

## IV Workshop de Dissertações de Mestrado do PPgSI (2017)

### FICHA DA PESQUISA<sup>i</sup>

DADOS GERAIS					
<b>Título do projeto de pesquisa</b>	Análise em tempo real de como os eventos da Cidade de São Paulo impactam o <i>Transporte Público</i>				
<b>Orientando</b>	Felipe Cordeiro Alves Dias				
<b>Orientador(es)</b>	Daniel de Angelis Cordeiro				
<b>Semestre no curso, na data do workshop</b>	<input type="checkbox"/> 2º semestre	<input checked="" type="checkbox"/> 3º semestre	<input type="checkbox"/> 4º semestre	<input type="checkbox"/> 5º semestre	<input type="checkbox"/> 6º semestre
<b>Qualificação</b>	<input type="checkbox"/> Qualificação já realizada em <input checked="" type="checkbox"/> Realização da qualificação planejada para: 20/10/2017				
<b>Defesa</b>	Prazo máximo para depósito: 29/01/2019 Realização da defesa planejada para: 08/06/2018				
<b>Linha e Área de pesquisa</b>	Gestão e desenvolvimento de Sistemas: <input type="checkbox"/> BD <input type="checkbox"/> Gestão de SI <input type="checkbox"/> Eng. de Software		<input type="checkbox"/> IHC	Inteligência de Sistemas: <input checked="" type="checkbox"/> IA de Padrões <input type="checkbox"/> Proc. Gráfico	
<b>Área de aplicação</b>	<input type="checkbox"/> Ambientes Corporativos <input type="checkbox"/> Bioinformática <input type="checkbox"/> Biometria <input type="checkbox"/> Economia	<input type="checkbox"/> Educação <input type="checkbox"/> Educação a Distância <input type="checkbox"/> Internet <input type="checkbox"/> Jogos	<input type="checkbox"/> Linguagem Natural <input type="checkbox"/> Linguística <input type="checkbox"/> Processos de Negócio <input type="checkbox"/> Química	<input checked="" type="checkbox"/> Redes Sociais <input type="checkbox"/> Robótica <input type="checkbox"/> Saúde <input type="checkbox"/> [outro – escrever]	
<b>Publicações associadas ao projeto de mestrado</b>					

DESCRIÇÃO DO PROJETO DE PESQUISA	
<b>Contextualização / motivação</b>	Hoje é possível utilizar as Redes Sociais na caracterização de problemas do transporte público. Apesar disso, não há estudos que por meio de <i>tweets</i> explorem as <i>features</i> relacionadas ao Transporte Público de São Paulo (SP).
<b>Problema de pesquisa</b>	Como extrair <i>features</i> de <i>tweets</i> provenientes de <i>profiles</i> oficiais sobre eventos de SP, capazes de contribuir com o melhor entendimento de como eventos da cidade impactam o Transporte Público de SP; abordando problemas como processamento em tempo real de grandes volumes de dados, geolocalização e confiabilidade.
<b>Objetivo geral da pesquisa</b>	Processar em tempo real grandes volumes de dados não estruturados, em Português Brasileiro, tendo o <i>Twitter</i> como fonte, visando extrair <i>features</i> relacionadas ao Transporte Público de São Paulo, contribuindo no melhor entendimento de como os eventos da cidade de São Paulo podem impactar a mobilidade urbana.
<b>Trabalhos relacionados</b>	<i>PairFac: Event Analytics through Discriminant Tensor Factorization</i> (WEN; LIN; PELECHRINIS, 2016), sobre impacto dos ataques terroristas em Paris no uso do Transporte Público. <i>Dynamic Cluster-Based Over-Demand Prediction in Bike Sharing Systems</i> (CHEN et al., 2016), sobre o impacto de eventos relacionados ao tráfego na demanda por bicicletas, em Nova Iorque e Washington D.C, USA. <i>Visual Exploration of Changes in Passenger Flows and Tweets on Mega-City Metro Network</i> (ITOH et al., 2016), sobre o impacto dos eventos anormais nas tomadas de decisão dos passageiros do Metrô de Tokyo. Todos limitam-se ao processamento em <i>batch</i> .
<b>Justificativa e relevância</b>	Com base numa Revisão Sistemática de 229 estudos primários, publicados entre 2011 e 2016, não foram encontrados trabalhos que utilizam <i>features</i> na caracterização de problemas relacionados ao Transporte Público de São Paulo. Abordar esse problema é relevante para melhor entendimento, em tempo real, dos eventos capazes de impactar a mobilidade urbana, auxiliando assim nas tomadas de decisão referentes ao planejamento, gestão e operações do Transporte Público de São Paulo.
<b>Proposta para Solução</b>	Uma abordagem para extração de localização de eventos por meio de Expressão Regular e obtenção dos dados de geolocalização, consultando a API do <i>Google Maps Geocoding</i> ; (ii) extração das demais <i>features</i> por meio de <i>PLN aplicado aos tweets</i> ; (iii) correlação com a base histórica da SPTRANS para análise do impacto dos eventos.
<b>Dados</b>	O projeto de pesquisa prevê a coleta de dados da conta oficial do Twitter da CET (dentre outras a serem definidos), além de dados históricos oriundos da SPTRANS e do aplicativo de mobilidade urbana chamado Coletivo.
<b>Validação</b>	A validação do proposta será feita através de comparações com os dados históricos da SPTRANS e do aplicativo de mobilidade urbana Coletivo, além disso, pretende-se avaliar a confiabilidade estatisticamente por meio da

	técnica <i>FI Score</i> .
<b>Limitações</b>	Processamento de dados apenas no idioma Português Brasileiro.
<b>Resultados esperados</b>	Contribuições científicas: <i>Features</i> relacionadas ao Transporte Público de SP; estudo de caso sobre como os eventos da cidade o impacta; base de dados contendo eventos relacionados ao Transporte Público de SP. Contribuições tecnológicas: API para ingestão de dados de fontes heterogêneas, capaz de manter a integridade do modelo utilizado, e consulta em tempo real.

<b>MÉTODO DE PESQUISA</b> [para cada item, marque todas as opções adequadas para seu projeto] (basear-se em: <a href="http://www.ufrgs.br/cursopgdr/downloadsSerie/derad005.pdf">http://www.ufrgs.br/cursopgdr/downloadsSerie/derad005.pdf</a> )			
<b>Gênero</b>	<input type="checkbox"/> Pesquisa teórica	<input checked="" type="checkbox"/> Pesquisa prática	<input type="checkbox"/> Pesquisa empírica <input type="checkbox"/> Pesquisa metodológica
<b>Natureza</b>	<input type="checkbox"/> Pesquisa básica/pura	<input checked="" type="checkbox"/> Pesquisa aplicada	
<b>Objetivo</b>	<input checked="" type="checkbox"/> Pesquisa exploratória	<input type="checkbox"/> Pesquisa descritiva	<input type="checkbox"/> Pesquisa explicativa <input type="checkbox"/> [outro – escrever]
<b>Abordagem</b>	<input checked="" type="checkbox"/> Pesquisa quantitativa	<input type="checkbox"/> Pesquisa qualitativa	<input type="checkbox"/> Pesquisa mista (quali-quant)
<b>Procedimento(s) técnico(s)</b>	<input type="checkbox"/> Pesquisa experimental <input type="checkbox"/> Pesquisa bibliográfica <input type="checkbox"/> Pesquisa documental <input type="checkbox"/> Pesquisa <i>ex-post-facto</i> <input type="checkbox"/> Pesquisa de levantamento	<input type="checkbox"/> Pesquisa com <i>survey</i> <input checked="" type="checkbox"/> Estudo de caso <input type="checkbox"/> Pesquisa participante <input type="checkbox"/> Pesquisa-ação <input type="checkbox"/> Pesquisa etnográfica	<input type="checkbox"/> Pesquisa netnográfica <input type="checkbox"/> Teoria fundamentada em dados ( <i>grounded theory</i> ) <input type="checkbox"/> Ciência do projeto ( <i>Design science research</i> ) <input type="checkbox"/> [outro – escrever]
<b>Fonte(s) de dados</b>	<input checked="" type="checkbox"/> pesquisa de laboratório	<input type="checkbox"/> pesquisa de campo	<input type="checkbox"/> pesquisa bibliográfica <input type="checkbox"/> [outro – escrever]
<b>Técnica(s) / Instrumento(s) de coleta de dados</b>	<input type="checkbox"/> medição <input type="checkbox"/> questionário <input type="checkbox"/> entrevista <input type="checkbox"/> grupos focais	<input type="checkbox"/> formulário <input type="checkbox"/> <i>benchmark</i>	<input type="checkbox"/> observação (direta / participante) <input type="checkbox"/> diário de campo / notas de campo <input type="checkbox"/> análise documental (ou de artefatos) <input checked="" type="checkbox"/> artefatos de software
<b>Técnica(s) de análise de dados</b>	Análise quantitativa: <input type="checkbox"/> Estatística descritiva <input checked="" type="checkbox"/> Estatística inferencial <input type="checkbox"/> [outro – escrever]	Análise qualitativa: <input type="checkbox"/> Análise de conteúdo <input type="checkbox"/> Análise do discurso <input type="checkbox"/> [outro – escrever]	

<b>CRONOGRAMA</b> [altere conforme necessário]																																					
	2016												2017												2018												
	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	
Estudo da literatura													X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Revisão sistemática													X	X	X	X	X	X	X																		
Definição da proposta								X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X													
Execução da proposta																				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Coleta/obtenção de dados														X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Análise dos dados													X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Validação da proposta													X	X	X	X	X	X	X	X	X	X	X	X													
Preparo da qualificação													X	X	X	X	X	X	X	X	X	X	X	X													
Exame de qualificação																									X												
Escrita da dissertação													X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Escrita de artigo																									X	X	X										
Depósito da dissertação																																					X
Crédito das disciplinas								X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X														
MVP													X	X	X	X	X	X																			

1 Esta ficha é uma adaptação da usada no “VIII Workshop de Teses e Dissertações em Sistemas de Informação (WTDSI 2015)” realizado como parte do “XI Simpósio Brasileiro de Sistemas de Informação (SBSI 2015)”