

# I - Workshop de Dissertações do Programa de Pós-graduação em Sistemas de Informação (PPgSI)

## Desenvolvimento de técnica para recomendar atividades em workflows científicos: uma abordagem baseada em ontologias

**Aluno:** Adilson Khouri NUSP 6411994, **Orientador:** Dr. Luciano Digiampietri

### Contexto

*Workflows* científicos auxiliam cientistas a criar experimentos com blocos de código ou serviços representados graficamente (atividades) permitindo construir experimentos arrastando ícones. O grande número de atividades exige que os cientistas conheçam muitas atividades de forma a conseguirem efetivamente reaproveitar as existentes. Para reduzir esta exigência foram propostas técnicas para recomendação.

### Objetivo

Este mestrado visa a desenvolver uma técnica de recomendação de atividades em *workflows* científicos baseada em ontologias; organizar uma base de workflows para comparação de soluções; e comparar a técnica proposta com outras presentes na literatura correlata.

### Método

A técnica proposta está sendo desenvolvida incrementalmente combinando conceitos de técnicas simples (baseadas em *itemsets* e frequência) com o uso de ontologias. Adicionalmente algoritmos de agrupamento e redução de dimensionalidade dos dados estão sendo testados para verificar a capacidade de melhoria dos resultados. A técnica proposta será comparada com a literatura correlata utilizando as métricas  $S@K$ ,  $MRR$  e número de acertos para verificar qual apresenta melhor desempenho. Foi organizada uma base de workflows para testes e validação utilizando *workflows* reais obtidos do repositório *myExperiment* (<http://www.myexperiment.org/>).

### Resultados

Foram implementados os experimentos iniciais (utilizando recomendadores simples), destes o melhor resultado foi do experimento baseado em frequência, com acertos de 55%,  $MRR = 5,05$ ,  $S@10 = 0,47$  e  $S@5 = 0,43$ . Para tratar a esparsidade dos dados foram implementados os algoritmos: AGNES e ROCK para agrupar o conjunto de dados, após o agrupamento observa-se uma melhoria nas métricas  $MRR = 1,14$ ,  $S@10 = 0,44$  e  $S@5 = 0,43$ .

## **Conclusões**

Pode-se concluir que ao acrescentar mais critérios para recomendar, como ordem das atividades, a recomendação melhora, espera-se que ao acrescentar semântica em conjunto com a ordem das atividades seja obtida uma nova técnica mais eficiente do que as correlatas.