

Alineación de glosarios específicos de dominio

Paola Grijalva-Arriaga¹ 3[0000-0003-2616-417X], Galo Cornejo-Gómez² [[0000-0003-3523-2795],
Leandro Antonelli³ 5[0000-0003-1388-0337], Pablo Thomas⁴ [[0000-0001-9861-987X]

¹Computación e Informática, Universidad Agraria del Ecuador, Guayaquil, Ecuador

²Ciencias de la Computación, Universidad Católica de Guayaquil, Guayaquil, Ecuador

³Lifia, Fac. de Informática, UNLP, La Plata, Bs As, Argentina

⁴Lidi, Fac. de Informática, UNLP, La Plata, Bs As, Argentina

⁵CAETI - Facultad de Tecnología Informática - Universidad Abierta Interamericana

paola.grijalva@info.unlp.edu.ar, galo.cornejo.uc@ucsg.edu.ec,
leandro.antonelli@lifia.info.unlp.edu.ar,
pablo.thomas@lidi.info.unlp.edu.ar

Resumen. Los glosarios son una parte importante de cualquier documento de requerimientos de software, ya que hace explícitos los términos técnicos en un dominio y proporciona definiciones, ayudando a mitigar la imprecisión y la ambigüedad. El Léxico Extendido del Lenguaje (LEL) es un glosario que describe el vocabulario del dominio caracterizando a cada expresión a través de dos atributos (a diferencia de un glosario tradicional que tiene una sola descripción). En la actualidad, es muy común la interoperación de los sistemas informáticos, en donde cierto sistema brinda servicios para que consuman otros sistemas. En este marco, es crucial comprender el lenguaje de ambos sistemas, para que la interacción sea efectiva. Este artículo propone un método de alineación de glosarios LEL del dominio. El método de alineación propuesto se basa en identificación de discrepancias y similitudes, para luego acordar el significado de los términos similares. Este artículo presenta una evaluación preliminar con resultados promisorios utilizando como caso de estudio un modelo de calidad para instituciones de educación superior.

Palabras Claves: Alineación de Dominios. Requerimientos. Léxico Extendido del Lenguaje.

1 Introducción

En la elicitación de requerimientos, una de las herramientas utilizadas son los glosarios. Martin Glinz [1] define un Glosario como una colección de definiciones de términos que son relevantes en algún dominio, que contiene referencias cruzadas, sinónimos, homónimos, siglas y abreviaturas. Arora et al. [2] indican que un glosario es una parte importante de cualquier documento de requerimientos de software, ya que hace explícitos los términos técnicos en un dominio y proporciona definiciones, ayudando a mitigar la imprecisión y la ambigüedad.

Leite et al. [3] definen el LEL (Léxico Extendido del Lenguaje) como un glosario que describe el vocabulario de la aplicación, sin necesidad de comprender la funcionalidad del proceso del negocio también llamado dominio. El LEL es una técnica utilizada para capturar y estructurar el conocimiento sobre un dominio específico, esencialmente construyendo un glosario detallado. Los glosarios de dominio establecen definiciones claras y estándar para los términos utilizados en el contexto del proyecto. Los glosarios permiten la aclaración del significado de conceptos del dominio, unificando los términos empleados en los diferentes modelos, mejorando la comunicación entre los involucrados. Los glosarios de dominio son utilizados por los equipos de desarrollo de software para garantizar que todos los participantes del proyecto utilicen el mismo significado de los términos.

En la actualidad, es muy común la interoperación de los sistemas informáticos, en este caso los LEL al proporcionar un vocabulario consistente y bien definido, simplifican la integración de sistemas heterogéneos. Cuando los sistemas utilizan los mismos términos y definiciones, se reduce la necesidad de mapeo complejo de datos y traducción de conceptos entre sistemas. Esto facilita la creación de interfaces de comunicación estándar y la interoperabilidad entre sistemas que, de otro modo, podrían tener estructuras de datos o modelos conceptuales diferentes.

A nivel gubernamental, o entre empresas la garantía de que los sistemas sean interoperables significa que podamos acceder a más información y funcionalidades útiles desde un único entorno de manera práctica y fiable. Proporcionando eficiencia, rentabilidad, garantía de las conexiones abiertas de los productos de la empresa, mayor flexibilidad en los formatos de los documentos, soporte a los estándares de la industria y mayor competitividad [4]. Al realizar una alineación de los dominios de estos sistemas asegura que todos los aspectos relacionados con un dominio específico, como conceptos, terminología, y procesos, estén claramente comprendidos y coordinados entre los diferentes participantes del proyecto, encontrando similitudes y diferencias. Tradicionalmente, en la ingeniería de requerimientos mediante las entrevistas con expertos del dominio, el análisis del dominio con diagramas de flujo, casos de uso, UML, modelado de procesos del negocio, y validación constante con los stakeholders del dominio, se ha buscado entender el dominio ya que garantizan una comunicación efectiva y un entendimiento compartido entre los diversos actores involucrados en el desarrollo de sistemas o software como desarrolladores, diseñadores y stakeholders [5][6][7].

En este marco, si cada organización cuenta con glosarios, en donde cada uno describe su dominio, pero con perspectiva de que haya un borde de conexión, facilita la identificación de fronteras entre los dos dominios con la finalidad de permitir la interoperabilidad. En este caso, es crítico identificar la intersección, superposición y solapamiento de los glosarios del dominio, para que los sistemas se comuniquen y puedan intercambiar información. Por ejemplo, entre una institución académica (IES) y una institución gubernamental de certificación de calidad, la IES debe de garantizar que sus sistemas proporcionen toda la información y documentación basados en los estándares e indicadores de calidad solicitados por el organismo gubernamental. Por ejemplo, para la institución acreditadora de calidad el término "Planificación de Proyectos de Investigación" requiere como evidencia un listado de todos los proyectos de investigación que la IES debe ejecutar dentro de un mediano plazo. Sin embargo, para la IES este término representa los cronogramas de trabajo para la ejecución de cada proyecto de

investigación, lo que generaría que entreguen informes que no cubren con las expectativas del organismo de acreditación. Para que el sistema de la institución interopere con el sistema del organismo gubernamental, se debe asegurar que todos los sistemas involucrados en una red de interoperabilidad compartan una comprensión común de los conceptos clave, por lo que es necesario que los glosarios de ambos dominios se encuentren alineados para mejorar el análisis de requisitos al proporcionar un marco estructurado para capturar y organizar el conocimiento del dominio.

Por este motivo, este artículo propone un método de alineación utilizando glosarios del dominio representados a partir de los glosarios LEL. Este método de alineación de glosarios de dominio tiene por objetivo identificar las similitudes, diferencias u omisiones entre los glosarios de dominio de las organizaciones, aportando de esta manera con información que permita mejorar la interoperabilidad de sus sistemas.

La sección 2 describe el glosario LEL. La sección 3 presenta los trabajos relacionados. La sección 4 describe el método propuesto para la alineación de glosarios de dominio. La sección 5 presenta un caso de estudio. La sección 6 describe la herramienta de soporte y la sección 7 expone conclusiones y futuros trabajos.

2 Glosarios LEL

Un LEL es un glosario que tiene como finalidad registrar la definición de términos (símbolos) que pertenecen a un dominio [8]. Cada símbolo tiene un “nombre” que lo identifica, una “noción” y un “impacto” que lo describen [3]. La noción es el significado, la descripción del símbolo. Y el impacto, es la relación entre él y otros símbolos [9].

Cada símbolo del LEL puede pertenecer a una de las siguientes categorías: sujeto, objeto, verbo o estado. Los sujetos son elementos activos dentro del dominio. Los objetos son elementos pasivos en el dominio y son recursos o elementos que los sujetos utilizan. Los verbos son acciones realizadas por sujetos utilizando objetos. Y los estados son situaciones en que sujetos, objetos o verbos pueden estar involucrados [10]. La tabla 1 muestra cada categoría con sus características y cómo describirlas.

Tabla 1. Categorías del LEL.

Categoría	Características	Noción	Impacto
Sujeto	Elemento activo que realiza una acción	Característica o condición que el sujeto satisface	Acciones que el sujeto realiza
Objeto	Elemento pasivo con la que el sujeto realiza una acción	Características o atributos que ese objeto tiene.	Acciones que son realizadas con el objeto.
Verbo	Acciones que los sujetos realizan	Objetivo que persigue el verbo	Procedimiento necesario para completar la acción
Estado	Situaciones en las que sujetos y objetos pueden encontrarse.	Situación representada	Acciones que deben ser realizadas para cambiar de estado.

3 Trabajos relacionados

Entre los trabajos relacionados se encuentran aquellos donde analizan el glosario LEL buscando similitudes, conflictos, omisiones, así como también la construcción de este de manera colaborativa, con la finalidad de alinearlos al dominio.

Sebastian et al. [8] proponen una variante de la técnica de inspección para detectar defectos en código fuente, modelos de diseño y modelos de requerimientos, basada en la construcción de mapas conceptuales para verificar un modelo léxico (LEL), debido a que suele presentar problemas de completitud, y de ambigüedad por el uso del lenguaje natural. La inspección propuesta detecta principalmente omisiones y ambigüedades. Esta propuesta de inspección se enmarcó en la construcción de mapas conceptuales, a través del cual se identifican defectos.

Litvak et al. [11] analizan todo el glosario LEL buscando conflictos de acuerdo a la siguiente clasificación desarrollada por los autores: semánticos, estructurales y sintácticos. Desarrollaron también un proceso y guías para su resolución. En la evaluación preliminar mostraron la importancia de identificar los conflictos y las soluciones para los conflictos propuestos. Mostrando la importancia de resolverlos para llegar a modelos de mejor calidad. El proceso constó de tres fases: 1) construcción de un mapa conceptual por cada símbolo del LEL; 2) análisis de conceptos de los mapas conceptuales; y 3) análisis de relaciones de cada mapa conceptual y entre mapas. A través de mapas conceptuales se pudo detectar a través de 4 modelos LEL diferentes: omisión de símbolos, omisión total o parcial de nociones e impactos, ambigüedad gramatical, ambigüedad por términos redundantes, por oraciones compuestas, por no cumplir los principios de circularidad y vocabulario mínimo, y símbolos no relevantes.

Hadad et al. [12] proponen una mejora en las heurísticas para facilitar la creación del LEL. Estas heurísticas se elaboraron en base a la identificación de las causas que provocaban las omisiones detectadas tanto reales como aparentes, estas últimas provocando otras falencias, tales como disminución en la comprensión del modelo, falta de homogeneidad en las descripciones contenidas en el mismo, ocultamiento de información relevante y dificultades en su validación.

Sebastián et al. [13] realizaron inspecciones en modelos léxicos de lenguaje natural en ingeniería de requisitos, aplicando tres variantes de listas de inspección: lista de comprobación, utilizando checklist; lectura de procedimientos, utilizando formularios y lectura constructiva, utilizando mapas conceptuales. Los resultados obtenidos demostraron que mediante la lectura constructiva se logró obtener una mayor tasa de detección de defectos, enfocándose en omisiones y ambigüedades. La lectura basada en procedimientos consumió más tiempo, pero capturo menos defectos en promedio.

Y. Liu et al. [14] proponen un enfoque para construir y desarrollar una red semántica de conceptos y frases de ingeniería del software, comenzando por un conjunto de glosarios existentes, utiliza los términos y asociaciones definidas explícitamente como punto de partida, utiliza técnicas basadas en aprendizaje automático para identificar y documentar dinámicamente asociaciones adicionales entre términos, aprovecha la red para interpretar consultas en lenguaje natural en el glosario, y finalmente aumenta la red semántica resultante con la retroalimentación proporcionada por los usuarios, permitiendo mejorar las consultas a través de la interfaces de lenguaje natural.

Otro grupo importante, es la relacionada con estudios sobre la alineación de ontologías que hacen uso de la semántica para conocer si dos conceptos son semejantes en cuanto a su significado en una ontología de dominio.

El método desarrollado por Alhassan BB [15] sugiere convertir semánticamente un par de ontologías de entrada al encontrar correspondencias semánticas. Amplían el proyecto Falcon-AO para encontrar sinónimos en dos ontologías diferentes y combinar alineaciones para el resultado final de este estudio, incorporando WordNet, una fuente de conocimiento léxico, como extensión. El Falcon-AO tiene cinco partes que se agregaron para encontrar términos. El primer paso es el pre-procesamiento; el segundo paso es la búsqueda de WordNet, el tercero es la creación del modelo, el cuarto es la especificación de la regla y el paso final es la estructura de datos y la fusión. El análisis comparativo reveló que el Falcon-AO extendido es mejor en términos de precisión.

Ngo D, Bellahsene Z [16] presenta un trabajo de verificación semántica para identificar de manera efectiva los conflictos entre un conjunto de mapeos, particularmente con respecto a la coincidencia de ontología a gran escala; realizó indexación estructural para emparejar ambas ontologías. Propone una medida de similitud semántica para determinar si dos clases en una gran ontología son potencialmente disjuntas. Luego, se establecen patrones para encontrar mapas de conflicto y se utiliza un algoritmo para eliminar la inconsistencia una vez que se identifica el conjunto de asignaciones de conflicto. YAM ++, un prototipo, implementó estas mejoras y participó en OEAI2013[15], obteniendo las primeras posiciones verificando los resultados de la evaluación sobre eficacia y eficiencia.

Hu X et al. [17] proponen un marco de alineación de ontologías de cuatro etapas (FOMF) para mejorar el rendimiento de las coincidencias. Se basa en la afirmación solamente aceptada de que una base de conocimiento exhaustiva externa se puede utilizar como puente semántico entre ontologías de dominio para el emparejamiento de ontologías. FOMF mapea semánticamente ontologías de dominio a una base de conocimiento y luego produce diferentes tipos de alineaciones, incluidas equivalencia, subclase y alineaciones de instancia. Las similitudes entre dos ontologías de dominio se emplean para mejorar la equivalencia y el descubrimiento de alineaciones. Finalmente, con base en las alineaciones adquiridas, las alineaciones inferidas se deducen para garantizar la integridad de los resultados de la coincidencia.

4 Método propuesto de alineación entre LELs

El propósito de un método de alineación entre glosarios LELs es encontrar las similitudes, omisiones y diferencias, entre dos glosarios contextualizados sobre dominios relacionados. Para la alineación, se utilizan dos glosarios, el primer glosario corresponde a un glosario máster identificado como LEL1 y el segundo, a un glosario secundario identificado como LEL2, este último se comparará contra el máster, en el proceso, obteniendo similitudes, diferencias y omisiones entre los glosarios. La Fig. 1 resume el método propuesto.

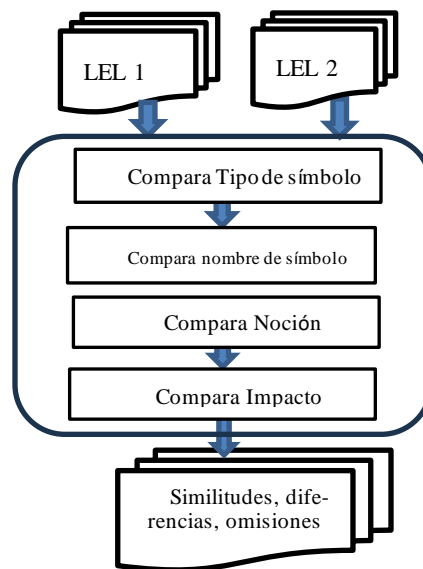


Fig. 1. Representación del método de alineación de glosarios

El método para la alineación entre los glosarios de dominio consiste en la ejecución de cuatro actividades i) la comparación de tipo de símbolos, ii) la comparación de los nombres de los símbolos, iii) la comparación de nociones, y iv) comparación de impactos.

Cada símbolo del glosario secundario es comparado contra todos los símbolos del glosario máster, obteniendo para cada uno de los símbolos del glosario secundario los siguientes estados: similitud, diferencia u omisión. Una *similitud*, es cuando el símbolo del glosario secundario coincide con el tipo de símbolo, nombre, noción e impacto de un símbolo en el glosario máster. Una *diferencia*, se establece cuando coincide en el tipo de símbolo, nombre del símbolo, sin embargo, puede existir variaciones en la noción o en el impacto descrito. De la misma manera, si la noción y el impacto del símbolo tienen coincidencia, sin embargo, el nombre o el tipo de símbolo difieren. Una *omisión*, es cuando el nombre del símbolo del glosario secundario no es encontrado en el glosario máster y no tiene coincidencias en nociones o impacto con ningún otro símbolo. En

caso de que existan símbolos en el glosario máster que no tuvieron su par en el glosario secundario, lo considerará como un *faltante*, como se describe en la Fig. 2.

El método presenta como resultados los pares de símbolos dentro de cada estado similitud, diferencias, omisiones y faltantes. Posteriormente, un experto en el dominio revisa, el listado de términos clasificados de acuerdo con estos estados, y decidirá si incluye o no los términos en la elaboración del glosario alineado, valorando así los resultados.

5 Caso de Estudio

Esta sección describe un caso de estudio que permite una evaluación preliminar del método para la alineación de los glosarios. Con la ejecución del procedimiento se realiza la comparación de los símbolos del LEL2 con relación al LEL1.

Los Glosarios corresponden al área de Investigación de una Institución de Educación Superior; por un lado un glosario elaborado a partir del manual de procesos y el otro elaborado a partir del Manual de Calidad del Consejo de Aseguramiento de la Calidad de la Educación Superior [18], ambos glosarios fueron elaborados por personal de sistemas de la institución de educación superior. Se procede posteriormente a ejecutar el método para alinear ambos glosarios, considerando como glosario secundario (LEL2) al manual de procesos y como glosario máster (LEL1), el manual de calidad. Los glosarios completos del área contienen entre 40 y 50 símbolos respectivamente. En la Tabla 2 se presentan algunos símbolos representativos del LEL1 y en la Tabla 3 se observa algunos símbolos representativos del LEL2.

La evaluación preliminar del método fue realizada con 11 símbolos del glosario máster y 8 símbolos del glosario secundario. Fue ejecutado por un ingeniero de requerimientos y los resultados fueron verificados por el responsable del área de calidad, quien actúa como experto humano.

Tabla 2. Extracto de los Símbolos del Glosario del Manual de Calidad CACES 2019

Tipo	Nombre	Noción	Impacto
Sujeto	La Institución	Es una Universidad o Escuela Politécnica que brinda servicios de educación superior y otorga títulos de tercer y cuarto nivel.	Desarrolla normativas y/o procedimientos para el desarrollo de la investigación. Aprueba las normativas y procedimientos para el desarrollo de la investigación.
Sujeto	Profesor Investigador	Persona que puede ser hombre o mujer. Posee un vínculo contractual con la institución.	Tiene asignado carga horaria en investigación. Planifica un proyecto de investigación. Elabora proyectos de investigación.

Objeto	Proyecto de Investigación	Documento que contiene una necesidad o un problema, hipótesis estrategias o acciones. Está relacionado a las líneas de investigación y/o dominios académicos institucionales.	El profesor investigador elabora el Proyecto de Investigación. El Profesor investigador propone el proyecto de investigación. Es seleccionado a través de un proceso de arbitraje.
Verbo	Planificación de los procesos de Investigación	Acción de planificar programas y/o proyectos de investigación científica y/o tecnológica, en coherencia con las líneas.	La institución planifica el desarrollo de programas de investigación científica. La institución planifica el desarrollo de proyectos de investigación.

Tabla 3. Extracto de símbolos del glosario del área de Investigación de la Institución

Tipo	Nombre	Noción	Impacto
Sujeto	Investigador Proponente	Persona con nombramiento de la institución que se dedica profesionalmente a la enseñanza, realiza investigación y actividades de vinculación.	Planifica un proyecto de investigación. Elabora proyectos de investigación.
Objeto	Proyecto de Investigación	Documento que contiene una necesidad o un problema, hipótesis estrategias o acciones.	El profesor investigador elabora el Proyecto de Investigación. El Profesor investigador propone el proyecto de investigación
Estado	Proyecto Aprobado	Inicia con la decisión favorable del Honorable Consejo Universitario para la asignación del financiamiento y recursos	Pasa al estado de ejecución del proyecto de investigación
Sujeto	Instituto de Investigación	Área de la institución que se encarga de la planificación, control y seguimiento a las actividades de investigación	Identifica entidades públicas y privadas para conseguir fondos internos o externos para los proyectos de investigación.

Al realizar la comparación de los símbolos, el símbolo "Profesor proponente" del glosario secundario, no tiene un símbolo con el mismo nombre en el glosario máster. Sin embargo, con el símbolo "profesor investigador", aunque no tienen exactamente el mismo nombre, coinciden en el tipo de símbolo, en la noción, ya que se expresa que realizan actividades de investigación y tienen algunas funciones similares, como por ejemplo la planificación y la elaboración de proyectos. La tabla 4 muestra la comparación con el símbolo más coincidente.

Tabla 4. Comparación símbolo Investigador Proponente

LEL 2 (Institución)	LEL 1 (Máster)	Nombre	Tipo	Noción	Impacto
Investigador Proponente	Profesor Investigador	No	Si	Si	Si

La comparación realizada del símbolo “Proyecto de Investigación” tiene similitud con el símbolo “Proyecto de investigación” como se observa en la Tabla 5, ambos coinciden en el tipo de símbolo, tienen similitud en la noción ya que ambos comparten la naturaleza de ser documentos que guían las actividades de investigación. Difieren en quien los realiza, propone, aprueba y ejecuta.

Tabla 5. Comparación símbolo Proyecto de Investigación

LEL 2 (Institución)	LEL 1 (Máster)	Nombre	Tipo	Noción	Impacto
Proyecto de Investigación	Proyecto de Investigación	Si	Si	Si	Si

El símbolo “Planificar proyecto de investigación” tiene mayor similitud con el símbolo “Planificación de los procesos de investigación” del LEL 1, como se observa en la Tabla 6, indica que ambos son verbos que describen acciones o procesos relacionados con la planificación de investigación. En relación con las nociones, el LEL 1 está relacionado con la acción de planificar programas y proyectos, así como mantener normativas y procedimientos, en cambio el LEL 2 está más relacionado a la definición de fechas y responsabilidades de los investigadores. Por otro lado, al comparar el impacto se enfoca en la planificación de obras literarias y el segundo en las acciones que realiza el investigador. Aunque ambos casos comparten similitudes en el nombre y tipo de símbolo, difieren en los detalles de la noción y el impacto.

Tabla 6. Comparación símbolo Planificar Proyectos de Investigación

LEL 2 (Institución)	LEL 1 (Máster)	Nombre	Tipo	Noción	Impacto
Planificar Proyectos de Investigación	Planificación de los Procesos de Investigación	Si	Si	No	No

En el caso del símbolo “Proyecto Aprobado”, no se encuentra similitud con ningún símbolo del LEL máster, por lo tanto, se consideraría una omisión, como se observa en la Tabla 7.

Tabla 7. Comparación del símbolo PROYECTO APROBADO

LEL 2 (Institución)	LEL 1 (Más-ter)	Nombre	Tipo	Noción	Impacto
Proyecto Aprobado	-----	No	No	No	No

Al realizar la ejecución del método se observa que existen símbolos del glosario máster que no se encuentran en el glosario secundario, lo que se considerarían como “Faltantes”, como fue el caso con el símbolo “Revisión por Pares” y “Normativa y/o procedimientos para el desarrollo de la investigación” que se encuentra en el LEL1, pero no en el LEL 2. En la Tabla 8, se puede visualizar el resultado global del método.

Tabla 8. Resumen de las comparaciones entre algunos símbolos de los glosarios

LEL1 \ LEL2	Profesor Investigador				Proyecto de Investigación				Planificación de los Procesos de Investigación			
	Nombre	Tipo	Noción	Impacto	Nombre	Tipo	Noción	Impacto	Nombre	Tipo	Noción	Impacto
Investigador Proponente	No	Si	Si	Si	No	No	No	Si	No	No	No	Si
Proyecto de Investigación Planificar	No	No	No	No	Si	Si	Si	Si	Si	No	No	No
Proyecto de Investigación Instituto de Investigación	No	No	No	Si	Si	No	No	No	Si	Si	Si	No
Comité de Investigación	No	Si	Si	No	No	No	No	No	No	No	Si	No
Proyecto Aprobado	No	Si	No	No	No	No	No	No	No	No	No	No
Proyecto Aprobado	No	No	No	No	No	No	No	No	No	No	No	No

Como resultado de la ejecución del método realizado con un extracto de 11 símbolos del glosario máster y 8 símbolos del glosario secundario, se obtuvo la siguiente información: *Similitudes*: Proyecto de Investigación – Proyecto de Investigación. *Omisiones*: Instituto de Investigación; Comité de Investigación; Publicación de artículos científicos. *Diferencias*: Investigador proponente – Profesor Investigador; Planificar Proyecto de Investigación – Planificación de los Procesos de Investigación; Proyecto Aprobado; Líneas de Investigación. *Faltantes*: Revisión por pares, Normativa y/o procedimientos para el desarrollo de la investigación; Dominios académicos; Procedimientos para el desarrollo de investigación; Devolución de resultados de investigación;

Normativa para el desarrollo de investigación; Producción académica y científica; Revisión por pares; Institución.

6 Herramienta de Soporte

Se prototipó una herramienta de software para automatizar la comparación sintáctica y léxica entre los glosarios LEL. El prototipo es un conjunto de scripts desarrollados en Python [19] utilizando bibliotecas y programas para el procesamiento de lenguaje natural de NLTK [20][21]. El prototipo recibe como entrada ficheros en formato CSV. El CSV se procesa para crear los diccionarios que representan a cada glosario LEL. Así, se identifica el tipo, el nombre, las nociones y los impactos de cada símbolo. Luego, se realiza una comparación entre diccionarios, donde cada símbolo del LEL secundario se compara con los símbolos del LEL Máster. Para esto, se vectoriza el contenido en base al estándar empleando TF-IDF[22] y se analiza las similitudes léxicas con aplicación de la métrica Similitud Coseno y Jaccard [23][24] para medir la similitud entre ellos. Estas similitudes se obtienen y muestran a través de un gráfico mapa de calor escalado de 0 a 1, como se observa en la Fig. 2 que representa la similitud entre dos símbolos, en este caso Proyecto de Investigación LEL2 con Proyecto de Investigación LEL1, se observa una alta similitud.

En la Fig. 3, se puede observar el resumen de la comparación de los símbolos de los glosarios. Resaltan por ejemplo, el símbolo “Proyecto de investigación” del LEL2 posee una alta similitud con el símbolo “Proyecto de investigación” con el LEL1, lo que indicaría *Similitud*; por otro lado el símbolo “investigador proponente” del LEL2 tiene una similitud media con el símbolo “profesor investigador” del LEL2, lo que indicaría una *Diferencia*; el símbolo “dominios académicos” del LEL1 posee similitud muy baja con datos de los símbolos del LEL2, lo que se consideraría como un *Faltante*, resultados que coinciden con la comparación manual indicada en Tabla 8.

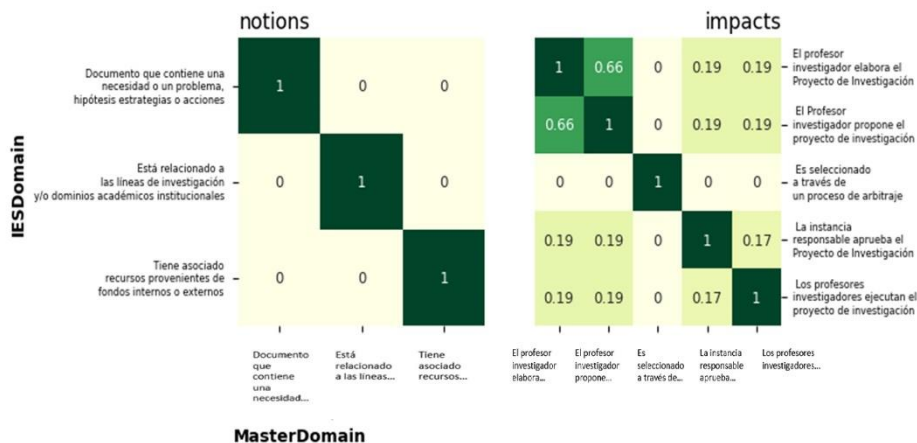


Fig. 2. Mapa de calor del símbolo Proyecto de Investigación



Fig. 3. Resumen de similitud de los símbolos de los glosarios LEL

7 Conclusiones y Trabajos Futuros

Este artículo propuso un método para alinear dos glosarios LEL de dominios diferentes pero relacionados.

El método utiliza un análisis sintáctico que se enfoca en la estructura y el orden de las palabras para verificar la conformidad con reglas de la elaboración del LEL, por ejemplo, cuando se verifica la estructura de las definiciones para asegurarse de que siguen un formato predefinido [25]. En el caso de estudio se compara no solo los nombres de los símbolos, sino también sus tipos, nociones e impactos, lo que implica una evaluación del significado y el uso de cada término en su contexto. En la comparación de los tipos de Símbolos, estos tienen un significado específico en el contexto del glosario, por lo que su comparación es semántica. Al comparar las Nociones e Impactos, estas son descripciones que reflejan el significado y el rol de los símbolos en el dominio, lo que implicará una evaluación semántica. El Análisis Léxico, se centra en la identificación de tokens o unidades léxicas en el texto, como palabras o símbolos, sin considerar su significado, por ejemplo, al identificar y contar palabras claves en el glosario. El método permitió revelar similitudes notables entre ciertos términos de los glosarios LEL utilizados como caso de estudio, como “Proyecto de Investigación”, así como entre “Planificar Proyecto de Investigación” y “Planificar proyectos de investigación”.

Estas similitudes sugieren una superposición en la conceptualización de ciertos términos entre los glosarios LEL1 y LEL2. Por otro lado, se identificaron también

Omisiones y faltantes entre los glosarios, como fue con los símbolos “Proyectos Aprobados” y “Revisión por pares”. Estas disparidades pueden afectar la integridad y la completitud de los glosarios. Otro aspecto importante para destacar es la identificación de las diferencias en los símbolos de los glosarios como “Profesor proponente” y “Profesor investigador”. Estas diferencias resaltan la necesidad de una alineación precisa de los términos para evitar interpretaciones erróneas.

La herramienta de soporte basada en scripts de Python y técnicas de procesamiento de lenguaje natural permite la automatización sintáctica-léxica entre los glosarios LEL y demuestra ser un apoyo para facilitar el proceso de comparación y alineación de los símbolos de los glosarios.

Como trabajo futuro es necesario continuar con la ejecución del método aplicándolo a un dominio más extenso, agregando similitud semántica, integrando la búsqueda de faltantes y una base semántica asociada al dominio, así como también la creación de una interfaz gráfica para mejorar la usabilidad del prototipo, y ofrecer al experto en el dominio una mejor toma de decisiones.

Referencias

1. Glinz, Martin. 2011. «A glossary of requirements engineering terminology». Standard Glossary of the Certified Professional for Requirements Engineering (CPRE) Studies and Exam, Version 1(May):56.
2. Arora, Chetan, Mehrdad Sabetzadeh, Lionel Briand, y Frank Zimmer. 2017. «Automated Extraction and Clustering of Requirements Glossary Terms». *IEEE Transactions on Software Engineering* 43(10):918-45.
3. Leite, J. C. S. do. P., y A. P. M. Franco. 1993. «A strategy for conceptual model acquisition». Pp. 243-46 en [1993] *Proceedings of the IEEE International Symposium on Requirements Engineering*. IEEE Comput. Soc. Press.
4. González, H. 2009. «Interoperabilidad entre los sistemas informáticos». VI Encuentro Internacional de Contabilidad, Auditoría y Finanzas (June 2009):llevado a cabo en la ciudad de La Habana, Cuba.
5. Barcelos, Leonardo, y Rosângela Penteado. 2016. «Especificação de requisitos no domínio de sistemas de informação com o uso de padrões». *Anais do WER 2016 - Workshop em Engenharia de Requisitos*.
6. Outa, Camila Tiemi, y Victor F. A. Santander. 2019. «O uso de modelos de processos de negócio e de modelagem organizacional em metodologias ágeis: uma revisão sistemática da literatura». en *Anais do Workshop em Engenharia de Requisitos*. Recife, Brasil: Even3.
7. Martins, Luiz Eduardo G., y Tony Gorschek. 2016. «Requirements engineering for safety-critical systems: A systematic literature review». *Information and Software Technology* 75:71-89.
8. Sebastián, Alberto, Graciela D. S. Hadad, y Ezequiel Robledo. 2017. «Inspección centrada en Omisiones y Ambigüedades de un Modelo Léxico». *CIBSE 2017 - XX Ibero-American Conference on Software Engineering* 71-84
9. Antonelli, Leandro, Gustavo Rossi, y Alejandro Oliveros. 2016. «A Collaborative Approach to Describe the Domain Language through the Language Extended Lexicon.» *The Journal*

- of Object Technology 15(3):3:1.
10. Gila, Gustavo D., Daniel Arias Figueroa, y Alejandro Oliveros. 2000. «Producción del LEL en un Dominio Técnico. Informe de un caso». WER00 III Workshop em Engenharia de Requisitos 53-69.
 11. Litvak, Claudia, Gustavo Rossi, y Leandro Antonelli. 2018. «Conflict Management in the Collaborative Description of a Domain Language (S)». Pp. 524-77 en.
 12. Hadad, Graciela Dora Susana, Claudia Silvia Litvak, y Jorge Horacio Doorn. 2014. «Heurísticas para el modelado de requisitos escritos en lenguaje natural». CACIC 2014 - XX Congreso Argentino de Ciencias de la Computación.
 13. Sebastián, Alberto, Graciela D. S. Hadad, y Daniela Raffo. 2019. «Evaluación de Variantes de Inspección en un Modelo Léxico». Anais do WER 2019 - Workshop em Engenharia de Requisitos 1-6
 14. Liu, Yalin, Jinfeng Lin, Jane Cleland-Huang, Michael Vierhauser, Jin Guo, y Sugandha Lohar. 2020. «SENET: A Semantic Web for Supporting Automation of Software Engineering Tasks». Pp. 23-32 en 2020 IEEE Seventh International Workshop on Artificial Intelligence for Requirements Engineering (AIRE). IEEE.
 15. Alhassan, Baththama Bello. 2015. «Extending an Ontology Alignment System with a Lexical Database». Scientific Research Journal (SCIRJ) III(I):12-17.
 16. Ngo, Duyhoa, y Zohra Bellahsene. 2015. «Efficient Semantic Verification of Ontology Alignment». Pp. 141-48 en 2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT). Vol. 1. IEEE.
 17. Hu, Xiaocao, Zhiyong Feng, Shizhan Chen, Keman Huang, Jianqiang Li, y Mengchu Zhou. 2017. «Accurate Identification of Ontology Alignments at Different Granularity Levels». IEEE Access 5:105-20.
 18. CACES. 2019. «Modelo de Evaluación Externa de Universidades y Escuelas Politécnicas 2019».
 19. Blank, Julian, y Kalyanmoy Deb. 2020. «Pymoo: Multi-Objective Optimization in Python». IEEE Access 8:89497-509.
 20. Millstein, F. 2020. Natural language processing with python: natural language processing using NLTK. Frank Millstein. editado por Frank Millstein.
 21. Project, NLTK. 2023. «NLTK Documentation». Recuperado 27 de febrero de 2024 (<https://www.nltk.org/>).
 22. Aizawa, Akiko. 2003. «An information-theoretic perspective of tf-idf measures». Information Processing & Management 39(1):45-65.
 23. García Monsalve, Luz. 2012. «Experimento de recuperación de información usando las medidas de similitud coseno, jaccard y dice». Tecciencia 6(12):14-24.
 24. Grijalva-Arriaga, Paola, Galo Cornejo-Gómez, Raquel Gómez-Chabla, Leandro Antonelli, y Pablo Thomas. 2022. «Alignment Techniques in Domain-Specific Models». Pp.45-61 en Technologies and Innovation.
 25. Carlos, Juan, Obando Roldán, ; José, Arturo Pulido Díaz, y Alberto Gómez Ávila. 2020. «Revista CIENCIA Y TECNOLOGÍA V.16 N.1 Procesamiento del Lenguaje Natural para Reconocer Mensajes de Textos Extorsivos a través del Análisis Sintáctico y Lematización Natural Language Processing to Recognize Extortive Text Messages through Syntactic Analys». Cienc. Tecnol 16(1):33-42.