

# ***Twittraffic*: Uma plataforma de monitoração, visualização e identificação de ocorrências no trânsito**

**Lucas Schmidt<sup>1</sup>, Jeferson Ferreira<sup>1</sup>, Vicente Amorim<sup>1</sup>,  
Ricardo Oliveira<sup>1</sup>, Fabricio Benevenuto<sup>2</sup>**

<sup>1</sup>Departamento de Computação – Universidade Federal de Ouro Preto (UFOP)  
Ouro Preto – MG – Brasil

<sup>2</sup>Departamento de Computação – Universidade Federal de Minas Gerais (UFMG)  
Belo Horizonte – MG – Brasil

{lucasscf, jerf21, vjpmorim, rrabelo, benevenuto}@gmail.com

**Abstract.** *In the current context, the traffic is one of the most critical factors of urban mobility in cities. In this perspective, this paper proposes **Twittraffic**: a platform for monitoring, visualization and identification of traffic events. The platform enables the sharing of information on social networks like Twitter, so pervasive and collaborative. During the period from December 2012 to March 2013, data were collected (tweets) with content related to transit in the cities of Belo Horizonte, Rio de Janeiro and São Paulo. After the analysis and processing of data collected, it was possible to profile the traffic of the cities studied, identifying time, place (streets) and days of the week when there is more traffic events. The study showed results of collecting close to 90% confidence in the context of traffic, which demonstrates the efficiency of the operation of the platform.*

**Resumo.** *No contexto atual, o trânsito é um dos fatores mais críticos da mobilidade urbana nas grandes cidades. Nesta perspectiva, o presente trabalho propõe a **Twittraffic**: uma plataforma de monitoração, visualização e identificação de ocorrências no trânsito. A plataforma permite, o compartilhamento das informações nas redes sociais como o Twitter, de forma pervasiva e colaborativa. Durante o período de dezembro de 2012 a março de 2013, foram coletados dados (tweets) com conteúdo relacionado ao trânsito das cidades de Belo Horizonte, Rio de Janeiro e São Paulo. Ao final da análise e tratamento dos dados coletados, foi possível traçar o perfil do trânsito das cidades estudadas, identificando horário, local (vias) e dias da semana em que há mais ocorrências de trânsito. O estudo apresentou resultados de coleta próximos a 90% de confiança no contexto de trânsito, fato que demonstra a eficiência do funcionamento da plataforma.*

## **1. Introdução**

Com o crescimento da frota de veículos e conseqüente aumento de ocorrências de trânsito nas principais capitais do país, desvela-se a necessidade de buscar alternativas gerenciais. Nesta perspectiva, estudos e pesquisas constituem-se como ferramentas importantes para identificação de pontos críticos no trânsito.

Um dos instrumentos utilizados com fins de estudo e pesquisa são as redes sociais. Redes sociais como *Twitter* e *Facebook*, possibilitam que usuários comuns propaguem e compartilhem suas experiências e opiniões em tempo real, através da *web*. No entanto, a maior quantidade de conteúdo disponível, gera dificuldades de visualização e monitoração das informações disponibilizadas.

Dessa forma, com base nas informações compartilhadas no *Twitter*, propõe-se a *Twittraffic*: uma plataforma de monitoração, visualização e identificação de ocorrências no trânsito. Através dessa ferramenta, é possível alertar o usuário sobre possíveis eventos no trânsito e sua incidência de acordo com a rota pré-definida. Paralelamente, busca-se o compartilhamento da informação em redes sociais, de forma pervasiva e colaborativa, extraindo informações de sensores móveis no contexto de um veículo móvel.

Além disso, é interessante abordar que esta proposta visa atender um dos desafios da comunidade em geral (referente ao trânsito) e principalmente da gestão governamental de uma cidade. Nesta perspectiva, a plataforma permite ainda traçar o perfil do trânsito em grandes centros urbanos, tais como Belo Horizonte, Rio de Janeiro e São Paulo, foco do estudo em questão.

O artigo encontra-se organizado da seguinte forma: na seção 2 estão os trabalhos relacionados com o tema. Na seção 3 as fases de concepção e implementação da plataforma. Já na seção 4 são apresentados os estudos de caso realizados com a plataforma. A seção 5 apresenta a forma de coleta de dados e seus resultados são mostrados na seção 6. Na seção 7 apresentamos os trabalhos futuros e na seção 8 as conclusões do artigo.

## **2. Trabalhos Relacionados**

As redes sociais, como o *Twitter*, adquirem no cenário atual, destaque e crescente importância na sociedade, principalmente por possibilitar a propagação e compartilhamento de informações, experiências e opiniões em tempo real. Como exemplo, um estudo realizado no Japão [Sakaki et al. 2010] mostrou que através do monitoramento de *tweets*, que caracterizavam eventos de terremotos, foi possível com a probabilidade de 96%, detectar e alertar usuários, em tempo real, sobre os terremotos no país. Em outros estudos o *Twitter* também foi utilizado como fonte de informação para o monitoramento de epidemias, como a dengue [Gomide et al. 2011], H1N1 [Chew and Eysenbach 2010] e eleições [Tumasjan et al. 2010].

No contexto do trânsito, baseado em informações das redes sociais, trabalhos como [Endarnoto et al. 2011] e [Albuquerque et al. 2012] apresentam abordagens interessantes na extração dos dados de trânsito do *Twitter* através de técnicas de PLN (Processamento de Linguagem Natural). Enquanto o primeiro estudo considera informações do tempo, origem, destino e condição do trânsito para definir regras de busca, o segundo utiliza de recursos mais detalhados como local, ponto de referência, direção da pista, intensidade do tráfego e a condição do trânsito. Técnicas como essas que não foram aprofundadas neste trabalho. Já o trabalho [Kosala et al. 2012], propõe um algoritmo para medir o nível de confiança das informações de trânsito, extraídas no *Twitter*, para um sistema de mapeamento em tempo real. Neste caso, tal confiança também pode ser questionada pois trata-se de um sistema específico, e não considera condições e características especiais de determinadas regiões, que podem ser ou não críticas.

Além disso, as redes sociais podem explorar o conteúdo captado por sen-

sores móveis (GPS, acelerômetros, sensores de localização ou de monitoramento de tráfego ou veículos), que são capazes de identificar situações e ocorrências de acordo com um determinado contexto [Aggarwal and Abdelzaher 2011]. Nos trabalhos [Demirbas et al. 2010], [Baquer 2009] e [Rosi et al. 2011], são apresentados projetos de infraestrutura capazes de integrar diretamente dados captados por sensores às redes sociais, através de métodos avançados de colaboração e coordenação.

### **3. Concepção e implementação da plataforma**

#### **3.1. Estágio I: Modelo Conceitual**

O modelo conceitual engloba cinco definições básicas: contexto, entidade, fonte, ocorrência e intensidade. No estudo, o contexto é o trânsito; a entidade correlata é o local da ocorrência de trânsito em determinada cidade (Belo Horizonte, Rio de Janeiro e São Paulo); a fonte é o *Twitter* e a ocorrência pode ser atropelamento, congestionamento, acidentes e outros. A intensidade é a relevância de ocorrências relacionadas ao trânsito, variando de acordo com a entidade.

#### **3.2. Estágio II: *Twittraffic* - Plataforma de trânsito**

##### **3.2.1. Coleta de dados**

A coleta de dados envolve especificamente o microblog *Twitter*. O tipo de informação proveniente dessa mídia representa tanto a visão de cidadãos quanto a de formadores de opinião da sociedade. Toda a coleta de dados é baseada em um dicionário controlado. A definição das palavras utilizadas no dicionário é considerada uma das fases mais importantes da execução da plataforma e está detalhada na seção 5.

##### **3.2.2. Filtragem e Extração**

A fase de filtragem e extração começa com a padronização da codificação dos caracteres, uma vez que diferentes fontes podem utilizar diferentes padrões (por exemplo, UTF-8, ISO8859-1). Além disso, são realizados tratamentos de data e hora, já que esses dados não são disponibilizados no formato ideal.

Posteriormente, são aplicados métodos tradicionais de pré-processamento de textos que extraem os radicais das palavras, tais como remoção de *stop words*, palavras comuns, como “o”, “de”, “no”, “para”, etc e *stemming*.

Finalmente, são aplicados métodos simplistas de processamento de linguagem natural na separação de locais, ruas, avenidas, bairros e regiões, além da condição do trânsito, de cada *tweet* coletado. Isso ocorre com a finalidade de identificar as entidades de interesse no contexto de trânsito.

##### **3.2.3. Identificação geográfica**

Os dados provenientes da fase de filtragem e extração servem de entrada para a identificação geográfica das entidades selecionadas. Ainda em formato textual, ocorre a consulta desta entidade em uma base de dados de localidades e coordenadas geográficas.

Quando esta não é localizada, ocorre a consulta e verificação diretamente na API do *Google Maps*. Uma vez encontrada, esta é armazenada na base de dados.

### 3.2.4. Exportação

Através da identificação geográfica das ocorrências de trânsito, o *web-service* disponibiliza a cada trinta minutos uma nova versão de documentos. Esses documentos possuem formatos leves e são utilizados para o intercâmbio de dados computacionais (JSON e KML). O conteúdo dos dados computacionais são as ocorrências de trânsito, bem como as indicações de intensidade de cada ocorrência.

Os referidos documentos podem ser lidos e representados por qualquer plataforma, dispositivos ou sistema operacional. Além disso, podem ser visualizados através dos *links*:

#### Ocorrências de trânsito - KML

- <http://www.twittrafficbh-tweet.vai.la> (Belo Horizonte)
- <http://www.twittrafficrj-tweet.vai.la> (Rio de Janeiro)
- <http://www.twittrafficsp-tweet.vai.la> (São Paulo)

#### Intensidades no trânsito - KML

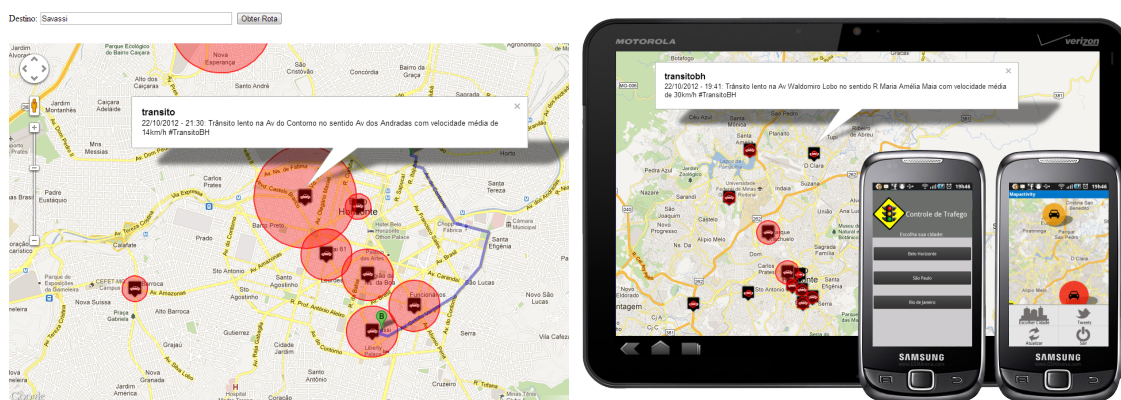
- <http://www.twittrafficbh-alert.vai.la> (Belo Horizonte)
- <http://www.twittrafficrj-alert.vai.la> (Rio de Janeiro)
- <http://www.twittrafficsp-alert.vai.la> (São Paulo)

### 3.2.5. Visualização

Uma vez que a intenção final do projeto é possuir uma plataforma de monitoração, visualização e identificação de ocorrências no trânsito, a visualização dos dados é parte imprescindível, sendo essencial para o êxito do trabalho. Em um mapa, apresenta-se geograficamente as ocorrências de trânsito e suas intensidades em forma de alerta. O objetivo é facilitar a visualização para usuários gerenciais e motoristas que acessam a ferramenta dentro de um veículo automotor ou através de dispositivos móveis.

O mapa geográfico e a aplicação para dispositivo móvel Android, que permite a visualização de ocorrências e intensidades de trânsito, podem ser visualizados através da Figura 1 ou utilizados através dos *links*:

- <http://www.twittrafficbh.vai.la> (Belo Horizonte)
- <http://www.twittrafficrj.vai.la> (Rio de Janeiro)
- <http://www.twittrafficsp.vai.la> (São Paulo)



**Figura 1. Twittraffic: Execução da plataforma na web e em dispositivos móveis**

## 4. Estudo de caso

O estudo de caso tem como objetivo traçar o perfil do trânsito nas cidades estudadas, com base nos dados obtidos na coleta. Com isso, torna-se possível confirmar e complementar dados estatísticos e gerenciar sobre o trânsito. Além disso, através desse estudo, torna-se possível verificar a eficiência do funcionamento da plataforma na coleta, extração, filtragem e identificação geográfica dos dados.

### 4.1. Estudo 1 - Relevância das palavras utilizadas no dicionário de busca

Para verificar a relevância das palavras utilizadas no dicionário de busca, foram realizados cálculos amostrais. Ocorreu posteriormente, a seleção de uma amostra de *tweets* para verificar se o conteúdo desses estava de acordo com a palavra selecionada.

### 4.2. Estudo 2 - Seleção e avaliação de novas palavras para o dicionário de busca

Investigou-se quais as palavras eram mais recorrentes nos *tweets* coletados em cada cidade. Além disso, foram verificados os *bi-grams* (termos com palavras duplas) com mais ocorrência nos *tweets* coletados.

### 4.3. Estudo 3 - Perfil do trânsito nas cidades

Foram realizadas análises que permitiram traçar o perfil do trânsito de cada cidade estudada. Essas, são importantes por servirem de base estratégica para melhorar a distribuição e organização do trânsito da cidade.

## 5. Coleta de dados

### 5.1. Instrumento de Coleta

Para iniciar o processo de coleta foi preciso definir as palavras que iriam fazer parte do dicionário de busca no *Twitter*. A escolha dessas ocorreu através de uma busca simples no *Twitter* com o termo “transito”. Com o resultado dessa busca, foram coletados 50 *tweets* de forma aleatória. Desses, foram selecionadas algumas palavras com maior ocorrência nos *tweets*, e que faziam parte do contexto de trânsito. Com isso, inicialmente, o dicionário de busca foi formado pelos seguintes termos: “transito(uf)”<sup>1</sup>, “transito”, “agarrado”, “garrado”, “parado”, “congestionado” e “lotado”.

<sup>1</sup>A uf é a sigla para Unidade da Federação. Unidade da Federação refere-se aos Estados brasileiros. Exemplo: São Paulo (SP)

## 5.2. Período de Coleta

O período da coleta iniciou-se no dia 10/12/2012 e finalizou-se no dia 10/03/2013. Nesses três meses (91 dias), a coleta de dados ocorreu de forma ininterrupta para as cidades de Belo Horizonte, Rio de Janeiro e São Paulo. O *Crawler* foi preparado para fazer, por dia, um ciclo da execução à cada 30 minutos, porém, o período de execução é variável e pode ser facilmente configurado.

## 5.3. Disponibilização dos dados coletados

Todos os dados coletados foram organizados e disponibilizados no *link*: <http://www.twittraffic-log.vai.la>

## 6. Resultados

Após o período de 91 dias, os dados coletados através da plataforma, foram analisados e tratados. Foram coletados 24.895 *tweets* para a cidade de Belo Horizonte, 27.919 *tweets* para a cidade de Rio de Janeiro e 17.530 *tweets* para a cidade de São Paulo. Após o tratamento de dados, com a exclusão de *tweets* redundantes, foi observado que o total de *tweets* únicos coletados para Belo Horizonte foi de 2.320, 2.395 para o Rio de Janeiro e 2.284 para São Paulo.

### 6.1. Estudo 1 - Relevância das palavras utilizadas no dicionário de busca

O processo que determina as palavras mais relevantes e a elaboração do melhor dicionário de busca é um dos mais importantes para o funcionamento adequado da plataforma. Obviamente, a ferramenta utiliza deste dicionário para realizar pesquisas sucessivas no *Twitter*.

Após o tratamento dos dados, foram realizados cálculos da relevância de cada palavra buscada para cada cidade estudada. A realização desses cálculos é imprescindível, uma vez que evidencia o nível de confiança das palavras utilizadas no dicionário de busca, além de mostrar o quão confiável é o próprio dicionário.

A Tabela 1 mostra detalhadamente a relevância de cada palavra. Para cada cidade, foi verificado se o conteúdo de uma determinada amostra de *tweets* estava de acordo com a palavra buscada, no contexto de trânsito (denominado “*Tweets* corretos”). Como amostra, foram selecionados, de forma aleatória, 100 *tweets* da população total, relacionados a cada palavra buscada. No caso de palavras com população de *tweets* abaixo de 100, a amostra selecionada corresponde ao total dessa população.

Desta forma, os cálculos da relevância de cada palavra são realizados através de cálculos de porcentagem simples para cada palavra  $R(p)$ . Já, os cálculos da relevância do dicionário consideram os valores totais, de cada dicionário, para cálculo da porcentagem  $R(d)$ . Posteriormente, foi calculado a média, considerando as relevâncias de cada dicionário, para definir a eficiência dos termos de coleta da plataforma.

$$R(p) = \left( \frac{\textit{Tweets corretos}}{\textit{Amostra}} \right) * 100 \quad R(d) = \left( \frac{\textit{Total de Tweets corretos}}{\textit{Total da Amostra}} \right) * 100$$

É possível observar que a relevância do dicionário de busca, foi de 92,51% para a cidade de Belo Horizonte, 92,83% para a cidade do Rio de Janeiro e 83,75% para a cidade de São Paulo. Com isso, a confiança dos termos de coleta é de aproximadamente 90%.

Palavra	Belo Horizonte				Rio de Janeiro				São Paulo			
	População de Tweets	Amostra	Tweets corretos	Relevância(%)	População de Tweets	Amostra	Tweets corretos	Relevância(%)	População de Tweets	Amostra	Tweets corretos	Relevância(%)
transito	1476	100	100	100	1595	100	100	100	1328	100	100	100
transito(uf)	602	100	100	100	661	100	100	100	706	100	100	100
garrado	19	19	16	84,21	3	3	3	100	2	2	1	50
agarrado	21	21	21	100	0	0	0	0	0	0	0	0
parado	141	100	92	92	104	100	93	93	150	100	89	89
lotado	27	27	8	29,62	22	22	5	22,72	92	92	39	42,39
congestionado	34	34	34	100	10	10	10	100	6	6	6	100
<b>Total</b>	<b>2320</b>	<b>401</b>	<b>371</b>	<b>92,51</b>	<b>2395</b>	<b>335</b>	<b>311</b>	<b>92,83</b>	<b>2284</b>	<b>400</b>	<b>335</b>	<b>83,75</b>

**Tabela 1. Relevância das palavras buscadas**

Belo Horizonte				Rio de Janeiro				São Paulo			
Palavra	Quant.	Bi-gram	Quant.	Palavra	Quant.	Bi-gram	Quant.	Palavra	Quant.	Bi-gram	Quant.
transito	1471	sentido BH	359	transito	1423	Av Brasil	348	SP	1173	Km lentidão	376
TransitoBH	964	BR 381	224	Av	1228	transito lento	581	transito	1118	sentido Centro	112
BH	838	BR 040	158	sentido	1148	sentido Centro	250	Av	1052	sentido Marginal	101
Av	673	lento Av	145	transitoRJ	772	lento sentido	228	sentido	805	Av Bandeirantes	86
lento	474	velocidade média	142	altura	681	lento Av	171	TransitoSP	729	sentido SP	82
BR	471	trânsito lento	112	lento	649	deixa transito	156	lentidão	458	Acidente Resgate	67
Acidente	276	BH Shopping	104	transitoRJO	516	Centro altura	111	Km	391	Resgate local	67
381	275	Cristiano Machado	104	Acidente	482	causa lentidao	98	Vd	213	Av Brig	61
parado	219	sent BH	86	RJ	383	Brasil sentido	97	Avenida	201	sentido Av	60
transito_bh	181	Av Amazonas	85	lentidao	319	Linha Amarela	97	Rua	183	Juliano Costa	60
40	175	Grande BH	85	Avenida	297	Brasil altura	76	São	139	Vd João	60
Transito98FM	163	sentido Av	84	Centro	272	Francisco Bicalho	73	parado	130	transito sp	59
intenso	148	BH trânsito	70	Brasil	227	lento Avenida	71	Marginal	128	sentido Santana	58
sent	145	trânsito intenso	67	causa	202	deixa trânsito	68	Acidente	122	Luís Antônio	57
média	142	BH Segundo	63	carro	196	pista central	67	João	121	Brig Luís	57
velocidade	142	381 sentido	61	BR-101	187	Ocorrencia Onibus	63	Centro	104	Av Estado	55
Amazonas	139	deixa trânsito	60	Onibus	187	Onibus enguicado	61	Bandeirantes	94	Lentidão sentido	52
km	139	040 sentido	50	intenso	180	Av Francisco	61	Luís	88	sentido Bairro	51
Grande	135	Av Cristiano	50	pista	164	central Av	60	Santo	84	Av Rebouças	51
altura	125	sentido centro	47	Sao	155	Brasil transitoRJ	59	Túnel	74	transito lento	48

**Tabela 2. As 20 palavras e os 20 bi-grams mais recorrentes em cada cidade**

Nas três cidades estudadas, a palavra “lotado” obteve valores de relevância insignificantes, no contexto de trânsito. Assim como a palavra “agarrado”, nas cidades de Rio de Janeiro e São Paulo. Como o objetivo do estudo é mostrar que a ferramenta permite traçar o perfil do trânsito das cidades, essas palavras foram eliminadas dos demais estudos apresentados neste trabalho.

## 6.2. Estudo 2 - Seleção e avaliação de novas palavras para o dicionário de busca

De acordo com o Estudo 1, um dos fatores mais importantes para o melhor funcionamento da plataforma é a definição das palavras a serem utilizadas no dicionário de busca. Com isso, a partir dos *tweets* coletados em cada cidade, foram verificadas as palavras e os *bi-grams* (termos com palavras duplas) mais recorrentes. O objetivo é a elaboração de uma ferramenta de busca com um maior nível de confiança. Além disso, um número menor de *tweets* redundantes possibilita uma busca mais abrangente. Também é importante salientar que devem ser definidos os melhores termos específicos correspondentes à linguagem e gírias de cada região.

A Tabela 2 apresenta as 20 palavras e 20 *bi-grams* mais encontrados para cada cidade.

## 6.3. Estudo 3 - Perfil do trânsito nas cidades

Com os dados dos *tweets* coletados, tratados e organizados, realizou-se o agrupamento de informações que permitem traçar o perfil do trânsito das cidades de Belo Horizonte, Rio de Janeiro e São Paulo. A análise dessas informações é pertinente já que podem ser utilizadas como base estratégica e gerencial para melhorar a distribuição e organização do trânsito nas cidades.

### 6.3.1. Belo Horizonte

Os gráficos correspondentes à Figura 2 representam o total de *tweets* coletados na cidade de Belo Horizonte no período de 91 dias. No gráfico (a) é possível visualizar picos e vales. Os picos mais elevados representam dias de grandes ocorrências no trânsito. Como exemplo, o maior pico (104 *tweets*) no dia 1 (que representa o dia 10/12/2012), é caracterizado pelas manifestações de vários usuários sobre alagamentos e pancadas de chuva na cidade (<http://vai.la/2Yr7>). Além disso, são visíveis os vales (marcados em vermelho) que representam os dias do final de semana, onde há menor incidência no trânsito. Esse declínio nos eventos de trânsito também pode ser observado no período do Natal e final de ano (marcado de verde). No gráfico (b) é representado o acúmulo de *tweets* (referentes ao trânsito) por dia.

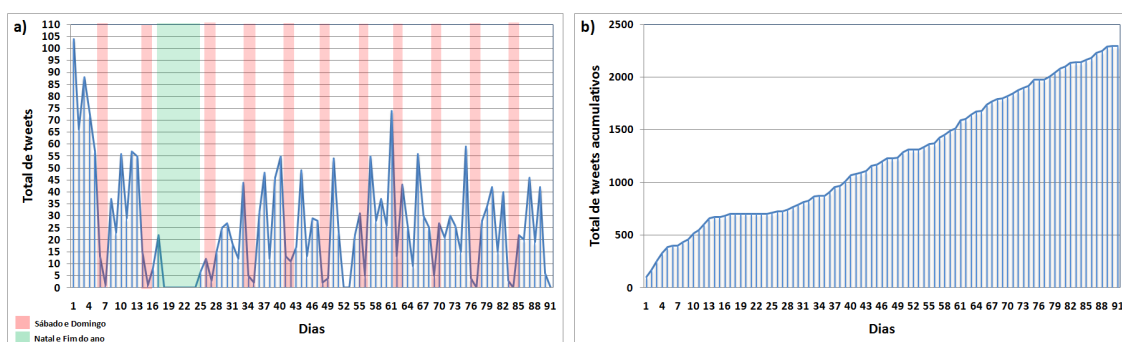


Figura 2. (a) Total de Tweets por Dia (b) Acumulativo de Tweets por Dia

Nos gráficos correspondentes à Figura 3 são apresentadas informações que caracterizam os dias da semana e as horas com maior intensidade de ocorrências no trânsito. No gráfico (a) é visível a grande quantidade de ocorrências no trânsito nas segundas-feiras, quartas-feiras e principalmente às sextas-feiras. Ao final dos 91 dias, a quantidade de *tweets* das sextas-feiras, ultrapassou o número de 500.

No gráfico (b) é apresentado um perfil dos principais horários de pico da cidade. O horário de pico na parte da manhã é das 07 horas até pouco mais das 11 horas. Na parte da tarde, a partir das 17 horas até pouco mais das 20 horas. Em horários de pico, o agrupamento da quantidade de *tweets* aproxima-se de 260, durante os 91 dias.

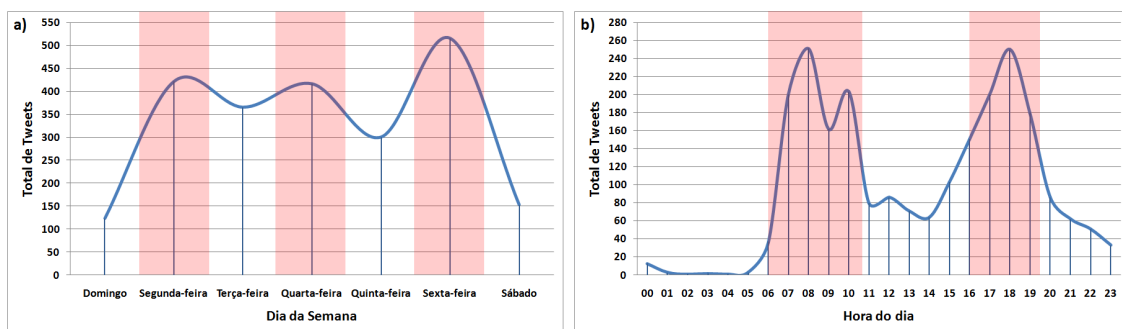


Figura 3. (a) Total de Tweets por Dia da Semana (b) Total de Tweets por Hora

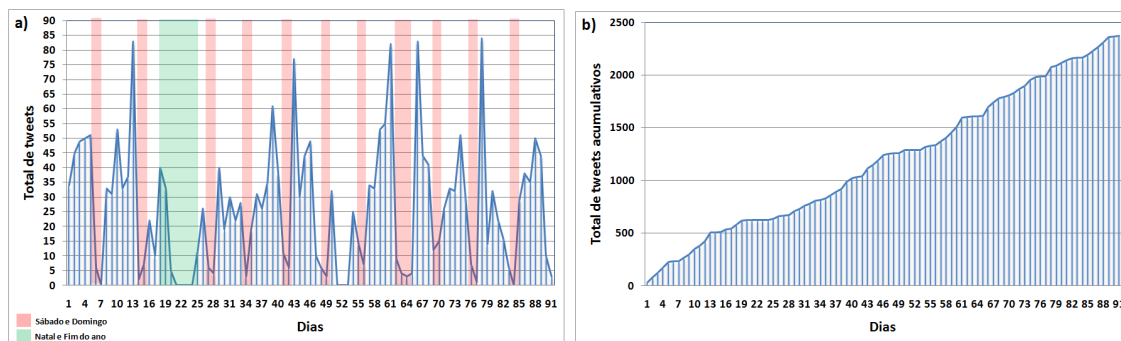
Além dessas informações, a plataforma também possibilitou apontar, de acordo com os *tweets* coletados, as vias, ruas e locais com maior intensidade de ocorrências



no trânsito. Em Belo Horizonte os pontos identificados como mais críticos são: BR-040, Avenida Presidente Antônio Carlos, Avenida Brasil, Avenida Cristiano Machado e Avenida Francisco Sales.

### 6.3.2. Rio de Janeiro

Assim como apresentado na cidade de Belo Horizonte, os gráficos correspondentes à Figura 4 representam o total de *tweets* coletados na cidade do Rio de Janeiro no período de 91 dias de coleta. No gráfico (a) um perfil interessante e perceptível, é o grande número de *tweets* de trânsito próximos aos finais de semana (aproximado de 80 *tweets*), o que indica maior quantidade de ocorrências de trânsito nas segundas-feiras e sextas-feiras. No gráfico (b) é representado o acúmulo de *tweets* (referentes ao trânsito) por dia.



**Figura 4. (a) Total de Tweets por Dia (b) Acumulativo de Tweets por Dia**

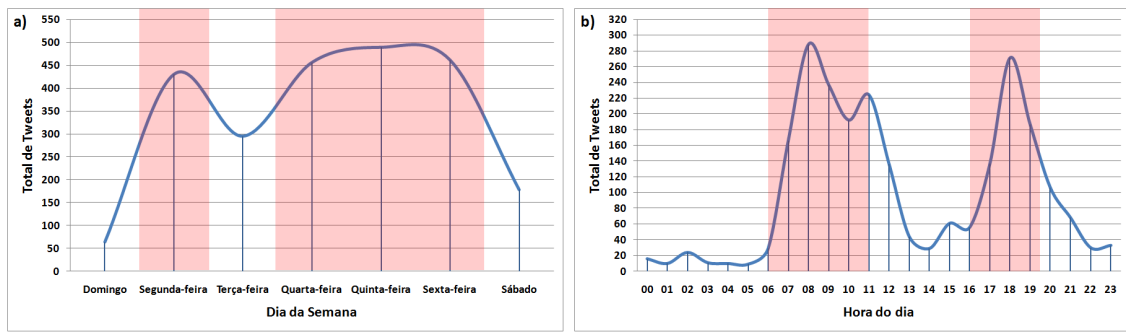
Nos gráficos correspondentes à Figura 5 são apresentadas informações que caracterizam os dias da semana e as horas com maior intensidade de ocorrências no trânsito. No gráfico (a) é visível a grande quantidade de eventos no trânsito nas segundas-feiras, quartas-feiras e principalmente quintas-feiras e sextas-feiras, o que comprova a observação realizada no gráfico correspondente à Figura 4 (a). Para as quintas-feiras e sextas-feiras, o agrupamento dos *tweets* aproximou da quantidade de 500 *tweets* durante os 91 dias.

No gráfico (b) é apresentado um perfil dos principais horários de pico da cidade. O horário de pico na parte da manhã, assim como em Belo Horizonte, também é a partir das 07 horas e permanece até pouco mais das 12 horas. Na parte da tarde, a partir das 17 horas até pouco mais das 20 horas. Em horários de pico, o agrupamento da quantidade de *tweets* aproxima-se de 280, durante os 91 dias.

A plataforma também conseguiu indicar, de acordo com os *tweets* coletados, que as vias, ruas e locais com pontos mais críticos da cidade do Rio de Janeiro são: Avenida Brasil, BR-101, Avenida Presidente Vargas, Avenida Dom Helder Camara e Avenida Niemeyer.

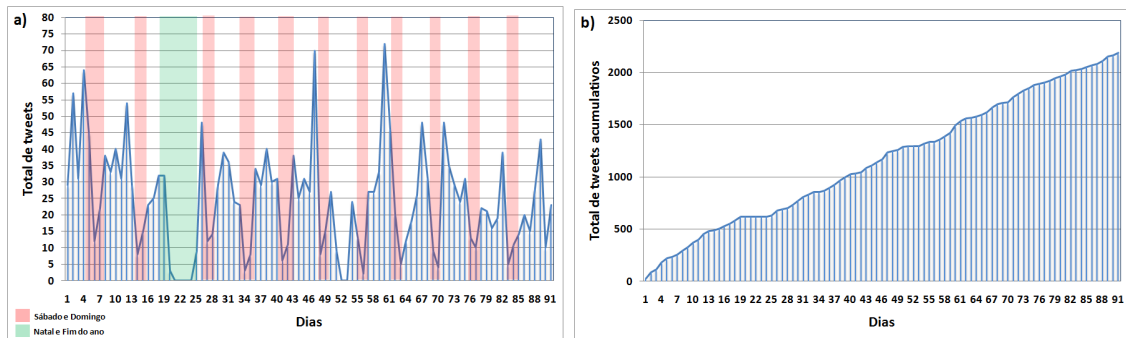
### 6.3.3. São Paulo

Os *tweets* coletados na cidade de São Paulo, no período de 91 dias, estão representados pelo gráfico correspondente à Figura 6. É possível ver um perfil do trânsito semelhante



**Figura 5. (a) Total de Tweets por Dia da Semana (b) Total de Tweets por Hora**

ao encontrado na cidade do Rio de Janeiro: grande número de *tweets* de trânsito próximo aos finais de semana (aproximadamente entre 50 e 70 *tweets*), o que indica grande quantidade de ocorrências de trânsito nas segundas-feiras e sextas-feiras. No gráfico (b) é representado o acúmulo de *tweets* (referentes ao trânsito) por dia.



**Figura 6. (a) Total de Tweets por Dia (b) Acumulativo de Tweets por Dia**

Os gráficos correspondentes à Figura 7 também apresentam informações que caracterizam os dias da semana e as horas com maior intensidade de ocorrências no trânsito na cidade de São Paulo. No gráfico (a) é visível a grande quantidade de ocorrências no trânsito nas segundas-feiras e principalmente quintas-feiras e sextas-feiras, o que também comprova a observação realizada no gráfico correspondente à Figura 6 (a). Nas quintas-feiras e sextas-feiras, o agrupamento dos *tweets* aproximou da quantidade de 500 *tweets* durante os 91 dias.

No gráfico (b) é apresentado um perfil dos principais horários de pico da cidade. O horário de pico na parte da manhã, assim como em Belo Horizonte e Rio de Janeiro, também é a partir das 07 horas e permanece até pouco mais das 13 horas. Na parte da tarde, a partir das 17 horas até pouco mais das 21 horas. Em horários de pico, o agrupamento da quantidade de *tweets* aproxima-se de 200, durante os 91 dias.

No gráfico (b) também é visível uma maior distribuição das ocorrências de trânsito nos demais horários de um dia. Esta característica sugere um perfil de trânsito com grandes índices de ocorrência de trânsito, não só nos horários de pico.

De acordo com os *tweets* coletados, as vias, ruas e locais com pontos mais críticos da cidade de São Paulo são: Avenida Engenheiro Billings, Avenida Brigadeiro Luís Antônio, Rodoanel (SP-021), Avenida do Estado e Avenida Brasil.

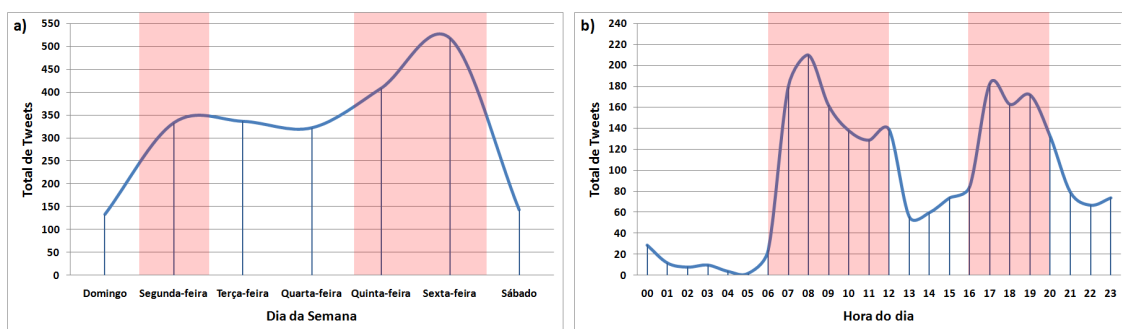


Figura 7. (a) Total de Tweets por Dia da Semana (b) Total de Tweets por Hora

## 7. Trabalhos Futuros

Uma coleta de dados cada vez mais apurada e confiável da plataforma, é um dos resultados esperados neste estudo. Com as palavras e *bi-grams* apresentados na Tabela 2, foi definido um novo dicionário. O dicionário será constituído pelos termos: “transito”, “lento”, “acidente”, “parado”, “intenso”, “velocidade”, “sentido”, “lentidão”, “pista”, “sentido (uf)”, “transito lento”, “transito intenso” e “transito centro”. Esses termos serão utilizados nas próximas execuções da plataforma. Posteriormente, os cálculos de relevância para esses novos termos e novas análises do perfil do trânsito serão realizados, como no Estudo 1, 2 e 3.

Também é necessário aplicar técnicas aprofundadas em processamento de linguagem natural. Essas técnicas irão permitir melhor busca, extração e filtragem dos dados de acordo com cada cidade.

Além disso, os resultados obtidos neste trabalho serão validados com dados de empresas especialistas na monitoração e gerenciamento do trânsito nas cidades.

## 8. Conclusões

O trabalho propôs a *Twittraffic*: uma plataforma de monitoração, visualização e identificação de ocorrências no trânsito. Com esta ferramenta é possível alertar o usuário sobre possíveis eventos no trânsito e sua incidência de acordo com uma rota pré-definida.

Além disso, foi possível traçar o perfil do trânsito das cidades estudadas, através de análises dos dados coletados. Pôde-se também verificar a eficiência dos termos de coleta da plataforma, através do cálculo da confiança do dicionário utilizado para busca em cada cidade. Obteve-se como resultado 92,51% de confiança para a cidade de Belo Horizonte, 92,83% para a cidade do Rio de Janeiro e 83,75% para a cidade de São Paulo, resultando em uma confiança de aproximadamente 90% para os termos de coleta da plataforma. Conclui-se que quanto maior o nível de confiança, melhor e mais abrangente é a coleta e filtragem de dados no contexto de trânsito.

## 9. Agradecimentos

Agradecemos a Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) pelo suporte no Projeto Universal - Processo APQ-00577-09 - e a empresa SEVA Engenharia Eletrônica S/A pelo suporte ao projeto KITT, beneficiada pela lei de informática: 8248, 10.176 e 11.077, Brasil.

## Referências

- Aggarwal, C. C. and Abdelzaher, T. F. (2011). Integrating sensors and social networks. In *Social Network Data Analytics*, pages 379–412.
- Albuquerque, F. C., Bacelar, F. C., Tapia, X. A. C., and Carvalho, M. T. M. (2012). Extrator de fatos relacionados ao tráfego. *SBB'D'12 Simpósio Brasileiro de Bancos de Dados*, pages 169 – 176.
- Baqer, M. (2009). S-sensors: Integrating physical world inputs with social networks using wireless sensor networks. In *5th International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2009*, pages 213–218. IEEE Computer Society.
- Chew, C. and Eysenbach, G. (2010). Pandemics in the age of twitter: Content analysis of tweets during the 2009 h1n1 outbreak. *PLoS ONE*, 5(11):e14118.
- Demirbas, M., Bayir, M. A., Akcora, C. G., Yilmaz, Y. S., and Ferhatosmanoglu, H. (2010). Crowd-sourced sensing and collaboration using twitter. In *Proceedings of the 2010 IEEE International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM), WOWMOM '10*, pages 1–9, Washington, DC, USA. IEEE Computer Society.
- Endarnoto, S. K., Pradipta, S., Nugroho, A. S., and Purnama, J. (2011). Traffic condition information extraction visualization from social media twitter for android mobile application. In Syaichu-Rohman, A., Hamdani, D., Akbar, S., Adiprawita, W., Razali, R., and Sahari, N., editors, *ICEEI*, pages 1–4. IEEE.
- Gomide, J., Veloso, A., Jr., W. M., Almeida, V., Benevenuto, F., Ferraz, F., and Teixeira, M. (2011). Dengue surveillance based on a computational model of spatio-temporal locality of twitter. In *ACM Web Science Conference (WebSci)*.
- Kosala, R., Adi, E., and Steven (2012). Harvesting real time traffic information from twitter. *Procedia Engineering*, 50(0):1 – 11. International Conference on Advances Science and Contemporary Engineering 2012.
- Rosi, A., Mamei, M., Zambonelli, F., Dobson, S., Stevenson, G., and Ye, J. (2011). Social sensors and pervasive services: Approaches and perspectives. In *PerCom Workshops*, pages 525–530.
- Sakaki, T., Okazaki, M., and Matsuo, Y. (2010). Earthquake shakes twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web, WWW '10*, pages 851–860, New York, NY, USA. ACM.
- Tumasjan, A., Sprenger, T. O., Sandner, P. G., and Welp, I. M. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. In *ICWSM'10*, pages 1–1.