área de investigación.

Tabla 2-1. Comparativa de antecedentes.

Trabajo Antecedente	Uso de ontologías	Uso del lenguaje natural	Procesos relacionados con el dominio problema y requerimientos de software	Búsqueda sistemática de información
Generación de Modelos Conceptuales a partir de Modelos de Requisitos: un Enfoque Basado en Tratamiento Automático del Lenguaje Natural (Sánchez Pacheco, 2008)	No aplica	Procesamiento del lenguaje natural de las especificaciones textuales de los casos de uso basado en heurística y detección de patrones sintácticos preposicionales en el texto.	Creación de herramienta “Enoch” para la generación automática de modelo conceptual orientado a objetos partiendo de especificaciones textuales de casos de uso.	No aplica
Traductor de Consultas en Lenguaje Natural a SPARQL para Realizar Búsquedas sobre Ontologías (Vázquez Vázquez, 2010)	Las tripletas que genera la herramienta, después de analizar la entrada en lenguaje natural, las utiliza para construir la consulta en lenguaje SPARQL equivalente.	Esta herramienta recibe, analiza y transforma una consulta en lenguaje natural para detectar las entidades, acciones y el predicado para crear tripletas.	No aplica	No aplica
Generación de ontologías en OWL a partir de diagramas de	Creación de herramienta para la generación de	No aplica	No aplica	No aplica

clases de UML mediante MDA (Arquitectura Dirigida por Modelos) (Galindo García, 2011)	ontologías mediante diagramas de clases UML para presentar modelos de conocimiento con el objetivo de proveer una notación visual al inicio del desarrollo.			
Almacenamiento y manipulación de ontologías que describen patrones de análisis y su semántica (Candia Zavaleta, 2017)	Generación de una estructura para almacenar descripciones de patrones de análisis utilizando una jerarquía de ontologías alojados en un repositorio disponible a través de una aplicación web.	No aplica	No aplica	No aplica
Revisión del estado del arte de los algoritmos K-means y sus Mejoras (Vega Villalobos, 2017)	No aplica	No aplica	No aplica	Realiza una revisión del estado del arte mediante los principios de revisión sistemática y <i>surveys</i> para obtener las

				publicaciones más relevantes con respecto a las etapas de K-means.
--	--	--	--	--

2.2 Trabajos relacionados

2.2.1 The state of the art in automated requirements elicitation

En el trabajo de (Meth, Brhel, & Maedche, 2013) se reporta el estado del arte acerca de la automatización en la elicitación de requerimientos, donde se seleccionaron 36 trabajos relevantes. Este trabajo de investigación realizó el estudio del estado del arte mediante una revisión sistemática de literatura. Los autores realizaron un esquema de clasificación de los artículos seleccionados, de acuerdo a la categoría de la herramienta, el grado de automatización, reuso de conocimiento, enfoque de evaluación y evaluación de conceptos.

Este trabajo de investigación resulta relevante debido al esquema de clasificación realizado que proporciona un panorama del área de investigación respecto a la elicitación de requerimientos. En este trabajo de tesis se comparten algunas clasificaciones con este trabajo relacionado, como es el grado de automatización y el enfoque de automatización.

2.2.2 Formal Rule Representation and Verification from Natural Language Requirements Using an Ontology

En este artículo de investigación (Sadoun, Dubois, Ghamri-Doudane, & Grau, 2014) mencionan que el desarrollo de un sistema de software está basado principalmente en los requerimientos, que a menudo son escritos en lenguaje natural, por lo tanto, la revisión de la consistencia y completos de los requerimientos de software, necesita tener una representación formal. En este artículo, los autores se enfocan en los requerimientos de usuario que describen el comportamiento del sistema, es decir, las reglas de comportamiento. Dicha investigación muestra cómo transformar a una especificación formal (Maude), las reglas de comportamiento expresadas en lenguaje natural y representadas en una ontología de OWL. Maude es una especificación formal mediante términos algebraicos, permite la verificación de propiedades y transformaciones sobre modelos; y permite describir la dinámica de un sistema.

La propuesta reportada en el artículo es mapear los elementos de la ontología, es decir, las instancias, a una representación formal Maude. Mediante una ontología, se representa el comportamiento de un sistema y esta ontología permite cerrar la brecha entre lenguajes formales

e informales, para automatizar la transformación de reglas en lenguaje natural a una especificación Maude.

Debido a que el objetivo de este tema de tesis es la recopilación de evidencia de la utilización de ontologías, este trabajo relacionado es relevante porque utiliza ontologías como representación forma de los requerimientos en lenguaje natural, de forma que apoya el análisis de calidad dentro de la etapa de análisis.

2.2.3 Software Requirement Elicitation Using Natural Language Processing

En este artículo de investigación (Mohan & Philip, 2016) hacen mención que los requerimientos de software usualmente están escritos en lenguaje natural, por lo tanto, se propone un método para la transformación de la especificación de requerimientos de usuario y diseño de negocios escritos en lenguaje natural, en modelos orientados a objetos. Desarrollaron la herramienta OpenNLP, para la detección de oraciones, la tokenización, el etiquetado de partes del habla (POS) y el análisis sintáctico de las especificaciones de requisitos. OpenNLP pertenece a un área de investigación con el objetivo de obtener información sobre cómo la computadora entiende y procesa el lenguaje natural. La herramienta comienza a ejecutarse extrayendo *tokens*, luego continúa analizando la sintaxis y los análisis semánticos de cada frase.

Dicha investigación recalca que es muy relevante el POS, dado que a través de partes del habla se identifican elementos orientados a objetos como son: clases, atributos y relaciones entre clases. Una vez obtenida la información, a SBVR (vocabulario y reglas semánticas de negocios) es aplicado para identificar y extraer los elementos orientados a objetos de la especificación de requerimientos, como son conceptos individuales, verbos y cuantificadores. SBVR es un estándar adoptado por la OMG que pretende ser la base para la descripción declarativa del lenguaje natural formal y detallado, es decir, describe el vocabulario deseado y las reglas para proporcionar la documentación semántica, hechos y reglas de negocios. SBVR proporciona una capacidad multilingüe, inequívoca y rica de idiomas que utilizan los diseñadores de software y gente de negocios en varios dominios.

Este trabajo relacionado aporta el uso de técnicas de procesamiento de lenguaje natural, la cual recalca la técnica POS para obtener elementos orientados a objetos de los requerimientos expresados en lenguaje natural, de manera que aporta un enfoque semiautomático en la etapa de análisis. En este trabajo de tesis tiene como uno de sus objetivos definir un esquema de clasificación de técnicas NLP utilizadas en los últimos años en la comunidad científica para el procesamiento de lenguaje natural, y este trabajo relacionado hace aporte de algunas técnicas.

2.2.4 Survey of Works that Transform Requirements into UML Diagrams

En este artículo de investigación (Abdouli, Karaa, & Ghezala, 2016) realizaron un estudio de trabajos relacionados con el proceso de transformación de requerimientos de software a diagramas UML. Que abarca desde técnicas manuales en los inicios de 1976 hasta las automatizadas en 2015. Este artículo presenta diferentes enfoques evaluando sus ventajas y desventajas. Por medio de este estudio se pueden evaluar los diferentes enfoques como alternativas para la Ingeniería de Requisitos. Se clasifica 5 enfoques en este estudio: Inspección, Procesamiento de Lenguaje Natural, Heurística (Reglas, Métricas), Patrones/Grafos y Ontologías.

El propósito de este estudio es presentar una vista general de artículos dedicados al análisis de requerimientos y compararlos. Se concluye que en todos los trabajos se hace uso, como entrada, del lenguaje natural y solo muy pocos trabajos generan diagrama UML. Los artículos se organizan en tres principales categorías: Heurística, Procesamiento de Lenguaje Natural y Ontologías. Se concluyó que el uso de ontologías produce muy pocos errores y por lo tanto son útiles para satisfacer los requerimientos textuales.

Este trabajo de investigación realizó un estado del arte sobre trabajos de investigación que se enfocan en la transformación de requerimientos de software a diagramas UML, resulta relevante dado que, similar a este trabajo de tesis, presentó un panorama general de los trabajos relacionados, sin embargo, en esta tesis se centra más en la etapa de entendimiento del dominio del problema y la identificación de requerimientos e usuario, además de llevar el proceso de búsqueda de trabajos relacionados mediante un estudio de mapeos sistemático. Por otro lado, este trabajo relacionado presenta las ventajas y desventajas de los distintos enfoques identificados, mientras que en este trabajo de tesis da un panorama visual de la evidencia existente sobre el uso de ontologías y procesamiento de lenguaje natural en el dominio del problema y la elicitación de requerimientos.

2.2.5 Use of Ontologies in Embedded Systems: A Systematic Mapping

En la investigación de (Sousa, Couto, Agra, & Alencar, 2016) se reporta un estudio de mapeo sistemático para identificar trabajos acerca del uso de ontologías en sistemas embebidos. La especificación de conceptos a menudo complica el análisis de datos y la identificación de relaciones entre ellos (etapa de análisis). Se simplifica esta labor usando ontologías. En este estudio se seleccionaron 19 artículos para analizarse a detalle, y como resultado se muestra que la mitad de los artículos usan el lenguaje OWL, recomendado por el W3C. Por otra parte se observó que la mayoría de las ontologías se desarrollaron para dominios específicos, lo que hace difícil su estandarización.

La principal contribución de este trabajo relacionado es la realización de un estudio de mapeo sistemático para identificar trabajos que usen ontologías en sistemas embebidos, aunque su objetivo va dirigido a sistemas embebido, se identificó que principalmente se utiliza en la especificación de conceptos y sus relaciones (entender el dominio del problema). En este tema de tesis, se abarca más que a sistemas embebidos y aporte de un a clasificación de tipos de ontologías y propósitos de las mismas, además del esquema de clasificación definido para las técnicas de procesamiento de lenguaje natural.

2.2.6 The Applications of Natural Language Processing (NLP) for Software Requirement Engineering - A Systematic Literature Review

En este artículo (Nazir, Butt, Anwar, & Khattak, 2017) mencionan que el NLP es una técnica muy conocida en la inteligencia artificial para la extracción de elementos de texto plano. Se utiliza para el procesamiento de requerimientos de software para especificar la solución propuesta, así como su clasificación (en requerimientos funcionales y requerimientos no funcionales). Se reporta un estudio del uso del procesamiento del lenguaje natural en el contexto de la Ingeniería de Requisitos de Software. Presenta una revisión sistemática de literatura en la que se seleccionaron 27 estudios publicados entre 2002-2016. Se identificaron seis técnicas de NLP, 2 algoritmos NLP, 14 herramientas NLP (por ejemplo, GATE tool, Open NLP, NLKT, entre otros) y otras nueve herramientas NLP que son aportes de los investigadores. Se concluyó que las técnicas de NLP acelera el proceso de Ingeniería de Requisitos de Software. Sin embargo, se requieren todavía operaciones manuales en los requerimientos de texto plano, por ejemplo limpiar el texto para procesarlo, antes de aplicar el enfoque NLP.

Este trabajo relacionado realizó una revisión sistemática de literatura (SLR) de la aplicación del procesamiento de lenguaje natural en la ingeniería de requerimientos de software, es un proceso de búsqueda diferente al de un SMS, principalmente dado que su objetivo es identificar las mejores prácticas, mientras que en un SMS es abarcar toda la evidencia existente. En este trabajo de tesis se realizó la búsqueda de evidencia mediante un SMS, principalmente que utilicen el procesamiento de lenguaje natural para entender el dominio del problema y extracción de requerimientos de usuario (elicitación de requerimientos), de manera que por medio de un esquema de clasificación se provea un panorama de las técnicas existentes.

2.2.7 A Framework using NLP to automatically convert User-Stories into Use Cases in Software Projects

En este artículo de investigación (Azzazi, 2017) menciona que para el desarrollo de software de calidad se encuentran los factores tiempo y costo, y existen diferentes enfoques para construirlo. Dentro de la etapa de análisis de software se utilizan historias de usuario (*User-Stories*) que permiten capturar lo que el usuario requiere, es decir, muestran lo que el sistema debe hacer. El

uso de historias de usuario es una manera efectiva de comunicar los requerimientos y facilita el desarrollo de software, pues son descripciones de funcionalidad cortas, sencillas y simples. En esta investigación se reporta la conversión de las *User-Stories* en diagramas de casos de uso. Que expresan una secuencia de pasos para lograr cierta funcionalidad u objetivo de un sistema de software, donde existe una interacción usuario-sistema. Se presenta un caso de estudio que valida la funcionalidad de la herramienta al detectar los actores y casos de uso, por lo que se concluyó que la herramienta es útil dado que el porcentaje de mapeo de actores fue de 75% y de casos de uso fue de 82%.

En este trabajo relacionado se observa el uso de procesamiento de lenguaje natural para convertir los *User-Stories* (lenguaje natural) en diagramas de casos de uso. Este trabajo de investigación resulta relevante para este trabajo de tesis debido a que utiliza NLP en la etapa de análisis para, en parte, entender el dominio del problema a partir de *User-Stories*. En esta tesis se identificó el uso del NLP no solo en *User-Stories* si no también en requerimientos de usuario en lenguaje natural, descripciones de casos de uso, texto, diagramas y modelos.

2.2.8 Software patterns and requirements engineering activities in real-world settings: A systematic mapping study

En este artículo (Barros-Justo, Benitti, & Cravero-Leal, 2017) mencionan que el uso de patrones de software se ha investigado en lo últimos años. Sin embargo, existe muy poca evidencia del uso de patrones de software en las primeras etapas del desarrollo de software y pocas evidencias han reportado su aplicación en el área industrial. Por lo tanto, para atender este problema, se estableció un estudio de la aplicación de patrones de software en el contexto de actividades en ingeniería de requerimientos.

Dicho estudio se realizó mediante un estudio de mapeo sistemático (SMS) donde se reunieron trabajos acerca del dominio, en el periodo 2002-2017. Mediante la búsqueda de estudios relacionados al tema se seleccionaron sólo 22 trabajos, dados ciertos criterios de exclusión e inclusión. Se identificaron 327 patrones de software, siendo los casos de uso la mayor fuente, para identificar patrones. El 50 por ciento de los estudios usaron patrones para la elicitación de requerimientos (RE), mientras que el otro 50 por ciento los usa en otras actividades de la ingeniería de requisitos. Además, se observó que usar patrones ofreció beneficios como mejorar calidad y reducir el tiempo de proyecto. Los autores concluyen que, en RE, los patrones afectan de manera beneficiosa y que son útiles para los desarrolladores novatos.

La principal aportación de este trabajo relación para este trabajo de tesis fue el proceso del estudio de mapeo sistemático (SMS) y el reporte de resultados presentados. Aunque este SMS su objetivo es identificar los patrones de software en actividades de la ingeniería de requerimientos, el mapa sistemático presentado es similar a este trabajo de tesis, debido a que el objetivo de un SMS es presentar panorama visual del área de investigación, mediante

clasificaciones de tipo informativo como: año de publicación, tipo de artículos (conferencia/revista), base de datos, identificación de congresos y revistas, país de origen; que son datos informativos presentes en todos los SMS. Lo que realmente cambia es la clasificación de acuerdo al área de investigación, para este trabajo de tesis fue la clasificación de: tipo de investigación, tipos de entradas y salidas, tipo de evaluación, tipo de contribución, nivel de automatización y principalmente las técnicas de NLP y uso de ontologías.

2.2.9 Discusión de trabajos relacionados

En la Tabla 2-2, se muestra un cuadro comparativo de trabajos relacionados a esta investigación. Como se puede observar existen trabajos relacionados con el procesamiento de lenguaje natural y ontologías en la ingeniería de requisitos. Dentro de este conjunto de trabajos relacionados, hay que resaltar los trabajos que realizaron un estudio del estado del arte, como en la investigación de (Meth et al., 2013), donde se presenta un estado del arte acerca de la automatización en la elicitación de requerimientos. Este trabajo relacionado contribuyó a la investigación reportada en esta tesis, mostrando un esquema de clasificación donde se categorizaron los trabajos de investigación recuperados.

En la investigación de (Abdouli et al., 2016) se realizó un estado del arte sobre trabajos de investigación que transforman los requerimientos de software a diagramas UML. Por otro lado, en el trabajo de (Nazir et al., 2017) se estudió el uso del lenguaje natural en la ingeniería de requisitos mediante la técnica de revisión sistemática de literatura de (Kitchenham, 2004). La principal aportación de los estos trabajos relacionados es el proceso para poder llevar a cabo el estado del arte como además de algunos esquemas de clasificación de la evidencia recuperada donde algunos se tomaron para este tema de tesis.

En la investigación de (Barros-Justo, Benitti, et al., 2017) se realizó un estudio de mapeo sistemático para obtener un panorama de la aplicación de los patrones de software en el contexto de la ingeniería de requisitos. Resulta muy relevante, dado que el proceso de realización de estudio de mapeo sistemático está basado en la metodología de (Petersen, Vakkalanka, & Kuzniarz, 2015) y es un ejemplo útil del proceso a seguir.

Tabla 2-2. Comparativa de trabajos relacionados.

Trabajo Relacionado	Enfoques de ontologías	de Enfoques del procesamiento de lenguaje natural	del Proceso de búsqueda de información
The state of the art in automated requirements elicitation	Se recuperaron trabajos respecto al uso de ontologías en la elicitación de	Se recuperaron trabajos respecto al uso de procesamiento de lenguaje natural de documento de texto	Proceso de búsqueda de realizado mediante un el proceso de revisión sistemática de literatura

(Meth et al., 2013)	requerimientos, principalmente ontologías de dominio.	para análisis de calidad de ambigüedad, inconsistencia y el entendimiento del dominio.	definido por (Kitchenham, 2004).
Formal Rule Representation and Verification from Natural Language Requirements Using an Ontology (Sadoun et al., 2014)	Hace uso de ontologías para la representación del comportamiento de un sistema de software.	Los requerimientos de usuario están representadas en lenguaje natural las cuales se representan formalmente mediante ontologías.	No aplica
Software Requirement Elicitation Using Natural Language Processing (Mohan & Philip, 2016)	No aplica	Hace uso del procesamiento de lenguaje natural (OpenNLP) de requerimientos de software para la detección de oraciones, tokenización, etiquetados de partes del habla y análisis sintáctico.	No aplica
Survey of Works that Transform Requirements into UML Diagrams (Abdoui et al., 2016)	En este estudio se encontraron trabajos relacionados al uso de ontologías.	En este estudio se encontraron trabajos relacionados con el procesamiento de lenguaje natural.	Hace un estudio de trabajos relacionados a la transformación de requerimientos en diagramas UML mediante la técnica <i>survey</i> .
Use of Ontologies in Embedded	Recuperación de artículo de trabajos de	No aplica	Realizo el proceso de búsqueda de trabajos relacionados bajo las

Systems: A Systematic Mapping (Sousa et al., 2016)	investigación de uso de ontología en la etapa de análisis para representar el dominio.		directrices de (Kitchenham, 2004).
The Applications of Natural Language Processing (NLP) for Software Requirement Engineering - A Systematic Literature Review (Nazir et al., 2017)	No aplica	En esta investigación se hace un estudio de trabajos de aplicación del procesamiento del lenguaje natural en la ingeniería de requerimientos.	El estudio realizado se realizó mediante la técnica de revisión sistemática de literatura con la cual se contiene los trabajos con las mejores prácticas.
A Framework using NLP to automatically convert User-Stories into Use Cases in Software Projects (Azzazi, 2017)	No aplica	Realiza procesamiento de lenguaje natural de User-Stories a diagramas de casos de uso.	No aplica
Software patterns and requirements engineering activities in real-world settings: A systematic mapping study (Barros-Justo, Benitti and	No aplica	No aplica	Realización de estudio de mapeo sistemático del uso patrones de software en el contexto de ingeniería de requisitos.

Cravero-Leal, 2017)			
------------------------	--	--	--

2.3 Resumen del Capítulo

En este capítulo se mostraron algunos trabajos realizados en CENIDET y en otras instituciones respecto al uso de procesamiento de lenguaje natural, ontologías y realización de estado del arte mediante procesos sistemáticos en la etapa de dominio del problema y elicitación de requerimientos de la fase de análisis del proceso del desarrollo de software. En el siguiente capítulo se presentan los conceptos necesarios que permitan entender el dominio de este trabajo de tesis.

Capítulo 3.

Marco

Conceptual

En este capítulo se muestra el marco conceptual, donde se abordan conceptos relacionados al lenguaje natural, ontologías y de estudios sistemáticos, los cuales son conceptos que permiten comprender la terminología utilizada en esta investigación. Como primer concepto se describe dominio del problema en el apartado 3.1 y después el concepto dominio de la solución en el apartado 3.2 debido a que son dos conceptos relacionados, pero se debe tener en claro la diferencia dado que este tema de tesis se enfoca en el dominio del problema. En el apartado 3.3 se explica las principales actividades de la ingeniería de software, entre las cuales se describe la elicitación de requerimientos, debido a que este trabajo de tesis se enfoca en esta etapa. En los apartados 3.4, 3.5 y 3.6 se describen los términos reconocimiento del habla, procesamiento de lenguaje natural y ontologías, que son las áreas de investigación aplicada en el dominio del problema y elicitación de requerimientos en los que se centró este trabajo de tesis. Por último, en los apartados 3.7 y 3.8 se explican brevemente los términos *survey paper* y estudio sistemático.

3.1 Dominio del problema

Un dominio de problema se refiere al área temática para la que se especificará una solución a crear mediante software. El diseño de sistema-software, dirigido por dominio, hace hincapié en la necesidad de centrarse en el dominio (que abarca el problema y solución), por encima de cualquier otra cosa, cuando se trabaja en la creación de software para sistemas empresariales complejos y de gran escala. Los expertos en el dominio del problema trabajan con el equipo de desarrollo para centrarse en las áreas del dominio que son útiles para poder producir software valioso. Por ejemplo, cuando se escribe software para la industria de la salud para registrar el tratamiento del paciente, no es importante aprender a ser médico. Lo que es importante comprender es la terminología de la industria de la salud, cómo los diferentes departamentos ven a los pacientes y la atención que se les proporciona, qué información recopilan los médicos y qué hacen con ella (Millet & Tune, 2015).

Área de experiencia o aplicación que necesita conocerse para resolver un problema. En el ámbito de los sistemas de información, el dominio del problema es el conjunto de conceptos interrelacionados que es necesario conocer para entender el negocio del cliente, y, por lo tanto, para poder entender sus necesidades y proponer una solución adecuada. El objetivo principal del "estudio del dominio del problema" es obtener un conocimiento suficiente para poder comunicarse eficazmente con clientes y usuarios, comprender su negocio, entender sus necesidades y poder proponer una solución adecuada (Desarrollo de Marco de la junta de Andalucía, 2019).

3.2 Dominio de solución

El dominio de la solución es el dominio en el que los ingenieros utilizan su ingenio para resolver problemas. La característica principal que diferencia el dominio de la solución del

dominio del problema es que, invariablemente, la ingeniería de requisitos se ubica en el dominio de la solución y describe la estructura y elementos adicionales a los del problema, que soportan una solución y consta de un conjunto dado de requisitos. En el dominio del problema, la ingeniería de requisitos comienza con un objetivo vago o una lista de deseos.

Del grado en que los requisitos de entrada para el dominio de la solución están "bien formados" depende la calidad de la solución que utilizarán las personas, dentro de la organización del cliente que utilizará dicha solución. En un mundo ideal, todos los requisitos serían claramente articulados, requisitos que se pueden probar individualmente (Hull et al., 2017).

3.3 Ingeniería de requerimientos

En la investigación de (Farfeleder, Moser, Krall, Ståalthane, et al., 2011) se menciona que la ingeniería de requerimientos es una disciplina que trata de comprender, documentar, comunicar e implementar las necesidades de los clientes. Por lo tanto, la comprensión y la gestión insuficientes de los requerimientos pueden considerarse como la principal causa de fracaso del proyecto. Las principales actividades de un proceso de ingeniería de requisitos se pueden definir de la siguiente manera:

1. Elicitación de requerimientos. Obtención de requerimientos, que involucra al personal técnico que trabaja con los clientes para averiguar sobre el dominio de la aplicación, los servicios que debe proporcionar el sistema y las limitaciones operativas del sistema.
2. Análisis de requerimientos y negociación. Es una actividad que tiene como objetivo descubrir problemas y conflictos con los requisitos y llegar a un acuerdo sobre los cambios para satisfacer a todos los interesados del sistema (personas afectadas por el sistema propuesto). El objetivo final es alcanzar un entendimiento común de los requisitos entre todos los participantes del proyecto.
3. Documentación de requerimientos y validación. Los requisitos definidos, escritos en una especificación de requisitos de software, se validan según criterios como corrección, integridad, consistencia, verificabilidad, no ambigüedad, trazabilidad, etc.
4. Gestión de requerimientos. La gestión de requisitos consiste en gestionar los cambios de requisitos (manteniéndolos coherentes), por ejemplo, asegurando la trazabilidad de los requisitos (identificación de las interdependencias entre los requisitos y con los artefactos resultantes).

3.4 Reconocimiento del habla

El habla es el principal modo de comunicación entre los seres humanos y también la forma más natural y eficiente de intercambiar información entre humanos. Por lo tanto, es lógico

que el próximo desarrollo tecnológico sea el reconocimiento de voz en lenguaje natural para el área de interfaz humano computadora. El reconocimiento de voz se puede definir como el proceso de conversión de la señal de voz a una secuencia de palabras mediante un algoritmo implementado como un programa de computadora. El procesamiento del habla es una de las áreas emocionantes del procesamiento de señales (Gaikwad, Gawali, & Yannawar, 2010).

3.5 Procesamiento de Lenguaje Natural

(Rodríguez Correa & Benavides Cañón, 2007) menciona que el lenguaje natural se entiende como el lenguaje hablado y escrito con el propósito que exista comunicación entre varias personas. La comunicación es importante en el lenguaje natural debido a que este proceso involucra la transmisión y recepción de información.

El Procesamiento del Lenguaje Natural (NLP, *Natural Language Processing*) es una subdisciplina de la inteligencia artificial y una rama de la ingeniería lingüística computacional; NLP construye sistemas y mecanismos que permitan la comunicación entre personas y máquinas por medio de lenguajes naturales. NLP busca poder crear programas que puedan analizar, entender y generar lenguajes que los humanos utilizan habitualmente, de manera que el usuario pueda llegar a comunicarse con la computadora de la misma forma que lo haría con un ser humano. NLP tiene diversas aplicaciones:

1. Corrección de textos.
2. Traducción automática.
3. Recuperación de la Información.
4. Extracción de información y resúmenes.
5. Búsqueda de documentos.
6. Sistemas inteligentes para la educación y el entrenamiento.

3.6 Ontología

(T. R. Gruber, 1993) define ontología como: “Una especificación explícita de una conceptualización”. Años después, (Borst, 1997) la define como: “Una ontología es una especificación formal de una conceptualización compartida”. Dadas estas definiciones, (Studer, Benjamins, & Fensel, 1998) utilizó estas dos definiciones para expresar una mejorada: “Una ontología es una especificación formal y explícita de una conceptualización compartida”. Para explicar esta definición se presenta a continuación, el significado los conceptos que la forman:

1. Conceptualización: Se refiere a un modelo abstracto de algún fenómeno en el mundo, construido mediante la identificación de los conceptos relevantes de dicho fenómeno.

2. Explícito: Significa que el tipo de concepto utilizado, y las restricciones de su uso se definen explícitamente.
3. Formal: Se refiere a que la ontología debe ser legible por una computadora, lo que excluye el lenguaje natural.
4. Compartida: Refleja la noción de que una ontología captura el conocimiento consensual, es decir, el conocimiento no es solo de un individuo, sino que es aceptado por un grupo.

Las ontologías suelen especificarse en idiomas que permiten la abstracción de estructuras de datos y estrategias de implementación; en la práctica, los lenguajes de las ontologías tienen un poder expresivo más cercano a la lógica de primer orden que los lenguajes utilizados para modelar bases de datos. Por esta razón, se dice que las ontologías están en el nivel "semántico", mientras que el esquema de la base de datos son modelos de datos en el nivel "lógico" o "físico".

En el contexto de la computación, una ontología define un conjunto de primitivas representacionales para modelar un dominio de conocimiento. Estas primitivas representacionales son las clases(conjuntos), atributos(propiedades) y relaciones (entre los miembros de la clase). Las definiciones de las primitivas representacionales incluyen información sobre su significado y restricciones en su lógica de aplicación consistente (T. Gruber, 2009).

3.7 Artículo encuesta o sondeo

Un artículo encuesta (*survey paper*, *review paper* o *overview paper*) se define como un documento que resume y organiza los resultados de investigaciones recientes en una forma novedosa de tal manera que, integra e incluye entendimiento del trabajo en el área. Un artículo tipo *survey* asume un conocimiento general del área; enfatiza en la clasificación de la literatura existente, desarrolla una perspectiva, evalúa nuevas tendencias en el área y tiene como principal objetivo proveer al lector una vista de trabajos existentes en el tema de manera organizada y fácil de comprender. (ACM Computing Surveys, n.d.).

Un *survey paper* reúne información de una manera organizada y metódica acerca de características de interés sobre algunos o todas las unidades de una población usando conceptos, métodos y procedimientos bien definidos, y compila dicha información en una forma de resumen útil. Un *survey* usualmente inicia con la necesidad de datos donde no hay o es insuficiente la que existe (Canada, 2010).

Un *survey paper* clasifica los diferentes enfoques existentes para el problema, utilizando un buen conjunto de criterios de clasificación. La idea principal detrás de los trabajos del *survey* es construir un marco de un tema de investigación basado en la literatura existente. El trabajo de la encuesta no publica nuevas investigaciones. Examinar documentos

de investigación significa recopilar datos y resultados de otros trabajos de investigación (Mazumdar, 2013).

3.8 Estudio sistemático

Un estudio sistemático es un método de investigación bien definido (guía) cuyo objetivo es interpretar conocimiento de un área en particular mediante el análisis de la evidencia existente. Es un tipo de estudio secundario porque se enfoca en analizar investigaciones previas (estudios primarios). Actualmente, existen dos tipos de estudio sistemático, estas son: Revisión sistemática de literatura y estudio de mapeo sistemático.

3.8.1 Tipos de estudios sistemáticos

3.8.1.1 Revisión Sistemática de Literatura (SLR)

Una revisión sistemática de literatura (también conocido como revisión sistemática) es una forma de estudio que usa una metodología bien definida para identificar, analizar e interpretar toda la evidencia disponible relacionada a una pregunta de investigación específica, de una manera que sea imparcial. Las revisiones sistemáticas deben llevarse a cabo de acuerdo con una estrategia de búsqueda predefinida (Kitchenham & Charters, 2007).

La principal desventaja de esta metodología es que se requiere un esfuerzo considerablemente mayor que el requerido por las revisiones bibliográficas tradicionales. El proceso, presentado por (Kitchenham & Charters, 2007), se muestra en la Figura 3.1. En la fase 1: 1) se definen las preguntas de investigación que formulan el tema de interés y que sus respuestas serán el contenido de la revisión sistemática; 2) se describen los protocolos que sirven de guía para la búsqueda, así como; 3) la validación del protocolo. En la fase 2: 4) se realiza la identificación de investigaciones relevantes de acuerdo a ciertos parámetros establecidos; 5) de estas investigaciones se seleccionan los estudios primarios (documentos a revisarse); 6) se evalúa la calidad de los estudios, 7) se realiza la selección de los más relevantes; 8) se extraen de los datos que serán la base del estudio. Por último, en la fase 3: 9) se escribe el reporte y 10) se valida el reporte resultante (Kitchenham, 2004).

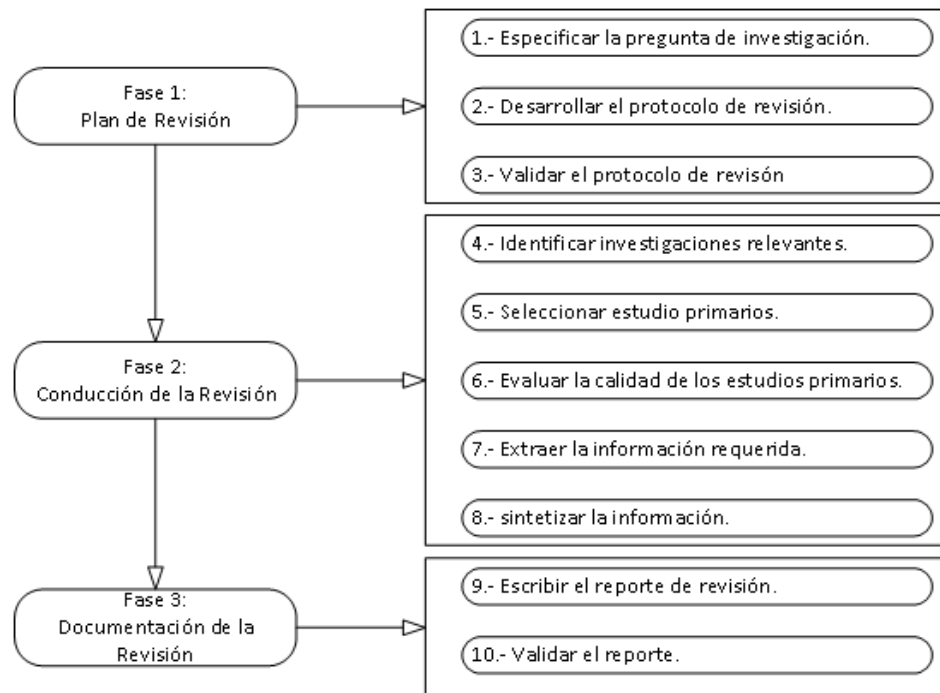


Figura 3.1. Proceso de revisión sistemática de literatura.

3.8.1.2 Estudio de Mapeo Sistemático (SMS)

Al pasar de los años, las áreas de investigación maduran debido a que muchos investigadores publican una gran cantidad de informes y resultados, por lo que es de suma importancia tener disponible un resumen que proporcione un panorama del avance que se tiene hasta el momento. Un estudio de mapeo sistemático (también conocido como mapa sistemático) brinda un panorama general de un área de investigación, a través de la clasificación y conteo de contribuciones (artículos) que tratan temas de las categorías definidas. Con el propósito de identificar las áreas de investigación que no se ha abordado mucho, para trabajarlas. Este proceso implica buscar literatura acerca de áreas trabajadas y que se han reportado los resultados en publicaciones.

Un estudio de mapeo sistemático proporciona una estructura para el informe de investigación, que incluye resultados que se han publicado, categorizados y presentados en un resumen visual o “mapa”. A menudo su construcción requiere menos esfuerzo y además proporciona una visión general. Se han recomendado estudios de mapeo sistemáticos en ingeniería de software principalmente para áreas de investigación donde faltan estudios primarios relevantes y de alta calidad.

El proceso del mapeo sistemático implica las cinco actividades de la Figura 3.2 que se describen a continuación (Petersen et al., 2008).

1. Definición de Preguntas de Investigación (Alcance de Investigación): El principal objetivo de un estudio de mapeo sistemático es proveer un resumen visual del área de investigación. Para lograr este objetivo es necesario detectar la frecuencia de estudios publicados. Como objetivo secundario de los mapeos sistemáticos, se identifican los foros de publicación. Los objetivos mencionados orientan la definición de las preguntas de investigación.
2. Búsqueda (De todos los artículos relacionados): Para poder identificar la evidencia literaria en bases de datos científicas se requiere proponer una cadena de búsqueda, que permita obtenerlas. Para definir una cadena de búsqueda se estructura en términos de los tipos de: población, intervención, comparación y resultado (Kitchenham, 2004), y orientada a las preguntas de investigación establecida.
3. Revisión de documentos (Para identificar estudios relevantes): Los criterios de exclusión e inclusión nos permiten discriminar estudios que no son relevantes o que no contestan las preguntas de investigación. La exclusión de artículos se basa en la lectura de títulos y resúmenes, como también el acceso al artículo completo.
4. Palabras clave usando resúmenes (Para estructurar el esquema de clasificación): En este paso, se deben leer los resúmenes para buscar palabras clave y conceptos que reflejen la contribución del documento. Al hacerlo, se identifica el contexto de la investigación. Con el conjunto de palabras clave identificadas en documentos diferentes, se construyen combinaciones que permita un entendimiento de alto nivel, sobre la naturaleza y la contribución de la investigación. Esto ayuda a los revisores a encontrar un conjunto de categorías representativas de la población subyacente.
5. Extracción de datos y proceso de mapeo (Creación del mapa sistemático): Una vez obtenido el esquema de clasificación, los artículos relevantes se ordenan y colocan dentro de esta categorización. Así se extraen los datos, por ejemplo, tipo de investigación, tipo de contribución, año de publicación y base de datos. Este proceso se puede realizar mediante una tabla donde se van recopilando los estudios. Para presentar un resumen visual que muestre un panorama del área de investigación, que permita identificar las tendencias, las ubicaciones donde hay ausencia de datos en un área, así como los aportes de cada congreso o revista de publicación.

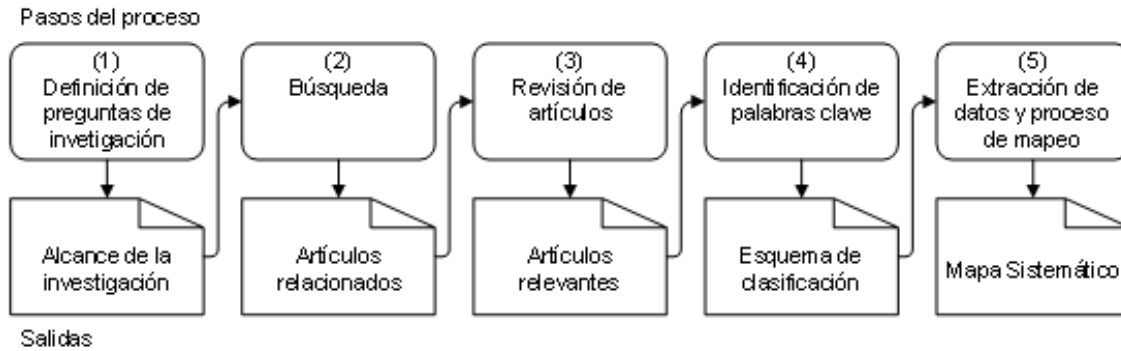


Figura 3.2. Proceso del estudio de mapeo sistemático.

3.8.1.3 Principales diferencias entre SLR y SMS

En el trabajo de (Petersen et al., 2008) se presenta una comparación entre estos dos tipos de estudios. A continuación, se presentan brevemente las principales diferencias:

Diferencia de objetivos: Ambos estudios se centran en identificar brechas de investigación. Sin embargo, una revisión sistemática se enfoca en identificar las mejores prácticas basados en estudios empíricos. Este no es el objetivo de un mapa sistemático, debido a que no se estudian los artículos seleccionados con suficiente detalle. En lugar de ello, su principal enfoque es la clasificación conducida por el análisis temático e identificación de foros de publicación.

Diferencia en proceso: Los dos estudios tiene procesos muy diferentes. En mapas sistemáticos, los artículos no son evaluados respecto a su calidad. Segundo, el método de extracción de datos es diferente. El análisis temático de un estudio de mapeo sistemático es un interesante método de análisis debido a que ayuda a identificar que categorías están muy bien cubiertas en términos de numero de publicación. En revisiones sistemáticas, el método de análisis requiere otro nivel de extracción de datos para continuar trabajando con datos cuantitativos en los estudios primarios.

Diferencia de amplitud y profundidad: En un estudio de mapeo sistemático, más artículos pueden ser considerados debido a que no son evaluados con suficiente detalle. Por lo tanto, un gran campo puede ser estructurado. Por otro lado, una revisión sistemática establece el resultado y la evaluación de calidad de los artículos como un enfoque principal, lo que aumenta la profundidad y, por lo tanto, el esfuerzo requerido. Esto podría requerir un enfoque más específico del estudio y, por lo tanto, se incluirían menos estudios.

Clasificación del enfoque de investigación: En estudios de mapeos sistemático se usan categorías generales para evaluar el artículo en términos de novedad y evaluación. Esto es debido a que los artículos no son evaluados con detalle. Por otro lado, se debe utilizar un esquema de clasificación diferente para las revisiones sistemáticas, ya que el enfoque de

investigación empírica se evalúa con mucho más detalle. Para evaluar un artículo respecto a estas categorías requiere un análisis mucho más profundo del artículo.

3.9 Resumen del Capítulo

En este capítulo se presentaron los conceptos básicos para entender el contexto del presente trabajo de tesis. Se describieron los conceptos dominio del problema y dominio de la solución, las principales actividades de la ingeniería de requerimientos, los términos reconocimientos del habla, procesamiento de lenguaje natural, ontologías, *survey paper* y las diferencias entre los tipos de estudio SLR y SMS. Este trabajo de tesis identificó artículos de investigación que usan reconocimientos del habla, procesamiento de lenguaje natural y ontologías en elicitación de requerimientos y dominio del problema. El proceso de realización del estado del arte fue realizado mediante un estudio de mapeo sistemático. El siguiente capítulo presenta el proceso realizado (estudio de mapeo sistemático) para la identificación de artículos de investigación como objeto de este estudio.

Capítulo 4.

Metodología de

búsqueda

En el presente capítulo, se representan las actividades realizadas para construir el estudio de mapeo sistemático reportado en el presente documento de tesis, siguiendo la metodología propuesta por (Petersen et al., 2008), representada en la Figura 3.2. Donde primero (actividad 1, sección 4.1), se crearon las preguntas de investigación para definir el alcance de la investigación, después (actividad 2, sección 4.2) la selección de bases de datos científicas y creación de cadena de búsqueda para la obtención de literatura, luego (actividad 3, sección 4.3) se definieron los criterios de inclusión y exclusión para seleccionar las investigaciones relevantes, a continuación (actividad 4, sección 4.4) se realiza el correspondiente análisis de estos artículos seleccionados para obtener el esquema de clasificación y finalmente (actividad 5, sección 4.5), se presenta brevemente el proceso de realización del mapa sistemático.

4.1 Definición de preguntas de investigación

En el desarrollo de cualquier estudio sistemático, ya sea mapeo sistemático (SMS) o revisión sistemática de literatura (SLR), es necesario fijar rumbo hacia el tema que se requiere estudiar, para lograrlo es necesario establecer objetivos. El objetivo principal de este estudio de mapeo sistemático se muestra a continuación:

“Identificar técnicas de reconocimiento del habla, procesamiento de lenguaje natural y uso de ontologías para la identificación del dominio del problema y requerimientos de solución en el contexto de la ingeniería de software”

Dado el objetivo anteriormente descrito, se planteó una pregunta de investigación principal, que define el alcance de la investigación.

“¿Qué investigaciones se han reportado sobre el uso de reconocimiento del habla, procesamiento del lenguaje natural y uso de ontologías para identificar el dominio del problema y requerimientos de solución en la ingeniería de software?”

Como se menciona en la investigación de (Barros-Justo, Cravero-Leal, Benitti, & Capilla-Sevilla, 2017), la pregunta de investigación principal tiende a ser bastante amplia, por lo tanto se divide el espacio de investigación en dos dimensiones: espacio de investigación (*Research Space*) y espacio de publicación (*Publication Space*). El espacio de investigación del tema tiende a ser variable, esto quiere decir que depende en gran medida del objetivo principal y la pregunta de investigación principal, y por consecuencia, las preguntas de investigación (*Research Questions*) están sujetas a cambios en cada revisión sistemática. Por otra parte, las preguntas dentro del espacio de publicación (*Publication Questions*) son fijas de antemano debido a que son muy similares en muchas revisiones sistemáticas.

A continuación, se explican las dos dimensiones, para la investigación que se reporta en esta tesis.

4.1.1 Espacio de publicación

Para nuestra investigación, se definieron un conjunto de preguntas de publicación que permiten clasificar la literatura con respecto al espacio de publicación, para proveer un panorama del estado del arte, respecto al área de investigación.

- PQ1.- ¿Qué tipo de publicación es? (Conferencia/Revista)
- PQ2.- ¿Cuál es la cantidad de artículos o conferencias publicados por año?
- PQ3.- ¿Cuál es la cantidad de artículos o conferencias por país de procedencia? (De acuerdo a la institución de procedencia del autor)
- PQ4.- ¿Cuál es el ámbito del autor? (Académico/Industrial)
- PQ5.- ¿Cuáles son los congresos o revistas en los que se publicaron los artículos?

4.1.2 Espacio de investigación

Dentro de esta dimensión, para nuestra investigación, se definieron las preguntas enfocándonos en el tema de investigación. Estas preguntas se derivaron de la pregunta de investigación principal. A continuación, se muestran las preguntas definidas para este estudio y en la Tabla 4-1 se describen cada una:

- RQ1.- ¿Qué tema se trata en la elicitación de requerimientos y/o dominio del problema?
- RQ2.- ¿Qué aporte se presenta en el artículo/conferencia? (Contribución).
- RQ3.- ¿Qué tipo de entrada se procesa en el trabajo reportado?
- RQ4.- ¿Qué tipo de salidas o resultados presenta el trabajo reportado?
- RQ5.- ¿Qué tipo de validación se utilizó en el trabajo reportado?
- RQ6.- ¿Qué tipo de investigación se publicó?
- RQ7.- ¿Cuál es el nivel de automatización en el trabajo reportado?
- RQ8.- ¿Qué técnicas de procesamiento de lenguaje natural y ontologías se presentan en la elicitación de requerimientos y/o dominio del problema?

Tabla 4-1. Descripción de preguntas de investigación.

Pregunta de Investigación	Descripción
RQ1.- ¿Qué tema se trata en la elicitación de requerimientos y/o dominio del problema?	Se refiere al tema que se trata. Para esta investigación los posibles son: <ul style="list-style-type: none"> • Reconocimiento del habla • Procesamiento de Lenguaje Natural • Ontología

	<ul style="list-style-type: none"> • Enfoques híbridos (combinación de los anteriores)
RQ2.- ¿Qué aporte se presenta en el artículo/conferencia? (Contribución).	<p>Se refiere al tipo de contribución de la investigación, los tipos posibles son:</p> <ul style="list-style-type: none"> • Método/Procedimiento/Técnica • Herramienta • Marco de trabajo(<i>Framework</i>)
RQ3.- ¿Qué tipo de entrada se procesa en el trabajo reportado?	<p>Tipo de entrada (verbal, escrita o modelo) que requiere el enfoque para su correspondiente procesamiento. Para esta investigación se calificaron en:</p> <ul style="list-style-type: none"> • Diagramas • Documento con lenguaje natural controlado • Documentos en lenguaje natural • Ontologías • Voz • No aplica
RQ4.- ¿Qué tipo de salidas o resultados presenta el trabajo reportado?	<p>Se refiere a la finalidad del enfoque propuesto, la clasificación fue con base a la propuesta en la investigación de (Meth et al., 2013). A dicha clasificación se agregó la categoría “Clasificación de requerimientos” y “Otro”, como sigue:</p> <ul style="list-style-type: none"> • Análisis de calidad • Clasificación de requerimientos • Generación de modelos • Identificación de abstracción • Identificación de requerimientos • Otro
RQ5.- ¿Qué tipo de validación se utilizó en el trabajo reportado?	<p>Se refiere al tipo de validación del trabajo de investigación, para ello se utilizó como base la clasificación definida por (Meth et al., 2013). Se agregó la categoría “Otro” y “No aplica”:</p> <ul style="list-style-type: none"> • Caso de estudio • Experimento controlado • Prueba de concepto

	<ul style="list-style-type: none"> • Simulación • Otro (No entra en las categorías anteriores) • No aplica(No se reporta evaluación en el artículo)
<p>RQ6.- ¿Qué tipo de investigación se publicó?</p>	<p>Se utilizó la clasificación propuesta por (Wieringa, Maiden, Mead, & Rolland, 2006) donde describe vario tipos de investigación, entre los cuales son los siguientes, y se agregó la categoría “Otro” para aquellas investigaciones que no pertenecían a ninguna categoría anterior:</p> <ul style="list-style-type: none"> • Investigación de evaluación • Propuesta de solución • Investigación de validación • Artículo filosófico • Otro(No entra en las clasificaciones anteriores)
<p>RQ7.- ¿Cuál es el nivel de automatización en el trabajo reportado?</p>	<p>Se refiere al nivel de automatización que presenta el enfoque, para esto se clasificó en tres:</p> <ul style="list-style-type: none"> • Automático • Semi-automático • Manual
<p>RQ8.-¿Qué técnicas de reconocimiento del habla, procesamiento de lenguaje natural y/o ontologías se presentan?</p>	<p>El objetivo de esta pregunta es identificar las distintas técnicas presentes en las investigaciones. Debido a la gran variedad de técnicas dichas técnicas se clasificaron de acuerdo al flujo de trabajo del NLP mencionado en (Sarkar, 2018). Para abarcar el objetivo de investigación definido, se completó esta clasificación agregando “Reconocimiento del habla”:</p> <ul style="list-style-type: none"> • Reconocimiento del habla • Pre-procesamiento de texto • Análisis de texto y análisis de datos exploratorios • Representación de texto e ingeniería de características • Modelado y extracción de patrones

	<ul style="list-style-type: none"> • Evaluación y despliegue <p>Respecto a la clasificación para enfoques ontológicos se utilizó la clasificación mencionada en (Farfeleder, Moser, Krall, Ståalthane, et al., 2011), que muestra las fases de la ingeniería de requerimientos en que se utilizan las ontologías. A esta clasificación se agregó la fase de “Modelado del dominio del problema”:</p> <ul style="list-style-type: none"> • Modelado del dominio del problema • Elicitación de Requerimientos • Análisis y Negociación de Requerimientos • Documentación y Validación de Requerimientos • Administración de Requerimientos <p>Por otra parte también se clasificaron los distintos tipo de ontologías propuesto de acuerdo a (Castañeda, Ballejos, Caliusco, & Galli, 2010):</p> <ul style="list-style-type: none"> • Ontología de requerimientos; • Ontología de documento de especificación de requerimientos • Ontología del dominio de aplicación
--	--

4.2 Búsqueda

En esta sección se presentan las estrategias a seguir para realizar el proceso de búsqueda de trabajos publicados en revistas, dado que el objetivo del estudio es identificar la evidencia existente.

Una de las primeras actividades a realizar es la selecciona de la base de datos, y estas fueron seleccionadas de acuerdo al área de investigación, una vez seleccionadas dichas bases de datos, fue necesario crear una cadena de búsqueda derivada del objetivo de este trabajo de tesis, para posteriormente ejecutar dicha cadena en cada una de las bases de datos y obtener los trabajos relacionados.

4.2.1 Selección de base de datos científica

Se seleccionaron las bases de datos de acuerdo al contexto de la investigación, en este caso como se buscan reportes de investigaciones sobre el uso de técnicas de reconocimiento del habla, procesamiento de lenguaje natural y ontologías en la elicitación de requerimientos y/o

dominio del problema. Dentro del área de ingeniería de requisitos y/o dominio del problema, que pertenece al proceso de desarrollo de software o ingeniería de software, que a su vez se incluyen en contexto de las ciencias de la computación. Por lo tanto, se seleccionaron las siguientes bases de datos que almacenan trabajos de investigación en el área de la computación:

- IEEE Xplore Digital Library
- ScienceDirect (Elsevier)
- ACM Digital Library
- SpringerLink
- Redalyc (cuatro revistas)

Cabe recalcar que en la base de Redalyc se buscaron artículo sólo en cuatro revistas mexicanas (“Computación y Sistemas”, “Ingeniería Investigación y Tecnología”, “Journal Of Applied Research And Technology” y “Polibits”), las cuales pertenecen al área de computación y se encuentran dentro del “Sistema de Clasificación de Revistas Mexicanas de Ciencia y Tecnología” (revistas reconocidas por el Consejo Nacional de Ciencia y Tecnología, CONACYT).

4.2.2 Creación de cadena de búsqueda

4.2.2.1 Búsquedas avanzadas en bases de datos científicas

Las bases de datos científicas como ScienceDirect, ACM Digita Library, IEEE Xplore y Springlink brindan la opción de realizar búsquedas avanzadas que permiten especificar más la información para encontrar los artículos de investigación relacionados. Estas búsquedas avanzadas se realizan por medio del uso de cadenas de búsqueda que utilizan comandos de búsqueda (operadores booleanos y cláusulas anidadas mediante paréntesis).

Los comandos de búsqueda varían dependiendo la base de datos científica. Por esta razón, antes de realizar una consulta con una cadena de búsqueda verifique los comandos permitidos. A continuación, se presenta los comandos de búsqueda más utilizados y permitidos en los repositorios antes mencionados:

Operadores booleanos: Los principales operadores booleanos utilizados son AND, OR y NOT. Deberán utilizarse en mayúsculas. A excepción del repositorio ScienceDirect, este no utiliza la palabra NOT, en su lugar usa el símbolo “-“(por ejemplo, “*use case*” -*diagrams*). Se permite el uso de paréntesis para anidar cláusulas y agrupar los términos para ser más claro. Por ejemplo:

user OR elicitation AND requirements

es equivalente a

(user OR elicitation) AND requirements

Comillas: Las comillas son utilizadas para especificar palabras compuestas (términos que aparecen uno después del otro). Por ejemplo:

("heart attack" OR "myocardial infarction") AND diabetes AND NOT cancer

4.2.2.2 Definición de cadena de búsqueda

La cadena de búsqueda se utiliza para identificar los trabajos de investigación para el estudio y debe de ingresarse en cada una de las bases de datos. La cadena debe reflejar el objetivo del estudio, es decir, se deriva del objetivo principal de la investigación (ver sección 4.1), por lo tanto, ciertas palabras clave se encuentran en el texto del objetivo. Dado que buscamos artículos publicados en su mayoría en inglés, se realizó la traducción correspondiente.

Del objetivo antes descrito, se extrajeron las siguientes palabras clave:

"Speech Recognition" "Natural Language" "Ontology", "Requirement Elicitation" y "Problem Domain".

Obtenidas estas palabras claves se agregaron otras que permiten refinar la cadena de búsqueda, como fue la palabra clave "Requirement Engineering" en lugar de "Software Engineering", debido a que esta segunda es un término más general que abarca no solo la etapa de análisis (en el contexto de desarrollo de software). Además, se agregó la palabra clave "Voice Recognition" para completar mejor la búsqueda. La cadena de búsqueda se definió de la siguiente manera:

("Requirements Engineering" OR "Requirements Elicitation" OR "Problem Domain")

AND

("Natural Language" OR Ontology OR Ontologies OR "Speech Recognition" OR "Voice Recognition")

4.2.3 Ejecución de cadena de búsqueda

La cadena de búsqueda se utilizó en cada una de las bases de datos científicas antes mencionadas los resultados se pueden observar en la Tabla 4-2, donde se observa el número de artículos obtenidos de cada base de dato. Se muestra también la fecha de ejecución de la cadena de búsqueda correspondiente, debido a que con el transcurso del tiempo algunas bases de datos agregaban más artículos.

Tabla 4-2. Resultados de las diferentes bases de datos científicas al ejecutar la cadena de búsqueda.

Base de datos	Resultados	Fecha Consulta
ACM Digital Library	138	01/10/2018
IEEE Explorer	536	24/10/2018
ScienciDirect	81	26/09/2018
SpringerLink	8330	10/11/2018
Redalyc	6	11/11/2018
Total	9783	

4.3 Revisión de artículos

La selección de artículos de investigación se llevó a cabo mediante la aplicación de los criterios de inclusión y exclusión. Estos criterios se aplicaron al conjunto de artículos resultante de las bases de datos científicas, con el propósito de excluir estudios que no son relevantes y responder las preguntas de investigación definidas anteriormente. Las aplicaciones de los criterios de inclusión y exclusión se aplicaron a títulos, resúmenes, y conclusiones. Los criterios para excluir literatura, fueron los siguientes:

- Son estudios que se encuentran fuera del contexto de la ingeniería de software.
- Libros, presentaciones, reportes técnicos, tutoriales o algún otro tipo de literatura gris (solo se incluirán artículos de revistas y conferencias).
- Estados del arte o artículos de opinión.
- Estudios que no se encuentren disponibles en texto completo.
- Estudios duplicados en las distintas bases de datos.

Los criterios de inclusión se presentan a continuación:

- Son estudios que se centran en la elicitación de requerimientos y/o dominio del problema.
- Son estudios que se enfocan en la aplicación de procesamiento de reconocimiento del habla, lenguaje natural u ontologías.

Como último paso, sólo se seleccionaron los artículos de investigación a partir del 2010, para realizar solo la lectura completa de los artículos seleccionados, después de aplicar dichos criterios. De esta manera se obtuvieron 64 artículos como selección final.

El resumen del proceso de selección de artículos se presenta en la Figura 4.1, donde después de utilizar la cadena de búsqueda, en las bases de datos, se obtuvieron 9783 resultados. A estos se aplicaron los criterios de inclusión y exclusión, reduciéndose la selección a 1247. Después se utilizaron los criterios de inclusión obteniendo 138 artículos, luego se eliminaron aquellos publicados antes del 2010 quedando 81. Por último, se realizó la lectura completa de los artículos, dado que el análisis de los artículos a fondo permite correctamente seleccionar el artículo (si se enfoca en la elicitación de requerimientos y modelado del dominio). Mediante la lectura completa se llegó a una selección final de 64 artículos de investigación para este estudio de mapeo sistemático. La lista de los artículos se encuentra indexada mediante un identificador, por ejemplo ART01, y se incluyen en la Tabla A-1, en el anexo A (Lista de artículos seleccionados para este estudio de mapeo sistemático).

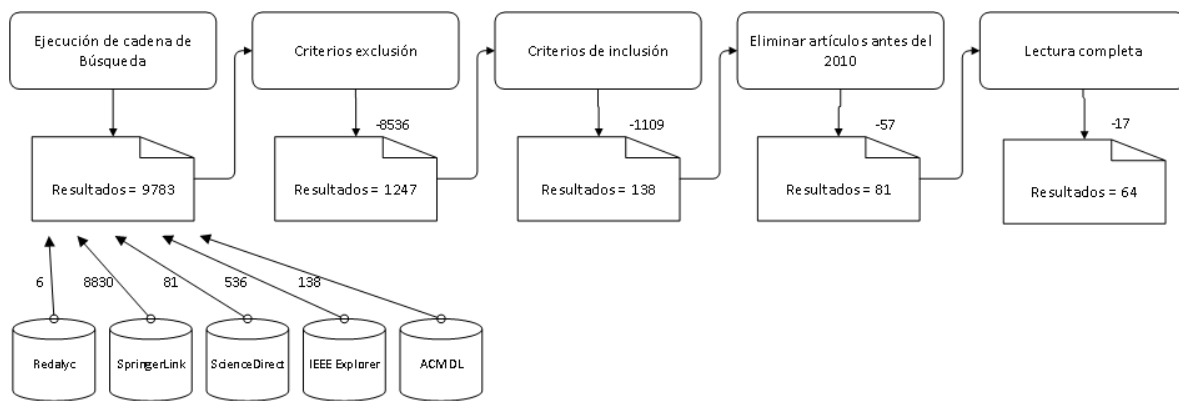


Figura 4.1. Proceso de selección de artículos de investigación.

4.4 Identificación de palabras clave

Una de las principales actividades de este estudio es la extracción de datos, para ello fue necesario crear un formato para la captura de información. Dicho formato lleva la estructura que se presenta en la Tabla 4-3. Se debe recalcar que esta clasificación fue realiza después del análisis de lectura de los artículos y respecto a otros estudios sistemáticos (mencionados en el apartado 2.2 Trabajos relacionados).

Tabla 4-3. Plantilla para captura clasificación de artículos.

Dato	Descripción	Pregunta
Id	Número de identificación	
Título	Título del artículo	
Documento	Tipo de artículo (conferencia o revista).	PQ1
Año	Año de publicación del artículo.	PQ2

País	País de origen	PQ3
Ámbito	Se refiere si pertenece al ámbito industrial o académico.	PQ4
Revista/Congreso	Nombre de la revista o congreso de publicación.	PQ5
Elemento	Tipo de información que contiene el artículo (NLP, reconocimiento del habla, Ontología o Híbrido).	RQ1
Contribución	Tipo de contribución del artículo (Método, Herramienta o Framework)	RQ2
Entrada	Tipo de entrada que recibe el enfoque (Documento CNL, Documento en NL, diagrama, voz, ontología u otro).	RQ3
Salida/Resultado	Tipo de salida o resultado que genera el enfoque (análisis de calidad, clasificación de requerimientos, generación de modelos, identificación de abstracción, identificación de requerimientos u otro).	RQ4
Evaluación	Tipo de evaluación aplicado (caso de estudio, experimento controlado, simulación, prueba de concepto, no aplica u otro)	RQ5
Tipo de Investigación	Tipo de investigación (investigación de evaluación, propuesta de solución, investigación de validación, artículo filosófico u otro).	RQ6
Nivel de Automatización	Grado de automatización del enfoque propuesto (automática, semi-automática, manual)	RQ7
Técnicas/Enfoques	Las técnicas de reconocimiento del habla, procesamiento de lenguaje natural y tipos de ontologías.	RQ8

Mediante este formato se obtiene un panorama acerca de los avances de las investigaciones, dado que este es el propósito de los SMS, en el área de conocimiento definida para este estudio, como son las técnicas NLP, enfoques ontológicos utilizados, tipos de evaluación, niveles de automatización, tipo de entradas que procesan y tipos de salidas de

los enfoques. En las Tabla 4-4 y Tabla 4-5 se presentan un ejemplo del artículo ART01 con datos de las preguntas de publicación e investigación.

Tabla 4-4. Artículo con datos de preguntas de publicación.

Id	RQ1	RQ2	RQ3	RQ4	RQ5
ART01	Revista	2018	Jordania	Académico	Information and Software Technology

Tabla 4-5. Artículos con datos de preguntas de investigación

Id	RQ1	RQ2	RQ3	RQ4	RQ5	RQ6	RQ7	RQ8
ART01	NLP	Método / Procedimiento/ Técnica	Doc. en NL	Identificación de abstracción	Caso de Estudio	Investigación de evaluación	Semi-automática	Análisis de texto y análisis de datos exploratorios; Modelado y extracción de patrones; Evaluación y Despliegue

El formato es un cuadro comparativo de los artículos recuperados. En el Tabla A-2 se presenta la clasificación de todos los artículos de acuerdo a las preguntas de publicación y en la Tabla A-3 se presentan todos los artículos clasificados de acuerdo a las preguntas de investigación.

4.5 Análisis de estos artículos y creación del mapa resultante

Una vez llenado la información de los artículos seleccionados en el formato, se procedió a realizar la elaboración de los gráficos visuales que permiten analizar la información recopilada en el formato. Estos gráficos fueron elaborados mediante la herramienta Excel de Microsoft Office. El resultado del SMS se presenta detallado en el Capítulo 5.

4.6 Resumen del Capítulo

En este capítulo se presentó el proceso sistemático realizado para la búsqueda y selección de los artículos de investigación que representan la evidencia existente de uso de reconocimiento

del habla, procesamiento del lenguaje natural y ontologías en la identificación del dominio del problema y requerimientos de solución de la etapa de análisis del desarrollo de software. En el siguiente capítulo se presenta el resultado de este estudio (mapa sistemático) donde se muestra el panorama general acerca de los avances en esta área de investigación.

Capítulo 5.

Resultados del

estudio

En este capítulo se presenta un resumen visual (mapa sistemático) acerca del uso de técnicas de reconocimiento del habla, procesamiento de lenguaje natural y enfoques ontológicos en la elicitación de requerimientos y/o modelado del dominio del problema. Mediante este resumen se presenta el panorama general acerca de la evidencia existente respecto a esta área de investigación. Esta sección se divide en dos subsecciones: la primera subsección presenta los resultados de los datos relacionados a las preguntas de publicación y la segunda presenta los resultados de los datos relacionados a las preguntas de investigación.

A continuación, se describen los hallazgos relacionados con las preguntas de publicación:

5.1 Preguntas de publicación

5.1.1 PQ1.- ¿Qué tipo de publicación es? (Conferencia/Revista)

En este estudio se identificaron 19 artículos en revista y 45 artículos en conferencia, con un total de 64 artículos. Más del 70% de artículos acerca del uso de modelos de lenguaje natural y ontologías se publican en conferencias, véase la Figura 5.1. Para complementar la información, en la Tabla 5-1 que se presenta la relación de artículos de acuerdo al tipo de publicación. La mayoría de las publicaciones en conferencias proviene de la base de datos de IEEE Xplorer.

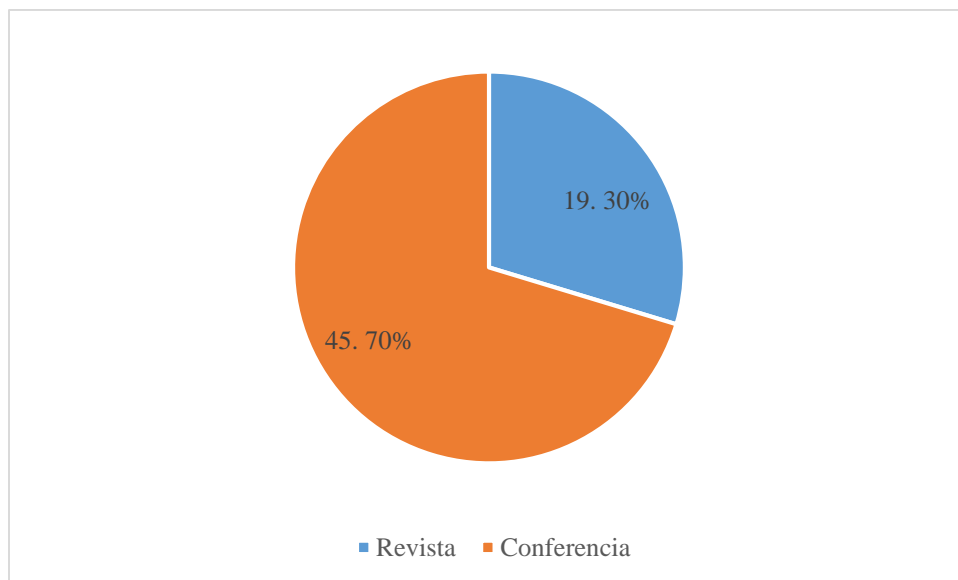


Figura 5.1. Porcentaje de artículos de investigación de acuerdo a su tipo.

Tabla 5-1. Referencia de artículos de acuerdo al tipo de publicación.

Tipo de Artículo	Artículos	Número de artículos
Conferencia	ART04, ART06, ART07, ART08, ART09, ART10, ART11, ART12, ART13, ART14, ART15, ART16, ART17, ART18, ART19, ART21, ART22, ART23, ART24, ART25, ART26, ART27, ART32, ART33, ART34, ART35, ART38, ART41, ART42, ART44, ART45, ART46, ART49, ART50, ART51, ART52, ART56, ART57, ART58, ART59, ART60, ART61, ART62, ART63, ART64	45
Revista	ART01, ART02, ART03, ART05, ART20, ART28, ART29, ART30, ART31, ART36, ART37, ART39, ART40, ART43, ART47, ART48, ART53, ART54, ART55	19

5.1.2 PQ2.- ¿Cuál es la cantidad de artículos o conferencias publicados por año?

Dado nuestro conjunto seleccionado de 64 artículos, se observó que hay un número de trabajos relacionados publicados por año, véase la Figura 5.2. Se puede notar que la publicación del 2010 al 2013 fue decreciendo, pero del año 2014 al 2018 se incrementó la generación de trabajos. De este hecho, se deduce que está tomando más importancia, dado que en el 2018 hay 12 trabajos al respecto. La referencia de los artículos se puede observar en la Tabla 5-2.

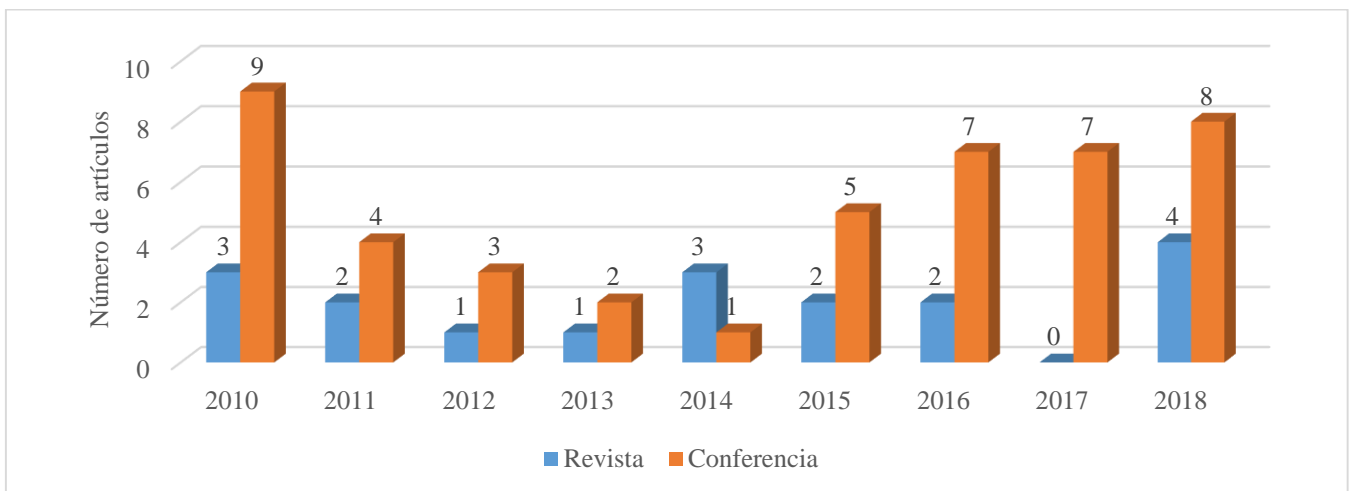
**Figura 5.2. Artículos por año.**

Tabla 5-2. Referencia de artículos por año.

Año de publicación	Tipo de artículo	Artículo	Número de artículo
2010	Revista	ART53, ART54, ART55	3
	Conferencia	ART56, ART57, ART58, ART59, ART60, ART61, ART62, ART63, ART64	9
2011	Revista	ART47, ART48	2
	Conferencia	ART49, ART50, ART51, ART52	4
2012	Revista	ART43	1
	Conferencia	ART44, ART45, ART46	3
2013	Revista	ART40	1
	Conferencia	ART41, ART42	2
2014	Revista	ART36, ART37, ART39	3
	Conferencia	ART38	1
2015	Revista	ART30, ART31	2
	Conferencia	ART29, ART32, ART33, ART34, ART35	5
2016	Revista	ART20, ART28	2
	Conferencia	ART21, ART22, ART23, ART24, ART25, ART26, ART27	7
2017	Revista	-----	0
	Conferencia	ART13, ART14, ART15, ART16, ART17, ART18, ART19	7
2018	Revista	ART01, ART02, ART03, ART05	4
	Conferencia	ART04, ART06, ART07, ART08, ART09, ART10, ART11, ART12	8

5.1.3 PQ3.- ¿Cuál es la cantidad de artículos o conferencias por país de procedencia? (De acuerdo a la institución de procedencia del autor)

Con base al conjunto de artículos seleccionados, se obtuvieron aportes de distintos países, para ser exactos de 33 países diferentes. Sin embargo, los países con más cantidad de trabajos al respecto son: con seis artículos Alemania, Canadá, China e India, que sumados aportan 24 artículos (37.50% de los 64 artículos seleccionados); con tres artículos Estados Unidos y Reino Unido, que sumados aportan 6 artículos (9.37% de los 64 artículos seleccionados); y con dos artículos Australia, Corea, España, Japón, Jordania, Luxemburgo y Nigeria, que sumados aportan 14 artículos (21.88% de los 64 artículos seleccionados) y de países con un artículo publicado, con 20 artículos (31.25% de los 64 artículos seleccionados). En la Figura 5.3 se observa que los países con seis artículos, en conjunto representa la mayor parte de contribución con un 37.50% de los trabajos. En la Tabla 5-3 se observa la referencia de artículos por país con mayor contribución.

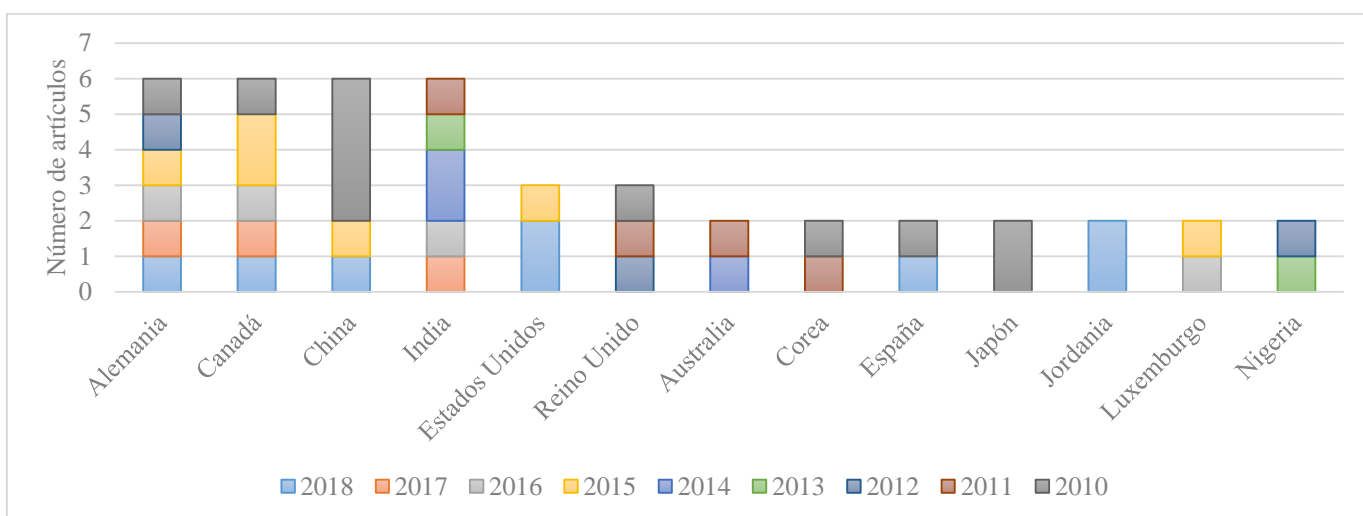


Figura 5.3. Países con más aportes de trabajos relacionados.

Tabla 5-3. Referencia de artículos con mayor contribución por país.

País	Artículos	Número de artículos
Alemania	ART11, ART16, ART25, ART29, ART43, ART64	6
Canadá	ART08, ART18, ART22, ART33, ART35, ART63	6
China	ART03, ART32, ART55, ART56, ART58, ART61	6
India	ART17, ART27, ART36, ART37, ART40, ART48	6

Estados Unidos	ART09, ART12, ART30	3
Reino Unido	ART46, ART50, ART54	3
Australia	ART38, ART51	2
Corea	ART52, ART59	2
España	ART05, ART53	2
Japón	ART57, ART60	2
Jordania	ART01, ART07	2
Luxemburgo	ART20, ART34	2
Nigeria	ART42, ART45	2

5.1.4 PQ4.- ¿Cuál es el ámbito del autor? (Académico/Industrial)

El objetivo de esta pregunta es identificar si los autores de los artículos de investigación pertenecen al sector académico o industria, de acuerdo al instituto o empresa al que pertenece. Para contestar dicha pregunta se consideró al primer autor de cada artículo. Solo 3 trabajos (el 4.69%) pertenecen al ámbito industria mientras que el resto pertenece al ámbito académico (95.31%). En la Tabla 5-4 se muestra las referencias de artículos respecto al ámbito del autor.

Tabla 5-4. Referencia de artículos de acuerdo al ámbito del autor.

Ámbito	Artículos	Número de artículos
Académico	ART01, ART03, ART04, ART05, ART06, ART07, ART08, ART09, ART10, ART11, ART12, ART14, ART15, ART16, ART17, ART18, ART19, ART20, ART21, ART22, ART23, ART24, ART25, ART26, ART27, ART28, ART29, ART30, ART31, ART32, ART33, ART34, ART35, ART36, ART37, ART38, ART40, ART41, ART42, ART43, ART44, ART45, ART46, ART47, ART48, ART49, ART50, ART51, ART52, ART53, ART54, ART55, ART56, ART57, ART58, ART59, ART60, ART61, ART62, ART63, ART64	61
Industrial	ART02, ART13, ART39	3

5.1.5 PQ5.- ¿Cuáles son los congresos o revistas donde se publicaron los artículos?

En este SMS se identificaron 32 congresos y 16 revistas, en la Tabla 5-5 se muestran los cinco congresos con más artículos publicados. (el congreso con el número mayor de publicaciones es “*IEEE International Requirements Engineering Conference (RE)*”). Por otro parte, las tres revistas con más publicaciones de artículos (con dos artículos publicados cada una: “*Information and Software Technology*”, “*Procedia Computer Science*”, “*Software and Systems Modelling*”).

Tabla 5-5. Congresos con más aportes de artículos.

Congresos	Artículos	Número de artículos
IEEE International Requirements Engineering Conference (RE)	ART11, ART12, ART15, ART16, ART17, ART18, ART25, ART26, ART35, ART38, ART46, ART58	12
International Workshop on Artificial Intelligence for Requirements Engineering (AIRE)	ART08, ART09, ART10	3
ACM SIGSOFT Software Engineering Notes	ART40, ART48	2
Asia Pacific Software Engineering Conference	ART13, ART57	2
IEEE International Conference on Computer and Information Technology (CIT)	ART23, ART59	2

5.2 Preguntas de investigación

La elicitación de requerimiento es el proceso de entender el dominio del problema e identificar las necesidades mediante la comunicación con los clientes para, posteriormente proponer una solución, que se describe mediante la especificación de requerimientos. Sin embargo, este proceso tiende a complicarse debido a que se realiza en lenguaje natural. Esto causa problemas de ambigüedad y captura de requerimientos incompletos, además de consumir tiempo del analista.

Uno de los elementos utilizados en el proceso de modelado del dominio y elicitación de requerimientos, es la automatización mediante técnicas de procesamiento de lenguaje natural y de otro tipo, como las que usan ontologías. En la literatura encontramos varias características comunes entre los artículos que tratan estos temas, en relación con las técnicas utilizadas. Utilizando estos hallazgos (las respuestas a las preguntas de investigación) se definió un esquema de clasificación de los artículos. En la Figura 5.4 se presenta dicho esquema general.

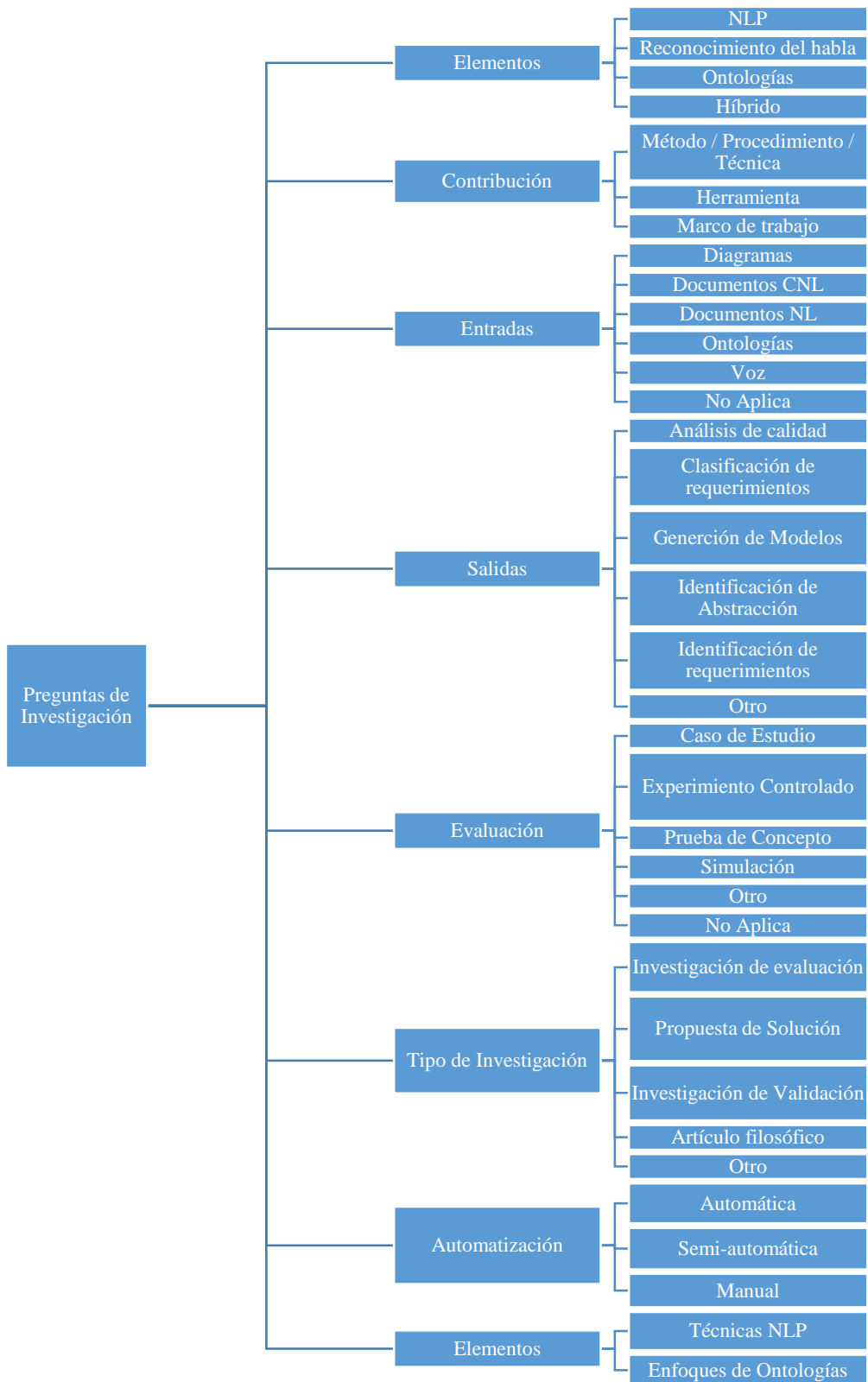


Figura 5.4. Esquema de clasificación dado el análisis de la literatura.

En la Figura 5.4 se presenta el esquema general de acuerdo a las preguntas de investigación anteriormente mencionadas, donde RQ1(Enfoque) se presentan los tipos de temas presentes en los artículos(NLP, reconocimiento del habla, ontologías e híbridos), RQ2(Contribución) se muestran los artículos clasificados de acuerdo al tipo de aportación, RQ3(Entradas) se clasifican los artículos de acuerdo al formato de entrada que procesa el enfoque, RQ4(Salida) es la clasificación con base a las salidas que generan los enfoques, RQ5(Evaluación) se refiere al tipo de validación utilizado, RQ6(Tipo de investigación) clasifica los artículos de acuerdo al tipo de investigación, RQ7(Automatización) se refiere al nivel de automatización del trabajo de investigación y RQ8(Elemento) se presentan las diferentes técnicas de NLP y tipos de ontologías.

A continuación, se describen los hallazgos relacionados con las preguntas de investigación:

5.2.1 RQ1.- ¿Qué tema se trata en la elicitación de requerimientos y/o dominio del problema?

Esta pregunta de investigación se centra en identificar qué tipo de enfoques (NLP, reconocimiento del habla, ontologías, híbridos) se manejan dentro de los trabajos recuperados. Aunque el procesamiento de lenguaje natural engloba texto y voz, para este estudio se considera NLP a los artículos que procesan información textual y reconocimiento del habla a los que procesan la voz. Por otro lado, se identificaron artículos con enfoques ontológicos en la elicitación de requerimientos y modelado del dominio del problema, es decir, no utilizan técnicas NLP y reconocimiento del habla. Aquellos artículos que combinen NLP y ontologías se consideran como enfoques híbridos (en este estudio sólo se identificaron dos artículos que combinan NLP y ontologías). En la Figura 5.5 se muestra la cantidad de artículos por cada tópico.

Se recuperaron dos artículos del reconocimiento del habla, 30 acerca del NLP, 19 acerca de ontologías y 13 que usan NLP y ontologías para llevar a cabo la elicitación de requerimientos y/o dominio del problema. Se determinó que los trabajos que presentan uso tanto de NLP y ontologías(híbridos), son trabajos con enfoque semi-automático en aplicaciones en la industria. En la Figura 5.6 se muestra la tendencia de mayor investigación del uso del NLP, que pasa de 2 publicaciones en 2014 a 10 publicaciones en 2018. Además, se observa que son recientes los trabajos relacionados al reconocimiento del habla en el contexto de la elicitación de requerimientos (2015 y 2016). Por último, en la Tabla 5-6 se presenta la referencia de artículos de acuerdo al enfoque utilizado en dicho trabajo de investigación.

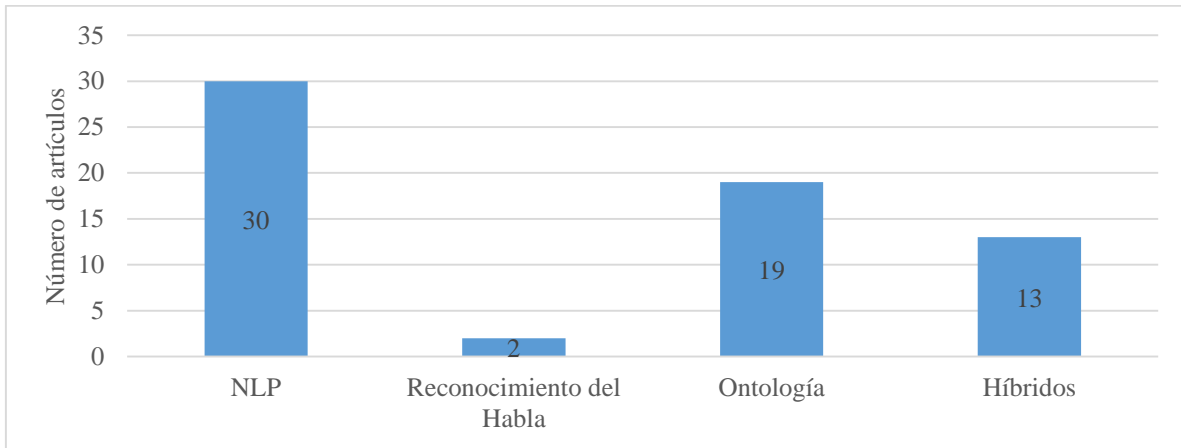


Figura 5.5. Distribución de artículos respecto a al tipo de enfoque.

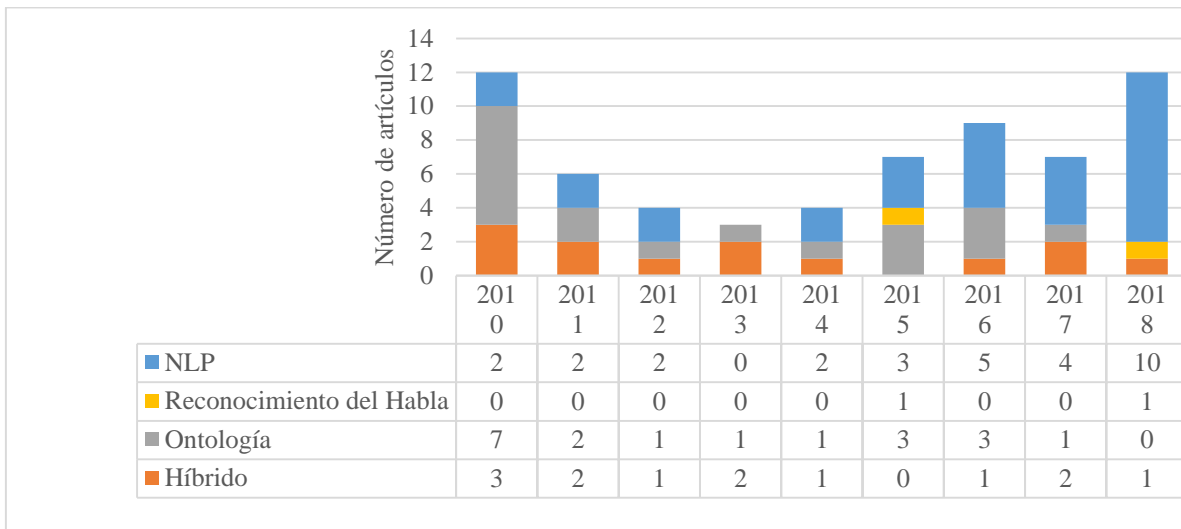


Figura 5.6. Tipos de artículos por año.

Tabla 5-6. Referencia de artículos de acuerdo al tipo de enfoque.

Elemento que contiene el artículo	Artículos	Número de artículos
NLP	ART01, ART02, ART03, ART04, ART05, ART06, ART07, ART09, ART10, ART11, ART15, ART17, ART18, ART19, ART20, ART24, ART25, ART26, ART28, ART29, ART34, ART35, ART36, ART39, ART43, ART46, ART47, ART48, ART63, ART64	30

Reconocimiento del Habla	ART08, ART31	2
Ontologías	ART13, ART21, ART23, ART27, ART30, ART32, ART33, ART38, ART41, ART44, ART49, ART50, ART53, ART54, ART55, ART56, ART58, ART59, ART60	19
Híbrido	ART12, ART14, ART16, ART22, ART37, ART40, ART42, ART45, ART51, ART52, ART57, ART61, ART62	13

5.2.2 RQ2.- ¿Qué aporte se presenta en el artículo/conferencia? (Contribución)

El objetivo de esta pregunta de investigación es saber el tipo de resultado que se presenta en los artículos recopilados. Los valores recopilados se clasificaron en 3 tipos de contribución, que son:

1. Método/Procedimiento/Técnica
2. Herramienta
3. Framework

En la Figura 5.7 se presenta la distribución de artículos que pertenece a cada categoría. Las investigaciones con aporte de Métodos/Procedimiento/Técnica representan el 65.625 %, el 25% contribuyen con una herramienta y el 9.375% propone marcos de trabajo. En la Tabla 5-7 se presentan las referencias de artículos de los tres tipos de contribución.

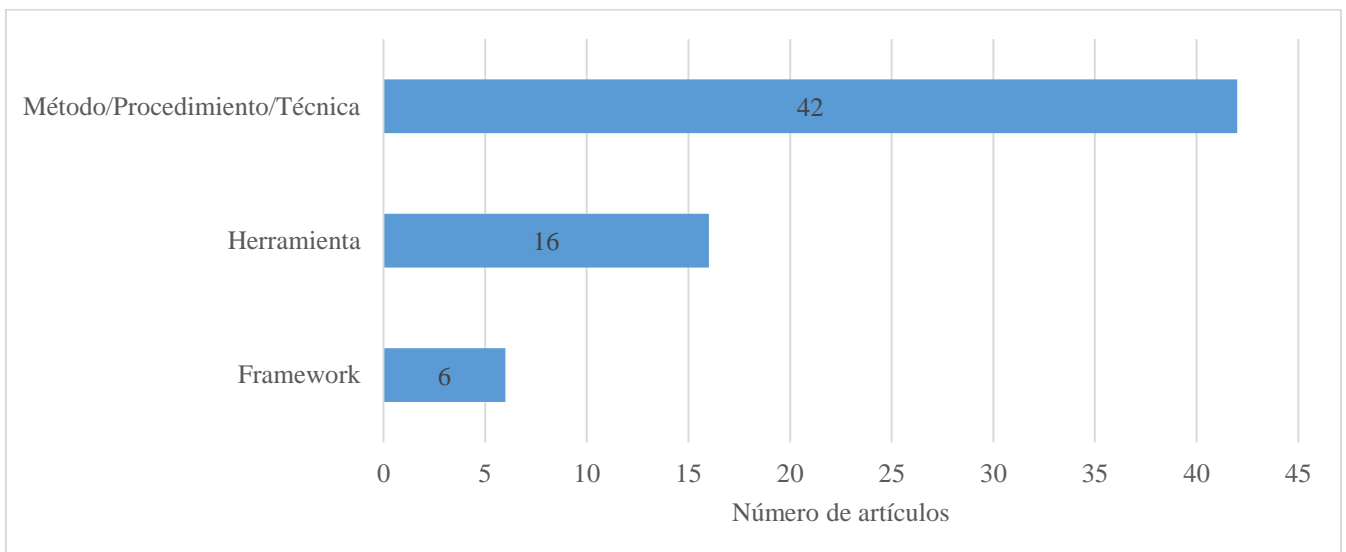


Figura 5.7. Tipos de contribuciones en los artículos.

Tabla 5-7. Referencia de artículos de acuerdo al tipo de contribución.

Contribución	Artículos	Número de artículos
Método/ Procedimiento/ Técnica	ART01, ART02, ART03, ART04, ART07, ART09, ART10, ART11, ART13, ART15, ART16, ART20, ART21, ART22, ART23, ART24, ART25, ART28, ART29, ART30, ART32, ART33, ART34, ART36, ART37, ART39, ART40, ART41, ART44, ART46, ART47, ART49, ART51, ART52, ART54, ART55, ART56, ART57, ART61, ART62, ART63, ART64	42
Herramienta	ART05, ART06, ART08, ART12, ART17, ART18, ART19, ART26, ART35, ART38, ART43, ART45, ART48, ART58, ART59, ART60	16
Framework	ART14, ART27, ART31, ART42, ART50, ART53	6

5.2.3 RQ3.- ¿Qué tipo de entrada se procesa en el trabajo reportado?

En este SMS se identificaron varios tipos de entrada, por ejemplo: documento de requerimientos, descripciones de casos de uso, reportes de entrevistas, documentos del dominio del problema, ontologías, diagramas de procesos de negocio, gráfico de metas, reglas de negocio, entre otros. Por lo tanto, fue necesario crear una clasificación entre todos estos tipos de entrada. En la Tabla 5-8 se describen los tipos de entrada.

Tabla 5-8. Clasificación de tipos de entrada identificados.

Entradas de los enfoques	Descripción
Documentos en lenguaje natural (<i>Natural Language</i> , NL)	Son aquellos enfoques que procesan información textual, procesan el lenguaje natural sin ninguna restricción en el mismo.
Documentos con lenguaje natural controlado (<i>Controlled Natural Language</i> , CNL)	Los enfoques que procesan información textual, pero el lenguaje natural está restringido o controlado, es decir, se tiene una serie de reglas y estructura para escribir dicha información (por ejemplo, lenguaje Gherkin y plantillas Restricted User Case Modeling). La gramática y el vocabulario se restringen para eliminar o reducir la ambigüedad y complejidad.

Voz	Se refiere a aquellos artículos que realizan el reconocimiento de voz y producen una representación apropiada (texto), para su procesamiento posterior.
Diagramas	Se refiere a aquellos enfoques que reciben como entrada algún diagrama o modelo y producen una representación apropiada (otro modelo como casos de uso, proceso de negocio, actividades u ontologías), para su procesamiento.
Ontología	Se recibe ontologías, como entrada, para poder procesarlos.
No Aplica	Se engloban métodos o procedimientos para llevar a cabo la elicitación de requerimientos, que no procesan tal cual una entrada. Presentan algún tipo de directriz para llevar dicha actividad (enfoque manual), por ejemplo, una guía para la construcción de ontología para representar el dominio del problema, o un procedimiento para captura de requerimientos con el apoyo de una ontología ya definida.

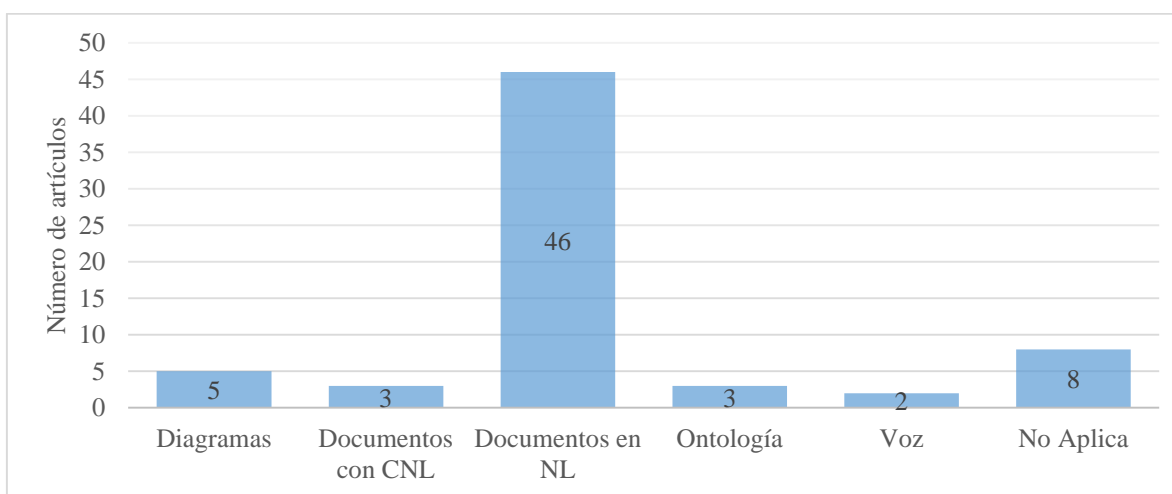


Figura 5.8. Tipos de entrada.

En la Figura 5.8 se identifican 46 artículos que procesan documentos en lenguaje natural, que representan casi el 68% de los artículos recuperados. Para interpretar la figura, se requiere considerar que existen artículo que procesan tanto voz y texto o diagramas y texto, por lo que un mismo artículo puede estar en dos categorías (debido a esto, la suma de los artículos en las categorías es mayor a 64). En Tabla 5-9 se presentan las referencias de artículos de acuerdo a tipo de entrada que procesa el enfoque.

Tabla 5-9. Referencia de artículos de acuerdo al tipo de entrada que procesa el enfoque.

Entradas del enfoque	Artículos	Número de artículos
Diagrama	ART40, ART49, ART60, ART16, ART34	5
Documentos CNL	ART06, ART20, ART63	3
Documento NL	ART01, ART02, ART03, ART04, ART05, ART07, ART08, ART09, ART10, ART11, ART12, ART14, ART15, ART16, ART17, ART18, ART19, ART21, ART23, ART24, ART25, ART26, ART28, ART29, ART30, ART34, ART35, ART36, ART37, ART38, ART39, ART41, ART42, ART43, ART45, ART46, ART47, ART48, ART50, ART51, ART52, ART56, ART59, ART61, ART62, ART64	46
Ontología	ART22, ART53, ART57	3
Voz	ART08, ART31	2
No Aplica	ART13, ART27, ART32, ART33, ART44, ART54, ART55, ART58	8

5.2.4 RQ4.- ¿Qué tipo de salidas o resultados presenta el trabajo reportado?

Esta pregunta de investigación está relacionada con la anterior, debido a que las entradas se procesan para generar una salida, por ejemplo: generación de modelos, identificación de requerimientos o análisis de calidad de los requerimientos. En la Tabla 5-10 se muestra la clasificación de (Meth et al., 2013), modificada agregando los tipos de salida “Clasificación de Requerimientos” e “Identificación de Requerimientos”, y se agregó la clasificación “Otro”, para agregar los artículos que tiene como resultado un tipo de salida que no entran en algún tipo de los considerados.

Tabla 5-10. Clasificación de tipos de salida.

Tipo de salida/finalidad del enfoque	Descripción
Análisis de calidad	Se enfoca en la detección de incertidumbre, ambigüedad o detección de duplicidad de funcionalidad en los requerimientos de usuario.
Clasificación de Requerimientos	Aquellos trabajos cuyo propósito es clasificar los requerimientos de un documento en lenguaje natural.
Generación de Modelos	Se centra en la generación de modelos, diagramas o esquemas.
Identificación de abstracción	Se refiere a aquellos trabajos cuyo objetivo es identificar metas, actores, tareas, casos de uso, glosario de términos, generación de ontologías o simplemente entender el dominio.
Identificación de Requerimientos	Es la identificación de posibles requerimientos en un documento de lenguaje natural.
Otro	Son aquellos artículos que no entran en las clasificaciones anteriores.

La distribución de artículos con base a la categoría antes mencionada, se presenta en la Figura 5.9. Ahí se observa que la mayoría de los trabajos seleccionados en este estudio generan modelos y la identificación de requerimientos. En la Tabla 5-11 se encuentra la relación de artículos respecto a los distintos tipos de salida.

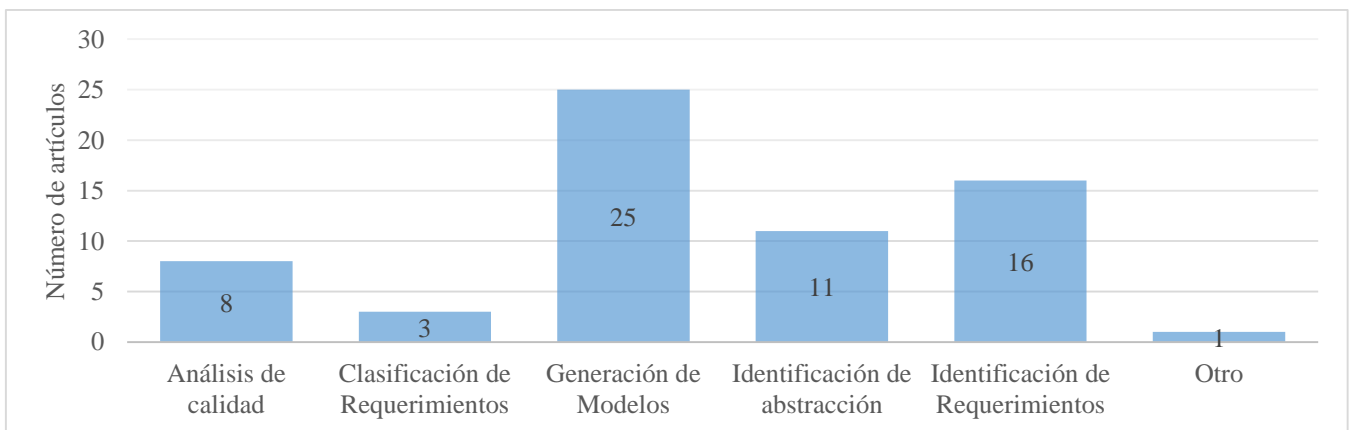
**Figura 5.9. Distribución de artículos con base en su tipo de salida.**

Tabla 5-11. Referencia de artículos con base al tipo de salida o finalidad del artículo.

Tipo de salida o finalidad	Artículos	Número de artículos
Análisis de Calidad	ART10, ART12, ART13, ART17, ART24, ART28, ART46, ART60	8
Clasificación de Requerimientos	ART02, ART03, ART48	3
Generación de Modelos	ART05, ART14, ART16, ART19, ART20, ART21, ART22, ART23, ART26, ART30, ART31, ART33, ART34, ART37, ART38, ART39, ART40, ART47, ART49, ART53, ART57, ART58, ART62, ART63, ART64	25
Identificación de Abstracción	ART01, ART07, ART11, ART15, ART35, ART36, ART43, ART44, ART50, ART52, ART61	11
Identificación de Requerimientos	ART04, ART08, ART09, ART18, ART25, ART27, ART29, ART32, ART41, ART42, ART45, ART51, ART54, ART55, ART56, ART59	16
Otro	ART06	1

5.2.5 RQ5.- ¿Qué tipo de validación se utilizó en el trabajo reportado?

Para presentar los tipos de evaluaciones realizadas para validar los resultados de los enfoques o herramientas, se utilizó la clasificación implementada en (Meth et al., 2013). Modificada agregando el tipo “Otro”, para incluir artículos que realizaron algún otro tipo de evaluación y el tipo “No Aplica”, para cuando no se presenta algún tipo de evaluación. Los tipos de evaluación, incluidos en esta clasificación, se explican en la Tabla 5-12. La distribución porcentual de artículos, cuando se ubican en esta clasificación se muestra en la Figura 5.10.

Tabla 5-12. Descripción de tipos de evaluación.

Tipo de evaluación	Descripción
Caso de estudio	Evaluación en la práctica, , mediante un ejemplo en el contexto de un entorno real.

Experimento controlado	Un experimento controlado implica estudiar el método o artefacto presentado, incluido en un entorno controlado, con recopilación y análisis de datos. Por ejemplo, comparar el desempeño de un analista que usa el artefacto, con el desempeño de un analista que no lo usa.
Prueba de concepto	Una demostración del artefacto o método presentado, sin recopilación y análisis de datos. Por ejemplo, utilización de una aplicación a un ejemplo del mundo real.
Simulación	Ejecución del artefacto con datos de prueba. Donde se compara la salida del artefacto con un conjunto de requisitos estándar, que es la salida creada manualmente por un experto o un grupo de expertos.
Otro	Evaluación que no entran en la categorías anteriores.
No aplica	No se presenta evaluación alguna.

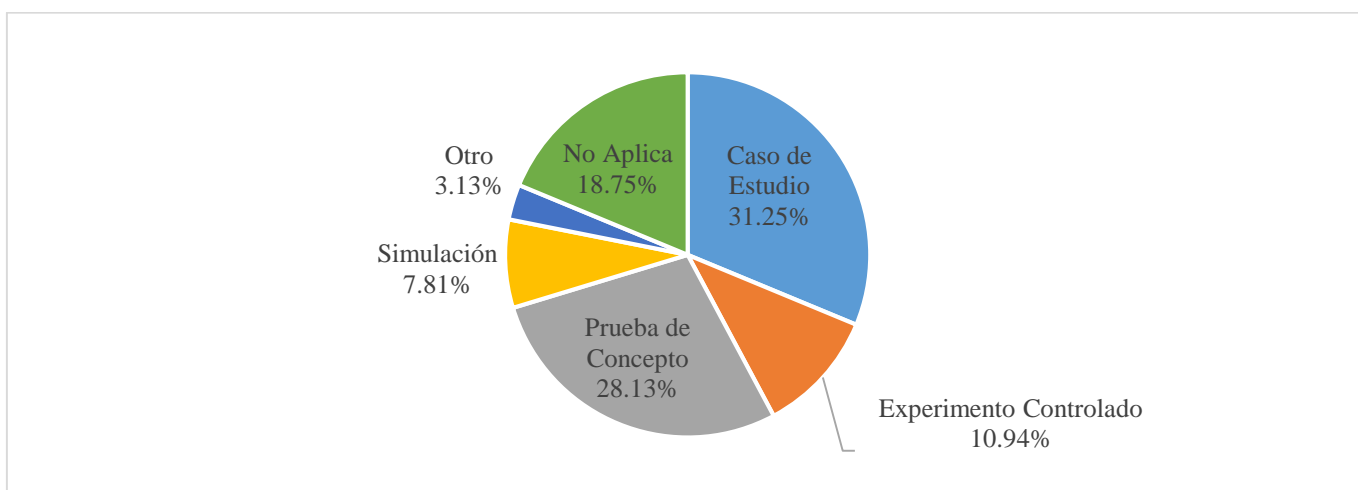


Figura 5.10. Distribución porcentual de artículos sobre sus tipos de evaluación.

En la Figura 5.10 se presenta que 31.25% (20) de los artículo usan caso de estudio para evaluar su enfoque. 28.13% (18) de los artículos realizan la evaluación prueba de concepto, es decir, que sólo presenta un ejemplo del funcionamiento de su enfoque. La evaluación mediante experimentos controlados representa 10.94% (7 artículos), que controlan las variables que influyen en sus experimentos. Sólo 3.13% (dos) artículos no entraron en los tipos de evaluación anteriores, por ejemplo, personas que evalúan la herramienta o enfoque propuesto de manera cualitativa. Por últimos 18.78% (12) artículos no presentan algún tipo

de evaluación. En la Tabla 5-13 se presentan las referencias de los artículos, de acuerdo al tipo de evaluación empleado.

Tabla 5-13. Referencia de artículos de acuerdo tipo de evaluación empleado.

Enfoque de evaluación	Artículos	Número de artículos
Caso de estudio	ART01, ART03, ART05, ART08, ART09, ART11, ART19, ART20, ART24, ART26, ART28, ART32, ART34, ART36, ART37, ART38, ART39, ART50, ART51, ART62	20
Experimento Controlado	ART02, ART21, ART35, ART41, ART45, ART49, ART57	7
Prueba de Concepto	ART07, ART10, ART15, ART22, ART23, ART29, ART30, ART31, ART33, ART40, ART44, ART47, ART54, ART55, ART59, ART60, ART61, ART64	18
Simulación	ART04, ART14, ART25, ART46, ART48	5
Otro	ART43, ART63	2
No aplica	ART06, ART12, ART13, ART16, ART17, ART18, ART27, ART42, ART52, ART53, ART56, ART58	12

5.2.6 RQ6.- ¿Qué tipo de investigación se publicó?

De acuerdo a la clasificación de tipos de investigación que propone (Wieringa et al., 2006), sólo se identificaron artículos que se encuentran en la categoría de:

- **Investigación de evaluación:** Es la investigación de un problema en la práctica de ingeniería de requerimientos o en la implementación práctica de una técnica.
- **Propuesta de solución:** Es la investigación donde se propone una técnica de solución y defiende su relevancia, sin una validación completa.
- **Investigación de validación:** En este tipo, se investigan las propiedades de una propuesta de solución que todavía no se implementa en la práctica de ingeniería de requerimientos.

- **Artículo filosófico:** Este tipo de publicación describe una nueva forma de mirar las cosas, un marco de trabajo conceptual nuevo.

Se añadió la categoría “Otro” para agregar aquellos artículos que no entraron en ninguna de las categorías anteriores. De los trabajos de investigación recuperados, se identificaron 17 artículos de tipo “Investigación de evaluación”. 39 artículos se clasificaron como “Propuesta de solución”, donde se proponen nuevos enfoques para llevar a cabo la elicitación de requerimientos. Por último, dos artículos se categorizaron como “Investigación de validación” y un artículo se identificó como “Artículo filosófico”. Esto se observa en la Figura 5.11 donde se muestra la distribución de artículos y en la Tabla 5-14 se muestra la referencia de artículos respecto al tipo de investigación.

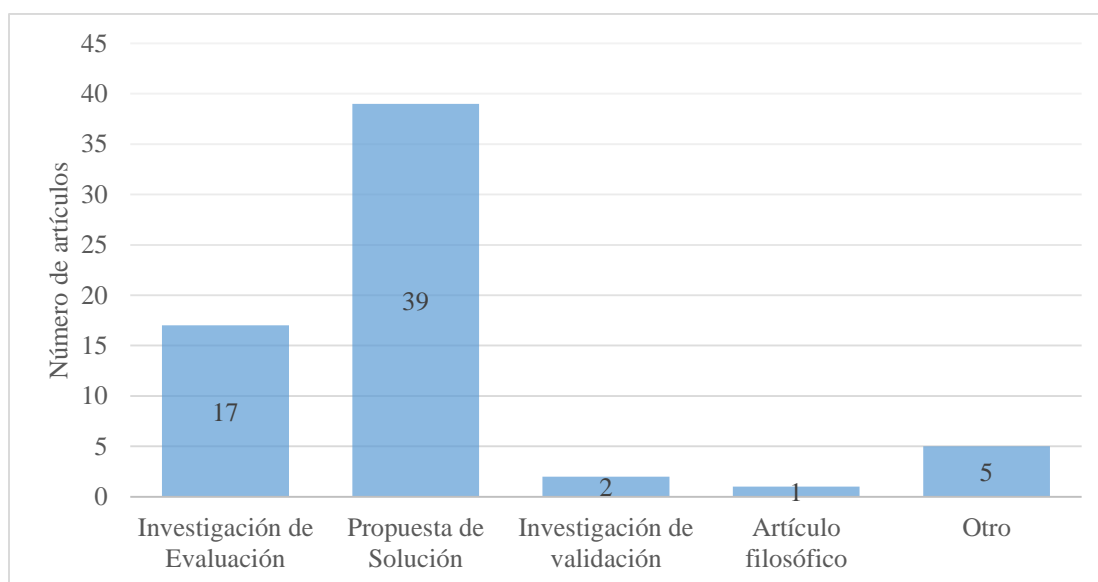


Figura 5.11. Distribución de artículos de acuerdo al tipo de investigación.

Tabla 5-14. Referencia de artículos respecto al tipo de investigación.

Tipo de Investigación	Artículos	Número de artículos
Investigación de evaluación	ART01, ART02, ART05, ART23, ART26, ART30, ART31, ART33, ART34, ART36, ART38, ART39, ART40, ART46, ART50, ART62, ART64	17
Propuesta de solución	ART03, ART04, ART06, ART07, ART08, ART09, ART10, ART11, ART12, ART13, ART14, ART15, ART16, ART19, ART20, ART21, ART22, ART24, ART27, ART28, ART29, ART32, ART35, ART37,	39

	ART41, ART42, ART44, ART45, ART47, ART49, ART51, ART52, ART54, ART55, ART56, ART57, ART59, ART61, ART63	
Investigación de validación	ART25, ART48	2
Artículo filosófico	ART53	1
Otro	ART17, ART18, ART43, ART58, ART60	5

5.2.7 RQ7.- ¿Cuál es el nivel de automatización en el trabajo reportado?

Un punto importante que este estudio quiere evaluar, es el avance en la automatización del modelado del problema y la elicitación de requerimientos. Mediante el uso del procesamiento del lenguaje natural de los clientes, para entender el dominio del problema y especificar los requerimientos. Se dividió el grado de automatización en tres categorías: automático, semi-automático (donde el usuario interactúa para completar algún resultado) y manual. Se identificaron 13 artículos con enfoques y herramientas automatizadas, 29 con un grado de semi-automatización y 22 con enfoques manuales, en la Figura 5.12 se muestra la distribución porcentual y en la Tabla 5-15 se muestra la referencia de los artículos.

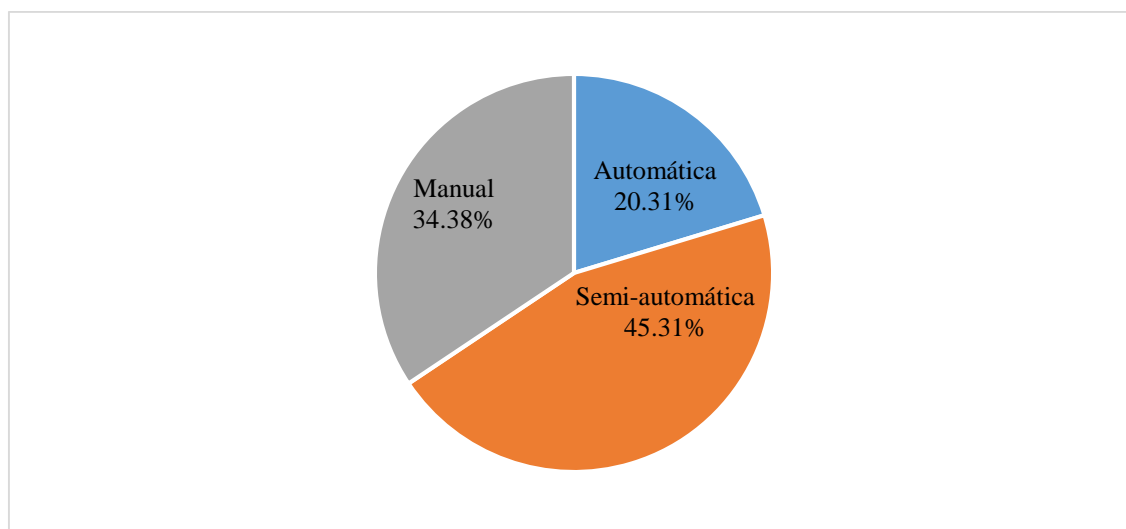


Figura 5.12. Distribución porcentual de artículos de acuerdo al grado de automatización.

Tabla 5-15. Referencia de artículos de acuerdo al grado de automatización.

Grado de automatización	Artículos	Número de artículos
Automático	ART07, ART08, ART11, ART16, ART19, ART24, ART26, ART29, ART38, ART43, ART46, ART61, ART63	13
Semi-automático	ART01, ART02, ART03, ART04, ART06, ART09, ART10, ART12, ART14, ART15, ART17, ART18, ART20, ART21, ART28, ART31, ART34, ART35, ART36, ART37, ART39, ART40, ART45, ART47, ART48, ART51, ART57, ART62, ART64	29
Manual	ART05, ART13, ART22, ART23, ART25, ART27, ART30, ART32, ART33, ART41, ART42, ART44, ART49, ART50, ART52, ART53, ART54, ART55, ART56, ART58, ART59, ART60	22

5.2.8 RQ8.-¿Qué técnicas de reconocimiento del habla, procesamiento de lenguaje natural y/o ontologías se presentan?

Lo que respecta a las técnicas de procesamiento de lenguaje natural, se identificaron gran variedad de estas, las cuales son: *lowercasing, sentence splitter, tokenization, punctuation removal, stop word removal, lemmatization, stemming, co-reference resolution, chunking, collocation extraction methods and lexical association measure, intention recognition, POS Tagging, named entity recognition, dependency parsing, constituency parsing, semantic role labelling, regular expression, JAPES Rules, bag of words, word embedding, topic modelling, similarity measure, Artificial Neuronal Networks, Machine Learning algorithms, statistical filtering, heuristic rules, algorithms, transformation rules, analogy-based reasoning, sequence alignment, precision, recall, F-measure, mean reciprocal rank, Fleiss 'kappa, average precision, mean average precision*. Además se observó un patrón de utilización de estas técnicas en distintas etapas de los enfoques propuestos por los artículos, por lo que se decidió clasificarlas de acuerdo al flujo de trabajo que sigue el NLP definido en (Sarkar, 2018).

En este estudio se identificaron artículos que realizan reconocimiento del habla, procesan el lenguaje natural textual, implementan el uso de ontología y enfoques híbridos (que combinan estos dos últimos). La clasificación de los artículos en estas categorías se muestra en la Tabla 5-16. Debido a que el mayor número de artículos tratan acerca de NLP y del uso de ontologías, se decidió describir los resultados del análisis en las dos subsecciones siguientes: Técnicas de procesamiento de lenguaje natural y Enfoques de ontologías.

Tabla 5-16. Artículos de acuerdo al tipo de enfoque.

Enfoque	Número de artículos
NLP (texto)	30
Reconocimiento del Habla (voz)	2
Ontologías	20
Híbrido (NLP y Ontologías)	12

En este estudio se identificaron 44 artículos con enfoques de NLP y reconocimiento del habla, en dichos artículos, se identificaron varias técnicas, modelos y métricas. Se realizó la clasificación de acuerdo al flujo de trabajo del procesamiento de lenguaje natural estándar, definida en (Sarkar, 2018), sin embargo, para este estudio se añadió otra fase para incluir las técnicas de reconocimiento del habla(voz). Esta clasificación se detalla en la sección 6.2 Técnicas de procesamiento de lenguaje natural.

En la Figura 5.13 se visualiza la distribución de artículos que utilizan técnicas de NLP de acuerdo a las clasificación modificada de (Sarkar, 2018). En la Tabla 5-17 se muestra la referencia de artículos.

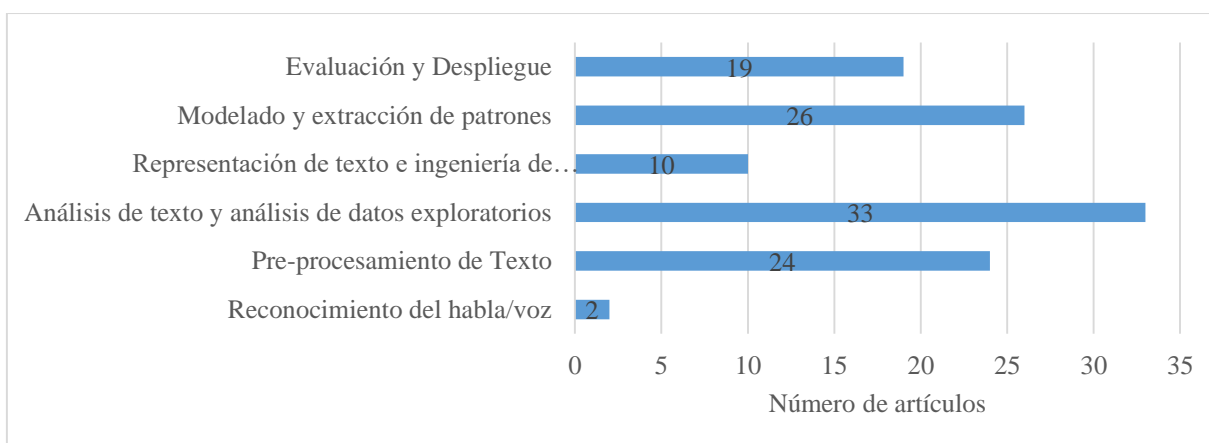


Figura 5.13. Distribución de artículos de acuerdo la clasificación de técnicas de procesamiento de lenguaje natural.

Tabla 5-17. Referencia de artículos de acuerdo al flujo de trabajo del NLP.

Fase del flujo de trabajo de NLP	Artículos	Número de artículos
Reconocimiento del habla	ART08, ART31	2
Pre-Procesamiento de texto	ART03, ART04, ART05, ART06, ART08, ART11, ART14, ART15, ART17, ART18, ART20, ART24, ART25, ART28, ART29, ART34, ART37, ART45, ART46, ART47, ART48, ART51, ART61, ART62	24
Análisis de texto y análisis de datos exploratorios	ART01, ART02, ART04, ART05, ART06, ART07, ART11, ART12, ART14, ART16, ART17, ART18, ART19, ART20, ART22, ART24, ART25, ART26, ART28, ART29, ART31, ART34, ART35, ART36, ART37, ART39, ART42, ART45, ART46, ART47, ART48, ART51, ART62	33
Representación de texto e ingeniería de características	ART03, ART04, ART05, ART08, ART09, ART10, ART25, ART29, ART57, ART61	10
Modelado y Extracción de patrones	ART01, ART02, ART03, ART04, ART05, ART08, ART09, ART10, ART11, ART14, ART16, ART17, ART18, ART19, ART20, ART25, ART26, ART28, ART29, ART35, ART36, ART42, ART46, ART57, ART63, ART64	26
Evaluación y Despliegue	ART01, ART02, ART03, ART04, ART05, ART08, ART09, ART11, ART14, ART24, ART25, ART26, ART28, ART35, ART36, ART37, ART46, ART61, ART63	19

De nuestro conjunto seleccionado de artículos se identificaron 32 trabajos de investigación que utilizan ontologías en sus enfoques. Dentro de estos trabajos, se observó diferentes tipos de ontologías propuestos que, a su vez, apoyan las distintas actividades de la ingeniería de requerimientos. Derivado de esta información identificada, se decidió clasificar los artículos por los tipos de ontologías propuesto y por la actividad (de la ingeniería de

requerimientos) que apoya dicha ontología. En la siguiente sección presentamos esta clasificación.

5.2.8.1 Enfoques de ontologías identificados

En este estudio se identificaron 32 artículos que proponen enfoques de ontologías en las distintas fases de la ingeniería de requerimientos y el modelado del problema. (Farfeleder, Moser, Krall, Ståalhane, et al., 2011) menciona que la ingeniería de requerimientos es una disciplina que trata de comprender, documentar, comunicar e implementar las necesidades del cliente. Dicho autor define las principales actividades del proceso de ingeniería de requerimientos en: “elicitación de requerimientos”; “análisis y negociación de requerimientos”; “documentación y validación de requerimientos”; así como la “administración de requerimientos”. A estas actividades, se agregó la actividad de “modelado del dominio del problema” (modelado del negocio), para abarcar completamente la temática de la fase de análisis, del desarrollo de software, una descripción breve de las fases de la ingeniería se muestra en la Tabla 5-18. Se utilizó esta clasificación dado que el enfoque de las ontologías lo determina la fase de la ingeniería de requerimientos en la que se utiliza. La distribución de artículos respecto a esta clasificación se presenta en la Figura 5.14, donde se identificó que la principal implementación de ontologías es en la etapa de elicitación de requerimientos como apoyo para la captura de requerimientos. En la Tabla 5-19 se presenta la referencia de artículos de acuerdo a esta clasificación.

Tabla 5-18. Descripción de las fases de la ingeniería de requerimientos.

Fase de la Ingeniería de Requerimientos	Descripción
Modelado del dominio del problema	Se realiza la asistencia en el modelado explícito del dominio del problema.
Elicitación de Requerimientos	La obtención de requerimientos involucra al personal técnico que trabaja con los clientes para averiguar sobre el dominio de la aplicación, los servicios que debe proporcionar el sistema y las limitaciones operativas del sistema. El objetivo es reunir los requisitos en bruto.
Análisis y Negociación de Requerimientos	El análisis y la negociación de requerimientos es una actividad que tiene como objetivo descubrir problemas y conflictos con los requisitos y llegar a un acuerdo sobre los cambios para satisfacer a todos los interesados del sistema (personas afectadas por el sistema propuesto). El objetivo final es alcanzar un

	entendimiento común de los requisitos entre todos los participantes del proyecto.
Documentación y Validación de Requerimientos	Los requerimientos definidos, escritos en una especificación de requisitos de software, se validan según criterios como corrección, integridad, consistencia, verificabilidad, no ambigüedad, trazabilidad, etc.
Administración de Requerimientos	La gestión de requerimientos consiste en gestionar los cambios de requisitos (manteniéndolos coherentes), por ejemplo, asegurando la trazabilidad de los requisitos (identificación de las interdependencias entre los requisitos, otros requisitos y artefactos).

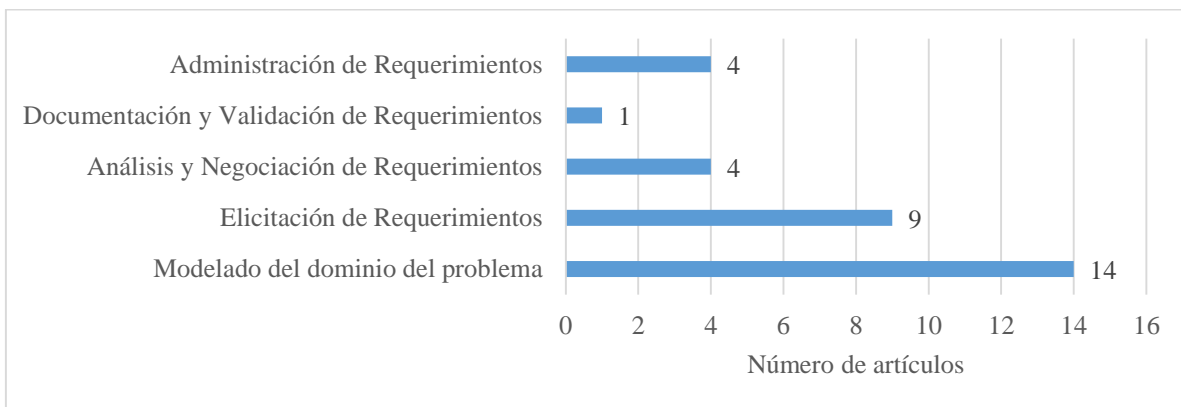


Figura 5.14. Fases de la ingeniería de requerimientos soportados por ontologías.

Tabla 5-19. Referencia de artículos de acuerdo a la fase de la RE que apoya el enfoque ontológico.

Fase de la ingeniería de requerimientos	Artículos	Número de artículos
Modelado del dominio del problema	ART14, ART22, ART23, ART30, ART40, ART44, ART49, ART50, ART52, ART53, ART54, ART57, ART60, ART62	14
Elicitación de Requerimientos	ART12, ART16, ART27, ART32, ART41, ART42, ART51, ART58, ART59	9
Análisis y Negociación de Requerimientos	ART13, ART45, ART55, ART61	4

Documentación y Validación de Requerimientos	ART38	1
Administración de Requerimientos	ART21, ART33, ART37, ART56	4

Por otro lado, se identificó el uso de distintos tipos de ontologías dentro de los procesos de modelado del dominio del problema y la ingeniería de requerimientos. En este estudio se utilizó la clasificación propuesta por (Castañeda et al., 2010) donde se identificaron tres tipos de ontologías: “ontología de requerimientos”; “ontología de documento de especificación de requerimientos”; y “ontología del dominio de aplicación” (aquí es donde se tienen dos modelos de dominio: modelo de dominio del problema (relacionado a los patrones de análisis) y modelo de dominio de la solución (relacionado a los patrones de requerimientos)). En la Tabla 5-20 se muestra una breve descripción de los tres tipos de ontología de acuerdo a (Castañeda et al., 2010). La distribución de artículos con base a este esquema de clasificación se muestra en la Figura 5.15. Para interpretar correctamente la información mostrada en la Figura 5.15, se menciona que se identificaron trabajos de investigación que utilizan dos ontologías u ontologías híbridas, por lo tanto, un artículo puede estar en dos categorías. Por último, en la Tabla 5-21 se presenta la referencia de artículos de acuerdo al tipo de ontología que manejan.

Tabla 5-20. Descripción de los tipos de ontología de acuerdo a(Castañeda et al., 2010).

Tipos de ontología	Descripción
Ontología de requerimientos	Las especificaciones de requisitos son las descripciones de las características de software deseadas especificadas por los clientes. Este modelo se puede definir utilizando una ontología de nivel superior. Esta ontología se puede utilizar durante el proceso de obtención para reducir los requisitos ambiguos y evitar definiciones de requisitos incompletas. Las restricciones sobre los requisitos se pueden definir en esta ontología. Ayudan en la validación y verificación de requisitos.
Ontología de documento de especificación de requerimientos	En la ingeniería de requerimientos se utilizan diferentes enfoques como pasos intermedios para obtener los requisitos. Uno de estos enfoques es la técnica de los escenarios, que son descripciones ejemplares del uso del sistema planificado para alcanzar un objetivo definido. El uso de ontologías para describir la estructura de los documentos de especificación de requisitos

	<p>reduce las especificaciones de requisitos insuficientes. Además, pueden ayudar enormemente en la definición de varias estructuras para mostrar el mismo conocimiento, para, por ejemplo, involucrar a todas las partes interesadas en el análisis de los requisitos obtenidos. Además, también pueden ayudar a reutilizar la representación estructurada para objetivos o proyectos diversos, solo cambiando su contenido.</p>
<p>Ontología del dominio de aplicación</p>	<p>Esta ontología representa el conocimiento del dominio de la aplicación y la información comercial requerida para crear aplicaciones de software en un dominio específico. También incluye las relaciones semánticas establecidas entre sus conceptos desde un punto de vista del mundo real. Una ontología de dominio de aplicación es útil para identificar requisitos dinámicos y cambiantes, ya que ayuda a comprender el dominio.</p>

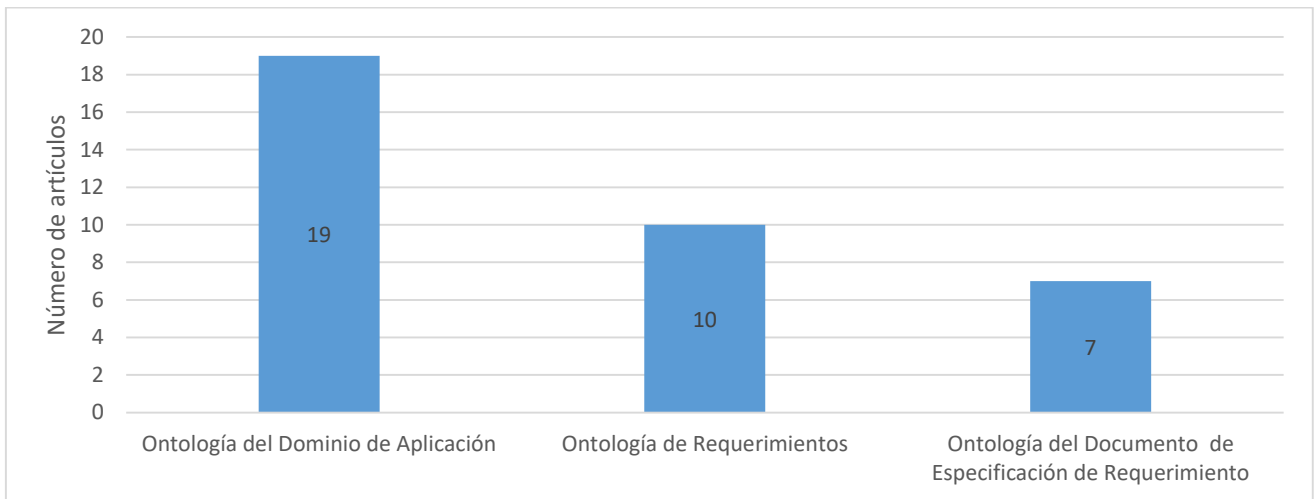


Figura 5.15. Distribución de artículos de acuerdo el tipo de ontología.

Tabla 5-21. Referencia de artículos de acuerdo al tipo de ontología implementado.

Tipo de Ontología	Artículos	Número de artículos
Ontología de requerimientos	ART14, ART21, ART27, ART30, ART33, ART37, ART41, ART56, ART59, ART60	10

Ontología de documento de especificación de requerimientos	ART12, ART13, ART33, ART37, ART49, ART54, ART61	7
Ontología del dominio de aplicación	ART16, ART22, ART23, ART30, ART32, ART38, ART40, ART41, ART42, ART44, ART45, ART50, ART51, ART52, ART53, ART55, ART57, ART58, ART62	19

5.3 Resumen del capítulo

En este capítulo se presentó el mapa sistemático resultado de este estudio, donde se muestra un panorama visual del estado de arte sobre uso de reconocimiento del habla, procesamiento de lenguaje natural y ontologías para modelado del dominio del problema e identificación de requerimientos de solución (parte de la actividad de la elicitación de requerimientos dentro de la etapa de análisis del desarrollo de software). Con este resumen visual (mapa sistemático) se observó el estado del arte correspondiente sobre esta área de investigación.

Capítulo 6.

Principales

hallazgos

En este capítulo se presentan los principales hallazgos a que se identificaron en resultados obtenidos en este estudio de mapeo sistemático.

6.1 Aspectos importantes

En este estudio de mapeo sistemático se desarrollaron ocho clasificaciones de acuerdo al análisis temático de los artículos de investigación, véase la sección 4.1.2. Estas clasificaciones permiten analizar desde diferentes puntos de vista la información recopilada e intentar obtener conclusiones en cada una de estas perspectivas. Las clasificaciones son acordes a las ocho preguntas de investigación definidas, las cuales son: RQ1 (clasificación de tipos de técnica), RQ2 (tipo de contribución del artículo), RQ3 (tipo de entrada al proceso), RQ4 (tipo de resultados reportados), RQ5 (tipo de enfoque de evaluaciones de los resultados), RQ6 (tipo de investigación reportada), RQ7 (nivel de automatización del resultado) y RQ8 (técnicas NLP y enfoque de las ontologías utilizadas).

En lo que respecta a la RQ1, se identificó que la mayor aportación de artículos sobre esta área de investigación es en congresos, en donde la principal fuente es la base de datos IEEE Xplorer. Con base a la pregunta de investigación RQ2, se identificó que la importancia sobre la automatización en la etapa de análisis del desarrollo de software tomó más importancia a partir del 2015 con base a nuestros artículos obtenidos en este SMS, esto implica que es un área en crecimiento. El objetivo de la RQ3 es identificar aquellos países con mayor contribución, entre los cuales son: Alemania, Canadá China e India; aunque hay que agregar que Estados Unidos, España, Jordania y Luxemburgo han contribuido recientemente (a partir del 2015). Lo que corresponde a la RQ4 es identificar el ámbito del autor, donde se identificó que el 95.31% corresponde al ámbito académico, debido a que estas investigaciones provienen de instituciones y universidades. Por último, la RQ5 tiene como objetivo identificar las principales revistas y congresos acerca de esta área de investigación, donde el congreso con más aportaciones (12 artículos) es *IEEE International Requirements Engineering Conference (RE)*, por lo que debe considerar para búsqueda de más artículo de investigación.

Dentro de este SMS se identificaron 32 artículos con respecto al uso de Procesamiento de Lenguaje Natural y sólo dos artículos utilizan el reconocimiento del habla. Por otra parte, dado que uno de los objetivos de este estudio es identificar evidencia del uso de técnicas de Procesamiento de Lenguaje Natural en la elicitación de requerimientos, se seleccionaron artículos que abarcan todo el proceso de la ingeniería de requerimientos dado que incluyen la etapa de elicitación. Se debe resaltar que en lo que respecta a técnicas de reconocimiento del habla, sólo se recuperaron dos artículos, por lo que se tuvo que realizar la búsqueda extra de artículos que aporten sobre este tema, entre los cuales resaltan las investigaciones de (Ajit Singh, Taruna Panchal, 2013), (Nisha, 2017) y (Alasadi & Deshmukh, 2018), donde coinciden con técnicas de reconocimiento del habla las cuales se identificaron. Dentro de los dos artículos que utilizan reconocimiento del habla, sólo uno menciona el uso de la técnica Modelos de Markov Ocultos mientras que el otro artículo no menciona la técnica.

Para las técnicas de la fase “pre-procesamiento”, se identificaron que 24 artículos utilizan técnicas de esta fase (de acuerdo al flujo de trabajo del NLP) para limpieza de información ya sea para lenguaje natural o lenguaje natural controlado. Dentro de la fase de “análisis de texto y análisis de datos exploratorios” son técnicas que se enfocan más en la semántica y estructuración de las oraciones como son las técnicas de construcción de árboles sintácticos (*Dependency parsing* y *Constituency parsing*), etiquetado de rol semántico (*Semantic Role Labelling*), etiquetado de palabras (*POS Tagging*), identificación de entidades o sustantivos compuestos por dos palabras (*Named Entity Recognition* y *Chunking*) y por último, técnicas de extracción de combinación de palabras que ocurren con frecuencia (*Collocation Extraction*).

La siguiente fase, “Representación de texto e ingeniería de características” tiene como objetivo representar el texto mediante características (numéricas) para procesarlos. Se identificaron varias técnicas como son: *word frequency*, *n-gramas models* y *TF-IDF* que son categorizados como bolsa de palabras, además de *word2vec* y *paragraph vector* categorizados como palabras embebidas. La peculiaridad de estas técnicas es que son utilizadas en la ingeniería de características, es decir que, para procesar la información es necesario transformar el texto (lenguaje natural) en características (atributos cuantitativos) de manera que las ML, ANN o técnica de alineación de secuencia puedan trabajar y procesar dichos datos.

En la fase “Modelado y extracción de patrones” se identificaron distintas estrategias para realizar ciertas tareas como: identificación de requerimientos o posibles requerimientos, clasificación de requerimientos, identificación de actores, casos de uso, metas, acciones (verbos), detección de términos ambiguos, extracción de glosario de términos, generación de modelos (modelos conceptuales, esquemas de estrella, cuadros de estado, ontologías, modelos de metas, modelo de características, diagramas de casos de uso y modelos de procesos de negocio). En este estudio se identificaron ocho que usan ML y tres artículos que usan ANN, para identificación y clasificación de requerimientos de documentos en lenguaje natural. Por otro lado, también se identificó el uso de reglas de heurística en ocho artículos y uso de algoritmos en siete artículos. Sin embargo, la tendencia de los últimos tres años (2016 al 2018) presenta uso de ML y ANN, por lo que deben de investigarse más a fondo el uso de estas técnicas.

Por último, la fase de “Evaluación y desarrollo” se utilizan métricas para medir el nivel de efectividad de los enfoques NLP de manera, dentro de las técnicas más utilizadas son *Presicion*, *Recall* y *Accuracy*, pero además se encontraron las métricas *F-measure*, *Mean Reciprocal Rank*, *Fleiss' kappa*, *Average Presicion* y *Mean Average Presicion*, que también son métricas para evaluar este tipo de enfoques.

Respecto al uso del NLP se identificaron el uso de recursos, como es el uso de listas de entidades, listas de acciones, bases de datos léxicas (como WordNet y FrameNet), corpus y ontologías con el propósito de utilizarlos como diccionarios, al momento de identificar ciertos elementos en texto (lenguaje natural).

Respecto a los distintos enfoques de uso de ontologías, en 32 artículos, se utilizó la clasificación de (Farfeleder, Moser, Krall, Ståalhan, et al., 2011), donde se definen las principales actividades de la

ingeniería de requerimientos. Así, se observó que los artículos seleccionados abarcan todo el proceso de la ingeniería de requerimientos, pero las ontologías propuestas atacaban sólo una o dos actividades en concreto de dicho proceso, donde la mayoría (59.38%) se ocupaban en la elicitación de requerimientos (que abarca entender el dominio del problema y obtener los requerimientos).

Así mismo, se identificaron tres tipos de ontologías, que concuerdan con la clasificación de (Farfeleder, Moser, Krall, Ståalhanne, et al., 2011): ontologías del dominio, ontologías de requerimientos y ontologías del documento de especificación de requerimientos.

- *La ontología de dominio* contiene conceptos acerca del dominio del problema y se identificó el uso de este tipo de ontología en 19 artículos, los cuales tratan la actividad de elicitación.
- La “ontología de requerimientos”, trata el “análisis y negociación de requerimientos”, “documentación y validación de requerimientos” y la “administración de requerimientos”, de los cuales se identificaron 10 artículos.
- La “ontología de documento de especificación de requerimientos”, que se usa como intermediarios para obtener los requerimientos donde se tiene información como metas, actores y escenarios. Este tipo de ontología ataca tanto la elicitación como el análisis y negociación de requerimientos, este estudio identificó siete de este tipo.

Se debe mencionar que se identificaron enfoques con dos tipos de ontologías o combinación de estas, pero no una combinación de los tres tipos, por lo tanto, es un área de oportunidad.

6.2 Técnicas de procesamiento de lenguaje natural

En este estudio se identificaron 44 artículos con enfoques de NLP y reconocimiento del habla, entre los cuales, se identificaron varias técnicas, modelos y métricas con el objetivo de automatizar la etapa de elicitación de requerimientos. Se realizó la clasificación de acuerdo al flujo de trabajo del procesamiento de lenguaje natural estándar, definida en (Sarkar, 2018), sin embargo, para este estudio se agregó otra fase para incluir las técnicas de reconocimiento del habla(voz). A continuación, se explica cada fase, véase la Figura 6.1.

1. **Reconocimiento del habla:** Proceso de convertir la señal de voz en una secuencia de palabras mediante un algoritmo implementado como un programa de computadora (Gaikwad et al., 2010). La clasificación de los distintos métodos/técnicas es de acuerdo a (Ajit Singh, Taruna Panchal, 2013), (Nisha, 2017) y (Alasadi & Deshmukh, 2018), los cuales coinciden en ciertas técnicas.
2. **Pre-procesamiento de texto:** Se enfoca en realizar un proceso de limpieza y/o alteración del documento de texto para que este pueda ser manipulados por el proceso posterior.
3. **Análisis de texto y análisis de datos exploratorios:** Es el proceso de realizar el análisis de sintáctico y semántico de texto en lenguaje natural.

4. **Representación de texto e ingeniería de características:** Es la etapa donde se utiliza el conocimiento de dominio de los datos(texto) para seleccionar o crear nuevas características (valores numéricos) que permitan a los algoritmos de aprendizaje automático trabajar.
5. **Modelado y extracción de patrones:** Se enfoca en la extracción de patrones mediante modelos (modelado de tópicos, medidas de similitud, algoritmos, reglas de transformación (inferencia), de aprendizaje automático, redes neuronales, alineación de secuencias, y reglas heurísticas).
6. **Evaluación y despliegue:** Etapa donde se utilizan métricas para medir el nivel de interpretación o efectividad de un enfoque de NLP.

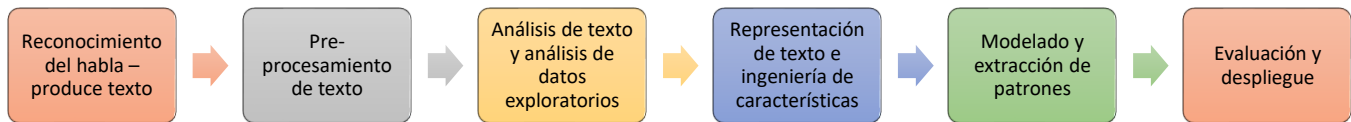


Figura 6.1. Flujo de trabajo del NLP.

Se realizó la búsqueda exhaustiva de todas las técnicas en NLP tanto en los artículos recuperados como de otras fuentes (otros artículos, libros, información de la web). En la Figura 6.2 se muestra un esquema general de las técnicas identificadas categorizadas en cada una de las fases de la clasificación y una breve descripción de las técnicas en la Tabla 6-1.

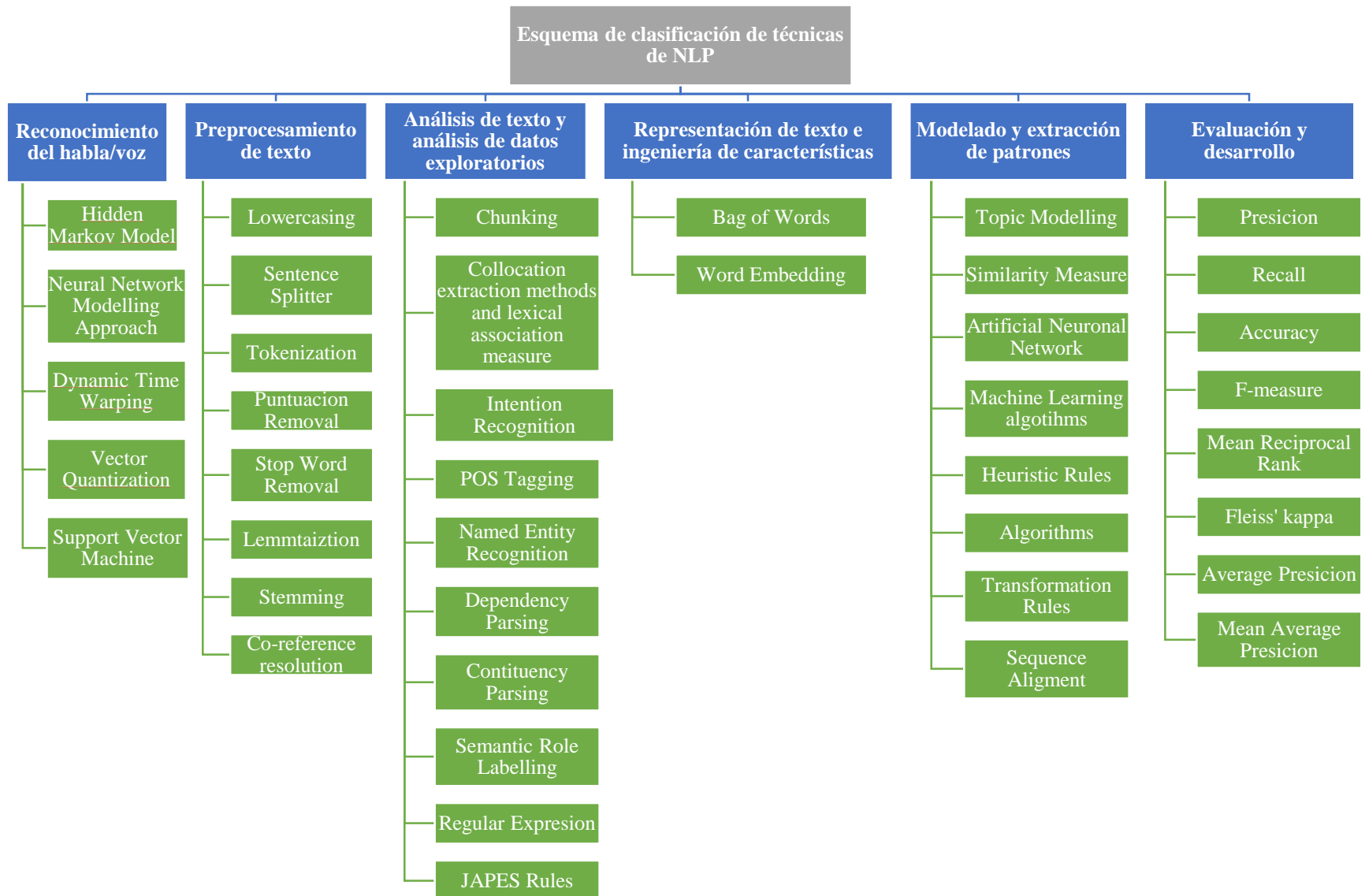


Figura 6.2. Clasificación de técnicas de procesamiento de lenguaje natural.

Tabla 6-1. Descripción de técnicas de Procesamiento de Lenguaje Natural.

Fase de flujo de trabajo de NLP	Técnicas	Descripción
Reconocimiento del habla	Modelo de Markov Oculto (<i>Hidden Markov Model, HMM</i>)	HMM es un enfoque estocástico. Es fácil, computacionalmente práctico y puede ser entrenado rutinariamente, por lo que son modelos muy modernos en el reconocimiento de altavoces. Este modelo se caracteriza por un modelo de Markov de estado finito y un conjunto de distribuciones de salida. Se basa en un amplio vocabulario de sistemas de reconocimiento de voz y se entrena automáticamente en grandes datos de voz durante horas (Ajit Singh, Taruna Panchal, 2013).
	Enfoque de modelado de red neuronal (<i>Neural Network Modelling Approach</i>)	Red neuronal también se han utilizado para el sistema de reconocimiento de voz. Se están utilizando en la resolución de tareas de identificación complejas. La ventaja de este enfoque es que pueden controlar datos ruidosos y de baja calidad y son independientes del hablante. La desventaja del enfoque de la Red Neuronal es que la selección de la configuración óptima no es fácil de seleccionar. Enfoque basado en Red Neuronal utilizado en el reconocimiento de fonemas. El sistema que utiliza el sistema basado en el enfoque de Red Neuronal proporciona una mejor precisión en comparación con HMM, que se utiliza para datos de entrenamiento y vocabulario limitados. El híbrido Red Neuronal-HMM usa la parte Red Neuronal para la identificación de fonemas y la parte HMM para el modelado de idiomas(Ajit Singh, Taruna Panchal, 2013).
	Alineación Dinámica en Tiempo (<i>Dynamic Time Warping, DTW</i>)	DTW es un algoritmo, que se utiliza para calcular la similitud entre dos series que pueden diferir en el tiempo o la velocidad. Ha sido útil para video, audio, gráficos y cualquier información que pueda ser retorcida en una representación lineal y analizada con DTW. El proceso de optimización se realiza mediante la programación dinámica, por lo que se denomina

		Dynamic Time Warping. La continuidad es menor en DTW en comparación con otros enfoques(Nisha, 2017).
	Cuantización vectorial (<i>Vector Quantization, VQ</i>)	VQ es una función de mapear las muestras de voz de un gran espacio vectorial a un número finito de regiones en ese espacio. Cada región se llama como un clúster y ese clúster se puede denotar por su centro denominado como palabras de Centroid / Code. El libro de códigos(<i>codebook</i>) es una colección de palabras de código. Para cada hablante se genera un libro de códigos en el método VQ. El libro de códigos luego sirve como palabras pregrabadas para el orador y se usa cuando el orador se prueba en el sistema. En el sistema de reconocimiento de voz, para lograr una alta tasa de reconocimiento de los hablantes, utiliza VQ como parámetro y este parámetro se utiliza para la velocidad de precisión, el tiempo de procesamiento, el número de altavoces y el tamaño de la base de datos de entrenamiento. El beneficio de esta técnica es que ahorra mucho tiempo a lo largo de la fase de prueba y disminuye el esfuerzo de almacenamiento y cálculo para resolver la semejanza de los vectores de análisis espectral (Ajit Singh, Taruna Panchal, 2013).
	Máquinas de Vector Soporte (<i>Support Vector Machine, SVM</i>)	La idea detrás del enfoque de SVM es el uso de diferentes tipos de funciones de kernel como lineal, cuadrática o exponencial. que mapean los vectores de características de la señal de voz a un espacio de características de mayor dimensión utilizando una transformación no lineal. SVM genera una puntuación y las puntuaciones se integran a lo largo de toda la emisión para obtener la puntuación de la decisión final. El reciente enfoque basado en SVM se ha aplicado a los super vectores que son vectores simples que representan una expresión completa(Ajit Singh, Taruna Panchal, 2013).
Pre-procesamiento	Conversión de mayúsculas a minúsculas (<i>Lowercasing</i>)	Conversión de texto a minúsculas. Ejemplo en inglés:

	<p>Entrada: “<i>Example of input text for Natural Language Processing</i>”</p> <p>Salida: “<i>example of input text for natural language processing</i>”</p>
Divisor de oraciones (<i>Sentence Splitter</i>)	<p>Segmentación del texto en oraciones. Ejemplo en inglés:</p> <p>Entrada: “<i>The sky is blue. The clouds are white.</i>”</p> <p>Salida: “<i>The sky is blue.</i>”, “<i>The clouds are white</i>”</p>
Tokenización (<i>Tokenization</i>)	<p>Segmentación de la oración en tokens. Ejemplo en inglés:</p> <p>Entrada: “<i>The house is pink</i>”</p> <p>Salida: “<i>The</i>”, “<i>house</i>”, “<i>is</i>”, “<i>pink</i>”</p>
Eliminación de signos de puntuación (<i>Punctuation Removal</i>)	<p>Eliminación de signos de puntuación. Ejemplo en inglés:</p> <p>Entrada: “<i><h1>Processing of Natural Language: ‘a novel approach’</h1></i>”</p> <p>Salida: “<i>Processing of Natural Language a novel approach</i>”</p>
Eliminación de palabras sin significado (<i>Stop Word Removal</i>)	<p>Eliminación de palabras sin significado como son artículos y preposiciones. Ejemplos en inglés:</p> <p>“<i>the</i>”, “<i>a</i>”, “<i>an</i>” e “<i>in</i>”</p>
Lematización (<i>Lemmatization</i>)	<p>Método para dejar en forma base las palabras (lema, unidad autónoma constituyente del léxico de un idioma). Por ejemplo:</p> <p>“<i>am</i>”, “<i>are</i>”, “<i>is</i>” => “<i>be</i>”</p> <p>“<i>working</i>”, “<i>worked</i>”, “<i>works</i>” => “<i>work</i>”</p>

	Reducción a estema (<i>Stemming</i>)	Método para reducir una palabra a un estema (forma base de una palabra después de que todos los afijos son eliminados). “friendship”, “boyfriend”, “friendly” => “friend”
	Resolución de co-referencia (<i>Co-reference Resolution</i>): • <i>Resolution of Anaphora Procedure (RAP)</i>	Tarea de encontrar todas las expresiones que se refieren a la misma entidad en un texto (sustitución de pronombres por el sustantivo correspondiente). Entrada: “ <i>Juan likes Sonia. She likes him.</i> ” Salida: “ <i>Juan likes Sonia. Sonia likes Juan.</i> ”
Análisis de texto y análisis de datos exploratorios	Identificación de palabras compuestas (<i>Chunking</i>)	Identificación de palabras compuestas (<i>noun phrase</i> en inglés). Por ejemplo: Entrada: “ <i>South Africa is a country</i> ” Salida: “ <i>South Africa</i> ”
	Métodos de extracción de colocaciones y medidas de asociación léxica (<i>Collocation extraction methods and lexical association measure</i>)	Las colocaciones son combinaciones de palabras que ocurren juntas con mayor frecuencia de lo que se esperaría por casualidad. Las medidas de asociación léxica son fórmulas que determinan el grado de asociación entre los componentes de la colocación. Ellos calculan una puntuación de asociación(métrica) para cada candidato de colocación extraído de un corpus. Ejemplos de colocaciones en inglés: “ <i>to make the bed</i> ” “ <i>to do homework</i> ” “ <i>to take a risk</i> ”

		<i>“to give someone advice”</i>
Reconocimiento de intenciones (<i>Intention Recognition</i>)	La tarea de reconocer las intenciones de un agente analizando algunas o todas sus acciones y / o analizando los cambios en el estado (entorno) que resultan de sus acciones. Por ejemplo, las expresiones: “Buenos días”, “buenas tardes”, “¡qué tal!” y “¡hola!” se reconocen como saludos.	
Etiquetado POS (<i>Part Of Speech Tagging</i>)	Proceso de asignar una etiqueta sintáctica a cada token en una oración basado en su significado léxico. Ejemplo en inglés: Entrada: <i>“Santiago likes playing football”</i> Salida: <i>“Santiago” => NOUN</i> <i>“likes” => VERB</i> <i>“playing” => VERB</i> <i>“football” => NOUN</i>	
Reconocimiento de entidades nombradas (<i>Named Entity Recognition</i>)	Identificación predefinida de tipo de entidades en una oración. Por ejemplo: Entrada: <i>“When Michael Jordan was at the peak of his powers as an NBA superstar, his Chicago Bulls team were moving down the completion, winning six National Basketball Association.”</i> Salida: <i>“Chicago”</i> <i>“Michael Jordan”</i>	

		“National Basketball Association”
<p>Generación de árboles de dependencia (<i>Dependency Parsing</i>)</p>	<p>Construcción de un árbol sintáctico de la oración. Un árbol de análisis de circunscripción se divide el texto en subfrases. Los no terminales en el árbol son tipos de frases, los terminales son las palabras en la oración y los bordes no están etiquetados. Para una oración simple "John see Bill", un análisis de distrito electoral es:</p>	<pre> graph TD sees[sees] --- subject[subject] sees --- object[object] subject --- John[John] object --- Bill[Bill] </pre>
<p>Generación de árboles de constituyentes (<i>Constituency Parsing</i>)</p>	<p>Construcción de un árbol sintáctico de la oración. Un análisis de dependencia conecta las palabras de acuerdo con sus relaciones. Cada vértice en el árbol representa una palabra, los nodos secundarios son palabras que dependen del padre y los bordes están marcados por la relación. Un análisis de dependencia de "John see Bill", se muestra a continuación:</p>	<pre> graph TD Sentence[Sentence] --- NP1[Noun Phrase] Sentence --- VP[Verb Phrase] NP1 --- John[John] VP --- V[Verb] VP --- NP2[Noun Phrase] V --- sees[sees] NP2 --- Bill[Bill] </pre>
<p>Etiquetado de Rol Semántico (<i>Semantic Role Labelling</i>)</p>	<p>Asigna etiquetas frases en una oración que indican su rol (agente, predicado, tema y ubicación) semántico en una oración. Por ejemplo:</p> <p>Entrada: “The police officer detained the suspect at the scene of the crime”</p> <p>Salida:</p>	

		<p>“The police officer” => AGENT</p> <p>“detained” => PREDICATE</p> <p>“the suspect” => THEME</p> <p>“at the scene of the crime” => LOCATION</p>
	Expresiones regulares (<i>Regular Expression</i>)	<p>Secuencia de caracteres que forma un patrón de búsqueda, utilizada para la búsqueda de patrones de cadenas de caracteres. Específicamente, las expresiones regulares se construyen utilizando los operadores unión, concatenación y clausura de Kleene.</p> <p>Por ejemplo, "ho+la" describe el conjunto infinito hola, hoola, hoolola, hooloolola y así sucesivamente.</p>
	Reglas JAPE (<i>Java Annotation Pattern Engine Rules</i>)	<p>Conjunto de fases donde cada uno consiste en un conjunto de reglas de patrones y acciones. Para más información consúltese el manual en https://gate.ac.uk/sale/thakker-jape-tutorial/GATE%20JAPE%20manual.pdf</p>
Representación de texto e ingeniería de características	<p>Bolsa de palabras (<i>Bag of Words</i>):</p> <ul style="list-style-type: none"> • Frecuencia de palabras (<i>Word Frequency</i>) • Modelo n-gramas (<i>N-grams models</i>) • Frecuencia de termino-frecuencia de documento 	<p>Son métodos que se utilizan para el procesamiento de lenguaje natural para representar documentos donde no se toma en cuenta el orden de las palabras(gramática).</p> <p>Los métodos identificados son:</p> <ul style="list-style-type: none"> • <i>Word Frequency</i>: Vector de características de valores donde se indica la frecuencia en que aparece un término. Ejemplo de corpus: <p>Entrada: “<i>Tom is a lying cheater. Roger is incompetent and has a history of lying.</i>”</p>

	<p>inverso (<i>Term Frequency – Inverse Document Frequency, TF-IDF</i>)</p> <ul style="list-style-type: none"> • Frecuencia de termino normalizado (<i>Normalized Term Frequency</i>) 	<p>Salida:</p> <p>“Tom” => 1</p> <p>“is” => 2</p> <p>“a” => 2</p> <p>“lying” => 2</p> <p>“cheater” => 1</p> <p>“Roger” => 1</p> <p>“incompetent” => 1</p> <p>“and” => 1</p> <p>“has” => 1</p> <p>“history” => 1</p> <p>“of” => 1</p> <ul style="list-style-type: none"> • <i>N-grams Models</i>: Subsecuencia de n elementos de una secuencia dada. Por ejemplo: <p>Entrada: “<i>I want to eat Chinese food</i>”</p> <p>Salida:</p> <p>Bigrama => {“<i>I want</i>”, “<i>want to</i>”, “<i>to eat</i>”, “<i>eat Chinese</i>”, “<i>Chinese food</i>”}</p> <p>Trigrama => {“<i>I want to</i>”, “<i>want to eat</i>”, “<i>to eat Chinese</i>”, “<i>eat Chinese food</i>”}</p>
--	--	---

		<ul style="list-style-type: none"> • <i>Term Frequency – Inverse Document Frequency</i>: Métrica que expresa que tan relevante es una palabra en un documento. $TF \times IDF(t, DS) = tf(DS, t) \times \left(\log \left(\frac{ DS }{df(DS, t)} \right) + 1 \right)$ <p>Donde</p> <p>$tf(DS, t)$ es la frecuencia de t en todos los documentos DS</p> <p>DS es el número de documentos en DS</p> <p>$df(DS, t)$ es el número de documentos en DS que contiene t</p> <ul style="list-style-type: none"> • <i>Normalized Term Frequency</i>: Métrica que calcula la frecuencia de términos en un documento. $NTF(sd, d) = \frac{\sum_{i=1}^N tf(c_i, d)}{\text{Count}(d)}$ <p>Donde</p> <p>$\text{Count}(d)$ es el número de palabras en un documento d</p> <p>N es el número de conceptos en un subdominio</p> <p>c_i es el i-ésimo concepto en sd</p> <p>$tf(c_i, d)$ es la frecuencia de c_i en d</p>
	<p>Embebido de palabra (<i>Word Embedding</i>)</p> <ul style="list-style-type: none"> • Word2Vec 	<p>Son métodos donde las palabras, frases u oraciones son mapeadas a vectores de números reales, es decir, son un tipo de representación que permite a las palabras tener un significado similar. Ejemplos son los siguientes:</p>

	<ul style="list-style-type: none"> • Paragraph Vector 	<ul style="list-style-type: none"> • Word2Vec: Transforma una palabra a vector de números reales. Por ejemplo: "hello" => [0.5289,2.18794,3.1514] • Paragraph Vector: Transforma un párrafo en vector de números reales. Por ejemplo: I see the blue sky =>[1.483,-0.56,4.36]
<p>Modelado y extracción de patrones</p>	<p>Modelado de tópicos (<i>Topic Modelling</i>):</p> <ul style="list-style-type: none"> • Asignación de Dirichlet Latente (<i>Latent Dirichlet Allocation, LDA</i>) • Análisis Semántico Latente (<i>Latent Semantic Analysis, LSA</i>) 	<p>Tipo de modelado estadístico para descubrir "temas/tópicos" abstractos que aparecen en una colección de documentos. Por ejemplo, un conjunto de documentos con palabras: "monitor", "software", "CPU", "RAM", "sistema operativo" y "teclado" pueden referirse mismo tópico de "computación"</p> <ul style="list-style-type: none"> • <i>Latent Dirichlet Allocation(LDA)</i>: LDA es un modelo probabilístico generativo de un corpus(documento). La idea es identificar el tópico de un documento con base a una distribución de palabras relacionadas. • <i>Latent Semantic Analysis(LSA)</i>: Es una técnica en el procesamiento del lenguaje natural, en particular la semántica distributiva, de analizar las relaciones entre un conjunto de documentos y los términos que contienen al producir un conjunto de conceptos relacionados con los documentos y términos.
	<p>Medidas de Similitud (<i>Similarity Measure</i>):</p> <ul style="list-style-type: none"> • METEOR • Lexical Similarity 	<p>Son métricas utilizados en procesamiento de lenguaje natural para medir el significado (similitud semántica) entre palabras, frases o documentos de texto.</p>

	<ul style="list-style-type: none"> • BLEU • Corley Mihalcea • Simpson Coefficient 	
	<p>Redes Neuronales Artificiales (<i>Artificial Neuronal Networks, ANN</i>)</p>	<p>Modelo computacional inspirado en el comportamiento biológico de las neuronas. El objetivo de la red neuronal es resolver los problemas de la misma manera que el cerebro humano.</p>
	<p>Algoritmos de máquinas de aprendizaje (<i>Machine Learning algorithms</i>):</p> <ul style="list-style-type: none"> • J48 • Sequential Minimal Optimization (SMO) • Random Forest • K-Nearest Neighbors (K-NN) • Naive Bayes (NB) • Naive Bayes Multinomial (NBM) • Logistic Regression (LR) • Support Vector Machine (SVM) 	<p>El aprendizaje automático es el estudio científico de algoritmos y modelos estadísticos que los sistemas informáticos utilizan para realizar de manera efectiva una tarea específica sin utilizar instrucciones explícitas, sino que se basan en patrones e inferencias.</p>

	<ul style="list-style-type: none"> • Stochastic Gradient Descent (SGD) • Decision Tree (DT) • Conditional Random Field (CRF) 	
	Reglas Heurísticas (<i>Heuristic Rules</i>)	Son reglas que por experiencia nos permiten tomar decisiones en situaciones complicadas sin necesidad de hacer un número detallado de cálculos en el proceso, en el diseño, entre otros.
	Algoritmos (<i>Algorithms</i>)	Conjunto ordenado de operaciones sistemáticas que permite hacer un cálculo y hallar la solución de un tipo de problemas.
	Reglas de transformación / Reglas de Inferencia (<i>Transformation Rules / Inference Rules</i>)	En lógica, una regla de inferencia o regla de transformación es una forma lógica que consiste en una función que toma premisas, analiza su sintaxis, y devuelve una conclusión
	Alineación de Secuencias (<i>Sequence Alignment</i>)	La alineación de secuencias básicamente compara los elementos constituyentes de dos secuencias diferentes (también denominadas símbolos o caracteres en la jerga bioinformática) y tiene como objetivo encontrar una subcadena que maximice una función de similitud dada.
Evaluación y Despliegue	Precisión (<i>Precision</i>)	Es la relación de las observaciones positivas predichas correctamente al total de observaciones positivas predichas. Precision = TP/(TP+FP) Donde:

	<p>TP: Es True Positive (Verdadero Positivo)</p> <p>FP: Es False Positive(Falso Positivo)</p>
Cobertura (<i>Recall</i>)	<p><i>Recall</i> es la proporción de observaciones positivas predichas correctamente a todas las observaciones en la clase real.</p> <p>$Recall = TP/(TP+FN)$</p> <p>Donde:</p> <p>TP es True Positive (Verdadero Positivo)</p> <p>FP es False Positive(Falso Positivo)</p>
Exactitud (<i>Accuracy</i>)	<p>Nivel de comparación entre el valor predictivo y el valor actual de la extensión del valor.</p> <p>$Accuracy = (TP+TN)/(TP+TN+FP+FN)$</p> <p>Donde:</p> <p>TP es True Positive (Verdadero Positivo)</p> <p>TN: es True Negative (Verdadero Negativo)</p> <p>FP es False Positive (Falso Positivo)</p> <p>FN es False Negative (Falso Negativo)</p>
Medida F (<i>F-measure/F1 score</i>)	<p>Medida F es una medida de la precisión de una prueba. Considera tanto la <i>precision</i> como el <i>recall</i> de la prueba para calcular la puntuación. El puntaje de F1 es el promedio armónico de la precisión y el recuerdo, donde un puntaje de F1 alcanza su mejor valor en 1 (precisión perfecta y recordatorio) y el peor en 0.</p>

		$F = 2 \frac{\textit{precision} \times \textit{recall}}{\textit{precision} + \textit{recall}}$
Rango Recíproco Media (<i>Mean Reciprocal Rank, MRR</i>)	<p>MRR, es una medida estadística para evaluar cualquier proceso que produzca una lista de posibles respuestas a una muestra de consultas, ordenada por probabilidad de corrección.</p> $\text{MRP} = \frac{1}{ Q } \sum_{i=1}^{ Q } \frac{1}{\text{rank}_i}$ <p>Donde:</p> <p> Q es el tamaño del conjunto de la prueba.</p> <p>rank_i es la posición de rango de la respuesta correcta para la i-esima consulta.</p>	
Fleiss 'kappa	<p>Fleiss 'kappa es una medida estadística para evaluar la confiabilidad del acuerdo entre un número fijo de calificadores al asignar calificaciones categóricas a un número de elementos o elementos de clasificación.</p> $k = \frac{\tilde{p} - \tilde{p}e}{1 - \tilde{p}e}$ <p>Donde:</p> <p>$\tilde{p} - \tilde{p}e$ proporciona el grado de acuerdo que probablemente esté por encima del azar.</p> <p>$1 - \tilde{p}e$ proporciona el grado de acuerdo que se logra exactamente por encima del azar.</p>	

	<p>Precisión Promedio (<i>Average Precision, AP</i>)</p>	<p>AP es una medida de valor único que refleja el rendimiento de la recuperación de todos los documentos relevantes al promediar el valor de precisión obtenido después de recuperar cada documento relevante.</p> $AP = \frac{\sum_{k=1}^n P(k) \times Rel(k)}{\# \text{ de documentos relevantes}}$ <p>Donde:</p> <p>P(k) es la precisión del documento k</p> <p>Rel(k) indica la relevancia del documento k.</p>
	<p>Media de precisión promedio (<i>Mean Average Precision, MAP</i>)</p>	<p>MNP para un conjunto de consultas es la media de las puntuaciones <i>Average Precision</i> para cada consulta, como se indica en la ecuación siguiente.</p> $MAP = \frac{\sum_{q=1}^n AP(q)}{n}$ <p>Donde:</p> <p>AP(q) representa la precisión promedio del comparador para la consulta q y n representa el número de consultas.</p>

6.3 Resumen del capítulo

En este capítulo se presentaron algunos de los hallazgos importantes identificados en esta investigación. Se describe aspectos relevantes identificados en este estudio. Se presentó el esquema de clasificación de las técnicas de procesamiento de lenguaje natural derivado del análisis de los artículos seleccionados, que es una de las grandes contribuciones de este trabajo de tesis.

Capítulo 7.

Conclusiones

En este capítulo se presentan las conclusiones a las que se llegaron y las aportaciones obtenidas como resultado de esta investigación, además, se sugieren algunos trabajos futuros para dar continuidad a la investigación.

8.1 Conclusiones

Con base en los resultados presentados en el Capítulo 5 se concluye el cumplimiento del objetivo general, que se menciona en la sección 1.3.1 del Capítulo 1:

“Realizar un estudio bibliográfico (estado del arte) que brinde un panorama general acerca del reconocimiento del habla, procesamiento del lenguaje natural y uso de ontologías en apoyo al modelado del dominio del problema y la elicitación de requerimientos.”.

Durante el proceso de desarrollo del estudio de mapeo sistemático, se tuvieron los siguientes hallazgos:

1. Se identificaron diferentes enfoques para realizar distintas tareas en la elicitación de requerimientos y modelado del dominio del problema como son: el análisis de calidad (detección de incertidumbre, ambigüedad o detección de duplicidad de funcionalidad en los requerimientos de usuario), la clasificación de requerimientos (propósito de clasificar los requerimientos de un documento en lenguaje natural), la generación de modelos (como diagramas de casos de uso, diagramas de actividad, diagramas de proceso de negocios y mapas conceptuales), identificación de abstracción (identificar actores, tareas y términos del dominio del dominio del problema) e identificación de requerimientos (propósito de posibles requerimientos en un documento de lenguaje natural).
2. Respecto a las técnicas de “Modelado y extracción de patrones” (etapa del flujo de trabajo de NLP), se identificaron varias estrategias para realizar el análisis del lenguaje natural como: clasificación e identificación de requerimiento. Las técnicas identificadas son: modelado de tópicos, medidas de similitud, algoritmos, reglas de transformación (inferencia), de aprendizaje automático, redes neuronales, alineación de secuencias, y reglas heurísticas. Sin embargo, en los últimos tres años (2016-2018) se han presentado una tendencia de utilización de aprendizaje automático y redes neuronales para dichas tareas.
3. El resultado de este estudio provee un panorama general actualizado de trabajos de investigación en esta área. Esta información resumida aportará a investigadores de la misma área (elicitación de requerimientos y modelado del dominio del problema) a analizar las diferentes estrategias existentes y, de esta manera, seguir contribuyendo a mejorando las etapas tempranas del desarrollo de análisis.
4. Una de las principales contribuciones de este estudio es la clasificación de las técnicas de procesamiento de lenguaje natural. Esta clasificación provee un panorama general de todas las técnicas NLP identificadas en este estudio, con el propósito de ver las distintas estrategias existentes para apoyar la automatización, no solo en la elicitación de requerimientos, sino también en otras áreas de investigación.

8.2 Aportaciones

Las principales aportaciones de este trabajo de tesis son:

1. Un mapa (panorama visual) de la evidencia bibliográfica actual, existente acerca del uso de reconocimiento del habla, Procesamiento del Lenguaje Natural y ontología en la elicitación de requerimientos y/o dominio del problema (Capítulo 5). Un estudio de mapeo sistemático proporciona una estructura para el informe de investigación, que incluye resultados que se han publicado, categorizados y presentados en un resumen visual.
2. La clasificación de técnicas utilizadas en el flujo de trabajo del Procesamiento de Lenguaje Natural (Sección 6.2). Esta clasificación fue de acuerdo al flujo de trabajo del procesamiento de lenguaje natural definida en (Sarkar, 2018) que cuenta con cinco fases (Pre-procesamiento de texto, Análisis de texto y análisis de datos exploratorios, Representación de texto e ingeniería de características, Modelado y extracción de patrones, por último, Evaluación y despliegue), pero en este trabajo de tesis se agregó una fase más (reconocimiento del habla). Cada técnica identificada se categoriza en una fase de este esquema.

8.3 Trabajos futuros

Para trabajos futuros se propone:

1. Investigar a fondo las técnicas de reconocimiento del habla (*Hidden Markov Model*, *Neuronal Network Modelling*, *Dynamic Time Warping*, *Vector Quantization* y *Support Vector Machine*) y las herramientas que se utilizan en la actualidad para su implementación (software o API que permita utilizar estas técnicas para el reconocimiento del habla).
2. Revisar la posible existencia de más técnicas de Procesamiento de Lenguaje Natural para completar el esquema de clasificación, así como las distintas herramientas para el NLP (software o API especializado en NLP), debido a que el análisis de este estudio sólo considero cuatro bases de datos científicas.
3. Realizar una comparación y evaluación de las distintas técnicas de Procesamiento de Lenguaje Natural. Por ejemplo, evaluar los enfoques que utilicen modelado de tópicos, medidas de similitud, ANN, ML, reglas heurísticas, algoritmos, reglas de transformación y alineación de secuencias. Esto con el propósito de evaluar las ventajas y desventajas de cada una.
4. Llevar a cabo la investigación e implementación del reconocimiento del habla para representar el dominio del problema en la etapa de análisis debido a la poca evidencia existente. Es decir, que, a partir del habla (voz), se identifiquen conceptos y relaciones o se construyan modelos (proceso de negocio, ontologías o modelos de requerimientos) para contribuir en el entendimiento del dominio del problema.

Referencias

- Abad, Z. S. H., Gervasi, V., Zowghi, D., & Barker, K. (2018). ELICA: An Automated Tool for Dynamic Extraction of Requirements Relevant Information. *Fifth International Workshop on Artificial Intelligence for Requirements Engineering Requirements Engineering (AIRE'18)*, (1). <https://doi.org/10.1109/AIRE.2018.00007>
- Abdouli, M., Karaa, W. B. A., & Ghezala, H. Ben. (2016). Survey of works that transform requirements into UML diagrams. *2016 IEEE/ACIS 14th International Conference on Software Engineering Research, Management and Applications, SERA 2016*, 117–123. <https://doi.org/10.1109/SERA.2016.7516136>
- ACM Computing Surveys. (n.d.). <https://csur.acm.org/>.
- Ajit Singh, Taruna Panchal, M. S. (2013). A Review of Automatic Speaker Recognition System. *International Journal of Soft Computing and Engineering*, 3(2), 350–354.
- Al-Hroob, A., Imam, A. T., & Al-Heisa, R. (2018). The use of artificial neural networks for extracting actions and actors from requirements document. *Information and Software Technology*, 101(March), 1–15. <https://doi.org/10.1016/j.infsof.2018.04.010>
- Alabdulkareem, F., Cercone, N., & Liaskos, S. (2015). Goal and Preference Identification through natural language. *2015 IEEE 23rd International Requirements Engineering Conference, RE 2015 - Proceedings*, 56–65. <https://doi.org/10.1109/RE.2015.7320408>
- Alasadi, A. A. A., & Deshmukh, R. R. (2018). Automatic Speech Recognition: A Review. *Signal Processing and Computer Vision*, 464–470. <https://doi.org/10.5120/9722-4190>
- Azzazi, A. (2017). *A Framework using NLP to automatically convert User-Stories into Use Cases in Software Projects*. 17(5), 71–76.
- Bargui, F., Ben-Abdallah, H., & Feki, J. (2011). A decision-making ontology for analytical requirements elicitation. *Computer and Information Sciences II - 26th International Symposium on Computer and Information Sciences, ISCIS 2011*, 495–501. <https://doi.org/10.1007/978-1-4471-2155-8-63>
- Barros-Justo, J. L., Benitti, F. B. V., & Cravero-Leal, A. L. (2017). Software patterns and requirements engineering activities in real-world settings: A systematic mapping study. *Computer Standards and Interfaces*, 000(August), 1–20. <https://doi.org/10.1016/j.csi.2017.12.001>
- Barros-Justo, J. L., Cravero-Leal, A. L., Benitti, F. B. V., & Capilla-Sevilla, R. (2017). *Systematic Mappig Protocol*. Retrieved from <https://arxiv.org/ftp/arxiv/papers/1701/1701.05747.pdf>
- Borst, W. N. (1997). Construction of engineering ontologies for knowledge sharing and reuse. In *Technology*. <https://doi.org/10.1006/ijhc.1996.0096>
- Canada, S. (2010). Survey Methods and Practices. In *Statistics Canada*. <https://doi.org/12-587-X>
- Candia Zavaleta, G. (2017). *Almacenamiento y manipulación de ontología que describen patrones de análisis y su semántica*. Centro Nacional de Investigación y Desarrollo Tecnológico.

- Casagrande, E., Woldeamlak, S., Woon, W. L., Zeineldin, H. H., & Svetinovic, D. (2014). NLP-KAOS for Systems Goal Elicitation: Smart Metering System Case Study. *Ieee Transactions on Software Engineering*, 40(10), 941–956. <https://doi.org/10.1109/TSE.2014.2339811>
- Castañeda, V., Ballejos, L., Caliusco, M. L., & Galli, M. R. (2010). The Use of Ontologies in Requirements Engineering. *Global Journal of Researches in Engineering*, 10(6), 2–8. Retrieved from <http://engineeringresearch.org/index.php/GJRE/article/view/76>
- Chen, H., He, K., Liang, P., & Li, R. (2010). Text-based requirements preprocessing using nature language processing techniques. *2010 International Conference on Computer Design and Applications, ICCDA 2010, 1(Iccda)*. <https://doi.org/10.1109/ICCDA.2010.5540935>
- Chen, X., Yin, B., & Jin, Z. (2010). DPTool: A tool for supporting the problem description and projection. *Proceedings of the 2010 18th IEEE International Requirements Engineering Conference, RE2010*, 401–402. <https://doi.org/10.1109/RE.2010.58>
- Choe, H., & Moon, S. (2011). A methodology for accurate and redundancy-free business requirements description using ontology. *2011 International Conference on Information Science and Applications, ICISA 2011*. <https://doi.org/10.1109/ICISA.2011.5772352>
- Daramola, O., Sindre, G., & Stalhane, T. (2012). Pattern-based security requirements specification using ontologies and boilerplates. *2012 2nd IEEE International Workshop on Requirements Patterns, RePa 2012 - Proceedings*, 54–59. <https://doi.org/10.1109/RePa.2012.6359973>
- Deeptimahanti, D. K., & Sanyal, R. (2011). Semi-automatic generation of UML models from natural language requirements. *Proceedings of the 4th India Software Engineering Conference on - ISEC '11*, 165–174. <https://doi.org/10.1145/1953355.1953378>
- Desarrollo de Marco de la junta de Andalucía. (2019). Estudiar el dominio del problema. Retrieved from www.juntadeandalucia.es website:
<http://www.juntadeandalucia.es/servicios/madeja/contenido/libro-pautas/175#libro-pautas-toc>
- Dimanidis, A., Chatzidimitriou, K. C., & Symeonidis, A. L. (2018). A Natural Language Driven Approach for Automated Web API Development. *Companion of the The Web Conference 2018 on The Web Conference 2018 - WWW '18*, 1869–1874. <https://doi.org/10.1145/3184558.3191654>
- Dutta, A., Bhattacharjee, S., Kanjilal, A., Dasgupta, R., & Bhattacharya, S. (2013). Engineering of requirements for a distributed teleteaching system. *ACM SIGSOFT Software Engineering Notes*, 38(6), 1–12. <https://doi.org/10.1145/2532780.2532814>
- Elamin, E., Alshomrani, S., & Feki, J. (2017). SSReq: A method for designing Star Schemas from decisional requirements. *Proceedings - 2017 International Conference on Communication, Control, Computing and Electronics Engineering, ICCCCCEE 2017*. <https://doi.org/10.1109/ICCCCEE.2017.7867645>
- Farfeleder, S., Moser, T., Krall, A., Ståalhane, T., Omoronyia, I., & Zojer, H. (2011). Ontology-driven guidance for requirements elicitation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6643 LNCS(PART 2), 212–226. https://doi.org/10.1007/978-3-642-21064-8_15

- Farfeleder, S., Moser, T., Krall, A., Stålhane, T., Zojer, H., & Panis, C. (2011). DODT: Increasing requirements formalism using domain ontologies for improved embedded systems development. *Proceedings of the 2011 IEEE Symposium on Design and Diagnostics of Electronic Circuits and Systems, DDECS 2011*, 271–274. <https://doi.org/10.1109/DDECS.2011.5783092>
- Ferrari, A., Esuli, A., & Gnesi, S. (2018). Identification of Cross-domain Ambiguity with Language Models. *2018 5th International Workshop on Artificial Intelligence for Requirements Engineering*. <https://doi.org/10.1109/AIRE.2018.00011>
- Gaikwad, S. K., Gawali, B. W., & Yannawar, P. (2010). A Review on Speech Recognition Technique. *International Journal of Computer Applications*, 10(3), 16–24. <https://doi.org/10.5120/1462-1976>
- Galindo García, J. A. (2011). *Generación de ontologías en OWL a partir de diagramas de clases de UML mediante MDA (Arquitectura Dirigida por Modelos)*. Centro Nacional de Investigación y Desarrollo Tecnológico.
- Gemkow, T., Conzelmann, M., Hartig, K., & Vogelsang, A. (2018). Automatic Glossary Term Extraction from Large-Scale Requirements Specifications. *2018 IEEE 26th International Requirements Engineering Conference Automatic*, 412–417. <https://doi.org/10.1109/RE.2018.00052>
- Gruber, T. (2009). <http://tomgruber.org/writing/ontology-definition-2007.htm>.
- Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2), 199–220. <https://doi.org/10.1006/knac.1993.1008>
- Guo, H., Kafal, Ö., & Singh, M. P. (2018). Extraction of Natural Language Requirements from Breach Reports Using Event Inference. *Fifth International Workshop on Artificial Intelligence for Requirements Engineering*. <https://doi.org/10.1109/AIRE.2018.00009>
- Hajri, I., Goknil, A., Briand, L. C., & Stephany, T. (2015). Applying product line Use case modeling in an industrial automotive embedded system: Lessons learned and a refined approach. *2015 ACM/IEEE 18th International Conference on Model Driven Engineering Languages and Systems, MODELS 2015 - Proceedings*, 338–347. <https://doi.org/10.1109/MODELS.2015.7338265>
- Hajri, I., Goknil, A., Briand, L. C., & Stephany, T. (2016). Configuring use case models in product families. *Software and Systems Modeling*, 17(3), 939–971. <https://doi.org/10.1007/s10270-016-0539-8>
- Harmain, H. M., & Gaizauskas, R. (2000). CM-Builder: An automated NL-based CASE tool. *Proceedings ASE 2000: 15th IEEE International Conference on Automated Software Engineering*, (May 2014), 45–53. <https://doi.org/10.1109/ASE.2000.873649>
- Hull, E., Dick, J., Jackson, K., Dick, J., Hull, E., & Jackson, K. (2017). Requirements Engineering in the Solution Domain. *Requirements Engineering*, 135–158. https://doi.org/10.1007/978-3-319-61073-3_6
- Jain, R., Ghaisas, S., & Sureka, A. (2014). SANAYOJAN: a framework for traceability link recovery between use-cases in software requirement specification and regulatory documents. *Proceedings of the 3rd International Workshop on Realizing Artificial Intelligence Synergies in Software*

- Engineering - RAISE 2014*, 12–18. <https://doi.org/10.1145/2593801.2593804>
- Jebril, E. M. (2018). An Algorithmic Approach to Extract Actions and Actors. *Proceedings of the International Conference on Geoinformatics and Data Analysis ICGDA '18*, 13–17.
- Kaiya, H., Shimizu, Y., Yasui, H., Kaijiri, K., & Saeki, M. (2010). Enhancing domain knowledge for requirements elicitation with Web mining. *Proceedings - Asia-Pacific Software Engineering Conference, APSEC*, 3–12. <https://doi.org/10.1109/APSEC.2010.11>
- Kifle, M. (2012). Multi-perspective ontology to understand organizational requirements. *Proceedings - African Conference on Software Engineering and Applied Computing, ACSEAC 2012*, 67–74. <https://doi.org/10.1109/ACSEAC.2012.16>
- Kitchenham, B. (2004). Procedures for performing systematic reviews. *Keele, UK, Keele University*, 33(TR/SE-0401), 28. <https://doi.org/10.1.1.122.3308>
- Kitchenham, B., & Charters, S. (2007). Guidelines for performing Systematic Literature Reviews in Software Engineering. *Engineering*, 2, 1051. <https://doi.org/10.1145/1134285.1134500>
- Kof, L., & Garching, M. (2010). Scenario Analysis : Generation of Possible Scenario Interpretations and their. *2009 Fourth International Workshop on Requirements Engineering Visualization (Rev'09)*.
- Kundi, M., & Chitchyan, R. (2017). Use case elicitation with FrameNet frames. *Proceedings - 2017 IEEE 25th International Requirements Engineering Conference Workshops, REW 2017*, (section 2), 224–231. <https://doi.org/10.1109/REW.2017.53>
- Li, C., Huang, L., Ge, J., Luo, B., & Ng, V. (2018). Automatically classifying user requests in crowdsourcing requirements engineering. *Journal of Systems and Software*, 138, 108–123. <https://doi.org/10.1016/j.jss.2017.12.028>
- Li, R., Ma, S., & Yao, W. (2015). Ontology-based requirements generation for credibility validation of safetycritical system. *Proceedings - 15th IEEE International Conference on Computer and Information Technology, CIT 2015, 14th IEEE International Conference on Ubiquitous Computing and Communications, IUCC 2015, 13th IEEE International Conference on Dependable, Autonomic and Se*, 849–854. <https://doi.org/10.1109/CIT/IUCC/DASC/PICOM.2015.126>
- Li, Y., Guzman, E., Tsiamoura, K., Schneider, F., & Bruegge, B. (2015). Automated requirements extraction for scientific software. *Procedia Computer Science*, 51(1), 582–591. <https://doi.org/10.1016/j.procs.2015.05.326>
- Liang, P., Avgeriou, P., He, K., & Xu, L. (2010). From collective knowledge to intelligence: Pre-Requirements Analysis of Large and Complex Systems. *Proceedings of the 1st Workshop on Web 2.0 for Software Engineering - Web2SE '10*, 26. <https://doi.org/10.1145/1809198.1809208>
- Mazumdar, S. (2013). How to Write a Good Survey Paper. *Journal of Chemical Information and Modeling*, 53(9), 1689–1699. <https://doi.org/10.1017/CBO9781107415324.004>
- Merten, T., Falis, M., Hübner, P., Quirchmayr, T., Bürsner, S., & Paech, B. (2016). Software Feature Request Detection in Issue Tracking Systems. *Proceedings - 2016 IEEE 24th International Requirements Engineering Conference, RE 2016*, 166–175. <https://doi.org/10.1109/RE.2016.8>

- Meth, H., Brhel, M., & Maedche, A. (2013). The state of the art in automated requirements elicitation. *Information and Software Technology*, 55(10), 1695–1709. <https://doi.org/10.1016/j.infsof.2013.03.008>
- Meth, H., Mädche, A., & Einöder, M. (2012). Exploring design principles of task elicitation systems for unrestricted natural language documents. *Proceedings of the 4th ACM SIGCHI Symposium on Engineering Interactive Computing Systems*, 205–210. <https://doi.org/10.1145/2305484.2305518>
- Miah, S. J., Islam, H., & Samsudin, A. Z. H. (2016). Ontology techniques for representing the problem of discourse: Design of solution application perspective. *Proceedings - 2016 16th IEEE International Conference on Computer and Information Technology, CIT 2016, 2016 6th International Symposium on Cloud and Service Computing, IEEE SC2 2016 and 2016 International Symposium on Security and Privacy in Social Netwo*, 148–153. <https://doi.org/10.1109/CIT.2016.115>
- Millet, S., & Tune, N. (2015). *Patterns, Principles, and Practices of Domain-Driven Design*.
- Mohamed, K. A., Ellatif, M. A., & Farhan, M. S. (2017). Using ontology-based concept maps for requirements engineering: A case study. *ICENCO 2017 - 13th International Computer Engineering Conference: Boundless Smart Societies, 2018-Janua*, 366–371. <https://doi.org/10.1109/ICENCO.2017.8289816>
- Mohanan, M., & Philip, S. (2016). Software Requirement Elicitation Using Natural Language Processing. *Springer*, 424, 197–208. <https://doi.org/10.1007/978-3-319-28031-8>
- Morales-Ramirez, I., Kifetew, F. M., & Perini, A. (2018). Speech-acts based analysis for requirements discovery from online discussions. *Information Systems*, 0, 1–17. <https://doi.org/10.1016/j.is.2018.08.003>
- Nazir, F., Butt, W. H., Anwar, M. W., & Khattak, M. A. K. (2017). The Applications of Natural Language Processing (NLP) for Software Requirement Engineering - A Systematic Literature Review. *Information Science and Applications*, 424. <https://doi.org/10.1007/978-981-10-4154-9>
- Networking and Information Technology Research and Development. (2011). *Future of Software Engineering Research*. (December), 65.
- Nguyen, T. H., Grundy, J., & Almorsy, M. (2014). GUITAR: An ontology-based automated requirements analysis tool. *2014 IEEE 22nd International Requirements Engineering Conference, RE 2014 - Proceedings*, 315–316. <https://doi.org/10.1109/RE.2014.6912274>
- Nisha. (2017). Voice Recognition Technique : A Review. *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, 5(V), 262–268.
- Noaen, M., Abad, Z. S. H., & Far, B. H. (2017). Let's Hear it from RETTA: A Requirements Elicitation Tool for TrAffic Management Systems. *Proceedings - 2017 IEEE 25th International Requirements Engineering Conference, RE 2017*, 450–451. <https://doi.org/10.1109/RE.2017.78>
- Omoronyia, I., Sindre, G., Stålhane, T., Biffel, S., Moser, T., & Sunindyo, W. (2010). A domain ontology building process for guiding requirements elicitation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*,

- 6182 LNCS, 188–202. https://doi.org/10.1007/978-3-642-14192-8_18
- Onyeka, E. (2013). A process framework for managing implicit requirements using analogy-based reasoning: Doctoral consortium paper. *Proceedings - International Conference on Research Challenges in Information Science*. <https://doi.org/10.1109/RCIS.2013.6577726>
- Petersen, K., Feldt, R., Mujtaba, S., & Mattsson, M. (2008). Systematic Mapping Studies in Software Engineering. *12Th International Conference on Evaluation and Assessment in Software Engineering*, 17, 10. <https://doi.org/10.1142/S0218194007003112>
- Petersen, K., Vakkalanka, S., & Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology*, 64, 1–18. <https://doi.org/10.1016/j.infsof.2015.03.007>
- Portugal, R. L. Q., Li, T., Silva, L., Almentero, E., & do Prado Leite, J. C. S. (2018). NFRfinder: A Knowledge Based Strategy for Mining Non-functional Requirements. *Proceedings of the XXXII Brazilian Symposium on Software Engineering*, 102–111. <https://doi.org/10.1145/3266237.3266269>
- Provenzano, L., Hanninen, K., Zhou, J., & Lundqvist, K. (2017). An Ontological Approach to Elicit Safety Requirements. *2017 24th Asia-Pacific Software Engineering Conference (APSEC)*, 5, 713–718. <https://doi.org/10.1109/APSEC.2017.91>
- Rago, A., Marcos, C., & Diaz-Pace, J. A. (2016). Identifying duplicate functionality in textual use cases by aligning semantic actions. *Software and Systems Modeling*, 15(2), 579–603. <https://doi.org/10.1007/s10270-014-0431-3>
- Robeer, M., Lucassen, G., Van Der Werf, J. M. E. M., Dalpiaz, F., & Brinkkemper, S. (2016). Automated Extraction of Conceptual Models from User Stories via NLP. *Proceedings - 2016 IEEE 24th International Requirements Engineering Conference, RE 2016*, 196–205. <https://doi.org/10.1109/RE.2016.40>
- Rodríguez Correa, S., & Benavides Cañón, P. A. (2007). Procesamiento Del Lenguaje Natural En La Recuperación De Información. In *Procesamiento Del Lenguaje Natural En La Recuperación De Información*. Retrieved from http://eprints.rclis.org/9598/1/PROCESAMIENTO_DEL LENGUAJE NATURAL EN LA RECUPERACION_DE INFORMACION.pdf
- Romero, O., & Abelló, A. (2010). A framework for multidimensional design of data warehouses from ontologies. *Data and Knowledge Engineering*, 69(11), 1138–1157. <https://doi.org/10.1016/j.datak.2010.07.007>
- Sadoun, D., Dubois, C., Ghamri-Doudane, Y., & Grau, B. (2014). Formal rule representation and verification from natural language requirements using an ontology. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8620 LNCS(October), 226–235. https://doi.org/10.1007/978-3-319-09870-8_17
- Saeki, M., Hayashi, S., & Kaiya, H. (2010). An integrated support for attributed goal-oriented requirements analysis method and its implementation. *Proceedings - International Conference on Quality Software*, 357–360. <https://doi.org/10.1109/QSIC.2010.19>

- Salini, P., & Kanmani, S. (2016). A novel method: Ontology-based security requirements engineering framework. *1st International Conference on Emerging Trends in Engineering, Technology and Science, ICETETS 2016 - Proceedings*. <https://doi.org/10.1109/ICETETS.2016.7602982>
- Sánchez Pacheco, E. H. (2008). *Generación de Modelos Conceptuales a partir de Modelos de Requisitos: un Enfoque Basado en Tratamiento Automático del Lenguaje Natural*. Centro Nacional de Investigación y Desarrollo Tecnológico.
- Sarkar, D. (DJ). (2018). A Practitioner's Guide to Natural Language Processing (Part I) — Processing & Understanding Text. Retrieved from Towards Data Science website: <https://towardsdatascience.com/a-practitioners-guide-to-natural-language-processing-part-i-processing-understanding-text-9f4abfd13e72>
- Sawant, K. P., Roy, S., Sripathi, S., Plesse, F., & Sajeev, A. S. M. (2014). Deriving requirements model from textual use cases. *Companion Proceedings of the 36th International Conference on Software Engineering - ICSE Companion 2014*, 235–244. <https://doi.org/10.1145/2591062.2591193>
- Sharma, A., & Kushwaha, D. S. (2011). Natural language based component extraction from requirement engineering document and its complexity analysis. *ACM SIGSOFT Software Engineering Notes*, 36(1), 1. <https://doi.org/10.1145/1921532.1921547>
- Sim, W. W., & Brouse, P. (2015). Developing ontologies and persona to support and enhance requirements engineering activities-A case study. *Procedia Computer Science*, 44(C), 275–284. <https://doi.org/10.1016/j.procs.2015.03.060>
- Sitthithanasakul, S., & Choosri, N. (2017). Using ontology to enhance requirement engineering in agile software process. *SKIMA 2016 - 2016 10th International Conference on Software, Knowledge, Information Management and Applications*, 181–186. <https://doi.org/10.1109/SKIMA.2016.7916218>
- Soares, F., & Wanderley, F. (2015). VoiceToModel : An Approach to Generate Requirements Models from Speech Recognition Mechanisms. *Proceedings of the 30th Annual ACM Symposium on Applied Computing SAC '15*, 1350–1357.
- Souag, A., Salinesi, C., Wattiau, I., & Mouratidis, H. (2013). Using security and domain ontologies for security requirements analysis. *Proceedings - International Computer Software and Applications Conference*, 101–107. <https://doi.org/10.1109/COMPSACW.2013.124>
- Sousa, A., Couto, T., Agra, C., & Alencar, F. (2016). Use of Ontologies in Embedded Systems: A Systematic Mapping. *2016 10th International Conference on the Quality of Information and Communications Technology (QUATIC)*, 1–8. <https://doi.org/10.1109/QUATIC.2016.011>
- Sree-kumar, A., Planas, E., & Clarisó, R. (2018). Extracting Software Product Line Feature Models from Natural Language Specifications. *Proceedings of the 22nd International Systems and Software Product Line Conference - Volume 1 SPLC '18*.
- Studer, R., Benjamins, V. R., & Fensel, D. (1998). Knowledge engineering: Principles and methods. *Data & Knowledge Engineering*, 25(1–2), 161–197. [https://doi.org/10.1016/S0169-023X\(97\)00056-6](https://doi.org/10.1016/S0169-023X(97)00056-6)

- Sun, L., & Mushi, C. J. (2010). Case-based analysis in user requirements modelling for knowledge construction. *Information and Software Technology*, 52(7), 770–777. <https://doi.org/10.1016/j.infsof.2010.03.010>
- Tiwari, S., & Laddha, M. (2017). UCAalyzer: A Tool to Analyze Use Case Textual Descriptions. *Proceedings - 2017 IEEE 25th International Requirements Engineering Conference, RE 2017*, 448–449. <https://doi.org/10.1109/RE.2017.39>
- Tuz Zahra, F., Rahim, H., & Iqbal, S. (2016). Sematic analysis of business rules using discourse representation theory. *2016 6th International Conference on Innovative Computing Technology, INTECH 2016*, 210–215. <https://doi.org/10.1109/INTECH.2016.7845068>
- Umoh, E., Sampaio, P. R. F., & Theodoulidis, B. (2011). REFINTO: An ontology-based requirements engineering framework for buisness-IT alignment in financial services organizations. *Proceedings - 2011 IEEE International Conference on Services Computing, SCC 2011*, 600–607. <https://doi.org/10.1109/SCC.2011.104>
- Van Rooijen, L., Bäumer, F. S., Platenius, M. C., Geierhos, M., Hamann, H., & Engels, G. (2017). From user demand to software service: Using machine learning to automate the requirements specification process. *Proceedings - 2017 IEEE 25th International Requirements Engineering Conference Workshops, REW 2017*, 379–385. <https://doi.org/10.1109/REW.2017.26>
- Vázquez Vásquez, C. M. (2010). *Traductor de Consultas en Lenguaje Natural a SPARQL para Realizar Búsquedas sobre Ontologías*. Centro Nacional de Investigación y Desarrollo Tecnológico.
- Vega Villalobos, A. (2017). *Revisión del estado del arte de los algoritmos K-means y sus Mejoras* (Centro Nacional de Investigación y Desarrollo Tecnológico). Retrieved from https://www.cenidet.edu.mx/subaca/web-elec/tesis_mc/87MC_jlil.pdf
- Wang, T., Si, Y., Xuan, X., Wang, X., Yang, X., Li, S., & Kav, A. J. (2010). A QoS ontology cooperated with feature models for non-functional requirements elicitation. *Proceedings of the Second Asia-Pacific Symposium on Internetware - Internetware '10*, 1–4. <https://doi.org/10.1145/2020723.2020740>
- Wieringa, R., Maiden, N., Mead, N., & Rolland, C. (2006). Requirements engineering paper classification and evaluation criteria: A proposal and a discussion. *Requirements Engineering*, 11(1), 102–107. <https://doi.org/10.1007/s00766-005-0021-6>
- Williams, I. (2018). An Ontology Based Collaborative Recommender System for Security Requirements Elicitation. *26th International Requirements Engineering Conference*, 448–453. <https://doi.org/10.1109/RE.2018.00060>
- Yang, H., De Roeck, A., Gervasi, V., Willis, A., & Nuseibeh, B. (2012). Speculative requirements: Automatic detection of uncertainty in natural language requirements. *2012 20th IEEE International Requirements Engineering Conference, RE 2012 - Proceedings*, 11–20. <https://doi.org/10.1109/RE.2012.6345795>
- Yuan, X., & Tripathi, S. (2016). An approach of dynamically combining ontologies for interactive Requirements Elicitation. *Proceedings of the IEEE International Conference on Software Engineering and Service Sciences, ICSESS*, 11–16. <https://doi.org/10.1109/ICSESS.2016.7883006>

- Yuan, X., & Zhang, X. (2015). An ontology-based requirement modeling for interactive software customization. *5th International Model-Driven Requirements Engineering Workshop, MoDRE 2015 - Proceedings*, 21–30. <https://doi.org/10.1109/MoDRE.2015.7343872>
- Yue, T., Briand, L. C., & Labiche, Y. (2010). An Automated Approach to Transform Use Cases into Activity Diagrams. *Modelling Foundations and Applications*, 337–353. https://doi.org/10.1007/978-3-642-13595-8_26
- Zhang, T., & Lee, B. (2010). Complementary classification techniques based personalized software requirements retrieval with semantic ontology and user feedback. *Proceedings - 10th IEEE International Conference on Computer and Information Technology, CIT-2010, 7th IEEE International Conference on Embedded Software and Systems, ICESS-2010, ScalCom-2010, (Cit)*, 1358–1363. <https://doi.org/10.1109/CIT.2010.243>

Anexos

Anexo A) Lista de artículos seleccionados para este estudio de mapeo sistemático.

A continuación, se muestra la lista de artículos seleccionados por este estudio de mapeo sistemático. La Tabla A-1 muestra los artículos con un identificador y la referencia del trabajo de investigación.

Tabla A-1. Referencia de artículos seleccionados para este estudio.

ID	TÍTULO	BASE DE DATOS	AÑO	ENFOQUE	REF.
ART01	The use of artificial neural networks for extracting actions and actors from requirements document	ScienceDirect	2018	NLP	(Al-Hroob, Imam, & Al-Heisa, 2018)
ART02	Speech-acts based analysis for requirements discovery from online discussions	ScienceDirect	2018	NLP	(Morales-Ramirez, Kifetew, & Perini, 2018)
ART03	Automatically classifying user requests in crowdsourcing requirements engineering	ScienceDirect	2018	NLP	(C. Li, Huang, Ge, Luo, & Ng, 2018)
ART04	NFRFinder: A Knowledge Based Strategy for Mining Non-Functional Requirements NFRFinder: A Knowledge Based Strategy for Mining Non-Functional	ACM Library	2018	NLP	(Portugal, Li, Silva, Almentero, & do Prado Leite, 2018)
ART05	Extracting Software Product Line Feature Models from Natural Language Specifications	ACM Library	2018	NLP	(Sree-kumar, Planas, & Clarisó, 2018)
ART06	A Natural Language Driven Approach for Automated Web API Development	ACM Library	2018	NLP	(Dimanidis, Chatzidimitriou, & Symeonidis, 2018)
ART07	An Algorithmic Approach to Extract Actions and Actors	ACM Library	2018	NLP	(Jebri, 2018)
ART08	ELICA: An Automated Tool for Dynamic Extraction of Requirements Relevant Information	IEEE Xplore	2018	Reconocimiento del habla	(Abad, Gervasi, Zowghi, & Barker, 2018)

ART09	Extraction of Natural Language Requirements from Breach Reports Using Event Inference	IEEE Xplore	2018	NLP	(Guo, Kafal, & Singh, 2018)
ART10	Identification of Cross-domain Ambiguity with Language Models	IEEE Xplore	2018	NLP	(Ferrari, Esuli, & Gnesi, 2018)
ART11	Automatic Glossary Term Extraction from Large-Scale Requirements Specifications	IEEE Xplore	2018	NLP	(Gemkow, Conzelmann, Hartig, & Vogelsang, 2018)
ART12	An Ontology Based Collaborative Recommender System for Security Requirements Elicitation	IEEE Xplore	2018	NLP, Ontología	(Williams, 2018)
ART13	An Ontological Approach to Elicit Safety Requirements	IEEE Xplore	2017	Ontología	(Provenzano, Hanninen, Zhou, & Lundqvist, 2017)
ART14	Using Ontology-Based Concept Maps for Requirements Engineering: A Case Study	IEEE Xplore	2017	NLP, Ontología	(Mohamed, Ellatif, & Farhan, 2017)
ART15	Use Case Elicitation with FrameNet Frames	IEEE Xplore	2017	NLP	(Kundi & Chitchyan, 2017)
ART16	From User Demand to Software Service: Using Machine Learning to Automate the Requirements Specification Process	IEEE Xplore	2017	NLP, Ontología	(Van Rooijen et al., 2017)
ART17	UCAnalyzer: A Tool to Analyze Use Case Textual Descriptions	IEEE Xplore	2017	NLP	(Tiwari & Laddha, 2017)
ART18	Let's hear it from RETTA: A Requirements Elicitation Tool for TrAffic management systems	IEEE Xplore	2017	NLP	(Noaen, Abad, & Far, 2017)
ART19	SSReq: A Method for Designing Star Schemas from Decisional Requirements	IEEE Xplore	2017	NLP	(Elamin, Alshomrani, & Feki, 2017)
ART20	Configuring use case models in product families	SpringerLink	2016	NLP	(Hajri, Goknil, Briand, & Stephany, 2016)

ART21	Using Ontology to Enhance Requirement Engineering in Agile Software Process	IEEE Xplore	2016	Ontología	(Sithithanasakul & Choosri, 2017)
ART22	An approach of dynamically combining ontologies for interactive Requirements Elicitation	IEEE Xplore	2016	NLP, Ontología	(Yuan & Tripathi, 2016)
ART23	Ontology Techniques for Representing the Problem of Discourse: Design of Solution Application Perspective	IEEE Xplore	2016	Ontología	(Miah, Islam, & Samsudin, 2016)
ART24	Sematic Analysis of Business Rules using Discourse Representation Theory	IEEE Xplore	2016	NLP	(Tuz Zahra, Rahim, & Iqbal, 2016)
ART25	Software Feature Request Detection in Issue Tracking Systems	IEEE Xplore	2016	NLP	(Merten et al., 2016)
ART26	Automated Extraction of Conceptual Models from User Stories via NLP	IEEE Xplore	2016	NLP	(Robeer, Lucassen, Van Der Werf, Dalpiaz, & Brinkemper, 2016)
ART27	A Novel Method: Ontology-based Security Requirements Engineering Framework	IEEE Xplore	2016	Ontología	(Salini & Kanmani, 2016)
ART28	Identifying duplicate functionality in textual use cases by aligning semantic actions	SpringerLink	2016	NLP	(Rago, Marcos, & Diaz-Pace, 2016)
ART29	Automated Requirements Extraction for Scientific Software	ScienceDirect	2015	NLP	(Y. Li, Guzman, Tsiamoura, Schneider, & Bruegge, 2015)
ART30	Developing Ontologies and Persona to Support and Enhance Requirements Engineering Activities – A Case Study	ScienceDirect	2015	Ontología	(Sim & Brouse, 2015)
ART31	VoiceToModel: An Approach to Generate Requirements Models from Reconocimiento del habla Mechanisms	ACM Library	2015	Reconocimiento del habla	(Soares & Wanderley, 2015)

ART32	Ontology-based Requirements Generation for Credibility Validation of Safetycritical System	IEEE Xplore	2015	Ontología	(R. Li, Ma, & Yao, 2015)
ART33	An Ontology-Based Requirement Modeling for Interactive Software Customization	IEEE Xplore	2015	Ontología	(Yuan & Zhang, 2015)
ART34	Applying Product Line Use Case Modeling in an Industrial Automotive Embedded System: Lessons Learned and a Refined Approach	IEEE Xplore	2015	NLP	(Hajri, Goknil, Briand, & Stephany, 2015)
ART35	Goal and Preference Identification through natural language	IEEE Xplore	2015	NLP	(Alabdulkareem, Cercone, & Liaskos, 2015)
ART36	SANAYOJAN: A Framework for Traceability Link Recovery between Use-Cases in Software Requirement Specification and Regulatory Documents	ACM Library	2014	NLP	(Jain, Ghaisas, & Sureka, 2014)
ART37	Deriving Requirements Model from Textual Use Cases	ACM Library	2014	NLP, Ontología	(Sawant, Roy, Sripathi, Plesse, & Sajeev, 2014)
ART38	GUITAR: An Ontology-based Automated Requirements Analysis Tool	IEEE Xplore	2014	Ontología	(Nguyen, Grundy, & Almorsy, 2014)
ART39	NLP-KAOS for Systems Goal Elicitation: Smart Metering System Case Study	IEEE Xplore	2014	NLP	(Casagrande, Woldeamlak, Woon, Zeineldin, & Svetinovic, 2014)
ART40	Engineering of Requirements for a Distributed Teleteaching System: A Conceptual Graph-based Approach	ACM Library	2013	NLP, Ontología	(Dutta, Bhattacharjee, Kanjilal, Dasgupta, & Bhattacharya, 2013)
ART41	Using Security and Domain ontologies for Security Requirements Analysis	IEEE Xplore	2013	Ontología	(Souag, Salinesi, Wattiau, & Mouratidis, 2013)
ART42	A Process Framework for Managing Implicit Requirements Using Analogy-Based Reasoning	IEEE Xplore	2013	NLP, Ontología	(Onyeka, 2013)

ART43	Exploring Design Principles of Task Elicitation Systems for Unrestricted Natural Language Documents	ACM Library	2012	NLP	(Meth, Mädche, & Einöder, 2012)
ART44	Multi-Perspective Ontology to Understand Organizational Requirements	IEEE Xplore	2012	Ontología	(Kifle, 2012)
ART45	Pattern-Based Security Requirements Specification Using Ontologies and Boilerplates	IEEE Xplore	2012	NLP, Ontología	(Daramola, Sindre, & Stalhane, 2012)
ART46	Speculative Requirements: Automatic Detection of Uncertainty in Natural Language Requirements	IEEE Xplore	2012	NLP	(Yang, De Roeck, Gervasi, Willis, & Nuseibeh, 2012)
ART47	Semi-automatic Generation of UML Models from Natural Language Requirements	ACM Library	2011	NLP	(Deeptimahanti & Sanyal, 2011)
ART48	Natural Language based Component Extraction from Requirement Engineering Document and its Complexity Analysis	ACM Library	2011	NLP	(Sharma & Kushwaha, 2011)
ART49	A Decision Making Ontology Building Process for Analytical Requirements Elicitation	IEEE Xplore	2011	Ontología	(Bargui, Ben-Abdallah, & Feki, 2011)
ART50	REFINTO: An Ontology-Based Requirements Engineering Framework for Buisness-IT Alignment in Financial Services Organizations	IEEE Xplore	2011	Ontología	(Umoh, Sampaio, & Theodoulidis, 2011)
ART51	DODT: Increasing Requirements Formalism using Domain Ontologies for Improved Embedded Systems Development	IEEE Xplore	2011	NLP, Ontología	(Farfeleder, Moser, Krall, Stalhane, et al., 2011)
ART52	A Methodology for Accurate and Redundancy-free Business Requirements Description Using Ontology	IEEE Xplore	2011	NLP, Ontología	(Choe & Moon, 2011)
ART53	A framework for multidimensional design of data warehouses from ontologies	ScienceDirect	2010	Ontología	(Romero & Abelló, 2010)

ART54	Case-based analysis in user requirements modelling for knowledge construction	ScienceDirect	2010	Ontología	(Sun & Mushi, 2010)
ART55	A QoS Ontology Cooperated with Feature Models for Non-functional Requirements Elicitation	ACM Library	2010	Ontología	(Wang et al., 2010)
ART56	From Collective Knowledge to Intelligence: Pre-Requirements Analysis of Large and Complex Systems	ACM Library	2010	Ontología	(Liang, Avgeriou, He, & Xu, 2010)
ART57	Enhancing Domain Knowledge for Requirements Elicitation with Web Mining	IEEE Xplore	2010	NLP, Ontología	(Kaiya, Shimizu, Yasui, Kaijiri, & Saeki, 2010)
ART58	DPTool: A Tool for Supporting the Problem Description and Projection	IEEE Xplore	2010	Ontología	(X. Chen, Yin, & Jin, 2010)
ART59	Complementary Classification Techniques based Personalized Software Requirements Retrieval with Semantic Ontology and User Feedback	IEEE Xplore	2010	Ontología	(Zhang & Lee, 2010)
ART60	An Integrated Support for Attributed Goal-Oriented Requirements Analysis Method and Its Implementation	IEEE Xplore	2010	Ontología	(Saeki, Hayashi, & Kaiya, 2010)
ART61	Text-based Requirements Preprocessing using Nature Language Processing Techniques	IEEE Xplore	2010	NLP, Ontología	(H. Chen, He, Liang, & Li, 2010)
ART62	A Domain Ontology Building Process for Guiding Requirements Elicitation	SpringerLink	2010	NLP, Ontología	(Omoronyia et al., 2010)
ART63	An Automated Approach to Transform Use Cases into Activity Diagrams	SpringerLink	2010	NLP	(Yue, Briand, & Labiche, 2010)
ART64	Scenario Analysis: Generation of Possible Scenario Interpretations and their Visualization	IEEE Xplore	2010	NLP	(Kof & Garching, 2010)

Anexo B) Lista de artículos de acuerdo a las preguntas de publicación

A continuación, se presenta en la Tabla A-2 la clasificación de los artículos de acuerdo a las preguntas de publicación.

Tabla A-2. Clasificación de artículo de acuerdo a las preguntas de publicación.

ID	PQ1 Tipo	PQ2 Año de publicación	PQ3 País	PQ4 Ámbito del autor	PQ5 Revista/ Congreso
ART01	Revista	2018	Jordania	Académico	Information and Software Technology
ART02	Revista	2018	México	Industrial	Information Systems
ART03	Revista	2018	China	Académico	Journal of Systems and Software
ART04	Conferencia	2018	Brasil	Académico	Brazilian Symposium on Software Engineering
ART05	Revista	2018	España	Académico	International Systems and Software Product Line Conference
ART06	Conferencia	2018	Grecia	Académico	The 2018 Web Conference Companion
ART07	Conferencia	2018	Jordania	Académico	International Conference on Geoinformatics and Data Analysis
ART08	Conferencia	2018	Canadá	Académico	International Workshop on Artificial Intelligence for Requirements Engineering (AIRE)

ART09	Conferencia	2018	Estados Unidos	Académico	International Workshop on Artificial Intelligence for Requirements Engineering (AIRE)
ART10	Conferencia	2018	Italia	Académico	International Workshop on Artificial Intelligence for Requirements Engineering (AIRE)
ART11	Conferencia	2018	Alemania	Académico	IEEE International Requirements Engineering Conference (RE)
ART12	Conferencia	2018	Estados Unidos	Académico	IEEE International Requirements Engineering Conference (RE)
ART13	Conferencia	2017	Suecia	Industrial	Asia Pacific Software Engineering Conference
ART14	Conferencia	2017	Egipto	Académico	International Computer Engineering Conference (ICENCO)
ART15	Conferencia	2017	Inglaterra	Académico	IEEE International Requirements Engineering Conference (RE)
ART16	Conferencia	2017	Alemania	Académico	IEEE International Requirements Engineering Conference (RE)
ART17	Conferencia	2017	India	Académico	IEEE International Requirements Engineering Conference (RE)
ART18	Conferencia	2017	Canadá	Académico	IEEE International Requirements Engineering Conference (RE)

ART19	Conferencia	2017	Sudán	Académico	International Conference on Communication, Control, Computing and Electronics Engineering (ICCCCEE)
ART20	Revista	2016	Luxemburgo	Académico	Software and Systems Modelling
ART21	Conferencia	2016	Tailandia	Académico	International Conference on Software, Knowledge, Information Management & Applications (SKIMA)
ART22	Conferencia	2016	Canadá	Académico	IEEE International Conference on Software Engineering and Service Science (ICSESS)
ART23	Conferencia	2016	Malasia	Académico	IEEE International Conference on Computer and Information Technology (CIT)
ART24	Conferencia	2016	Pakistan	Académico	Sixth International Conference on Innovative Computing Technology (INTECH)
ART25	Conferencia	2016	Alemania	Académico	IEEE International Requirements Engineering Conference (RE)
ART26	Conferencia	2016	Holanda	Académico	IEEE International Requirements Engineering Conference (RE)
ART27	Conferencia	2016	India	Académico	International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS)
ART28	Revista	2016	Argentina	Académico	Software and Systems Modelling

ART29	Revista	2015	Alemania	Académico	Procedia Computer Science
ART30	Revista	2015	Estados Unidos	Académico	Procedia Computer Science
ART31	Revista	2015	Portugal	Académico	Annual ACM Symposium on Applied Computing SAC
ART32	Conferencia	2015	China	Académico	IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing
ART33	Conferencia	2015	Canadá	Académico	IEEE International Model-Driven Requirements Engineering Workshop (MoDRE)
ART34	Conferencia	2015	Luxemburgo	Académico	ACM/IEEE International Conference on Model Driven Engineering Languages and Systems (MODELS)
ART35	Conferencia	2015	Canadá	Académico	IEEE International Requirements Engineering Conference (RE)
ART36	Revista	2014	India	Académico	International Workshop on Realizing Artificial Intelligence Synergies in Software Engineering
ART37	Revista	2014	India	Académico	International Conference on Software Engineering

ART38	Conferencia	2014	Australia	Académico	IEEE International Requirements Engineering Conference (RE)
ART39	Revista	2014	Emiratos Arabes	Industrial	IEEE Transactions on Software Engineering
ART40	Revista	2013	India	Académico	ACM SIGSOFT Software Engineering Notes
ART41	Conferencia	2013	Francia	Académico	IEEE Annual Computer Software and Applications Conference Workshops
ART42	Conferencia	2013	Nigeria	Académico	IEEE International Conference on Research Challenges in Information Science (RCIS)
ART43	Revista	2012	Alemania	Académico	ACM SIGCHI symposium on Engineering interactive computing systems
ART44	Conferencia	2012	Etopía	Académico	African Conference for Software Engineering and Applied Computing
ART45	Conferencia	2012	Nigeria	Académico	IEEE International Workshop on Requirements Patterns
ART46	Conferencia	2012	Reino Unido	Académico	IEEE International Requirements Engineering Conference (RE)
ART47	Revista	2011	Irlanda	Académico	India Software Engineering Conference

ART48	Revista	2011	India	Académico	ACM SIGSOFT Software Engineering Notes
ART49	Conferencia	2011	Túnez	Académico	IEEE International Conference on Trust, Security and Privacy in Computing and Communications
ART50	Conferencia	2011	Reino Unido	Académico	IEEE International Conference on Services Computing
ART51	Conferencia	2011	Australia	Académico	IEEE International Symposium on Design and Diagnostics of Electronic Circuits and Systems
ART52	Conferencia	2011	Corea	Académico	International Conference on Information Science and Applications
ART53	Revista	2010	España	Académico	Data & Knowledge Engineering
ART54	Revista	2010	Reino Unido	Académico	Information and Software Technology
ART55	Revista	2010	China	Académico	Asia-Pacific Symposium on Internetware
ART56	Conferencia	2010	China	Académico	Web 2.0 for Software Engineering
ART57	Conferencia	2010	Japón	Académico	Asia Pacific Software Engineering Conference

ART58	Conferencia	2010	China	Académico	IEEE International Requirements Engineering Conference (RE)
ART59	Conferencia	2010	Corea	Académico	IEEE International Conference on Computer and Information Technology (CIT)
ART60	Conferencia	2010	Japón	Académico	International Conference on Quality Software
ART61	Conferencia	2010	China	Académico	International Conference On Computer Design and Applications
ART62	Conferencia	2010	Noruega	Académico	International Working Conference on Requirements Engineering: Foundation for Software Quality
ART63	Conferencia	2010	Canadá	Académico	European Conference on Modelling Foundations and Applications
ART64	Conferencia	2010	Alemania	Académico	International Workshop on Requirements Engineering Visualization

Anexo C) Lista de artículos de acuerdo a las preguntas de investigación

A continuación, en la Tabla A-3 se presenta la clasificación de los artículos de acuerdo a las preguntas de publicación.

Tabla A-3. Clasificación de artículo de acuerdo a las preguntas de publicación

ID	RQ1 Enfoque	RQ2 Contribución	RQ3 Entrada	RQ4 Salidas	RQ5 Evaluación	RQ6 Tipo de investigación	QR7 Automatización	RQ8 NLP	R8 Fase de RE	R8 Tipo de ontología
ART01	NLP	Método/ Procedimiento / Técnica	Documentos en NL	Identificación de abstracción	Caso de Estudio	Investigación de evaluación	Semi- automática	Análisis de texto y análisis de datos exploratorios; Modelado y extracción de patrones; Evaluación y Despliegue	NA	NA
ART02	NLP	Método/ Procedimiento / Técnica	Documentos en NL	Clasificación de Requerimientos	Experimento Controlado	Investigación de evaluación	Semi- automática	Análisis de texto y análisis de datos exploratorios; Modelado y extracción de patrones; Evaluación y Despliegue	NA	NA
ART03	NLP	Método/ Procedimiento / Técnica	Documentos en NL	Clasificación de Requerimientos	Caso de Estudio	Propuesta de solución	Semi- automática	Pre- procesamiento de texto; Representación de texto e ingeniería de características; Modelado y extracción de patrones;	NA	NA

								Evaluación y Despliegue		
ART04	NLP	Método/ Procedimiento / Técnica	Documentos en NL	Identificación de Requerimientos	Simulación	Propuesta de solución	Semi- automática	Pre- procesamiento de texto; Representación de texto e ingeniería de características; Análisis de texto y análisis de datos exploratorios; Modelado y extracción de patrones; Evaluación y Despliegue;	NA	NA
ART05	NLP	Herramienta	Documentos en NL	Generación de Modelos	Caso de Estudio	Investigación de evaluación	Manual	Pre- procesamiento de texto; Análisis de texto y análisis de datos exploratorios; Representación de texto e ingeniería de características; Modelado y extracción de patrones; Evaluación y Despliegue;	NA	NA

ART06	NLP	Herramienta	Documentos con CNL	Otro	No Aplica	Propuesta de solución	Semi-automática	Pre-procesamiento de texto; Análisis de texto y análisis de datos exploratorios;	NA	NA
ART07	NLP	Método/ Procedimiento / Técnica	Documentos en NL	Identificación de abstracción	Prueba de Concepto	Propuesta de solución	Automática	Análisis de texto y análisis de datos exploratorios;	NA	NA
ART08	Reconocimiento del habla	Herramienta	Documentos en NL, Voz	Identificación de Requerimientos	Caso de Estudio	Propuesta de solución	Automática	Reconocimiento del habla, Pre-procesamiento de texto; Representación de texto e ingeniería de características; Modelado y extracción de patrones; Evaluación y Despliegue	NA	NA
ART09	NLP	Método/ Procedimiento / Técnica	Documentos en NL	Identificación de Requerimientos	Caso de Estudio	Propuesta de solución	Semi-automática	Representación de texto e ingeniería de características; Modelado y extracción de patrones; Evaluación y Despliegue;	NA	NA
ART10	NLP	Método/ Procedimiento / Técnica	Documentos en NL	Análisis de Calidad	Prueba de Concepto	Propuesta de solución	Semi-automática	Representación de texto e ingeniería de características; Modelado y	NA	NA

								extracción de patrones;		
ART11	NLP	Método/ Procedimiento / Técnica	Documentos en NL	Identificación de abstracción	Caso de Estudio	Propuesta de solución	Automática	Pre- procesamiento de texto; Análisis de texto y análisis de datos exploratorios; Modelado y extracción de patrones; Evaluación y Despliegue;	NA	NA
ART12	NLP, Ontología	Herramienta	Documentos en NL	Análisis de Calidad	No Aplica	Propuesta de solución	Semi- automática	Análisis de texto y análisis de datos exploratorios;	Elicitación de Requerimie ntos	Ontología del Documento de Especificación de Requerimien to
ART13	Ontología	Método/ Procedimiento / Técnica	No Aplica	Análisis de calidad	No Aplica	Propuesta de solución	Manual		Análisis y Negociación de Requerimie ntos	Ontología del Documento de Especificación de Requerimien tos

ART14	NLP, Ontología	Framework	Documentos en NL	Generación de Modelos	Simulación	Propuesta de solución	Semi-automática	Pre-procesamiento de texto; Análisis de texto y análisis de datos exploratorios; Modelado y extracción de patrones; Evaluación y Despliegue;	Modelado del dominio del problema	Ontología de Requerimientos
ART15	NLP	Método/ Procedimiento / Técnica	Documentos en NL	Identificación de abstracción	Prueba de Concepto	Propuesta de solución	Semi-automática	Pre-procesamiento de texto	NA	NA
ART16	NLP, Ontología	Método/ Procedimiento / Técnica	Documentos en NL, Diagramas	Generación de Modelos	No Aplica	Propuesta de solución	Automática	Análisis de texto y análisis de datos exploratorios; Modelado y extracción de patrones;	Elicitación de Requerimientos	Ontología del Dominio
ART17	NLP	Herramienta	Documentos en NL	Análisis de Calidad	No Aplica	Otro	Semi-automática	Pre-procesamiento de texto; Análisis de texto y análisis de datos exploratorios; Modelado y extracción de patrones;	NA	NA
ART18	NLP	Herramienta	Documentos en NL	Identificación de Requerimientos	No Aplica	Otro	Semi-automática	Pre-procesamiento de texto; Análisis de texto y análisis de datos	NA	NA

								exploratorios; Modelado y extracción de patrones;		
ART19	NLP	Herramienta	Documentos en NL	Generación de Modelos	Caso de Estudio	Propuesta de solución	Automática	Análisis de texto y análisis de datos exploratorios; Modelado y extracción de patrones;	NA	NA
ART20	NLP	Método/ Procedimiento / Técnica	Documentos con CNL	Generación de Modelos	Caso de Estudio	Propuesta de solución	Semi- automática	Pre- procesamiento de texto; Análisis de texto y análisis de datos exploratorios; Modelado y extracción de patrones;	NA	NA
ART21	Ontología	Método/ Procedimiento / Técnica	Documentos en NL	Generación de Modelos	Experi- mento Controlad o	Propuesta de solución	Semi- automática		Administra- ción de Requerimie- ntos	Ontología de Requerimien- tos
ART22	NLP, Ontología	Método/ Procedimiento / Técnica	Ontología	Generación de Modelos	Prueba de Concepto	Propuesta de solución	Manual	Análisis de texto y análisis de datos exploratorios;	Modelado del dominio del problema	Ontología del Dominio
ART23	Ontología	Método/ Procedimiento / Técnica	Documentos en NL	Generación de Modelos	Prueba de Concepto	Investigación de evaluación	Manual		Modelado del dominio del problema	Ontología del Dominio

ART24	NLP	Método/ Procedimiento / Técnica	Documentos en NL	Análisis de Calidad	Caso de Estudio	Propuesta de solución	Automática	Pre- procesamiento de texto; Análisis de texto y análisis de datos exploratorios; Evaluación y Despliegue;	NA	NA
ART25	NLP	Método/ Procedimiento / Técnica	Documentos en NL	Identificación de Requerimientos	Simulación	Investigación de validación	Manual	Pre- procesamiento de texto; Análisis de texto y análisis de datos exploratorios; Representación de texto e ingeniería de características; Modelado y extracción de patrones; Evaluación y Despliegue;	NA	NA
ART26	NLP	Herramienta	Documentos en NL	Generación de Modelos	Caso de Estudio	Investigación de evaluación	Automática	Análisis de texto y análisis de datos exploratorios; Modelado y extracción de patrones; Evaluación y Despliegue;	NA	NA
ART27	Ontología	Framework	No Aplica	Identificación de requerimientos	No Aplica	Propuesta de solución	Manual		Elicitación de Requerimientos	Ontología de Requerimientos

ART28	NLP	Método/ Procedimiento / Técnica	Documentos en NL	Análisis de Calidad	Caso de Estudio	Propuesta de solución	Semi- automática	Pre- procesamiento de texto; Análisis de texto y análisis de datos exploratorios; Modelado y extracción de patrones; Evaluación y Despliegue;	NA	NA
ART29	NLP	Método/ Procedimiento / Técnica	Documentos en NL	Identificación de Requerimientos	Prueba de Concepto	Propuesta de solución	Automática	Pre- procesamiento de texto; Análisis de texto y análisis de datos exploratorios; Representación de texto e ingeniería de características; Modelado y extracción de patrones;	NA	NA
ART30	Ontología	Método/ Procedimiento / Técnica	Documentos en NL	Generación de Modelos	Prueba de Concepto	Investigación de evaluación	Manual		Modelado del dominio del problema	Ontología del Dominio, Ontología de Requerimien tos
ART31	Reconoci miento del habla	Framework	Voz	Generación de Modelos	Prueba de Concepto	Investigación de evaluación	Semi- automática	Reconocimiento del habla; Análisis de texto y análisis de datos exploratorios;	NA	NA

ART32	Ontología	Método/ Procedimiento / Técnica	No Aplica	Identificación de requerimientos	Caso de Estudio	Propuesta de solución	Manual		Elicitación de Requerimie ntos	Ontología del Dominio
ART33	Ontología	Método/ Procedimiento / Técnica	No Aplica	Generación de Modelos	Prueba de Concepto	Investigación de evaluación	Manual		Administrac ión de Requerimie ntos	Ontología del Documento de Especificació n de Requerimien to, Ontología de Requerimien tos
ART34	NLP	Método/ Procedimiento / Técnica	Documentos en NL, Diagramas	Generación de Modelos	Caso de Estudio	Investigación de evaluación	Semi- automática	Pre- procesamiento de texto; Análisis de texto y análisis de datos exploratorios;	NA	NA
ART35	NLP	Herramienta	Documentos en NL	Identificación de abstracción	Experime nto Controlad o	Propuesta de solución	Semi- automática	Análisis de texto y análisis de datos exploratorios; Modelado y extracción de patrones; Evaluación y Despliegue;	NA	NA
ART36	NLP	Método/ Procedimiento / Técnica	Documentos en NL	Identificación de abstracción	Caso de Estudio	Investigación de evaluación	Semi- automática	Análisis de texto y análisis de datos exploratorios; Modelado y extracción de patrones;	NA	NA

								Evaluación y Despliegue;		
ART37	NLP, Ontología	Método/ Procedimiento / Técnica	Documentos en NL	Generación de Modelos	Caso de Estudio	Propuesta de solución	Semi-automática	Pre-procesamiento de texto; Análisis de texto y análisis de datos exploratorios; Evaluación y Despliegue;	Administración de Requerimientos	Ontología del Documento de Especificación de Requerimiento, Ontología de Requerimientos
ART38	Ontología	Herramienta	Documentos en NL	Generación de Modelos	Caso de Estudio	Investigación de evaluación	Automática		Documentación y Validación de Requerimientos	Ontología del Dominio
ART39	NLP	Método/ Procedimiento / Técnica	Documentos en NL	Generación de Modelos	Caso de Estudio	Investigación de evaluación	Semi-automática	Análisis de texto y análisis de datos exploratorios	NA	NA
ART40	Ontología	Método/ Procedimiento / Técnica	Diagramas	Generación de Modelos	Prueba de Concepto	Investigación de evaluación	Semi-automática		Modelado del dominio del problema	Ontología del Dominio

ART41	Ontología	Método/ Procedimiento / Técnica	Documentos en NL	Identificación de Requerimientos	Experime nto Controlad o	Propuesta de solución	Manual		Elicitación de Requerimie ntos	Ontología del Dominio, Ontología de Requerimien tos
ART42	NLP, Ontología	Framework	Documentos en NL	Identificación de Requerimientos	No aplica	Propuesta de solución	Manual	Análisis de texto y análisis de datos exploratorios; Modelado y extracción de patrones;	Elicitación de Requerimie ntos	Ontología del Dominio
ART43	NLP	Herramienta	Documentos en NL	Identificación de abstracción	Otro	Otro	Automática	No menciona	NA	NA
ART44	Ontología	Método/ Procedimiento / Técnica	No Aplica	Identificación de abstracción	Prueba de Concepto	Propuesta de solución	Manual		Modelado del dominio del problema	Ontología del Dominio
ART45	NLP, Ontología	Herramienta	Documentos en NL	Identificación de Requerimientos	Experime nto Controlad o	Propuesta de solución	Semi- automática	Pre- procesamiento de texto; Análisis de texto y análisis de datos exploratorios;	Análisis y Negociación de Requerimie ntos	Ontología del Dominio
ART46	NLP	Método/ Procedimiento / Técnica	Documentos en NL	Análisis de Calidad	Simulació n	Investigación de evaluación	Automática	Pre- procesamiento de texto; Análisis de texto y análisis de datos exploratorios; Modelado y extracción de patrones;	NA	NA

								Evaluación y Despliegue;		
ART47	NLP	Método/ Procedimiento / Técnica	Documentos en NL	Generación de Modelos	Prueba de Concepto	Propuesta de solución	Semi- automática	Pre- procesamiento de texto; Análisis de texto y análisis de datos exploratorios;	NA	NA
ART48	NLP	Herramienta	Documentos en NL	Clasificación de Requerimientos	Simulación	Investigación de validación	Semi- automática	Pre- procesamiento de texto; Análisis de texto y análisis de datos exploratorios	NA	NA
ART49	Ontología	Método/ Procedimiento / Técnica	Diagramas	Generación de Modelos	Experimento Controlado	Propuesta de solución	Manual		Modelado del dominio del problema	Ontología del Documento de Especificación de Requerimiento
ART50	Ontología	Framework	Documentos en NL	Identificación de abstracción	Caso de Estudio	Investigación de evaluación	Manual		Modelado del dominio del problema	Ontología del Dominio
ART51	NLP, Ontología	Método/ Procedimiento / Técnica	Documentos en NL	Identificación de Requerimientos	Caso de Estudio	Propuesta de solución	Semi- automática	Pre- procesamiento de texto; Análisis de texto y análisis de	Elicitación de Requerimientos	Ontología del Dominio

								datos exploratorios;		
ART52	NLP, Ontología	Método/ Procedimiento / Técnica	Documentos en NL	Identificación de abstracción	No aplica	Propuesta de solución	Manual	No menciona	Modelado del dominio del problema	Ontología del Dominio
ART53	Ontología	Framework	Ontología	Generación de Modelos	No aplica	Artículo filosófico	Manual		Modelado del dominio del problema	Ontología del Dominio
ART54	Ontología	Método/ Procedimiento / Técnica	No Aplica	Identificación de requerimientos	Prueba de Concepto	Propuesta de solución	Manual		Modelado del dominio del problema	Ontología del Documento de Especificación de Requerimiento
ART55	Ontología	Método/ Procedimiento / Técnica	No Aplica	Identificación de requerimientos	Prueba de Concepto	Propuesta de solución	Manual		Análisis y Negociación de Requerimientos	Ontología del Dominio
ART56	Ontología	Método/ Procedimiento / Técnica	Documentos en NL	Identificación de Requerimientos	No aplica	Propuesta de solución	Manual		Administración de Requerimientos	Ontología de Requerimientos

ART57	NLP, Ontología	Método/ Procedimiento / Técnica	Ontología	Generación de Modelos	Experimento Controlado	Propuesta de solución	Semi-automática	Representación de texto e ingeniería de características; Modelado y extracción de patrones;	Modelado del dominio del problema	Ontología del Dominio
ART58	Ontología	Herramienta	No Aplica	Generación de Modelos	No aplica	Otro	Manual		Elicitación de Requerimientos	Ontología del Dominio
ART59	Ontología	Herramienta	Documentos en NL	Identificación de Requerimientos	Prueba de Concepto	Propuesta de solución	Manual		Elicitación de Requerimientos	Ontología de Requerimientos
ART60	Ontología	Herramienta	Diagramas	Análisis de calidad	Prueba de Concepto	Otro	Manual		Modelado del dominio del problema	Ontología de Requerimientos
ART61	NLP, Ontología	Método/ Procedimiento / Técnica	Documentos en NL	Identificación de abstracción	Prueba de Concepto	Propuesta de solución	Automática	Pre-procesamiento de texto; Representación de texto e ingeniería de características; Evaluación y Despliegue;	Análisis y Negociación de Requerimientos	Ontología del Documento de Especificación de Requerimiento
ART62	NLP, Ontología	Método/ Procedimiento / Técnica	Documentos en NL	Generación de Modelos	Caso de Estudio	Investigación de evaluación	Semi-automática	Pre-procesamiento de texto; Análisis de texto y análisis de datos exploratorios;	Modelado del dominio del problema	Ontología del Dominio

ART63	NLP	Método/ Procedimiento / Técnica	Documentos con CNL	Generación de Modelos	Otro	Propuesta de solución	Automática	Modelado y extracción de patrones; Evaluación y Despliegue;	NA	NA
ART64	NLP	Método/ Procedimiento / Técnica	Documentos en NL	Generación de Modelos	Prueba de Concepto	Investigación de evaluación	Semi- automática	ext Parsing and Exploratory Data Analysis; Modelado y extracción de patrones;	NA	NA