

IV Workshop de Dissertações de Mestrado do PPgSI (2017)

FICHA DA PESQUISAⁱ

DADOS GERAIS				
Título do projeto de pesquisa	Inferência das áreas de atuação de pesquisadores			
Orientando	Felipe Penhorate Carvalho da Fonseca			
Orientador(es)	Luciano Antonio Digiampietri			
Semestre no curso, na data do workshop	<input type="checkbox"/> 2º semestre	<input type="checkbox"/> 3º semestre	<input checked="" type="checkbox"/> 4º semestre	<input type="checkbox"/> 5º semestre
Qualificação	<input checked="" type="checkbox"/> Qualificação já realizada em: 13/03/2017 <input type="checkbox"/> Realização da qualificação planejada para: dd/mm/aaaa			
Defesa	Prazo máximo para depósito: 19/08/2018		Realização da defesa planejada para: 30/01/2018	
Linha e Área de pesquisa	Gestão e desenvolvimento de Sistemas: <input type="checkbox"/> BD <input type="checkbox"/> Gestão de SI <input type="checkbox"/> Eng. de Software		Inteligência de Sistemas: <input checked="" type="checkbox"/> IA de Padrões <input type="checkbox"/> Proc. Gráfico	
Área de aplicação	<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 checked="" 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]
Publicações associadas ao projeto de mestrado	Publicada: Análise da relação entre obtenção de bolsas de produtividade do CNPq e medidas bibliométricas e de análise de redes sociais. Conferência: V Brazilian Workshop on Social Network Analysis and Mining (BraSNAM 2016). Em Elaboração: Uso do LDA para representação de textos por meio de bases de dados externa. Revista: Revista Iberoamericana de Inteligência Artificial			

DESCRIÇÃO DO PROJETO DE PESQUISA

Contextualização / motivação	O contexto do projeto envolve expandir e melhorar um trabalho feito anteriormente pelo orientador do aluno, de forma a se fazer um estudo mais abrangente sobre como fazer inferência das áreas de atuação de pesquisadores.
Problema de pesquisa	O problema de pesquisa está relacionado em encontrar, estudar e avaliar diversas formas de inferência das áreas de atuação de pesquisadores.
Objetivo geral da pesquisa	O objetivo geral deste é o desenvolvimento de uma estratégia para a inferência das áreas de atuação de pesquisadores com base em seus Currículos Lattes usando um gold standard de bolsistas produtividade.
Trabalhos relacionados	Vo e Ock (2005) propõem uma forma de fazer enriquecimento de texto, por meio do LDA, para auxiliar na classificação de títulos de artigos científicos, sendo esta a abordagem testada no projeto para melhor a inferência das áreas de atuação. Gabrilovich e Markovitch (2007) introduzem o ESA, uma forma de representação de texto usando somente conceitos do Wikipédia, sendo esta utilizada como uma das alternativas para representação ao TFIDF e, por fim, Miyata, Kano e Digiampietri (2013) realizaram a classificação automática das áreas de atuação dos pesquisadores que estão presentes na Plataforma Lattes, sendo está o baseline do projeto de pesquisa.
Justificativa e relevância	Diversos trabalhos de análise bibliométrica ou análise de grupos de pesquisa, bem como trabalhos de identificação de especialistas necessitam (ou se beneficiariam) de contar com informações sobre a provável área de atuação de pesquisadores. Por isso estudar técnicas de inferência das áreas de atuação é importante. Além disso, as soluções atuais se utilizam de abordagens consideravelmente simples para fazer a inferência, sendo possível melhorar o desempenho por meio da aplicação de abordagens como enriquecimento do texto, por exemplo.
Proposta para Solução	A solução proposta utiliza de uma base externa de dados como representativa dos diferentes domínios/áreas e utiliza de LDA para a representação dos documentos (no caso, os títulos das publicações) para, com base nessa representação, realizar a inferência das áreas de atuação utilizando algoritmos de inteligência artificial.
Dados	Os dados utilizados no projeto de pesquisa são os currículos da plataforma Lattes de (que já foram obtido anteriormente em outra pesquisa) e um conjunto de dados externo, que foi construído usando páginas da Wikipédia relacionadas as áreas de atuação presentes na taxonomia de Grandes Áreas, Áreas e Subáreas do CNPq.
Validação	A avaliação dos resultados obtidos no projeto de pesquisa está sendo feita por testes estatísticos em cima da acurácia das abordagens testadas, comparando-as de forma a se delinear qual abordagem tem o melhor

