

Estudio de Técnicas de Procesamiento de Lenguaje Natural en la Validación de Requerimientos

Sonia Santana¹ [0000-0003-2163-120X], Lucrecia Perero¹ [0000-0002-7422-9479], Noelia Rodriguez¹ [0009-0009-4981-6792]

Leandro Antonelli² [0000-0003-1388-0337], Alejandro Fernandez² [0000-0002-7968-6871]

¹ Facultad de Ciencias de la Administración - Universidad Nacional de Entre Ríos
{sonia.santana, ramona.perero}@uner.edu.ar, romi2022@gmail.com

² Laboratorio de Investigación y Formación en Informática Avanzada (LIFIA),
Facultad de Informática, Universidad Nacional de La Plata
{alejandro.fernandez, leandro.antonelli}@lifia.info.unlp.edu.ar

Abstract. La validación de los requerimientos es una tarea fundamental en cualquier proyecto de Ingeniería de Software. La calidad de los requerimientos tiene un impacto significativo sobre los artefactos durante el desarrollo del software, debido a que los requerimientos se expresan generalmente en lenguaje natural, lo cual puede ocasionar distintos tipos de defectos como la ambigüedad, integridad y atomicidad. Por otra parte, el principal objetivo de la validación de los requerimientos es confirmar que los requerimientos especificados sean representaciones de las necesidades y expectativas de los usuarios y deben ser completos, correctos y consistentes entre otras características, evaluar estos atributos de calidad puede resultar una tarea difícil. En este trabajo se muestra el relevamiento de las principales tendencias de las técnicas de procesamiento de lenguaje natural (PLN) en la validación de requerimientos publicados en los últimos 13 años. Se seleccionaron y analizaron a partir de un enfoque basado en un esquema de caracterización para conocer información asociadas a los mismos.

Keywords: Ingeniería de requerimientos, validación de requerimientos, lenguaje de procesamiento natural, técnicas de procesamiento de lenguaje natural

1. Introducción

El objetivo de la validación de requerimientos es certificar que los requerimientos documentados sean un reflejo del sistema del mundo real para el cual se va a desarrollar el software. Las especificaciones de requerimientos representan las expectativas, necesidades y deseos de los interesados del sistema propuesto.

En el marco de la Ingeniería de Requerimientos la validación de los requerimientos es una tarea fundamental en cualquier proyecto de Ingeniería de Software y debe ser un proceso continuo en el ciclo de vida del desarrollo del sistema. Los defectos de los requerimientos se encuentran entre las principales causas del fracaso de los proyectos [1][2]. Un motivo importante es que los requerimientos están generalmente escritos en lenguaje natural. Esta técnica se utiliza desde los primeros tiempos de la ingeniería de software y aún hoy es una de las más utilizadas [3], por lo tanto, los problemas del lenguaje natural son inherentes a los requerimientos [4].

La comprensión del lenguaje natural es el área principal en el campo del PLN [5]. Además, el PLN ocupa un lugar destacado en la Ingeniería de Requerimientos (IR),

particularmente cuando se deben diseñar sistemas [6]. En esa situación crítica, las barreras de comunicación e idioma entre los clientes y el equipo del proyecto causan una gran cantidad de problemas que reducen la calidad de la aplicación y el seguimiento de los requerimientos. [6].

Este trabajo se enfoca en las principales tendencias de las técnicas de PLN en la validación de requerimientos del software y presenta un análisis de enfoques para conocer información asociadas a los mismos.

El resto del artículo está organizado de la siguiente manera. La sección 2 describe una revisión de la literatura para identificar los distintos enfoques de las técnicas relacionadas con el PLN en la validación de requerimientos. La sección 3 describe un análisis que consiste en obtener información de los enfoques sobre las características asociadas a técnicas de PLN en la validación de requerimientos. Además, se presentan los resultados del análisis comparativo realizado a los enfoques, donde se identificaron las fortalezas y debilidades de las técnicas de PLN en la validación de requerimientos. La sección 4 discute los resultados obtenidos.

Finalmente, la sección 5 expresa las conclusiones y trabajos futuros.

2. Revisión bibliográfica

En esta sección se realizó un proceso de tres fases: búsqueda, selección y evaluación de trabajos para una revisión bibliográfica de enfoques relacionados con el PLN en la validación de requerimientos. La pregunta a responder para el proceso de revisión es ¿Cuál es la tendencia de las prácticas relacionadas de PLN en la validación de requerimientos? Para las búsquedas se desarrollaron frases, en inglés y español, con el fin de encontrar respuestas a la pregunta inicial, que son las siguientes:

- (“requirements”) y (“validation” o “validate” o “validity”) y (“technique”) y (“natural language” o “NLP” o “natural language processing”)
- (“requisitos” o “requerimientos”) y (“validación” o “validar” o “validez”) y (“técnica”) y (“lenguaje natural” o “PLN” o “procesamiento lenguaje natural”)

Las búsquedas se realizaron en artículos publicados en las editoriales de información IEEE, Springer y ACM Digital Library, publicados entre 2010 y 2023.

Los criterios de inclusión y exclusión son los siguientes: a) Los artículos seleccionados deben estar directamente relacionados con el tema de PLN en la validación de requerimientos en el área de IR. b) Los artículos no deben discutir la validación de requerimientos en la fase de prueba del software desarrollado o en la implementación en prueba con respecto a los requerimientos. c) Los artículos no deben utilizar términos como "Revisión sistemática de la literatura", "Revisión de la literatura", "Análisis sistemático".

Luego, se aplicaron los criterios de evaluación presentados en la Tabla 1. La lista de criterios de evaluación fue adaptada de [7] y permite seleccionar los artículos finales en base a cuatro criterios que se enumeran.

Tabla 1. Criterios de evaluación.

<i>Sección</i>	<i>Criterios</i>
Introducción	1. ¿La introducción proporciona una descripción general del aporte para PLN en la validación de requerimientos? 2. ¿Está claramente definido el propósito / objetivo de la investigación?
Metodología	3. ¿Está claramente definida la metodología de investigación? 4. ¿Se establece una sucesión clara de pruebas a partir de las observaciones a las conclusiones? 5. ¿Se indican claramente los procedimientos de análisis?
Resultados	6. ¿Se indica adecuadamente el caso y su contexto? 7. ¿El estudio de caso se basa en la teoría y está vinculado a la literatura? 8. ¿Están los hallazgos claramente establecidos? ¿Los resultados ayudan a resolver los problemas de validación de requerimientos?
Conclusión	9. ¿Contiene conclusiones claras, implicaciones para la práctica y la investigación futura? 10. ¿Las conclusiones, implicaciones para la práctica y futuras investigaciones son derivadas adecuadamente para su audiencia? 11. ¿Existen límites o restricciones impuestas a la afirmación de conclusión?

En la búsqueda inicial aplicando la estrategia de búsqueda se encontraron 6045 artículos, luego de eliminar los artículos duplicados y aplicar los criterios de inclusión y exclusión definidos se preseleccionaron 576 artículos. Después de leer los títulos, los resúmenes, palabras claves y las conclusiones se preseleccionaron 85 artículos relevantes. Por último, luego de aplicar los criterios de evaluación definidos en la Tabla 1 se preseleccionaron 32 documentos, de los cuales posteriormente se seleccionaron solo aquellos que aportan una técnica de PLN para el proceso de validación de requerimientos por su modo de pensamiento, modelamiento, conceptos y trabajo. El resultado se presenta en la Tabla 2.

Tabla 2. Artículos seleccionados de la revisión bibliográfica

<i>Ref. Bibliog.</i>	<i>Ref. Estudio</i>	<i>Año</i>	<i>Nombre</i>
[8]	T1	2011	Automatic Validation and Correction of Formalized, Textual Requirements
[9]	T2	2012	Using Norm Analysis Patterns for Automated Requirements Validation
[10]	T3	2018	Automated Validation of Requirement Reviews: A Machine Learning Approach
[11]	T4	2019	A Scalable Operational Framework for Requirements Validation Using Semantic and Functional Models

3. Análisis de técnicas

En esta sección se comparan las técnicas seleccionadas con un enfoque basado en un esquema de caracterización [12] para conocer sus características, necesidades de información y restricciones de las técnicas de PLN en la validación de requerimientos.

3.1 Definición del esquema de caracterización para analizar las técnicas.

Para la comparación de las técnicas de PLN en la validación de requerimientos se adaptó el esquema de categorización definido por Vegas [12]. El esquema presenta tres niveles: nivel operacional, nivel de uso y nivel histórico, cada uno contiene distintos atributos que definen los elementos de cada nivel. Fue desarrollado para recopilar información relevante sobre técnicas de prueba de software que contribuye a un repositorio donde almacena la descripción de cada técnica. En este contexto, el esquema fue adaptado al análisis propuesto, se eliminó el nivel histórico relacionado con el repositorio y se modificaron algunos atributos del nivel operacional y de uso.

Cuando el esquema es instanciado en una variedad de técnicas, proporciona a los desarrolladores suficiente información para poder seleccionar las técnicas más adecuadas en un proyecto determinado. Esto asegura que las decisiones que toman se basan en el conocimiento objetivo de las técnicas. La información sobre las técnicas se organiza en dos niveles, los mismos fueron adecuados para la comparación de técnicas para la validación de requerimientos:

Nivel operacional. Este nivel está relacionado con las condiciones óptimas de operatividad de la técnica, una vez determinadas las características dadas del entorno en el que se va a aplicar la técnica. Esto significa que puede o no ser adecuado aplicar una determinada técnica en función de los conocimientos, experiencia del personal y existencia de herramientas. Este nivel contiene cuatro elementos objetivo, técnica, personal y resultados que muestran la información disponible.

Nivel de uso. Este nivel especifica las experiencias de los sujetos en el uso de la técnica y contiene dos elementos: proyecto, que especifica información sobre proyectos en los que se utilizó la técnica y satisfacción, que especifica qué opinión merece la técnica entre las personas que la utilizaron.

La Tabla 3 muestra el contenido del esquema de categorización adaptado.

Tabla 3. Esquema de categorización.

<i>Nivel</i>	<i>Elemento</i>	<i>Atributo</i>	<i>Descripción</i>
Operativo	Objetivo	Función	Función de la técnica.
		Etapa	Etapa de desarrollo del software vinculada a la técnica.
		Elementos	Elementos del desarrollo del sistema vinculados
	Técnica	Comprensibilidad	Facilidad de la técnica.
		Entradas	Entradas requeridas de la técnica.
		Adecuación	Ajuste de la técnica en el caso de estudio.
		Identificación de datos	Esfuerzo para identificar los datos de la técnica.
		Dependencias	Relaciones de la técnica con otros recursos.
		Fuentes de información	Información sobre la técnica.
		Herramientas	Herramientas utilizadas por la técnica.

<i>Nivel</i>	<i>Elemento</i>	<i>Atributo</i>	<i>Descripción</i>
	Personal	Experiencias	Experiencia requerida para usar la técnica.
		Conocimientos	Conocimientos requeridos para utilizar la técnica
	Resultados	Entorno	Entorno proporcionado por el conjunto de casos de estudio de la técnica.
		Eficacia	Capacidad del conjunto de datos para detectar defectos con la técnica.
		Propiedades	Propiedades del modelo de requerimiento propuesta por la técnica.
<i>Uso</i>	Proyectos	Referencia	Proyectos donde se ha utilizado la técnica.
		Personal	Personal que ha utilizado la técnica.
		Lenguaje	Lenguaje utilizado por la técnica.
	Satisfacción	Opinión	Opinión general sobre la utilización de la técnica.
		Beneficios	Beneficios de usar la técnica.
		Problemas	Problemas con el uso de la técnica.

El análisis consiste en responder al esquema de categorización presentado en la Tabla 3. Las Tablas 4,5,6 y 7 presentan las respuestas para cada técnica.

Tabla 4. Análisis de la técnica T1

<i>Nivel</i>	<i>Elemento</i>	<i>Atributo</i>	<i>Descripción</i>
<i>Operativo</i>	Objetivo	Función	Busca patrones estructurales simples dentro del Gráfico de Sintaxis Abstracta (ASG) generado a partir de la especificación de requerimientos.
		Etapa	Etapa inicial del desarrollo del software.
		Elementos	Requerimientos formales de software.
	Técnica	Comprensibilidad	Medio.
		Entradas	Especificación de requerimientos
		Adecuación	Describe una jerarquía de funciones en el dominio automotriz mediante la búsqueda de patrones simples dentro del ASG generado a partir de la especificación de requerimientos.
		Identificación de datos	Medio.
		Dependencias	Se puede relacionar con modelos gráficos basados en el Modelo de Lenguaje Unificado (UML).
		Fuentes de información	https://ieeexplore.ieee.org/abstract/document/5954453 .
		Herramientas	Reclipse Tool Suite, detecta patrones de diseño estáticos y dinámicos utilizando ingeniería inversa con eclipse.

<i>Nivel</i>	<i>Elemento</i>	<i>Atributo</i>	<i>Descripción</i>
			https://www.researchgate.net/publication/22155383 3 Reverse Engineering with the Reclipse Tool Suite.
	Personal	Experiencias	Lenguaje Natural Controlado (CNL).
		Conocimientos	Técnica ASG.
	Resultados	Entorno	Industria Automotriz.
		Eficacia	Control de funciones de confort de vehículos. Se logró validar los requerimientos de las funcionalidades del bloqueo de puertas del sistema interconectando con las funcionalidades ya existentes.
		Propiedades	Inconsistencia, completitud y defectos en los requerimientos.
<i>Uso</i>	Proyectos	Referencia	Sistema eléctrico de vehículos.
		Personal	Ingeniero de Requerimientos.
		Lenguaje	Inglés.
	Satisfacción	Opinión	La técnica puede ser factible para otros dominios de aplicación.
		Beneficios	Verifica automáticamente si los requerimientos formalizados violan ciertas reglas, que consisten en restricciones lógicas y pautas para especificaciones de alta calidad. Además, ofrece operaciones para la corrección de requerimientos.
		Problemas	Es necesario tener un conocimiento previo de la técnica utilizada para poder adaptar los artefactos que contienen información específica del dominio.

Tabla 5. Análisis de la técnica T2

<i>Nivel</i>	<i>Elemento</i>	<i>Atributo</i>	<i>Descripción</i>
<i>Operativo</i>	Objetivo	Función	Aplica patrones de análisis semántico de normas para la validación de requerimientos.
		Etapa	Etapa inicial del desarrollo del software.
		Elementos	Requerimientos formales de software.
	Técnica	Comprensibilidad	Media
		Entradas	Especificación de requerimientos
		Adecuación	Propone generar las normas de una biblioteca para la validación de los requerimientos del sistema.
		Identificación de datos	Medio
		Dependencias	Casos de uso y modelado de procesos comerciales.
		Fuentes de información	https://ieeexplore.ieee.org/document/6359965

<i>Nivel</i>	<i>Elemento</i>	<i>Atributo</i>	<i>Descripción</i>
		Herramientas	CommonRules de IBM para el PLN. Analizador de Stanford para el análisis de oraciones y normas. https://nlp.stanford.edu/software/lex-parser.shtml Etiquetador de Stanford para el etiquetado de partes del discurso https://nlp.stanford.edu/software/tagger.shtm
	Personal	Experiencias	Lenguaje Natural Controlado (LNC).
		Conocimientos	Armado de patrones de normas comerciales.
	Resultados	Entorno	Educativo.
		Eficacia	No proporciona información.
		Propiedades	Defectos en los requerimientos.
<i>Uso</i>	Proyectos	Referencia	Sistema biblioteca.
		Personal	Analistas de requerimientos.
		Lenguaje	Inglés.
	Satisfacción	Opinión	La técnica se aplicó sobre un escenario de normas de biblioteca, lo que resultó relativamente simple, debido a que casi todos los actores estaban en conocimiento del dominio de la biblioteca.
		Beneficios	La técnica brinda una ventaja adicional sobre las revisiones e inspecciones periódicas, ya que genera inferencias que se pueden comparar con el resultado esperado de un escenario en el mundo real. Esto ayuda a solucionar cualquier discrepancia en los requerimientos en una etapa temprana en lugar de esperar a que la fase de prueba descubra los defectos, lo que reduce tanto el costo como el tiempo dedicado al desarrollo de software.
		Problemas	La técnica se vuelve confusa en oraciones complejas para identificar normas que no se expresan como patrones simples de “si-entonces”.

Tabla 6. Análisis de la técnica T3

<i>Nivel</i>	<i>Elemento</i>	<i>Atributo</i>	<i>Descripción</i>
<i>Operativo</i>	Objetivo	Función	Aplica análisis semántico y minería de grafos para encontrar patrones comunes e identificar requerimientos propensos a errores.
		Etapa	En la etapa inicial del desarrollo del software.
		Elementos	Requerimientos formales de software.
	Técnica	Comprensibilidad	Media.
		Entradas	Especificación de requerimientos.

<i>Nivel</i>	<i>Elemento</i>	<i>Atributo</i>	<i>Descripción</i>
		Adecuación	Menú interactivo de restaurante. Valida el error del requerimiento verdadero como útil y el requerimiento falso positivo como no útil. Sistema de control de estacionamiento. Analiza las prioridades de los requerimientos y determinar el error del requerimiento verdadero y requerimientos falsos positivos luego del análisis semántico. Sistema de organización de préstamos. Clasifica los requerimientos para aislar los requerimientos erróneos de los requerimientos falsos positivos.
		Identificación de datos	Medio.
		Dependencias	No proporciona información.
		Fuentes de información	https://www.researchgate.net/publication/328313448
		Herramientas	Algoritmo LSA (Análisis Semántico Latente), determina matemáticamente temas dentro de un documento https://github.com/loana-P/pca_and_clustering_for_edu_purposes/blob/master/newsgroups_LSA.ipynb . Algoritmo LDA (Asignación de Dirichlet Latente), encuentra temas a los que pertenece el documento en función de las palabras que contiene. https://github.com/lda-project/Lda
	Personal	Experiencias	Lenguaje Natural Controlado (LNC).
		Conocimientos	Análisis semántico y métricas de evaluación. Minería de grafos.
	Resultados	Entorno	Comercial.
		Eficacia	Sistema interactivo de restaurante. Los dos modelos de entrenamiento probados obtienen una tendencia similar a las métricas de evaluación, el primero con el 65% y el segundo con el 45%. Sistema de control de estacionamiento. Modelo de películas se desempeñó mejor que el modelo de revisiones. Sistema de organización de préstamos. Los clasificadores de conjuntos separaron las fallas de no fallas con una precisión muy alta, en algunos casos entre el 85% y el 89%.
		Propiedades	Defectos en los requerimientos.
<i>Uso</i>	Proyectos	Referencia	Sistema interactivo de restaurante. Sistema de control de estacionamiento. Sistema de organización de préstamos.
		Personal	Profesionales industriales. Ingeniero de requerimientos.
		Lenguaje	Inglés.
		Satisfacción Opinión	No existe literatura reportada sobre el uso de análisis semántico y minería de grafos para mejorar la calidad de los requerimientos en lenguaje natural registrados en la especificación de requerimientos. El uso de técnicas LSA y LDA

<i>Nivel</i>	<i>Elemento</i>	<i>Atributo</i>	<i>Descripción</i>
			para escribir requerimientos de alta calidad es un enfoque novedoso.
		Beneficios	La solución de minería ayuda a identificar los requerimientos que están más conectados y pueden haber sido afectados por las correcciones durante el ciclo de inspección.
		Problemas	No proporciona información.

Tabla 7. Análisis de la técnica T4

<i>Nivel</i>	<i>Elemento</i>	<i>Atributo</i>	<i>Descripción</i>
Operativo	Objetivo	Función	Aplica modelos semánticos y encuentra similitudes o relaciones entre los requerimientos.
		Etapa	En cualquier etapa del desarrollo de software.
		Elementos	Requerimientos informales de software.
	Técnica	Comprensibilidad	Media.
		Entradas	Requerimientos informales de software y los conjuntos de datos de dominio relacionado.
		Adecuación	La técnica está en los primeros pasos del desarrollo se aplicó en un sistema de gestión de cursos donde los profesores pueden crear contenido en línea para los estudiantes y administrar cursos en línea.
		Identificación de datos	Medio.
		Dependencias	Metodología Scrum.
		Fuentes de información	https://dl.acm.org/doi/10.1145/3305160.3305166
		Herramientas	Algoritmos Naive Bayes y Support Vector Machine (SVM) e Integration Definition for Function Modeling (IDEF0)
Personal	Experiencias	Lenguaje Natural Controlado (LNC).	
	Conocimientos	Modelos semánticos.	
Resultados	Entorno	Educativo.	
Uso		Eficacia	La técnica ofrece una lista de probabilidades de más de 300 entradas por cada palabra en el vocabulario de requerimiento.
		Propiedades	Inconsistencia en los requerimientos.
		Proyectos	Referencia
		Personal	Experto en ciencia de datos, ingenieros de software senior, clientes y gerente de proyecto de software senior.
		Lenguaje	Inglés.
	Satisfacción	Opinión	La técnica encuentra defectos en cualquier fase del desarrollo de software basado en los requerimientos textuales, los conjuntos de datos de dominio y los modelos semánticos.

<i>Nivel</i>	<i>Elemento</i>	<i>Atributo</i>	<i>Descripción</i>
		Beneficios	La técnica es independiente del dominio del software y del idioma de los requerimientos textuales.
		Problemas	La restricción de la técnica es el conjunto limitado de datos disponibles, por ejemplo, leyes, reglamento, estándares de la organización entre otros.

4. Resultados

Del análisis de los resultados obtenidos de los enfoques indicados en la Tabla 2 y los aspectos evaluados en los enfoques en las Tablas 4, 5, 6, y 7 se extraen diferentes características que son calificadas como puntos de valoración para los enfoques seleccionados. En la Tabla 8 se enumeran las características identificadas, en la cual se indica con el vocablo SI que el enfoque contribuye a la característica, mientras que el vocablo NO indica que el enfoque no contribuye a la característica.

Tabla 8. Comparación de los enfoques de validación de requerimientos.

<i>Técnica</i>	T1	T2	T3	T4
<i>Característica</i>				
Requerimientos formales	SI	SI	SI	NO
Requerimientos informales	NO	NO	NO	SI
Especificación de requerimientos	SI	SI	SI	NO
Correctitud	SI	SI	SI	NO
Consistencia	SI	NO	NO	SI
Completitud	SI	NO	NO	NO
Análisis sintáctico	SI	NO	NO	NO
Análisis semántico	NO	SI	SI	SI
Algoritmos de aprendizaje automático	SI	SI	SI	SI
Uso de patrones	SI	SI	SI	NO
Ciclo de vida del software	NO	NO	NO	SI
Dominio específico	SI	SI	SI	NO
Uso de métricas	NO	NO	SI	NO
Experiencia en LNC	SI	SI	SI	SI
Participación cliente	NO	NO	NO	SI
Idioma específico	SI	SI	SI	NO

Del análisis comparativo realizado entre los enfoques seleccionados se extraen los siguientes resultados:

1. Las técnicas T1, T2, T3 validan requerimientos formales de software, mientras que T4 valida requerimientos informales de software. Las entradas necesarias para aplicar las técnicas T1, T2 y T3 es la especificación de requerimientos, mientras

- que la técnica T4 se aplica en los requerimientos informales de software y los conjuntos de datos de dominio relacionado.
2. La técnica T1 aplica las propiedades de consistencia, completitud y correctitud en los requerimientos, la técnica T2 y T3 aplica la propiedad de correctitud en los requerimientos y finalmente la técnica T4 aplica la propiedad de consistencia en los requerimientos.
 3. Las técnicas T2, T3 y T4 utilizan el análisis semántico, mientras que T1 utiliza el análisis sintáctico en la validación de requerimientos.
 4. Las técnicas T1, T2, T3 y T4 emplean algoritmos de aprendizaje automático como herramientas para la validación de requerimientos y requieren personal con experiencia en Lenguaje Natural Controlado.
 5. Las técnicas T1, T2, T3 validan las propiedades de los requerimientos utilizando patrones, mientras que T4 valida las propiedades de los requerimientos mediante la relación de categorías semánticas.
 6. Las técnicas T1, T2 y T3 se utilizan en la etapa inicial del desarrollo del software, mientras que la técnica T4 se puede utilizar en cualquier etapa del desarrollo del software.
 7. Las técnicas T1, T2 y T3 se desarrollaron para un dominio específico, mientras que T4 es una técnica independiente del dominio del software.
 8. La técnica T3 emplea métricas para evaluar los resultados de la validación de los requerimientos.
 9. Las técnicas T1, T2, T3 y T4 requieren personal con experiencia en LNC para su implementación.
 10. La técnica T4 es la única que incorpora al cliente en la validación de requerimientos.
 11. Las técnicas T1, T2 y T3 se implementan en un idioma específico de requerimientos, mientras que T4 es independiente del idioma.

5. Conclusiones y Trabajos Futuros

Este trabajo se ha focalizado en un proceso fundamental de la Ingeniería de Requerimientos: la Validación. Se han preseleccionado 32 trabajos para posteriormente centralizarse en 4 de ellos. Para responder a las preguntas de evaluación de las cuatro técnicas de PLN seleccionadas se identificaron las contribuciones de los trabajos en el proceso de Validación de Requerimientos.

Las técnicas T2, T3 y T4 se centran en el análisis semántico, es decir, en el significado contextual de las palabras en la validación de requerimientos, mientras que la técnica T1 en el análisis sintáctico, es decir, analiza la disposición de las palabras en una oración para que tengan sentido gramatical en la validación de requerimientos.

Adicionalmente, las técnicas T1, T2 y T3 validan la propiedad de correctitud de los requerimientos utilizando patrones en la especificación de requerimientos y se aplican en la etapa inicial del desarrollo del software. La técnica T4 valida la propiedad de consistencia de los requerimientos mediante la relación de categorías semánticas en todo el ciclo de vida del desarrollo del software, esto se debe a que la técnica es

independiente del dominio del software y del idioma de los requerimientos textuales. Además, hay que considerar que la técnica T4 está en los primeros pasos del desarrollo.

Las técnicas proporcionan diferencias en el enfoque de dominio en la validación de requerimientos con diversos grados de éxito. Este éxito depende de la naturaleza de la organización en sí y del conocimiento profundo del negocio para adaptar la técnica a las necesidades del negocio y del usuario.

Se evidencia el escaso uso de indicadores de rendimiento para el control y seguimiento de los defectos en los requerimientos, así como también objetivos a ser evaluados por dichos indicadores.

Todas las técnicas requieren personal con experiencia en LNC debido a que aplican algoritmos de aprendizaje automático en la validación de requerimientos, es decir, el personal debe tener conocimiento sobre lenguaje natural y la aplicación de restricciones de vocabulario, gramática y/o semántica.

Para fortalecer el proceso de validación de requerimientos, como trabajo futuro se propone crear un proceso de validación a partir de la técnica de PLN seleccionada, donde se definen las actividades del proceso y se describen objetivos, tareas, roles, productos de entradas y salidas para cada actividad utilizando un entorno colaborativo, que permitiría a las diferentes partes interesadas en un proyecto validar requerimientos. De esta manera, se reduciría el tiempo y esfuerzo necesarios para validar los requerimientos y se aseguraría que todos los stakeholders estén de acuerdo en cuanto a las entidades relevantes y requerimientos del proyecto.

Referencias

1. R. Chillarege, W. L. Kao, R. G. Condit: Defect Type and its Impact on the Growth Curve, in the 13th IEEE International Conference on Software Engineering, Austin, Texas, May 13-17, (1991).
2. Reyes Juárez Ramírez: Towards improving user interfaces: a proposal for integrating functionality and usability since early phases, IEEE, Indonesia, (2011).
3. Nikora Alen P.: Experiments in Automated Identification of Ambiguous Natural-Language Requirements, (2011).
4. D. Berry, E. Kamsties: The syntactically dangerous all and plural in specification, IEEE Software, pp. 55-57, (2005).
5. xAli Memon Kamran, Xiaoling Xia: Deciphering and Analyzing Software Requirements employing the techniques of Natural Language Processing, 4th International Conference on Mathematics and Artificial Intelligence (ICMAI 2019), Association for Computing Machinery, New York, NY, USA, 153–156, (2019).
6. S. Peshterliev, J. Kearney, A. Jagannatha, I. Kiss, S. Matsoukas: Active Learning for New Domains in Natural Language Understanding, (2018).
7. Nor Aiza Mokhtar, Massila Kamalrudin, Mohd Yusof Mokhtar, Sidek Safiah: A review on requirements validation for software development, Journal of Theoretical and Applied Information Technology, (2018).
8. J. Holtmann, J. Meyer, M. von Detten: Automatic Validation and Correction of Formalized, Textual Requirements, IEEE Fourth International Conference on Software Testing, Verification and Validation Workshops, Berlin, Germany, 2011, pp. 486-495, (2011).
9. R. Sharma, K. K. Biswas: Using norm analysis patterns for automated requirements validation, Second IEEE International Workshop on Requirements Patterns (RePa), Chicago, IL, USA, pp. 2328, (2012).

10. M. Singh: Automated Validation of Requirement Reviews: A Machine Learning Approach, IEEE 26th International Requirements Engineering Conference (RE), Banff, AB, Canada, pp. 460-465, (2018).
11. Issa Atoum: A Scalable Operational Framework for Requirements Validation Using Semantic and Functional Models, in Proceedings of the 2nd International Conference on Software Engineering and Information Management (ICSIM '19), Association for Computing Machinery, New York, NY, USA, 1-6, (2019).
12. S. Vegas, V. Basili: A characterisation schema for software testing techniques, Empirical Softw. Engg., vol. 10, no. 4, pp. 437-466, (2005).