Avaliação qualitativa da consulta Espaço-Textual Top-k

Alternative Title: Qualitative Evaluation of the Top-k Spatial-Textual Query

Luiz Felipe Naziazeno Neto
Instituto Federal de Alagoas
luiz.felipe@ifal.edu.br

João B. Rocha-Junior
Universidade Estadual de Feira de Santana
joao@uefs.br

Rodrigo Tripodi Calumby
Universidade Estadual de Feira de Santana
rtcalumby@ecomp.uefs.br

RESUMO

O número de pesquisas relacionas à consulta Espaço-Textual Top-k aumentou nos últimos anos. Isso se deve ao volume de dados com informação espacial (latitude e longitude) na Internet, o que gera necessidade de criação de métodos de busca eficientes. A maioria dos artigos que tratam este problema focam na eficiência dos métodos de busca, no entanto, é necessário avaliar também a eficácia da consulta no que se refere à relevância dos documentos recuperados para o usuário. Este artigo descreve uma metodologia para avaliar qualitativamente a consulta Espaço-Textual Top-k, e para criar coleções de referência espaço-textuais adaptadas de coleções tradicionais existentes, como a coleção Reuters-21578. Os testes realizados avaliaram as duas funções de ranqueamento existentes e indicaram que o balanceamento entre a relevância textual e a distância espacial é crucial para atingir melhores resultados.

Palavras-Chave

Função de ranqueamento; Avaliação da eficácia; Consulta Espaço-Textual.

ABSTRACT

The number of researches related to the Top-k Spatio-Textual Query has increased in recent years. This is due to the volume of data with spatial information (latitude and longitude) on the Internet, which necessitate the creation of efficient search methods. Most of the researches that address this problem focus on the search methods efficiency, however, It is also necessary to evaluate the queryt's effectiveness. This article describes the methodology for qualitatively evaluating a Top-k Spatio-Textual Query, and one method to create space-textual reference collections adapted from traditional collections, such as the Reuters-21578 collection. The tests performed in the two ranking functions indicate that the balance between textual relevance and spatial distance is crucial for better results.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SBSI 2017 June 5^{th} – 8^{th}, 2017, Lavras, Minas Gerais, Brazil

Copyright SBC 2017.

CCS Concepts

•**Information systems** → Information retrieval; Evaluation of retrieval results; Retrieval effectiveness;

Keywords

Ranking Function; Efficacy Assessment; Space-textual query.

1. INTRODUÇÃO

E cada vez mais comum encontrar grandes bases de dados com informação textual e espacial (latitude e longitude). Isto explica o interesse por novas técnicas para extrair informações relevantes destas bases de dados. As consultas espaciais por palavras-chave estão auxiliando diversas aplicações, tais como o Google Maps[1], na qual pontos de interesse podem ser encontrados; o Foursquare[2], na qual documentos georreferenciados com recomendações de lugares de interesse (Ex: bares e restaurantes) podem ser recuperados; o Twitter[3], na qual tweets com informação geográfica (latitude e longitude) podem ser retornados.

Uma maneira de recuperar informação com a característica espaço-textual, que vem destacando-se recentemente, é o uso da consulta espaço-textual top-k, que retorna os k melhores documentos ordenados por uma pontuação, que é calculada levando-se em consideração a distância espacial e a relevância textual dos documentos em relação ao local da consulta [9].

Em um sistema de recuperação de informações ordenado, a consulta espaço-textual top-k apresenta um desafio fundamental, que é estimar quais documentos os usuários considerarão relevantes e quais os usuários considerarão irrelevantes. Essa tarefa é executada por uma função de ranqueamento [18, 20, 2], cujo objetivo principal é incluir os documentos mais relevantes no topo do ranking.

A função de ranqueamento da consulta espaço-textual top-k tem como objetivo fazer o balanceamento entre a proximidade espacial e a relevância textual, com o intuito de ranquear os melhores documentos para uma determinada consulta [3]. A depender do valor desse balanceamento, o resultado obtido pela consulta pode ser prejudicado qualitativamente, retornando documentos indesejados para o usuário.

Em sistemas de recuperação de informação a relevância dos documentos retornados por uma consulta tem um papel primordial, que é maximizar a recuperação de documentos

[1]https://maps.google.com/
[2]https://foursquare.com/
[3]https://twitter.com/

relevantes e minimizar a recuperação de documentos irrelevantes. Assim, para verificar como um sistema de recuperação de informação está se comportando em relação a uma necessidade do usuário, é necessário uma avaliação da qualidade dos resultados retornados.

As abordagens de avaliação de um sistema de recuperação de informação podem ser centradas no usuário e/ou centradas no sistema. As abordagens centradas no usuário focam em descobrir como as preferências dos usuários podem ser afetadas pelas características da interface de um sistema de recuperação de informaçãoo e pela sua facilidade de uso [24]. As centradas no sistema são as mais tradicionais e dominantes na literatura e são chamadas de abordagens Cranfield [2]. Dispensa-se a influência de aspectos humanos. Os experimentos são facilmente escaláveis e reproduzíveis, o que facilita uma comparação quantitativa entre distintas implementações. E para facilitar a comparação emprega-se coleções de referência padronizadas (ex: Crafielnd [13] e TREC [23, 2]).

Apesar de oferecerem um tratamento mais completo da relevância, os métodos centrados no usuário são difíceis de reproduzir, de projeto complexo e caros [10]. Isso acontece, pois podem ser muito dependentes da interface do sistema, da experiência do usuário com aplicações de recuperação de informação, entre outras coisas. Devido a isso e levando em consideração o custo de implementação, este artigo utiliza a abordagem centrada no sistema para avaliar a consulta espaço-textual top-k.

O objetivo deste é apresentar uma metodologia de criação de uma coleção espaço-textual sitética a partir da coleção Reuters-21578 com a finalidade de realizar uma avaliação qualitativa da consulta espaço-textual top-k [9]. Para o melhor do nosso conhecimento, esta avaliação nunca foi relizada anteriormente.

2. TRABALHOS RELACIONADOS

A avaliação dos sitemas de recuperação tem sido sempre uma questão importante no âmbito acadêmico. Consequentemente, muitos trabalhos científicos têm focado em modelos de avaliação de sistemas de recuperação de informação, resultando em um grande número de métricas de avaliação, tais como, *Average Precision*, *Discounted Cumulative Gain* (DCG) [11], *Expected Reciprocal Rank* (ERR) [6], *Expected Browsing Utility* (EBU) [25], etc.

Apesar das métricas citadas acima serem importantes para avaliação de sistemas de recuperação, elas foram originadas de modelos de avaliação utilizando julgamentos de especialistas. O custo para aplicar esses julgamentos é muito alto [5]. Como resultado, várias abordagens foram desenvolvidas, incluindo a construção de conjunto de testes de avaliação de baixo custo e a avaliação de coletas de teste nos casos em que os julgamentos de especialistas estão ausentes. Os métodos de avaliação de baixo custo podem ser vistos nas pesquisas de Carterette e Alan [4] e Sanderson e Joho [22].

Este artigo utiliza uma abordagem de avaliação de baixo custo, utilizando o paradigma de Cranfield [8]. Esse paradigma pressupõe um conjunto de documentos, um conjunto de consultas e um conjunto de julgamentos de relevância para cada par consulta-documento. A união desses conjuntos é chamada de coleção e diversas iniciativas de apoio e fomento à pesquisa para disponibilizar essas coleções foram criadas, tais como: CLEF, NTCIR, OHSUMED, Reuters-21578, INEX, entre outras.

No entanto, os julgamentos dessas coleções são baseados somente entre os termos da consulta e os textos dos documentos. Não existe um julgamento quanto à localização do documento em relação à consulta, ou seja, não existe uma avaliação de qual documento é mais relevante espacialmente que o outro. Neste sentido, a presente pesquisa utiliza uma das coleções tradicionais existentes e gera, sinteticamente, coleções espaço-textuais através de critérios que serão definidos em seções posteriores.

3. REVISÃO DA LITERATURA

Esta Seção aborda a revisão da literatura utilizada nesta pesquisa, consistindo em estudos sobre a consulta espaciais por palavra-chave, que dão a base para o entendimento da consulta espaço-textual top-k. Além disso, é feito um estudo sobre a área de Recuperação da Informação, abordando, principalmente, aspectos que dão suporte para avaliação qualitativa de uma consulta.

3.1 Consultas Espaciais Por Palavra-Chave

Com o aumento de objetos online com informação textual e espacial (latitude e longitude) a internet torna-se um ambiente dimensionamente espacial [7]. Os usuários com smartphones, tablets e outros aparelhos com GPS (*Global Positioning System*) e os conteúdos da web estão cada vez mais geo-posicionados e geocodificados. Além disso, pontos de interesse estão cada vez mais disponíveis na web.

Esse desenvolvimento necessita de técnicas que permitam a indexação de dados que contenham descrições textuais e espaciais. Uma dessas técnicas é a consulta espacial por palavra-chave que tem como arqumentos a localização espacial e um conjunto de palavras-chave, retornando conteúdo levando em consideração a informação textual e a localização espacial [3].

São três os tipos de consulta espacial por palavra-chave mais importantes: A consulta kNN (*k-Nearest-Neighbor*) booleana, a consulta *range* booleana e a consulta kNN top-k [7].

Consulta kNN Booleana. Retorna os k objetos mais próximos ao local do usuário (representado por um ponto), tal que cada descrição textual contém as palavras-chave da consulta.

Consulta kNN Top-k. Retorna os k objetos ordenados por um ranking levando em consideração a distância dos objetos em relação à localização da consulta e a relevância textual dos objetos em relação às palavras-chave da consulta. Essa consulta possui o mesmo número de argumentos que a consulta kNN booleana, no entanto os critérios de ranqueamento são definidos levando em consideração a proximidade espacial e a relevância textual.

Consulta *Range* Booleana. Retorna todos os objetos cuja descrição textual contém as palavras-chave da consulta e cuja localização fica a menos de uma determinanda distância (especificada pelo usuário - *range*) do local da consulta.

3.2 Recuperação de Informação

A área de Recuperação de Informação (RI) tem como foco principal o usuário e suas necessidades para, principalmente, oferecer de maneira eficaz, o acesso à informação. Recuperar informação é encontrar material (comumente documentos) de um ambiente não estruturado (normalmente texto) dentro de grandes coleções, que satisfaça uma necessidade de informação [16].

O problema de RI. A idéia de um sistema de RI é recuperar mais documentos relevantes e, consequentemente, menos documentos irrelevantes. Neste sentido, entender a relevância é fundamental.

Uma questão importante é que a relevância é um julgamento pessoal, que deriva de um problema a ser resolvido e do seu contexto. Por exemplo, à medida que novas informações vão sendo disponibilizadas, a relevância pode mudar [2]. Ou seja, tempo é uma propriedade importante relacionada à relevância. Além disso, o local de interesse tem um impacto significativo na relevância (e.g., um restaurante mais próximo pode ser a resposta mais relevante).

3.3 Avaliação de Sistemas de Recuperação

O surgimento de novas técnicas de Recuperação de Informação demandou como resultado a criação de novas técnicas de avaliação dos resultados. Para avaliar um sistema de Recuperação de Informação é fundamental estimar o quão bem o sistema atende às necessidades de informação do usuário.

É díficil fazer esse cálculo, pois o mesmo conjunto resposta pode ter diversas interpretações por diferentes usuários. No entanto, pode ser feita definição de métricas aproximadas que, na média, podem ter uma relação entre as preferências de uma amostra e a população de usuários [2]. Algumas destas métricas serão apresentadas na Seção 3.6.

3.4 Paradigma de Cranfield

O paradigma de Cranfield [8] é uma abordagem, que, apesar de antiga, é muito usada nos dias de hoje [5]. Ela tem como vantagens: o custo relativamente baixo, a avaliação de um sistema de Recuperação de Informação pode ser feita rapidamente e seus resultados podem ser reproduzidos posteriormente para fins de verificação (repetibilidade). Além disso, pode ser aplicado focando tipos particulares de necessidades de informação (faces, plantas, imagens de satélite, imagens médicas e *Web*).

Os experimentos de Cleverdon [8] culminaram em métricas modernas de avaliação da recuperação de informação, que serão abordadas na Seção 3.6. Para avaliar os resultados das consultas é importante saber a relevância dos objetos em relação à consulta. O conhecimento dos resultados dos julgamentos de relevância é fundamental uma vez que para avaliação de um sistema de Recuperação de Informação, utilizando o paradigma de Cranfield, é necessário coleções de referência com 3 conjuntos fundamentais: um conjunto de documentos, um conjunto de consultas e um conjunto de julgamentos de relevância.

3.5 Coleções de Referência

Uma coleção de referência é constituida de: um conjunto de documentos; um conjunto de consultas; e, para cada consulta, um conjunto de julgamentos de relevância. Esse conjunto de julgamentos de relevância é identificado manualmente através de um processo que envolve um esforço humano significativo [2]. Esse conjunto pode ser entendido através da função representada pela Equação 1.

$$rel(d, q) \in [0, 1] \tag{1}$$

onde d é um documento do conjunto de documentos e q é uma consulta. De maneira geral, a Equação 1 é definida em um intervalo de 0 a 1. 0 siginifica que o documento é completamente irrelevante para a consulta e 1 significa que o documento é completamente relevante para a mesma. Outros valores dentro desse intervalo representa o nível de relevância.

Coleções de Referência Existentes. Atualmente existem inúmeras iniciativas de apoio e fomento à pesquisa, campanhas de avaliação de sistemas, desafios de pesquisa, etc. relacionadas à área de RI, tais como: CLEF, NTCIR, OHSUMED, Reuters, INEX, entre outras. Uma das coleções mais utilizadas para realização de experimentos em recuperação de informações é a *Reuters-21578*. São ao todo 21578 artigos da agência de notícias *Reuters* [15]. Essa coleção possui aproximadamente 20Mb de tamanho e não fornece consultas para avaliação. Uma abordagem de criar tais consultas *Reuters-21578* foi proposta por [21] e está disponível para download no endereço [17].

3.6 Métricas de Avaliação

Como foi descrito na Seção 3.3, para avaliar até que ponto um sistema de Recuperação de Informação atende às necessidades do usuário, devem ser usadas métricas quantitativas para estimar a eficácia. Levando-se em consideração um conjunto de necessidades de informação, seja R o conjunto de documentos relevantes e A o conjunto-resposta gerado pela execução de uma consulta que processa a necessidade de informação [2].

Revocação. Revocação ou também chamado de Sensibilidade (*Sensitivity*) é a fração do número de documentos relevantes recuperados pelo número total de documentos relevantes que existem na coleção [2].

$$Revocacão = \frac{|R \cap A|}{|R|} \tag{2}$$

Precisão. Precisão é a fração de documentos relevantes recuperados em relação número de documentos recuperados [2].

$$Precisão = \frac{|R \cap A|}{|A|} \tag{3}$$

Precisão Média. É computada como a média das precisões em um ponto do conjunto-resposta, para cada documento relevante. Ou seja, através de uma lista de documentos recuperados por uma consulta, calcula-se a precisão em um determinado ponto para cada documento relevante e em seguida calcula-se a média aritmética dessa precisão [16, 2].

$$PrecisãoMédia = \frac{\sum_{n=1}^{|A|} (Precisao@n \times rel(n))}{|R \cap A|} \tag{4}$$

onde, n é a posição de um objeto no ranking, $Precisão@n$ é a precisão desse objeto e $rel(n) = 1$, caso o objeto na posição n seja relevante, e $rel(n) = 0$, caso contrário.

MAP - *Mean Average Precision*. O MAP é uma métrica de eficácia que resume em um único valor as precisões médias de cada consulta. O cálculo é feito como uma média aritmética das precisões médias. Assim, calcula-se as precisões médias de múltiplas consultas e depois realiza-se uma média aritmética dessas precisões médias [2].

$$MAP = \frac{\sum_{q=1}^{|Q|} PrecisaoMedia(q)}{|Q|} \tag{5}$$

4. DEFINIÇÃO DO PROBLEMA

A consulta espaço-textual top-k é uma consulta espacial por palavra-chave [7], onde utiliza-se a localização do usuário, um conjunto de palavras-chave, fornecidas por ele, e o número de resultados como parâmetros. A consulta identifica objetos que são espacialmente próximos à localização do usuário, e textualmente relevantes às palavras-chave, retornando os k objetos que melhor atendem a estas duas características (proximidade espacial e relevância textual). Uma função de ranqueamento avalia a proximidade espacial entre um objeto e o usuário, além da relevância textual da descrição do objeto considerando o conjunto de palavras-chave. A resposta desta consulta é ordenada a partir dos valores de escore gerados para cada objeto pela função de ranqueamento [3]. Essa função possui parâmetros que, a depender dos seus valores, podem afetar qualitativamente a relevância do resultado para o usuário.

Uma das funções de ranqueamento utilizada na execução de uma consulta espaço-textual top-k é a função representada pela Equação 6 [9].

$$rank(p,q) = \alpha \times \delta(p.l, q.l) + (1-\alpha) \times \theta(p.d, q.d) \quad (6)$$

onde p é o objeto, q é a consulta, $p.l$ é a localização do objeto, $q.l$ é a localização da consulta, $p.d$ é o texto vinculado ao objeto e $q.d$ representa as palavras-chave vinculadas à consulta.

A função interna $\delta(p.l, q.l)$ calcula a proximidade espacial Euclidiana entre a localização da consulta e a localização do objeto. A função interna $\theta(p.d, q.d)$ calcula a relevância textual entre o texto vinculado ao objeto e as palavras-chave vinculadas à consulta. Tanto a função de proximidade espacial quanto a função de relevância textual retornam valores entre 0 e 1. O parâmetro de balanceamento $\alpha \in [0,1]$ e é utilizado para definir a importância de um critério (proximidade espacial ou relevância textual) sobre a outra.

O cálculo da proximidade espacial Euclidiana da função de ranqueamento, representada pela Equação 6, é calculado através da Equação 7. A Equação 6 normaliza o valor da distância entre a localização do objeto e a localização da consulta. O d_{max} é a maior distância entre dois pontos no espaço e normaliza da distância Euclidiana para que os valores fiquem no intervalo de [0,1] [9].

$$\delta(p.l, q.l) = 1 - \frac{d(p.l, q.l)}{d_{max}} \quad (7)$$

Além da função de ranqueamento representada pela Equação 6, uma outra função de ranqueamento pode ser utilizada para uma consulta espaço textual top-k, representada pela Equação 8 [19].

$$rank(p,q) = \frac{\theta(p.d, q.d)}{1 + \alpha \times \delta(p.l, q.l)} \quad (8)$$

Relevância Textual (θ). A relevância textual das duas equações pode ser computada utilizando a Equação 9 [9]. A função adota uma conhecida maneira de calcular a relevância textual, chamada de cosseno [1].

$$\theta(p.d, q.d) = \frac{\sum_{t \in q.d} w_{t,p.d} \times w_{t,q.d}}{\sqrt{\sum_{t \in p.d}(w_{t,p.d})^2 \times \sum_{t \in q.d}(w_{t,q.d})^2}} \quad (9)$$

onde $w_{t,p.d}$ representa o peso do termo t no objeto $p.d$. A Equação 9 calcula o cosseno entre os dois vetores ($w_{t,p.d}$ e $w_{t,q.d}$), de forma que quanto mais próximo de 1 seja o cosseno, mais similares são os objetos.

O objetivo deste trabalho é avaliar a consulta espaço-textual no que se refere a qualidade dos objetos retornados para o usuário e comparar as funções de ranqueamento para identificar o desempenho qualitativo das duas funções.

5. MÉTODOS DE AVALIAÇÃO

Nesta Seção será apresentado os métodos de avaliação de uma consulta espaço-textual por palavra-chave com as etapas necessária e em que ordem essas etapas devem ser executadas para realização de testes, dentro de um modelo do mundo real.

5.1 Etapas para Avaliação

Na Figura 1, pode-se verificar as etapas de avaliação proposta. Os experimntos e os testes são geralmente destinados a responder a perguntas específicas. Neste trabalho em particular, a pergunta a ser respondida é: Qual é a eficácia da consulta espaço-textual top-k?

Segundo [12], eficácia é o quão bem o sistema faz o que é suposto fazer. Assim, seus benefícios são os ganhos decorrentes do que o sistema faz e sua eficiência é o quão barato ele faz o que é para fazer, documentos relevantes e suprimindo documentos irrelevantes.

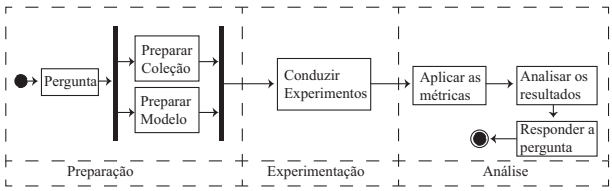


Figura 1: Etapas para avaliação.

A fase de preparação envolve todos os componentes necessários para a realização de uma avaliação adequada. Isso inclui, preparar uma coleção de referência para que o sistema seja testado, bem como escolher um modeo de recuperação de informação que execute as consultas nos testes. Portanto, é apresentado a seguir a maneira utilizada neste trabalho para preparar a coleção espaço-textual, atribuindo uma informação espacial em cada documento da coleção.

Preparar Coleção Espaço-Textual. Cada coleção espaço-textual deve conter uma consulta, um conjunto de documentos espaço-textuais e um conjunto de julgamentos de relevância para cada par documento-consulta. Essas coleções podem ser criadas utilizando coleções existentes. Para demonstrar a metodologia de inclusão de informação espacial em coleções tradicionais, foi escolhida a coleção Reuters-21578.

Coleção Reuters-21578. Uma das coleções mais utilizadas para realização de testes em recuperação de informações [15] e se diferencia das outras coleções pelo fato de não possuir um conjunto de consultas com seus respectivos documentos relevantes [21]. No entanto, cada documento na coleção Reuters-21578 está inserido em uma categoria e são essas categorias que fazem com que a Reuters-21578 seja utilizada como uma coleção. Esse uso, utilizando categorias, foi primeiro descrito por Lewis [14] e pouco tempo depois por

Sanderson [21].

Os documentos desta coleção são notícias divulgadas no período entre 1987 e 1991 [15], formando uma coleção composta por 21578 documentos organizados em 5 categorias e suas respectivas sub-categorias. Esses documentos contém texto, título, autores, data da publicação, entre outras informações. A Tabela 1 apresenta as categorias e o número de sub-categorias.

Tabela 1: Categorias e Número de Sub-Categorias da Coleção *Reuters*-21578

Categorias	Número de sub-categorias
EXCHANGES	39
ORGS	56
PEOPLE	267
PLACES	175
TOPICS	135

Para este trabalho, cada coleção espaço-textual é criada a partir de uma categoria da coleção Reuters-21578. Para criação das consultas, é feita uma leitura em todos os 21578 artigos da coleção Reuters-21578 para verificar as combinações de sub-categorias existentes dentro da categoria. Se um documento estiver em uma sub-categoria apenas, a consulta terá somente uma palavra-chave. Exemplo: se um documento estiver na sub-categoria "acq" a palavra-chave da consulta será "acq". Portanto, foram selecionadas as combinações existentes nos documentos e agrupadas em categorias: com uma, duas, três, quatro e com cinco palavras-chave.

Tomando como base o exemplo anterior, a partir da subcategoria "acq" é gerada uma coleção, verificando todos os documentos existentes na coleção Reuters-21578 que pertecem somente à categoria "acq". Em seguida, é criado um ranking desses documentos para compor o conjunto de julgamentos de relevância da consulta "acq". Os documentos que não pertecem à sub-categoria "acq" são considerados irrelevantes. Além disso, os documentos pertencentes ao conjunto de julgamento de relevância e os documentos não pertencentes a este conjunto recebem uma informação espacial, que será descrito a seguir. Logo, a coleção espaço-textual "acq" possui: uma consulta com a palavra-chave "acq", um conjunto de documetos espaço-textual formados por documentos pertencentes ao conjunto de julgamentos de relevância e documentos não pertencentes a este conjunto e um conjunto de documentos de relevância.

O ranking estabelecido em cada conjunto de julgamento de relevância é baseado no cosseno como função de similaridade [26].

A Tabela 2 apresenta as 237 coleções criadas com 1, 2, 3, 4 e 5 palavras-chave cada uma. Além disso, cada uma destas coleções é composta por 50 consultas, exceto a coleção com 5 palavras-chave que tem apenas 37 consultas, que é o número de artigos na coleção Reuters-21578 com 5 palavras-chave distintias.

Inclusao da informação espacial nas coleções de referência. Uma coleção tradicional não possui uma informação espacial (latitude e longitude), fazendo-se necessário

Tabela 2: Número de coleções espaço-textuais agrupadas pela quantidade de palavras-chave.

Número de palavras-chave	Número de coleções
1	50
2	50
3	50
4	50
5	37

incluir essa informação em todo documento da coleção para que a consulta espaço-textual top-k seja executada. Para isso, inicialmente, faz-se uma varredura em todos os documentos da coleção e, para cada documento, coordenadas geográaficas são geradas de maneira aleatória.

Para cada documento pertencente ao conjunto de julgamentos de relevância, as coordenadas são associadas a um raio r, cujo ponto central é formado pelas coordenadas onde está sendo realizada a consulta. O raio aumenta na proporção de 1000 em 1000, critério estabelecido neste trabalho, a medida que novos pontos relevantes são encontrados. Utilizou-se o Teorema de Pitágoras para calcular a distância Euclidiana do ponto aleatório (P_a) até o ponto da consulta (P_c) com base nas coordenadas cartesianas (x_3,y_3)e (x_1,y_1), como mostra a Figura 2.

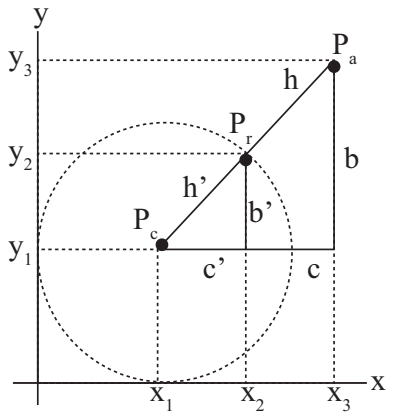


Figura 2: Representação matemática da aplicação das teorias de triângulos para os pontos relevantes, sendo P_a (Ponto aleatório), P_r (Ponto relevante) e P_c (Ponto da consulta).

No segundo momento, foi utilizado o Teorema de Tales para estabelecer uma relação matemática entre os pontos aleatório (P_a) e o relevante (P_r). A Equação 10 e a Equação 11 mostram como o cálculo é feito em relação a Figura 2.

$$\frac{h'}{h} = \frac{b'}{b} = \frac{c'}{c} \tag{10}$$

$$\frac{h'}{h} = \frac{x_2 - x_1}{x_3 - x_1} = \frac{y_2 - y_1}{y_3 - y_1} \tag{11}$$

A Figura 3 apresenta a criação de uma coleção de referência espaço-textual a partir da adaptação de uma coleção existente. Primeiro os objetos espaço-textuais relevantes, que foram ranqueados utilizando a função cosseno [26], são posicionados aleatoriamente em uma distância de 1000 em 1000.

Os objetos espaço-textuais não relevantes são distribuidos aleatoriamente na circuferência gerada, com raio 3000 como mostra a Figura 3. Isso é feito para garantir que os itens não relevantes estejam no mesmo raio de acesso que os itens Relevantes, garantindo que existam tanto objetos relevantes quanto objetos não relevantes nos arredores do Ponto da Consulta e com similar distribuição espacial. Isso não necessariamente ocorre em um contexto real, mas foi adotado como solução para evitar desbalanceamento na distribuição espacial dos itens relevantes e não relevantes.

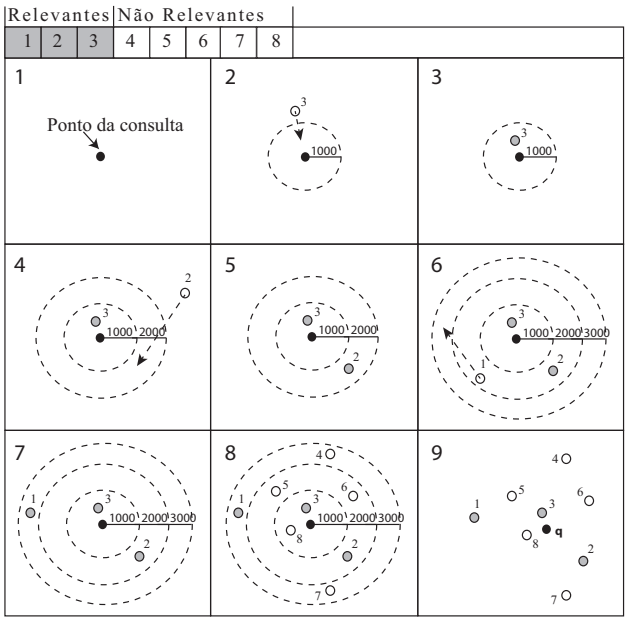


Figura 3: Representação da criação de uma coleção espaço-textual a partir da adaptação de uma coleção existente.

6. RESULTADOS E DISCUSSÃO

As duas funções de ranqueamento utilizadas nos experimentos são as representadas pela Equações 6 e pela Equação 8. Foram selecionadas, para apresentação dos resultados, 5 bases de dados (das 237 coleções utilizadas nos testes experimentais) com características distintas, que proporcionaram resultados mais significativos que os demais em cada agrupamento (grupo de coleções com 1 palavra-chave, 2, 3, 4 e 5 palavras-chave). A Base1 pertence ao grupo de coleções com uma palavra-chave (cocoa), a Base2 ao grupo com duas (acq e ship), a Base3 ao grupo com 3 (grain, corn e oat), a Base4 ao grupo com 4 (cocoa, coffee, sugar e heat) e a Base5 ao grupo de coleções com 5 palavras-chave (grain, oat, corn, oilseed e soybean). Cada base de dados possui 21578 documentos espaço-textuais. Além disso, cada base de dados possui um número de documentos relevantes, como mostra a Tabela 3.

Tabela 3: Número de documentos relevantes das bases de dados.

Banco de Dados	N. de documentos relevantes
Base1	55
Base2	181
Base3	32
Base4	11
Base5	5

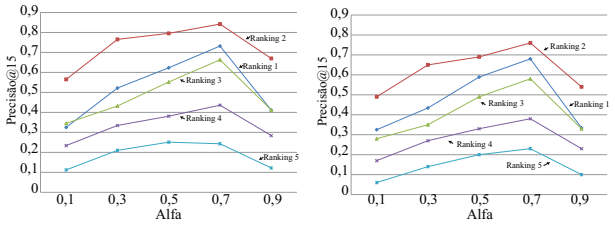
6.1 Avaliação da consulta: utilização das métricas

A Tabela 4 apresenta o resultado avaliativo de 5 consultas utilizando a métrica Precisão com a função de ranqueamento representada pela Equação 6. O Ranking 2 apresenta-se com melhores valores de precisão que os demais. Avaliando os gráficos da Figura 4(a) percebe-se que os maiores resultados da Precisão ocorreram quando utilizou-se valores de $\alpha = 0.3$ e $\alpha = 0.7$. Como é conhecido, a Precisão indica o percentual de documentos relevantes recuperados, assim os resultados mais relevantes foram recuperados ao se utilizar valores de α intermediários, o que indica que o balanceamento entre a relevância textual e a distância espacial determina uma maior satisfação do usuário. Os α's 0.1 e 0.9 representam os extremos; ao executar a consulta utilizando $\alpha = 0.1$, as informações textuais do documento estão sendo consideradas em detrimento da distância espacial. Em contrapartida, ao utilizar o $\alpha = 0.9$ a distância espacial assume maior importância. O destaque está na utilização do $\alpha = 0.7$, o que indica um percentual de importância maior para a distância espacial. Aplicando os valores de α na Equação 8 observa-se

Tabela 4: Eficácia de Ranking utilizando Precision@k (P@k)

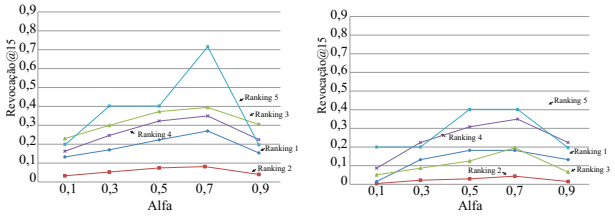
Rankings/ Base	P@15/ $\alpha = 0.1$	P@15/ $\alpha = 0.3$	P@15/ $\alpha = 0.5$	P@15/ $\alpha = 0.7$	P@15/ $\alpha = 0.9$
1	0.3250	0.5220	0.6230	0.7320	0.4110
2	0.5654	0.7654	0.7954	0.8420	0.6700
3	0.3450	0.4320	0.5520	0.6632	0.4120
4	0.2340	0.3345	0.3809	0.4360	0.2840
5	0.1120	0.2100	0.2510	0.2430	0.1220

no gráfico da Figura 4(b) uma discreta queda da eficácia na recuperação dos resultados relevantes. Nesta equação não ocorre normalização da função, visto que, esta operação influencia negativamente nos resultados da aplicação para qual esta função foi orginalmente desenvolvida. Ao se aplicar a métrica Revocação, no intuito de verificar o número de documentos relevantes recuperados em relação a todos os documentos relevantes da base, verifica-se que o $\alpha = 0.7$ também gerou os melhores resultados qualitativos. Este resultado se confirmou nas duas funções de ranqueamento analisadas. No entanto, vale ressaltar que o valor da revocação depende do núméto de documentos relevantes da base, uma vez que quanto maior for esse número menor será o resultado obtido. Para demonstrar o desempenho qualitativo das duas funções de ranqueamento nas 237 coleções geradas a partir da adaptação da coleção Reuters-21578, foi aplicado a métrica MAP (Mean Average Precision). A Figura 6 apresenta o resultado dos 20 primeiros documentos das 237 consultas realizadas. Embora mais uma vez a Equação 6 (EQ1) tenha apresen-



(a) Precisão utilizando Equação 6. (b) Precisão utilizando Equação 8.

Figura 4: Aplicação da métrica Precisão utilizando as duas funções de ranqueamento.



(a) Revocação utilizando Equação 6. (b) Revocação utilizando Equação 8.

Figura 5: Aplicação da métrica Revocação utilizando as duas funções de ranqueamento.

tado uma maior eficácia, ressalta-se que a Equação 8 (EQ2) não sofre normalização, o que influência na diminuição da eficácia quando comparada com a Equação 6, uma vez que esta foi desenvolvida para aplicação em rodovias, quando a normalização não é possível sob pena de definir com falhas as tragetórias .

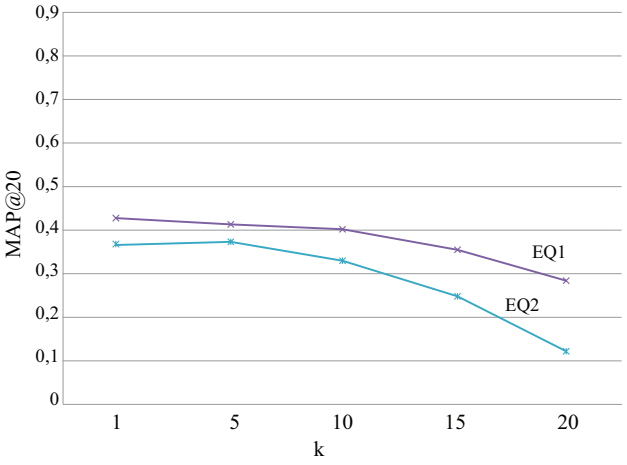


Figura 6: MAP@20 das funções EQ1 (Equação 6) e EQ2 (Equação 8).

7. CONCLUSÕES E TRABALHOS FUTUROS

Para avaliar uma consulta espaço-textual top-k é necessário coleções de referências que contenham conjuntos de consulta, documentos espaço-textuis e julgamentos de relevância para cada par consulta-documento. Este artigo demonstrou uma metodologia para avaliar a consulta e criar coleções de referência a partir da adaptação de coleções tradicionais. Através das 237 coleções geradas foi possível aplicar as métricas selecionadas e avaliar a consulta qualitativamente, utilizando as duas funções de ranqueamento existentes.

Os valores de α para os melhores resutlados das métricas aplicadas foi 0.7, demonstrando a importância do balanceamento entre a relevância textual e distância espacial. Esse valor é alterado de acordo com as variáveis utilizadas em cada consulta: número de palavras-chave, valores de k, distância Euclidiana, etc. Além disso, os experimentos realizados indicaram que a Equação 6 [9] demonstrou maior eficácia para as métricas aplicadas. No entanto, para fins de comparação a Equação 8 [19] por não sofrer normalização tem seu desempenho prejudicado porque os valores utilizados nos testes eram todos menores que um.

Para trabalhos futuros pode ser sugerido o aprimoramento da avaliação utilizando métricas que façam a harmonia entre uma métrica de distância e uma métrica de relevância textual. Outro ponto a ser sugerido, é a aplicação de técnicas de otimização para encontrar α's significativos, ou seja, que proporcionem melhores resultados. Essas técnicas de otimização podem ser, por exemplo, o uso de algoritmos genéticos ou programação linear.

8. REFERÊNCIAS

[1] R. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval - the concepts and technology behind search, Second edition*. Pearson Education Ltd., Harlow, England, 2011.

[2] R. Baeza-Yates and B. Ribeiro-Neto. Modern information retrieval. *ACM Press. New York*, 2013.

[3] X. C. Cao, G. Cong, C. S. Jensen, Q. Qu, A. Skovsgaard, D. Wu, and M. L. Yiu. Spatial keyword querying. *Er*, 7532(1):16–29, 2012.

[4] B. Carterette and J. Allan. Incremental test collections. In *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, CIKM '05, pages 680–687, New York, NY, USA, 2005. ACM.

[5] O. Chapelle, T. Joachims, F. Radlinski, and Y. Yue. Large-scale validation and analysis of interleaved search evaluation. *ACM Trans. Inf. Syst.*, 30(1):6:1–6:41, Mar. 2012.

[6] O. Chapelle, D. Metlzer, Y. Zhang, and P. Grinspan. Expected reciprocal rank for graded relevance. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, CIKM '09, pages 621–630, New York, NY, USA, 2009. ACM.

[7] L. Chen, G. Cong, C. S. Jensen, and D. Wu. Spatial keyword query processing: An experimental evaluation. *Proceedings of the VLDB Endowment*, 6(3):217–228, 2013.

[8] C. W. Cleverdon. The significance of the cranfield tests on index languages. *Proceedings of the 14th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 3–12, 1991.

[9] G. Cong, C. S. Jesen, and D. Wu. Efficient retrieval of the top-k most relevant spatial web objects. *Proceedings of the VLDB Endowment*, 2(1):337–348, 2009.

[10] A. GÃűker and H. Myrhaug. Evaluation of a mobile information system in context. *Pergamon Press*, 44(1):39–65, 2008.

[11] K. Järvelin and J. Kekäläinen. Cumulated gain-based evaluation of ir techniques. *ACM Trans. Inf. Syst.*, 20(4):422–446, Oct. 2002.

[12] K. S. Jones. *Information Retrieval Experiment*. Butterworth-Heinemann Newton, MA, USA, 1981.

[13] G. Kazai, N. GÃűvert, M. Lalmas, and N. Fuhr. The inex evaluation initiative. *INtelligent Search on XML Data*, 2818(3):279–293, 2003.

[14] D. D. Lewis. *Representation and learning in information retrieval*. PhD thesis, Department of Computer and Information Science, University of Massachusetts, Amherst, 1992. UMI Order No. GAX92-19460.

[15] D. D. Lewis. Reuters-21578 text categorization text collection, 2004. http://www.daviddlewis.com/resources/testcollections/reuters21578/. [Online; acessado 19-maio-2016].

[16] C. D. Manning, P. Raghavan, and H. Shutze. *Introduction to Information Retrieval*. Cambridge University Press., 2008.

[17] Reuters-21578. Reuters-21578 test collection 2006, 2006. [Online; acessado 23-setembro-2016].

[18] C. J. V. Rijsbergen. *Information Retrieval*. Butterworth-Heinemann Newton, MA, USA, 1979.

[19] J. B. Rocha-Junior and K. NÃÿrvag. *Top-k spatial keyword queries on road networks*. Proceedings of the 15th International Conference on Extending Database Technology - EDBT 12, Norwegian University of Science and Technology, 2012.

[20] G. Salton and M. J. McGill. *Introduction to Modern Information Retrieval*. McGraw Hill Book Co., 1986.

[21] M. Sanderson. Reuters test collection. Technical report, Proceedings of the Sixteenth Research Colloquium of the British Computer Society Information Retrieval Specialist Group, Drymen, 1994.

[22] M. Sanderson and H. Joho. Forming test collections with no system pooling. In *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '04, pages 33–40, New York, NY, USA, 2004. ACM.

[23] E. M. Voorhees and D. K. Harman. *TREC: Experiment and Evaluation in Information Retrieval (Digital Libraries and Electronic Publishing)*. The MIT Press, 2005.

[24] R. Wilkinson and M. Wu. Evaluation experiments and experience from the perspective of interactive information retrieval. *the Proceedings of the Third Workshop on Empirical of Adaptive Systems*, pages 23–26, 2004.

[25] E. Yilmaz, M. Shokouhi, N. Craswell, and S. Robertson. Expected browsing utility for web search evaluation. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, CIKM '10, pages 1561–1564, New York, NY, USA, 2010. ACM.

[26] J. Zobel and A. Moffat. Inverted files for text search engines. *ACM Comput. Surv.*, 38(2), July 2006.