

Interactive Trace Clustering

Agrupamento interativo aplicado à mineração de processos

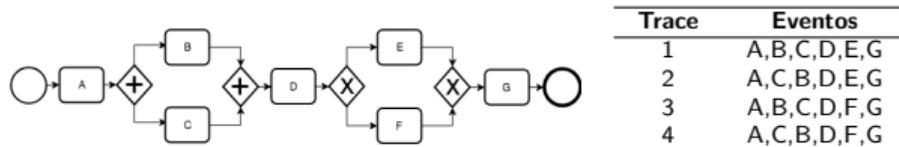
Thais Rodrigues Neubauer

Orientadora: Profa. Dra. Sarajane Marques Peres

Co-orientador: Prof. Dr. Marcelo Fantinato

Agosto de 2019

- Mineração de processos: contextualização
- Problema de pesquisa, objetivos e hipótese
- Agrupamento interativo
- Método de pesquisa
- Atividades realizadas - experimentos
- Andamento do projeto



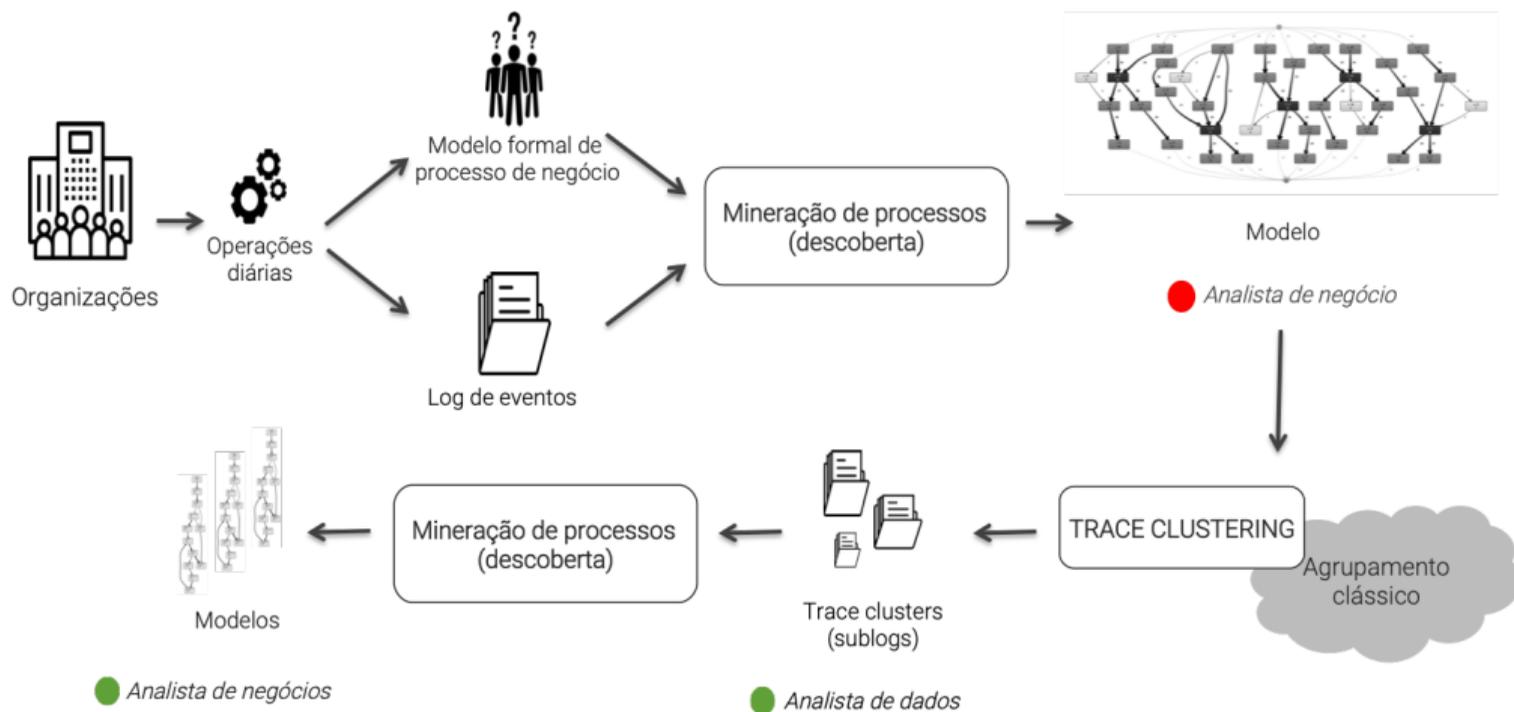
Modelo de processo

- Conjunto de atividades e restrições de execução.
- Conhecer o processo real de operações diárias → sucesso na gestão de negócios.
- Comumente não formalizado por razões culturais ou falta de recursos humanos ou materiais adequados.

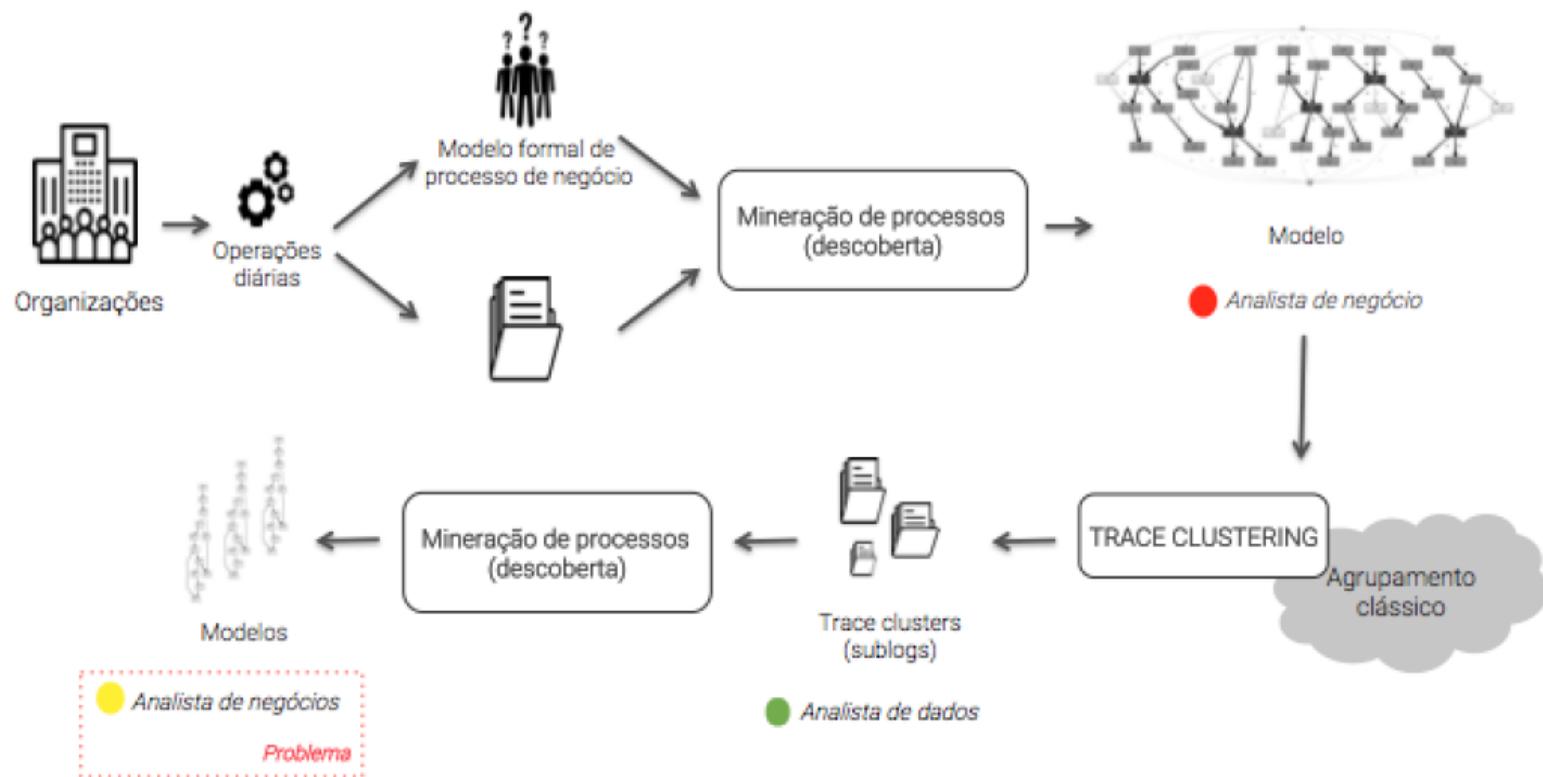
Mineração de processos

- Extração de conhecimento a partir de logs de eventos.
- Processos organizacionais automaticamente descobertos, analisados e aprimorados.
- Usa conhecimentos de *Mineração de Dados*, especialmente técnicas de **agrupamento** - *Trace clustering*.

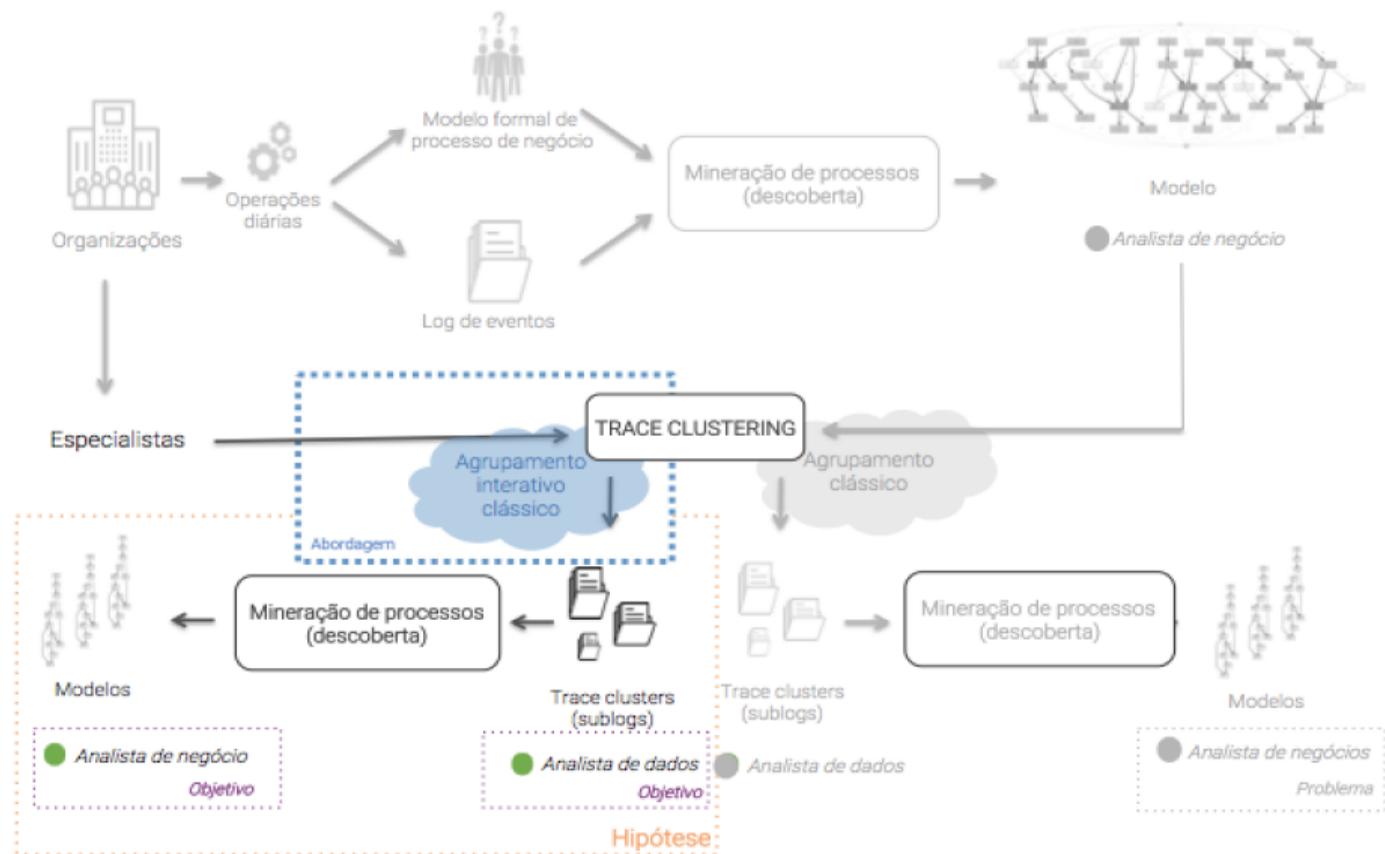
Problema de pesquisa e proposta



Problema de pesquisa e proposta

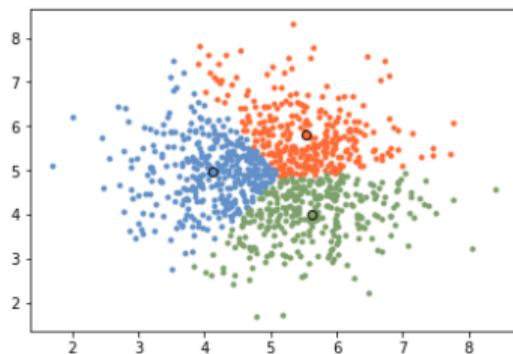


Problema de pesquisa e proposta

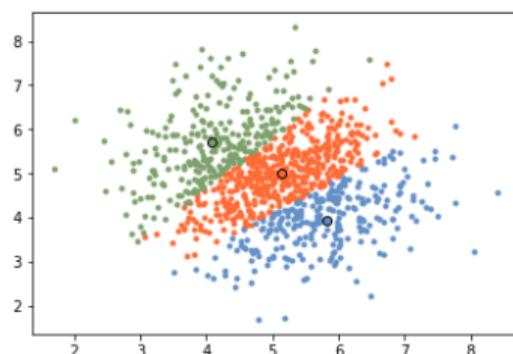


Agrupamento interativo

- Agrupamento: identificação de padrões (sem informação prévia sobre eles) com base em *similaridade*.
- A função de similaridade e/ou a representação escolhidas podem levar a ambiguidade e/ou inadequação de resultados em relação à expectativa humana.
- Sem supervisão, usualmente resolvidos por meio de tomadas de decisão e suposições estatísticas.



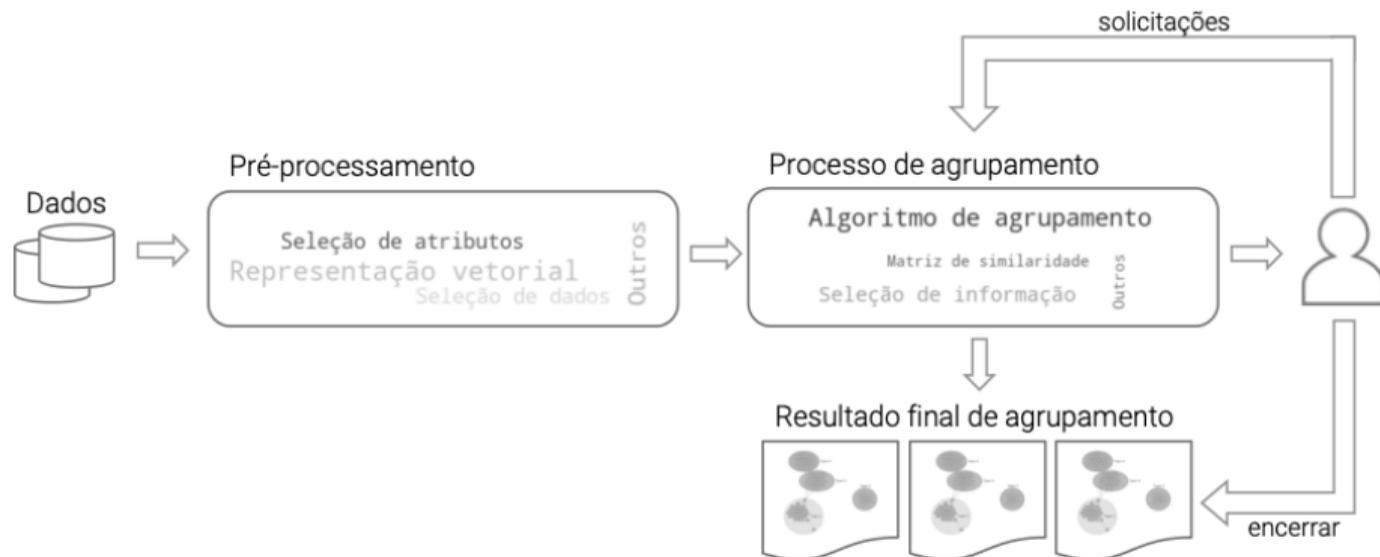
(c) Distância euclidiana



(d) Similaridade cosseno

Agrupamento interativo

- Além da função de similaridade, outras decisões também são tomadas, como a representação vetorial e o algoritmo de agrupamento.
- Ideia central do **agrupamento interativo**: substituição ou complementação de suposições e decisões pelo conhecimento de especialista.



- Natureza: aplicada.
- Objetivo metodológico: pesquisa explicativa.
 - Pesquisa na qual a preocupação está em identificar fatores que determinam ou contribuem para a ocorrência de fenômenos.
 - Fenômeno: melhoria da qualidade de resultados com a inserção de especialistas no processo de agrupamento de *traces*.
- Estudo empírico – conclusões baseadas em análise de resultados obtidos em experimentos.
 - Procedimentos de revisão exploratória da literatura e revisão sistemática da literatura também estão sendo adotados.

Experimentos

- Aplicação de *trace clustering* e *interactive trace clustering* em log de eventos sintéticos e em log de eventos reais;
 - Algoritmo k-means e suas variações - mais popular.
 - Abordagens interativas *must-link/cannot-link* e *split-merge*.
- Aplicação de soluções para problemas de mineração de processos nos logs originais e reduzidos por meio de *trace clustering*;

Avaliação

- Agrupamento: índices de validação internos e externos - *Silhouette* e *Ajusted Rand*.
- Qualidade de processos: medidas de aspectos de qualidade dos modelos de processo descobertos - *Completeness*, *Precisão*, *Simplicidade* e *Generalização*.
- Tarefa de predição de tempo de resolução de incidentes: estatística descritiva.



Apromore

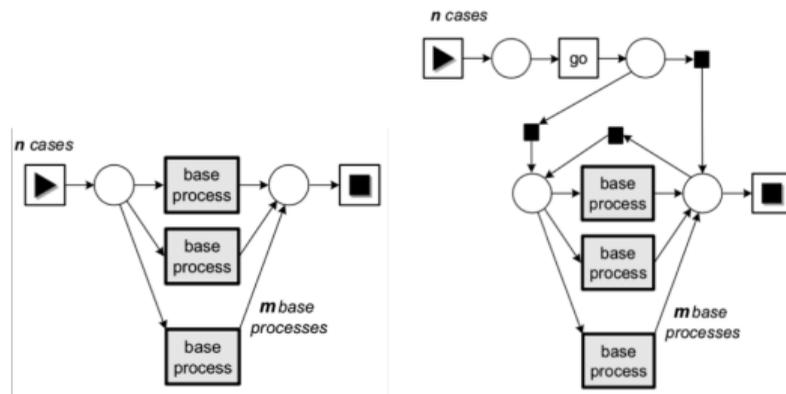
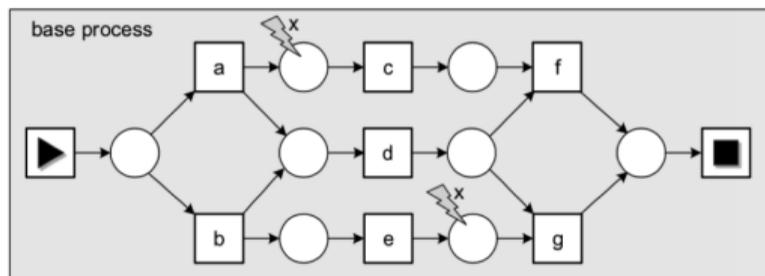


Disco

Experimentos - log de eventos sintéticos

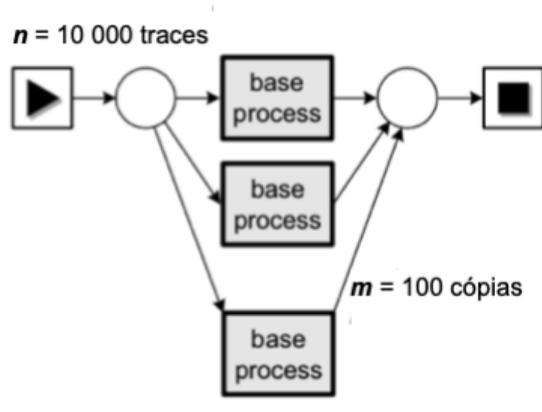
Objetivo do experimento

Ambiente de teste independente de contexto → ambiente controlado, abstração de questões semânticas.



Baseado no conjunto "Benchmarking logs to test scalability of process discovery algorithm" (4TU.ResearchData, W. M. P. Van der Aalst, 2017)

Experimentos - log de eventos sintéticos



100 GRUPOS

$K = 100$

$x[m]$ $a[m]$ $b[m]$
 $d[m]$ $d[m]$ $e[m]$
 $f[m]$ $g[m]$

Expectativa
do especialista

f_{17}
 d_{17} 17 17
 e_{17} x_{17} ∞ b_{17} c_{17}
 a_{17}

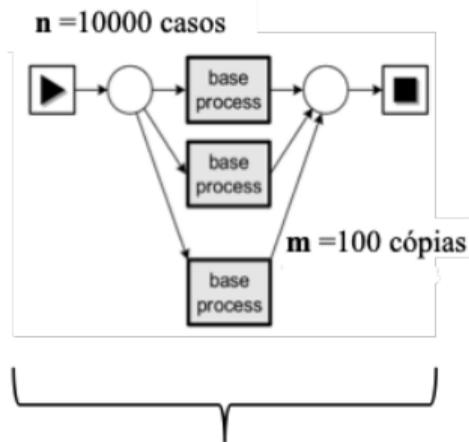
g_{30} b_{30} d_{30} e_{30}
 x_{30}

c_{30} d_{30}
 x_{30} a_{30} f_{30}

b_{37} a_{51} e_{48} c_{31} g_{51} c_{37} e_{37}
 a_{48} f_{10} d_{37} d_{10}
 c_{51} d_{37} e_{10} g_{10} g_{37} f_{31} e_{51}
 x_{10} b_{51} d_{51} d_{31} a_{37} b_{48}
 a_{31} d_{51} c_{10} b_{31} g_{48} b_{10} f_{37} f_{51} a_{37} a_{10}
 f_{48} c_{48}

Resultados
alcançados

Experimentos - log de eventos sintéticos

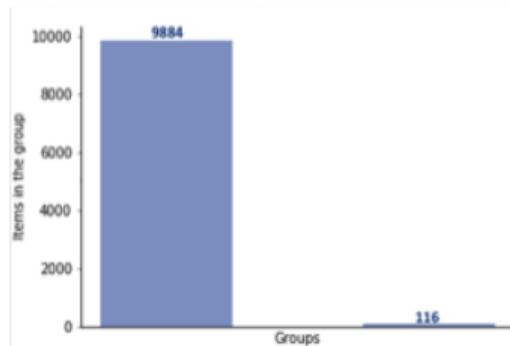


2 GRUPOS
K = 2

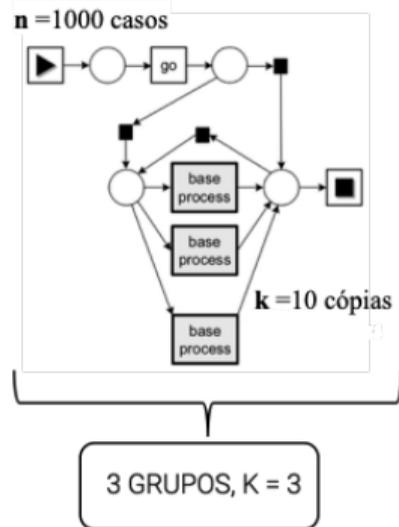
Resultado esperado pelo especialista



Resultado alcançado - Silh: 0.0102



Experimentos - log de eventos sintéticos



Resultado esperado pelo especialista



Resultado alcançado - Silh: 0.27



Experimento - log de eventos reais

O log de eventos reais

- Relata eventos de um processo de gerenciamento de incidentes extraído de uma instância da plataforma *ServiceNow*TM utilizada por uma empresa de TI.
- Composto por 24.918 incidentes, extraídos de 03/2016 – 02/2017.
- **Tarefa de mineração de processos:** Predição de tempo de resolução de incidentes.

	number	incident_state	active	reassignment_count	...	close_code	resolved_by	resolved_at	closed_at
0	INC0000045	New	True	0	...	code 5a	Resolved by 149	29/2/2016 11:29	2016-03-05 12:00:00
1	INC0000045	Resolved	True	0	...	code 5a	Resolved by 149	29/2/2016 11:29	2016-03-05 12:00:00
2	INC0000045	Resolved	True	0	...	code 5a	Resolved by 149	29/2/2016 11:29	2016-03-05 12:00:00

Experimentos - log de eventos reais

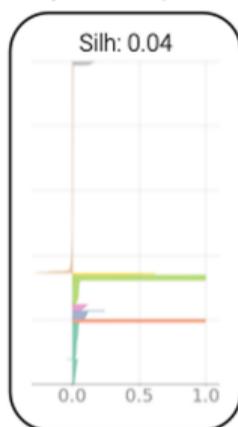
Objetivo - trace clustering

Obter grupos com menor variação de tempo de resolução de incidentes (maior precisão).

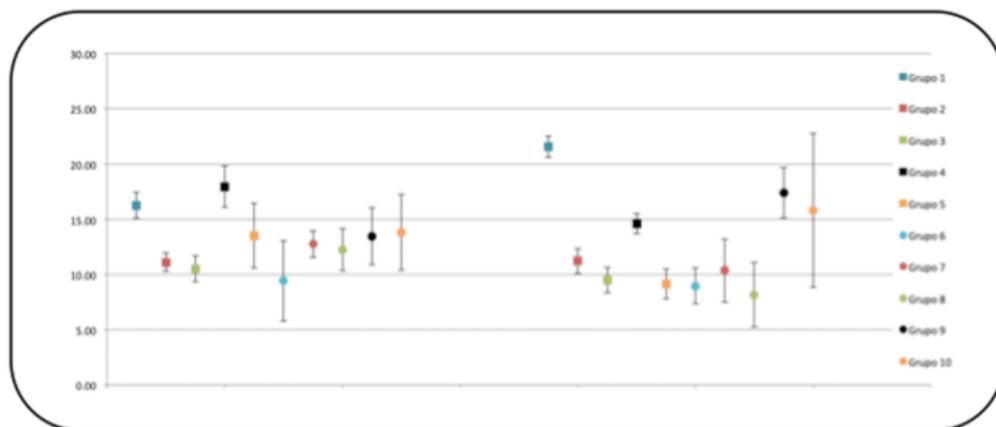
Seleção do algoritmo



Seleção do especialista



Seleção do algoritmo



- Revisão da literatura.
- Estudo dos logs a serem utilizados.
- Definição da proposta de pesquisa.
- RS sobre agrupamento interativo.
- Mapeamento dos logs para vetores.

- Experimentos log de eventos sintéticos.
 - Experimentação inicial.
 - Design dos experimentos junto aos especialistas.
 - Execução dos experimentos [Ago/2019].
 - Avaliação dos experimentos [Ago/2019].

- Exame de qualificação.
- Divulgação - publicações.
 - WTDSI - SBSI 2019.
 - Artif. Intellig. Review, Springer [Ago/2019].
- Escrita e entrega da dissertação [Nov/2019].

- Experimentos log de eventos reais.
 - Experimentação inicial.
 - Design dos experimentos junto aos especialistas [Ago/2019].
 - Execução dos experimentos [Set/2019].
 - Avaliação dos experimentos [Set/2019].

- Revisão da literatura.
- Estudo dos logs a serem utilizados.
- Definição da proposta de pesquisa.
- RS sobre agrupamento interativo.
- Mapeamento dos logs para vetores.

- Experimentos log de eventos sintéticos.
 - Experimentação inicial.
 - Design dos experimentos junto aos especialistas.
 - Execução dos experimentos [Ago/2019].
 - Avaliação dos experimentos [Ago/2019].

- Exame de qualificação.
- Divulgação - publicações.
 - WTDSI - SBSI 2019.
 - Artif. Intellig. Review, Springer [Ago/2019].
- Escrita e entrega da dissertação [Nov/2019].

- Experimentos log de eventos reais.
 - Experimentação inicial.
 - Design dos experimentos junto aos especialistas [Ago/2019].
 - Execução dos experimentos [Set/2019].
 - Avaliação dos experimentos [Set/2019].

*O presente trabalho foi realizado com apoio da
Coordenação de Aperfeiçoamento de Pessoal de Nível
Superior (CAPES) - Código de Financiamento 001.*



Interactive Trace Clustering

Agrupamento interativo aplicado à mineração de processos

Thais Rodrigues Neubauer

Orientadora: Profa. Dra. Sarajane Marques Peres

Co-orientador: Prof. Dr. Marcelo Fantinato

Agosto de 2019



VI Workshop de Dissertações do PPgSI

<http://ppgsi.each.usp.br/>