



Evento	Salão UFRGS 2014: FEIRA DE INOVAÇÃO TECNOLÓGICA DA UFRGS – FINOVA
Ano	2014
Local	Porto Alegre
Título	Extraindo Informações de Trânsito a partir de Textos da WEB
Autor	THIAGO BELL FELIX DE OLIVEIRA
Orientador	ANA LUCIA CETERTICH BAZZAN

Extraindo Informações de Trânsito a partir de Textos da WEB

Simulações de trânsito ajudam a compreender melhor o impacto causado por ações que visam melhorar a mobilidade urbana. No entanto, a escassez de informações sobre o tráfego em grandes cidades dificultam sua execução com precisão. Existem dados de pesquisas que tentaram determinar as principais áreas de origem e destino das viagens realizadas em uma cidade. Entretanto, eles não oferecem informação sobre a distribuição do tráfego em função do tempo nem das rotas adotadas, por exemplo.

Para poder melhorar estudar os efeitos de comunicação interveicular em tempos de viagem e congestionamentos, o projeto visa a obtenção de dados de tráfego urbano a partir de textos em linguagem natural que descrevem, ao longo do tempo, o estado de vias de uma cidade e de locais específicos das mesmas. Uma fonte desses textos são aqueles similares a *tweets*, pois, são curtos e possuem estrutura simples. O jornal O Estado de São Paulo mantém um *blog* contendo informações de trânsito da cidade de São Paulo referenciadas em relação à data e ao horário. Elas contêm o estado atual - congestionado, liberado, por exemplo - de uma ou mais vias e usa outras para restringir esse estado para um trecho da primeira. Adquiriu-se, então, as publicações do *blog* entre agosto de 2010 e outubro de 2013.

Para reconhecer os padrões, categorias de palavras (entidades) foram definidas as quais são chamadas de: localização (LOC) – usadas para indicar nomes de logradouros – e *status* (STA) que indica o estado do fluxo do trânsito. Dos textos obtidos, mil foram classificados manualmente para permitir a evolução do projeto. A figura 1 ilustra uma classificação de um texto.

Considerando os padrões que compõe um texto, pôde-se detectar as seguintes informações de interesse a partir de entidades de localização: **logradouro principal** – uma entidade que define o logradouro referido pelo texto – e **restrição** a qual demarca uma região no logradouro principal.

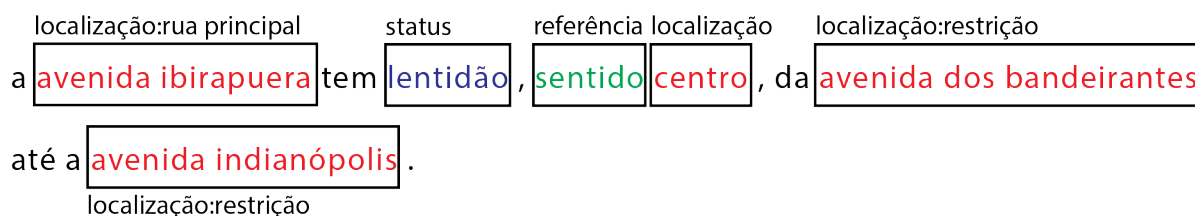


Figura 1: Exemplo de texto classificado

Cada informação é restrita a um trecho da rua principal, usando duas outras vias que a intersectam e associa um *status* ao mesmo. Com esses dados, pode-se replicar esse *status* no trecho e momento equivalente em uma simulação. No entanto, ainda é necessário transformar essa informação descrita em pontos em um mapa. Para resolver esse problema, foi criado um geocodificador; uma ferramenta capaz de determinar as coordenadas geográficas de um texto com um determinado formato, por exemplo, como um endereço.

O geocodificador criado, também, atende a necessidades essenciais para uma solução robusta. Ele compensa por variações dos nomes de ruas, abreviações e erros gramaticais. Ao receber duas ruas, ele as procura em sua base de dados e analisa geometricamente suas componentes a procura por intersecções, incluindo aquelas que não estão em nível, isto é, considera casos como o de um viaduto cruzando uma avenida.

As atividades desenvolvidas no projeto consistem em um primeiro passo no desenvolvimento de métodos que possibilitarão a obtenção de dados de trânsito atualmente de escassos. Isso é essencial para a realização de simulações com fidelidade.