**PROPUESTA DE TEMA DE TESIS DE DOCTORADO**

**DOCTORADO EN CIENCIAS DE LA INGENIERÍA**

**TÍTULO DEL TEMA DE TESIS PROPUESTO**

**Interfaz de Lenguaje Natural para la Consulta de Información en un Lago de Datos Hospitalario**

**PROPONENTE**

**Jonathan Zavala Díaz**

**DIRECTOR DE LA TESIS**

**Dr. Juan Carlos Olivares Rojas**

**Palabras clave:** Lenguaje natural, NPL, Interfaz, Español, Hospitalario.

**Índice**

**1. Resumen 1**

**2. Introducción 2**

**3. Antecedentes 2**

**4. Marco Teórico 2**

**5. Objetivos 2**

***5.1 Objetivo General* 2**

***5.2 Objetivos Particulares* 2**

**6. Metas 2**

**7. Impacto 2**

**8. Metodología 2**

**9. Programa de actividades, calendarización 2**

**10. Productos Entregables 2**

**11. Vinculación con otras instituciones, empresas o sectores 2**

**12. Referencias 2**

**1. Resumen**

El área de la salud es una de las áreas más beneficiadas con el uso de las TICs en los procesos de cuidado de la salud y diagnóstico de los pacientes. Al igual que muchas otras áreas, los datos de salud crece de forma rápida y en grandes volúmenes en poco tiempo. Esto presenta enormes ventajas, pero también muchos retos como el detalle de poder realizar consultas y predicciones en poco tiempo. Los lenguajes de consultas generalmente no son amigables ni dinámicos para los usuarios finales. En los últimos años el avance en el procesamiento del lenguaje natural y de la inteligencia artificial a permitiendo la masificación de aplicaciones como los chatbots y asistentes digitales que permiten una comunicación de forma natural en diversos idiomas como el inglés. Desafortunadamente, la información médica es de un contexto muy específico, así como los desarrollos existentes están muy enfocado en otros idiomas como el inglés, dejando a los especialistas de la salud hispanoparlantes muy rezagado.

**2. Introducción**

Con los avances y crecimiento en las Tecnologías de la Información y la Comunicación (TICs), grandes volúmenes de datos se siguen acumulando velozmente en los bancos de datos de cada organización. Mas sin embargo la mayor parte de esta información nunca se puede utilizar para generar un beneficio real para la organización, para ello se debe ser capaz de convertir la información en conocimiento. El conocimiento se puede utilizar para aliviar el proceso de toma de decisiones por parte de los expertos en el área dominio [1].

En los últimos 20 años, la recopilación y el almacenamiento de datos hospitalarios se ha incrementado enormemente con el uso generalizado de los sistemas de información clínica, los cuales contienen grandes cantidades de datos sobre la salud y atención medica de los pacientes [2]. Una amplia gama de estos datos se encuentra comúnmente dentro de narrativas clínicas las cuales son del tipo de datos no estructurados. Los informes narrativos permiten la flexibilidad de expresión como dudas, negaciones o hipótesis diagnósticas y la representación compleja de enfermedades, examen clínico, historial del paciente y antecedentes médicos familiares [3]. La mayoría de los registros médicos actuales conservan un gran elemento de texto libre. Si bien esto es atractivo para la mayoría de los usuarios finales debido a la flexibilidad de expresión, crea desafíos para el uso continuo de la información contenida en las notas [4]. El procesamiento del lenguaje natural (NLP, por sus siglas en inglés) una forma de aprendizaje automático se puede utilizar en este contexto para procesar y analizar elementos de texto libre, por lo cual puede ayudar en la predicción de los resultados de los pacientes, aumentar los sistemas de clasificación de hospitales y generar modelos de diagnóstico que detectan enfermedades crónicas en etapa temprana [4].

El NPL ahora se usa cada vez más en la medicina para mejorar la utilización de registros de salud electrónicos no estructurados y para proporcionar una forma de comunicación con los pacientes para responder preguntas y realizar consultas [5]. Otro problema surge cuando se requieren hacer consultas a los datos debido a que las personas que no saben utilizar un lenguaje de consulta no podrán hacer consultas de información, por lo cual se requiere desarrollar sistemas de interfaz de lenguaje natural para bases de datos para realizar consultas en lenguaje natural con las bases de datos.

Usar el lenguaje natural para recuperar información de la base de datos es una forma más fácil. Las computadoras no pueden entender el lenguaje natural por lo que necesitan una interfaz; esa es la razón para desarrollar una interfaz de lenguaje natural para la base de datos. La interfaz de lenguaje natural es capaz de traducir la consulta de lenguaje natural dada por el usuario a una equivalente en lenguaje de consulta de base de datos. Por lo tanto, se desarrollaron interfaces de lenguaje natural para bases de datos (NLIDB) para convertir lenguaje natural a consulta SQL y obtener el resultado correspondiente del base de datos. Todavía hay mucho trabajo de investigación en el campo de la interfaz de lenguaje natural y se están desarrollando nuevas interfaces para las bases de datos que brindan respuestas más precisas. Con el avance en el poder de procesamiento del hardware, muchos NLIDBS brindan buenos resultados. Aunque el desarrollo del sistema de base de datos de lenguaje natural ha avanzado en los últimos años, todavía no se usa comúnmente. Se necesitan algoritmos para desarrollar e incorporar para la optimización de consultas. [6]

En la literatura de los últimos años han propuesto herramientas basadas en lenguaje natural para consulta de información, más sin embargo dichas herramientas están desarrolladas para idiomas principalmente en inglés y algunos otros idiomas específicos como francés, afgano y cingalés [2], [7], [8]. Por lo cual nace la necesidad de crear una herramienta especial para el idioma español.

**3. Antecedentes**

En el trabajo publicado por Rencis [9] se presenta un lenguaje natural controlado el cual cumple la función de un lenguaje tan fácil de aprender y también es muy eficiente al dar la respuesta a una consulta típica en menos de un cuarto de segundo. Después de desarrollar el lenguaje de consulta basado en el lenguaje natural controlado, se probó con varios tipos de expertos del ámbito hospitalario. Sin embargo, el estudio de caso descrito en su trabajo mostró que el idioma aún no es totalmente utilizable por expertos en el dominio que deseen poder escribir sus consultas en él. Concluyen que este lenguaje sirva como un trampolín en el camino hacia una interfaz basada en lenguaje completamente natural para consultar datos hospitalarios. Solo se presenta un lenguaje natural controlado, más sin embargo solo es el primer paso para la creación de una interfaz basada en lenguaje natural para consulta de información.

En los trabajos publicados por Rencis [1][10] propone un sistema que permite formular consultas en lenguaje natural, sus experimentos muestran que este tipo de consulta podría proporcionar una mejora en el proceso de toma de decisiones de los profesionales de la salud. Se desarrolla un sistema que permite al usuario escribir consultas en palabras clave contenidas en lenguaje natural. Estas consultas luego se traducen automáticamente en consultas válidas en nuestro lenguaje de consulta base existente. Cabe mencionar que la herramienta desarrollada está diseñada para la necesidad de los hospitales de Letonia. Así, el sistema estaría basado en el idioma letón. El autor menciona que un trabajo a futuro que aún no se inspecciona es involucrar el aprendizaje automático en el proceso de traducción de consultas.

El trabajo publicado por Pech-May et al. [11] menciona que una de las técnicas de aprendizaje profundo más utilizadas para el procesamiento de lenguaje natural son las redes neuronales recurrentes. Además, describe dichas redes, así como sus variantes. Asimismo, se analizan distintos trabajos que hacen uso de ellas para el procesamiento de texto. Los autores concluyen que los trabajos analizados han demostrado las ventajas sobre los métodos tradicionales de aprendizaje automático en algunos campos. También concluyen que el aprendizaje profundo se ha aplicado al procesamiento del lenguaje natural con cierto éxito, y los resultados parecen prometedores; sin embargo, los resultados son preliminares y aún están lejos de ser satisfactorios.

En el trabajo publicado por Álvarez et al. [12] se realiza un análisis de datos en retrospectiva obtenidos del SIMF-IMSS, el cual busca desarrollar modelos predictivos de complicaciones crónicas de la diabetes mellitus tipo 2 con base en la información capturada por el personal médico. Destacamos este trabajo debido a que una de las fases de este proyecto se trabaja con técnicas de procesamiento de lenguaje natural como base de su algoritmo de desidentificación para eliminar datos personales de las historias clínicas y así proteger su identidad. Para dicha fase de desidentificación de notas médicas se utiliza la librería flair, la cual es una red neuronal profunda, entrenada con texto en español. Plantean utilizar modelos predictivos para predecir riesgos en las personas con diabetes a través de los datos clínicos de cada paciente, más sin embargo solo está planteado como trabajo a futuro y una oportunidad de investigación.

En el trabajo publicado por Akbik et al. [13] presentan FLAIR, un framework de NLP diseñado para facilitar la capacitación y distribución de modelos de lenguaje, clasificación de texto y etiquetado de secuencias de última generación. La idea central del framework es presentar una interfaz simple y unificada para tipos conceptualmente muy diferentes de word embeddings y documentos. Esto oculta efectivamente toda la complejidad de ingeniería específica de la integración y permite a los investigadores "mezclar y combinar" varias funciones de integración con poco esfuerzo. El marco también implementa entrenamiento de modelos estándar y rutinas de selección de hiperparámetros, así como un módulo de obtención de datos que puede descargar conjuntos de datos NLP disponibles públicamente y convertirlos en estructuras de datos para una configuración rápida de experimentos. La ventaja de poder usar este framework es que cuenta con la capacidad de entrenar redes neuronales con texto en español.

En el trabajo publicado por Setlur et al. [14] presentan Eviza, una interfaz de lenguaje natural para un cuadro de diálogo de consulta interactivo con una visualización existente en lugar de comenzar desde una hoja en blanco y hacer preguntas cerradas que devuelven una sola respuesta de texto o una visualización estática. El sistema emplea un enfoque basado en la gramática probabilística con reglas predefinidas que se actualizan dinámicamente en función de los datos de la visualización, a diferencia de los enfoques basados en el conocimiento o el aprendizaje profundo computacionalmente intensivo.

En el trabajo publicado por Pressat-Laffouilhère et al. [2] presentan Doc'EDS, una herramienta de preselección basada en análisis textual y semántico. El sistema proporciona una interfaz fácil de usar para buscar documentos en francés. Dicha herramienta de búsqueda semántica que se ocupa de las sutilezas del idioma para mejorar un motor de búsqueda avanzado de texto completo dedicado a los documentos de salud en francés. Cabe mencionar que esta herramienta de búsqueda fue construida sobre el almacén de datos clínicos desarrollado en el Hospital Universitario de Rouen. También se realizó una evaluación formal para medir los algoritmos de procesamiento del lenguaje natural implementados.

En el trabajo publicado por Peduru-Hewa et al. [7] presentan SinSQLFinder, una interfaz de usuario en lenguaje natural cingalés para generar consultas SQL. El objetivo principal de esta investigación ha sido establecer un fuerte vínculo entre el idioma cingalés y el lenguaje de consulta estructurado. Debido a la complejidad del idioma, se desarrolló un método único para traducir preguntas en cingalés a lenguajes de consulta estructurados.

En el trabajo publicado por S. Karimi, A. A. Rasel & M. S. Abdullah [8] presentan la interfaz de control y consulta del idioma afgano (ALQCI) la cual se basa en el enfoque NLIDB. El idioma afgano se utiliza en este sistema para trabajar con datos elegidos por el usuario de una base de datos. La recuperación de información de la base de datos requiere conocimiento del lenguaje de consulta SQL o una interfaz bien diseñada. Es conveniente que los usuarios expertos en el dominio recuperen datos de la base de datos. Sin embargo, puede ser un desafío para los usuarios no expertos acceder a la base de datos a través de SQL Query en ausencia de una interfaz de usuario flexible. Este sistema ayuda a superar el desafío para quienes hablan persa en todo el mundo. El sistema puede complementar o sustituir interfaces de usuario (UI) menos amigables. El sistema genera un lenguaje de consulta estructurado (SQL) cuando un usuario ingresa una consulta de lenguaje natural hablado en persa. El sistema funciona al permitir que un usuario ingrese una consulta en persa, que luego el sistema traduce a una consulta SQL utilizando la técnica de mapeo de palabras. El SQL se genera a través del análisis semántico. Posteriormente, el SQL generado se utiliza para extraer conocimiento utilizando el idioma afgano de una base de datos.

En el trabajo publicado por Yu & Silva [15] presentan, FlowSense, una interfaz de lenguaje natural para la exploración visual de datos, la cual emplea un analizador semántico con etiquetado de expresiones especiales y marcadores de posición de expresiones especiales para generalizar a diferentes conjuntos de datos y diagramas de flujo de datos. Presenta explícitamente conjuntos de datos reconocidos y expresiones especiales de diagramas para que el usuario conozca el contexto del flujo de datos. Con FlowSense, el usuario puede expandir y ajustar los diagramas de flujo de datos de manera más conveniente en un lenguaje sencillo. Aplicamos FlowSense al sistema de visualización de flujo de subconjunto VisFlow para mejorar su usabilidad. Evaluamos FlowSense mediante un estudio de caso con expertos en el dominio sobre un problema de análisis de datos del mundo real y un estudio de usuario formal.

En el trabajo publicado por KraljevicYu et al. [16] presentan un el kit de herramientas de anotación de conceptos médicos (MedCAT) de código abierto que proporciona: (a) un novedoso algoritmo de aprendizaje automático autosupervisado para extraer conceptos utilizando cualquier vocabulario de conceptos, incluido UMLS/SNOMED-CT; (b) una interfaz de anotaciones rica en funciones para personalizar y entrenar modelos IE; y (c) integraciones al ecosistema CogStack más amplio para la implementación de sistemas de salud independientes del proveedor.

En el trabajo publicado por Das & Balabantaray [17] presentan un enfoque para construir una Interfaz de lenguaje natural a la base de datos (NLIDB). Se están realizando investigaciones desde hace muchos años para desarrollar varios métodos para abordar este desafiante problema. El sistema propuesto MyNLIDB hace hincapié en el mapeo de palabras clave. Es independiente del dominio y de la base de datos. Funciona muy bien para consultas simples. Actualmente no estamos utilizando ningún método de aprendizaje automático o aprendizaje profundo y no nos enfocamos en la inferencia de ruta de unión; que dejamos para trabajos futuros.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Trabajo |  | interfaz | idioma | ambito | herramienta |
| Rencis [9] | lenguaje natural controlado | no | leton | hospitalario |  |
| Álvarez et al. [12] | técnicas de procesamiento de lenguaje natural como base de su algoritmo de desidentificación | no | español | historias clínicas | librería flair, |
| Setlur et al. [14] | una interfaz de lenguaje natural para un cuadro de diálogo de consulta interactivo | si |  |  | gramática probabilística |
| Pressat-Laffouilhère et al. [2] | Doc'EDS, una herramienta de preselección basada en análisis textual y semántico | si | frances | documentos de salud en francés. | algoritmos de procesamiento del lenguaje natural |
| Peduru-Hewa et al. [7] | SinSQLFinder, una interfaz de usuario en lenguaje natural cingalés para generar consultas SQL | si | cingalés |  |  |
| S. Karimi, A. A. Rasel & M. S. Abdullah [8] | la interfaz de control y consulta del idioma afgano | si | afgano |  | técnica de mapeo de palabras |
| Yu & Silva [15] | FlowSense, una interfaz de lenguaje natural para la exploración visual de datos | si |  |  | un analizador semántico con etiquetado de expresiones especiales y marcadores de posición |
| KraljevicYu et al. [16] | kit de herramientas de anotación de conceptos médicos (MedCAT) | si |  |  |  |
| Das & Balabantaray [17] | presentan un enfoque para construir una Interfaz de lenguaje natural a la base de datos | no |  |  |  |

**4. Marco teórico**

El lenguaje natural se refiere a la forma en que las personas se comunican y el procesamiento de lenguaje natural se encarga de analizar y procesar el lenguaje humano a través de herramientas y tecnologías de software. Esto involucra distintas áreas de la computación, tales como inteligencia artificial, lingüística computacional, etc. Mediante el procesamiento del lenguaje natural es posible procesar documentos de texto, mensajes SMS, email, páginas web, etc. y organizar el conocimiento para realizar tareas como análisis de sentimientos, análisis de contexto, generación de resúmenes, traducción automática, sistemas de diálogos, etc. [11]

Un concepto importante en el procesamiento del lenguaje natural son las **word embeddings** ya que se ha demostrado que si son previamente entrenadas son de gran utilidad para tareas posteriores de NLP, tanto por su capacidad para ayudar al aprendizaje y la generalización con información aprendida de datos no etiquetados, como por la relativa facilidad de incluirlos en cualquier proceso de aprendizaje [13].

En la última década, junto con el crecimiento del aprendizaje profundo, las representaciones (embeddings) basadas en redes neuronales han reemplazado casi por completo a los modelos convencionales basados en conteo y han dominado el campo. Dado que las **word embeddings** neuronales generalmente se entrenan con algún tipo de objetivo de modelado de lenguaje, como predecir una palabra faltante en un contexto, también se conocen como modelos predictivos [18]. Las incrustaciones de palabras fueron popularizadas por Word2vec, en [19] utilizan Word2Vec y BERT.

Existen herramientas avanzadas para el procesamiento del lenguaje natural como Apache OpenNLP [20], Stanford CoreNLP [21] y Stanza [22].

OpenNLP:

La biblioteca Apache OpenNLP es un conjunto de herramientas basado en el aprendizaje automático para el procesamiento de texto en lenguaje natural. OpenNLP es compatible con las tareas de NLP más comunes, como la tokenización, la segmentación de oraciones, el etiquetado de partes del discurso, la extracción de entidades nombradas, la fragmentación, el análisis, la detección de idiomas y la resolución de correferencias.

CoreNLP:

Stanford CoreNLP es nuestro conjunto de herramientas de Java que proporciona una amplia variedad de herramientas de PNL. CoreNLP es su ventanilla única para el procesamiento de lenguaje natural en Java. CoreNLP permite a los usuarios derivar anotaciones lingüísticas para el texto, incluidos límites de tokens y oraciones, partes del discurso, entidades nombradas, valores numéricos y de tiempo, análisis de dependencia y circunscripción, correferencia, sentimiento, atribuciones de citas y relaciones. CoreNLP actualmente admite 8 idiomas: árabe, chino, inglés, francés, alemán, húngaro, italiano y español.

Stanza:

Stanza es una nueva biblioteca de NLP de Python que incluye una canalización de NLP neuronal multilingüe y una interfaz para trabajar con Stanford CoreNLP en Python. Stanza es una colección de herramientas precisas y eficientes para el análisis lingüístico de muchos lenguajes humanos. Desde texto sin procesar hasta análisis sintáctico y reconocimiento de entidades, Stanza trae modelos de NLP de última generación a los idiomas de su elección.

**5. Objetivos**

**5.1 Objetivo general**

Desarrollar un interfaz de Lenguaje Natural para la Consulta de Información en un Lago de Datos Hospitalario

**5.2 Objetivos particulares.**

1. Investigar diversas técnicas y herramientas de lenguaje natural en español para el área de la salud.
2. Estudiar de Técnicas de conversión a lenguajes de consulta de datos.
3. Implementar un lago de datos hospitalarios provenientes de diversas fuentes y en diversos formatos.
4. Desarrollar interfaces en lenguaje natural (español) que permita la obtención de datos de sistemas hospitalarios en forma sencilla y funcional.
5. Realizar modelos de aprendizaje automático y de análisis de datos que permitan mostrar información de pronósticos.
6. Desarrollar una interfaz de usuario para visualizar la información proveniente de la interfaz de lenguaje natural

**6. Metas.**

1. 1 tabla con al menos 10 técnicas y herramientas de lenguaje natural en español para el área de la salud, resaltando sus características, evaluarlas, para elegir 3 de ellas con las cuales trabajar.
2. 1 tabla con diferentes Técnicas de conversión a lenguajes de consulta de datos y elegir la que tenga mayores prestaciones y se adapte a nuestros requerimientos.
3. 1 implementación lago de datos hospitalarios provenientes de diversas fuentes y en diversos formatos.
4. 1 interfaz en lenguaje natural en el idioma que permita la obtención de datos de sistemas hospitalarios en forma sencilla y funcional.
5. 1 implementación de modelos de aprendizaje automático y de análisis de datos que permitan mostrar información de pronósticos.
6. 1 interfaz de usuario para visualizar la información proveniente de la interfaz de lenguaje natural

**7. Impacto.**

Impacto o beneficio en la solución a un problema relacionado con el sector productivo o la generación del conocimiento científico o tecnológico. Sustente la realización de su proyecto respecto a la magnitud del problema, la trascendencia de su estudio, su factibilidad, vulnerabilidad e impacto (social, económico, ambiental, etc.) en congruencia con la línea de investigación del programa de doctorado (½ cuartilla).

**8. Metodología.**

Explique el o los procedimientos científico-metodológicos a seguir para cumplir los objetivos y metas del proyecto, indicando las pruebas estadísticas, diseño experimental y técnicas a utilizar (máximo dos cuartillas).

**9. Programa de actividades, calendarización.**

Gráfico

Descripción generada automáticamente

**10. Productos entregables.**

* Software - Interfaces en lenguaje natural (español) que permita la obtención de datos de sistemas hospitalarios en forma sencilla y funcional.
* Método para modelos de aprendizaje automático y de análisis de datos que permitan mostrar información de pronósticos.
* Software - Interfaz de Lenguaje Natural para la Consulta de Información en un Lago de Datos Hospitalario
* Presentación en congreso internacional
* 1 publicación en el índice JCR
* Documento de tesis
* Certificado Toefl con 550 puntos

**11. Vinculación con otras instituciones, empresas o sectores.**

Pendiente…

**12. Referencias.**

[1] E. Rencis, “Application of a Configurable Keywords-Based Query Language to the Healthcare Domain,” *J. Adv. Inf. Technol.*, vol. 12, no. 2, pp. 142–147, 2021, doi: 10.12720/jait.12.2.142-147.

[2] T. Pressat-Laffouilhère *et al.*, “Evaluation of Doc’EDS: a French semantic search tool to query health documents from a clinical data warehouse,” *BMC Med. Inform. Decis. Mak.*, vol. 22, no. 1, p. 34, Dec. 2022, doi: 10.1186/s12911-022-01762-4.

[3] N. Garcelon, A. Neuraz, V. Benoit, R. Salomon, and A. Burgun, “Improving a full-text search engine: the importance of negation detection and family history context to identify cases in a biomedical data warehouse,” *J. Am. Med. Informatics Assoc.*, vol. 24, no. 3, pp. 607–613, May 2017, doi: 10.1093/jamia/ocw144.

[4] S. Locke, A. Bashall, S. Al-Adely, J. Moore, A. Wilson, and G. B. Kitchen, “Natural language processing in medicine: A review,” *Trends Anaesth. Crit. Care*, vol. 38, pp. 4–9, Jun. 2021, doi: 10.1016/j.tacc.2021.02.007.

[5] J. Wang *et al.*, “Systematic Evaluation of Research Progress on Natural Language Processing in Medicine Over the Past 20 Years: Bibliometric Study on PubMed,” *J. Med. Internet Res.*, vol. 22, no. 1, p. e16816, Jan. 2020, doi: 10.2196/16816.

[6] E. U. Reshma and P. C. Remya, “A review of different approaches in natural language interfaces to databases,” in *2017 International Conference on Intelligent Sustainable Systems (ICISS)*, Dec. 2017, pp. 801–804. doi: 10.1109/ISS1.2017.8389287.

[7] D. S. Peduru Hewa and C. Farook, “A Sinhala Natural Language Interface for Querying Databases Using Natural Language Processing,” in *2021 21st International Conference on Advances in ICT for Emerging Regions (ICter)*, Dec. 2021, pp. 213–218. doi: 10.1109/ICter53630.2021.9774794.

[8] S. Karimi, A. A. Rasel, and M. S. Abdullah, “Natural Language Query and Control Interface for Database Using Afghan Language,” in *2022 International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, Aug. 2022, pp. 1–8. doi: 10.1109/INISTA55318.2022.9894168.

[9] E. Rencis, “Towards a natural language-based interface for querying hospital data,” in *Proceedings of 2018 International Conference on Big Data Technologies - ICBDT ’18*, 2018, pp. 25–28. doi: 10.1145/3226116.3226133.

[10] E. Rencis, “Knowledge Extraction from Healthcare Data Using User-Adaptable Keywords-Based Query Language,” in *Proceedings of the 2020 the 4th International Conference on Information System and Data Mining*, May 2020, pp. 128–131. doi: 10.1145/3404663.3406876.

[11] F. Pech-May, L. A. López-Gómez, and J. Magaña-Govea, “Procesamiento de lenguaje natural con aprendizaje profundo,” *Komput. Sapiens*, vol. 2, pp. 56–61, 2019, [Online]. Available: http://komputersapiens.smia.mx/publicaciones.php#KSXI-II

[12] C. Álvarez *et al.*, “Estudio longitudinal para el desarrollo de modelos predictivos de complicaciones crónicas de la diabetes mellitus tipo 2,” *Komput. Sapiens*, vol. 3, pp. 10–15, 2022, [Online]. Available: http://komputersapiens.smia.mx/publicaciones.php#KSXIV-III

[13] A. Akbik, T. Bergmann, D. Blythe, K. Rasul, S. Schweter, and R. Vollgraf, “FLAIR: An easy-to-use framework for state-of-the-art NLP,” in *Proceedings of the 2019 Conference of the North*, 2019, pp. 54–59. doi: 10.18653/v1/N19-4010.

[14] V. Setlur, S. E. Battersby, M. Tory, R. Gossweiler, and A. X. Chang, “Eviza: A natural language interface for visual analysis,” in *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, Oct. 2016, pp. 365–377. doi: 10.1145/2984511.2984588.

[15] B. Yu and C. T. Silva, “FlowSense: A Natural Language Interface for Visual Data Exploration within a Dataflow System,” *IEEE Trans. Vis. Comput. Graph.*, vol. 26, no. 1, pp. 1–11, Jan. 2020, doi: 10.1109/TVCG.2019.2934668.

[16] Z. Kraljevic *et al.*, “Multi-domain clinical natural language processing with MedCAT: The Medical Concept Annotation Toolkit,” *Artif. Intell. Med.*, vol. 117, p. 102083, Jul. 2021, doi: 10.1016/j.artmed.2021.102083.

[17] A. Das and R. C. Balabantaray, “MyNLIDB: A Natural Language Interface to Database,” in *2019 International Conference on Information Technology (ICIT)*, Dec. 2019, pp. 234–238. doi: 10.1109/ICIT48102.2019.00048.

[18] M. T. Pilehvar and J. Camacho-Collados, “Word Embeddings,” 2021, pp. 25–40. doi: 10.1007/978-3-031-02177-0\_3.

[19] J.-C. Hernández-Hernández, D. Juárez-Morales, J.-J. Guzmán-Landa, and G. J. Hoyos-Rivera, “Análisis de Sentimientos en Twitter,” *Komput. Sapiens*, vol. 2, pp. 59–63, 2022, [Online]. Available: http://komputersapiens.smia.mx/publicaciones.php#KSXIV-II

[20] “Apache OpenNLP.” https://opennlp.apache.org/

[21] “CoreNLP.” https://stanfordnlp.github.io/CoreNLP/

[22] “Stanza.” https://stanfordnlp.github.io/stanza/