

PERÍODO DE ABRANGÊNCIA DO PROGRAMA: 1º/08/2018 a 31/07/2019

Obs.: o bolsista deve enviar cópia do Certificado do EAIC/2019 para o e-mail do seu orientador(a) e este(a) deverá anexá-lo no SGPEX (www.sgp.uem.br).

1. **BOLSISTA:** William Rodrigues da Silva

2. **ORIENTADOR:** Marcos Aurélio Domingues

3. **DEPARTAMENTO:** DIN

4. **CO-ORIENTADOR:**

5. **DEPARTAMENTO:**

6. **TÍTULO DO PROJETO:** Concepção e desenvolvimento de mecanismos para obtenção de pontos de interesse em cidades brasileiras

7. **RELATÓRIO CONTENDO OS RESULTADOS DA PESQUISA** Em anexo

8. **COMPROVANTE DE APRESENTAÇÃO DOS RESULTADOS DA PESQUISA EM EVENTOS CIENTÍFICOS**
(Se houve premição, informar o evento, a classificação e anexar comprovante).

9. **COMPROVANTE DE PUBLICAÇÃO, COM A PARTICIPAÇÃO DOS BOLSISTAS, EM PERIÓDICOS INDEXADOS E/OU COM CORPO EDITORIAL.**

10. **AVALIAÇÃO DO ORIENTADOR SOBRE O DESEMPENHO DO BOLSISTA NO PROJETO.**

O bolsista William Rodrigues da Silva desenvolveu todas as atividades do projeto com grande motivação e dedicação, tendo atingido todos os objetivos propostos para este projeto de iniciação científica.

11. **AVALIAÇÃO DO ORIENTADOR SOBRE O PROGRAMA INSTITUCIONAL DE BOLSAS DE INICIAÇÃO CIENTÍFICA.**

O programa institucional de bolsas de iniciação científica é de grande importância para o despertar científico dos alunos de graduação, uma vez que oferece a estes a oportunidade de participar em projetos de pesquisa e adquirir novos conhecimentos. A participação em projetos de pesquisa também pode despertar no aluno o interesse futuro em um curso de pós-graduação.

12. **AVALIAÇÃO DO ACADÊMICO SOBRE O PROGRAMA INSTITUCIONAL DE BOLSAS DE INICIAÇÃO CIENTÍFICA.**

Excelente, tendo em vista que o Programa de Iniciação Científica proporcionou a oportunidade de aquisição de novos conhecimentos, sobre o funcionamento de um crawler, o uso de API's, o crawler de páginas na internet, assim como outros recursos utilizados no desenvolvimento dos sistemas web, como por exemplo a manipulação de mapas em páginas web, que não são abordados profundamente durante o curso. Também permitiu o aprendizado da sistematização de conhecimentos, assim como das normas de um trabalho científico.

**UNIVERSIDADE ESTADUAL DE MARINGÁ
PROGRAMA INSTITUCIONAL DE BOLSAS DE INICIAÇÃO
CIENTÍFICA – PIBIC/CNPq-Fundação Araucária-UEM
DEPARTAMENTO DE INFORMÁTICA
ORIENTADOR(A): Prof. Marcos Aurélio Domingues
Bolsista: William Rodrigues da Silva**

**CONCEPÇÃO E DESENVOLVIMENTO DE MECANISMOS PARA
OBTENÇÃO DE PONTOS DE INTERESSE EM CIDADES BRASILEIRAS**

Maringá, 31 de julho de 2019.

**UNIVERSIDADE ESTADUAL DE MARINGÁ
PROGRAMA INSTITUCIONAL DE BOLSAS DE INICIAÇÃO
CIENTÍFICA – PIBIC/CNPq-Fundação Araucária-UEM
DEPARTAMENTO DE INFORMÁTICA
ORIENTADOR(A): Prof. Marcos Aurélio Domingues
Bolsista: William Rodrigues da Silva**

**CONCEPÇÃO E DESENVOLVIMENTO DE MECANISMOS PARA
OBTENÇÃO DE PONTOS DE INTERESSE EM CIDADES BRASILEIRAS**

Relatório contendo os
resultados finais do
projeto de iniciação
científica vinculado ao
PIBIC/CNPq-Fundação
Araucária-UEM.

Maringá, 31 de julho de 2019.

Resumo: O setor de turismo no Brasil vem crescendo nos últimos anos. Apesar do crescimento do setor, este ainda apresenta vários problemas, entre os quais, vale ressaltar a falta de informações para turistas. Uma possível solução para a falta de informações turísticas no que se refere a pontos de interesse (serviços e locais turísticos) consiste na proposta de uma plataforma que faça uso de sistemas de recomendação de pontos de interesse. Entretanto, para o uso da plataforma, se faz necessário à obtenção de pontos de interesse para então recomendá-los. Este projeto teve como objetivo a investigação e proposta de mecanismos para obtenção de pontos de interesse em cidades brasileiras para serem utilizados pela plataforma de recomendação de pontos de interesse.

Palavras-chave: Coleta de informações da web, Pontos de interesse, *Crawler* Java.

Introdução

O setor de turismo no Brasil cresceu a uma taxa de 13,1% no ano de 2012 [1]. Eventos como a Copa das Confederações de 2013, a Copa do Mundo de 2014 e as Olimpíadas de 2016 impulsionaram este setor, que mesmo em época de crise se mantém em crescimento. Apesar do crescimento do setor de turismo, este ainda apresenta vários problemas, entre os quais, vale ressaltar a falta de informações para turistas [2].

Uma possível solução para a falta de informações turísticas no que se refere a pontos de interesse (serviços e locais turísticos) consiste na concepção e desenvolvimento de uma plataforma para um sistema de recomendação sensível ao contexto capaz de recomendar pontos de interesse por meio de aparelhos de telefones celulares inteligentes (*smartphones*).

Um sistema de recomendação sensível ao contexto é uma tecnologia de filtragem de informação que, além do perfil e interesse do usuário (turista), utiliza também informação contextual para poder recomendar a este uma lista ordenada de itens, por exemplo, locais turísticos, que lhe sejam potencialmente interessantes [3]. Informação contextual pode ser definida como um conjunto de fatores externos que influenciam o comportamento e interesse de um usuário, como, por exemplo, local, tempo, data, hora e condições meteorológicas [4]. Os sistemas de recomendação sensíveis ao contexto compõe um tema com elevado potencial de desenvolvimento de ciência, tecnologia e inovação, sendo que projetos demonstrativos podem ser encontrados ao redor do mundo em diversos domínios de aplicação, como, por exemplo, recomendação de fotos, recomendação de música, recomendação de nomes próprios, recomendação de programas de TV, recomendação de filmes e recomendação de restaurantes.

Já com relação ao uso de telefones celulares inteligentes (*smartphones*), a popularização destes aparelhos tem sido considerada por muitos a revolução tecnológica de maior impacto nos últimos tempos após a revolução causada pela Internet e pelas redes sociais. Esses dispositivos podem ser considerados como computadores de bolso que acompanham usuários 24 horas por dia. Em 2013 foram vendidos mais *smartphones* que celulares convencionais [5]. No Brasil, apenas em 2012 foram vendidos 16 milhões desses dispositivos e há expectativas de que o Brasil se torne o quarto maior mercado desse setor no mundo [6]. A massificação dos *smartphones* tem gerado oportunidades que agregam valores a serviços e produtos de diversos setores. Esses dispositivos processam e carregam uma grande quantidade de informação relacionada a seus usuários e que pode ser utilizada para oferecer experiências personalizadas em diversos domínios.

Uma das plataformas de recomendação mais conhecida atualmente é a *Foursquare* [7]. Esta é uma plataforma colaborativa baseada em uma rede social e mecanismos de *check-in* [8]. Nesta plataforma, os usuários da rede social *Foursquare* marcam os lugares já visitados por eles, e as recomendações são feitas a partir dos relacionamentos dentro da rede social. Contudo, neste paradigma, as recomendações se tornam imparciais quando os amigos/relacionamentos de um dado usuário nunca estiveram no local que ele está visitando no momento ou quando o usuário possui poucos amigos/relacionamentos.

O cenário mais comum para o uso de um sistema de recomendação de pontos de interesse sensível ao contexto é quando um usuário solicita recomendações, através de seu *smartphone*, no momento em que ele está visitando um local desconhecido. Neste cenário, muito provavelmente, ele não terá amigos/relacionamentos que visitaram tal local e recomendações baseadas em sistemas como o *Foursquare* serão pouco precisas.

Diante destes fatos, o orientador deste projeto de iniciação científica propôs um projeto de pesquisa intitulado “Uma Plataforma Avançada para Recomendação Automática de Pontos de Interesse em Cidades Brasileiras”. Este projeto foi aprovado no edital 01/2016-PPG – Pesquisa Básica e Aplicada, e está sendo desenvolvido no Departamento de Informática da Universidade Estadual de Maringá.

Para que a plataforma de recomendação de pontos de interesse, proposta no projeto de pesquisa, não necessite de mecanismo de *check-in* e produza recomendações interessantes para o usuário, se faz necessário que ela utilize mecanismos que obtenham e armazenem automaticamente as informações sobre pontos de interesse de uma cidade. Assim, o objetivo principal deste projeto de iniciação científica foi à investigação e proposta de mecanismos para obtenção de pontos de interesse em cidade brasileira para serem utilizados como entrada em sistemas de recomendação de pontos de interesse, proporcionando o desenvolvimento de uma plataforma que possa produzir recomendações de maior interesse para o usuário e evitar a necessidade de mecanismos de *check-in*.

Objetivos

O objetivo principal deste projeto de iniciação científica foi a investigação, proposta e desenvolvimento de mecanismos para a obtenção de pontos de interesse que serão utilizados como entrada em sistemas de recomendação de pontos de interesse. Neste projeto foram exploradas as seguintes linhas de trabalho para a obtenção de pontos de interesse:

1. Uso de APIs existentes de sites que fornecem dados sobre pontos de interesse [9];
2. Pesquisa em recuperação de informação [11-12], como, por exemplo, recuperação automática de coleções de documentos web baseado em tópicos específicos (pontos de interesse) [10], e subsequente extração de informação;
3. Concepção e desenvolvimento de um sistema web para preenchimento de informações sobre pontos de interesse pelos próprios proprietários dos pontos.

Materiais e Métodos

Desenvolvimento do crawler Java

O desenvolvimento de mecanismos para a obtenção de pontos de interesse iniciou-se com a criação de um *crawler* (ferramenta para obtenção de informações) utilizando a linguagem de programação Java, que faz o levantamento de informações sobre pontos de interesse em cidades brasileiras, cria objetos com estas informações e faz a persistência dos objetos criados no banco de dados da plataforma. As informações que um objeto registrado pelo *crawler* pode possuir são: nome, endereço formatado, número de telefone formatado, latitude, longitude, URL, website, nota de avaliação, tipo, comentário sobre o local e autor do comentário. O *crawler* realiza também uma categorização dos pontos de interesse com base no tipo do local, desta forma criando uma generalização dos tipos.

O *crawler* desenvolvido pode ser usado em dois diferentes modos de operação, sendo eles: “modo API” e “modo WEB”, e cada modo extrai informações de fontes diferentes.

No “modo API”, o *crawler* realiza a coleta de informações de pontos de interesse utilizando a API Google Places¹, partindo de uma coordenada inicial e buscando os pontos de interesse em um dado raio (em metros) com base no ponto inicial especificado. Neste modo, o *crawler* faz inicialmente uma requisição a API e recebe respostas no formato JSON contendo os pontos de interesse encontrados na região determinada, e a partir desta resposta, realiza para cada ponto uma requisição específica solicitando mais informações sobre o local em questão. O *crawler* faz então uma validação de coerência com base no nome e categoria do local, e se aprovada, cria e persiste um objeto no banco de dados com as informações do ponto de interesse.

No “modo WEB”, o *crawler* tem também como objetivo a obtenção de informações de pontos de interesse, porém, as informações obtidas provêm de outra fonte de informação. Neste modo, o *crawler* navega automaticamente pelo kekanto², um site web que lista pontos de interesse em cidades brasileiras, e a partir dos parâmetros recebidos, extrai da estrutura do site as informações necessárias para formar um objeto referente a um ponto de interesse. O *crawler* então realiza este processo para todos os pontos de interesse registrados em uma determinada cidade, criando e persistindo objetos no banco de dados para cada um destes.

A Figura 1 mostra o diagrama de entidade-relacionamento do banco de dados que o *crawler* salva os dados obtidos referentes aos pontos de interesse.

¹ <https://developers.google.com/places/web-service/intro>

² <http://kekanto.com.br>

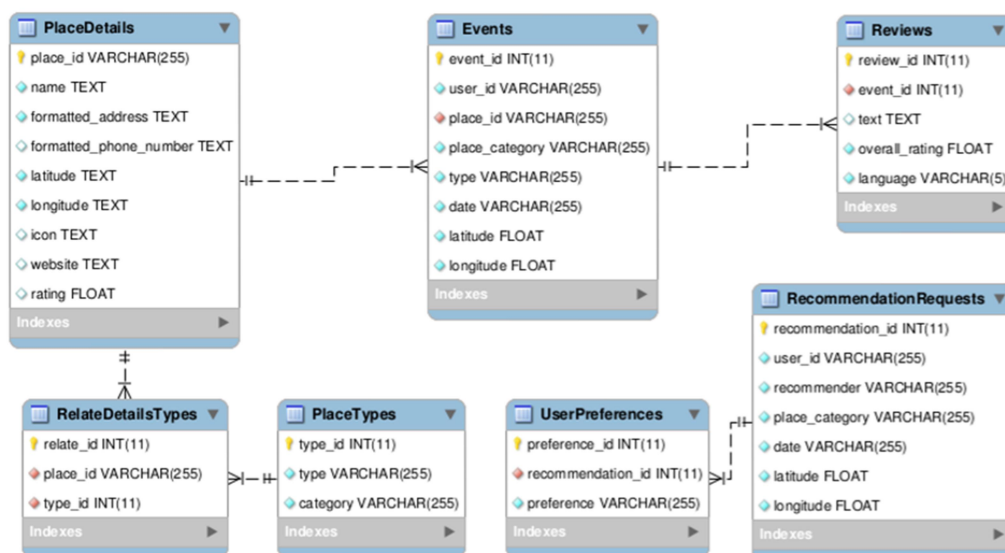
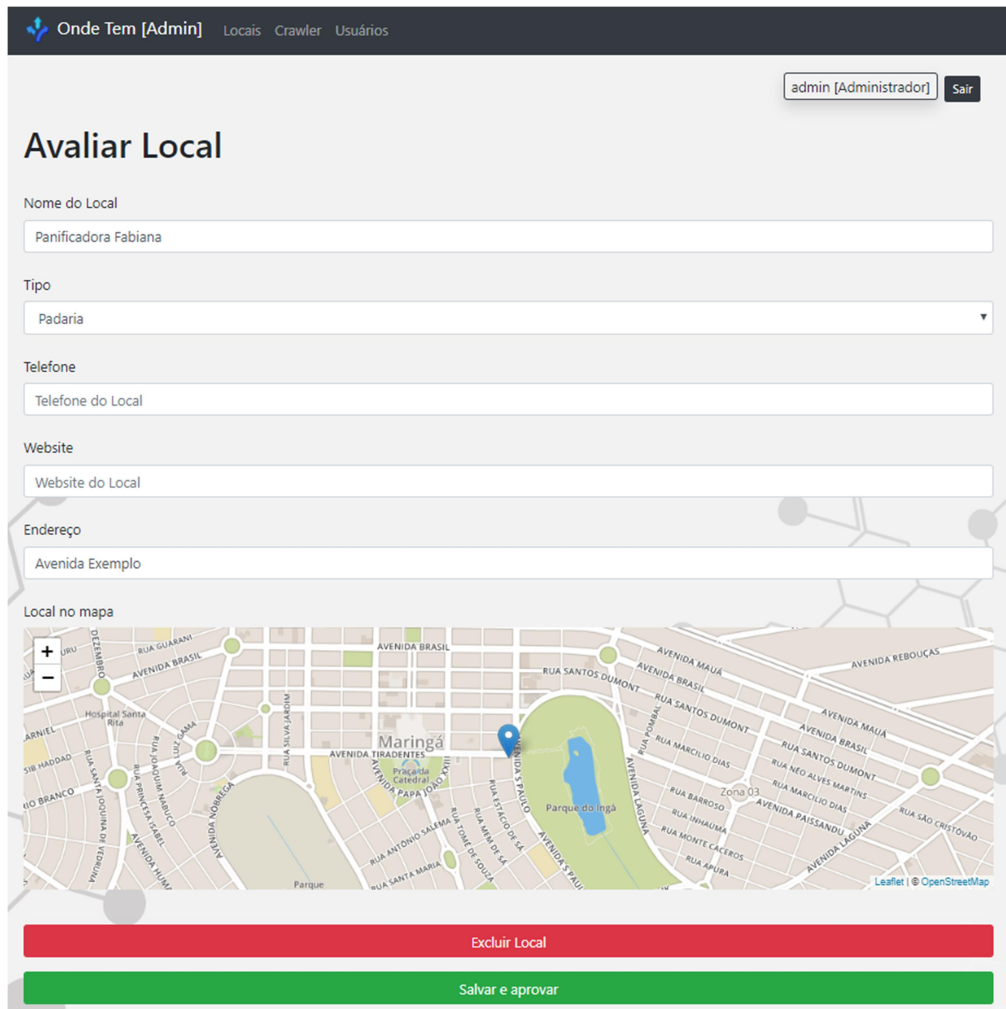


Figura 1 – Diagrama de entidade-relacionamento do banco de dados.

Desenvolvimento do sistema Web

A segunda etapa do desenvolvimento ocorreu com a criação de um sistema web composto de dois módulos. O primeiro módulo, nomeado como “OndeTem Web”, tem como objetivo permitir que proprietários de pontos de interesse realizem um cadastro com informações referentes aos seus pontos de interesse, e estas informações ficam então registradas em um banco de dados auxiliar para que possam ser avaliadas posteriormente por um administrador do sistema. Se aprovados por um administrador do sistema, estes pontos de interesse são removidos do banco de dados auxiliar e armazenados no banco de dados definitivo da aplicação. A Figura 2 mostra a tela de cadastro de um novo ponto de interesse, que pode ser realizado por um usuário (por exemplo, o proprietário de um estabelecimento) na internet.

Neste sistema ao selecionar a opção avaliar sobre um ponto de interesse cadastrado, o administrador do sistema pode ter acesso a todos os dados cadastrados pelo usuário referentes aquele ponto de interesse, e pode então realizar alterações necessárias, aprovar o local ou até mesmo reprová-lo. A Figura 4 mostra a tela de avaliação de um ponto de interesse cadastrado.



Onde Tem [Admin] Locais Crawler Usuários

admin [Administrador] Sair

Avaliar Local

Nome do Local
Panificadora Fabiana

Tipo
Padaria

Telefone
Telefone do Local

Website
Website do Local

Endereço
Avenida Exemplo

Local no mapa

Excluir Local

Salvar e aprovar

Figura 4 – Tela de avaliação de um ponto de interesse cadastrado.

Unificação dos sistemas desenvolvidos

Por fim, para possibilitar um acesso mais direto e unificado dos mecanismos de obtenção de pontos de interesse, as funcionalidades dos modos do *crawler* foram adicionadas ao módulo “OndeTem Admin”, permitindo que um administrador além de realizar as funções já citadas, possa também executar os modos “API” e “WEB” do *crawler* Java por meio de uma interface Web. A Figura 5 mostra a tela de seleção do *crawler* presente no mesmo sistema que se faz a avaliação de pontos de interesse cadastrados.



Figura 5 – Tela do sistema unificado.

Resultados e Discussão

Para demonstrar os dados obtidos pelo sistema *crawler* Java, este foi executado no “modo WEB” para a cidade de Maringá-PR. A Tabela 1 apresenta uma pequena amostra das informações obtidas e registradas no banco de dados da plataforma.

Tabela 1 – Amostra das informações obtidas com o *crawler* Java no “modo WEB”.

formatted_address	latitude	longitude	name
Av. Tiradentes, 500, Zona 01 - Maringá	-23.426240000	-51.938172000	Catedral de Maringá
Av. Tamandaré, 200, Zona 01 - Maringá	-23.418971000	-51.934994000	Café Cremoso
Av. São Paulo, 1099, Zona 09 - Maringá	-23.418286000	-51.932920600	Maringá Park Shopping
Av. Herval, 26, Zona 50 - Zona Central - Maringá	-23.425390000	-51.936368000	Hotel Deville Maringá

Conclusões

Com a realização deste projeto de iniciação científica foram desenvolvidos mecanismos para obtenção de pontos de interesse em cidades brasileiras que podem extrair informações de três diferentes fontes: *Crawler* utilizando API do Google Places, *crawler* que extrai informações de uma site Web, e um sistema web que permite o cadastro de pontos de interesse, colaborando assim para que a plataforma de recomendação de pontos de interesse tenha acesso a estas informações e possa realizar de modo adequado as suas recomendações.

Referências

1. SEBRAE-SP. **Turismo cresce 13,1% em 2012**. Disponível em: <<http://www.sebraesp.com.br/index.php/76-noticias/multissetorial/8867-pesquisa-setor-de-turismo-cresce-13-1-em-2012>>. Acesso em: 21 ago. 2016.
2. UOL Mais. **Turismo no Brasil tem falta de informação**. Disponível em: <<http://mais.uol.com.br/view/s70pk4i6az2h/turismo-no-brasil-tem-falta-de-informacao-04020D9A3166E0914326>>. Acesso em: 21 ago. 2016.
3. ADOMAVICIUS, G.; TUZHILIN, A. **Context-Aware Recommender Systems**. In: Recommender Systems Handbook, 2011, p. 217-253.
4. DEY, A. K. “**Undertanding and Using Context**”. Personal and Ubiquitous Computing, 5 (1), 2001, pp. 4-7.
5. Business Wire. “**Venda de smartphones supera celulares comuns em 2013 no País**” Disponível em: <<http://blogs.estadao.com.br/link/venda-de-smartphones-supera-celulares-comuns-em-2013-no-pais/>>. Acesso em: 21 ago. 2016.
6. IDC Analyze the Future. “**China to Become the Largest Market for Smartphones in 2012 with Brazil and India Forecast to Join the Top 5 Country-Level Markets by 2016, Acording to IDC**” Disponível em: <<http://www.idg.com/www/pb.nsf/ByID/MBEN-8SENF>>. Acesso em: 21 ago. 2016.
7. FOURSQUARE. Disponível em: <<https://foursquare.com>> Acesso em: 21 ago. 2016.
8. YING, J. J. C.; LU, E. H. C.; KUO, W. N.; VINCENT, S. T. “**Urban point-of-interest recommendation bt mining user check-in behaviors**”. In: International Workshop on Urban Computing (UrbComp ‘12). Pp. 63-70. 2012.
9. Programabel Web. “**146 Location APIs: Foursquare, Panoramio and Geocoder**”. Disponível em: <<http://blog.programmableweb.com/2012/06/20/146-location-apis-foursquare-panoramio-and-geocoder/>>. Acesso em 21 ago. 2016.
10. ESCUDEIRO, N. F.; JORGE, A. M. “**Semi-automatic Creation and Maintenance of Web Resources with webTopic**”. In: EWMF/KDO. Pp. 82-102. 2005.
11. BAEZA-YATES, R. A.; RIBEIRO-NETO, B. A. “**Modern Information Retrieval – the concepts and techonology behind search (Second edition)**”. Pearson Education Ltd., 2011.
12. RAE, A.; MURDOCK, V.; POPESCU, A.; BOUCHARD, H. “**Mining the web for point of interest**”. In: 35th International Conference on Research and Developoment in Information Retrieval (SIGIR’12), 2012, pp. 711-720.