

ARQUITETURA DE UM SISTEMA SPATIO-TEXTUAL

Fellipe de Lima Fonseca¹; João Batista Rocha-Junior²

1. Bolsista CNPq, Graduando em Engenharia de Computação, Universidade Estadual de Feira de Santana, e-mail: fellipefonseca9@gmail.com

1. Orientador, Departamento de Exatas, Universidade Estadual de Feira de Santana, e-mail: joao@uefs.br

PALAVRAS-CHAVE: banco de dados espaciais, busca spatio-textual. aplicativo.

INTRODUÇÃO

Devido ao grande desenvolvimento na área de vendas de dispositivos móveis, cada vez mais usuários acessam a Web utilizando recursos de geolocalização. Como resultado disso, temos a grande proliferação de recurso de geo-posicionamento (GPS, Wi-Fi, 3G, 2G), o que tem como consequência uma Web onde conteúdos podem ser associados a locais usando esses serviços de localização. Isso faz com que um número considerado de consultas tenham como alvo objetos espaciais.

Uma pesquisa realizada em 2012 aponta que 20% das buscas realizadas no Google são direcionadas a objetos espaciais (Google Places. Stats & Facts, 2012) e, uma outra, mostra que em 53% das consultas no Bing, realizadas à partir de dispositivos móveis, os usuários estão interessados em encontrar um local (Search Engine Land, 2012). Com isso, faz-se necessário um sistema onde esses pontos de interesse contenham um conteúdo ou descrição relacionados a eles (Chen *et al.*, 2013). Um protótipo de uma consulta spatio-textual tem um local de consulta e palavras-chave como argumentos e retorna os objetos que são relevantes a estes argumentos (Cao. X *et al.* 2012). Então, visando atender à essas necessidades de consultas por locais contendo descrição textual, foi elaborado uma arquitetura para a elaboração de um sistema de busca de objetos spatio-textuais.

Existem vários sites especializados em coletar e compartilhar a localização de objetos espaciais (como o OpenStreetMap¹ e GoogleMaps²). Por outro lado, há sites que contam com descrição para esses objetos (como o Wikipédia³). Essas informações podem fazer parte de uma base de dados de objetos spatio-textuais. Então, faz-se necessário a criação de um índice com essas informações para melhor recuperação de dados, criando assim, uma nova base de dados onde é possível realizar buscas através de dispositivos móveis e pela web. Nesse sistema, os usuários podem contribuir com informações para a base de dados dando sua votação e fazendo comentários.

METODOLOGIA

Os principais passos para a realização dessa pesquisa foram: 1) levantamento bibliográfico nas áreas de Banco de Dados Espaciais e Buscas Textuais, 2) aquisição de base de dados spatio-textual, 3) indexação dos objetos e 4) desenvolvimento de uma aplicação para web e smartphone utilizando-se dos resultados obtidos na pesquisa.

Um sistema spatio-textual requer a utilização de técnicas aplicadas em Banco de Dados Espaciais (Guting, 1994) e técnicas aplicadas em Sistemas de Busca Textual (Zobel &

1 <http://www.openstreetmap.org/>

2 <https://maps.google.com.br/>

3 <http://www.wikipedia.org/>

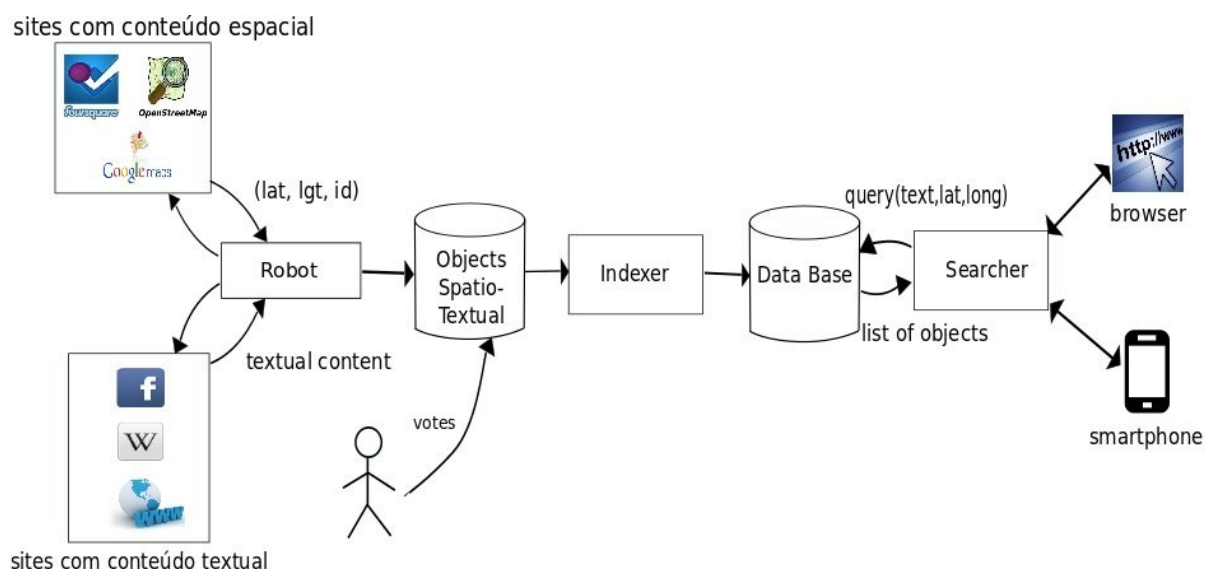


Figura 1. arquitetura de um sistema spatio-textual

Moffat, 2006). Portanto, fez-se necessário o estudo de artigos referentes a essas duas áreas de pesquisa.

Para colocar esse sistema em prática, fez-se necessário a obtenção de uma base de dados contendo objetos spatio-textuais. Para isso, foram utilizados objetos espaciais extraídos do OpenStreetMap e objetos textuais extraídos do Wikipédia, e a partir dessas informações, foi feito o processo de anotação de objetos espaciais (Fonseca & Rocha-Junior, 2013). Este processo consiste em anotar objetos espaciais visando obter uma descrição para esses objetos.

Esses objetos são indexados visando a recuperação de dados de maneira eficiente. Com isso, dois tipos de buscas são disponibilizadas: a primeira, considerando os objetos mais próximos à uma dada localização, e a segunda, considerando os objetos mais relevantes referentes a uma consulta por palavras-chave. Foi criado, ainda, uma aplicação cliente-servidor, onde o usuário tem a possibilidade de realizar essas buscas.

RESULTADOS

Como resultado da pesquisa, foi construído uma arquitetura de um sistema de busca spatio-textual, e em cima disso, foi desenvolvido uma aplicação para Web e Smartphone. A Figura 1 ilustra a arquitetura de um sistema spatio-textual, e os seus componentes são descritos a seguir:

O componente Robot tem a função de recuperar informações espaciais de sites especializados nesses tipos de conteúdo, assim como extrair conteúdo textual pela Web. A princípio foi utilizado apenas o OpenStreetMap para a obtenção de conteúdo espacial, e o Wikipédia para conteúdo textual. Existem muitos desses objetos espaciais que contêm páginas no Wikipédia que os descrevem. Tendo isso em vista, foi feito o processo de anotação de objetos espaciais (Fonseca & Rocha-Junior, 2013), onde é feita uma anotação automática nos objetos espaciais, procurando, para cada um deles, um artigo correspondente no Wikipédia. Desse processo foram obtidos mais de 10.000 objetos spatio-textuais.

Nesse sistema, o usuário pode contribuir para o conteúdo da base de dados a partir de



Figura 2. A - Lista de pontos próximos. B - Pagina de detalhamento com informações sobre o ponto de interesse. C - Rota até o ponto.

comentários sobre objetos. O usuário também tem o papel de ajudar a descrever os objetos que não foram anotados no processo de anotação automática de objetos espaciais, aumentando assim, a base de dados spatio-textual.

O componente Indexer é responsável por indexar a base de dados, onde dois tipos de indexação são realizadas. Na primeira delas, os objetos spatio-textuais são indexados em uma R*-tree (Beckmann *et al.*, 1990). Com isso, é possível atribuir um índice apropriado para o armazenamento de cada objeto, permitindo recuperar, de forma eficiente, objetos próximos a uma dada localização. Porém, essa busca é feita apenas por pontos próximos de uma dada localização, não havendo uma consulta por palavras-chave.

Então, para possibilitar uma busca por pontos de interesse, foi utilizado o índice s2i (Rocha-Junior *et al.*, 2011). Este, faz a indexação dos objetos, de modo a recuperar, de forma eficiente, os melhores objetos spatio-textuais ranqueados de acordo com sua proximidade com o local de consulta e relevância para a consulta por palavras-chave.

O componente Searcher é responsável por acessar a base de dados indexada, conforme uma requisição. Esta requisição pode ser feita tanto através do browser quanto por smartphones e deve conter o local a partir do qual deve ser realizada a busca (latitude e longitude), assim como as palavras-chave. Dois tipos de buscas são disponibilizadas: a primeira, quando não há palavras-chave na requisição, onde são retornados os objetos mais próximos ao local de busca. E uma segunda, onde são indicadas palavras-chave. Nesta busca, é feito um ranqueamento entre os objetos, considerando a proximidade do local de busca, e a relevância do conteúdo textual desses objetos, levando em conta as palavras-chave indicadas.

Baseado na arquitetura mostrada na Figura 2, foi criado o PerTA⁴ (Personal Travel Assistant), que é um assistente de viagens pessoal, disponível para smartphone e web, que auxilia o turista, mostrando pontos de interesses próximo a ele, assim como informações sobre esses locais.

4 <https://sites.google.com/a/ecomp.uefs.br/perta/>

O PerTA é uma aplicação cliente-servidor, onde o servidor tem o papel de indexar os objetos spatio-textuais de forma a possibilitar uma consulta de forma eficiente, e disponibilizar os resultados para diferentes clientes (interface Web ou dispositivos móveis). De posse dessas informações indexadas, o aplicativo, a partir da localização do usuário (adquirida através do GPS ou redes móveis), retorna os pontos mais próximos a ele de acordo com o seu interesse (através de uma busca por palavras-chave). A partir desses pontos retornados, o usuário tem a possibilidade de solicitar a rota até esses locais, assim como, obter informações sobre eles. A Figura 2 ilustra essas funcionalidades.

A Figura 2 mostra como o usuário pode obter informações sobre um ponto de interesse a partir de uma lista de pontos próximos. Neste exemplo, foi escolhido a Universidade Estadual de Feira de Santana como ponto de interesse. Assim, o aplicativo permite ao usuário traçar uma rota entre ele e o local desejado. O usuário ainda tem a opção de obter informações desse local, como fotos e uma descrição sobre ele.

CONCLUSÃO

Este artigo abordou o problema das consultas espaciais onde há a necessidade de obter conteúdo textual para pontos de interesse. Foi apresentada uma arquitetura spatio-textual, e como resultado, foi criado um sistema para buscas por pontos de interesse baseado nessa arquitetura. Esse sistema conta com dados do OpenStreetMap e Wikipédia formando sua base de dados, podendo futuramente usar outras fontes de dados espaciais e textuais disponíveis na web.

REFERÊNCIAS

- FONSECA, F. L.; ROCHA-JUNIOR, J. B. Anotação Automática de Objetos Espaciais. In Proceedings of ERBASE, 2013.
- BECKMAN, N.; KRIEGEL, H. P.; SCHNEIDER, R.; SEEGER, B. The R*-tree: an efficient and robust access method for points and rectangles. ACM International Conference on Management of Data (SIGMOD), 1990.
- ROCHA-JUNIOR, J. B.; GKORGKAS, O.; JONASSEN, S.; NORVAG, K. Efficient Processing of Top-k Spatial Keyword Queries. Proc. of the International Symposium on Spatial and Temporal Databases, Springer, LNCS 2011.
- CAO, X.; CHEN, L.; CONG, G.; JENSEN, C. S.; QU, Q.; SKOVSGAARD, A.; WU, D.; YIU, M. L. Spatial keyword querying. In ER, 2012.
- CHEN, L.; CONG, G.; JENSEN, C.; WU, D. (2013) Spatial keyword query processing: an experimental evaluation, In Proceedings of the International Conference on Very Large Data Bases (VLDB).
- GUTING, R. H. . An introduction to spatial database systems. *VLDB Journal*, 1994.
- ZOBEL, J.; MORFFAT, A. Inverted files for text search engines. ACM Computing. Surveys, 2006.
- Search Engine Land: Microsoft: 53 percent of mobile searches have local intent (2012), searchengineland.com/microsoft-53-percent-of-mobile-searches-have-local-intent-55556. Acessado em 18/10/2013.
- Google: Google Places. Stats & Facts (2012), sites.google.com/a/pressatgoogle.com/googleplaces/metrics. Acessado em 18/10/2013.

