# TweeTraffic: ferramenta de análise das condições de trânsito baseado nas informações do *Twitter*.

**Bernardo Pereira Lauand[1], Jonice Oliveira[1,2]**

[1]Programa de Pós Graduação em Informática (PPGI) – Universidade Federal do Rio de Janeiro (UFRJ)
[2] Departamento de Ciência da Computação (DCC) – Universidade Federal do Rio de Janeiro (UFRJ)
Caixa Postal 68.530 – 21.945-970 – Rio de Janeiro – RJ – Brazil

belauand@gmail.com, jonice@dcc.ufrj.br

***Abstract.*** *In large cities, traffic is becoming increasingly heavy at a fast pace. As a result, many cities, as Rio de Janeiro, face traffic problems. This project aims at showing how Twitter can be used as a resourceful data source for planning daily drive. Comments posted by drivers and news from officials agencies are processed and we infer the traffic conditions. We executed two assessments – one qualitative and other quantitative – and both results were satisfactory.*

***Resumo.*** *O trânsito nas grandes metrópoles tem se intensificado num ritmo cada vez mais rápido. Como consequência, muitas cidades, como o Rio de Janeiro, sofrem com o problema de congestionamentos. Este projeto tem como objetivo mostrar que o Twitter pode ser uma importante fonte de dados no planejamento do deslocamento diário. Os relatos dos motoristas, bem como notícias de fontes oficiais, são processados e a partir deles inferimos as condições do trânsito. Duas avaliações foram executadas - uma qualitativa e outra quantitativa - e em ambos os casos os resultados se mostraram satisfatórios.*

## 1. Introdução

Como em toda grande cidade que teve um crescimento acelerado e sem planejamento, o Rio de Janeiro precisa lidar com o problema de tráfego, tendo em média 94 quilômetros de trânsito lento diariamente.

O projeto TweeTraffic visa utilizar os dados do *Twitter* para prever as condições de trânsito na cidade do Rio de Janeiro, de maneira que os cidadãos possam planejar melhor as suas rotas.

## 2. Arquitetura Conceitual

O projeto consiste em um cliente e um servidor. O cliente, mostrado na Figura 1, é um aplicativo *Android* onde o usuário pode verificar as condições de trânsito nas ruas da cidade do Rio de Janeiro, avaliar a resposta dada pelo sistema, verificar se há novos

*tweets* de seus seguidores ou "tuitar"[1] qualquer mensagem. O servidor é responsável por extrair e analisar informações provenientes do *Twitter*. Ele se conecta periodicamente ao *Twitter* em busca de informações relevantes sobre as principais ruas, seleciona o conteúdo, as salva no banco de dados e realiza a análise de sentimentos. Esta é a etapa chamada "análise estática", que é descrita na seção 3.3. O servidor, quando consultado pelo cliente, realiza a "análise dinâmica" (seção 3.4), inferindo a condição de trânsito da via consultada. Na Figura 2 temos a arquitetura conceitual do projeto TweeTraffic.

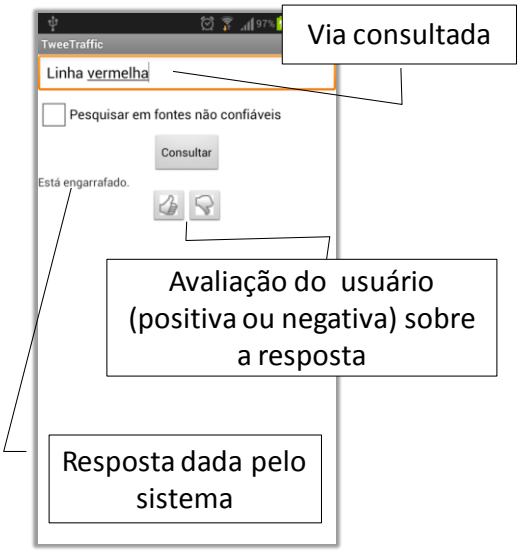


**Figura 1 – Interface do Cliente**

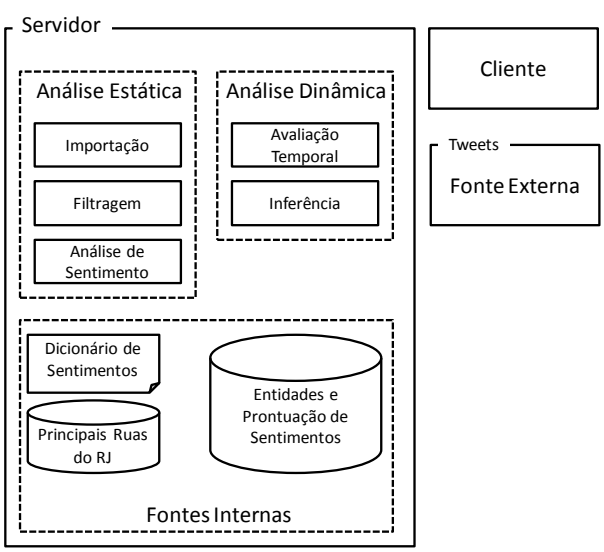


**Figura 2 – Arquitetura Conceitual**

## 3. Verificação do estado atual do trânsito

Para realizar uma consulta, o cliente se conecta ao servidor e envia a rua a ser consultada (Figura 1). A seguir são realizadas as seguintes ações:

### 3.1 Identificação da Rua

Ao receber o nome da rua a ser consultada, o servidor procura por abreviaturas comuns e as substitui pela sua palavra correspondente (figura 3.b), já que diferentes logradouros possuem nomes iguais e o que os distinguem são as denominações de via. Por exemplo: Avenida Rui Barbosa (bairro Flamengo, zona sul do Rio de Janeiro), Rua Rui Barbosa (bairro Guaratiba, zona oeste) e Travessa Rui Barbosa (zona norte).

### 3.2 Verificação da Região

Após a fase de Identificação da Rua (figura 3.b), o servidor verifica se o nome recebido é de uma rua da cidade do Rio de Janeiro (figura 3.c). Caso a rua pesquisada não seja uma rua principal, será feita uma Análise Estática (seção 3.3), caso contrário, a Análise Dinâmica será feita (seção 3.4).

---

[1] Tuitar refere-se ao ato de postar no *Twitter* comentários, informações, fotos, etc.

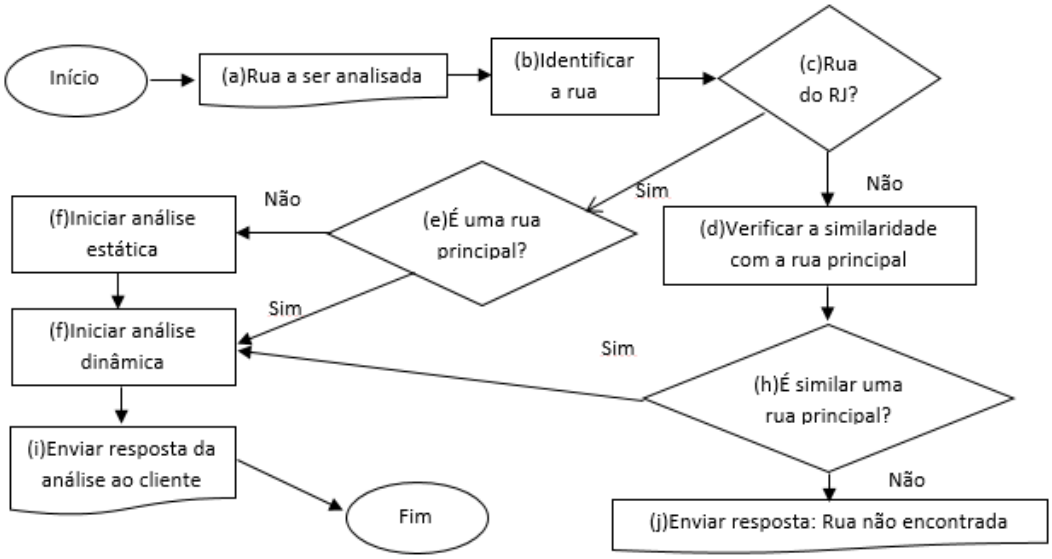


**Figura 3. Fluxograma do procedimento para a identificação do estado**

## 3.3 Análise Estática

A análise estática é realizada continuamente sobre os dados captados periodicamente, processando-os e armazenando o resultado da análise de sentimento sobre cada entidade. Caso a via a ser consultada não seja uma via principal, a análise estática é acionada após o usuário informar a rua procurada.

Esta análise é composta de três etapas: importação, filtragem e análise de sentimento. **Importação**: O servidor importa do Twitter todas as mensagens que contenham o nome da rua, postadas nos últimos 60 minutos e que não estejam armazenados no banco de dados.

**Filtragem**: Caso o usuário opte por fazer a pesquisa considerando apenas os autores confiáveis, os tweets dos autores que não constam na relação de autores confiáveis são eliminados nessa etapa. Consideramos um "autor confiável" os perfis do Twitter de importantes veículos de comunicação como perfis oficiais de revistas, jornais, estações de rádio e televisão, além de perfis de entidades da prefeitura, como CET-RIO, Rio Trânsito e o Centro de Operações da Prefeitura do Rio de Janeiro.

**Análise de Sentimento**: O servidor possui um Dicionário de Sentimentos, que é composto por um conjunto de palavras e suas respectivas classificações. Uma palavra pode ser classificada em: positiva, negativa ou modificadora. Uma palavra positiva é aquela que sugere uma boa condição de trânsito (ex: livre, bom, rápido). Já uma negativa denota uma má condição de trânsito (ex: ruim, parado, batida). Uma palavra modificadora por sua vez é aquela que altera a classificação de uma palavra de positiva para negativa ou vice versa (ex: sem, não, nenhum). A primeira fase da análise de sentimento separa o texto dos *tweets* em frases e as analisam individualmente, descartando as frases interrogativas. Para cada frase, contabiliza-se o número de palavras positivas e negativas e verifica-se a existência de palavras modificadoras. Caso haja, altera-se o valor das palavras positivas ou negativas próximas (distância de até 2

palavras) a palavra modificadora. O sistema classifica um *tweet* de acordo com a quantidade de palavras negativas e positivas**.**

### 3.4 Análise Dinâmica

Ela consiste em avaliar os *tweets* da última hora levando em consideração o resultado da análise estática e o horário em que o *tweet* foi publicado. A partir daí é feita uma inferência, baseando-se nas quatro situações distintas:

**Caso 1:** Caso não haja nenhum *tweet* sobre a via consultada nos últimos 60 minutos, o sistema retorna: "Não há informações suficientes sobre essa via.".

**Caso 2:** Caso haja *tweets*, na última hora, com classificações divergentes de acordo com a análise estática, calcula-se o intervalo de tempo entre os *tweets* mais recentes e conflitantes. Se esse intervalo for superior a 15 minutos, o conflito é ignorado e considera-se a classificação do *tweet* mais recente como certa. Caso este seja positivo, o sistema retorna: "Não está engarrafado.". Caso contrário: "Está engarrafado.".

**Caso 3:** Caso haja *tweets*, na última hora, com classificações divergentes de acordo com a análise estática e o intervalo de tempo entre os *tweets* conflitantes seja menor do que 15 minutos, é contabilizado o total de *tweets* positivos e negativos. Caso haja mais *tweets* positivos, o sistema retorna: "Provavelmente não está engarrafado.". Caso haja mais *tweets* negativos, o sistema retorna: "Provavelmente está engarrafado.". Caso o número de *tweets* positivos seja igual ao número de *tweets* negativos o sistema considera a classificação do *tweet* mais recente como certa. Se positivo, retorna: "Provavelmente não está engarrafado.". Caso contrário: "Provavelmente está engarrafado.".

**Caso 4:** Caso haja *tweets* na última hora e não haja divergência de classificação de acordo com a classificação estática, o sistema retorna: i) "Não está engarrafado" caso haja ao menos um *tweet* positivo na última hora, ii) "Está engarrafado." caso haja ao menos um *tweet* negativo na última hora e iii) "Provavelmente não está engarrafado" caso todos os *tweets* da última hora sejam neutros.

## 4. Trabalhos Correlatos

Como trabalhos correlatos podemos mencionar: *Google Maps Brasil* [2012], a solução de Endarnoto [2011], *Waze* [2012] e a abordagem de Ribeiro et al. [2012].Ao compararmos nosso projeto com os demais, identificamos diferenças na forma de como as condições de trânsito nas vias das cidades são extraídas e inferidas.

## 5. Avaliação

A avaliação foi dividida em duas etapas. A primeira composta por uma avaliação qualitativa, através de um estudo de caso para verificar o grau de utilidade e usabilidade da aplicação (seção 5.1). A segunda, através da monitoração do trânsito em diferentes horários, a verificação do resultado dado pelo TweeTraffic e o cálculo da precisão e acurácia (seção 5.2).

## 5.1 Estudo de Caso

Os participantes utilizaram o aplicativo durante a semana de 19 e 26 de agosto de 2012 e a amostra deste estudo era composta por 7 pessoas. Os testes foram realizados por indivíduos de ambos os gêneros que possuíam, em sua maioria, idade entre 22 e 25 anos e níveis de escolaridade diversos e todos eles tinham alguma experiência com o *Twitter*. Todos os usuários acharam entre fácil e muito fácil a checagem de trânsito, e 80% deles o usariam durante o cotidiano.

## 5.2 Acurácia e Precisão

Durante o período 20 a 24 de agosto o trânsito de algumas ruas foi monitorado em tempo real através das imagens de câmeras de trânsito da [CET-Rio 2012], nos horários: 8h, 12h e 18h. Para cada observação, realizamos as consultas no TweeTraffic: uma considerando apenas as fontes confiáveis e a outra considerando todas as fontes, como mostra a Figura 4.

| Dia | Horário | Imagem CET-Rio | TweeTraffic: Uso de todas as fontes | TweeTraffic: Uso apenas das fontes confiáveis |
|---|---|---|---|---|
| 22/08/12 | 12:00 | Condição real, extraída a partir da análise da imagem: "Não está engarrafado" | Provavelmente não está engarrafado | Não está engarrafado |

**Figura 4. – Exemplo de uma análise realizada, para a via com nome Linha Vermelha**

Para verificar a eficiência do sistema foram utilizadas duas métricas: precisão (fórmula 1) e acurácia (fórmula 2).

$$\text{Precisão} = \frac{\text{Positivo verdadeiro}}{\text{Positivo verdadeiro} + \text{positivo falso}}$$

**Fórmula 1. Precisão**

$$\text{Acurácia} = \frac{\text{Positivo verdadeiro} + \text{negativo verdadeiro}}{\text{Positivo verdadeiro} + \text{positivo falso} + \text{negativo verdadeiro} + \text{negativo falso}}$$

**Fórmula 2. Acurácia**

Consideramos como **positivo verdadeiro** a situação na qual a imagem coletada indicava trânsito na via e a consulta feita no TweeTraffic retornava a mensagem "Está

engarrafado”. **Positivo falso** é a situação na qual a imagem coletada indicava trânsito na via e a consulta feita no TweeTraffic retornava a mensagem “Não está engarrafado”. Já **negativo verdadeiro** é a situação na qual a imagem coletada indicava que não havia trânsito na via e a consulta feita no TweeTraffic retornava a mensagem “Não está engarrafado”. **Negativo falso** a situação na qual a imagem coletada indicava que não havia trânsito na via e a consulta feita no TweeTraffic retornava a mensagem “Está engarrafado”.

Os resultados estão na Tabela 1.

**Tabela 1. Resultados**

| | **Precisão** | **Acurácia** |
|---|---|---|
| **Todas as fontes** | 0,36 | 0,53 |
| **Fonte confiável** | 0,71 | 0,79 |

Com base nos resultados obtidos na Tabela 1, podemos observar um melhor desempenho de resultados ao se utilizar apenas fontes confiáveis.

## 6. Conclusão

De acordo com os resultados obtidos nas duas avaliações, podemos observar um interesse das pessoas em um aplicativo como esse e um grau de acerto satisfatório por parte dele. Vale ressaltar também que os resultados obtidos ao se fazer pesquisas usando todas as fontes não foram tão bons. Uma das explicações é a dificuldade em se distinguir ruas com o mesmo nome de ruas cariocas, mas que sejam de outra cidade. O pouco uso do *Twitter* para informar o trânsito em ruas menores torna difícil a verificação das condições de trânsito nessas vias. O mesmo problema ocorre com vias extensas e de mão dupla, nas quais é um obstáculo descobrir em que região e sentido da rua se encontram os problemas. Outra dificuldade encontrada foi o fato de os usuários do *Twiiter* não georeferenciarem seus *tweets,* tornando impossível descobrir a real localização dos usuários na hora de filtra-los.

## Agradecimento

Ao CNPq.

## Referências

Endarnoto, S. K. (2011) "Traffic Condition Information Extraction & Visualization from Social Media Twitter for Android Mobile Application" in Electrical Engineering and Informatics (ICEEI), 2011

Google Maps (2012), http://maps.google.com.br, Setembro

Ribeiro, S. S., Rennó, D., Gonçalves, T. S., Davis, C. A., Meira W. (2012) " Observatório do Trânsito: sistema para detecção e localização de eventos de trânsito no Twitter " in Simpósio Brasileiro de Bancos de Dados - SBBD, 2012

Waze (2012) "Free GPS Navigation with Turn by Turn - Waze", http://www.waze.com