Trabajo Terminal 1

Modelo generativo de SQL a partir de consultas en español

PRESENTAN

Víctor Ulises Miranda Chávez
Adair Nicolás Hernández
Ives Lancelote Pérez Sánchez
Zury Yael Rubio López

DIRECTORES:

Enrique Alfonso Carmona García
Ituriel Enrique Flores Estrada

Contenido

ANTECEDENTES

- Origen de SQL (Structured Query Language):
  - Creado en 1974 por **Donald Chamberlin y Raymond Boyce** como un lenguaje para manipular y gestionar datos en bases de datos relacionales.
  - Inicialmente conocido como "SEQUEL" (Structured English Query Language)[1].

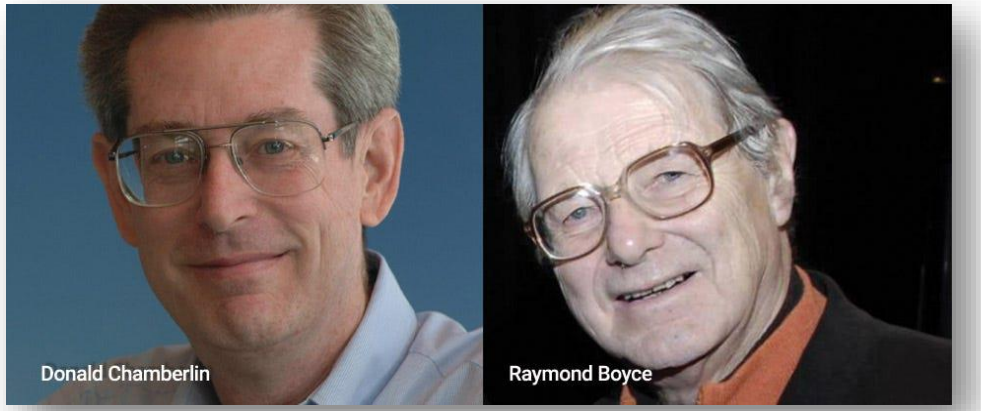


Fig 1. Creadores de SQL. De [2]

[1] D. D. Chamberlin, “Early History of SQL”, IEEE Ann. Hist. Comput., vol. 34, núm. 4, pp. 78–82, oct. 2012, doi: 10.1109/MAHC.2012.61
[2] “SQL Starter Pack. Overview | by Nate Tsegaw | Medium”. Consultado: el 23 de noviembre de 2023. [En línea]. Disponible en: https://ntsegaw.medium.com/sql-starter-pack-286561037697

SITUACIÓN PROBLEMÁTICA

- Complejidad de SQL:
  - Aunque SQL es intuitivo, acceder a bases de datos requiere conocimientos técnicos.
  - Consultas complejas demandan comprensión profunda del lenguaje y de la estructura de la base de datos.
  - Una consulta sencilla requiere que se invierta más tiempo conforme se tienen más tablas y atributos. [3]

[3] SQL: A Beginner's Guide, Third Edition 3rd edition by Oppel, Andy, Sheldon, Robert (2008) Paperback.

JUSTIFICACIÓN

Propósito

Según Woods en [1], la necesidad de acceso a datos almacenados en bases de datos por parte de científicos, expertos en su campo, incrementaría conforme aumentaran la cantidad de redes de computadoras.

El problema que encuentra Woods es que dichos expertos necesitan acceder no solo a una base de datos, sino a varias, las cuales pueden tener distintas estructuras [1].

En este sentido, aquellos que deseen acceder a datos en una base de datos relacional no solo deben saber SQL, sino que además deben invertir tiempo adicional analizando cada base de datos con la que trabajarán para identificar los datos que les sean relevantes [1].

[1] W. A. Woods, “Progress in natural language understanding: an application to lunar geology”, en *Proceedings of the June 4-8, 1973, national computer conference and exposition on - AFIPS ’73*, New York, New York: ACM Press, 1973, p. 441. doi: 10.1145/1499586.1499695.

JUSTIFICACIÓN

- Desarrollo de *text to SQL*:
  - Surge la necesidad de interactuar con bases de datos a partir de lenguaje natural.
  - Anteriormente, proyectos como LUNAR se enfocaron en bases de datos específicas, limitando su adaptabilidad [4].

Progress in natural language understanding—An application to lunar geology

by W. A. WOODS

Bolt Beranek and Newman Inc.
Cambridge, Mass.

INTRODUCTION

The advent of computer networks such as the ARPA net (see e.g., Ornstein et al.[1]) has significantly increased the opportunity for access by a single researcher to a variety of different computer facilities and data bases, thus raising expectations of a day when it will be a common occurrence rather than an exception that a scientist will casually undertake to use a computer facility located 3000 miles away and whose languages, formats, and conventions are unknown to him. In this foreseeable future, learning and remembering the number of different languages and conventions that such a scientist would have to know will require significant effort—much greater than that now required to learn the conventions of his local computing center (where other users and knowledgeable assistance is readily available). The Lunar Sciences Natural Language Information System (which we will hereafter refer to as LUNAR) is a research prototype of a system to deal with this and other man-machine communication problems by adapting the machine to the conventions of ordinary natural English rather than requiring the man to adapt to the machine.

English as a query language

There are two important reasons why one might want to use English as a mode of communication between a man and a machine. First, the man already knows his natural language and if he is to use a particular computer system or data base only occasionally or as a minor part of his work, then he may not have the time or inclination to learn and remember a formal machine language. Second, the human thinks in his native language, and if the mode of communication involves the free and immediate communication of ideas to the machine which the user is conceiving in the course of the interaction, then the additional effort required for him to translate his ideas into another language more suitable to the machine may slow down or otherwise interfere with the interaction. English is therefore an attractive medium because the human can express his ideas in the form in which they occur to him.

State of the art

Although the state of the art in "understanding" natural language by machine is still very limited, significant advances in this area have been made in recent years. Since Simmons' first survey of question answering systems,[4] our understanding of the mysterious "semantic interpretation" component has been made more clear by work such as Woods,[7,8] and the techniques for mechanically parsing natural language sentences have been advanced by the advent of transition network grammars and their parsing algorithms.[9,10,11] Recently, Terry Winograd's blocks world system[6] has demonstrated the potential of some of these techniques—especially those of procedural semantics. The field is now at the point where prototype applications to real problems can make significant contributions to our understanding of the problems of natural language communication with machines. It must be realized, that such applications are still essentially research vehicles, since the problems of mechanical understanding of natural language remain far from solution. However, by using real problems, rather than imaginary toy problems, as the vehicles for such research, one cannot only focus the effort on problems in need of solution, but may also reap the additional benefit of producing a system which will perform a task which someone really wants done.

Natural language understanding

I want to distinguish here between the objectives of this research, which I will call "natural language understanding" research, and the development of so called "English-like" languages and querying systems. There have been a number of computer question answering systems developed for special applications or for general purpose data management tasks which use English words and English syntactic constructions and call their languages "English-like." By this they mean that, although the statements in the language may look more or less like ordinary English sentences, the language makes no attempt to encompass the totality of English constructions that a user might want to use. Also, the interpretations of those sentences by the machine may differ from those which a user would assume without in-

Fig 2. Portada del artículo donde se presenta LUNAR [4]

1973 → ACTUALIDAD

[4] W. A. Woods, "Progress in natural language understanding: an application to lunar geology", en *Proceedings of the June 4-8, 1973, national computer conference and exposition on - AFIPS '73*, New York, New York: ACM Press, 1973, p. 441. doi: 10.1145/1499586.1499695.

JUSTIFICACIÓN

Público:

- Al mes de noviembre del 2023, se aproxima que alrededor del 72.1% de las bases de datos son relacionales [5], lo cual nos habla de la cantidad de personas que pueden beneficiarse de una herramienta que les facilite su flujo de trabajo a la hora de acceder a una base de datos.

Escasez de herramientas para el lenguaje español:

- La comunidad de la inteligencia artificial se ha centrado en realizar avances en *text to SQL* para el inglés.
- Comparados con la comunidad de la inteligencia artificial de habla inglesa, quienes han aprovechado los avances traídos por el aprendizaje profundo, su contraparte en español ha tenido pocos avances.

[5] “DB-Engines Ranking per database model category”. Consultado: el 29 de noviembre de 2023. [En línea]. Disponible en: https://db-engines.com/en/ranking_categories

ESTADO DEL ARTE

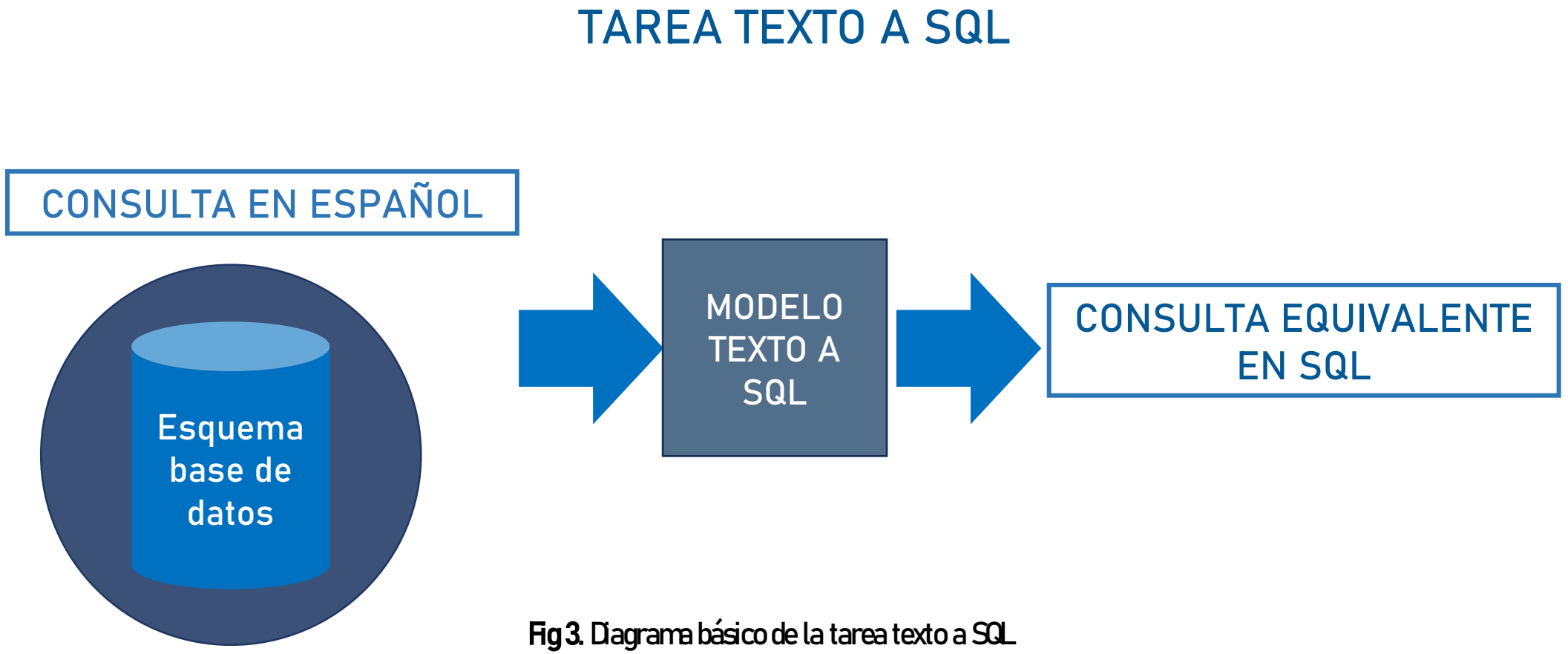


Fig 3. Diagrama básico de la tarea texto a SQL

ESTADO DEL ARTE

NIVELES DE CONSULTAS

Según Özcan en [6], existen 4 niveles de consultas dependiendo de su complejidad:

	NIVEL 1	NIVEL 2	NIVEL 3	NIVEL 4
Consultas con atributos, selección de tabla y condición	✓	✓	✓	✓
Uso de agregadores, ordenamiento y/o agrupamiento	✗	✓	✓	✓
Involucra multiples tablas	✗	✗	✓	✓
Involucra consultas anidadas	✗	✗	✗	✓

Fig 4. Nveles de consulta descritos en las investigaciones.

[6] F. Őzcan, A. Quamar, J. Sen, C. Lei, y V. Efthymiou, "State of the Art and Open Challenges in Natural Language Interfaces to Data", en Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data, Portland OR USA: ACM, jun. 2020, pp. 2629–2636. doi: 10.1145/3318464.3383128.

ESTADO DEL ARTE

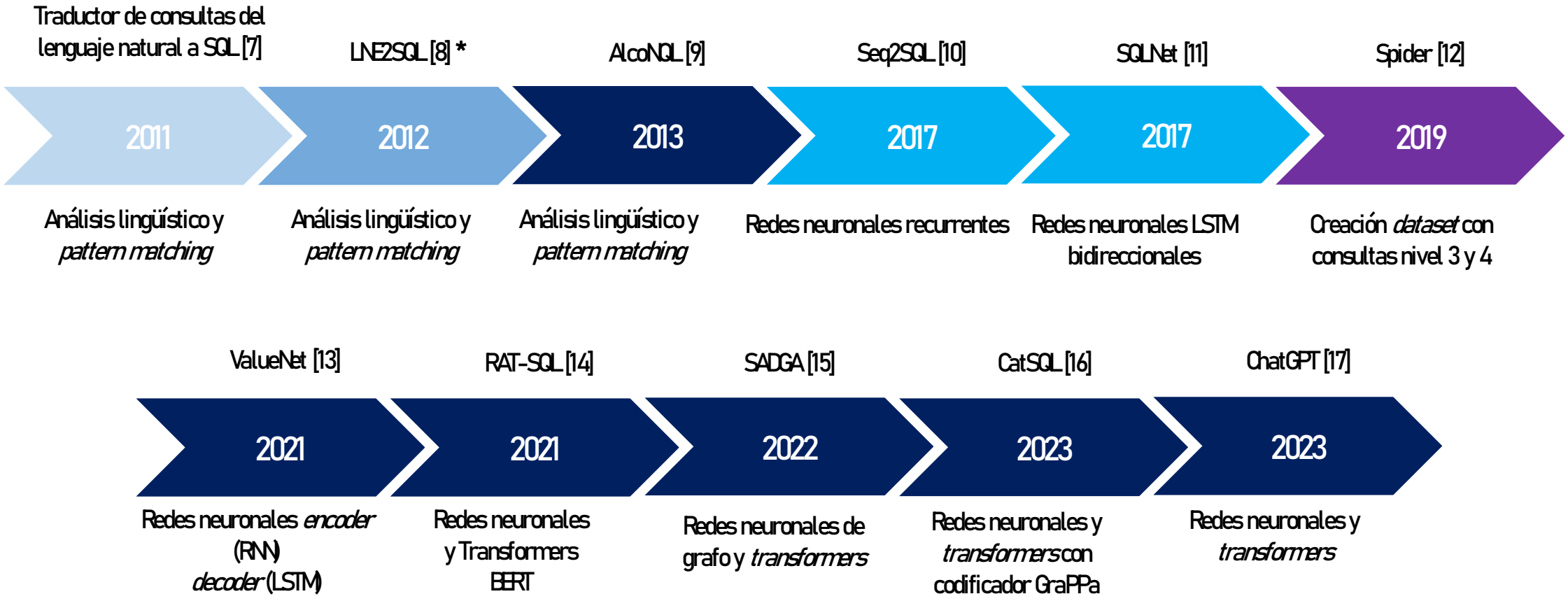


Fig 5. Línea del tiempo en la tarea text to SQL

Nivel 1
Nivel 2
Nivel 3
Nivel 4
Evento relevante

*Alcanza el nivel 3 parcialmente, no cumple con el nivel 2

OBJETIVOS

Objetivo general:

- Diseñar un modelo generativo de SQL a partir de consultas en español. El modelo podrá ser utilizado como parte de una herramienta que facilite a los usuarios el proceso de consulta de información en bases de datos relacionales.

OBJETIVOS

Objetivos específicos:

1.- Realizar una investigación exhaustiva de las estructuras sintácticas y semánticas correctas de SQL y de cómo se relacionan con las expresiones naturales humanas.

2.- Construir un conjunto de ejemplos etiquetados para entrenar al modelo, incluyendo una gran cantidad de consultas en español, esquemas de bases de datos, y sus correspondientes representaciones en SQL.

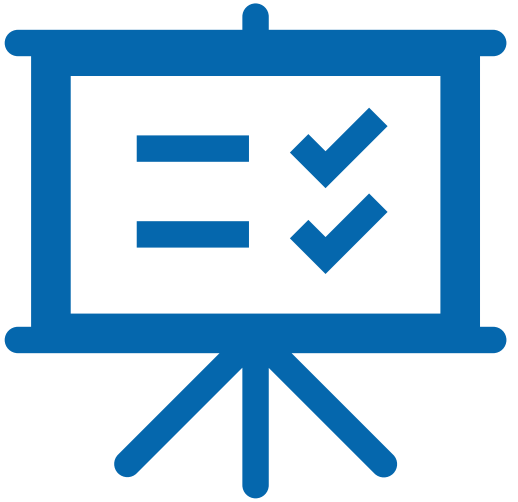
3.- Seleccionar una arquitectura con la capacidad suficiente para capturar patrones complejos de dependencias entre palabras y partes de la frase en el texto en español y sus equivalentes en SQL.

4.- Crear e implementar un modelo de aprendizaje automático para convertir consultas escritas en español a consultas SQL.

OBJETIVOS

Objetivos específicos:

5.- Asegurar la efectividad y corrección del modelo generado, evaluando y validando a través de pruebas exhaustivas.



MARCO TEÓRICO

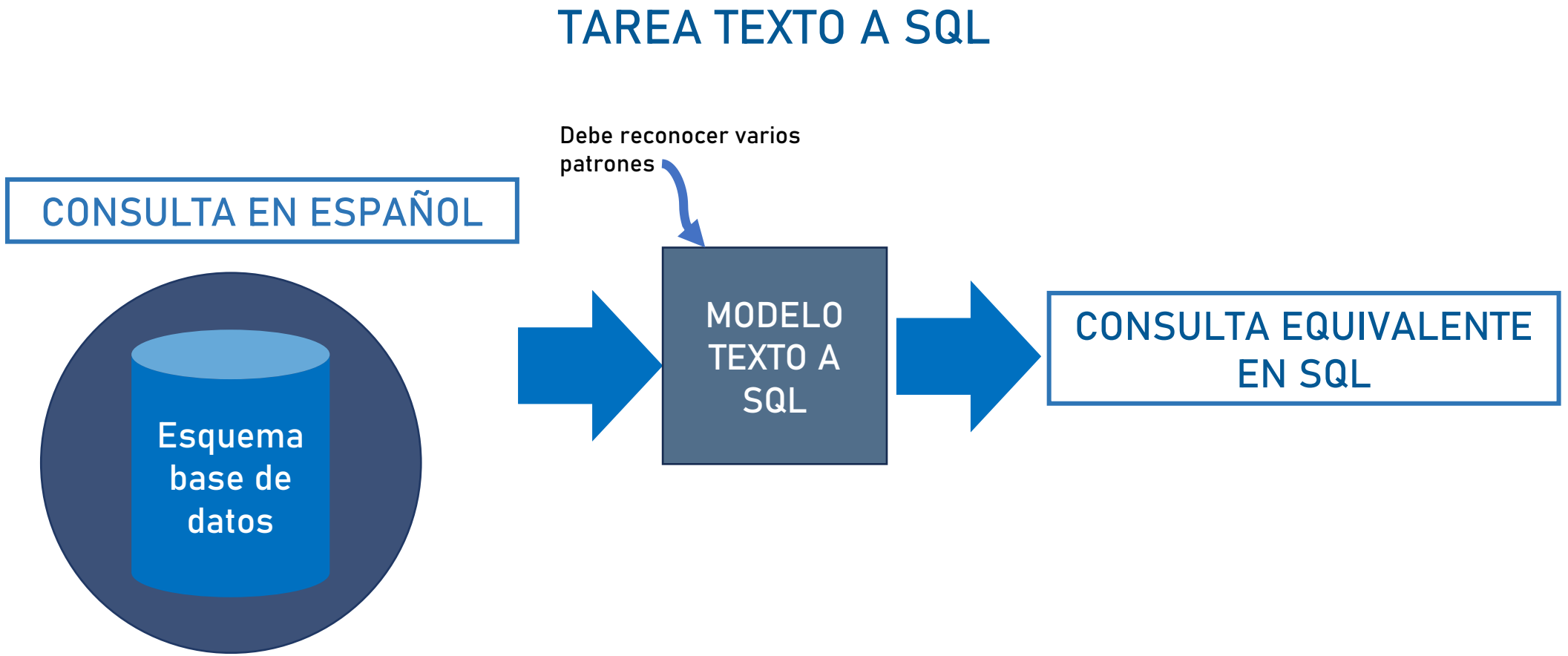


Fig 3. Diagrama básico de la tarea texto a SQL

MARCO TEÓRICO

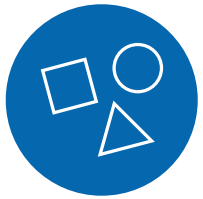
APRENDIZAJE MÁQUINA

Definición:

- Es un campo de la inteligencia artificial (IA) que se centra en el desarrollo de sistemas capaces de aprender y mejorar su rendimiento sin intervención humana directa [18].

[18] B. Mahesh, Machine Learning Algorithms -A Review. 2019. doi: 10.21275/ART20203995.

APRENDIZAJE MÁQUINA



Inducción de reglas:

- Descubre patrones significativos en conjuntos de datos
- Construye de reglas lógicas o condiciones
- Busca representar relaciones y regularidades presentes [19].

La inducción por reglas implica:

- La capacidad de aprender dichas reglas a partir de ejemplos y datos disponibles
- Permite descubrir conexiones y tendencias inherentes [19].

[19] B. Arinze, “Selecting appropriate forecasting models using rule induction”, Omega, vol. 22, núm. 6, pp. 647–658, nov. 1994, doi: 10.1016/0305-0483(94)90054-X.

CONSIDERACIONES DEL MODELO

01

Ambigüedad de las consultas:

Retos con la interpretación de estructuras lingüísticas o ambigüedades en el habla

02

Tipo de consulta:

NIVEL 1

03

Operador (uno por consulta):

Se presenta uno de los siguientes operadores lógicos: <,>,=,<=, >=, !=

CONSIDERACIONES DEL MODELO

04

Generalización

Límite en la cantidad y diversidad de ejemplos usados en el entrenamiento

05

Ausencia de memoria

No se tiene información de consultas pasadas para la generación de las nuevas consultas

CRONOGRAMA

Presentamos el cronograma de actividades que abarca el proyecto. A continuación, explicaremos en que consistió cada una de las actividades para TT-1 y como fueron abordadas

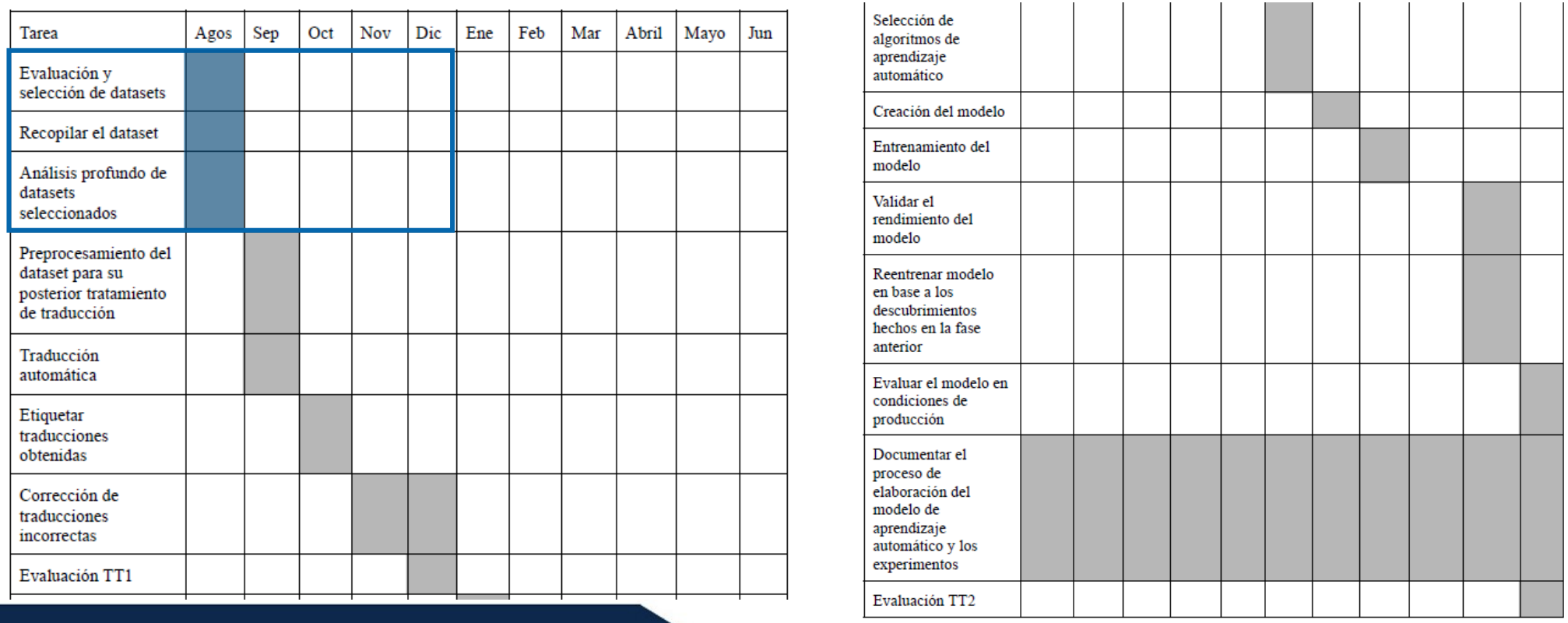


Fig 6. Cronograma

ACTIVIDADES DESARROLLADAS

Evaluación y selección del dataset

Análisis rápido sobre WikiSQL

WikiSQL [10]

SELECT Notes FROM table
WHERE Current_slogan =
"SOUTH AUSTRALIA"

CONSULTAS DE HASTA NIVEL 2

80,654

PREGUNTAS EN LENGUAJE NATURAL

80,654

CONSULTAS EN SQL

24,241

BASES DE DATOS

[10] V. Zhong, C. Xiong, y R. Socher, “Seq2SQL: Generating Structured Queries from Natural Language using Reinforcement Learning”, arXiv.org. Consultado: el 12 de noviembre de 2023. [En línea]. Disponible en: https://arxiv.org/abs/1709.00103v7

ACTIVIDADES DESARROLLADAS

Análisis rápido sobre Spider:

Spider [12]

```
SELECT DISTINCT T1.creation FROM
department AS T1 JOIN
management AS T2 ON
T1.department_id =
T2.department_id JOIN head AS T3
ON T2.head_id = T3.head_id WHERE
T3.born_state = 'Alabama'
```

CONSULTAS DE HASTA NIVEL 4

10,181

PREGUNTAS EN LENGUAJE NATURAL

5,693

CONSULTAS EN SQL

200

BASES DE DATOS

[12] T. Yu et al., “Spider: A Large-Scale Human-Labeled Dataset for Complex and Cross-Domain Semantic Parsing and Text-to-SQL Task”. arXiv, el 2 de febrero de 2019. doi: 10.48550/arXiv.1809.08887.

ACTIVIDADES DESARROLLADAS

Evaluación y selección de dataset

WikiSQL [10]

- Cubre consultas de nivel 1 y 2
- Sus consultas al igual que sus bases de datos, son variadas y pertenecientes a diferentes contextos
- Sus bases de datos están conformadas únicamente por una tabla

Spider [12]

- Cubre consultas de todos los niveles
- Sus consultas al igual que sus bases de datos, son variadas y pertenecientes a diferentes contextos
- Sus bases de datos están conformadas por múltiples tablas

Ambos están en el idioma inglés

[10] V. Zhong, C. Xiong, y R. Socher, "Seq2SQL: Generating Structured Queries from Natural Language using Reinforcement Learning", arXiv.org. Consultado: el 12 de noviembre de 2023. [En línea]. Disponible en: https://arxiv.org/abs/1709.00103v7

[12] T. Yu et al., "Spider: A Large-Scale Human-Labeled Dataset for Complex and Cross-Domain Semantic Parsing and Text-to-SQL Task". arXiv, el 2 de febrero de 2019. doi: 10.48550/arXiv.1809.08887.

ACTIVIDADES DESARROLLADAS

Análisis del conjunto de datos Spider:

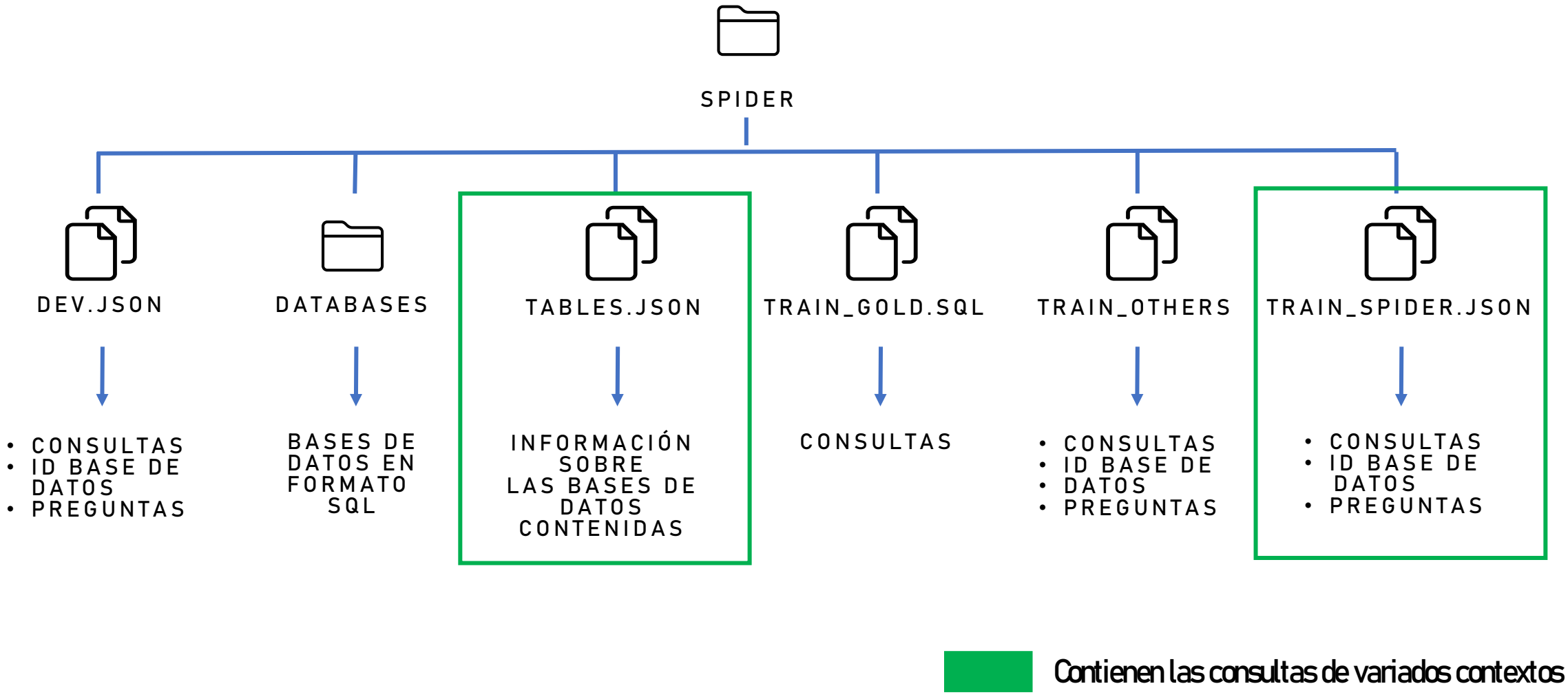
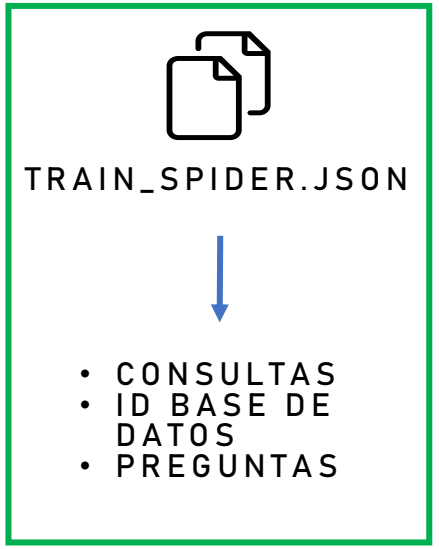


Fig 7. Diagrama de archivos incluidos en SPIDER

ACTIVIDADES DESARROLLADAS

En este conjunto de datos (dataset), Spider, la representación de la información está contenida en formato JSON. Entre los datos que podemos encontrar son:

Clave	Descripción
db_id	Identificador de la tabla
query	Consulta en SQL
query_toks	Tokens por los que está conformada la consulta SQL
query_toks_no_value	Valores de query_toks sin incluir valores específicos, solo indicando dónde se encuentra un valor
question	Pregunta en lenguaje natural equivalente a la consulta en SQL
question_toks	Tokens por los que está conformada la pregunta en lenguaje natural
sql	Información relacionada a la consulta SQL
groupBy	Booleano que indica si se utiliza el modificador GROUP BY en la consulta SQL
having	Booleano que indica si se utiliza el modificador HAVING en la consulta SQL
orderBy	Booleano que indica si se utiliza el modificador ORDER BY en la consulta SQL
limit	Booleano que indica si se utiliza el modificador LIMIT en la consulta SQL
intersect	Booleano que indica si se utiliza el modificador INTERSECT en la consulta SQL
union	Booleano que indica si se utiliza el modificador UNION en la consulta SQL
except	Booleano que indica si se utiliza el modificador EXCEPT en la consulta SQL



CRONOGRAMA

Presentamos el cronograma de actividades que abarca el proyecto. A continuación, explicaremos en que consistió cada una de las actividades para TT-1 y como fueron abordadas

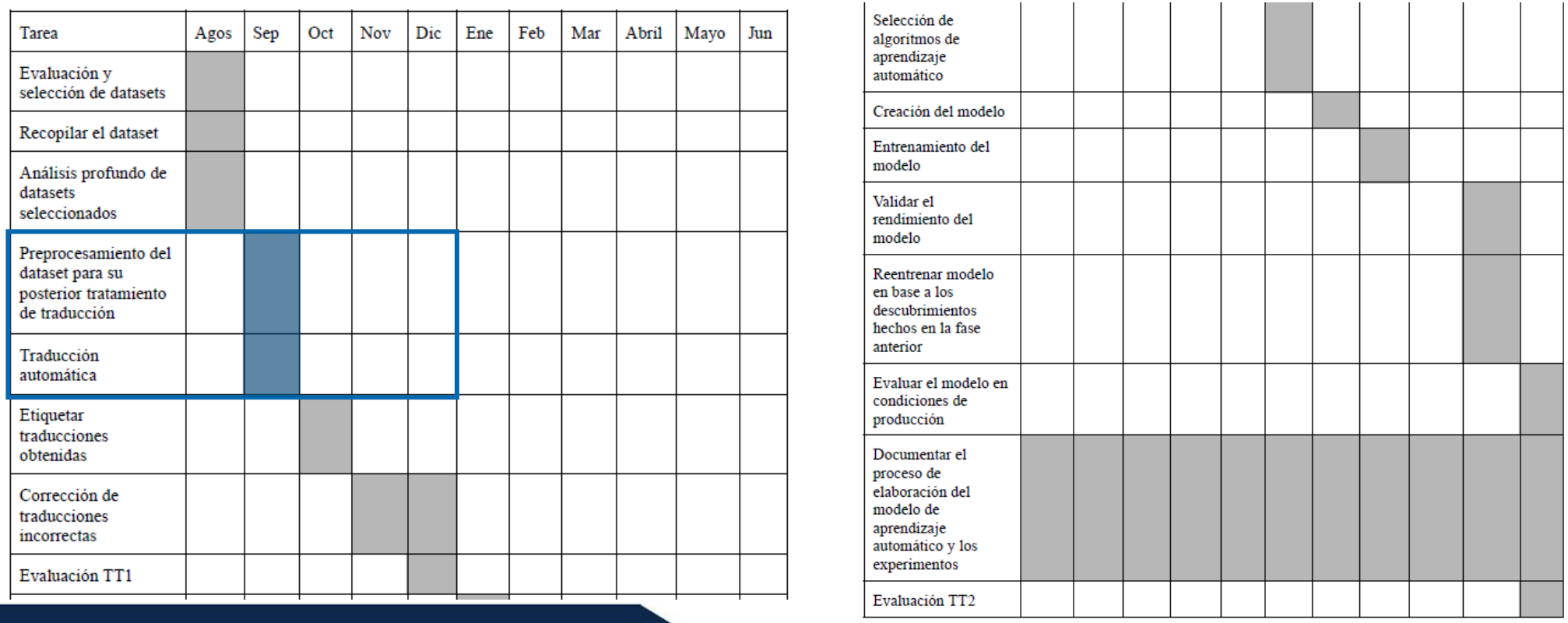


Fig 6. Cronograma

ACTIVIDADES DESARROLLADAS

Preprocesamiento del conjunto y traducción automática supervisada:

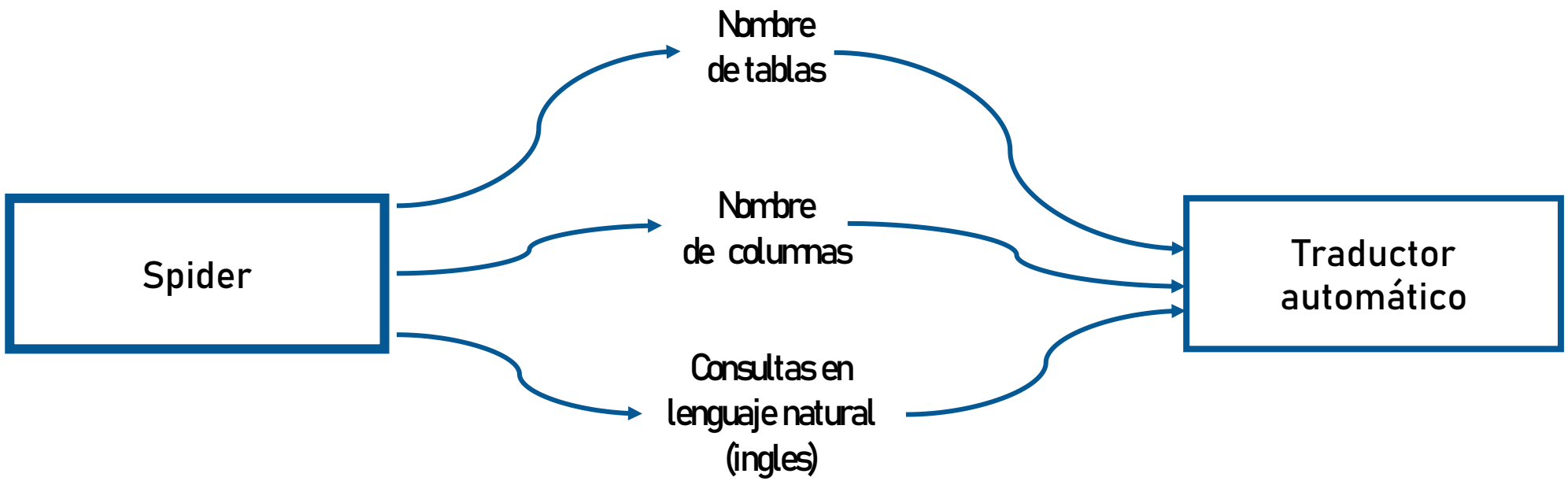


Fig 8. Diagrama descriptivo de la traducción automática: Enfoque de traducción parcial

[10] V. Zhong, C. Xiong, y R. Socher, "Seq2SQL: Generating Structured Queries from Natural Language using Reinforcement Learning", arXiv.org. Consultado: el 12 de noviembre de 2023. [En línea]. Disponible en: https://arxiv.org/abs/1709.00103v7

[12] T. Yu et al., "Spider: A Large-Scale Human-Labeled Dataset for Complex and Cross-Domain Semantic Parsing and Text-to-SQL Task". arXiv, el 2 de febrero de 2019. doi: 10.48550/arXiv.1809.08887.

ACTIVIDADES DESARROLLADAS

Problemas de este enfoque al evaluar el desempeño de un modelo:

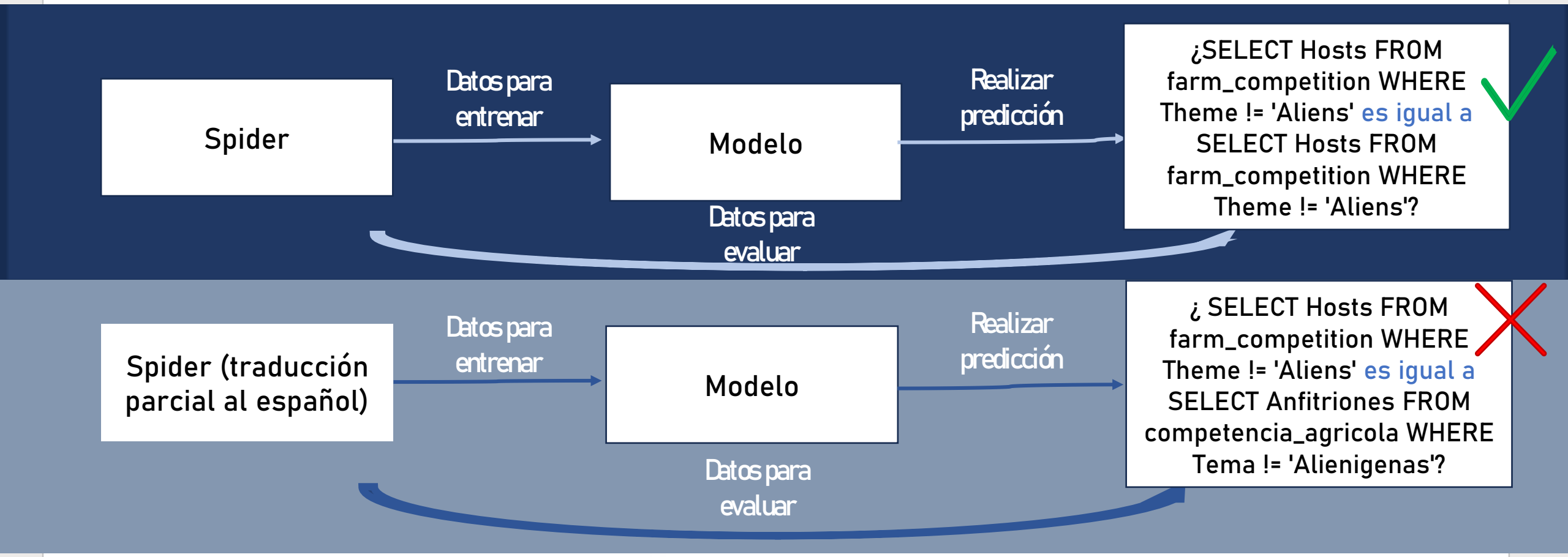


Fig 9. Problema de la comparación entre la sentencia original (inglés) y la sentencia resultante (español)

ACTIVIDADES DESARROLLADAS

Problemas de este enfoque al evaluar el desempeño de un modelo:

¿Cuáles son los anfitriones de concursos cuyo tema no es "Alienígenas"?

What are the hosts of competitions whose theme is not "Aliens"?

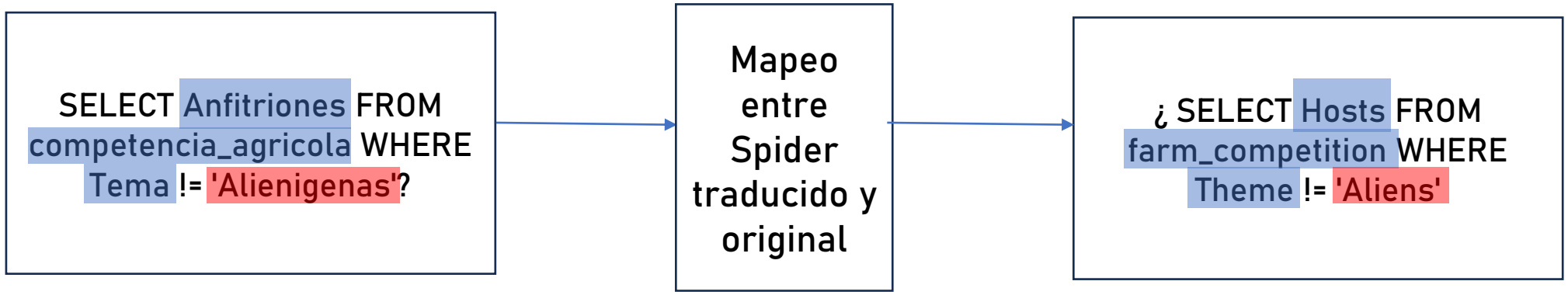


Fig 10. Problema de la comparación entre la sentencia original (inglés) y la sentencia resultante (español) a detalle

ACTIVIDADES DESARROLLADAS

Problemas durante la traducción:

```
cual el estado tiene la Ohio río
qué estados tener ríos llamado Ohio
a través de cual estados hace el Ohio fluir
qué los estados son los siguientes a el Ohio
a través de cual estados hace el carrera de ohio
qué estados hace el Ohio río ir a través de
qué el estado tiene la más grande población
qué es el mayoría populoso estado
qué estado es el mayor en población
cual el estado tiene la más grande población
```

CRONOGRAMA

Tarea	Agos	Sep	Oct	Nov	Dic	Ene	Feb	Mar	Abril	Mayo	Jun
Evaluación y selección de datasets											
Recopilar el dataset											
Análisis profundo de datasets seleccionados											
Preprocesamiento del dataset para su posterior tratamiento de traducción											
Traducción automática											
Etiquetar traducciones obtenidas											
Corrección de traducciones incorrectas											
Evaluación TT1											
Selección de algoritmos de aprendizaje automático											
Creación del modelo											
Entrenamiento del modelo											
Validar el rendimiento del modelo											
Reentrenar modelo en base a los descubrimientos hechos en la fase anterior											
Evaluar el modelo en condiciones de producción											
Documentar el proceso de elaboración del modelo de aprendizaje automático y los experimentos											
Evaluación TT2											

Fig 6. Cronograma

ARQUITECTURA DEL MODELO

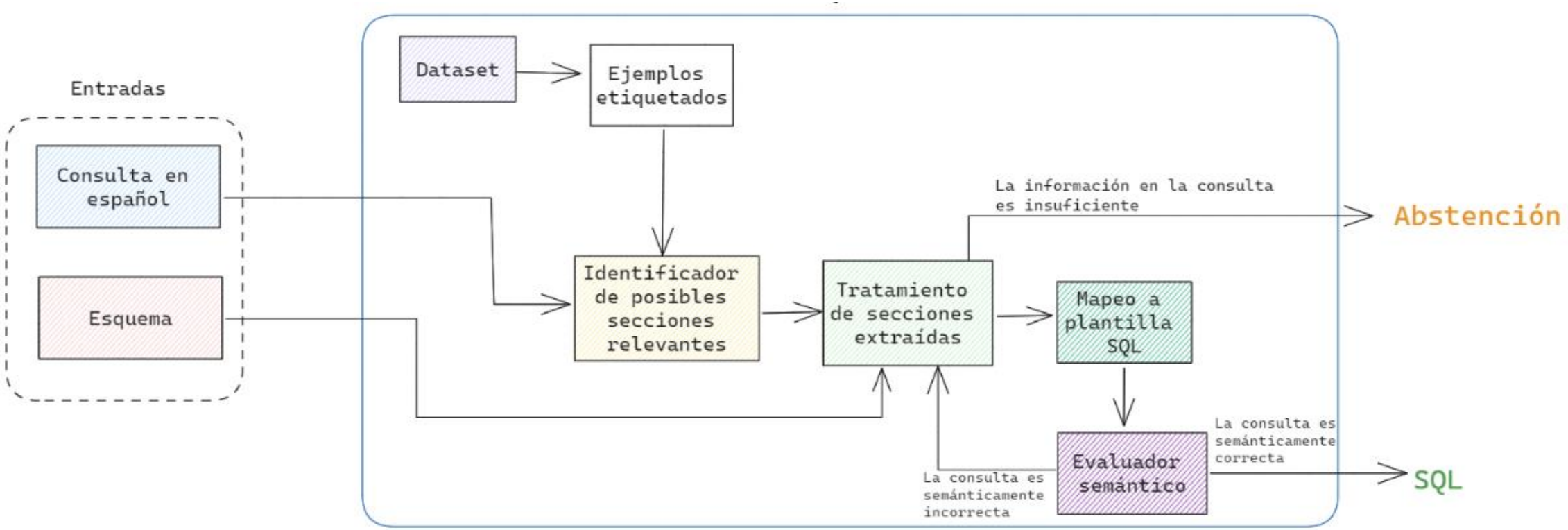
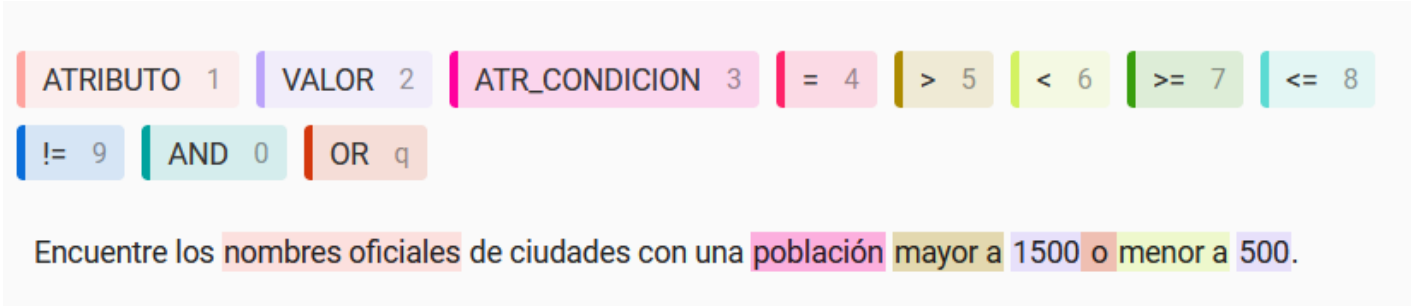


Fig 12. Diagrama de la arquitectura del modelo que proponemos

Nota: Este diagrama no constituye la versión final y puede estar sujeto a actualizaciones y refinamientos a medida que se realicen experimentos y se obtenga una comprensión más detallada de las necesidades del sistema.

ACTIVIDADES DESARROLLADAS

(a)



(b)

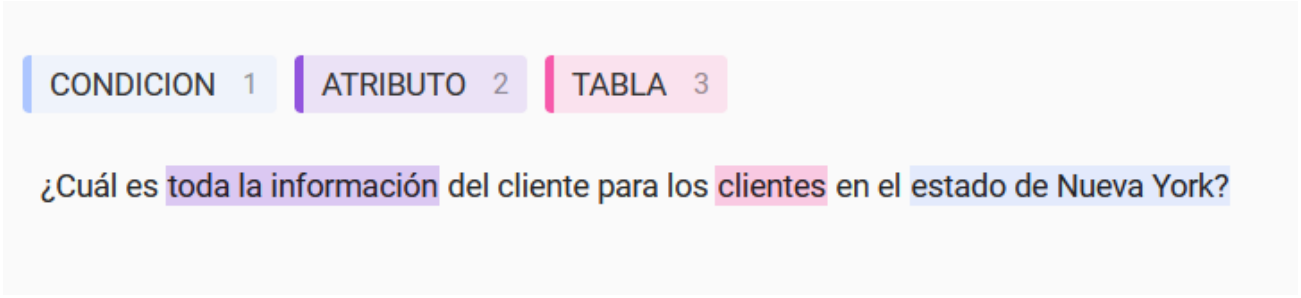


Fig 13. Etiquetado propuesto para las consultas en español. (a) Específico. (b) General

ETIQUETAR TRADUCCIONES OBTENIDAS

- Realizado con label-studio
- Estas nuevas etiquetas aportan en el estado del arte
- Producto extra obtenido

ACTIVIDADES DESARROLLADAS

¿Qué ha funcionado en problemas similares?

- Word2Vec utiliza el contexto de una palabra para adquirir una representación numérica significativa de la misma, lo que se puede transferir al aprendizaje de una categoría

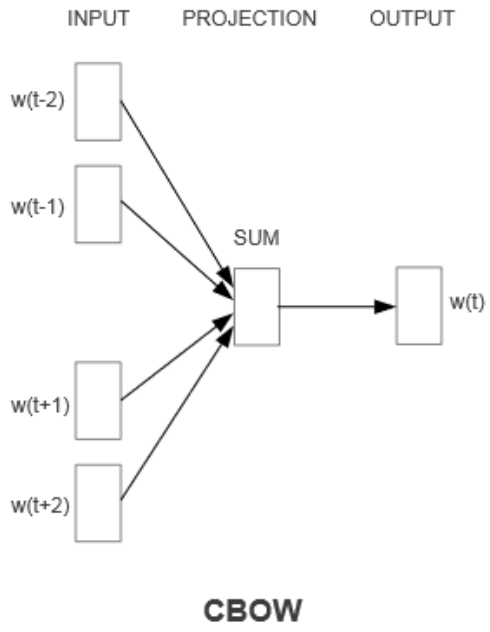


Fig 14. Idea general de generación de embeddings con metodo CBOW, tomado de [20]

[23] M. T. Pilehvar y J. Camacho-Collados, Embeddings in Natural Language Processing: Theory and Advances in Vector Representations of Meaning. Springer Nature, 2022.

ACTIVIDADES DESARROLLADAS

Facilita la generalización, permitiendo que el modelo funcione mejor en datos nuevos y no vistos previamente.

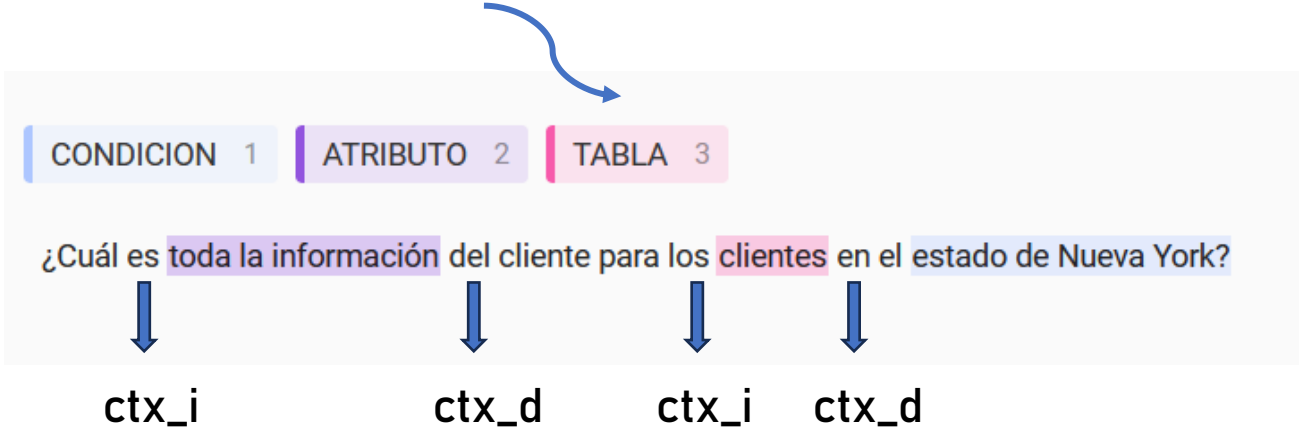


Fig 15. Obtención de contexto de las etiquetas

OBTENER CONTEXTO DE LAS ETIQUETAS

CRONOGRAMA

Tarea	Agos	Sep	Oct	Nov	Dic	Ene	Feb	Mar	Abril	Mayo	Jun
Evaluación y selección de datasets											
Recopilar el dataset											
Análisis profundo de datasets seleccionados											
Preprocesamiento del dataset para su posterior tratamiento de traducción											
Traducción automática											
Etiquetar traducciones obtenidas											
Corrección de traducciones incorrectas											
Evaluación TT1											
Selección de algoritmos de aprendizaje automático											
Creación del modelo											
Entrenamiento del modelo											
Validar el rendimiento del modelo											
Reentrenar modelo en base a los descubrimientos hechos en la fase anterior											
Evaluar el modelo en condiciones de producción											
Documentar el proceso de elaboración del modelo de aprendizaje automático y los experimentos											
Evaluación TT2											

Fig 6. Cronograma

ACTIVIDADES DESARROLLADAS

Identificando el origen de las traducciones incorrectas:

```
¿Cuál es el número máximo y mínimo de vacas en todas las granjas?
Devuelve el número máximo y mínimo de vacas en todas las granjas.
¿Cuántos estados diferentes tienen las ciudades?
Cuente el número de estados diferentes.
Enumere los nombres oficiales de las ciudades en orden descendente de población.
```

```
cual el estado tiene la Ohio río
qué estados tener ríos llamado Ohio
a través de cual estados hace el Ohio fluir
qué los estados son los siguientes a el Ohio
a través de cual estados hace el carrera de ohio
qué estados hace el Ohio río ir a través de
qué el estado tiene la más grande población
qué es el mayoría populoso estado
qué estado es el mayor en población
cual el estado tiene la más grande población
```

ACTIVIDADES DESARROLLADAS

Identificando el origen de las traducciones incorrectas:

```
which states do ohio river flow through
what states does the ohio river run through
what states border the ohio river
which states border the ohio river
what states does the ohio run through
where is the ohio river
which states does the ohio river run through
which states does the ohio run through
which states does the ohio river pass through
what are the states that the ohio run through
which state has the ohio river
```

```
which states do ohio river flow through?
what states does the ohio river run through?
what states border the ohio river?
which states border the ohio river?
what states does the ohio run through?
where is the ohio river?
which states does the ohio river run through?
which states does the ohio run through?
which states does the ohio river pass through?
what are the states that the ohio run through?
which state has the ohio river?
```

ACTIVIDADES DESARROLLADAS

Identificando el origen de las traducciones incorrectas:

```
which states do ohio river flow through?
what states does the ohio river run through?
what states border the ohio river?
which states border the ohio river?
what states does the ohio run through?
where is the ohio river?
which states does the ohio river run through?
which states does the ohio run through?
which states does the ohio river pass through?
what are the states that the ohio run through?
which state has the ohio river?
```

```
¿Por qué estados fluye el río Ohio?
¿Por qué estados pasa el río Ohio?
¿Qué estados bordean el río Ohio?
¿Qué estados bordean el río Ohio?
¿Por qué estados pasa Ohio?
¿Dónde está el río Ohio?
¿Por qué estados pasa el río Ohio?
¿Por qué estados pasa Ohio?
¿Por qué estados pasa el río Ohio?
¿Cuáles son los estados por los que pasa Ohio?
¿Qué estado tiene el río Ohio?
```

RESUMEN

1. Determinamos la importancia de la tarea de texto a SQL.

2. Comparamos los trabajos existentes para el inglés y el español.

3. Encontramos como problemática secundaria la ausencia de un conjunto de datos que cumpla con las especificaciones requeridas, por lo cual partimos de Spider y realizamos una traducción al español.

4. Presentamos nuestra primera propuesta de modelo para resolver alguna de las problemáticas detectadas en los trabajos para el español, la cual desarrollaremos y ajustaremos durante TT 2.

REFERENCIAS

- [1] D. D. Chamberlin, “Early History of SQL”, IEEE Ann. Hist. Comput., vol. 34, núm. 4, pp. 78–82, oct. 2012, doi: 10.1109/MAHC.2012.61.
- [2]“SQL Starter Pack. Overview | by Nate Tsegaw | Medium”. Consultado: el 23 de noviembre de 2023. [En línea]. Disponible en: https://ntsegaw.medium.com/sql-starter-pack-286561037697
- [3] SQL: A Beginner’s Guide, Third Edition 3rd edition by Oppel, Andy, Sheldon, Robert (2008) Paperback.
- [4] W. A. Woods, “Progress in natural language understanding: an application to lunar geology”, en Proceedings of the June 4-8, 1973, national computer conference and exposition on - AFIPS ’73, New York, New York: ACM Press, 1973, p. 441. doi: 10.1145/1499586.1499695.
- [5] “DB-Engines Ranking per database model category”. Consultado: el 29 de noviembre de 2023. [En línea]. Disponible en: https://db-engines.com/en/ranking_categories
- [6] F. Őzcan, A. Quamar, J. Sen, C. Lei, y V. Efthymiou, “State of the Art and Open Challenges in Natural Language Interfaces to Data”, en Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data, Portland OR USA: ACM, jun. 2020, pp. 2629–2636. doi: 10.1145/3318464.3383128.
- [9] D. Alconada, “AlcoNQL : Herramienta de consulta SQL por medio de lenguaje natural”, Universitat Oberta de Catalunya, 2013. [En línea]. Disponible en: https://openaccess.uoc.edu/bitstream/10609/18849/6/diealcoTFC0113memoria.pdf
- [7] M. Bonilla, “Traductor de consultas del lenguaje natural a SQL”, Universidad Central “Marta Abreu” de Las Villas, 2011. [En línea]. Disponible en: https://dspace.uclv.edu.cu/bitstream/handle/123456789/9225/%5b06-28%5d%20Trabajo%20de%20Diploma_Marlen%20FINAL%20OK%20.pdf?sequence=1&isAllowed=y
- [8] F. Reyes García, “LNE2SQL: traductor de consultas del lenguaje natural a SQL v2.0”, Universidad Central “Marta Abreu” de Las Villas, 2012. [En línea]. Disponible en: https://dspace.uclv.edu.cu/bitstream/handle/123456789/6067/Frank%20Reyes%20Garcia--Tesis.pdf?sequence=1&isAllowed=y
- [10] V. Zhong, C. Xiong, y R. Socher, “Seq2SQL: Generating Structured Queries from Natural Language using Reinforcement Learning”, arXiv.org. Consultado: el 12 de noviembre de 2023. [En línea]. Disponible en: https://arxiv.org/abs/1709.00103v7
- [11] X. Xu, C. Liu, y D. Song, “SQLNet: Generating Structured Queries From Natural Language Without Reinforcement Learning”. arXiv, el 13 de noviembre de 2017. Consultado: el 11 de abril de 2023. [En línea]. Disponible en: http://arxiv.org/abs/1711.04436
- [12] T. Yu et al., “Spider: A Large-Scale Human-Labeled Dataset for Complex and Cross-Domain Semantic Parsing and Text-to-SQL Task”. arXiv, el 2 de febrero de 2019. doi: 10.48550/arXiv.1809.08887.

REFERENCIAS

- [13] U. Brunner y K. Stockinger, “ValueNet: A Natural Language-to-SQL System that Learns from Database Information”. arXiv, el 22 de febrero de 2021. Consultado: el 12 de abril de 2023. [En línea]. Disponible en: http://arxiv.org/abs/2006.00888
- [14] B. Wang, R. Shin, X. Liu, O. Polozov, y M. Richardson, “RAT-SQL: Relation-Aware Schema Encoding and Linking for Text-to-SQL Parsers”. arXiv, el 24 de agosto de 2021. Consultado: el 19 de agosto de 2023. [En línea]. Disponible en: http://arxiv.org/abs/1911.04942
- [15] R. Cai, J. Yuan, B. Xu, y Z. Hao, “SADGA: Structure-Aware Dual Graph Aggregation Network for Text-to-SQL”. arXiv, el 17 de enero de 2022. Consultado: el 23 de abril de 2023. [En línea]. Disponible en: http://arxiv.org/abs/2111.00653
- [16] H. Fu, C. Liu, B. Wu, F. Li, J. Tan, y J. Sun, “CatSQL : Towards Real World Natural Language to SQL Applications”, Proc. VLDB Endow., vol. 16, núm. 6, pp. 1534–1547, feb. 2023, doi: 10.14778/3583140.3583165.
- [17] A. Liu, X. Hu, L. Wen, y P. S. Yu, “A comprehensive evaluation of ChatGPT’s zero-shot Text-to-SQL capability”. arXiv, el 11 de marzo de 2023. Consultado: el 23 de abril de 2023. [En línea]. Disponible en: http://arxiv.org/abs/2303.13547
- [18] B. Mahesh, Machine Learning Algorithms -A Review. 2019. doi: 10.21275/ART20203995.
- [19] B. Arinze, “Selecting appropriate forecasting models using rule induction”, Omega, vol. 22, núm. 6, pp. 647–658, nov. 1994, doi: 10.1016/0305-0483(94)90054-X.
- [20] M. T. Pilehvar y J. Camacho-Collados, Embeddings in Natural Language Processing: Theory and Advances in Vector Representations of Meaning. Springer Nature, 2022.
- [21] A. C. Vasquez, “Procesamiento de lenguaje natural”, Rev. Investig. Sist. E Inform., ene. 2009, Consultado: el 12 de noviembre de 2023. [En línea]. Disponible en: https://www.academia.edu/66213908/Procesamiento_de_lenguaje_natural
- [22] BBVA, “¿Qué es la explicabilidad de la IA? Cómo quitarle misterio a la tecnología”, BBVA NOTICIAS. Consultado: el 3 de diciembre de 2023. [En línea]. Disponible en: https://www.bbva.com/es/innovacion/que-es-la-explicabilidad-de-la-ia-como-quitarle-misterio-a-la-tecnologia/
- [23] “Neural Machine Translation by Jointly Learning to Align and Translate”. Consultado: el 3 de diciembre de 2023. [En línea]. Disponible en: https://arxiv.org/abs/1409.0473
- [24] “1.10. Decision Trees — scikit-learn 1.3.2 documentation”. Consultado: el 3 de diciembre de 2023. [En línea]. Disponible en: https://scikit-learn.org/stable/modules/tree.html
- [25] G. Cervone, P. Franzese, y A. P. K. Keesee, “Algorithm quasi-optimal (AQ) learning”, *WIREs Computational Stats*, vol. 2, núm. 2, pp. 218–236, mar. 2010, doi: 10.1002/wics.78.

AVANCES EN TAREAS PARA TT-2

A continuación, mostraremos la implementación de un algoritmo quasi-optimal (AQ) para un obtener un primer panorama de las reglas inducidas:

Tarea	Agos	Sep	Oct	Nov	Dic	Ene	Feb	Mar	Abril	Mayo	Jun
Selección de algoritmos de aprendizaje automático											
Creación del modelo											
Entrenamiento del modelo											
Validar el rendimiento del modelo											
Reentrenar modelo en base a los descubrimientos hechos en la fase anterior											
Evaluar el modelo en condiciones de producción											
Documentar el proceso de elaboración del modelo de aprendizaje automático y los experimentos											
Evaluación TT2											

Fig 6. Cronograma

Trabajo Terminal 1

Modelo generativo de SQL a partir de consultas en español

PRESENTAN

Víctor Ulises Miranda Chávez
Adair Nicolás Hernández
Ives Lancelote Pérez Sánchez
Zury Yael Rubio López

DIRECTORES:

Enrique Alfonso Carmona García
Ituriel Enrique Flores Estrada

APÉNDICES

Referencias

ESTADO DEL ARTE (PARA EL ESPAÑOL)

AlcoNQL [9]

- Necesita que el usuario escriba con gramáticas poco flexibles
- La herramienta está sujeta a una base de datos por diseño
- Requiere que se haga mención explícita de los atributos

	NIVEL 1	NIVEL 2	NIVEL 3	NIVEL 4
Consultas con atributos, selección de tabla y condición	✓	✓	✓	✓
Uso de agregadores, ordenamiento y/o agrupamiento	✗	✓	✓	✓
Involucra múltiples tablas	✗	✗	✓	✓
Involucra consultas anidadas	✗	✗	✗	✓

Ejemplo de consulta válida:
Muestra Nombre y Apellido de personas con DNI incluido en Muestra DNI de asistentes con código igual a Muestra código evento con descripción igual a comida_febrero y confirmado igual 0

Delimitadores

Atributos

Traductor de consultas del lenguaje natural a SQL [7]

- Necesita que el usuario escriba con gramáticas poco flexibles
- Las columnas pueden no ser mencionadas explícitamente, pero para ello debe construirse un diccionario de dominio de forma manual
- Las consultas se hacen sobre una sola tabla, por lo que no se sabe si tiene funcionalidades multi-tabla

	NIVEL 1	NIVEL 2	NIVEL 3	NIVEL 4
Consultas con atributos, selección de tabla y condición	✓	✓	✓	✓
Uso de agregadores, ordenamiento y/o agrupamiento	✗	✓	✓	✓
Involucra múltiples tablas	✗	✗	✓	✓
Involucra consultas anidadas	✗	✗	✗	✓

Ejemplo de consulta válida:

Listar el nombre, primer apellido, fecha de ingreso y la fecha de egreso de pacientes donde el nombre sea igual Antonio

Delimitadores

Atributos

LNE2SQL: traductor de consultas del lenguaje natural a SQL v2.0 [8]

- Permite el uso de gramáticas más flexibles en comparación a las propuestas anteriores
- Tiene la capacidad de trabajar en bases de datos con más de una tabla
- Las columnas pueden no ser mencionadas explícitamente, pero para ello debe construirse un diccionario de dominio de forma manual

	NIVEL 1	NIVEL 2	NIVEL 3	NIVEL 4
Consultas con atributos, selección de tabla y condición	✓	✓	✓	✓
Uso de agregadores, ordenamiento y/o agrupamiento	✗	✓	✓	✓
Involucra múltiples tablas	✗	✗	✓	✓
Involucra consultas anidadas	✗	✗	✗	✓

Ejemplo de consulta válida:

Dame los nombres de los estudiantes y nombres de las asignaturas que están en el nivel mayor que 3 y el grupo 3

Delimitadores

Atributos

DIAGRAMA DE CLASES

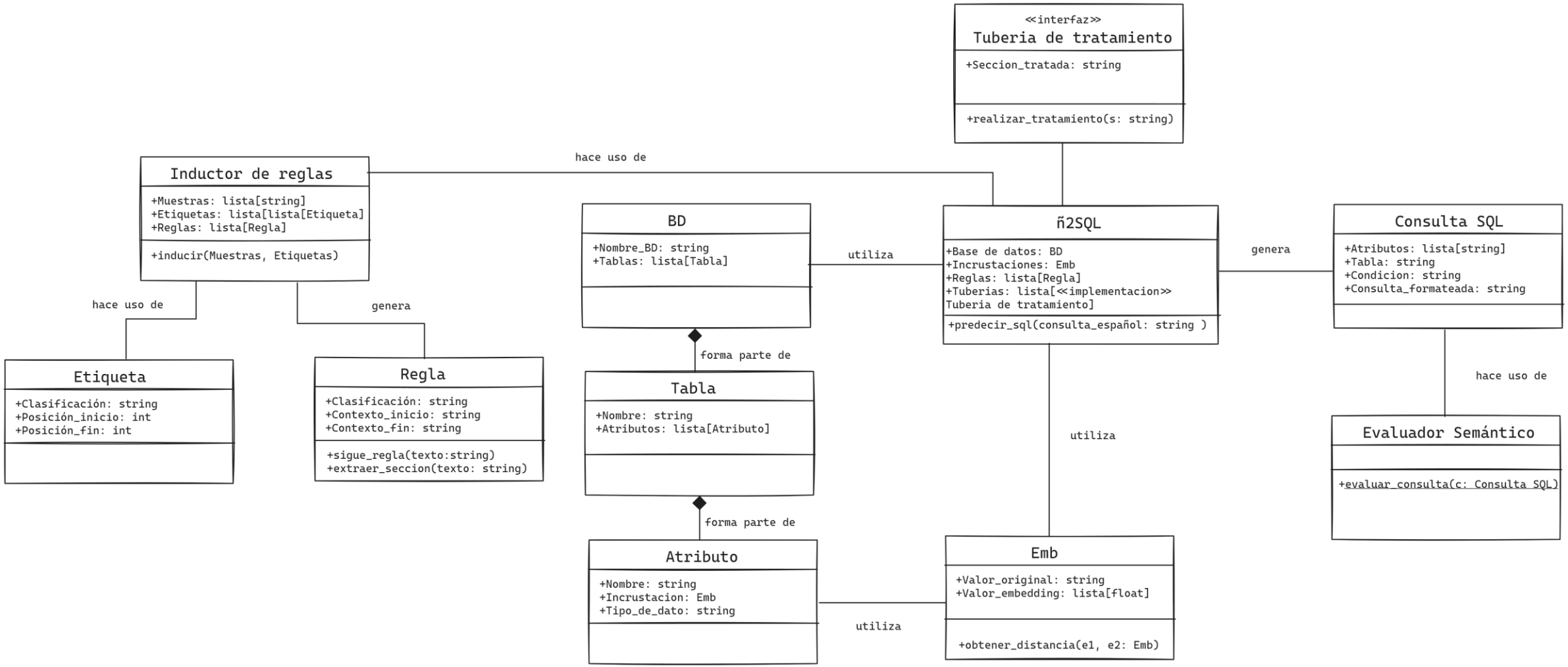


Fig 16. Diagrama de clases

DIAGRAMA DE CLASES

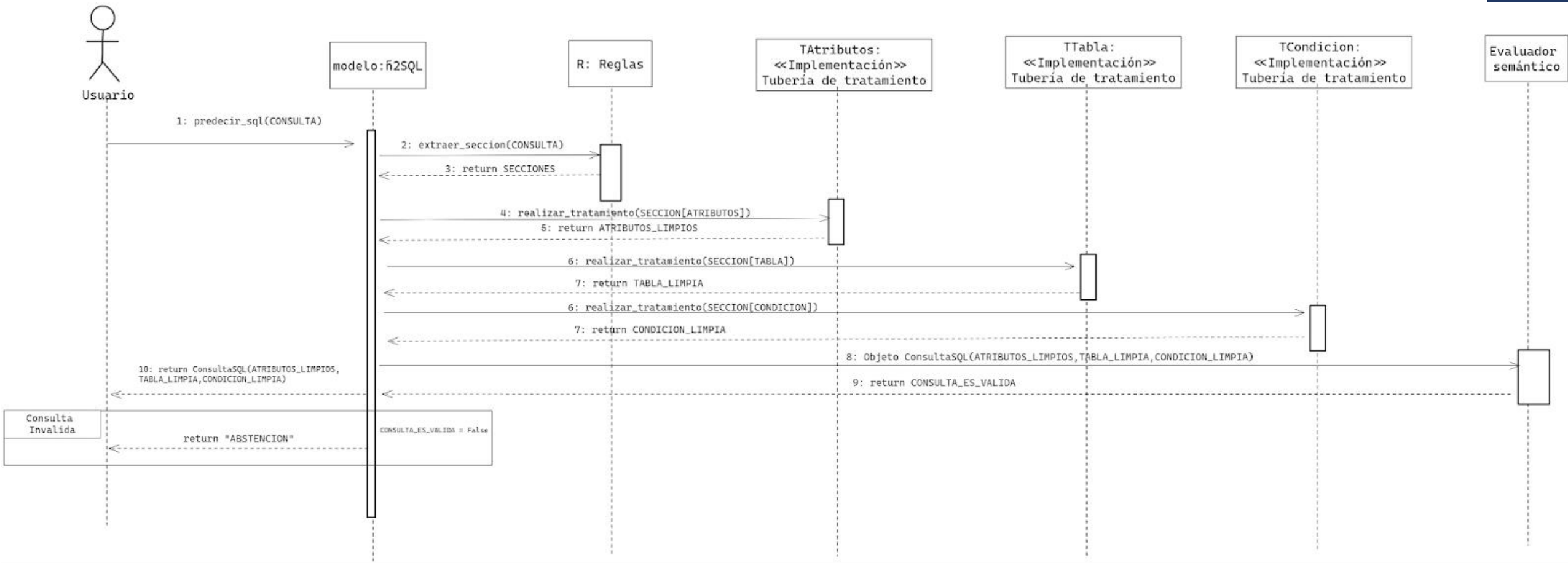


Fig 17. Diagrama de secuencia

REQUERIMIENTOS FUNCIONALES

1.1 El modelo podrá recibir una entrada en lenguaje natural (español) y el esquema de la base de datos.

1.2 El modelo deberá entregar consultas SQL válidas

1.3 En caso de que el modelo obtenga más de una respuesta válida el sistema entregará todas las posibles respuestas

1.4 En caso de no poder generar una consulta SQL válida, el modelo deberá tener la capacidad de abstenerse

1.5 El modelo soportara consultas que involucren la selección de uno o más atributos.

REQUERIMIENTOS FUNCIONALES

1.6 El modelo inferirá la tabla con la que se está interactuando sin necesidad de que esta sea mencionada explícitamente cuando esto le sea posible.

1.7 El modelo deberá ser capaz de inferir los atributos consultados sin necesidad de que sean escritos en la consulta en español exactamente como aparecen en el esquema de la base de datos (cuando esto sea posible)

1.9 El modelo detectará condiciones simples (<,>,=,<=, >=, !=) cuando sea necesario (en caso de ser posible)

REQUERIMIENTOS NO FUNCIONALES

1

Rendimiento

El modelo debe ser capaz de generar consultas SQL en un tiempo razonable

2

Mantenimiento

El modelo debe ser fácil de mantener y actualizar.

3

Eficiencia

El modelo deberá hacer uso óptimo de los recursos computacionales.

4

Mantenimiento

El modelo buscará mantener la consistencia en la generación de consultas SQL tratando de minimizar errores.

SQL

Es el lenguaje principal para realizar consultas y manipular datos en tablas [3].

Con su fórmula de consulta descriptiva, SQL solo requiere la especificación de las condiciones de selección deseadas en la cláusula WHERE [3].

Estructura básica de una consulta SQL [3]:

SELECT creation FROM department

Atributo Cláusula Tabla

[3] SQL: A Beginner's Guide, Third Edition 3rd edition by Oppel, Andy, Sheldon, Robert (2008) Paperback.

CARACTERÍSTICAS DE SQL

Modificadores del Predicado en SQL [3]:

Modificador	Descripción	Ejemplo
WHERE	Filtra filas basado en condiciones específicas.	SELECT * FROM empleados WHERE salario > 50000;
ORDER BY	Ordena resultados según columnas especificadas.	SELECT * FROM productos ORDER BY precio DESC;
GROUP BY	Agrupa filas con valores similares y aplica funciones de agregación.	SELECT SUM(ventas) FROM reporte_ventas GROUP BY mes;
HAVING	Filtra grupos de filas luego de la agrupación (usado con GROUP BY).	SELECT AVG(edad) FROM empleados GROUP BY departamento HAVING AVG(edad) > 30;
DISTINCT	Elimina duplicados de los resultados de la consulta.	SELECT DISTINCT categoria FROM productos;

[3] SQL: A Beginner's Guide, Third Edition 3rd edition by Oppel, Andy, Sheldon, Robert (2008) Paperback.

CARACTERÍSTICAS DE SQL

Operadores de Comparación en SQL [3]:

Operador	Descripción	Ejemplo
=	Igualdad	SELECT * FROM clientes WHERE edad = 25;
!= o <>	Desigualdad	SELECT * FROM empleados WHERE departamento != 'Ventas';
>	Mayor que	SELECT * FROM ventas WHERE cantidad > 100;
<	Menor que	SELECT * FROM productos WHERE precio < 20;
>=	Mayor o igual que	SELECT * FROM empleados WHERE edad >= 30;
<=	Menor o igual que	SELECT * FROM clientes WHERE puntos <= 1000;

[3] SQL: A Beginner's Guide, Third Edition 3rd edition by Oppel, Andy, Sheldon, Robert (2008) Paperback.

CARACTERÍSTICAS DE SQL

Funciones de Agregación en SQL [3]:

Función	Descripción	Ejemplo
MAX()	Valor máximo de una columna	SELECT MAX(precio) FROM productos;
MIN()	Valor mínimo de una columna	SELECT MIN(edad) FROM empleados;
SUM()	Suma de valores de una columna	SELECT SUM(ventas) FROM reporte_ventas;
AVG()	Promedio de valores de una columna	SELECT AVG(edad) FROM empleados;
COUNT()	Número de filas o valores distintos	SELECT COUNT(*) FROM clientes;

[3] SQL: A Beginner's Guide, Third Edition 3rd edition by Oppel, Andy, Sheldon, Robert (2008) Paperback.

CARACTERÍSTICAS DE SQL

Operadores Adicionales en SQL [3]:

Operador	Descripción	Ejemplo
BETWEEN	Selecciona valores dentro de un rango específico.	SELECT * FROM productos WHERE precio BETWEEN 10 AND 50;
LIKE	Busca un patrón en una columna.	SELECT * FROM empleados WHERE nombre LIKE 'Mar%';
IN	Compara un valor con una lista de valores posibles.	SELECT * FROM productos WHERE categoria IN ('Electrónica', 'Ropa', 'Hogar');
NOT	Niega una condición.	SELECT * FROM clientes WHERE NOT edad > 30;

[3] SQL: A Beginner's Guide, Third Edition 3rd edition by Oppel, Andy, Sheldon, Robert (2008) Paperback.

MARCO TEÓRICO

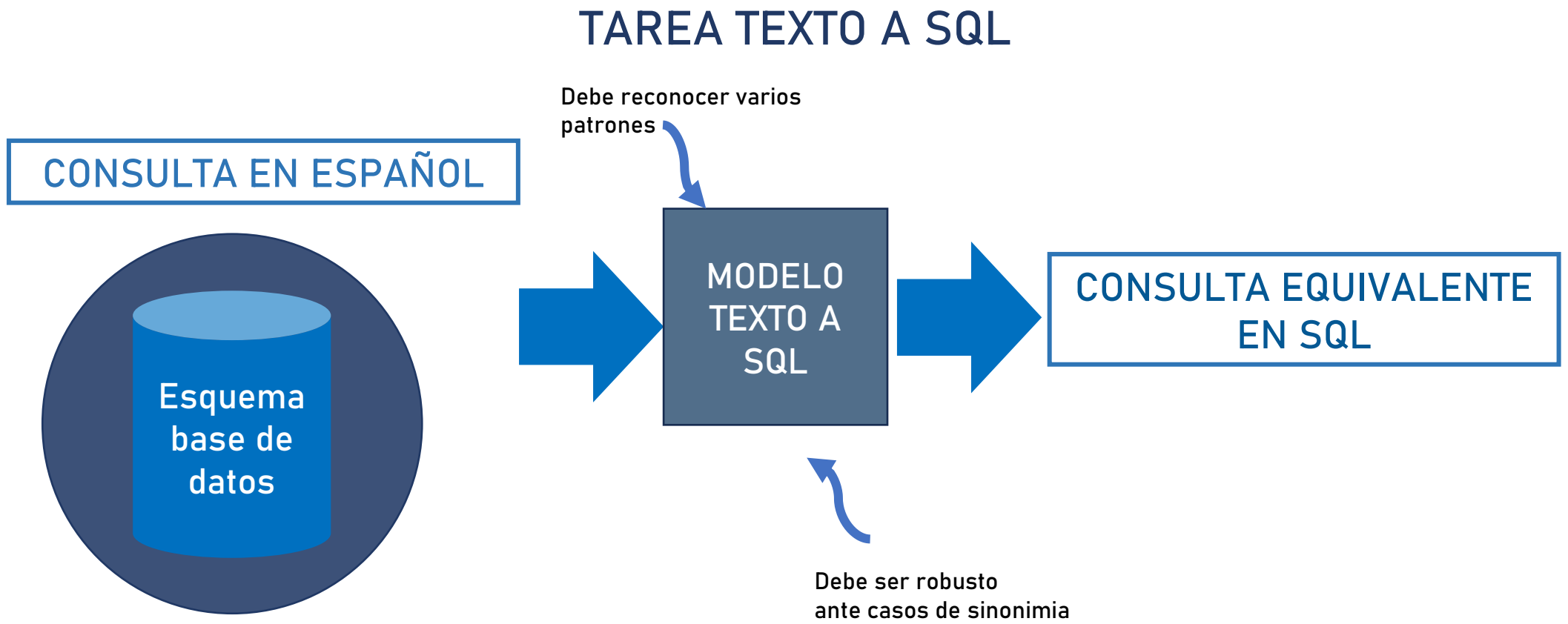
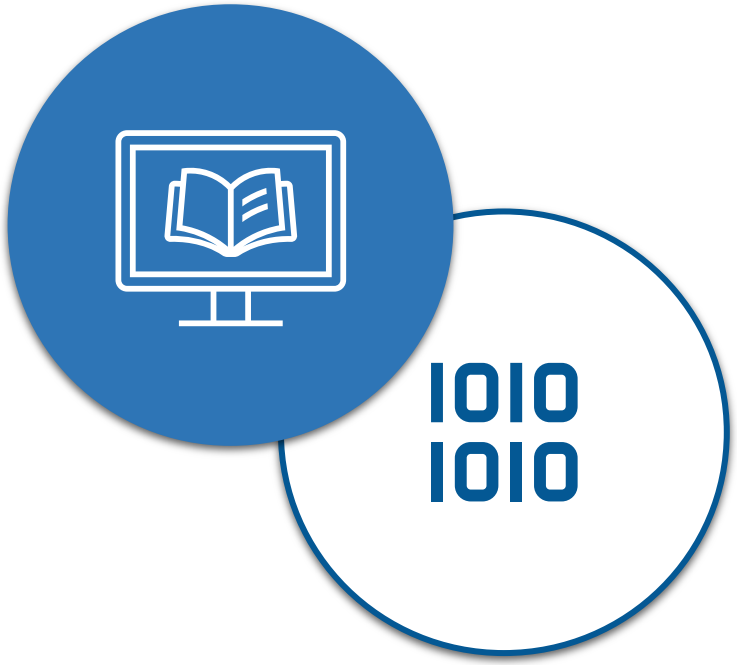


Fig 3. Diagrama básico de la tarea texto a SQL

PROCESAMIENTO DE LENGUAJE NATURAL

Definición:

- El Procesamiento del Lenguaje Natural (PLN) se refiere al uso de lenguaje natural para la comunicación con las computadoras. En esta disciplina, las computadoras deben ser capaces de comprender las oraciones proporcionadas por los usuarios [20].

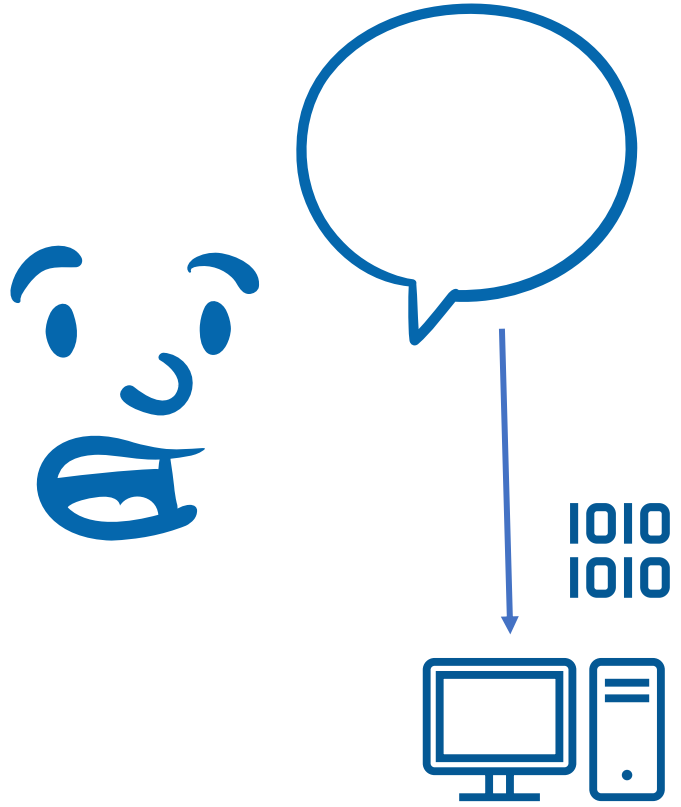


[21] A. C. Vasquez, “Procesamiento de lenguaje natural”, Rev. Investig. Sist. E Inform., ene. 2009, Consultado: el 12 de noviembre de 2023. [En línea]. Disponible en: https://www.academia.edu/66213908/Procesamiento_de_lenguaje_natural

PROCESAMIENTO DE LENGUAJE NATURAL

Problemas al procesar lenguaje natural

- La ambigüedad dificulta comprender la necesidad real del usuario [20].
- Se pueden generar diversas interpretaciones válidas.
- Algunas de ellas no reflejan exactamente lo que el usuario pretendía [20].

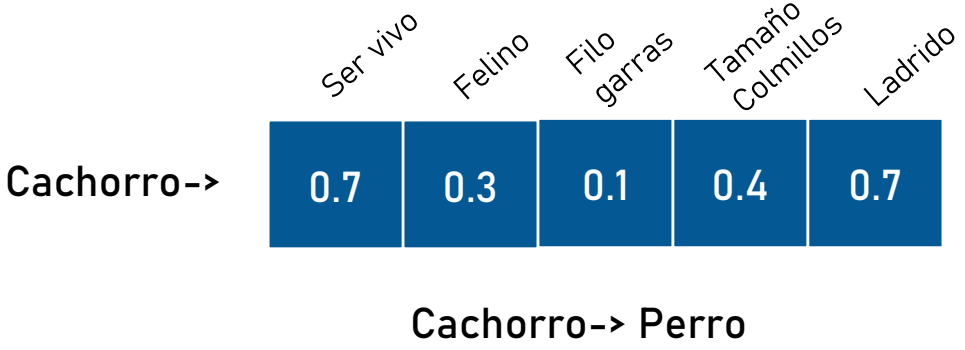


[21] A. C. Vasquez, “Procesamiento de lenguaje natural”, Rev. Investig. Sist. E Inform., ene. 2009, Consultado: el 12 de noviembre de 2023. [En línea]. Disponible en: https://www.academia.edu/66213908/Procesamiento_de_lenguaje_natural

WORD EMBEDDINGS

Word embeddings o incrustaciones:

- Métodos utilizados en el campo en PLN para crear representaciones numéricas de palabras y textos que capturan la semántica y el contexto [23].



[23] M. T. Pilehvar y J. Camacho-Collados, Embeddings in Natural Language Processing: Theory and Advances in Vector Representations of Meaning. Springer Nature, 2022.

MARCO TEORICO

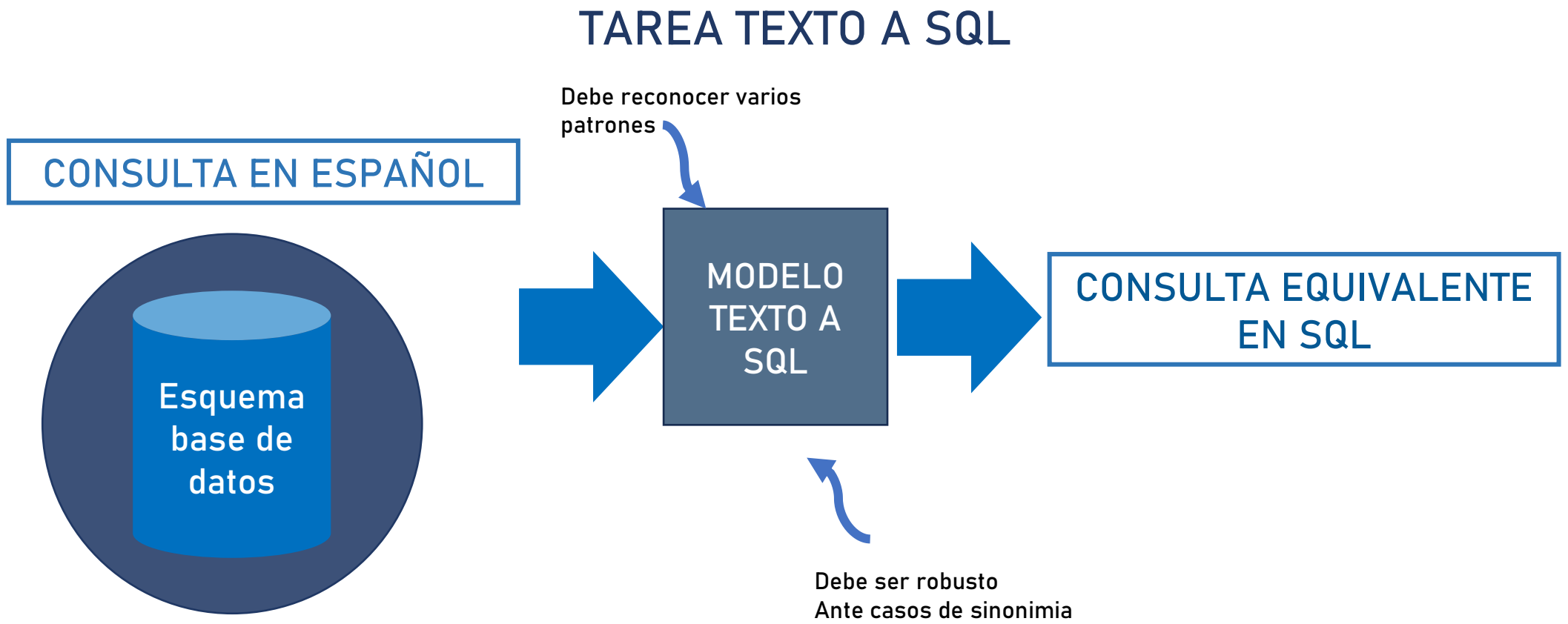
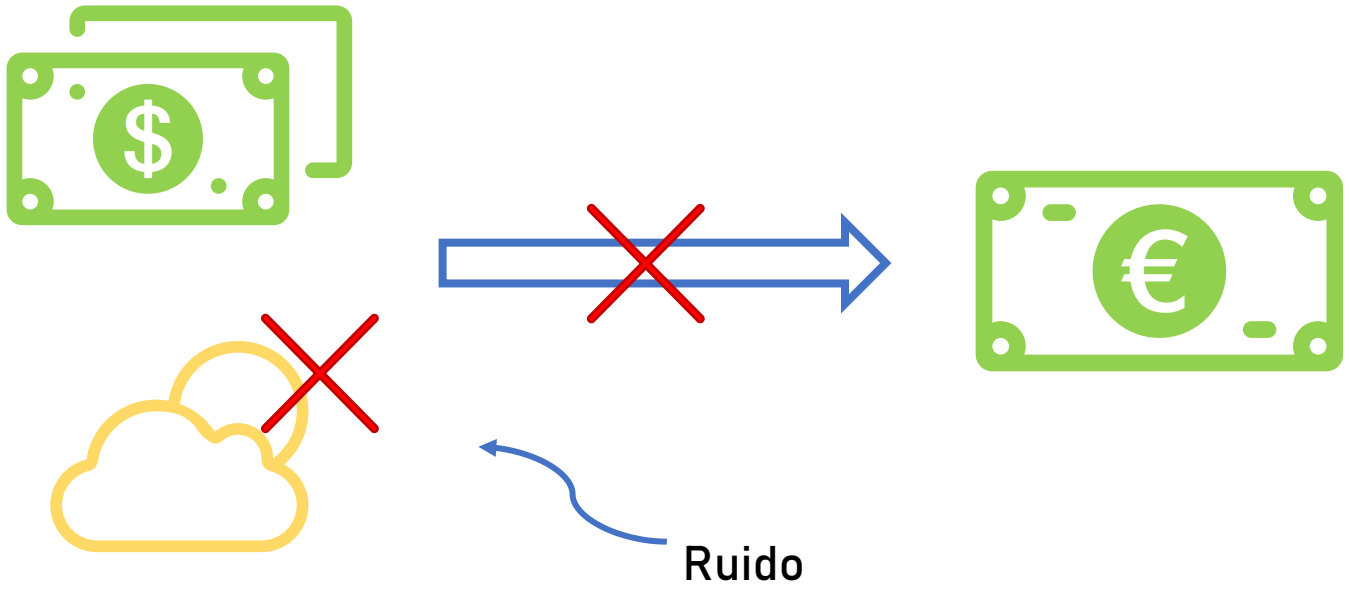


Fig 3. Diagrama básico de la tarea texto a SQL

ACTIVIDADES DESARROLLADAS

¿Que hace a las etiquetas… “buenas etiquetas”?

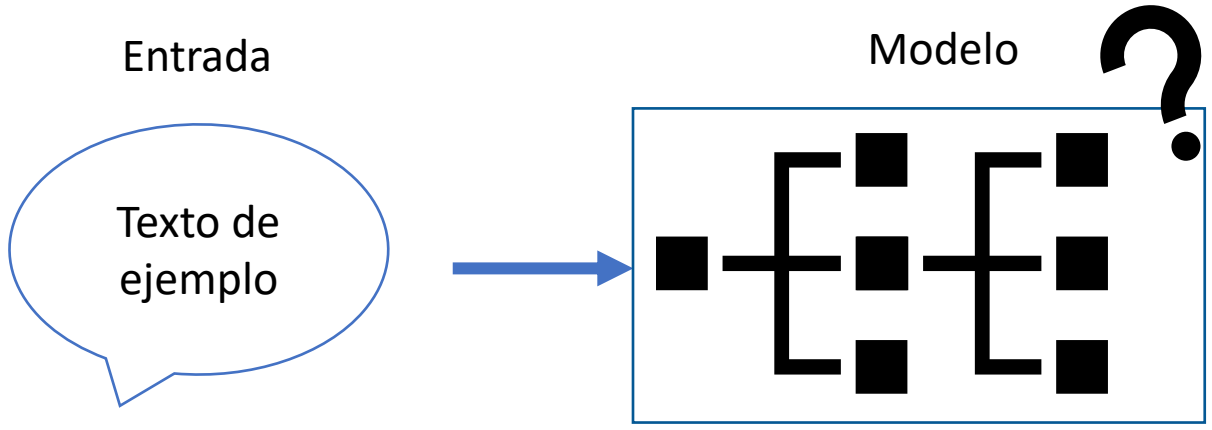
- Las etiquetas deben ser relevantes para resolver el problema
- Evitar datos irrelevantes o que no aporten información (Ruido)



ACTIVIDADES DESARROLLADAS

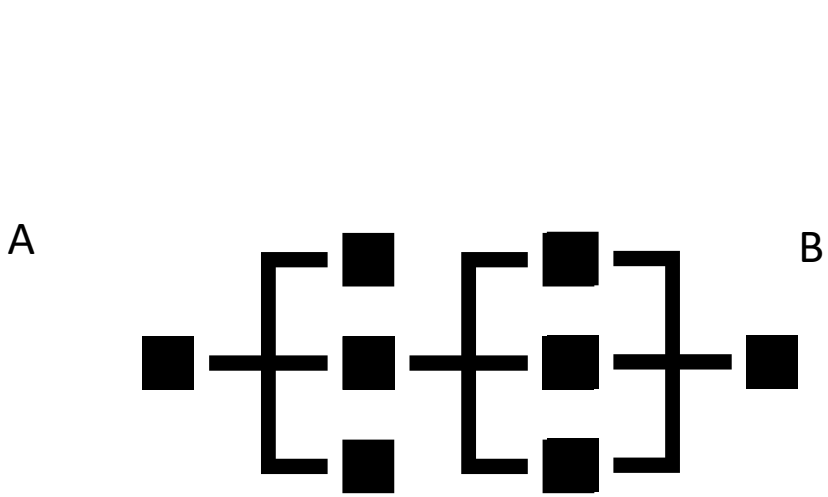
¿Que hace a las etiquetas… “buenas etiquetas”?

- ¿Pueden ser procesadas por el modelo en cuestión?
- Por lo cual, antes de etiquetar, debemos saber que etiquetas necesita



ACTIVIDADES DESARROLLADAS

¿Qué salida(s) buscamos producir?



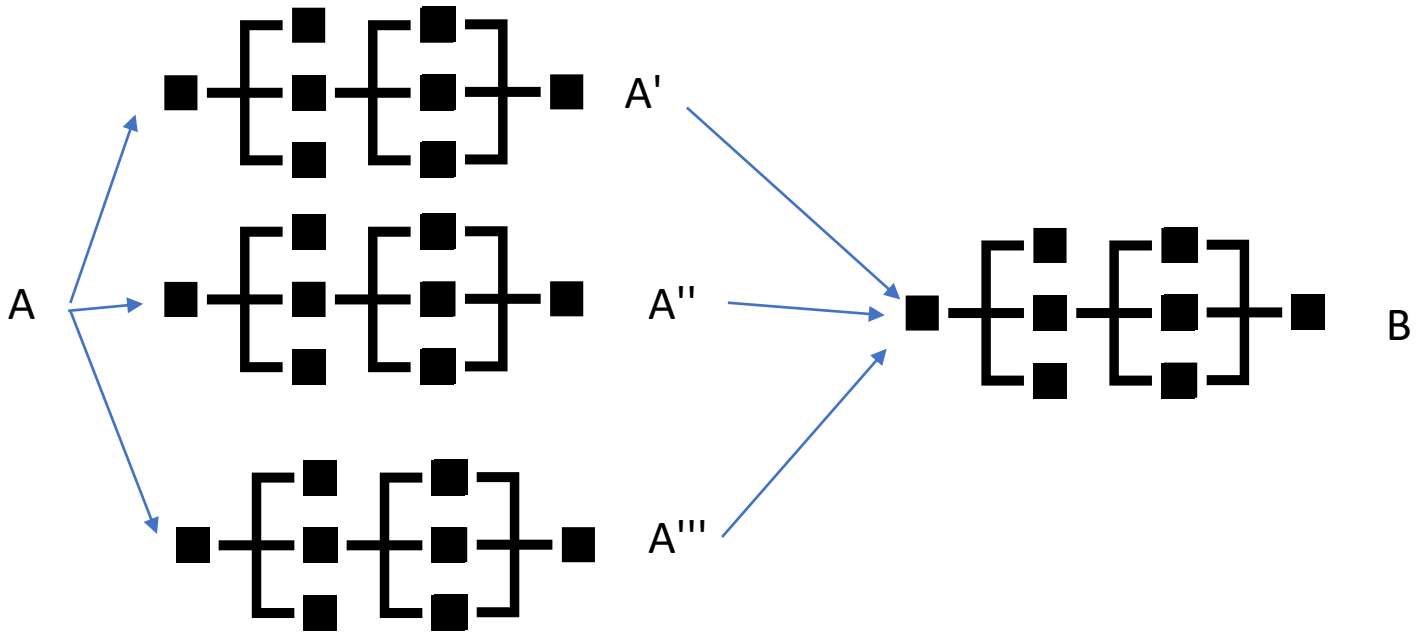


Fig 11. Tipos de salidas analizadas.

ACTIVIDADES DESARROLLADAS

¿Por qué usar inducción de reglas sobre las redes neuronales en la tarea *text to sql?*

Ventajas del uso de reglas

- Explicabilidad [20]
- Menos Dependencia de Datos
- Generalización Transparente [22]

Desventajas del uso de redes neuronales

- Falta de Interpretabilidad [21]
- Requieren Grandes Conjuntos de Datos [21]
- Requerimientos de Recursos Computacionales [21]

[22] BBVA, "¿Qué es la explicabilidad de la IA? Cómo quitarle misterio a la tecnología", BBVA NOTICIAS. Consultado: el 3 de diciembre de 2023. [En línea]. Disponible en: https://www.bbva.com/es/innovacion/que-es-la-explicabilidad-de-la-ia-como-quitarle-misterio-a-la-tecnologia/
[23] "Neural Machine Translation by Jointly Learning to Align and Translate". Consultado: el 3 de diciembre de 2023. [En línea]. Disponible en: https://arxiv.org/abs/1409.0473
[24] "1.10. Decision Trees — scikit-learn 1.3.2 documentation". Consultado: el 3 de diciembre de 2023. [En línea]. Disponible en: https://scikit-learn.org/stable/modules/tree.html

ALGORITMO QUASI - OPTIMAL

Descripción del algoritmo:

- Es una técnica de aprendizaje automático que busca aprender reglas de decisión simbólicas a partir de ejemplos y contraejemplos.
- Este algoritmo ha evolucionado para abordar desafíos más amplios de cobertura.

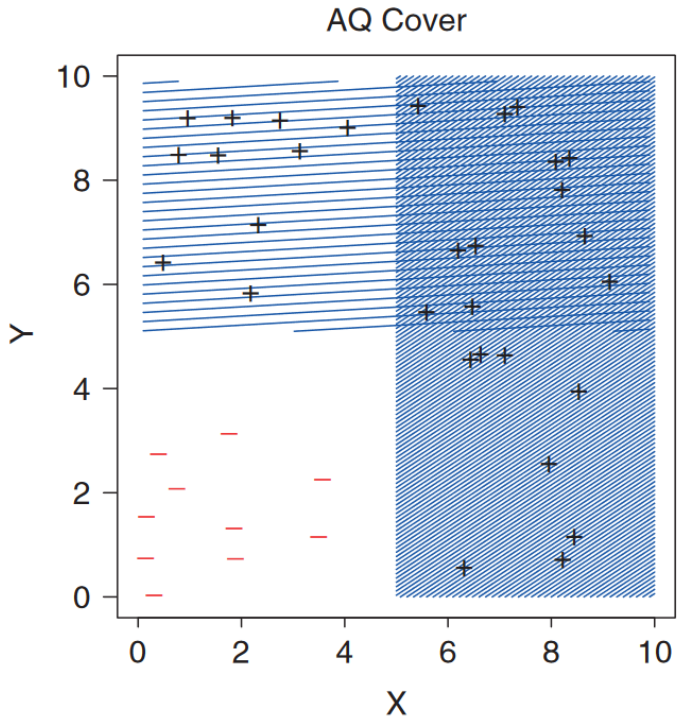


Fig 12. Cobertura del algoritmo AQ [25].

[25] G. Cervone, P. Franzese, y A. P. K. Keesee, "Algorithm quasi-optimal (AQ) learning", *WIREs Computational Stats*, vol. 2, núm. 2, pp. 218–236, mar. 2010, doi: 10.1002/wics.78.

REFERENCIAS

- [1] D. D. Chamberlin, “Early History of SQL”, IEEE Ann. Hist. Comput., vol. 34, núm. 4, pp. 78–82, oct. 2012, doi: 10.1109/MAHC.2012.61.
- [2]“SQL Starter Pack. Overview | by Nate Tsegaw | Medium”. Consultado: el 23 de noviembre de 2023. [En línea]. Disponible en: https://ntsegaw.medium.com/sql-starter-pack-286561037697
- [3] SQL: A Beginner’s Guide, Third Edition 3rd edition by Oppel, Andy, Sheldon, Robert (2008) Paperback.
- [4] W. A. Woods, “Progress in natural language understanding: an application to lunar geology”, en Proceedings of the June 4-8, 1973, national computer conference and exposition on - AFIPS ’73, New York, New York: ACM Press, 1973, p. 441. doi: 10.1145/1499586.1499695.
- [5] “DB-Engines Ranking per database model category”. Consultado: el 29 de noviembre de 2023. [En línea]. Disponible en: https://db-engines.com/en/ranking_categories
- [6] F. Őzcan, A. Quamar, J. Sen, C. Lei, y V. Efthymiou, “State of the Art and Open Challenges in Natural Language Interfaces to Data”, en Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data, Portland OR USA: ACM, jun. 2020, pp. 2629–2636. doi: 10.1145/3318464.3383128.
- [9] D. Alconada, “AlcoNQL : Herramienta de consulta SQL por medio de lenguaje natural”, Universitat Oberta de Catalunya, 2013. [En línea]. Disponible en: https://openaccess.uoc.edu/bitstream/10609/18849/6/diealcoTFC0113memoria.pdf
- [7] M. Bonilla, “Traductor de consultas del lenguaje natural a SQL”, Universidad Central “Marta Abreu” de Las Villas, 2011. [En línea]. Disponible en: https://dspace.uclv.edu.cu/bitstream/handle/123456789/9225/%5b06-28%5d%20Trabajo%20de%20Diploma_Marlen%20FINAL%20OK%20.pdf?sequence=1&isAllowed=y
- [8] F. Reyes García, “LNE2SQL: traductor de consultas del lenguaje natural a SQL v2.0”, Universidad Central “Marta Abreu” de Las Villas, 2012. [En línea]. Disponible en: https://dspace.uclv.edu.cu/bitstream/handle/123456789/6067/Frank%20Reyes%20Garcia--Tesis.pdf?sequence=1&isAllowed=y
- [10] V. Zhong, C. Xiong, y R. Socher, “Seq2SQL: Generating Structured Queries from Natural Language using Reinforcement Learning”, arXiv.org. Consultado: el 12 de noviembre de 2023. [En línea]. Disponible en: https://arxiv.org/abs/1709.00103v7
- [11] X. Xu, C. Liu, y D. Song, “SQLNet: Generating Structured Queries From Natural Language Without Reinforcement Learning”. arXiv, el 13 de noviembre de 2017. Consultado: el 11 de abril de 2023. [En línea]. Disponible en: http://arxiv.org/abs/1711.04436
- [12] T. Yu et al., “Spider: A Large-Scale Human-Labeled Dataset for Complex and Cross-Domain Semantic Parsing and Text-to-SQL Task”. arXiv, el 2 de febrero de 2019. doi: 10.48550/arXiv.1809.08887.

REFERENCIAS

- [13] U. Brunner y K. Stockinger, “ValueNet: A Natural Language-to-SQL System that Learns from Database Information”. arXiv, el 22 de febrero de 2021. Consultado: el 12 de abril de 2023. [En línea]. Disponible en: http://arxiv.org/abs/2006.00888
- [14] B. Wang, R. Shin, X. Liu, O. Polozov, y M. Richardson, “RAT-SQL: Relation-Aware Schema Encoding and Linking for Text-to-SQL Parsers”. arXiv, el 24 de agosto de 2021. Consultado: el 19 de agosto de 2023. [En línea]. Disponible en: http://arxiv.org/abs/1911.04942
- [15] R. Cai, J. Yuan, B. Xu, y Z. Hao, “SADGA: Structure-Aware Dual Graph Aggregation Network for Text-to-SQL”. arXiv, el 17 de enero de 2022. Consultado: el 23 de abril de 2023. [En línea]. Disponible en: http://arxiv.org/abs/2111.00653
- [16] H. Fu, C. Liu, B. Wu, F. Li, J. Tan, y J. Sun, “CatSQL : Towards Real World Natural Language to SQL Applications”, Proc. VLDB Endow., vol. 16, núm. 6, pp. 1534–1547, feb. 2023, doi: 10.14778/3583140.3583165.
- [17] A. Liu, X. Hu, L. Wen, y P. S. Yu, “A comprehensive evaluation of ChatGPT’s zero-shot Text-to-SQL capability”. arXiv, el 11 de marzo de 2023. Consultado: el 23 de abril de 2023. [En línea]. Disponible en: http://arxiv.org/abs/2303.13547
- [18] B. Mahesh, Machine Learning Algorithms -A Review. 2019. doi: 10.21275/ART20203995.
- [19] B. Arinze, “Selecting appropriate forecasting models using rule induction”, Omega, vol. 22, núm. 6, pp. 647–658, nov. 1994, doi: 10.1016/0305-0483(94)90054-X.
- [20] M. T. Pilehvar y J. Camacho-Collados, Embeddings in Natural Language Processing: Theory and Advances in Vector Representations of Meaning. Springer Nature, 2022.
- [21] A. C. Vasquez, “Procesamiento de lenguaje natural”, Rev. Investig. Sist. E Inform., ene. 2009, Consultado: el 12 de noviembre de 2023. [En línea]. Disponible en: https://www.academia.edu/66213908/Procesamiento_de_lenguaje_natural
- [22] BBVA, “¿Qué es la explicabilidad de la IA? Cómo quitarle misterio a la tecnología”, BBVA NOTICIAS. Consultado: el 3 de diciembre de 2023. [En línea]. Disponible en: https://www.bbva.com/es/innovacion/que-es-la-explicabilidad-de-la-ia-como-quitarle-misterio-a-la-tecnologia/
- [23] “Neural Machine Translation by Jointly Learning to Align and Translate”. Consultado: el 3 de diciembre de 2023. [En línea]. Disponible en: https://arxiv.org/abs/1409.0473
- [24] “1.10. Decision Trees — scikit-learn 1.3.2 documentation”. Consultado: el 3 de diciembre de 2023. [En línea]. Disponible en: https://scikit-learn.org/stable/modules/tree.html
- [25] G. Cervone, P. Franzese, y A. P. K. Keesee, “Algorithm quasi-optimal (AQ) learning”, *WIREs Computational Stats*, vol. 2, núm. 2, pp. 218–236, mar. 2010, doi: 10.1002/wics.78.

Trabajo Terminal 1

Modelo generativo de SQL a partir de consultas en español

PRESENTAN

Víctor Ulises Miranda Chávez
Adair Nicolás Hernández
Ives Lancelote Pérez Sánchez
Zury Yael Rubio López

DIRECTORES:

Enrique Alfonso Carmona García
Ituriel Enrique Flores Estrada