

II Workshop de Dissertações de Mestrado do PPgSI

FICHA DA PESQUISAⁱ

DADOS GERAIS	
Título do projeto de pesquisa	Desenvolvimento de técnica para recomendar atividades em workflows científicos: uma abordagem baseada em ontologias
Orientando	Adilson Lopes Khouri
Orientador(es)	Luciano Antonio Digiampietri
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: 11/12/2014 <input type="checkbox"/> Qualificação planejada para: ____/____/____
Defesa	Prazo máximo para depósito: 05/07/2016 Depósito planejado para: 05/12/2015
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 checked="" type="checkbox"/> IA <input type="checkbox"/> Rec. de Padrões <input type="checkbox"/> Proc. Gráfico
Área de aplicação	<input type="checkbox"/> Ambientes Corporativos <input type="checkbox"/> Educação <input type="checkbox"/> Linguagem Natural <input type="checkbox"/> Redes Sociais
	<input checked="" 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 <input type="checkbox"/> [outro – escrever]

DESCRIÇÃO DO PROJETO DE PESQUISA	
Contextualização / motivação	Este projeto de mestrado é uma extensão/complemento do projeto de doutorado do meu orientador. Durante a construção de experimentos científicos, especialmente no formato de workflows, o cientista precisa combinar diferentes atividades. Esta combinação é muitas vezes complexa, pois exige que o cientista conheça um grande número de atividades. Sistemas de recomendação de atividades visam a facilitar esta atividade, sugerindo aos cientistas uma lista de atividades potencialmente úteis para completar o workflow em construção.
Problema de pesquisa	<i>Melhorar o desempenho de sistemas de recomendação de atividades para workflows científicos utilizando ontologias de domínio.</i>
Objetivo geral da pesquisa	<i>Desenvolver uma técnica para recomendar atividades em workflows científicos que utilize informações sobre frequência de atividades, entrada/saída e ontologias de domínio. Estabelecer um processo padronizado de testes para técnicas de recomendação. Dessa forma, espera-se obter um resultado superior a literatura correlata.</i>
Trabalhos relacionados	<i>A literatura correlata propõe três grandes estratégias para recomendar atividades. A combinação entre entrada/saída de atividades sua principal deficiência é desconsiderar a semântica dos dados (ZHANG, 2006). O uso da proveniência de informação, usando o log de modelagem para recomendar atividades, necessita de proveniência dos dados (KOOP et al., 2008). Por fim, a frequência de ocorrência de pares atividades, sua principal deficiência é desconsiderar uso de informação semântica (OLIVEIRA, 2010).</i>
Justificativa e relevância	<i>Auxiliar cientistas a construir experimentos mais rapidamente filtrando o grande número de atividades, da ordem de milhares, por meio de uma técnica mais precisa de recomendação. A literatura correlata é escassa sobre métodos para validar diferentes técnicas de recomendação de atividades.</i>
Proposta para Solução	<i>Desenvolver uma técnica para recomendar atividades em workflows científicos que utilize informações sobre frequência de atividades, entrada/saída e ontologias de domínio. Além dessa nova técnica é proposta uma nova modelagem para o problema de recomendação de atividades como um problema de classificação e regressão. Por fim, todas as técnicas serão comparadas com um mesmo conjunto de dados e com as mesmas métricas de avaliação.</i>
Dados	<i>Dados obtidos no repositório myExperiment (http://www.myexperiment.org/home)</i>
Forma de validação	<i>10-fold cross validation utilizando as métricas $S@K$ (probabilidade do item desejado estar entre os K primeiros na lista de recomendações) e MRR (posição média do item desejado na lista de recomendações) para selecionar os melhores parâmetros de cada técnica.</i>
Limitações	<i>A solução proposta não considera o uso de proveniência de informação (execução e modelagem) para recomendar atividades.</i>
Resultados esperados	<i>Contribuições científicas esperadas: nova técnica para recomendar atividades mais precisa que a literatura correlata, metodologia para testes de diferentes sistemas de recomendação para um mesmo conjunto de dados. Nova modelagem do problema de recomendação como um problema de classificação e regressão.</i>
	<i>Contribuições tecnológicas: o sistema de recomendação implementado (propriamente dito).</i>

Í 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)”