# Detección Temprana de Escenarios Duplicados en Español mediante Enriquecimiento Semántico basado en LEL

Gabriela Pérez[1,2], Catalina Mostaccio[1], and Leandro Antonelli[1,3]

[1] LIFIA, Facultad de Informática, Universidad Nacional de La Plata
[2] UNAJ IIyA, Universidad Nacional Arturo Jauretche
[3] CAETI, Facultad de Tecnología Informática, Universidad Abierta Interamericana
{gperez, catty, lanto}@lifia.info.unlp.edu.ar

**Resumen** La detección temprana de escenarios duplicados es un problema relevante en ingeniería de requisitos, especialmente cuando distintos analistas describen una misma situación con vocabulario diferente. Los modelos de embeddings capturan regularidades lingüísticas generales, pero no conocimiento conceptual específico del dominio. En este trabajo se propone una estrategia de enriquecimiento semántico que integra el Léxico Extendido del Lenguaje (LEL), ampliando cada escenario con definiciones del dominio antes de generar sus embeddings. La estrategia se evalúa sobre un dataset de 126 pares de escenarios agrícolas en español anotados por tres analistas. Los resultados muestran mejoras significativas en la correlación con las evaluaciones humanas, alcanzando un Spearman de 0.83 en la configuración título-objetivo, frente a 0.57 del baseline sin enriquecimiento. Los resultados sugieren que reutilizar artefactos propios del proceso de ingeniería de requisitos como fuente de conocimiento estructurado constituye una estrategia efectiva y de bajo costo de adopción, particularmente valiosa en etapas tempranas donde los escenarios contienen poca información y los modelos generales no disponen de suficiente contenido textual para capturar el vocabulario del dominio.

**Keywords:** Ingeniería de Requerimientos · Escenarios · LEL · Similaridad Semántica · NLP · Sentence BERT

## 1. Introducción

En ingeniería de requisitos, los escenarios son artefactos ampliamente utilizados para describir el comportamiento esperado de un sistema [1], [2], [3], ya que se expresan en lenguaje natural y no requieren formalismos complejos, lo que facilita su producción y comprensión tanto por desarrolladores como por clientes. Además, la especificación de requisitos suele ser una actividad colaborativa en la que diferentes integrantes del equipo describen distintos aspectos del sistema. Esta dinámica puede dar lugar a la creación de escenarios superpuestos o duplicados. Esto puede deberse a que se utilizó terminología diferente para expresar

Proceedings of the 29th Workshop on Requirements Engineering (WER2026),
August 11-13, 2026, La Plata, Argentina.
https://doi.org/10.29327/1884517.29-5

una misma situación, o a la necesidad de crear un escenario adicional como una extensión de uno ya existente.

En este contexto, es fundamental contar con herramientas que permitan la detección temprana de escenarios similares, mediante análisis semántico que funcione incluso cuando se utiliza terminología diferente. Estas herramientas deben adaptarse al contexto regional, y estar disponibles como software de código abierto, facilitando su implementación en proyectos locales, particularmente en América Latina donde existe una escasez significativa de herramientas especializadas de ingeniería de requisitos.

En los últimos años, los grandes modelos de lenguaje (LLMs) han revolucionado el procesamiento del lenguaje natural. Entre las técnicas más utilizadas para medir la similitud semántica se encuentra el uso de embeddings, que representan textos como vectores en un espacio semántico. Modelos como BERT y Sentence-BERT [5] han mostrado buenos resultados en diversas tareas de similitud, pero al ser generales, no siempre capturan conocimiento específico del dominio. Por ejemplo, en agricultura, expresiones como "regar" y "distribuir agua" pueden ser equivalentes según el contexto, algo que no siempre reflejan los embeddings genéricos.

Por otro lado, el Léxico Extendido del Lenguaje (LEL) es un artefacto de Ingeniería de Requisitos que organiza el conocimiento del dominio mediante la definición de términos clave. Cada término se describe por su significado en el contexto (noción) y sus efectos o relaciones (impactos). Así, el LEL establece un vocabulario común entre clientes y desarrolladores, reduciendo ambigüedades e incorporando conocimiento específico del proyecto de forma incremental.

A partir de estos dos artefactos surge la pregunta de si el enriquecimiento de escenarios con definiciones provenientes del LEL puede mejorar la detección temprana de escenarios duplicados, especialmente en etapas iniciales donde los escenarios suelen contener información limitada.

En este trabajo proponemos una estrategia de enriquecimiento semántico basada en el LEL que integra conocimiento explícito del dominio con técnicas modernas de representación vectorial para mejorar la detección temprana de escenarios similares. Los escenarios se enriquecen incorporando las definiciones asociadas a los términos del LEL antes de generar sus embeddings, lo que permite que las representaciones resultantes capturen no solo regularidades lingüísticas generales, sino también conocimiento conceptual específico del dominio. La estrategia no requiere reentrenamiento ni ajuste fino del modelo de lenguaje. Además, se define un protocolo de evaluación alineado con la naturaleza incremental de la ingeniería de requisitos, distinguiendo entre configuraciones de comparación homogénea e incremental, y se construye un dataset en español del dominio agrícola anotado por expertos para analizar el impacto del conocimiento del dominio en la detección temprana de escenarios duplicados. A lo largo del trabajo se utiliza el término analista de requisitos para referirse al profesional responsable de las actividades de elicitación y especificación, rol que en otros contextos puede denominarse ingeniero de requisitos.

El resto del trabajo se organiza de la siguiente manera. La sección 2 describe los trabajos relacionados. La sección 3 presenta brevemente los conceptos fundamentales. La sección 4 propone la estrategia de enriquecimiento con LEL. La sección 5 presenta los resultados de la evaluación. La sección 6 discute los resultados. Por último, la sección 7 presenta las conclusiones y el trabajo futuro.

## 2. Trabajos Relacionados

La estimación de similitud semántica entre textos ha sido ampliamente estudiada en el campo del Procesamiento del Lenguaje Natural. En términos generales, los enfoques pueden agruparse en tres categorías: (i) métodos basados en similitud léxica, (ii) métodos basados en representaciones distribucionales y (iii) métodos basados en conocimiento estructurado.

Los métodos léxicos estiman la cercanía entre textos a partir del solapamiento de términos o caracteres, utilizando métricas como Jaccard o TF-IDF con similitud del coseno. Sin embargo, su dependencia de coincidencias explícitas limita la detección de equivalencias conceptuales cuando se emplea terminología diferente, y esto sucede en ambientes colaborativos.

Los métodos basados en representaciones distribucionales permiten superar parcialmente estas limitaciones. La introducción de Google BERT [4] representó un avance significativo en la obtención de representaciones contextuales profundas. Posteriormente, [5] propusieron Sentence-BERT, optimizado para generar embeddings comparables directamente mediante similitud del coseno. Si bien estos modelos capturan relaciones semánticas más complejas, las comparaciones basadas únicamente en representaciones generales no resultan suficientes en dominios especializados.

Finalmente, los enfoques basados en conocimiento estructurado incorporan información explícita mediante ontologías, recursos léxicos o grafos semánticos. Un ejemplo clásico es Princeton University WordNet [6], que permite enriquecer la comparación textual a través de relaciones como sinonimia o hiperonimia. Sin embargo, estos enfoques suelen depender de recursos externos de propósito general o requerir procesos de integración complejos, y no reutilizan necesariamente artefactos producidos durante el propio proceso de ingeniería de requisitos.

En este contexto, el Léxico Extendido del Lenguaje (LEL), introducido por [7], [8], constituye un mecanismo específico para capturar y estructurar conocimiento del dominio. El LEL ha sido utilizado para reducir ambigüedades y favorecer la coherencia conceptual del proyecto. Sin embargo, no se han reportado trabajos previos que integren explícitamente el LEL con modelos modernos de embeddings para la detección temprana de escenarios duplicados, particularmente en español y considerando la naturaleza incremental del proceso de especificación.

La estrecha relación entre el LEL y los escenarios ha sido reconocida desde los primeros trabajos sobre estos artefactos. [9] muestra cómo los escenarios se construyen a partir del LEL, que permite comprender el vocabulario específico del dominio, y cómo ambos artefactos se retroalimentan durante la elicitación.

Esta relación fundamenta la estrategia propuesta: si los escenarios se redactan utilizando el vocabulario del LEL, resulta natural enriquecerlos con sus definiciones antes de generar sus representaciones vectoriales.

En [10] evaluaron diferentes modelos de embeddings mediante similitud coseno, comparando su desempeño en un conjunto de escenarios. Posteriormente, en [11] extendieron este trabajo, comparando técnicas de whitening, prompting y reformulación textual del título del escenario con combinación ponderada de scores. Aunque se obtuvieron mejoras en las métricas de correlación, estas fueron acotadas y se mantuvieron dentro del marco de representaciones generales, sin incorporar conocimiento explícito del dominio ni contemplar configuraciones incrementales propias del proceso de especificación.

El presente trabajo avanza sobre estas limitaciones mediante una estrategia de enriquecimiento semántico que integra conocimiento estructurado proveniente del LEL antes de la generación de embeddings. El proceso comienza con la identificación automática de los sintagmas presentes en el escenario, para lo cual se utiliza el método propuesto en [12], basado en patrones morfosintácticos sobre textos cortos. A diferencia de los enfoques basados en reformulación o ajuste fino, la propuesta no modifica el modelo ni delega la interpretación conceptual exclusivamente en el LLM, sino que incorpora definiciones construidas y validadas por expertos del dominio.

## 3. Background

En un proceso de especificación de requisitos, el conocimiento del dominio suele recopilarse inicialmente a partir de documentación existente, entrevistas con los clientes y otras fuentes. A partir de este material se construye el LEL, que captura y unifica el vocabulario del dominio, reduciendo ambigüedades y malentendidos entre las partes. Los términos definidos se utilizan posteriormente en la redacción de escenarios, que describen el comportamiento esperado del sistema en lenguaje natural. Asimismo, la propia elaboración de escenarios puede enriquecer el LEL mediante la incorporación de nuevos términos del dominio aún no definidos. A continuación, se presentan estos conceptos junto con las técnicas empleadas para la comparación semántica de escenarios.

### 3.1. Léxico Extendido del Lenguaje (LEL)

El Léxico Extendido del Lenguaje (LEL) es un modelo utilizado en ingeniería de requisitos que surge de la idea de comprender el lenguaje del dominio sin necesidad de comprender completamente el problema [7], [13]. Su propósito es capturar y estructurar el vocabulario propio de la aplicación, preservando el significado que los actores asignan a los términos utilizados.

Los símbolos del LEL se clasifican en cuatro categorías: Sujetos (entidades activas), Objetos (entidades pasivas), Verbos (acciones realizadas por los sujetos) y Estados (condiciones de sujetos u objetos). Su construcción sigue los principios

clásicos del LEL, como la circularidad y el vocabulario mínimo, que favorecen la coherencia semántica del léxico.

Una vez definidos los primeros términos centrales del dominio en el LEL, el equipo puede comenzar a redactar los escenarios. Ambos artefactos evolucionan de manera incremental a lo largo del proceso.

### 3.2. Escenarios

Los escenarios describen el funcionamiento de un sistema mediante narraciones, facilitando su comprensión. Son fáciles de usar tanto para desarrolladores como para expertos del dominio, mejoran la comunicación y pueden aplicarse en distintas etapas del desarrollo de software. Leite [14] define un escenario con los siguientes atributos: (i) un título; (ii) un objetivo que debe ser alcanzado a través de la ejecución del escenario; (iii) un contexto que establece el punto de partida; (iv) los recursos, que son objetos físicos o información que debe estar disponible; (v) los actores, que son agentes que realizan las acciones; y (vi) el conjunto de episodios. Cada episodio representa acciones que son realizadas por los actores utilizando los recursos disponibles. Dado que los escenarios se redactan de forma incremental, la comparación semántica con los existentes debe funcionar incluso cuando sólo se dispone del título. Las técnicas de embeddings permiten hacerlo y se presentan a continuación.

### 3.3. Modelos de lenguaje y embeddings

Los modelos de embeddings permiten representar textos como vectores en un espacio de alta dimensionalidad, de modo que textos con significados similares se ubican cercanos entre sí. En el campo del procesamiento del lenguaje natural, los Sentence Transformers están diseñados para generar representaciones vectoriales de oraciones completas, optimizadas para tareas de similitud semántica y basadas en la arquitectura Transformer [15]. Estos modelos capturan el significado global de una oración, lo que los hace particularmente adecuados para comparar textos cortos como los títulos y objetivos de los escenarios.

Una vez obtenidos los embeddings, la similitud entre dos escenarios se calcula mediante la similitud del coseno, que mide el ángulo entre dos vectores y produce valores entre 0 y 1, donde los valores cercanos a 1 indican mayor similitud semántica. Para evaluar en qué medida estas similitudes se corresponden con las valoraciones humanas, se emplean métricas de correlación estadística, que se presentan a continuación.

### 3.4. Métricas de evaluación

Para evaluar el rendimiento de los modelos se utilizan los coeficientes de correlación de Pearson ($r$) y Spearman ($\rho$), que miden el grado de asociación entre dos variables. En este caso, se comparan las similitudes predichas por el modelo con las evaluaciones de referencia etiquetadas en el dataset.

El coeficiente de Pearson mide la relación lineal entre las puntuaciones de similitud generadas por el modelo y las del dataset. Por su parte, el coeficiente de Spearman evalúa la relación monótona entre ambas variables, centrándose en el orden o ranking de las similitudes más que en sus valores absolutos.

## 4. Estrategia de Enriquecimiento Semántico basada en LEL

La estrategia propuesta busca detectar escenarios duplicados de forma temprana, idealmente cuando un desarrollador comienza a redactar uno nuevo. En ese momento, el único contenido disponible suele ser el título, o bien el título y el objetivo, ya que el contexto, los episodios, los recursos y los actores se incorporan de manera incremental. Por este motivo, la comparación se realiza utilizando únicamente estos campos, calculando la similitud del coseno entre los embeddings de los escenarios.

### 4.1. Identificación de términos

Para identificar los términos del LEL en el texto de un escenario, se implementó un extractor de sintagmas basado en patrones morfosintácticos implementado con spaCy. El proceso comienza con un análisis lingüístico, donde spaCy segmenta el texto en tokens y asigna a cada uno su categoría gramatical (part-of-speech tagging). A partir de este análisis, se identifican secuencias de tokens que coinciden con patrones definidos manualmente para capturar construcciones sintácticas típicas del español. Dado que distintos patrones pueden solaparse en un mismo fragmento de texto, se aplica una estrategia de resolución de conflictos que privilegia los sintagmas de mayor longitud.

Posteriormente, un módulo de postprocesamiento elimina secuencias que comienzan con preposición o finalizan en elementos funcionales sin carga semántica relevante (por ejemplo, determinantes, conjunciones o pronombres), garantizando que los constituyentes extraídos sean lingüísticamente coherentes.

La identificación a nivel de sintagma resulta fundamental, ya que el significado de un término del dominio no siempre puede inferirse a partir de sus componentes individuales. Por ejemplo, “distribuir agua” denota una actividad específica del dominio agrícola cuyo significado no se deriva simplemente de “distribuir” o “agua” considerados de forma aislada. Los sintagmas extraídos se comparan posteriormente con los términos definidos en el LEL para identificar correspondencias con el vocabulario del dominio.

### 4.2. Estrategia de enriquecimiento

La estrategia propuesta consiste en enriquecer el texto de cada escenario con conocimiento específico del dominio proveniente del LEL antes de generar su representación vectorial. El proceso comienza con la identificación automática de los sintagmas presentes en el escenario. Una vez extraídos, se comparan con los

símbolos definidos en el LEL mediante similitud semántica. Este enfoque permite identificar coincidencias incluso cuando el sintagma no coincide exactamente con el nombre del símbolo, por ejemplo reconociendo ”controlar plagas” como similar a ”control de plaga”. Si la similitud supera el umbral de 0.95, se considera que el sintagma corresponde a ese símbolo del LEL. Este valor fue seleccionado para garantizar alta precisión en la identificación, ya que un umbral elevado reduce los falsos positivos y asegura que solo se incorporen definiciones del LEL cuando la correspondencia conceptual es clara. Luego, las definiciones asociadas (noción e impactos) se incorporan al texto mediante concatenación. De esta manera, los embeddings generados reflejan tanto la descripción original como la información vinculada a los términos del dominio que aparecen en él. En aquellos casos en que un término identificado no esté definido en el LEL, el sistema genera automáticamente una definición preliminar utilizando un modelo de lenguaje (LLM), Gemma3. Esta definición se presenta al analista para su revisión y aprobación, pero no se incorpora al léxico del proyecto de forma automática. El analista puede editarla, refinarla y, si la considera adecuada, integrarla formalmente al LEL. De este modo, el conocimiento incorporado al léxico continúa siendo validado por el usuario, preservando su corrección y relevancia para el dominio. Cabe señalar que, en los experimentos reportados, la mayoría de los sintagmas identificados en los escenarios del dataset contaban con entrada en el LEL construido; los casos de generación automática fueron minoritarios y su impacto individual en los resultados no fue analizado de forma aislada, constituyendo una línea de trabajo futuro.

## 5. Evaluación Experimental

Esta sección presenta los resultados de la evaluación de la estrategia de enriquecimiento con LEL sobre el dataset de escenarios del dominio de cultivo de tomate.

### 5.1. Dataset

El dataset utilizado consiste en 126 pares de escenarios en español, pertenecientes al dominio del cultivo de tomate. Cada par incluye dos escenarios y una etiqueta de similitud calculada como el promedio de las evaluaciones realizadas por tres analistas de requisitos en una escala de 1 a 5. Un puntaje 5 indica que los escenarios son equivalentes, 4 que son mayormente equivalentes con diferencias menores, 3 que son aproximadamente equivalentes pero con diferencias en aspectos clave, 2 que no son equivalentes pero comparten algunos detalles, y 1 que no son equivalentes aunque pertenecen al mismo dominio. La escala no incluye el valor 0, ya que todos los escenarios pertenecen al mismo dominio de aplicación y comparten inevitablemente actores, recursos o contexto general. Incluir pares completamente diferentes no representaría situaciones reales en un proceso de especificación de requisitos, donde todos los escenarios

describen actividades del mismo proyecto. Cada escenario está estructurado siguiendo la notación de Leite, con los campos título, objetivo, contexto, recursos, actores y episodios. El dataset incluye escenarios sobre actividades agrícolas como fumigar, regar, podar, cosechar, trasplantar, desmalezar, sembrar e injertar, entre otras. Los pares cubren combinaciones tanto de escenarios idénticos como de escenarios funcionalmente distintos dentro del mismo dominio, lo que permite evaluar la capacidad de discriminación semántica de los modelos. El dataset está disponible en [16].

Un ejemplo de par etiquetado con 1 corresponde a un escenario de siembra comparado con otro de control de enfermedades. La calidad del proceso de anotación fue evaluada mediante métricas de acuerdo entre anotadores. Se obtuvo un Krippendorff de 0.854 (métrica de intervalo), lo que indica un nivel de acuerdo sustancial entre los tres analistas. La etiqueta final de similitud de cada par se calculó como el promedio de las tres evaluaciones, lo que produce valores continuos en el rango [1.0, 5.0].

La distribución presenta una leve asimetría hacia los extremos. En particular se observan 47 pares en el rango bajo (1.0–2.33, 37.3 %), 31 en el rango medio (2.67–3.33, 24.6 %) y 48 en el rango alto (3.67–5.0, 38.1 %). El dataset está compuesto por 41 escenarios únicos, cubriendo distintos niveles de similitud dentro del dominio, desde escenarios que apenas comparten el contexto general hasta escenarios idénticos. La menor representación del rango medio refleja la naturaleza del dominio, donde los escenarios tienden a describir actividades claramente distintas o claramente relacionadas, siendo menos frecuentes los casos de similitud parcial. Cabe destacar que aun los pares del rango bajo, aunque funcionalmente distintos, pertenecen al mismo dominio y comparten inevitablemente vocabulario, actores y contexto general.

### 5.2. Construcción del LEL

El LEL utilizado en los experimentos contiene 163 símbolos, de los cuales 112 son verbos, 31 objetos, 13 estados y 7 sujetos. Los verbos predominan ya que el dominio del cultivo de tomate se caracteriza por una alta densidad de actividades (regar, podar, fumigar, trasplantar, injertar, desmalezar, cosechar, entre otras). Como se describió en la sección anterior, cuando un escenario contiene sintagmas sin entrada asociada en el LEL, el sistema puede generar definiciones preliminares mediante Gemma3 12b, que el analista puede revisar y decidir si incorpora formalmente al léxico o utiliza únicamente como apoyo semántico para la comparación. Las definiciones capturan conocimiento conceptual del dominio agrícola y no reproducen el contenido específico de los episodios, por lo que el enriquecimiento no introduce información trivialmente redundante.

### 5.3. Configuraciones evaluadas

Para evaluar la estrategia de enriquecimiento con LEL, se consideraron cuatro niveles de completitud de los escenarios organizados en dos formas de comparación, homogénea e incremental. Los niveles de completitud son:

- Título (T): solamente se tiene en cuenta el título del escenario.
- Título y Objetivo (TO): representa el estado inicial de un escenario.
- Título, Objetivo y Contexto (TOC): representa un escenario en desarrollo.
- Título, Objetivo, Contexto, Episodios (TOCE): representa un escenario terminado.

Los campos Actores y Recursos se omitieron ya que suelen mencionarse explícitamente dentro de los episodios y, por lo tanto, su incorporación como sección separada introduce redundancia sin aportar información adicional. Asimismo, dado que múltiples escenarios comparten los mismos recursos (por ejemplo, herramientas o insumos agrícolas), su inclusión podría incrementar artificialmente la similitud entre escenarios funcionalmente distintos. La comparación homogénea evalúa el caso en que ambos escenarios se representan considerando los mismos niveles de completitud (por ejemplo, T, TO, TOC o TOCE). La comparación incremental, en cambio, evalúa un caso de uso más cercano a la práctica real, donde un desarrollador comienza a redactar un nuevo escenario y lo compara, utilizando una representación parcial (T, TO o TOC), contra escenarios del repositorio que se representan considerando todos sus campos. De este modo, se analiza el comportamiento del modelo cuando se enfrentan representaciones con distinta cantidad de información disponible.

### 5.4. Resultados: comparación homogénea

La Tabla 1 (A) presenta los resultados de la comparación homogénea, tomando como baseline la versión sin enriquecimiento con LEL. Enriquecer únicamente uno de los escenarios produce resultados inferiores al baseline, debido a la desalineación semántica que introduce. La ganancia al enriquecer ambos es más pronunciada en TO, donde la correlación pasa de 0.57 a 0.83, y es menor a medida que el escenario cuenta con más texto, dado que el propio contenido ya incorpora conocimiento del dominio.

Se evaluaron variantes incorporando únicamente la noción, únicamente los impactos, o ambos componentes, sin observarse diferencias significativas. Los resultados reportados corresponden a la variante con entrada completa del LEL.

### 5.5. Resultados: comparación incremental

La Tabla 1 (B) muestra los resultados de la comparación incremental. Cuando el escenario en construcción contiene únicamente el título, los baselines son muy bajos, reflejando que un título aislado ofrece poca información semántica para la comparación. El enriquecimiento con LEL compensa esta limitación de manera notable, y se maximiza cuando se enriquecen ambos lados. Cuando el escenario en construcción ya cuenta con título y objetivo (TO), la mejora al enriquecer ambos lados sigue siendo significativa, aunque enriquecer el escenario del repositorio cuando este ya es muy completo puede introducir redundancia y reducir levemente la correlación.

A medida que el escenario en construcción cuenta con más texto, las mejoras se vuelven más modestas. Cuando ya incluye título y objetivo, enriquecer ambos lados sigue siendo beneficioso. Enriquecer un escenario completo puede introducir redundancia y reducir levemente la correlación. Con título, objetivo y contexto, el LEL prácticamente no aporta, confirmando que el enriquecimiento es más valioso cuanto menos información tiene el escenario en construcción.

**Tabla 1.** Resultados de las configuraciones presentadas sobre el dataset de agricultura.

| **(A) Configuración homogénea** | | **Spearman** | **Pearson** |
|---|---|---|---|
| T | T | 0.6634 | 0.6607 |
| T + LEL | T | 0.4751 | 0.4677 |
| T + LEL | T + LEL | 0.7487 | 0.7540 |
| TO | TO | 0.5733 | 0.5436 |
| TO + LEL | TO | 0.6156 | 0.5625 |
| TO + LEL | TO + LEL | 0.8307 | 0.8143 |
| TOC | TOC | 0.6534 | 0.6575 |
| TOC + LEL | TOC | 0.6560 | 0.6525 |
| TOC + LEL | TOC + LEL | 0.8180 | 0.8033 |
| TOCE | TOCE | 0.7542 | 0.7567 |
| TOCE + LEL | TOCE | 0.7649 | 0.7642 |
| TOCE + LEL | TOCE + LEL | 0.7831 | 0.7805 |
| **(B) Configuración incremental** | | **Spearman** | **Pearson** |
| T | TO | 0.1674 | 0.1566 |
| T + LEL | TO | 0.4193 | 0.3991 |
| T + LEL | TO + LEL | 0.6549 | 0.6583 |
| T | TOC | 0.0351 | 0.0436 |
| T + LEL | TOC | 0.4182 | 0.4183 |
| T + LEL | TOC + LEL | 0.6072 | 0.5923 |
| T | TOCE | 0.0601 | 0.0464 |
| T + LEL | TOCE | 0.4224 | 0.4391 |
| T + LEL | TOCE + LEL | 0.4312 | 0.4392 |
| TO | TOC | 0.6540 | 0.6614 |
| TO + LEL | TOC | 0.6703 | 0.6678 |
| TO + LEL | TOC + LEL | 0.8301 | 0.8120 |
| TO | TOCE | 0.6841 | 0.6675 |
| TO + LEL | TOCE | 0.7392 | 0.7105 |
| TO + LEL | TOCE + LEL | 0.7077 | 0.6999 |
| TOC | TOCE | 0.7062 | 0.6827 |
| TOC + LEL | TOCE | 0.7120 | 0.6869 |
| TOC + LEL | TOCE + LEL | 0.7003 | 0.6896 |

Los experimentos se realizaron utilizando el modelo paraphrase-multilingual-mpnet-base-v2 [17], disponible en Hugging Face [18], que soporta más de 50 idiomas incluido el español y genera embeddings de 768 dimensiones. Es un modelo Sentence Transformer optimizado para similitud semántica y se utiliza exclusivamente para la generación de representaciones vectoriales. La búsqueda de términos del LEL se realizó mediante similitud semántica con un umbral de 0.95. La similitud entre escenarios se calculó mediante similitud del coseno, y el rendimiento se evaluó con los coeficientes de correlación de Pearson y Spearman.

### 5.6. Aplicando la estrategia en un caso real

Para ilustrar el impacto del enriquecimiento con LEL en un caso concreto, se propone un análisis cualitativo en el que se simula la creación de nuevos escenarios. El experimento consiste en comparar cinco títulos nuevos contra un conjunto de 14 escenarios previamente definidos detallados en la Tabla 2.

**Tabla 2.** Escenarios seleccionados para el análisis.

| **id** | **Título** | **id** | **Título** |
|---|---|---|---|
| 1 | Eliminar las malezas | 8 | Cosechar los tomates de forma manual |
| 2 | Quitar las malas hierbas | 9 | Realizar el podado de las plantas |
| 3 | Controlar las plagas | 10 | Controlar las plagas e insectos |
| 4 | Despuntar las inflorescencias | 11 | Regar las plántulas de tomate |
| 5 | Regar las plantas de tomate | 12 | Cosechar los tomates en racimos |
| 6 | Controlar las enfermedades bacterianas | 13 | Controlar las enfermedades virales |
| 7 | Prevención de enfermedades fungosas | 14 | Realizar la poda de forma manual |

Los títulos nuevos fueron diseñados para asegurar diversidad sintáctica y, en varios casos, no comparten vocabulario con los escenarios existentes, lo que representa el escenario más desafiante para los modelos de embeddings. Para cada título nuevo se devuelven los escenarios más similares, ordenados por similitud del coseno. Los resultados esperados fueron validados por expertos y se presentan en la Tabla 3. Se debe destacar que no existe una única respuesta correcta, ya que la interpretación puede variar según el criterio del analista.

**Tabla 3.** Resultados de la encuesta realizada a expertos.

| | **Título nuevo escenario** | **Resultados esperados** |
|---|---|---|
| Título 1 | Realizar fumigación para controlar plagas | id 3, id 6, id 7, id 10, id 13 |
| Título 2 | Recortar ramas de la planta | id 4, id 9, id 14 |
| Título 3 | Distribuir agua en los cultivos | id 5, id 11 |
| Título 4 | Erradicar vegetación indeseada | id 1, id 2 |
| Título 5 | Recolectar los tomates maduros | id 8, id 12 |

Dado que el título es el primer elemento disponible al comenzar a redactar un escenario, este análisis se restringe a la configuración T, comparando el baseline T|T contra T+LEL|T+LEL. Los resultados se presentan en la Tabla 4, donde se muestran los 5 resultados mas similares y se destacan en negrita los escenarios que coinciden con los resultados esperados. El enriquecimiento con LEL mejora la comparación, especialmente en los casos en los que el título nuevo no comparte vocabulario con los escenarios existentes. Para ”Distribuir agua en los cultivos”, la configuración sin LEL no recupera ningún resultado esperado entre los cinco más similares, mientras que con LEL los escenarios id 11 y id 5, ambos relacionados con riego, aparecen en los dos primeros lugares. Un comportamiento similar se observa en “Erradicar vegetación indeseada”, donde con LEL los escenarios id 1 e id 2 pasan a ocupar las primeras posiciones. En “Recolectar los tomates maduros”, ambas configuraciones recuperan los resultados esperados, pero con LEL aparecen en las dos primeras posiciones en lugar de la tercera y cuarta. En “Recortar ramas de la planta”, la configuración sin LEL

recupera 2 de 3 resultados esperados, mientras que con LEL recupera los 3. El único caso donde ambas configuraciones muestran un comportamiento similar es "Realizar fumigación para controlar plagas", donde el vocabulario compartido con los escenarios existentes ya es suficiente para una buena recuperación. Estos resultados confirman que el LEL actúa como puente semántico en situaciones donde el vocabulario utilizado difiere entre escenarios que describen actividades equivalentes.

**Tabla 4.** Resultados obtenidos con T-T y T+LEL-T+LEL.

| | **Título 1** (5 rtas) | **Título 2** (3 rtas) | **Título 3** (2 rtas) | **Título 4** (2 rtas) | **Título 5** (2 rtas) |
|---|---|---|---|---|---|
| **T \| T** | **id 3 (0.88)**<br>**id 10 (0.83)**<br>**id 7 (0.71)**<br>**id 6 (0.66)**<br>id 1 (0.60) | **id 9 (0.81)**<br>id 2 (0.69)<br>**id 4 (0.65)**<br>id 11 (0.58)<br>id 5 (0.57) | id 9 (0.54)<br>**id 5 (0.48)**<br>id 2 (0.47)<br>id 4 (0.44)<br>**id 11 (0.41)** | id 4 (0.77)<br>**id 2 (0.76)**<br>id 9 (0.67)<br>**id 1 (0.57)**<br>id 5 (0.56) | id 11 (0.91)<br>id 5 (0.89)<br>**id 8 (0.84)**<br>**id 12 (0.82)**<br>id 9 (0.59) |
| **T+LEL \| T+LEL** | **id 3 (0.84)**<br>**id 10 (0.80)**<br>**id 7 (0.74)**<br>**id 6 (0.67)**<br>id 1 (0.65) | **id 9 (0.74)**<br>id 2 (0.67)<br>**id 4 (0.67)**<br>id 14 (0.64)<br>id 12 (0.59) | **id 11 (0.67)**<br>**id 5 (0.67)**<br>id 1 (0.57)<br>id 7 (0.54)<br>id 2 (0.52) | **id 1 (0.85)**<br>**id 2 (0.83)**<br>id 9 (0.70)<br>id 4 (0.66)<br>id 14 (0.65) | **id 8 (0.80)**<br>**id 12 (0.78)**<br>id 11 (0.58)<br>id 5 (0.58)<br>id 9 (0.56) |

## 6. Discusión

El enriquecimiento semántico basado en el LEL resulta particularmente efectivo en etapas tempranas de redacción, cuando el escenario contiene poca información. En estos casos, los modelos preentrenados de propósito general no disponen de suficiente contenido textual para capturar el conocimiento conceptual del dominio, mientras que las definiciones del LEL incorporan conocimiento estructurado y validado por expertos. Todas las definiciones utilizadas en el enriquecimiento, ya sea provenientes del LEL o generadas automáticamente por el modelo de lenguaje, fueron revisadas y validadas por un experto del dominio antes de su uso, garantizando la confiabilidad del proceso.

En contraste, cuando los escenarios están completamente desarrollados, la ganancia es menor, ya que el enriquecimiento no actúa simplemente como 'más texto' sino como incorporación de conocimiento que resulta más valioso cuando dicho conocimiento no está explícitamente presente en el escenario.

Asimismo, los resultados de la comparación incremental refuerzan la viabilidad del enfoque en un contexto realista de especificación. El LEL actúa como puente semántico entre artefactos en diferentes niveles de madurez, permitiendo que un escenario parcial enriquecido se compare efectivamente con escenarios completos. Esto es especialmente relevante porque refleja la situación real en la que un analista comienza a redactar un nuevo escenario y necesita detectar duplicados antes de contar con toda la información, momento en que el LEL aporta el contexto conceptual necesario para que la comparación sea efectiva.

El trabajo evidencia el valor de reutilizar artefactos propios de la ingeniería de requisitos como fuente de conocimiento, sin depender de recursos externos de propósito general. No obstante, los resultados deben interpretarse considerando ciertas limitaciones. El dataset pertenece a un único dominio, el cultivo de tomate, por lo que no es posible afirmar que se replicará el mismo comportamiento. Además, la calidad del enriquecimiento depende de la precisión del extractor de sintagmas y de la calidad del LEL construido, ya que una identificación incompleta de términos o definiciones poco precisas podría reducir el impacto observado. Finalmente, la estrategia se basa en una concatenación textual simple, lo que abre interrogantes sobre posibles mejoras mediante técnicas más sofisticadas de selección o ponderación del contenido incorporado. Además, el umbral de similitud utilizado para la identificación de términos del LEL (0.95) fue fijado por criterio de precisión y no explorado sistemáticamente, por lo tanto, variaciones en este parámetro podrían afectar los resultados en dominios con terminología más ambigua. Estas limitaciones delimitan el alcance de las conclusiones y motivan el trabajo futuro mostrado en la siguiente sección.

## 7. Conclusiones y Trabajo Futuro

Este trabajo propone y evalúa una estrategia de enriquecimiento semántico para mejorar la detección de similitud entre escenarios escritos en español, ampliando cada escenario con las definiciones de los términos del dominio presentes en el LEL antes de generar su representación vectorial. Los experimentos sobre un dataset de 126 pares de escenarios agrícolas, anotados por tres analistas, muestran mejoras consistentes en la correlación con las evaluaciones humanas. La mejor configuración alcanzó un Spearman de 0.83 frente a 0.57 del baseline, evidenciando que el LEL introduce conocimiento específico del dominio que los modelos preentrenados de propósito general no capturan. Enriquecer ambos lados de la comparación permite obtener mejoras significativas. Hacerlo únicamente en el escenario nuevo desplaza su representación hacia una región distinta del espacio semántico, mientras que enriquecer ambos permite que los vectores se aproximen y la similitud coseno recupere su capacidad discriminativa. La estrategia no requiere corpus previo y puede aplicarse desde el inicio del proyecto, lo que la hace adecuada para equipos que trabajan en español, donde la disponibilidad de herramientas para ingeniería de requisitos es limitada.

Como trabajo futuro se propone explorar la construcción automática del LEL a partir de entrevistas con clientes, reduciendo el esfuerzo inicial requerido. Asimismo, se prevé una evaluación formal del mecanismo de generación automática de definiciones implementado en este trabajo, incluyendo la evaluación de la calidad de las definiciones producidas y la validación de la estrategia en dominios adicionales para determinar el alcance de su generalización.

## Referencias

1. Alexander, I., Maiden, N.: Scenarios, stories, and use cases: the modern basis for system development. Computing Control Engineering Journal 15(5), 24–29 (2004).

2. Carroll, J. M.: Five reasons for scenario-based design. Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences, (1999).
3. Antonelli, L., Delle Ville, J., Dioguardi, F., Fernandez, A., Tanevitch, L., Torres, D.: An Iterative and Collaborative Approach to Specify Scenarios using Natural Language. Workshop on Requirements Engineering (WER) 2022.
4. Devlin, J., Chang, M., Lee, K., Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. 2018. http://arxiv.org/abs/1810.04805 (2018)
5. Reimers, N., Gurevych, I. : Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). (2019)
6. Miller, G. A. (1995). WordNet: A lexical database for English. *Communications of the ACM*, 38(11), 39–41. https://dl.acm.org/doi/10.1145/219717.219748
7. J. C. S. d. P. Leite and A. P. M. Franco, "A strategy for conceptual model acquisition," [1993] Proceedings of the IEEE International Symposium on Requirements Engineering, San Diego, CA, USA, 1993, pp. 243-246, doi: 10.1109/ISRE.1993.324851.
8. Leite, J.C.S.d.P., Rossi, G., Balaguer, F., et al.: Enhancing requirements baseline with scenarios. Requirements Engineering Journal, vol. 2, no. 4, pp. 184–198 (1997).
9. Thomas, P. J. y Oliveros, A. Identificación de objetivos a partir de LEL y escenarios. En: *V Workshop de Investigadores en Ciencias de la Computación (WICC)*, 2003.
10. Pérez, G., Mostaccio, C., Maltempo, G., Antonelli, L. Evaluation of natural language processing models to measure similarity between scenarios written in Spanish. Cadernos do IME: Série Informática: Vol. 50, pp. 43-56. e-ISSN: 2317-2193(online). DOI: 10.12957/cadinf.2024.87935. Diciembre 2024
11. Pérez, G., Mostaccio, C., Antonelli, L. Análisis comparativo de arquitecturas de NLP para detectar similitudes entre escenarios en español. Proceedings of the 28th Workshop on Requirements Engineering (WER2025), Agosto 20-22, 2025, Rio de Janeiro-RJ, Brazil. https://doi.org/10.29327/1588952.28-3
12. Pérez, G., Mostaccio, C., Maltempo, G., Bogado, J., Antonelli, L. Automatic identification of domain expressions to build a LEL glossary from Spanish texts. ICSE 2026 - 3rd Workshop on Multi-disciplinary, Open, and integRatEd Requirements Engineering (MO2RE) - Abril, 2026.
13. Antonelli, L. , Rossi, G., Leite, J.C.S.d.P., Oliveros, A. Buenas prácticas en la especificación del dominio de una aplicación, 2013. Workshop in Requirements Engineering (WER), Montevideo, Uruguay, April 8 – 10 (2013)
14. Leite, J.C.S.P., Hadad, G., Doorn, J.H., Kaplan, G.: A Scenario Construction Process. Requirements Engineering, 5(1), 38–61 (2000). https://doi.org/10.1007/PL00010342
15. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, L., Polosukhin, I. : Attention is all you need - 31st Conference on Neural Information Processing Systems, pages 5998–6008 (NIPS 2017), Long Beach, CA, USA. (2017).
16. Perez, G., Mostaccio, C., Antonelli, L. (2026). Dataset de Similitud entre Escenarios en Español para Ingeniería de Requisitos. Dominio de Cultivo de Tomate (1.0). Zenodo. https://doi.org/10.5281/zenodo.20371156
17. Reimers, N., Gurevych, I.: Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation. EMNLP 2020.
18. Hugging Face, https://huggingface.co/, Accedido febrero 2026.