



JOÃO VICTOR TEÓFIO SALGADO

**APRIMORAMENTO DO PLANEJAMENTO DE
REDES SEM FIO:
VALIDAÇÃO AUTOMATIZADA DE DADOS
ESPACIAIS**

1^a edição revista, atualizada e ampliada

LAVRAS – MG

2024

JOÃO VICTOR TEÓFIO SALGADO

APRIMORAMENTO DO PLANEJAMENTO DE REDES SEM FIO:

VALIDAÇÃO AUTOMATIZADA DE DADOS ESPACIAIS

1^a edição revista, atualizada e ampliada

Artigo apresentado à Universidade Federal de Lavras, como parte das exigências do Curso de Ciência da Computação, para obtenção do título de Bacharel.

Prof. DSc. Renata Lopes Rosa

Orientadora

LAVRAS – MG

2024

**Ficha catalográfica elaborada pela Coordenadoria de Processos Técnicos
da Biblioteca Universitária da UFLA**

Salgado, João Victor Teófilo

Aprimoramento do Planejamento de Redes sem Fio : Validação
Automatizada de Dados Espaciais / João Victor Teófilo Salgado. –
Lavras : UFLA, 2024.

45 p. : il.

TCC–Universidade Federal de Lavras, 2024.

Orientadora: Prof. DSc. Renata Lopes Rosa.

Bibliografia.

1. Redes sem fios. 2. Geoprocessamento. 3. Engenharia de
Dados. Trabalho Científico – Normas. I. Universidade Federal de
Lavras. II. Título.

CDD-808.066

JOÃO VICTOR TEÓFIO SALGADO

**APRIMORAMENTO DO PLANEJAMENTO DE REDES SEM FIO:
VALIDAÇÃO AUTOMATIZADA DE DADOS ESPACIAIS
ENHANCING WIRELESS NETWORK PLANNING: AUTOMATED
VALIDATION OF SPATIAL DATA**

Artigo apresentado à Universidade Federal de Lavras, como parte das exigências do Curso de Ciência da Computação, para obtenção do título de Bacharel.

APROVADA em [DATA DA DEFESA].

Profa. Renata Lopes Rosa DCC - UFLA
Prof. Demóstenes Zegarra Rodríguez DCC - UFLA
Prof. Vinicius Vitor dos Santos Dias DCC - UFLA

Prof. DSc. Renata Lopes Rosa
Orientadora

**LAVRAS – MG
2024**

AGRADECIMENTOS

Agradeço a minha família, principalmente meus pais Carlos e Sandra, que sempre estiveram ao meu lado em todas as etapas da minha formação. Seu incentivo e apoio incondicional foram essenciais para que eu pudesse superar os desafios e chegar até aqui.

Agradeço aos meus amigos que compartilharam comigo a jornada acadêmica, em especial: Sérgio Garcia e Thiago Takaki, que me acompanharam desde o início da graduação.

Agradeço a minha namorada, Dandara, por estar sempre ao meu lado e ser o meu porto seguro durante a reta final desta jornada acadêmica.

Por fim, agradeço à Comp Júnior, por auxiliar na minha formação profissional como desenvolvedor e acreditar no meu potencial desde o começo da graduação.

RESUMO

Este trabalho aborda o desafio de manter a integridade dos dados espaciais em bancos de dados utilizados para o planejamento e otimização de redes sem fio. Os autores propõem uma aplicação automatizada para validar geometrias em ambientes de banco de dados, complementada por uma interface web de fácil uso. A solução visa eliminar problemas causados por geometrias inválidas, aumentando a confiabilidade e eficiência das operações de bancos de dados espaciais. Avaliações de desempenho e precisão demonstram a eficácia da abordagem, melhorando a experiência do usuário e a confiança nos processos de planejamento de redes sem fio com esforços mínimos de implementação. O estudo destaca a importância da integridade dos dados espaciais, especialmente em consultas complexas, e sugere pesquisas futuras para desenvolver ferramentas automatizadas que detectem e corrijam geometrias inválidas de forma preventiva.

Palavras-chave: Redes sem fios; Geoprocessamento; Engenharia de Dados.

ABSTRACT

This work addresses the challenge of maintaining the integrity of spatial data in databases used for planning and optimizing wireless networks. The authors propose an automated application to validate geometries in database environments, complemented by an easy-to-use web interface. The solution aims to eliminate problems caused by invalid geometries, increasing the reliability and efficiency of spatial database operations. Performance and accuracy evaluations demonstrate the effectiveness of the approach, improving the user experience and confidence in wireless network planning processes with minimal implementation efforts. The study highlights the importance of spatial data integrity, especially in complex queries, and suggests future research to develop automated tools that detect and correct invalid geometries preventively.

Keywords: Wireless networks; geoprocessing; Data engineering.

LISTA DE FIGURAS

Figura 3.1 – Mapa da cidade de Lavras, Minas Gerais, Brasil. Área de interesse, em vermelho, utilizada para recordar o <i>dataset</i> original.	31
Figura 3.2 – Representação de uma geometria inválida contendo auto-intersecções.	32
Figura 3.3 – Representação de uma geometria com orientação de vértices inválida.	33
Figura 3.4 – Os dados geométricos inválidos oriundos de várias fontes (à esquerda) são impedidos de serem acessados por usuários finais e aplicativos por meio do uso de <i>views</i> (à direita).	36
Figura 3.5 – Desempenho médio para cada consulta e conjunto de dados disponível. O eixo y representa cada conjunto de dados por sua taxa de integridade. O eixo x representa o tempo de execução de cada consulta (em segundos).	39
Figura 3.6 – Desvio médio dos valores originais para cada consulta. O eixo y representa cada código de erro analisado. O eixo x representa o valor total do desvio na unidade especificada.	40

LISTA DE TABELAS

Tabela 3.1 – Conjuntos de dados planejados	30
--	----

SUMÁRIO

1	INTRODUÇÃO	17
1.1	Objetivos	17
1.2	Estrutura do documento	18
2	REFERENCIAL TEÓRICO	19
2.1	Dados Espaciais	19
2.2	Sistemas de Informações Geográficas	20
2.3	Bancos de dados	20
2.3.1	Sistema Gerenciador de Banco de Dados	21
2.3.2	Armazenamento de Dados Espaciais	21
2.4	Planejamento de rede sem fio	22
3	ARTIGO 1 - Aprimoramento do Planejamento de Redes sem Fio: Validação Automatizada de Dados Espaciais	25
3.1	Introdução	25
3.2	Trabalhos relacionados	27
3.3	Metodologia	29
3.3.1	Ambiente proposto	29
3.3.2	Conjuntos de dados	29
3.3.2.1	Obtenção de dados geométricos válidos	30
3.3.2.2	Geração de dados geométricos inválidos	31
3.3.3	Detalhes de implementação	33
3.3.3.1	Consulta para testes de performance	33
3.3.3.2	Consultas para testes de precisão	34
3.3.4	Solução proposta	35
3.3.5	Simulação de dispositivos sem fio e IoT	37
3.4	Resultados e discussões experimentais	37
3.5	Considerações finais	38
4	CONCLUSÃO	41

REFERÊNCIAS 43

1 INTRODUÇÃO

As redes sem fio modernas são a espinha dorsal da comunicação global, permitindo a troca de informações em tempo real e a conectividade contínua. No entanto, o planejamento e a otimização eficazes dessas redes são desafiadores, especialmente quando se trata de manter a integridade dos dados espaciais envolvidos no processo (MACHADO et al., 2012). Estes dados são cruciais para a localização precisa e o posicionamento de dispositivos em uma rede, impactando diretamente o desempenho e a eficiência do serviço (SILVA, 2011). A saúde dos dados espaciais é, portanto, um aspecto fundamental para garantir a confiabilidade das operações de rede.

Além dos dados obtidos de forma inerente ao processo, o planejamento de redes inclui também o uso de dados obtidos de terceiros. Estes terceiros podem ser integrações específicas de negócios (por exemplo, *GE Smallworld*), conjuntos de dados de arruamento proprietários (por exemplo, *ArcGIS StreetMap Premium*) e conjuntos de dados abertos (por exemplo, *OpenStreetMap*). A consolidação destes conjuntos envolve, muitas vezes, a conversão entre formatos de armazenamento de geometrias utilizados por cada provedor. Essa conversão, quando problemática, pode acarretar na inserção de feições inválidas do ponto de vista da base de dados de destino, degradando a integridade geral (TEIXEIRA et al., 2017).

Diante destes desafios, neste trabalho, apresentaremos uma nova metodologia para a análise dos impactos causados justamente pela presença de geometrias corrompidas numa base de dados. Além disso, sugerimos uma abordagem para prevenir os impactos mencionados anteriormente baseada numa aplicação *web* amigável de fácil implementação.

1.1 Objetivos

O objetivo principal deste trabalho foi mensurar os impactos na performance e precisão de consultas causados pela integridade de um conjunto de dados

espaciais. Através deste objetivo, diversos objetivos específicos foram traçados sendo os principais descritos a seguir.

- **Obtenção de dados para realização de *benchmark***

No decorrer do artigo, discutimos sobre a escolha de um *dataset* base para a realização dos testes que represente a variabilidade de dados gerados por comportamento humano de forma fiel. Além disso, demonstramos a necessidade do desenvolvimento de uma ferramenta para artificialmente deformar as geometrias do *dataset* original de forma arbitrária dado a ausência de conjuntos de dados satisfatórios contendo geometrias inválidas para este fim.

- **Construção de consultas de *benchmark***

Visando apresentar testes representativos do caso de uso original deste artigo, o planejamento de redes sem fio, consultas de variados níveis de complexidade foram construídas. Estas consultas foram então aplicadas a conjuntos de dados de variados níveis de integridade, discutidos no tópico anterior, para obter os resultados apresentados no artigo.

1.2 Estrutura do documento

O presente documento é estruturado de forma a seguir. A Seção 2 é dedicada à revisão dos aspectos teóricos relevantes abordados durante o desenvolvimento do artigo, fornecendo uma base para o entendimento dos assuntos tratados. Já na Seção 3, o artigo é disposto em sua completude. Por fim, a Seção 4 é dedicada à discussão das atividades, resultados, desafios encontrados e as lições aprendidas. Apresentando uma conclusão geral do trabalho, incluindo uma reflexão sobre os aspectos mais importantes e as contribuições do tema estudado para o desenvolvimento profissional do aluno.

2 REFERENCIAL TEÓRICO

Este capítulo descreve os principais conceitos e tecnologias utilizadas ao longo de trabalho, a saber: dados espaciais, sistemas de informações geográficas e bancos de dados.

2.1 Dados Espaciais

Dados espaciais, também conhecidos como dados geoespaciais, referem-se a informações que possuem uma localização geográfica associada (LONGLEY et al., 2015). Esses dados são fundamentais em diversas áreas de pesquisa e aplicações práticas, incluindo geografia, planejamento de redes sem fio, ciências ambientais e sistemas de informação geográfica (SIG) (GOODCHILD, 2010).

Os dados espaciais podem ser classificados em duas categorias principais: dados vetoriais e dados *raster* (BOLSTAD, 2012). Dados vetoriais são representados por pontos, linhas e polígonos, sendo úteis para representar características discretas do mundo real, como estradas e edifícios. Por outro lado, os dados *raster* são representados por *pixels* e são mais adequados para representar fenômenos contínuos, como elevação do terreno ou temperatura (SMITH; GOODCHILD; LONGLEY, 2018).

Neste trabalho, apenas os dados espaciais no formato vetorial serão abordados visto que são amplamente utilizados nas trocas de informações geográficas relacionadas a redes sem fio, como por exemplo: posicionamento de dispositivos móveis, delimitação de infraestrutura cabeada e demarcação de áreas de cobertura. Correspondendo aos tipos geométricos mencionados acima, respectivamente: pontos, linhas e polígonos.

2.2 Sistemas de Informações Geográficas

Os Sistemas de Informações Geográficas (SIG) são sistemas computadorizados voltados à armazenagem, gerenciamento e manipulação de dados geográficos, hoje com uso disseminado em diversos campos científicos, tiveram suas primeiras aplicações nas Ciências Ambientais (SANTOS; BRITO; SILVA-NETO, 2022). Esses sistemas têm se desenvolvido e popularizado tanto nas últimas duas décadas que hoje em dia são aceitos como ferramentas essenciais definitivas para a manipulação das informações geográficas (NAJAR; MARQUES, 1998). Permitem reunir uma grande quantidade de dados convencionais de expressão espacial, estruturando-os adequadamente, de modo a otimizar o tratamento integrado de seus três componentes: posição, topologia e atributos, na execução de análises e aplicações gráficas (NAJAR; MARQUES, 1998).

No campo do planejamento de redes sem fio, os Sistemas de Informações Geográficas (SIG) têm um papel crucial. Eles são utilizados para otimizar a localização e a cobertura das estações base, levando em consideração a topografia do terreno e os padrões de uso dos usuários (NAKAMURA, 2003). Além disso, os SIG podem ajudar a identificar áreas de baixa cobertura e sugerir locais ideais para a instalação de novas antenas. Portanto, o uso de SIG em conjunto com outras técnicas de geoprocessamento pode melhorar significativamente a eficiência e a eficácia do planejamento e da gestão de redes sem fio (NAKAMURA, 2003).

2.3 Bancos de dados

Os Bancos de Dados são uma parte essencial da infraestrutura tecnológica moderna, facilitando o armazenamento, a recuperação e a manipulação de dados de maneira eficiente e segura. A seguir, aprofundaremos a discussão sobre os sistemas gerenciadores de bancos de dados (SGDB) e como o armazenamento de dados espaciais é implementado nestas plataformas.

2.3.1 Sistema Gerenciador de Banco de Dados

Os sistemas de gerenciamento de banco de dados (SGBD) fornecem uma interface entre os dados brutos armazenados e as aplicações que precisam desses dados (ELMASRI; NAVATHE, 2011). Além disso, estes sistemas podem implementar extensões para permitir o desenvolvimento de rotinas e aplicações que executam no mesmo contexto da instância do banco de dados. Devido a ausência de perdas de rede inerentes a comunicação com SGBDs, estas aplicações possuem vantagens no que tange a performance das operações. Devido a isso, neste trabalho, sugerimos uma solução baseada em PL/SQL, uma extensão da linguagem padrão SQL para Oracle que permite a realização dos desenvolvimentos customizados necessários.

Os SGBDs modernos suportam uma variedade de modelos de dados, incluindo relacional, objeto-relacional e NoSQL, cada um com suas próprias vantagens e desvantagens (SILBERSCHATZ; KORTH; SUDARSHAN, 2020). O modelo relacional, por exemplo, é amplamente utilizado devido à sua simplicidade e eficácia na representação de dados estruturados, enquanto os modelos NoSQL são preferidos para lidar com dados não estruturados ou semi-estruturados em grande escala.

A pesquisa em bancos de dados continua a ser uma área ativa, com foco em tópicos como escalabilidade, consistência, tolerância a falhas e segurança dos dados (STONEBRAKER; HELLERSTEIN, 2015). À medida que avançamos para uma era cada vez mais digital, a importância dos bancos de dados só tende a aumentar.

2.3.2 Armazenamento de Dados Espaciais

O armazenamento de dados espaciais em bancos de dados é uma área de pesquisa em constante evolução. Segundo Longley et al. 2005, os bancos de dados espaciais são projetados para armazenar e gerenciar dados que estão associados a

objetos no espaço, como localizações geográficas. Esses bancos de dados são capazes de realizar consultas espaciais complexas e fornecer resultados precisos, o que é essencial em muitos campos, incluindo geografia, planejamento de redes sem fio e ciências ambientais.

No entanto, o armazenamento de dados espaciais apresenta vários desafios. De acordo com Stonebraker et al. 2007, os bancos de dados relacionais tradicionais não foram projetados para lidar com dados espaciais nativamente, o que pode resultar em problemas de compatibilidade com as regras do negócio. Para superar esses desafios, foram desenvolvidas extensões espaciais para bancos de dados relacionais, como o Oracle Spatial and Graph para o Oracle, que permitem o armazenamento e a manipulação eficientes de dados espaciais. Neste trabalho, as análises e testes realizados utilizarão justamente a extensão mencionada visto que empregamos o banco de dados Oracle na versão 19.

2.4 Planejamento de rede sem fio

O planejamento de redes sem fio é uma área fundamental na engenharia de redes, especialmente com o crescente uso de dispositivos móveis e a demanda por conectividade em qualquer lugar. O planejamento adequado visa garantir uma cobertura confiável e de alta qualidade, minimizando interferências e maximizando a capacidade da rede.

Um dos principais desafios no planejamento de redes sem fio é a otimização da cobertura, levando em consideração diversos fatores como a topografia do local, materiais de construção, obstáculos físicos e a distribuição esperada dos usuários. Modelos de propagação de sinal são frequentemente utilizados para prever a intensidade do sinal em diferentes áreas, auxiliando na determinação da localização ideal dos pontos de acesso.

Além da cobertura, o planejamento também deve considerar a capacidade da rede. Isso inclui a distribuição adequada dos pontos de acesso para atender à

demanda de tráfego esperada, bem como a escolha das tecnologias e protocolos mais adequados para garantir um desempenho satisfatório, como o uso de bandas de frequência adequadas e a implementação de técnicas de acesso múltiplo.

Outro aspecto importante do planejamento de redes sem fio é a mitigação de interferências.

O uso de um banco de dados com dados em forma de grafos é extremamente importante no planejamento de redes sem fio, pois permite representar de forma eficiente as diversas relações e interações entre os elementos da rede, como pontos de acesso, dispositivos e obstáculos físicos.

No entanto, é crucial que os dados nesse banco sejam limpos e claros para garantir a precisão e confiabilidade das análises e decisões tomadas com base neles. Dados sujos ou mal formatados podem levar a resultados imprecisos e a decisões inadequadas no planejamento da rede.

Portanto, é essencial que os dados sejam consistentes, livres de erros e bem estruturados, para que possam ser facilmente interpretados e utilizados de forma eficaz no planejamento da rede sem fio. Isso inclui a verificação da qualidade dos dados, a eliminação de duplicatas e a padronização da forma como os dados são registrados e armazenados. Dessa forma, é possível garantir um planejamento mais preciso e eficiente da rede sem fio, levando a uma melhor experiência do usuário e a uma maior eficiência na utilização dos recursos da rede.

3 ARTIGO 1 - APRIMORAMENTO DO PLANEJAMENTO DE REDES SEM FIO: VALIDAÇÃO AUTOMATIZADA DE DADOS ESPACIAIS

3.1 Introdução

Os bancos de dados espaciais são componentes vitais no âmbito das comunicações sem fio, servindo como repositórios de dados espaciais para várias aplicações incluindo planejamento de rede, otimização e serviços baseados em localização. Esses bancos de dados, normalmente implementados em sistemas de gerenciamento de banco de dados relacionais (RDBMS) (SHARMA et al., 2022), costumam incorporar conjuntos de geometrias de diferentes fontes como: integrações específicas de negócios (por exemplo, *GE Smallworld*), conjuntos de dados proprietários (por exemplo, *ArcGIS StreetMap Premium*) e conjuntos de dados abertos (por exemplo, *OpenStreetMap*), o que é particularmente relevante no contexto do planejamento da infraestrutura de comunicação sem fio (YADAV; SHARMA, 2023).

No domínio das comunicações sem fio, as técnicas precisas de localização e posicionamento desempenham um papel fundamental na otimização do desempenho da rede e na viabilização de serviços avançados (OBEIDAT et al., 2021). O aproveitamento de conjuntos de dados espaciais enriquecidos com dados geográficos precisos de fontes confiáveis facilita a implementação de algoritmos sofisticados de localização e estratégias de planejamento de rede. Além disso, a integração do sensoriamento remoto e comunicações em bancos de dados espaciais abre caminho para novos aplicativos e serviços, permitindo uma interação contínua entre dispositivos sem fio e o ambiente ao redor (DECURNINGE et al., 2018).

É importante destacar que esses conjuntos de dados geralmente armazenam e descrevem geometrias usando formatos diferentes e, muitas vezes, proprietários (MAMOULIS, 2022). Essa situação impõe muitos desafios durante a integração e a consolidação de diferentes fontes de dados em um único *warehouse*. É

comum encontrar geometrias inválidas e malformadas em um banco de dados espacial depois de consolidar dados de diferentes fontes (BRUNSDON; COMBER, 2021).

Neste artigo, avaliaremos os impactos dessas geometrias inválidas e proporemos uma aplicação projetada especificamente para garantir a integridade dos dados espaciais usados, por exemplo, no planejamento e na otimização de redes de comunicação sem fio. Ao aproveitar algoritmos de alto desempenho consolidados, a aplicação tem como objetivo aprimorar a precisão e a confiabilidade dos bancos de dados espaciais, contribuindo para o avanço dos sistemas e serviços de comunicação sem fio (RODRÍGUEZ et al., 2019). Portanto, em resumo, o aprimoramento do planejamento de redes sem fio proposto neste trabalho oferece as seguintes contribuições:

- Nova metodologia para a criação de um ambiente de *benchmarking* robusto destinado a fornecer testes de desempenho confiáveis e isolados, por meio de uma descrição detalhada do processo de seleção e pré-processamento de conjuntos de dados de teste, enfatizando a importância dos índices de integridade.
- Utilização de instâncias de banco de dados de Processamento de Transação Autônoma (ATP) e Infraestrutura em Nuvem (CI) para execução de testes de desempenho, garantindo acessibilidade e reprodutibilidade.
- Dados geométricos gerados por comportamento humano para *benchmarking*, oferecendo *insights* realistas sobre os impactos da integridade do *dataset* no desempenho de consultas.
- Integração de dispositivos de IoT e simulação sem fio na metodologia, fornecendo uma avaliação abrangente do desempenho do banco de dados espacial em cenários de IoT.

As seções seguintes deste documento estão estruturadas da seguinte forma: Trabalhos relacionados tratam dos trabalhos existentes na área, enquanto a seção Metodologia aborda os aspectos práticos da análise efetuada. A seção Resultados trata dos resultados dos experimentos realizados. Finalmente, A seção Conclusões resume os resultados do trabalho proposto.

3.2 Trabalhos relacionados

Nesta seção, são tratados os trabalhos relacionados aos métodos tradicionais sobre dados espaciais.

Vários estudos abordaram os desafios associados aos bancos de dados espaciais e sua utilização no contexto das comunicações sem fio (MAO, 2023). Embora esses estudos tenham feito contribuições significativas, eles geralmente encontram limitações e deficiências que nossa abordagem proposta visa superar.

Um problema comum no gerenciamento de bancos de dados espaciais é a presença de geometrias inválidas (MAUS et al., 2020), o que pode levar a comportamentos inesperados e à degradação do desempenho durante a execução de consultas (BADHRUDEEN et al., 2022). Várias abordagens foram propostas para resolver esse problema, incluindo processos de validação manual e *scripts ad-hoc*. No entanto, esses métodos geralmente consomem muito tempo, são propensos a erros e não são escaláveis, o que os torna impraticáveis para bancos de dados espaciais de larga escala (JAIN et al., 2023).

Por exemplo, Meng et al. (MENG et al., 2023) propuseram um procedimento de validação manual para detectar e corrigir geometrias inválidas em bancos de dados espaciais. Embora seja eficaz na identificação de erros, essa abordagem depende muito da intervenção humana, o que a torna ineficiente e propensa a falhas, especialmente em ambientes com dados espaciais extensos. Da mesma forma, um *script* personalizado para validação de geometria (SALLEH; UJANG; AZRI, 2023) é introduzido em ambientes de banco de dados *Oracle*. Embora seja

capaz de automatizar o processo de validação até certo ponto, essa abordagem carece de robustez e escalabilidade, pois não oferece uma solução abrangente para lidar com vários tipos de erros geométricos.

Um desafio comum nos sistemas de comunicação sem fio é garantir a precisão das coordenadas espaciais ingeridas em um banco de dados. Essas coordenadas geralmente são obtidas de dispositivos remotos de uso pessoal, normalmente com precisão e exatidão insuficientes devido à intensidade de sinal, à qualidade do hardware etc. A determinação precisa dessas coordenadas é essencial para otimizar o desempenho da rede e garantir a conectividade contínua em sistemas de comunicação sem fio (LI et al., 2020). No entanto, as imprecisões nas coordenadas espaciais, geralmente decorrentes de geometrias inválidas, podem afetar significativamente a eficácia das operações de comunicação sem fio (CHEN; LIU, 2022). Essas imprecisões podem levar a comportamentos inesperados e à degradação do desempenho durante as atividades (BOUKERCHE; KANTARCI; KAPTAN, 2020). Conforme mencionado anteriormente, apesar dos esforços para resolver esse problema por meio de vários métodos como processos de validação manual e *scripts ad-hoc*, estes geralmente se mostram ineficientes e impraticáveis, especialmente quando aplicados a grandes bancos de dados espaciais (ALAWAD; KRAEMER, 2022). Para enfrentar esses desafios, a abordagem proposta enfatiza a automatização e escalabilidade do processo.

Nosso aplicativo de validação de geometria automatizada proposto oferece uma solução abrangente e dimensionável para enfrentar os desafios associados ao gerenciamento de bancos de dados espaciais em ambientes de comunicação sem fio. Além disso, nossa abordagem sugere uma interface de usuário acessível baseada na *web*, o que facilita a implantação e uso mesmo para usuários com conhecimento técnico limitado. Com nosso trabalho, pretendemos aprimorar a experiência geral do usuário e a confiança nos bancos de dados espaciais existentes, minimi-

zando os esforços de implementação e melhorando o desempenho e a eficiência do planejamento e otimização de redes de comunicação sem fio.

3.3 Metodologia

Nesta seção, explicaremos as principais etapas que seguimos para criar um ambiente de *benchmarking* confiável e seguro para avaliar os efeitos de geometrias inválidas. Em seguida, examinaremos os detalhes referentes à seleção e ao processamento dos conjuntos de dados de teste. mais tarde, discutiremos os detalhes específicos da elaboração de cada teste de *benchmark*. Por fim, apresentaremos uma discussão sobre como implementamos uma solução para evitar a disseminação de dados geométricos inválidos em um banco de dados espacial.

3.3.1 Ambiente proposto

Todos os testes de desempenho deste trabalho foram executados em instâncias do banco de dados *Oracle Autonomous Transaction Processing (ATP)* na versão 19.22.0.1.0 (19c). O motivo por trás desta versão específica é sua alta base de usuários e suporte ativo. Por fim, considerando normalização, relação custo-benefício e o licenciamento, todas as instâncias serão executadas em um ambiente de nuvem fornecido pela *Oracle Cloud Infrastructure (OCI)*.

3.3.2 Conjuntos de dados

Um dos objetivos deste documento é destacar os impactos no desempenho de geometrias inválidas em um conjunto de dados espaciais. Dito isso, planejamos quatro conjuntos de dados com proporções variadas de feições inválidas/válidas para realizar *benchmarks*. Um detalhamento é apresentado na Tabela 3.1.

Agora, será apresentado uma descrição detalhada de como os conjuntos de dados de geometrias válidas e inválidas foram adquiridos e processados.

Tabela 3.1 – Conjuntos de dados planejados

Nome do conjunto de dados	Integridade (% de feições válidas)
BUILDING	100%
BUILDING_80_PERCENT	80%
BUILDING_50_PERCENT	50%
BUILDING_20_PERCENT	20%

3.3.2.1 Obtenção de dados geométricos válidos

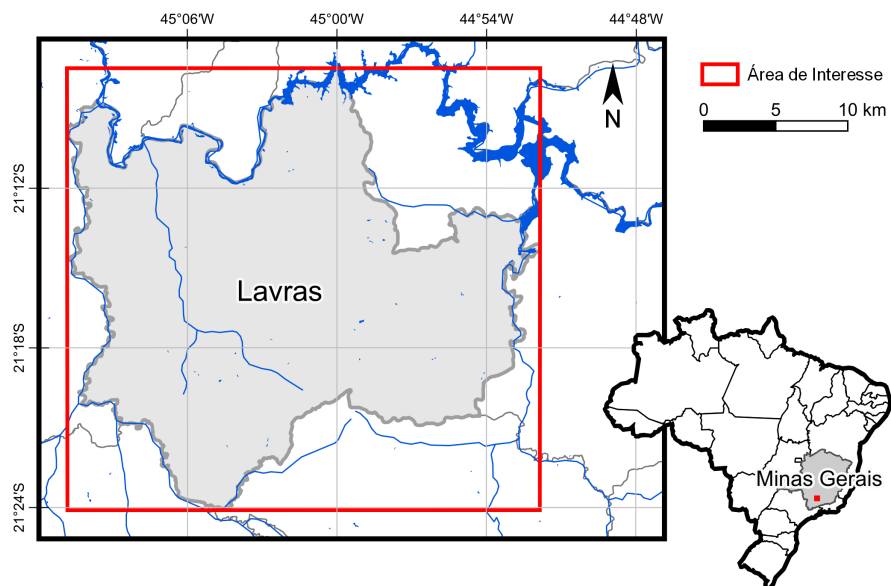
Com relação aos dados geométricos válidos, consideramos muitos aspectos antes de escolher um conjunto de dados. Os mais importantes deles, que levaram à nossa decisão final, foram:

- **Dados reais gerados por humanos:** Para aumentar a fidelidade com cenários do mundo real, foi dada preferência aos dados gerados por comportamento humano.
- **Acessibilidade e disponibilidade:** Foi dada preferência a conjuntos de dados abertos e gratuitos, permitindo uma alta reprodutibilidade deste trabalho.
- **Tamanho:** Novamente, para estar alinhado com cenários do mundo real, o conjunto de dados deve ter um tamanho considerável.

Com isso em mente, foi escolhido o conjunto de dados do *Google Open Building footprint*. Esse conjunto de dados contém 1,8 bilhão de *footprints* de edifícios em uma área de inferência de 58 milhões de km² que inclui regiões como: África, Sul da Ásia, Sudeste Asiático, América Latina e Caribe.

Para executar os testes em um tempo viável, os dados foram recortados por um quadrado de 1.000 km² ao redor de uma parte da região sul do estado de Minas Gerais, Brasil. A área de interesse é mostrada na Fig. 3.1.

Figura 3.1 – Mapa da cidade de Lavras, Minas Gerais, Brasil. Área de interesse, em vermelho, utilizada para recordar o *dataset* original.



Depois de extrair e recortar o conjunto de dados utilizado o *Google Earth Engine*, foi obtido um total de 106.628 feições únicas. Por fim, o *SQL Loader* foi utilizado para realizar a ingestão dos dados.

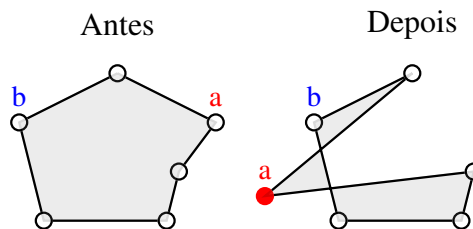
3.3.2.2 Geração de dados geométricos inválidos

Embora existam muitos conjuntos de dados geométricos gratuitos e abertos disponíveis na *internet* atualmente, quase não há conjuntos de dados contendo dados geométricos inválidos para fins de testes. Com isso em mente, uma ferramenta para deformar artificialmente geometrias válidas em inválidas foi criada exclusivamente para este artigo. Depois de analisar a extensa lista de códigos de erros do *Oracle* relacionados à validação de geometrias, três das condições mais comuns foram selecionadas para serem geradas por nossa ferramenta: 13349, "limite do polígono cruza a si mesmo"; 13356, "pontos adjacentes em uma geometria são redundantes"; 13367, "orientações inválidas para anéis internos e/ou externos".

Abaixo, estão os detalhes de implementação que descrevem como a ferramenta gera geometrias inválidas para cada código de erro.

- **13349:** Esse código de erro é causado por geometrias que se autointersec-tam. Esse comportamento é gerado artificialmente pela translação de um determinado vértice para fora dos limites da geometria original, sobrepondo a área existente. Conforme apresentado na Fig. 3.2, o vértice *a* foi transla-dado para a esquerda do vértice mais à esquerda da geometria (vértice *b*).

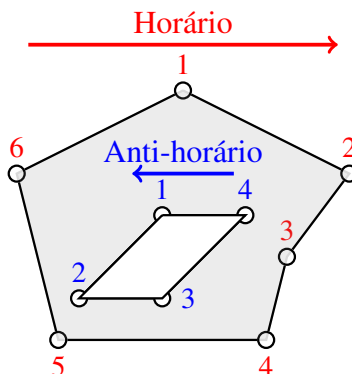
Figura 3.2 – Representação de uma geometria inválida contendo auto-intersecções.



- **13356:** Esse código de erro é causado por geometrias com pontos adjacentes redundantes. Esse comportamento é gerado artificialmente duplicando um vértice aleatório em uma geometria e colocando sua cópia após o vértice original.
- **13367:** Esse código de erro é causado por geometrias com anéis internos e/ou externos com orientações inválidas. Os vértices em geometrias válidas devem ser orientados no sentido anti-horário no anel mais externo enquanto seus furos devem ser declarados no sentido horário. A figura 3.3 representa uma geometria inválida devido à orientação invertida dos vértices, ou seja: o anel mais externo é definido no sentido horário e os furos estão definidos no sentido anti-horário.

Para gerar esse comportamento, a ferramenta simplesmente inverte a ordem dos vértices do anel mais externo em uma determinada geometria válida.

Figura 3.3 – Representação de uma geometria com orientação de vértices inválida.



3.3.3 Detalhes de implementação

Para abordar os impactos no desempenho e na precisão causados pelas geometrias inválidas nos conjuntos de dados apresentados acima (Tabela 3.1), foram desenvolvidas várias consultas SQL que buscam representar da melhor forma possível cenários de uso do mundo real. A execução de cada consulta foi iniciada e gerenciada por procedimentos armazenados (i.e. *stored procedures*) para atenuar quaisquer latências e imprecisões relacionadas à rede normalmente causadas pela execução em plataformas externas. Essas consultas são divididas em dois subgrupos: consultas para teste de desempenho e consultas para teste de precisão. A seguir, apresentamos uma explicação detalhada das consultas de cada grupo.

3.3.3.1 Consulta para testes de performance

Quatro consultas SQL distintas foram desenvolvidas para analisar os efeitos no desempenho causados por feições inválidas em um conjunto de dados. As consultas têm custos de execução estimados variados. Abaixo, um detalhamento de cada consulta.

- **Consulta 1:** Essa consulta simplesmente calcula a área total de cada feição no conjunto de dados, retornando a soma de todos os resultados em metros quadrados. **Custo estimado: 455**
- **Consulta 2:** Essa consulta simplesmente calcula o perímetro de cada feição no conjunto de dados, retornando a soma de todos os resultados em metros. **Custo estimado: 455**
- **Consulta 3:** Essa consulta calcula a área total de polígonos sobrepostos no conjunto de dados. Em outras palavras, para cada par de polígonos que se cruzam, a área de sobreposição será adicionada ao resultado final. **Custo estimado: 510.104**
- **Consulta 4:** Para um determinado polígono a no conjunto de dados, é criado um polígono de casco convexo b que abrange todos os polígonos no raio de 10 metros do polígono a . O resultado dessa consulta é a área total de cada casco convexo (b) somada. **Custo estimado: 510.327**

3.3.3.2 Consultas para testes de precisão

Diferentemente da seção anterior, apenas as geometrias com códigos de erro de validação 13356 (3.3.2.2) e 13367 (3.3.2.2) foram consideradas pois representam topologicamente a mesma superfície que suas contrapartes válidas. Por outro lado, as geometrias que apresentam o erro de validação 13349 (3.3.2.2) são topologicamente diferentes de seus originais e apresentarão variações na área total e no perímetro da geometria.

Um detalhamento de cada consulta é apresentado a seguir.

- **Consulta 5:**

Essa consulta compara a área de cada polígono no conjunto de dados de geometrias válidas com sua cópia inválida. A diferença entre os dois valores é retornada.

- **Consulta 6:**

Essa consulta compara o perímetro de cada polígono no conjunto de dados de geometria válidas com sua cópia inválida. A diferença entre os dois valores é retornada.

3.3.4 Solução proposta

A implementação da solução proposta nesta seção engloba três etapas principais, descritas a seguir.

1. Instalação da estrutura de validação principal

Inicialmente, os *scripts* de implantação criarão uma tabela de consulta para manter referências como nomes de tabelas, *IDs* e outras informações relevantes de cada feição no banco de dados que contenha geometrias inválidas. Essa tabela desempenha um papel fundamental na observação da integridade de cada classe de feição armazenada no banco de dados. É altamente recomendável implementar técnicas de controle de versão nessa tabela de negócios para executar relatórios sobre a integridade do banco de dados durante um determinado período de tempo. Em seguida, é implantada uma *stored procedure* responsável pela execução regular de tarefas de validação de geometrias em todo o banco de dados. Essa *procedure* interage com a tabela de consulta mencionada acima, reportando o status de cada feição que contenha geometrias inválidas. Essa *procedure* também fornece informações detalhadas sobre como corrigir possíveis problemas em uma determinada geometria.

Em um ambiente de banco de dados *Oracle* as informações necessárias para a *procedure* em questão podem ser obtidas através da função `VALIDATE_GEOMETRY_WITH_CONTEXT` (do pacote `SDO_GEOM`).

2. Criação de views

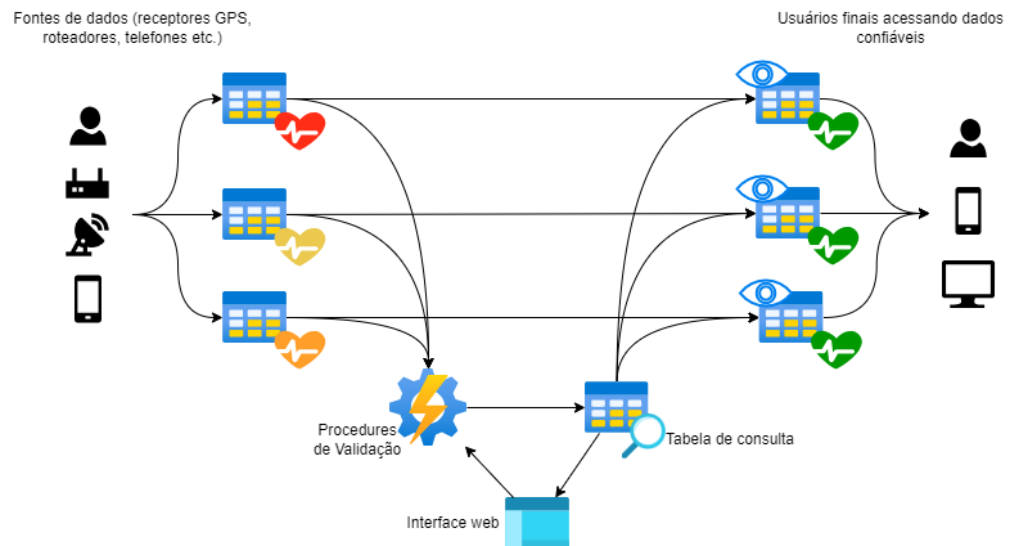
É criada uma *view* para filtrar geometrias inválidas para cada classe de feição no banco de dados. Essas *views* têm os mesmos campos que suas tabelas de origem. A filtragem das geometrias inválidas é realizada através de uma consulta simples na tabela de consulta, impedindo que as feições listadas com problemas apareçam na *view*. A partir desse ponto, os usuários desse banco de dados só devem consumir dessas visualizações protegendo-os de possíveis dados problemáticos.

3. Instalação dos componentes web

Por fim, é implantada uma interface *web* na qual o usuário tem controle total do processo. Por meio dessa interface o usuário deve ser capaz de analisar a integridade de cada classe de feição no banco de dados, executar tarefas de validação e analisar relatórios para cada feição inválida etc.

Um diagrama da arquitetura proposta é apresentado na Fig. 3.4.

Figura 3.4 – Os dados geométricos inválidos oriundos de várias fontes (à esquerda) são impedidos de serem acessados por usuários finais e aplicativos por meio do uso de *views* (à direita).



3.3.5 Simulação de dispositivos sem fio e IoT

Nesta subseção, integramos dispositivos de IoT e simulação sem fio em nossa metodologia para fornecer uma avaliação abrangente do desempenho do banco de dados espacial. Usando um ambiente de simulação adaptado para implantações de IoT, emulamos o comportamento dos dispositivos de IoT que interagem com bancos de dados espaciais em canais de comunicação sem fio. Ao modelar características realistas de dispositivos, topologias de rede e protocolos de comunicação, pretendemos avaliar a eficiência e a confiabilidade do banco de dados espacial no tratamento de consultas e transmissões de dados em cenários de IoT. Essa integração permite uma avaliação holística das soluções de banco de dados espaciais, garantindo sua adequação às implantações de IoT no mundo real.

3.4 Resultados e discussões experimentais

Nesta seção, serão apresentados os impactos das geometrias inválidas no desempenho e na precisão nos conjuntos de dados mencionados anteriormente (Tabela 3.1).

Na Fig. 3.5, os impactos das geometrias inválidas no desempenho das consultas são representados. A taxa de integridade de cada conjunto de dados, que representa a proporção das geometrias válidas, é refletida no eixo y, enquanto o tempo de execução de cada consulta é representado no eixo x. Por exemplo, na Consulta 1, à medida que a taxa de integridade do conjunto de dados diminui (passando de 100% para 20%), o tempo de execução tende a aumentar. Essa tendência é consistente em todas as consultas, indicando que os conjuntos de dados com baixa integridade resultam em tempos de execução de consultas mais longos. Essa relação ressalta o efeito prejudicial das geometrias inválidas sobre a eficiência das operações em um banco de dados espacial. Além disso, observamos diferentes graus de impacto em diferentes consultas. Por exemplo, na Consulta 4, o aumento no tempo de execução é mais acentuado em comparação com outras consultas à

medida que a taxa de integridade diminui. Isso sugere que as consultas mais complexas podem ser mais sensíveis à presença de geometrias inválidas, possivelmente devido ao grande número de operações complexas envolvidas.

Na Fig. 3.6, são mostrados os impactos das geometrias inválidas na precisão dos resultados das operações nos conjuntos de dados. O eixo y exibe os códigos de erro de validação estudados enquanto o eixo x mostra o valor médio de divergência em todo o conjunto de dados. É possível observar que os vértices duplicados (representados pelo código de erro 13356) não afetam a precisão geral de um determinado conjunto de dados. Entretanto, a ordem dos vértices em um recurso (representada pelo código de erro 13367) afeta a precisão do conjunto de dados. Considerando que o polígono médio nesse conjunto de dados tem uma área de $89,41\text{m}^2$, isso representa uma divergência média de 7,7% por polígono.

3.5 Considerações finais

Neste estudo, apresentamos uma metodologia abrangente para a criação de um ambiente de *benchmarking* confiável para avaliar os impactos das geometrias inválidas no desempenho de bancos de dados espaciais. Ao projetar meticulosamente o ambiente proposto e selecionar conjuntos de dados apropriados, nosso objetivo foi abordar os desafios associados à validação e retificação de dados espaciais. A seleção de conjuntos de dados de teste de variados níveis de integridade nos permitiu simular cenários do mundo real e avaliar o desempenho de bancos de dados espaciais em diferentes condições.

Como fonte de geometrias válidas, utilizamos o conjunto de dados abertos *Google Open Building footprint dataset*, garantindo feições acessíveis geradas por humanos. Para geometrias inválidas, devido a carência de conjuntos de dados com esta característica, desenvolvemos uma ferramenta personalizada capaz de deformar artificialmente geometrias válidas, replicando condições de erro encontradas comumente em conjuntos de dados espaciais.

Figura 3.5 – Desempenho médio para cada consulta e conjunto de dados disponível. O eixo y representa cada conjunto de dados por sua taxa de integridade. O eixo x representa o tempo de execução de cada consulta (em segundos).

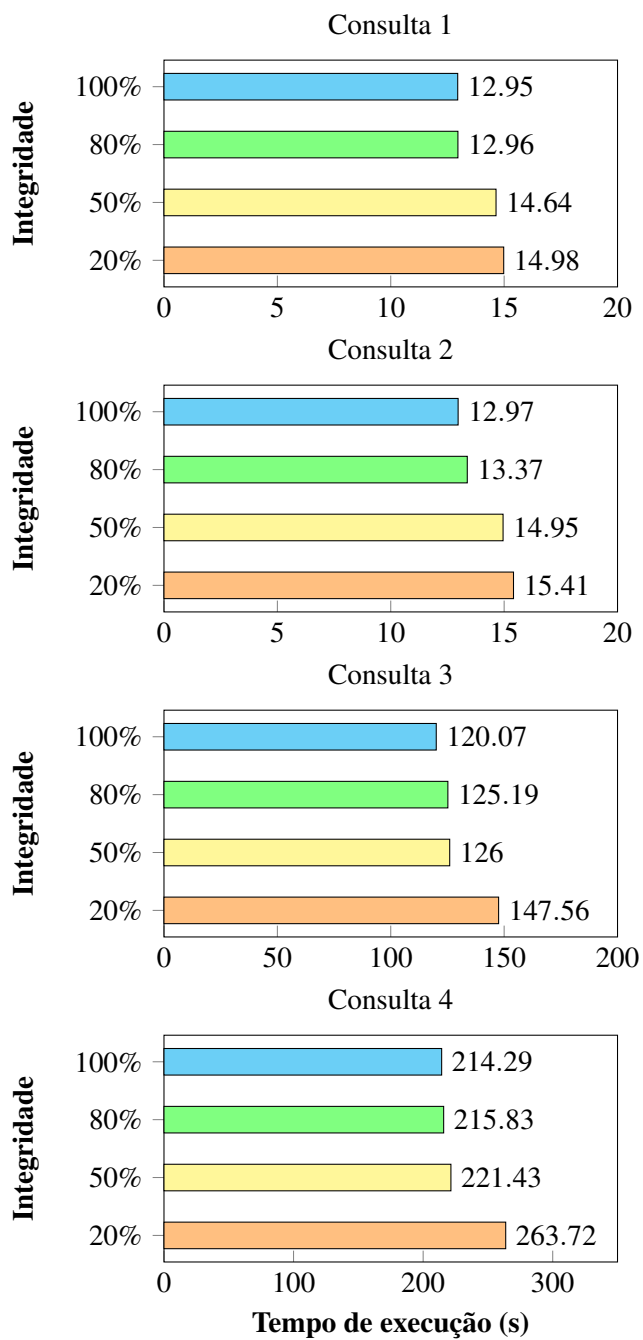
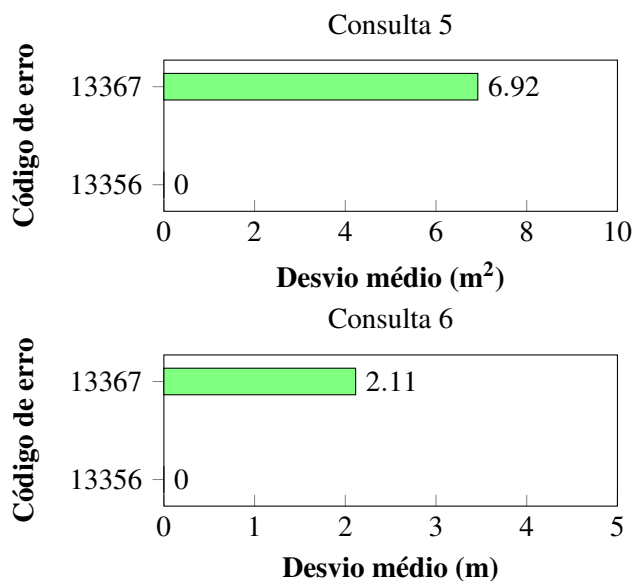


Figura 3.6 – Desvio médio dos valores originais para cada consulta. O eixo y representa cada código de erro analisado. O eixo x representa o valor total do desvio na unidade especificada.



Nossas descobertas destacam a importância de garantir a integridade dos dados espaciais especialmente no que tange a performance de consultas complexas e de longa duração. Além disso, evidenciamos os impactos causados na precisão dos dados retornados em consultas que envolvem cálculos de área e perímetro.

Pesquisas futuras podem incluir uma maior exploração do desenvolvimento de ferramentas automatizadas para detectar e corrigir geometrias inválidas de forma preventiva. Além disso, por meio de experimentação sistemática e análise de desempenho, avaliar o impacto das restrições da comunicação sem fio nas operações espaciais e identificar otimizações para aprimorar o desempenho geral do sistema em ambientes de IoT.

4 CONCLUSÃO

A pesquisa apresentada neste estudo destaca a importância da integridade dos dados espaciais no desempenho dos bancos de dados espaciais. A metodologia abrangente adotada para criar um ambiente de benchmarking confiável revelou que a presença de geometrias inválidas pode impactar significativamente a eficiência das operações em um banco de dados espacial. Os resultados mostraram que, à medida que a integridade dos conjuntos de dados diminui, o tempo de execução das consultas tende a aumentar, especialmente em consultas mais complexas.

Além disso, a análise da precisão dos resultados das operações nos conjuntos de dados destacou que certos tipos de geometrias inválidas, como a ordem dos vértices em um recurso, podem afetar a precisão dos dados retornados em consultas, indicando a necessidade de garantir a validade dos dados espaciais.

Para o futuro, sugere-se uma maior exploração no desenvolvimento de ferramentas automatizadas para detectar e corrigir geometrias inválidas de forma preventiva. Além disso, é importante realizar experimentos adicionais para avaliar o impacto das restrições da comunicação sem fio nas operações espaciais e identificar otimizações para melhorar o desempenho geral do sistema em ambientes de IoT. Essas melhorias podem contribuir significativamente para a eficiência e precisão dos bancos de dados espaciais em aplicações do mundo real.

REFERÊNCIAS

- ALAWAD, F.; KRAEMER, F. A. Value of information in wireless sensor network applications and the iot: A review. **IEEE Sensors Journal**, IEEE, v. 22, n. 10, p. 9228–9245, 2022.
- BADHRUDEEN, M. et al. The c2g framework to convert infrastructure data from computer-aided design (cad) to geographic information systems (gis). In: **MDPI. Informatics**. [S.l.], 2022. v. 9, p. 42.
- BOLSTAD, P. **GIS Fundamentals: A First Text on Geographic Information Systems**. [S.l.: s.n.], 2012.
- BOUKERCHE, A.; KANTARCI, B.; KAPTAN, C. Towards ensuring the reliability and dependability of vehicular crowd-sensing data in gps-less location tracking. **Pervasive and Mobile Computing**, Elsevier, v. 68, p. 101248, 2020.
- BRUNSDON, C.; COMBER, A. Opening practice: supporting reproducibility and critical spatial data science. **Journal of Geographical Systems**, v. 23, n. 4, p. 477–496, Oct 2021. ISSN 1435-5949. Disponível em: <<https://doi.org/10.1007/s10109-020-00334-2>>.
- CHEN, X.; LIU, X. quantitative analysis of urban spatial morphology based on gis regionalization and spatial syntax. **Journal of the Indian Society of Remote Sensing**, Springer, p. 1–10, 2022.
- DECURNINGE, A. et al. Csi-based outdoor localization for massive mimo: Experiments with a learning approach. In: IEEE. **2018 15th International Symposium on Wireless Communication Systems (ISWCS)**. [S.l.], 2018. p. 1–6.
- ELMASRI, R.; NAVATHE, S. B. **Fundamentals of Database Systems**. [S.l.]: Pearson, 2011.
- GOODCHILD, M. F. Twenty years of progress: Giscience in 2010. **Journal of Spatial Information Science**, 2010.
- JAIN, A. et al. Information retrieval using data mining. In: **Machine Vision and Augmented Intelligence: Select Proceedings of MAI 2022**. [S.l.]: Springer, 2023. p. 287–293.
- LI, X. et al. A unified framework for hs-uav noma networks: Performance analysis and location optimization. **IEEE Access**, IEEE, v. 8, p. 13329–13340, 2020.
- LONGLEY, P. et al. **Geographic Information Systems and Science**. [S.l.]: John Wiley & Sons, 2005.

LONGLEY, P. A. et al. **Geographic information science and systems**. [S.l.]: John Wiley & Sons, 2015.

MACHADO, V. A. et al. Uma abordagem para a garantia de qualidade em redes sem fio em malha. In: **Conferência Brasileira de Redes Sem Fio em Malha**. Rio de Janeiro, Brazil: [s.n.], 2012.

MAMOULIS, N. **Spatial Data Management**. Springer International Publishing, 2022. (Synthesis Lectures on Data Management). ISBN 9783031018848. Disponível em: <<https://books.google.com.br/books?id=gYdyEAAAQBAJ>>.

MAO, H. Urban planning management information integration platform system based on big data analysis. In: IEEE. **2023 2nd International Conference on Artificial Intelligence and Autonomous Robot Systems (AIARS)**. [S.l.], 2023. p. 540–544.

MAUS, V. et al. A global-scale data set of mining areas. **Scientific data**, Nature Publishing Group UK London, v. 7, n. 1, p. 289, 2020.

MENG, Y. et al. A data-driven method for constructing the spatial database of traditional villages—a case study of courtyard residential typologies in yunnan, china. **Buildings**, MDPI, v. 13, n. 12, p. 2956, 2023.

NAJAR, A.; MARQUES, E. **Saúde e Espaço: Estudos Metodológicos e Técnicas de Análise**. Editora da Fundação Oswaldo Cruz, 1998. ISBN 9788575412954. Disponível em: <<https://books.google.com.br/books?id=MGXRBAQAQBAJ>>.

NAKAMURA, F. G. **Planejamento Dinâmico para Controle de Cobertura e Conectividade em Redes de Sensores Sem Fio Planas**. Dissertação (Mestrado) — Universidade Federal de Minas Gerais, Belo Horizonte, dez. 2003.

OBEIDAT, H. et al. A review of indoor localization techniques and wireless technologies. **Wireless Personal Communications**, Springer, v. 119, p. 289–327, 2021.

RODRÍGUEZ, D. Z. et al. Speech quality assessment in wireless communications with mimo systems using a parametric model. **IEEE Access**, v. 7, p. 35719–35730, 2019.

SALLEH, S.; UJANG, U.; AZRI, S. Topology models and rules: a 3d spatial database approach. **The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, Copernicus Publications Göttingen, Germany, v. 48, p. 117–122, 2023.

SANTOS, L.; BRITO, T.; SILVA-NETO, C. Uso dos sistemas de informação geográficas (sig) nas ciências ambientais entre 2009 e 2019: uma análise cienciométrica. **Revista Brasileira de Geografia Física**, v. 15, p. 1715–1731, 07 2022.

SHARMA, A. et al. Database management systems—an efficient, effective, and augmented approach for organizations. In: SPRINGER. **ICT with Intelligent Applications: Proceedings of ICTIS 2021, Volume 1**. [S.l.], 2022. p. 465–478.

SILBERSCHATZ, A.; KORTH, H. F.; SUDARSHAN, S. **Database System Concepts**. [S.l.]: McGraw-Hill, 2020.

SILVA, M. d. **Planejamento de Redes em Malha Sem Fio com Atendimento a Parâmetros de Qualidade de Serviço**. Dissertação (Mestrado) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2011. Disponível em: <<http://urlib.net/8JMKD3MGP7W/39HAD9H>>.

SMITH, M. J. D.; GOODCHILD, M. F.; LONGLEY, P. **Geospatial analysis: a comprehensive guide**. [S.l.: s.n.], 2018.

STONEBRAKER, M. et al. Scidb: A database management system for applications with complex analytics. **Computing in Science & Engineering**, v. 15, n. 3, p. 54–62, 2007.

STONEBRAKER, M.; HELLERSTEIN, J. M. **Readings in Database Systems**. [S.l.]: MIT Press, 2015.

TEIXEIRA, I. L. da S. et al. Avaliação da qualidade de dados geoespaciais: uma abordagem moderna. **Anais do XVIII Simpósio Brasileiro de Sensoriamento Remoto - SBSR**, INPE Santos - SP, Brasil, p. 28–31, Maio 2017.

YADAV, P.; SHARMA, S. Unveiling the cutting edge: A comprehensive survey of localization techniques in wsn, leveraging optimization and machine learning approaches. **Wireless Personal Communications**, Springer, v. 132, n. 4, p. 2293–2362, 2023.