

Redes de Compartilhamento de Conhecimento

Analise Exploratória dos Dados

01. Introdução

O desenvolvimento de software apresenta grandes desafios que vão além do conhecimento técnico. Um problema muito comum ocorre durante a escolha de quais tecnologias e ferramentas utilizar. Diversos estudos são realizados na busca da combinação perfeita. Porém, alguns fatores importantes como os aspectos sociais e culturais são ignorados.

Nenhum sistema é criado sem a utilização de mão de obra especializada. Assim, o custo do projeto e consequentemente o seu sucesso, são influenciados pela demanda desses profissionais. Portanto, compreender como está a distribuição do conhecimento técnico nas diferentes regiões tem grande importância para essa tomada de decisão.

A grande utilização das redes sociais de compartilhamento de conhecimento, permite que seja realizada uma nova abordagem na análise de seus dados. Para isso, utilizaremos dados geográficos na visualização do comportamento de duas das maiores redes que foram selecionadas para esse trabalho devido a sua complementaridade proposta por Vasilescu, Filkov e Serebrenik (2013).

GitHub

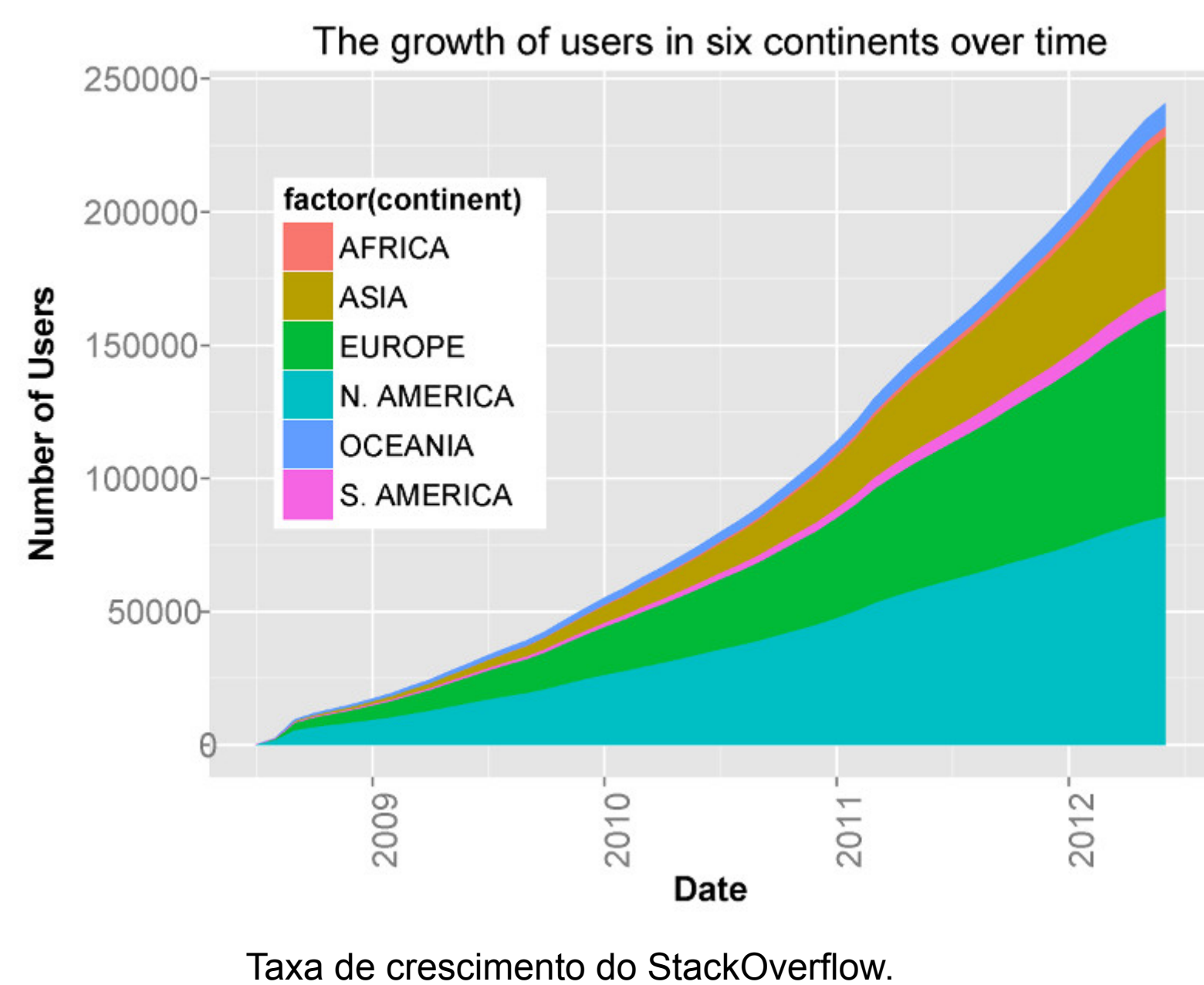
stackoverflow

02. Referencial Teórico

Com o surgimento de redes sociais como StackOverflow e GitHub, uma vasta quantidade de informações de desenvolvimento é criada diariamente. Tais dados de contexto pessoal e social tem um enorme potencial para apoiar na avaliação automática e eficaz da capacidade do desenvolvedor (HUANG et al.,). A complementaridade das ferramentas foi constatada num estudo que mostra que "committers" ativos do GitHub perguntam menos e fornecem mais respostas no StackOverflow. Assim como, "askers" ativos do StackOverflow distribui seu trabalho de uma maneira menos uniforme que os outros (VASILESCU; FILKOV; SEREBRENIK, 2013). Bem como comportamentos semelhantes entre usuários das redes diferentes foram relatados por Badashian et al. (2014) ao verificar que desenvolvedores ativos contribuem para as principais atividades da plataforma (ou seja, realizar commits no GitHub e responder no StackOverflow), mas também se envolvem em outras atividades gerenciais (como gerenciamento de problemas no GitHub e voto de qualidade no StackOverflow).

Geocodificação é um conjunto de métodos capazes de transformar descrições em coordenadas geográficas. Endereços urbanos são uma das principais formas de expressão da localização geográfica em cidades. Muitos sistemas de informação incluem atributos para receber endereços e, assim, contam com uma referência espacial indireta. A obtenção de coordenadas a partir de endereços é um dos métodos de geocodificação mais importantes, mas é dificultada por variações comuns no endereço, como abreviações e omissão de componentes. No caso de endereços, existe uma expectativa de detalhamento hierárquico, com componentes que indicam o país, o estado e a cidade. O formato de apresentação desses componentes varia de país para país, e em muitas situações, alguns componentes são intencionalmente omitidos ou simplificados (MARTINS; JR; FONSECA, 2012).

03. Visão Tradicional x Objetivos



04. Metodologia

Obter informações das bases de dados de ambas as redes sociais. Fazer o download de um backup do banco mais recente.

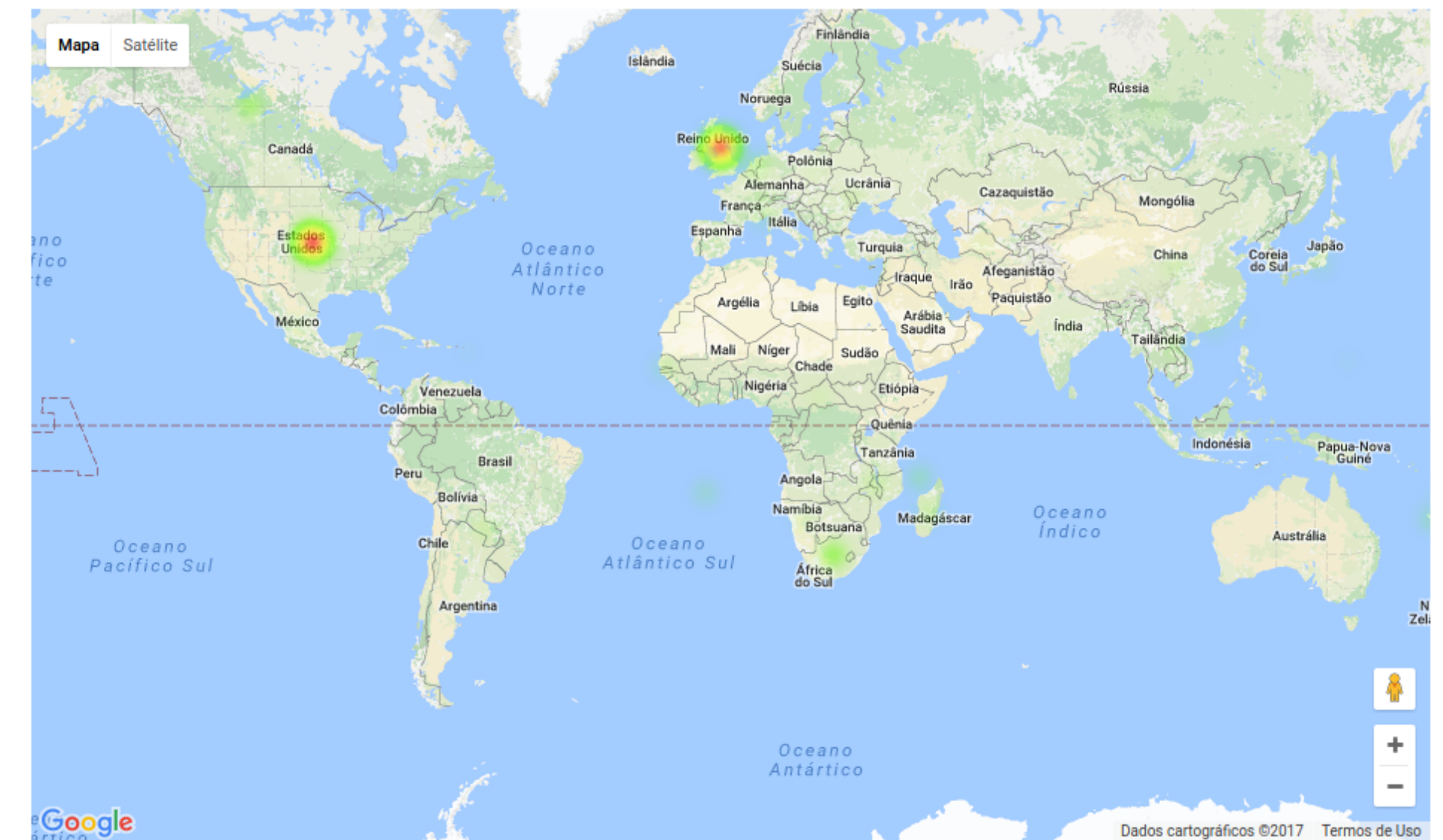
Mapear e extrair os dados pertinentes ao problema. Identificar a estrutura e elaborar consultas para relacionar e extrair apenas as informações que serão utilizadas.

Realizar um pré-processamento dos dados. Obter a localização geográfica dos usuários. Neste passo utilizaremos conceitos identificados no referencial teórico.

Desenvolver uma interface para visualização dos dados na forma geográfica.

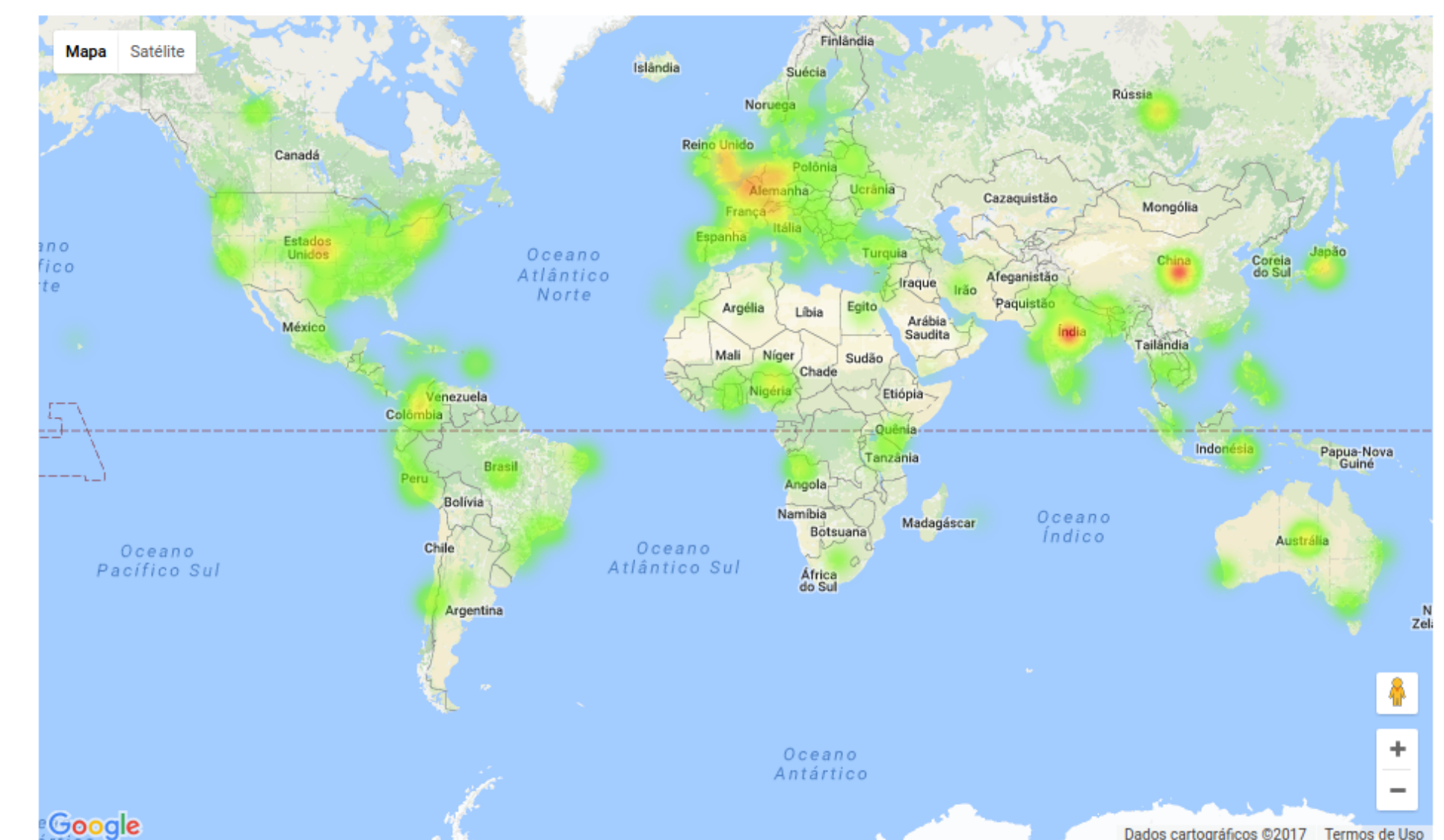
Estudar a correlação dos conteúdos para algumas determinadas regiões.

05. Resultados



256 tecnologias do StackOverflow.

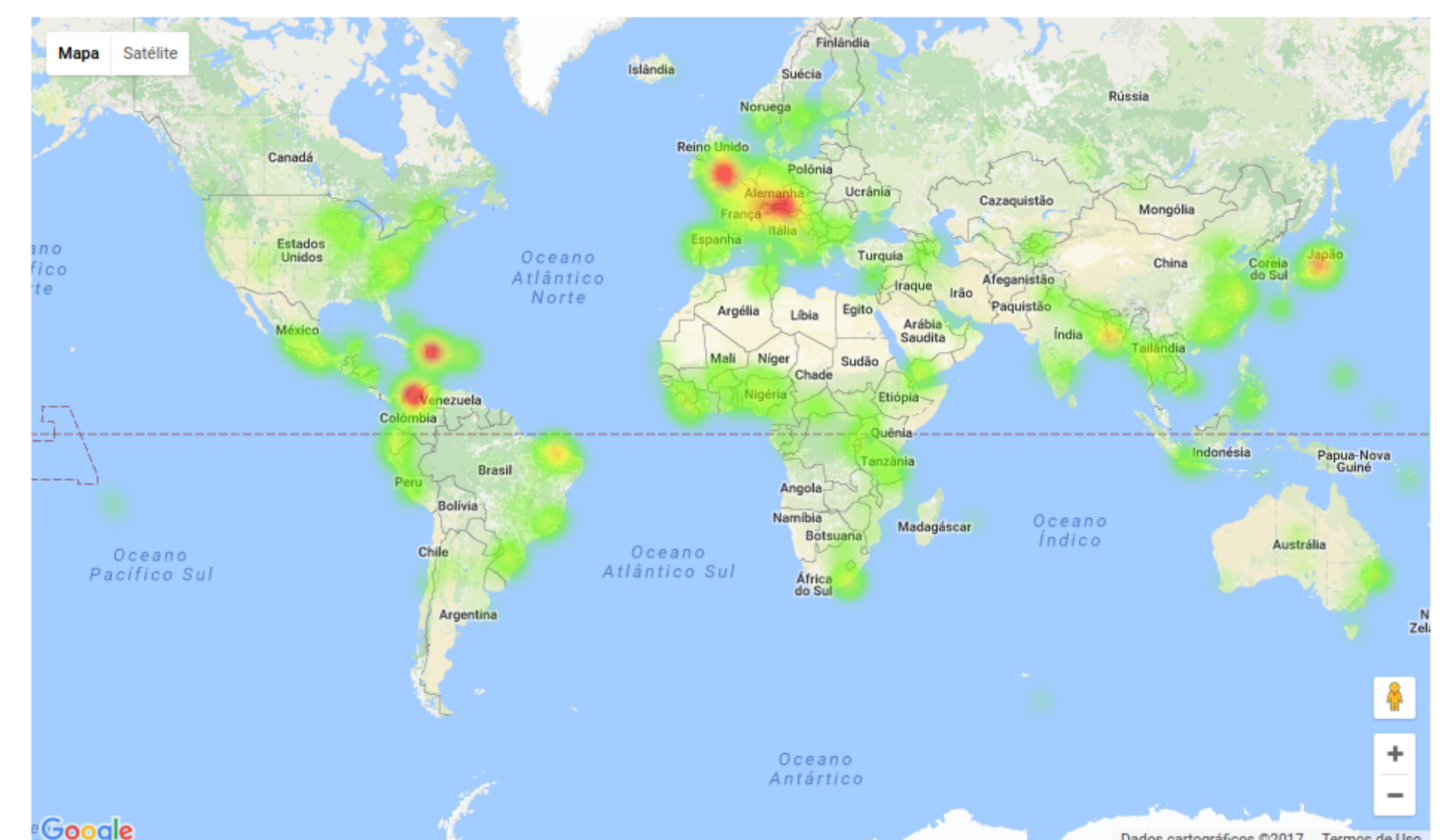
Seleciona usuários por casamento exato de sigla do país.



Seleciona usuários por selecao de parte do nome regex.



Seleciona usuários por selecao de parte de toponimos regex.



Seleciona usuários por distancia editavel do nome.

06. Referências

BADASHIAN, A. S. et al. Involvement, contribution and influence in github and stack overflow. In: IBM CORP. Proceedings of 24th Annual International Conference on Computer Science and Software Engineering. [S.I.], 2014. p. 19–33.

HUANG, W. et al. Cpdscorer: Modeling and evaluating developer programming ability across software communities.

MARTINS, D.; JR, C. A. D.; FONSECA, F. T. Geocodificação de endereços urbanos com indicação de qualidade. Proceedings XIII GEOINFO, p. 36–41, 2012.

VASILESCU, B.; FILKOV, V.; SEREBRENIK, A. Stackoverflow and github: Associations between software development and crowdsourced knowledge. In: IEEE. Social Computing (SocialCom), 2013 International Conference on. [S.I.], 2013. p. 188–195.