

UNIVERSIDADE FEDERAL DO RIO GRANDE
CENTRO DE CIÊNCIAS COMPUTACIONAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO
CURSO DE MESTRADO EM ENGENHARIA DE COMPUTAÇÃO

Dissertação de Mestrado

**Simulação de Logs de Eventos com Foco na Análise de
Processos de Construção na Indústria Naval Brasileira**

Thales Vaz Maciel

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Computação da Universidade Federal do Rio Grande, como requisito parcial para a obtenção do grau de Mestre em Engenharia de Computação

Orientador: Prof. Dr. Nelson Lopes Duarte Filho
Co-orientador: Prof^a. Dr^a. Silvia Silva Botelho

Rio Grande, 2016

Ficha catalográfica

M152s Maciel, Thales Vaz.

Simulação de logs de eventos com foco na análise de processos de construção na indústria naval brasileira / Thales Vaz Maciel. – 2016. 96 f.

Dissertação (mestrado) – Universidade Federal do Rio Grande – FURG, Programa de Pós-graduação em Engenharia de Computação, Rio Grande/RS, 2016.

Orientador: Dr. Nelson Lopes Duarte Filho.

Coorientadora: Dr^a. Silvia Silva Botelho.

1. Construção naval 2. Simulação 3. Mineração 4. Processos
I. Duarte Filho, Nelson Lopes II. Botelho, Silvia Silva II. Título.

CDU 004.6



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO RIO GRANDE
CENTRO DE CIÊNCIAS COMPUTACIONAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO
CURSO DE MESTRADO EM ENGENHARIA DE COMPUTAÇÃO

ATA DE SESSÃO DE DEFESA DE DISSERTAÇÃO DE MESTRADO

Ata nº 10 /2016

Na data de 30 de agosto de 2016, às 14 horas, ocorreu a Sessão de Defesa de Dissertação de Mestrado do discente Thales Vaz Maciel, que apresentou a dissertação intitulada "*Simulação de Logs de Eventos com foco na análise de processos de construção na Indústria Naval Brasileira*", realizada sob a orientação do Prof. Dr. Nelson Lopes Duarte Filho. A banca examinadora foi constituída pelos Profs. Drs. Luiz Felipe Assis (UFRJ), Eduardo Nunes Borges (FURG) e Silvia Silva da Costa Botelho (FURG) sob a presidência do orientador. Após a apresentação do trabalho, a banca arguiu o candidato e, a seguir, deliberou pela

- aprovação da Dissertação
- aprovação da Dissertação, sugerindo modificações no texto
- reprovação da Dissertação

Rio Grande, 30 de agosto de 2016



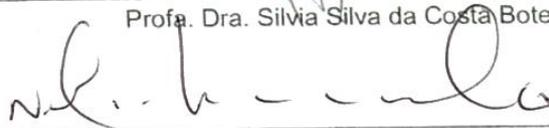
Prof. Dr. Luiz Felipe Assis



Prof. Dr. Eduardo Nunes Borges



Profa. Dra. Silvia Silva da Costa Botelho



Prof. Dr. Nelson Lopes Duarte Filho
Orientador

*À todas pessoas que torcem pela
minha felicidade e sucesso e
entendem por que fico "na
frente do computador".*

AGRADECIMENTOS

À Universidade Federal do Rio Grande, especificamente no Centro de Ciências Computacionais, pela infraestrutura e oportunidades disponibilizadas.

Ao Prof. Dr. Nelson Lopes Duarte Filho, pela paciência na orientação e incansável compartilhamento de conhecimento.

À Prof. Dra. Silvia Silva de Botelho, pela coorientação e eventual compartilhamento de ideias "mirabolantes".

Aos demais professores do Programa de Pós-Grauação Computação, pela docência.

Ao CNPq pelo apoio financeiro.

À minha família, embora muitos deles não estejam mais entre nós. É de suma importância reconhecer as pessoas que me fizeram ser quem sou. Sei que, de onde estão, torcem por mim.

À Deus, pela saúde, força, bençãos que me concede e pessoas boas que encontrei em meu caminho.

*With the lights out,
it's less dangerous.*
— KURT COBAIN

RESUMO

MACIEL, Thales Vaz. **Simulação de Logs de Eventos com Foco na Análise de Processos de Construção na Indústria Naval Brasileira.** 2016. 97 f. Dissertação (Mestrado) – Programa de Pós-Graduação em Computação. Universidade Federal do Rio Grande, Rio Grande.

Há muito se trata da necessidade de melhoramento na competitividade dos estaleiros da indústria brasileira de construção naval em relação a concorrentes no âmbito da indústria internacional. Em grande parte, a baixa eficiência verificada neste setor da indústria se dá pela baixa priorização do emprego tecnológico para metodologias automatizadas para controle e diagnóstico de processos de construção, por exemplo. Neste contexto, a mineração de processos vem sendo consolidada como solução para descoberta de modelos, análise de conformidade e melhoramento de processos. Contudo, estas atividades não são triviais, tendo como principal problemática a qualidade dos dados contidos nos logs de eventos. Este trabalho propõe uma metodologia para melhoramento de qualidade em logs de eventos originalmente caracterizados pela baixa granularidade das atividades nos aspectos quantitativos e temporal através do emprego de distribuições de probabilidades com a implementação de um novo software capaz de sintetizar um novo log de eventos, então livre de tais problemáticas de qualidade. Foi realizado um estudo de caso em estaleiro da indústria brasileira, onde foram possibilitados experimentos de descoberta de modelos de processos com algoritmos livres e proprietários, bem como a utilização de uma ferramenta de animação para detecção de gargalos no processo. Estes testes foram realizados com base no log de eventos original, provido pelo estaleiro e também sobre o log de eventos sintético, gerado pelo software de simulação, para fins de validação da abordagem. Os resultados mostraram sucesso ao revelar a fragmentação oculta das atividades, possibilitando a descoberta de modelos fidedignos e abrindo precedente à trabalhos futuros.

Palavras-chave: Construção naval, simulação, log de eventos, mineração de processos.

RESUMO

MACIEL, Thales Vaz. **Event Log Simulation with Focus on Analysing Processes from the Brazilian Shipbuilding Industry**. 2016. 97 f. Dissertação (Mestrado) – Programa de Pós-Graduação em Computação. Universidade Federal do Rio Grande, Rio Grande.

The need for improvement in the competitiveness on the Brazilian shipbuilding industry's shipyards, in relation to its competitors from abroad, is not a novice issue. The low efficiency that can be verified in this section of the national industry is greatly caused by the low prioritization of technological usage in automated methodologies for controlling and diagnosing the assembly process, for example. In that context, process mining has been consolidated as the solution for discovering models, conformance analysis and enhancement of business processes. However, such activities are far from trivial oftenly facing log event data quality issues. This work proposes a novice methodology for the improvement of data quality in event logs that are originally described as ungranular in the quantitative and temporal aspects, by using probability distributions with a new software implementation that is capable of synthesising a new event log, which is then free of such quality problematics. A case study has been performed in a shipyard from within the Brazilian industry, where various process discovery experiments have been executed with both free and proprietary algorithms. Also, a process model animation tool has been applied for bottleneck detection purposes. Such experiments were conducted based on the original event log that was provided by the shipyard's administration office and also on the event log that has been generated by the simulation software, for validating the presented approach. The results showed success in revealing the hidden fragmentation in the activities, enabling the discovery of trustworthy process models and opening precedents for future work.

Palavras-chave: shipbuilding, simulation, event logs, process mining.

LISTA DE FIGURAS

Figura 1	Comparação dos cenários ideal e real do registro de atividades de construção de blocos.	19
Figura 2	Modelo empírico de um processo de construção de blocos na indústria naval brasileira.	26
Figura 3	Cenário do parque de armazenamento de chapas de aço (SANTOS, 2012).	27
Figura 4	Painéis de aço na seção de armazenamento (SANTOS, 2012).	27
Figura 5	Composição de constructos na construção naval, segundo (SHA; MISRA; GUPTA, 2009).	30
Figura 6	Etapas da Descoberta de Conhecimento em Bancos de Dados (MACIEL et al., 2015)	32
Figura 7	Exemplo de notações utilizadas em redes de Petri (WESKE, 2012)	39
Figura 8	Exemplo de notações utilizadas em EPC (WESKE, 2012)	40
Figura 9	Exemplo de notações utilizadas em redes de fluxo (WESKE, 2012)	41
Figura 10	Exemplo de notações utilizadas em YAWL (WESKE, 2012)	42
Figura 11	Exemplo de notações utilizadas em BPMN (WESKE, 2012)	43
Figura 12	Utilização de Técnicas para Mineração de Processos (CLAES; POELS, 2012)	48
Figura 13	Utilização de Ferramentas pra Mineração de Processos (CLAES; POELS, 2012)	49
Figura 14	Modelo ER Parcial	55
Figura 15	Exemplos de distribuições de probabilidades que podem ser utilizadas para geração de valores aleatórios.	56
Figura 16	Modelo ER Final	62
Figura 17	Fluxograma do processo de geração do log de eventos sintético.	64
Figura 18	Modelo do processo antes da aplicação da metodologia	72
Figura 19	Exemplo de modelagem errônea de processo cujo log de eventos denota sobreposição de atividades.	75
Figura 20	Exemplo de modelagem de processos onde ordens de execução são registradas de forma confusa.	76
Figura 21	Consulta em linguagem SQL utilizada para extrair o log de eventos sintético a partir do conjunto de dados de eventos proveniente da simulação com o software implementado.	80
Figura 22	Modelo do processo após aplicação da metodologia	84
Figura 23	Quadro de animação do modelo do processo para identificação de gargalos	85

Figura 24	Comparação entre modelos gerados com o algoritmo Heuristics Miner sobre o log de eventos original (a) e sintético (b).	87
Figura 25	Comparação entre modelos gerados com o algoritmo Fuzzy Miner sobre o log de eventos original (a) e sintético (b).	88
Figura 26	Comparação entre modelos gerados na ferramenta Disco sobre o log de eventos original (a) e sintético (b).	89

LISTA DE TABELAS

Tabela 1	Comparação de Características de Business Intelligence e Mineração de Processos	17
Tabela 2	Tipos de Atividades da Construção Naval	30
Tabela 3	Propriedades Admitidas em Logs de Eventos, segundo (AALST, 2015)	37
Tabela 4	Descrição da entidade destinada ao armazenamento dos tipos de constructo admitidos no processo	54
Tabela 5	Descrição da entidade destinada ao armazenamento dos tipos de atividade admitidos no processo	54
Tabela 6	Descrição da entidade destinada ao armazenamento dos tipos de constructo tidos como possíveis requisitos para a execução dos tipos de atividade admitidos	54
Tabela 7	Descrição da entidade destinada ao armazenamento dos tipos de constructo tidos como possíveis produtos da execução dos tipos de atividade admitidos	55
Tabela 8	Descrição da entidade destinada ao armazenamento do peso total dos artefatos fabricados em cada instância do processo, conforme log de eventos original	55
Tabela 9	Descrição da entidade destinada ao armazenamento provisório das das posições temporais das atividades	57
Tabela 10	Descrição da entidade destinada ao armazenamento dos constructos simulados.	60
Tabela 11	Descrição da entidade destinada ao armazenamento das atividades simuladas para concretização de cada constructo simulado.	60
Tabela 12	Descrição das atividades do processo de construção de blocos	67
Tabela 13	Descrição da entidade destinada ao armazenamento do conjunto de dados originalmente disponibilizado ao estudo	68
Tabela 14	Amostra de dados do processo, em seu formato original.	69
Tabela 15	Descrição da entidade original_event_log, destinada ao armazenamento do log de eventos original	70
Tabela 16	Sobreposição parcial de atividades.	74
Tabela 17	Exemplo de log de eventos com atividades ocorrendo em paralelo.	75
Tabela 18	Exemplo de log de eventos com atividades ocorrendo em paralelo.	76
Tabela 19	Dados inseridos na entidade construct_type para parametrização parcial do algoritmo.	78

Tabela 20	Dados inseridos na entidade activity_nature para parametrização parcial do algoritmo.	78
Tabela 21	Dados inseridos na entidade activity_requirement para parametrização parcial do algoritmo.	79
Tabela 22	Dados inseridos na entidade activity_product para parametrização parcial do algoritmo.	79
Tabela 23	Descrição da entidade synthetic_event_log	79
Tabela 24	Log de Eventos Original para o Bloco L05C	81
Tabela 25	Log de Eventos Sintético para o Bloco L05C (parcial)	82
Tabela 26	Evidências da desfragmentação das atividades para 18 blocos de exemplo	83

LISTA DE ABREVIATURAS E SIGLAS

BAM	Business Activity Monitoring
BAMS	Block Assembly Monitoring System
BI	Business Intelligence
BPMN	Business Process Modeling Notation
CEP	Complex Event Processing
CPI	Continuous Process Improvement
CPM	Corporate Performance Management
EPC	Event-driven Process Chains
ER	Entidades-Relacionamentos
ERP	Enterprise Resource Planning
PROM	Process Mining Framework
RFID	Radio-Frequency Identification
SRCODES	Sistema para Rastreamento e Coleta de Dados para Estaleiros
SQL	Structure Query Language
UML	Unified Modeling Language
WEKA	Waikato Environment for Knowledge Analysis
YAWL	Yet Another Workflow Language

SUMÁRIO

1	INTRODUÇÃO	16
1.1	Problema a Ser Resolvido	17
1.2	Solução Proposta	20
1.3	Delimitação de Escopo	20
1.4	Justificativa	21
1.5	Objetivos	22
1.5.1	Objetivo Geral	22
1.5.2	Objetivos Específicos	22
2	METODOLOGIA	24
2.1	Visita de Campo	25
2.2	Revisão Bibliográfica	28
2.2.1	A Pluralidade de Especificações do Processo de Construção na Indústria Naval	28
2.2.2	Suporte Operacional com Base em Dados e Inteligência Artificial	31
2.2.3	Mineração de Dados	31
2.2.4	Mineração de Processos	35
2.2.5	Modelagem e Análise de Processos de Negócio	38
2.2.6	Abordagens e Algoritmos para Descoberta de Processos	44
2.2.7	Ferramentas para Mineração de Processos	47
2.2.8	Plataformas para Coleta e Armazenamento de Dados de Atividades	50
2.3	Projeto e Implementação do Software de Simulação	52
3	ESTUDO DE CASO	65
3.1	O Modelo do Processo de Construção Atual	66
3.2	Análise Semântica do Log de Eventos Original	67
3.3	Observação dos Resultados sobre o Log de Eventos Original	71
3.4	O Contexto Oculto das Atividades do Processo em Relação ao Log de Eventos	73
3.4.1	Sobreposição de Atividades da Mesma Instância do Processo	74
3.4.2	Registro Temporal de Baixa Granularidade e Imprecisão de Ordenação	75
3.4.3	Fragmentação Oculta das Atividades	77
3.5	Aplicação da Metodologia de Simulação do Processos	77
3.6	Análise do Log de Eventos Sintético	81
3.7	Mineração de Processos sobre o Log de Eventos Sintético	83

4	CONSIDERAÇÕES FINAIS	86
4.1	Resultados Obtidos	86
4.2	Conclusão	90
4.3	Trabalhos Futuros	91
	REFERÊNCIAS	92

1 INTRODUÇÃO

Há muito se trata da necessidade de melhoramento na competitividade dos estaleiros da indústria brasileira de construção naval em relação a concorrentes no âmbito da indústria internacional (SOUZA, 2009; DUARTE FILHO; BOTELHO; DOS SANTOS, 2013; DUARTE FILHO; BOTELHO; SANTOS, 2014; WEIS; BOTELHO; NAGY, 2015).

Em geral, a literatura inerente ao assunto converge sobre a problemática da baixa eficiência da construção naval do país em relação ao mercado internacional. Entende-se que, embora não seja a única problemática encontrada neste segmento industrial, ela é consequência direta da conduta de não priorização do emprego tecnológico adequado de metodologias automatizadas, tais quais sistemas computacionais, para controle, análise e diagnóstico de processos de construção e montagem, em estaleiros do país.

(SOUZA, 2009) considera que esta característica da indústria brasileira de construção naval seja pertinente a uma diferença fundamental entre ela e outros setores da indústria. O autor afirma que, normalmente, os sistemas utilizados para gestão de operações, que são incorporados nas atividades executadas nos estaleiros, são desenvolvidos para indústrias cujas metodologias se aproximam da construção ou da manufatura, definidas respectivamente, de forma breve, como orientada a projetos e orientada à produção em série.

Assim, considerando um suposto espectro de metodologias industriais composto por aquelas da construção e da manufatura, a indústria naval, entretanto, estaria posicionada em nenhuma das tipologias, mas em uma intersecção dos dois modelos, apresentando uma denotação híbrida com o compartilhamento de características de construção, de manufatura e, ainda, outras específicas em seu ramo de atuação.

Conseqüentemente, quando da verificação dos resultados apresentados por sistemas tradicionalmente utilizados, eles tendem a retornar respostas insatisfatórias, o que acontece devido à sua inadequação ao contexto específico do processo de construção naval e à inerente incapacidade em tratar particularidades deste cenário.

A mineração de processos surgiu com o objetivo de interseccionar as áreas de conhecimento da mineração de dados e da modelagem e análise de processos, provendo meios para descoberta, monitoramento e melhoramento de modelos de processos reais,

não apenas idealizados, através da extração de conhecimento a partir de logs de eventos provenientes dos sistemas de informação utilizados nas organizações (AALST, 2011a).

Com objeto correlato, ferramentas para Business Intelligence (BI) surgiram como propostas para a transformação de dados brutos em informação significativa e útil para melhoramento de estratégias, suporte operacional e tomada de decisões nas organizações (SANTOS; RAMOS, 2006).

Business Activity Monitoring (BAM), Complex Event Processing (CEP), Corporate Performance Management (CPM) e Continuous Process Improvement (CPI), entre outros, são termos já existentes dentro do escopo de BI, mas estas tecnologias têm em comum características que as diferenciam fundamentalmente da mineração de processos (Tabela 1) (AALST, 2011b):

Tabela 1: Comparação de Características de Business Intelligence e Mineração de Processos

Características	Business Intelligence	Mineração de Processos
Requisitos de Entrada	Esquemas Relacionais de Dados	Log de Eventos
Foco	Em Dados	Em Processos
Fundamentação	OLAP e Agregação de Dados	Inteligência Artificial

Desta forma, fica evidente que a mineração de processos, embora hoje seja referenciada dentro do escopo de BI, é fundamentalmente diferente do que havia sido agregado a esta área quando do seu surgimento, trazendo, então, o viés da inteligência artificial, foco em processos e um formato específico para entrada de dados.

Dentro deste contexto, esta dissertação visa instituir uma contribuição no âmbito da indústria brasileira de construção naval com viés no melhoramento da compreensão dos processos de construção utilizados nos estaleiros do país.

Isto é proposto através do desenvolvimento de uma nova metodologia para produção de dados sintéticos que sejam pertinentes às atividades de montagem desempenhadas, conforme descrito em um log de eventos preliminar, tendo em vistas a extração de um novo log de eventos, então isento de anomalias referentes à qualidade dos dados e, portanto, passível de utilização em tarefas de mineração de processos.

1.1 Problema a Ser Resolvido

O sucesso na aplicação de tecnologias da mineração de dados depende, essencialmente e sobremaneira, da proveniência de dados com a qualidade adequada às questões a serem respondidas. Portanto, é importante que hajam meios capazes de apresentar esses dados com correção e precisão compatíveis com o tipo de estudo proposto.

A problemática para aplicação da mineração de processos no contexto da indústria brasileira de construção naval ocorre como consequência da dificuldade ou impossibili-

dade da obtenção de dados que sejam isentos de incoerências que os tornam anômalos.

Especificamente no processo de construção de artefatos metálicos de grande porte, é muitas vezes inviável o monitoramento direto das atividades realizadas sobre todos os constructos que são consumidos e àqueles que são produzidos.

Isto acontece devido ao fato de que eles, independentemente de sua tipologia e tamanho, são geralmente compostos por números diversos de constructos menores de tipologia variada. Da mesma forma, é possível que estes também façam parte da composição de constructos maiores, em fases posteriores do mesmo processo de construção.

(EYRES; BRUCE, 2012) afirma que não existe uma definição típica para navios, tampouco para processos de construção naval. Contudo, os autores apresentam definições para os tipos de constructos que tipicamente fazem parte dos processos de construção de navios e plataformas:

- Constructos menores: perfis, baluartes, peças que são construídas a partir do processamento de chapas de aço, vigas e outros com peso até duas toneladas;
- Submontagens: painéis planos ou curvos, estruturas de envólucros ou outros com até 20 toneladas;
- Unidades: comumente a associação entre painéis e estrutura interior com peso até 60 toneladas;
- Blocos: Estrutura formada por unidades, com peso em torno de 200 toneladas.

Neste contexto, o rastreamento da composição e decomposição de constructos desde o início do processo de construção dos blocos torna-se trabalhoso ou invasivo demais para as práticas atuais no cenário da indústria brasileira de construção naval (WEIS; BOTELHO; NAGY, 2015).

Desta forma, a solução ordinariamente utilizada para a organização dos dados do monitoramento de atividades no perímetro de construção, é a da abstração de tais variâncias, em uma tentativa de solução paliativa para o registro de execução das atividades. Por conseguinte, o resultado implica em uma consolidação resumida, agrupada e de baixa granularidade dos registros de eventos do processo, podendo ocultar a fragmentação das atividades.

Por exemplo, uma atividade comum no processo de construção naval é o jateamento de chapas de aço com material abrasivo, com o intuito de remover impurezas e outros resíduos da superfície do metal que posteriormente é utilizado para o corte de peças e composição de painéis (EYRES; BRUCE, 2012). A figura 1 ilustra como um suposto registro deste tipo de atividade aconteceria no cenário ideal (a) e como se daria no cenário proposto como realidade atual na indústria brasileira (b).

constructo	atividade	tempo de ocorrência	bloco
1000	jateamento	80	B8
1001	jateamento	83	B8
1002	jateamento	84	B8
1003	jateamento	84	B8
1004	jateamento	84	B8
1005	jateamento	85	B8
1006	jateamento	95	B8

(a)

bloco	atividade	tempo de início	tempo de término
B8	jateamento	80	95

(b)

Figura 1: Comparação dos cenários ideal e real do registro de atividades de construção de blocos.

Adicionalmente, é possível que as fontes de dados mantidas pela empresa não tenham sido projetadas ou populadas com vistas na realização de tarefas de descoberta, análise ou melhoramento de processos. Em muitos casos, os conjuntos de dados acessíveis são, na verdade, apenas o subproduto de atividades de controle financeiro, planejamento de recursos (Enterprise Resource Planning, ERP) ou depuração de sistemas computacionais, por exemplo.

Em outras palavras, embora haja o registro das atividades executadas sob alguma forma, isto não significa que tenha acontecido com o propósito da realização de mineração de processos ou provendo qualidade dos dados em consideração da possibilidade de que estudos desta natureza venham a ser conduzidos.

A seção 3.4 provê detalhamento no contexto das problemáticas de qualidade em logs de eventos que são tratadas no âmbito do presente trabalho.

Dada a subsistente inviabilidade da aplicação de softwares próprios para coleta e armazenamento de logs de eventos das atividades desempenhadas no âmbito do processo de construção naval no país, torna-se inviável, também, a obtenção de resultados proveitosos a partir da mineração de processos diretamente sobre logs de eventos extraídos de fontes de dados tais como as descritas nesta seção - isto é, sem que hajam quaisquer tratamentos para modificação de métricas de qualidade do conjunto de dados para uma situação otimizada para contexto da realização de mineração de processos.

Embora dados formatados sob esta circunstância não denotem quaisquer anomalias quando analisados sob a perspectiva de seu propósito original, a mudança dessa perspectiva pode revelar a existência de contextos ocultos que podem ser utilizados no melhoramento em métricas como precisão e granularidade dos dados.

Hoje em dia, é diversa a tecnologia para pré-processamento em conjuntos de dados e log de eventos (HALL et al., 2009; DONGEN, 2005; VERBEEK et al., 2010). Contudo, não há propostas de tratamento para problemas de baixa granularidade dos atributos ou capacidade de solucionar problemas de insuficiência de precisão. Desta forma, é evidenciada necessidade da criação de metodologia original que seja capaz de solucionar tais anomalias de qualidade.

1.2 Solução Proposta

Diante da dificuldade na obtenção de logs de eventos satisfatórios sob a perspectiva dos requisitos para mineração de processos, esta seção apresenta a proposta de solução para a problemática ilustrada na seção 1.1.

Trata-se da utilização do sistema de software projetado e implementado no âmbito deste estudo, cujo produto é a simulação de processos em ambiente computacional e geração de logs de eventos a partir da mesma simulação.

Na abordagem proposta, a simulação é realizada com base em um log de eventos caracterizado pela baixa granularidade na descrição das atividades, porém acompanhado de conjuntos de dados complementares devidamente estruturados, que devem conter informações estatísticas inerentes ao domínio de negócio do processo analisado. A finalidade é que esses dados sejam utilizados na parametrização dos mecanismos de simulação possibilitados pelos software.

A hipótese estudada nesta dissertação é a de que, com base em aplicações de aprendizado de máquina, na forma de mineração de processos, pode ser fundamentada uma solução para o diagnóstico de processos de construção e montagem adotados na indústria naval brasileira.

Dentro deste contexto, há a segunda hipótese a ser trabalhada no estudo; a de que logs de eventos caracterizados pela baixa granularidade na descrição das atividades, com a complementaridade de conjuntos de dados inerentes e estruturados de acordo com uma especificação, podem ser aproveitados por mecanismos de simulação de processos em ambiente computacional para geração de logs de eventos isentos de anomalias, compatíveis com as restrições definidas e, portanto, aptos à proveniência conhecimento relevante quando da entrada em mineração de processos.

1.3 Delimitação de Escopo

Este trabalho tem caráter multidisciplinar e é composto por estudos nas áreas de mineração de dados, mineração de processos, análise e modelagem de processos de negócio, construção naval, estatística, bancos de dados, programação de computadores e simulação de dados.

A convergência destas áreas se dá com a concretização da proposta de resolução da problemática apresentada e avaliação de eficácia através de aplicação em estudo de caso sobre o cenário de um estaleiro brasileiro.

No escopo deste projeto estão a análise, projeto, implementação e testes do software para simulação de dados parametrizada com distribuições de probabilidades estatísticas, bem como aplicação e análise dos resultados coletados no estudo de caso descrito. Desta forma, o trabalho visa contribuição mista entre as áreas de mineração de processos, onde é proposta a nova metodologia de pré-processamento de dados e, também, na construção

naval, onde a abordagem é aplicada, sugerindo a possibilidade da mineração de processos sem a necessidade da aplicação de alternativas invasivas para coleta de dados no perímetro de construção.

1.4 Justificativa

Justifica-se a formalização da proposta de solução para a problemática apresentada no presente trabalho diante da dificuldade da implantação, no âmbito da indústria brasileira de construção naval, de sistemas específicos para a automatização da coleta e armazenamento de logs de eventos com características de qualidade satisfatórias sob a perspectiva dos requisitos para realização de atividades de mineração de processos.

Block Assembly Monitoring Systems (BAMS) são sistemas de informação para monitoramento da montagem de blocos na construção naval que não são propostos e implementados no meio acadêmico e no âmbito da indústria internacional (LEE; PARK; BAE, 2013; PARK; LEE; ZHU, 2014; KIM et al., 2015).

A utilização de dispositivos de identificação por radiofrequência (RFID) também é proposta no âmbito de projetos de sistemas de rastreamento automatizados (PRUYN; MARTINEZ; MOLENAAR, 2010; DUARTE FILHO; BOTELHO; DOS SANTOS, 2013; SILVA et al., 2015), embora sofram resistência de certos segmentos do mercado devido à alta sensibilidade à interferência do sinal o qual emitem para comunicação. Isto é, muitas vezes, provocado pela proximidade com materiais metálicos de alta densidade, o que é inerente a perímetros de construção naval e limita sobremaneira a aplicabilidade de tecnologias de radiofrequência em estaleiros desta indústria (WEIS; BOTELHO; NAGY, 2015).

Contudo, embora estes tipos de soluções especializadas sejam, em teoria, capazes de produzir meios para coleta de dados com flexibilidade adequada, elas têm em comum a especificação de arquiteturas e infraestruturas dedicadas que devem ser mandatoriamente implantadas para obtenção de funcionamento adequado do sistema.

Conseqüentemente, é dificultada a adoção de tais abordagens no contexto da indústria brasileira de construção naval pela problemática dos custos inerentes à implantação, treinamento de pessoal e a questão prática da invasividade que as abordagens propõem sobre as atividades do processo de construção.

Em resposta a esta problemática, a simulação de processos de construção naval já foi proposta em trabalhos como (SOUZA, 2009; KÖNIG et al., 2007; KIM et al., 2003). Contudo, entende-se que o diferencial da abordagem proposta neste trabalho, seja o aproveitamento de logs de eventos caracterizados pela baixa granularidade na descrição das atividades, com a complementação por dados estatísticos do próprio processo na parametrização do processo de simulação.

O escopo e granularidade dos dados contidos em um log de eventos deve ser com-

patível com a profundidade das questões a serem compreendidas, embora, em muitos sistemas de informação, dados de eventos sejam tratados apenas como uma consequência de atividades de depuração, ou outras diretamente relacionadas com a atividade-fim do negócio.

Neste contexto, a solução proposta para adequação da granularidade e detalhamento dos dados apresentado em logs de eventos altamente abstratos mostra relevância e inovação ao abordar a incorporação de conhecimento do processo analisado, em formato devidamente estruturado, na simulação de um novo log de eventos, então, livre de anomalias também no aspecto de granularidade.

Trata-se de uma abordagem complexa, dada a multidisciplinaridade da metodologia, abrangendo aspectos de inteligência artificial, estudo de processos da indústria de construção naval, compreensão profunda das atividades envolvidas, análise de dados de forma manual e automatizada, análise de resultados preliminares, identificação da problemática do registro de eventos no perímetro de construção naval, estudo de regras de negócio do domínio da construção naval, extração de dados estatísticos, geração de dados distribuídos probabilisticamente, projeto de banco de dados, extração de logs de eventos a partir de bancos de dados e programação do software de simulação de logs de eventos com base em distribuições de probabilidades parametrizadas.

A abordagem proposta neste trabalho foi projetada, implementada e, posteriormente, validada através de estudo de caso com objeto no processo de construção de blocos na indústria naval brasileira. Contudo, nota-se que a metodologia proposta é aplicável ao contexto geral de processos de construção e montagem, tendo em vista que o software produzido fora projetado em desprezo da natureza específica do processo ao qual teve aplicação em estudo de caso.

1.5 Objetivos

Esta seção formaliza os objetivos do trabalho, conforme descrito a seguir.

1.5.1 Objetivo Geral

Instituir uma contribuição original no âmbito da indústria brasileira de construção naval com viés no melhoramento da compreensão dos processos de construção utilizados nos estaleiros do país, com envolvimento de mineração de processos.

1.5.2 Objetivos Específicos

1. Identificar um método para melhoramento de qualidade em logs de eventos da indústria naval caracterizados pela baixa granularidade na descrição das atividades.
2. Propor uma metodologia de simulação de processos da indústria naval em ambiente computacional, com vistas na construção de logs de eventos sintéticos com

qualidade adequada à mineração de processos.

3. Realizar estudo de caso, comparando os modelos de processos inferidos com base no log de eventos original e no log de eventos sintetizado pela simulação, de forma a provar a eficácia da metodologia descrita.

2 METODOLOGIA

Este capítulo formaliza os procedimentos realizados como parte da metodologia de pesquisa da presente dissertação.

O trabalho foi iniciado por uma visita de campo à um estaleiro da indústria brasileira para a observação presencial sobre as atividades inerentes à montagem dos constructos, os tipos de constructos envolvidos, as relações entre eles nas atividades de construção e como cada tipo de constructo é produzido dentro do processo de montagem de blocos.

Posteriormente, foi iniciada a revisão bibliográfica, através de um estudo sobre o atual estado da arte em mineração de dados, mineração de processos, notações gráficas para análise e modelagem de processos de negócio e ferramentas utilizadas para estes tipos de atividades. Nesta etapa, foi buscada fundamentação para aplicação de mineração de processos no suporte à diagnóstico de processos e suporte operacional.

Foram estudados diversos algoritmos utilizados no âmbito da mineração de processos para descoberta de modelos. Assim, foi possível compreender seus mecanismos de funcionamento, requisitos para inferência de modelos de processos, vantagens e desvantagens em relação aos outros algoritmos existentes e como suas abordagens falham ao tratar a problemática apresentada nesta dissertação.

O domínio da construção naval, especificamente no processo de montagem de blocos, também foi investigado através de pesquisa na literatura relacionada ao tema. Desta forma, foi formada a relação entre as informações coletadas de forma empírica no início do estudo e as práticas apresentadas na bibliografia científica.

A partir da aquisição dos dados de eventos providos pelo estaleiro que foi alvo do estudo de caso na presente pesquisa, foi possibilitada a realização de testes preliminares para tomada de amostras da prática de mineração de processos sobre logs de eventos caracterizados pela baixa granularidade na descrição das atividades.

Neste contexto, foram inferidos modelos de processo com a aplicação dos algoritmos para descoberta de processos e analisada a representatividade de cada um em relação ao comportamento do processo na realidade. Desta forma, foi possível fundamentar a justificativa do presente estudo, e apresentada a incapacidade das abordagens até então existentes em tratar a problemática da baixa granularidade das informações contidas em

logs de eventos.

Com o objetivo de tratar a problemática apresentada, foi projetado e implementado um produto de software, utilizando a linguagem de programação Java e a linguagem SQL para manipulação de um banco de dados SQLite, também concebido no escopo do projeto.

O sistema teve como objetivo realizar a simulação do processo de construção de blocos, porém sendo capaz de estimar a ocorrência das atividades do processo real, com vistas na produção sintética de um log de eventos melhor adequado ao contexto da mineração de processos.

Com base em dados do log de eventos apresentado e informações complementares sobre a execução das atividades, devidamente estruturadas, o sistema é capaz de simular a posição temporal em que cada instância de atividade ocorre. Isto é concebido através da simulação de dados gerados aleatoriamente por distribuições de probabilidade devidamente tipificadas e parametrizadas e o registro dos dados simulados em tuplas do banco de dados.

Nota-se que as informações adicionais citadas são referentes a dados estatísticos do processo descrito no conjunto de dados e são, a partir desta proposta, utilizadas na parametrização das distribuições de probabilidades instanciadas pelo sistema e devem ser providas por profissionais com conhecimento do domínio do processo e do log de eventos estudados.

Após a simulação da execução das instâncias do processo, o sistema provê um esquema de banco de dados contendo entidades adicionais, enquadrando dados que possibilitam a extração de um novo log de eventos com o nível de granularidade das atividades adequado ao contexto das questões cuja atividade de mineração de processos tenha o objetivo de responder.

Ao final do experimento, foi possível comparar os modelos descobertos antes e depois da aplicação da metodologia proposta. Os resultados são analisados nas considerações finais, juntamente com a conclusão do estudo.

2.1 Visita de Campo

Esta seção visa relatar a realização de visita de campo a um estaleiro da indústria de construção naval brasileira, ocorrido no âmbito inicial da presente pesquisa, bem como documentar o produto da visita.

A principal finalidade do procedimento foi a da compreensão ds atividades envolvidas e o próprio processo utilizado para construção de blocos, conforme as práticas atuais no cenário da indústria naval brasileira.

Isto foi possível através da observação presencial da execução de atividades de jateamento, soldagem, corte e armazenamento de constructos nas oficinas de construção e armazéns de estocagem. Também foi observada a metodologia utilizada para o transporte

dos constructos, com esteiras e veículos específicos para deslocamento de cargas com peso que pode variar no âmbito das toneladas.

Foi possibilitada a análise de viabilidade do projeto de pesquisa referente à esta dissertação, no tocante à identificação de fontes de dados existentes e à possibilidade da criação de novas fontes, de onde pudessem ser extraídos conjuntos de dados para posterior aplicação de técnicas de mineração de dados e outras do âmbito do aprendizado de máquina, como a mineração de processos.

Neste contexto, foi obtido, junto à administração do estaleiro, um esboço de modelo gráfico do processo de montagem de blocos. Ele foi concebido de forma empírica com notações particulares e está disposto na figura 2.

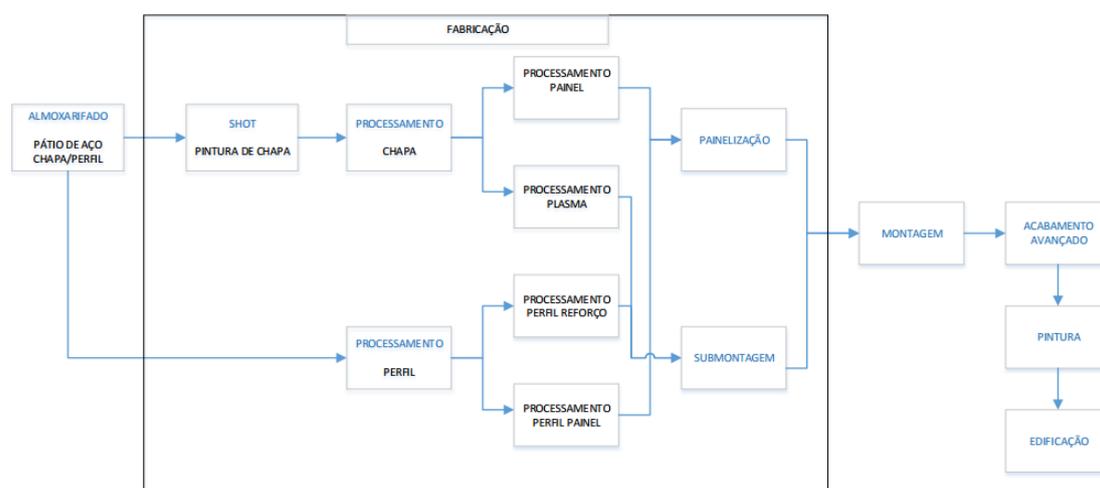


Figura 2: Modelo empírico de um processo de construção de blocos na indústria naval brasileira.

Através da observação presencial das atividades nas oficinas de montagem, foi possível identificar os tipos de constructos que são envolvidos no processo de montagem de outros constructos.

Em outras palavras, foi possível observar que a composição e decomposição dos materiais ocorre em diversas etapas desde a aquisição da matéria prima utilizada até a edificação de blocos já em estágio avançado de montagem.

Desta forma, foi possibilitado um entendimento mais refinado sobre o processo de construção de blocos e, portanto, melhor compreendida a denotação do modelo do processo idealizado que fora disponibilizado.

As figuras 3 e 4 apresentam imagens capturadas durante a visita no perímetro de construção do estaleiro. A primeira ilustra o parque de armazenamento de chapas de aço, enquanto a segunda ilustra uma pilha de painéis, possivelmente prontos para serem utilizados em uma tarefa de montagem.

Em face da obtenção de um modelo idealizado do processo de construção de blocos e evidências positivas em relação à viabilidade de coleta de dados para mineração de



Figura 3: Cenário do parque de armazenamento de chapas de aço (SANTOS, 2012).



Figura 4: Painéis de aço na seção de armazenamento (SANTOS, 2012).

dados e de processos, foi iniciado o estudo da literatura inerente à mineração de processos e estudos de caso inerentes no âmbito das indústrias de construção naval nacional e internacional.

Este modelo foi utilizado no estudo de caso presente nesta dissertação e é descrito detalhadamente na seção 3.1.

2.2 Revisão Bibliográfica

Esta seção caracteriza a revisão da literatura realizada no escopo da presente pesquisa. A revisão teve seu início com a investigação do processo de montagem de blocos na indústria de construção naval em âmbitos nacional e internacional, de modo à buscar compreensão sobre como os processos são relatados e referenciados no meio científico.

Foram buscadas referências sobre metodologias para suporte operacional com base em dados e inteligência artificial. Foram revisadas práticas e o estado da arte em mineração de dados e mineração de processos, além de notações gráficas utilizadas para análise e modelagem de processos.

Foram estudados algoritmos para descoberta de processos e analisados seus desempenhos e capacidade de tratamento perante a problemática abordada na motivação deste estudo. Também foram estudadas ferramentas, então consideradas o estado da arte em mineração de processos.

Finalmente, foram revisadas propostas acadêmicas para plataformas para coleta e armazenamento de dados de atividades industriais e averiguadas as ferramentas existentes para simulação de processos industriais, com vistas na simulação de logs de eventos dos processos simulados.

2.2.1 A Pluralidade de Especificações do Processo de Construção na Indústria Naval

No âmbito da indústria de construção naval, é possível observar que não há padronização no modelo de processo de construção de blocos que é adotado pelos estaleiros. Ao revisar a literatura, fica evidente a dissonância entre um perímetro de construção e outro, no que se trata da nomenclatura das atividades envolvidas no processo de construção e da relação de encadeamento entre elas, ao se considerar os constructos produto e aqueles necessários para execução de cada atividade.

(EYRES; BRUCE, 2012)

Os autores especificam o processo de construção naval em 17 fases: armazenamento de chapas de aço, armazenamento de seções, jateamento de chapas de aço, jateamento de seções, corte, marcação e forma, curva e corte, entrada de materiais para acabamento, curva de canos, fabricação de acabamento, submontagem, painelização, montagem de

matrizes, montagem de unidades, acabamento e modularização, montagem de blocos, edificação de blocos e acabamento final.

(PARK; LEE; ZHU, 2014)

Os autores descrevem um processo de construção de blocos onde não são abrangidas somente tarefas referentes à montagem física dos constructos: divisão de blocos, plano de aninhamento, pré-tratamento, corte, prensa, montagem de componente, montagem de unidade, montagem, pré-acabamento, pré-pintura, pré-edificação, edificação, acabamento, pintura e teste, totalizando a descrição de 15 atividades.

(LEE; PARK; BAE, 2013)

O trabalho propõe inicialmente um processo bastante abstrato, composto por 5 etapas: montagem, pré-acabamento, pré-pintura, pré-edificação e edificação. Contudo, o autor apresenta uma abordagem capaz de extrair logs de eventos a partir de logs de transporte dos constructos pelo perímetro do estaleiro estudado. Neste contexto, a abordagem se mostra capaz de inferir um processo de construção de blocos com 16 fases: montagem, montagem + pré-edificação, jateamento, docagem, inspeção, acabamento, acabamento exterior, pintura, pintura exterior, pré-edificação, pré-edificação pesada, teste, recebimento, armazenamento e corte, além de uma etapa denominada etc para representação de tarefas adicionais de menor relevância no processo.

(SOUZA, 2009)

Com uma organização mais minuciosa do processo de construção de blocos, (SOUZA, 2009) descreve 19 estágios da produção dos constructos, especificamente no âmbito da indústria brasileira. Segundo o autor, o processo compreende as atividades de armazenagem, tratamento, fabricação/corte, fabricação/perfis/conformação, submontagem, montagem, pré-edificação, edificação, pintura, acabamento na montagem, acabamento na pré-edificação, acabamento na edificação, instalação de módulos, armazenagem de blocos, acabamento na área de armazenagem, inspeção de blocos, fabricação de acabamento e montagem de módulos.

Ao tratar-se de um processo de construção incremental, tal qual o da montagem de blocos na indústria naval, é possível notar a distinção entre atividades do processo pela observação da conjuntura de constructos que são requisitos necessários para suas execuções, bem como aqueles que são produzidos em cada atividade.

É notório que os diferentes constructos produzidos em cada atividade são posteriormente consumidos por outra atividade do processo, assim como os diferentes constructos consumidos por cada atividade foram anteriormente produzidos ou incorporados ao pro-

cesso durante outra atividade.

Este contexto de interdependência entre as atividades que acontece através de relações produto(s)/requisito(s) é impossível de ser retratado explicitamente em logs de eventos que sejam conformados de acordo com o formato padronizado para tais conjuntos de dados (AALST, 2011a).

Contudo, é de suma importância, para o desenvolvimento da presente dissertação, a formalização da classificação das atividades do processo de construção naval em função da natureza de suas aplicações: decomposição, composição e tratamento (tabela 2).

Tabela 2: Tipos de Atividades da Construção Naval

Tipo de Atividade	Descrição
Decomposição	Referente às tarefas de corte, nas quais um constructo é recebido e, dele, são derivados (cortados) outros constructos menores.
Composição	Referente às tarefas de solda, onde diversos constructos são recebidos e agrupados fisicamente, de forma a compor um constructo maior.
Tratamento	Referente às tarefas de tratamento de constructos, onde o mesmo constructo que é recebido é o produto, porém, após sofrer preparação específica para outra etapa do processo, como jateamento com produto abrasivo ou pintura.

A figura 5 ilustra um exemplo sobre como pode acontecer a composição de constructos no processo de montagem de blocos.

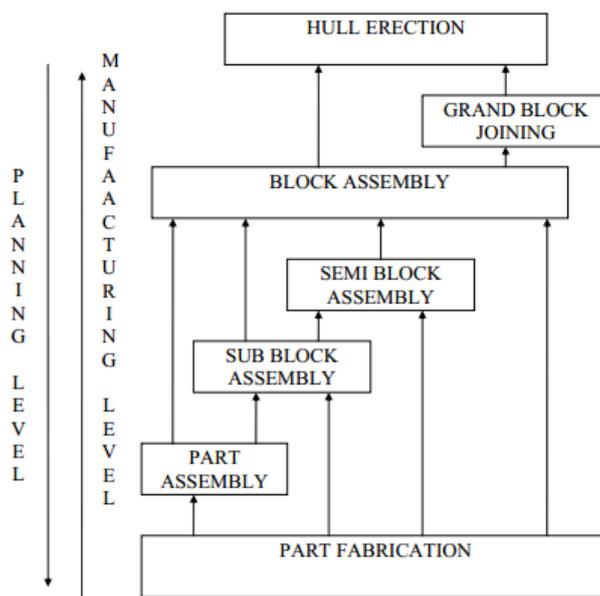


Figura 5: Composição de constructos na construção naval, segundo (SHA; MISRA; GUPTA, 2009).

Em vista do exposto nesta subseção, é possível observar grande diversidade na

especificação de processos de construção. Consequentemente, é admitida a situação em que a qualidade do registro dessas atividades e dos dados coletados e armazenados também varie entre um estaleiro e outro, de acordo com as regras de negócio inerentes às empresas mantenedora dos perímetros.

2.2.2 Suporte Operacional com Base em Dados e Inteligência Artificial

O aumento da competição entre as empresas, principalmente nos setores mais lucrativos da economia, faz com que as elas recorram às suas bases de dados objetivando conhecer melhor os seus clientes e se manter em prontidão para suprir as demandas dos mesmos (CAMARGO, 2002).

Devido ao crescente volume no registro eletrônico de atividades organizacionais e consequente maior detalhamento sobre o histórico de processos, existe a possibilidade de melhorar e dar suporte aos processos de negócio de forma cada vez mais competitiva e provendo agilidade na resposta à mudanças (AALST, 2011b).

A mineração de dados é proposta como uma solução à descoberta de conhecimento através da análise de conjuntos de dados com o objetivo da identificação de padrões como regras de associação, predição de valores, classificação e o agrupamento de instâncias com bases em suas características (HAN; KAMBER; PEI, 2012).

Com viés voltado para descoberta de conhecimento em processos de negócio, a mineração de processos propõe a investigação automatizada dos logs de eventos - registros de atividades mantidos por meio dos sistemas de informação ativamente presentes nas organizações, com o objetivo da extração de conhecimento relevante a tais processos (AALST, 2011a).

Neste contexto, a mineração de processos apresenta as funcionalidades de descoberta, monitoramento e melhoramento de processos, que podem ser utilizados em diversos domínios de aplicação (RUBIN et al., 2007).

2.2.3 Mineração de Dados

A mineração de dados pode ser definida como a extração de informações implícitas, até então desconhecidas e potencialmente úteis, a partir de dados ou, simplesmente, como o processo automatizado da descoberta de padrões em dados (WITTEN; FRANK; HALL, 2011).

Sob uma perspectiva mais estrita, abordagem pertinente neste trabalho, a mineração de dados é tratada como uma etapa do processo da descoberta de conhecimento, onde a análise de dados é automatizada em busca de padrões, sendo apoiada por algoritmos de aprendizado de máquina, estatística e técnicas de visualização (BORGES, 2013).

Ainda considerada como uma área de pesquisa em expansão, a mineração de dados passou por desenvolvimento significativo desde seu surgimento na década de 1980 e hoje é utilizada em um vasto espectro de áreas de conhecimento como ciência, medicina, en-

genharias, biologia, demografia, finanças e marketing, por exemplo (HAN; KAMBER; PEI, 2012).

Um esboço de processo de descoberta de conhecimento em bancos de dados é constituído por tarefas de pré-processamento, mineração de dados e pós-processamento (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996). A figura 6 apresenta como a mineração de dados fica disposta no contexto deste processo (MACIEL et al., 2015).

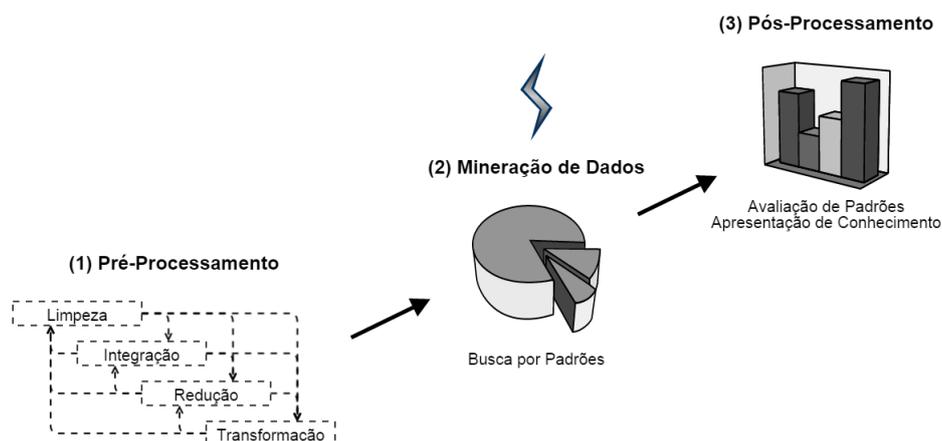


Figura 6: Etapas da Descoberta de Conhecimento em Bancos de Dados (MACIEL et al., 2015)

A etapa de pré-processamento é caracterizada pela execução de tarefas que devem preceder a execução dos algoritmos para descoberta de padrões. Este tipo de atividade tem o objetivo de adequar ou melhorar a qualidade dos dados que são posteriormente tidos como objeto de análise por algoritmos de aprendizado de máquina. As atividades de pré-processamento ocorrem nas formas de (HAN; KAMBER; PEI, 2012):

Limpeza de Dados

Trata-se o preenchimento de valores faltantes, remoção de dados ruidosos e a correção de possíveis inconsistências no conjunto de dados. Mesmo considerando que diversos algoritmos de mineração de dados possuem mecanismos capazes de tratar esses tipos de problemas de qualidade dos dados durante sua própria execução, muitas vezes, tais mecanismos não são robustos o suficiente, de forma que isto pode gerar confusão no processo de identificação de padrões e, conseqüentemente, resultados inconfiáveis. De forma a tratar estes problemas, considera-se a limpeza de dados como uma das mais importantes tarefas de pré-processamento na mineração de dados.

Integração de Dados

Onde ocorre a consolidação coerente de múltiplas fontes de dados em uma única, se necessário. Sistemas gerenciadores de bancos de dados, planilhas eletrônicas, arquivos de texto e documentos físicos são exemplos de fontes de dados cuja integração pode ser

necessária quando da execução de projetos de descoberta de conhecimento. Esta tarefa trata, ainda, da resolução de inconsistências resultantes da própria integração, de forma que é frequentemente seguida de medidas de limpeza de dados, para a eliminação de redundâncias e dados conflitantes, por exemplo.

Redução de Dados

Onde o objetivo é obter uma representação reduzida do conjunto de dados, mas que produza o mesmo resultado na análise, em comparação com o conjunto completo. Para tal, podem ser utilizadas estratégias de redução dimensional, com a escolha conveniente da codificação de caracteres e a seleção de subconjuntos dos dados originais, visando a eliminação de atributos irrelevantes ou redundantes, ou de redução numérica, com a aplicação de técnicas de amostragem aleatória simples com/sem substituição, estratificação, etc.

Transformação de Dados

Tem como finalidade a obtenção de formatos de dados mais adequados à execução específica da tarefa de mineração de dados. Pode ocorrer na forma da remoção de dados ruidosos, cálculo de valores para adição de novos atributos, assim como a normalização e discretização de dados numéricos.

A etapa de pós-processamento é caracterizada pela execução de tarefas que sucedem a mineração de dados, geralmente para a avaliação e apresentação dos padrões ou modelos descobertos pelos algoritmos aplicados no processamento.

A avaliação de padrões trata da identificação de padrões que apresentem maior relevância estatística dentre os descobertos e é considerada uma atividade essencial para a realização de progressos em atividades de mineração de dados (WITTEN; FRANK; HALL, 2011).

Algumas métricas utilizadas na avaliação de padrões são: a percentagem de acertos e erros, valor da estatística de kappa, avaliação da matriz de confusão (em atividades de classificação), homogeneidade dos agrupamentos (em clustering), confiança e suporte (em tarefas de associação), entre outras (HALL et al., 2009).

Embora a utilização de métricas baseadas em estatística possibilite a avaliação objetiva da relevância dos resultados encontrados no processamento, existe a subjetividade proposta pela avaliação dos resultados por usuários com conhecimento do domínio de negócio em estudo (HAN; KAMBER; PEI, 2012).

Devido à natureza experimental da mineração de dados, é possível que a relevância estatística de um resultado obtido seja incompatível com sua significância para o domínio de negócio (WITTEN; FRANK; HALL, 2011).

Finalmente, a visualização dos resultados da mineração de dados trata da apresentação gráfica do conhecimento obtido. Tais formatos podem incluir diagramas, árvores de decisão, regras de associação, agrupamentos, indicações de valores discrepantes, e outras estruturas genéricas (HAN; KAMBER; PEI, 2012).

A mineração de dados é o núcleo do processo de descoberta de conhecimento, onde são aplicados algoritmos para a extração de padrões a partir de dados pré-processados (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

É na fase da descoberta ocorre o processamento do conjunto de dados e descobertos os modelos que, posteriormente, são disponibilizados para avaliação e então apresentação do que for julgado relevante.

Em aprendizado de máquina, as atividades são categorizadas como aprendizado supervisionado, onde são processadas instâncias já rotuladas para atividades preditivas, ou aprendizado não supervisionado, onde não há rotulação prévia de instâncias (WITTEN; FRANK; HALL, 2011).

De forma mais aprofundada, também são referidos o aprendizado semi-supervisionado, onde tanto instâncias rotuladas quanto não-rotuladas podem ser utilizadas no treinamento, e o aprendizado ativo, onde um ator humano assume um papel ativo no aprendizado de máquina (HAN; KAMBER; PEI, 2012).

As funcionalidades da mineração de dados determinam quais tipos de padrões devem ser procurados e quais tipos de modelos devem ser apresentados como resultado (HAN; KAMBER; PEI, 2012). As aplicações destas funcionalidades se dão pela execução de algoritmos específicos sob respectivas implementações e suas variantes (HALL et al., 2009).

As funcionalidades de mineração de dados mais comumente referenciadas são (WITTEN; FRANK; HALL, 2011):

Classificação

O método de aprendizado utiliza um conjunto de instâncias já classificadas, as quais são utilizadas como base para inferir um modelo capaz de classificar novas instâncias. É comumente referido como aprendizado supervisionado devido ao método de funcionamento, onde a atuação do algoritmo é supervisionada através do prévio conhecimento do resultado da classificação para cada instância utilizada no treinamento.

O sucesso na atividade pode ser medido, de forma objetiva, pela proporção de acertos sobre um conjunto de dados de teste. Contudo, muitas aplicações práticas de mineração de dados utilizam critérios ainda subjetivos em relação a aceitação de um modelo no mundo real.

Tem, como objetivo, ser capaz de, após o treinamento com instâncias conhecidas, determinar a classe de instâncias novas.

Associação

Qualquer tipo de associação entre atributos é considerada, não apenas aquelas referentes à predição do valor de uma determinada classe. Neste caso, o objetivo é encontrar quaisquer associações que sejam considerada interessantes. Diferentemente da classificação, é capaz de prever valores de qualquer atributo, não somente a classe, e é capaz de prever os valores de mais de um atributo por regra.

Por isto, é comum a identificação de contagens imensas de regras e, portanto, a utilização de métodos para a supressão de resultados que não contemplem requisitos de aplicação relativa ao conjunto de dados e relevância. Geralmente envolve apenas atributos categóricos.

Agrupamento

São formados grupos de instâncias que compartilham características semelhantes. Quando não há especificação de classes para instâncias de treinamento, a tarefa também conhecida como clustering é utilizada para agrupar elementos que aparentam se aproximar naturalmente. O objetivo é o de descobrir estes agrupamentos e determinar qual deles abriga cada instância, além de ser capaz de atribuir novas instâncias a eles.

O sucesso na aplicação de clustering pode ser medido subjetivamente por quão útil o resultado parece no mundo real. Pode preceder tarefas de classificação, onde pode ser gerado um modelo inteligível sobre como novas instâncias serão alocadas nos grupos.

Regressão

É uma variante do aprendizado por classificação em um processo semelhante, com a exceção de que não prediz valores categóricos, mas valores numéricos (predição numérica). Em problemas de regressão, assim como em outras situações em aprendizado de máquina, o valor predito para novas instâncias é comumente menos interessante do que a estrutura que leva até dada conclusão. Isto acontece devido a possibilidade de expressar em termos quais atributos são importantes e como eles se relacionam ao resultado numérico.

Ferramentas de software para mineração de dados e descoberta de conhecimento existem no mercado, tanto privado como software livre. Alguns exemplos são o Waikato Environment for Knowledge Analysis (WEKA) (HALL et al., 2009), o RapidMiner (JUNGERMANN, 2009) e o sistema R (R Core Team, 2014).

2.2.4 Mineração de Processos

Sistemas de informação têm se tornado cada vez mais intrínsecos aos processos de negócios os quais suportam operacionalmente e, como resultado disto, grandes quanti-

dades de eventos do domínio do negócio são registradas por estes sistemas, embora ainda haja dificuldade, por parte das organizações, em extrair valor informacional a partir deste tipo de dado (AALST, 2011a).

Inicialmente, a mineração de processos surgiu com a premissa de prover um método para expor as discrepâncias entre modelos de processos envisioned e aqueles que seriam observados na prática (WEIJTERS; AALST, 2001). Na última década, evoluiu como um importante campo de pesquisa onde são investigados os diferentes tipos de relação entre modelos de processos e informações sobre a execução desses processos (WESKE, 2012).

A mineração de processos tem o objetivo de extrair informações sobre processos a partir do acesso a registros de transações (ou logs de eventos) assumindo que (DONGEN, 2005; WEIJTERS; AALST W.; MEDEIROS, 2006):

1. cada evento esteja relacionado a uma atividade ou etapa de processo;
2. cada evento esteja relacionado a uma instância de processo;
3. cada evento esteja relacionado a uma executante ou ator do sistema que realizou ou iniciou uma atividade; e
4. cada evento possua registro de data e hora em que aconteceram (timestamp) e que estejam ordenados.

Neste contexto, há a definição da estrutura padrão que é requerida em um log de eventos para que seja admitido como entrada em tarefas de mineração de processos. Um log de eventos deve estar conformado como uma coleção de registros de eventos, onde cada evento contém as propriedades descritas na tabela 3 (AALST, 2015).

Desta forma, fica evidente que a mineração de processos aparece como um método mais objetivo para esta descoberta de conhecimento, em comparação com a abordagem tradicional da especificação de processos de negócio, onde pessoas em atividade de determinadas tarefas seriam consultadas diretamente em busca da descrição de um processo.

Grande parte dos dados armazenados digitalmente se encontra de forma desestruturada e as organizações podem ter problemas ao tratá-los em grande quantidade. O desafio em mineração de processos é o de explorar logs de eventos visando melhorar a compreensão sobre processos, disponibilizar meios para identificação de gargalos de execução e violações de políticas internas, além de possibilitar a recomendação de contra-medidas em antecipação à eventuais problemas (AALST, 2011a).

Algumas características relevantes da mineração de processos são (AALST, 2011b):

- Não limitação à descoberta de controles de fluxo. Embora esta seja a atividade mais visada pelas práticas na área, ela representa apenas um dos três formatos básicos de tarefas de mineração de processos e apenas uma das três principais perspectivas da prática;

Tabela 3: Propriedades Admitidas em Logs de Eventos, segundo (AALST, 2015)

Propriedade	Descrição	Obrigatório
Instância	Indicação da instância do processo à qual o registro do evento pertence. Caso um evento compreenda múltiplas instâncias do processo, deve ser replicado de acordo.	Sim
Atividade	Indicação da atividade que ocorreu.	Sim
Timestamp	Indicação temporal de ocorrência do evento, no âmbito da instância do processo. Utilizada para ordenação temporal e avaliação de desempenho, é conceitualmente obrigatória, embora alguns algoritmos sejam capazes tratar sua ausência ou imprecisão através da identificação da ordenação física no log de eventos apresentado.	Sim
Recurso	Identificação da pessoa, máquina ou software que tenha executado o evento.	
Tipo	Tipologia transacional do evento, como início, fim, suspensão, retomada, etc.	
Custo	Custo associado à execução do evento.	
Cliente	Informação da pessoa ou organização para quem o evento esteja sendo executado.	

- Não se trata apenas de um tipo específico de mineração de dados. É considerada a conexão, até então inexistente, entre mineração de dados e gerenciamento de processos de negócio. Modelos de processos propõem estruturas mais complexas do que as resultantes de tarefas tradicionais de mineração de dados. Desta forma, foi aberto precedente para a criação de novos tipos de algoritmos e representações centrados em logs de eventos e processos de negócio.
- Não limitação ao processamento offline. Apesar de extrair conhecimento de logs e histórico de eventos, os resultados podem ser aplicados a instâncias em andamento, de forma que o tempo de execução de determinada instância de processo possa ser estimado a partir da descoberta do modelo do processo.

Tarefas de mineração de processos incluem (AALST, 2011a) a descoberta de processos, a análise de conformidade e o melhoramento de processos:

Descoberta de Processos

Recebe um log de eventos e produz um modelo de processo sem a utilização de informações previamente conhecidas. Um exemplo disto é o algoritmo Alpha (AALST; WEIJTERS; MARUSTER, 2004), que, com base em logs de eventos, é capaz de produzir redes de Petri aptas a explicar o comportamento registrado nos logs sem a necessidade de

informações adicionais. Também pode ser utilizado para a construção de modelos relacionados a recursos do negócio, caso os logs utilizados contem informações sobre estes.

Análise de Conformidade

Realiza a comparação entre um modelo de processo existente e um log de eventos do mesmo processo. A análise de conformidade pode ser utilizada para verificar se a realidade, conforme registrada nos logs de eventos, estão em conformidade com a especificação de um modelo, e vice-versa.

Exemplos de possibilidades deste tipo de análise é a verificação do efetivo cumprimento das regras do negócio, a verificação de casos onde atividades específicas dentro de um processo devem ser desempenhadas por indivíduos diferentes, e a detecção de atividades fraudulentas, com base na leitura de logs de eventos e um modelo que especifique requisitos. Portanto, a análise de conformidade pode ser utilizada para detectar, localizar e explicar e medir o impacto de desvios na execução de processos de negócio.

Melhoramento de Processos

Tem o objetivo de estender ou melhorar um modelo de processo já existente com base na utilização de registros de logs de eventos do próprio processo. Enquanto a análise de conformidade visa verificar o alinhamento entre um modelo de processo e a realidade, atividades de melhoramento procuram a modificação ou a extensão e um modelo já conhecido.

Atividades de melhoramento são definidas como reparo, onde há a modificação de um modelo para uma melhor reflexão da realidade, ou de extensão, onde é adicionada uma nova perspectiva a um modelo já existente.

Desta forma, hoje em dia é notório que a mineração de processos provê uma importante conexão entre modelagem e análise de processos de negócio, do gerenciamento de processos de negócio, e mineração de dados, da inteligência artificial (AALST, 2011b).

2.2.5 Modelagem e Análise de Processos de Negócio

Considerada uma atividade essencial da engenharia de processos de sistemas e indiscutivelmente presente em quaisquer áreas de atividade humana, a modelagem de processos trata da representação da realidade, com o objetivo de projetar, controlar e otimizar o comportamento de sistemas (CAMERON; HANGOS, 2001).

No contexto da modelagem de processos de negócio, os modelos especificam as atividades, conforme realizadas dentro das organizações, e suas relações entre si, provendo uma visão detalhada sobre as operações dos processos e suas restrições de execução (WESKE, 2012).

Diversos e importantes padrões abertos para modelagem de processos de negócio têm sido disponibilizados e eles vêm recebendo amplo e crescente suporte dentro da indústria de software voltado ao gerenciamento de processos de negócio, onde também existe a iniciativa para a criação de um conjunto unificado de notações visuais para a representação de fluxo de processos (JESTON; NELIS, 2014).

Quando da apresentação de um modelo de processo, é necessário considerar o viés representacional da notação gráfica utilizada para este propósito. Em outras palavras, é necessária a definição de uma linguagem para modelagem que tenha capacidade de representar os tipos de controle de fluxo requisitados na formalização de um modelo. Neste contexto, são destacadas as linguagens para modelagem de processos de negócio mais comumente utilizadas (WESKE, 2012):

Redes de Petri

São uma das mais conhecidas técnicas para a especificação formal e abstrata de processos e consideradas uma importante base para as demais linguagens de modelagem de processos. Podem ser utilizadas para modelar sistemas dinâmicos com estrutura estática, onde a rede de Petri representa os componentes estáticos um recurso (token) representa o comportamento dinâmico.

São compostas por locais, transições e arcos direcionados que conectam locais e transições na forma de grafos bipartidos, considerando que seus arcos nunca conectam dois lugares ou duas transições. Em notações gráficas, locais são representados por círculos, transições por retângulos e conexões por arcos direcionados. A figura 7 ilustra as notações gráficas utilizadas em redes de Petri.

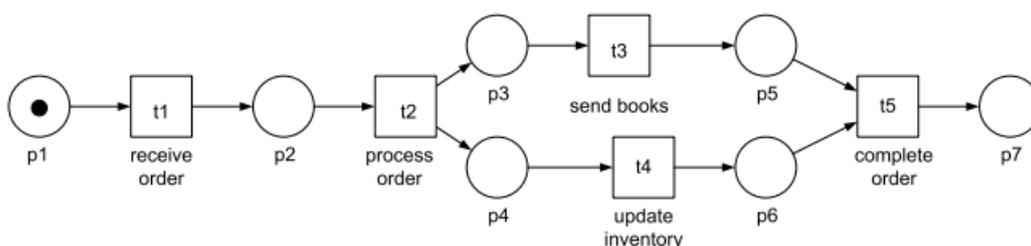


Figura 7: Exemplo de notações utilizadas em redes de Petri (WESKE, 2012)

Cadeias de Processos Orientadas a Eventos (Event-driven Process Chains - EPC)

São uma importante notação para a modelagem de aspectos de domínio de processos do negócio. Considerada uma notação informal, tem seu foco na representação de conceitos e processos de domínio ao invés de seus aspectos formais ou detalhes técnicos.

Elaborada como parte da abordagem holística de modelagem chamada Arquitetura de Sistemas de Informação Integrados (Architecture of Integrated Information Systems -

ARIS), propõe uma abordagem em três níveis de abstração: conceitual, arquitetural e de implementação, respectivamente definidos como definição de requisitos, especificação de projeto e descrição de implementação.

O nível conceitual representa maior abstração e trata da modelagem de dados, controle e funções, dando ênfase a requisitos não técnicos dos processos do negócio e seu ambiente de execução. Sua arquitetura tem o objetivo principal de criar associação entre o nível conceitual e o nível de implementação, onde são tomadas medidas concretas em busca da realização dos processos de negócio.

A figura 8 ilustra as notações gráficas utilizadas em cadeias de processos orientadas a eventos.

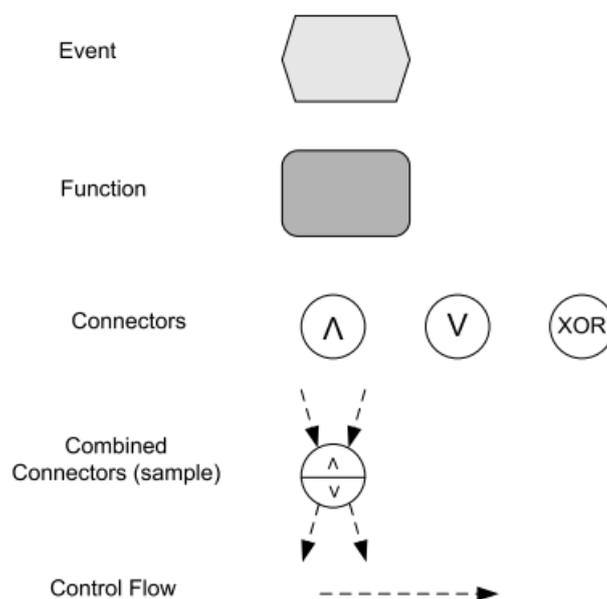


Figura 8: Exemplo de notações utilizadas em EPC (WESKE, 2012)

Redes de Fluxo

Embora redes de Petri sejam consideradas muito úteis para a representação de processos simples, processos mais complexos, como processos de negócio das empresas, necessitam de mecanismos avançados para modelagem.

Cadeias de eventos dirigidas a eventos, por sua vez, são capazes de prover apenas notações informais para a representação de processos e ambientes de negócio. Redes de fluxo apresentam uma abordagem de aperfeiçoamento das redes de Petri neste sentido, com conceitos e notações que facilitam a representação de processos de negócio, ao mesmo tempo que permitem a adição de restrições estruturais adequadas a modelagem de processos.

Redes de Petri e, conseqüentemente, redes de fluxo permitem que processos de negócio sejam definidos formalmente, o que se aplica particularmente aos aspectos de

controle de fluxo. Nelas, as atividades dos processos e suas restrições de execução são expressas graficamente, o que facilita a comunicação sobre os processos quando do envolvimento de diferentes stakeholders (mesmo que muitos deles prefiram notações mais informais).

A semântica formal expressa em redes de Petri permite a análise de propriedades dos processos. Embora muitas ferramentas de gerenciamento de processos de negócio sejam baseadas em redes de Petri, seu formalismo é independente de tecnologias ou plataformas. A figura 9 apresenta as notações gráficas utilizadas em redes de fluxo.

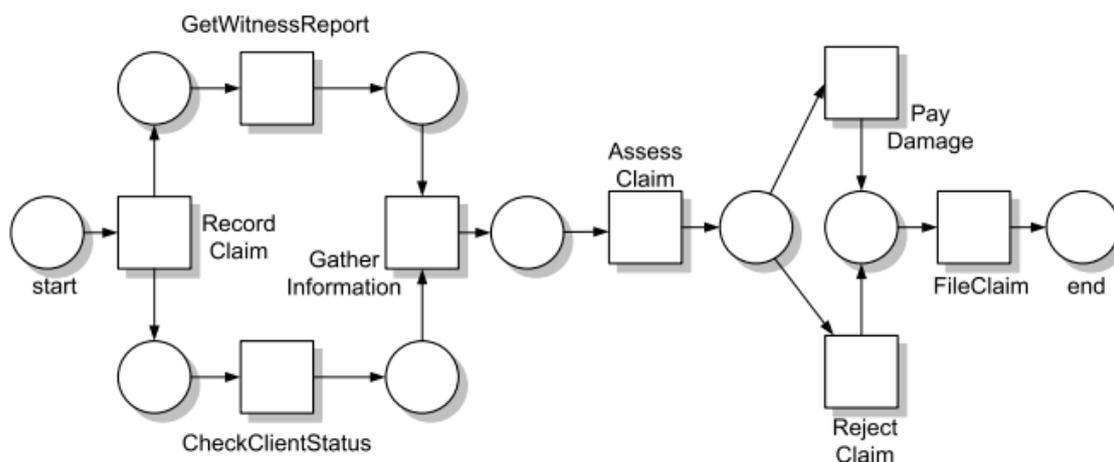


Figura 9: Exemplo de notações utilizadas em redes de fluxo (WESKE, 2012)

YAWL

A motivação para o desenvolvimento da Yet Another Workflow Language (YAWL), "Mais uma Linguagem de Fluxo" em tradução não-literal, foi a falta de uma linguagem de processos capaz de suportar todos padrões de controle de fluxo.

Embora tenha sua base em redes de fluxo, a semântica de execução dos processos é especificada por sistemas de transição de estados, não redes de Petri. Ainda que redes de Petri provenham bom formalismo para a expressão da maioria dos padrões de fluxo, possuem deficiências que impedem a expressão de todos eles.

Em relação a suas notações, a YAWL é baseada numa variante das redes de fluxo, chamadas redes YAWL, que também foram base para a especificação da linguagem. Redes YAWL provêm melhoramentos em redes de fluxo tradicionais com a adição de arcos que conectam transições, comportamento de junções de disjunções explícitas que podem ser associadas a transições, comportamento não-local e o suporte a tarefas com múltiplas instâncias.

A figura 10 apresenta as notações gráficas utilizadas na linguagem YAWL.

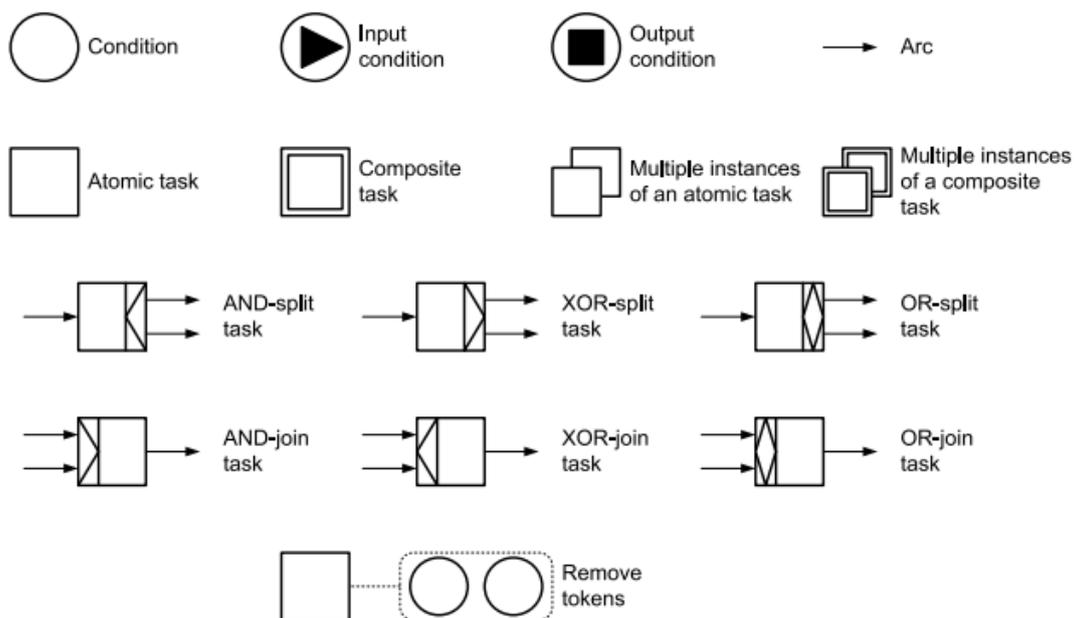


Figura 10: Exemplo de notações utilizadas em YAWL (WESKE, 2012)

BPMN

Desenvolvida pelo Object Management Group (OMG), este padrão internacional de modelagem de processos, em sua segunda versão, inclui diversos aprimoramentos desde a primeira. Antigamente Business Process Modeling Notation (BPMN), o agora Business Process Metal Model and Notation (BPMN) inclui, também, a definição de um meta-modelo para processos de negócio.

O objetivo da utilização de BPMN para modelagem de processos de negócio é similar ao da utilização da Unified Modeling Language (UML) para análise e projeto de software orientado a objetos, ou seja, identificar as melhores práticas das abordagens existentes e combiná-las em uma única, com maior abrangência e aceitação no mercado.

Enquanto outras linguagens procuram dar foco individualmente a cada nível de abstração, desde o nível de domínio do negócio até o mais técnico, a BPMN visa suportar todo o âmbito de níveis de abstração, do domínio do negócio até a implementação técnica. Portanto, a BPMN cria um vínculo entre o projeto de processos e suas implementações de forma padronizada.

A figura 11 apresenta as notações gráficas utilizadas na linguagem BPMN.

Modelos de processos de negócio, independentemente da linguagem e notações com as quais sejam construídos, além de ter o propósito da especificação das atividades que os constituem, podem ser amplamente utilizados no suporte da análise e otimização destes processos (HAMMER, 2010).

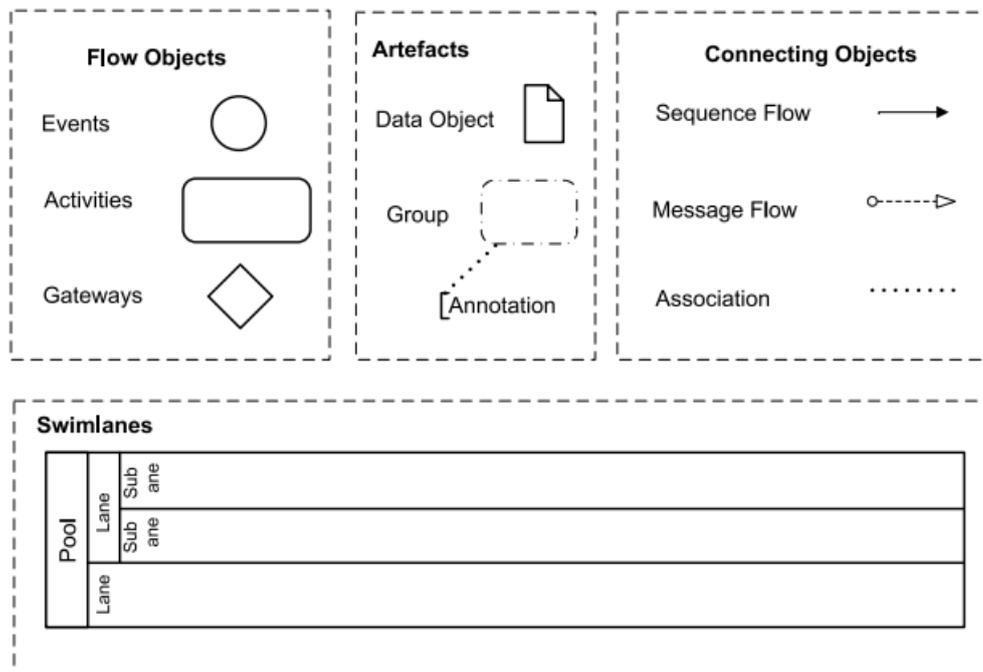


Figura 11: Exemplo de notações utilizadas em BPMN (WESKE, 2012)

(HOLT, 2009) propõe diferentes perspectivas à análise de modelos (HOLT, 2009): enquanto a verificação trata-se de garantir que o processo funcione corretamente, ou seja, que responda às possíveis entradas de maneira previsível, a validação deve certificar que o processo realmente realiza o que foi especificado, ou seja, supre seus requisitos.

Técnicas de simulação também são utilizadas no âmbito da análise de modelos de processos de negócio, onde certas sequências de execução indesejadas podem ser visualizadas em atividade simulada, com o objetivo de mostrar deficiências no modelo de processo, potencialmente apoiando atividades de validação (WESKE, 2012).

A avaliação de desempenho de processos de negócio ocorre pela identificação (no contexto do processo) de indicadores para este propósito e que permitam a especificação detalhada de seu desempenho, sendo estes indicadores divididos em quatro tipos ou grupos (HECKL; MOORMANN, 2010):

- **Qualidade:** descreve em que nível as propriedades e atributos de um produto estão em conformidade com as especificações reais deste produto. Embora indicadores referentes a custos fossem confundidos com qualidade no passado, nos dias de hoje é a satisfação do cliente, um indicador de maior relevância na medição de qualidade de produtos e serviços. Conseqüentemente, pode haver a definição de indicadores baseados nos requisitos que influenciam no desempenho;
- **Tempo:** é considerado um indicador de competitividade e desempenho de processos. Mesmo assim, o aspecto temporal pode ser avaliado de maneira subjetiva,

onde, em uma abordagem com foco em produção sob demanda, por exemplo, as situações de produção e entrega do produto de forma prematura ou tardia são ambos considerados desperdício de tempo. A otimização de processos neste aspecto trata da minimização do tempo de processo e os indicadores de desempenho incluem tempo de produção, tempo real de processamento, tempo de espera, tempo de transporte e tempo de entrega;

- **Custo:** é baseado em diversos fatores para a determinação de indicadores, incluindo custos de mão de obra, custos com tecnologia da informação, custos de produção, custos de serviços e custos de falhas. A distinção entre eles é feita sob a determinação de se o custo é fixo ou variável;
- **Flexibilidade:** é referente ao quanto um produto ou serviço pode ser modificado, incluindo valores de tempo e custos associados à reestruturação do processo de tal produto ou serviço. Um indicador de flexibilidade referencia a quantidade de componentes de produtos ou serviços que podem ser modificados ou trocados dentro de um determinado período de tempo.

Atividades de verificação, validação, simulação e análise de desempenho dependem fortemente na disponibilidade de modelos de alta qualidade, ou seja, na eventualidade de os modelos construídos e a realidade terem pouco em comum, a análise baseada em modelos perde sua fundamentação e motivação (AALST, 2011a). Desta forma, fica explícita e importância da utilização de boas práticas e da própria modelagem de processos de negócio.

2.2.6 Abordagens e Algoritmos para Descoberta de Processos

Diversas abordagens para a descoberta de processos vêm sido propostas e elas aparecem como soluções puramente algorítmicas em variações ou extensões do algoritmo Alpha (AALST; WEIJTERS; MARUSTER, 2004), envolvimento de heurística e tratamento de processos difusos, além de outras, baseadas em abordagens completamente diferentes, como inteligência computacional e abordagens divididas em duas fases, por exemplo (AALST, 2011a).

(TIWARI; TURNER; MAJEED, 2008) propõe uma taxonomia para as metodologias de descoberta de processos, que pode ser complementada pela proposta de (VAN DONGEN; DE MEDEIROS; WEN, 2009).

Abordagem Algorítmica Geral

São algoritmos originais projetados e propostos por autores diversos para minerar processos (TIWARI; TURNER; MAJEED, 2008).

(VAN DONGEN; DE MEDEIROS; WEN, 2009) referencia este tipo de abordagem como baseada em abstração e cita o algoritmo Alpha (AALST; WEIJTERS;

MARUSTER, 2004) como o primeiro algoritmo à implementá-la. Os autores consideram que todos algoritmos desenvolvidos subseqüentemente com base nesta metodologia foram extensões do algoritmo Alpha.

Abordagem Heurística

Propõe que apenas relações de ordenação frequentes o suficiente, de acordo com métricas de frequência identificadas heurísticamente a partir da análise do próprio log de eventos, sejam mantidas quando da construção de grafos de dependência para o modelo de processo.

Abordagens tradicionais para mineração de processos, como a utilizada pelo algoritmo Alpha (AALST; WEIJTERS; MARUSTER, 2004), tendem a funcionar bem em processos que tem um estrutura bem definida, mas têm problemas ao tratar processos não estruturados e, desta forma, muitas vezes, falham em prover resultados relevantes (GUNTHER; VAN DER AALST, 2007).

Conseqüentemente, modelos descobertos com tais abordagens apresentam todos detalhes do processo, sem distinguir o que é relevante e o que não é, de forma que um modelo pode representar a realidade, mas não prover um nível de abstração que permita seu uso na prática (DONGEN; ADRIANSYAH, 2010).

Abordagem Genética

Representada por algoritmos projetados com inspiração no processo darwiniano de seleção natural (TIWARI; TURNER; MAJEED, 2008).

A abordagem evolucionária provida por algoritmos da biologia computacional busca reproduzir computacionalmente o processo de evolução conforme observado na natureza, sendo capaz de tratar logs de eventos ruidosos ou incompletos (AALST, 2011a).

No contexto da descoberta de processos, segundo (MEDEIROS, 2006), estes algoritmos são adaptados para: interpretar o log de eventos; calcular as relações de ordenação entre atividades; construir a população inicial de modelos de processo gerados aleatoriamente, com base nas atividades do log de eventos interpretado; calcular o encaixe de cada modelo de processo gerado, em relação ao log de eventos; decidir entre retornar os modelos com melhor encaixe ou criar nova população através da utilização de operadores genéticos e retornar à etapa (4) do cálculo do encaixe de cada modelo de processo gerado.

Segundo (AALST, 2011a), uma desvantagem da abordagem genética é que esta não apresenta comportamento muito eficiente para logs de eventos ou modelos extensos, considerando que o alto consumo de tempo e poder computacional, que ainda é inerente à algoritmos da biologia computacional.

Abordagem Baseada em Regiões (de Linguagens)

A mineração de processos baseada em regiões utiliza uma abordagem dividida em duas etapas: (1) um sistema de transições, que é a representação mais básica possível para um processo, é induzido e utilizado como representação intermediária do processo e; então, (2) transformado, com base na aplicação da "teoria das regiões"(BERGENTHUM et al., 2007), em uma representação mais refinada, no formato de rede de Petri, podendo ser convertidas em outras notações de alto nível, como BPMN, YAWL, EPC, etc. (AALST et al., 2007).

Nota-se que esta abordagem é capaz de tratar problemas como o da descoberta de dependências complexas, paralelismo e repetições aninhadas, além da ocorrência de eventos similares em diferentes fases do processo, gerando modelos consistentes, apesar de que, para o tratamento de dados ruidosos e incompletos, sugere-se a integração deste com métodos de heurística (AALST, 2011a).

Algoritmos para Descoberta de Processos

O algoritmo Alpha, descrito por (AALST; WEIJTERS; MARUSTER, 2004), utiliza da metodologia baseada em abstração para mineração de processos. O processo de descoberta ocorre em três etapas (VAN DONGEN; DE MEDEIROS; WEN, 2009): abstração, indução e construção.

Na etapa de abstração, todos eventos são interpretados e as relações básicas de ordenação entre as atividades são estipuladas, de forma a retratar a quantidade de diferentes sucessões entre as tarefas do processo. Na fase de indução, relações de ordenação avançadas, como dependência causal, paralelismo e exclusão mútua são inferidas a partir das relações básicas já processadas. Na etapa de construção, o modelo final é construído a partir das relações de ordenação avançadas, que são utilizadas na determinação das restrições impostas pelo modelo do processo.

O Alpha+ (MEDEIROS et al., 2004) foi a primeira extensão do algoritmo Alpha, adicionando suporte à loops curtos em modelos baseados em redes de fluxo.

O Tsinghua Alpha (WEN et al., 2009) adicionou ao algoritmo Alpha a capacidade de tratar eventos não atômicos, ou seja, eventos que contém informações transacionais, como indicação de início e término.

O Alpha++ (WEN et al., 2007) propôs ao algoritmo Alpha a capacidade de descobrir construtores de fluxo de escolha não-livres. Como resultado, o modelo é uma extensão sem denominação específica de uma rede de fluxo.

O Alpha# (WEN; WANG; SUN, 2007) é uma extensão do algoritmo Alpha+ que permite a descoberta de modelos com notações diferenciadas daquelas propostas por redes de fluxo, como por exemplo, a representação de transições invisíveis.

Uma importante habilidade, para o âmbito da análise automatizada de dados, que não é implementada no algoritmo Alpha e suas extensões mais conhecidas é o trata-

mento de ruído, ou seja, detecção de dados que tenham sido registrados de forma incorreta ou que representem situações excepcionais no log de eventos (VAN DONGEN; DE MEDEIROS; WEN, 2009). Consequentemente, estes algoritmos são incapazes de considerar a frequência das relações de ordenação básicas quando da inferência das relações de ordenação avançadas e posterior construção do modelo do processo.

Neste contexto, foi proposto o algoritmo Heuristics Miner (WEIJTERS; AALST W.; MEDEIROS, 2006), capaz de tratar eventos ruidosos e pode ser utilizado para expressar o comportamento principal que fora registrado em um log de eventos, de modo que nem todos detalhes e comportamentos excepcionais sejam denotados graficamente.

(VAN DONGEN; DE MEDEIROS; WEN, 2009) afirma que os algoritmos baseados em abstração e aqueles baseados em heurísticas são incapazes de retratar todos construtores comuns de fluxo ao mesmo tempo em que apresentam robustez no tratamento de ruído.

Em resposta a esta problemática, foi idealizada a abordagem genética, implementada através dos algoritmos Genetic Algorithm Miner (MEDEIROS, 2006) e Duplicates Genetic Algorithm Miner (MEDEIROS; WEIJTERS; AALST, 2007), que buscam simular o comportamento da evolução natural na seleção de um modelo de processo ótimo.

Outra situação, descrita por (VAN DONGEN; DE MEDEIROS; WEN, 2009), é a do erro na admissão da hipótese de que todas informações contidas em um log estejam registradas no mesmo nível de abstração. Segundo o autor, o que se observa na prática é justamente o contrário, onde o nível de abstração das informações varia entre diferentes níveis.

Neste contexto, há a proposta dos algoritmos fuzzy, capazes de tratar processos altamente indefinidos e confusos. No âmbito de algoritmos como o Fuzzy Miner (GUNTHER; VAN DER AALST, 2007) e Disco Miner (GUNTHER; ROZINAT, 2012), técnicas de clustering são empregadas para agrupar eventos, de forma a criar abstrações na representação do modelo do processo em uma maneira adaptativa e flexível.

Desta forma, os algoritmos fuzzy são capazes de efetivamente representar o processo em diferentes níveis de abstração, obedecendo à parametrização feita pelo usuário do algoritmo.

Em (CLAES; POELS, 2012) é realizado um estudo exploratório acerca da prática de mineração de processos onde, entre outros resultados, são elencadas as ferramentas mais populares, bem como as técnicas de mineração de processos mais utilizadas.

A figura 12 apresenta o resultado parcial da pesquisa sobre a utilização de técnicas de mineração de processos (CLAES; POELS, 2012).

2.2.7 Ferramentas para Mineração de Processos

Esta subseção apresenta ferramentas para mineração de processos estudadas e utilizadas no âmbito deste trabalho para fins da análise de logs de eventos.

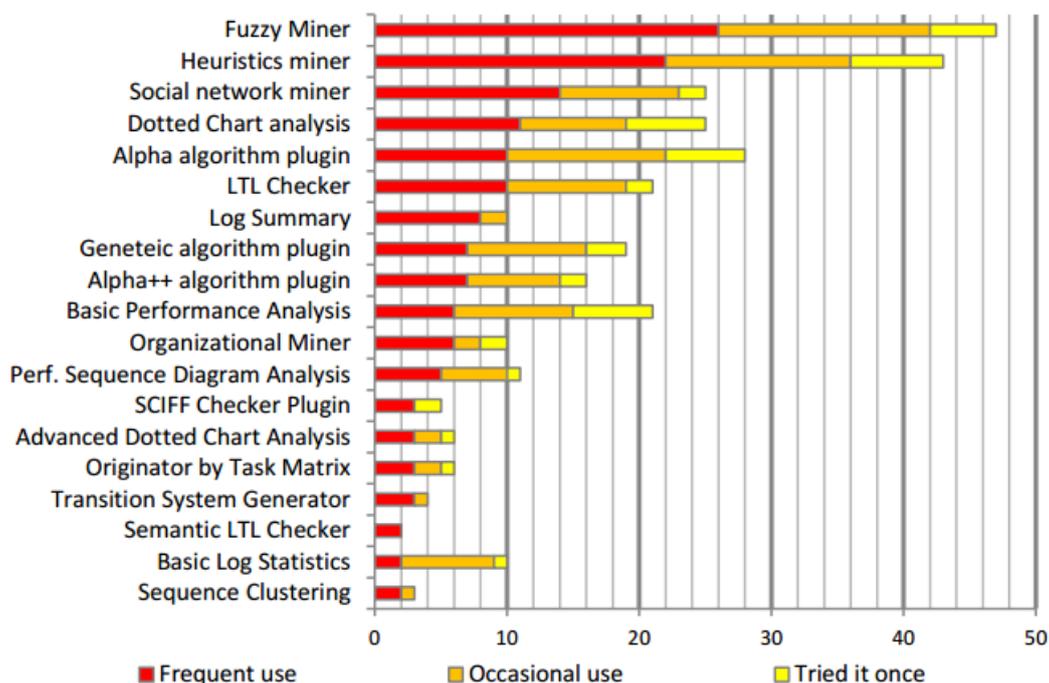


Figura 12: Utilização de Técnicas para Mineração de Processos (CLAES; POELS, 2012)

O Framework ProM

No contexto da implementação de algoritmos de descoberta de processos, destaca-se o Process Mining Framework (ProM) (DONGEN, 2005), um framework que provê uma variedade de algoritmos sob interface unificada e suporta atividades da mineração de processos no sentido amplo.

Em outras palavras, o ProM não provê somente funcionalidades de descoberta de processos com base em logs de eventos, mas também métodos para filtragem de dados, identificação de gargalos de execução no processo e verificação de regras de negócio, etc.

Além destas, o ProM também disponibiliza interfaces gráficas para a importação e extração de dados de diferentes fontes de dados (RUBIN et al., 2007).

Outras ferramentas para a extração de conhecimento a partir de logs de eventos já existiam nos meios acadêmico e comercial antes da disponibilização do ProM, como EMiT (DONGEN; AALST, 2004), Little Thumb (AALST; WEIJTERS, 2004), ARIS PPM (VAN DER AALST; TER HOFSTEDE; WESKE, 2003) e ILOG Jviews (SANDER; VASILIU, 2002).

Contudo, elas utilizavam formatos diferentes para recuperar e armazenar dados e expressavam seus resultados sob diferentes representações, o que dificultava suas utilizações simultâneas em um mesmo log de eventos para a comparação de resultados, além de que cada uma destas ferramentas implementavam conceitos que poderiam ser úteis em outras ferramentas, mas, devido à inexistência de interface entre elas, pesquisadores que estivessem trabalhando em novos projetos de mineração de processos deveriam construir

toda infraestrutura necessária para realizar testes de forma isolada desde as bases, sem conexão com outras implementações (DONGEN, 2005).

O ProM permite que pesquisadores implementem algoritmos de mineração de processos em um ambiente padronizado, integrado, genérico e de código aberto, o que possibilita diversas facilidades para o suporte no desenvolvimento de novas abordagens, como o aproveitamento da infraestrutura existente para a realização de experimentos (BOLT; LEONI; AALST, 2015).

Sendo flexível em relação à formatos de entrada e saída e a reutilização de código, o ProM abrange as funcionalidades de várias ferramentas já existentes e provê diversos outros recursos, suportando diversos formatos de arquivos, linguagens de modelagem de processos e implementações de algoritmos, que podem ser combinados e utilizados em situações de casos de uso reais (DONGEN, 2005).

O ProM é, por muitos, considerado a ferramenta padrão para a mineração de processos, contendo a implementação do estado da arte no que se trata de descoberta e melhoramento de processos (BOLT; LEONI; AALST, 2015).

A figura 13 apresenta um resultado parcial do estudo realizado por (CLAES; POELS, 2012), onde o framework ProM é identificado como a ferramenta mais popular e mais utilizada para estudos de mineração de processos.

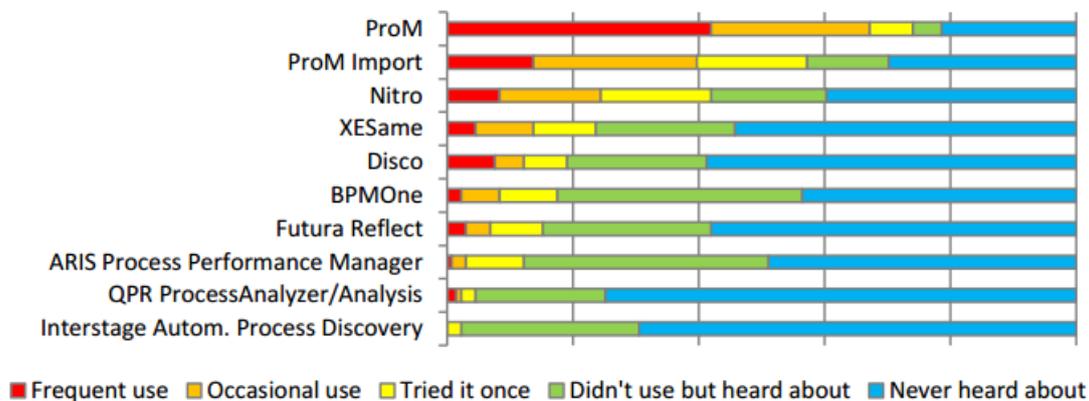


Figura 13: Utilização de Ferramentas pra Mineração de Processos (CLAES; POELS, 2012)

Fluxicon Disco

O Fluxicon Disco é uma ferramenta de software para mineração de processos, cujo algoritmo de descoberta implementa uma extensão do algoritmo Fuzzy Miner (GUNTHER; VAN DER AALST, 2007), presente no framework ProM (DONGEN, 2005). Em relação ao seu predecessor, adiciona diversas informações estatísticas ao mapa do processo no resultado da descoberta (GUNTHER; ROZINAT, 2012).

Enquanto o ProM é um framework para mineração de processos que tem no meio

científico o seu foco de desenvolvimento, o Disco é um produto de software proprietário com foco em usabilidade, provendo funcionalidades de importação de dados, identificação de atributos relevantes, filtragem, descoberta automática de processos e estatísticas inerentes, análise de desempenho, e animações de processos, entre outras (GUNTHER; ROZINAT, 2012).

Ao tratar-se de uma ferramenta proprietária, não há a publicação total da metodologia utilizada pelo algoritmo de descoberta de processos que ela implementa.

2.2.8 Plataformas para Coleta e Armazenamento de Dados de Atividades

Esta subseção em o objetivo de formalizar a revisão bibliográfica realizada sobre propostas de plataformas para coleta e armazenamento automatizado de atividades de produção, no âmbito dos processos de construção na indústria naval brasileira.

(DUARTE FILHO; BOTELHO; DOS SANTOS, 2013)

Em (DUARTE FILHO; BOTELHO; DOS SANTOS, 2013), é proposta uma abordagem original em um sistema computacional para rastreamento de pessoas, materiais e equipamentos da indústria da construção naval.

No trabalho, aplicações de RFID foram utilizadas para a coleta, armazenamento e processamento de dados de localização de recursos, com viés na assistência automatizada de processos de tomada de decisão.

O sistema denominado Sistema de Rastreamento e Coleta de Dados para Estaleiros - SRCODE, que também abrange a infraestrutura de rede necessária para implantação da tecnologia, foi dividido em três componentes: gerenciador com interface web, interface de comunicação entre dispositivos de RFID e banco de dados e o modelo virtual tridimensional.

Os testes práticos concluíram que o sistema é capaz de prover dados de localização dos recursos, desde que as tags RFID fossem posicionadas com intervalo de, no máximo, 15 metros de distância.

(DE GRANDI; PARZIANELLO, 2014)

A obra de (DE GRANDI; PARZIANELLO, 2014) descreve o processo de desenvolvimento de um framework para rastreamento através de RFID capaz de abstrair os diferentes tipos de marcas de equipamentos, estimar a posição de determinados objetos, armazenar os dados coletados e fornecer visualização dos mesmos.

O projeto do framework propôs uma arquitetura modular, separando as responsabilidades por comunicação, armazenamento, posicionamento e visualização. O software foi escrito em linguagem C++ e a validação do funcionamento foi feita em testes controlados em laboratório, onde os framework se mostrou adequado para a aplicação a qual se propõe.

(FORTUNATO; SOUZA; BARBOSA JR., 2015)

Em (FORTUNATO; SOUZA; BARBOSA JR., 2015), os autores abordam a automação do acompanhamento da produção em indústrias de construção e montagem através do uso de visão computacional. O trabalho propõe uma arquitetura distribuída em uma rede de câmeras aéreas, instaladas perpendicularmente ao solo, de forma a prover cobertura sobre os perímetros de construção e montagem das oficinas.

O sistema possibilita o acompanhamento em tempo real do avanço das atividades realizadas e rastreamento de recursos humanos, com a finalidade da composição de um banco de dados contendo informações que podem ser analisadas por sistemas de logística dos estaleiros.

(SILVA et al., 2015)

(SILVA et al., 2015) propõe a utilização de identificação por radiofrequência (RFID) no monitoramento da localização de insumos e o controle sobre o investimento de horas de trabalho nas atividades desempenhadas.

Através de uma arquitetura baseada em tags ativas ou passivas e leitores RFID, o trabalho propõe a identificação das atividades de processo de construção, de modo a oferecer formas de melhorar o fluxo nas linhas de produção, reduzir custos e tempo de construção.

De forma abstrata, o trabalho versa sobre a possibilidade da construção de um banco de dados para promover melhorias contínuas nos processos da indústria, além da adoção de novos processos, melhoramento de qualidade e controle de logística.

Aplicabilidade na Indústria Naval Brasileira

Embora apresentem abordagens originais e resultados promissores em suas aplicações ou casos de uso, (WEIS; BOTELHO; NAGY, 2015) classifica a aplicação de automação nas atividades de rastreamento no cenário da indústria de construção naval brasileira como inexpressiva.

(DUARTE FILHO; BOTELHO; SANTOS, 2014) reconhece a problemática da interferência no sinal de radiofrequência, capaz de inviabilizar a utilização de tecnologias que tenham nela a fundação de seu funcionamento. Segundo os autores, é inerente ao perímetro de construção naval que haja a presença de grandes objetos metálicos, fortes campos magnéticos causados por equipamentos de soldagem e redes wireless, por exemplo, que tornam os estaleiros ambientes hostis para a aplicação de tais tecnologias. É possível que esta situação seja a fundamentação da afirmação feita por (WEIS; BOTELHO; NAGY, 2015).

2.3 Projeto e Implementação do Software de Simulação

Esta seção descreve o processo de desenvolvimento do software utilizado como parte fundamental da solução para a problemática de logs de eventos quanto à qualidade dos dados. Esta problemática está descrita na seção 1.4 do presente trabalho.

Embora a existência de dados em quantidade volumosa e organização coerente à um domínio de aplicação seja, abstratamente, suficiente para se considerar o emprego da mineração de dados - ou log de eventos no caso da mineração de processos, é necessário considerar também a qualidade dos dados providos para análise.

Conforme disposto na seção 1.4, o domínio da indústria naval de grande porte provê alta complexidade dos processos e características inerentes, de forma que pode ser inviabilizada a coleta de dados das atividades diretamente com a granularidade e nível de abstração adequados à prática de mineração de processos.

A partir da visita de campo documentada na seção 2.1, ficou evidente que a relação de composição/decomposição entre os diversos tipos de constructos utilizados na montagem de qualquer tipo de constructo (não somente blocos) pode ficar oculta quando da descrição do processo no perímetro de construção dos estaleiros.

A subseção 2.2.1 trouxe referências sobre como as atividades do processo de construção de blocos variam de acordo com a logística de construção de cada estaleiro. Em decorrência deste cenário, a nomenclatura dos tipos de constructos manipulados durante as atividades de montagem também varia, bem como a relação de encadeamento entre as atividades do processo.

O contexto subjetivo das regras de negócio inerentes às atividades do processo de montagem de blocos mostra que o processo trata de diferentes atividades que consomem diferentes recursos de diferentes tipos em diferentes quantidades durante o tempo de execução das instâncias do processo.

Esta é uma dimensão de circunstâncias externas ou contexto oculto, conforme descrito por (LINDSTROM; DELANY; MAC NAMEE, 2008), que não é admitida na composição padronizada de um log de eventos (AALST, 2011a). Isto influencia diretamente na descoberta de conhecimento sobre processos, que ocorre com base em algoritmos que são incapazes de detectar tais subjetividades.

Por conseguinte, é necessário um método para que estas regras de negócio sejam integradas à tarefa de mineração de processos. A solução aqui apresentada parte do princípio de que o log de eventos inicial expressa atividades atribuídas no maior nível de abstração apenas e consequentes anomalias referentes à baixa granularidade das atividades e seus atributos.

O problema é tratado a partir da utilização de um sistema de software projetado e implementado no contexto deste estudo com a finalidade da simulação de atividades de construção naval e sintetização de um log de eventos que representa a execução das ativi-

dades na simulação.

O log de eventos original serve de entrada para a execução do software, com a adição de dados complementares que contenham informações estatísticas sobre o processo sendo analisado. Também é sugerida a presença de dados que denotem noção de volume dos constructos no mesmo nível de abstração e granularidade descritos no log de eventos, como, por exemplo, o peso dos blocos relacionados.

Estes dados devem ser providos por profissionais com conhecimento técnico sobre o domínio de aplicação estudado e devem passar por estruturação lógica antes da importação para o banco de dados. São eles:

1. tipos de constructo manipulados no processo de construção e respectivos limites de peso inferiores e superiores;
2. tipos de atividade admitidas no processo e respectiva natureza: decomposição em vários constructos, composição de um constructo ou tratamento de constructo;
3. relação entre os tipos de constructo com as atividades que os têm como prováveis requisitos e a quantificação desta probabilidade se concretizar durante a simulação;
4. relação entre os tipos de constructos com as atividades que os tem como produto da execução;
5. relação entre os constructos, descritos como instâncias do processo no log de eventos original, e o peso atribuído para cada um, no respectivo nível de abstração;

Com formatação adequada, estes dados servem na parametrização do método de geração de um novo conjunto dados, que é o propósito do software.

O sistema propõe a importação do log de eventos original para um esquema de banco de dados relacional, de forma a prover meios flexíveis para acesso aos atributos do log de eventos. Assim, também é facilitada a exportação de eventuais resultados de consultas para formatos manipuláveis em outras aplicações para análise de dados (MACIEL; FILHO, 2015).

A tabela 3, disposta na seção 2.2.4, exhibe a conformação dos dados genericamente requeridos para a composição de um log de eventos. As tabelas 4, 5, 6, 7 e 8 apresentam, respectivamente, a estrutura proposta para os conjuntos de dados suplementares ao log de eventos principal.

É requisito que os dados suplementares informados sejam estruturados conforme formalizado nesta seção, antes de serem integradas à simulação. O diagrama entidade-relacionamento (ER) disposto na figura 14 promove o entendimento das relações entre as entidades utilizadas na parametrização da metodologia de simulação apresentada.

No âmbito deste trabalho, em relação aos diagramas ER apresentados, nota-se que os tipos de dados dos atributos das entidades foram suprimidos. Isto fora sugerido com

vistas na melhor legibilidade do modelo e tal abstração foi justificada pela irrelevância de tal informação no contexto deste estudo.

Tabela 4: Descrição da entidade destinada ao armazenamento dos tipos de constructo admitidos no processo

Entidade construct_type	
Nome do Campo	Descrição do Campo
type	Nomenclatura do tipo de constructo.
weight_minimum	Peso mínimo admitido para constructos do respectivo tipo.
weight_maximum	Peso máximo admitido para constructos do respectivo tipo.

Tabela 5: Descrição da entidade destinada ao armazenamento dos tipos de atividade admitidos no processo

Entidade activity_nature	
Nome do Campo	Descrição do Campo
activity	Nomenclatura da atividade.
nature	Natureza da atividade: decomposição, composição ou apenas tratamento.

Tabela 6: Descrição da entidade destinada ao armazenamento dos tipos de constructo tidos como possíveis requisitos para a execução dos tipos de atividade admitidos

Entidade activity_requirement	
Nome do Campo	Descrição do Campo
activity	Nomenclatura do tipo de atividade.
construct_type	Nomenclatura do tipo de constructo requerido para a atividade.
composition	Porcentagem do respectivo tipo de constructo nos requisitos da respectiva atividade.

A solução projetada propõe que o sistema seja capaz de interpretar um log de eventos e, com a informação de conjuntos de dados suplementares, disponha de fundamentação para a instanciação, parametrização e utilização de distribuições de probabilidades diversas que são utilizadas na estimativa de características de cada evento virtualizado na simulação.

De acordo com (NIST/SEMATECH, 2012), uma das principais aplicações de distribuições de probabilidades acontecem no âmbito da análise exploratória de dados, em estudos de simulações que envolvem a geração de valores aleatórios que obedeçam determinado padrão ou conjunto de restrições para geração de valores.

Distribuições de probabilidades possuem importância fundamental no âmbito desta pesquisa devido ao fato de que a solução para simulação do processo utiliza de suas aplicações em diversas situações, dentro do ambiente de execução.

Tabela 7: Descrição da entidade destinada ao armazenamento dos tipos de constructo tidos como possíveis produtos da execução dos tipos de atividade admitidos

Entidade activity_product	
Nome do Campo	Descrição do Campo
activity	Nomenclatura do tipo de atividade.
construct_type	Nomenclatura do tipo de constructo que se dá como produto da respectiva atividade.

Tabela 8: Descrição da entidade destinada ao armazenamento do peso total dos artefatos fabricados em cada instância do processo, conforme log de eventos original

Entidade original_weight	
Nome do Campo	Descrição do Campo
caseid	Identificação da instância do processo, conforme relacionados no log de eventos original
weight	Peso do respectivo constructo, conforme inferido do log de eventos original

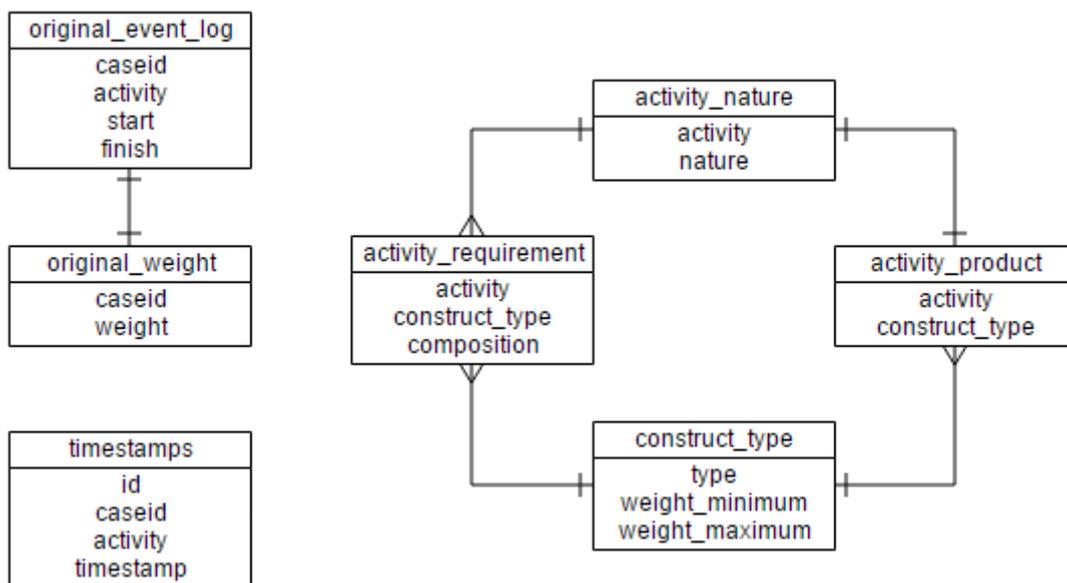


Figura 14: Modelo ER Parcial

O padrão da concentração de ocorrências de valores gerados reflete a definição da distribuição de probabilidade utilizada, e a determinação dos tipos de distribuições de probabilidades a serem utilizadas em cada aspecto possibilitado pelo software faz parte da parametrização do sistema.

A figura 15 apresenta exemplos de distribuições de probabilidades conhecidas que podem ser utilizadas para descrever a concentração de ocorrências de valores sobre determinada dimensão das características disponíveis em um contexto, como peso, posição temporal ou custo, por exemplo.

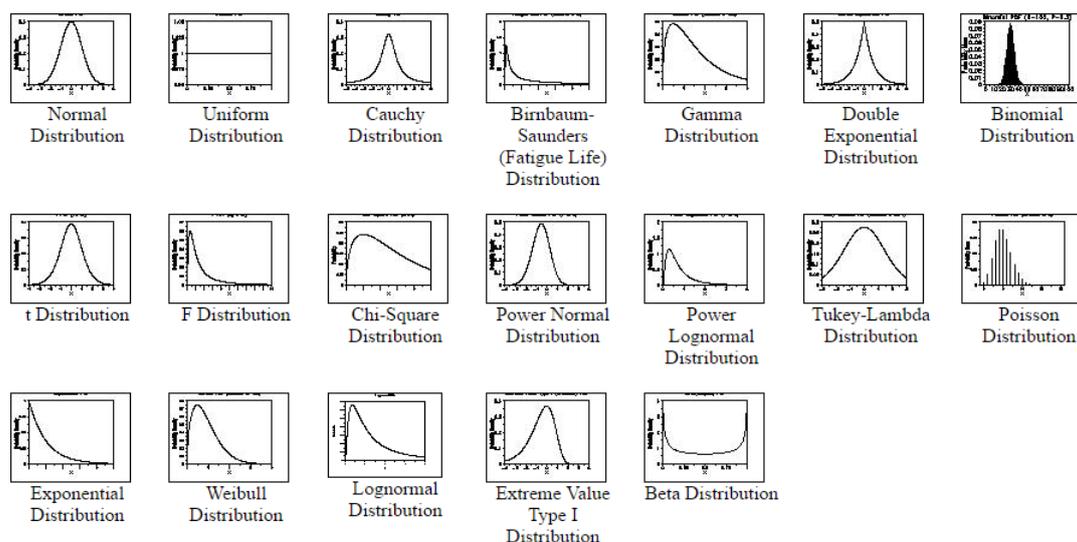


Figura 15: Exemplos de distribuições de probabilidades que podem ser utilizadas para geração de valores aleatórios.

Desde que devidamente quantificadas e estruturadas, é concebida a integração de circunstâncias externas ao log de eventos original, ao mesmo tempo em que é provida base estatística para a satisfação das restrições temporais impostas pelo log de eventos original. Consequentemente, é possibilitado que a simulação do processo virtualize atividades e constructos em níveis de abstração e granularidade de forma flexível, embora dentro de um contexto controlado.

Realizada a importação do log de eventos original e conjuntos de dados suplementares para o esquema de banco de dados utilizado no software de simulação, este dispõe de mecanismos para a estimativa da quantidade de cada tipo de atividade e respectiva posição temporal de cada uma delas na instância do processo.

A quantidade de realizações de cada atividade é determinada através da razão entre o total do peso do aço utilizado na instância do processo e a média aritmética do peso mínimo e máximo para o construto informado como produto da execução da respectiva atividade.

Para determinação da posição temporal de cada execução de atividade na simulação, é instanciada uma distribuição de probabilidade que é parametrizada com valor mínimo

correspondente ao início da atividade para a respectiva instância do processo no log de eventos original e como valor máximo correspondente ao dado de término da atividade no log de eventos original.

Com base na instância do processo denotada no log de eventos original, a atividade manipulada, a quantidade de atividades a serem simuladas e a instância uma distribuição de probabilidade pronta para ser utilizada, são simuladas as marcações temporais (timestamps) para a ocorrência de cada atividade dentro da simulação.

Neste contexto, é de suma importância notar que, embora o log de eventos original denote as unidades temporais em semanas, para fins de tratamento das problemáticas de qualidade quando da geração dos dados sintéticos, a granularidade da unidade de tempo foi aumentada para dias. Consequentemente, um suposto espectro de tempo de uma unidade (semana) no log original, passar a ter sete unidades (dias) no log sintético, e assim sucessivamente.

O armazenamento lógico destas informações ocorre através de inserções na entidade timestamps, conforme a estrutura exposta na tabela 9.

Tabela 9: Descrição da entidade destinada ao armazenamento provisório das das posições temporais das atividades

Entidade timestamps	
Nome do Campo	Descrição do Campo
id	Identificador único da timestamp.
caseid	Instância do processo, conforme o log de eventos original, à qual a respectiva timestamp foi atribuída.
activity	Identificador da atividade atribuída à timestamp.
timestamp	Valor da posição temporal.

Em consequência da atribuição prévia das atividades à uma marcação no espectro temporal disponível, é criado precedente lógico para que o método de simulação das atividades evolua gradualmente através do espectro temporal especificado no log de eventos original.

Isto possibilita a associação da disponibilidade de determinados constructos com a ocorrência das atividades que os têm como requisitos para própria execução. Da mesma forma, também é primordial para esta abordagem que haja a associação contínua das atividades com os constructos que são produto de suas execuções.

Após a definição das posições temporais para ocorrência de cada atividade, é simulada a aquisição dos constructos básicos utilizados no processo de construção, que ocorre através de inserções na tabela construct, com a população dos dados referentes à identificação do constructo, respectivo tipo e constructo original, de acordo com o log de eventos original.

Segundo (EYRES; BRUCE, 2012), no processo real de construção de blocos, há a

estimativa do peso final para cada um destes constructos. De acordo com os autores, há também a definição usual de que aproximadamente 5% da matéria-prima utilizada na construção de tais constructos seja perdida entre atividades de corte, que podem produzir retalhos de chapas de aço inutilizáveis, por exemplo. Neste contexto, a medida a ser tomada é a da aquisição e incorporação de matéria-prima com peso adicional ao estimado como peso final do constructo em 5%.

Para fins da metodologia de simulação, foi adotado o mesmo critério, sendo adicionados 5% do peso definido para os blocos ao peso total da matéria-prima a ser incorporada ao processo de construção do respectivo bloco. É importante notar que este dado pode ser definido via parametrização do sistema, dando flexibilidade ao seu contexto na simulação.

A partir da simulação da incorporação da matéria-prima a ser utilizada, o software é capaz de iniciar a simulação do processo de construção dos blocos. Para cada instância do processo, a simulação evolui no espectro temporal definido pela população da entidade timestamps. Para cada timestamp de cada instância do processo é realizada uma tentativa de execução da atividade correspondente. Durante a execução de cada atividade, são verificados os constructos tidos como requisitos à atividade a ser executada, bem como a probabilidade desta relação ser realizada na simulação. Da mesma forma, é identificado o tipo de constructo que é produto da atividade correspondente e a natureza da atividade.

Estes dados são recuperados do banco de dados utilizado como entrada no sistema, para onde foram importados os dados suplementares do processo, e aqueles referentes aos requisitos da respectiva atividade são utilizados para parametrização da distribuição de probabilidade que determina os constructos que serão consumidos na simulação da respectiva execução.

A partir da determinação da natureza da atividade simulada, o processo de simulação executa lógicas distintas. Para fins de registro dos dados resultantes da simulação, o sistema dispõe das entidades estruturadas conforme especificado nas tabelas 10 e 11.

Caso a atividade corresponda ao tratamento de um constructo, é verificada a disponibilidade do constructo a ser tratado no período de tempo determinado e, em caso positivo, realizados os registros da execução da atividade, com o consumo e produção de novo constructo.

Tecnicamente no sistema de simulação, isto acontece respetivamente através de inserção na entidade `construct_events`, populando os dados da identificação do constructo que foi produto da atividade, a atividade executada e a timestamp de execução e inserção na entidade `construct` populando os dados de identificação do constructo, respectiva tipologia, identificação do constructo que foi decomposto (ou tratado) para composição do novo, a instância do processo, e o peso do novo constructo (neste caso, o peso do produto é igual ao do requisito).

Em caso de a atividade se tratar da disjunção de um constructo em diversas partes, após verificação positiva de disponibilidade do constructo tido como requisito à atividade

no indicador temporal determinado, são recuperados, dos parâmetros da simulação, os valores mínimo e máximo admitidos para constructos que sejam produtos da atividade correspondente. Estes dados são utilizados na parametrização de uma nova distribuição de probabilidade que é usada na determinação do peso de cada constructo que seja produto da atividade na simulação.

Para fins de registro das atividades no banco de dados da simulação, para cada constructo produzido na atividade são realizadas duas inserções: uma na entidade `construct_events`, com a população da identificação do constructo, atividade e timestamp, e outra na entidade `construct`, com a população da identidade do constructo produzido, respectiva tipologia, identificação do constructo que foi decomposto, para a composição dos novos, a instância do processo, e o respectivo peso do constructo produzido.

Em cada criação de novo constructo que é criado a partir da decomposição de um maior, é realizada uma atualização na entidade `construct`, a fim de atualizar o peso do constructo que foi decomposto, diminuindo do valor o peso estipulado para o novo constructo.

Para atividades de junção de constructos, são recuperados, dos parâmetros da simulação, os valores mínimo e máximo admitidos para constructos que sejam produtos da atividade correspondente. Com base nestes atributos, é instanciada uma nova distribuição de probabilidade, responsável por determinar o peso do constructo que será gerado na simulação.

Até que seja completo o peso determinado para o constructo simulado, são adicionados constructos menores, então disponíveis, à composição do novo constructo, produto da atividade.

O registro da execução da atividade é feito através da atualização da entidade `construct` à cada incorporação de determinado constructo no produto da atividade, de forma a informar que o respectivo item, então faz parte de outro item. Além disto, são realizadas inserções nas tabelas `construct_events` e `construct`, respectivamente para registro da identificação do constructo produzido, a atividade que o produziu e registro temporal da ocorrência da atividade, e registro o constructo gerado, com identificação, respectivo tipo, identificação da instância do processo, e o respectivo peso do constructo produzido na simulação, que é atribuído com base na soma dos pesos dos constructos que o compõem.

Através desta metodologia de sistema de informação, é possível manter os relacionamentos das atividades com os constructos que foram requisitos e produtos das execuções. Consequentemente, é mantida a coerência na identificação da composição dos constructos por outros constructos menores e a identificação das atividades que os originaram.

O produto da execução do software proposto é um esquema de banco de dados relacional de onde seja possibilitada a extração de um log de eventos oriundo da simulação dos eventos, porém, livre de quaisquer vícios referentes à granularidade dos dados e níveis de abstração das atividades.

Tabela 10: Descrição da entidade destinada ao armazenamento dos constructos simulados.

Entidade construct	
Nome do Campo	Descrição do Campo
id	Identificador único do constructo simulado.
type	Nomenclatura do tipo atribuído ao respectivo constructo.
decompoundid	Identificador do constructo o qual tenha sido fisicamente fragmentado para a concretização do respectivo constructo.
parentid	Identificador do constructo o qual o respectivo compôs, no processo de montagem.
highestid	Identificador do constructo original, conforme o log de eventos original.

Tabela 11: Descrição da entidade destinada ao armazenamento das atividades simuladas para concretização de cada constructo simulado.

Entidade construct_event	
Nome do Campo	Descrição do Campo
id	Identificador único do constructo simulado.
activity	Identificador da atividade simulada.
timestamp	Identificador da posição temporal da ocorrência da respectiva atividade para montagem do respectivo constructo.

A finalidade é para a constituição sintética de um novo log de eventos, então isento de anomalias referentes à granularidade e abstração dos dados contidos nele e compatível com as restrições temporais descritas no log de eventos original.

Com a simulação da execução das instâncias do processo, o sistema alimenta continuamente um esquema de banco de dados contendo entidades adicionais, enquadrando dados que possibilitam a extração de um novo log de eventos, então com o nível de granularidade das atividades adequado ao contexto das questões cuja atividade de mineração de processos tenha o objetivo de responder.

A intenção desta forma de organizar os dados é que, a partir da população destas duas entidades dentro do esquema relacional proposto, uma consulta escrita em linguagem SQL seja capaz de extrair um log de eventos na conformação desejada. Desta forma é alcançada a flexibilidade na determinação do nível de abstração em que o log é apresentado, onde este se encontra livre das problemáticas referentes à qualidade dos dados, conforme tratadas neste estudo.

Finalmente, através a geração de um novo log de eventos, então adequado à tarefa de mineração de processos a ser desempenhada, estudos inerentes à prática podem ser conduzidos, tendo o log de eventos sintético como fonte de dados.

O diagrama ER exposto na figura 16 formaliza o esquema de banco de dados utilizado pelo sistema em sua totalidade, complementando o que havia sido exposto de forma par-

cial na figura 14.

Para fins de armazenamento do log de eventos sintético, o esquema de banco de dados proposto disponibiliza a entidade `synthetic_event_log`. Contudo, nota-se que sua descrição tem aplicação específica no estudo de caso abordado no capítulo 3. Ela é descrita de forma apropriada no capítulo correspondente.

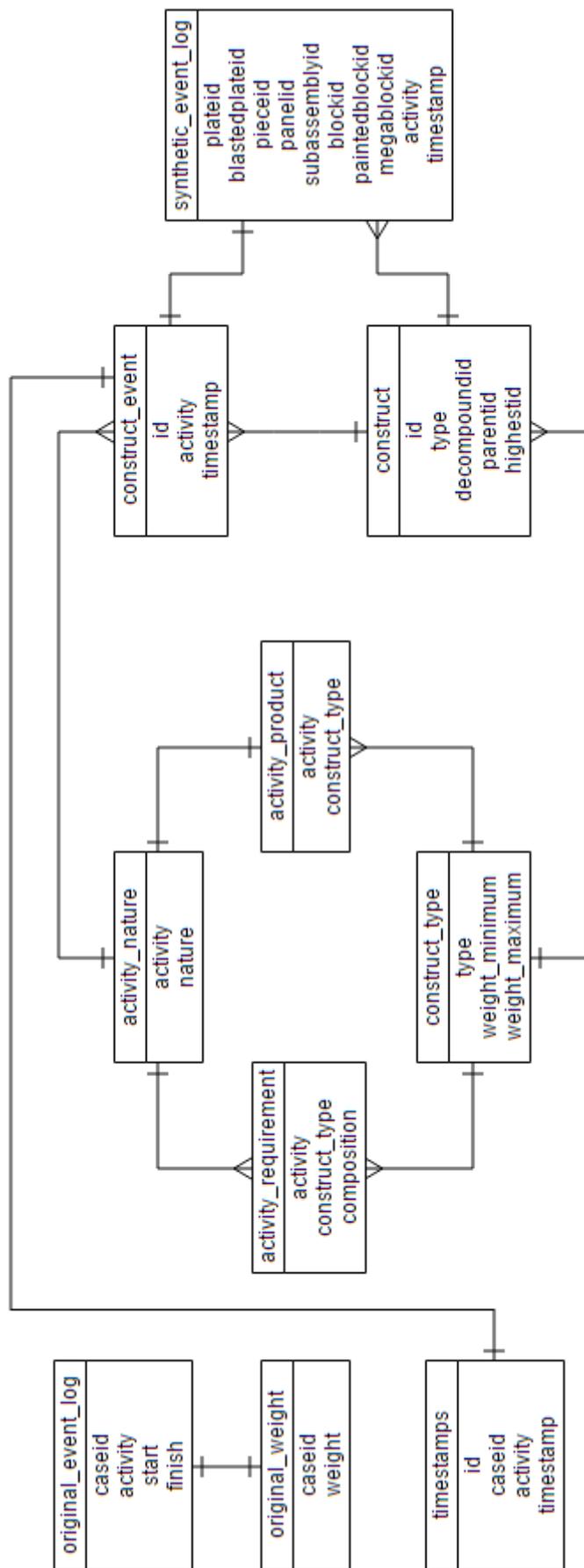


Figura 16: Modelo ER Final

As tecnologias envolvidas na implementação do software de simulação descrito foram:

1. linguagem de programação Java;
2. biblioteca Apache Commons Math para a linguagem Java;
3. interface com banco de dados SQLite para a linguagem Java;
4. um banco de dados SQLite, esquematizado conforme os diagramas das figuras 14 e 16.

O projeto da camada de aplicação do software propôs uma arquitetura orientada a serviços, onde foram mapeados conceitos dos executores de atividades no perímetro de construção em artefatos conhecidos como classes, através da orientação a objetos. A especificação formal da camada de aplicação do software se encontra fora do escopo deste estudo. Contudo, um fluxograma é apresentado na figura 17 com o intuito de facilitar o entendimento sobre o processo de geração do log de eventos sintético, conforme descrito nesta seção.

A metodologia foi aplicada em um estudo de caso onde foram analisados dados oriundos do processo de construção de blocos em um estaleiro brasileiro. Ao final do estudo foi possível comparar os modelos descobertos antes e depois da aplicação da metodologia proposta e os resultados são apresentados na próxima seção.

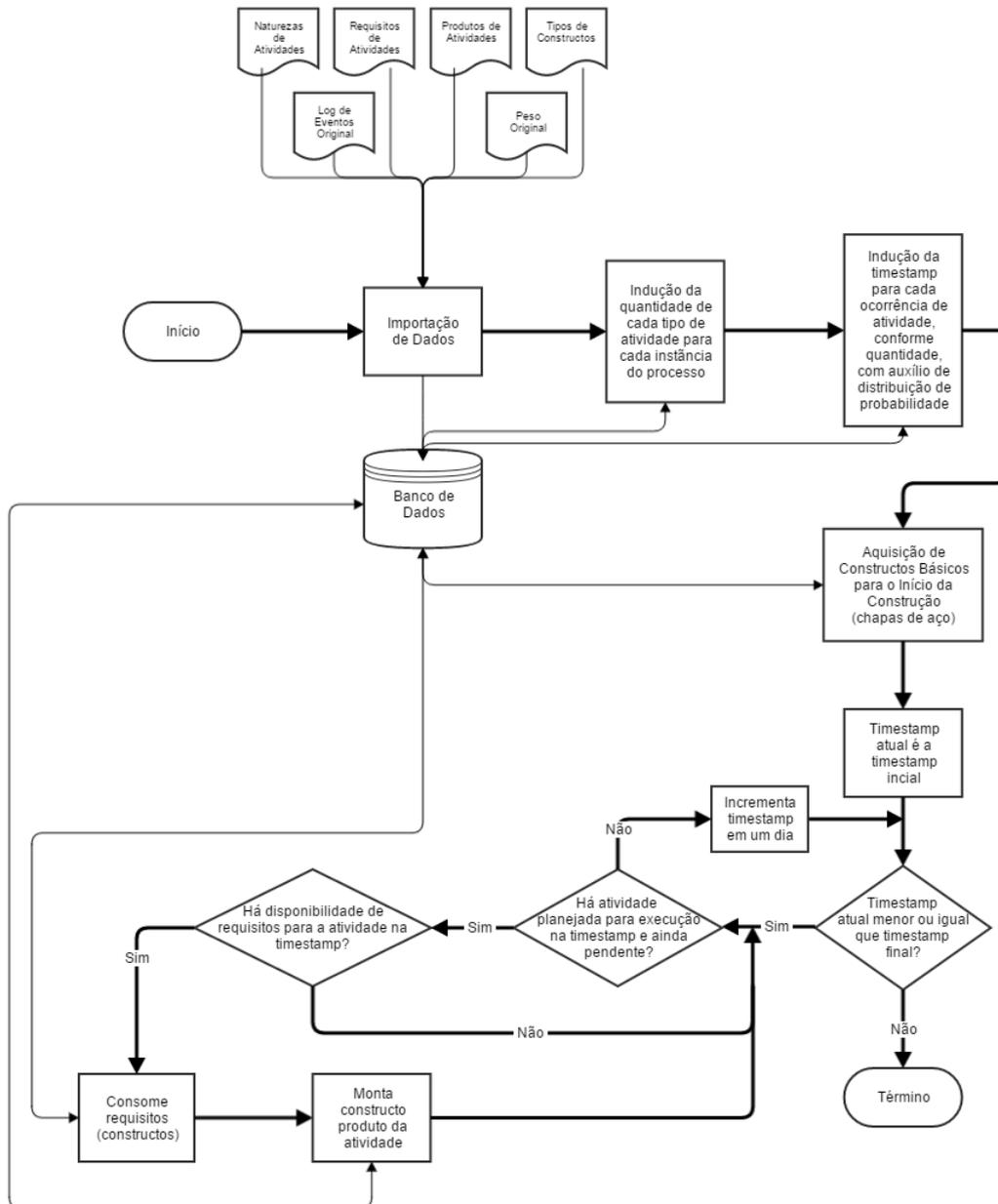


Figura 17: Fluxograma do processo de geração do log de eventos sintético.

3 ESTUDO DE CASO

Este capítulo documenta o estudo de caso realizado no âmbito deste trabalho, com o intuito de validar a proposta para solução do problema de pesquisa apresentada.

Trata-se da aplicação da metodologia de geração de log de eventos sintético, no contexto do processo de construção de blocos em um estaleiro da indústria naval brasileira.

A empresa mantenedora do estaleiro cujo processo de construção de blocos foi analisado neste estudo, concordou em fornecer um modelo empírico de seu processo e um conjunto de dados inerente que corresponde às atividades ocorridas na construção de uma plataforma de exploração petrolífera.

O estudo de caso foi dividido em etapas:

1. análise do modelo de processo atual;
2. análise da semântica dos dados contidos no log de eventos;
3. observação dos resultados de tarefas de mineração de processos sobre o log de eventos original;
4. análise sobre a adequação do estudo de caso à problemática da qualidade dos dados do log de eventos e o contexto oculto do domínio;
5. descrição da parametrização e utilização do sistema de geração de dados sintéticos;
6. análise da semântica dos dados contidos no log de eventos sintético produzido pelo sistema de geração;
7. observação dos resultados de tarefas de mineração de processos sobre o conjunto de dados sintético.

Para o detalhamento de cada uma destas etapas, foi destinada uma seção distinta no presente capítulo.

3.1 O Modelo do Processo de Construção Atual

Diante da inexistência de um modelo formal do processo de construção de blocos no estaleiro, foi obtida uma representação informal junto aos colaboradores técnicos responsáveis pelo planejamento das atividades no perímetro de construção. A representação é ilustrada na figura 2, disposta na seção 2.1.

É importante notar que o modelo idealizado pelo estaleiro não é composto por notações formais de controles de fluxo, como redes de Petri, EPC, YAWL, redes de fluxo ou BPMN.

Desta forma, é dificultado o entendimento do modelo do processo sem acompanhamento de profissionais envolvidos na sua concepção. Este problema de semântica, que poderia ser eliminado com a utilização de notações formais na modelagem, em conjunto com a conformação dos dados de entrada na pesquisa, sem indicação de metadados com a significância de cada campo, são essencialmente importantes para a concretização da problemática de qualidade dos dados abordada neste estudo.

A partir do modelo empírico apresentado na figura 2, é possível inferir 10 etapas distintas do processo, sendo elas: almoxarifado, shot (shotblasting), processamento, painelização, submontagem, montagem, acabamento avançado, pintura e edificação.

Contudo, analisando a notação gráfica utilizada para relacionar as etapas do processo neste modelo empírico, é impossível determinar, com certidão absoluta, os tipos de controles de fluxo aplicados nas relações.

Ao contrário de notações formais para representação de controles de fluxo (vide subseção 2.2.5), o modelo apresentado não é capaz de ilustrar, especificamente, características de paralelismo, ordenação ou exclusão mútua entre as atividades do processo, por exemplo.

Desta forma, analisando-se visualmente o modelo informal apresentado na figura 2, entende-se que o processo começa no almoxarifado, sendo posposto pela atividade de “shot” e processamento ou, diretamente, por processamento. Após o processamento, determinados constructos seguem à painelização, enquanto outros seguem à submontagem. Da painelização e submontagem, os constructos seguem à montagem e, subsequentemente, ao acabamento avançado, pintura e edificação, onde o processo de construção do bloco é terminado.

Notam-se múltiplas representações gráficas para atividade de processamento e, isto é devido à natureza dos constructos que são supostamente relacionados como entrada em cada um das variantes (chapa, perfil, painel, plasma, perfil-reforço e perfil-painel).

Contudo, para atividades como submontagem, montagem, pintura e edificação, torna-se impossível determinar a natureza e os produtos de cada uma delas sem conhecimento aprofundado sobre as regras de negócio do estaleiro específico.

Desta forma, através da revisão bibliográfica descrita na subseção 2.2.1, foi possível

definir com maior certidão o escopo real da realização das atividades descritas no modelo. A tabela 12 consolida este entendimento aprimorado.

Tabela 12: Descrição das atividades do processo de construção de blocos

Atividade	Descrição
Almoxarifado	Armazenamento dos constructos e matéria-prima utilizados na construção de blocos.
Shotblasting	Jateamento de chapas de aço com material abrasivo para remoção de ferrugem e outras impurezas.
Processamento	Corte de chapas de aço que já passaram pelo jateamento para fabricação de peças.
Painelização	Construção de painéis a partir de chapas de aço jateadas e peças fabricadas.
Submontagem	Construção de subblocos a partir de peças fabricadas e chapas de aço jateadas.
Montagem	Construção de blocos a partir de subblocos, painéis, peças individuais e, excepcionalmente, chapas de aço jateadas.
Acabamento Avançado	Também conhecido como pré-edificação, trata do acoplamento de acabamento avançado na estrutura dos blocos.
Pintura	Pintura dos blocos construídos.
Edificação	Içamento do bloco para edificação do megabloco.

3.2 Análise Semântica do Log de Eventos Original

O conjunto de dados disponibilizado para a pesquisa tem a estrutura conforme exposto na tabela 13 e um trecho de seu conteúdo é disposto na tabela 14 para fins de amostragem.

Tabela 13: Descrição da entidade destinada ao armazenamento do conjunto de dados originalmente disponibilizado ao estudo

Nome do Campo	Descrição do Campo
Bloco	Bloco em construção, ou instância do processo
JATI	Semana de início do jateamento para as peças do bloco
JATT	Semana de término do jateamento para as peças do bloco
PRCI	Semana de início do processamento para as peças do bloco
PRCT	Semana de término do processamento para as peças do bloco
PNLI	Semana de início da painelização para as peças do bloco
PNLT	Semana de término da painelização para as peças do bloco
SMNI	Semana de início da submontagem do bloco
SMNT	Semana de término da submontagem do bloco
MONI	Semana de início da montagem do bloco
MONT	Semana de término da montagem do bloco
PNTI	Semana de início da pintura do bloco
PNTT	Semana de término da pintura do bloco
EDFI	Semana de início da edificação do bloco
EDFT	Semana de término da edificação do bloco

Tabela 14: Amostra de dados do processo, em seu formato original.

Bloco	JATI	JATT	PRCI	PRCT	PNLI	PNLT	SMNI	SMNT	MONI	MONT	PNTI	PNTT	EDFI	EDFT
B01C	1133	1133	1137	1138	1139	1141	1138	1140	1139	1151	1201	1201	1244	1308
B02C	1130	1130	1134	1135	1137	1138	1137	1139	1138	1147	1148	1148	1235	1409
B03C	1125	1125	1130	1131	1132	1134	1132	1134	1133	1143	1146	1146	1244	1325
B04C	1131	1131	1138	1139	1137	1139	1137	1139	1138	1148	1150	1150	1244	1330
B05C	1134	1134	1137	1138	1137	1139	1139	1141	1140	1150	1151	1151	1239	1330
B06C	1136	1136	1140	1142	1144	1146	1143	1145	1144	1204	1208	1208	1244	1247
B07C	1134	1134	1141	1142	1144	1146	1144	1146	1145	1204	1205	1205	1244	1327
B08C	1144	1144	1144	1145	1145	1147	1145	1147	1146	1206	1210	1210	1248	1308
B09C	1201	1201	1151	1210	1152	1201	1216	1217	1220	1234	1240	1240	1244	1334
B10C	1144	1144	1144	1145	1145	1147	1145	1147	1146	1204	1206	1206	1248	1308
B11C	1145	1145	1143	1145	1149	1151	1148	1150	1148	1209	1212	1212	1237	1334
B12C	1149	1149	1147	1149	1149	1151	1149	1151	1147	1208	1214	1214	1248	1308
B13C	1150	1150	1149	1151	1152	1202	1152	1201	1202	1213	1221	1221	1240	1334
B14C	1202	1202	1201	1203	1201	1203	1203	1204	1204	1214	1219	1219	1302	1327
B15C	1201	1201	1205	1208	1210	1212	1210	1212	1222	1229	1244	1244	1248	1339
B16C	1201	1201	1203	1208	1203	1205	1204	1205	1204	1216	1246	1246	1302	1326
B17C	1204	1204	1203	1204	1203	1204	1204	1205	1205	1216	1217	1217	1249	1333
B18C	1204	1204	1202	1203	1203	1204	1205	1206	1205	1217	1227	1227	1302	1326
T01P	1142	1142	1141	1142	1143	1144	1144	1145	1145	1202	1314	1314	1313	1337
T01S	1144	1144	1143	1144	1144	1145	1143	1144	1145	1151	1312	1312	1311	1334
T02P	1149	1149	1148	1149	1149	1150	1150	1151	1151	1205	1310	1310	1309	1334
T02S	1147	1147	1146	1147	1148	1149	1149	1150	1150	1205	1304	1304	1303	1339
T03P	1147	1147	1148	1149	1149	1150	1150	1151	1151	1204	1313	1313	1312	1334
T03S	1142	1142	1141	1142	1142	1143	1144	1145	1145	1202	1312	1312	1311	1339
T04P	1144	1144	1143	1144	1143	1144	1144	1145	1145	1204	1319	1319	1318	1342
T04S	1149	1149	1148	1149	1149	1150	1150	1151	1150	1206	1318	1318	1317	1351
T05P	1147	1147	1146	1147	1148	1149	1149	1150	1150	1205	1322	1322	1321	1343
T05S	1147	1147	1148	1149	1149	1150	1150	1151	1151	1204	1318	1318	1317	1338

A tabela 14 apresenta registros compostos por 16 colunas, onde são expostas a identificação do bloco construído, seu peso e as posições temporais de início e fim para as atividades de jateamento, processamento, painelização, submontagem, montagem, pintura e edificação para cada bloco, respectivamente, denotando as semanas do projeto pelas quais as atividades ocorreram.

Outrossim, é notório, em relação ao modelo empírico proposto pelo estaleiro, que estão ausentes no conjunto de dados as informações sobre as supostas atividades denominadas como almoxarifado, acabamento avançado e as seis variantes para a atividade de processamento, sendo esta abstraída em nomenclatura única.

No total, o conjunto de dados obtido originalmente conteve informações referentes à execução de 276 instâncias do processo de construção de blocos no estaleiro. Contudo, o formato disponibilizado não é compatível com aquele tido como requisito para entrada em ferramentas de mineração de processos (vide tabela 3, seção 2.2.4).

Outrossim, torna-se necessária a realização de uma tarefa de pré-processamento que objetive a extração dos dados inerentes a composição de um log de eventos no formato adequado.

Para este propósito, o conjunto de dados original foi importado ao esquema de banco de dados SQLite utilizado no estudo, sendo, então, armazenado sob entidade descrita de forma similar ao disposto na tabela 13.

Conforme exposto anteriormente, a utilização de um esquema relacional de banco de dados para armazenamento do log de eventos tem o objetivo de prover flexibilidade na seleção, formatação e exportação dos dados resultantes de pesquisas. Nesta situação, a abordagem facilita a extração do log de eventos a partir de conjunto de dados original, para conformação adequada.

A tabela 15 apresenta a estrutura da entidade utilizada no armazenamento do log de eventos inicialmente extraído do conjunto de dados original. O produto desta extração foi, a partir de então, utilizado como log de eventos base do estudo e referenciado como tal.

Tabela 15: Descrição da entidade original_event_log, destinada ao armazenamento do log de eventos original

Entidade original_event_log	
Nome do Campo	Descrição do Campo
caseid	Identificação do bloco correspondente à atividade executada
activity	Atividade executada
start	Semana de início da atividade correspondente
finish	Semana de término da atividade correspondente

Sob a conformação descrita na tabela 15, foi possível isolar cada atividade em um registro único, de forma a obedecer o formato proposto por (AALST, 2015) (tabela 3).

3.3 Observação dos Resultados sobre o Log de Eventos Original

Embora tenha sido possível extrair um log de eventos do conjunto de dados original, isto não implica em que o log extraído esteja imediatamente livre de problemáticas referentes à qualidade dos dados coletados.

Esta seção visa apresentar o resultado errático que pôde ser verificado quando da realização de tarefas de mineração de processos sobre o log de eventos original, que foi extraído do conjunto de dados apresentado para o estudo.

Os algoritmos para descoberta de processos existentes propõem abordagens eficazes para determinar relações de ordenação entre as atividades de processos sob diferentes metodologias (AALST, 2011a). Contudo, apesar da existência de diversas propostas deste tipo na bibliografia inerente, há espaço para o tratamento de cenários ainda não previstos por elas, como é o caso da problemática dos dados identificada nesta pesquisa.

A análise do log de eventos utilizado como base do estudo possibilitou o entendimento de que as instâncias nele contidas descrevem situações implícitas, que são compreendidas somente com entendimento aprofundado do processo sendo analisado.

Desta forma, as abordagens atualmente existentes para descoberta de processos e mineração de dados são incapazes de prever o contexto oculto do processo que, indiretamente, gerou log de eventos. Consequentemente, a inferência de relações entre atividades, no que se refere a ordenação e causalidade, são apresentadas em formato errôneo, ou inadequado ao contexto do processo real.

Em outras palavras, embora o processamento ocorra de acordo com as premissas da utilização de determinados algoritmos, o resultado apresentado se torna insatisfatório para fins de representação do processo factual.

Para fins de experimentação sobre o comportamento geral de algoritmos para descoberta de processos, foi observada a utilização do algoritmo Disco Miner (GUNTHER; ROZINAT, 2012), integrado a ferramenta de software de mesmo nome.

Foram analisados os modelos gráficos inferidos para o processo de construção de blocos a partir da entrada do log de eventos na ferramenta Disco (GUNTHER; ROZINAT, 2012). O resultado preliminar apresentou o modelo ilustrado na figura 18.

No modelo, os nós denotam atividades, enquanto as arestas denotam as relações de causalidade entre elas, conforme descoberta efetuada pelo algoritmo. A numeração atribuída para as arestas significam o número de instâncias do processo que utilizaram o respectivo fluxo, quando em execução. De forma semelhante, as atividades também recebem esta informação. Em ambos casos, seus valores estão vinculados ao nível de abstração selecionado pelo usuário da ferramenta.

Através da comparação visual entre o modelo empiricamente descrito por profissionais técnicos (figura 2) e aquele descoberto pela ferramenta de software Disco pela análise automatizada do log de eventos (figura 18), foi notória a discrepância entre os dois mod-

elos.

De fato, o log de eventos original não registra todas as atividades apresentadas no modelo empírico. Contudo, em comparação ao modelo descrito pelos profissionais do estaleiro, foi possível inferir uma quantidade maior de relações entre as atividades registradas e retratá-las no modelo descoberto com a mineração de processos, com base no log de eventos do processo.

Além disto, a complexidade demasiada no modelo descoberto causou estranheza por denotar relações de ordenação e causalidade entre as atividades que não são usuais ou que são impossíveis de acontecer, de acordo com profissionais técnicos da construção naval.

Por exemplo, é possível que uma instância do processo de construção de blocos tenha o início com as atividades de jateamento e processamento na mesma semana. Porém, não é usual que o processamento aconteça antes do jateamento, pois, teoricamente, tarefas de processamento recebem chapas de aço já jateadas.

Outra irregularidade inferida no modelo descoberto é uma suposta exclusão mútua entre a atividade de submontagem e montagem, enquanto, na verdade, a submontagem produz sub-blocos que, por sua vez, é um dos tipos de constructo geralmente consumidos na montagem de blocos.

Ao final do processo, de acordo com o modelo descoberto, é denotada a possibilidade do término do processo nas atividades de edificação ou pintura. Contudo, somente a primeira possibilidade é verdadeira. Na realidade, atividades de edificação recebem blocos pintados para incorporação em megablocos.

Outrossim, esta seção expõe a problemática de qualidade do log de eventos estudado, no que se refere ao contexto implícito do processo e o que os dados, de fato, significam. Este detalhamento se faz necessário para o entendimento sobre como a problemática de dados apresentada na justificativa do estudo se faz presente no log de eventos estudado.

3.4 O Contexto Oculto das Atividades do Processo em Relação ao Log de Eventos

Com base no modelo apresentado na descoberta do processo de construção de blocos, foram identificadas falhas ao representar as relações entre as atividades expostas. Conforme descrito anteriormente, isto acontece devido à existência de um contexto do domínio de negócio da construção naval que é desconhecido pela ferramenta de análise de log de eventos.

Consequentemente, embora hajam dados que retratam a execução das atividades sob determinada perspectiva, para fins da descoberta do processo que os rege, eles se tornam inadequados devido à alta abstração dos atributos. Esta seção expõe a justificativa pela qual o conjunto de dados analisado se encaixa ao contexto da problemática abordada e da solução proposta.

3.4.1 Sobreposição de Atividades da Mesma Instância do Processo

Analisando os indicadores temporais de início e término das atividades descritas no log de eventos, foi possível observar que além de instâncias diferentes do processo ocorrerem em paralelo, atividades pertencentes à mesma instâncias do processo compartilham o mesmo período de tempo de execução.

Esta situação gera um problema, baseado no viés representacional de algumas notações de modelagem de processos que não são capazes de denotar concorrência de forma explícita, como a de mapas de processo, utilizada pelo algoritmo Fuzzy Miner (GUNTHER; VAN DER AALST, 2007).

Conceitualmente, um modelo de processo baseado em mapeamento não admite que dois nós compartilhem o mesmo nome. Desta forma, quando da necessidade de denotar paralelismo, o algoritmo cria relações de ordenação que podem não ser reais (MACIEL; FILHO, 2015).

O log de eventos disposto na tabela 16 apresenta, para fins de exemplificação da problemática abordada, uma instância de processo hipotética. Nela ocorre a sobreposição parcial entre as atividades de jateamento, executada do tempo 10 ao tempo 20, e processamento, executada do tempo 15 ao 25, de forma que ambas atividades são executadas concorrentemente do tempo 15 ao tempo 20.

Com a utilização do algoritmo Disco Miner sobre o conjunto de dados descrito, fora gerado o modelo disposto na figura 19.

Tabela 16: Sobreposição parcial de atividades.

Instância	Atividade	Início	Término
xpt0	jateamento	10	20
xpt0	processamento	15	25

Com base na análise do log de eventos e no respectivo modelo de processo, foi possível observar que:

- No log de eventos, o processo começa com jateamento, enquanto o modelo permite que comece diretamente com a ocorrência de jateamento e processamento de forma concorrente.
- No log de eventos, após o jateamento, ocorre a concorrência entre esta e processamento, enquanto, no modelo, esta concorrência ocorre logo após o início do processo.
- No log de eventos, após a concorrência entre jateamento e processamento, ocorre processamento de forma individual, e, só então, o processo termina. No modelo gerado, após a concorrência entre jateamento e processamento, o processo pode terminar.

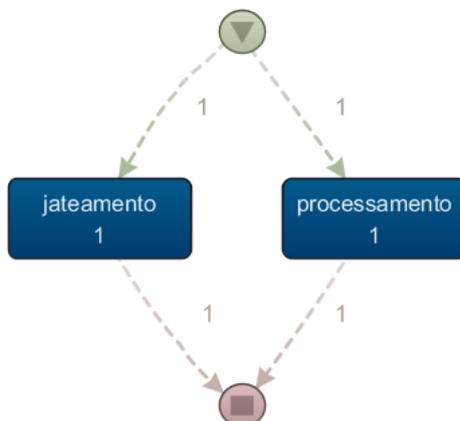


Figura 19: Exemplo de modelagem errônea de processo cujo log de eventos denota sobreposição de atividades.

Desta forma, fica evidente a atuação equivocada do algoritmo ao inferir relações de causalidade entre atividades que se sobrepõem parcialmente na mesma instância do processo.

3.4.2 Registro Temporal de Baixa Granularidade e Imprecisão de Ordenação

A situação de baixa granularidade, evidente nos registros temporais das atividades constantes no log de eventos em estudo conforme exposto na seção 1.1, vem a causar imprecisão na indução de relações de ordenação entre atividades do processo, quando da análise por algoritmos de descoberta de processos.

Este tipo de ocorrência é exemplificada por uma situação hipotética, descrita pelos logs de eventos dispostos nas tabelas 17 e 18, onde ambos expressam a situação em que, para determinada instância do processo, as atividades de jateamento e processamento ocorrem em concorrência no tempo 10.

A diferença entre eles ocorre pela ordem em que aparecem, fisicamente, em cada um dos logs: no primeiro, o evento de processamento aparece no registro imediatamente posterior ao evento de jateamento, enquanto, no segundo, justamente o inverso.

Tabela 17: Exemplo de log de eventos com atividades ocorrendo em paralelo.

Instância	Atividade	Tempo de Execução
xpt0	jateamento	10
xpt0	processamento	10

A figura 20 mostra os modelos de processo gerados a partir da utilização da ferramenta Disco sobre os conjuntos de dados.

Tabela 18: Exemplo de log de eventos com atividades ocorrendo em paralelo.

Instância	Atividade	Tempo de Execução
xpt0	processamento	10
xpt0	jateamento	10

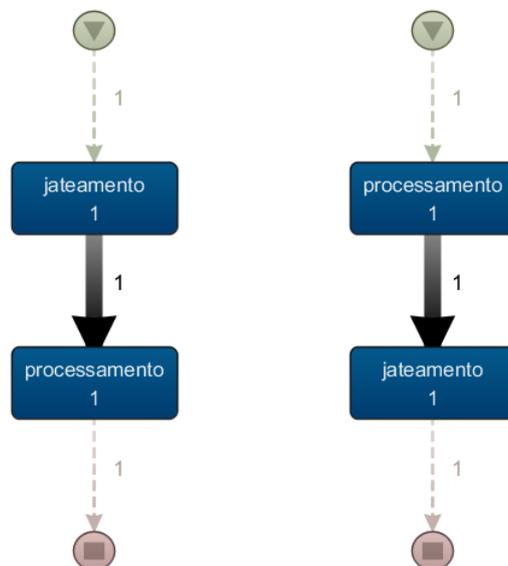


Figura 20: Exemplo de modelagem de processos onde ordens de execução são registradas de forma confusa.

Com a comparação visual entre os dois modelos, fica evidente que, embora os logs de eventos tenham registrado que os eventos ocorreram em concorrência no igual período de tempo, um algoritmo de descoberta de processos pode induzir a relação de ordenação entre os eventos com base na ordem em que as atividades estavam dispostas fisicamente no log de eventos.

Conforme descrito anteriormente, o log de eventos original, utilizado como entrada neste estudo, tem a unidade de tempo apresentada sob a forma de semana, o que causa imprecisão quando da inferência de relações entre atividades que ocorrem em períodos mais granulares, como dias, turnos ou horas.

3.4.3 Fragmentação Oculta das Atividades

Em análise empírica do conjunto de dados, junto a representantes técnicos com conhecimento de processos do domínio da indústria naval brasileira, fora identificado que o log de eventos analisado contém uma problemática referente à ocultação da fragmentação de atividades.

Na prática, isto significa que embora o conjunto de dados registre um indicador de tempo para início e término de todas as atividades em todas as instâncias do processo, falha em registrar que, dentro do período de tempo determinado para a execução de cada atividade, na verdade, elas foram executadas múltiplas vezes com durações menores e desconhecidas dentro do período indicado, ao invés de uma única vez com a duração de todo o período.

A tabela 16 exemplifica um log de eventos nesta situação onde, em primeiro momento, é explícito que a atividade de jateamento ocorreu entre os tempos 10 e 20 e a atividade de processamento ocorreu entre os tempos 15 e 25, para determinada instância do processo. A figura 1, exposta na seção 1.1, também auxilia no entendimento da situação descrita.

A seção 3.5 provê detalhamento sobre o processo de simulação de processos capaz da constituição do log de eventos sintético que esteja livre das problemáticas de qualidade apresentadas.

3.5 Aplicação da Metodologia de Simulação do Processos

Na aplicação da metodologia de simulação de processos e constituição de logs de eventos sintéticos apresentada nesta dissertação, foi necessária a parametrização do software com dados estatísticos do próprio processo de construção, conforme descrito na seção 2.3.

Aqui, são descritos estes parâmetros, que foram coletados junto à profissionais com conhecimento do domínio de negócio da construção naval, conforme os padrões da indústria brasileira.

Com vistas em amenizar as problemáticas de qualidade do log de eventos original, conforme descritas nas subseções 3.4.1, 3.4.2 e 3.4.3, foi incorporado ao processo de

simulação do log de eventos sintético, especificamente na determinação dos registros temporais para execução das atividades, uma conversão do indicador de semanas para dias, de modo a aumentar a granularidade do espectro de possibilidades para geração de valores aleatórios.

Nos testes realizados, esta forma de tratamento de dados temporais mostrou potencial para diminuição da ocorrência de sobreposições de atividades da mesma instância do processo, bem como a consequente imprecisão de ordenação de tais atividades.

As tabelas 19, 20, 21 e 22 exibem os conjuntos de dados utilizados na parametrização do software desenvolvido no âmbito deste trabalho para sintetização de logs de eventos.

Tabela 19: Dados inseridos na entidade `construct_type` para parametrização parcial do algoritmo.

construct_type		
type	weight_minimum	weight_maximum
chapa	0.5	6.0
chapa_jateada	0.5	6.0
peca	0.2	1.5
painel	5.0	25.0
sub_bloco	5.0	25.0
bloco	10.0	300.0
bloco_pintado	10.0	300.0
mega_bloco	900.0	5800.0

Tabela 20: Dados inseridos na entidade `activity_nature` para parametrização parcial do algoritmo.

activity_nature	
activity	nature
jateamento	Tratamento
processamento	Disjunção
painelizacao	Junção
submontagem	Junção
montagem	Junção
pintura	Tratamento
edificacao	Tratamento

Após aplicação da metodologia apresentada nesta dissertação, através da execução do software de simulação projetado conforme disposto na seção 2.3, pôde ser verificada a população das entidades `construct` e `construct_event`.

Desta forma, através da execução de uma consulta escrita em linguagem SQL, foi possível obter o log de eventos com as informações em nível de abstração e granularidade adequados. O código-fonte da consulta está disposto na figura 21. A descrição da entidade `synthetic_event_log` é feita na tabela 23.

Tabela 21: Dados inseridos na entidade `activity_requirement` para parametrização parcial do algoritmo.

activity_requirement		
activity	construct_type	composition
jateamento	chapa	1.00
processamento	chapa_jateada	1.00
painelizacao	chapa_jateada	0.66
painelizacao	peca	0.34
submontagem	chapa_jateada	0.33
submontagem	peca	0.67
montagem	chapa_jateada	0.12
montagem	peca	0.22
montagem	painel	0.33
montagem	sub_bloco	0.33
pintura	bloco	1.00
edificacao	bloco_pintado	1.00

Tabela 22: Dados inseridos na entidade `activity_product` para parametrização parcial do algoritmo.

activity_product	
activity	construct_type
jateamento	chapa_jateada
processamento	peca
painelizacao	painel
submontagem	sub_bloco
montagem	bloco
pintura	bloco_pintado
edificacao	mega_bloco

Tabela 23: Descrição da entidade `synthetic_event_log`

Nome do Campo	Descrição do Campo
plateid	Indicação de chapa de aço.
bastedplateid	Indicação de chapa de aço jateada.
pieceid	Indicação de peça.
panelid	Indicação de painel.
subassemblyid	Indicação de subbloco.
blockid	Indicação de bloco.
paintedblockid	Indicação de bloco pintado.
megablockid	Indicação de megabloco.
activity	Indicação da atividade.
timestamp	Indicação temporal de ocorrência da atividade.

```

1  select  plate.id          plateid      ,
2          blastedplate.id  blastedplateid ,
3          piece.id         pieceid       ,
4          panel.id         panelid        ,
5          subassembly.id   subassemblyid ,
6          block.id         blockid        ,
7          paintedblock.id  paintedblockid ,
8          megablock.id     megablockid    ,
9          events.activity   activity       ,
10         events.timestamp  timestamp
11 from    (select * from construct where type = 'plate'      ) plate
12 left join (select * from construct where type = 'blasted_plate' ) blastedplate
13         on ( plate.id in ( blastedplate.decompoundid ) )
14 left join (select * from construct where type = 'piece'      ) piece
15         on ( blastedplate.id in ( piece.decompoundid ) )
16 left join (select * from construct where type = 'panel'      ) panel
17         on ( panel.id in ( blastedplate.parentid,
18                          piece.parentid
19                          ) )
19 left join (select * from construct where type = 'subassembly' ) subassembly
20         on ( subassembly.id in ( blastedplate.parentid,
21                                piece.parentid
22                                ) )
22 left join (select * from construct where type = 'block'      ) block
23         on ( block.id in ( blastedplate.parentid,
24                          panel.parentid,
25                          subassembly.parentid
26                          ) )
26 left join (select * from construct where type = 'painted_block' ) paintedblock
27         on ( block.id in ( paintedblock.decompoundid ) )
28 left join (select * from construct where type = 'megablock'   ) megablock
29         on ( paintedblock.id in ( megablock.decompoundid ) )
30 join construct_events events
31     on ( events.constructid in (
32         plate.id          ,
33         blastedplate.id  ,
34         piece.id         ,
35         panel.id         ,
36         subassembly.id   ,
37         block.id         ,
38         paintedblock.id  ,
39         megablock.id     ) )
39 where block.id is not null;

```

Figura 21: Consulta em linguagem SQL utilizada para extrair o log de eventos sintético a partir do conjunto de dados de eventos proveniente da simulação com o software implementado.

A estrutura da entidade `synthetic_event_log` denota a possibilidade para identificação dos constructos envolvidos direta e indiretamente na mesma atividade.

Por exemplo, ao manter a identificação das peças em um registro que contém, também, a identificação das chapas de aço que foram cortadas e vieram a fazer parte de determinado bloco. Desta forma, é alcançada flexibilidade no rastreamento dos constructos e das atividades executadas, de forma que, quando da realização de estudos de mineração de processos, qualquer um dos níveis disponíveis pode ser definido como identificação da instância do processo analisado.

3.6 Análise do Log de Eventos Sintético

Para fins de análise do log de eventos sintético, esta seção apresenta a tabela 24 com os dados do processo de produção conforme exposto no log de eventos original para o bloco identificado como L05C, enquanto a tabela 25 ilustra os dados para o processo de construção do mesmo bloco, porém, conforme disposto no log de eventos sintético, gerado pelo processo de simulação concebido neste estudo.

Tabela 24: Log de Eventos Original para o Bloco L05C

caseid	activity	start	finish
L01C	jateamento	1123	1124
L01C	processamento	1124	1124
L01C	painelizacao	1124	1125
L01C	submontagem	1124	1124
L01C	montagem	1125	1132
L01C	pintura	1249	1249
L01C	edificação	1248	1325

Tabela 25: Log de Eventos Sintético para o Bloco L05C (parcial)

plateid	blastedplateid	pieceid	panelid	subassemblyid	blockid	paintedblockid	megablockid	activity	timestamp
1	32			42	112	115	116	jateamento	7866
1	32			42	112	115	116	submontagem	7869
1	32			42	112	115	116	montagem	7876
1	32			42	112	115	116	pintura	8749
1	32			42	112	115	116	edificacao	9026
2	105	107	110		112	115	116	jateamento	7875
2	105	107	110		112	115	116	processamento	7875
2	105	107	110		112	115	116	painelizacao	7875
2	105	107	110		112	115	116	montagem	7876
2	105	107	110		112	115	116	pintura	8749
2	105	107	110		112	115	116	edificacao	9026
3	35	50		65	112	115	116	jateamento	7867
3	35	50		65	112	115	116	processamento	7870
3	35	50		65	112	115	116	submontagem	7871
3	35	50		65	112	115	116	montagem	7876
3	35	50		65	112	115	116	pintura	8749
3	35	50		65	112	115	116	edificacao	9026
3	35	52		65	112	115	116	jateamento	7867
3	35	52		65	112	115	116	processamento	7870
3	35	52		65	112	115	116	submontagem	7871
3	35	52		65	112	115	116	montagem	7876
3	35	52		65	112	115	116	pintura	8749
3	35	52		65	112	115	116	edificacao	9026
3	35	57		65	112	115	116	jateamento	7867
3	35	57		65	112	115	116	processamento	7870

Através da comparação entre os dois trechos dos logs de eventos, foi possível notar que o log sintético retrata o mesmo processo que o log original, porém em maior granularidade das atividades, revelando o contexto do processo que, antes da simulação, estava oculto.

A tabela 26 evidencia a desfragmentação das atividades a partir da sintetização de um novo log de eventos, apresentando uma comparação na quantidade de atividades envolvidas na construção de 18 blocos aleatórios, agrupadas por tipo de atividade, conforme dados do log de eventos original e naquele sintetizado.

Tabela 26: Evidências da desfragmentação das atividades para 18 blocos de exemplo

Atividade	Quantidade Original	Quantidade Sintetizada
Jateamento	18	472
Processamento	18	776
Painelização	18	72
Submontagem	18	48
Montagem	18	18
Pintura	18	18
Edificação	18	18

Outrossim, a simulação do log de eventos, conforme o método descrito, foi capaz de aumentar o nível de detalhamento das atividades, no que diz respeito à quantidade de instâncias de cada atividade, maior detalhamento no espectro temporal do processo e no posicionamento das atividades sobre este espectro, e a criação de uma relação de dependência entre as atividades desempenhadas, através da identificação dos tipos de constructos e atividades envolvidos na montagem de blocos, e os tipos de constructos que são requisitos e produtos para cada atividade.

Consequentemente, foi possível melhorar o cenário para a realização de tarefas de aprendizado de máquina com mineração de processos sobre os registros das atividades de construção naval.

A seção 3.7 apresenta resultados de atividades de descoberta de processos e identificação de gargalos de produção com base no log de eventos sintetizado através da ferramenta de simulação desenvolvida neste estudo.

3.7 Mineração de Processos sobre o Log de Eventos Sintético

A simulação do processo e geração de log de eventos sintético foi capaz de gerar um conjunto de dados livre das problemáticas de qualidade inerentes à imprecisão dos dados, sobreposição de atividades e baixa granularidade dos atributos contidos nele.

Finalmente, foi possibilitada a obtenção de um log de eventos adequado à prática de tarefas de mineração de processos, de forma que seja possível inferir resultados relevantes para o domínio de negócio do processo estudado.

A utilização da ferramenta Disco para descoberta do processo sobre o log de eventos sintético retornou o modelo disposto na figura 22.

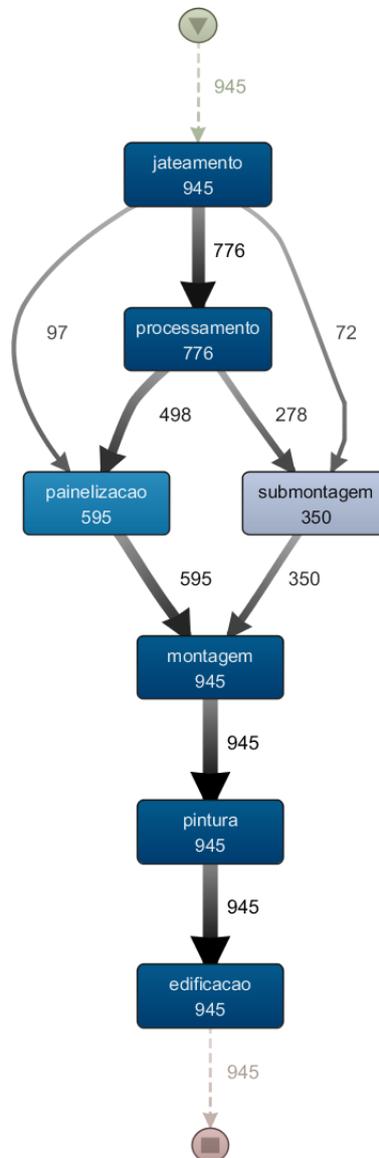


Figura 22: Modelo do processo após aplicação da metodologia

Em conformação mais granular do log de eventos, ou seja, de modo que a identificação da instância do processo não foi baseada apenas na identificação do bloco construído, mas, sim, das peças que constituíram constructos maiores, o algoritmo de descoberta de processos foi capaz de inferir relações de ordenação e causalidade fidedignas à realidade. Da mesma forma, não ocorreu a identificação errônea de sobreposições de atividades.

Estudos com a animação gráfica da ilustração do processo são utilizados para identificação de gargalos de produção e outras deficiências que ficam evidentes sob uma

óptica abstrata do processo. A figura 23 apresenta quadros oriundos da animação correspondente à execução do processo conforme o log de eventos sintético gerado e disposto parcialmente na tabela 25.

Na figura, os círculos vermelhos preenchidos em amarelo denotam a acumulação de atividades em execução de forma simultânea, permitindo o entendimento sobre onde ocorrem gargalos no fluxo do processo.

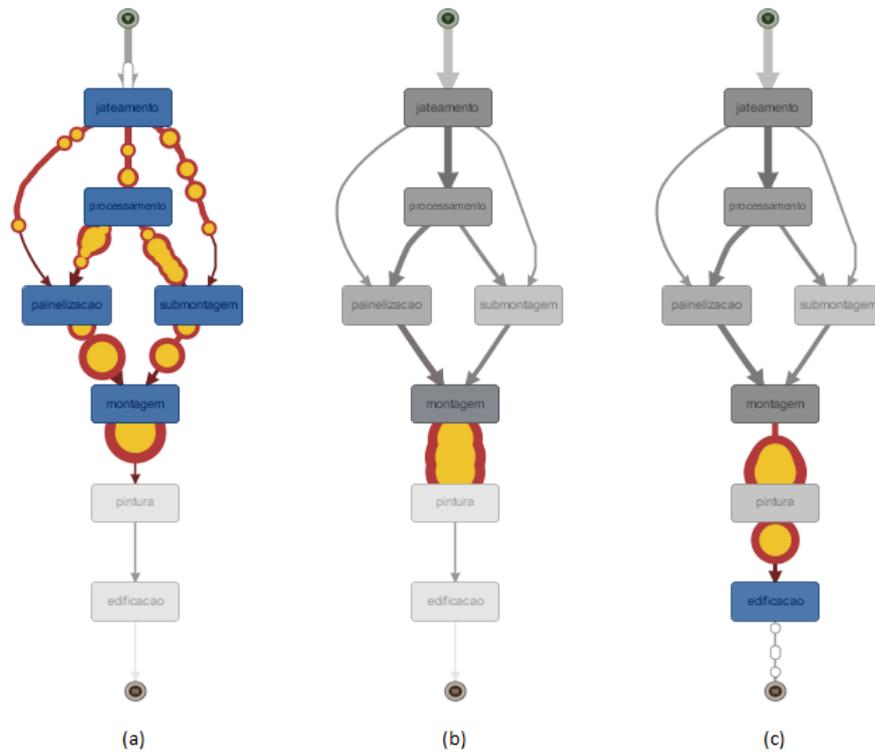


Figura 23: Quadro de animação do modelo do processo para identificação de gargalos

O quadro (a) demonstra o andamento das atividades precedentes à montagem, enquanto o quadro (b) evidencia um atraso na produção pela situação de espera, nas atividades de pintura, pelo término da última atividade de montagem. O quadro (c) mostra a retomada do processo após a satisfação da suposta restrição.

4 CONSIDERAÇÕES FINAIS

Este capítulo apresenta as considerações finais da presente dissertação. São ilustrados os resultados obtidos com a aplicação da metodologia, é feita a conclusão do trabalho e propostos trabalhos futuros para aprimoramento do produto desenvolvido.

4.1 Resultados Obtidos

De forma a registrar os resultados obtidos com a aplicação da metodologia de simulação de processos com vistas na geração de logs de eventos sintéticos da construção naval, esta seção apresenta e compara modelos do processo analisado no estudo de caso presente nesta dissertação, conforme descoberta feita por diferentes algoritmos.

Foram gerados modelos sobre o log de eventos original (a) e o log de eventos sintético (b) com os algoritmos Heuristics Miner (WEIJTERS; AALST W.; MEDEIROS, 2006), Fuzzy Miner (GUNTHER; VAN DER AALST, 2007) e Disco Miner (GUNTHER; ROZINAT, 2012), retratados respectivamente nas figuras 24, 25 e 26.

Nos modelos referentes às atividades dos algoritmos sobre o log de eventos original, é possível notar grande confusão nas relações entre as atividades do processo, bem como grande discrepância entre um modelo e outro, no que se trata destas relações.

Na figura 24(a), são retratados ciclos nas atividades de jateamento e pintura, bem como a relação de causalidade entre a montagem de bloco e a edificação, ignorando a pintura do bloco, que, no ambiente ideal, deve ocorrer entre elas. O modelo gerado pelo algoritmo Heuristics Miner com base no log de eventos original também retratou ciclos entre as atividades de processamento e submontagem, bem como entre as atividades de submontagem e montagem, embora estas sejam situações que não ocorrem no processo real.

A figura 24(b) mostra a atuação do mesmo algoritmo, porém, sobre o conjunto de dados gerado pelo método de simulação especificado nesta dissertação. Neste caso, as relações de causalidade e ordenação inferidas mostraram fidedignidade em relação ao processo real de construção, mostrando um fluxo conforme especificado na parametrização do sistema de simulação utilizado.

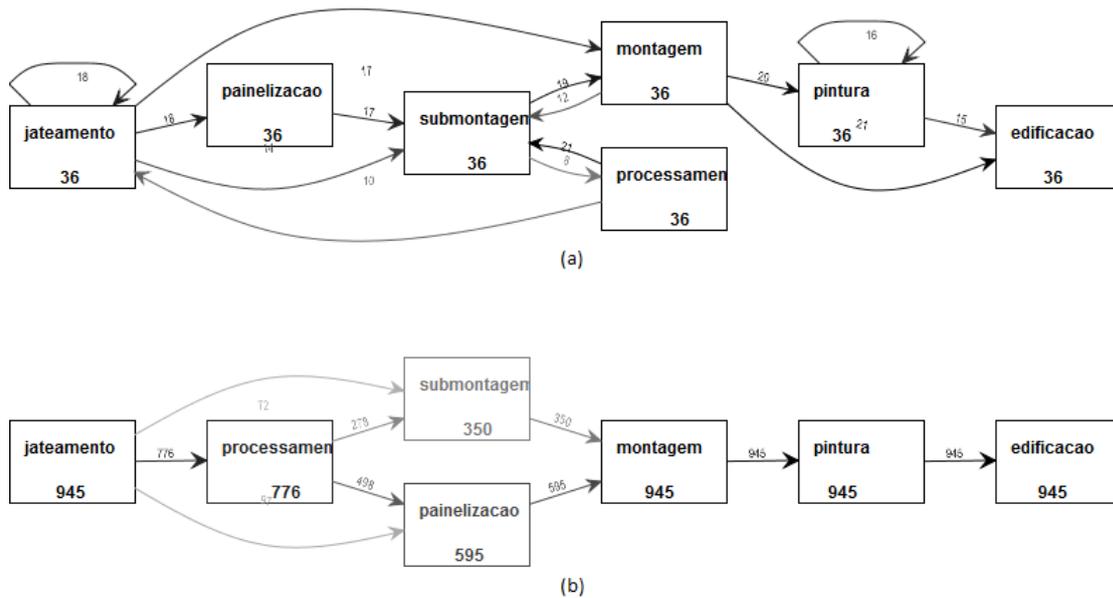


Figura 24: Comparação entre modelos gerados com o algoritmo Heuristics Miner sobre o log de eventos original (a) e sintético (b).

De forma mais acentuada, o modelo gerado pelo algoritmo Fuzzy Miner sobre o conjunto de dados original (figura 25 (a)) apresenta confusão nas relações de causalidade e ordenação entre as atividades do processo. De fato, o modelo apresentado denota um processo do tipo que é caracterizado por (AALST, 2011b) como "espaguete", em analogia à desordem caótica nas relações entre as atividades.

Em contraste ao modelo ilustrado na figura 25 (a), o modelo apresentado na figura 25 (b) mostra um processo mais claramente estruturado, muito semelhante àquele inferido pelo algoritmo Heuristics Miner (figura 24 (b)). Este modelo foi inferido sobre o log de eventos sintetizado nos experimentos deste estudo.

Por se tratar de uma extensão do Fuzzy Miner, é natural que o algoritmo implementado na ferramenta Disco (GUNTHER; ROZINAT, 2012) apresente modelos semelhantes ao que seria apresentado por aquele, em contexto semelhante.

Contudo, há a adição de notações gráficas que denotam importantes informações.

Desta forma, são apresentados os modelos que foram inferidos a partir deste algoritmo, respectivamente, sobre os logs de eventos original e sintético, nas figuras 26 (a) e 26 (b).

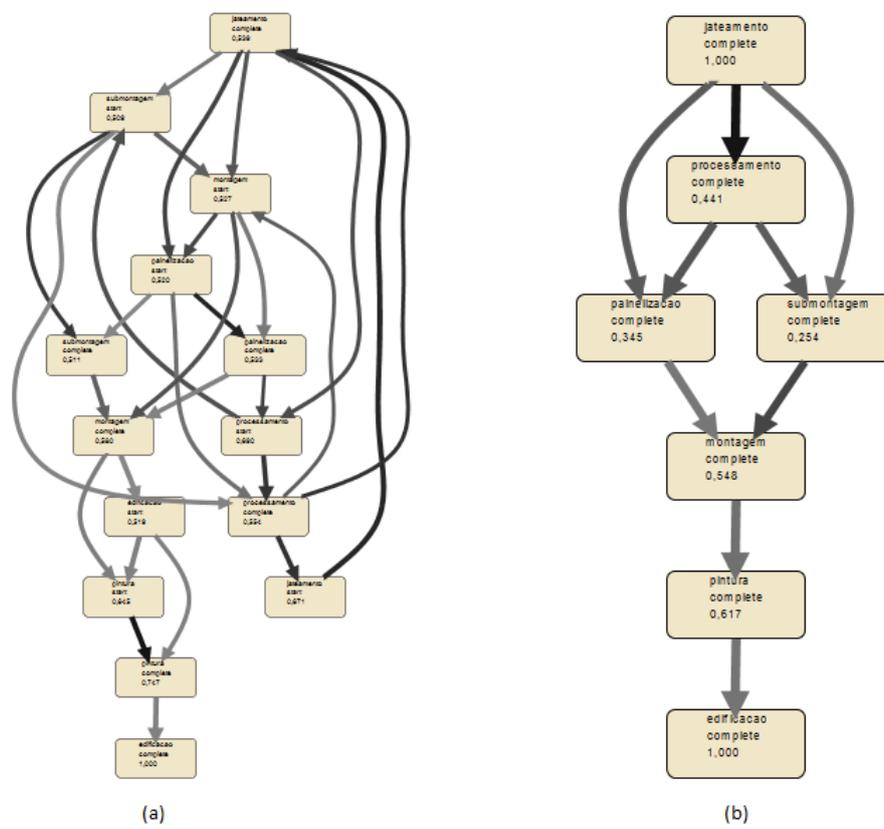


Figura 25: Comparação entre modelos gerados com o algoritmo Fuzzy Miner sobre o log de eventos original (a) e sintético (b).

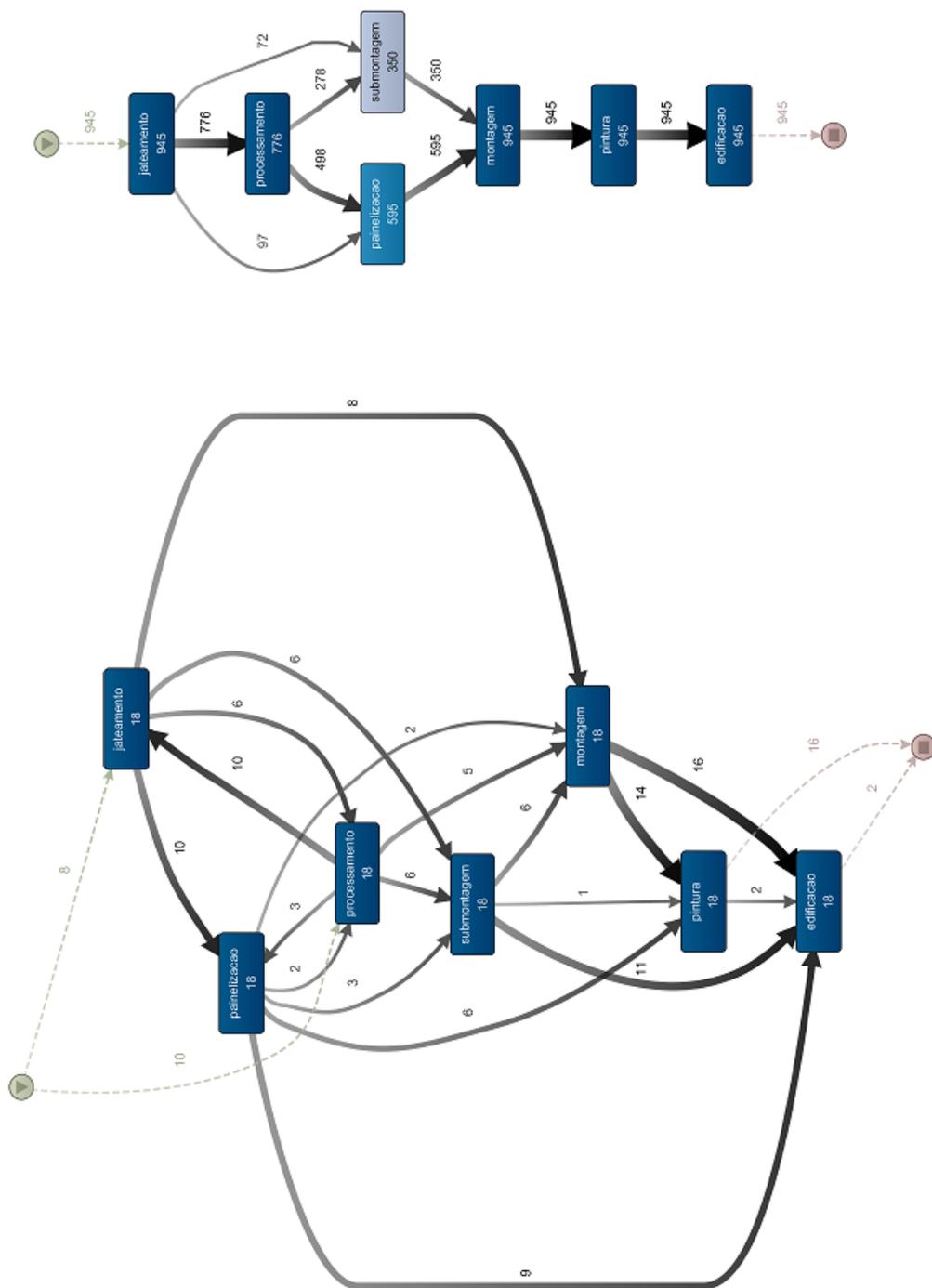


Figura 26: Comparação entre modelos gerados na ferramenta Disco sobre o log de eventos original (a) e sintético (b).

Nestes modelos, há a indicação explícita das atividades por onde o processo pode iniciar e terminar. Há, também a denotação da quantidade absoluta de instâncias do processo que utilizaram cada rota do processo.

No caso do primeiro modelo, além da caoticidade evidente, é notória a possibilidade do início do processo nas atividades de jateamento e processamento, embora conhecimento técnico sobre o processo de construção indique que somente o primeiro caso é admitido. No modelo da figura 26 (a) também é expressa a possibilidade de término do processo nas atividades de pintura e edificação, embora somente a última seja admitida na realidade.

O modelo expresso na figura 26 (b) apresenta, basicamente, o mesmo modelo também expresso na figura 25 (b), porém, com a indicação da quantidade absoluta de instâncias simuladas do processo que aderiram à cada caminho deste.

4.2 Conclusão

Esta dissertação visou a proposta de um método original para a sintetização de logs de eventos capazes de retratar um processo de construção naval.

De forma a suprir o primeiro objetivo específico do presente trabalho, foi proposto que, com base em logs de eventos caracterizados de baixa granularidade das atividades nos aspectos quantitativo e temporal, além de parametrização específica, fosse criado um sistema para geração de um novo log de eventos sintético, então isento das anomalias de qualidade apresentadas neste estudo.

O segundo objetivo foi tratado a partir da implementação do sistema descrito na seção 2.3, com a utilização da linguagem de programação Java, interface SQL SQLite e a biblioteca Apache Commons Math para implementação de distribuições de probabilidades diversas. O resultado desta atividade foi o sistema de software para simulação do processo de construção naval e sintetização do log de eventos correspondente.

A realização de estudo de caso em estaleiro da indústria brasileira de construção naval supriu o terceiro objetivo do trabalho, possibilitando a validação da abordagem proposta através da obtenção de modelos do processo de construção idealizado e conjuntos de dados para a extração do log de eventos original utilizado no estudo.

Com base nos resultados obtidos, foi possível comprovar a hipótese de que um log de eventos caracterizado pela baixa granularidade na descrição das atividades, com a adição de conjuntos de dados estatísticos inerentes, estruturados de acordo com uma especificação, podem ser utilizados por mecanismos de simulação de processos em ambiente computacional para sintetização de um novo log de eventos isento de anomalias e compatível com as restrições do processo.

A partir do sucesso na aplicação do método de geração do log de eventos sintético e, posteriormente, na inferência de modelos do processo de construção livres de anomalias,

foi possível fundamentar a solução para diagnóstico de processos adotados na indústria naval brasileira através da mineração de processos, suprimindo o objetivo geral do presente trabalho, bem como comprovando a hipótese principal apresentada.

4.3 Trabalhos Futuros

Esta seção apresenta ideias para a continuação deste trabalho, conforme a listado abaixo.

1. Implementação de interface gráfica para a parametrização do sistema de software concebido neste trabalho, que atualmente é feita via código-fonte em linguagem Java;
2. Estudo de viabilidade de aplicação da metodologia de sintetização de logs de eventos em outros segmentos da indústria, além da construção naval;
3. Experimento de extensão do algoritmo Alpha (AALST; WEIJTERS; MARUSTER, 2004) para integração do contexto oculto do processo, de forma a incorporar o tratamento específico à problemática da baixa granularidade das atividades na execução do algoritmo;
4. Especificação de metodologia para coleta de dados de atividades no perímetro de construção que seja viável no contexto da indústria naval brasileira sob os vieses financeiro e de organização de recursos humanos;
5. Proposta de reorganização do sistema de produção e montagem com vistas na coleta dos logs de eventos adequados à mineração de processos no momento em que as atividades ocorrem no processo real no perímetro de construção.

REFERÊNCIAS

AALST, W. M. Van der; WEIJTERS, A. Process mining: a research agenda. **Computers in industry**, [S.l.], v.53, n.3, p.231–244, 2004.

AALST, W. van der. **Process Mining: Discovery, Conformance and Enhancement of Business Processes**. [S.l.]: Springer, 2011.

AALST, W. van der. Process Mining Manifesto. **Business Process Management Workshops 2011**, [S.l.], 2011.

AALST, W. van der. Extracting event data from databases to unleash process mining. In: **BPM-Driving Innovation in a Digital World**. [S.l.]: Springer, 2015. p.105–128.

AALST, W. van der; RUBIN, V.; DONGEN, B. van; KINDLER, E.; GUNTHER, C. Process Mining: A Two-Step Approach using Transition Systems and Regions. **BPM-06-30, BPM Center Report**, [S.l.], 2007.

AALST, W. van der; WEIJTERS, A.; MARUSTER, L. Workflow Mining: Discovering Process Models from Event Logs. **IEEE Transactions on Knowledge and Data Engineering**, [S.l.], 2004.

BERGENTHUM, R.; DESEL, J.; LORENZ, R.; MAUSER, S. Process mining based on regions of languages. In: INTERNATIONAL CONFERENCE ON BUSINESS PROCESS MANAGEMENT, 2007. **Anais...** [S.l.: s.n.], 2007. p.375–383.

BOLT, A.; LEONI, M. de; AALST, W. M. van der. Scientific Workflows for Process Mining: Building Blocks, Scenarios, and Implementation. **BPMcenter.org**, [S.l.], 2015.

BORGES, E. **Um Método para Deduplicação de Metadados Bibliográficos Baseado no Empilhamento de Classificadores**. 2013. Tese (Doutorado) — Universidade Federal do Rio Grande do Sul.

CAMARGO, S. **Mineração de Regras de Associação no Problema da Cesta de Compras Aplicada ao Comércio Varejista de Confeção**. 2002. Tese (Doutorado) — Universidade Federal do Rio Grande do Sul.

CAMERON, I.; HANGOS, K. **Process Modelling and Model Analysis**. [S.l.]: Academic Press, 2001.

CLAES, J.; POELS, G. Process mining and the ProM framework: an exploratory survey. In: INTERNATIONAL CONFERENCE ON BUSINESS PROCESS MANAGEMENT, 2012. **Anais...** [S.l.: s.n.], 2012. p.187–198.

DE GRANDI, E.; PARZIANELLO, L. Itracking - Um Framework para Rastreamento através de RFID. **13º Mostra de Produção Universitária - FURG**, [S.l.], 2014.

DONGEN, B. F. van; AALST, W. M. Van der. EMiT: A process mining tool. In: **Applications and Theory of Petri Nets 2004**. [S.l.]: Springer, 2004. p.454–463.

DONGEN, B. F. van; ADRIANSYAH, A. Process mining: fuzzy clustering and performance visualization. In: BUSINESS PROCESS MANAGEMENT WORKSHOPS, 2010. **Anais...** [S.l.: s.n.], 2010. p.158–169.

DONGEN, B. van. The ProM Framework: a New Era in Process Mining Tool Support. **ICATPN 2005**, [S.l.], 2005.

DUARTE FILHO, N.; BOTELHO, S.; DOS SANTOS, E. Improving the Shipbuilding Industry through RFID. **NAVTEC 2013**, [S.l.], 2013.

DUARTE FILHO, N.; BOTELHO, S.; SANTOS, E. dos. Improving Shipbuilding by Tracking People and Things Through RFID. **2014 International Conference on RFID**, [S.l.], 2014.

EYRES, D. J.; BRUCE, G. J. **Ship Construction**. [S.l.]: Butterworth-Heinemann, 2012.

FAYYAD, M.; PIATETSKY-SHAPIRO, G.; SMYTH, P. **From Data Mining to Knowledge Discovery: an Overview**. [S.l.]: MIT Press/AAAI Press, 1996.

FORTUNATO, T.; SOUZA, I.; BARBOSA JR., A. Sistema de Acompanhamento da Produção Baseado em Visão para Indústrias de Construção e Montagem - Estudo de Caso em Estaleiro Naval e Offshore. **14º Mostra de Produção Universitária - FURG**, [S.l.], 2015.

GUNTHER, C. W.; ROZINAT, A. Disco: Discover Your Processes. **BPM (Demos)**, [S.l.], 2012.

GUNTHER, C. W.; VAN DER AALST, W. M. Fuzzy mining—adaptive process simplification based on multi-perspective metrics. In: **Business Process Management**. [S.l.]: Springer, 2007. p.328–343.

HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTERMANN, P.; WITTEN, I. H. The WEKA Data Mining Software: An Update. **SIGKDD Explorations**, [S.l.], 2009.

HAMMER, M. What is Business Process Management? In: **Handbook on Business Process Management I**. [S.l.]: Springer, 2010.

HAN, J.; KAMBER, M.; PEI, J. **Data Mining: Concepts and Techniques**. [S.l.]: Morgan Kaufman, 2012.

HECKL, D.; MOORMANN, J. Process Performance Management. In: **Handbook on Business Process Management II**. [S.l.]: Springer, 2010.

HOLT, J. **A Pragmatic Guide to Business Process Modelling**. [S.l.]: British Informatics Society, 2009.

JESTON, J.; NELIS, J. **Business Process Management: Practical Guidelines to Successful Implementations**. 3a.ed. [S.l.]: Routledge, 2014.

JUNGERMANN, F. Information extraction with rapidminer. In: GSCL SYMPOSIUM SPRACHTECHNOLOGIE UND EHUMANITIES, 2009. **Proceedings...** [S.l.: s.n.], 2009. p.50–61.

KIM, H.; LEE, J.-G.; LEE, S.-S.; PARK, J. H. A simulation-based shipbuilding system for evaluation of validity in design and manufacturing. In: SYSTEMS, MAN AND CYBERNETICS, 2003. IEEE INTERNATIONAL CONFERENCE ON, 2003. **Anais...** [S.l.: s.n.], 2003. v.1, p.522–529.

KIM, M.; CHOI, W.; KIM, B.-C.; KIM, H.; SEOL, J. H.; WOO, J.; KO, K. H. A vision-based system for monitoring block assembly in shipbuilding. **Computer-Aided Design**, [S.l.], v.59, p.98–108, 2015.

KÖNIG, M.; BEISSERT, U.; STEINHÄUER, D.; BARGSTÄDT, H.-J. Constraint-based simulation of outfitting processes in shipbuilding and civil engineering. In: EUROSIM CONGRESS ON MODELING AND SIMULATION, 6., 2007. **Proceedings...** [S.l.: s.n.], 2007.

LEE, D.; PARK, J. H.; BAE, H. Comparison between planned and actual data of block assembly process using process mining in shipyards. **Journal of Society for e-Business Studies**, [S.l.], v.18, n.4, 2013.

LINDSTROM, P.; DELANY, S. J.; MAC NAMEE, B. Autopilot: Simulating changing concepts in real data. In: CONFERENCE PAPERS, 2008. **Anais...** [S.l.: s.n.], 2008. p.21.

MACIEL, T.; SEUS, V.; SANTOS, K. dos; BORGES, E. Mineração da Dados em Triagem de Risco de Saúde. **Revista Brasileira de Computação Aplicada** v. 7 n.2, [S.l.], 2015.

MACIEL, T. V.; FILHO, N. D. **Process Mining at Glance in the Brazilian Shipbuilding Industry**.

MEDEIROS, A. A. de; DONGEN, B. F. van; AALST, W. M. Van der; WEIJTERS, A. **Process mining**: Extending the α -algorithm to mine short loops. [S.l.]: BETA working paper series, WP 113, Eindhoven University of Technology, Eindhoven, 2004.

MEDEIROS, A. **Genetic Process Mining**. 2006. Tese (Doutorado) — Eindhoven University of Technology.

MEDEIROS, A. K. A. de; WEIJTERS, A. J.; AALST, W. M. van der. Genetic process mining: an experimental evaluation. **Data Mining and Knowledge Discovery**, [S.l.], v.14, n.2, p.245–304, 2007.

NIST/SEMATECH. **e-Handbook of Statistical Methods**. [S.l.]: <http://www.itl.nist.gov/div898/handbook/>, 2012.

PARK, J.; LEE, D.; ZHU, J. An integrated approach for ship block manufacturing process performance evaluation: Case from a Korean shipbuilding company. **International Journal of Production Economics**, [S.l.], v.156, p.214–222, 2014.

PRUYN, J.; MARTINEZ, J. C.; MOLENAAR, W. Can RFID be Successfully Used in Shipbuilding? **Journal of Ship Production and Design**, [S.l.], v.26, n.2, p.111–116, 2010.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria: [s.n.], 2014.

RUBIN, V.; GUNTHER, C. W.; VAN DER AALST, W. M.; KINDLER, E.; VAN DONGEN, B. F.; SCHAFER, W. Process mining framework for software processes. **Software Process Dynamics and Agility**, [S.l.], 2007.

SANDER, G.; VASILIU, A. The ILOG JViews graph layout module. In: GRAPH DRAWING, 2002. **Anais...** [S.l.: s.n.], 2002. p.438–439.

SANTOS, E. B. dos. **Rastreamento em ambientes de construção e montagem naval**. 2012. ção (Mestrado em Ciência da Computação) — Universidade Federal do Rio Grande, Rio Grande.

SANTOS, M. Y.; RAMOS, I. **Business Intelligence**: Tecnologias da informação na gestão de conhecimento. [S.l.]: FCA-Editora de Informática, 2006.

SHA, O.; MISRA, S.; GUPTA, A. Simulation of Block Assembly Process in Shipbuilding by Petri-nets. **WMTC2009**, [S.l.], 2009.

SILVA, V.; LAVOURA, G.; FORTUNATO, T.; FURLAN, T. RFID para Rastreamento de Ativos - Um Estudo de Caso em Estaleiro Naval e Offshore. **14º Mostra de Produção Universitária - FURG**, [S.l.], 2015.

SOUZA, C. M. de. **Técnicas Avançadas em Planejamento e Controle da Construção Naval**. 2009. Tese (Doutorado) — Universidade Federal do Rio de Janeiro.

TIWARI, A.; TURNER, C. J.; MAJEED, B. A review of business process mining: state-of-the-art and future trends. **Business Process Management Journal**, [S.l.], v.14, n.1, p.5–22, 2008.

VAN DER AALST, W. M.; TER HOFSTEDÉ, A. H.; WESKE, M. Business process management: A survey. In: **Business process management**. [S.l.]: Springer, 2003. p.1–12.

VAN DONGEN, B. F.; DE MEDEIROS, A. A.; WEN, L. Process mining: Overview and outlook of petri net discovery algorithms. In: **Transactions on Petri Nets and Other Models of Concurrency II**. [S.l.]: Springer, 2009. p.225–242.

VERBEEK, H.; BUIJS, J.; VAN DONGEN, B.; AALST, W. M. van der. Prom 6: The process mining toolkit. **Proc. of BPM Demonstration Track**, [S.l.], v.615, 2010.

WEIJTERS, A.; AALST, W. M. van der. Process mining: discovering workflow models from event-based data. In: BELGIUM-NETHERLANDS CONF. ON ARTIFICIAL INTELLIGENCE, 2001. **Anais...** [S.l.: s.n.], 2001.

WEIJTERS, A.; AALST W. van der; MEDEIROS, A. **Process Mining with the Heuristic Miner Algorithm**. [S.l.: s.n.], 2006.

WEIS, Á. A.; BOTELHO, M. P. de; NAGY, I. Gyártási termék nyomkövetése RFID technológiával: esettanulmány egy hajógyári környezetben. **Erdélyi Múzeum-Egyesület**, [S.l.], 2015.

WEN, L.; AALST, W. M. van der; WANG, J.; SUN, J. Mining process models with non-free-choice constructs. **Data Mining and Knowledge Discovery**, [S.l.], v.15, n.2, p.145–180, 2007.

WEN, L.; WANG, J.; AALST, W. M. van der; HUANG, B.; SUN, J. A novel approach for process mining based on event types. **Journal of Intelligent Information Systems**, [S.l.], 2009.

WEN, L.; WANG, J.; SUN, J. Mining invisible tasks from event logs. In: **Advances in Data and Web Management**. [S.l.]: Springer, 2007. p.358–365.

WESKE, M. **Business Process Management: Concepts, Languages, Architectures**. [S.l.: s.n.], 2012.

WITTEN, I.; FRANK, E.; HALL, M. **Data Mining: Practical Machine Learning Tools and Techniques**. [S.l.]: Morgan-Kauffman, 2011.