

V Workshop de Dissertações de Mestrado do PPgSI (2018)

FICHA DE PESQUISA

DADOS GERAIS				
Título do projeto de pesquisa	Biagrupamento e coagrupamento aplicado a dados textuais			
Orientando	Alexandra Katiuska Ramos Diaz			
Orientador(es)	Profa. Dra. Sarajane Marques Peres			
Semestre no curso, na data do workshop	<input type="checkbox"/> 2º semestre	<input type="checkbox"/> 3º semestre	<input type="checkbox"/> 4º semestre	<input checked="" type="checkbox"/> 5º semestre
Qualificação	<input type="checkbox"/> Qualificação já realizada em: 07/07/2018		<input type="checkbox"/> Realização da qualificação planejada para: dd/mm/aaaa	
Defesa	Prazo máximo para depósito: 29/08/2018		Realização da defesa planejada para: 10/10/2018	
Linha e Área de pesquisa	Gestão e Desenvolvimento de Sistemas: <input type="checkbox"/> BD <input type="checkbox"/> Engenharia de Software <input type="checkbox"/> Gestão de TI <input type="checkbox"/> IHC		Inteligência de Sistemas: <input checked="" type="checkbox"/> IA <input type="checkbox"/> Processamento Gráfico <input type="checkbox"/> Reconhecimento de Padrões	
Área de aplicação	<input type="checkbox"/> Ambientes corporativos <input type="checkbox"/> Bioinformática <input type="checkbox"/> Biometria <input type="checkbox"/> Dispositivos móveis <input type="checkbox"/> Educação	<input type="checkbox"/> Educação a distância <input type="checkbox"/> Governo eletrônico <input checked="" type="checkbox"/> Internet <input type="checkbox"/> Jogos <input type="checkbox"/> Jogos sérios	<input type="checkbox"/> Língua Natural <input type="checkbox"/> Linguística <input type="checkbox"/> Processos de Negócio <input type="checkbox"/> Químioinformática	<input type="checkbox"/> Redes Sociais <input type="checkbox"/> Robótica <input type="checkbox"/> Saúde <input type="checkbox"/> [outro – escrever]
Publicações associadas ao projeto de mestrado	-Diaz, A.K.R. et. al. Uma análise comparativa das ferramentas de pré-processamento de dados textuais: NLTK, PreText e R. Relatório Técnico (PPgSI 001/2018) - publicado -Biagrupamento e coagrupamento: conceitos, algoritmos e viabilidade para mineração de texto: Parte I (Revista de Informática Teórica e Aplicada - RITA) - em elaboração -Biagrupamento e coagrupamento: conceitos, algoritmos e viabilidade para mineração de texto: Parte II (Revista de Informática Teórica e Aplicada - RITA) - em elaboração			

DESCRIÇÃO DO PROJETO DE PESQUISA	
Contextualização / motivação	Esta pesquisa está inserida no contexto de um grupo que têm interesse em aplicações de coagrupamento para dados textuais. A representação computacional para dados textuais é geralmente uma representação esparsa que dificulta trabalho dos algoritmos de coagrupamento e biagrupamento. Assim o contexto deste trabalho é explorar tais dificuldades.
Problema de pesquisa	A qualidade da análise de textos, fornecida na resolução de tarefas de mineração de textos, pode ser enriquecida com o uso de coagrupamento e biagrupamento, já que tal estratégia permite a extração da informação com base em similaridades parciais. Contudo, a resolução do problema de coagrupamento é de alta complexidade computacional, difícil de ser resolvido em sua otimalidade. Ainda a característica da matriz esparsa proveniente da representação dos dados textuais, agrava a dificuldade do problema. O problema tratado nesta pesquisa é a análise de dois algoritmos pertencentes ao coagrupamento e biagrupamento, capazes de fornecer soluções sub-ótimas para o problema em tempo aceitável, e que são adequados para a realização de soluções em espaços caracterizados pela esparsidade dos dados.
Objetivo geral da pesquisa	Prover um arcabouço da análise do comportamento do algoritmo para biagrupamento Cheng e Church e do algoritmo de coagrupamento NBVD, aplicado ao contexto de dados textuais.
Trabalhos relacionados	Biclustering Algorithms for Biological Data Analysis: A Survey: analisa diferentes estruturas ou organizações de bigrupos e vários algoritmo de biagrupamento. Orthogonal nonnegative matrix tri-factorization for co-clustering: Multiplicative updates on Stiefel manifolds: Referencial teórico dos algoritmos de coagrupamento. Biclustering of Expression Data e Co-clustering by block value decomposition: trabalhos que apresentam os algoritmos usados de biagrupamento e coagrupamento.
Proposta para solução	Foi escolhido o algoritmo de biagrupamento Cheng e Church e o algoritmo de coagrupamento NBVD. Ambos foram estudados e analisados em ambientes sintéticos sem/com esparsidade, e aplicados no contexto de dados textuais, com uso de representações binárias, tf e tf-idf.

