UNIVERSIDADE FEDERAL DE CAMPINA GRANDE

CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA

UNIDADE ACADÊMICA DE SISTEMAS E COMPUTAÇÃO

COORDENAÇÃO DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

GABRIEL JOSEPH RAMOS RAFAEL

BUSCA POR GRUPOS DE PONTOS DE INTERESSE USANDO PROCESSAMENTO QUALITATIVO DE REGIÕES ESPACIAIS

CAMPINA GRANDE - PB

2021

GABRIEL JOSEPH RAMOS RAFAEL

BUSCA POR GRUPOS DE PONTOS DE INTERESSE USANDO PROCESSAMENTO QUALITATIVO DE REGIÕES ESPACIAIS

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Campina Grande, pertencente à linha de pesquisa de Sistemas de Informação Geográfica, como requisito para a obtenção do Título de Mestre em Ciência da Computação.

Orientador(a): Prof. PhD. Cláudio Elízio Calazans Campelo

Co-orientador(a): Prof. Dr. Carlos Eduardo Santos Pires

CAMPINA GRANDE - PB

2021

R136b Rafael, Gabriel Joseph Ramos.

Busca por grupos de pontos de interesse usando processamento qualitativo de regiões espaciais / Gabriel Joseph Ramos Rafael. – Campina Grande, 2021.

77 f. : il. color.

Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática, 2021.

"Orientação: Prof. Dr. Cláudio Elízio Calazans Campelo, Prof. Dr. Carlos Eduardo Santos Pires".

Referências.

1. Sistemas de Informação Geográfica. 2. Recuperação da Informação. 3. Busca Espacial. 4. Relações Espaciais. I. Campelo, Cláudio Elízio Calazans. II. Pires, Carlos Eduardo Santos. III. Título.

CDU 004.78:025.4.036:911(043)

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECÁRIA SEVERINA SUELI DA SILVA OLIVEIRA CRB-15/225

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
POS-GRADUACAO CIENCIAS DA COMPUTACAO
Rua Aprigio Veloso, 882, - Bairro Universitario, Campina Grande/PB, CEP 58429-900

FOLHA DE ASSINATURA PARA TESES E DISSERTAÇÕES

GABRIEL JOSEPH RAMOS RAFAEL

BUSCA POR GRUPOS DE PONTOS DE INTERESSE USANDO PROCESSAMENTO QUALITATIVO DE REGIÕES ESPACIAIS

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação como pré-requisito para obtenção do título de Mestre em Ciência da Computação.

Aprovada em: 19/08/2021

Prof. Dr. CLÁUDIO ELÍZIO CALAZANS CAMPELO, Orientador, UFCG

Prof. Dr. CARLOS EDUARDO SANTOS PIRES, Orientador, UFCG

Prof. Dr. NAZARENO FERREIRA DE ANDRADE, Examinador Interno, UFCG

Prof. Dr. YURI ALMEIDA LACERDA, Examinador Externo, IFCE

Documento assinado eletronicamente por **CARLOS EDUARDO SANTOS PIRES**, **PROFESSOR 3 GRAU**, em 23/09/2021, às 09:52, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da Portaria SEI nº 002, de 25 de outubro de 2018.

Documento assinado eletronicamente por **CLAUDIO ELIZIO CALAZANS CAMPELO**, **PROFESSOR(A) DO MAGISTERIO SUPERIOR**, em 23/09/2021, às 10:14, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da Portaria SEI nº 002, de 25 de outubro de 2018.

Documento assinado eletronicamente por **NAZARENO FERREIRA DE ANDRADE**, **PROFESSOR(A) DO MAGISTERIO SUPERIOR**, em 23/09/2021, às 16:51, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da Portaria SEI nº 002, de 25 de outubro de 2018.

Documento assinado eletronicamente por **Yuri Almeida Lacerda**, **Usuário Externo**, em 27/09/2021, às 13:53, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da Portaria SEI nº 002, de 25 de outubro de 2018.

A autenticidade deste documento pode ser conferida no site https://sei.ufcg.edu.br/autenticidade, informando o código verificador **1791081** e o código CRC **BFBFE91E**.

Referência: Processo nº 23096.050674/2021-71 SEI nº 1791081

Resumo

Para minimizar dificuldades de locomoção, criar roteiros de viagens ou economizar tempo, as pessoas comumente se deparam com a necessidade de encontrar Pontos de Interesse (POI) que compartilhem a mesma extensão espacial ou estejam localizados em regiões interconectadas. As buscas por POI usando ferramentas web se concentram exclusivamente em consultas por um único tipo de estabelecimento (e.g. restaurante ou hotel) ou por palavras-chave que se referem ao nome de um local (e.g. starbucks ou subway). A recuperação de um grupo de lugares usando palavras-chave e relações de conectividade entre suas regiões é um desafio atual para ferramentas de busca, pois não consideram a representação do POI como uma região, mas como um ponto no espaço. As principais soluções existentes baseiam-se apenas no cálculo da distância entre estes pontos. Poucas são capazes de avaliar as relações de conectividade entre as extensões espaciais dos POI. Neste contexto, este trabalho propõe uma técnica de busca textual por grupo de POI, baseada nas relações qualitativas entre regiões espaciais. Com a técnica, é possível, por exemplo, encontrar estabelecimentos de diferentes tipos que são vizinhos ou estão localizados no mesmo prédio. A solução, denominada Topo-MSJ, define um padrão de consultas espaciais qualitativas, utilizando a combinação de um algoritmo do estado-da-arte, o "Multi-Star-Join" (MSJ), juntamente com um modelo espacial de relações qualitativas, denominado "Region Connection Calculus" (RCC). O Topo-MSJ, em uma única consulta, pode explorar até quatro tipos de relações espaciais de conectividade diferentes, sendo sobretudo adequado ao cenário de *Big Spatial Data*. A eficiência do algoritmo proposto é avaliada através de uma comparação com outros trabalhos que utilizam soluções de indexação qualitativa, além de uma avaliação comparativa das consultas em formato SQL. As bases utilizadas na avaliação experimental incluem aproximadamente 900 mil POI dos estados americanos da Califórnia e Nova Iorque, além de bases de dados textuais e geográficos da Agência Ambiental Europeia (AAE), utilizadas pelos trabalhos de indexação qualitativa comparados a esta pesquisa. Os resultados experimentais apontam que o algoritmo proposto é mais eficiente, em tempo de execução, do que consultas SQL realizadas em bancos de dados espaciais. Além disso, é mostrado que, mesmo possibilitando a realização de consultas de maior complexidade, é possível obter um tempo similar ao das soluções de indexação qualitativa.

Palavras-chave: Busca Espacial, Relações Espaciais, Sistemas de Informação Geográfica, Recuperação da Informação.

Abstract

In order to decrease mobility difficulties, create travel itineraries or save time, people usually face the need to find Points of Interest (POI) that share the same spatial extent or are located in interconnected regions. POI searches using web tools focus exclusively on queries for a single type of establishment (e.g., restaurant or hotel) or for keywords referring to a place's name (e.g., Starbucks or Subway). Retrieving a group of places by using keywords and connectivity relationships between their regions is a current challenge for search tools, as they do not consider POI's representation as a region, but as a point in space. The main existing solutions are based only on the distance's calculation between these points. However, few tools are able to assess the connectivity relationships between POIs' spatial extensions. In this context, the present study proposes a textual search technique for a group of POIs, based on the qualitative relationships between spatial regions. With the technique, it is possible, for example, to find different types of establishments that are neighbors or are located in the same building. The solution, named Topo-MSJ, defines a pattern of qualitative spatial queries by using the combination of a state-of-the-art algorithm, the "Multi-Star-Join" (MSJ), along with a spatial model of qualitative relationships, entitled "Region Connection Calculus" (RCC). Topo-MSJ, in a single query, retrieves up to four different types of spatial connectivity relationships and is particularly suited to the Big Spatial Data scenario. The algorithm's efficiency is evaluated through the proposed solution's comparison with other works that use qualitative indexing solutions, in addition to a comparative evaluation of the queries in SQL format. The databases used in the experimental evaluation include approximately 900,000 POIs from the American states of California and New York, as well as textual and geographic databases from the European Environment Agency (EEA), which are used by the qualitative indexing works compared to this research. The experimental results indicate that the proposed algorithm is more efficient (in terms of execution time) than SQL queries performed on spatial databases. Furthermore, it is shown that even allowing the execution of more complex queries, it is possible to achieve similar execution times compared to other existing qualitative indexing solutions.

Keywords: Spatial Search, Spatial Relations, Geographic Information Systems, Information Retrieval.

Agradecimentos

A minha mãe Kalina, uma grande mulher, um exemplo de coragem e força na minha história.

Ao meu irmão Carlos, irmão de sangue e de fé, meu melhor amigo, você sempre será um guia pra mim.

A minha esposa Maria Beatriz, por sua presença, companheirismo e amor em minha vida.

Aos meus orientadores, Cláudio Campelo e Carlos Eduardo, pela paciência e por todo aprendizado transmitido desde o começo, agradeço imensamente todo o apoio e incentivos prestados.

Conteúdo

Lista de Símbolos

9IM - *9-Intersection Model*

AID - *Average Intersection Degree*

AOI - *Area of Interest*

CCEQ - *Consulta de Configuração Espacial Qualitativa*

CD - *Cardinal Direction*

CEQ - *Cálculo Espacial Qualitativo*

GPM - *Graph Pattern Matching*

GPS - *Global Positioning System*

IGV - *Informações Geográficas Voluntárias*

ISO - *International Organization for Standardization*

MBR - *Minimum Boundig Rectangle*

MPJ - *Multi-Pair-Join*

MSJ - *Multi-Star-Join*

OGC - *Open Geospatial Consortium*

PEQ - *Padrão Espacial Qualitativo*

POI - *Ponto de Interesse*

RCC - *Region Connection Calculus*

RI - *Recuperação da Informação*

RQE - *Raciocínio Qualitativo Espacial*

RRQ - *Rede de Restrição Qualitativa*

SBL - *Serviços Baseados em Localização*

SGBD - *Sistema de Gerenciamento de Banco de Dados*

SIG - *Sistemas de Informação Geográfica*

SQL - *Structured Query Language*

UML - *Unified Modeling Language*

Lista de Figuras

Lista de Tabelas

Lista de Códigos Fonte

Capítulo 1

Introdução

O ser humano em todos os momentos da história necessitou se localizar no ambiente em que vive, assim como descobrir os lugares que contribuíam para a manutenção de sua existência. Na antiguidade, os locais que concentravam comércios, instituições religiosas ou fontes de recursos naturais eram objeto de registro em sistemas cartográficos.

Um Ponto de Interesse (em inglês, *Point of Interest* - POI) pode ser definido como todo local que possui alguma utilidade ou importância para um determinado grupo de pessoas. Por exemplo, igrejas, escolas e supermercados. Trata-se de um entendimento humano, associado ao lazer ou à utilidade. Para atender as necessidades dos dias atuais, uma maior quantidade de POI estão surgindo. Locais como bares, academias e farmácias são exemplos destes pontos relevantes para o ser humano [31]. A busca por novas formas de armazenar e pesquisar esses tipos de lugares ajudou a promover o avanço das geotecnologias.

Com o intuito de criar uma representação para os POI, diferentes tipos de dados são utilizados por novas tecnologias. Os tipos que possuem forma geométrica associada à posição geográfica, também chamados de geometrias, podem ser caracterizados em diversos formatos, tais como: pontos, linhas e polígonos. Um ponto é definido como uma coordenada geográfica no espaço; neste caso, um valor de latitude e longitude no planeta. Usualmente a representação por ponto é a mais comum, porém a menos precisa, por utilizar apenas o centroide ou ponto central de uma área como representação de um POI. Pode-se entender a linha como uma sequência de pontos que comumente pode simbolizar ruas ou divisões político-administrativas. O polígono especificamente representa a extensão espacial do POI, contém a região do objeto e usualmente simboliza o prédio ou edifício em que o POI está situado.

Algumas ferramentas como GoogleMaps[1], OpenStreetMap[2] e Foursquare[3] manipulam esses tipos de geometrias e permitem a realização de buscas a partir de outros atributos pertencentes ao POI, como descrições detalhadas, horários de funcionamento e avaliações de qualidade do serviço. As bases de dados da maioria desses serviços são enriquecidas pelos próprios usuários, que criam os POI e inserem grande parte das informações em formato textual. Na Figura 1.1, é possível visualizar a interface de busca da ferramenta Foursquare.

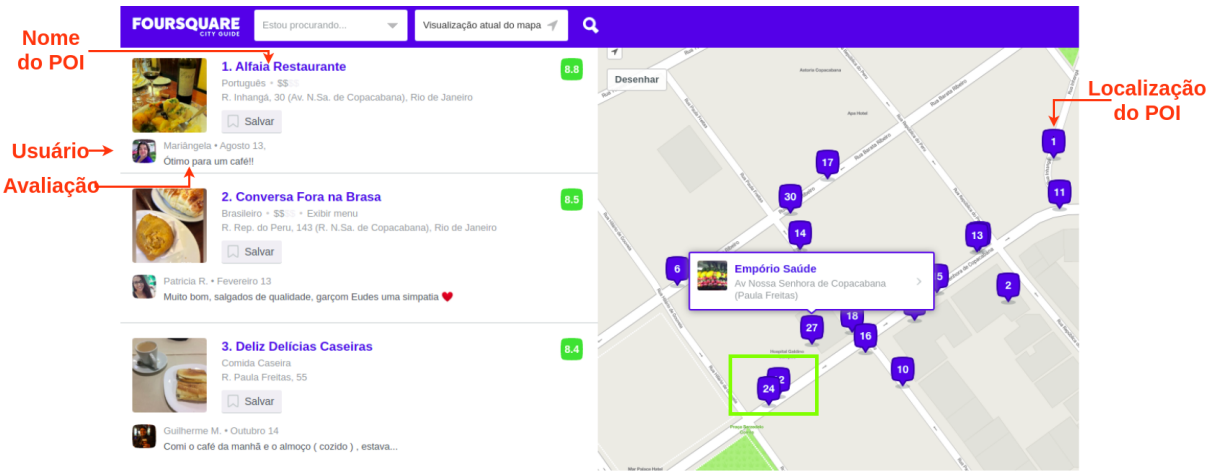


Figura 1.1: Resultado da busca por POI na ferramenta Foursquare.

A busca ilustrada na Figura 1.1 foi realizada utilizando a palavra-chave “restaurante” como parâmetro de entrada e a extensão espacial da cidade do Rio de Janeiro foi definida como espaço de busca. Um aspecto a ser observado é que um único POI pode concentrar um conjunto variado de informações, como: nome, endereço e comentário do usuário. Dessa forma, observa-se a presença de dados no formato textual em vários elementos na busca por POI.

Além disso, é possível observar na Figura 1.1 a presença de uma lista de POI que estão próximos uns aos outros; em alguns casos são vizinhos, como os números 24 e 22, destacados por um retângulo verde. A busca por POI que se encontram vizinhos poderia ser feita em linguagem natural da seguinte forma: “busque POI que sejam do tipo restaurante ou possuam a palavra restaurante em sua descrição e compartilhem alguma fronteira de sua extensão espacial”. Não foi localizado na literatura nenhum sistema que permita este tipo de busca por relações de conectividade, obrigando o usuário a criar múltiplas consultas espaciais ou

[1]https://www.google.com.br/maps/
[2]https://www.openstreetmap.org/
[3]https://foursquare.com/

privando-o desta informação no momento da pesquisa.

1.1 Motivação

Com o surgimento do conceito de Serviços Baseados em Localização (SBL), a necessidade de realizar pesquisas por POI aumentou amplamente na área de Sistemas de Informações Geográficas (SIG). As pessoas usualmente realizam buscas por esse tipo de local utilizando palavras-chave, que indicam o nome do estabelecimento ou a categoria a qual ele pertence. Em alguns casos, existe a necessidade de pesquisar um grupo de objetos espaciais em uma única consulta.

Por exemplo, vamos considerar que uma pessoa deseja: i) comprar um livro, ii) visitar um museu e iii) fazer uma refeição no mesmo passeio. Se esses lugares estiverem localizados em edifícios diferentes ou forem de difícil acesso a pé, o usuário terá que lidar com perda de tempo por deslocamento, procurar estacionamentos e/ou enfrentar problemas de acessibilidade. Esses problemas podem ser resolvidos encontrando POI que possuam edifícios conectados ou localizados no mesmo prédio. Em outras palavras, em algumas situações existe a necessidade de encontrar um grupo de POI que possua uma distribuição específica no espaço, com relações de conexão definidas entre estes itens, tais como relações de interseção ou compartilhamento da mesma região espacial.

Em geral, as pesquisas por locais em ferramentas na web se concentram exclusivamente em consultas com um único valor textual; neste caso, uma palavra-chave que se refere ao tipo de POI (e.g. Restaurante, Dentista e Hotel). Recuperar um grupo de lugares com palavras associadas e relações de conectividade entre suas regiões é um desafio atual para ferramentas de pesquisas em SIG. Essencialmente, estas ferramentas não criam uma representação de POI como polígonos, mas apenas como pontos no espaço, permitindo apenas uma comparação por distância entre esses locais.

A solução proposta nesta dissertação pode auxiliar no desenvolvimento de ferramentas para recomendação de lugares, por exemplo, para fins turísticos. Para esta aplicação no turismo, é importante a utilização do conceito de *Area of Interest* (AOI), que pode ser definido como uma área urbana, ao nível da cidade ou bairro, com uma alta concentração de POI. Comumente, a AOI é localizada ao longo de uma região de grande importância espacial. A busca por regiões do tipo AOI pode ser melhorada utilizando aspectos qualitativos, principalmente na avaliação da concentração entre os tipos de POI procurados.

Um outro campo de possível aplicação desta solução é na geração de roteiros automatizados, tendo em vista que alguns trabalhos desenvolvidos na área, como em [33], apresentam conceitos de pacotes de viagens. Os pacotes de viagem são definidos, na área de busca espacial, como um conjunto de objetos pertencentes a uma região e compõem um itinerário de viagem para um turista. Estes pacotes são criados de forma automática por sistemas de busca, baseados no histórico de preferências do usuário. Os tipos de POI buscados pelas soluções propostas por estes trabalhos são distintos e pesquisados na forma de grupo de objetos espaciais, deste modo, semelhantes ao tipo de busca utilizada na solução proposta por esta dissertação.

Para realizar uma busca contendo regras de conectividade entre regiões, é importante definir uma representação de cenários espaciais, ou cenas espaciais, como apresentado em [6]. Uma cena espacial é um conjunto de objetos geográficos, juntamente com suas relações espaciais (conectividade, distância e/ou direção) e opcionalmente outros tipos de características espaciais, a fim de criar uma estrutura de consulta que corresponda à expectativa do usuário. Por exemplo, um museu que possui um restaurante dentro de seu prédio e esteja vizinho a uma livraria pode ser um bom resultado para o problema de busca supracitado. A Figura 1.2 representa a visualização deste resultado.

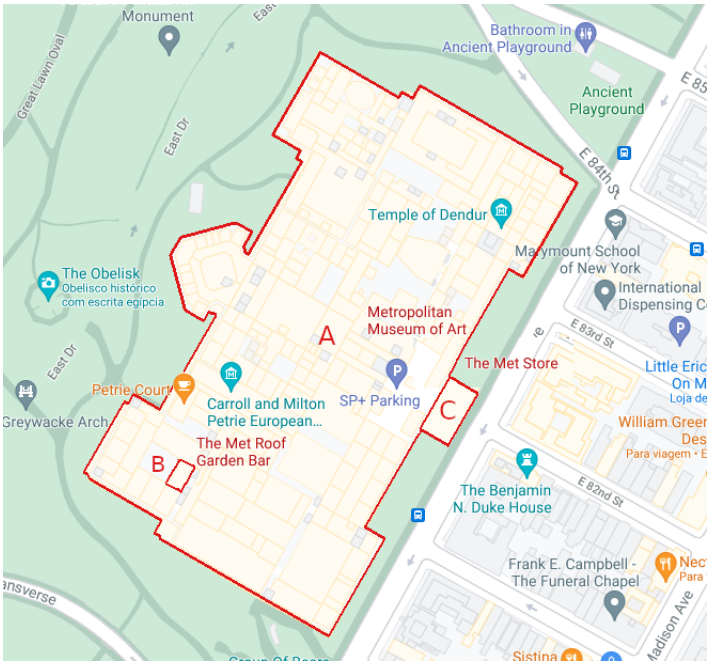


Figura 1.2: Visualização de POI e suas respectivas geometrias na ferramenta GoogleMaps. A: Museu, B: Restaurante e C: Livraria.

Na Figura 1.2, cada POI é identificado por um polígono na cor vermelha, representando a área correspondente ao seu edifício. O polígono maior que possui a letra A em seu centro caracteriza o museu. Simbolizado pela letra B, o restaurante se localiza no interior do museu, em um polígono de tamanho menor. A letra C indica a livraria, localizada vizinha ao museu com suas respectivas geometrias se tocando. A descoberta dessa distribuição de POI no mapa só é possível através de uma busca que utilize, no seu processamento, as relações de conectividade entre as regiões dos POI.

1.2 Objetivos

Visando a melhoria dos sistemas de busca por POI que utilizam o processamento de relações espaciais em suas consultas, o objetivo geral deste trabalho é propor uma técnica de busca por grupos de POI, baseada em parâmetros textuais e em relações espaciais qualitativas.

Pretendendo alcançar este objetivo geral, os seguintes objetivos específicos são estabelecidos:

- Definir um modelo de consulta que representa padrões de conectividade entre regiões espaciais utilizando palavras-chave;
- Utilizar regiões espaciais como representação dos Pontos de Interesse no processamento do algoritmo;
- Evoluir um algoritmo de busca e indexação de regiões espaciais a fim de permitir o uso de funções de conectividade;
- Avaliar o desempenho do algoritmo desenvolvido, comparando com abordagens que propõem a indexação de relações qualitativas entre regiões.

Fundamentado nesse propósito, propõe-se a seguinte hipótese geral a ser considerada: *a combinação de técnicas de consultas por palavras-chave espaciais com um modelo de conectividade de regiões permite a busca por padrões qualitativos em consultas por grupos de objetos espaciais*.

1.3 Relevância

A solução proposta por este trabalho permite a aplicação em vários cenários do mundo real, essencialmente onde a busca por POI através de consultas textuais se faz necessária. É

possível destacar o uso em áreas de planejamento urbano, SIG e detecção de anomalias em bases de dados.

O conceito de Informação Geográfica Voluntária (IGV) se refere a todo conteúdo que possui dados geoespaciais, em sua maioria, criado por não profissionais e de forma colaborativa [44]. Esse tipo de informação possui campos que comumente necessitam ser classificados e validados, uma vez que qualquer usuário pode atribuir o valor que julgar correto para um determinado dado.

No contexto de mapas, algumas geometrias podem receber o rótulo de um estabelecimento ou de uma vegetação de forma incorreta, o meio de detecção destas anomalias pode ser através de restrições pré-determinadas, algumas vezes em aspectos qualitativos, como em [2; 3]. Exemplos: uma região classificada como um rio não deve cruzar a geometria de um hospital; uma escola usualmente não deve possuir um posto de gasolina no seu interior. Todos esses casos de violação de restrições podem ser encontrados por meio de consultas envolvendo relações qualitativa entre POI.

No cenário de pandemia da doença Coronavírus 2019 (COVID-19), alguns estudos como [10] estão avaliando a disseminação do vírus em humanos que visitaram determinados POI. O objetivo é informar qual o índice de uma possível contaminação baseado-se no tipo de estabelecimento ou local de importância que a população frequenta. Na Figura 1.3, é possível observar, através dos gráficos, qual a proporção de infecção diária em determinados POI de quatro cidades americanas.

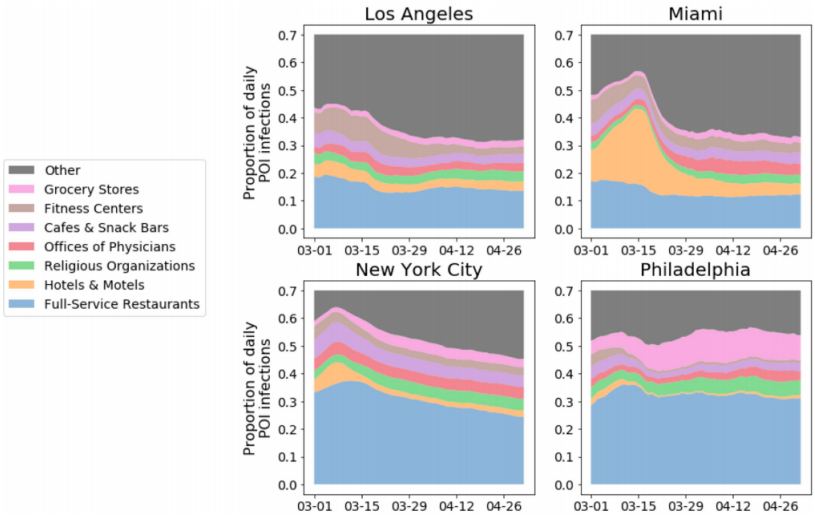


Figura 1.3: Proporções diárias de infecções por COVID-19 em humanos de acordo com o POI. Fonte: Mobility network models of COVID-19 explain inequities and inform reopening, p.15.

A solução proposta nesta pesquisa pode ser usada como detector de áreas de alto risco de infecção, ao identificar as regiões que possuem um ou mais POI no mesmo prédio ou com grande proximidade entre seus edifícios, considerando esse tipo de conexão como um indicador severo de possibilidade de aglomeração.

Uma pesquisa por cena espacial pode ajudar na área de planejamento urbano a encontrar padrões em centros comerciais como em [7], auxiliando a identificar quais tipos de estabelecimentos se concentram em determinadas cidades ou regiões, além de permitir a aplicação de algum recurso público ou mecanismo de fiscalização, dependendo do tipo de POI. Por exemplo, a sede de uma agência de vigilância sanitária pode ser instalada próximo a locais que oferecem serviços de saúde ou a produção de alimentos.

Por último, para um usuário comum que possui alguma limitação de mobilidade ou precisa economizar tempo em uma viagem, a busca por locais que possuam uma específica distribuição de POI também pode ser útil. A realização de uma consulta por POI que compartilham a mesma geometria, por exemplo, pode permitir a descoberta por locais que estejam em um mesmo prédio ou edifício.

1.4 Contribuições

Neste trabalho, é apresentada uma técnica para realização de consultas por grupos de POI, capaz de avaliar as relações de conexão entre suas extensões espaciais, ou seja, a busca deve considerar o modo como os polígonos que representam cada estabelecimento se conectam.

Propõe-se combinar uma abordagem que utiliza consultas por POI baseadas palavras-chave, utilizando um algoritmo do estado-da-arte denominado Multi-Star-Join (MSJ), desenvolvido em [23], juntamente com um modelo de relações qualitativas chamado de Region Connection Calculus (RCC) [39]. A solução utiliza uma árvore de índice invertido, uma IR-Tree [32], para a indexação das consultas qualitativas.

Desta forma, as principais contribuições deste trabalho são:

- Um modelo para consultas espaciais chamado Padrão Espacial Qualitativo (PEQ), que define as relações de conectividade entre as regiões espaciais;
- Um algoritmo que realiza consultas espaciais utilizando palavras-chave e relações de conectividade espacial;
- Uma ferramenta de consulta textual e visualização em mapas das áreas relativas aos POI.

A solução proposta faz uso de um modelo de consulta textual e um algoritmo que permite realizar a busca com tempo satisfatório em relação as soluções existentes e mais eficiente, em tempo de execução, do que consultas SQL em bancos de dados espaciais. A ferramenta de busca espacial desenvolvida utilizando a solução proposta por este trabalho é chamada de Topokey. As capturas de telas e exemplos de consultas realizadas utilizando a ferramenta podem ser encontradas no Apêndice E deste documento.

1.5 Histórico da Pesquisa

As questões iniciais levantadas nesta pesquisa se concentravam na busca por soluções para o problema de similaridade em grupos de regiões espaciais. O objetivo foi identificar diferentes regiões que possuíam a mesma distribuição de itens com atributos heterogêneos. As soluções pertencentes à área de cenas espaciais foram as primeiras a serem avaliadas.

Neste contexto de cenas espaciais, os atributos das regiões eram definidos por alguma característica natural, como vegetação ou um corpo aquático. A busca por bases de dados

que possuíam esse tipo de dado não encontrou nenhuma quantidade substancial de regiões, nem a aplicabilidade para este tipo de consulta se mostrou relevante em alguma área. A partir deste momento, o uso de POI foi avaliado como uma possível alternativa de consulta, principalmente para aplicações no campo do turismo ou soluções para problemas de acessibilidade.

Foi possível observar que o método de entrada para realização de consultas espaciais qualitativas também deveria ser levado em consideração na pesquisa. Alguns trabalhos que usam desenhos em forma de *sketches*, como em [43], também foram analisados como possíveis soluções. Somente após a escolha do uso de POI como objeto de consulta, observou-se que os aspectos textuais eram de grande importância na consulta deste tipo de dado. Neste momento, surgiu a necessidade de criação do PEQ, a fim de considerar as palavras-chave como método de entrada na busca.

A área de *matching* de dados espaciais foi explorada em trabalhos como [47]. Este tipo de informação foi utilizado como fundamento para o entendimento de comparações entre regiões espaciais. A área de cálculo espacial qualitativo também foi explorada e os conceitos aplicados nesta pesquisa foram definidos dentre os trabalhos existentes nesta área.

Um ponto de importância no processo de amadurecimento na implementação da solução foi a descoberta da base de dados. Os dados foram disponibilizados pelo provedor durante um contexto da pandemia do COVID-19, e foram decisivos no uso experimental do algoritmo. O tamanho considerável da base permitiu avaliar cenários relevantes em relação ao processamento de consultas espaciais.

A busca por um algoritmo que permitisse a busca textual de dados espaciais iniciou-se com o conceito de *travel packages* e *composite itens*. Esperou-se utilizar métodos de agrupamento, como o *K-means*, para criação dos grupos de objetos espaciais. Ao término destes estudos, observou-se que não existiam trabalhos que permitissem o uso de regiões para a avaliação de relações qualitativas.

Por fim, a investigação dos métodos textuais permitiu a busca por soluções que utilizavam palavras-chave espaciais, permitindo que o algoritmo de referência para o Topo-MSJ fosse encontrado e utilizado na implementação da solução proposta.

1.6 Organização do Trabalho

A estrutura deste documento está organizada como segue: no Capítulo 2 são apresentados os conceitos fundamentais e o arcabouço teórico necessário à compreensão deste trabalho, como as definições de relações qualitativas entre itens espaciais e as consultas coletivas por palavras-chave. O Capítulo 3 exibe os trabalhos relacionados a esta pesquisa. No Capítulo 4, é apresentada a solução proposta, com detalhamento do algoritmo e implementação técnica. O Capítulo 5 descreve o processo de avaliação experimental e os resultados obtidos. No Capítulo 6, as conclusões da pesquisa e as perspectivas para trabalhos futuros são discutidas.

Capítulo 2

Fundamentação Teórica

Neste capítulo, são apresentados os principais conceitos que fundamentam o desenvolvimento desta pesquisa. Na Seção 2.1, são apresentados os formatos dos dados que representam as entidades espaciais, especificamente os POI. A Seção 2.2 define os conceitos de cálculo espacial entre regiões geográficas, assim como as funções utilizadas nesse processamento. Na Seção 2.3, as consultas espaciais qualitativas são apresentadas. A Seção 2.4 descreve o problema de satisfação de restrições. Na Seção 2.5, as consultas por palavras-chave são abordadas e exemplificadas de acordo com suas estruturas de indexação. Por fim, na Seção 2.7, são feitas as considerações finais acerca do arcabouço teórico desta pesquisa.

2.1 Representação Espacial

O uso de novas tecnologias que possuem o recurso de *Global Positioning System* (GPS) fez com que smartphones, tablets e outros dispositivos permitissem o acesso a mapas que antes existiam apenas em formato físico. Ao mesmo tempo em que a leitura e escrita de conteúdos textuais nestes dispositivos foi aumentando, o uso de palavras-chave espaciais surgiu para a identificação de lugares.

Como definido em [36], o tipo de dado chamado "palavra-chave espacial" é representado como um objeto o com o seguinte formato: $o = [oid, loc, text]$. Nessa definição, o oid representa uma chave única de identificação, o loc simboliza o atributo de localização, geralmente sua coordenada geográfica. Representado pelo atributo $text$, o dado textual pode corresponder a uma ou várias palavras associadas à entidade, representando o nome ou o tipo do objeto. A Figura 2.1 possui uma conjunto de POI distribuídos em um espaço bidimensional com palavras associadas ao seu tipo de estabelecimento.

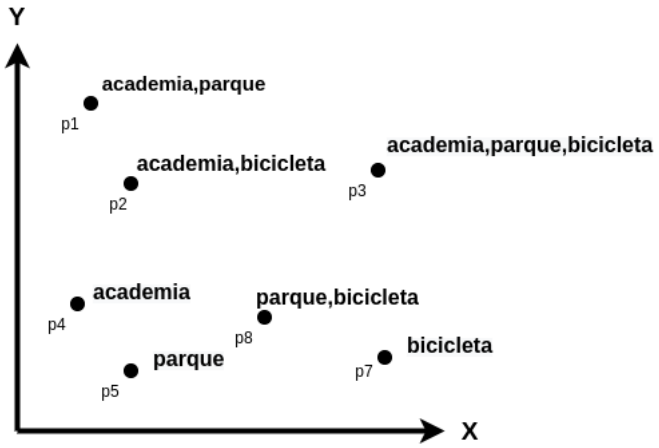


Figura 2.1: Palavras-chave espaciais distribuídas em duas dimensões.

Considere o cenário em que um turista está visitando uma nova cidade. Ele(a) deseja ir a um parque que possua uma área com equipamentos semelhantes aos de uma academia, mas não busca praticar nenhuma atividade relativa a ciclismo. De acordo com a Figura 2.1, os POI p1, p4 e p5 representam os locais mais adequados ao tipo de necessidade existente, pois apresentam as palavras mais correlatas à sua preferência. Para resolução desse problema, a pesquisa por *top-k* palavras-chave é necessária para alcançar um resultado mais próximo da expectativa da consulta.

A representação de um POI em um sistema SIG usualmente possui uma descrição textual e uma coordenada geográfica. Os valores textuais, neste caso, podem auxiliar as ferramentas de buscas no processo de indexação e de recuperação da informação. Os dados geográficos, antes do uso em ferramentas de buscas na web, já eram utilizados nas ferramentas de SIG. Neste período, surgiu a necessidade de criar uma metodologia, como apresentada em [45], para o mapeamento de determinadas áreas e o gerenciamento no meio digital destas informações.

No contexto de SIG, os dados geográficos ganharam representações de acordo com a evolução das tecnologias de armazenamento e consulta. Os sistemas atuais utilizam dados padronizados pelo *Open Geospatial Consortium (OGC)*[1] e pela *International Organization for Standardization (ISO)*[2]. Os padrões definidos por estas entidades para geometrias bidimensionais são: ponto, curva e superfície.

Na Figura 2.2, é possível observar um diagrama no formato UML (*Unified Modeling Language*), o qual representa a hierarquia de classes no contexto de geometrias para dados

[1]https://www.ogc.org/

[2]https://www.iso.org/

espaciais. No topo do diagrama, é possível observar que o ponto, a curva e a superfície possuem uma relação de herança com a classe geometria. Os demais tipos são derivados dessas estruturas, assim como as coleções de entidades, acompanhadas usualmente pela palavra 'Multi', como o MultiPolígono ou o MultiPonto.

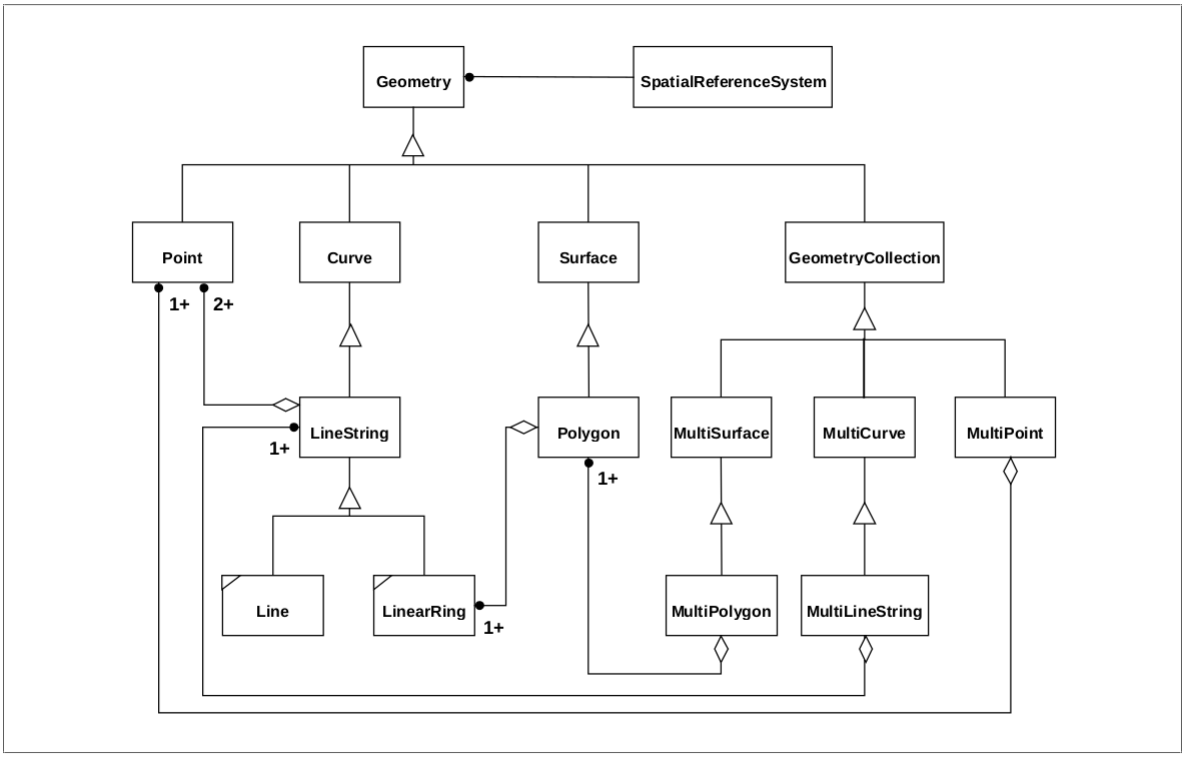


Figura 2.2: Hierarquia de classes de geometrias (OGC,1999).

Um ponto é comumente representando e armazenado pelas coordenadas de latitude e longitude. No início do uso de SIG, este formato era o mais empregado para representar POI [28]. A linha ou curva pode ser entendida como um segmento de reta que usualmente define uma rua ou fronteira no contexto de SIG. O polígono, como apresentado anteriormente, representa a extensão espacial de uma determinada entidade. Utiliza-se nesta pesquisa especificamente a geometria do tipo polígono para representação dos POI.

2.2 Cálculo Espacial

O raciocínio espacial de acordo com Cohn [39] consiste em derivar novos conhecimentos de informações fornecidas, verificar a consistência destas informações, atualizar o conhecimento ou encontrar uma representação mínima para atributos espaciais. O conceito de

relação é um dos mais fundamentais na área de raciocínio espacial, e está diretamente associado à ideia de conexão entre objetos no espaço, comumente definida como um ambiente de duas dimensões.

O cálculo de relacionamentos entre objetos no espaço é uma forma de modelar a configuração espacial. O Cálculo Espacial Qualitativo (CEQ) é baseado em relações binárias ou ternárias entre regiões. Trata-se de uma construção de conhecimento espacial e temporal sobre pontos, linhas ou regiões geralmente referidas como entidades.

Como citado em [6], as relações de conectividade frequentemente capturam a essência de uma configuração espacial. No Raciocínio Qualitativo Espacial (RQE), a conectividade fornece o modelo para relações espaciais. Os dois modelos de conectividade mais adotados no RQE são o *9-Intersection Model (9IM)* [19] e o *Region Connection Calculus (RCC)* [39].

O modelo de relação usado neste trabalho é derivado do modelo de cálculo RCC8. Trata-se de uma sub-álgebra deste modelo, cobrindo as relações: *Disconnected (DC)*, *Externally Connected (EC)*, *Partially Overlap (PO)* e *Equal (EQ)*, destacadas em verde na Figura 2.3.

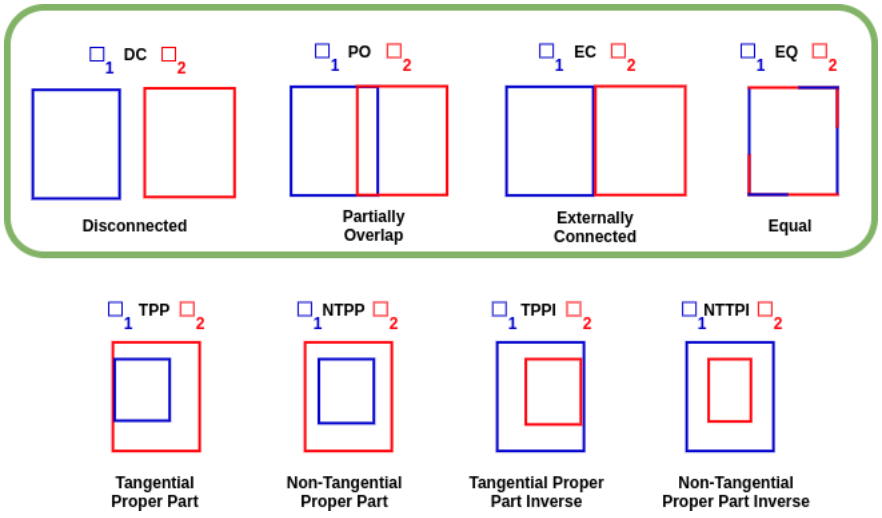


Figura 2.3: Relações de Region Connection Calculus

Outro tipo de cálculo qualitativo entre regiões é o *Cardinal Direction* (CD), que representa o conhecimento acerca da localização de um objeto em relação ao outro no espaço, baseando-se essencialmente nas orientações cardeais, como apresentado na Figura 2.3. As estruturas internas dos objetos não influenciam na análise destas direções.

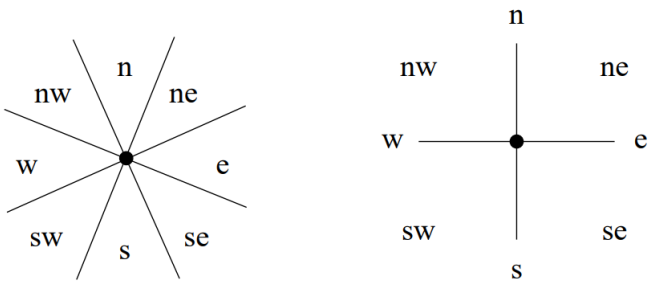


Figura 2.4: Direções cardeais.

2.2.1 Funções Espaciais

Para criar uma representação das relações qualitativas, existe um conjunto de funções pertencentes aos SGBDs para formular cálculos de domínio. Na Figura 2.5, é possível observar o conjunto de funções utilizadas pela extensão PostGis[3], pertencentes ao SGBD PostgreSQL[4]. Essas funções correspondem às relações qualitativas avaliadas no escopo desta pesquisa.

PostGIS SQL-MM Compliant Functions

```
boolean ST_Intersects( geometry geomA , geometry geomB );

boolean ST_Disjoint( geometry A , geometry B );

boolean ST_Equals(geometry A, geometry B);

boolean ST_Touches(geometry g1, geometry g2);
```

Figura 2.5: Funções espaciais da biblioteca PostGis Fonte: https://postgis.net/docs

As funções apresentadas na Figura 2.5 utilizam duas geometrias como parâmetros para comparação. A função *Equals* é utilizada no contexto de busca por POI que estão em um mesmo prédio ou polígono. A função *Intersects* define a relação de interseção entre regiões, ou seja, avalia se duas áreas compartilham algum ponto em comum, ou não. O significado da função *Touches* está associado ao toque, ou seja, se duas geometrias tocam em algum ponto de suas arestas ou vértices. Por último, a função *Disjoint* está associada à desconexão, quando uma região não toca nem compartilha nem um ponto com a outra. Para este trabalho, um raio de limite foi definido para esta distância em caso de desconexão de regiões.

[3]https://postgis.net/docs
[4]https://www.postgresql.org//

2.3 Consultas Espaciais Qualitativas

O Modelo Espacial Qualitativo (MEQ) é um modelo computacional de regiões, usado para raciocinar sobre relações direcionais e de conectividade. Várias representações consideram direções cardeais, relações de conectividade ou abordagens combinadas. Para alcançar abstração qualitativa, este modelo espacial usa uma linguagem relacional para formular representações de domínio.

O MEQ possui um conjunto de técnicas para representar e manipular conhecimento espacial e temporal por meio de linguagens relacionais que utilizam um conjunto finito de símbolos. Esses símbolos representam classes de propriedades semanticamente significativas do domínio representado (posições, direções, etc.).

O conceito de Consulta de Configuração Espacial Qualitativa (CCEQ) foi usado em [24] para definir um tipo de consulta que busca regiões para uma ou várias relações qualitativas. O conceito reforça a ideia de que, apesar de suas diferenças semânticas, as consultas espaciais compartilham uma semelhança básica: todas codificam predicados espaciais que, tipicamente, incorporam algum tipo de relação espacial qualitativa.

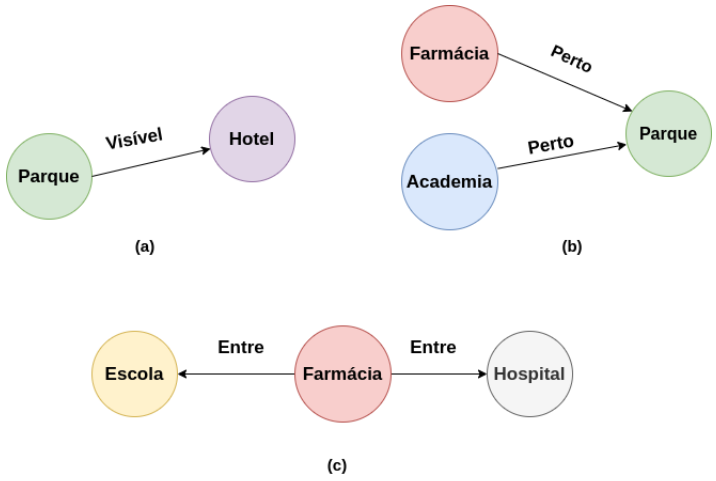


Figura 2.6: Predicados espaciais entre POI.

A Figura 2.6 apresenta um conjunto de relações qualitativas entre POI que usualmente são definidas por seres humanos. Por exemplo, a relação (a) descrita pela palavra "visível" representa a visibilidade de um POI em relação a outro. A relação definida por (b) exemplifica um conceito relativo de distância entre regiões; a expressão "perto" pode ter interpretações de acordo com o contexto em que é avaliada. Por último, a relação "entre" abrange o caso em que é necessário buscar regiões que estão entre outros tipos de regiões

definidas. No exemplo da Figura 2.6, uma farmácia está localizada entre uma escola e um hospital.

A Figura 2.6 mostra diferentes predicados espaciais entre regiões, os quais podem ter um grau de especificidade de acordo com preferências espaciais que os seres humanos expressam. Esses predicados espaciais fazem parte da definição das consultas espaciais qualitativas.

As consultas espaciais qualitativas podem ser classificadas de acordo com o grau de indeterminação dos predicados espaciais. Na Tabela 2.1, é possível identificar dois tipos de consultas entre regiões espaciais.

O tipo de consulta descrita como "Checagem de Relação" possui, de forma pré-definida, os valores correspondentes às regiões e relações qualitativas. Por exemplo, uma consulta realizada no formato *overlap(escola, parque)* busca identificar se existe uma relação de *overlap* (sobreposição) entre as regiões dos POI do tipo escola e parque, presentes na base de dados.

O tipo de consulta chamado de "Recuperação de Relação" possui apenas as regiões espaciais definidas na consulta e busca identificar a relação existente entre estas regiões. Um exemplo deste tipo de consulta é tentar recuperar a relação qualitativa existente entre os POI do tipo escola e parque.

Consulta	**Objeto A**	**Relação Qualitativa**	**Objeto B**	**Requisição espacial**
Checagem de Relação	informado	informado	informado	Existe a relação informada para os objetos dados?
Recuperação de Relação	informado	não informado	informado	Que relação existe entre os objetos dados?

Tabela 2.1: Classificação de consultas espaciais qualitativas de acordo com o grau de indeterminação dos predicados

2.4 Problema de Satisfação de Restrições

O Problema de Satisfação de Restrições (PSR) foi definido em [37] e discutido pelo trabalho [30]. Trata-se de um conjunto de variáveis que representam entidades e possuem relações definidas por um grupo de restrições. Estas restrições são pertencentes a um domínio discreto e finito, também apresentadas como predicados na Seção 2.3, são usadas para definir as

relações entre entidades espaciais.

O PSR está presente nos campos de Inteligência Artificial e Pesquisa Operacional. Possui um campo específico de satisfabilidade de restrições booleanas, conhecido como problemas SAT. Para o contexto de relações entre regiões, a definição de relações qualitativas desenvolvida em [39] foi importante para determinar o tipo de restrição entre entidades espaciais.

As técnicas de busca se baseiam em uma representação por grafos, tendo por base as relações de conexão definidas pelo padrão RCC e por direções cardinais. Na área de teoria dos grafos, os conceitos de isomorfismo definem uma relação de equivalência entre grafos e são utilizados como técnicas de busca. Segundo os autores de [40], o problema de satisfabilidade para RCC8 (e RCC5) é NP-completo.

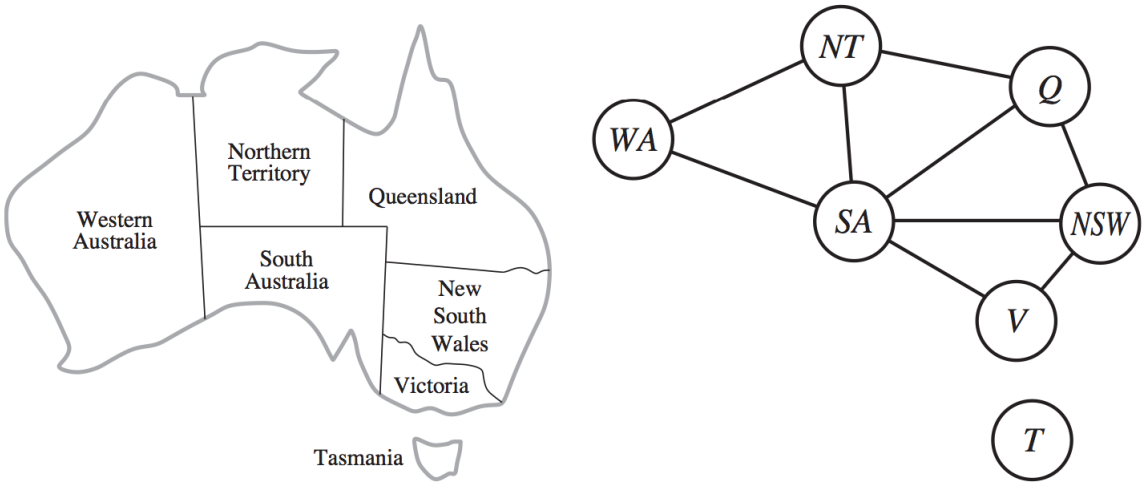


Figura 2.7: Mapa da Austrália representado por uma estrutura de grafo.

A Figura 2.7 extraída de [42] mostra em (a) um mapa da Austrália dividido em regiões e em (b) a sua representação em formato de grafo. No grafo, as adjacências ou fronteiras espaciais são representadas pelos arcos e as regiões pelos nós. Essa representação é importante para definição de uma estrutura de rede. Conforme apresentado no trabalho [18], o tipo de raciocínio baseado em restrições definido para cálculos espaciais qualitativos é chamado de geração de rede de restrições.

2.4.1 Redes de Restrições Qualitativas

A Rede de Restrição Qualitativa (RRQ) é definida como um grafo rotulado direcionado, com nós representando um domínio específico de entidades e as arestas caracterizando restrições binárias entre estes nós. Cada restrição é uma relação qualitativa e os nós representam um

domínio de entidade espacial. No âmbito desta dissertação, o domínio de um POI é utilizado para representação destas entidades.

Considerando uma RRQ de relações RCC, o problema que este trabalho busca solucionar é o de satisfabilidade. O objetivo é decidir se existe algum arranjo espacial das variáveis do RRQ que satisfaz todos os predicados. Por exemplo, um POI do tipo igreja que possui um restaurante dentro de seu prédio e está localizado ao lado de uma farmácia, pode ser representando no contexto RRQ através de um grafo, onde os nós são os POI e as arestas as relações RCC. A busca por registros que correspondem a todas as relações e nós do arranjo é uma forma de satisfazer a estrutura de RRQ.

A solução de um PSR qualitativo é encontrar um conjunto de variáveis que satisfaça todos ou um limite definido de predicados. Subjetivamente, o conceito de uma solução parcial de PSR pode ser interpretado como apenas uma parte dos predicados sendo satisfeitos no resultado. Essa definição é descrita como relaxamento de correspondência, ao ser usada no contexto de cenas espaciais [6].

2.5 Consultas Espaciais por Palavras-chave

A busca por palavras-chave é um dos tipos de consultas mais conhecidos na área de Recuperação da Informação. Esse tipo de busca recupera conjuntos de objetos espaciais com localizações próximas umas das outras e cujas descrições são relevantes para a definição fornecida pelo usuário, chamadas de conjunto de palavras-chave.

O conjunto de palavras-chave reflete os tipos de objetos procurados pelo usuário. A definição de busca por grupos empregada nesse trabalho se refere a esse tipo de busca, realizada à procura de um grupo heterogêneo de entidades. Existem vários tipos de consultas por palavras-chave [8; 15]. As consultas são categorizadas essencialmente pelo tipo de predicado. A seguir, apresenta-se uma lista destes tipos de predicados:

- Predicados de Seleção
- Predicados de Agregação
- Predicados de Junção
- Predicados de Grupo

Os predicados de seleção são aqueles que buscam por entidades detentoras de determinadas palavras-chave e se concentram em uma área específica, como a pesquisa de um único POI em um mapa de uma cidade, por exemplo.

Os predicados de agregação buscam associar um número maior de palavras à busca, baseando-se nos termos frequentes em um contexto temporal, ou seja, inserem palavras associadas que podem surgir para um mesmo contexto de busca.

Os predicados de junção buscam unir informações de diferentes fontes que possuem um formato textual definido. Por exemplo, uma informação pertencente a uma postagem em uma rede social como o Twitter[5] pode ser combinada com uma descrição ou avaliação de um POI para agregar valor ao resultado de uma consulta.

Os predicados de grupo são os tipos de predicados abordados nesta pesquisa, eles não consideram especificamente as propriedades dos objetos individuais, mas sim as propriedades coletivas de um grupo de pontos ou regiões. Buscar as características do arranjo do grupo é o objetivo da consulta, também chamada de busca por padrões espaciais [36]. É importante ressaltar que os padrões espaciais ou restrições espaciais precisam ser total ou parcialmente satisfeitas na resposta da consulta.

2.5.1 Indexação Espacial de Palavras-chave

Um dos tipos de indexação mais utilizados para dados espaciais são as R-Tree[4], trata-se de um índice hierárquico similar a uma B-Tree[14]. Cada nó da R-Tree possui um retângulo que delimita todos os dados espaciais dentro do nó, chamado *Minimum Boundig Rectangle* (MBR). Inicialmente, os dados espaciais são inseridos em um nó, denominado nó raiz, quando o número de elementos dentro de um nó da R-Tree excede sua capacidade, o nó é dividido em vários nós filhos.

Na solução proposta por esta pesquisa, é utilizado um tipo de índice que deriva da R-Tree, chamado de IR-Tree. Ele é utilizado para facilitar a busca por documentos relevantes em uma determinada base textual, no contexto deste trabalho, ele realiza a indexação de textos referentes aos tipos de POI utilizados nesta pesquisa.

[5]https://twitter.com/

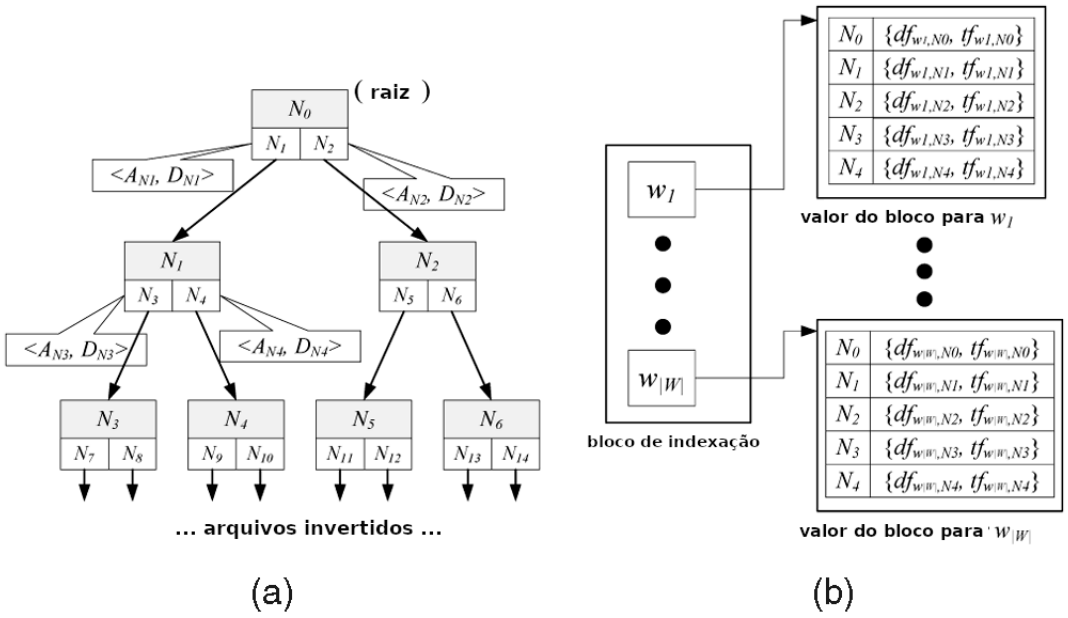


Figura 2.8: IR-Tree: (a) Estrutura da árvore e (b) Estrutura de documentos. Traduzida de [33].

A Figura 2.8 apresenta uma IR-Tree, um tipo diferente de R-Tree que possui uma estrutura de lista invertida de objetos em seu nós. A parte da figura identificada pela letra (a) mostra a estrutura da árvore de índices e a divisão de seus nós, a figura representada pela letra (b) mostra a lista de palavras presentes em cada nó do bloco de índices. A pesquisa por palavras-chave recupera um subconjunto dos objetos indexados que contêm todas as palavras-chave da consulta.

2.6 Considerações Finais

Este capítulo buscou apresentar os principais conceitos para entendimento da solução proposta. Com relação aos aspectos teóricos, as definições das relações qualitativas entre regiões foram discutidas para facilitar o entendimento da criação de consultas com padrões espaciais.

Em termos técnicos, algumas funções pertencentes as bibliotecas que foram implementadas nos experimentos da solução também foram apresentadas, juntamente com o detalhamento do formato dos dados manipulados. Por último, as técnicas de indexação de palavras-chave foram exibidas como etapa de processamento do algoritmo.

Capítulo 3

Trabalhos Relacionados

São apresentados neste capítulo os trabalhos que abordam os principais temas associados a esta pesquisa. A Seção 3.1 discorre sobre consultas coletivas utilizando palavras-chave, o que corresponde ao tipo de busca utilizado no algoritmo proposto. Em seguida, na Seção 3.2, são apresentados os trabalhos que realizam a indexação e consulta usando relações qualitativas. As fontes de pesquisa utilizadas para busca dos trabalhos utilizados foram: *ACM Digital Library*[1], *IEEE Xplore Digital Library*[2] e *Digital Bibliography Library Project*[3].

As principais palavras-chave utilizadas como parâmetro de busca foram: *"qualitative constraint network"*, *"spatial search"*, *"top-k spatial search"*, *"collective spatial search"* e *"spatial keyword queries"*, em língua portuguesa foram consideradas as palavras "busca coletiva espacial", "palavras-chave espaciais" e "relações espaciais qualitativas".

Aproximadamente 43 trabalhos sobre consultas por palavras-chave e redes de restrições qualitativas foram encontrados. Destes, 21 fazem referência a buscas de itens coletivos e foram selecionados de acordo com a viabilidade de implementação na área de redes de restrições qualitativas.

3.1 Consultas por Grupos de Palavras-Chave Espaciais

Para recuperar grupos de objetos espaciais com palavras-chave, diversos estudos como [5; 41; 49] definiram inicialmente o problema de consultas de palavras-chave espaciais coletivas. O objetivo foi associar uma localização geográfica à descrição textual de objetos. Um tipo de

[1] https://dl.acm.org/

[2] https://ieeexplore.ieee.org/

[3] https://dblp.org/

consulta que pode representar essa categoria são as consultas de *top-k* palavras-chave, que se refere a um grupo de objetos que contêm coletivamente todas as palavras-chave de consulta, conforme apresentado em [25; 12; 9; 26; 13].

Consultas de palavras-chave espaciais, como as apresentadas em [48; 17], fornecem uma pesquisa por diferentes entidades quando procuramos por tipos de POI. Cada palavra-chave na consulta pode representar um nome de local, descrição textual ou uma categoria do objeto de interesse do usuário. Estudos relacionados a consultas de palavras-chave espaciais, como os apresentados em [12; 25; 5], focam na distribuições dos pontos avaliadas por distância.

A abordagem desenvolvida em [17] pretende melhorar o processo de busca por palavras-chave, ela utiliza a classificação de qualidade do POI fornecida pelo usuário como parâmetro para seleção de itens relevantes. Em alguns casos, existe a necessidade de criar algum processo de pesquisa aliada a movimentação espacial do usuário que realiza a consulta. O algoritmo proposto em [27] permite que múltiplas consultas sejam executadas durante a mudança de um trajeto a partir de um ponto inicial.

A combinação de palavras-chave e padrões espaciais foi trazida em [25; 21]. Na pesquisa realizada em [22], é proposta uma definição de padrão espacial, utilizando a distância como restrição para a solução de um problema na área de *Graph Pattern Matching* (GPM).

O trabalho desenvolvido nesta dissertação é baseado no algoritmo proposto em [20], que foi avaliado como uma solução da área de RRQ. As principais diferenças existentes no novo algoritmo proposto nesta dissertação em relação ao [20] são: i) o uso de regiões poligonais ao invés de pontos para representação espacial e ii) as restrições de cálculo qualitativo que definem as relações entre estas entidades.

Um sistema denominado SpaceKey [21] também foi implementado para permitir a execução de consultas espaciais e fazer uma comparação entre várias abordagens com consultas espacias por palavras-chave. Posteriormente, outra solução foi proposta com o mesmo algoritmo de [20], usando um método de indexação Quad-Tree [11].

3.2 Indexação de Relações Espaciais Qualitativas

Tomando como base os estudos de SIG mostrados em [3], problemas do mundo real que afetam as relações qualitativas requerem o uso extensões espacias das entidades avaliadas. Embora seja possível encontrar vários estudos que abrangem as restrições de distância e direções cardeais [34; 29], o uso de conceitos de conectividade é o mais adotado em soluções de

RRQ, permitindo que ferramentas, como o SparQ [38], gerem e calculem redes de restrições complexas.

A abordagem do índice de agrupamento espacial de [23] fez melhorias em termos de cálculo e armazenamento, mas, como já mencionado, seu desempenho depende fortemente da qualidade da construção do próprio índice. Posteriormente, [23] sugere o uso da enumeração de correspondência de sub-grafos do algoritmo de Ullmann [46] para resolver o problema de CEQ para objetos espaciais. O problema abordado nesta dissertação também é classificado como um problema de correspondência de sub-grafo.

A Figura 3.1 foi elaborada para servir como um referencial teórico na construção de um mecanismo de busca por CCEQ, ela representa um esboço de arquitetura proposta por [23] para a realização das buscas qualitativas em bases espaciais. As imagens que estão representadas pela cor azul simbolizam a proposta de implementação do trabalho citado.

O fluxo de execução inicia com a entrada de descrições de relações espaciais por parte do usuário. Em seguida, um conjunto de predicados espaciais é definido para cada relação descrita. Conversões de predicados para consultas são realizadas na penúltima etapa. Por fim, consultas são executadas nas bases de dados espaciais e os resultados são enviados ao usuário através de uma visualização em mapa.

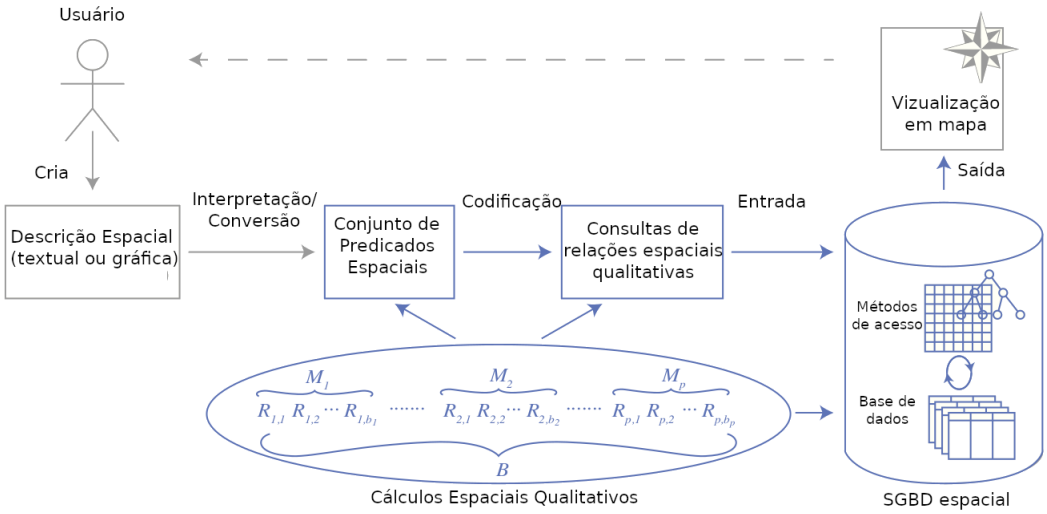


Figura 3.1: Arquitetura de mecanismo de busca por CCEQ. Fonte: Imagem traduzida de Fogliaroni(2016)[23]

Foi desenvolvido por [1] uma outra variante de indexação por agrupamento espacial para objetos, que visa obter um melhor *cluster* usando uma estratégia de *clustering* mais sofisti-

cada.

Em um estudo subsequente, a pesquisa realizada em [34], propõe uma nova abordagem para resolver o método de indexação para armazenar relações qualitativas, os conjuntos de dados usados em seus experimentos também são aplicados neste artigo. O objetivo é calcular o tempo de consulta usando as construções do experimento de busca em [35], da mesma forma, computando relações espaciais qualitativas, mas usando relações de conectividade ao invés de direções cardinais.

3.3 Resumo dos Trabalhos Relacionados

A Tabela 3.1 oferece em um quadro comparativo das soluções apresentadas na Seção 3.1. É possível observar que todas as soluções apresentadas realizam a busca por palavras-chave e consideram, em seus respectivos cálculos, as restrições de distância entre pontos .

	Busca por Grupos de Palavras-chave	**Restrição de Distância**	**Índice IR-Tree**	**Restrição de conectividade**	**Uso de Polígonos como objetos espaciais**
m-CK [25]	X	X			
min-SK [12]	X	X			
CO-SKQ[26]	X	X	X		
MSJ [20]	X	X	X		
Topo-MSJ	X	X	X	X	X

Tabela 3.1: Quadro comparativo de trabalhos relacionados

Todas as abordagens apresentadas na Tabela 3.1 utilizam algum tipo de índice para a estruturação dos dados, mas apenas os trabalhos [26; 20] fazem uso do índice IR-Tree. As abordagens que não utilizam este tipo de índice buscam criar uma indexação de forma híbrida, combinando o uso de listas invertidas com algoritmos de busca espacial.

O trabalho representado pelo MSJ [20] foi referência para criação do algoritmo Topo-MSJ. O uso de restrições de conectividade no cálculo das relações entre objetos espaciais é unicamente realizado no Topo-MSJ, assim como o uso de polígonos para a representação de objetos espaciais.

É importante observar que o algoritmo Topo-MSJ também considera as restrições de

distância, especificamente a relação qualitativa do tipo “disjoint”, onde é permitido definir um valor para o raio de afastamento entre polígonos no espaço.

3.4 Considerações Finais

Este capítulo apresenta os principais trabalhos relacionados com a solução proposta nesta dissertação. Pode-se observar que o algoritmo Topo-MSJ propõe uma nova técnica de busca e também explora uma nova área no campo de consultas por grupos de palavras-chave. O uso de polígonos como objetos espaciais, neste caso representando os POI, promove a aplicabilidade deste tipo de geometria que está cada vez mais presente nas bases de dados atuais. Além disso, a criação de consultas compostas por relações qualitativas preenche uma lacuna importante na literatura desta área de estudo. A seguir, é apresentado o algoritmo Topo-MSJ e sua aplicação no contexto de SIG.

Capítulo 4

Topo - Multi Star Join

Neste capítulo, é apresentada a solução proposta para a busca por grupos de POI utilizando o processamento qualitativo de regiões espaciais. A Seção 4.1 mostra uma visão geral da solução proposta e os conceitos de restrições qualitativas utilizados. Na Seção 4.2, é apresentado o Padrão Espacial Qualitativo (PEQ), ou seja, o formato de consultas a ser fornecido ao algoritmo de busca. A Seção 4.3 descreve o algoritmo e suas etapas. Por fim, a Seção 4.4 discute as considerações finais do capítulo.

4.1 Solução Proposta

Diferente de outras abordagens que utilizam pontos ou coordenadas geográficas de um mapa, a solução proposta nesta dissertação usa a relação entre os polígonos dos POI como parâmetro de restrição para a realização de buscas. O uso de geometrias espaciais visa expressar a extensão espacial dos prédios pertencentes aos POI e permitir o cálculo de suas relações de conectividade, sem considerar unicamente a distância entre entidades.

Na Figura 4.1, é possível observar o uso do algoritmo Topo-MSJ aplicado a um SIG. O usuário realiza requisições mediante consultas textuais utilizando o PEQ, que é definido na etapa “Consulta”. O algoritmo Topo-MSJ faz uso de uma base de dados indexados por uma IR-Tree, com o objetivo de fornecer dados a serem visualizados em um serviço de mapas.

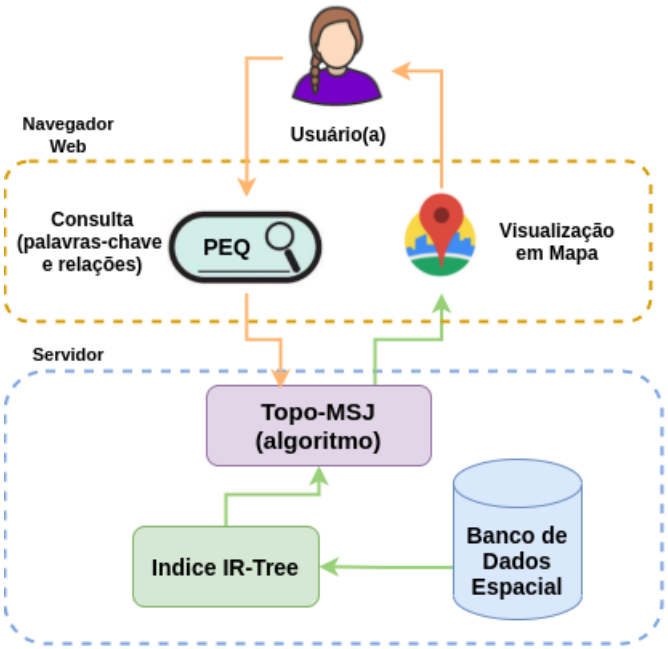


Figura 4.1: Arquitetura de um SIG utilizando o Topo-MSJ.

Baseando-se nos tipos de relações presentes nas bases utilizadas nesta solução, são usados quatro predicados espaciais *(Equal, Partially Overlap, Externally Connected e Disconnected)* oriundos de relações espaciais binárias de conectividade, apresentados na Seção 2.2 desta dissertação. Os predicados são usados na implementação da busca por POI e instanciados com as seguintes funções: *equals*, *overlap*, *touches* e *disjoint*.

A Figura 4.2 apresenta, em um mesmo cenário, os quatro tipos de relações abordadas neste trabalho. As cinco regiões **[a, b, c e d]** estão rotuladas na cor preta e as relações são apresentadas na cor vermelha. Várias combinações entre regiões e relações são possíveis. Por exemplo, caso o usuário esteja buscando por POI que possuam as áreas de todos os prédios conectados, todas as relações podem ser definidas apenas com o tipo *touches*. As possibilidades podem variar de acordo com a quantidade de regiões e relações existentes.

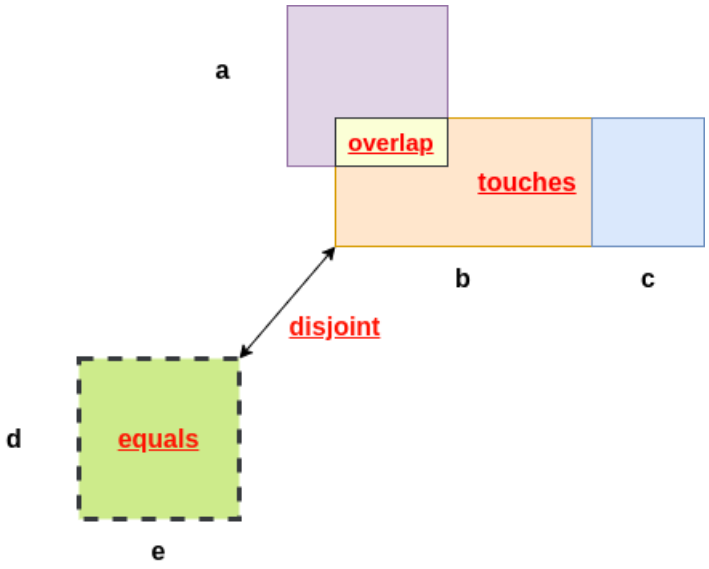


Figura 4.2: Relações qualitativas utilizadas pelo algoritmo Topo-MSJ.

A função *equals* retorna um par de POI que compartilham exatamente o mesmo prédio, sendo possível identificar esta relação entre as regiões **d** e **e** na Figura 4.2. A função *touches* é criada a partir de um limiar de conexão entre prédios pertencentes aos POI, as regiões **b** e **c** na Figura 4.2 simbolizam esta relação.

Em uma comparação entre a representação geométrica do POI em uma base de dados e sua área real, percebe-se que a função *touches* representa, em alguns casos, o toque das regiões que são de propriedade dos POI, mas não necessariamente o perímetro exato do prédio. Esta diferença existe devido ao fato que estes espaços no mundo real são criados como limites de segurança para casos de incêndios ou como áreas para estacionamentos.

A relação *disjoint* representa edifícios próximos que não se tocam, mas estão contidos em um limite definido por um raio calculado através de distância euclidiana, caracterizada entre as regiões **e** e **b** na Figura 4.2. A função *overlap*, existente entre **a** e **b**, simboliza a interseção de regiões que compartilham áreas em comum.

É importante salientar que a função *contains*, também pertencente ao modelo RCC, não foi utilizada no contexto desta pesquisa. Esta função permite identificar se uma determinada geometria está contida no interior de outra. No contexto do CEQ, apresentado na Seção 2.2, ela é definida como uma relação pertencente ao grupo *Proper Part*, que abrange as relações *Tangential Proper Part, Non-Tangential Proper Part, Tangential Proper Part Inverse e Non-Tangential Proper Part Inverse*.

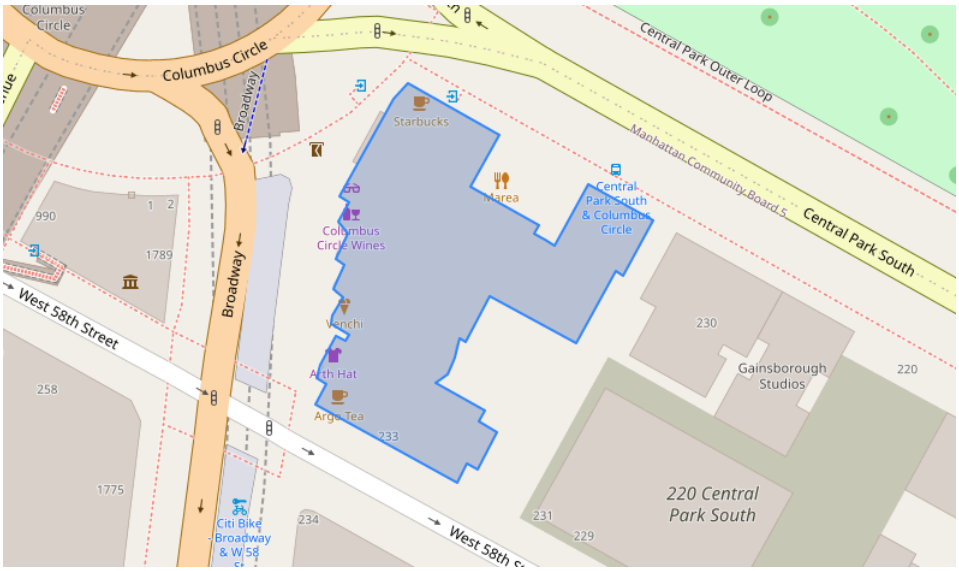


Figura 4.3: Visualização de geometria na ferramenta OpenStreetMap referente a um shopping center e um grupo de POI em seu interior

A região apresentada na Figura 4.3 mostra, em destaque na cor azul, uma geometria que pertence a 12 tipos de POI diferentes. Por características da construção da base de dados, os POI podem possuir geometrias iguais, uma vez que compartilhem a mesma extensão espacial.

No contexto da Figura 4.3, pode-se entender a região em destaque como o local de um shopping center, que possui algumas lojas em seu interior. A não delimitação destes estabelecimentos faz com que a relação existente entre a geometria de um restaurante e uma loja de departamento, por exemplo, seja avaliada pela função *equal*.

Embora tenha sido implementada e pudesse ser adicionada ao contexto do cálculo qualitativo, a função *contains* foi retirada do contexto experimental deste trabalho, em razão das bases de dados utilizadas nos experimentos não possuírem nenhuma ocorrência deste tipo de relação entre suas regiões. Ademais, funções que utilizam direções cardeais (e.g. Norte, Leste e Sudeste) foram consideradas como trabalhos futuros.

O objetivo principal da solução proposta nesta pesquisa é encontrar, em uma base de dados, padrões como os apresentados na Figura 4.2. Para tal, é utilizado o Padrão Espacial Qualitativo (PEQ) para definir a estrutura das consultas que podem ser realizadas.

Os desafios associados a esta solução estão presentes na representatividade da criação de consultas espaciais. As consultas qualitativas podem possuir diferentes métodos de entrada para o contexto de busca espacial. Alguns trabalhos sugerem até mesmo o uso de desenhos

do tipo *sketch* para facilitar a representação de itens espaciais.

A elaboração do PEQ visa facilitar o método de entrada de consultas qualitativas através de representação textual. É possível, portanto, destacar dois grandes benefícios desta solução: o modelo PEQ e o algoritmo Topo-MSJ. Ambos permitem a realização de consultas espaciais qualitativas em um tempo satisfatório, especificamente para o uso em aplicações no contexto de SIG, comparando com consultas SQL e estruturas de indexação qualitativa.

4.2 Padrão Espacial Qualitativo

O Padrão Espacial Qualitativo (PEQ) visa definir quais relações espaciais, representadas pelo modelo RCC [39], existem entre POI no espaço. O intuito deste modelo é permitir que a relação possa ser verificada em uma base de dados, usando a busca por palavras-chave como primeira etapa do processamento.

O primeiro objetivo da busca por PEQ é encontrar as entidades que o compõem. No contexto da pesquisa por POI, as palavras-chave são o atributo definidor dos tipos de entidades, as quais podem ser expressas como nomes de categorias (e.g. restaurantes, dentistas e hotéis). Além da busca por entidades, o PEQ procura definir a relação entre POI. Na Definição 1, é possível observar como se caracteriza esta correspondência.

Definição 1: A correspondência entre dois vértices (v_i,v_j) $\in V$ que representam regiões ocorre quando os dois polígonos no espaço têm a relação qualitativa definida pela restrição no PEQ. Chamaremos essa correspondência de *match* daqui para frente.

Seja *O* um conjunto de dados de objetos geo-textuais. Cada objeto $o \in O$ é um polígono em um espaço euclidiano 2D e está associado a um conjunto de palavras-chave $o_w = \{w_1, w_2, w_3, ..., w_{|i|}\}$, descrevendo as categorias de objetos ou seus nomes associados.

O Padrão Espacial Qualitativo (PEQ) é um grafo P = G (V, E), onde V é o conjunto de vértices e E é o conjunto de arestas. É possível observar na Figura 4.4 uma representação em grafo do conjunto de relações apresentadas na Figura 4.2 em forma de PEQ.

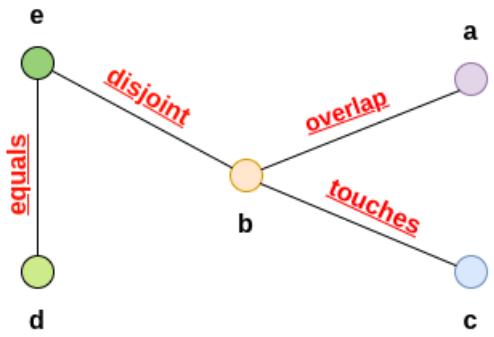


Figura 4.4: Exemplo de grafo do Padrão Espacial Qualitativo.

Os vértices em V e as arestas em E possuem as seguintes restrições:

- Cada vértice $v_i \in V$ é rotulado por uma palavra-chave $w_i \in W$;
- Cada aresta $(v_i, v_j) \in E$ é descrita por uma relação qualitativa $rel(v_i, v_j)$;
- As relações qualitativas são definidas pelas seguintes funções de conectividade: $disjoint(v_i, v_j)$, $overlap(v_i, v_j)$, $equals(v_i, v_j)$ e $touches(v_i, v_j)$.

O PEQ é formado por consulta espacial qualitativa baseada em palavras-chave feita por um usuário. Supõe-se que uma pessoa busca por um shopping center que possua um restaurante no mesmo prédio. Além disso, deseja-se que este restaurante esteja conectado a um supermercado. Por fim, o usuário pode estar procurando uma drogaria também nas proximidades As funções utilizadas no padrão determinado pela busca mencionada podem ser definidas da seguinte forma:

$$\textit{equals}('shopping','restaurante') \;\&$$
$$\textit{touches}('restaurante','supermercado') \;\&$$
$$\textit{disjoint}('supermercado','drogaria')$$

O formato da consulta proposta pelo PEQ permite inserir textos que representem diferentes atributos dos POI. Por exemplo, a criação de consultas por palavras-chave pode representar, não apenas a categoria do POI como *equals* ('escola', 'restaurante'), mas o nome do estabelecimento, na forma *equals* ('shell', 'burger').

4.3 Algoritmo

O algoritmo proposto nesta pesquisa, chamado de Topo-MSJ, é derivado do algoritmo Multi-Star-Join, proposto em [20; 21]. O Topo-MSJ soluciona uma variação do problema de correspondência de sub-grafo em bancos de dados espaciais, inicialmente definido em [16]. O problema chamado *Graph Pattern Matching* (GPM) busca encontrar um conjunto de sub-grafos S em um dado grafo G, no qual todos os vértices e relações de $s \in S$ possuam correspondência a um certo padrão P.

O Topo-MSJ, essencialmente, adaptou o algoritmo *Multi Star Join* (MSJ), apresentado na Seção 3.1, para realizar cálculos qualitativos utilizando as extensões espaciais dos POI. O Topo-MSJ utiliza dois sub-algoritmos de forma aninhada em sua implementação. Um é chamado de Topo-MPJOrder, que define a ordem de associação dos *matches* entre pares de regiões. O outro algoritmo é o Topo-PJ, responsável pela formação de listas de pares de itens que satisfazem as restrições das funções qualitativas, ou seja, ele identifica os *matches* entre os POI.

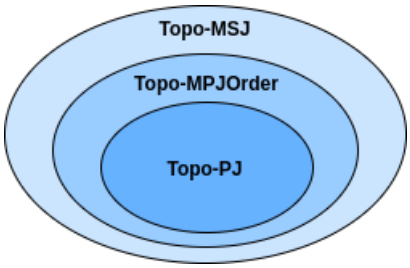


Figura 4.5: Ordem de execução dos algoritmos.

Na Figura 4.3, é possível identificar a ordem de execução dos algoritmos. A execução inicia-se com o algoritmo Topo-MSJ, realizando uma chamada ao sub-algoritmo Topo-MPJOrder que, por sua vez, chama o sub-algoritmo Topo-PJ.

4.3.1 IR-Tree

Antes da execução do algoritmo Topo-PJ, é necessária a realização da busca por palavras-chave para a identificação dos *matches* entre regiões. Esta busca é realizada utilizando o índice IR-Tree. Por este motivo, o primeiro procedimento realizado pela solução é a criação deste índice.

Para construção de uma IR-Tree, uma R-Tree é construída inicialmente e, em seguida, uma lista de documentos é associada a cada nó de forma invertida. Em cada nó folha, a

palavra-chave é associada a uma lista de objetos espaciais que a contém. Observa-se que os nós da árvore podem armazenar as palavras de forma individual, permitindo a busca de regiões utilizando uma correspondência textual parcial, com apenas uma palavra pertencente ao nome do POI.

Como discutido no Capítulo 3 desta dissertação, as relações espaciais qualitativas usualmente não são armazenadas em um SIG. Para este tipo de problema, em geral, uma estrutura de indexação do SGBD é utilizada para auxiliar o cálculo qualitativo. Nesta solução, as funções qualitativas são calculadas por meio de uma checagem nas listas invertidas dos nós da IR-Tree, verificando assim, o *match* entre regiões.

4.3.2 Topo-PJ

O Algoritmo 4.3.2, Topo-PJ, é um sub-algoritmo do algoritmo Topo-MPJOrder, responsável pelo *match* de pares de objetos. O Topo-PJ tem como objetivo recuperar um conjunto de objetos espaciais que atendam as restrições definidas pela consulta. Em outras palavras, concentra-se em utilizar as propriedades do algoritmo para definir a gama de candidatos para operação de junção no algoritmo Topo-MSJOrder.

O Topo-PJ recebe quatro entradas representadas por: $root$, rel $u_{i,j}$, w_i e w_j. A entrada $root$ representa o índice IR-Tree, a relação entre os pares de objetos p e q é definida pela entrada rel. O $u_{i,j}$ corresponde ao valor de *upper bound*, que indica o raio máximo de distância em uma relação de *disjoint*. O *upper bound* é calculado com base na distância máxima entre os MBRs, citados na Seção 2.5 do Capítulo 2. Por último, as entradas w_i e w_j simbolizam as palavras-chaves correspondentes aos POI.

A busca por pares de candidatos é realizada efetivamente da linha 3 até a linha 17. A função ϕ.keySet() retorna todas chaves de ϕ e a função ϕ.getKey(p) retorna o valor de uma chave para p. É possível perceber na linha 11 que o algoritmo Topo-PJ realiza operações com pares de geometrias. Neste momento, o cálculo de distância é substituído por um cálculo qualitativo, buscando identificar as restrições qualitativas entre as regiões espaciais dos POI. Para adaptar a solução buscando resolver consultas qualitativas, é necessário utilizar um limite superior para filtrar e definir uma gama de candidatos no grafo.

Algorithm 1 Topo - PJ

Input: root, rel, $u_{i,j}$, w_i, w_j;

Output: ϕ, all the matches

```
1: h ← height(root), φ ← ∅;
2: Λ.add(root), φ.add(root, Λ);
3: for i ← 1 to h do
4:     φ' ← ∅;
5:     for p ∈ φ.keySet() do
6:         Λ ← ∅, flag ← false;
7:         for p' ∈ p.invFile(w_i) do
8:             for q ∈ φ.getKey(p) do
9:                 for q' ∈ q.invFile(w_j) do
10:                    if MaxDist(p'.mbr, q'.mbr) ≤ ui,j then
11:                        if TopoRel(p'.geometry,q'.geometry) == rel; then
12:                            Λ.add(q');
13:                    else
14:                        flag ← true;break;
15:                if flag = true then break
16:            if flag = false then φ'.add(p', Λ);
17:    φ ← φ'
18: return φ
```

4.3.3 Topo-MPJOrder

O Algoritmo 4.3.3 Topo-MPJOrder é responsável por receber os pares de regiões que possuem *match* e definir a ordem de checagem dos vértices no grafo que representa o PEQ.

O fluxo de execução do Topo-MPJOrder é realizado da seguinte forma: recebe-se como entrada uma árvore IR-Tree e um padrão P. Inicia-se as variáveis na linha 1, onde Γ é uma lista criada para inserção dos resultados finais. A variável Q é uma fila de prioridade, em ordem crescente, que determina quais arestas são classificadas por seus números estimados de *matches*. Por fim, a lista U mantém o registro dos vértices visitados.

Em seguida, executa-se o Topo-PJ para arestas nas linhas 2 a 4, seleciona-se um vértice v e adiciona-se suas arestas em Q nas linhas 6 e 7. É possível observar que a função

$estimate(v-u)$ realiza uma amostragem para estimar o número existente de pares combinados. No laço de repetição da linhas 8 a 17, adicionamos a aresta com o número mínimo de *matches* para Γ.

Na linha 13, continua-se a retirar uma aresta da fila em Q, caso contrário, adicionamos à fila os vizinhos do vértice v, da linha 14 até a linha 18. O novo vértice v é marcado como visitado e, finalmente, Γ é retornado na linha 20.

Algorithm 2 Topo - MPJOrder

Input: root, P, δ, ϵ, θ;

Output: Γ, the join order of Topo-MPJ;

1: $\Gamma \leftarrow \emptyset, Q \leftarrow \emptyset, U \leftarrow \emptyset, \Upsilon \leftarrow \emptyset$;
2: **for** each edge (v_i, v_j) of P **do**
3: $\quad \Phi \leftarrow$ perform **Topo-PJ** for this edge;
4: $\quad \Upsilon.add((v_i, v_j), \Phi)$;
5: randomly select a vertex $v \in P$, and add it to U;
6: **for** $u \in nd(v)$ **do**
7: $\quad Q.add((v, u), estimate(v, u))$;
8: **while** Q.size > 0 **do**
9: $\quad (v_i, v_j) \leftarrow Q.pop()$;
10: $\quad \Gamma.add((v_i, v_j))$;
11: $\quad$ **if** $v_i \in U$ *and* $v_j \in U$ **then** $continue$;
12: $\quad v \leftarrow$ a newly considered vertex in (v_i, v_j) and U;
13: $\quad$ **for** $u \in nb(v) \wedge U$ **do**
14: $\quad\quad \Gamma.add((v_i, v_j))$
15: $\quad$ **for** $u \in nb(v) \setminus U$ **do**
16: $\quad\quad Q.add((v, u), estimate(v, u))$;
17: $\quad U.add(v)$;
18: return Γ;

4.3.4 Topo-MSJ

O algoritmo MSJ, modificado para a criação do Algoritmo 4.3.4 Topo-MSJ, utiliza tipos de relações de exclusão/inclusão no cálculo da distância entre pontos. O uso de parâmetros de

lower e *upper bound* cria um filtro de exclusão de acordo com a distância que um ponto se encontra de um referencial. No Topo-MSJ não há processo de exclusão de nós, o foco da solução é na geração de candidatos para todos os tipos de consultas, sendo construído para computar apenas relações de não exclusão.

As entradas do MSJ são: um IR-Tree e um padrão P. As saídas são todas as correspondências de P. Para cada aresta de Γ, encontra-se a ordem de todos os $matches$ na linha 1, através da chamada do *Topo-MSJOrder*. Depois disso, realiza-se busca por vértices âncoras, são aqueles que possuem mais de uma aresta conectada.

O processo de junção é realizado a partir da linha 6, onde a navegação entre os itens do grafo é realizada, adicionando os vértices que possuem uma correspondência entre pares de *matches*. Dois tipos de podas são realizados neste laço de execução. Uma poda para vértices âncoras e outra parcial, esta última, é aplicada para os itens que se encontram nos extremos do grafo. Por último, todos os *matches* são recuperados na linha 14.

Algorithm 3 Topo - MSJ

Input: root, P;

Output: Ψ, all the matches;

1: run **Topo-MSJOrder** and get Γ
2: find a set Π of anchors vertices from the 2-core of P
3: $\Psi \leftarrow \emptyset, \Phi_1 \leftarrow \emptyset, \Phi_2 \leftarrow \emptyset, ..., \Phi_m \leftarrow \emptyset$;
4: **for** $i \leftarrow 1$ to m **do**
5: $\quad \Phi_i \leftarrow$ run **Topo-PJ** for the edge e_i;
6: **for** $k \leftarrow 1$ to m **do**
7: $\quad$ let $e_k = (v_i, v_j)$ be the k-th edge in Γ ;
8: $\quad$ **if** e_k is a forward edge **then**
9: $\quad\quad \Psi \leftarrow \Psi.link(\Phi_k)$;
10: $\quad\quad$ let v be the latest considered vertex in e_k ;
11: $\quad\quad$ **if** $v \in \Pi$ **then** perform anchor-pruning;
12: $\quad$ **else**
13: $\quad\quad$ prune some partial matches in Ψ;
14: return Ψ';

4.3.5 Exemplo de Execução

Assumindo que um padrão de relações espaciais qualitativas foi criado em formato PEQ da seguinte forma:

$$\textit{equals}('academia','restaurante') \;\&$$
$$\textit{touches}('restaurante','padaria') \;\&$$
$$\textit{disjoint}('padaria','dentista')$$

O padrão PEQ apresentado será utilizado como parâmetro de entrada para o Topo-MSJ juntamente com uma IR-Tree. O índice é criado utilizando as palavras-chave da base de dados. Na primeira etapa do processamento do algoritmo é realizada uma consulta utilizando o Topo-PJ, a fim de encontrar os *matches* relativos as restrições espaciais. As relações equal('academia', 'restaurante'), touches(restaurante', 'padaria') e disjoint('padaria', 'dentista') serão chamadas nesta Seção de A, B e C, respectivamente.

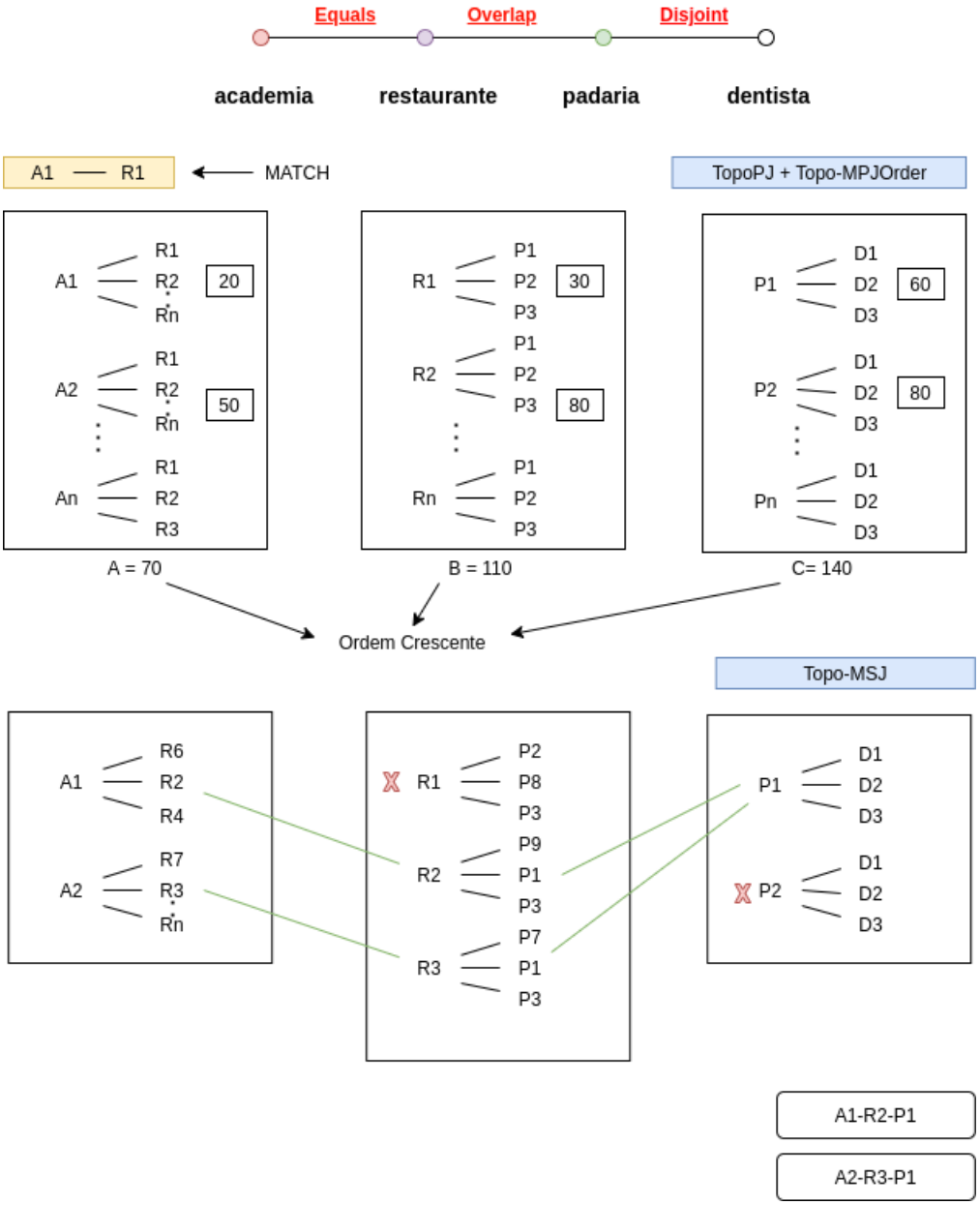


Figura 4.6: Exemplo de execução do Topo-MSJ

A Figura 4.3.5 apresenta de forma simplificada o processamento realizado pelo algoritmo Topo-MSJ. Na parte superior da figura é possível identificar o padrão PEQ, na sequência, a execução do Topo-PJ é realizada para à geração de *matches*. A ordenação na lista de grupos de *matches* é do feita pelo Topo-MPJOrder.

A pesquisa é realizada textualmente nas listas invertidas, o objeto espacial que possui a palavra buscada é identificado e comparado com um grupo de candidatos próximos. Para cada uma das relações comparadas, o Topo-PJ adiciona na lista de pares o *match* realizado nesta etapa do processamento.

No passo seguinte, estas listas de *matches* existentes são ordenadas de acordo suas quantidades de elementos, supõe-se que as listas de pares de relações possuem os seguintes tamanhos: A=70, B=110 e C=140. Após identificar o tamanho das listas de pares, o Topo-MPJOrder é responsável por criar uma ordem crescente do tamanho das listas, a serem avaliadas na junção destes pares.

Por ultimo, o Topo-MSJ é responsável por criar as ligações entre os pares de palavras. Para tal, ele navega por cada vértice da lista baseando-se no padrão de entrada, cada passo nesta etapa é calculado de acordo com a quantidade de possíveis candidatos para comparação.

Uma operação importante realizada nesta etapa é a poda, que consiste em eliminar grupos de candidatos que não realizam *match* com nenhum dos vértices, subsequentes. Por exemplo, na Figura 4.3.5 pode-se observar que não existe *match* entre R1 e nenhum objeto do tipo A, sabendo que o P2 só possui *match* com R1, logo, ele será eliminado em forma de poda. A eliminação baseia-se em *matches* de pares não realizados em passos anteriores.

Quando o Topo-MSJ consegue navegar em todo o padrão de entrada e encontra uma correspondência entre os candidatos da junção realizada entre pares, ele assume aquele grupo de objetos como uma solução da busca realizada.

4.4 Considerações Finais

Este capítulo apresenta um detalhamento do algoritmo Topo-MSJ e a definição do modelo PEQ. No começo do capítulo é possível identificar a função do Topo-MSJ em um cenário de aplicação SIG. O Topo-MSJ, juntamente com uma IR-Tree, pode ser utilizado de forma simples em sistemas de busca por dados espaciais. Os passos de execução do algoritmo Topo-MSJ também são descritos neste capítulo, assim como suas principais operações. É possível observar que a o algoritmo Topo-MSJ é estruturado de forma modularizada, contendo o sub-algoritmo Topo-MSJOrder e o Topo-PJ. Este tipo de modularização permite a implementação e manutenção de maneira mais eficiente, bem como a adição de novos tipos comparações entre relações.

A seguir, no Capítulo 5, são apresentados os experimentos, o detalhamento da implementação e os resultados obtidos. As questões de pesquisa também são definidas e discutidas no próximo capítulo.

Capítulo 5

Avaliação Experimental

Neste capítulo, é apresentada a avaliação experimental da solução proposta. O objetivo experimental principal é avaliar a eficiência do algoritmo Topo-MSJ em relação ao tempo de execução de uma consulta, além da eficácia no uso do PEQ para funções qualitativas espaciais. Foram realizados dois experimentos com bases de dados distintas, que buscam avaliar o desempenho da solução em relação à consultas SQL em um Sistema de Gerenciamento de Bancos de Dados (SGBD) e à algumas técnicas de indexação qualitativa.

Na Seção 5.1, são apresentadas as questões de pesquisa com um detalhamento dos objetivos de cada avaliação. A Seção 5.2 descreve a configuração dos experimentos, o formato das bases de dados e as bibliotecas utilizadas para implementação. Na Seção 5.3, é apresentado o Experimento 1 e seus resultados. A Seção 5.4 apresenta o Experimento 2 e sua avaliação comparativa. Por último, a Seção 5.5 apresenta as considerações finais do capítulo.

5.1 Questões de Pesquisa

Para orientar a avaliação da solução proposta, as seguintes questões de pesquisa foram definidas para dois cenários de avaliação.

O primeiro cenário de avaliação aborda aspectos de **eficiência** do algoritmo:

- **QP1:** O algoritmo Topo-MSJ permite realizar consultas com um tempo de execução inferior a uma consulta escrita em Structured Query Language (SQL) realizada em um SGBD Relacional?
- **QP2:** O algoritmo Topo-MSJ permite realizar consultas com um tempo de execução inferior ao das abordagens que utilizam técnicas de indexação espacial?

O segundo cenário avaliativo aborda as aspectos de **eficácia** do algoritmo:

- **QP3:** A busca por grupos de POI utilizando relações qualitativas recupera todos os dados existentes referentes à consulta?
- **QP4:** Um modelo para consultas espaciais baseado em Padrão Espacial Qualitativo (PEQ) permite representar as relações de conectividade entre as regiões espaciais?
- **QP5:** É possível utilizar dados espaciais no formato de polígono para representar POI em consultas qualitativas?
- **QP6:** O agrupamento de vários tipos de relações de conectividade entre regiões pode ser realizado no contexto de consultas espaciais?

5.2 Configuração dos Experimentos

A solução proposta nesta pesquisa é avaliada em dois experimentos. O Experimento 1 busca avaliar o tempo de execução de consultas qualitativas usando o algoritmo Topo-MSJ, comparando-o ao tempo de uma consulta SQL executada em um SGBD Relacional. Numa outra abordagem, o Experimento 2 tem como objetivo comparar o tempo de execução de consultas realizadas utilizando o Topo-MSJ e abordagens que utilizam indexação espacial.

Para cada experimento, é executado um tipo diferente de consulta qualitativa, definidos na Seção 2.3. O tipo utilizado no Experimento 1 é a "Checagem de Relação". Esse tipo de consulta possui os objetos e as relações espaciais pré-definidas. Neste caso, a consulta irá identificar, por exemplo, se na base de dados existe dois objetos (POI) que possuam uma relação qualitativa específica.

No Experimento 2, o algoritmo Topo-MSJ foi modificado para permitir a realização do tipo de consulta chamado "Recuperação de Relação". O objetivo desta modificação é explorar o mesmo tipo de consulta utilizado em [34], que possui as abordagens de indexação qualitativa usadas na comparação realizada pelo experimento. Neste caso, o algoritmo consegue identificar qual a relação existente entre dois objetos informados.

A implementação do algoritmo foi realizada na linguagem de programação Java[1], juntamente com a biblioteca de funções espaciais JTS Topology Suite[2] e o SGBD PostgreSQL[3]

[1] https://www.java.com/pt-BR/
[2] https://www.osgeo.org/projects/jts/
[3] https://www.postgresql.org/

com a extensão PostGIS[4], a qual permite manipular dados geográficos.

5.2.1 Bases de Dados

Duas bases de dados foram selecionadas para as avaliações realizadas nos experimentos. A base de dados utilizada no Experimento 1 refere-se aos POI dos estados americanos de Nova York e da Califórnia, pertencentes à empresa SafeGraph[5], uma empresa de dados que agrega dados de localização anônimos de vários aplicativos para fornecer insights sobre lugares físicos, por meio da Comunidade SafeGraph. Para aumentar a privacidade, o SafeGraph exclui as informações do grupo de blocos do censo se menos de dois dispositivos visitaram um estabelecimento em um mês de um determinado grupo de blocos do censo, os dados utilizados neste trabalho foram coletados no mês de maio de 2021.

O SafeGraph é um provedor comercial de dados massivos de POI, ele possui registros com chaves únicas chamadas de *placekeys*. Os nomes e tipos dos POI são representados em formato textual e as geometrias definem os polígonos, que possuem a extensão espacial dos prédios pertencentes aos POI.

Base de Dados	**#Regiões**	**#Nomes de POI**	**#Tipos de POI**
Nova Iorque	257.252	194.084	145
Califórnia	650.798	484.536	158

Tabela 5.1: Base de dados do Safegraph.

A Tabela 5.1 apresenta um detalhamento da quantidade de POI em cada base de dados, sendo a base de dados da Califórnia a maior. A tabela também apresenta a quantidade de nomes de POI distintos presentes em cada base, seguida da quantidade de tipos de POI.

O Experimento 2 utiliza cinco bases de dados de regiões administrativas de Áreas Administrativas Globais (AAG), chamadas neste experimento de Real-1. Além delas, utiliza-se cinco bases de dados sobre habitats naturais da Agência Ambiental Europeia, denominadas de Real-2. No total, são 10 bases[6] de dados diferentes usadas na comparação do experimento.

A Tabela 5.2 apresenta um resumo das bases de dados, se dividindo em dois grupos, chamados Real-1 e Real-2. É possível observar a quantidade de regiões presentes em cada

[4]https://postgis.net/

[5]https://www.safegraph.com/

[6]Disponível em: http://zhiguolong.github.io/file/datasets.zip

Base de Dados	#Regiões	AID
Germany	434	6,57
Ukraine	629	6,17
Australia	1.395	7,12
China	2.411	6,86
USA	3.145	6,42
Real-2.1	600	4,22
Real-2.2	610	5,60
Real-2.3	605	1,37
Real-2.4	611	8,43
Real-2.5	604	4,77

Tabela 5.2: Bases de dados Real-1 e Real-2.

base de dados, assim como os dados de *Average Intersection Degree* (AID), que representam a média de interseção dos MBRs em uma determinada base de dados.

5.3 Experimento 1: Análise comparativa com consultas SQL

O Experimento 1 visa comparar o tempo de consultas executadas pelo algoritmo Topo-MSJ em relação ao tempo das consultas SQL realizadas em um SGBD. O SGBD utilizado neste experimento foi o PostgreSQL. O objetivo deste experimento é mostrar que, para recuperação de dados espaciais, a solução proposta pode ser mais eficiente do que a execução de consultas SQL.

As consultas SQL realizadas de forma direta a um SGBD, em sua maioria, possuem um tempo de execução menor que as consultas executadas por uma camada de aplicação. A redução do tráfego de rede é um dos principais motivos deste ganho de desempenho. A depender da consulta, ela também pode possuir a vantagem de ter um plano de execução definido e armazenado no próprio SGBD. Como a solução proposta neste trabalho visa sugerir uma técnica a ser implementada em SIG, o objetivo deste experimento é buscar o melhor cenário comparativo, considerando o desempenho mais eficiente em tempo de execução de consultas espaciais qualitativas.

Um ponto importante deste experimento é o tamanho da base de dados avaliada. Criou-se

uma aproximação do tamanho real das bases de dados de SIG, entendendo que a densidade de POI em grandes cidades é elevada. Por exemplo, dados mostram que no ano de 2017 a cidade de Nova Iorque possuía 26.697 POI do tipo restaurante, para uma área de 783 km².

A base de dados utilizada na execução das consultas SQL foi indexada por meio de uma *Generalized Search Tree* (GiST). Uma GiST é uma estrutura de dados usada para construir uma variedade de índices baseados em árvores. Essa abordagem tem o intuito de agilizar o acesso aos dados e comparar a eficiência do algoritmo proposto com um índice adequado ao tipo de dado espacial.

Em um primeiro momento, são definidas as consultas utilizadas na comparação, baseadas no modelo proposto pelo PEQ. Os padrões de consulta foram criados seguindo uma metodologia semelhante aos experimentos de trabalhos relacionados, especificamente, os que realizam consultas por palavras-chave espaciais, como em [20]. Seis palavras-chave referentes aos tipos de POI mais frequentes na base de dados foram escolhidos para compor uma lista, estas são: “Offices of Physicians”, “Automotive Repair and Maintenance”, “Personal Care Services”, “Restaurants and Other Eating Places”, “Offices of Dentists” e “Religious Organizations”. Selecionou-se aleatoriamente desta lista os objetos espaciais a fim de formar cada item da consulta. As relações também foram escolhidas de forma aleatória, dentre as pertencentes ao PEQ.

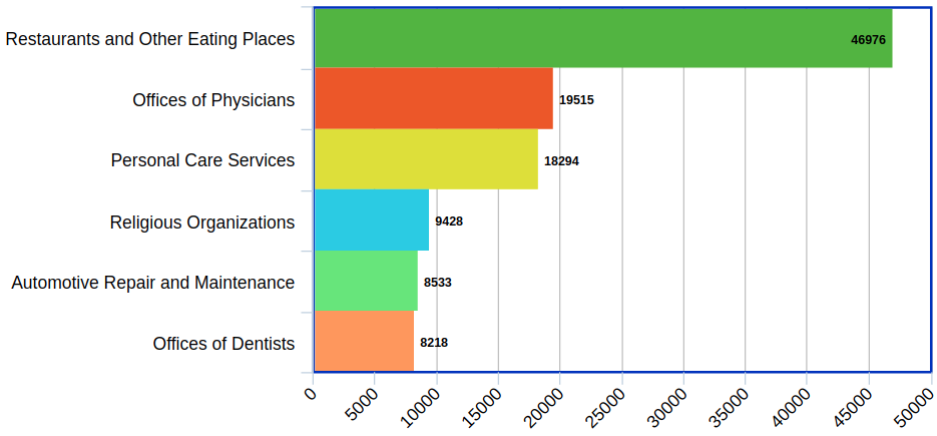


Figura 5.1: Histograma dos tipos de POI na base de dados de Nova Iorque.

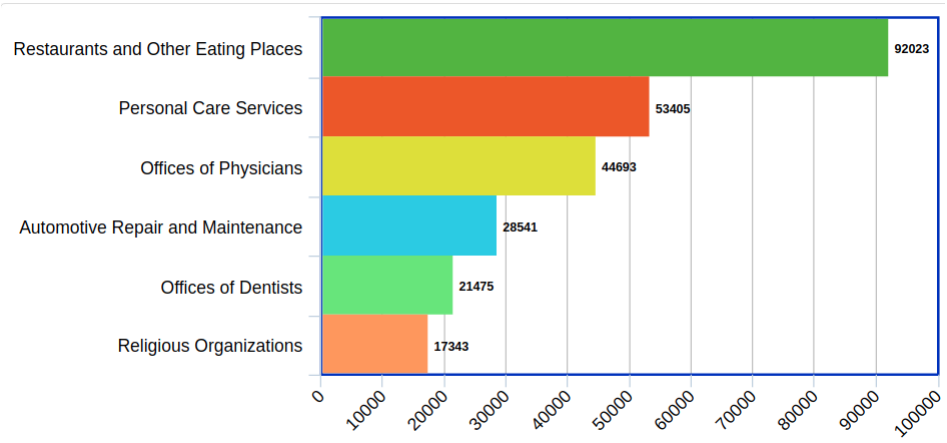


Figura 5.2: Histograma dos tipos de POI na base de dados da Califórnia.

Nas Figuras 5.1 e 5.2, são apresentadas as distribuições de frequências referentes aos POI utilizados na construção dos padrões das consultas. Os dados correspondem às bases de dados da Califórnia e Nova Iorque. Observa-se que o POI do tipo "Restaurants and Other Eating Places" está em maior quantidade em ambas as bases de dados. No caso da base de Nova Iorque, este tipo de POI possui mais que o dobro de registros em relação ao tipo "Offices of Physicians", que é segundo POI mais frequente.

Para definir a lista de consultas, foram criados oito grafos com diferentes ordens e tipos de arestas, apresentados na Figura 5.3. Cada vértice representa um POI, definido pelo nome da categoria na cor preta. As arestas representam a relação de conectividade entre os POI, ilustrados na cor vermelha.

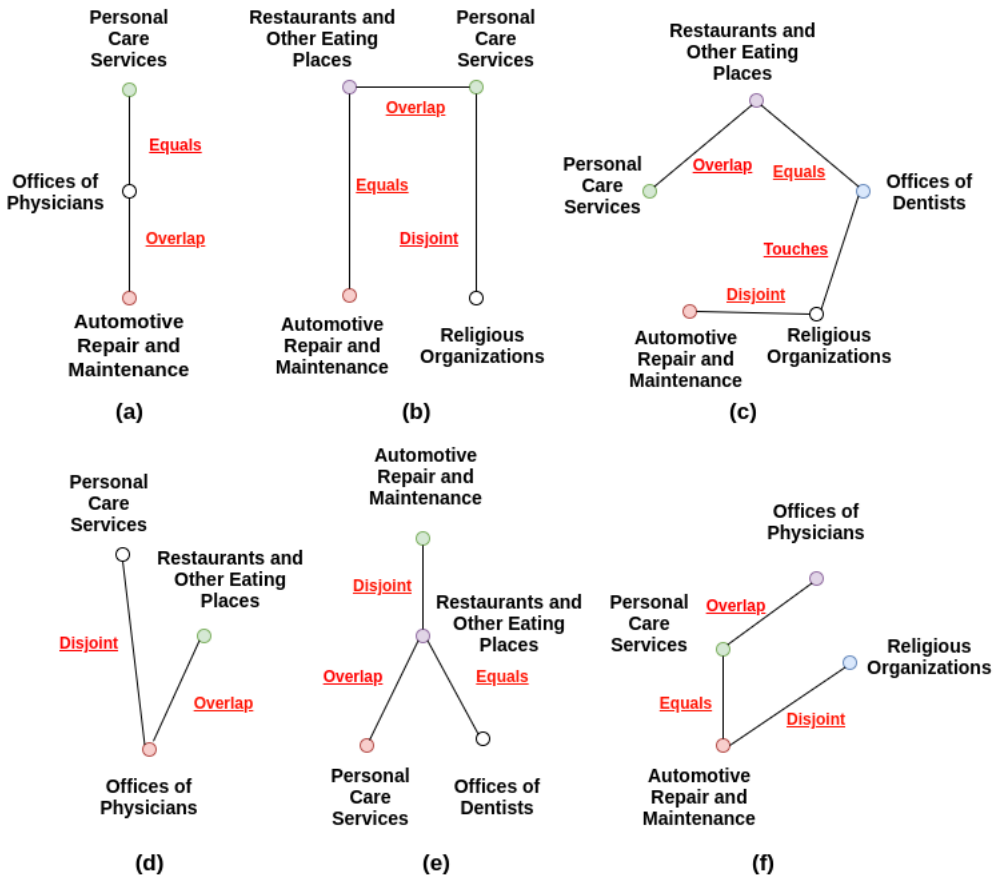


Figura 5.3: Consultas em PEQ utilizadas no experimento.

Os padrões de consultas criados na Figura 5.3 são baseados em padrões de palavras-chave espaciais definidos em trabalhos anteriores, como em [20], nos quais as arestas representavam as restrições de distância. Nesta solução, porém, as arestas simbolizam as relações de conectividade entre POI.

A Tabela 5.3 apresenta os resultados referentes ao tempo de execução de consultas em um SGBD, assim como as consultas realizadas usando o algoritmo Topo-MSJ. As colunas da tabela representam os padrões de **(a)** até **(f)**, definidos na Figura 5.3 e mapeados para as consultas presentes no Apêndice A.

As consultas possuem diferentes graus de complexidade de acordo com a quantidade de POI e relações presentes. Na consulta (a), especificamente, é definida apenas as relações "Overlap" e "Equal"entre os POI do tipo “Offices of Physicians” e “Automotive Repair and Maintenance”. A consulta (a) possui um grau de complexidade baixo e seu tempo de execução se apresenta similar ao tempo de consulta do Topo-MSJ.

Três consultas de maior complexidade, como **(c)**, **(e)** e **(f)**, possuem quatro tipos de POI

e três diferentes tipos de relações. Neste caso, são utilizadas funções da biblioteca PostGIS para definir as relações qualitativas em SQL. Em virtude desse grau de complexidade mais elevado, é possível notar que as consultas em SQL são significativamente mais lentas que as consultas realizadas pelo Topo-MSJ.

As consultas (b) até (f) possuem entre seus elementos a relação "Disjoint", esta relação necessita de um cálculo de distância entre regiões, uma vez que o PEQ define um raio entre os objetos que possuem este tipo de predicado. É possível observar, na Tabela 5.3, que o tempo de execução destas consultas é também mais elevado quando realizado em SQL, sendo necessário definir um limite de distância na criação da consulta.

Pode-se observar que as consultas SQL, além de cálculos qualitativos entre regiões, realizam também uma busca textual baseando-se nos tipos de POI. Por exemplo, a expressão "top_category = Religious Organizations" permite encontrar tipos de POI que correspondem ao formato exato desta entrada textual.

	(a)	**(b)**	**(c)**	**(d)**	**(e)**	**(f)**
Consulta SQL (NY)	2s	8m 18s	15m 47s	11m 33s	12m 55s	6m 53s
Consulta SQL (CA)	5s	44m 26s	40min 38s	36m 17s	43m 18s	36m 12s
Consulta Topo-MSJ (NY)	1s	4s	3s	2s	4s	3s
Consulta Topo-MSJ (CA)	2s	8s	6s	4s	7s	5s

Tabela 5.3: Tempo de execução de consultas.

A Tabela 5.4 apresenta os resultados referentes as execuções utilizando as consultas presentes no Apêndice B. A diferença entre as consultas definidas no Apêndice A e Apêndice B está na comparação textual realizada entre os tipos de POI.

As consultas realizadas no Apêndice B utilizam apenas uma palavra para definir o valor textual da consulta, o que amplia a possibilidade da busca por palavras que estejam contidas em tipos de POI diferentes. Por exemplo, a consulta (b) possui a seguinte linha: "top_category LIKE %Automotive%", esta forma de pesquisa permite que POI do tipo "Automotive Equipment Rental and Leasing", "Automotive Repair and Maintenance"e "Automotive Parts, Accessories, and Tire Stores", sejam recuperados.

É possível observar na Tabela 5.4 que o tipo de comparação textual utilizando o operador **LIKE** tem influência no tempo de processamento da consulta SQL, tornando este tipo de consulta mais lenta em comparação com as consultas do Apêndice A, enquanto o Topo-MSJ

sofre pouca diferença no tempo de execução com essa mudança.

	(a)	**(b)**	**(c)**	**(d)**	**(e)**	**(f)**
Consulta SQL (NY)	4s	16m 22s	27m 26s	25m 33s	15m 02s	11m 12s
Consulta SQL (CA)	7s	1h 7min	1h 26m	1h 18m	56m 12s	38m 34s
Consulta Topo-MSJ(NY)	2s	5s	6s	4s	8s	4s
Consulta Topo-MSJ(CA)	4s	10s	13s	8s	12s	7s

Tabela 5.4: Tempo de execução de consultas (operador LIKE).

Foi realizado uma verificação da corretude dos resultados gerados pelo algoritmo, pode-se assumir, portanto, que a consulta recupera todos os registros existentes na base de dados referentes aos predicados definidos. Em relação às consultas realizadas pelo algoritmo Topo-MSJ, é possível encontrar o código da implementação do PEQ no Apêndice D desta dissertação.

O Código Fonte 5.1 apresenta a Consulta C realizada nesta avaliação experimental. A consulta possui boa representatividade no uso do PEQ por conter todas as relações qualitativas (*equals*, *overlap*, *touches* e *disjoint*) implementadas na solução. O tempo de execução de uma consulta como esta no SGBD é de aproximadamente 2 minutos e 16 segundos, enquanto que usando o algoritmo Topo-MSJ são gastos aproximadamente três segundos.

Código Fonte 5.1: Consultas SQL - C

```
1
2  SELECT * FROM ca_test n1, ca_test n2, ca_test n3, ca_test n4
3  WHERE ST_Equals(n1.geom, n2.geom)
4  AND ST_Intersects(n2.geom, n3.geom)
5  AND ST_Disjoint(n3.geom, n4.geom)
6  AND ST_Distance(n3.geom, n4.geom) < 0.001
7  AND n1.top_category LIKE '%Museum%'
8  AND n2.top_category LIKE '%Restaurants%'
9  AND n3.top_category LIKE '%Book%'
10 AND n4.top_category LIKE '%Automotive%';
```

Alguns fatores influenciam no grau de complexidade de uma consulta em termos de processamento. A comparação textual utilizando “LIKE”, ao invés do uso do simbolo de igualdade “=”, faz com que o termo buscado seja procurado dentro do texto que representa o tipo de POI, consumindo assim mais tempo de execução. Dentre as relações qualitativas utiliza-

das na consulta, a relação que possui o maior custo computacional é a função *STDistance*; ela realiza um cálculo de distância euclidiana entre regiões e compara em uma unidade de graus.

A Figura 5.4, em seu item (b), representa o padrão utilizado para construção da consulta apresentada pelo Código Fonte 5.1.

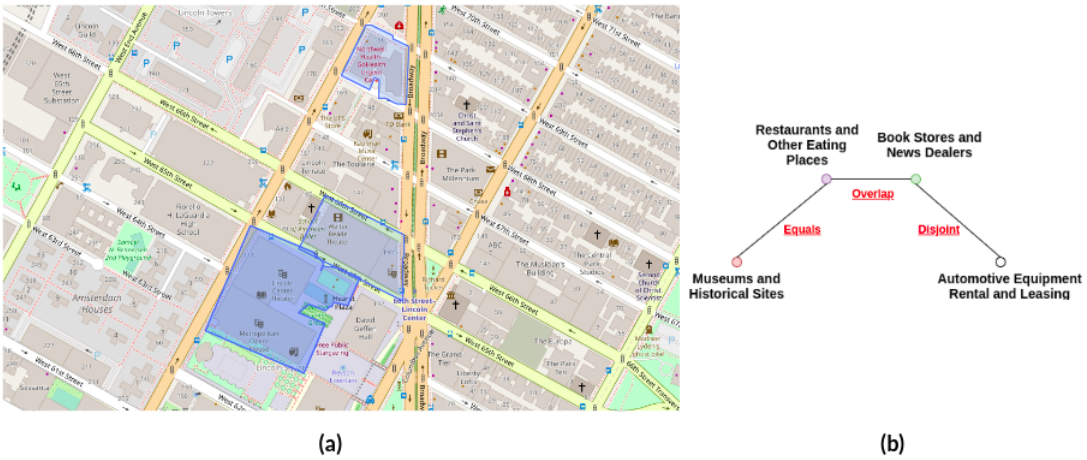


Figura 5.4: (a) Visualização do resultado da consulta na ferramenta OpenStreetMap. (b) Representação do padrão PEQ utilizado na consulta

No item (a) da Figura 5.4, identifica-se um dos resultados encontrados pela consulta realizada utilizando o padrão do item (b). Estes dados são visualizados pela ferramenta *OpenStreetMap* e representam uma busca semelhante ao cenário apresentado na Seção 1.1. Neste contexto, supõe-se que uma pessoa realizando uma viagem necessita encontrar um restaurante que esteja dentro de um museu, este, conectado a uma livraria, que se encontra em uma determinada distância de uma locadora de carros. O resultado desta consulta consegue encontrar com exatidão o padrão procurado para este tipo de busca.

Na Figura 5.5 são apresentados os resultados de 120 consultas realizadas pelo algoritmo Topo-MSJ. As consultas foram criadas utilizando os padrões presentes na Figura 5.3. Os tipos de POI para cada execução foram escolhidos de forma aleatória, provenientes dos seis tipos de POI utilizados no experimento, as relações se mantiveram as mesmas para cada padrão.

O objetivo desse experimento é identificar se o tempo de execução do algoritmo Topo-MSJ possui significativas diferenças, de acordo com os tipos de POI utilizados na busca.

Na Figura 5.5 é possível observar que o maior intervalo de tempo que todas as execuções conseguem variar é de 4 segundos. Estes resultados mostram que o algoritmo Topo-MSJ se mantém em uma faixa de tempo de execução, mesmo com variações nos parâmetros utilizados em sua busca.

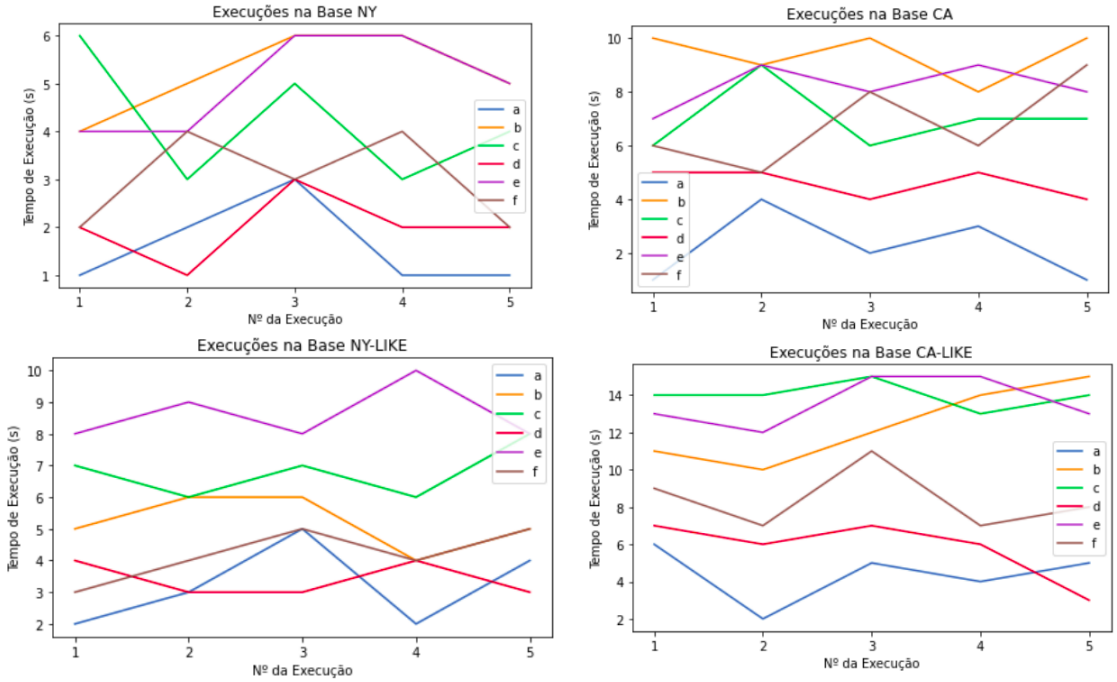


Figura 5.5: Tempo de execução de consultas

Observando a forma como as consultas SQL são elaboradas, é possível afirmar que a solução sugerida por este trabalho, além de possuir um desempenho superior, possui vantagens em relação ao tempo e a complexidade da criação destas consultas SQL.

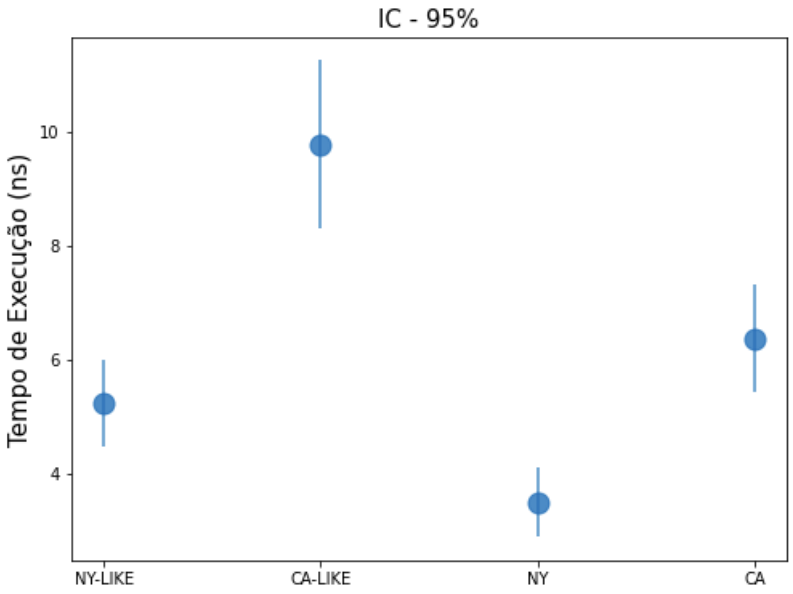


Figura 5.6: Intervalo de confiança do tempo de execução das consultas

Na Figura 5.6, é possível encontrar um gráfico que apresenta os intervalos de confiança referentes ao tempo de execução das consultas. Os valores apresentados são relativos à múltiplas execuções dos diferentes tipos de consulta utilizados na avaliação experimental. Apesar das consultas variarem na quantidade de relações, elas se encontram em uma faixa de 2 segundos em relação ao tempo médio de consulta.

A criação de consultas em SQL com padrões espaciais qualitativos pode ser demorada, prolixa e suscetível a erros, mesmo usando funções já implementadas em extensões em sistemas SGBD. Além disso, observa-se nos Apêndices A e B, que as consultas podem ter um tamanho extenso, o que torna difícil a sua manutenção.

Os exemplos de resultados das consultas executadas e definidas pelos padrões da Figura 5.3 podem ser vistos no Apêndice C deste documento. Estas imagens apresentam os mapas com as regiões dos POI da cidade de Nova Iorque e suas respectivas relações de conectividade definidas na consulta.

5.4 Experimento 2: Análise comparativa com outras abordagens

O Experimento 2 consiste em comparar o tempo de execução do algoritmo proposto com soluções na área de indexação qualitativa, mostrando que o Topo-MSJ pode ser uma alternativa à realização de consultas com restrições qualitativas em um tempo de execução satisfatório.

Como apresenta-se na Seção 5.2 deste Capítulo, o Experimento 2 foi elaborado utilizando um tipo de consulta chamado "Recuperação de Relação". Para tornar possível a realização deste tipo de consulta, o algoritmo Topo-PJ foi modificado em seu processamento na etapa de checagem de relações qualitativas, verificando a relação existente entre duas regiões ao invés de buscar por uma relação pré-definida pela consulta.

As soluções comparadas neste experimento foram avaliadas inicialmente em [34] e são divididas em *Direct Computation (Direct)*, método de uso de *MBRs*, indexação baseada em *Grid*, uma estrutura baseada em *R-tree* e um método denominado *Complete*.

A primeira abordagem avaliada é a *Direct Computation (Direct)*, que representa o cálculo geométrico realizado de forma direta entre duas regiões. A segunda estrutura é o índice de agrupamento por *Grid*. Neste tipo de índice, é construída uma estrutura de agrupamento para regiões espaciais utilizando retângulos conectados como grade. Estes retângulos cobrem o conjunto de dados espaciais como blocos de índice.

O terceiro tipo de estrutura de indexação é o R-Tree, ele utiliza uma estrutura semelhante ao índice usado na solução proposta por esta dissertação, mas sem o uso de listas invertidas. Para o tipo de indexação baseado em MBR, a criação se concentra em identificar todos os MBRs que se cruzam (internamente) dentro da base de dados. Por último, o tipo *Complete* consiste no cálculo e armazenamento das relações de cada par de regiões espaciais.

É possível identificar na Figura 5.7 os valores do tempo de execução das consultas utilizando as abordagens de indexação qualitativa descritas.

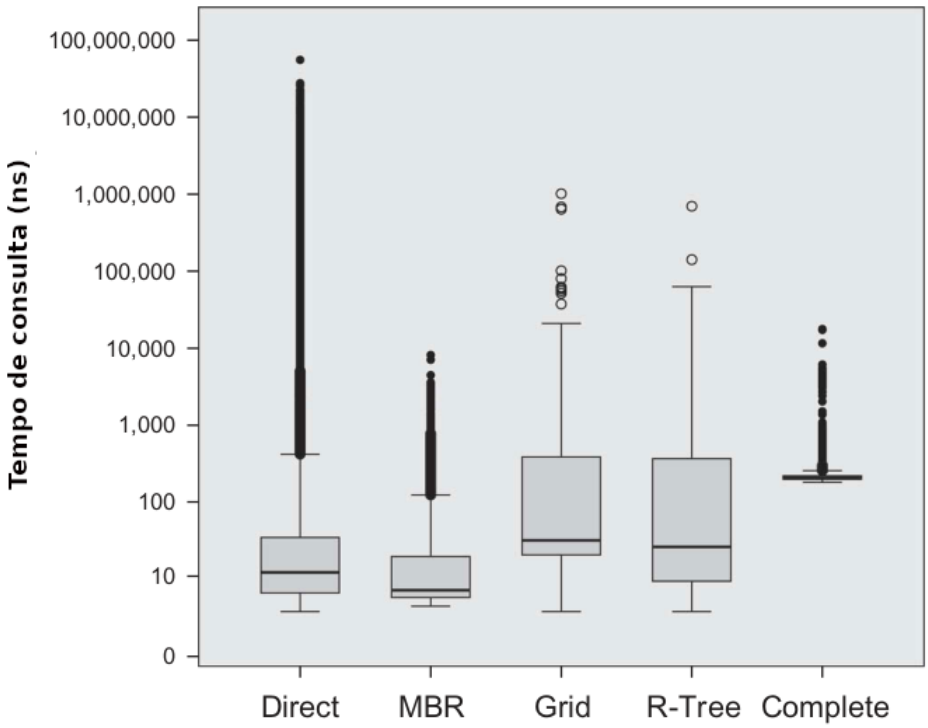


Figura 5.7: Gráfico com tempo de consulta utilizando técnicas de indexação qualitativa. Fonte:Imagem traduzida de Long (2016)[33].

Semelhante ao experimento realizado em [35], foram criadas 10.000 consultas à pares aleatórios de regiões para obter a relação entre esses itens. As consultas foram realizadas em 10 bases de dados distintas, divididas nos grupos Real-1 e Real-2, pertencentes a Tabela 5.2.

Para compor as consultas realizadas pelo experimento, os pares de objetos utilizados nas buscas foram definidos através de palavras-chave. Estas palavras correspondem ao nome das regiões existentes na base de dados e foram escolhidas de forma aleatória. O objetivo da consulta é recuperar da relação existente entre estes itens.

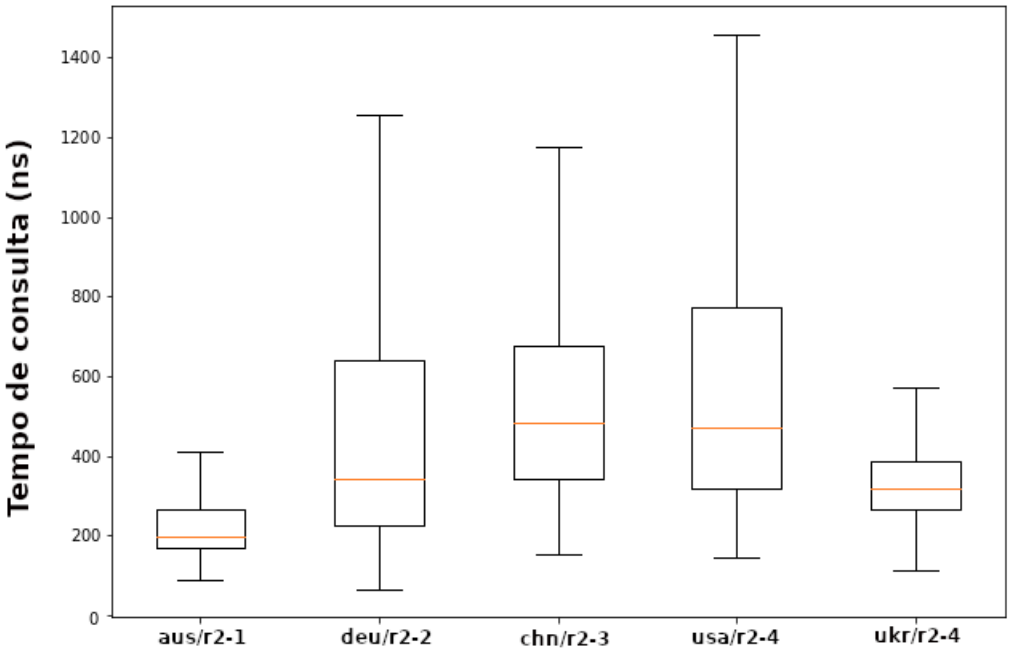


Figura 5.8: Gráfico com tempo de consulta do Topo-MSJ nos grupos Real-1 e Real-2.

A Figura 5.8 apresenta os resultados obtidos por execuções do Topo-MSJ nos grupos Real-1 e Real-2. Esta Figura contém resultados de 10 bases de dados apresentados na Tabela 5.2 com regiões politico-administrativa de países.

É possível observar que os resultados das execuções do algoritmo Topo-MSJ não conseguiram superar, em termos de desempenho, os resultados das abordagens comparadas na Figura 5.7. Os resultados apresentados na Figura 5.8 indicam que o algoritmo Topo-MSJ possui o tempo de execução de consulta mais alto que a maioria das abordagens comparadas, mas não apresenta uma elevada diferença em relação aos valores alcançados pelo *baseline*.

Comparando especificamente com a estrutura de indexação do tipo *complete*, é possível observar que o algoritmo Topo-MSJ, na base de dados da Austrália, possui um tempo de execução similar a esta estrutura. Conclui-se, portanto, que a solução proposta se aproxima, em termos de desempenho, de algumas técnicas do estado-da-arte na área de indexação qualitativa.

Na Figura 5.8, verifica-se que o tamanho das bases de dados possui uma relação com o tempo de execução das consultas. Uma maior quantidade de dados pode indicar um aumento da quantidade de relações existentes entre as regiões. No grupo Real-1, a base de dados dos Estados Unidos é a maior e também a que possui o tempo de execução mais elevado dentre as bases de dados avaliadas, seguido da China e Alemanha. As bases de dados do grupo

Real-2, possuem quantidades similares de regiões; em virtude disso, os números relativos às execuções foram semelhantes em todas as 5 execuções.

Uma possível sugestão de melhoria para as execuções realizadas pelo algoritmo Topo-MSJ seria o ajuste de parâmetros na construção da IR-Tree. A análise do uso de valores como *fanout* pode melhorar a distribuição e a quantidade de nós. Por exemplo, uma árvore com baixo *fanout* criada em memória é mais eficiente porque, embora seja mais profunda, precisa realizar menos comparações por pesquisa.

As execuções realizadas neste experimento utilizaram as mesmas condições técnicas da literatura. Uma máquina com processador Intel Core^{TM}-i7 3.6 GHz e memória RAM de 16GB foi utilizada para a realização das consultas. O equipamento possui as mesmas configurações usadas pelo trabalho considerado como *baseline*.

O experimento realizado nesta Seção busca entender se a solução do Topo-MSJ consegue, além de realizar consultas com grau de complexidade mais alto, ser uma alternativa na busca por relações mais simples no contexto de restrições qualitativas.

5.5 Considerações Finais

Esse capítulo apresentou a avaliação experimental da solução proposta. Dois experimentos foram realizados no contexto desta pesquisa. O Experimento 1 mostrou que o algoritmo Topo-MSJ é mais eficiente em relação às consultas em formato SQL, especificamente, no tempo de execução de consultas qualitativas. É possível constatar também, que o desempenho do Topo-MSJ é mais substancial em consultas com maior complexidade. No Experimento 2, o Topo-MSJ não conseguiu um tempo de consulta inferior em comparação com as soluções de indexação qualitativa, mas atingiu a média de alguns resultados. Na Tabela 5.5 é possível encontrar um resumo das questões de pesquisa e as conclusões obtidas com a realização dos experimentos.

Questão de Pesquisa	Conclusões Obtidas
QP1: O algoritmo Topo-MSJ permite realizar consultas em um tempo de execução inferior a uma consulta SQL realizada em um SGBD?	Sim, os resultados mostram que o tempo de execução de uma consulta realizada pelo Topo-MSJ é substancialmente menor comparado com consultas SQL em SGBDs. No Experimento 1 da Seção 5.3 é possível identificar isto.
QP2: A busca por grupos de POI utilizando relações qualitativas recupera todos os dados existentes possíveis à consulta?	Sim, os mesmos dados recuperados em consultas em SQL, são também obtidos pelo Topo-MSJ. É possível verificar no Experimento 1 da Seção 5.3.
QP3: O algoritmo Topo-MSJ permite realizar consultas com um tempo de execução inferior as abordagens que utilizam técnicas de indexação espacial?	Não, os resultados mostram que houve um tempo busca equivalente em relação as mesmas bases de dados utilizadas. Na Seção 5.4, o Experimento 2 apresenta essa comparação.
QP4: Um modelo para consultas espaciais baseado em Padrão Espacial Qualitativo (PEQ) permite representar as relações de conectividade entre as regiões espaciais?	Sim, o modelo é utilizado como padrão para criação da consulta qualitativa em todas as consultas da avaliação experimental. É possível verificar no Experimento 1 da Seção 5.3.
QP5: É possível utilizar dados espaciais no formato de polígono para representar POI em consultas qualitativas?	Sim, o uso da IR-Tree permite a indexação e consulta de regiões espaciais dentro do formato MBR. A configuração dos experimentos confirma isso na Seção 5.2
QP6: O agrupamento de vários tipos de relações de conectividade entre regiões pode ser realizado no contexto de consultas espaciais?	Sim, o algoritmo permite a realização de consultas com até quatro relações qualitativas diferentes de forma simultânea. As consultas do Experimento 1 (Seção 5.3) permitem comprovar isto.

Tabela 5.5: Quadro resumido contendo respostas às questões de pesquisa

Capítulo 6

Conclusão

Este capítulo apresenta as conclusões gerais deste trabalho. A Seção 6.1 apresenta uma discussão dos resultados e a Seção 6.2 discorre sobre as perspectivas para trabalhos futuros.

6.1 Discussão

Esta dissertação apresentou uma solução que permite a realização de consultas por POI utilizando relações qualitativas entre regiões espaciais. É possível constatar que o PEQ consegue definir um modelo para a criação de consultas por conectividade. Este padrão permite representar com eficiência as configurações espaciais qualitativas, sendo um novo modelo a ser considerado no contexto de técnicas para consultas de POI, baseadas em palavras-chave espaciais. O algoritmo apresentou flexibilidade quanto a implementação de novas funções para compor o modelo de PEQ.

No Experimento 1 (Capítulo 5), é possível observar que o grau de complexidade dos padrões em consultas tem impacto importante no tempo de execução das buscas em um SGBD. A vantagem no resultado do Topo-MSJ é mais evidente em um contexto superior a dois tipos de entidades (POI) espaciais diferentes. É importante destacar que o contexto dessa pesquisa considera a busca do POI como abordagem, mas a solução aqui apresentada pode ser generalizada para qualquer contexto de pesquisa por palavras-chave espaciais.

Provamos experimentalmente que o algoritmo apresentado nesta pesquisa possui um tempo de consulta similar às atuais soluções de indexação qualitativa. No Experimento 2 (Capítulo 5), foi utilizado apenas o tempo de execução das consultas como comparação. Observa-se, contudo, que a criação dos índices na maioria das soluções de RRQ necessitam de uma etapa chamada “qualificação”, que possui um custo de processamento mais lento que

a criação da IR-Tree utilizada pelo Topo-MSJ.

A solução proposta neste trabalho é adequada para o contexto de SIG, em cenários que permitem a criação de consultas com uma maior complexidade, considerando os tipos de relações entre entidades. O Topo-MSJ é também adequado ao uso em soluções com grandes bases de dados espaciais que possuam elementos textuais.

6.2 Trabalhos Futuros

Nesta seção, apresentam-se as sugestões e perspectivas de extensão desta pesquisa. São detalhados a seguir os pontos relevantes ao desenvolvimento destes objetivos futuros:

- **Ampliação dos tipos de relações espaciais do algoritmo.** Ampliar os tipos de relacionamentos utilizados nas consultas do Topo-MSJ é importante para abranger diferentes padrões espaciais. Pode-se destacar dois possíveis tipos de relações a serem implementados. O primeiro tipo é referente a uma categoria personalizada, pertencente ao campo de relações de conectividade entre regiões espaciais. Por exemplo, a pesquisa realizada em [24] sugere alguns tipos de relações em formatos personalizados, como: "visível", "entre" ou "próximo". Este tipo relação permite customizar a busca por padrões mais elaborados, que atendam à determinadas necessidades do contexto de uso. O segundo tipo de ampliação seria a combinação de relações já existentes na área de modelos qualitativos, como as relações direcionais e as relações de conectividade. Por exemplo, realizar a busca por um POI que, simultaneamente, esteja localizado vizinho e ao norte em relação a outro é uma forma de combinar tipos qualitativos em uma mesma consulta. A mensuração da área de regiões geométricas ou outras restrições quantitativas, geralmente combinadas em implementações de conectividade, são também alguns tipos possíveis de ampliação das relações espaciais.

- **Avaliação de correspondência parcial de relações para ranqueamento.** Existe uma definição de correspondência incompleta entre padrões espaciais chamada *Partial Spatial Pattern Matching*, este tipo de correspondência é avaliada quando a busca retorna um valor aproximado ao padrão utilizado na consulta. Sugere-se, portanto, a criação de uma ordem hierárquica nos resultados da consulta, com uma classificação baseada nas restrições de conectividade. Por exemplo, em uma busca por POI que estão no mesmo prédio, a relação *equals* é a restrição espacial a ser considerada na consulta.

Entretanto, a relação *overlap* pode ser apresentada como um tipo de predicado aproximado ao utilizado na consulta. Estes conceitos podem ser correlacionados com a definição de relaxamento em similaridade de cenas espaciais, permitindo o ranqueamento dos resultados em um contexto de RI. O atual estado do algoritmo Topo-MSJ permite apenas a busca por relações que possuam uma correspondência exata ao formato do PEQ.

- **Utilização do tempo como padrão de pesquisa.** A área de REQ define conceitos e modelos não apenas para o contexto espacial, como também para o contexto temporal. As bases de dados espaço-temporais são utilizadas amplamente nesta área. Um tipo de dado presente nestas bases é o *timestamp*, este tipo de dado permite a identificação da data que o dado foi inserido ou criado em um sistema de SIG. O uso de dados temporais na busca realizada pelo algoritmo pode permitir a visualização da mudança espacial em uma determinada região. Para o contexto de POI, se tornaria possível identificar em que ordem temporal os lugares surgiram ou quais estabelecimentos mudaram de localização com o passar do tempo.

- **Agregação de outras fontes textuais no contexto da pesquisa.** Como apresentado no Capítulo 2, existe um tipo de consulta espacial que utiliza predicados por junção. Este tipo de predicado permite a associação de diferentes dados textuais à busca. No cenário tecnológico atual, o uso de redes sociais e a geração constante de dados por parte dos usuários, possibilita a criação de bases com diferentes fontes de dados combinadas, permitindo assim, o enriquecimento do corpo textual presente no espaço de busca do algoritmo.

Bibliografia

[1] Rami Al-Salman. *Qualitative spatial query processing: Towards cognitive geographic information systems*. PhD thesis, Universität Bremen, 2014.

[2] Ahmed Loai Ali, Zoe Falomir, Falko Schmid, and Christian Freksa. Rule-guided human classification of volunteered geographic information. *ISPRS journal of photogrammetry and remote sensing*, 127:3–15, 2017.

[3] Ahmed Loai Ali, Nuttha Sirilertworakul, Alexander Zipf, and Amin Mobasheri. Guided classification system for conceptual overlapping classes in openstreetmap. *ISPRS International Journal of Geo-Information*, 5(6):87, 2016.

[4] Norbert Beckmann, Hans-Peter Kriegel, Ralf Schneider, and Bernhard Seeger. The r*-tree: An efficient and robust access method for points and rectangles. In *Proceedings of the 1990 ACM SIGMOD international conference on Management of data*, pages 322–331, 1990.

[5] Kenneth S Bøgh, Anders Skovsgaard, and Christian S Jensen. Groupfinder: a new approach to top-k point-of-interest group retrieval. *Proceedings of the VLDB Endowment*, 6(12):1226–1229, 2013.

[6] Tom Bruns and Max Egenhofer. Similarity of spatial scenes. In *Seventh international symposium on spatial data handling*, pages 31–42. Delft, The Netherlands, 1996.

[7] Fangjie CAO, Hanfa XING, Dongyang HOU, Haibin XU, Yuan MENG, and Xuan GUO. Research on identification and spatial patterns of commercial centers in beijing based on poi data. *Geomatics World*, page 01, 2019.

[8] Xin Cao, Lisi Chen, Gao Cong, Christian S Jensen, Qiang Qu, Anders Skovsgaard, Dingming Wu, and Man Lung Yiu. Spatial keyword querying. In *International Conference on Conceptual Modeling*, pages 16–29. Springer, 2012.

[9] Xin Cao, Gao Cong, Christian S Jensen, and Beng Chin Ooi. Collective spatial keyword querying. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data*, pages 373–384, 2011.

[10] Serina Chang, Emma Pierson, Pang Wei Koh, Jaline Gerardin, Beth Redbird, David Grusky, and Jure Leskovec. Mobility network models of covid-19 explain inequities and inform reopening. *Nature*, 589(7840):82–87, 2021.

[11] Hongmei Chen, Yixiang Fang, Ying Zhang, Wenjie Zhang, and Lizhen Wang. Espm: Efficient spatial pattern matching. *IEEE Transactions on Knowledge and Data Engineering*, 32(6):1227–1233, 2019.

[12] Dong-Wan Choi, Jian Pei, and Xuemin Lin. Finding the minimum spatial keyword cover. In *2016 IEEE 32nd International Conference on Data Engineering (ICDE)*, pages 685–696. IEEE, 2016.

[13] Dong-Wan Choi, Jian Pei, and Xuemin Lin. On spatial keyword covering. *Knowledge and Information Systems*, pages 1–36, 2020.

[14] Douglas Comer. Ubiquitous b-tree. *ACM Computing Surveys (CSUR)*, 11(2):121–137, 1979.

[15] Gao Cong and Christian S Jensen. Querying geo-textual data: Spatial keyword queries and beyond. In *Proceedings of the 2016 International Conference on Management of Data*, pages 2207–2212, 2016.

[16] Donatello Conte, Pasquale Foggia, Carlo Sansone, and Mario Vento. Thirty years of graph matching in pattern recognition. *International journal of pattern recognition and artificial intelligence*, 18(03):265–298, 2004.

[17] Ke Deng, Xin Li, Jiaheng Lu, and Xiaofang Zhou. Best keyword cover search. *IEEE Transactions on Knowledge and Data Engineering*, 27(1):61–73, 2014.

[18] Frank Dylla, Jae Hee Lee, Till Mossakowski, Thomas Schneider, André Van Delden, Jasper Van De Ven, and Diedrich Wolter. A survey of qualitative spatial and temporal calculi: Algebraic and computational properties. *ACM Computing Surveys (CSUR)*, 50(1):1–39, 2017.

[19] Max J Egenhofer and Robert D Franzosa. Point-set topological spatial relations. *International Journal of Geographical Information System*, 5(2):161–174, 1991.

[20] Yixiang Fang, Reynold Cheng, Gao Cong, Nikos Mamoulis, and Yun Li. On spatial pattern matching. In *2018 IEEE 34th International Conference on Data Engineering (ICDE)*, pages 293–304. IEEE, 2018.

[21] Yixiang Fang, Reynold Cheng, Jikun Wang, Lukito Budiman, Gao Cong, and Nikos Mamoulis. Spacekey: exploring patterns in spatial databases. In *2018 IEEE 34th International Conference on Data Engineering (ICDE)*, pages 1577–1580. IEEE, 2018.

[22] Yixiang Fang, Yun Li, Reynold Cheng, Nikos Mamoulis, and Gao Cong. Evaluating pattern matching queries for spatial databases. *The VLDB Journal*, 28(5):649–673, 2019.

[23] Paolo Fogliaroni and Heidelinde Hobel. Implementing naive geography via qualitative spatial relation queries. In *Proceedings of the 18th AGILE international conference on geographic information science*, 2015.

[24] Paolo Fogliaroni, Paul Weiser, and Heidelinde Hobel. Qualitative spatial configuration search. *Spatial Cognition & Computation*, 16(4):272–300, 2016.

[25] Tao Guo, Xin Cao, and Gao Cong. Efficient algorithms for answering the m-closest keywords query. In *Proceedings of the 2015 ACM SIGMOD international conference on management of data*, pages 405–418, 2015.

[26] Ramón Hermoso, Raquel Trillo-Lado, and Sergio Ilarri. Re-coskq: Towards pois recommendation using collective spatial keyword queries. In *CEUR workshop proc.*, number ART-2019-114041, 2019.

[27] Weihuang Huang, Guoliang Li, Kian-Lee Tan, and Jianhua Feng. Efficient safe-region construction for moving top-k spatial keyword queries. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 932–941, 2012.

[28] Hiroyuki Kanemitsu. Poix: Point of interest exchange language specification. *http://www. w3. org/TR/poix/*, 1999.

[29] Leif Harald Karlsen and Martin Giese. Qualitatively correct bintrees: an efficient representation of qualitative spatial information. *GeoInformatica*, 23(4):689–731, 2019.

[30] Vipin Kumar. Algorithms for constraint-satisfaction problems: A survey. *AI magazine*, 13(1):32–32, 1992.

[31] Huayu Li, Yong Ge, Defu Lian, and Hao Liu. Learning user's intrinsic and extrinsic interests for point-of-interest recommendation: A unified approach. In *IJCAI*, pages 2117–2123, 2017.

[32] Zhisheng Li, Ken CK Lee, Baihua Zheng, Wang-Chien Lee, Dik Lee, and Xufa Wang. Ir-tree: An efficient index for geographic document search. *IEEE Transactions on Knowledge and Data Engineering*, 23(4):585–599, 2010.

[33] Qi Liu, Yong Ge, Zhongmou Li, Enhong Chen, and Hui Xiong. Personalized travel package recommendation. In *2011 IEEE 11th International Conference on Data Mining*, pages 407–416. IEEE, 2011.

[34] Zhiguo Long, Matt Duckham, Sanjiang Li, and Steven Schockaert. Indexing large geographic datasets with compact qualitative representation. *International Journal of Geographical Information Science*, 30(6):1072–1094, 2016.

[35] Zhiguo Long, Hua Meng, Tianrui Li, and Sanjiang Li. Compact geometric representation of qualitative directional knowledge. *Knowledge-Based Systems*, 195:105616, 2020.

[36] Ahmed R Mahmood and Walid G Aref. Scalable processing of spatial-keyword queries. *Synthesis Lectures on Data Management*, 11(1):1–116, 2019.

[37] Ugo Montanari. Networks of constraints: Fundamental properties and applications to picture processing. *Information sciences*, 7:95–132, 1974.

[38] Jorge Pérez, Marcelo Arenas, and Claudio Gutierrez. Semantics and complexity of sparql. *ACM Transactions on Database Systems (TODS)*, 34(3):1–45, 2009.

[39] David A Randell, Zhan Cui, and Anthony G Cohn. A spatial logic based on regions and connection. *KR*, 92:165–176, 1992.

[40] Jochen Renz and Bernhard Nebel. On the complexity of qualitative spatial reasoning: A maximal tractable fragment of the region connection calculus. *Artificial Intelligence*, 108(1-2):69–123, 1999.

[41] Joao B Rocha-Junior, Orestis Gkorgkas, Simon Jonassen, and Kjetil Nørvåg. Efficient processing of top-k spatial keyword queries. In *International Symposium on Spatial and Temporal Databases*, pages 205–222. Springer, 2011.

[42] Stuart Russell and Peter Norvig. Artificial intelligence: a modern approach. 2002.

[43] Angela Schwering, Jia Wang, Malumbo Chipofya, Sahib Jan, Rui Li, and Klaus Broelemann. Sketchmapia: Qualitative representations for the alignment of sketch and metric maps. *Spatial cognition & computation*, 14(3):220–254, 2014.

[44] Daniel Sui, Sarah Elwood, and Michael Goodchild. *Crowdsourcing geographic knowledge: volunteered geographic information (VGI) in theory and practice*. Springer Science & Business Media, 2012.

[45] Roger F Tomlinson. *Thinking about GIS: geographic information system planning for managers*, volume 1. ESRI, Inc., 2007.

[46] Julian R Ullmann. An algorithm for subgraph isomorphism. *Journal of the ACM (JACM)*, 23(1):31–42, 1976.

[47] Emerson MA Xavier, Francisco J Ariza-López, and Manuel A Urena-Camara. A survey of measures and methods for matching geospatial vector datasets. *ACM Computing Surveys (CSUR)*, 49(2):1–34, 2016.

[48] Dongxiang Zhang, Yeow Meng Chee, Anirban Mondal, Anthony KH Tung, and Masaru Kitsuregawa. Keyword search in spatial databases: Towards searching by document. In *2009 IEEE 25th International Conference on Data Engineering*, pages 688–699. IEEE, 2009.

[49] Pengfei Zhang, Huaizhong Lin, Bin Yao, and Dongming Lu. Level-aware collective spatial keyword queries. *Information Sciences*, 378:194–214, 2017.

Apêndice A

Consultas SQL

Código Fonte A.1: Consultas SQL

```
1
2  //Consulta (a)
3  SELECT * FROM ny_test n1, ny_test n2, ny_test n3
4  WHERE ST_Intersects(n2.geom, n1.geom)
5  AND ST_Equals(n1.geom , n3.geom)
6  AND n1.top_category = 'Offices of Physicians'
7  AND n2.top_category = 'Automotive Repair and Maintenance'
8  AND n3.top_category = 'Personal Care Services';
9
10 //Consulta (b)
11 SELECT * FROM ca_test n1, ca_test n2, ca_test n3, ca_test n4
12 WHERE ST_Equals(n1.geom , n2.geom)
13 AND ST_Intersects(n2.geom,n3.geom)
14 AND ST_Disjoint(n3.geom, n4.geom)
15 AND ST_Distance(n3.geom, n4.geom)< 0.001
16 AND n1.top_category = 'Automotive Repair and Maintenance'
17 AND n2.top_category = 'Restaurants and Other Eating Places'
18 AND n3.top_category = 'Personal Care Services'
19 AND n4.top_category = 'Religious Organizations';
20
21 //Consulta (c)
22 SELECT * FROM ny_test n1, ny_test n2, ny_test n3, ny_test n4, ny_test n5
23 WHERE ST_Intersects(n1.geom,n2.geom)
24 AND ST_Equals(n2.geom , n3.geom)
25 AND ST_Touches(n3.geom, n4.geom)
26 AND ST_Disjoint(n5.geom, n4.geom)
```

```
27 AND ST_Distance(n5.geom, n4.geom)< 0.001
28 AND n1.top_category = 'Personal Care Services'
29 AND n2.top_category = 'Restaurants and Other Eating Places'
30 AND n3.top_category = 'Offices of Dentists'
31 AND n4.top_category = 'Religious Organizations'
32 AND n5.top_category = 'Automotive Repair and Maintenance';
33
34 //Consulta (d)
35 SELECT * FROM ny_test n1, ny_test n2, ny_test n3
36 WHERE ST_Disjoint(n3.geom, n1.geom)
37 AND ST_Distance(n3.geom, n1.geom)< 0.001
38 AND ST_Intersects(n2.geom , n1.geom)
39 AND n1.top_category = 'Offices of Physicians'
40 AND n2.top_category = 'Restaurants and Other Eating Places'
41 AND n3.top_category = 'Personal Care Services';
42
43 //Consulta (e)
44 SELECT * FROM ny_test n1, ny_test n2, ny_test n3, ny_test n4
45 WHERE ST_Equals(n1.geom , n2.geom)
46 AND ST_Intersects(n3.geom , n2.geom)
47 AND ST_Disjoint(n4.geom, n2.geom)
48 AND ST_Distance(n4.geom, n2.geom)< 0.001
49 AND n1.top_category = 'Offices of Dentists'
50 AND n2.top_category = 'Restaurants and Other Eating Places'
51 AND n3.top_category = 'Personal Care Services'
52 AND n4.top_category = 'Automotive Repair and Maintenance';
53
54 //Consulta (f)
55 SELECT * FROM ny_test n1, ny_test n2, ny_test n3, ny_test n4
56 WHERE ST_Intersects(n1.geom , n2.geom)
57 AND ST_Equals(n2.geom , n3.geom)
58 AND ST_Disjoint(n4.geom, n2.geom)
59 AND ST_Distance(n4.geom, n2.geom)< 0.001
60 AND n1.top_category = 'Offices of Physicians'
61 AND n2.top_category = 'Personal Care Services'
62 AND n3.top_category = 'Automotive Repair and Maintenance'
63 AND n4.top_category = 'Religious Organizations';
```

Apêndice B

Consultas SQL - LIKE

Código Fonte B.1: Consultas SQL (operador LIKE)

```
1
2  //Consulta (a)
3  SELECT * FROM ny_test n1, ny_test n2, ny_test n3
4  WHERE ST_Intersects(n2.geom, n1.geom)
5  AND ST_Equals(n1.geom , n3.geom)
6  AND n1.top_category LIKE '%Physicians%'
7  AND n2.top_category LIKE '%Automotive%'
8  AND n3.top_category LIKE '%Personal%';
9
10 //Consulta (b)
11 SELECT * FROM ca_test n1, ca_test n2, ca_test n3, ca_test n4
12 WHERE ST_Equals(n1.geom , n2.geom)
13 AND ST_Intersects(n2.geom,n3.geom)
14 AND ST_Disjoint(n3.geom, n4.geom)
15 AND ST_Distance(n3.geom, n4.geom)< 0.001
16 AND n1.top_category LIKE '%Automotive%'
17 AND n2.top_category LIKE '%Restaurants%'
18 AND n3.top_category LIKE '%Personal%'
19 AND n4.top_category LIKE '%Religious%';
20
21 //Consulta (c)
22 SELECT * FROM ny_test n1, ny_test n2, ny_test n3, ny_test n4, ny_test n5
23 WHERE ST_Intersects(n1.geom,n2.geom)
24 AND ST_Equals(n2.geom , n3.geom)
25 AND ST_Touches(n3.geom, n4.geom)
26 AND ST_Disjoint(n5.geom, n4.geom)
```

```
27 AND ST_Distance(n5.geom, n4.geom)< 0.001
28 AND n1.top_category LIKE '%Personal%'
29 AND n2.top_category LIKE '%Restaurants%'
30 AND n3.top_category LIKE '%Dentists'
31 AND n4.top_category LIKE '%Religious%'
32 AND n5.top_category LIKE '%Automotive%';
33
34 //Consulta (d)
35 SELECT * FROM ny_test n1, ny_test n2, ny_test n3
36 WHERE ST_Disjoint(n3.geom, n1.geom)
37 AND ST_Distance(n3.geom, n1.geom)< 0.001
38 AND ST_Intersects(n2.geom , n1.geom)
39 AND n1.top_category LIKE '%Physicians%'
40 AND n2.top_category LIKE '%Restaurants%'
41 AND n3.top_category LIKE '%Personal%';
42
43 //Consulta (e)
44 SELECT * FROM ny_test n1, ny_test n2, ny_test n3, ny_test n4
45 WHERE ST_Equals(n1.geom , n2.geom)
46 AND ST_Intersects(n3.geom , n2.geom)
47 AND ST_Disjoint(n4.geom, n2.geom)
48 AND ST_Distance(n4.geom, n2.geom)< 0.001
49 AND n1.top_category LIKE '%Dentists'
50 AND n2.top_category LIKE '%Restaurants%'
51 AND n3.top_category LIKE '%Personal%'
52 AND n4.top_category LIKE '%Automotive%';
53
54 //Consulta (f)
55 SELECT * FROM ny_test n1, ny_test n2, ny_test n3, ny_test n4
56 WHERE ST_Intersects(n1.geom , n2.geom)
57 AND ST_Equals(n2.geom , n3.geom)
58 AND ST_Disjoint(n4.geom, n2.geom)
59 AND ST_Distance(n4.geom, n2.geom)< 0.001
60 AND n1.top_category LIKE '%Physicians%'
61 AND n2.top_category LIKE '%Personal%'
62 AND n3.top_category LIKE '%Automotive%'
63 AND n4.top_category LIKE '%Religious%';
```

Apêndice C

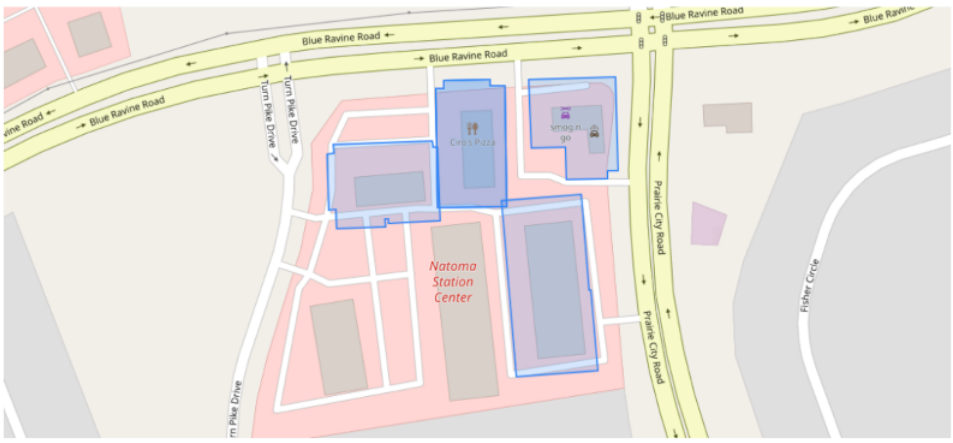
Exemplos de Resultados das Consultas

	id bigint	location_name character varying (1000)	top_category character varying (1000)	geom geography
1	113700	Summit Ridge Corporation	Automotive Repair and Maintenance	0103000020E6100000010000000500000004B995D68487F52C0AF57119F756044403FF97D42767...
2	202504	Katz Harvey MD	Offices of Physicians	0103000020E6100000010000000500000058D25ECE4F7F52C01D7ECB23616044406EDF439E7D...
3	202516	Chrystoph Marten	Personal Care Services	0103000020E6100000010000000500000058D25ECE4F7F52C01D7ECB23616044406EDF439E7D...

Figura C.1: Resultado da consulta (a) no PostgreSQL

	id bigint	location_name character varying (1000)	top_category character varying (1000)	geom geography
1	47903	Hungarian Reform Church	Religious Organizations	0103000020E6100000010000000500000003043706E017D52C0D7D0FC6254634440268
2	102027	Sang Yuan Body Works	Personal Care Services	0103000020E6100000010000000900000B323F1711F7D52C05B1BC84C6B6344404D9
3	189813	Technics Auto Repair	Automotive Repair and Maintenance	0103000020E6100000010000000700000002F31B1BEFF7C52C0B657BA1587634440667
4	189818	Jacques	Restaurants and Other Eating Places	0103000020E6100000010000000700000002F31B1BEFF7C52C0B657BA1587634440667

Figura C.2: Resultado da consulta (b) no PostgreSQL

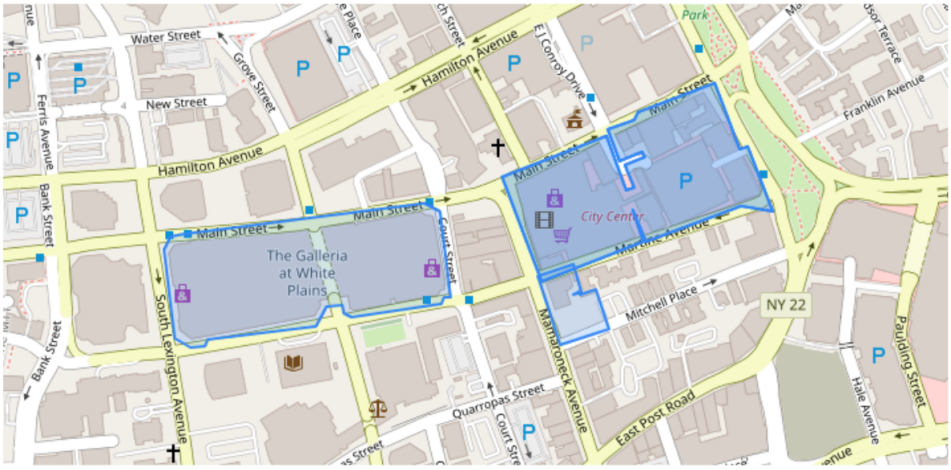


	id bigint	location_name character varying (1000)	top_category character varying (1000)	geom geography
1	133649	The Acne & Skin Care Clinic	Personal Care Services	0103000020E6100000010000000700000021257AC8D7...
2	397412	Ciros Pizza Cafe	Restaurants and Other Eating Places	0103000020E6100000010000009000000F3ACA415DF...
3	397413	Oleg Oliferuk DDS	Offices of Dentists	0103000020E6100000010000009000000F3ACA415DF...
4	557299	Smog N Go	Automotive Repair and Maintenance	0103000020E6100000010000009000008CDFD29ECE...
5	250626	New Song Christian Church	Religious Organizations	0103000020E6100000010000000F000000D5CF5B1AEA...

Figura C.3: Resultado da consulta (c) no PostgreSQL

	id bigint	location_name character varying (1000)	top_category character varying (1000)	geom geography
1	225478	Wolfgang Puck	Restaurants and Other Eating Places	0103000020E6100000010000000500000015EB867F988052C0CCE899A8
2	225484	Coffee Cuisine of Church St	Religious Organizations	0103000020E6100000010000000500000015EB867F988052C0CCE899A8

Figura C.4: Resultado da consulta (d) no PostgreSQL



	id bigint	location_name character varying (1000)	top_category character varying (1000)	geom geography
1	188951	Well Being Nails	Personal Care Services	0103000020E6100000010000000B0000003E086D92087152C0AA11C144068444409
2	9263	Freshii	Restaurants and Other Eating Places	0103000020E6100000010000001600000523EFABB077152C0CCF8315C048444401
3	9269	White Plains Dental	Offices of Dentists	0103000020E6100000010000001600000523EFABB077152C0CCF8315C048444401
4	256946	Sears Auto	Automotive Repair and Maintenance	0103000020E610000001000000 1A000000EAF68092667152C008767DF0168444407

Figura C.5: Resultado da consulta (e) no PostgreSQL

id bigint	location_name character varying (1000)	top_category character varying (1000)	geom geography
138847	Christ Apostolic Church	Religious Organizations	0103000020E6100000010000000B000000E1A2898F657C52C0F2230B03DB654440
88366	Fromer Eye Centers	Offices of Physicians	0103000020E6100000010000000900000FCD1404D697C52C08057C0908F654440
229846	Fancie Nail Beauty Spa	Personal Care Services	0103000020E6100000010000000B0000002162584971 7C52C09D509F35AB654440
229848	Cibao Auto Repair	Automotive Repair and Maintenance	0103000020E6100000010000000B0000002162584971 7C52C09D509F35AB654440

Figura C.6: Resultado da consulta (f) no PostgreSQL

Apêndice D

Código de implementação do PEQ

Código Fonte D.1: Código do PEQ em linguagem Java

```
1
2         HashSet<String> personal = new HashSet<>();
3         personal.add("Personal Care Services");
4
5         HashSet<String> physicians = new HashSet<>();
6         physicians.add("Offices of Physicians");
7
8         HashSet<String> restaurant = new HashSet<>();
9         restaurant.add("Restaurants and Other Eating Places");
10
11        HashSet<String> religious = new HashSet<>();
12        religious.add("Religious Organizations");
13
14        HashSet<String> dentist = new HashSet<>();
15        dentist.add("Offices of Dentists");
16
17        HashSet<String> automotive = new HashSet<>();
18        automotive.add("Automotive Repair and Maintenance");
19
20        //Criacao do PEQ
21        List<Link> queryPEQ = new ArrayList<Link>();
22
23        //Consulta (a)
24        queryPEQ.add(new Link(automotive, physicians, "overlap"));
25        queryPEQ.add(new Link(physicians, personal, "equal"));
26
```

```
27          //Consulta (b)
28          queryPEQ.add(new Link(automotive, restaurant, "equal"));
29          queryPEQ.add(new Link(restaurant, personal, "overlap"));
30          queryPEQ.add(new Link(personal, religious, "disjoint"));
31
32          //Consulta (c)
33          queryPEQ.add(new Link(personal, restaurant, "overlap"));
34          queryPEQ.add(new Link(restaurant, dentist, "equal"));
35          queryPEQ.add(new Link(dentist, religious, "touches"));
36          queryPEQ.add(new Link(religious, automotive, "disjoint"));
37
38          //Consulta (d)
39          queryPEQ.add(new Link(personal, physicians, "disjoint"));
40          queryPEQ.add(new Link(physicians, personal, "overlap"));
41
42          //Consulta (e)
43          queryPEQ.add(new Link(automotive, restaurant, "disjoint"));
44          queryPEQ.add(new Link(restaurant, dentist, "equal"));
45          queryPEQ.add(new Link(restaurant, personal, "overlap"));
46
47          //Consulta (f)
48          queryPEQ.add(new Link(physicians, personal, "overlap"));
49          queryPEQ.add(new Link(personal, automotive, "equal"));
50          queryPEQ.add(new Link(automotive, religious, "disjoint"));
```

Apêndice E

Ferramenta de Busca - Topokey

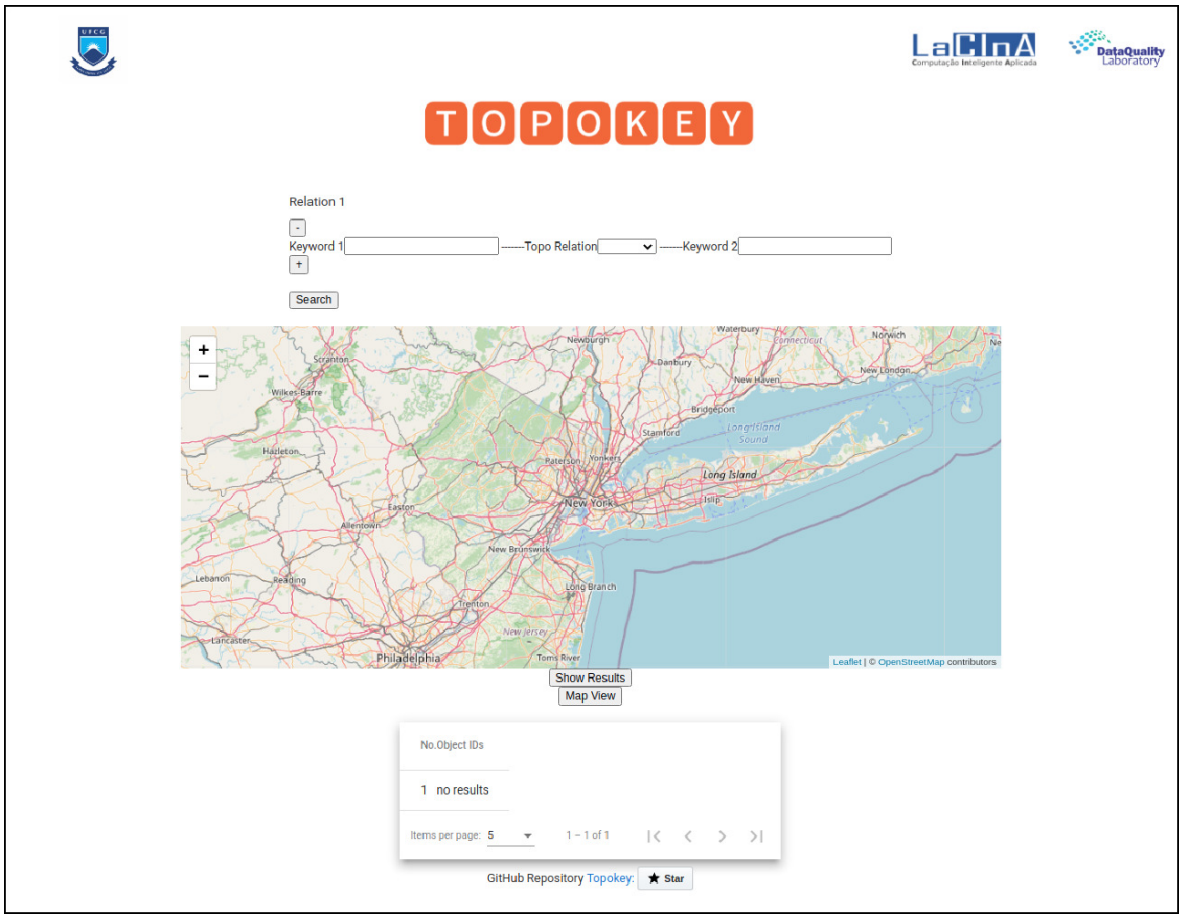


Figura E.1: Captura de tela da ferramenta Topokey.

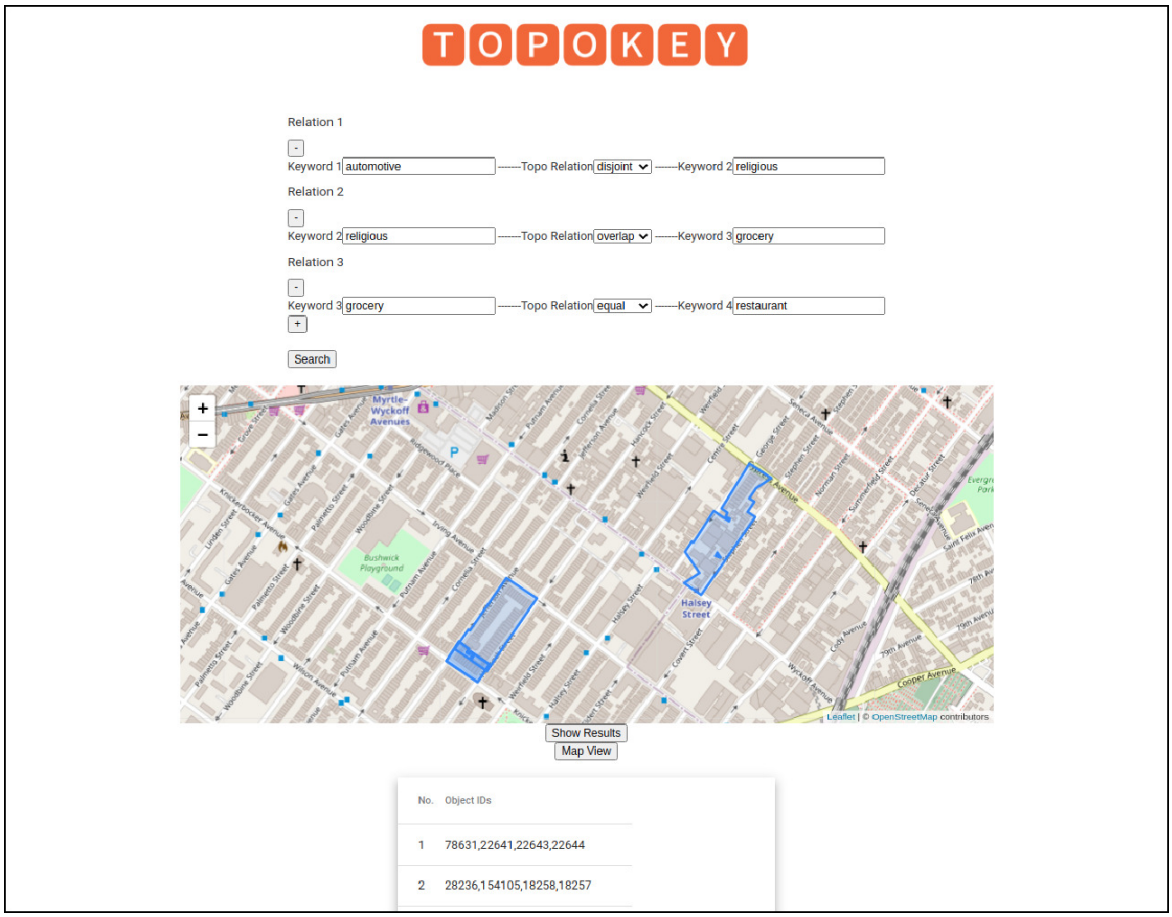


Figura E.2: Captura de tela da ferramenta Topokey com resultados da consulta.