

II Workshop de Dissertações de Mestrado do PPgSI

FICHA DA PESQUISA¹

DADOS GERAIS	
Título do projeto de pesquisa	Análise de viés em notícias na língua portuguesa
Orientando	Gabriel Domingos de Arruda
Orientador(es)	Norton Trevisan Roman
Momento atual	<input type="checkbox"/> 3º semestre <input type="checkbox"/> 4º semestre <input checked="" type="checkbox"/> 5º semestre <input type="checkbox"/> 6º semestre
Qualificação	<input checked="" type="checkbox"/> Qualificação já realizada em: 02/03/2015 <input type="checkbox"/> Qualificação planejada para: ___/___/___
Defesa	Prazo máximo para depósito: 15/07/2016 Depósito planejado para: 20/10/2015
Linha e Área de pesquisa	<input type="checkbox"/> 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 <input checked="" type="checkbox"/> Inteligência de Sistemas: <input type="checkbox"/> IA <input checked="" type="checkbox"/> Rec. de Padrões <input type="checkbox"/> Proc. Gráfico
Área de aplicação	<input type="checkbox"/> Ambientes Corporativos <input type="checkbox"/> Educação <input checked="" type="checkbox"/> Linguagem Natural <input type="checkbox"/> Redes Sociais <input type="checkbox"/> Bioinformática <input type="checkbox"/> Educação a Distância <input type="checkbox"/> Linguística <input type="checkbox"/> Robótica <input type="checkbox"/> Biometria <input type="checkbox"/> Internet <input type="checkbox"/> Processos de Negócio <input type="checkbox"/> Saúde <input type="checkbox"/> Economia <input type="checkbox"/> Jogos <input type="checkbox"/> Química

DESCRIÇÃO DO PROJETO DE PESQUISA	
Contextualização / motivação	É comum redes sociais e buscadores fornecerem conteúdo personalizado que seja do interesse do usuário, apesar dessa abordagem ser interessante para anúncios por exemplo, pode ser um problema para o consumo de notícias ao prender as pessoas em uma “câmara de eco”.
Problema de pesquisa	A internet oferece acesso a uma grande variedade de provedores de notícias, mas se faz necessário ferramentas que auxiliem os leitores a escolherem provedores que tenham linhas editoriais diversas.
Objetivo geral da pesquisa	Desenvolver um método para análise de viés de produtores de notícias em relação à determinadas entidades que seja sistemático e automatizável.
Trabalhos relacionados	“Social media news communities”: o trabalho analisava o viés a partir de produtores de notícias e comunidades do Twitter, ou seja, que entidades estão sendo abordadas, o quanto estão sendo abordadas e se a abordagem é positiva ou negativa. “Spatial analysis of news source”: produtores de notícia dos EUA foram monitorados e a diferença de destaque para algumas entidades é analisado em diversas regiões do país. “Sentiment bias detection in support of news credibility judgment”: utilizando análise de sentimentos, o autor apresenta um gráfico em linha do sentimento de diferentes jornais em relação a entidades.
Justificativa e relevância	O objetivo do nosso projeto é abordar a relação de diferentes meios de comunicação em relação a entidades políticas, fornecer uma visão comparativa como “Sentiment bias detection in support of news credibility judgment” mas utilizando uma abordagem multi-dimensional e uma metodologia clara para definição de viés.
Proposta para Solução	No projeto foi construído um corpus de notícias segmentado por parágrafo e classificados manualmente por anotadores voluntários. Utilizando esse corpus, técnicas de análise de sentimentos foram avaliadas com o objetivo de identificar a abordagem mais adequada para a língua portuguesa. Por fim, uma metodologia para identificação de viés baseada no conceito outliers foi proposta e analisada a partir do corpus extraído.
Dados	As notícias utilizadas no corpus foram extraídas dos perfis do Twitter de cinco produtores de notícias distintos e, após a coleta, esses dados foram anotados por voluntários.
Forma de validação	Os resultados de anotação do corpus foi avaliado com base em métricas de concordância entre anotadores, os resultados das técnicas de análise de sentimentos foi avaliado com métricas de desempenho de algoritmos de classificação e o viés dos meios de comunicação a partir da aplicação da metodologia proposta sobre o corpus.
Limitações	O volume de notícias extraídos seguindo a metodologia não pode ser realizada por um grande período de tempo, tornando-se um recorte reduzido do problema.

Resultados esperados	Contribuições científicas: técnicas de análise de sentimentos serão avaliados em língua portuguesa no domínio de notícias, um corpus será disponibilizado para uso em outros trabalhos de análise de sentimentos e, por fim, será proposta uma metodologia para análise de viés automática.
-----------------------------	---

MÉTODO DE PESQUISA			
Gênero	<input type="checkbox"/> Pesquisa teórica	<input type="checkbox"/> Pesquisa prática	<input type="checkbox"/> Pesquisa empírica <input checked="" type="checkbox"/> Pesquisa metodológica
Natureza	<input type="checkbox"/> Pesquisa básica/pura	<input checked="" type="checkbox"/> Pesquisa aplicada	
Objetivo	<input type="checkbox"/> Pesquisa descritiva	<input checked="" type="checkbox"/> Pesquisa exploratória	<input type="checkbox"/> Pesquisa explicativa
Abordagem	<input checked="" type="checkbox"/> Pesquisa quantitativa	<input type="checkbox"/> Pesquisa qualitativa	<input type="checkbox"/> Pesquisa mista (quali-quant)
Procedimento(s) técnico(s)	<input checked="" type="checkbox"/> Pesquisa experimental <input checked="" 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 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>)
Fonte(s) de dados	<input type="checkbox"/> pesquisa de laboratório	<input checked="" type="checkbox"/> pesquisa de campo	<input checked="" type="checkbox"/> pesquisa bibliográfica
Técnica(s) / Instrumento(s) de coleta de dados	<input checked="" 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)
Técnica(s) de análise de dados	<input checked="" type="checkbox"/> Análise quantitativa: <input checked="" type="checkbox"/> Estatística descritiva <input checked="" type="checkbox"/> Estatística inferencial		<input type="checkbox"/> Análise qualitativa: <input type="checkbox"/> Análise de conteúdo <input type="checkbox"/> Análise do discurso

CRONOGRAMA																																						
	2014												2015												2016													
	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																																						
Revisão sistemática	X	X	X	X	X	X																																
Execução da proposta				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																							
Análise dos dados													X	X	X	X	X	X	X	X	X																	
Validação da proposta						X	X	X	X	X	X																											
Preparo da qualificação						X	X	X	X	X	X																											
Exame de qualificação															X																							
Escrita da dissertação																		X	X	X	X	X	X															
Escrita de artigo						X	X				X	X	X																									
Depósito da dissertação																							X															
(Outros)																																						

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)”