

# Integrated Requirements Engineering Framework with Intelligent Systems and Semantic Reasoning for the Extraction of Application Vocabularies

Alexandra Corral<sup>1</sup>[0000-0002-9225-1346], Luis E. Sanchez<sup>2</sup>[0000-0003-0086-1065] and Leandro Antonelli<sup>3</sup>[0000-0003-1388-0337]

<sup>1</sup> University of the Armed Forces ESPE, Sangolquí, Ecuador  
[macorral@espe.edu.ec](mailto:macorral@espe.edu.ec)

<sup>2</sup> GSyA Research Group, University of Castilla-La Mancha, Ciudad Real, Spain  
[luise.sanchez@uclm.es](mailto:luise.sanchez@uclm.es)

<sup>3</sup> Lifia – Facultad de Informatica, National, University of La Plata, La Plata, Argentina  
[lanto@lifia.info.unlp.edu.ar](mailto:lanto@lifia.info.unlp.edu.ar)

**Abstract.** Domain discovery is an essential activity in the early stages of software development, problems of understanding and communication are common in development teams, due to the excessive handling of information from different sources, generating great effort in: consensus, interpretation and transfer of useful information for the application. For this reason, it has been seen the need to merge artificial intelligence with requirements engineering processes that mimic human behavior, through the creation of a vocabulary for the application, improving the Extended Lexicon of Language LEL, using Natural Language Processing, ontologies for semantic reasoning and heuristic rules to improve completeness.

**Keywords.** Software Engineering, Software Requirements, Artificial Intelligence Techniques, Natural Language Processing.

## 1 Introducción

Los problemas comunes en el equipo de desarrollo, es la falta de comunicación y comprensión del dominio del Discurso. Integrar y procesar excesiva información de diferentes fuentes puede generar errores de interpretación del problema.

El conocimiento humano es una de las fuentes principales para obtener información preliminar en el conjunto de requerimientos iniciales. Elicitar el conocimiento humano es una tarea que implica de mucho esfuerzo en equipos de desarrollo de software para entender el dominio de la aplicación.

Dominios como por ejemplo la agricultura, requiere en gran medida del conocimiento experto, siendo varios factores que influyen en la diversidad del conocimiento para el desarrollo de aplicaciones como: conocimiento humano, conocimiento ambiental, conocimiento tecnológico y de infraestructura, convertidos en sistemas con terminología compleja de entender y consensuar entre los diferentes stakeholders.

El descubrimiento y conocimiento del dominio generalmente están dados en fases tempranas del desarrollo de software, por lo que la integración de aplicaciones no convencionales provenientes de la inteligencia artificial que imiten un comportamiento similar al de la inteligencia humana, podría beneficiar al equipo de desarrollo en la resolución de problemas comunes en la ingeniería de requisitos.

De esta manera se pretende generar un Framework para la construcción de un vocabulario de la aplicación utilizando técnicas de inteligencia artificial sobre dominios específicos que permitan acortar brechas entre el conocimiento humano y el conocimiento de la aplicación

## 2 Estado del Arte

La creación de vocabularios para la aplicación es relativamente nueva desde el punto de vista semántico sobre conceptos de un determinado dominio. Algunas investigaciones han dado sus aportes sobre el uso del PLN para extraer conceptos relevantes de especificaciones de requerimientos para ayudar a generar modelos UML como [1] sin establecer semántica. Estudios recientes, sobre la generación de vocabularios están en [2], cuyo aporte se basa en la extracción de entidades, frases nominales y verbales utilizando el Lenguaje natural, para la transformación de modelos ya existentes, genera algoritmos basados en patrones y cuya evaluación de precisión es a través de métricas que no supera al 0,69 %.

## 3 Declaración del Problema y Contribuciones

Estudios anteriores extraen conceptos relevantes utilizando PLN, el análisis semántico se centra en la identificación de sinónimos, homónimos, antónimos, hipónimos, entre otros, sin establecer una semántica en un contexto determinado, que permitan aproximar al impacto y connotación de la aplicación.

En este sentido la contribución pretende resolver dicho problema dando mejoras sobre:

- La generación de patrones para el análisis complejo de las estructuras gramaticales, extraídas del lenguaje natural de los *stakeholders*, daría lugar a la transformación de conceptos del dominio del problema a conceptos de la aplicación mediante la utilización de PLN. Dichos patrones permitirán:

- Analizar la completitud sobre conceptos como sujetos y objetos constituidos en el LEL
- Analizar acciones del discurso del dominio a través de la categorización de verbos, permitiendo ampliar símbolos del LEL.
- Determinar la semántica del vocabulario utilizando ontologías sobre dominios específicos y a través de relaciones aproximar a escenarios, generando connotación sobre el impacto de la aplicación.
- Generar Heurísticas para determinar la completitud del vocabulario.

## **4 Metodología y enfoque**

La metodología de la investigación, para el desarrollo del Framework de construcción del vocabulario en sistemas inteligentes con razonamiento semántico, se basa en la aplicación del método científico, mismo que permitirá obtener los resultados esperados a través de la deducción e inducción sobre la temática, ya que revisando estados del arte, se seleccionó los más relevantes, para disponer de los insumos necesarios a fin de procesar el lenguaje natural, generando información para el análisis lingüístico de patrones de estructuras construyendo de este modo el vocabulario extendido. En la elaboración de los procesos y/o protocolos es indispensable la observación directa previa de comportamiento del proceso manual y automatizado del uso del lenguaje, para que antes de concretar su formalización se establezca las métricas de evaluación que finalmente logren concretar el entendimiento del dominio y entendimiento del lenguaje común del Stakeholder.

## **5 Plan de Evaluación**

El presente informe de simposio doctoral, es de etapa temprana; el cumplimiento de la hipótesis se validará en función de la calidad de resultados obtenidos en cumplimiento de los patrones establecidos que logren las mejores estructuras gramaticales, a través del uso de ontologías de dominio extraer el vocabulario adecuado, para que sea validado a través de métricas y revisión de expertos.

## **6 Conclusiones**

La contribución pretende a través de un Framework la construcción de vocabularios para la aplicación sobre dominios específicos, mediante la incorporación de inteligencia y semántica en las etapas iniciales de la ingeniería de requisitos.

Estudios preliminares no han demostrado aún una precisión en la extracción de conceptos del dominio ni han establecido la semántica por completo para generar vocabularios, lo que nuestra propuesta tiene pertinencia para dar solución a estos problemas.

## References

1. A. A. Almazroi, L. Abualigah, M. A. Alqarni, E. H. Houssein, A. Q. M. AlHamad, y M. A. Elaziz, «Class Diagram Generation from Text Requirements: An Application of Natural Language Processing», en *Deep Learning Approaches for Spoken and Natural Language Processing*, Springer, 2021, pp. 55-79.
2. P. Danenas, T. Skersys, y R. Butleris, «Natural language processing-enhanced extraction of SBVR business vocabularies and business rules from UML use case diagrams», *Data & Knowledge Engineering*, vol. 128, p. 101822, 2020.