

UNIVERSIDAD AGRARIA DE LA HABANA
“Fructuoso Rodríguez Pérez”

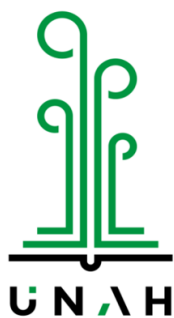
FACULTAD DE CIENCIAS TÉCNICAS

Método para la realización de consultas en un Sistema de Información Geográfica. Caso de estudio UDP “El Guayabal”.

Autor: Ivett Sosa Franco

TESIS EN OPCIÓN AL TÍTULO DE MÁSTER EN BIOMATEMÁTICA

MAYABEQUE, 2022



UNIVERSIDAD AGRARIA DE LA HABANA
“Fructuoso Rodríguez Pérez”

FACULTAD DE CIENCIAS TÉCNICAS

Método para la realización de consultas en un Sistema de Información Geográfica. Caso de estudio UDP “El Guayabal”.

Autor: Ivett Sosa Franco

Tutoras: Dra.C Neili Machado García

Dra.C María Elena Ruiz Pérez

TESIS EN OPCIÓN AL TÍTULO DE MÁSTER EN BIOMATEMÁTICA

MAYABEQUE, 2022

*“No existe una manera fácil.
No importa cuán talentoso seas,
tu talento te va a fallar
sino lo desarrollas.
Si no estudias,
si no trabajas duro,
si no te dedicas a ser
mejor cada día.”*

- Will Smith

Agradecimientos

Primero quiero darle gracias a Dios por permitirme estar hoy aquí.

A mi mamá, mi mamita por ser mi motor. Por apoyarme siempre. Por alentarme a superarme y salir adelante. Por estar siempre presente en cada paso que doy, en las buenas y en las malas. Por ayudarme a resolver mis problemas y no permitir que me rinda. Por sacrificarse tanto por mí como por mi hermano y pensar en nosotros antes que en ella. La persona que soy hoy, en gran medida es gracias a ti. Te amo mucho !!!

A mi padre, mi verdadero padre, muchas gracias por todo, por tu ayuda por tu cariño, por tu preocupación, por tus consejos. Por ser un ejemplo para mí, como persona y como profesional. Por siempre estar ahí cuando lo he necesitado en momentos buenos y malos. Por compartir conmigo tus conocimientos, mi profesor de programación y de todas las ciencias y por aguantar mis berrinches cuando ya no quería estudiar. Y si hoy estoy aquí, también es gracias a ti. Te quiero mucho !!!

A mi hermano, Rorry, te quiero muchísimo !!! gracias por estar presente en mi vida, por tus malcriadeces y tu cariño sincero. Siempre quise tener un hermano, y me tocó el mejor. Espero con ansias un día estar del otro lado escuchando tus agradecimientos cuando seas tú el que se gradúe.

A mi novio Gabriel, mi amor muchas gracias por ayudarme y apoyarme siempre. Gracias por haber llegado a mi vida hace ya casi 3 años y poco a poco convertirme en una mejor persona. Por animarme y darme esperanzas cada vez que te decía entre lágrimas que no aguantaba más. Por ayudarme a superarme y siempre creer en mí. Gracias por ser mi amigo, mi compañero, mi peleón. Por todo lo bueno y malo que hemos vivido. Esta es otra meta que hoy vencemos juntos. Te amo muchísimo !!!

A los cuatro gracias por estar hoy aquí, compartir mis logros que son también de ustedes y aguantar mis malcriadeces. Todo lo que pueda decir hoy entre lágrimas se queda pequeño comparado con lo que siento. Infinitas gracias, mi vida no sería lo mismo sin ustedes en ella. Los quiero mucho a todos !!!

A mi abuela Gilda, gracias por tanto cariño y apoyo incondicional en estos 24 años. Por siempre estar presente y pendiente hasta en la distancia.

A mi tía Yuley, mi madrina Tania y mi prima Gretter muchas gracias por su apoyo y cariño.

A la pequeña Rocy, mi niña, te quiero muchísimo !!!

A la familia de Gaby, a su mamá, a su papá, a sus hermanitas y a su hermano, a Brian y a su familia y a su tía Nana, gracias por su apoyo y preocupación y aunque, todos no estén de forma física hoy aquí nos apoyan desde la distancia.

A mi tutora Neili, muchas gracias por aceptar tutorarme desde la tesis de grado, sin vacilar, cuando Yulier se fue de la universidad y me quedé sin tutor comenzando la tesis. Gracias por seguir apoyándome, por brindarme tus conocimientos y ayuda hasta el final. Por todas tus críticas constructivas y por tu guía incondicional.

A mi tutora María Elena, muchas gracias por toda su ayuda profe, por sus conocimientos, por su apoyo y por sus consejos. Gracias por no dudar al aceptar tutorarme, aunque faltaba mucho por hacer. Sin su empujón en la última recta esto no fuera posible. Gracias por recibirme, abrirme las puertas de su casa, por las charlas. Por todo el tiempo dedicado, hoy otro logro es haberla conocido.

Sin ustedes dos esto no fuera posible, muchas gracias !!!

A la profe Lucy por siempre estar pendiente, por su preocupación. De forma general a todo el claustro que nos impartió las asignaturas en esta edición, Norlan, Alexis, Ileana, Yusney, Lilibeth, Nelson, Gustavo. Muchas gracias a todos !

A mi oponente, mi profe de física, profe usted sabe que lo aprecio mucho y para mí fue un honor y un placer tenerlo cerca nuevamente.

A mis compañeros de trabajo y amigos, a mis compañeros de la maestría, gracias por estar pendientes y brindarme su apoyo y preocupación. A todos muchas gracias!

A los trabajadores de la UDP “El Guayabal” por el tiempo, la ayuda y la atención brindada hasta el último momento.

A mis padres, mi hermano y mi novio, esto es por y para ustedes.

Resumen

Una gran parte de la información geoespacial se encuentra almacenada en bases de datos relacionales. Las mismas proporcionan un método confiable para gestionar grandes cantidades de datos al tiempo que ofrecen una combinación de rendimiento del sistema y facilidad de implementación. Los usuarios que interactúan con estos sistemas le realizan consultas con el principal objetivo de recuperar información, sin embargo, no todos poseen habilidades informáticas para realizar los complejos procesos que esta operación requiere. Lo conveniente sería realizar estas consultas en lenguaje natural. En esta investigación se abordan los principales temas relacionados con los Sistemas de Información Geográfica y sus herramientas más populares a nivel mundial. También se explican aspectos relevantes sobre los Sistemas de Gestión de Bases de Datos y se establece una comparación entre los Relacionales y los NoSQL. Además, se plantean los principales tipos de consultas, así como el lenguaje en el que se realizan y se muestran algunos ejemplos de las mismas. Finalmente, se desarrolla un método informático para ayudar a los usuarios inexpertos en SIG y SQL a transformar las preguntas de lenguaje natural a consultas SQL.

Palabras clave: Sistemas de Información Geográfica, Sistemas de Gestión de Bases de datos, consultas SQL.

Abatract

Much geospatial information is stored in relational databases. Relational databases provide a reliable method for managing large amounts of data while offering a combination of system performance and ease of implementation. Users interacting with these systems perform queries on them with the main objective of retrieving information, however not all of them have the computer skills to perform the complex processes that this operation requires. It would be convenient to perform these queries in natural language. This research addresses the main topics related to Geographic Information Systems and their most popular tools worldwide. Relevant aspects about Database Management Systems are also explained and a comparison between Relational and NoSQL is established. In addition, the main types of queries are presented, as well as the language in which they are made and some examples of them are shown. Finally, a computational method is developed to help inexperienced GIS and SQL users to transform natural language queries into SQL queries.

Keywords: Geographic Information Systems, Database Management Systems, SQL queries.

Índice

Introducción.....	1
Capítulo I: Revisión bibliográfica	5
1.1. Introducción del capítulo	5
1.2. Sistemas de Información Geográfica: conceptos y principales herramientas.....	5
1.2.1. Selección entre QGIS y ArcGIS.....	6
1.2.2. Tipos de datos en un SIG: datos espaciales y de atributos	8
1.2.2.1. Datos espaciales.....	8
1.2.2.2. Datos de atributos	9
1.3. Bases de datos	11
1.3.1. Sistema de Gestión de Bases de Datos	11
1.3.1.1. Selección entre Sistemas de Gestión de Base de Datos Relacionales y NoSQL	12
1.3.1.2. PostgreSQL.....	14
1.3.2. Bases de datos espaciales	14
1.3.2.1. PostGIS extensión de PostgreSQL	15
1.3.3. Consultas SQL.....	15
1.3.3.1. SQL, lenguaje de consulta estructurado	16
1.4. Consultas en SIG	17
1.4.1. Tipos de consultas en SIG.....	18
1.4.1.1. Selección.....	18
1.4.1.2. Consulta por atributo	19
1.4.1.3. Consulta espacial.....	19

1.4.1.3.1. Relaciones espaciales.....	20
1.5. Conclusiones parciales	20
Capítulo II: Materiales y métodos	22
2.1 Introducción del capítulo	22
2.2 Caracterización del área de estudio	22
2.3 Tecnologías y herramientas utilizadas	24
2.4 Procesamiento de la consulta en lenguaje natural con LeNatSQL	28
2.5 Conclusiones parciales	32
Capítulo III: Resultados y discusión	33
3.1 Introducción capítulo	33
3.2 Consultas en LeNatSQL.....	33
3.3 Visualización de la consulta obtenida en LeNatSQL a través de postgisQueryBuilder y QGIS	36
3.4. Dificultades del uso de postgisQueryBuilder	38
3.2 Conclusiones parciales	41
Conclusiones.....	42
Recomendaciones.....	43
Referencias Bibliográficas	44

Índice de Figuras

Figura 1: Datos vectoriales (izquierda) y datos ráster (Derecha). Fuente: (Dempsey, 2022).....	9
Figura 2: Principales áreas productivas de la UDP "El Guayabal". Fuente: elaboración propia.....	23
Figura 3: Conexión de QGIS con PostgreSQL. Fuente: elaboración propia.	24
Figura 4: Constructor de consultas por defecto de QGIS. Fuente: elaboración propia.	26
Figura 5: Complemento para realizar análisis espaciales en PostGIS. Fuente: elaboración propia.....	27
Figura 6: Flujo de interacción del usuario con LeNatSQL para la obtención de la consulta SQL a partir de una en lenguaje natural. Fuente: elaboración propia.	28
Figura 7: Procesamiento de la búsqueda de la pregunta. Fuente: elaboración propia.	29
Figura 8: Consulta para convertir los elementos geográficos al sistema de referencia SRID 4326. Fuente: elaboración propia.	32
Figura 9: Interfaz principal de LeNatSQL con la pregunta insertada “¿Cuáles fueron las áreas de producción agrícola disponibles en el mes de febrero cuyo suelo es Ferralítico Rojo Típico?”. Fuente: elaboración propia.	34
Figura 10: Respuesta de LeNatSQL para la pregunta “¿Cuáles fueron las áreas de producción agrícola disponibles en el mes de febrero cuyo suelo es Ferralítico Rojo Típico?”. Fuente: elaboración propia.	35
Figura 11: Respuesta de LeNatSQL cuando la pregunta insertada no coincide con las almacenadas en el dataset. Fuente: elaboración propia.	35
Figura 12: Ayuda de LeNatSQL. Fuente: elaboración propia.....	36
Figura 13: Consulta SQL obtenida, insertada en postgisQueryBuilder para la pregunta “¿Cuáles fueron las áreas de producción agrícola disponibles en el mes de febrero cuyo suelo es Ferralítico Rojo Típico?”. Fuente: elaboración propia. ..	37

Figura 14: Capa resultante visualizada en QGIS para la pregunta “¿Cuáles fueron las áreas de producción agrícola disponibles en el mes de febrero cuyo suelo es Ferralítico Rojo Típico?”. Fuente: elaboración propia.....	38
Figura 15: Consulta para la pregunta “¿Cuáles son las áreas de producción agrícola que cuentan con riego?”. Fuente: elaboración propia.....	39
Figura 16: Consulta para la pregunta “¿Qué cultivos se encuentran bajo riego?”. Fuente: elaboración propia.....	40

Introducción

La informática juega un papel fundamental en el mundo actual. Principalmente su mayor aporte está dado en facilitar la comunicación entre las personas desde casi cualquier lugar del planeta. Esto es posible mediante sistemas o programas que realizan complejas operaciones de forma instantánea (Rasulovich, 2020).

Un sistema informático permite almacenar y procesar información. Es el conjunto de partes interrelacionadas: hardware, software y personal informático. Los mismos han transformado la manera en que las organizaciones trabajan apuntando siempre hacia importantes mejoras: mayor productividad y competitividad, automatización y control de los procesos críticos, apoyo al proceso de tomas de decisiones, entre otras (Li et al., 2020) .

Actualmente, es difícil concebir un área que no esté vinculada o requiera del uso y apoyo de los sistemas informáticos. La informática se aplica a numerosas y variadas áreas del conocimiento y la actividad humana, por ejemplo: gestión de negocios, monitorización y control de procesos, industria, robótica, comunicaciones, investigación, ingeniería, medicina, agricultura, física, meteorología, diseño computarizado, almacenamiento y consulta de información, por solo mencionar algunos ejemplos (Rainer y Prince, 2021).

En la agricultura se utilizan los sistemas informáticos con el principal objetivo de aumentar la eficiencia de las cosechas. Las nuevas tecnologías sustituyen a los antiguos instrumentos de cultivo en todas las áreas: desde el análisis del campo y del suelo, hasta la monitorización y promoción de un crecimiento efectivo de las plantas en todas sus etapas. Proporcionan recomendaciones precisas sobre el tratamiento del campo, ayudan a manejar la documentación, recopilan mapas de un área determinada, procesan imágenes espaciales y monitorizan posibles riesgos climáticos (Hess, 2022).

El mercado a nivel mundial ofrece una amplia gama de herramientas para los agricultores, que les permite optimizar las gestiones agrícolas y mejorar la precisión al tomar decisiones. La utilización de satélites y aviones teledirigidos

para recoger datos sobre la vegetación, las condiciones del suelo, el clima y el terreno son tecnologías muy valiosas en la agricultura. Las imágenes de drones ayudan a definir con gran precisión la biomasa de los cultivos, la altura de las plantas, la presencia de malas hierbas y la saturación de agua en ciertas zonas del campo. Los datos obtenidos de satélites espaciales ayudan a predecir el rendimiento del campo, así como realizar una vigilancia sobre el terreno casi en tiempo real para diferentes tipos de amenazas (Talawar et al., 2020).

Dado que los campos están basados en la localización, los Sistemas de Información Geográfica (SIG) se convierten en otra herramienta increíblemente útil en términos de tecnologías agrícolas. Los SIG permiten a los agricultores cartografiar los cambios actuales y futuros en las precipitaciones, la temperatura, el rendimiento de los cultivos, la salud de las plantas, entre otros (Wang et al., 2019).

Uno de los problemas fundamentales al trabajar con estos sistemas está dado por la complejidad que representa la realización de determinadas funciones como la búsqueda y recuperación de información espacial condicionada por el nivel de conocimiento y experiencia que presentan los usuarios (Pinto Molina, 2018). Esta información mayormente se almacena en bases de datos relacionales, de ahí que sea necesario que el usuario también domine conocimientos sobre los lenguajes de consulta estructurados y las relaciones espaciales existentes (Valverde et al., 2019).

En muchas ocasiones los usuarios carecen de la experiencia técnica necesaria lo que trae consigo que a lo largo del tiempo los programadores y diseñadores hayan desarrollado mecanismos informáticos para facilitar la recuperación de información espacial. Permitiendo así que los usuarios no necesiten comprender los esquemas de la base de datos, ni tampoco la sintaxis del lenguaje de consulta (Bareth y Waldhoff, 2018).

A nivel mundial existen diferentes mecanismos para facilitar la búsqueda y recuperación de información espacial a usuarios inexpertos en estos campos.

Estos sistemas transforman preguntas realizadas en lenguaje natural en consultas para bases de datos. Están basados principalmente en, machine learning, redes neuronales, deep learning y procesamiento del lenguaje natural (Anisha et al., 2019). Ejemplos de estos sistemas son Text2SQL (Bonde, 2022), SQLNet (Anisha et al., 2019), NLSQL (Khamis y Shatnawi, 2010), Seq2SQL (Anisha et al., 2019), entre otros.

Cuba no queda exenta de la informatización, ni de los avances tecnológicos y un ejemplo de esto es el desarrollo de sistemas con características similares en la universidad central “Marta Abreu” de las Villas. Ejemplos de estos sistemas son Traductor LN a SQL (Bonilla Hernández, 2011) y LNE2SQL (Reyes García, 2012).

En la Unidad docente Productiva (UDP) “El Guayabal” se proyecta desplegar el SIG QGIS. Dicho sistema va a ser utilizado por diferentes usuarios, que pueden tener o no conocimiento y experiencia en la búsqueda y recuperación de información espacial. Esta información es almacenada en una base de datos relacional, de ahí que sea necesario que el usuario también domine conocimientos sobre los lenguajes de consulta estructurados y las relaciones espaciales existentes. En muchas ocasiones los usuarios carecen de la experiencia técnica necesaria en la recuperación de información en bases de datos espaciales lo que trae consigo que esto se convierta en una dificultad para los mismos.

Lo anterior descrito revela el siguiente **problema** ¿Cómo facilitar la realización de consultas en el SIG que se proyecta desplegar en la UDP “El Guayabal”?

Se ha identificado como **objeto de estudio** las consultas especializadas para Sistemas de Información Geográfica.

Tras la realización de la presente investigación se tiene como **hipótesis** que el desarrollo de un nuevo método para el procesamiento de consultas geoespaciales facilitará la realización de consultas de los usuarios en el SIG que se proyecta desplegar en la UDP “El Guayabal”.

La presente investigación se sustenta en la **línea de investigación** de Agrofísica y Geoinformática en la Agricultura.

Para resolver el problema se traza como **objetivo general** desarrollar un método que facilite la realización de consultas en el SIG que se proyecta desplegar en la UDP “El Guayabal”.

Para darle cumplimiento al objetivo general se proponen los siguientes **objetivos específicos**:

- Establecer el marco teórico – conceptual sobre los SIG y los Sistemas de Gestión de Bases de Datos.
- Seleccionar el mecanismo para la realización de consultas espaciales en un SIG.
- Desarrollar un método para transformar preguntas proporcionadas por los usuarios en lenguaje natural, en consultas al SIG.
- Implementar una herramienta informática para transformar las preguntas en lenguaje natural confeccionadas por los usuarios en consultas a SIG.

Capítulo I: Revisión bibliográfica

1.1. Introducción del capítulo

En este capítulo se analizan los temas más relevantes afines con los Sistemas de Información Geográfica y sus herramientas más populares a nivel mundial. También se exponen aspectos importantes sobre los Sistemas de Gestión de Bases de Datos y se realiza una comparación entre los Relacionales y los NoSQL.

1.2. Sistemas de Información Geográfica: conceptos y principales herramientas

Los mapas y la tecnología cartográfica están prácticamente en todas partes. Aunque los modos y medios de hacer y distribuir mapas se han revolucionado con los recientes avances informáticos y tecnológicos (Bartel, 2022). El arte y la ciencia de hacer mapas se remontan a siglos atrás. Esto se debe a que los humanos son organismos inherentemente espaciales, y para que puedan vivir en el mundo, primero deben relacionarse con él de alguna manera (Restrepo, 2015).

Actualmente existen herramientas y programas informáticos aplicados a la geografía que permiten visualizar datos geográficos y alfanuméricos de manera integrada. Asimismo, permiten administrar la información en forma de capas de diferentes formatos y elaborar análisis espaciales con fines específicos (Mariani, 2012).

Un ejemplo de esto son los Sistemas de Información Geográfica (SIG), según (Goodchild y Kemp, 1990) son sistemas de información compuestos por hardware, software y procedimientos para capturar, manejar, manipular, analizar, modelar y representar datos georreferenciados, con el objetivo de resolver problemas de gestión y planificación.

Los SIG son herramientas eficaces para recopilar, almacenar y gestionar datos espaciales mediante mapas. Los métodos y herramientas SIG se utilizan ampliamente en aplicaciones de ingeniería, especialmente en el diseño, la planificación y la gestión de sistemas de transporte/logística, el diseño de instalaciones, la especificación y asignación de ubicaciones de recursos

(específicamente para recursos agrícolas) y la determinación de regiones potenciales con alta demanda. Otras áreas que dependen de la recopilación, el análisis y la visualización de datos espaciales son las telecomunicaciones, la agricultura, la minería, entre otras (Balaman, 2019).

Un SIG permite gestionar una gran cantidad de datos e información. Desde la tarea relativamente simple de mapear la ruta que debe seguir un tractor para transportar los alimentos recién recogidos del campo hasta la tarea más compleja de determinar el proceso de siembra más eficiente, un SIG se usa en los sectores público y privado. Los servicios basados en la ubicación, la navegación y los mapas en línea y móviles también están personalizando y democratizando los SIG al llevar los mapas y la cartografía a las masas (Dastrup, 2020).

1.2.1. Selección entre QGIS y ArcGIS

En la actualidad existen diferentes aplicaciones SIG tanto comerciales como de código abierto, utilizadas por una gran cantidad de usuarios. ArcGIS y QGIS son dos de estas aplicaciones mayormente usadas cuando se habla de SIG (Maurya et al., 2015).-

Según Khan y Mohiuddin (2018) QGIS es una herramienta SIG totalmente competente y de fácil acceso, es un programa de código abierto, gratuito a diferencia de ArcGIS. QGIS puede ser introducido en diferentes marcos de trabajo, por ejemplo, Windows, Mac OS X, Linux (Ubuntu), y Unix mientras que ArcGIS solo es soportado por Windows.

En su trabajo Maurya et al. (2015) afirma que existen muchos software SIG de código abierto disponibles a nivel mundial, pero QGIS es el más popular entre ellos. Asimismo, plantea que ArcGIS es el software propietario más notorio en el campo de los SIG, por lo que en varias ocasiones son comparados ambos SIG. ArcGIS tiene una excelente interfaz gráfica de usuario en comparación con cualquier software de código abierto. Sin embargo, este solo es soportado por sistemas operativos Windows, mientras que QGIS puede ser introducido en disímiles marcos de trabajo como Linux. ArcGIS cuenta con una aceptada

documentación, mientras que QGIS presenta una vasta y bien descrita documentación, con numerosos tutoriales y videos introductorios muy útiles y de fácil comprensión.

Según Flenniken et al. (2020) QGIS es una herramienta de creciente utilidad para la visualización y el análisis espacial, lo que la convierte en una alternativa viable a otros costosos paquetes de software. El mismo es un software de código abierto que se beneficia de las contribuciones de expertos y usuarios de todo el mundo. La naturaleza de código abierto de QGIS, junto con la creciente disponibilidad de datos gratuitos de fuentes en línea, ofrece una oportunidad para que los usuarios poco experimentados aprendan e incorporen la tecnología SIG en sus investigaciones.

Benduch (2017) sugiere que QGIS podría ser una alternativa aceptable del software propietario ArcGis, en el proceso de enseñanza de los SIG, teniendo en cuenta la posibilidad financiera de las instituciones educativas y también la complejidad de sus funcionalidades.

Según Pérez Betancourt et al. (2016) QGIS representa una herramienta de gran impacto por su aplicación en ramas como la agricultura, la meteorología, el turismo y la salud pública por solo mencionar algunas. El uso de QGIS en la rama de la salud tiene cada día mayor utilidad, su empleo contribuye al fortalecimiento de la capacidad de análisis en materia de salud pública y epidemiología, brindando información de utilidad para la toma de decisiones. También facilita la identificación de la información geográfica de establecimientos de salud y grupos de población que presentan mayor riesgo de enfrentar o de morir prematuramente y que por tanto requieren de mayor atención preventiva, curativa o de promoción de la salud. Además, se utiliza en el análisis de la distribución espacial de enfermedades.

Perez-Garcia et al. (2019) afirman que el uso de QGIS tiene un gran impacto en la agricultura posibilitando la creación de mapas digitales formados por diversas capas, como son: el tipo de suelo, la distribución de nutrientes, la topografía del

terreno, la humedad del suelo y la cobertura vegetal. Siendo todas estas la base de información para el empleo de nuevas técnicas de gestión de parcelas basadas en la variación intraparcilaria de los cultivos. Además, proporciona a los usuarios nuevas bases para fomentar la toma de decisiones administrativas y la planificación de tareas.

Después de consultar varias bibliografías y analizar las opiniones de diferentes autores (Khan y Mohiuddin, 2018; Maurya et al., 2015; Flenniken et al., 2020; Benduch, 2017) se puede afirmar que tanto QGIS como ArcGIS son software SIG de alta calidad, ambos son sistemas estables y buenas propuestas al elegir un SIG. Pero, aunque ArcGIS destaca en varios aspectos y tiene el respaldo de una gran empresa detrás, su alto coste constituye una barrera para muchos usuarios. La gran ventaja de QGIS está en ser una propuesta económica ya que no se necesita de una licencia para poder utilizar todas las herramientas que este ofrece. Además, el mismo es un software libre de código abierto, por lo que está al alcance de cualquier usuario, con las mismas funcionalidades y prestaciones de un SIG de pago.

1.2.2. Tipos de datos en un SIG: datos espaciales y de atributos

QGIS al igual que otros SIG admite la visualización y el análisis de datos geográficos (Khan y Mohiuddin, 2018). Además, ofrece múltiples facilidades, como la captura, gestión y consulta de datos. Permite a los usuarios visualizar y comprender las relaciones entre los datos geográficos en forma de mapas, informes y gráficos. Los tipos de datos básicos disponibles definen los datos habituales en un mapa. Existen principalmente dos tipos de datos: los de atributos y los espaciales (Lithmee, 2019).

1.2.2.1. Datos espaciales

Los datos espaciales describen información geométrica, como la ubicación de una entidad en el espacio o la posición relativa a otras entidades. Además, los mismos son multidimensionales y abarcan varios datos diferentes, aunque correlacionados (Howari y Ghrefat, 2021).

Los mismos comprenden la información geográfica relativa sobre la tierra y sus características. Un par de coordenadas de latitud y longitud define una ubicación específica en la tierra. Estos son de dos tipos, según la técnica de almacenamiento: datos ráster y datos vectoriales (Janipella et al., 2019):

- Los datos ráster se componen de celdas de cuadrícula identificadas por fila y columna. Toda el área geográfica se divide en grupos de celdas individuales, que representan una imagen. Imágenes de satélite, fotografías e imágenes escaneadas, son ejemplos de datos ráster (Figura 1, derecha).
- Los datos vectoriales se componen de puntos, polilíneas y polígonos. Los pozos, las casas, entre otros, se representan mediante puntos. Las carreteras, ríos y arroyos, se representan mediante polilíneas. Los pueblos y ciudades están representados por polígonos (Figura 1, izquierda).

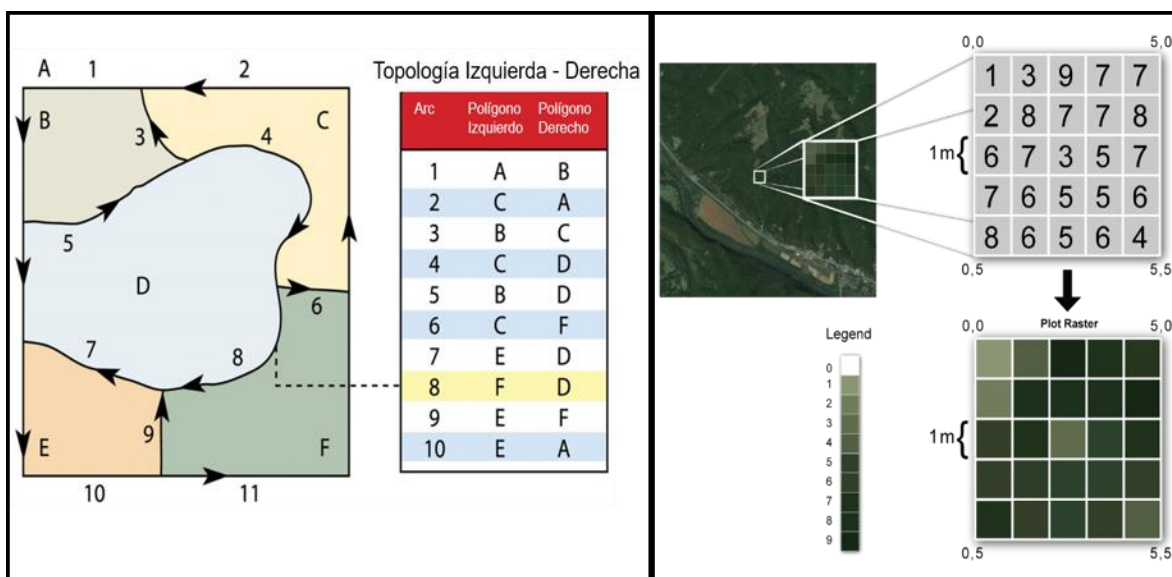


Figura 1: Datos vectoriales (izquierda) y datos ráster (Derecha). Fuente: (Dempsey, 2022).

1.2.2.2. Datos de atributos

Los datos de atributos son descripciones o medidas de características geográficas en un mapa. Se refiere a datos detallados que se combinan con datos geoespaciales. Cada característica tiene propiedades que se pueden describir. Por ejemplo, un edificio: este tiene un año de construcción, el número de pisos,

entre otras características o atributos. También pueden ser hechos que se conocen pero que no son visibles, como el año de construcción. Además, pueden representar la falta de una característica (Lithmee, 2019).

Estos datos comprenden la información pertinente sobre los datos espaciales. La función de consulta trabaja en función de los datos de atributos, es decir, se adjunta a los datos geoespaciales. Los tipos de datos de atributos son (Janipella et al., 2019):

- Datos nominales: describen diferentes tipos de categorías de datos, como los tipos de uso de la tierra o los tipos de suelo.
- Datos ordinales: diferencian los datos por su relación de clasificación. Por ejemplo, las ciudades pueden agruparse en grandes, medianas y pequeñas según el tamaño de la población.
- Datos de intervalo: tienen intervalos conocidos entre los valores, como la lectura de la temperatura. Por ejemplo, una lectura de temperatura de 30 °C es más caliente que 20 °C por 10 °C.
- Datos de proporción: son los mismos que los datos de intervalo, excepto que los datos de proporción se basan en un valor significativo o cero absolutos. Las densidades de población son un ejemplo de datos de relación, porque una densidad de 0 es un cero absoluto.

Para el almacenamiento y organización de los datos geográficos se utilizan bases de datos. Los SIG utilizan entre otras fuentes de datos las bases de datos espaciales las cuales permiten a los usuarios tener acceso a información actualizada sin necesidad de tener los ficheros espaciales físicamente. Si el usuario tiene acceso a la base de datos espacial tendrá acceso a la información espacial almacenada en esta. Las mismas permiten minimizar la posibilidad de errores en el manejo de la información ya que facilitan la eliminación de redundancias e inconsistencias. Además, ayudan a estructurar y organizar los datos de una manera eficiente y funcional (Nur et al., 2018).

1.3. Bases de datos

Una base de datos es un listado o una agrupación de datos, normalmente reunidos de forma que se pueda acceder a ellos rápidamente. Sin embargo, especialmente en el mundo actual, una base de datos puede ser tan simple como una libreta de direcciones que contenga una lista de nombres y direcciones hasta algo tan sofisticado como un conjunto o banco de datos completo de miles de registros de clientes (Maliakal, 2019).

Las bases de datos son colecciones de datos relacionados organizados de manera que se pueda acceder, administrar y actualizar fácilmente los datos. Las mismas pueden estar basadas en software o hardware, con un único propósito, almacenar datos (Cai y Vasilakos, 2017).

1.3.1. Sistema de Gestión de Bases de Datos

Un sistema de gestión de bases de datos (SGBD), es un software para ayudar a mantener y utilizar grandes colecciones de datos, y la necesidad de estos sistemas, así como su uso, está creciendo rápidamente. Los SGBD se diseñaron para facilitar el almacenamiento y la recuperación de grandes colecciones de datos (Rawat y Purnama, 2021).

Los mismos ofrecen varias funciones además de la gestión de archivos; permiten la concurrencia, controlan la seguridad, mantienen la integridad de los datos, proporcionan una copia de seguridad y recuperación, controlan la redundancia, permiten la independencia de los datos y proporcionan un lenguaje de consulta no procedimental¹. Incluyen funciones de protección y seguridad de estos, mantenimiento de la coherencia de los datos almacenados y la disponibilidad de los mismos para múltiples usuarios al mismo tiempo (Singh, 2015).

¹ Existen lenguajes procedimentales y no procedimentales. En el primer caso el usuario da instrucciones al sistema para que realice una serie de procedimientos u operaciones en la base de datos para calcular un resultado final. Mientras que en el segundo caso el usuario describe la información deseada sin un procedimiento específico para obtener esa información.

1.3.1.1. Selección entre Sistemas de Gestión de Base de Datos Relacionales y NoSQL

Existen diversos tipos de SGBD: relacional, distribuido, NoSQL, orientado a objetos y orientado a gráficas, esto se debe a la variedad de la forma de trabajo que se requiere de ellos. Entre las más usadas para el trabajo con los QGIS se encuentran los Sistemas de Gestión de Base de Datos Relacionales (SGBDR) como PostgreSQL y los sistemas de gestión de bases de datos NoSQL como MongoDB (Nayak et al., 2013).

Según Guillot Jiménez y García (2016) los sistemas de gestión de bases de datos NoSQL constituyen una excelente opción para aplicaciones que requieran de la gestión eficiente de datos semiestructurados y no estructurados y de una alta disponibilidad por parte de los usuarios finales. Para el manejo de datos geoespaciales destacan los sistemas orientados a documentos y los orientados a grafos sobre aquellos que poseen un modelo de datos del tipo llave/valor y los orientados a columnas. Aunque el dominio más amplio en lo referido a la gestión de datos geoespaciales continúa estando en manos de los sistemas relacionales.

En su trabajo Guo y Onstein (2020) plantean que las bases de datos NoSQL como MongoDB son muy útiles para el trabajo de datos geoespaciales, ya que se cargan rápidamente y tienen un buen tiempo de ejecución, asimismo presentan un buen rendimiento de consulta y abundantes funciones geoespaciales y métodos de indexación. Dependiendo del escenario de la aplicación, las bases de datos gráficas, las bases de datos de valor clave y las bases de datos de columnas anchas también tienen sus propias ventajas. Además, el cálculo de superficies geométricas y el procesamiento de volúmenes no se manejan en las bases de datos NoSQL existentes.

Según Nasution (2021) un SGBDR es un software utilizado para almacenar, gestionar, consultar y recuperar los datos almacenados en una base de datos relacional. Este proporciona una interfaz entre los usuarios y las aplicaciones y la base de datos, así como funciones administrativas para gestionar el almacenamiento, el acceso y el rendimiento de los datos.

Twumasi (2002) plantea que uno de los modelos de base de datos bien establecidos a nivel lógico es el modelo relacional, que organiza los datos en tablas. Estas tablas son manipuladas por un SGBDR. Los SIG han utilizado tradicionalmente bases de datos relacionales para el almacenamiento de sus datos. A pesar de su limitada capacidad para captar el comportamiento de los objetos, la tecnología relacional está matemáticamente bien establecida en el campo de las bases de datos.

Según Yue y Tan (2017) los SGBDR que se basan en modelos relacionales de datos geográficos, han dominado el mercado empresarial durante mucho tiempo y desempeñan un papel importante en la moderna sociedad de la información. Los modelos relacionales se basan en conjuntos de álgebra con rigurosos fundamentos matemáticos y capacidades de análisis formal.

Hammersley (2018) opina que los SGBDR han estado a la vanguardia en el campo de los SIG como QGIS durante muchos años. En su investigación plantea que luego de una prueba de rendimiento de SGBD escalables comparando dos SGBDR con un SGBD de documentos y un SGBD de gráficos, los resultados muestran que los SGBDR son en general superiores, en cuanto a sentencias de unión, extracción de estadísticas significativas, actualizaciones parciales de registros e integridad referencial.

Luego del análisis y consulta de diferentes bibliografías (Guillot Jiménez y García, 2016; Guo y Onstein, 2020; Hammersley, 2018; Twumasi, 2002; Yue y Tan, 2017) se puede afirmar que tanto los SGBDR como los NoSQL son bases de datos recomendadas para utilizar a la hora de comenzar en proyectos SIG, cada una de ellas con sus ventajas y desventajas. Sin embargo, los SGBDR brindan un mayor soporte y más variedad de herramientas debido a que llevan más tiempo en el mercado. Los sistemas relacionales son considerados los más populares entre las organizaciones de todo el mundo ya que proporcionan un método confiable para almacenar y recuperar grandes cantidades de datos al tiempo que ofrece una combinación de rendimiento del sistema y facilidad de implementación. Además, son útiles para manejar y obtener datos geográficos.

1.3.1.2. PostgreSQL

Existen SGBDR con licencia privativa y comercial, entre estos se encuentran: Microsoft SQL Server, Oracle y DB2 siendo su coste significativo en algunos casos solo alcanzable por grandes organizaciones. Sin embargo, también se pueden encontrar software libres, de código abierto tan potentes y eficaces como los de pago (Garrido et al., 2021).

Según (Ismail Hossain et al., 2019; Wodyk y Skublewska-Paszkowska, 2020) afirman que entre los SGBDR de código abierto más populares del mercado a nivel mundial se encuentran SQLite, MySQL y PostgreSQL. Cada uno tiene sus propias características y limitaciones únicas, y sobresalen en escenarios particulares. SQLite destaca en aplicaciones pequeñas mientras que PostgreSQL resalta gracias a características como el procesamiento paralelo y la concurrencia. Además, garantiza la confiabilidad y la integridad de los datos y es ideal para proyectos científicos y de investigación.

PostgreSQL cuenta con un amplio conjunto de tipos de datos, permitiendo su extensión mediante tipos y operadores definidos y programados por el usuario. Su administración se basa en usuarios y privilegios y es altamente confiable en cuanto a estabilidad se refiere. Los mensajes de error pueden estar en español y hacer ordenaciones correctas con palabras acentuadas o con la letra 'ñ'. Puede extenderse con librerías externas para soportar encriptación, búsquedas por similitud fonética, entre otros. Posibilita el control de concurrencia multi-versión, lo que mejora sensiblemente las operaciones de bloqueo y transacciones en sistemas multi-usuario y es posible definir un nuevo tipo de tabla a partir de otra previamente definida (Ginestà y Mora, 2012).

1.3.2. Bases de datos espaciales

PostgreSQL al igual que otros SGBDR por sí solo no permite el almacenamiento de la ubicación física y la forma de los objetos geométricos dentro de las tablas, para esto se necesita una base de datos espacial (Hess, 2022).

Las bases de datos espaciales permiten la representación de objetos geométricos simples como puntos, líneas y polígonos. Algunas manejan estructuras más complejas como objetos 3D, coberturas topológicas, redes lineales y redes irregulares trianguladas. Dichas bases de datos requieren una funcionalidad adicional para procesar los tipos de datos espaciales de manera eficiente, y los desarrolladores a menudo han agregado tipos de datos geométricos o de características (Salleh et al., 2021).

Las bases de datos espaciales son extensiones de las bases de datos clásicas que contienen objetos, procedimientos y mecanismos especiales para almacenar y manipular datos geoespaciales (Maina et al., 2019).

1.3.2.1. PostGIS extensión de PostgreSQL

PostGIS es la extensión para el SGBDR PostgreSQL, este almacena el objeto espacial en el formato estándar Well Known Binary (WKB, por sus siglas en inglés). El mismo hereda automáticamente las características de las bases de datos empresariales, así como los estándares abiertos que implementa un SIG dentro del motor de base de datos (Belciu et al., 2014).

Entre los principales datos utilizados por PostGIS se encuentran: punto (point), cadena de líneas (linestring), polígono (polygon), multipunto (multipoint), multilínea (multilinestring) y multipolígono (multipolygon) (Leslie y Ramsey, 2022). Este ofrece la posibilidad de cargar masivamente los datos desde shapefiles². Se publicó bajo la licencia GNU GPL, lo que significa que está disponible como software de código abierto (Hsu y Obe, 2021).

1.3.3. Consultas SQL

PostgreSQL al igual que otros SGBDR les permite a los usuarios la introducción de consultas para recuperar o manipular los datos necesarios para funciones de trabajo específicas. El propósito principal de una consulta es recuperar información, esta función permite a los usuarios acceder a la información o

² Formato de archivos de ESRI que almacena datos geoespaciales.

cambiar los datos de alguna manera, como agregar o eliminar información (Baik et al., 2019).

El proceso de recuperación se lleva a cabo mediante consultas donde se almacena la información estructurada, mediante un lenguaje de interrogación adecuado. Es necesario tener en cuenta los elementos clave que permiten hacer la búsqueda, determinando un mayor grado de pertinencia y precisión, como son: los índices, palabras clave y los problemas que se pueden dar en el proceso (Kul et al., 2018).

Las consultas tienen otras funciones importantes, incluido el filtrado de datos, la presentación clara de resultados, la compilación de información y la adición de criterios (Uma et al., 2019).

1.3.3.1. SQL, lenguaje de consulta estructurado

PostgreSQL utiliza el Lenguaje de Consulta Estructurado (SQL, por sus siglas en inglés) como lenguaje de programación para realizar las consultas. (Shin et al., 2018). Cuando las sentencias que provee SQL son utilizadas de manera eficiente, permite la creación de sistemas altamente escalables, flexibles y manejables (Pino et al., 2018).

SQL implementa cláusulas para estructurar las consultas. Una cláusula es un elemento de lenguaje que incluye las instrucciones de consulta select (seleccionar), from (de), where (donde), order by (ordenar por) y having (teniendo) (Pratt, 2009).

- Select indica qué campos de la tabla de atributos desea ver.
- From denota la tabla de atributos en la que reside la información.
- Where denota los criterios definidos por el usuario para la información de atributo que se debe cumplir para que se incluya en el conjunto de salida.
- Order by denota la secuencia en la que se mostrará el conjunto de salida.
- Having indica el predicado utilizado para filtrar la salida de la cláusula order by.

Si bien las cláusulas select y from son declaraciones obligatorias en una consulta SQL, where es una cláusula opcional que se usa para limitar el conjunto de resultados. order by y having son cláusulas opcionales que se utilizan para presentar la información de manera interpretable (Shin et al., 2018).

Tabla 1: Ejemplos de consultas SQL. Fuente: elaboración propia.

Consultas SQL	Preguntas en lenguaje natural
select * from cultivos order by cultivo.area	Muestre los cultivos cuya área de sembrado sea mayor
select * from cultivos where cultivo.produccion > 20	Cultivo cuya producción sea mayor que 20 t
select * from cultivos where cultivos.nombre like 'arroz' order by cultivos.produccion descending	Mostrar las producciones de arroz de mayor a menor

Las consultas SQL se corresponden a preguntas confeccionadas por los usuarios donde la estructura de las sentencias SQL es similar a la del lenguaje natural (Taipalus et al., 2018).

Según Shin et al. (2018) las consultas SQL no son más que preguntas en lenguaje natural sobre el mundo y cómo las personas se relacionan con él. Estas preguntas pueden ser simples con un enfoque local o más complejas con una perspectiva más global.

1.4. Consultas en SIG

La presencia de un SGBDR como PostgreSQL dentro del SIG QGIS permite la realización de consultas a este, la misma es una llamada a dicho sistema gestor, el cual devuelve como respuesta una serie de elementos tomados de la información contenida en la base de datos. Es decir, del total de datos se obtiene como consecuencia de la consulta una parte de los mismos. La respuesta a la consulta es un conjunto de elementos, de la misma forma que si en un mapa impreso se pregunta ¿qué hay aquí? y se obtiene como respuesta los datos

correspondientes al punto que se señaló. Estos datos son una fracción particular del conjunto de todos los contenidos en dicho mapa (Olaya, 2014).

Uno de los análisis que se puede realizar sobre una capa de información geográfica dentro de QGIS es la consulta de esta. Se entiende por consulta a una operación en la cual se pregunta a los datos geográficos algún tipo de cuestión simple, generalmente basada en conceptos formales sencillos. Este tipo de análisis, aunque no implica el uso de conceptos analíticos complejos, es uno de los elementos clave de los SIG, pues es parte básica del empleo diario de estos (Pratt, 2009).

1.4.1. Tipos de consultas en SIG

Después de analizar varias bibliografías (Olaya, 2014; Shin et al., 2018; Pratt, 2009) se puede apreciar que cada autor tiene su forma particular para clasificar los tipos de consultas en un SIG. Independientemente de ello los autores hacen referencia a los mismos tipos de consultas, pudiéndose clasificar las mismas en tres tipos fundamentales: selección, consultas por atributos y consultas espaciales.

1.4.1.1. Selección

La selección representa la forma más fácil de buscar y consultar datos. La misma resalta la información de los datos seleccionados, tanto en pantalla como en la tabla de atributos, para su posterior visualización o análisis. Para lograr esto se seleccionan los datos en el mapa simplemente usando el cursor para “apuntar y hacer clic” en la característica de interés o usando el cursor para arrastrar un cuadro alrededor de esas características (Shin et al., 2018).

Las opciones avanzadas para seleccionar subconjuntos de datos del conjunto de datos más grande incluyen la creación de una nueva selección, la selección de las características seleccionadas actualmente, la adición a la selección actual y la eliminación de la selección actual (Olaya, 2014).

1.4.1.2. Consulta por atributo

La consulta de atributos requiere el tratamiento de los datos de atributos. En otras palabras, se trata de un proceso de selección de información mediante la formulación de preguntas lógicas. Las consultas de atributos adoptan efectivamente la forma de una pregunta. La manera más fácil de conceptualizarlo es que se está construyendo una frase utilizando una sintaxis especial que seleccionará un subconjunto de sus datos como resultado (Shnizai, 2011).

Las características del mapa y sus datos asociados se pueden recuperar a través de la consulta de información de atributos dentro de las tablas de datos. Por ejemplo, las herramientas de búsqueda y consulta le permiten a un usuario mostrar todos los terrenos productivos cuya área sea mayor que 5 ha, mostrar los terrenos sembrados con arroz con mayor producción, entre otros (Shin et al., 2018).

Los atributos pueden ser valores numéricos, cadenas de texto, valores booleanos (es decir, verdadero o falso) o fechas. Este tipo de consulta es similar a una consulta realizada a cualquier base de datos; sin embargo, las respuestas, es decir las características relacionadas con los registros seleccionados por el proceso se destacan en el mapa, así como en la tabla (Pratt, 2009).

1.4.1.3. Consulta espacial

Las consultas espaciales son consultas en una base de datos espacial que pueden responderse únicamente a partir de la información geométrica, es decir, la posición espacial y la extensión de los objetos implicados. Las mismas también conocidas como "consultas por geografía", permiten resaltar atributos particulares al examinar su posición en relación con otras características (Shin et al., 2018).

Las funciones de las consultas espaciales permiten encontrar, mostrar y/o aislar los registros de atributos conectados a las características del mapa situadas dentro de un área de interés definida. La consulta espacial implica la selección de características basadas en la ubicación o en las relaciones espaciales, lo que requiere el procesamiento de la información espacial. Las consultas espaciales

son una herramienta útil con la capacidad de encontrar cómo o dónde se relacionan espacialmente dos o más conjuntos de datos. (Shnizai, 2011).

1.4.1.3.1. Relaciones espaciales

La relación espacial es el componente central de una consulta espacial. Es una relación entre características espaciales con respecto a sus ubicaciones espaciales y arreglos espaciales. Existen tres tipos de relaciones espaciales en SIG: relaciones de proximidad, relaciones topológicas y relaciones de dirección (Dan et al., 2020).

- Relaciones topológicas: no se ven afectadas por la transformación bicontinua, como el estiramiento, el desplazamiento, la rotación o la flexión de las características espaciales involucradas.
- Relaciones de proximidad: se refieren a las relaciones espaciales basadas en distancias entre entidades.
- Relaciones de dirección: está basada en la separación angular de una característica en relación con otra característica en un sistema de coordenadas.

Las consultas pueden hacerse sobre la componente de atributos de los datos, sobre la espacial, o sobre ambas. En cualquier caso, las mismas se encuentran vinculadas, por lo que el resultado de la consulta afecta a ambas. La selección se hace patente sobre ambas componentes, con independencia de cuál de ellas haya sido la encargada de aplicar el criterio de selección. Tanto la tabla de atributos como la representación visual de la componente espacial se ven afectadas por la realización de una consulta.

1.5. Conclusiones parciales

Los SIG son herramientas eficaces para recopilar, almacenar y gestionar datos espaciales. Existen herramientas SIG de código abierto muy populares a nivel mundial por su fácil acceso y competitividad, entre ellas se encuentra QGIS. Esta utiliza bases de datos relacionales para almacenar y gestionar los datos geográficos.

Los SGBDR como PostgreSQL permiten minimizar la posibilidad de errores en el manejo de la información ya que facilitan la eliminación de redundancias e inconsistencias. QGIS junto a PostgreSQL a través de la extensión PostGIS para trabajar con bases de datos espaciales constituye una poderosa herramienta al procesar datos georreferenciados.

Para recuperar o manipular información las personas deben realizarle consultas a este sistema de base de datos. Estas no son más que preguntas en lenguaje natural, confeccionadas por los usuarios que trabajan con QGIS. Estas preguntas se realizan en lenguaje SQL y se clasifican en tres tipos fundamentales: consultas por selección, consultas de atributos y consultas espaciales. Cualquiera de las tres va a arrojar resultados verídicos no importa si se usan por separado o combinándolas.

Capítulo II: Materiales y métodos

2.1 Introducción del capítulo

En este capítulo se presentan los fundamentos teóricos que sostienen la presente investigación. Luego se expone una breve caracterización del área de estudio, donde se presentan las peculiaridades de la información utilizada durante la investigación. Más adelante se explican las tecnologías y herramientas seleccionadas y se plantea cómo se realiza el procesamiento de las consultas en lenguaje natural utilizando la solución propuesta.

2.2 Caracterización del área de estudio

La investigación se realiza en la Unidad Docente Productiva (UDP) “El Guayabal”. La misma está situada a la salida de San José de las Lajas, en la provincia Mayabeque, Cuba. Se encuentra situada específicamente en las coordenadas 22°59'57.49"N 82°10'09.42"W a unos 121m por encima del nivel del mar y unos 5,3 km de la Universidad Agraria de La Habana (UNAH) “Fructuoso Rodríguez Pérez”. Posee una extensión de 665,7 ha, equivalente a 49,6 caballerías. Su área agrícola es de 72,2 ha y sus suelos son predominantemente Ferralíticos Rojos. En este centro se desarrollan procesos de ingeniería, agrícolas, de salud y reproducción animal. Así mismo se trabaja en la ceba de toros, recría, apicultura, acuicultura y en la comercialización de estos. Su área productiva cuenta con cría de ganado mayor y menor como: vacas, cerdos, aves, conejos y peces. Dicho centro se caracteriza por la labor investigativa y científica, así como por la producción agropecuaria que se pretende sea cada vez mayor. Pertenece al Complejo Científico Docente (el cual está compuesto por la UNAH, Centro Nacional de Sanidad Agropecuaria, Instituto Nacional de Ciencias Agrícolas e Instituto de Ciencia Animal) y sobresale por la preparación a profesionales vinculados a la agricultura y a su desarrollo en la región occidental del país.

Para el desarrollo de la presente investigación los directivos de dicho centro facilitaron los siguientes documentos: Reporte Anual de la Producción Agrícola (2020 y 2021), Resumen Mensual de la Producción Agrícola en Proceso (2018,

2019, 2020 y 2021) y los Boletines Económicos (2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019).

Se realizó, además, un recorrido por el centro guiado por sus directivos permitiendo geolocalizar los terrenos de producción agrícola a partir del uso de la aplicación para dispositivos móviles Fields Area Measure Pro³, con la ayuda del GPS⁴ y las imágenes satelitales. Esto permitió identificar un total de 13 áreas de producción marcadamente definidas, como muestra la Figura 2.

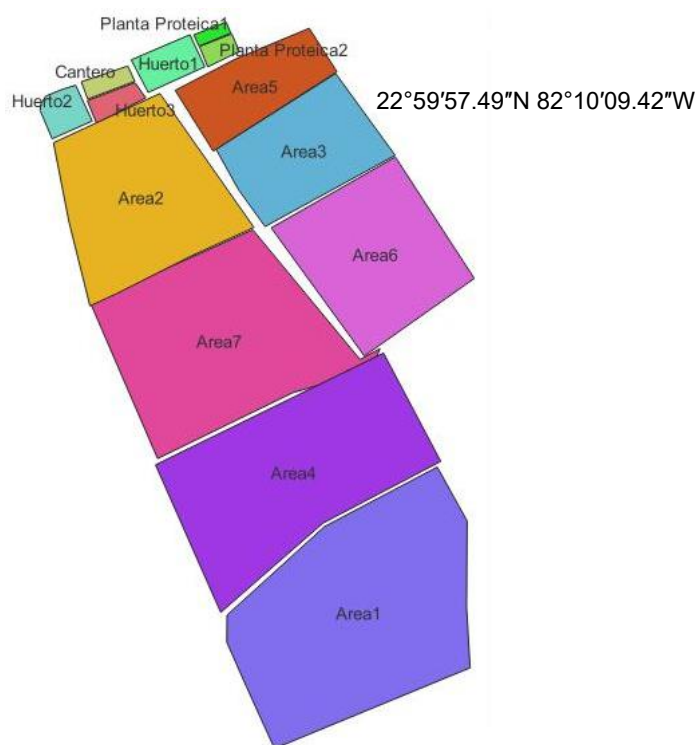


Figura 2: Principales áreas productivas de la UDP "El Guayabal". Fuente: elaboración propia.

Con los datos obtenidos se confeccionaron varios ficheros de información geográfica con los principales productos agrícolas sembrados en el centro a lo

³ Aplicación para dispositivos móviles de libre acceso desarrollada por la empresa Farmis dedicada a la medición de terrenos.

⁴ El Sistema de Posicionamiento Global (GPS, por sus siglas en inglés) es un sistema que permite localizar cualquier objeto sobre la Tierra.

largo de 12 meses, las áreas de siembra, los tipos de suelo presentes y el área que abarca la máquina de riego tipo Fregat.

2.3 Tecnologías y herramientas utilizadas

La información obtenida se introdujo en una base de datos espacial para facilitar el almacenamiento y la obtención de información de los datos. Para gestionar dicha base de datos se utilizó el SGBDR PostgreSQL permitiendo manipular de forma segura los mismos. Se utilizó, además, PostGIS como extensión de PostgreSQL para manejar los datos espaciales. Las investigaciones de (Ginestà y Mora, 2012), (Hsu y Obe, 2021) y (Nasution, 2021) evidencian las posibilidades que ofrece PostgreSQL con su extensión PostGIS en la recuperación y gestión de información espacial.

Esta información geoespacial se visualiza en forma de capas usando la herramienta SIG, QGIS. En sus investigaciones (Khan y Mohiuddin, 2018), (Maurya et al., 2015) y (Flenniken et al., 2020) afirman que QGIS es un potente SIG de código abierto comparable con herramientas SIG propietarias. Es un software económico y de fácil acceso. QGIS cuenta con soporte para realizar conexiones a bases de datos espaciales y así visualizar la información almacenada en esta.

De esta forma los usuarios de QGIS se conectan a la base de datos para acceder a las capas que se almacenan en esta. Teniéndose así un simple pero efectivo entorno de trabajo que permite el almacenamiento, gestión y muestreo de las capas geográficas. La Figura 3 ilustra la distribución de esta estructura. Cada tabla dentro de la base de datos que contenga al menos una columna con referencia a un objeto geográfico será visualizada en QGIS como una capa.

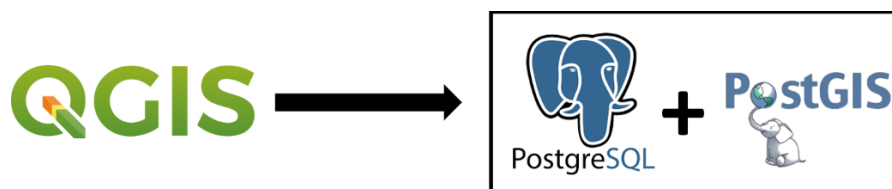


Figura 3: Conexión de QGIS con PostgreSQL. Fuente: elaboración propia.

Para obtener y recuperar información de las bases de datos relacionales se realizan consultas utilizando el lenguaje SQL. Las investigaciones realizadas por (Baik et al., 2019) y (Kul et al., 2018) destacan el uso de SQL en la realización de consultas en bases de datos relacionales. Pueden realizarse consultas SQL en PostgreSQL (incluyendo consultas espaciales), sin embargo, el resultado no puede ser visualizado correctamente. Por lo que es necesario la creación de consultas SQL dentro de QGIS para que el resultado de la operación sea visualizado como información espacial.

En su versión más sencilla QGIS admite la realización de consultas por defecto sobre una única capa a la vez utilizando el constructor de consultas de la capa, como ilustra la Figura 4. Este constructor de consultas solamente permite filtrar información de la propia capa y por ello no admite relaciones de ningún tipo con otras capas. Lo cual limita en gran medida las opciones de búsqueda y recuperación de información entre las mismas.

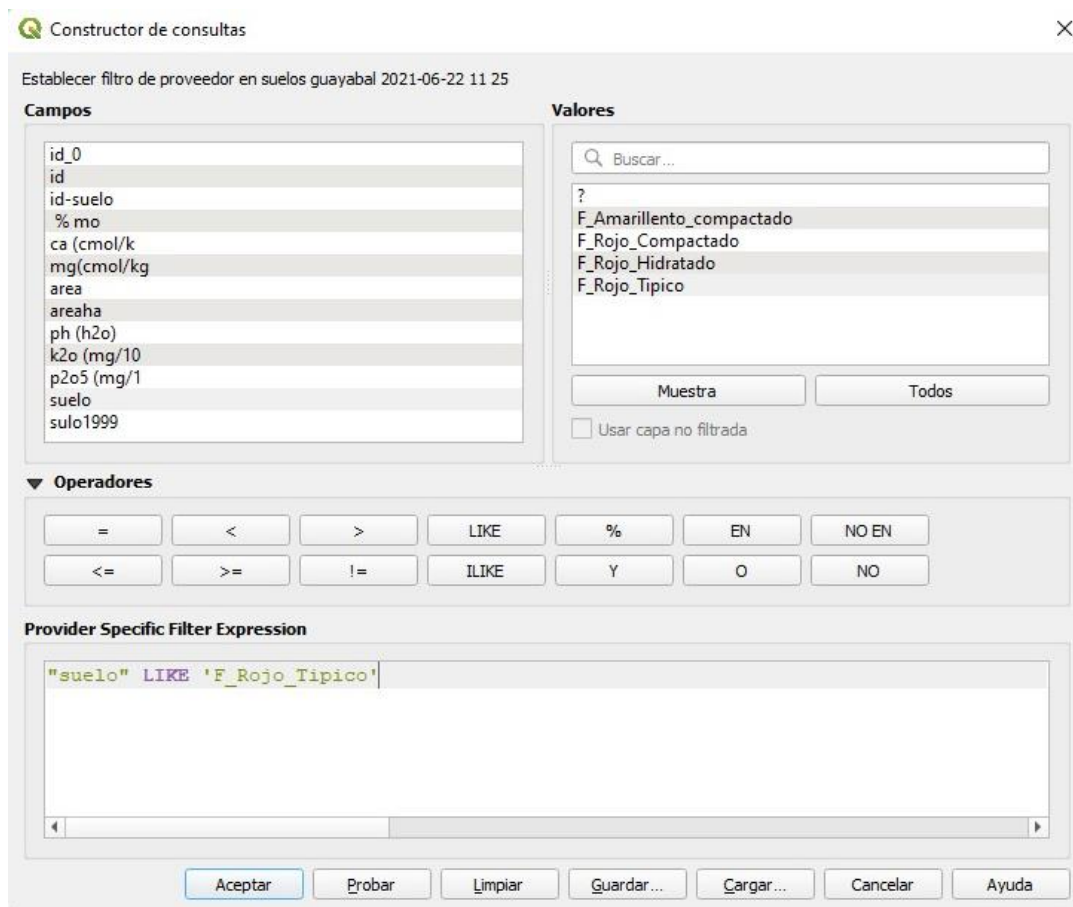


Figura 4: Constructor de consultas por defecto de QGIS. Fuente: elaboración propia.

Una investigación en los foros de la comunidad de desarrolladores y mantenedores de QGIS arrojó la existencia de un complemento instalable en QGIS que permite la realización de consultas SQL especializadas con la componente espacial. El complemento postgisQueryBuilder, como se muestra en la Figura 5 tiene como objetivo proporcionar una herramienta para realizar análisis espaciales en PostgreSQL/PostGIS mediante consultas SQL y a su vez crear nuevas capas espaciales con sus resultados.

postgisQueryBuilder puede utilizarse en entornos educativos para un análisis en vivo de la estructura de las consultas espaciales comunes. El complemento puede administrar vistas PostgreSQL y puede usarse para implementar marcos de análisis espacial basados en cadenas de relaciones espaciales (Ferreguti, 2022).

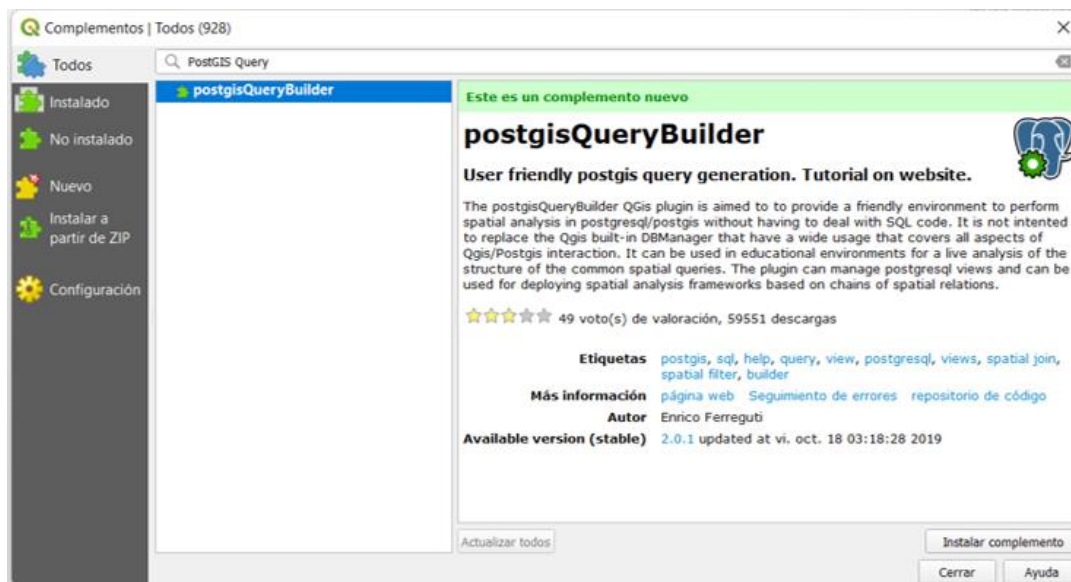


Figura 5: Complemento para realizar análisis espaciales en PostGIS. Fuente: elaboración propia.

La complejidad que trae consigo la realización de consultas en lenguaje SQL para usuarios con poca experiencia en recuperación de información de bases de datos relacionales conllevó al desarrollo de una herramienta que transforme preguntas en lenguaje natural en consultas en lenguaje SQL.

LeNatSQL es un sistema informático de escritorio multiplataforma que tiene como objetivo principal la transformación de preguntas en lenguaje natural a consultas en lenguaje SQL para facilitar la búsqueda y recuperación de información espacial.

Para realizar el sistema propuesto se utilizó JAVA como lenguaje de programación. El mismo es independiente de la plataforma, usando su máquina virtual, el programa puede ejecutarse en múltiples plataformas. Generalmente los programas escritos en Java se ejecutan más rápido que en otros lenguajes de programación como Python. El mismo es muy útil en el desarrollo de aplicaciones de escritorio, es fácil de aprender debido a que tiene un fuerte apoyo de su comunidad (Khoirom et al., 2020).

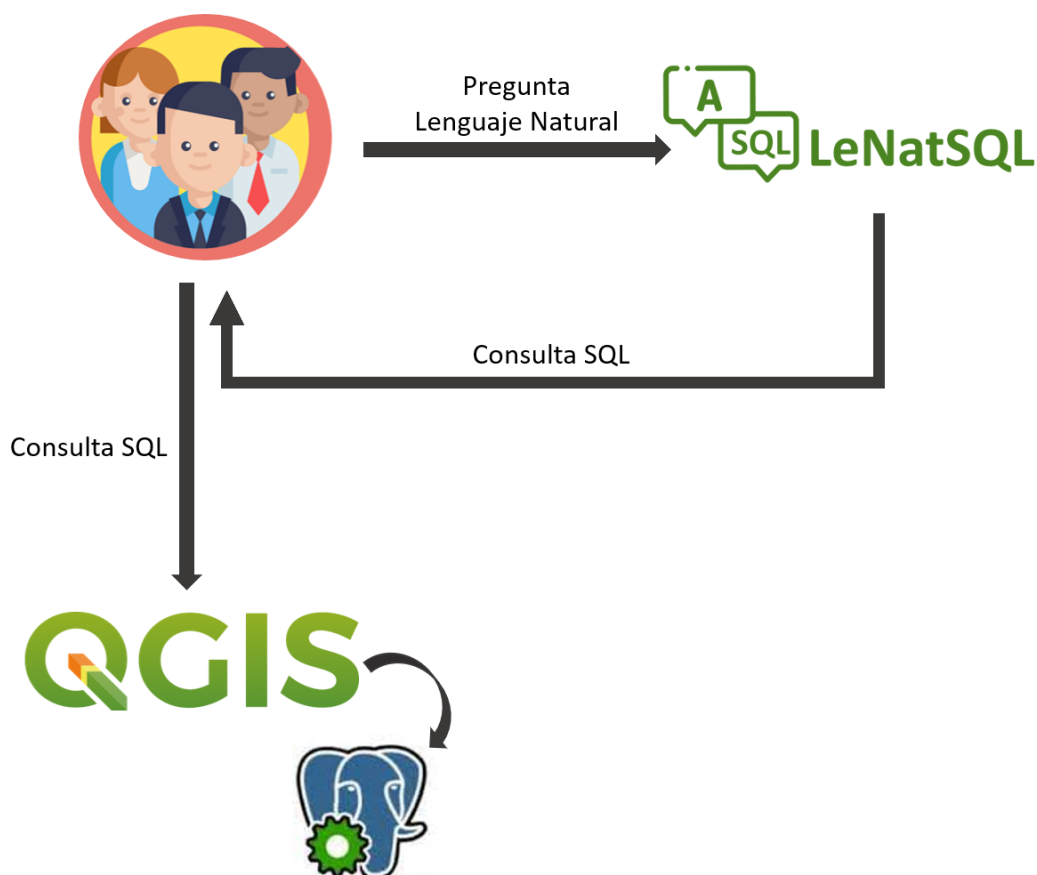


Figura 6: Flujo de interacción del usuario con LeNatSQL para la obtención de la consulta SQL a partir de una en lenguaje natural. Fuente: elaboración propia.

Los usuarios confeccionan una pregunta en lenguaje natural dentro de un campo de la interfaz de LeNatSQL. Este lo procesa y confecciona una consulta SQL que se corresponde con la búsqueda solicitada por el usuario. Esta consulta SQL deberá ser copiada y pegada en el postgisQueryBuilder para que QGIS muestre la capa resultante. La Figura 6 ilustra el proceso antes mencionado.

2.4 Procesamiento de la consulta en lenguaje natural con LeNatSQL

Para el proceso de conversión de las consultas de lenguaje natural a SQL se confeccionó un dataset⁵, el mismo contiene una lista de posibles preguntas que puede introducir el usuario y las respuestas que puede devolver el sistema

⁵ Un dataset es una colección de datos habitualmente tabulada. El conjunto de datos también puede estar conformado de una colección de documentos o de archivos.

respectivamente. El sistema verifica si la estructura de la pregunta realizada es similar a las que se encuentran en el dataset y con ello emite una respuesta.

Para determinar si la pregunta introducida puede ser respondida, el sistema revisa la pregunta introducida contra el dataset y si esta se encuentra busca la respuesta correspondiente. La Figura 7 muestra el proceso de revisión y respuesta.

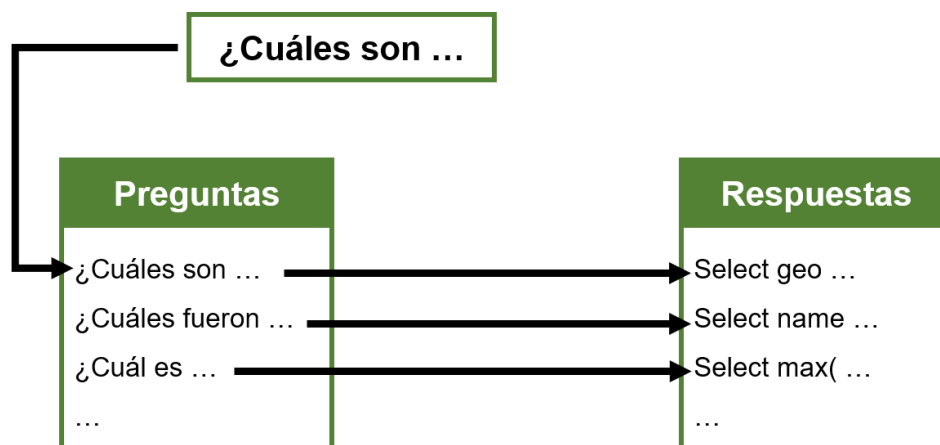


Figura 7: Procesamiento de la búsqueda de la pregunta. Fuente: elaboración propia.

El sistema provee al usuario de una ayuda que indica la estructura de las posibles consultas que pueden ser introducidas, teniendo en cuenta que el sistema no podrá responder a cualquier pregunta indicada por el usuario. De esta forma se garantiza que el usuario pueda introducir preguntas en un formato acertado

Para la confección de las preguntas se tuvo en cuenta las relaciones espaciales admitidas por QGIS. Estas relaciones espaciales permiten determinar el nivel de relación que pueden tener dos capas espaciales de acuerdo a su cercanía o solapamiento. La Tabla 2 muestra un resumen de estas relaciones espaciales.

Tabla 2: Ejemplo de las relaciones espaciales que acepta QGIS. Fuente: elaboración propia.

Ejemplo de uso de la relación espacial	Relación espacial	Función SQL asociada
El área sembrada de yuca está «junto» al área sembrada de boniato.	Junto	ST_Intersects ST_Contains
La calabaza está sembrada «en» el área 1.	En	ST_Within
El área sembrada con calabaza «interseca» el área de yuca.	Interseca	ST_Intersects
El área 3 «cruza» las áreas 6 y 7.	Cruza	ST_Crosses
El área de la máquina de riego «contiene» al área 2.	Contiene	ST_Contains
El área de la máquina de riego «cubre» 5 áreas.	Cubre	ST_Covers
El área de la máquina de riego «no interseca» el área sembrada con yuca.	No interseca	ST_Disjoint
El área que se encuentra sembrada con maíz «solapa» el área sembrada con boniato.	Solapa	ST_Overlaps

El área que está sembrada de yuca «es igual a» las áreas sembrada de boniato y de calabaza.	Es igual a	ST_Equals
La circunferencia de la máquina de riego «toca» las áreas sembradas con frijoles.	Toca	ST_Touches

Para conformar las preguntas de prueba se analizaron las interrogantes establecidas en el dataset GeoQuery y se utilizó su estructura siguiendo una homogeneidad. GeoQuery es un dataset de datos geográficos que contiene preguntas sobre la geografía de los Estados Unidos junto con las consultas SQL correspondientes (Davis y Meltzer, 2007).

Para garantizar que todas las capas se encuentren en el mismo sistema de referenciación geográfica fue necesario verificar y corregir todas aquellas capas que no se encontraban en este. Los sistemas de referenciación geográfica generan un marco espacial para una entidad geográfica, brindando la información y el comportamiento necesario para manipular los diferentes sistemas de coordenadas y unidades involucradas (Llitas, 2000).

Teniendo en cuenta posibles errores en el muestreo de los objetos geográficos por encontrarse en diferentes sistemas de referenciación (SRID⁶), se definió por defecto el uso del SRID 4326, siendo este un estándar a nivel internacional para la cartografía. La diferencia fundamental entre estos sistemas de referenciación consiste en la proyección de sus elementos en el sistema de coordenadas ya que manipulan sistemas de coordenadas diferentes. Debido a ello, dos capas que pertenecen al mismo lugar geográfico si fueron creados en sistemas de referenciación diferentes no serán mostrados correctamente. Para convertir un

⁶ Identificador de Referencia Espacial, es un identificador estándar único que hace referencia a un Sistema de Coordenadas concreto. Cada código, por tanto, se asocia de forma exclusiva a un Sistema de Coordenadas.

objeto geográfico de un sistema a otro puede utilizarse una consulta en lenguaje SQL donde debe indicarse el objeto geográfico, el sistema de referenciación en el que se encuentra y al que se desea convertir.

Algunas capas de información geográficas disponibles para la presente investigación, en este caso las proporcionadas por la facultad de agronomía se encontraban en el sistema de referenciación geográfica SRID 2085 utilizado por Cuba (actualmente en desuso). Por tal motivo, fue necesario realizar una conversión de sus elementos geográficos a la norma internacional. La Figura 8 muestra un ejemplo de la consulta realizada para convertir un objeto geográfico de tipo polígono del sistema SRID 2085 al SRID 4326.

```
SELECT ST_AsText(ST_Transform(ST_GeomFromText('MULTIPOLYGON(((380064.161558007
350992.547340859, 380182.906314644 350994.727614876, 380182.906314644
350994.727614876, 380276.956785687 350996.454476796, 380276.956785687
350996.454476796, 380431.209176398 350999.286706943, 380507.43549809
350683.740818214, 380507.43549809 350683.740818214, 380505.451747011
350683.104724118, 380359.76085776 350634.844748827, 380155.120176428
350590.789915716, 380154.95572773 350592.057989478, 380064.161558007
350992.547340859))))',2085),4326)) As wgs_geom;
```

Figura 8: Consulta para convertir los elementos geográficos al sistema de referencia SRID 4326. Fuente: elaboración propia.

Aunque en QGIS es posible al insertar las capas seleccionar el sistema de georreferenciación más adecuado para tratarla, en la base de datos pueden encontrarse en sistemas diferentes y ello puede traer consigo que las capas no sean ubicadas en la localización correcta. Esta consulta permite estandarizar los objetos geográficos a un mismo sistema de georreferenciación.

2.5 Conclusiones parciales

En el presente capítulo se realizó una descripción del área de estudio donde se realizó la investigación, así mismo se presentan las peculiaridades de la información utilizada. Luego se explicaron los fundamentos de la propuesta de solución y se evidenciaron las principales características del entorno de trabajo utilizado y las funciones que se implementaron en la propuesta de solución.

Capítulo III: Resultados y discusión

3.1 Introducción capítulo

Para la realización de consultas SQL especializadas con componente geográfico a partir de preguntas en lenguaje natural el usuario utiliza LeNatSQL para la obtención de dichas consultas. LeNatSQL utiliza una estructura predefinida para la emisión de sus respuestas por lo que no es posible responder a cualquier pregunta del usuario. El presente capítulo ilustra algunos ejemplos de preguntas que pueden ser realizadas a este software. Se mostrarán las interfaces por las que deberá transitar el usuario hasta llegar a la visualización de los resultados en QGIS. Además, se exponen las ventajas del uso de LeNatSQL y las dificultades que provocan para usuarios inexpertos en lenguaje SQL y consultas espaciales, no usar este sistema.

3.2 Consultas en LeNatSQL

LeNatSQL permite la realización de consultas con diferente nivel de complejidad. Algunas consultas se basan en los parámetros con valores nativos presentes en las tablas como, por ejemplo: “¿Cuáles fueron las áreas que se encontraban sembradas en el mes de febrero?” o “¿Qué cultivos se encontraban sembrados en el mes de junio?” En ambos casos las preguntas se centran en campos nativos, en el primer caso se refiere a “descriptio” (que contiene el identificador del área de producción agrícola) y en el segundo a “name” (que contiene el cultivo sembrado en dicha área).

Por otro lado, existe la posibilidad de realizar consultas donde se evidencian las relaciones espaciales como, por ejemplo: “¿Cuáles son las áreas agrícolas sembradas cuyo suelo sea Ferralítico Rojo Típico?” o “¿Cuáles fueron los cultivos sembrados en el mes de febrero sin riego?”. En estos casos se debe realizar una consulta donde se relacionen los elementos geográficos de ambas capas para así determinar el nivel de solapamiento existente entre las mismas. Para resolver estar interrogantes se utilizaría una relación espacial de tipo «intercepta». Esta verificará qué áreas de siembra son cubiertas por el tipo de suelo Ferralítico Rojo

Típico y qué áreas de cultivos sembradas en el mes de febrero se encontraban cubiertas por la máquina de riego.

La interfaz de LeNatSQL es simple, enfocada en facilitar la interacción con el usuario. Cuenta con un campo de entrada de texto donde el usuario colocará la pregunta y un botón que indicará el inicio del proceso de traducción y su posterior respuesta, como muestra la Figura 9.



Figura 9: Interfaz principal de LeNatSQL con la pregunta insertada “¿Cuáles fueron las áreas de producción agrícola disponibles en el mes de febrero cuyo suelo es Ferralítico Rojo Típico?”. Fuente: elaboración propia.

Una vez ejecutada la tarea de traducción el sistema emitirá dos tipos de respuestas:

- Tipo 1: el sistema ha conseguido traducir la pregunta, como muestra la Figura 10.
- Tipo 2: al sistema no le fue posible encontrar una respuesta válida, como se ilustra en la Figura 11.

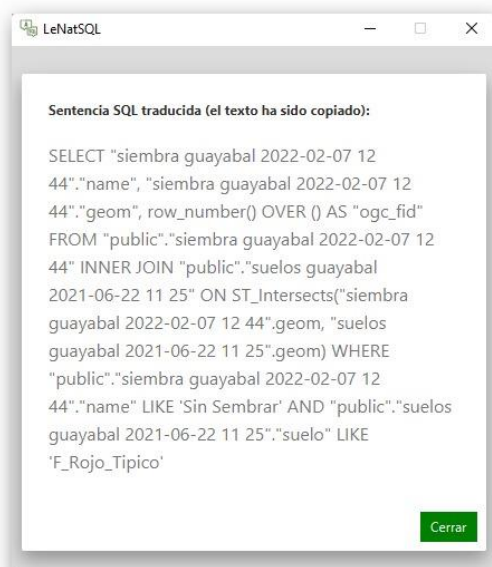


Figura 10: Respuesta de LeNatSQL para la pregunta “¿Cuáles fueron las áreas de producción agrícola disponibles en el mes de febrero cuyo suelo es Ferralítico Rojo Típico?”.

Fuente: elaboración propia.

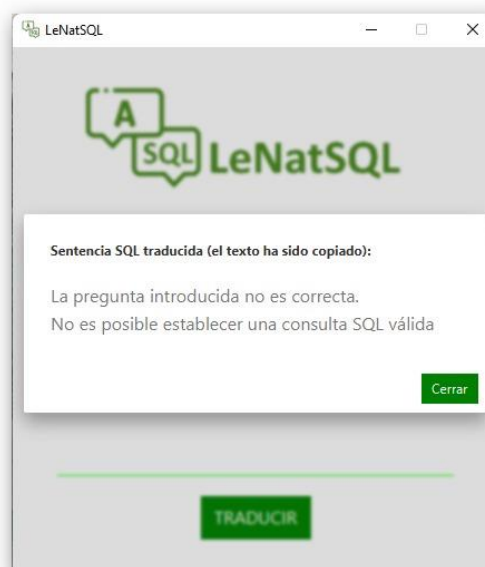


Figura 11: Respuesta de LeNatSQL cuando la pregunta insertada no coincide con las almacenadas en el dataset. Fuente: elaboración propia.

LeNatSQL cuenta con una ayuda que permite a sus usuarios conocer el formato de las consultas válidas disponibles, como muestra la Figura 12. El usuario deberá

utilizar este formato o de lo contrario el sistema devolverá que no es posible realizar la consulta.

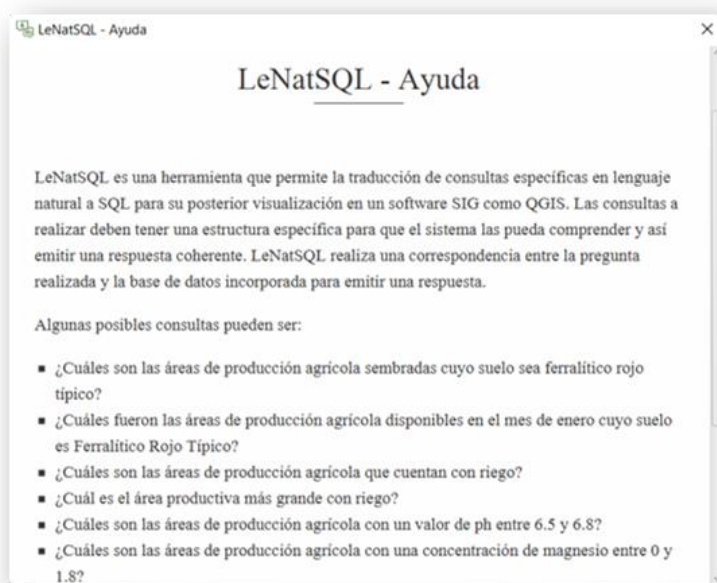


Figura 12: Ayuda de LeNatSQL. Fuente: elaboración propia.

3.3 Visualización de la consulta obtenida en LeNatSQL a través de postgisQueryBuilder y QGIS

Una vez obtenida la respuesta de LeNatSQL esta se colocará en la sección SQL de postgisQueryBuilder como muestra la Figura 13.

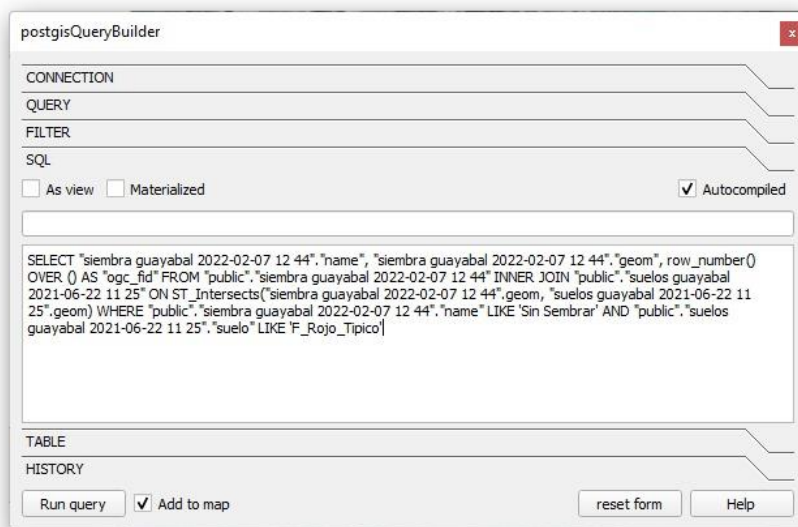


Figura 13: Consulta SQL obtenida, insertada en postgisQueryBuilder para la pregunta “¿Cuáles fueron las áreas de producción agrícola disponibles en el mes de febrero cuyo suelo es Ferralítico Rojo Típico?”. Fuente: elaboración propia.

Al ejecutar la consulta el sistema creará una capa con el resultado, como muestra la Figura 14. La capa obtenida permite al usuario la visualización desde el punto de vista geográfico del resultado a su pregunta.

A la capa resultante al igual que el resto de las capas en QGIS se le pueden aplicar filtros, realizar una categorización por colores o rango de valores y etiquetar los elementos que la componen. De esta forma se facilita la visualización de los elementos que componen la capa.

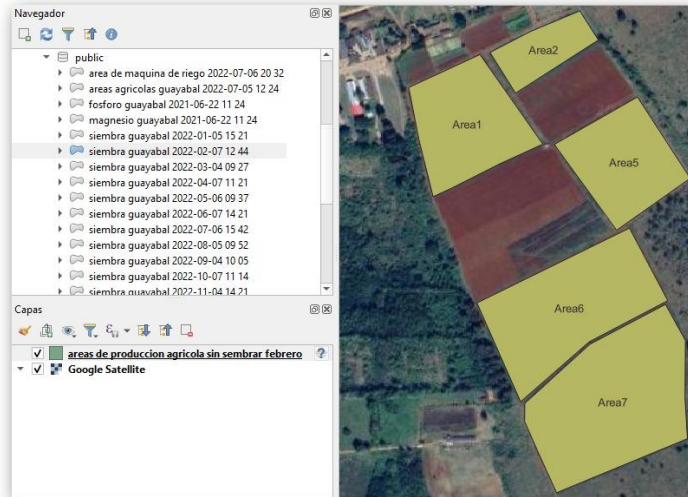


Figura 14: Capa resultante visualizada en QGIS para la pregunta “¿Cuáles fueron las áreas de producción agrícola disponibles en el mes de febrero cuyo suelo es Ferralítico Rojo Típico?”. Fuente: elaboración propia.

El uso de LenatSQL para obtener la consulta SQL está limitado a las preguntas disponibles en el dataset. Para insertar una nueva pregunta y su respectiva respuesta es necesario el trabajo de un experto en lenguaje SQL pues la obtención de esta respuesta parte de sus habilidades para la creación de consultas. Además, es importante destacar que la respuesta a una pregunta puede complejizarse en dependencia de la pregunta realizada.

3.4. Dificultades del uso de postgisQueryBuilder

PostgisQueryBuilder permite la realización de consultas SQL a través de un entorno visual con diversos formularios, con lo cual a medida que el usuario va avanzando se va conformando la consulta. Sin embargo, este se encuentra limitado a la realización de consultas SQL simples, o sea, consultas que no necesitan el resultado de otra consulta. Además de que la interfaz de usuario es compleja para usuarios con poca o ninguna experiencia en el lenguaje SQL y consultas espaciales.

Por ejemplo, para dar respuesta a la pregunta “¿Cuáles son las áreas de producción agrícola que cuentan con riego?” el usuario deberá dar los siguientes pasos en postgisQueryBuilder:

- Seleccionar la base de datos y el esquema al cual se realizará la consulta.
- Dirigirse a la sección Query para seleccionar los elementos de la consulta.
- Seleccionar el tipo de consulta a realizar (en este caso “JOIN Spatial), las tablas en las cuales se realizará la consulta (en este caso “areas agricolas guayabal 2022-07-05 12 24” y “area maquina de riego 2022-07-06 20 32”).
- Seleccionar el tipo de relación espacial que debe existir entre ambas capas (en este caso “ST_Intersects”).
- Seleccionar las columnas de la tabla que se desean obtener (en este caso “name” y “description”).
- Dirigirse a la sección SQL para verificar que la consulta sea correcta y corregir los nombres de los campos en caso de ser necesario.
- Seleccionar Run Query para obtener un resultado.

La Figura 15 ilustra la consulta SQL que da respuesta a la pregunta “¿Cuáles son las áreas de producción agrícola que cuentan con riego?”

```
SELECT "areas agricolas guayabal 2022-07-05 12 24"."name",  
       "areas agricolas guayabal 2022-07-05 12 24"."description",  
       "areas agricolas guayabal 2022-07-05 12 24"."geom",  
       row_number() OVER () AS "ogc_fid"  
FROM "public"."areas agricolas guayabal 2022-07-05 12 24"  
INNER JOIN "public"."area de maquina de riego 2022-07-06 20 32"  
ON ST_Intersects("areas agricolas guayabal 2022-07-05 12 24"."geom", "area de maquina de riego 2022-07-06 20 32"."geom")
```

Figura 15: Consulta para la pregunta “¿Cuáles son las áreas de producción agrícola que cuentan con riego?”. Fuente: elaboración propia.

Si la operación se realizó correctamente postgisQueryBuilder creará una capa en QGIS con los resultados de la operación realizada mostrándose al usuario, en caso contrario podría mostrar un error si los elementos introducidos no son correctos.

Si la consulta contiene elementos que condicionan el resultado (a través del parámetro WHERE) es necesario adicionar condiciones a la consulta a través del

apartado “Filter”, sin embargo, el uso de este apartado para la creación de condiciones se complejiza en gran medida pues postgisQueryBuilder solo está preparado para realizar filtros en la tabla seleccionada como principal.

Otro ejemplo de pregunta podría ser: “¿Qué cultivos se encuentran bajo riego?”.

Esta pregunta se diferencia de la anterior debido a su complejidad, teniendo en cuenta que en este caso se trata de una pregunta compuesta por una pregunta principal y una sub-pregunta. Inicialmente es necesario obtener el tamaño del área de mayor tamaño que se encuentra dentro del área abarcada por la máquina de riego y posteriormente comparar este resultado con el de todas las áreas de producción agrícola para identificarla y obtener sus datos.

La realización de este tipo de consultas en postgisQueryBuilder resulta imposible a través del formulario pues en el proceso de la realización de la segunda consulta para obtener el tamaño máximo de las áreas productivas bajo la máquina de riego es necesario colocarla dentro de la sección de filtros y no es posible realizar otra consulta en esta sección. Por lo que se requiere de la modificación de la consulta de forma manual en la sección SQL.

El resultado de esta consulta sería el siguiente:

```
SELECT "areas agricolas guayabal 2022-07-05 12 24"."name",
       "areas agricolas guayabal 2022-07-05 12 24"."geom",
       row_number() OVER () AS "None"
FROM "public"."areas agricolas guayabal 2022-07-05 12 24"
     INNER JOIN "public"."area de maquina de riego 2022-07-06 20 32"
           ON ST_Intersects("areas agricolas guayabal 2022-07-05 12 24"."geom",
                           "area de maquina de riego 2022-07-06 20 32"."geom")
where ST_Area("areas agricolas guayabal 2022-07-05 12 24"."geom") =
      (Select max(ST_Area("areas agricolas guayabal 2022-07-05 12 24"."geom"))
       FROM "public"."areas agricolas guayabal 2022-07-05 12 24"
           INNER JOIN "public"."area de maquina de riego 2022-07-06 20 32"
                 ON ST_Intersects("areas agricolas guayabal 2022-07-05 12 24"."geom",
                                   "area de maquina de riego 2022-07-06 20 32"."geom"))
```

Figura 16: Consulta para la pregunta “¿Qué cultivos se encuentran bajo riego?”.

Fuente: elaboración propia.

Como se puede apreciar por el tamaño y complejidad de esta consulta, un usuario sin capacitación suficiente en lenguaje SQL y consultas espaciales no sería capaz

de realizarla. Teniendo en cuenta además la imposibilidad de su realización por medio de un formulario. También se puede observar que la creación de las consultas SQL directamente en postgisQueryBuilder resulta engorroso debido a que no cuenta con auto completamiento durante el proceso de creación de la consulta SQL por lo que sería necesario la utilización de un tercer programa como pgAdmin⁷ o Navicat⁸ para facilitar el proceso.

Lo antes planteado demuestra la necesidad del uso de LeNatSQL para la realización de consultas SQL por parte de los usuarios con poca experiencia en lenguaje SQL y consultas espaciales debido principalmente a su fácil uso.

3.2 Conclusiones parciales

En el presente capítulo se describió el funcionamiento de las consultas en LeNatSQL, se mostraron algunos ejemplos de preguntas que pueden ser realizadas a este software. Se observaron las interfaces por las que transita el usuario hasta llegar a la visualización de los resultados en forma de capas en QGIS. Además, se plasmaron las ventajas del uso de LeNatSQL y las dificultades que provocan para usuarios inexpertos en lenguaje SQL y consultas espaciales, no utilizar este sistema.

⁷ pgAdmin es una aplicación para la gestión y administración de bases de datos en PostgreSQL. Es un sistema de código abierto, multiplataforma, disponible en Linux, Unix, Mac OS X y Windows.

⁸ Navicat es un administrador gráfico de base de datos, de pago, con soporte para PostgreSQL.

Conclusiones

1. El uso del SIG QGIS junto al SGBDR PostgreSQL a través de su extensión PostGIS para el trabajo con bases de datos espaciales, constituyen la herramienta adecuada para gestionar y visualizar la información geoespacial en la UDP “El Guayabal”.
2. El complemento postgisQueryBuilder resulta el mecanismo adecuado para la ejecución de consultas espaciales en QGIS. Sin embargo, la elaboración de consultas SQL directamente en postgisQueryBuilder resulta engorroso para usuarios sin experiencia en consultas SQL y relaciones espaciales.
3. Se obtiene un método para transformar las preguntas realizadas en lenguaje natural por los usuarios de la UDP “El Guayabal” en consultas en lenguaje SQL.
4. Se implementa la herramienta informática LeNatSQL para facilitar la realización de consultas para los usuarios no expertos en lenguaje SQL y consultas espaciales en la UDP “El Guayabal”.

Recomendaciones

1. La actualización cada cierto tiempo del dataset, así como el enriquecimiento del mismo a través de nuevas preguntas y sus respuestas correspondientes, resultantes de la experiencia obtenida en el trabajo con el sistema.
2. Desarrollar un complemento para QGIS utilizando inteligencia artificial que permita la transformación de las preguntas en lenguaje natural a consultas en SQL.

Referencias Bibliográficas

ANISHA: ; RAFEEQUE: ; REENA: Text to SQL Query Conversion Using Deep Learning: A Comparative Analysis, [en línea] 15 de agosto de 2019, Disponible en: <https://papers.ssrn.com/abstract=3437616> [Consulta: 26 de octubre de 2022].

BAIK, C.: ; JAGADISH, H.V.: ; LI, Y.: "Bridging the semantic gap with SQL query logs in natural language interfaces to databases", ISBN-1-5386-7474-2, pp. 374-385, 2019.

BALAMAN, Ş.Y.: "Chapter 7 - Modeling and Optimization Approaches in Design and Management of Biomass-Based Production Chains", [en línea] edit. Şebnem Yılmaz Balaman, Decision-Making for Biomass-Based Production Chains, ISBN-978-0-12-814278-3, pp. 185-236, Ed. Academic Press, 1 de enero de 2019. Disponible en: <https://www.sciencedirect.com/science/article/pii/B9780128142783000078> [Consulta: 20 de agosto de 2022].

BARETH, G.: ; WALDHOFF, G.: "2.01 - GIS for Mapping Vegetation", [en línea] edit. Bo Huang, Comprehensive Geographic Information Systems, ISBN-978-0-12-804793-4, pp. 1-27, Ed. Elsevier, Oxford, 1 de enero de 2018. Disponible en: <https://www.sciencedirect.com/science/article/pii/B9780124095489096366> [Consulta: 26 de octubre de 2022].

BARTEL, C.: "Computer Art, Technology, and the Medium", Being and Value in Technology, pp. 141-161, Ed. Springer, 2022.

BELCIU, A.: ; DIACONITA, V.: ; LUNGU, I.: ; VIRGOLICI, A.-M.: "Development of Spatial Database for Regional Development in Romania", Informatica Economica, 18(3): 55, 2014.

BENDUCH, P.: "Comparison of the results of spatial analyses performed in arcgis and qgis software", Infrastruktura i Ekologia Terenów Wiejskich, , 2017.

BONDE, Y.: "Text2SQL", [en línea] Python, 23 de octubre de 2022. Disponible en: <https://github.com/yashbonde/text2sql> [Consulta: 26 de octubre de 2022].

BONILLA HERNÁNDEZ, M.: Traductor de consultas del lenguaje natural a SQL, Universidad Central "Marta Abreu" de Las Villas, 2011.

CAI, H.: ; VASILAKOS, A.V.: "Chapter 12 - Web of Things Data Storage", [en línea] edit. Quan Z. Sheng, Yongrui Qin, Lina Yao, y Boualem Benatallah, Managing the Web of Things, ISBN-978-0-12-809764-9, pp. 325-354, Ed. Morgan Kaufmann, Boston, 1 de enero de 2017. Disponible en: <https://www.sciencedirect.com/science/article/pii/B9780128097649000159> [Consulta: 22 de agosto de 2022].

DAN, S.: ; KORDJAMSHIDI, P.: ; BONN, J.: ; BHATIA, A.: ; CAI, J.: ; PALMER, M.: ; ROTH, D.: "From spatial relations to spatial configurations", arXiv preprint arXiv:2007.09557, , 2020.

DAVIS, S.: ; MELTZER, P.S.: "GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor", [en línea] Bioinformatics, ISSN-1367-4803, DOI-10.1093/bioinformatics/btm254, 23(14): 1846-1847, 15 de julio de 2007.

DEMPSEY, C.: Types of GIS Data Explored: Vector and Raster, GIS Lounge, [en línea] 1 de julio de 2022, Disponible en: <https://www.gislounge.com/geodatabases-explored-vector-and-raster-data/> [Consulta: 21 de agosto de 2022].

FERREGUTI, E.: postgisQueryBuilder — Repositorio de complementos Python de QGIS, [en línea] 14 de julio de 2022, Disponible en: <https://plugins.qgis.org/plugins/postgisQueryBuilder/> [Consulta: 14 de octubre de 2022].

FLENNIKEN, J.M.: ; STUGLIK, S.: ; IANNONE, B.V.: "Quantum GIS (QGIS): An introduction to a free alternative to more costly GIS platforms: FOR359/FR428, 2/2020", EDIS, 2020(2): 7-7, 2020.

GINESTÀ, M.G.: ; MORA, O.P.: "Bases de datos en PostgreSQL", SI:[sn], , 2012.

GOODCHILD, M.F.: ; KEMP, K.K.: NCGIA core curriculum: introduction to GIS, Ed. National Center for Geographic Information and Analysis, University of ..., vol. 1, 1990.

GUILLOT JIMÉNEZ, J.: ; GARCÍA, L.: "Bases de datos NoSQL para la gestión de datos geoespaciales", [en línea] DOI-10.13140/RG.2.1.3803.0481, 2016.

GUO, D.: ; ONSTEIN, E.: "State-of-the-art geospatial information processing in NoSQL databases", ISPRS International Journal of Geo-Information, 9(5): 331, 2020.

HAMMERSLEY, B.: "SQL vs NoSQL: A Spatial Benchmark for Selected NoSQL and Relational Database Management Systems", , 2018.

HESS, P.T.: Building a Spatial Database for Agricultural Record Keeping and Management on a Regenerative Farm, University of Southern California, 2022.

HOWARI, F.M.: ; GHREFAT, H.: "Chapter 4 - Geographic information system: spatial data structures, models, and case studies", [en línea] edit. Abdel-Mohsen O. Mohamed, Evan K. Paleologos, y Fares M. Howari, Pollution Assessment for Sustainable Practices in Applied Sciences and Engineering, ISBN-978-0-12-809582-9, pp. 165-198, Ed. Butterworth-Heinemann, 1 de enero de 2021. Disponible en:

<https://www.sciencedirect.com/science/article/pii/B9780128095829000049>

[Consulta: 21 de agosto de 2022].

HSU, L.S.: ; OBE, R.: PostGIS in action, Ed. Simon and Schuster, ISBN-1-61729-669-4, 2021.

ISMAIL HOSSAIN, M.: ; MAHMUD, S.: ; SANTA, T.D.: "Oracle, MySQL, PostgreSQL, SQLite, SQL Server: Performance based competitive analysis", , 2019.

JANIPELLA, R.: ; GUPTA, V.: ; MOHARIR, R.V.: "Chapter 8 - Application of Geographic Information System in Energy Utilization", [en línea] edit. Sunil Kumar, Rakesh Kumar, y Ashok Pandey, Current Developments in Biotechnology and Bioengineering, ISBN-978-0-444-64083-3, pp. 143-161, Ed. Elsevier, 1 de enero de 2019. Disponible en: <https://www.sciencedirect.com/science/article/pii/B9780444640833000087> [Consulta: 21 de agosto de 2022].

KHAMIS, R.: ; SHATNAWI, S.: "Toward enhanced natural language processing to databases: building a specific domain ontology derived from database conceptual model", En: 2010 The 7th International Conference on Informatics and Systems (INFOS), ISBN-1-4244-5828-5, pp. 1-8, 2010.

KHAN, S.: ; MOHIUDDIN, K.: "Evaluating the parameters of ArcGIS and QGIS for GIS Applications", Int. J. Adv. Res. Sci. Eng, 7: 582-594, 2018.

KHOIROM, S.: ; SONIA, M.: ; LAIKHURAM, B.: ; LAISHRAM, J.: ; SINGH, T.D.: "Comparative analysis of Python and Java for beginners", Int. Res. J. Eng. Technol, 7(8): 4384-4407, 2020.

KUL, G.: ; LUONG, D.T.A.: ; XIE, T.: ; CHANDOLA, V.: ; KENNEDY, O.: ; UPADHYAYA, S.: "Similarity metrics for SQL query clustering", IEEE Transactions on Knowledge and Data Engineering, 30(12): 2408-2420, 2018.

LESLIE, M.: ; RAMSEY, P.: "Introduction to PostGIS", , 2022.

LI, Z.: ; HUANG, D.: ; LIU, X.: ; XIE, N.: ; ZHANG, G.: "Information structures in a covering information system", Information Sciences, 507: 449-471, 2020.

LITHMEE What is the Difference Between Attribute Data and Spatial Data, Pediaa.Com, [en línea] 3 de marzo de 2019, Disponible en: <https://pediaa.com/what-is-the-difference-between-attribute-data-and-spatial-data/> [Consulta: 20 de agosto de 2022].

LLITERAS, B.: "Sistemas de Referencias para Aplicaciones SIG's", : 10, 2000.

MAINA, J.: ; OUMA, P.O.: ; MACHARIA, P.M.: ; ALEGANA, V.A.: ; MITTO, B.: ; FALL, I.S.: ; NOOR, A.M.: ; SNOW, R.W.: ; OKIRO, E.A.: "A spatial database of health facilities managed by the public health sector in sub Saharan Africa", Scientific data, 6(1): 1-8, 2019.

MALIAKAL, G.: "Chapter Two - Database Management", [en línea] edit. Michael Levin-Epstein, Careers in Biomedical Engineering, ISBN-978-0-12-814816-7, pp. 19-36, Ed. Academic Press, 1 de enero de 2019. Disponible en: <https://www.sciencedirect.com/science/article/pii/B9780128148167000029> [Consulta: 22 de agosto de 2022].

MARIANI, G.: "Los Sistemas de Información Geográfica (SIG) - Caja de herramientas para un Nuevo Urbanismo", 2012.

MAURYA, S.: ; OHRI, A.: ; MISHRA, S.: "Open Source GIS: A Review", 2015.

NASUTION, M.: "Relational Database Management Systems", 2 de marzo de 2021.

NAYAK, A.: ; PORIYA, A.: ; POOJARY, D.: "Type of NOSQL databases and its comparison with relational databases", International Journal of Applied Information Systems, 5(4): 16-19, 2013.

NUR, W.H.: ; KUMORO, Y.: ; SUSILOWATI, Y.: "GIS and Geodatabase Disaster Risk for Spatial Planning", [en línea] IOP Conference Series: Earth and Environmental Science, ISSN-1755-1315, DOI-10.1088/1755-1315/118/1/012046, 118: 012046, febrero de 2018.

OLAYA, V.: Sistemas de información geográfica. Un libro libre de Víctor Olaya, Ed. Obtenido de: [http://volaya.github.io/librosig/chapters ...](http://volaya.github.io/librosig/chapters...), 2014.

PÉREZ BETANCOURT, Y.G.: ; GONZÁLEZ POLANCO, L.: ; MORALES PÉREZ, R.: ; TORRES VEGA, Y.: "Estratificación de territorios basada en indicadores de

salud sobre el Sistema de Información Geográfica QGIS", Revista Cubana de Ciencias Informáticas, ISSN-2227-1899, 10: 163-175, 2016.

PEREZ-GARCIA, C.A.: ; PÉREZ-ATRAY, J.J.: ; HERNÁNDEZ-SANTANA, L.: ; GUSTABELLO-COGLE, R.: ; BECERRA-DE ARMAS, E.: "Sistema de Información Geográfica para la agricultura cañera en la provincia de Villa Clara", Revista Cubana de Ciencias Informáticas, 13(2): 30-46, 2019.

PILICITA GARRIDO, A.: ; BORJA LÓPEZ, Y.: ; GUTIÉRREZ CONSTANTE, G.: "Rendimiento de MariaDB y PostgreSQL", , 2021.

PINO, A.V.: ; PONCE, J.: ; MORETA, O.E.: "Análisis de la optimización de sentencias SQL del Lenguaje de Consulta Estructurado usando grandes volúmenes de datos", Revista Publicando, 5(16 (1)): 70-79, 2018.

PINTO MOLINA, M.: "Búsqueda y recuperación de información", , 2018.

PRATT, M.: "Asking the Right Questions", , 2009.

R. ADAM DASTRUP, M.A.: "1.4 The Future of GIS", [en línea] Introduction to Geographic Information Systems, 2020. Disponible en: <https://gist.pressbooks.com/chapter/1-4/> [Consulta: 29 de julio de 2022].

RAINER, R.K.: ; PRINCE, B.: Introduction to information systems, Ed. John Wiley & Sons, ISBN-1-119-76146-8, 2021.

RASULOVICH, H.R.: "The importance of information technology in training medical personnel", European Journal of Research and Reflection in Educational Sciences, 8: 12, 2020.

RAWAT, B.: ; PURNAMA, S.: "MySQL Database Management System (DBMS) On FTP Site LAPAN Bandung", International Journal of Cyber and IT Service Management, 1(2): 173-179, 2021.

RESTREPO, J.C.G.: "La cartografía en el ámbito de las tecnologías de la información y la comunicación (TIC)", DIM: Didáctica, Innovación y Multimedia, ISSN-1699-3748, (31): 1-18, 2015.

REYES GARCÍA, F.: LNE2SQL: traductor de consultas del lenguaje natural a SQL v2.0, [en línea] Tesis (en opción al Thesis), Universidad Central "Marta Abreu" de Las Villas, 2012. Disponible en: <http://dspace.uclv.edu.cu:8089/xmlui/handle/123456789/6067> [Consulta: 25 de octubre de 2022].

SALLEH, S.; UJANG, U.; AZRI, S.: "3D TOPOLOGICAL SUPPORT IN SPATIAL DATABASES: AN OVERVIEW.", International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, 46, 2021.

SHIN, M.E.; CAMPBELL, J.; BURKHART, N.: Essentials of Geographic Information Systems, Ed. FlatWorld, ISBN-1-4533-9080-4, 2018.

SHNIZAI, Z.: Spatial and Attribute Querying, [en línea] 12 de noviembre de 2011.

SINGH, S.: "DATABASE MANAGEMENT SYSTEM", Journal of Management Research and Analysis, ISSN-2394-2770, 2(1): 72-80, 15 de abril de 2015.

TAIPALUS, T.; SIPONEN, M.; VARTIAINEN, T.: "Errors and complications in SQL query formulation", ACM Transactions on Computing Education (TOCE), 18(3): 1-29, 2018.

TALAWAR, A.R.; TALAWAR, A.R.; KUMBALAVATI, S.B.; MALLAPUR, J.D.: "One Step Technological Solution for Agriculture", ICT for Competitive Strategies, ISBN-1-00-305209-6, pp. 713-719, Ed. CRC Press, 2020.

TWUMASI, B.O.: "Modelling spatial object behaviours in object-relational geodatabase", 2002.

UMA, M.: ; SNEHA, V.: ; SNEHA, G.: ; BHUVANA, J.: ; BHARATHI, B.: "Formation of SQL from natural language query using NLP", ISBN-1-5386-9471-9, pp. 1-5, 2019.

VALVERDE, V.: ; PORTALANZA, N.: ; MORA, P.: "Análisis descriptivo de base de datos relacional y no relacional", Revista Atlante: Cuadernos de Educación y Desarrollo, 3, 2019.

WANG, H.: ; PAN, Y.: ; LUO, X.: "Integration of BIM and GIS in sustainable built environment: A review and bibliometric analysis", Automation in construction, 103: 41-52, 2019.

WODYK, R.: ; SKUBLEWSKA-PASZKOWSKA, M.: "Performance comparison of relational databases SQL Server, MySQL and PostgreSQL using a web application and the Laravel framework", Journal of Computer Sciences Institute, 17: 358-364, 2020.

YUE, P.: ; TAN, Z.: "GIS Databases and NoSQL Databases", [en línea] Reference Module in Earth Systems and Environmental Sciences, ISBN-978-0-12-409548-9, 31 de diciembre de 2017.