

**RELATÓRIO PARCIAL**

**Desenvolvimento de Aplicação para Pré-processamento de Log de Eventos na área da Saúde**

#### Anderson Fuzino

##### Ciência da Computação

**Desenvolvimento de Aplicação para Pré-Processamento de Log de Eventos na Área da Saúde**

|  |  |
| --- | --- |
|  | Relatório Parcial apresentado à Pontifícia Universidade Católica do Paraná, Pró-Reitoria de Pesquisa, Pós-Graduação e Inovação, Programa Institucional de Bolsas de Iniciação Científica, sob orientação do:  Prof. Edson Emilio Scalabrin e  Msc. Luiz Southier – colaborador (doutorando) |

**Curitiba  
2021**

**SUMÁRIO**

[1 INTRODUÇÃO 4](#_heading=h.gjdgxs)

[2 OBJETIVO(S) 4](#_heading=h.30j0zll)

[3 MATERIAIS E MÉTODOS 5](#_heading=h.1fob9te)

[4 RESULTADOS PARCIAIS 5](#_heading=h.3znysh7)

[5 ETAPAS FUTURAS 5](#_heading=h.2et92p0)

[6 REFERÊNCIAS 6](#_heading=h.tyjcwt)

# 1 INTRODUÇÃO

Atualmente, observa-se que há uma pressão crescente sobre as organizações de saúde. Elas precisam melhorar a produtividade e reduzir a entrada no sistema e os tempos de espera; o que deve diminuir tempo e custos. Uma abordagem é focar nos processos de longo termo e não triviais que são realizados, por exemplo, no interior de hospitais (e.g., preparação e execução de uma cirurgia, tratamento de pacientes com câncer). A impressão de qualidade e satisfação de um paciente está diretamente ligada ao seu tempo de espera e permanência no percurso de tratamento. Portanto, para dar sugestões objetivas para melhorar e redesenhar os processos, é necessário explorar os dados do evento prontamente disponíveis. Tal análise é possível usando mineração de processo (AALST, 2016).

A mineração de processos foi aplicada com sucesso em diversas áreas, incluindo serviços bancários, produção, educação, diagnósticos inteligentes etc. (GARCIA et al., 2019). A mineração de processos relaciona o comportamento real de pessoas, máquinas e organizações ao comportamento modelado (AALST; WEIJTERS, 2004). Tal relacionamento muitas vezes leva a descobertas surpreendentes colocando em evidência que a realidade é muito diferente das percepções, opiniões e crenças das partes interessadas.

Para isso, a mineração de processos utiliza um log de eventos que é um arquivo que registra informações sobre os eventos do processo, tais como, *timestamp*, atividade, caso, recursos utilizados, entre outras. (AALST, 2011; BATISTA; SOLANAS, 2018). Porém, muitas vezes a obtenção do log das bases de dados é uma tarefa de alta complexidade, visto que tais bases podem se apresentar das mais diversas maneiras. Ainda, o log criado com as informações resgatas das bases deve passar por verificações de qualidade a fim de se determinar se devem ser aplicadas técnicas de filtragem, enriquecimento ou correção de erros (VANBRABANT, 2019; SURIADI, 2017). Assim, o processo de verificação de qualidade do log de eventos é de extrema necessidade para a adequada execução do pré-processamento do log.

# 2 OBJETIVO(S)

Almeja-se construir uma aplicação em software para auxiliar e facilitar o processo de avaliação de log de eventos. Essa aplicação deve dar suporte a profissionais, não necessariamente programadores, para que realizem o pré-processamento.

Objetivos específicos:

1. Realizar um estudo da literatura sobre métodos de avaliação de dados de sistemas de informações com vista a construção de logs de eventos;
2. Selecionar um subconjunto de métodos de avaliação de logs de eventos;
3. Implementar o subconjunto de métodos selecionados; e
4. Testar os métodos implementados.

# 3 MATERIAIS E MÉTODOS

Neste subprojeto está sendo utilizado um computador PC Quad Core, executando Windows 10 e/ou Linux disponível no laboratório LAS (Laboratório de Agentes de Software) do PPGIA (Programa de Pós-Graduação em Informática Aplicada). Para a implementação dos algoritmos serão usados softwares livres e outros licenciados/disponíveis na PUCPR como Java.

# 4 RESULTADOS PARCIAIS

Considerando o cronograma proposto no plano de trabalho, as seguintes atividades foram desenvolvidas:

1. Realizar estudo sobre as técnicas e abordagens de avaliação de log de eventos com vistas a mineração de processos.
2. Realizar seleção e avaliação de técnicas de avaliação de log de eventos.
3. Realizar estudo sobre arquitetura e abordagem de programação a ser utilizada.

Para estudos, avaliação e seleção de técnicas de log de eventos utilizou-se os trabalhos de Dixit (2018) e Vanbrabant (2019).

Dixit (2018) apresenta técnicas de avaliação e detecção de problemas na ordem dos eventos em um determinado log. A abordagem proposta neste artigo visa detectar e reparar imperfeições de ordenação de eventos que surgem devido a problemas relacionados ao *timestamp*. Como primeira etapa, os pesquisadores detectam exatamente onde podem existir problemas de carimbo de *timestamp* em um log, reconhecendo as características comumente associadas à imperfeição de ordenação de eventos.

Três técnicas de avaliação de *timestamp* são apresentadas. A primeira está relacionada à existência de granularidade grosseira em *timestamp* ou granularidade de *timestamp* mista (por exemplo, log de eventos inclui eventos de vários sistemas onde cada sistema registra *timestamps* de forma diferente). A segunda descreve a detecção de eventos que exibem uma ordem de acontecimentos incomum ou eventos repetidos. A terceira técnica detecta anomalias estatísticas mais genéricas, como aprender a posição temporal de uma determinada atividade no contexto de outras atividades, ou a distribuição de valores de *timestamp* de todos os eventos em um log.

Vanbrabant (2019) propõe diversas técnicas de avaliação de log de eventos. As técnicas estão relacionadas à detecção de dados faltantes como atributos, entidades e valores. Além disso, são explorados problemas de informação errada, como violação de ordem lógica, violação de dependência mútua, inexatidão de *timestamp*, erros de digitação ou valores fora de intervalos pré-definidos. Ainda, são detectados problemas de formatação incorreta nos dados, valores não explícitos, abreviações, entre outros.

Dentre as técnicas apresentadas, algumas foram selecionadas para implementação:

* Detecção de diferentes granularidades de *timestamp;*
* Detecção de valores ausentes;
* Detecção de violação de ordem lógica; e
* Detecção de anomalias temporais.

# 5 ETAPAS FUTURAS

Considerando o cronograma proposto no plano de trabalho, as seguintes atividades serão desenvolvidas:

* Realizar a implementação das técnicas selecionadas;
* Realizar testes com uma base de dados artificial a ser gerada;
* Realizar correções de erros; e
* Redigir o relatório final da iniciação científica.

# 6 REFERÊNCIAS

AALST, W. V. D. Process mining: discovery, conformance and enhancement of business processes. Heidelberg: Springer, 2011.

AALST, W. V. D. et al. Process mining manifesto. In: International Conference on Business Process Management. Springer, Berlin, Heidelberg, 2011. p. 169-194.

AALST, W. V. D. Data science in action. In: Process mining. [S.l.]: Springer, 2016. p. 3–23.

AALST, W. M. Van der; WEIJTERS, A. J. Process mining: a research agenda. [S.l.]: Elsevier, 2004.

BATISTA, E.; SOLANAS, A. Process Mining in Healthcare: A Systematic Review. In: 2018 9th International Conference on Information, Intelligence, Systems and Applications (IISA). [S.l.: s.n.], 2018. p. 1–6

DIXIT, Prabhakar M. et al. Detection and interactive repair of event ordering imperfection in process logs. In: International Conference on Advanced Information Systems Engineering. Springer, Cham, 2018. p. 274-290.

GARCIA, C. d. S. et al. Process mining techniques and applications – A systematic mapping study. Expert Systems with Applications, v. 133, p. 260–295, nov. 2019. ISSN 0957-4174.

SURIADI, S. et al. Event log imperfection patterns for process mining: Towards a systematic approach to cleaning event logs. Information Systems, v. 64, p. 132–150, mar. 2017. ISSN 0306-4379

VANBRABANT, L. et al. Quality of input data in emergency department simulations: Framework and assessment techniques. Simulation Modelling Practice and Theory, v. 91, p. 83–101, fev. 2019. ISSN 1569190X.