



UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

Verificação e classificação de não conformidades de processos
de negócios na indústria farmacêutica através de técnicas de
mineração de processos

Guilherme Neves Lopes

Orientadores

Flavia Maria Santoro
Fernanda Baião

Rio de Janeiro, RJ – Brasil

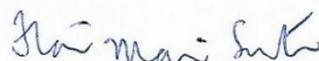
Março de 2015

Verificação e classificação de não conformidades de processos de negócios na indústria farmacêutica através de técnicas de mineração de processos.

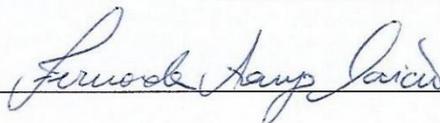
Guilherme Neves Lopes

DISSERTAÇÃO APRESENTADA COMO REQUISITO PARCIAL PARA OBTENÇÃO DO TÍTULO DE MESTRE PELO PROGRAMA DE PÓSGRADUAÇÃO EM INFORMÁTICA DA UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO (UNIRIO). APROVADA PELA COMISSÃO EXAMINADORA ABAIXO ASSINADA.

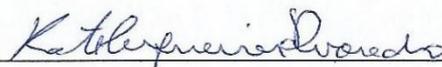
Aprovada por:



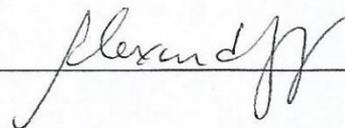
Flávia Maria Santoro, D.Sc. - UNIRIO



Fernanda Baião D.Sc. – UNIRIO



Kate Cerqueira Revoredo D.Sc - UNIRIO



Alexandre Gonçalves Evsukoff D.Sc. - UFRJ

Rio de Janeiro, 06 de Março, de 2015

L864 Lopes, Guilherme Neves.
Verificação e classificação de não conformidade de processos de negócios na indústria farmacêutica através de técnicas de mineração de processos / Guilherme Neves Lopes, 2015.
179 f. ; 30 cm

Orientadora: Flavia Maria Santoro.

Coorientadora: Fernanda Baião.

Dissertação (Mestrado em Informática) - Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, 2015.

1. Mineração de dados (Computação). 2. Mineração de processos.
3. Verificação automática de conformidade. 4. Qualidade dos produtos.
5. Indústria farmacêutica. I. Santoro, Flavia Maria. II. Baião, Fernanda.
III. Universidade Federal do Estado do Rio de Janeiro. Centro de Ciências Exatas e Tecnológicas. Curso de Mestrado em Informática.
IV. Título.

CDD – 005.74

Dedicatória

Dedico este trabalho a minha esposa Edna, minha companheira de todas as horas que suportou estes dois anos de ausência, noites mal dormidas, escutou pacientemente as minhas teorias e a leitura da dissertação. Sem o seu apoio e amor nada disto teria sido possível.

Agradecimentos

Agradeço às minhas orientadoras Flavia Santoro e Fernanda Baião pela infinita paciência e dedicação, aos professores sempre atentos e que abriram os meus olhos para um mundo novo de conhecimento, aos meus colegas de mestrado que também emprestaram seu tempo e conhecimento em intermináveis apresentações de artigos, aos funcionários da UNIRIO, sempre solícitos e eficientes.

Agradeço também aos meus amigos Haroldo Fernandes por ter me incentivado a voltar para faculdade e começar um novo ciclo de aprendizado, ao meu professor e companheiro Frederico Sauer que me inspirou e incentivou a seguir a área acadêmica, incentivando a fazer este mestrado e agora o doutorado que está por vir, aos meus pais que apesar de estarem longe do Rio de Janeiro, sempre estiveram presentes com seu incentivo.

E por fim agradeço a Deus que apesar de todas as dificuldades e provações me deu o fardo no peso certo, para que eu não caísse em exaustão e nem ficasse preguiçoso.

LOPES, Guilherme Neves. **Verificação e classificação de não conformidades de processos de negócios na indústria farmacêutica através de técnicas de mineração de processos**. UNIRIO, 2015. 179 páginas. Dissertação de Mestrado. Departamento de Informática Aplicada, UNIRIO.

RESUMO

O movimento mundial de regulamentação da economia a partir da década de 1990 aumentou a demanda de auditoria e verificação de conformidade nas empresas. No Brasil, com a criação das agências reguladoras, vários setores da economia tiveram que se adaptar rapidamente para adequar seus processos de negócios e ferramentas de controle para garantir a conformidade com as normas. Um setor foi especialmente afetado, a indústria farmacêutica, que a cada dia fica sujeita a novas regras. Procedimentos de teste e auditoria em processos de negócios normalmente são executados em uma amostra de dados. No entanto apesar do aumento do controle, o recolhimento de remédios tem aumentado ao longo dos anos, mostrando que a escolha da amostra para teste de qualidade não está sendo realizada de maneira satisfatória. Entretanto, informações detalhadas sobre a execução destes processos podem estar disponíveis em logs de eventos de sistemas de informação, o que torna inaceitável continuar a verificar apenas uma pequena massa de dados off-line. O problema a ser tratado nesta pesquisa é como realizar a verificação de conformidade e classificação de não conformidades das instâncias do processo de fabricação de medicamentos de forma sistemática e automática. A solução proposta é a utilização de técnicas mineração de processos para identificar e classificar instâncias do processo que têm desvios em relação ao processo padrão. Um estudo de caso com dados de um processo real foi realizado mostrando indícios da viabilidade da proposta.

Palavras Chave: Mineração de processos, verificação automática de conformidade, qualidade do produto, regulamentação farmacêutica, classificação de não conformidade.

ABSTRACT

The global trend for economy regulation since the 1990s raised the demand for auditing and compliance verification procedures in enterprises, which are usually performed on a sample of data. In Brazil, the creation of formal regulatory agencies forced various economic sectors to quickly adapt their business processes and control tools; in particular, the pharmaceutical sector has been majorly impacted due to the frequent creation of new rules. Despite all these quality control efforts, however, drug recall has increased, which may indicate that sampling for quality testing is not being performed satisfactorily. On the other hand, detailed information on the implementation of these processes has been made available in event logs of information systems, which makes it unacceptable to perform quality testing on a small amount of offline data. The problem addressed in this research is how to automatically and systematically perform conformity assessment and classification of non-compliant instances of manufacturing drugs. The proposed solution applies process mining techniques to identify and classify process instances that deviate from the standard process. A case study on a real process was performed, showing the feasibility of this proposal.

Keywords: Process mining, conformance check, pharmaceutical regulation, product quality, deviations classification.

Índice

1.	Introdução	1
1.1.	Cenário e motivação	1
1.2.	Problema e hipótese	6
1.3.	Enfoque de Solução e Metodologia de Pesquisa	6
1.4.	Organização do trabalho	7
2.	Referencial teórico	9
2.1.	Processo de Negócio	9
2.2.	Modelagem e Descoberta de Processos	9
2.2.1.	Variante de Processo	12
2.2.2.	Play in	12
2.2.3.	Play out	13
2.2.4.	Replay	13
2.2.5.	Algoritmo Fuzzy Miner	13
2.2.6.	Redes de Petri	14
2.3.	Dimensões da qualidade de um modelo de processo de negócio	15
2.3.1.	Fitness	15
2.3.2.	Precisão	16
2.3.3.	Generalização	17
2.3.4.	Simplicidade	17
2.4.	Verificação de conformidade de processos	18
2.5.	Principais técnicas para a verificação de conformidade de processos	21
2.5.1.	Verificação de conformidade utilizando alinhamento entre observado e modelado	21
2.5.2.	Verificação de conformidade utilizando Token-Based Replay	23
2.6.	Conceitos importantes na Indústria Farmacêutica	31
2.6.1.	Desvio	31
2.6.2.	Risco	32
2.6.3.	Detectabilidade do Risco	33
3.	Trabalhos relacionados	35
3.1.	Audit 2.0 usando a mineração de processos para melhorar os processos de auditoria (Wil M.P. van der Aalst, 2010).	35
3.2.	Descoberta de anomalias em fluxo de processos de sistemas de automação (UNISIST, 2012).	37
3.3.	Verificação de conformidade dos processos de atendimento à Saúde (R.S. Mans, 2008)	39

4.	Método de descoberta e classificação de não conformidades na indústria farmacêutica.	45
4.1.	Primeira Fase – Levantamento de informações sobre o domínio.	47
4.2.	Segunda Fase – Levantamento dos atributos obrigatórios nos logs e pré-processamento.	51
4.2.1.	Identificação do Case_id	53
4.2.2.	Identificação do TimeStamp	53
4.2.3.	Identificação da Activity	53
4.2.4.	Identificação do Resource	53
4.2.5.	Identificação de outros atributos do log	53
4.3.	Terceira Fase – Mineração de Processos.	54
4.4.	Quarta Fase – Verificação de conformidade.	54
4.4.1.	Identificação de instâncias conformes	55
4.4.2.	Identificação de instâncias com desvios não regulatórios e incompletas	55
4.4.3.	Identificação dos Desvios regulatórios com atividades proibidas.	56
4.4.4.	Identificação dos Desvios Regulatórios devido à restrição de valor de atributo.	56
5.	Avaliação da proposta	58
5.1.	O Estudo Exploratório	58
5.1.1.	Mineração com Heuristic Miner	67
5.1.2.	Mineração com Mine for a Fuzzy Model utilizando o PROM.	68
5.1.3.	Verificação De regras com LTL	69
5.1.4.	Mineração Fuzzy Model utilizando o DISCO	72
5.2.	O Estudo de Caso	74
5.2.1.	Primeira fase – Levantamento de informações sobre o domínio	75
5.2.2.	Segunda Fase – Levantamento dos atributos obrigatórios nos logs e pré-processamento.	82
5.3.	Terceira Fase – Mineração de Processos.	86
5.4.	Quarta Fase – Verificação de conformidade.	87
5.4.1.	Identificação das instâncias conformes	87
5.4.2.	Identificação de instâncias com desvios completas e incompletas	87
5.4.3.	Identificação das instâncias com desvios Regulatórios com atividades proibidas.	91
5.5.	Discussão sobre o estudo de caso	93
5.5.1.	Resultados obtidos	96
6.	Conclusões	97

6.1. Contribuições	98
6.2. Restrições da pesquisa	100
6.3. Trabalhos futuros	100
7. Referências bibliográficas	101
8. Apêndices	105
8.1. Pesquisa sobre qualidade na fabricação de medicamentos.	105
8.2. Exemplos de preenchimento do formulário de informações sobre o domínio	107
8.3. Exemplo de aplicação do filtro de atributos para descobrir instâncias completas	114
8.4. Exemplo de aplicação de filtro para descobrir instâncias com desvio regulatório com atividades proibidas.	125
8.5. Exemplo de aplicação de filtro para descobrir instâncias com desvio regulatório devido à restrição de valor.	130
8.6. Estudo de caso, aplicação de filtros para descoberta de instâncias completas no log com filtro de Resources.	133
8.7. Estudo de caso, Aplicação do filtro de atributo para descoberta de instâncias completas no log original com todas as instâncias.	147

Índice de figuras

Figura 1-2 - Sistema Computadorizado GAMP	6
Figura 2-1 Exemplo de Variante na ferramenta Disco	12
Figura 2-2- Exemplo de uma Rede de Petri	15
Figura 2-3 - 4 dimensões de qualidade do modelo de processo (Aalst, 2011)	15
Figura 2-4 - Modelos de processo	16
Figura 2-5 - Modelo de processo precisão alta	17
Figura 2-6 - Modelo de processo generalização	17
Figura 2-7 - Modelo de processo com alta simplicidade	18
Figura 2-8 - Visão da Verificação da conformidade (Aalst, 2011)	19
Figura 2-9 - Processos minerados por diferentes algoritmos (Aalst, 2011)	20
Figura 2-10 - Modelo de processo descoberto por diferentes algoritmos	21
Figura 2-11 - Terminologia utilizada no alinhamento de log e processo (Aalst, 2011).	22
Figura 2-12 – Função Padrão de Custo para alinhamento (Aalst, 2011)	23
Figura 2-13 - Custo padrão para determinar o melhor alinhamento (Aalst, 2011)	23
Figura 2-14 - Exemplo de replay (Rozinat, 2010)	25
Figura 2-15 - Replay de instância não conforme (Rozinat, 2010)	26
Figura 2-16 - Verificação de Conformidade utilizando Replay do Log em Redes de Petri (Aalst, 2011)	28
Figura 2-17 - Cálculo do fitness (Aalst, 2011)	29
Figura 2-18 - Modelos de descoberta com a visão aos 4 parâmetros, fitness, precisão, generalização e simplicidade (Aalst, 2011)	30
Figura 2-19 - Exemplo do diagnóstico do processo utilizando fitness (Aalst, 2011)	30
Figura 2-20 - Separação das instâncias não conformes utilizando drilling down (Aalst, 2011)	31
Figura 3-1 - Processo obtido pelo Heuristic Miner (R.S. Mans, 2008)	41
Figura 3-2 - Modelo de processos para os maiores clusters (R.S. Mans, 2008)	42
Figura 4-1 – Esquema do método produzido pelo autor	45
Figura 4-2 - Processo do método	46
Figura 5-1 - Resultado da pesquisa	61
Figura 5-1 - Processo padrão minerado com Fuzzy Miner no Disco	64
Figura 5-2 - Processo padrão minerado com Fuzzy Miner no PROM	65
Figura 5-3 - Log original do sistema	66
Figura 5-4 - Resultado da mineração com Heuristic Miner	68
Figura 5-5 - Modelo resultante Fuzzy Model	69
Figura 5-6 - Seleção de tipos de regra no LTL do PROM	70

Figura 5-7 - Regra de atributo no LTL	70
Figura 5-8 - Resultado da mineração de regras utilizando o LTL.....	71
Figura 5-9 - Modelo de processo usando Software Disco.....	72
Figura 5-10 - Modelo de processo de uma instância	73
Figura 5-11 - Resultado da aplicação do filtro	74
Figura 5-12 - Visão do sistema túnel de despirogenização (manual sistema).....	76
Figura 5-13 - Processo padrão minerado com Fuzzy Miner no Disco	80
Figura 5-14 - Instância Padrão com atividades do sistema	81
Figura 5-15 - Instância padrão.....	82
Figura 5-16 - Log original do sistema	83
Figura 5-17 - Log após a identificação do case_id.....	84
Figura 5-18 - Identificação do TimeStamp.....	85
Figura 5-19 - Modelo de processo minerado.....	86
Figura 5-20 - Resultado inicial da mineração de processos	87
Figura 5-21 - Instâncias sem as atividades obrigatórias	90
Figura 5-22 - Instâncias resultantes após aplicação do filtro.....	92
Figura 5-23 - Estatísticas após a aplicação do filtro	92
Figura 8-1 - Tela todos os programas	115
Figura 8-2 - Tela com a pasta de instalação do disco.....	115
Figura 8-3 - Tela com o aplicativo Disco	116
Figura 8-4 - Tecla abrir arquivo no Disco	116
Figura 8-5 - Seleção do arquivo de log.....	117
Figura 8-6 - De-para da coluna do log com o atributo case_id.....	117
Figura 8-7 - De-para da coluna do log com Activity.....	118
Figura 8-8 - De-para da coluna do log com o atributo TimeStamp.....	118
Figura 8-9 - Tecla de configuração de formato do timestamp	119
Figura 8-10 - Selecionando o padrão do TimeStamp	119
Figura 8-11 - De-para da coluna do log com o atributo Resource	120
Figura 8-12 - Marcando as outras colunas do log como outros.....	121
Figura 8-13 - Tecla de Start Import.	121
Figura 8-14 - Modelo de processo Minerado e Tecla de filtro	122
Figura 8-15 - Tecla Adicionar filtro	122
Figura 8-16 - Selecionando o tipo de filtro.....	123
Figura 8-17 - Selecionando Evento a ser filtrado	123
Figura 8-18 - Filtro em cascata.....	124

Figura 8-19 - Modelo de processo após a aplicação do filtro.....	125
Figura 8-20 - Selecionando os campos e fazendo o relacionamento de-para.....	126
Figura 8-21 – Importando o log de eventos.....	126
Figura 8-22 - Modelo de processo gerado	127
Figura 8-23 - Adicionando o filtro	127
Figura 8-24 - Selecionando o filtro de atributo	128
Figura 8-25 - Selecionando a atividade e o tipo de filtro	128
Figura 8-26 - Resultado da aplicação do filtro	129
Figura 8-27 - Verificando as instâncias resultantes.....	129
Figura 8-28 - Adicionando Filtro.....	130
Figura 8-29 - Seleção do filtro Desempenho.....	131
Figura 8-30 - Configurando o intervalo de tempo.....	131
Figura 8-31 - Aplicando o filtro	132
Figura 8-32 - Estatísticas do resultado da aplicação do filtro	132
Figura 8-33 - Instâncias resultantes após a aplicação do filtro.....	133
Figura 8-34 - Aplicação correta do filtro para descoberta de instâncias completas.....	134
Figura 8-35 - Aplicação correta do filtro para descoberta de instâncias completas sequência 1	134
Figura 8-36 - Aplicação correta do filtro para descoberta de instâncias completas sequência 2	135
Figura 8-37 - Aplicação correta do filtro para descoberta de instâncias completas sequência 3	135
Figura 8-38 - Aplicação correta do filtro para descoberta de instâncias completas sequência 4	136
Figura 8-39 - Aplicação correta do filtro para descoberta de instâncias completas sequência 5	136
Figura 8-40 - Aplicação correta do filtro para descoberta de instâncias completas sequência 6	137
Figura 8-41 - Aplicação correta do filtro para descoberta de instâncias completas sequência 7	137
Figura 8-42 - Aplicação correta do filtro para descoberta de instâncias completas sequência 8	138
Figura 8-43 - Aplicação correta do filtro para descoberta de instâncias completas sequência 9	138
Figura 8-44 - Aplicação correta do filtro para descoberta de instâncias completas sequência 9	139
Figura 8-45 - Aplicação correta do filtro para descoberta de instâncias completas sequência 10	139

Figura 8-46 - Aplicação correta do filtro para descoberta de instâncias completas sequência 11	140
Figura 8-47 - Aplicação correta do filtro para descoberta de instâncias completas sequência 12	140
Figura 8-48 - Aplicação correta do filtro para descoberta de instâncias completas sequência 13	141
Figura 8-49 - Aplicação correta do filtro para descoberta de instâncias completas sequência 14	141
Figura 8-50 - Aplicação correta do filtro para descoberta de instâncias completas sequência 15	142
Figura 8-51 - Aplicação correta do filtro para descoberta de instâncias completas sequência 16	142
Figura 8-52 - Aplicação correta do filtro para descoberta de instâncias completas sequência 17	143
Figura 8-53 - Aplicação correta do filtro para descoberta de instâncias completas sequência 18	143
Figura 8-54 - Aplicação correta do filtro para descoberta de instâncias completas sequência 19	144
Figura 8-55 - Aplicação correta do filtro para descoberta de instâncias completas sequência 19	144
Figura 8-56 - Aplicação correta do filtro para descoberta de instâncias completas sequência 20	145
Figura 8-57 - Aplicação correta do filtro para descoberta de instâncias completas sequência 21	145
Figura 8-58 - Aplicação correta do filtro para descoberta de instâncias completas sequência 22	146
Figura 8-59 - Aplicação correta do filtro para descoberta de instâncias completas sequência 23	147
Figura 8-60 - Filtro para eliminar atividades proibidas	148
Figura 8-61 - Estatísticas após a aplicação do filtro	148
Figura 8-62 - Carregando um novo log	149
Figura 8-63 - Selecionando a pasta para importação	149
Figura 8-64 – selecionando o arquivo de log	150
Figura 8-65 - Carregando o arquivo de log	150
Figura 8-66 - De-para Case_id	151
Figura 8-67 - De-para outros atributos	151
Figura 8-68 - Importando o log	152
Figura 8-69 - Modelo de processo minerado	152
Figura 8-70 - Adicionando o filtro	153

Figura 8-71 - Escolhendo o tipo de filtro	154
Figura 8-72 - Selecionando o modo do filtro	154
Figura 8-73 - Selecionando a primeira atividade.....	155
Figura 8-74 - Selecionando a segunda atividade	155
Figura 8-75 - Selecionando a terceira atividade	156
Figura 8-76 - Selecionando a quarta atividade	156
Figura 8-77 - Selecionando a quinta atividade	157
Figura 8-78 - Selecionando a sexta atividade	157
Figura 8-79 - Selecionando a sétima atividade.....	158
Figura 8-80 - Selecionando a oitava atividade	158
Figura 8-81 - Selecionando a nona atividade	159
Figura 8-82 - Selecionando a última atividade	159
Figura 8-83 - Aplicando o filtro	160
Figura 8-84 - Modelo de processo gerado após aplicação do filtro.....	161
Figura 8-85 – Estatísticas após a aplicação do filtro	161
Figura 8-86 - Instâncias Resultantes.....	162
Figura 8-87 - Instância A priori.....	163
Figura 8-88 - Demais instâncias completas com desvio não regulatório	163

Índice de tabelas

Tabela 3-1 - Comparativo entre trabalhos relacionados	43
Tabela 4-1 - Formulário de informações sobre o domínio	50
Tabela 5-1 - Formulário de informações sobre o processo túnel de despirogenização	62
Tabela 5-2 - Processo A priori	63
Tabela 5-3 - Tabela de regras do túnel	77
Tabela 5-4 - Formulário de informações sobre o domínio entrevista com especialista	78
Tabela 5-5 - processo padrão	79
Tabela 8-1 - Pesquisa sobre qualidade de fabricação de medicamentos	106
Tabela 8-2 Formulário de informações sobre o domínio equipamento xyz	107
Tabela 8-3 Formulário de informações sobre o domínio sistema XYZ	111

1. Introdução

Neste capítulo, são apresentados o cenário, a motivação do trabalho, a metodologia de pesquisa, o problema e o enfoque na solução e como o trabalho está organizado.

1.1. Cenário e motivação

A motivação deste trabalho é o crescente aumento da exigência regulatória na indústria farmacêutica e a visão de que os meios utilizados para escolha de amostra de produtos para testes de controle de qualidade não são mais considerados adequados para garantir que os problemas de qualidade sejam detectados antes de serem enviados para a venda. Com os escândalos financeiros ocorridos na década dos anos 1990, os governos de todo o mundo perceberam a necessidade de regulamentar as atividades econômicas, até então sujeitas apenas à auto-regulamentação. Leis foram criadas e agências reguladoras passaram a fiscalizar as empresas. Marcos da regulamentação foram a lei SOX (Sarbanes & Oxley, 2002) nos EUA e a lei No 9.986, DE 18 DE JULHO DE 2000 (Presidência da República do Brasil, 2000) no Brasil criaram as agências reguladoras, obrigando as empresas a se adequarem a regras mais rígidas, fazendo com que a verificação de conformidade passasse a ser uma atividade fundamental para a sua sobrevivência.

Atividades reguladas são sujeitas a auditorias constantes. No caso da indústria farmacêutica a cada dois anos, é emitido um certificado de boas práticas de fabricação, após uma auditoria completa nos processos de fabricação pelas agências reguladoras. O termo auditoria refere-se à avaliação das organizações e seus processos. Auditorias são executadas para verificar a validade e fiabilidade das informações relativas às organizações e seus processos. Este procedimento é feito para verificar se

os processos são executados dentro de certos limites definidos pelos gestores, governos e outras partes interessadas.

O cenário que se apresenta é um aumento em escala mundial no número de recolhimentos de medicamentos devido a problemas de qualidade. Em boa parte os problemas são associados a softwares mal instalados ou configurados, o que indica que os processos gerenciados por estes softwares estão produzindo produtos com qualidade abaixo do aceitável.

Para prevenir tais problemas, a Agência Nacional da Vigilância Sanitária do Brasil (ANVISA) exige que a qualificação e a validação de processos e sistemas não sejam consideradas exercícios únicos. Após a aprovação do relatório de qualificação e/ou validação, deve haver um programa contínuo de monitoramento, embasado em uma revisão periódica.

Também segundo o Food and Drug Administration (FDA) dos Estados Unidos, qualquer software usado para automatizar o processo de produção ou do sistema de qualidade deve ser validado para o uso pretendido, conforme exigido pelo regulamento 21 do CFR (Code of Federal Regulations) (FDA, 2002). Esta exigência aplica-se a qualquer software usado para automatizar um projeto de dispositivos, testes, aceitação de componentes, fabricação, rotulagem, embalagem, distribuição, tratamento de reclamações, ou para automatizar qualquer outro aspecto do sistema de qualidade (FDA, 2002). Sobre a análise de 3140 recolhimentos de dispositivos médicos realizados entre 1992 e 1998, o FDA revela que 242 deles (7,7%) são atribuíveis a falhas de software. Desses recolhimentos relacionados a software, 192 (79%) foram causados por defeitos de software, que foram introduzidos devido a mudanças feitas após a sua produção e distribuição inicial. Validação e outras práticas de engenharia de software associadas a boas práticas de desenvolvimento de software são o principal meio para evitar defeitos e recolhimentos (FDA, 2014).

Da mesma forma, para a Agência Regulatória Francesa, a documentação e relatórios de validação devem abranger as medidas pertinentes da vida útil do sistema. Os fabricantes devem ser capazes de justificar as suas normas, os seus protocolos, os seus critérios de aceitação, procedimentos e registros com base na sua avaliação de risco (ANSM - Agence Nationale de sécurité du médicament et des produits de santé, 2013).

Os sistemas automatizados devem ser avaliados periodicamente para garantir que eles permaneçam válidos. Estas avaliações devem incluir se necessário, as funções, os registros de desvios, incidentes, problemas, histórico de atualizações e relatórios de desempenho, confiabilidade, segurança e validação (ANSM - Agence Nationale de sécurité du médicament et des produits de santé, 2013).

Todos os incidentes, e não apenas aqueles relacionados a falhas no sistema e erros de dados, devem ser comunicados às agências reguladoras e avaliados internamente pelas empresas. A origem de um incidente crítico deve ser identificada e servir como base para ações corretivas e preventivas (ANSM - Agence Nationale de sécurité du médicament et des produits de santé, 2013). Quando dados críticos são inseridos manualmente, é necessário fornecer uma verificação adicional de precisão, e este controle pode ser feito por um segundo operador, ou por meios eletrônicos validados. A criticidade e as possíveis consequências de dados errados ou digitados incorretamente devem ser abrangidas pela gestão de riscos (ANSM - Agence Nationale de sécurité du médicament et des produits de santé, 2013).

Na fabricação de remédios, existem basicamente dois fatores que podem influenciar na qualidade do produto: a matéria prima e o processo de fabricação. Neste cenário, produtos fabricados em instâncias não conformes têm maior chance de também estarem não conformes.

A verificação de conformidade em processos de negócios é uma atividade fundamental para empresas sujeitas à regulamentação, pois a não observância dos padrões exigidos pode levar a sérias sanções legais, como multas, suspensão das atividades e até o fechamento da empresa. Após essas suspensões, a agência irá investigar as queixas técnicas e avaliar eventuais penalidades a serem aplicadas. Tais punições variam desde a advertência até o cancelamento da autorização de funcionamento da empresa ou do registro do produto. “Está prevista ainda a aplicação de multas, que podem oscilar entre R\$ 2 mil e R\$ 1,5 milhão” (DOU, 2014), de acordo com a ANVISA (Agência Nacional de Vigilâncias Sanitária). Os exemplos a seguir ilustram este cenário.

O lote 1998101 (validade 11/2015) do medicamento paracetamol 500 miligramas (mg) comprimido, produzido pelo Laboratório Teuto Brasileiro, foi suspenso após denúncia feita ao Procon (Fundação de proteção e defesa do consumidor). Um consumidor identificou que, em uma das cartelas do medicamento,

havia um parafuso no lugar do comprimido. Segundo a ANVISA, o fabricante já iniciou o recolhimento voluntário do lote, que foi distribuído em Goiás, Minas Gerais, no Rio Grande do Sul e na Bahia (DOU, 2014).

O lote 1048105 (validade 6/2015) do medicamento cetoconazol 200mg comprimido, também produzido pelo Laboratório Teuto Brasileiro, foi suspenso após queixa de um consumidor ao SAC da empresa. Na denúncia, o analista informou que, ao abrir a embalagem, constatou a presença de outro produto – o medicamento Atenolol 100mg. Segundo a ANVISA, o fabricante também iniciou o recolhimento voluntário do lote, que foi distribuído em Goiás, no Amazonas, em Alagoas, na Bahia, em Minas Gerais, no Pará, no Rio de Janeiro e em São Paulo (DOU, 2014).

O lote 8910019 (validade 2/2016) do medicamento nistatina 25.000 unidades internacionais por grama (UI/g) 60g, também produzido pelo Laboratório Teuto Brasileiro, foi suspenso depois que um analista relatou que, na cartonagem do medicamento, havia outro produto – neomicina+bacitracina. Segundo a ANVISA, o fabricante informou que o lote em questão foi distribuído no Distrito Federal, Espírito Santo, em Goiás, Minas Gerais e São Paulo (DOU, 2014).

Em 2014, o FDA fez 41 recolhimentos, como por exemplo: 11/24/2014 Aurobindo Pharma EUA, Inc. emite Recolhimentos voluntário em todo o país, de Northstar rótulo gabapentina cápsulas, USP 300mg devido às queixas de cápsulas vazias; 11/20/2014 Baxter inicia recolhimentos voluntário de um lote de injeção altamente concentrada de cloreto de potássio nos EUA devido à falta de rótulo; 10/20/2014 Contrato embalagem Resources, Inc. emite um recolhimento voluntário em todo o país de marca garantida naproxeno sódio comprimidos devido à confusão de embalagens.

Os dados citados demonstram que muitos destes problemas foram frutos de não conformidades no processo de fabricação. Podemos ver casos de remédios com cápsulas vazias, erro na concentração de componentes, produtos misturados na mesma embalagem, vazamento por tampa mal fechada, problemas de esterilidade, contaminação cruzada, dentre outras. Vários destes problemas poderiam ter sido detectados se fosse feita a verificação de conformidade em todas as instâncias do processo.

Apesar de toda regulamentação criada, observa-se que os procedimentos de teste e auditoria nos processos regulamentados ainda são executados utilizando uma

amostra dos dados, mesmo quando informações detalhadas sobre a execução destes processos estão disponíveis em logs de eventos de sistemas de informação, o que torna inaceitável verificar apenas uma amostra de dados (Aalst, 2011).

Evidências mostram que a qualidade do processo de fabricação influencia diretamente a qualidade do produto, mas, no entanto, seria prematuro fazer esta afirmação sem base em dados. Para ser possível fazer esta afirmação, foi preciso buscar o conhecimento dos especialistas do domínio.

De acordo com as respostas dos especialistas foi possível concluir de que há indícios de que a conformidade do processo de fabricação influencia diretamente na qualidade do produto, portanto produtos fabricados em instâncias não conformes do processo de fabricação, tem mais chance de estarem não conformes.

Na indústria farmacêutica, todo processo de negócio que possa impactar a qualidade do produto, a segurança do paciente ou a rastreabilidade dos dados, precisa passar por um processo de testes, chamado validação de sistemas. Nenhum processo é liberado para uso sem este conjunto de testes.

A validação de sistemas computadorizados para a indústria farmacêutica segundo a ANVISA (Agência Nacional de Vigilância Sanitária, 2010) é definida como:

- Ato documentado que atesta que qualquer procedimento, processo, equipamento, material, atividade ou sistema realmente e consistentemente leva aos resultados esperados;
- Validação de processo (VP): evidência documentada que atesta com um alto grau de segurança que um processo específico produzirá um produto de forma consistente, que cumpra com as especificações pré-definidas e características de qualidade;
- Validação de sistemas computadorizados: evidência documentada que atesta com um alto grau de segurança que uma análise de sistema computadorizado, controles e registros são realizados corretamente e que o processamento dos dados cumpre com especificações pré-determinadas;

Segundo o GAMP – Good Automated Manufacturing Practice (ISPE - International Society of Pharmaceutical Engineer, 2008), um sistema computadorizado é um conjunto de hardware, software e componentes de rede, junto com funções de controle e documentação associada, como mostrado na Figura 1-1.

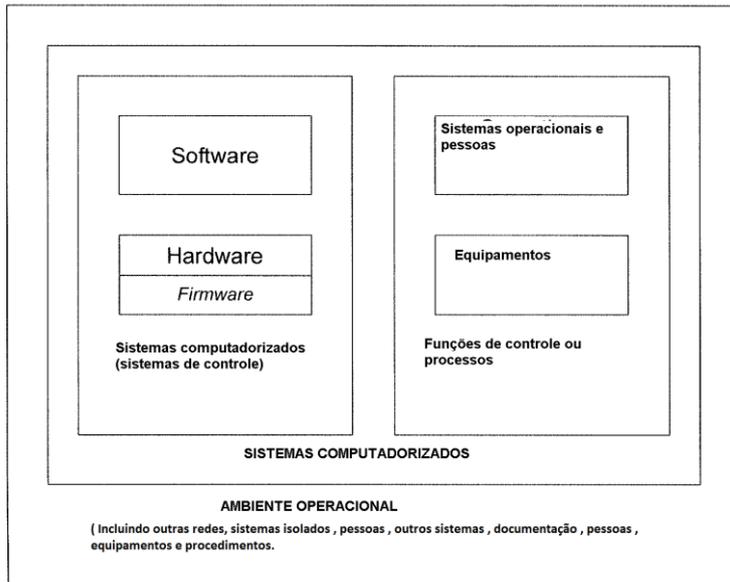


Figura 1-1 - Sistema Computadorizado GAMP

1.2. Problema e hipótese

Apesar de os dados para verificar todas as instâncias do processo estarem disponíveis, não é possível fazer o mesmo com o produto, pois o teste de 100% do produto significaria destruí-lo, então como podemos auxiliar o especialista do domínio a escolher quais os produtos têm mais chance de ter problemas de fabricação? Neste cenário, o problema a ser tratado nesta pesquisa é: Como identificar dentre um conjunto de medicamentos fabricados, aqueles que tem mais chance de ter problemas de qualidade, sem que para isto tenha que se destruir todo o produto?

A hipótese é: SE o método baseado em mineração de processos proposto for aplicado, ENTÃO as instâncias do processo serão classificadas como conformes ou não conformes, e as não conformidades serão classificadas. Se esta hipótese se confirmar, o analista terá uma fonte de informação mais segura para identificar os produtos com mais chances de ter problemas de qualidade diminuindo a aleatoriedade da escolha das amostras para os testes de qualidade. Esta hipótese é falseável, se através da aplicação do método proposto possível classificar as instâncias do processo.

1.3. Enfoque de Solução e Metodologia de Pesquisa

A solução proposta é um método para descobrir e classificar as não conformidades nas instâncias do processo de forma a indicar quais não conformidades

têm mais chance de ter gerado produtos fora do padrão, de forma sistemática e automática. Esta classificação pode ajudar o especialista do domínio a escolher quais os produtos testar para evitar que produtos fora do padrão cheguem ao mercado podendo provocar danos à saúde dos pacientes e à imagem da empresa.

Através da mineração de processos é possível descobrir, monitorar e melhorar processos reais, extraindo-se conhecimento a partir de logs de eventos disponíveis gerados pelos sistemas de informação. A mineração inclui a descoberta automatizada de processos, verificação de conformidade, descoberta de rede social, mineração organizacional, construção de modelos de simulação, extensão de modelo, reparação de modelo, predição, e recomendações (Process Mining Task Force, 2014).

A abordagem metodológica adotada foi primeiramente um estudo exploratório para verificar se o método proposto seria aplicável e se poderia ser melhorado. Como resultado, algumas mudanças foram feitas no método principalmente no que se refere à identificação única do processo, pois, no início do estudo considerava-se que o número do lote seria o identificador, no entanto verificou-se que é necessário que o especialista defina em cada processo, pois alguns processos usam outros identificadores. Após a verificação de que o método proposto poderia ser aplicado no domínio e do refinamento do mesmo com a experiência obtida neste estudo exploratório, foi iniciado um estudo de caso com uma massa de dados de um ano de fabricação do medicamento x fabricado neste processo, aplicando o método, com o objetivo de obter indícios de que é possível identificar e classificar instâncias não conformes do processo.

1.4. Organização do trabalho

O trabalho está organizado em cinco capítulos, além desta introdução.

No Capítulo 2, são apresentados os principais conceitos que fundamentam a pesquisa. Por se tratar de um estudo de caso utilizando técnicas de mineração na indústria farmacêutica alguns conceitos têm interpretação específica para este domínio, e portanto, podem ser diferentes do conceito tradicional.

No Capítulo 3, são apresentados trabalhos relacionados com o tema verificação de conformidade utilizando mineração de processos e um comparativo entre eles e o método proposto.

No Capítulo 4, é apresentado o método proposto descrito em quatro fases, levantamento de informações sobre o domínio e o processo, levantamento dos atributos obrigatórios no log e o pré-processamento, terceira fase a mineração do processo, quarta fase a verificação de conformidade e classificação.

No Capítulo 5, é apresentada a avaliação de resultados obtida com o estudo exploratório e o estudo de caso.

Por fim, o Capítulo 6 resume o trabalho, apresenta as conclusões, contribuições do trabalho, as limitações do método e as perspectivas futuras de pesquisa.

2. Referencial teórico

O referencial teórico tem como objetivo apresentar os conceitos que vão embasar a dissertação. As definições apresentadas são específicas para o domínio da indústria farmacêutica e da mineração de processos, portanto alguns conceitos são diferentes dos empregados normalmente em sistemas de informação.

2.1. Processo de Negócio

Processos de negócios são atividades que as empresas executam sempre que entregam um serviço ou um produto para os seus clientes. Os processos são projetados e executados de maneira a impactar tanto a qualidade de serviço percebida pelos clientes, quanto a eficiência com que os serviços são prestados. Uma organização pode superar outra que oferece os mesmos tipos de serviços se ela tiver melhores processos e executá-los melhor. Isto é verdade não só para os processos voltados diretamente ao cliente, mas também para os processos internos, tais como, o processo de pagamento, que é realizado com a finalidade de cumprir uma necessidade interna da organização (Dumas, et al., 2013).

Processo de negócio consiste em um conjunto de atividades executadas em uma organização. Estas atividades são coordenadas para juntas atingirem o objetivo do negócio. Cada processo de negócio é inato de uma organização, mas pode interagir com processos de negócios de outras organizações (Aalst & Sthal, 2011).

2.2. Modelagem e Descoberta de Processos

A modelagem de processos de negócio consiste na técnica de representar, de forma abstrata, como o processo acontece no mundo real, compreendendo características intrínsecas, como os recursos, controles, papéis e responsabilidades, entre outras (Santos, 2014). Modelos de processos de negócios podem ser descritivos ou prescritivos. Modelos descritivos tentam capturar os processos existentes sem ser normativos. Um exemplo é um processo de um hospital cujo objetivo é reagir a uma situação urgente, entretanto, a flexibilidade para divergir do fluxo normal das ações é crucial. Outro exemplo pode ser o modelo que é feito para documentar certos

procedimentos no sistema financeiro (com o log de eventos as atividades que é executado sem um modelo de processo explícito). Claramente isto é desejável para manter o modelo alinhado com o processo atual do sistema financeiro, através do uso de técnicas de verificação de conformidade.

Modelos Prescritivos descrevem a maneira que o processo pode ser executado. Em um Sistema de Workflow (WFM - WORKFLOW MODELS), modelos prescritivos são usados para reforçar um caminho particular de trabalho em sistemas de informação. Entretanto, analistas muitas vezes precisam desviar de seu fluxo normal em relação aos modelos mais rígidos dos sistemas WFM. Na maioria das situações, modelos prescritivos não são usados diretamente pelos sistemas de informação. Por exemplo, modelos de referência de contexto em SAP (SAP, 2015) e ARIS (ARIS, 2015) descrevem o caminho preferencial que o processo pode ser executado, mas os analistas do SAP (SAP, 2015) podem adaptar os modelos de referência.

A descoberta de processos é definida como o ato de coleta de informações sobre um processo existente e sua organização em um modelo (Dumas, et al., 2013). Deste modo, é uma atividade muito mais ampla do que uma modelagem, que é claramente uma parte desta atividade. Na verdade, a coleta de informações, muitas vezes revela-se complicada, e na prática demorada (Dumas, La Rosa, Mendling, & A. Reijers, 2013). Os processos de negócio são dinâmicos e complexos exigindo técnicas mais sofisticadas do que simplesmente a modelagem. Este é um dos principais motivadores da pesquisa em mineração de processos.

A ideia fundamental da mineração de processos é a existência de um processo desconhecido, mas o ambiente e o log de eventos são conhecidos, com isto pode-se alimentar uma técnica de mineração de processos e no final obter um modelo de processos do ambiente observado (C.W. Günther, 2007).

As técnicas de mineração de processos podem ser utilizadas para extrair informações de logs de eventos relacionadas a processos, para fazer descoberta automática de modelos de processos de sistemas corporativos (Aalst, 2011). O log de eventos é o registro da execução do processo. O evento neste log refere-se especificamente a uma atividade do processo de negócio, deve indicar a instância do processo (caso) e deve estar totalmente ordenado (Rozinat & Aalst, 2008). A principal razão pela qual a mineração de dados tem atraído muita atenção na indústria nos

últimos anos é a disponibilidade de grandes quantidades de dados e a necessidade de transformar esses dados em informações úteis e conhecimento (Han & Kamber, 2000).

A ideia da mineração do processo é descobrir, monitorar e melhorar processos reais, extraíndo conhecimento de logs de eventos disponíveis. O processo de mineração inclui a descoberta do processo automatizado, extração de modelos de processo de um log de eventos, verificação de conformidade, monitoramento de desvios, comparando o modelo e log, descoberta automática de rede social ou organizacional, construção automatizada de modelos de simulação, extensão de modelo, reparação de modelo, previsão de caso e recomendações com base em história (van der Aalst, 2012). Nas seções seguintes são apresentados alguns conceitos importantes relacionados às técnicas de mineração de processos.

A mineração de regras tem como objetivo encontrar regras em logs de eventos, algumas técnicas podem ser utilizadas para isto, o LTL (linear temporal logic) CHECKER (F.M. Maggi, 2011) é um algoritmo de mineração de regras, que procura nos registros se o atributo assume um valor ou se o atributo nunca assume um determinado valor. A verificação de conformidade compara um modelo a priori com a realidade observada armazenada em um log, tornando possível visualizar e quantificar as discrepâncias detectadas. O LTL Checker pode ser usado no caso em que não existe um modelo de processo a priori completo, mas apenas um conjunto de requisitos (por exemplo, as regras de negócios) (Process Mining Task Force, 2014). Um exemplo de aplicação deste algoritmo encontra-se em 5.1.3.

O FILTRO DE ATRIBUTO (Fluxicon, 2013) filtra a Variante para uma determinada atividade podendo ser utilizada da seguinte forma, a atividade sempre está presente, a atividade nunca está presente. O resultado é o conjunto de instâncias que atendem ao filtro.

O FILTRO DE PRECEDÊNCIA (Fluxicon, 2013) é utilizado para cadastrar as regras de precedência, pode-se filtrar uma sequência das atividades. Por exemplo: Atividade A precede Atividade C e Atividade E precede Atividade G, assim por diante. O resultado é o conjunto de instâncias que atendem ao filtro.

O FILTRO DE TEMPO (Fluxicon, 2013) retorna as instâncias que atendem os requisitos de tempo cadastrados. Por exemplo para filtrar as atividades que ultrapassaram uma determinada duração indicada no filtro. Funciona com regras do

tipo SE a duração atividade $A \geq 10$ min então retorna as instâncias em que a regra é atendida. Pode-se também usar regras do tipo SE a atividade A ocorre entre 10:00 e 12:00 então retorna as instâncias em que a condição seja verdadeira.

2.2.1.Variante de Processo

Durante o processo de descoberta, são encontradas várias instâncias repetidas e para melhorar a visualização e análise, as ferramentas de mineração agrupam estas instâncias semelhantes em Variantes. Variante (variant) como aparece na ferramenta Disco (Fluxicon, 2013) é o nome dado ao conjunto de instâncias que contém as mesmas atividades na mesma sequência, sem levar em conta o TimeStamp (Aalst, 2011). Na ferramenta Disco (Fluxicon, 2013) estes conjuntos são identificados e numerados. A Figura 2-1 exemplifica como as instâncias semelhantes são agrupadas.

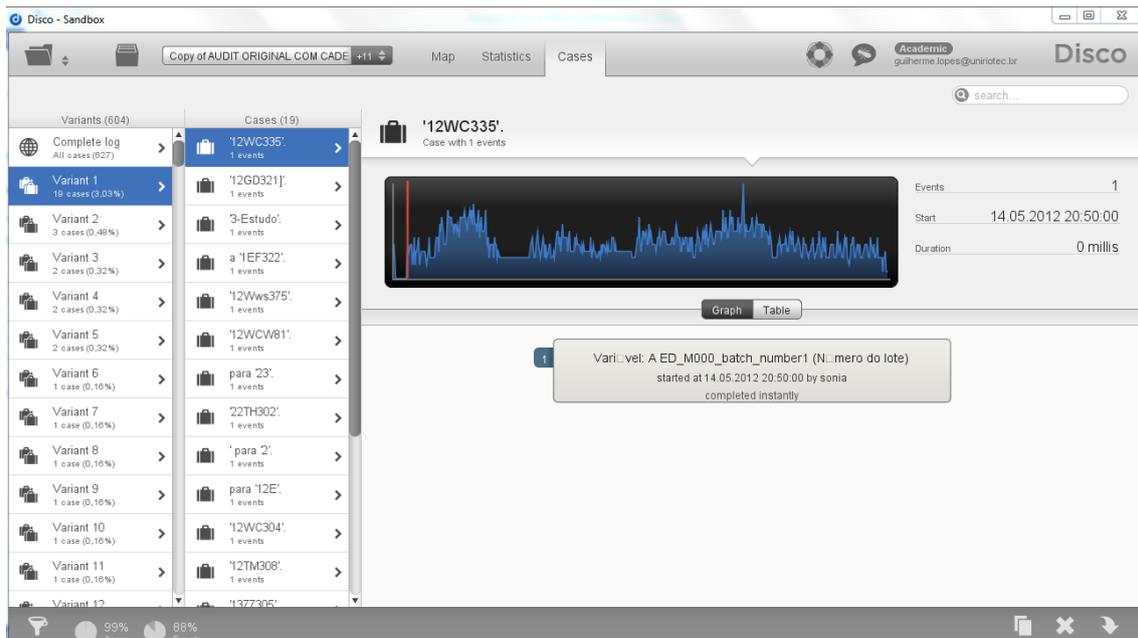


Figura 2-1 Exemplo de Variante na ferramenta Disco

2.2.2.Play in

É a técnica utilizada para descobrir processos através de um log de eventos. O log é minerado com um algoritmo de descoberta, e o resultado é o modelo do processo. (Aalst, 2011).

2.2.3.Play out

É a técnica utilizada para gerar o log de eventos a partir do modelo de processos. O processo é executado e cada possibilidade de execução gera uma log da instância do processo (Aalst, 2011).

2.2.4.Replay

A técnica de Replay utiliza o log de eventos e o modelo de processo para ser executada. O log é forçado a ser executado através do modelo do processo, mesmo que o modelo e log não sejam compatíveis, as atividades que faltam ou sobram durante a execução são contabilizadas, e a partir do cálculo do fitness é obtido um índice de conformidade entre o log e o modelo.

2.2.5.Algoritmo Fuzzy Miner

Um exemplo de técnica de mineração é o algoritmo FUZZY Miner (C.W. Günther, 2007). Este algoritmo utiliza técnicas de simplificação para evitar o resultado de processos do tipo “spagetti”, utiliza-se a agregação, abstração, ênfase e customização.

O trabalho de Günther explora como usar a simplificação e a complexidade na visualização em processos pouco estruturados. Para fazer isto é preciso um critério de decisão baseado na simplificação e a visualização dos modelos de processo, para isto foram identificadas duas métricas para a tomada de decisão, Importância e Correlação.

A Importância pode ser determinada tanto para as classes de eventos (atividades) quanto para relações binárias de precedência (as bordas). Mede a importância relativa do comportamento, ele especifica o nível de interesse que temos nos eventos, ou na sua ocorrência. Um exemplo para a medição de importância é a frequência, eventos ou relações de precedência que são observados mais frequentemente são considerados mais significativos.

Correlação por outro lado, só é relevante para as relações precedência sobre eventos. Ela mede o quanto são relacionados dois eventos seguidos. Exemplos para medição da correlação determina-se a sobreposição de atributos de dados associados a dois eventos que se sucedem, ou comparando a similaridade de seus nomes. Eventos são considerados correlacionados quando compartilham uma grande

quantidade de dados, ou têm semelhança expressa em seus nomes (por exemplo, "verificar a aprovação do pedido do cliente " e "aprovar o pedido do cliente"). Com base nestes dois parâmetros, os quais foram definidos especialmente para esta finalidade, a abordagem para o processo de simplificação é a seguinte:

- Comportamento altamente significativo é preservado, ou seja, contido no modelo simplificado.
- Comportamento menos significativo, mas altamente correlacionado é agregado, escondido em clusters no modelo simplificado.
- Comportamento menos significativo e menos correlacionado é removido do modelo simplificado.

Esta abordagem pode reduzir muito e concentrar o comportamento exibido, empregando os conceitos de agregação e abstração. Com base nesse modelo simplificado, podemos empregar o conceito de ênfase, destacando o comportamento mais significativo (C.W. Günther, 2007).

Concluindo, o algoritmo Fuzzy Miner é muito adequado para minerar logs com muitos ruídos, ou com muitas variações nas instâncias, devido à sua capacidade de simplificação do modelo, o que facilita o entendimento.

2.2.6.Redes de Petri

Redes de Petri é uma técnica de modelagem de processos, que usa um tipo especial de grafo contendo dois tipos de nós: posição e transição podem ser conectadas diretamente por arcos. Na teoria dos grafos, é chamado de bipartite. Em notação gráfica, a posição é representada por um círculo e a transição por um quadrado. Um arco somente pode conectar uma posição para uma transição ou vice-versa. Um arco entre duas posições ou duas transições é impossível (Aalst & Sthal, 2011). Na Figura 2-2, podemos ver um exemplo desta representação.

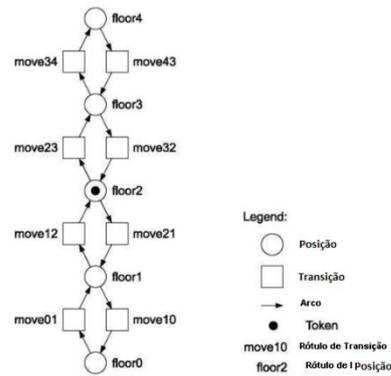


Figura 2-2- Exemplo de uma Rede de Petri

Pode-se utilizar a representação gráfica Redes de Petri também na descoberta de processos, por exemplo, utilizando o α algoritmo ou WF-Nets (Aalst & Sthl, 2011) para descobrir os processos através dos logs de eventos e depois representá-los como Redes de Petri.

2.3. Dimensões da qualidade de um modelo de processo de negócio

Várias abordagens têm sido utilizadas para enfrentar os problemas de completude e ruído. Falta de completude e ruído se referem à qualidade de log de eventos, mas não dizem muito sobre a qualidade do processo descoberto. Determinar a qualidade do resultado da mineração de processos é difícil e pode ser caracterizada por várias dimensões (Aalst, 2011). Neste trabalho são discutidas quatro dimensões: fitness, simplicidade, precisão e generalização, mostradas na Figura 2-3.

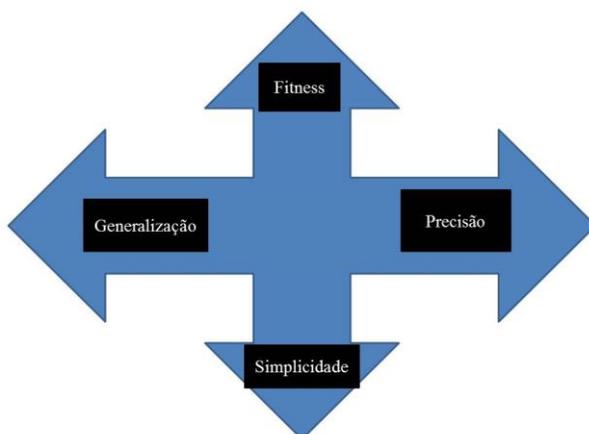


Figura 2-3 - 4 dimensões de qualidade do modelo de processo (Aalst, 2011)

2.3.1. Fitness

Um modelo com bom fitness permite que o comportamento seja compatível com o log de eventos. Um modelo tem o fitness perfeito se todas as

atividades das instâncias podem ser reproduzidas pelo modelo do início ao fim. O fitness pode ser definido como a porção das instâncias do processo no log de eventos que podem ser totalmente reproduzidas no modelo (Aalst, 2011).

Pode ser definido baseado na fração dos padrões de eventos representados pelo modelo ou baseados na fração de instâncias que podem ser repetidas no modelo (Dumas, et al., 2013).

O fitness indica o quanto o comportamento observado no log se encaixa no modelo de processos descoberto ou modelado. Por exemplo, se as instâncias do log ,forem executadas no modelo Figura 2-4, elas se encaixarão perfeitamente, então o fitness = 1. Quanto mais perto de 1 melhor o fitness, quanto mais perto de zero pior o fitness.

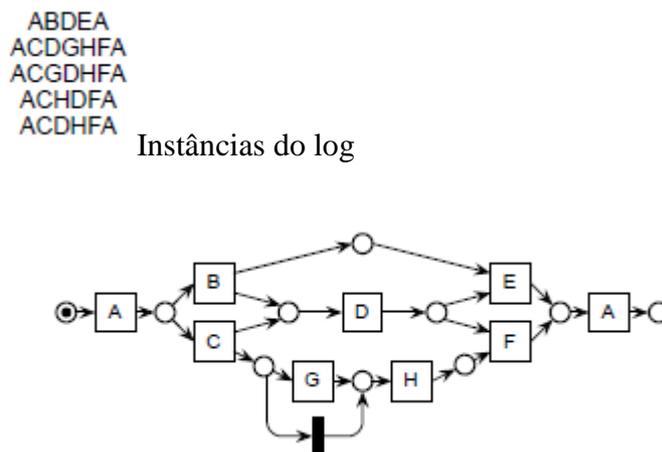


Figura 2-4 - Modelos de processo

2.3.2. Precisão

O modelo descoberto não deve permitir um comportamento completamente diferente do que foi visto no log de eventos (Aalst, 2011). Precisão é referente ao grau de comportamento que é permitido pelo modelo, mas não observado no log. Pode-se facilmente criar um modelo de processos que permite a execução de todas as tarefas em qualquer ordem arbitrária com repetição potencial, entretanto cada especificidade deve ser aprendida e estar refletida no modelo (Dumas, et al., 2013).

O log abaixo pode ser executado perfeitamente no modelo da Figura 2-5 no entanto a generalização é ruim, pois qualquer mudança de ordem na execução das

atividades não vai ser executada no modelo, ou instâncias com atividades diferentes não conseguirão ser executadas no modelo.

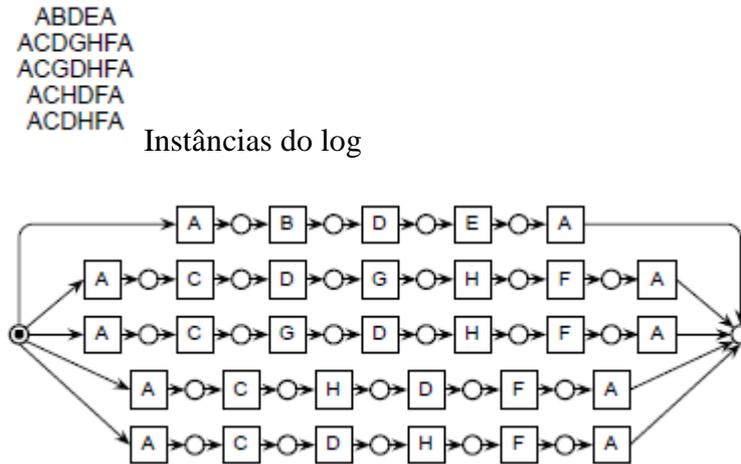


Figura 2-5 - Modelo de processo precisão alta

2.3.3.Generalização

O modelo descoberto deve generalizar o comportamento do exemplo visto no log de eventos (Aalst, 2011).

Generalização refere-se à capacidade de um modelo de processo resumir o comportamento que está documentado nos registros. Uma técnica de descoberta que é capaz de generalizar ajuda a trabalhar com comportamento incompleto (Dumas, et al., 2013).

No modelo da Figura 2-6 pode executar o log em qualquer sequência

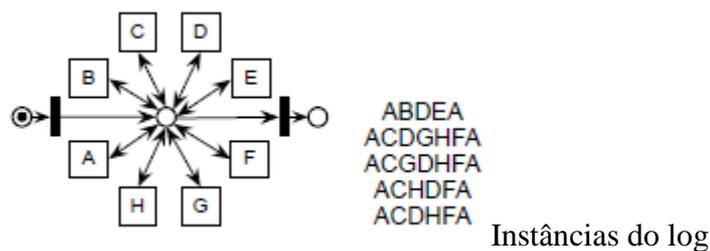


Figura 2-6 - Modelo de processo generalização

2.3.4.Simplicidade

O modelo descoberto deve ser o mais simples possível (Aalst, 2011). Simplicidade significa que o modelo de processo resultante deve ser facilmente

compreensível. Ela pode ser medida usando diferentes métricas de complexidade para os modelos de processo, como o tamanho do modelo ou grau de desestruturação (Dumas, et al., 2013).

O modelo da Figura 2-7 tem alta simplicidade, no entanto o fitness é ruim pois nem todas as instâncias do log conseguirão ser executadas no modelo.

ABDEA
ACDGHFA
ACGDHFA
ACHDFA
ACDHFA

Instâncias do log.

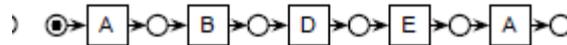


Figura 2-7 - Modelo de processo com alta simplicidade

2.4. Verificação de conformidade de processos

Segundo o trabalho realizado por Rozinat (Rozinat, 2010) muitas empresas têm adotado sistemas de TI para apoiar suas atividades de negócios. Estes sistemas tipicamente registram os eventos relacionados à execução dos processos de negócio, ao mesmo tempo, modelos explícitos de processos descrevem como os processos deveriam ser executados estão frequentemente disponíveis. Junto com os dados gravados, uma questão se coloca: “O modelo e o log estão conformes um com o outro?”. A verificação de conformidade refere-se à análise de conformidade e também a detecção de inconsistências entre o modelo de processos e o correspondente log de execução e a quantificação através de métricas (Rozinat, 2010).

A verificação de conformidade de processos de negócios tem como objetivo comparar um modelo existente com os dados de log extraídos da execução destes processos a partir dos sistemas de informação. Este modelo também pode ser um conjunto de regras de negócios. A classe de algoritmos de mineração de processos responsável pela análise de conformidade de processos é chamada de técnicas de verificação de conformidade, porque estes algoritmos verificam a conformidade de algum modelo já existente com a realidade (Rozinat, 2010). Um modelo pode ser construído manualmente ou descoberto, o processo pode ser normativo ou, descritivo,

a verificação de conformidade relaciona os eventos nos logs de eventos e o modelo de processos, e compara os dois (Aalst, 2011).

Na Figura 2-8, observa-se uma ciclovia com obstáculo feito para diminuir a velocidade dos ciclistas, este é o processo projetado, ao lado, existe um caminho alternativo que corresponde à realidade. O que a verificação de conformidade busca é a comparação do que foi projetado com o ambiente real, utilizando medidas de conformidade (Aalst, 2014).

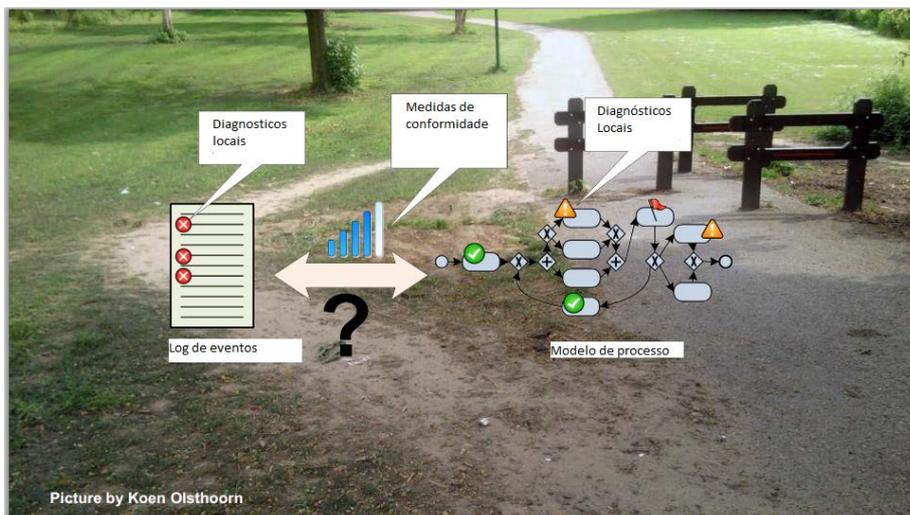


Figura 2-8 - Visão da Verificação da conformidade (Aalst, 2011)

Para ambos os modelos descritivo (1) ou prescritivo (2), é importante a capacidade de medir e localizar desvios. O problema da percepção da conformidade pode ser visto de dois ângulos:

1. O modelo assume ser correto porque representa o caminho que o processo de negócios deve ser realizado, e as não conformidades gravadas nos logs, correspondem a um comportamento indesejável.
2. O log de eventos pode assumir ser correto porque isto é o que realmente acontece, e o modelo de processo pode estar ultrapassado ou simplesmente não adaptado às necessidades dos funcionários para a execução das tarefas. Destacando esta questão, é mais fácil redesenhar o modelo e, portanto, aumentar a transparência.

Como consequência, uma interpretação final de um problema de conformidade, só pode ser dada por um especialista de domínio, mesmo se o modelo e o log estiverem aparentemente conformes. Pode ser útil entregar essa evidência para acelerar os processos de certificação. Além disso, um modelo validado através da

verificação de conformidade pode ser o ponto de partida para outros tipos de análise, como simulação. Os dados quantitativos extraídos do log podem ser projetados no modelo como frequências, probabilidades, gargalos, etc.

Hoje em dia, os algoritmos de descoberta de processos estão disponíveis para ajudar o analista de processos na concepção de um modelo de processo inicial para a implantação de um sistema WFM. O objetivo de um algoritmo de descoberta de processos é construir automaticamente um modelo de processo que reflete o comportamento observado no registro de eventos. Enquanto todas as abordagens de descoberta de processos visam a descoberta de um "bom" modelo de processo, muitas vezes são alvo os desafios como loops e tarefas duplicadas, e algumas vezes o algoritmo tem limitações específicas. Medidas de qualidade diferentes são utilizadas para determinar a qualidade do processo descoberto. Um mesmo processo é visto de forma diferente pelos algoritmos como mostra a Figura 2-9.

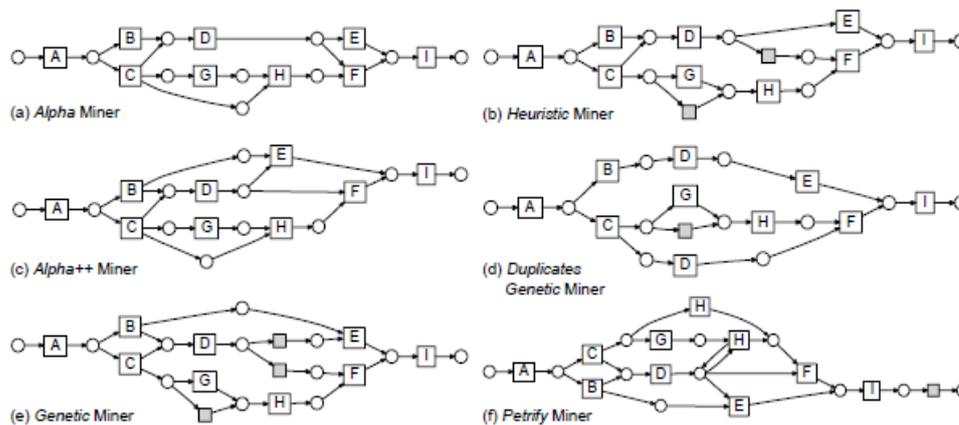


Figura 2-9 - Processos minerados por diferentes algoritmos (Aalst, 2011)

Para ilustrar o dilema, considera-se um exemplo simples de log, que contém apenas as seguintes cinco instâncias distintas: ABDEI, ACDGHFI, ACGDHFI, ACHDFI, e ACDHFI. Aplicam-se seis algoritmos de mineração de processos diferentes disponíveis no framework PROM (Process Mining Task Force, 2014) que obteve seis modelos de processos diferentes (para cada plug-in ou algoritmo, foi utilizada a configuração padrão no ProM) (Process Mining Task Force, 2014). A Figura 2-10 apresenta o resultado da mineração utilizando uma rede de Petri para os algoritmos Alpha, Heurística, a ++ Alpha, Genetic e Petrify. Os modelos parecem semelhantes, mas são todos diferentes. Eles são equivalentes? Se não, qual é o melhor "?

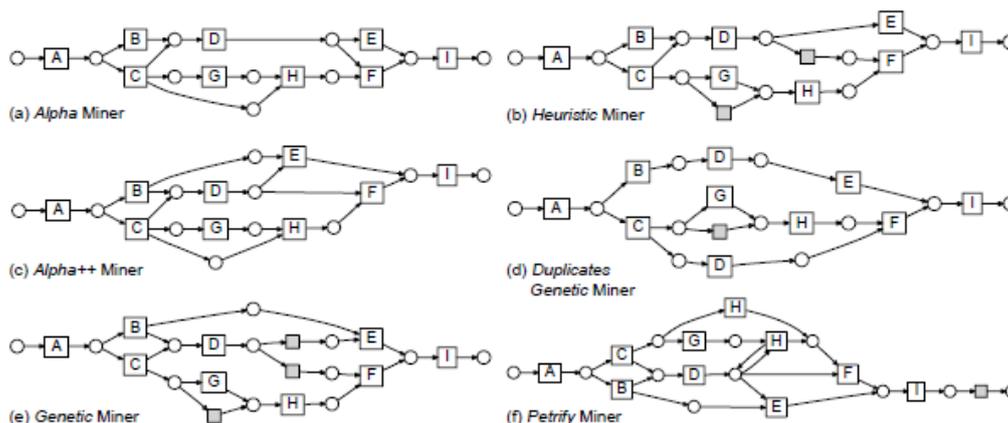


Figura 2-10 - Modelo de processo descoberto por diferentes algoritmos

Os analistas precisam saber se o modelo descoberto descreve de maneira correta a realidade. Por exemplo, se o processo minerado pode ser usado como uma ferramenta de base de conhecimento, deve ser possível estimar a acurácia do modelo descoberto, isto é a confiança com que o modelo reflete as entrelinhas do processo. Além disso, o analista precisa ser capaz de comparar os resultados obtidos por diferentes algoritmos de descoberta.

Resumindo, a verificação de conformidade é relevante tanto para processos modelados como para processos descobertos. Uma das métricas utilizadas para medir a conformidade é o fitness.

Existem várias técnicas de descoberta de processos e de verificação de conformidade. As principais técnicas serão descritas a seguir.

2.5. Principais técnicas para a verificação de conformidade de processos

Nos próximos tópicos serão enumeradas as principais técnicas de verificação de conformidade encontradas na literatura.

2.5.1. Verificação de conformidade utilizando alinhamento entre observado e modelado

A verificação de conformidade não deve impor restrições na notação de processos como, por exemplo, transições silenciosas, ou seja, duas transições com o mesmo nome. Deve-se prover um caminho mais próximo do perfeito para cada instância do processo e do modelo. Também deve ser feita uma análise de desempenho, além da análise de replay e fitness.

Uma vantagem do alinhamento é que ele é independente do modelo de descoberta. Um alinhamento perfeito é aquele em que o log se encaixa perfeitamente ao modelo, no entanto existem situações onde há sincronismo e outros onde não há correspondência dos eventos entre o log e o modelo (Aalst, 2014), como se pode observar na Figura 2-11.

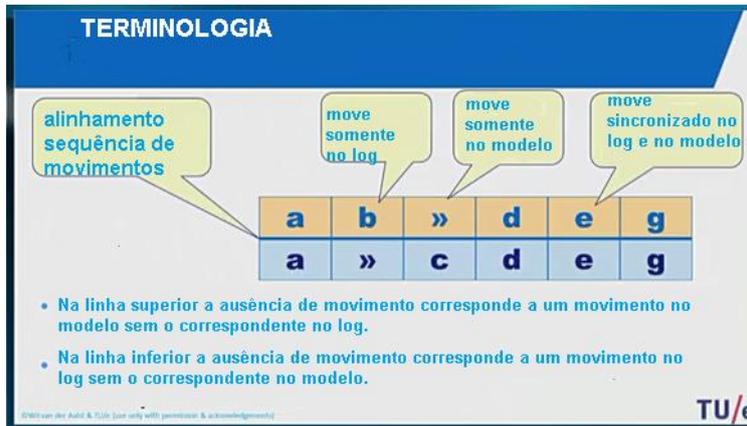


Figura 2-11 - Terminologia utilizada no alinhamento de log e processo (Aalst, 2011).

Visto que pode-se ter diferentes níveis de alinhamento, então é preciso saber qual alinhamento é melhor. Para isto cria-se uma função de custo para classificar o melhor alinhamento. Para cada situação de desvio entre o log e o modelo define-se um custo, e a função de custo é incrementada toda a vez que não há correspondência entre o log e o modelo. O símbolo » significa que não há correspondência entre o log e o modelo, ou seja o valor esperado não aconteceu ou no log ou no modelo. Na Figura 2-12 mostra como é feito o cálculo do custo , a linha de cima de mostra o log e a debaixo o modelo , quando um dos dois não coincide vê-se o símbolo >> e o custo é incrementado de uma unidade , o procedimento é repetido até o fim da execução do log.

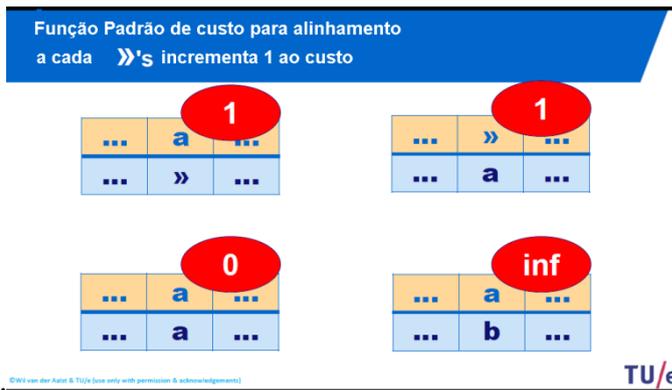


Figura 2-12 – Função Padrão de Custo para alinhamento (Aalst, 2011)

A partir do conhecimento do custo por movimento do token, calcula-se o custo do alinhamento entre cada instância do log e o modelo, conforme a Figura 2-13. Normalmente o alinhamento com o menor custo é o melhor, no caso de instâncias com o mesmo custo de alinhamento, o especialista do domínio pode tomar a decisão de qual alinhamento é melhor para o seu objetivo.

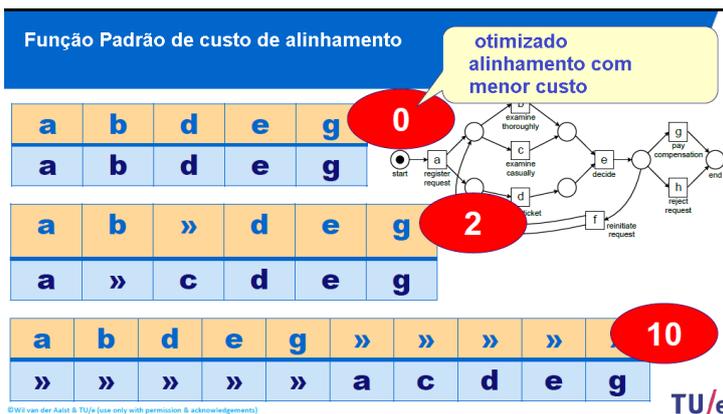


Figura 2-13 - Custo padrão para determinar o melhor alinhamento (Aalst, 2011)

As vantagens do alinhamento baseado em custo entre o log e o modelo são: comportamento observado é diretamente relacionado ao comportamento modelado; muito flexível, aceita qualquer tipo de estrutura de custo; provê diagnóstico detalhado; depois de alinhados o log e o modelo, as outras dimensões de qualidade (fitness, simplicidade, generalização, precisão) podem ser levadas em conta para desempatar quando o custo de alinhamento entre diferentes modelos é igual.

2.5.2. Verificação de conformidade utilizando Token-Based Replay

Na técnica Token-Based Replay, o log de eventos é minerado utilizando Redes de Petri. Após definido o processo padrão, um token é liberado no início do processo e vai percorrendo as atividades listadas em cada instância, se uma atividade

encontrada no log não é encontrada no processo, então temos um token perdido, se chegar a um ponto que o token é criado então temos a criação de um token. No final da execução do processo, contam-se os tokens perdidos e criados, e calcula-se o valor do fitness, ou seja o percentual de conformidade do processo (Aalst, 2014). Neste contexto, o cálculo do fitness é realizado da seguinte forma:

Event log = E
 Redes de Petri = PN
 S= conjunto de instâncias do log
 Para cada sequência de eventos $\sigma \in S = \alpha(E)$ tal que

m_σ é o número de tokens perdidos
 r_σ é o número de tokens remanescentes.
 c_σ é o número de tokens consumidos.
 p_σ é o número de tokens produzidos.

O fitness f é definido como:

Equação 1 - Fórmula do fitness

$$f = \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in S} S(\sigma) - m_\sigma}{\sum_{\sigma \in S} S(\sigma) - c_\sigma} \right) + \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in S} S(\sigma) - r_\sigma}{\sum_{\sigma \in S} p_\sigma} \right)$$

O fitness sempre vai assumir valores entre 0 e 1.

Para ilustrar o processo de replay, o exemplo na Figura 2-14, vê-se que a instância ABDEA é executada sem nenhum problema, não há tokens perdidos ou criados durante a execução.

O requisito dominante na verificação de conformidade é o fitness. Um log de eventos e uma Rede de Petri se encaixam se a Rede de Petri puder gerar cada instância no log. Em outras palavras, a Redes de Petri precisa ser capaz de gerar cada instância do log. É possível quantificar o fitness utilizando o log de eventos e a Redes de Petri, executando o log na rede de Petri e verificando as discrepâncias, para cada atividade existente no log e não existente na rede um token é produzido e a cada atividade existente na rede e não existente no log um token é perdido, utilizando estes valores na fórmula do fitness é possível calcular o índice de fitness para cada instância ou para todo o log.

Uma maneira de medir o fitness entre o log de eventos e o modelo de processos é fazendo o replay do log de eventos no modelo e medir as

incompatibilidades entre o modelo e o log. O replay de cada sequência inicia com uma marca inicial, uma posição no modelo, então a partir daí as transições são disparadas através dos tokens, à medida que as instâncias são executadas no modelo.

Existem diferentes maneiras de medir os erros no curso do replay, podem ser contados os tokens que foram criados artificialmente e o número de tokens perdidos no modelo.

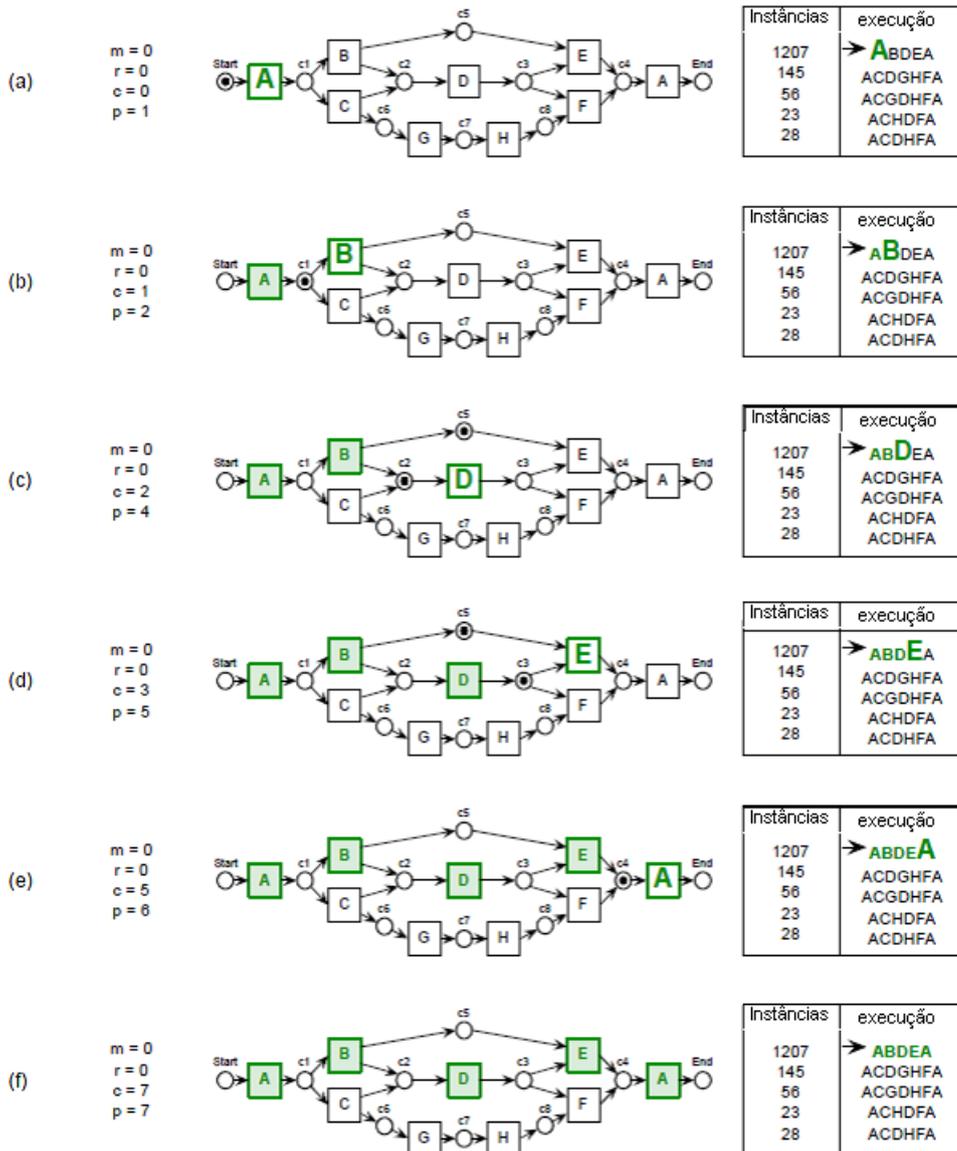


Figura 2-14 - Exemplo de replay (Rozinat, 2010)

No exemplo do replay, podemos observar que o $m=0$, $r=0$ e $c=p$, ou seja não há tokens perdidos ou remanescentes, e todos os tokens produzidos foram consumidos durante o replay da instância, isto caracteriza uma instância conforme.

Já no exemplo da Figura 2-15 podemos observar que o log não é executado perfeitamente no modelo a instância ACHDFA.

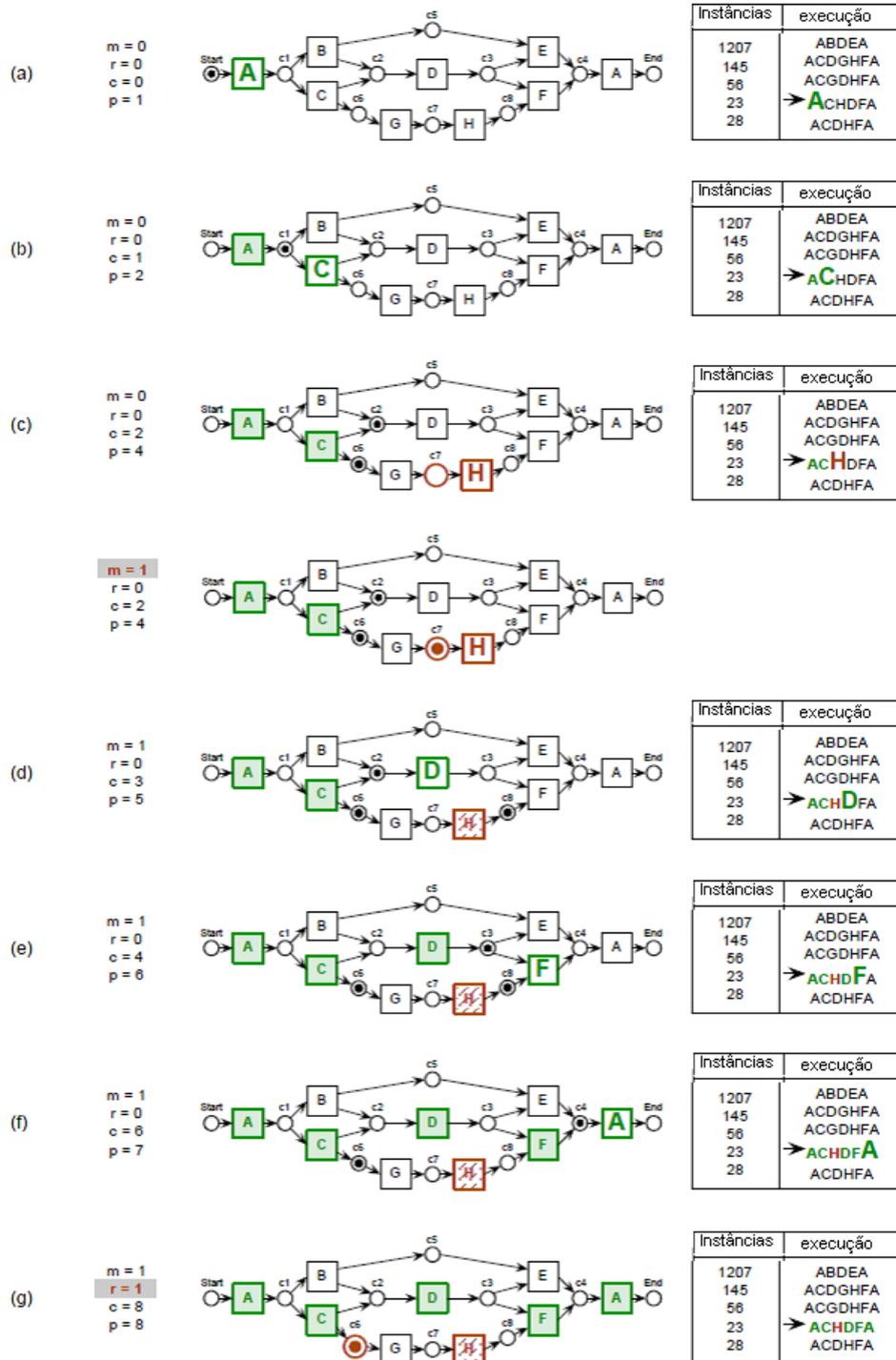


Figura 2-15 - Replay de instância não conforme (Rozinat, 2010)

Ao fim da execução da instância na Figura 2-15, observa-se que $m=1$ e $r=1$, apesar de $c=p$ temos tokens que foram perdidos e tokens que não foram consumidos no replay do processo o que indica que para ser executado tiveram que ser

forçados pois a atividade G não existia no log e H só poderia ser executada depois de G, para isto foi criado um token para que o disparo de execução pudesse ser feito, no entanto como G não é atividade do log, este token não é consumido. Por outro lado para que a atividade D pudesse ser executada, foi necessário produzir um token, pois não existe correspondente no modelo, já que esta atividade está fora da sequência de execução. No final, apura-se que em C6 tivemos um token não consumido e no fim do processo teremos um token remanescente, o que indica uma não conformidade, pois em uma execução conforme todos os tokens precisam ser consumidos, não podemos ter tokens faltando (senão o processo não termina), e nem sobrando (o que indica que atividades do modelo não existem na instância do log).

Portanto se executarmos todas as instâncias do log e aplicarmos na fórmula, teremos:

$$f(M1, L2) = \frac{1}{2} \left(1 - \frac{51}{10666}\right) + \frac{1}{2} \left(1 - \frac{51}{10666}\right) \approx 0.995.$$

Com este cálculo, temos um indicador de quanto o log se encaixa no modelo de processo, este indicador é muito útil para verificar se o processo modelado corresponde à realidade, ou no caso do processo descoberto o quanto os logs se encaixam em um processo padrão.

Na verificação de conformidade utilizando a Redes de Petri, a partir de um processo modelado ou descoberto, executam-se as instâncias do log pelo gráfico a partir do seu topo e verifica-se se existem tokens perdidos ou criados ao longo do caminho, como ilustrado na Figura 2-16 (Aalst, 2011).

O Replay pode detectar problemas

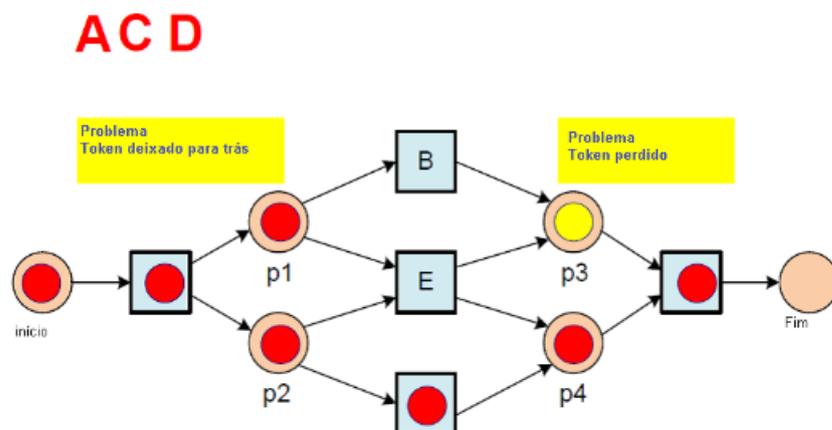


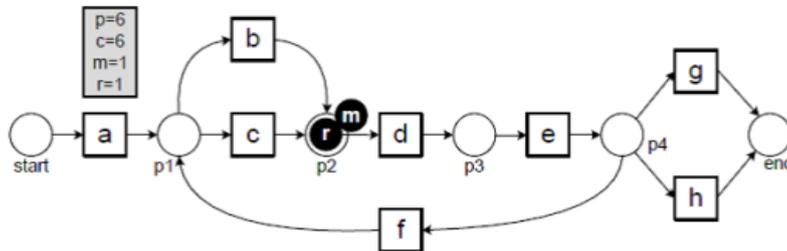
Figura 2-16 - Verificação de Conformidade utilizando Replay do Log em Redes de Petri (Aalst, 2011)

A cada instância ou trace submetido ao replay é calculado o fitness, no fim da execução de cada instância do processo, temos um diagnóstico de como está a execução em relação ao modelo de processo. Pode-se observar em cada posição a quantidade de tokens perdidos ou produzidos durante a execução. Na Figura 2-17 é mostrada a execução da instância, a apuração dos tokens produzidos e perdidos e a fórmula para o cálculo do fitness desta instância.

Problemas encontrados durante o Replay

σ_3 em N_2

$\sigma_3 = \langle a, d, c, e, h \rangle$



- Um token perdido (de 6 consumidos)
- Um token remanescente (de 6 produzidos)

$$fitness(\sigma, N) = \frac{1}{2} \left(1 - \frac{m}{c} \right) + \frac{1}{2} \left(1 - \frac{r}{p} \right)$$

Figura 2-17 - Cálculo do fitness (Aalst, 2011)

Na Figura 2-18, nota-se que o mesmo processo pode ser modelado de formas diferentes dependendo do algoritmo utilizado todos os modelos representam o mesmo processo, no entanto podemos verificar que o cálculo de fitness varia dependendo modelo, portanto precisa-se manter o mesmo padrão do início ao fim da análise e escolher o que melhor atende os requisitos de fitness, precisão, generalização e simplicidade. Cada modelo tem um custo computacional diferente e isto também deve ser levado em consideração na hora da escolha do modelo (Aalst, 2011).

Exemplo de valores do fitness

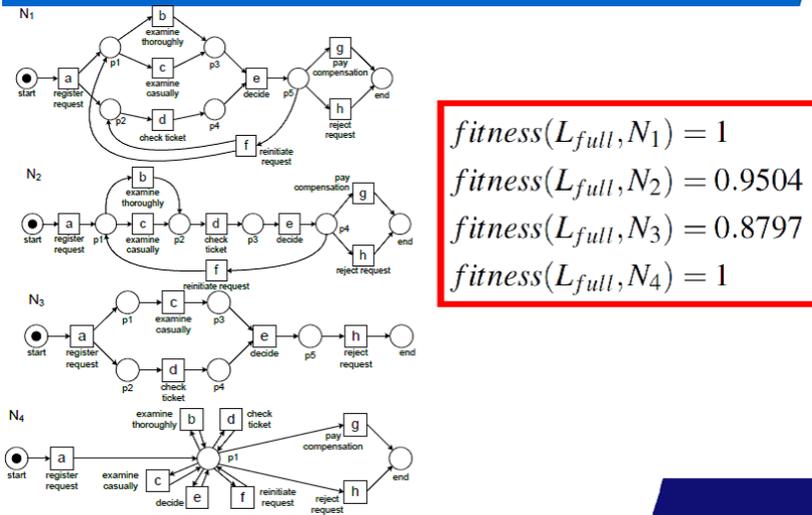


Figura 2-18 - Modelos de descoberta com a visão aos 4 parâmetros, fitness, precisão, generalização e simplicidade (Aalst, 2011)

A partir da execução das instâncias do processo no modelo, através da técnica de replay, é possível verificar as posições onde há problemas de execução, seja por que o modelo não tem a atividade existente no log, ou porque ela tem uma atividade e o log não. Estes desvios originam os tokens sobrando e faltando a cada execução, e no fim da execução de todas as instâncias é possível calcular um fitness global, que indica o quanto o log se encaixa no modelo, conforme mostrado na Figura 2-19.

Diagnóstico

$$(fitness(L_{full}, N_3) = 0.8797)$$

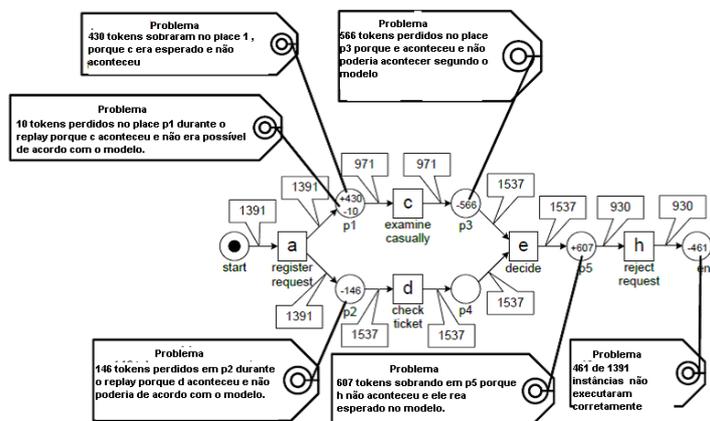


Figura 2-19 - Exemplo do diagnóstico do processo utilizando fitness (Aalst, 2011)

Para separar as instâncias conformes das não conformes utiliza-se a estratégia drilling down (executando de cima para baixo). Estipula-se o fitness aceitável, e abaixo deste valor, uma instância é considerada não conforme. A saída é um conjunto de instâncias que atendem o fitness neste caso consideradas conformes, e um conjunto de instâncias que não atingiram o valor mínimo de fitness, e portanto foram consideradas não conformes como ilustrado na Figura 2-20 (Aalst, 2011).

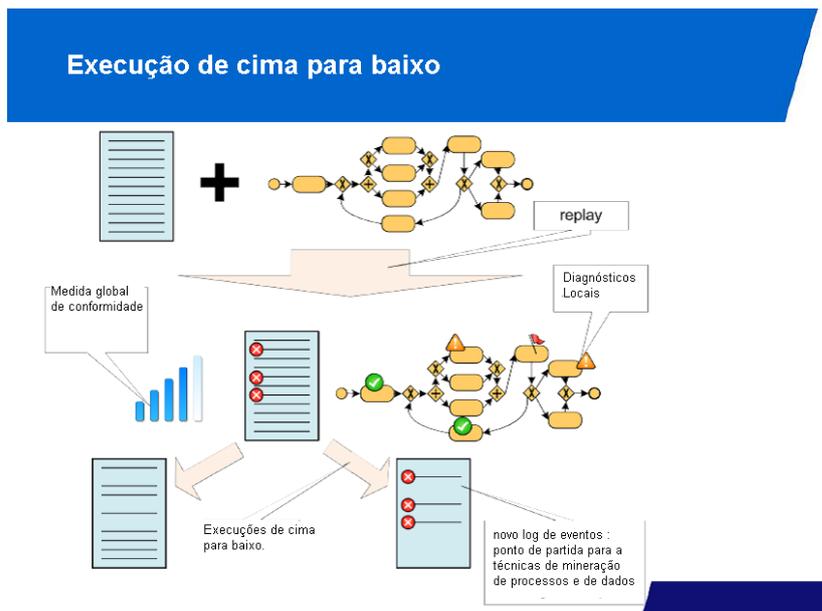


Figura 2-20 - Separação das instâncias não conformes utilizando drilling down (Aalst, 2011)

2.6. Conceitos importantes na Indústria Farmacêutica

Nesta seção são apresentados alguns conceitos de acordo com a sua aplicação específica na indústria farmacêutica. Estes conceitos são utilizados na proposta de dissertação.

2.6.1. Desvio

Desvio é definido como qualquer evento não planejado, que ocorra em sistemas computadorizados, podendo colocar em risco a qualidade e/ou a segurança de um produto ou dado. é uma instância diferente do processo padrão (Agência Nacional de Vigilância Sanitária, 2010).

Desvio não Regulatório: Qualquer desvio em que a instância execute atividades diferentes da instância padrão ou a priori, mas que nenhuma atividade executada infrinja a regulamentação.

Desvio Regulatório: Instância diferente do processo padrão e que tem atividades que ferem a regulamentação (Agencia Nacional de Vigilância Sanitária, 2010, pp. 10-14).

Instância Conforme: Instância com as mesmas atividades da instância padrão (Aalst, 2011).

2.6.2.Risco

Risco é a combinação da probabilidade de ocorrência e a severidade do dano ao paciente (Agencia Nacional de Vigilância Sanitária, 2010).

Detectabilidade é a capacidade de identificar o risco do processo de fabricação antes da liberação do produto fabricado para o mercado (Agencia Nacional de Vigilância Sanitária, 2010, p. 33).

Na abordagem da ANVISA, no seu guia de validação de sistemas computadorizados (Agencia Nacional de Vigilância Sanitária, 2010), o gerenciamento de riscos é uma sistemática de processos para a avaliação, controle, comunicação e revisão de riscos. É um processo interativo e complexo utilizado durante todo o ciclo de vida do sistema computadorizado contemplando, inclusive, o seu processo de descontinuidade. Basicamente, este gerenciamento está associado às metodologias de análise de riscos visando definir, identificar e eliminar as possíveis falhas, problemas ou riscos potenciais que estejam envolvidos nos processos relacionados aos sistemas computadorizados.

Em todo o mundo, não existe um consenso entre as agências reguladoras quanto ao nível aceitável de controles para a conformidade fabricação de compostos altamente perigosos (ou seja, aqueles compostos que podem causar efeitos adversos graves em baixas doses) em instalações multiprodutos, ou seja a mesma linha de produção fabricando medicamentos diferentes. Várias agências reguladoras permitem a produção de alta periculosidade de compostos por campanha, desde que a separação adequada e procedimentos de limpeza adequadamente validados estejam presentes, enquanto outros importantes órgãos reguladores não permitem certos compostos altamente perigosos em instalações multiprodutos (ISPE - Risk Based, 2010).

O regulamento do FDA, que muitas vezes é citado quando há preocupações com a contaminação cruzada - 21 CFR 211.42 (Referência 15, Seção 15.1), afirma: "as operações devem ser realizadas dentro do especificamente definido

em áreas de tamanho adequado. Haverá áreas separadas ou definidas ou quaisquer outros sistemas de controle na empresa de forma que as operações necessárias sejam executadas para evitar a contaminação no decurso da fabricação. ” (ISPE - Risk Based, 2010, p. 8).

O Guia da EU (União Europeia) para Boas Práticas de Fabricação (Seção 5.19) (Referência 14, Seção 15.1) afirma que "contaminação cruzada deve ser evitada por medidas técnicas ou organizacionais adequadas. As medidas adequadas devem ser determinadas após uma avaliação dos riscos de qualidade, tal como descrito no ponto 3.6. " (ISPE - Risk Based, 2010, p. 11).

No Brasil, ainda não há um conjunto de normas específicas, mas apenas regras isoladas, obrigando as indústrias a seguirem um conjunto de restrições para cada tipo de produto fabricado.

2.6.3. Detectabilidade do Risco

Detectabilidade baixa: Risco ou não conformidade improvável de ser detectado pelo operador, não existe uma verificação sistemática (Agencia Nacional de Vigilância Sanitária, 2010, p. 35).

Detectabilidade média: Risco ou não conformidade verificado por amostragem, existe um procedimento sistemático de verificação (Agencia Nacional de Vigilância Sanitária, 2010, p. 35).

Detectabilidade alta: Risco ou não conformidade verificado em 100% das instâncias, existe um procedimento sistemático de verificação (Agencia Nacional de Vigilância Sanitária, 2010, p. 35).

Perigo: fonte com potencial para provocar danos (Agencia Nacional de Vigilância Sanitária, 2010, p. 33).

Dano: dano para a saúde, incluindo danos que podem ocorrer devido a uma perda de qualidade do produto (Agencia Nacional de Vigilância Sanitária, 2010, p. 33).

Severidade: medida das possíveis consequências do perigo (Agencia Nacional de Vigilância Sanitária, 2010, p. 41).

Classe de Risco: é a classificação da severidade do risco e a sua probabilidade de ocorrência (Agencia Nacional de Vigilância Sanitária, 2010, p. 41).

Prioridade do Risco: combinação entre a classe do risco e a Detectabilidade (Agencia Nacional de Vigilância Sanitária, 2010, p. 41)

3. Trabalhos relacionados

Neste capítulo são apresentados os trabalhos relacionados que tratam da verificação de conformidade em processos de negócios, os problemas que cada proposta busca resolver, a solução sugerida por elas e suas vantagens e desvantagens em relação a esta dissertação. Em seguida, é feito um comparativo de cada trabalho com a proposta apresentada.

Será apresentado um quadro comparativo com as semelhanças e diferenças das abordagens em relação à proposta desta dissertação.

3.1. **Audit 2.0 usando a mineração de processos para melhorar os processos de auditoria (Wil M.P. van der Aalst, 2010).**

Verificar a quantidade de dados disponíveis nos logs de eventos dos sistemas de informação como ERP (Enterprise Resource Planning) é inviável de ser feito manualmente. A mineração de processos é uma poderosa ferramenta de auxílio à atividade de auditoria. A comparação do processo de jure com o processo minerado e a verificação de conformidade entre os dois vai indicar onde estão as não conformidades, e balizar o trabalho dos auditores. Outra técnica útil neste trabalho é a mineração de regras de negócio, a partir dela é possível descobrir se alguma atividade suspeita está sendo executada, ou se uma pessoa não autorizada está executando uma atividade de forma irregular.

Auditando os dados correntes, os auditores podem usar novas técnicas de mineração para influenciar o processo operacional antes de o evento ser completado. Estas técnicas envolvem a repetição da execução de instâncias do processo — por exemplo, através da execução das instâncias em modelos Redes de Petri para detectar desvios, prever resultados particulares e recomendar as medidas cabíveis. Mapeamento de regras de negócios em redes de Petri ou lógica temporal, tais como LTL, permite controles eficientes.

O trabalho sugere um método para comparação do processo de jure com o processo de facto para a verificação de conformidade. Os algoritmos utilizados para a descoberta podem ser vários, mas a representação do modelo utilizada é a Rede de

Petri. Através da técnica de Replay, onde o log é executado através do modelo e verificadas as discrepâncias.

Para a mineração de regras o algoritmo utilizado é o LTL (Process Mining Task Force, 2014) , onde as regras de negócio são buscadas no log de eventos.

Comparando os casos de execução com o modelo de jure, auditores podem detectar desvios como eles ocorrem e prever os desvios que tem chance de acontecer. Técnicas similares podem ser usadas para recomendar ações particulares por exemplo “Taking action X will minimize the risk of violating legal requirement Y ”. A verificação de conformidade através da técnica de replay, vai indicar o quanto o processo minerado (de facto) se encaixa com o processo de jure, e através do cálculo de fitness é possível apurar o índice de conformidade.

O trabalho proposto por Van der Aalst (van der Aalst,, et al., 2010) não é suficiente para solucionar o problema desta dissertação. A mera comparação do processo de jure com o processo de facto, e o cálculo do fitness apenas mostra as instâncias não conformes, mas não classifica os tipos de não conformidades. Duas instâncias não conformes com fitness 0,94 podem ser uma classificada como desvio regulatório e outra como desvio não regulatório, ou até mesmo incompleta, portanto este tipo de comparação não diminui a aleatoriedade da escolha das amostras.

O uso do LTL Checker o trabalho de Van der Aalst tem como objetivo priorizar regras de negócios para melhorar os controles, nesta dissertação o LTL é utilizado a partir de regras de negócios definidas, mineramos os valores que elas assumem, separamos as instâncias com valores proibidos e classificamos estas não conformidades como desvios regulatórios. O método Audit 2.0 não chega a este nível de detalhamento.

Concluindo, o método Audit 2.0 (Wil M.P. van der Aalst, 2010) não realiza a classificação da não conformidade, ele apenas verifica e quantifica o quanto o log se encaixa no modelo de processos, esta métrica sozinha não é capaz de classificar as instâncias de maneira que resolvam o problema da dissertação.

3.2. **Descoberta de anomalias em fluxo de processos de sistemas de automação (UNISIST, 2012).**

Na Samsung Eletro Mecanics (SES), o sistema SCADA¹ (Sistemas de Supervisão e Aquisição de Dados) em conjunto com o MES² (Manufacturing Execution System) foi desenvolvido para apoiar a gerência de produção a executar o processo de fabricação e melhorar o processo de decisão pelo fornecimento de informações úteis sobre o processo. A cada mês, a análise dos processos de fabricação é uma das funções mais importantes. Para analisar os dados provenientes do MES, o departamento Equipment Optimization Technology (EOT) na SEM usa estatísticas e mineração de dados. Entretanto, enquanto as técnicas trabalham bem para analisar dados específicos das tarefas de fabricação, elas têm dificuldade para fazer uma análise do resultado de processos, considerando o desempenho global e padrões de manufatura. Neste trabalho foram aplicadas técnicas de mineração de processos para analisar os processos na SEM.

Para o estudo de caso, foram usados os logs do mês. A unidade básica de manufatura é o lote, que foi utilizado como `case_id`, isto é um número de identificação de uma quantidade específica ou um lote de produto. O log de eventos obtido continha todas as tarefas das máquinas usadas para fabricar os lotes de maio a outubro de 2012, foram analisados 11226 casos, com 900542 eventos em 361 tarefas diferentes em 1217 equipamentos diferentes.

A proposta desta análise foi mostrar a aplicabilidade das técnicas de mineração de processos na análise de processos de manufatura. Após uma discussão aprofundada como o departamento EOT, as seguintes questões foram propostas:

- Qual o atual fluxo do processo?
- Existem não conformidades na produção do lote?
- Existem gargalos no processo?
- Como as máquinas são utilizadas?

¹SISTEMAS SCADA - É um sistema a supervisão à distância de processos industriais, podendo também realizar funções de operação e controle (Reginato, 2007).

² Um sistema MES tem como missão prover as ferramentas necessárias em forma estruturada para conseguir executar estas verificações. Sistemas MES devem ser capazes de reunir e ordenar dados de uma planta e providenciar informação mensurável. A partir de informações mensuráveis, indicadores de performance (KPIs – Key Performance Indicators) serão determinados (Mecatrônica Atual, 2013).

A verificação de conformidade fornece um valor de fitness derivada do modelo de processo. A conformidade mostra um fragmento do resultado da verificação de conformidade. O valor de fitness obtido de todo o processo foi 0.9962. Isso indicou que quase todas as instâncias se encaixavam com o modelo descoberto. Para os tokens restantes e desaparecidos, foram realizadas análises com o verificador de LTL que encontrou alguns fluxos anormais. Por exemplo, um dos principais fluxos anormais é o fluxo da tarefa de inspeção de entrega e a tarefa de teste final.

Os equipamentos de fabricação são recursos cruciais da fábrica, por isto foi realizada uma análise de desempenho na perspectiva dos recursos. Na análise, foi visualizada uma utilização desbalanceada dos equipamentos. Por exemplo, a frequência de uso de sete máquinas para a tarefa de embalagem, a máquina N28027 foi utilizada 5612 vezes enquanto a máquina N14074 só foi utilizada duas vezes.

O LTL foi utilizado nas instâncias não conformes para verificar a existência de ocorrências críticas, como por exemplo, o produto pular a fase do controle de qualidade e ir direto para o despacho de mercadorias. No caso dos gargalos, foram descobertas máquinas que realizavam as tarefas em mais tempo fazendo com que o plano de manutenção fosse alterado.

No trabalho da UNISAN na Samsung, o ambiente que foi utilizado para a mineração é muito parecido com o utilizado nesta dissertação, um sistema de automação industrial baseado em um sistema SCADA (Sistemas de Supervisão e Aquisição de Dado). Os dados gerados por este sistema são utilizados como log de eventos. Como em todo processo industrial automatizado, existe uma receita de fabricação que seria o nosso processo a priori, e durante a execução das instâncias podem ocorrer desvios por problemas do operador, defeitos do equipamento, falta de manutenção dentre outros.

Apesar de fazer um estudo de conformidade, o principal objetivo era comparar o desempenho entre equipamentos semelhantes, a conformidade do produto fabricado não foi tratada em nenhum momento, por isto não é possível utilizar este modelo para a solução do problema proposto nesta dissertação. No entanto, o trabalho é uma boa base de comparação nas técnicas de aquisição e pré processamento dos dados. A utilização do LTL Checker para a mineração de regras também tem paralelo nesta dissertação, no caso da Samsung, ele foi utilizado para a busca de instâncias incompletas, já na dissertação o LTL foi utilizado para buscar atividades proibidas. O

trabalho da UNISAN tem vários pontos em comum, mas também não é suficiente para realizar todas as classificações propostas na dissertação, pois falta a classificação dos desvios regulatórios e não regulatórios.

3.3. Verificação de conformidade dos processos de atendimento à Saúde (R.S. Mans, 2008)

Para melhorar a vantagem competitiva, os hospitais tentam simplificar seus processos, para isto é essencial que uma visão apurada dos processos de atendimento. Este trabalho apresenta a aplicação de mineração de processos para o aprendizado destes fluxos e descoberta dos caminhos seguidos para grupos particulares de pacientes. Esta não é uma tarefa trivial devido à natureza dinâmica dos processos de cuidados com a saúde. O artigo demonstra a aplicabilidade na mineração de processos usando casos reais dos processos na oncologia ginecológica no Dutch Hospital em Amsterdam. Utilizando uma variedade de técnicas de mineração de processos, foram analisados os processos de atendimento à saúde em três diferentes perspectivas: controle de fluxo, organizacional e desempenho. Para isto, foram extraídos logs relevantes dos sistemas de informação do hospital e analisados os logs usando o PROM (Process Mining Group, TU/E, 2013). Os resultados mostraram que a mineração de processos pode ser utilizada para mostrar uma nova visão que facilita a melhorias nos processos existentes (R.S. Mans, 2008).

O log continha dados brutos sobre 627 pacientes de oncologia ginecológica, tratados entre 2005 e 2006, e os diagnósticos e tratamentos. Não foi utilizado nenhum conhecimento a-priori sobre o processo de atendimento do grupo de pacientes e não foi obtido nenhum modelo de processos manualmente.

O log do Dutch Hospital contém uma enorme quantidade de atividades distintas, das quais muitas são atividades em um nível baixo de abstração. A mineração um log que contém muitas atividades distintas resultaria em um modelo muito detalhado, tipo espaguete, que é difícil de entender. Portanto, primeiro foram aplicadas algumas técnicas de pré-processamento sobre os logs para obter resultados interpretáveis durante a mineração. Durante o pré-processamento, o objetivo é "simplificar" o registro removendo o excesso de atividades de nível baixo, o objetivo é considerar apenas os eventos em nível de departamento. Desta forma, pode-se, focar nos caminhos do tratamento e interações entre departamentos.

O log também contém grupos de atividade de baixo nível para as quais não há nenhum representante. Por exemplo, no departamento de Radiologia muitas atividades podem ocorrer por exemplo, echo abdômen, tomográfica computadorizada de tórax e cérebro, mas os logs não contêm cada evento que ocorre em uma visita a este departamento. Foi aplicada a agregação para atividades de baixo nível em grupos sem um representante, (1) definindo um representante, (2) mapeando todas as atividades do grupo para esse representante, e (3) removendo repetições de eventos do log. Por exemplo, para o departamento de Radiologia foi definida "Radiologia" como representante. Um log que originalmente contém "..., ultrasound scan abdômen, peito raio-x de peito, Tomografia scan cerebral,", todas as atividades foram substituídas por "Radiologia", após o mapeamento de atividades de baixo nível para Radiologia todas as duplicatas foram removidas.

Neste trabalho, foi usado o algoritmo Heuristic mining, que trabalha com ruídos e exceções e habilita o analista a focar no fluxo principal do processo ao invés de verificar cada detalhe de comportamento que aparece no log. O resultado obtido através da mineração de todos os casos parece muito complexo (Figura 3-1), no entanto, os processos no domínio da saúde não têm um único tipo de fluxo, mas um grande número de variantes baseadas em pacientes e doenças, não é surpreendente que o modelo de processo obtido seja complicado.

Um dos métodos para lidar com este problema é quebrar o log em um ou mais sub logs até que ele seja simples o suficiente para ser analisado claramente. Podem ser aplicadas técnicas de clusterização, dividindo em clusters com propriedades semelhantes.

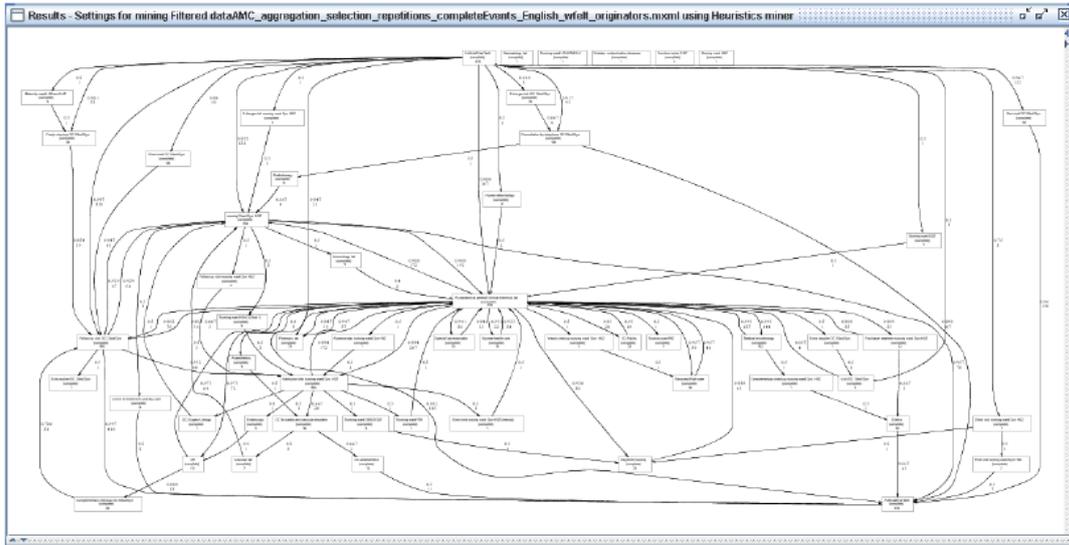


Figura 3-1 - Processo obtido pelo Heuristic Miner (R.S. Mans, 2008)

Usando esta abordagem, foram obtidos vários clusters com tamanho razoável. Neste trabalho foram mostrados apenas resultados dos maiores clusters, contendo 352 casos com propriedades semelhantes. A Figura 3-2 mostra a perspectiva do Heuristic Miner dividido em grandes clusters. O resultado é muito mais simples que a Miner, o fitness do modelo é bom. O modelo representa a maioria dos casos em um cluster, estes casos se encaixam no modelo descoberto. Uma inspeção mais detalhada no cluster principal pelos especialistas do domínio confirmou que este é o principal fluxo seguido pelo tratamento dos pacientes de ginecologia oncológica do hospital.

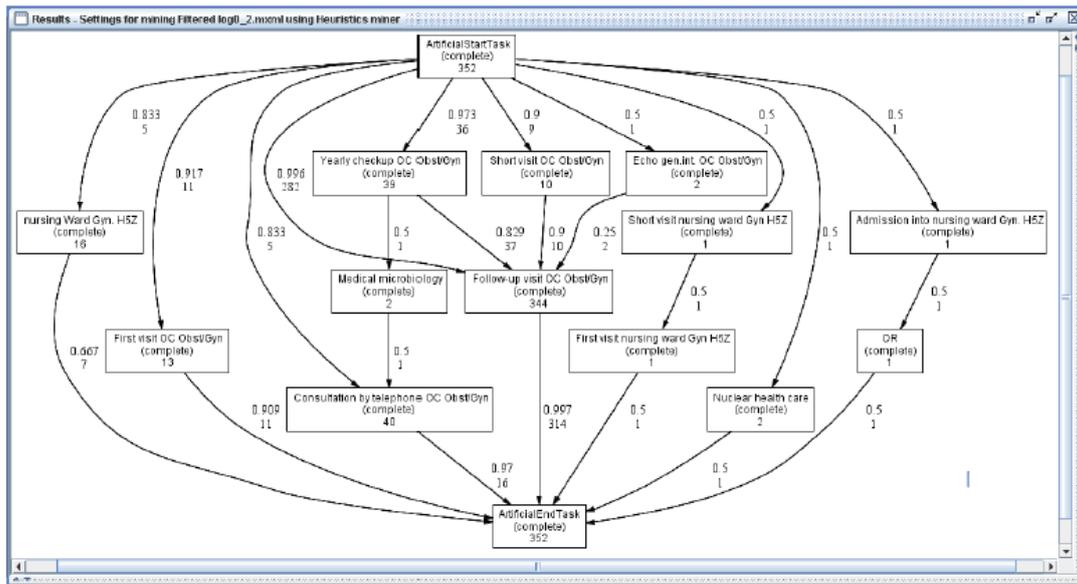


Figura 3-2 - Modelo de processos para os maiores clusters (R.S. Mans, 2008)

Outra abordagem para lidar com processos desestruturados é o Fuzzy miner, ele usa uma mistura de abstração com técnicas de cluster para gerar modelos mais simples e de fácil compreensão. Este algoritmo apresenta uma representação de alto nível eliminando os detalhes indesejáveis.

Este trabalho tem uma semelhança com esta dissertação principalmente na semelhança do domínio da área de cuidados com a saúde. As mesmas dificuldades encontradas devido à diversidade de atividades e fluxos são encontradas no domínio da indústria farmacêutica, da mesma forma em ambos os trabalhos, o log de eventos não tem uma visão de processo bem definida e precisa ser pré-processado para que possa ser utilizado. No entanto no início do trabalho do hospital existia uma pretensão de mapear os fluxos de trabalho em forma de processos com o intuito de depois verificar a sua conformidade, e também procurar a simplificação dos modelos, no entanto o resultado restringiu-se à descoberta dos processos, redes sociais e influência dos processos nos departamentos. A comparação dos algoritmos Heuristic Miner e Fuzzy Miner também foi feita nesta dissertação para averiguar qual deles teria melhor representação, nos dois casos o Fuzzy Miner foi o escolhido por tratar melhor ruídos e a diversidade de atividades deste tipo de domínio.

Além disso, foram comparados os resultados com um fluxograma para a trajetória de diagnóstico do processo de saúde de Oncologia Ginecológica, e onde uma abordagem top-down tinha sido usada para criar a fluxograma e obter os dados da

logística. No que se refere ao fluxograma, foram obtidos resultados comparáveis com muito menos esforço de forma semiautomática.

Apesar de se tratarem de domínios semelhantes o trabalho do hospital não avançou para a verificação de conformidade, portanto não é suficiente para solucionar o problema proposto nesta dissertação.

Tabela 3-1 - Comparativo entre trabalhos relacionados

Trabalho	Objetivo	Foco	Comparação com a proposta desta dissertação
Usando a mineração de processos para melhorar os processos de auditoria	Mostrar como as técnicas de mineração de processos e verificação de conformidade podem ser utilizadas para suportar atividades de auditoria. A proposta é auditar todos os eventos e não somente uma parte dos dados.	Verificação de conformidade através da comparação entre o processo de jure e o processo de facto, calculando o fitness entre eles, que dá um índice de conformidade e para atributos específicos minerar regras através do algoritmo LTL.	Esta abordagem não resolve o problema específico do domínio estudado, pois não consegue classificar as instâncias e tipos de desvios, apenas gera um índice de conformidade. Este índice não é suficiente para classificar a conformidade regulatória, pois uma instância com fitness alto pode executar uma atividade proibida e portanto é um desvio regulatório e uma instância com menor fitness pode não executar nenhuma atividade proibida. Neste caso a instância com maior fitness é pior do que a com maior fitness.
Descoberta de anomalias em fluxos de processos de sistemas de automação (UNISIST, 2012)	Mostrar a aplicabilidade das técnicas de mineração de processos na análise de processos de manufatura	Verificação de conformidade para medida de desempenho do processo de fabricação.	O trabalho de Song é baseado em um sistema SCADA, semelhante ao usado nesta dissertação, no entanto a verificação de conformidade não usa classificação apenas o índice de conformidade e a mineração de regras do trabalho de Song só buscam instâncias incompletas, portanto não é suficiente para solucionar o problema apresentado.
Verificação de conformidade dos processos de atendimento à Saúde	Verificar a aplicabilidade das técnicas de mineração de processos no domínio de cuidados com a saúde.	Mapear os processos de tratamento dos pacientes do departamento de oncologia ginecologia e sua interação com os outros departamentos nas perspectivas do fluxo do processo, organizacional e desempenho. Foram gerados modelos de processo, modelos de redes sociais e modelos de análise de desempenho dos processos.	O trabalho se assemelha ao trabalho desta dissertação por se tratar de um domínio da área de saúde, por este motivo o processo tem um nível de complexidade semelhante, ele busca descobrir o modelo do processo e verificar a conformidade dos fluxos como fazemos na dissertação, no entanto não chega a resultados de classificação das não conformidades.
Esta dissertação	Descobrir e classificar não conformidades no processo de fabricação de medicamentos, através da mineração de processos nos logs de eventos	Descobrir as não conformidades das instâncias do processo de fabricação e classificá-las se afetam ou não a regulamentação ou legislação. Criar um indicador para o especialista do domínio identificar que produtos têm mais chances	Esta dissertação propõe um método que sistematiza o tratamento do log de eventos, a verificação de conformidade e mineração de regras. Ela se diferencia dos trabalhos relacionados pois além de fazer a verificação de conformidades é capaz de classificar e segregar as instâncias com desvios.

	do sistema SCADA que registra os dados de fabricação.	de estarem não conformes e quais não conformidades podem impactar a qualidade do produto.	
--	---	---	--

Tanto nos trabalhos publicados na literatura científica e nos congressos do ISPE, não foram encontradas propostas similares , tratando de verificação de conformidade para a indústria farmacêutica. O problema da verificação de conformidade dos medicamentos, sem destruir todos os produtos fabricados, tem neste trabalho uma solução no método que permite classificar as instâncias, permitindo indicar ao especialista do domínio quais instâncias tem desvios e o tipo de desvio, propiciando uma escolha mais criteriosa da amostra dos produtos fabricados para os testes de qualidade.

4. Método de descoberta e classificação de não conformidades na indústria farmacêutica.

O método proposto descreve as etapas que devem ser seguidas para preparar, ler, processar e classificar as instâncias de um processo no domínio farmacêutico. Estas etapas foram elaboradas e refinadas com base em resultados de um estudo exploratório. A Figura 4-1 ilustra o método de maneira geral, com suas entradas e saídas.



Figura 4-1 – Esquema do método produzido pelo autor

O executor do método assume o papel de analista. O analista deve ser uma pessoa que tenha noções básicas do domínio, do processo e mineração de processos. O método é composto de quatro fases (Figura 4-2).

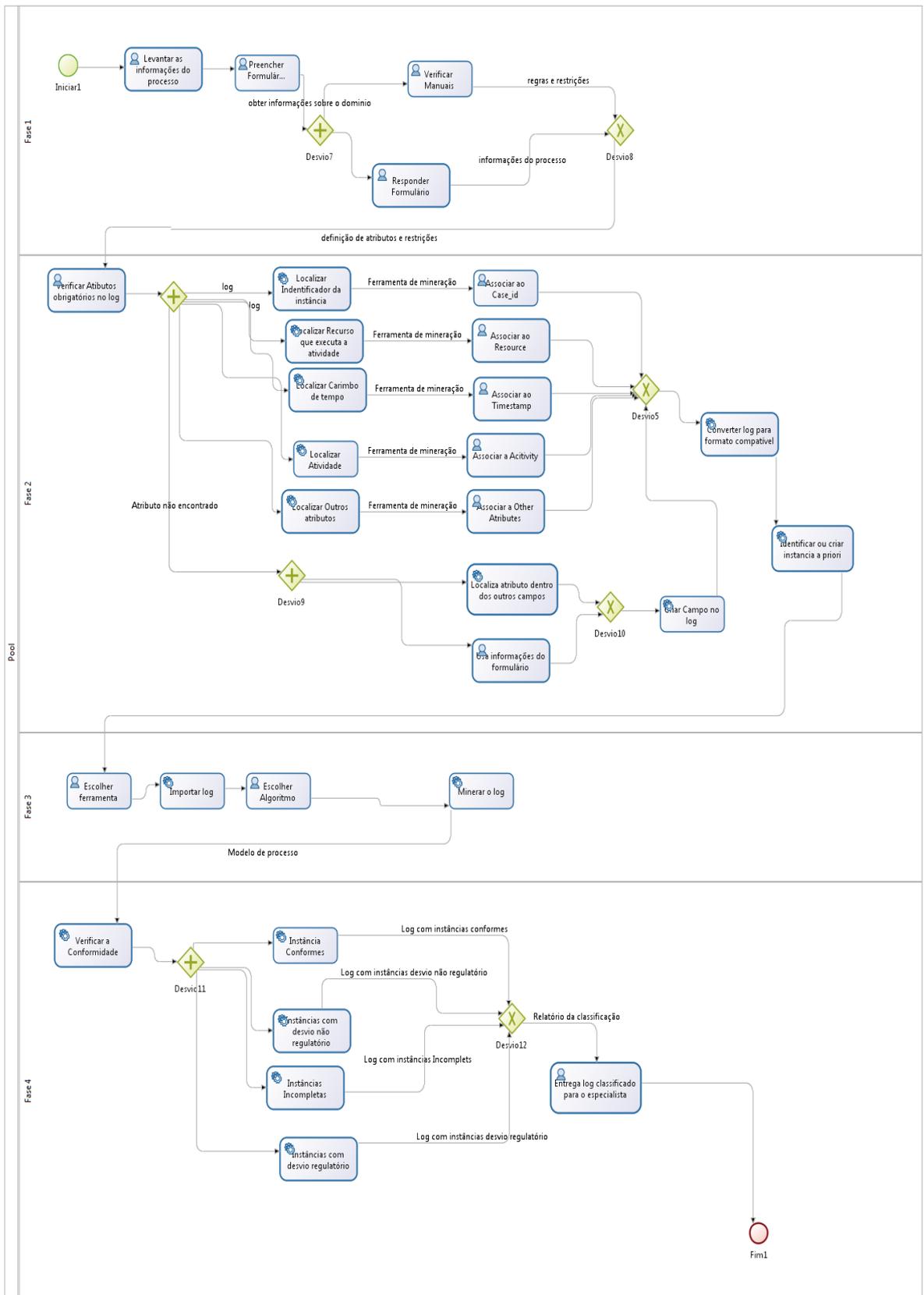


Figura 4-2 - Processo do método

Na primeira fase, são colhidas as informações sobre o domínio, as restrições e as regras que devem ser seguidas pelas instâncias do processo, a instância a priori que será usada como instância padrão para comparação com as outras instâncias do processo.

Na segunda fase, é feito o pré-processamento do log, o arquivo do log deve ser convertido para o formato CSV (texto separado por vírgula) para que seja compatível com as ferramentas de mineração, é verificado se os atributos mínimos necessários para a mineração de processos estão presentes, se será necessário criar atributos, ou modificá-los para que as informações contidas nos campos do log atendam aos requisitos descritos nas respostas do formulário da fase 1.

Na terceira fase, é feita a mineração do processo propriamente dita, com o log pré-processado, o arquivo de log deve ser importado para uma ferramenta de mineração escolhida, o algoritmo previamente testado para este domínio deve ser selecionado e executada a mineração. O resultado deve ser validado pelo especialista do domínio.

Na quarta fase, são feitas a descoberta e classificação de não conformidades, onde serão separadas as instâncias conformes e não conformes. As instâncias não conformes serão classificadas através da verificação das regras, levantadas no formulário da fase 1. Este resultado é encaminhado para o especialista do domínio para auxiliá-lo na escolha dos produtos que serão testados no controle de qualidade.

A seguir são descritas mais detalhadamente cada fase.

4.1. Primeira Fase – Levantamento de informações sobre o domínio.

O objetivo desta fase é levantar as informações do domínio necessárias para a obtenção da instância a priori do processo, que vai ser usada como instância padrão para comparação com as outras instâncias do processo. É importante que além do nome do atributo se faça uma pequena descrição do que significa e onde é utilizado. Dentre os atributos, devem existir obrigatoriamente um atributo que identifique de forma única uma instância do processo (*case_id*), um que identifique a hora e dia que iniciou a execução da atividade (*testam*), um para atividade (*activity*), e um para identificar o recurso (*resource*) que executou a atividade.

A primeira atividade a ser executada nesta fase é a entrevista com o especialista do domínio para coleta de informações sobre os procedimentos operacionais. As respostas da entrevista e os procedimentos coletados devem ser utilizados para preencher formulário de informações sobre o domínio (Formulário de informações sobre o domínio), que tem seus campos listados a seguir:

- Especialista do domínio: nome do entrevistado.
- Data e Hora: data e hora do início da entrevista.
- Cargo: cargo do entrevistado.
- Função: breve descrição das atividades que o entrevistado realiza no domínio.
- Empresa: nome da empresa do entrevistado.
- Analista: nome do responsável pelo preenchimento do formulário.
- Qual é o processo estudado: nome do processo estudado?
- Qual é o objetivo do processo: resultado esperado deste processo?
- Qual a descrição do processo: descrição das etapas do processo?
- Quais são os Atributos: nome e exemplo dos atributos utilizados neste processo (exemplo: número do lote, temperatura, tamanho do frasco, ciclo do processo, pressão diferencial, fluxo laminar, etc.)?
- Qual o formato do LOG: tipo de arquivo em que o log se encontra, por exemplo, CSV, XLS, XLSX, XML, XES, etc.
- Qual o atributo identifica unicamente uma instância do processo? Ex.: número do lote, número da requisição, número da análise, etc.
- Qual o atributo marca o início e fim do processo e qual o valor ele deve assumir? Nesta pergunta responda o nome do atributo e qual o valor que este atributo assume na atividade inicial do processo e qual o valor que o atributo assume na atividade final da instância do processo, e se são sempre as mesmas atividades e valores de início e fim de cada instância do processo.
- Qual a norma ou regulamento se aplicam a este processo? Responda quais as regulamentações são aplicadas a este processo, exemplo RDC 17, RDC 58, CFR PART 11 e onde podem ser encontradas as regulamentações, site, livro, documentos internos.

- Quais são os parâmetros críticos do processo em relação a regulamentação? Responda quais os valores críticos dos atributos das instâncias do processo. Escreva em forma de regras por exemplo? Se o atributo atividade for igual a erro 53 então valor proibido. Se o tempo da atividade aquecimento a $300\text{ }^{\circ}\text{C} < 30\text{ min}$ então valor proibido, se a condutividade da água for > 3 então valor proibido, vá elaborando todas as regras que ferem os parâmetros exigidos pelas regulamentações.
- Quais são as atividades críticas em relação a regulamentação que não podem acontecer? Responda quais as atividades ou sequência de atividades não podem acontecer, deve ser escrito como regras, exemplo:
 - Se o atributo atividade receber o valor tampa do tanque aberta, acontecer depois da atividade início da fabricação então atividade proibida,
 - Se o atributo atividade desliga o fluxo laminar acontecer depois de carregamento de frascos, então atividade proibida.
 - Escrever quantas regras forem necessárias até esgotar todas as restrições.
 - Quais são as atividades críticas em relação a regulamentação que precisam acontecer? Responda com a lista de atividades que precisam acontecer, caso exista uma sequência predeterminada especificar a sequência, por exemplo:

Ligar o equipamento.

Entrar com a senha do analista.

Especificar o ciclo

Digitar o número do lote.

Ligar a esteira.

Ligar o aquecimento.

Ligar o resfriamento.

Ligar o fluxo laminar e depois ligar o aquecimento.

- Quais são os critérios de decisão (definições para o que é conforme, desvio e regulatório)? Responda como classificar as instâncias, por exemplo:

- Instância conforme: toda a instância que contém as atividades da instância a priori e somente elas, e não infringem nenhuma das regras levantadas no Formulário de informações sobre o domínio.
- Instância com desvio não regulatório: toda a instância que contém todas as atividades da instância a priori, contém outras atividades e não infringem nenhuma das regras levantadas no Formulário de informações sobre o domínio.
- Instância com desvio regulatório: qualquer instância que infrinja uma ou mais regras levantadas no Formulário de informações sobre o domínio.

A Tabela 2 é o artefato utilizado para levantamento das informações sobre o domínio, sendo utilizada para entrevistar o especialista do domínio. Exemplos de preenchimento do formulário são apresentados no Apêndice 8.1.

Tabela 4-1 - Formulário de informações sobre o domínio

Especialista do Domínio:	
Data e Hora:	
Cargo:	
Função (descrição da função em relação ao domínio):	
Empresa	
Analista (nome do responsável por realizar a entrevista.)	
Processo estudado (nome do processo)	
Objetivos do Processo:	
Descrição do Processo:	
Atributos (descreva a lista de atributos que devem ser considerados no log)	
Formato do log (texto, banco de dados, planilha, etc.)	
Qual atributo marca o início e fim do processo e como identifica-lo?	
Qual a norma ou regulamento se aplicam a este processo?	
Quais são os parâmetros críticos do processo em relação a regulamentação?	
Quais são as atividades críticas em relação a regulamentação que não podem acontecer?	
Quais são as atividades críticas em relação a regulamentação que precisam acontecer?	

Critérios de decisão (conforme, desvio e regulatório)	
---	--

4.2. Segunda Fase – Levantamento dos atributos obrigatórios nos logs e pré-processamento.

O objetivo desta fase é a criação do relacionamento entre as respostas do formulário e os campos do log. O analista deve procurar no log os campos que têm o mesmo significado dos atributos levantados no formulário. O relacionamento deve ser um para um, fazendo uma relação de-para. Os atributos mínimos para ser possível a mineração de processos são os seguintes:

- Case_id: identificação única de uma instância do processo.
- Activity: atividade realizada em uma instância, associada a um único timestamp, Deve ser capaz de identificar a ação naquele instante.
- TimeStamp: o carimbo de tempo identifica o início de cada atividade em uma instância.
- Resource: Recurso, que pode ser humano, máquina ou de outro tipo, que executa a atividade.

Caso outros atributos sejam encontrados eles devem ser relacionados da mesma forma.

O log deve ser convertido em um arquivo CSV³ (Comma Separated Values), pois a maioria das ferramentas de mineração de processos, como por exemplo, o PROM (Process Mining Task Force, 2014) e o Disco (Fluxicon, 2013) aceitam este formato. Com o formulário preenchido e de posse do log de eventos, o analista deve importar o log para uma ferramenta de mineração. Após o carregamento, os atributos do log devem ser associados aos atributos da ferramenta de mineração escolhida. Esta atividade tem como objetivo adequar os nomes dos atributos do arquivo de log com os atributos utilizados em cada ferramenta e sua saída é o log importado pronto para ser minerado.

A próxima atividade é a mineração no log. O algoritmo a ser utilizado é o Fuzzy Miner. Conforme foi explicado no Capítulo 2, este algoritmo foi escolhido para uso no método devido à sua capacidade de gerar modelos com diversos níveis de

³ CSV – Comma Separated Values, padrão de arquivo texto onde os campos são separados por vírgula.

abstração, trata muito bem modelos do tipo “spagetti”, eliminando componentes irrelevantes e destacando componentes importantes, a visualização é simples e de fácil entendimento. Nesta atividade, a entrada é o log de eventos e a saída o processo minerado.

No questionário, foi descrita a instância padrão ou a priori. Com esta informação, deve ser identificada no log a instância que corresponde à instância a priori. Em ambas as ferramentas, a saída agrupa por variações, portanto deve ser escolhida a Variante que representa a instância a priori, que vai ser utilizada para a comparação, e esta deve ser validada com o especialista.

Caso nenhuma instância atenda aos requisitos do formulário, a amostra pode não ser significativa ou o identificador do processo pode estar errado. O questionário deve então ser revisto, bem como se a identificação da instância está correta e se foi associada de forma correta com o atributo `case_id` na ferramenta de mineração. Se mesmo assim não for possível encontrar uma instância que atenda aos requisitos mínimos, será preciso criar uma instância, para que se tenha uma base para comparação com as outras instâncias do processo.

Para gerar uma instância, deve ser criado um arquivo em um editor de textos com as seguintes informações separadas por vírgulas: identificador da instância, timestamp (pode ficar com valor 0 para a primeira atividade), a atividade, o recurso que executou a atividade. O arquivo deve ser salvo com extensão CSV e inserido no topo do arquivo de logs.

Deve-se garantir que as atividades constantes na instância padrão tenham correspondentes na atividade do log, muitas vezes o especialista responde em linguagem natural, e as atividades são diferentes daquelas encontradas no log do sistema. Neste caso, deve ser feita uma tabela de-para com as atividades informadas pelo especialista e as atividades contidas no log, a instância a priori utilizada para a mineração e comparação com as outras instâncias deve ter as atividades compatíveis com o log e não em linguagem natural. Este novo arquivo de log deve ser importado para a ferramenta de mineração.

Para trabalhos futuros pode-se utilizar técnicas de mineração de texto para automatizar este processo.

4.2.1. Identificação do Case_id

O objetivo desta etapa é a identificação do case_id no arquivo de log. A entrada é o arquivo de log e a saída é o arquivo de log ordenado case_id, cada registro do log deve ter um case_id. No caso do valor do case_id não estar presente no registro, deve-se preencher manualmente levando em conta as informações do Formulário de informações sobre o domínio.

4.2.2. Identificação do TimeStamp

O objetivo desta etapa é identificar e ordenar o TimeStamp. A entrada é o arquivo de logs e a saída é o arquivo de log ordenado por TimeStamp. O atributo TimeStamp deve estar ordenado em ordem crescente, todos os registros devem ter o mesmo formato até o fim do arquivo. Quando existir mais de um TimeStamp para a mesma atividade do processo, o analista deve escolher o atributo que contiver o menor valor, pois ele irá representar o início da atividade.

4.2.3. Identificação da Activity

Uma Activity é a ação realizada em cada registro da instância do processo. Caso um único campo do log não seja suficiente para representar a atividade podem ser unidos dois ou mais campos para formar uma Activity completa. Esta decisão deve ser tomada com base nas respostas do Formulário de informações sobre o domínio.

4.2.4. Identificação do Resource

O Resource é o atributo que representa o recurso que executa a ação, ou seja o ator que executa a Activity. Se um campo do arquivo de log não for suficiente para identificar de forma única o Resource, o especialista do domínio deve ser consultado para definir quais outros atributos podem ser utilizados para resolver os conflitos, podem-se combinar dois ou mais campos, por exemplo, Maria, e Maria operadora, Maria supervisora.

4.2.5. Identificação de outros atributos do log

Os outros campos do log não entram diretamente na mineração de processos, mas podem ser importantes na mineração de regras e para verificar a conformidade ou identificar características dos outros atributos.

Por exemplo no log temos o campo nome que vai ser associado ao atributo resource, no entanto pode ser que uma restrição do processo seja para mudar um parâmetro de engenharia o Resource tem que ser supervisor.

Quando foi feita a tabela de-para o campo cargo não foi incluído nos atributos para a mineração do processo e foi associado a outros atributos, mas quando for ser executada a mineração de regras para descobrir se o Resource que executou a mudança de parâmetro é supervisor este campo será utilizado.

4.3. Terceira Fase – Mineração de Processos.

O objetivo desta fase é gerar o modelo do processo e a separação das variações das instâncias. Tem como entrada o log pré-processado e como saída o processo minerado. A mineração do log deve ser feita utilizando uma ferramenta de mineração, por exemplo o software PROM (Process Mining Task Force, 2014) ou Disco (Fluxicon, 2013), escolhendo o algoritmo FUZZY-Miner, que foi o que apresentou melhor representação visual do domínio segundo os especialistas do domínio.

Uma discussão a respeito do Fuzzy Miner é a possibilidade de serem excluídas atividades obrigatórias do log na etapa de abstração. Este realmente é um risco deste algoritmo, no entanto o critério utilizado para a retirada de uma atividade é a sua frequência, e portanto atividades obrigatórias tem grande frequência e dificilmente serão eliminadas. Mesmo assim no caso da eliminação existe a possibilidade de se calibrar a importância e a correlação aumentando ou diminuindo, ajustando a descoberta as características do domínio e do log.

4.4. Quarta Fase – Verificação de conformidade.

A verificação de conformidade é feita comparando a instância padrão com as demais instâncias do processo, uma a uma utilizando as informações obtidas no formulário de informações do domínio. As regras são definidas pelo especialista do domínio e pela documentação do processo e são definidas no formulário de informação sobre o domínio.

As ferramentas de mineração de processos estudadas nesta dissertação, verificam as instâncias semelhantes, e as agrupa em variantes, que são um conjunto de instâncias com as mesmas atividades. Para classificar as instâncias do processo são

utilizados filtros, que é a tradução da regra para a ferramenta, o resultado da aplicação do filtro, são as instâncias que atendem à regra, portanto por exemplo se for aplicado um filtro para encontrar desvio regulatório, o resultado será apenas as instâncias com desvio regulatório.

O conceito de variante já visto anteriormente é fundamental nesta fase. A variante é o conjunto de instâncias que executam as mesmas atividades e são agrupadas em um local chamado variante. Normalmente as variantes são numeradas e as instâncias contidas na variante são idênticas em relação ao campo Activity, ou seja executam as mesmas atividades, na mesma ordem.

4.4.1. Identificação de instâncias conformes

As instâncias conformes são aquelas que executam as mesmas atividades da instância a priori e por isto agrupadas na mesma variante.

Para obter as instâncias conformes, o analista deve apenas observar a Variante onde está a instância a priori, todas as instâncias contidas nesta Variante estão conformes.

Para as instâncias conformes não há necessidade de nenhum filtro pois a ferramenta já faz a seleção automaticamente seguindo a regra 1.

Se todas as atividades contidas na instância a priori estiverem contidas na instância minerada **E** a instância pertence à variante onde está localizada a instância a priori
Então instância é classificada como conforme

Regra 1: Regra para instância conforme

4.4.2. Identificação de instâncias com desvios não regulatórios e incompletas

O analista deve identificar a instância padrão, pois esta instância contém a lista de atividades que precisam acontecer em todas as instâncias. Para obter as instâncias não conformes completas, ou seja que executam todas as atividades do processo a priori, o analista deve aplicar um filtro com a Regra 2:

Se todas as atividades contidas na instância a priori estiverem contidas na instância minerada **E** a instância **não** pertence à variante onde está localizada a instância a priori
Então instância é classificada como desvio
Senão a instância é classificada como incompleta.

Regra 2: Regra para instância não conforme e incompleta

O resultado será um log com todas as instâncias que têm desvio, mas realizam todas as atividades da instância a priori e mais algumas que ainda não se sabe se afetam a regulamentação.

Para eliminar deste log as instâncias que tem desvios regulatórios, deve-se aplicar o filtro de atributo com a atividade proibida, selecionado a opção Forbiden, esta opção apagará todas as instâncias que contiverem o valor selecionado. Como mostrado na sequência do Apêndice.8.7.

4.4.3. Identificação dos Desvios regulatórios com atividades proibidas.

Esta etapa tem como objetivo descobrir as instâncias que executam alguma atividade proibida pela regulamentação. As atividades proibidas foram levantadas no formulário de informações sobre o domínio. A entrada é o log obtido que contém todas as instâncias e a saída é o log com as instâncias que não atendem a regulamentação, ou seja, executam atividades proibidas.

O analista deve aplicar um filtro com a seguinte regra:

Se a instância minerada possui pelo menos uma atividade proibida
Então a instância é classificada como desvio regulatório.

Regra 3 – Regra para desvio regulatório

O exemplo de aplicação da regra é mostrado no apêndice 8.4.

4.4.4. Identificação dos Desvios Regulatórios devido à restrição de valor de atributo.

Esta atividade tem como objetivo aplicar as restrições de valor levantadas no formulário de informações sobre o domínio, para verificar se nas instâncias ocorrem valores de atributos que não atendem aos valores máximos e mínimos estabelecidos

na regulamentação. A entrada é o log com todas as instâncias e a saída é o log com as instâncias com desvio regulatório devido à restrição de valor de atributo.

Nesta atividade, deve-se procurar qual campo do log corresponde ao atributo que tem a restrição de valor aplicar a seguinte regra:

Se o valor do campo for menor que o valor mínimo OU o valor do campo for maior que o valor máximo

Então a instância é classificada como desvio regulatório.

Regra 4 – Regra de desvio regulatório por restrição de valor

Aplica-se a regra abaixo a todos os campos que tem alguma restrição de valor. É importante nesta fase observar que o valor pode estar nos outros campos, aqueles que não foram utilizados na mineração, mas que agora podem ser fundamentais para a aplicação da regra.

Os filtros e algoritmos de verificação de regras apresentados são capazes de encontrar as regras no log e gerar um novo log apenas com as instâncias que atendem à regra. As regras serão criadas de acordo com os requisitos levantados no formulário de informações sobre o domínio.

Para comprovar a efetividade do método foi realizado um estudo exploratório com um log de eventos reduzido, e posteriormente um estudo de caso com aproximadamente 1 ano de fabricação.

5. Avaliação da proposta

Neste capítulo é apresentada a avaliação da solução. Primeiro, é discutido um estudo exploratório realizado para coletar resultados preliminares sobre a viabilidade do método proposto, por fim, é apresentado o estudo de caso com dados de um ano de produção de uma empresa real, utilizando o método proposto.

Na Seção 5.1, será apresentado o estudo exploratório, este foi realizado utilizando uma base de dados reduzida, com o objetivo de testar todas as etapas do método, este estudo foi fundamental para ajustar a técnica de pré processamento e as perguntas do formulário de informações sobre o domínio.

Na Seção 5.2, é apresentado o estudo de caso, com um log completo de 1 ano de produção, o resultado da verificação de conformidade e classificação dos desvios. Os procedimentos para aplicação do método estão ilustrados nos apêndices 8.3 a 8.7.

O estudo de caso foi escolhido pois existe um problema definido, uma hipótese e um escopo. Um estudo de caso é comumente usado para investigar um fenômeno contemporâneo dentro de seu contexto de vida real, especialmente quando os limites entre fenômeno e contexto não são claramente evidentes (Recker, 2013).

5.1. O Estudo Exploratório

Os dados utilizados tanto no estudo exploratório e no estudo de caso são reais e por uma questão de confidencialidade os nomes das empresas, funcionários e de alguns equipamentos foram alterados.

Foi elaborada uma pesquisa, e enviada para os especialistas, membros da Associação Internacional de Engenharia Farmacêutica, por sua representatividade no cenário técnico da indústria farmacêutica, 96% consideram que a conformidade do processo de fabricação tem grande influência sobre a qualidade do produto. Com estes resultados, considerando 23 respostas em um universo de 220 especialistas.

A Associação Internacional de Engenharia Farmacêutica (ISPE) é uma associação sem fins lucrativos e foi fundada nos EUA em 1980. O Brasil conta atualmente com cerca de 220 sócios (ISPE, 2014). Os associados representam as indústrias farmacêuticas e de biotecnologia, os produtores de princípios ativos e de

produtos farmacêuticos, agências governamentais, universidades, empresas de engenharia, consultorias, prestadores de serviço, fabricantes de máquinas e equipamentos para as mais diversas aplicações (ISPE, 2014).

O ISPE possui uma série de publicações entre as quais se destaca a Revista Pharmaceutical Engineering, o Jornal ISPEAK, além de vários Baselines (guias técnicos criados em parceria com o FDA e tem como objetivo prover respostas práticas aos dinâmicos e complexos desafios que os projetistas têm que superar em seu trabalho diário). Os guias foram criados para servir como base técnica para profissionais e não tem a intenção de substituir nenhuma norma ou lei vigente (ISPE, 2014).

O objetivo da pesquisa foi verificar se a afirmação de que a qualidade do processo de fabricação de medicamentos influencia diretamente a qualidade do produto seria verdadeira. Foram obtidas 23 respostas, o que equivale a 6% dos associados, valor representativo para uma pesquisa de opinião (Lima, 2015). As perguntas feitas na pesquisa foram:

- Há quantos anos você trabalha na fabricação de medicamentos? Esta pergunta teve como finalidade avaliar a experiência do pesquisado no domínio.
- Qual o seu nível funcional na empresa? Esta pergunta tem como objetivo avaliar qual a influência do entrevistado no processo de fabricação se operacional, tático ou estratégico.
- Qual a influência da qualidade da matéria prima na qualidade final do medicamento? Esta pergunta tem como objetivo verificar a influência da matéria prima na conformidade final do produto.
- Qual a influência da conformidade do processo de fabricação na qualidade do produto? Esta pergunta quem como objetivo verificar o grau de influência da conformidade do processo de fabricação na conformidade do produto final.
- A qualidade do produto fabricado está diretamente relacionada com a conformidade do processo de fabricação?

O público que respondeu a pesquisa é composto de 46% de profissionais com mais de 10 anos de experiência, 23 % de profissionais entre 5 a 10 anos de experiência e 31 % de profissionais entre 3 a 5 anos de experiência, demonstrando que em sua maioria são profissionais sêniores.

O nível funcional do público que respondeu é composto de 8% de diretores, 23% de gerentes, 31% de supervisores, 31% de encarregados e 8% de

operadores, o que demonstra que a maioria dos entrevistados está no nível tático, ou seja, diretamente envolvidos na produção de medicamentos. As repostas foram processadas e são apresentadas na Figura 5-1.

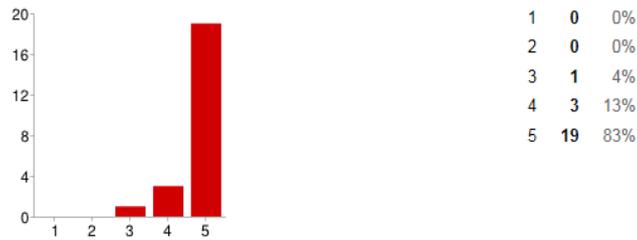
Na pergunta sobre a influência da qualidade da matéria prima na qualidade final do medicamento, 83% responderam que tem muito grande influência, 13% respondeu que tem grande influência e 4% respondeu que tem média influência, o que mostra que a maioria dos entrevistados considera a qualidade da matéria prima para a qualidade final do medicamento.

Na pergunta sobre a influência da conformidade do processo de fabricação na qualidade do produto? 83% responderam que tem muito grande influência, 9% respondeu que tem grande influência, e 9% respondeu que tem média influência, Isto mostra que a maioria dos pesquisados considera que a conformidade do processo tem grande influência na qualidade do produto.

Na pergunta sobre a qualidade do produto fabricado está diretamente relacionada com a conformidade do processo de fabricação? 87% responderam que tem muito grande influência, 9% que tem grande influência e 4% que tem média influência, isto é 96% considera que a conformidade do processo de fabricação tem grande influência sobre a qualidade do produto. Com estes resultados, considerando 23 repostas em um universo de 220 especialistas, e o percentual de 96% de grande influência ou maior, a amostra necessária para 95% de precisão da resposta é de 16,89 repostas, portanto a amostra é significativa (Lima, 2015).

Portanto, de acordo com os especialistas do domínio, existem indícios de que a conformidade do processo de fabricação tem influência direta na qualidade do produto. Na Figura 5-1 o eixo vertical mostra a quantidade de repostas e o eixo horizontal a resposta escolhida pelo especialista.

Qual a influência da qualidade da matéria prima na qualidade final do medicamento ?



Qual a influência da conformidade do processo de fabricação na qualidade do produto ?



A qualidade do produto fabricado está diretamente relacionada com a conformidade do processo de fabricação?



Figura 5-1 - Resultado da pesquisa

O estudo exploratório tem com objetivo de verificar a aplicação do método no log de eventos dos processos de negócio a partir do sistema SCADA Ifix (GE FANUC, n.d.) na empresa PharmaX, utilizado para coletar dados dos processos de fabricação de remédios. Para este estudo foram utilizados os logs do túnel de despirogenização (esterilização por ar seco), equipamento responsável pela esterilização de frascos de vidro. Foram utilizados um levantamento de requisitos do sistema através da documentação fornecida pela empresa, especificação funcional do sistema e instrução de trabalho, que contém o passo a passo da operação. Posteriormente foi realizada a modelagem do processo padrão (Figura 5-2) utilizando a ferramenta Disco (Fluxicon, 2013) e para tal, foram utilizados a documentação do sistema e entrevista com o especialista do domínio vide Tabela 5-1 - Formulário de informações sobre o processo túnel de despirogenização.

Tabela 5-1 - Formulário de informações sobre o processo túnel de despirogenização

Especialista do Domínio:	C.A.A.
Data e Hora:	05/10/2014
Cargo:	Diretor Industrial
Função (descrição da função em relação ao domínio):	Responsável técnico pela fabricação, e farmacêutico responsável.
Empresa	Pharma x
Analista (nome do responsável por realizar a entrevista.)	Guilherme Neves
Processo estudado (nome do processo)	Despirogenização
Objetivos do Processo:	O processo tem como o objetivo remover o resíduo de micro-organismos mortos na embalagem.
Descrição do Processo:	O processo consiste em passar através de um túnel de despirogenização os frascos de medicamento vazios, após a lavagem, o frasco é submetido à ventilação forçada a alta temperatura e depois resfriado para o enchimento.
Atributos (descreva a lista de atributos que devem ser considerados no log)	Número do lote Nome do operador.
Formato do Log (texto, banco de dados, planilha, etc.)	Arquivo texto separado por vírgula.
Qual atributo marca o início e fim do processo e como identifica-lo?	Início -Administração do operador e logo depois digitação do número do lote. Fim – Desligar esteira.
Qual a norma ou regulamento se aplicam a este processo?	RDC17
Quais são os parâmetros críticos do processo em relação à regulamentação?	Temperatura do aquecimento $\geq 300^{\circ}$
Quais são as atividades críticas em relação à regulamentação que não podem acontecer?	Temperatura baixa.
Quais são as atividades críticas em relação à regulamentação que precisam acontecer?	Temperatura de aquecimento $\geq 300^{\circ}$
Critérios de decisão (conforme, desvio e regulatório)	Se a embalagem não ficar aquecida a 300° por 30 min, então não conforme.

A instância a priori, do processo padrão descrito pelos especialistas do domínio e confirmado com a documentação do procedimento operacional padrão é apresentada na Tabela 5-2.

Tabela 5-2 - Processo A priori

case_id	activity
1	entrar nome do usuário
1	entrar com senha
1	clique enter
1	clique continuar
1	Entrar com número do lote
1	Ligar circulação de ar e exaustores
1	Ligar aquecimento
1	Reconhece falha
1	Ligar esteira
1	Selecionar o ciclo
1	Selecione a tecla ajuste.
1	Pressione a tecla atualização
1	Falha no empurrador
1	Reconhece falha
1	Inserir o número do lote.
1	Inserir comentário
1	Inserir senha
1	Parar lote
1	Desligar aquecimento
1	Desligar esteira

A instância padrão foi minerada pelo algoritmo Fuzzy miner utilizando a ferramenta Disco (Fluxicon, 2013), o modelo do processo é mostrado na Figura 5-2, também utilizando a ferramenta PROM mostrado na Figura 5-3 (Process Mining Group, TU/E, 2013). Foram utilizadas as duas ferramentas apenas para mostrar que o resultado não depende da ferramenta e sim do algoritmo de mineração.

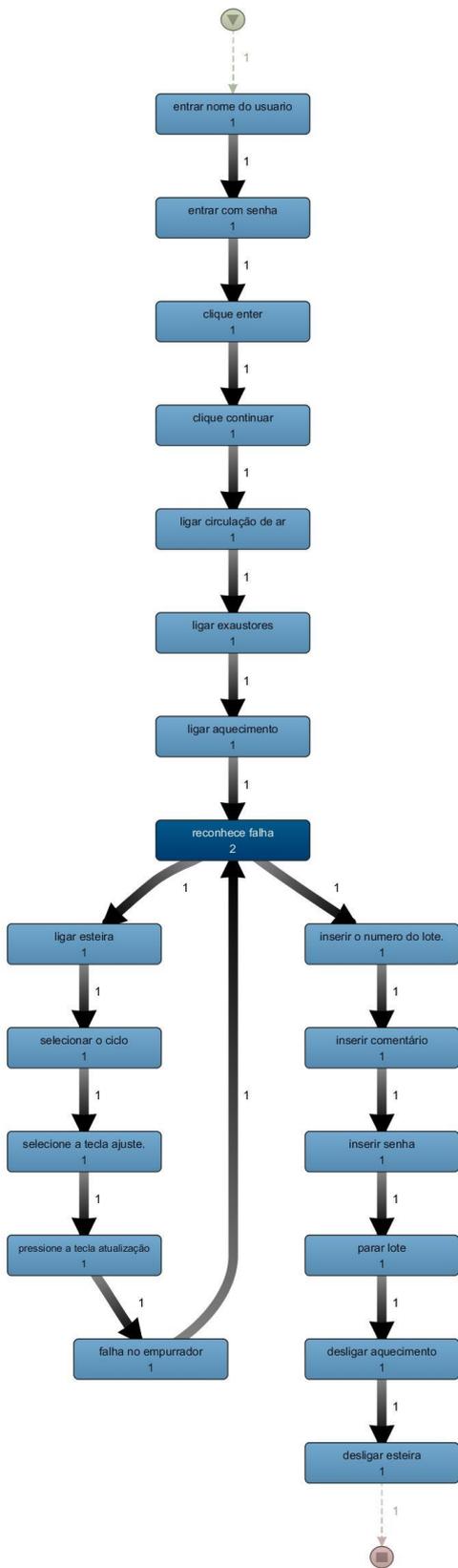


Figura 5-2 - Processo padrão minerado com Fuzzy Miner no Disco

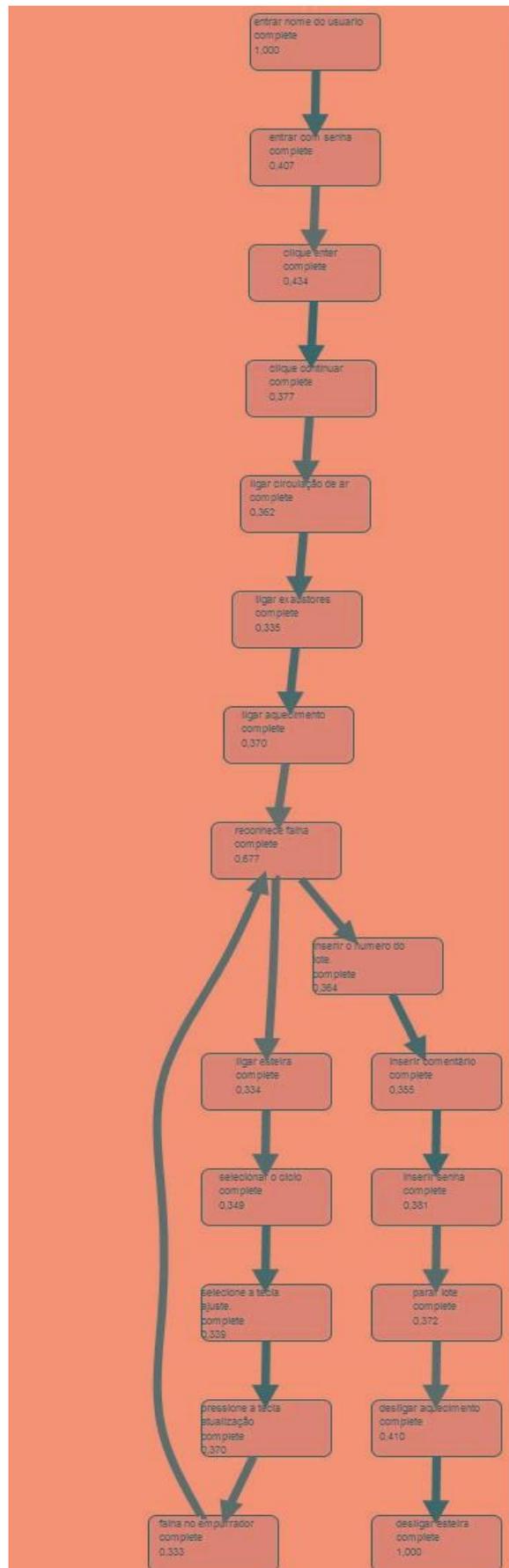


Figura 5-3 - Processo padrão minerado com Fuzzy Miner no PROM

O log original do sistema não continha todos os atributos que eram necessários para a mineração de processos. Como se pode ver na Figura 5-4, não há uma identificação do case_id (instância) do processo, os outros atributos estão presentes como timestamp, recurso, atividade

RecordID	TimeStamp	DeltaToU	UserID	ObjectID	Description	Comme	Checksun
0	29/04/2012 13:44 +2:00		Sistema	Aplicação	Algumas informações foram perdidas anteriormente na tentativa de escrevê-las no arquivo.		74wj00
0	29/04/2012 13:44 +2:00		Sistema	Aplicação	Início do Runtime de WinCC flexible RT 2008 SP2 no aparelho de comando HMI_Panel. Projecto		3p2G4L
48200	29/04/2012 13:59 +2:00		basim	Administração do utilizador	Utilizador 'basim' registado com o grupo 'BASIM'.		JTUJ0I
48201	29/04/2012 13:59 +2:00		basim	Mensagem: 26	Confirmação da mensagem '1010 Disjuntor desarmado' do comando PLC_1.		TUJj11
48202	29/04/2012 13:59 +2:00		basim	Mensagem: 133	Confirmação da mensagem '9990 Falta número do lote' do comando PLC_1.		1CQ1cR
48203	29/04/2012 13:59 +2:00		basim	Mensagem: 128	Confirmação da mensagem '3582 Empurrador de frascos não está em posição' do comando PLC		JqmftC
48204	29/04/2012 13:59 +2:00		basim	Mensagem: 119	Confirmação da mensagem '9972 Revise a velocidade dos filtros trox' do comando PLC_1.		mLqLkq
48205	29/04/2012 13:59 +2:00		basim	Mensagem: 114	Confirmação da mensagem '9001 Controle de tensão desligado' do comando PLC_1.		k7aP1L
48206	29/04/2012 13:59 +2:00		basim	Mensagem: 42	Confirmação da mensagem '3000 Chave de emergência acionada' do comando PLC_1.		#NOME?
48207	29/04/2012 13:59 +2:00		basim	Mensagem: 36	Confirmação da mensagem '3004 Botão de emergência pressionado no empurrador de frascos		MXmbX0
48208	29/04/2012 13:59 +2:00		basim	Mensagem: 13	Confirmação da mensagem '8400 Falha na tensão de alimentação' do comando PLC_1.		SFuQXH
48209	29/04/2012 13:59 +2:00		basim	Mensagem: 85	Confirmação da mensagem '9460 Temperatura muito baixa' do comando PLC_1.		SwLpkk
48210	29/04/2012 13:59 +2:00		basim	Mensagem: 128	Confirmação da mensagem '3582 Empurrador de frascos não está em posição' do comando PLC		OyM2cq
48211	29/04/2012 14:04 +2:00		basim	Mensagem: 114	Confirmação da mensagem '9001 Controle de tensão desligado' do comando PLC_1.		lZUWws
48212	29/04/2012 14:04 +2:00		basim	Mensagem: 114	Confirmação da mensagem '9001 Controle de tensão desligado' do comando PLC_1.		zpkafF
48213	29/04/2012 14:04 +2:00		basim	Mensagem: 114	Confirmação da mensagem '9001 Controle de tensão desligado' do comando PLC_1.		D/IsnO
48214	29/04/2012 14:04 +2:00		basim	Mensagem: 114	Confirmação da mensagem '9001 Controle de tensão desligado' do comando PLC_1.		ttFpSM
48215	29/04/2012 14:04 +2:00		basim	Mensagem: 114	Confirmação da mensagem '9001 Controle de tensão desligado' do comando PLC_1.		ZIN0dG
48216	29/04/2012 14:06 +2:00		basim	Mensagem: 114	Confirmação da mensagem '9001 Controle de tensão desligado' do comando PLC_1.		2KxrOj
48217	29/04/2012 14:06 +2:00		basim	Mensagem: 128	Confirmação da mensagem '3582 Empurrador de frascos não está em posição' do comando PLC		SL+L2b
48218	29/04/2012 14:06 +2:00		basim	Mensagem: 85	Confirmação da mensagem '9460 Temperatura muito baixa' do comando PLC_1.		hYLa3E
48219	29/04/2012 14:11 +2:00		Sistema	Administração do utilizador	Utilizador já fez log out.		twbXbf
48220	29/04/2012 14:22 +2:00		basim	Administração do utilizador	Utilizador 'basim' registado com o grupo 'BASIM'.		KD1Jgk

Figura 5-4 - Log original do sistema

A próxima atividade realizada foi a coleta e pré-processamento do log de eventos com o objetivo de criar um dataset compatível com as ferramentas PROM (Process Mining Group, TU/E, 2013) e Disco (Fluxicon, 2013). No log original, tínhamos os campos específicos do processo, recurso e *timestamp*, no entanto nenhum dos atributos era capaz de identificar de forma única uma instância do processo. No log pré-processado, foi utilizado o número do lote que estava contido originalmente no atributo descrição, e foi necessário criar uma nova coluna chamada case_id para colocar este valor. Como o número do lote só aparece uma vez no início da instância o valor foi repetido até que se achasse um novo número de lote. Este procedimento só é válido, porque os lotes são sequenciais e a linha de produção é única, portanto só é produzido um lote de cada vez, e o log está ordenado em ordem cronológica.

Para verificar o algoritmo que geraria um modelo de processo de melhor compreensão para o especialista, e levando em conta a experiência com os trabalhos relacionados foram testados dois algoritmos o Heuristic Miner e Fuzzy Miner, foram utilizados os softwares de mineração mais populares: Disco na sua versão acadêmica e PROM com licença aberta.

A mineração do processo foi feita utilizando o software Disco (Fluxicon, 2013) que gera um modelo baseado no algoritmo FUZZY Miner, pois este software só trabalha com este algoritmo e utilizando o software PROM 6.3 (Process Mining Group, TU/E, 2013) com os algoritmos Heuristic Mining e Fuzzy.

5.1.1.Mineração com Heuristic Miner

Foi realizado um estudo para verificar se o algoritmo Heuristic Miner poderia ser utilizado para gerar o modelo de processo para o domínio.

Foi utilizada a ferramenta de mineração PROM (Process Mining Task Force, 2014), após a seleção do arquivo, é preciso selecionar o tipo do arquivo para que a ferramenta leia corretamente o log, e após a importação deve ser escolhido o algoritmo de mineração ou a operação a ser realizada. Foi selecionada a Mine for a Causal Net using Heuristic Miner, pois no trabalho em 3.2 (UNISIST, 2012), o algoritmo foi utilizado com sucesso.

Após a seleção do algoritmo, é necessário que sejam definidos os parâmetros que vão definir o nível de abstração da mineração, optou-se por manter todos os valores padrão. O resultado obtido na mineração foi um modelo de processo muito detalhado e de difícil compreensão para o especialista do domínio, conforme mostrado na Figura 5-5. Os algoritmos de descoberta podem ser mais ou menos simples, no entanto com logs mais heterogêneos os modelos tendem a ficarem mais simples com algoritmos do tipo Fuzzy Miner por sua característica de tratamento de ruídos como mostrado em 2.2.5.

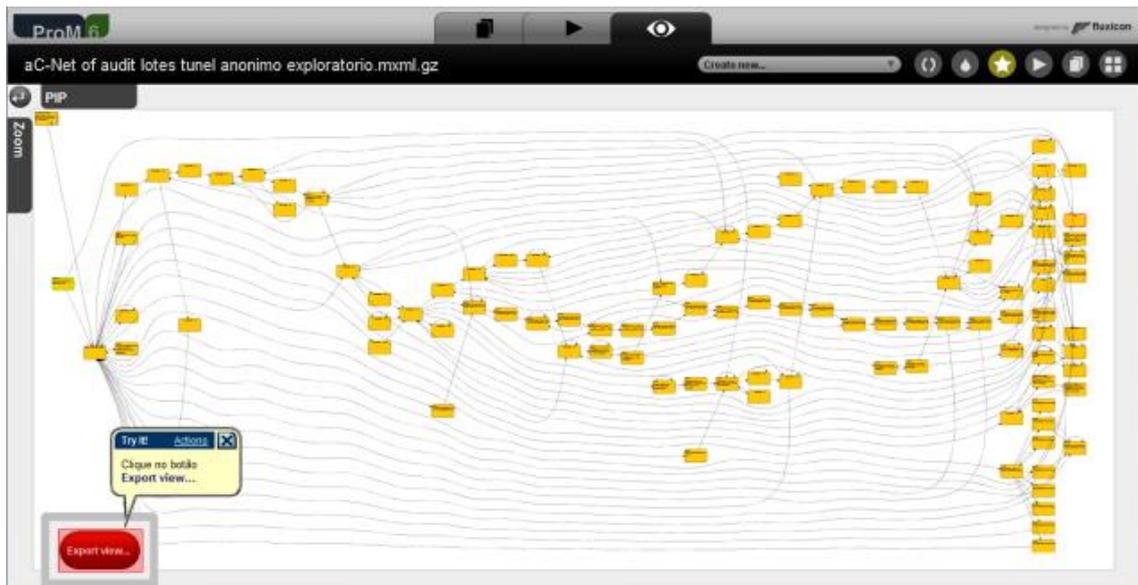


Figura 5-5 - Resultado da mineração com Heuristic Miner

Por este motivo a opção do uso deste algoritmo não foi aprovada, passando para o teste com outro algoritmo. O mesmo log foi minerado utilizando o algoritmo Fuzzy e o modelo encontrado foi melhor compreendido pelo especialista.

5.1.2. Mineração com Mine for a Fuzzy Model utilizando o PROM.

Utilizando o mesmo log e a mesma ferramenta, foi realizada a mineração, porém foi selecionado o algoritmo Mine for a Fuzzy Model, conforme discutido em 2.2.5. Este algoritmo funciona muito bem para processos mais complexos, pois consegue destacar as partes importantes e abstrair dos detalhes, também no trabalho relacionado em 3.3, foi utilizado este algoritmo com sucesso, e por se tratar de domínio semelhante o resultado tende a ser similar. O estudo exploratório teve o propósito de confirmar esta evidência (Figura 5-6).

As métricas para a mineração e os parâmetros de atenuação foram mantidas no padrão. O modelo descoberto apresentou uma quantidade de atividades menor, com destaque para as de maior frequência. De acordo com o especialista do domínio, este modelo é mais simples e fácil de compreender que o produzido pelo Heuristic Miner, e representando de forma mais clara o processo.

Utilizando o mesmo log, foi selecionado o plug-in LTL Checker Default, e após a seleção foi preciso selecionar um tipo de regra para ser verificado, conforme mostram a Figura 5-7 e Figura 5-8.

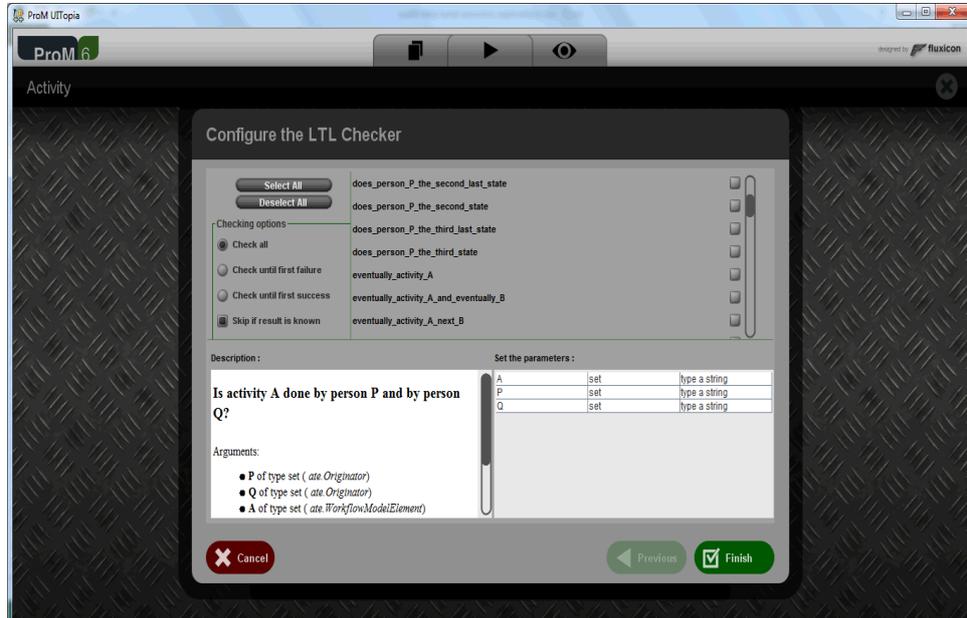


Figura 5-7 - Seleção de tipos de regra no LTL do PROM



Figura 5-8 - Regra de atributo no LTL

Escolhido o tipo de regra, é preciso definir qual o valor que é preciso ser buscado, no nosso caso foi escolhida a atividade do log **Mensagem: 141**, ou seja o LTL vai procurar no log uma atividade que corresponda a este valor.

O resultado encontrado foi 1 caso que atendia a regra e todos os outros que não atendem. No enfoque proposto, esta seria a busca por instâncias do processo que tem atividades proibidas, portanto a instância ou o case_id que atendeu a regra seria classificado como Instância não conforme regulatória. Com este resultado ficou evidenciado que para este domínio é possível minerar regras no log utilizando o LTL conforme mostra a Figura 5-9, desde que o log esteja no formato legível para a ferramenta e tenha os atributos obrigatórios para a mineração conforme descrito no método.

Qualquer regra pode ser minerada desde que ela seja compatível com os modelos pré-definidos na ferramenta. Isto significa que quando a regra for construída deve-se levar em conta a limitação dos modelos da ferramenta.

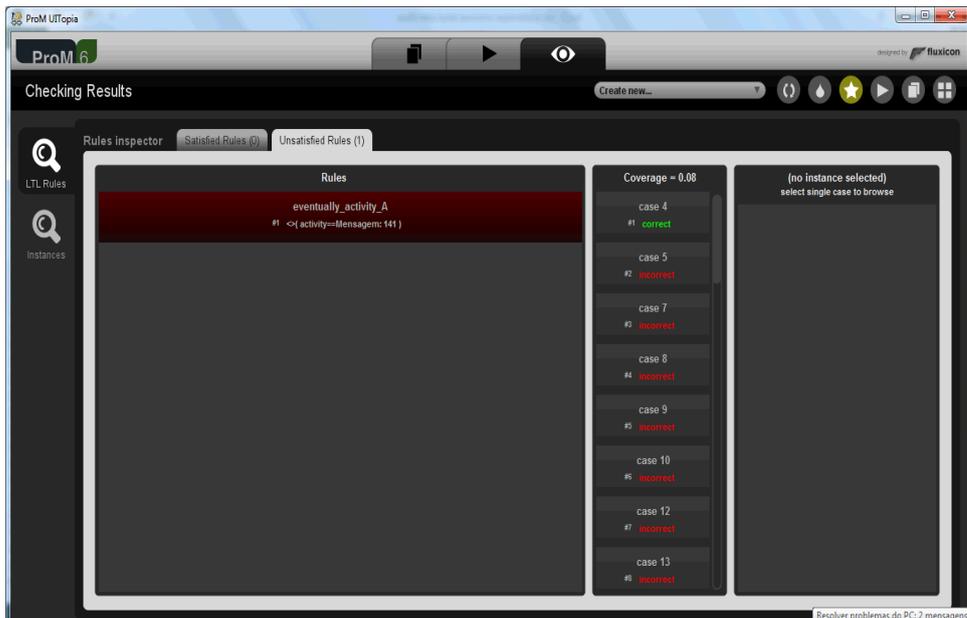


Figura 5-9 - Resultado da mineração de regras utilizando o LTL.

5.1.4. Mineração Fuzzy Model utilizando o DISCO

Outro teste realizado no estudo exploratório foi utilizar mais de uma ferramenta para o mesmo algoritmo, mostrando que os resultados encontrados são similares mesmo com o uso de ferramentas de mineração diferentes.

Foi realizada mineração do log utilizando o DISCO (Fluxicon, 2013), esta é uma ferramenta comercial, que é mais eficiente em termos de desempenho, e com interface mais amigável, no entanto só trabalha com o algoritmo Fuzzy. Foi realizada a importação do log no software Disco, utilizando o arquivo original no formato CSV (esta é outra vantagem da ferramenta, importa arquivos no formato CSV sem necessitar de nenhuma conversão). Para concluir a importação é necessário associar os atributos do DISCO aos campos do log, iniciando pelo case_id.

Feito isso, é necessário selecionar o TimeStamp e o seu formato. É necessário também selecionar o Recurso, que vai executar o evento. Por último é preciso selecionar a activity, ou seja o evento propriamente dito.

O processo foi minerado utilizando a configuração padrão, o modelo resultante é bem claro e de fácil entendimento pelo especialista do domínio que aprovou o modelo. A Figura 5-10 mostra o modelo gerado.

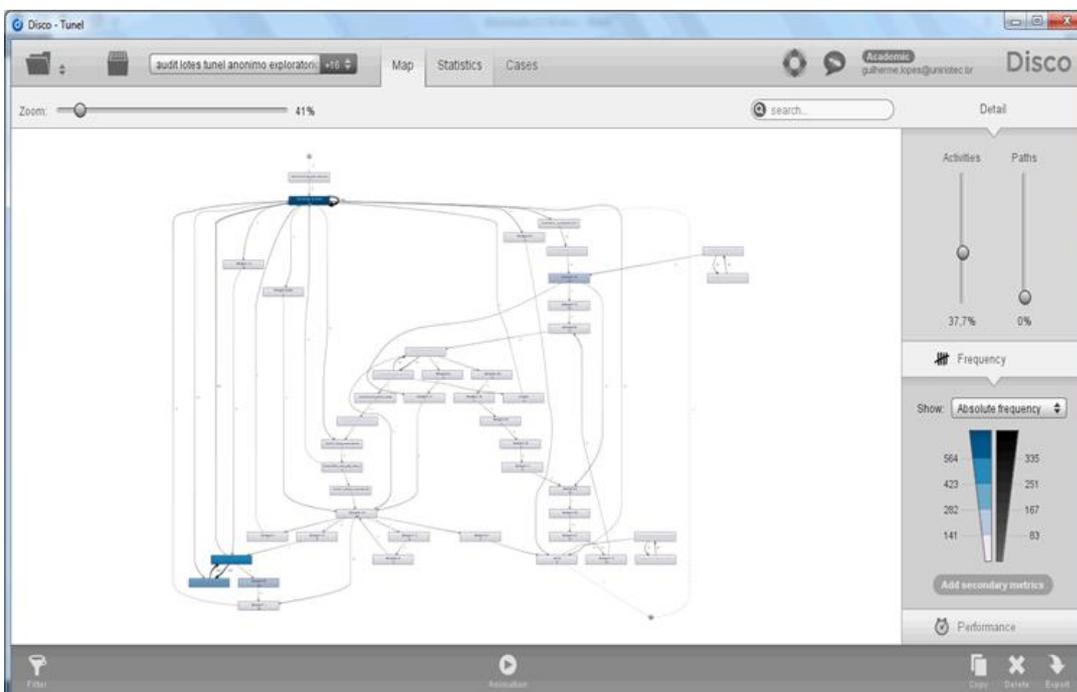


Figura 5-10 - Modelo de processo usando Software Disco

É possível verificar todas as instâncias do processo, e visualizar individualmente o modelo de processo da instância ou o seu log, conforme a Figura 5-11.

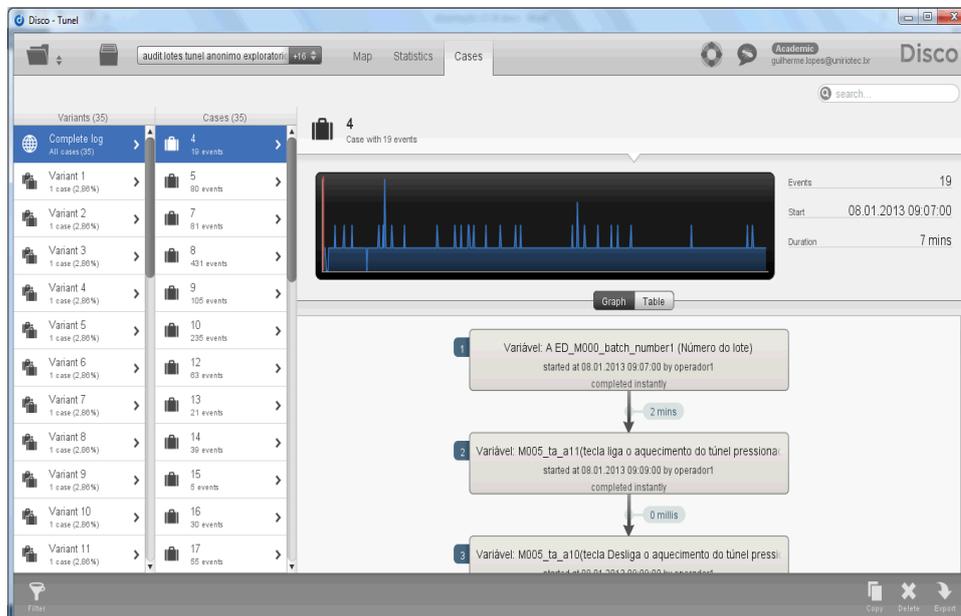


Figura 5-11 - Modelo de processo de uma instância

A mineração de regras no Disco é feita de outra forma, a função equivalente ao LTL é o filtro, existem filtros de tempo, de Variante, de atributo, de desempenho, de endpoints e de follower.

Para o estudo exploratório foi utilizado o filtro de atributo, buscando o mesmo valor utilizado no LTL, Mensagem: 141. O filtro pode assumir um dos 3 valores: Keep Select, Mandatory ou Forbiden. A opção Keep Select remove todos os eventos que não tenham sido selecionados, o Mandatory remove todas as instâncias que não tenham pelo menos um evento com um dos valores selecionados, e o Forbiden remove todas as instâncias que tiverem pelo menos uma das atividades dos valores selecionados. Foi selecionado Mandatory para selecionar apenas as instâncias onde o valor obrigatoriamente aparece.

O resultado encontrado foram três instâncias, mostradas na Figura 5-12.

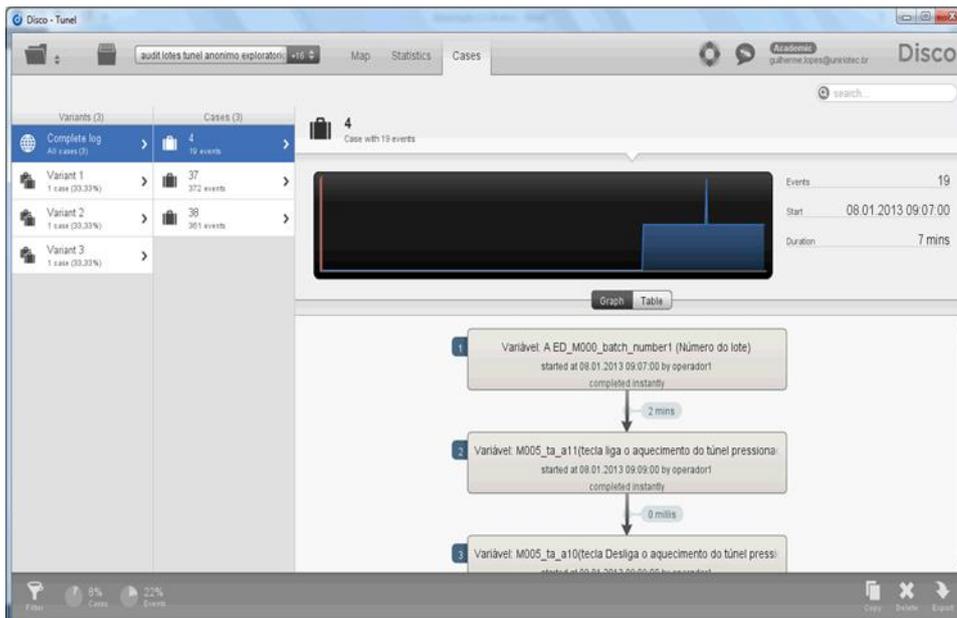


Figura 5-12 - Resultado da aplicação do filtro

Concluindo, as instâncias encontradas são as mesmas utilizando o LTL ou o filtro da ferramenta disco, a diferença entre as duas ferramentas é a interface e a maneira como a regra é montada.

5.2. O Estudo de Caso

Neste estudo de caso o nome da empresa e funcionários foram alterados para cumprir o acordo de confidencialidade firmado na autorização para o uso dos dados.

O estudo de caso teve como objetivo avaliar a aplicação do método no domínio e responder à questão formulada nesta pesquisa: Como identificar dentre um conjunto de medicamentos fabricados tem mais chance de ter problemas de qualidade para serem testados, sem que para isto tenha que se destruir todo o produto?

O papel do analista foi desempenhado pelo pesquisador e alguns especialistas participaram do estudo de caso. Os especialistas foram convidados de acordo com a sua função dentro da empresa, de encarregado de produção ao diretor industrial. A empresa PharmaX é líder mundial no seu segmento, no Brasil está instalada a 35 anos, tem cerca de 180 funcionários no brasil e 1200 no total, dispõe de 4 fábricas, 17 filiais e seus produtos são comercializados em 70 países.

O processo de fabricação de medicamentos faz parte do rígido domínio da indústria farmacêutica. O processo utilizado no estudo de caso consiste na esterilização e despirogenização de frascos de medicamentos, que é o processo que retira os resíduos

de micro organismos que ficam retidos após a lavagem dos frascos. O processo tem três fases: uma fase de aquecimento a 300° C e insuflação de ar por um período de tempo que depende do tamanho do frasco, a segunda fase de pré-resfriamento e a terceira fase de resfriamento. A não observância das fases pode provocar contaminação do frasco, ou não retirar os resíduos ou ainda colocar em risco o operador entregando um frasco quente que na próxima etapa vai receber um jato de água fria, podendo provocar a explosão do frasco com graves danos ao operador ou até leva-lo à morte no caso dos frascos maiores. Nenhum lote de medicamento pode ser liberado não conforme antes da devida avaliação do risco ao paciente.

A partir de agora é descrito o resultado do passo a passo do método proposto.

5.2.1.Primeira fase – Levantamento de informações sobre o domínio

O objetivo desta fase é levantar as informações do domínio necessárias para a obtenção da instância a priori do processo para ser usada como instância padrão para comparação com as outras instâncias do processo. É importante que além do nome do atributo se faça uma pequena descrição do que significa e onde é utilizado. Dentre os atributos, devem existir obrigatoriamente um atributo que identifique de forma única uma instância do processo (`case_id`), um que identifique a hora e dia que iniciou a execução da atividade (`TimeStamp`), um para atividade (`activity`), e um para identificar o recurso (`resource`) que executou a atividade.

A partir dos documentos operacionais padrões, foram levantados os seguintes requisitos: o sistema deve atender às exigências do FDA 21CFR Part11 (Registros Eletrônicos; Assinaturas Eletrônicas), a RDC 17 da ANVISA de 16 de abril de 2010 e, todo o ciclo de vida da documentação referente à validação do sistema deve seguir as diretrizes do GAMP5 e do Guia de Validação de Sistemas Computadorizados emitido pela ANVISA em abril de 2010.

O túnel trabalha com um sistema de esterilização, em conjunto com o fluxo laminar. Os frascos ainda molhados que saem da lavadora, passam gradativamente, numa esteira transportadora, por quatro zonas diferentes como mostra a Figura 5-13:

- Zona de Pré-aquecimento;
- Zona de Esterilização;

- Zona de Resfriamento 1;
- Zona de Resfriamento 2.

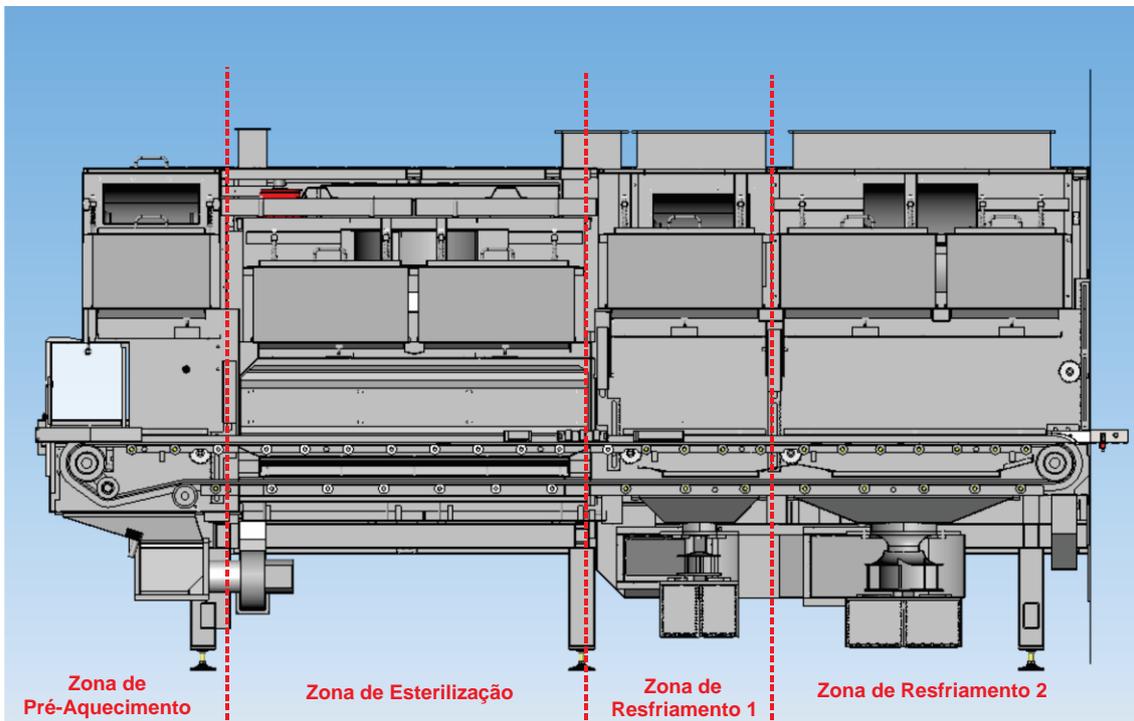


Figura 5-13 - Visão do sistema túnel de despirogenização (manual sistema)

ZONA DE PRÉ-AQUECIMENTO

Na zona de pré-aquecimento o fluxo laminar insufla ar através de um filtro esterilizado nos frascos que saem da lavadora com ar esterilizado. O ar quente que flui da zona de esterilização aquece a zona de pré-aquecimento e faz o processo de secagem dos frascos ao mesmo tempo. Quando o ar do pré-aquecimento aquece acima de uma temperatura de set point, resulta em uma troca de ar. Com isto entra ar fresco na parte superior do túnel e simultaneamente o ar quente desce para um canal de exaustão.

ZONA DE ESTERELIZAÇÃO

Na zona de esterilização, os frascos pré-aquecidos são completamente secos e esterilizados. O fluxo de ar quente que circula através de um duto no qual as resistências elétricas de aquecimento que elevam a temperatura do ar até 360°C. Este fluxo de ar passa por um filtro que dirige o ar aquecido garantindo assim um aquecimento uniforme nos frascos.

ZONA DE RESFRIAMENTO 1

Na zona de resfriamento 1, os frascos esterilizados são resfriados pelo fluxo laminar. O fluxo laminar suga o ar fresco resfriado e pressiona-o através de um filtro esterilizado para dentro dos frascos aquecidos. Simultaneamente o exaustor, por baixo dos frascos, suga o ar utilizado, expelindo-o para fora da máquina, através de um duto.

ZONA DE RESFRIAMENTO 2

Na zona de resfriamento, os frascos saem do túnel já resfriados à temperatura ambiente. Chegam à esteira de entrada da máquina enchedora através de uma esteira de transferência, para os frascos serem envazados. O fluxo laminar e o exaustor são adaptados um ao outro de tal forma, que seja mantida a alta pressão necessária dentro do túnel, para assegurar que não entre nenhum ar não esterilizado. As regras de funcionamento são mostradas na Tabela 5-3.

Tabela 5-3 - Tabela de regras do túnel

Zona de esterilização	Temperatura 360°
Zona de resfriamento 1	Fluxo Laminar Ativo
Zona de resfriamento 2	Fluxo Laminar Ativo
Zona de resfriamento 2	Exaustor Ativo
Zona de resfriamento 2	Pressão no interior do túnel

O levantamento de requisitos do sistema através da documentação fornecida pela empresa estudada, especificação funcional do sistema e instrução de trabalho, que contém o passo a passo da operação foram utilizados. Posteriormente foi gerado um log com uma instância padrão e minerada através do software Disco (Fluxicon, 2013), para tal foram utilizados a documentação do sistema e entrevistas com os analistas.

Tabela 5-4 - Formulário de informações sobre o domínio entrevista com especialista

Especialista do Domínio:	C.A.A.
Data e Hora:	05/10/2014
Cargo:	Diretor Industrial
Função (descrição da função em relação ao domínio):	Responsável técnico pela fabricação, e farmacêutico responsável.
Empresa	Pharma x
Analista (nome do responsável por realizar a entrevista.)	Guilherme Neves
Processo estudado (nome do processo)	Despirogenização
Objetivos do Processo:	O processo tem como o objetivo remover o resíduo de micro-organismos mortos na embalagem.
Descrição do Processo:	O processo consiste em passar através de um túnel de despirogenização os frascos de medicamento vazios, após a lavagem, o frasco é submetido a ventilação forçada a alta temperatura e depois resfriado para o enchimento.
Atributos (descreva a lista de atributos que devem ser considerados no log)	Número do lote Nome do operador.
Formato do Log (texto, banco de dados, planilha, etc.)	Arquivo texto separado por vírgula.
Qual atributo marca o início e fim do processo e como identificá-lo?	Início -Administração do operador e logo depois digitação do número do lote. Fim – Desligar esteira.
Qual a norma ou regulamento se aplicam a este processo?	RDC17
Quais são os parâmetros críticos do processo em relação à regulamentação?	Temperatura do aquecimento $\geq 300^{\circ}$
Quais são as atividades críticas em relação à regulamentação que não podem acontecer?	Temperatura baixa.
Quais são as atividades críticas em relação à regulamentação que precisam acontecer?	Temperatura de aquecimento $\geq 300^{\circ}$

Critérios de decisão (conforme, desvio e regulatório)	Se a embalagem não ficar aquecida a 360 ° C, então não conforme.
---	--

O processo padrão descrito pelos especialistas do domínio e confirmado com a documentação do procedimento operacional padrão é mostrado na Tabela 6.

Tabela 5-5 - processo padrão

case_id	activity		
1	entrar nome do usuario		
1	entrar com senha		
1	clique enter		
1	clique continuar		
1	ligar circulação de ar		
1	ligar exaustores		
1	ligar aquecimento		
1	reconhece falha		
1	ligar esteira		
1	selecionar o ciclo		
1	selecione a tecla ajuste.		
1	pressione a tecla atualização		
1	falha no empurrador		
1	reconhece falha		
1	inserir o numero do lote.		
1	inserir comentário		
1	inserir senha		
1	parar lote		
1	desligar aquecimento		
1	desligar esteira		

Após ser minerado pelo algoritmo Fuzzy Miner o modelo do processo é mostrado na Figura 5-14.

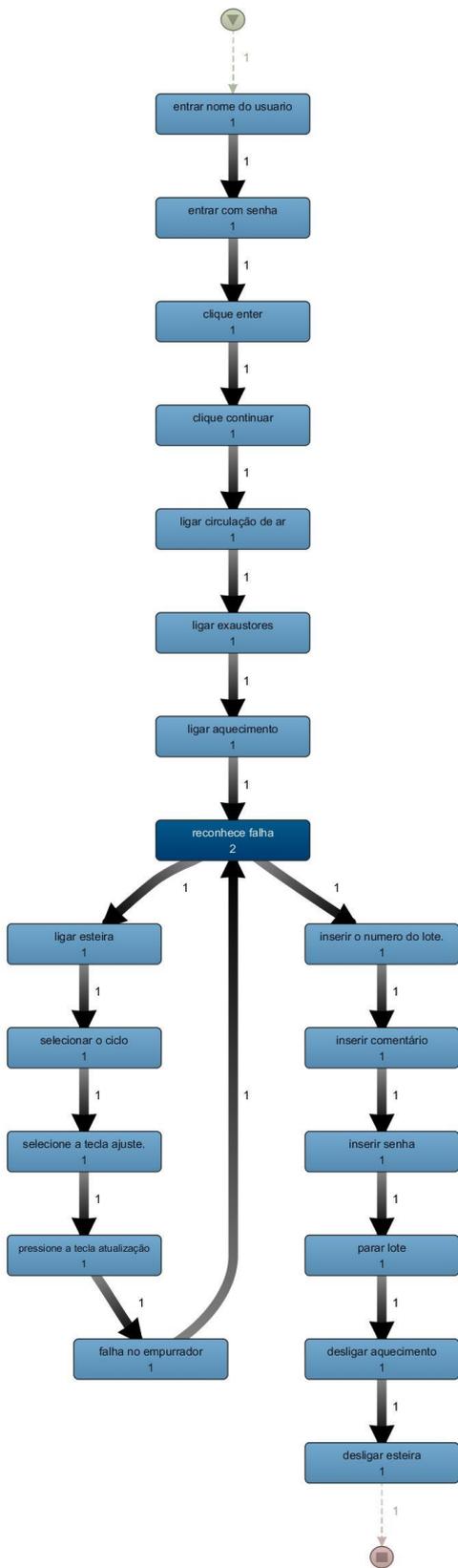


Figura 5-14 - Processo padrão minerado com Fuzzy Miner no Disco

Após uma primeira verificação, é possível perceber que cada esta atividade tem correspondentes no log gerado pelo equipamento, portanto foi feito um de-para associando as atividades descritas no formulário com a saída do log da máquina, conforme prescrito em 4.1. Este procedimento é necessário para a comparação das instâncias. O resultado foi um processo padrão mostrado na Figura 5-15.

activity	Activity sistema								
entrar nome do usuario / digitar a s	Administração do utilizador								
entrar com numero do lote	Variável: A ED_M000_batch_number1 (Número do lote)								
ligar circulação de ar e exaustores	Variável: M005_ta_a31(tecla liga o exaustor e a circulação de ar pressionada)								
ligar aquecimento	Variável: M005_ta_a11(tecla liga o aquecimento do túnel pressionada)								
ligar esteira	Variável: M005_ta_a01(tecla ligar o acionamento da esteira do túnel pressionada)								
selecionar o ciclo	Variável: M030_int_f_nr_neu(Número do Ciclo)								
selecione a tecla ajuste.	Variável: M030_Cicle_Selection(Botão de confirmação de seleção de ciclo)								
pressione a tecla atualização	Variável: M098_message_updating(Botão de confirmação de atualização de data-hora pressionado)								
desligar aquecimento	Variável: M005_ta_a10(tecla Desliga o aquecimento do túnel pressionada)								
desligar esteira	Variável: M005_ta_a00(tecla desliga o acionamento da esteira do túnel pressionada)								

Figura 5-15 - Instância Padrão com atividades do sistema

Após a nova mineração utilizando a ferramenta Disco (Fluxicon, 2013), a visualização do modelo de processo encontrado é mostrada na Figura 5-16.

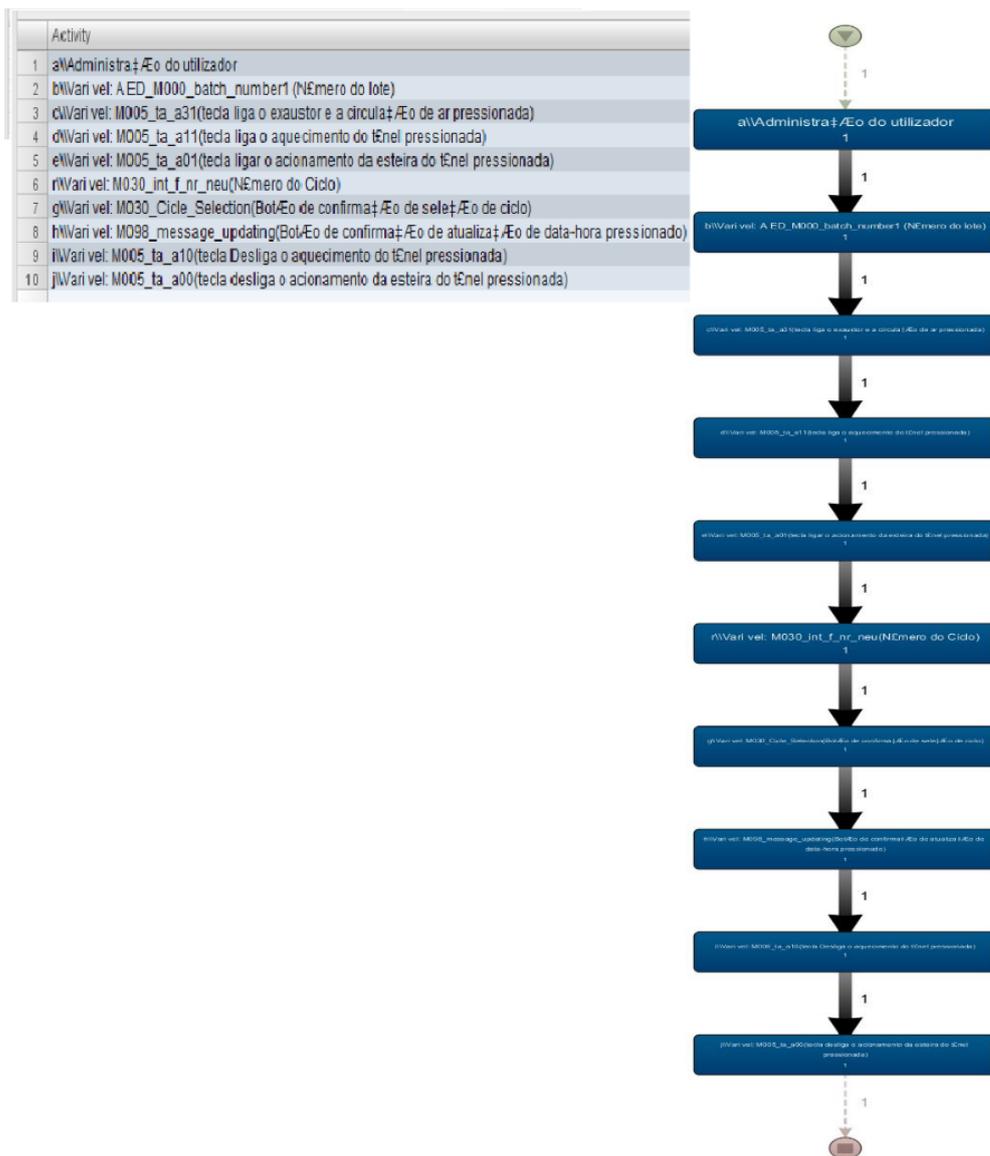


Figura 5-16 - Instância padrão

5.2.2.Segunda Fase – Levantamento dos atributos obrigatórios nos logs e pré-processamento.

O objetivo desta fase é a criação do relacionamento entre as respostas do formulário e os campos do log. O analista deve procurar no log os campos que tem o mesmo significado dos atributos levantados no formulário. O relacionamento deve ser um para um, fazendo uma associação do campo do log com o atributo.

- **Identificação do Case_id**

O objetivo desta etapa é a identificação do case_id no arquivo de log. A entrada é o arquivo de log e a saída é o arquivo de log ordenado pelo case_id. No caso do valor do case_id não estar presente no registro, deve-se preencher manualmente levando em conta as informações do Formulário de informações sobre o domínio.

Segundo a resposta do Formulário de informações sobre o domínio, o identificador único da instância do processo é o número do lote, portanto o número do lote foi associado ao case_id. Porém, o log original do sistema não tinha um campo explícito com o número do lote (Figura 5-17), portanto foi necessário busca-lo dentro do campo descrição.

RecordID	TimeStamp	DeltaToU	UserID	ObjectID	Description	Comm	Checksun
0	29/04/2012 13:44 +2:00		Sistema	Aplicação	Algumas informações foram perdidas anteriormente na tentativa de escrevê-las no arquivo.		74wjQ0
0	29/04/2012 13:44 +2:00		Sistema	Aplicação	Início do Runtime de WinCC flexible RT 2008 SP2 no aparelho de comando HMI_Panel. ProjectG3p2G4L		
48200	29/04/2012 13:59 +2:00		basim	Administração do utilizador	Utilizador 'basim' registado com o grupo 'BASIM'.		JTUJOI
48201	29/04/2012 13:59 +2:00		basim	Mensagem: 26	Confirmação da mensagem '1010 Disjuntor desarmado' do comando PLC_1.		TUjji1
48202	29/04/2012 13:59 +2:00		basim	Mensagem: 133	Confirmação da mensagem '9990 Falta número do lote' do comando PLC_1.		1CQ1cR
48203	29/04/2012 13:59 +2:00		basim	Mensagem: 128	Confirmação da mensagem '3582 Empurrador de frascos não está em posição' do comando PLC		JqmftC
48204	29/04/2012 13:59 +2:00		basim	Mensagem: 119	Confirmação da mensagem '9972 Revise a velocidade dos filtros trox' do comando PLC_1.		mlLkq
48205	29/04/2012 13:59 +2:00		basim	Mensagem: 114	Confirmação da mensagem '9001 Controle de tensão desligado' do comando PLC_1.		k7aP1L
48206	29/04/2012 13:59 +2:00		basim	Mensagem: 42	Confirmação da mensagem '3000 Chave de emergência acionada' do comando PLC_1.		#NOME?
48207	29/04/2012 13:59 +2:00		basim	Mensagem: 36	Confirmação da mensagem '3004 Botão de emergência pressionado no empurrador de frascos		MXmBX0
48208	29/04/2012 13:59 +2:00		basim	Mensagem: 13	Confirmação da mensagem '8400 Falha na tensão de alimentação' do comando PLC_1.		SFuQXH
48209	29/04/2012 13:59 +2:00		basim	Mensagem: 85	Confirmação da mensagem '9460 Temperatura muito baixa' do comando PLC_1.		SwLpkk
48210	29/04/2012 13:59 +2:00		basim	Mensagem: 128	Confirmação da mensagem '3582 Empurrador de frascos não está em posição' do comando PLC		OyM2cq
48211	29/04/2012 14:04 +2:00		basim	Mensagem: 114	Confirmação da mensagem '9001 Controle de tensão desligado' do comando PLC_1.		liZUMs
48212	29/04/2012 14:04 +2:00		basim	Mensagem: 114	Confirmação da mensagem '9001 Controle de tensão desligado' do comando PLC_1.		zpkafF
48213	29/04/2012 14:04 +2:00		basim	Mensagem: 114	Confirmação da mensagem '9001 Controle de tensão desligado' do comando PLC_1.		D/JsNO
48214	29/04/2012 14:04 +2:00		basim	Mensagem: 114	Confirmação da mensagem '9001 Controle de tensão desligado' do comando PLC_1.		ttFPSM
48215	29/04/2012 14:04 +2:00		basim	Mensagem: 114	Confirmação da mensagem '9001 Controle de tensão desligado' do comando PLC_1.		ZIN0dG
48216	29/04/2012 14:06 +2:00		basim	Mensagem: 114	Confirmação da mensagem '9001 Controle de tensão desligado' do comando PLC_1.		2KxrOj
48217	29/04/2012 14:06 +2:00		basim	Mensagem: 128	Confirmação da mensagem '3582 Empurrador de frascos não está em posição' do comando PLC		5L+L2b
48218	29/04/2012 14:06 +2:00		basim	Mensagem: 85	Confirmação da mensagem '9460 Temperatura muito baixa' do comando PLC_1.		hyLa3E
48219	29/04/2012 14:11 +2:00		Sistema	Administração do utilizador	Utilizador já fez log out.		twbXbf
48220	29/04/2012 14:22 +2:00		basim	Administração do utilizador	Utilizador 'basim' registado com o grupo 'BASIM'.		kD1jgk

Figura 5-17 - Log original do sistema

Foi utilizada uma função para preenchimento automático da nova coluna =SE (F3581="Variável: A ED_M000_batch_number1 (Número do lote)"; DIREITA (G3581;11); A3580), esta função aplicada a cada linha descobre o número do lote dentro do campo descrição, o número do lote é repetido manualmente até que se encontre um novo número do lote, o procedimento é repetido até o fim do log. Na Figura 5-18 é mostrado o log resultante com o case_id.

A	B	C	D	E	
case_id		RecordID	TimeStamp	UserID	ObjectID
b29042012'	1	48291	29/04/2012 15:34	basim	Vari vel: A ED_M000_batch_number1 (Número do lote)
b29042012'	1	48292	29/04/2012 15:34	basim	Mensagem: 133
b29042012'	1	48293	29/04/2012 15:34	basim	Mensagem: 85
b29042012'	2	48294	29/04/2012 15:39	Sistema	Administrat#Æo do utilizador
b29042012'	2	48295	29/04/2012 15:44	basim	Administrat#Æo do utilizador
b29042012'	3	48296	29/04/2012 15:52	Sistema	Administrat#Æo do utilizador
b29042012'	3	0	29/04/2012 16:10	Sistema	Arquivo
b29042012'	3	0	29/04/2012 16:10	Sistema	Aplicat#Æo
b29042012'	3	0	29/04/2012 16:10	Sistema	Arquivo
b29042012'	3	48298	29/04/2012 16:11	basim	Mensagem: 131
b29042012'	3	48299	29/04/2012 16:11	basim	Mensagem: 23
b29042012'	3	48300	29/04/2012 16:11	basim	Mensagem: 172
b29042012'	3	48301	29/04/2012 16:11	basim	Mensagem: 171
b29042012'	3	48302	29/04/2012 16:11	basim	Mensagem: 170
b29042012'	3	48303	29/04/2012 16:11	basim	Mensagem: 169
b29042012'	3	48304	29/04/2012 16:11	basim	Mensagem: 142
b29042012'	3	48305	29/04/2012 16:11	basim	Mensagem: 141
b29042012'	3	48306	29/04/2012 16:11	basim	Mensagem: 140

Figura 5-18 - Log após a identificação do case_id

Como o número do lote só aparece uma vez no início da instância o valor for repetido até que se achasse um novo número de lote. Este procedimento só é válido, porque os lotes são sequenciais e a linha de produção é única, portanto só é produzido um lote de cada vez, e o log está ordenado em ordem cronológica.

- **Identificação do TimeStamp**

O objetivo desta etapa é identificar e ordenar o TimeStamp. A entrada é o arquivo de logs e a saída é o arquivo de log ordenado por TimeStamp. O atributo TimeStamp deve estar ordenado em ordem crescente, todos os registros devem ter o mesmo formato até o fim do arquivo. Quando existir mais de um TimeStamp para a mesma atividade, o analista deve escolher o atributo que contiver o menor valor, pois ele irá representar o início da atividade.

O arquivo de log já estava com a coluna onde está o TimeStamp com o nome TimeStamp portanto a associação é automática conforme mostra a Figura 5-19.

case_id	Column 2	RecordID	TimeStamp	UserID	ObjectID	Description
1	b29042012	1	48291	29/04/2012 15:34	basim	Varivel: A ED_M000_batch_number1 (N:mero do
2	b29042012	1	48292	29/04/2012 15:34	basim	Mensagem: 133
3	b29042012	1	48293	29/04/2012 15:34	basim	Mensagem: 85
4	b29042012	2	48294	29/04/2012 15:39	Sistema	Administraç: do do utilizador
5	b29042012	2	48295	29/04/2012 15:44	basim	Administraç: do do utilizador
6	b29042012	3	48296	29/04/2012 15:52	Sistema	Administraç: do do utilizador
7	b29042012	3	0	29/04/2012 16:10	Sistema	Arquivo
8	b29042012	3	0	29/04/2012 16:10	Sistema	Aplicaç: o
9	b29042012	3	0	29/04/2012 16:10	Sistema	Arquivo
10	b29042012	3	48298	29/04/2012 16:11	basim	Mensagem: 131
11	b29042012	3	48299	29/04/2012 16:11	basim	Mensagem: 23
12	b29042012	3	48300	29/04/2012 16:11	basim	Mensagem: 172
13	b29042012	3	48301	29/04/2012 16:11	basim	Mensagem: 171
14	b29042012	3	48302	29/04/2012 16:11	basim	Mensagem: 170
15	b29042012	3	48303	29/04/2012 16:11	basim	Mensagem: 169
16	b29042012	3	48304	29/04/2012 16:11	basim	Mensagem: 142
17	b29042012	3	48305	29/04/2012 16:11	basim	Mensagem: 141
18	b29042012	3	48306	29/04/2012 16:11	basim	Mensagem: 140
19	b29042012	3	48307	29/04/2012 16:11	basim	Mensagem: 139
20	b29042012	3	48308	29/04/2012 16:11	basim	Mensagem: 104
21	b29042012	3	48309	29/04/2012 16:11	basim	Mensagem: 103
22	b29042012	3	48310	29/04/2012 16:11	basim	Mensagem: 65
23	b29042012	3	48311	29/04/2012 16:11	basim	Mensagem: 39
24	b29042012	3	48312	29/04/2012 16:11	basim	Mensagem: 17
25	b29042012	3	48313	29/04/2012 16:11	basim	Mensagem: 56
26	b29042012	3	48314	29/04/2012 16:11	basim	Mensagem: 56
27	b29042012	3	48315	29/04/2012 16:11	basim	Mensagem: 65
28	b29042012	3	48316	29/04/2012 16:11	basim	Mensanem: 54

Figura 5-19 - Identificação do TimeStamp

O próximo passo é identificar o formato da data e hora e associar ao modelo correto na ferramenta de mineração, neste arquivo o formato é dd/mm/yy hh:mm.

- **Identificação da Activity**

Uma Activity é a ação realizada em cada registro da instância do processo. Caso um único campo do log não seja suficiente para representar a atividade podem ser unidos dois ou mais campos para formar uma Activity completa. Esta decisão deve ser tomada com base nas respostas do Formulário de informações sobre o domínio.

Neste log não foi necessário utilizar mais de um campo. Se fosse necessário, a seleção seria feita no campo do log e no ícone Activity, quantos campos fossem necessários, a ferramenta faz o encadeamento dos campos na ordem em que foram selecionados, tratando-os como se fosse um campo único. Neste estudo de caso, o atributo ObjectID foi suficiente para representar a atividade realizada. Para a importação, foi associado o campo ObjectID ao campo Activity da ferramenta de mineração.

- **Identificação do Recurso**

O Recurso é o atributo que representa o recurso que executa a ação, ou seja, o ator que executa a Activity. Este atributo deve ser único, não são permitidos

homônimos, no arquivo de log não ocorreu nenhum homônimo, portanto o atributo pode ser usado sem alterações. O campo UserId do arquivo de log foi associado com o atributo Resource da ferramenta de mineração.

- **Outros atributos do log**

Os outros campos do log não entram diretamente na mineração de processos, mas podem ser importantes na mineração de regras e para verificar a conformidade ou identificar características dos outros atributos.

No caso estudado, foi executada uma mineração de regra que levava em conta a temperatura $\geq 300^\circ\text{C}$, mas no campo ObjectId não existia esta informação, ela estava contida no campo descrição. No entanto existia no campo Activity um valor “mensagem 143”, cujo significado era informar que a temperatura atinge valor inferior a 300°C . Por este motivo o campo descrição foi descartado, mas poderia ser utilizado, se não houvesse outra forma de encontrar a informação desejada na regra.

5.3. Terceira Fase – Mineração de Processos.

O objetivo desta fase é gerar o modelo do processo e a separação das variações das instâncias. Tem como entrada o log pré-processado e como saída o processo minerado. Para isso, foi utilizada a ferramenta de mineração Disco (Fluxicon, 2013), e após a identificação dos atributos necessários para a mineração, foi executada a importação do arquivo de log e a mineração dos processos. O resultado é um modelo de processo que representa todas as instâncias como mostrado na Figura 5-20.

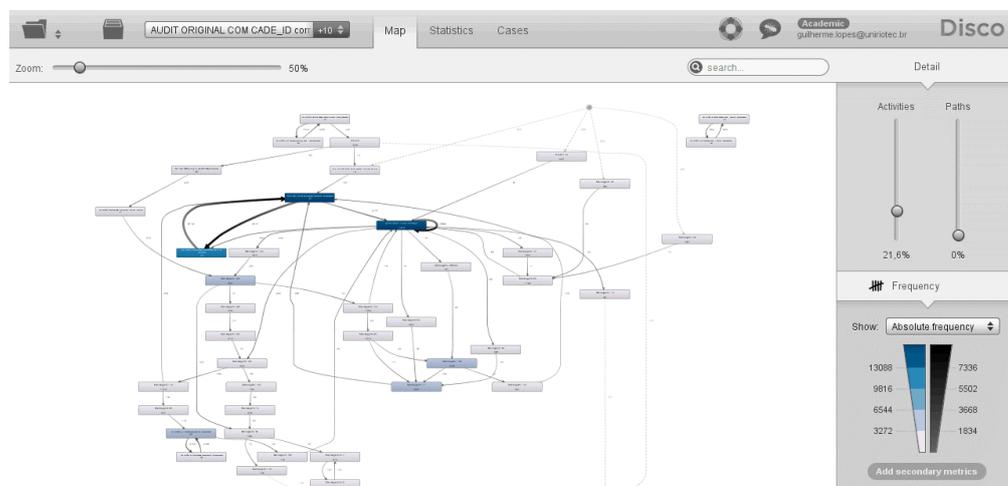


Figura 5-20 - Modelo de processo minerado

No lado esquerdo temos dois controles Activity e Path, estes controles são responsáveis pelo controle do nível de abstração que será mostrado na tela. O resultado da mineração de processos está resumido na Figura 5-21. Foram minerados 92047 eventos, em 629 instâncias, com 185 atividades diferentes e 29 recursos.

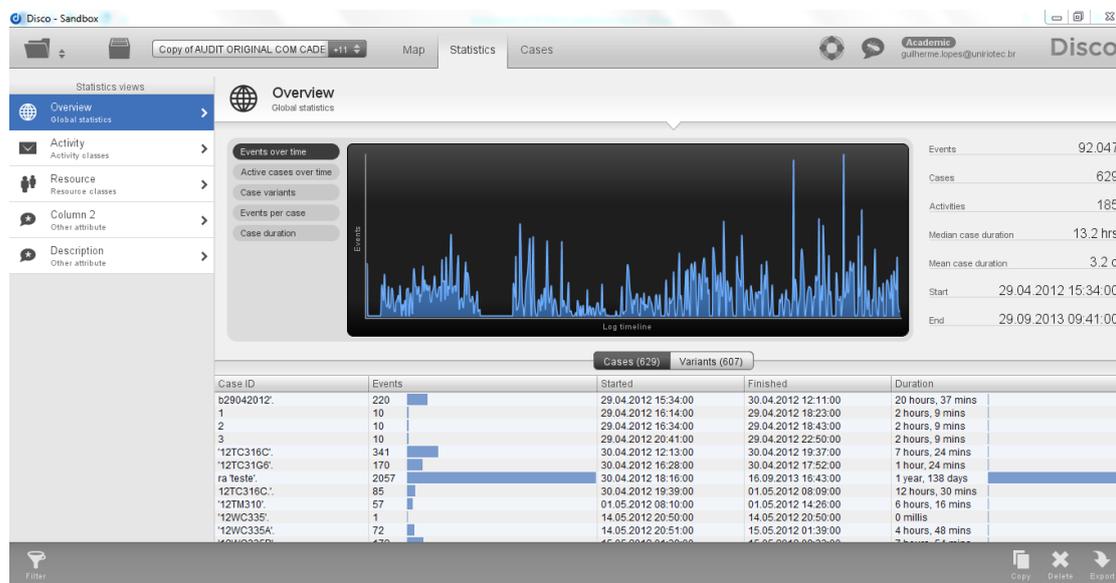


Figura 5-21 - Resultado inicial da mineração de processos

5.4. Quarta Fase – Verificação de conformidade.

A verificação de conformidade é feita comparando a instância padrão com as demais instâncias do processo, uma a uma utilizando as informações obtidas no formulário de informações do domínio.

5.4.1. Identificação das instâncias conformes

As instâncias conformes são agrupadas na Variante onde se encontra a instância padrão, portanto a classificação da instância conforme é direta, não depende de nenhum filtro. Todas as instâncias contidas na Variante são classificadas como conformes.

5.4.2. Identificação de instâncias com desvios completas e incompletas

O analista deve identificar a Variante que contém a instância padrão, pois nesta instância está a lista das atividades que precisam acontecer em todas as

instâncias. Para obter as instâncias que executam todas as atividades do processo a priori, o analista deve aplicar um filtro com a seguinte regra

Se a atividade Administração do utilizador

E

Variável: A ED_M000_batch_number1 (Número do lote)

E

Variável: M005_ta_a31(tecla liga o exaustor e a circulação de ar pressionada)

E

Variável: M005_ta_a11(tecla liga o aquecimento do túnel pressionada)

E

Variável: M005_ta_a01(tecla ligar o acionamento da esteira do túnel pressionada)

E

Variável: M030_int_f_nr_neu(Número do Ciclo)

E

Variável: M030_Cicle_Selection(Botão de confirmação de seleção de ciclo)

E

Variável: M098_message_updating(Botão de confirmação de atualização de data-hora pressionado)

E

Variável: M005_ta_a10(tecla Desliga o aquecimento do túnel pressionada)

E

Variável: M005_ta_a00(tecla desliga o acionamento da esteira do túnel pressionada)

estiverem contidas na instância e a variante onde está a instância padrão

Então instância é classificada como desvio

Senão a instância é classificada como incompleta.

O arquivo de log do estudo de caso corresponde a uma fase de implantação do equipamento, portanto existem instâncias de teste e algumas realizadas pelos técnicos do fabricante. Segundo o especialista, estas atividades deveriam ser eliminadas, pois não geraram produto. Estas instâncias foram identificadas através do Recurso que as executou. Assim, foram eliminadas todas as instâncias executadas por Sistema, Basin, Basin171, Mak, Man, teste e com valor em branco. Foi aplicado um filtro de atributo selecionando o atributo Resource e utilizando o filtro Keep select que elimina todos os eventos que foram executados pelos Recursos que não estiverem marcados.

Arquivos de log do mundo real nem sempre estão preparados para a mineração e podem ter ruídos e não informação relevante, por isto a interação com o especialista do domínio é fundamental para auxiliar na eliminação destes ruídos.

O resultado da aplicação do filtro é um log sem os eventos executados por estes Recursos. Após a aplicação do filtro, o log passou a ter 81243 eventos, 627 instâncias, 150 atividades e 22 Recursos.

Eliminados os eventos indesejáveis segundo o especialista do domínio e de posse do log limpo, foi possível passar para a próxima fase. A partir da entrevista e dos documentos fornecidos, as atividades que precisam acontecer para que a instância seja considerada completa estão listadas a seguir:

Administração do utilizador

Variável: A ED_M000_batch_number1 (Número do lote)

Variável: M005_ta_a31(tecla liga o exaustor e a circulação de ar pressionada)

Variável: M005_ta_a11(tecla liga o aquecimento do túnel pressionada)

Variável: M005_ta_a01(tecla ligar o acionamento da esteira do túnel pressionada)

Variável: M030_int_f_nr_neu(Número do Ciclo)

Variável: M030_Cicle_Selection(Botão de confirmação de seleção de ciclo)

Variável: M098_message_updating(Botão de confirmação de atualização de data-hora pressionado)

Variável: M005_ta_a10(tecla Desliga o aquecimento do túnel pressionada)

Variável: M005_ta_a00(tecla desliga o acionamento da esteira do túnel pressionada)

Para descobrir as instâncias completas é necessário aplicar um filtro de atributo, com o método Mandatory, este filtro elimina todas as instâncias que não tiverem pelo menos uma das atividades selecionadas. O primeiro impulso é marcar todas as atividades de uma só vez, no entanto o resultado não é satisfatório, pois o

filtro tem como critério se existe algum, portanto mesmo se achar uma correspondência verdadeira de qualquer atividade ele considera a instância, mesmo que as outras atividades não ocorram.

Esta forma de aplicar o filtro não atende à regra para encontrar instâncias completas, pois não considera Se todas as atividades, mas **Se existe uma** atividade.

Após aplicar o filtro, é gerado um novo modelo de processo. Analisando a Variante 4, observou-se que as duas instâncias 69 e 200 não têm todas as atividades, o que demonstra que o filtro não foi capaz de descobrir apenas as instâncias completas. Conforme mostrado na Figura 5-22.

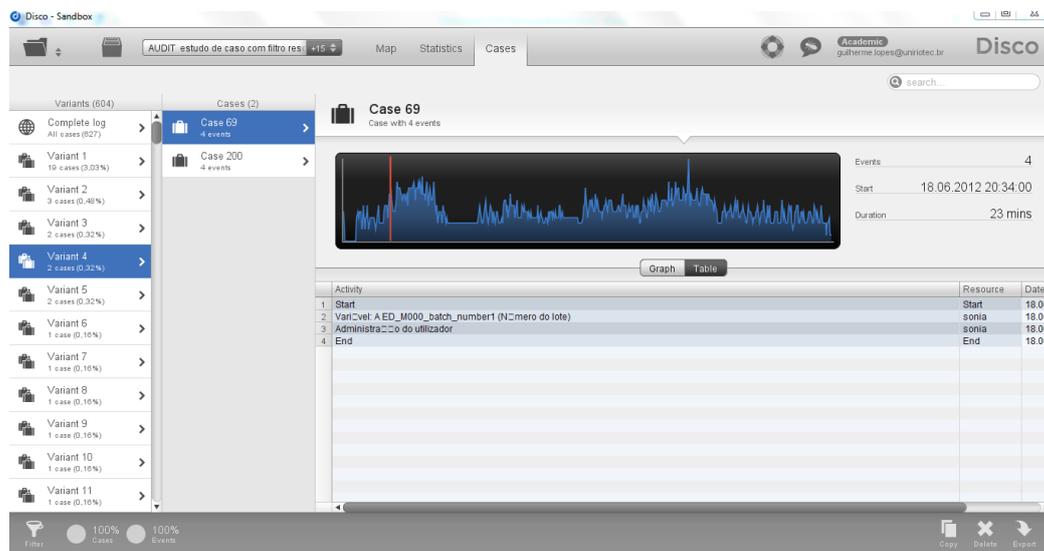


Figura 5-22 - Instâncias sem as atividades obrigatórias

Para cumprir a regra de descoberta das instâncias completas é preciso garantir que todas as atividades obrigatórias serão executadas, portanto com o filtro de atributos com a opção Mandatory tem como critério se existe pelo menos 1 valor que atende à regra, o correto é fazer filtros encadeados. No Apêndice 8.2, é mostrada a sequência de montagem do filtro passo a passo. No entanto, quando o filtro é aplicado, nota-se um erro, a mensagem diz que o filtro não pode ser aplicado, pois irá gerar um log em branco. Isto significaria que nenhuma instância seria completa inclusive a instância a priori.

Verificando o log, na aplicação do filtro de atributo para retirada dos Recursos, o filtro retirou inclusive instâncias válidas. O especialista informou que além de atividades de teste, estes Recursos ministraram treinamento e produziram produtos, portanto as atividades executadas não podiam ser todas consideradas como ruído. Para obter resultados válidos, o log original sem a aplicação do filtro de Resources foi

importado novamente e a montagem do filtro de atributo foi refeita, conforme mostrado no Apêndice 8.7.

5.4.3. Identificação das instâncias com desvios Regulatórios com atividades proibidas.

Esta etapa tem como objetivo descobrir se as instâncias com desvio executam alguma atividade proibida pela regulamentação. As atividades proibidas foram levantadas no formulário de informações sobre o domínio. A entrada é o log com todas as instâncias e a saída é o log com as instâncias que não atendem a regulamentação.

Neste estudo de caso foi levantado através do Formulário de informações sobre o domínio que a temperatura na fase de aquecimento do túnel deve ser $\geq 340^\circ \text{C}$, esta informação não consta dos atributos minerados, na realidade esta informação está no campo descrição, no entanto quando há um problema de temperatura abaixo de 340°C o equipamento gera um alarme com a mensagem 142 que consta do campo atividade, com a qual foi feita então a associação:

SE o atributo activity recebe o valor Mensagem 142

Então a instância é classificada como desvio regulatório

Com base nesta regra o log foi minerado a procura das instâncias que atendiam a esta regra, ou seja, que tivessem um activity com o valor Mensagem 142.

Após selecionar o tipo de filtro, seleciona-se o atributo que vai ser filtrado, neste caso Activity, seleciona-se qual a atividade que vai ser filtrada, e escolhe-se o modo de filtro, utilizou-se o Mandatory que elimina todas as instâncias que não contenham pelo menos uma das atividades selecionadas.

O resultado da aplicação do filtro são as instâncias que tem atividades proibidas, ou seja, Desvios Regulatórios como mostrado no filtro. Se houvesse mais de uma atividade para filtrar seria possível selecionar vários valores em um só filtro, pois o critério é se pelo menos um valor existir então é Desvio Regulatório. A Figura 5-23 mostra as instâncias resultantes após a aplicação do filtro.

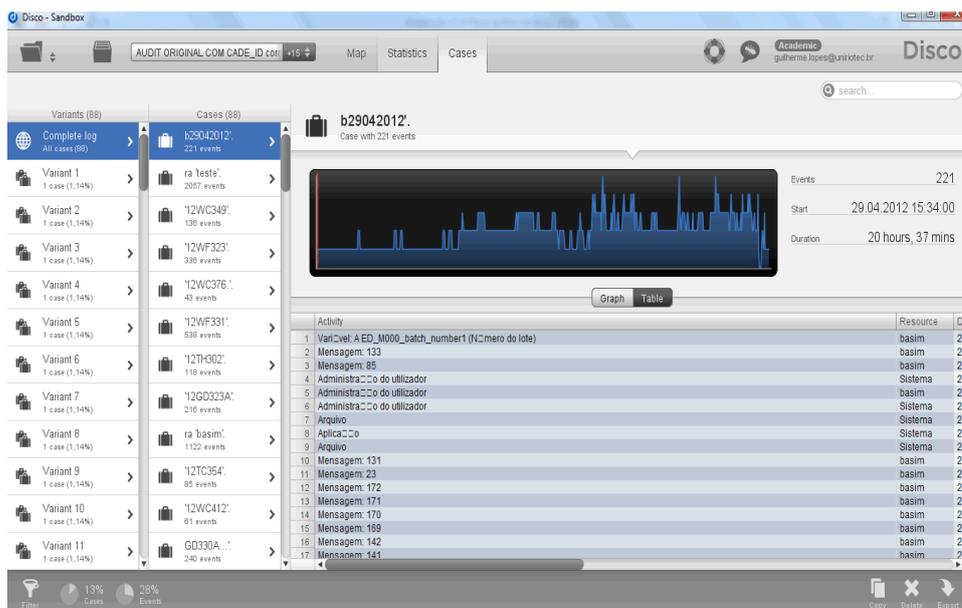


Figura 5-23 - Instâncias resultantes após aplicação do filtro

Os resultados mostraram que 89 instâncias foram classificadas como Desvio Regulatório (Figura 5-24). O produto fabricado por estas instâncias maior probabilidade de terem desvios regulatórios, como a severidade é alta e a probabilidade de ocorrência é alta, a classe de risco destes produtos será alta, conforme prescrito na RDC 17 (Agência Nacional de Vigilância Sanitária, 2010) o especialista deve tomar uma ação para mitigar o risco.

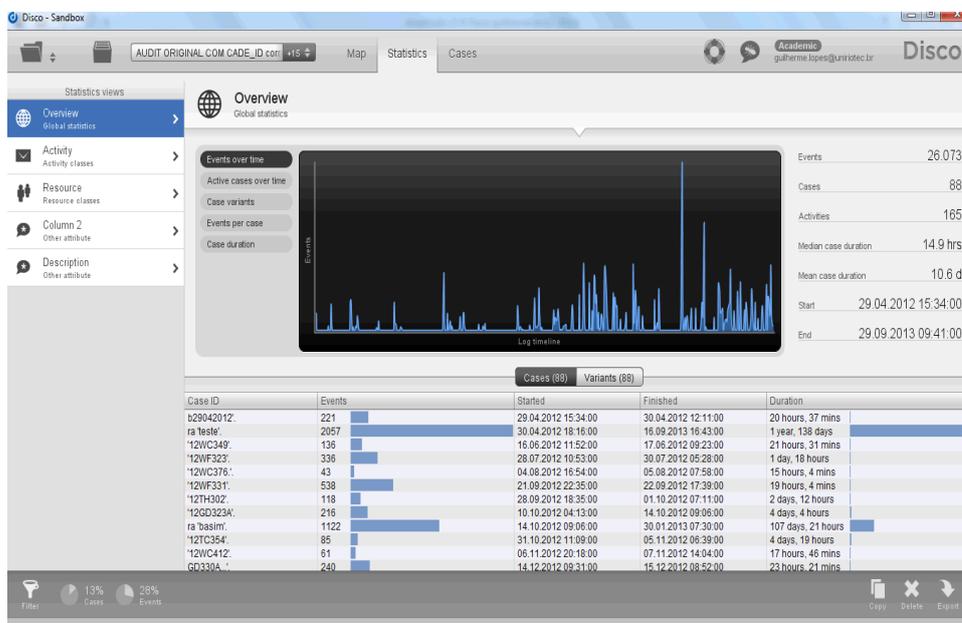


Figura 5-24 - Estatísticas após a aplicação do filtro

O resultado da mineração pode ser exportado em um arquivo de log que contenha apenas as instâncias com Desvio Regulatório, o que vai facilitar futuras análises pelo especialista do domínio.

5.5. Discussão sobre o estudo de caso

O estudo de caso demonstra que o método se seguido passo a passo consegue descobrir, classificar e separar as instâncias do processo.

Na primeira fase 5.2.1, foram levantadas as informações sobre o domínio e o processo. A entrevista com o especialista, apoiada com o formulário de informações sobre o domínio fornece uma visão rica do processo, descrevendo-o de forma que o analista possa entendê-lo, ficam também claros os regulamentos que devem ser seguidos, as restrições de atividades e valores de variáveis. Neste estudo de caso, também foi fornecido um manual de procedimento para o operador, o que ajuda em muito a verificação das informações passadas pelo especialista. É possível, analisando os dois documentos, verificar se não ficou algum requisito faltando ou regulamento.

Foi possível criar uma instância a priori baseada no manual e na entrevista com linguagem natural, que posteriormente foi confrontada com o log de eventos. Para cada atividade ou grupo de atividades da instância a priori foi associada uma atividade oriunda do sistema no log de eventos, com isto foi possível criar uma instância a priori, que tem correspondência no log, e portanto pode ser comparada com as outras instâncias do processo.

Concluindo, a primeira fase atingiu o seu objetivo que era descobrir ou gerar a instância a priori para servir de comparação para a verificação de conformidade.

Na segunda fase 5.2.2, foram levantados os atributos obrigatórios para realizar a mineração de processos, o arquivo de log tem o nome dos campos provenientes do sistema, portanto é preciso informar à ferramenta de mineração que campo do log corresponde a que atributo.

O primeiro atributo identificado foi o `case_id`, pois este atributo define as instâncias do processo. No estudo de caso, o log original não possuía nenhum campo que pudesse representar o `case_id`, esta informação estava contida dentro da descrição da atividade Figura 5-17, foi necessário criar um novo campo no log com o valor do

case_id, foi elaborada uma fórmula para procurar a string dentro da descrição, aplicada a fórmula, aprovada pelo especialista, o log pré processado passou a ter o case_id, conforme o prescrito no método.

O segundo atributo identificado foi o timestamp, neste estudo de caso o campo do log já se chamava timestamp, portanto já estava identificado, o único cuidado com este campo foi verificar se ele estava ordenado em ordem crescente e se todos os registros tinham o mesmo formato. No momento da associação com o atributo da ferramenta, foi feita a associação do formato da data com o modelo existente. Esta etapa também cumpriu o seu objetivo.

O terceiro atributo identificado foi a Activity, atividade que o processo executa, no log de eventos o campo tinha o nome de Object_id e foi associado ao campo Activity. Esta etapa também cumpriu o seu objetivo.

O quarto atributo identificado foi o recurso identificado no log pelo campo UserId, este campo foi associado ao atributo Resource, não foi necessária nenhuma outra configuração. Esta etapa também cumpriu o seu objetivo.

Apesar do arquivo de log possuir outros campos, eles foram marcados com outros atributos, pois não foram associados a nenhum dos atributos anteriores, e neste estudo de caso não existiam regras que buscassem informações nestes campos. Os outros atributos são utilizados quando as informações das regras a serem mineradas estão contidas nestes campos, portanto quando se aplica uma regra de atributo, o atributo escolhido para que seja aplicado o filtro resultante da regra é o de outros atributos.

Na terceira fase 5.3, foi realizada a mineração de processos propriamente dita, depois de associados os campos dos logs com os atributos, e, através das estatísticas geradas, foi possível observar que todo o log foi importado, e que não houve erros no processo de mineração. No caso de erro na importação, nenhuma instância do log será importada, neste caso deve-se verificar o arquivo em busca de linhas com erro, estas linhas devem ser reparadas ou eliminadas de acordo com orientação do especialista

Na quarta fase, foi feita a verificação e classificação das não conformidades, e para cada filtro aplicado foi gerado um novo log contendo apenas aquelas instâncias classificadas.

A primeira classificação é a mais simples é a das instâncias completas, a ferramenta de mineração agrupa as instâncias por semelhança em variantes, identificada a variante que contém a instância a priori, todas as instâncias contidas nesta variante estarão conformes.

A segunda classificação, o especialista informou que alguns recursos não faziam parte dos quadros da empresa, eram técnicos do fabricante e que deveriam ser eliminados do log. Através de um filtro de recursos, foram eliminadas todas as atividades executadas por estes recursos, no entanto quando este log foi minerado para verificar a existência de instâncias completas, ocorreu um resultado de erro, o log ficou vazio, indicando que estes recursos executaram instâncias completas, e retirando as atividades do log ele não representa mais o processo. Analisando o problema novamente o especialista e o encarregado de produção foram consultados, e eles informaram que alguns lotes de medicamentos foram produzidos utilizando a senha dos técnicos que acompanharam a produção por alguns meses. Este problema demonstra que o especialista tem uma forte influência sobre o resultado da mineração, e este procedimento de eliminação de informações do log deve ser feito com muita prudência e devidamente analisado e documentado. As atividades foram restabelecidas no log para que o trabalho pudesse continuar.

A segunda etapa desta fase é a identificação das instâncias com desvio que estão completas 5.4.2, e gerar um log de eventos somente com estas instâncias. Para realizar esta etapa, foi aplicado um filtro de atributos.

Nesta etapa deve-se ter muito cuidado com o falso positivo, pois a tendência é utilizar o filtro de atributo marcando todas as atividades em um único filtro, este tipo de configuração considera que se qualquer um dos valores for encontrado a instância é considerada completa, quando na realidade queremos uma regra do tipo se todos os valores acontecerem a instância é considerada completa.

Existiria um falso positivo, pois se apenas uma das atividades estivesse contida na instância, ela seria classificada com completa o que não é verdade, a forma correta de aplicação é mostrada em 8.7.

A última etapa executada foi a identificação das instâncias com desvio regulatório, esta identificação utiliza a resposta da pergunta: Quais são as atividades críticas em relação a regulamentação que não podem acontecer?

A partir desta resposta é elaborada a regra para o filtro, genericamente:

SE a instâncias contiver pelo menos uma atividade proibida
ENTÃO é classificada como não conforme.

Neste caso podemos utilizar o filtro que representa um OU como mostrado em 8.4. Nesta etapa o que pode ocorrer é o filtro não achar nenhuma instância que atenda a regra, neste caso não haverá nenhuma instância com desvio regulatório.

Todos os resultados de filtro podem ser exportados gerando logs apenas com as instâncias que correspondem àquela classificação.

Analisando o resultado, verifica-se que a hipótese foi cumprida, pois aplicando o método, foi possível classificar as instâncias conformes e não conformes, e as não conformidades foram classificadas como desvio regulatório, desvio não regulatório e instâncias incompletas.

O problema proposto que é identificar o conjunto de medicamentos que tem mais chance de estarem não conformes. Aplicando o método foi possível segregar e classificar as instâncias não conformes, com esta informação é possível fornecer ao especialista informações relevantes que diminuem a aleatoriedade das escolhas das amostras.

Neste cenário, o problema a ser tratado nesta pesquisa é: Como identificar dentre um conjunto de medicamentos fabricados os tem mais chance de ter problemas de qualidade para serem testados, sem que para isto tenha que se destruir todo o produto?

5.5.1. Resultados obtidos

Os resultados obtidos com o estudo de casos foram:

629 Instâncias mineradas

3 Instâncias conformes

25 Instâncias com desvios não regulatórios, 22variantes.

88 Instâncias com desvios regulatórios, 88 variantes.

513 Instâncias incompletas

Analisando os resultados pode-se estranhar a pouca quantidade de instâncias conformes, isto se deve ao fato que o processo estudado estava em fase de implantação, portanto em período de ajustes, este número é normal para esta fase.

As 25 instâncias com desvios não regulatórios estão agrupadas em 22 variantes, o que significa que pelos menos 3 instâncias repetiram, por se tratar de um defeito o fato de repetir ajuda no diagnóstico, e indica que o problema é recorrente.

As 88 instâncias com desvios regulatórios, estão dispostas em 88 variantes, o que indica que os desvios não se repetem, este grande número de desvios foi devido a um defeito no termostato. Através da análise das ordens de manutenção foi possível verificar a origem do problema que foi corrigido.

As 513 instâncias incompletas foram ocasionadas pelo fato do processo estar em implantação, portanto na fase de ajustes foram feitos inúmeros testes onde os lotes eram abortados para ajustes.

Os dados foram validados pelo especialista do domínio, e o resultado confirmado. As informações sobre o resultado dos testes do produto não estavam disponíveis quando do término desta dissertação e serão realizados em um estudo futuro.

6. Conclusões

A liberação de produtos farmacêuticos para venda passa por uma série de testes de controle de qualidade, no entanto é crescente o número de recolhimento de medicamentos por problemas no processo de fabricação, isto ocorre pois a quantidade produzida é cada vez maior, e os testes continuam sendo feitos por amostragem, devido ao custo e à característica destrutiva de determinados testes.

Na literatura observam-se várias técnicas de verificação de conformidade de processos, que na sua maioria, geram indicadores do grau de conformidade de uma instância do processo. Este tipo de abordagem é interessante, mas não é suficiente para

atender o problema da indústria farmacêutica onde qualquer índice diferente de 100% é considerado não conforme.

O principal critério para liberação de produtos na indústria de acordo com as boas práticas de fabricação (ISPE - International Society of Pharmaceutical Engineer, 2008) é a classe de risco e a detectabilidade do risco. A classe de risco mede a probabilidade de ocorrência do risco x severidade, e a detectabilidade do risco mede a chance de um produto não conforme ser liberado para a venda.

A classe de risco pode ser calculada utilizando agora a classificação das instâncias, cada classificação terá a sua própria classe de risco (baixa, média ou alta) e a detectabilidade do risco neste caso sempre passará de média (verificada por amostragem) para alta, pois pelo critério de classificação da detectabilidade do risco, todos os processos que são verificados 100% das instâncias tem detectabilidade alta.

O uso adequado destes novos indicadores pode auxiliar o especialista do domínio a diminuir a quantidade de produtos não conformes liberados ao mercado, melhorando a segurança do produto, diminuindo custos e melhorando a imagem da empresa.

6.1. Contribuições

A principal contribuição do trabalho é a sistematização da classificação das instâncias dos processos de fabricação de medicamentos. Como os produtos fabricados pelos processos de fabricação não podem ser testados 100% pois isto destruiria o produto, o teste de 100% do processo pode relacionar as não conformidades do processo com o risco da não conformidade do produto. De acordo com a pesquisa realizada com os especialistas do domínio, podemos relacionar a não conformidade do processo de fabricação com a não conformidade do produto. Sem o método é impossível testar 100% das instâncias, pois os testes tradicionais de controle de qualidade são destrutivos e portanto se fossem executados em todos os produtos fabricados, a empresa não teria produto para vender.

Os métodos de classificação tradicionais, como por exemplo os fitness não são capazes de classificar de forma correta os desvios com foco no risco da conformidade do produto gerado pelo processo de fabricação. A classificação tradicional se aplicada neste domínio induzirá ao erro, tendendo a classificar com risco menor de produzir produtos não conformes, as instâncias com maior índice de

conformidade , no entanto uma instância com 98 % de conformidade pode executar uma atividade proibida , portanto o risco do produto estar não conforme é enorme , e uma instância com 70% de conformidade não executar nenhuma atividade proibida e ser classificada como desvio não regulatório , com risco para a conformidade do produto muito menor.

Os resultados da classificação das não conformidades do processo de fabricação de medicamentos, é um indicativo de que instâncias do processo tem maior chance de ter fabricado produtos não conformes com a regulamentação, estes produtos têm mais chances de ter problemas de qualidade.

A escolha de produtos para teste de qualidade é feita de forma aleatória, com a classificação o especialista poderá fazer a escolha levando em consideração 4 tipos de instâncias:

- Conformes: fabricadas de acordo com as normas e procedimentos, executa todas as atividades obrigatórias e somente estas.
- Desvio Regulatório: fabricadas fora do padrão da regulamentação e executando atividades proibidas ou ferindo as restrições regulatórias.
- Desvio Não Regulatório: fabricadas executando todas as atividades obrigatórias, no entanto executam outras atividades que não ferem a regulamentação.
- Incompletas: fabricadas sem executar todas as atividades obrigatórias.

A classificação muda a perspectiva do especialista do domínio em relação à escolha das amostras para o controle de qualidade diminuindo a aleatoriedade da escolha. A escolha pode ser feita dentro de um grupo específico e de acordo com a classificação os testes de controle de qualidade podem ser mais ou menos elaborados.

Sem a classificação, a chance de escolha de um produto fabricado por uma instância conforme ou não conforme era de 50%, com a classificação a aleatoriedade foi eliminada, o especialista sabe exatamente que tipo de instância fabricou o produto e pode escolher de acordo com os critérios de proporcionalidade e risco quantas amostras tirar de cada grupo. Apesar do método classificar as instâncias em 4 conjuntos, a escolha dentro de cada conjunto continua sendo aleatório, portanto o controle estatístico do processo não é prejudicado, a grande diferença é que os tipos de desvio estão classificados.

A verificação e classificação de 100% das instâncias formam uma base de dados com resultados detalhados, auxiliando a análise de riscos realizada pelos especialistas para a liberação do produto, e evitando que produtos não conformes sejam liberados para a venda, pois com uma análise mais detalhada a detectabilidade do risco aumenta.

Outra contribuição é financeira, pois a classificação não destrói o produto e possibilita o teste de 100% das instâncias, gerando economia tanto no processo de controle de qualidade com a melhoria da acurácia e também para a imagem da empresa com a tendência à diminuição da liberação de produtos não conformes.

Como contribuição secundária, é possível relacionar os recursos humanos e equipamentos que fabricaram as instâncias conformes e não conformes, possibilitando ao especialista analisar a frequência da não conformidade relacionada com um determinado operador, equipamento, horário, auxiliando na identificação da causa raiz da não conformidade.

6.2. Restrições da pesquisa

O método foi baseado no domínio da indústria farmacêutica e como tal utiliza o modelo de regras deste domínio. É necessária a presença do especialista para fornecer as informações sobre o domínio.

Como o método não foi testado em outros domínios não é possível fazer uma generalização para qualquer domínio.

6.3. Trabalhos futuros

Para trabalhos futuros, aprimorar a coleta de dados do especialista automatizando o formulário de pesquisa, testar outros algoritmos de mineração de processos.

Como extensão do método, coletar as informações de como os especialistas usaram o resultado e verificar se diminuiu a quantidade de produtos enviados com problemas de conformidade para o mercado. Criar uma métrica de qual a distribuição de escolha das amostras por tipo de classificação.

Uma outra possibilidade é a aplicação deste método em outros domínios, existe uma grande demanda para este tipo de classificação na área de smart grid e da

internet das coisas. A segurança da informação de redes de infraestruturas inteligentes demanda uma abordagem de conformidade mais complexa para definição das ações automáticas de proteção aos ativos, evitando que uma não conformidade desligue, por exemplo, toda a eletricidade de uma região do país, a classificação de não conformidades do processo pode criar níveis de atuação evitando grandes transtornos.

Outra área que pode ser beneficiada com a adaptação do método é a área de cidades inteligentes, por exemplo no entendimento do processo de mobilidade urbana, a construção de modelos de deslocamento, e a análise de anomalias do deslocamento em relação ao modelo, podendo classifica-las em caso de acidentes, eventos, protestos, contribuindo para o planejamento diário e gestão de crises.

7. Referências bibliográficas

van der Aalst, W. M., van Hee, K. M., van der Werf, J. M. & Verdonk, M., 2010. Auditing 2.0: Using Process Mining to Support Tomorrow's Auditor. *IT SYSTEMS PERSPECTIVES - IEEE Computer Society*, Março, pp. 90-93.

Aalst, W. M. P. v. d., 2011. *Process Mining. Discovery, Conformance and Enhancement of Business Process*. Berlin: Springer.

Aalst, W. v. d., 2011. Process Discovery. Em: *Process Mining , discovery , conformance and enhancement of business processes*. Eindhoven: Springer.

Aalst, W. v. d., 2011. Process Discovery. In: *Process Mining , discovery , conformance and enhancement of business processes*. Eindhoven: Springer, p. Capítulo 5.

Aalst, W. v. d., 2014. *Process Mining Data science in Action*, Eindhoven: TU/E.

Aalst, W. v. d., 2014. *Process Mining: Data science in Action*, Eindhoven: TU/E.

Aalst, W. v. d., 2014. *Process Mining: Data science in Action* , Eindhoven: TU/E.

Aalst, W. v. d., 2014. *Process Mining: Data science in Action Lecture 4.5*, Eindhoven: TU/E.

- Aalst, W. v. d., 2014. *Process Mining: Data science in Action Lecture 4.7*, Eindhoven: TU/E.
- Aalst, W. v. d. & Sthål, C., 2011. *Modelin Business Process - A Petri Net-Oriented Approach*. Cambridge, Massachusetts: The MIT Press.
- Agência Nacional de Vigilância Sanitária, 2010. *RESOLUÇÃO - RDC Nº 17, DE 16/04/10*, Brasília: ANVISA.
- Agencia Nacional de Vigilância Sanitária, 2010. *Guia de validação de sistemas computadorizados*. Brasília, s.n.
- AICPA, 2006. *A Guide To The Sarbanes-Oxley Act*. [Online] Available at: <http://www.soxlaw.com/> [Accessed 10 jan 2014].
- Anon., 1975. *Ludwig. Teoria geral dos sistemas*. s.l.:Vozes.
- ANSM - Agence Nationale de sécurité du médicament et des produits de santé, 2013. *I Annexe 11: Systèmes informatisés*, Paris: ANSM.
- ARIS, 2015. *Workflow Management within the ARIS Framework*. [Online] Available at: http://www.pera.net/Methodologies/ARIS/ARIS_Paper_by_Ted_Williams.html [Acesso em 12 2015].
- Belohlavek, P., s.d. *UNICIST*. [Online] Available at: http://www.unicist-library.org/br/unicistwiki_library/index.php?title=Intelig%C3%A7%C3%A3o_Antica_Unicista_da_Natureza [Acesso em 19 05 2013].
- C.W. Günther, W. v. d. A., 2007. *Fuzzy Mining – Adaptive Process Simplification Based on Multi-perspective Metrics*. Brisbane, Australia, Springer, pp. 328-343.
- Davenport, T. H., 1993. *Process Innovation : reengineering work through information technology*. New York: Harvard Business School Press.
- DOU, 2014. *RESOLUÇÃO - , DE 19 DE AGOSTO DE 2014. Diário Oficial da União*, 20 08, p. 49.
- Dumas, M., La Rosa, M., Mendling, J. & A. Reijers, H., 2013. *Fundamentals of Business Process Management*. Berlin: Springer Heidelberg.
- Dumas, M., Rosa, M. L., Mendling, J. & Reijers, H. A., 2013. *Fundamentals of Business Process Management*. Verlag Berlin Heidelberg: Springer.
- European Medicine Agency, 2014. *European Medicine Agency*. [Online] Available at: http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/document_listing/document_listing_000162.jsp&mid=WC0b01ac058076ed73
- F.M. Maggi, M. W. M. v. d. A., 2011. *Runtime Verification of LTL-Based Declarative Process Models*, Eindhoven: s.n.
- FDA, 2002. *General Principles of Software Validation, Guidance for Industry and FDA staff*, Rockville: Food and Drug Administration.
- Fluxicon, 2013. *Fluxicom*. [Online] Available at: <http://www.fluxicon.com/disco/>

GE FANUC, n.d. *GE INTELLIGENT PLATAFORMS*. [Online] Available at: <http://www.ge-ip.com/products/proficy-hmi-scada-ifix/p3311> [Accessed 5 JAN 2014].

Han, J. & Kamber, M., 2000. *Data Mining Concepts and Technics*. Burnaby: Morgan Kaufmann Publishers.

ISPE - International Society of Pharmaceutical Engineer, 2008. *GAMP 5 - A RISC BASE APPROACH TO COMPLIANT GXP COMPUTERIZED SYSTEM*. s.l.:ISPE.

ISPE - Risk Based, 2010. *Risk-Based Manufacture of Pharmaceutical Products*. s.l.:ISPE BASELINE.

ISPE, 2014. *ISPE*. [Online] Available at: http://www.ispe.org.br/ispe_about/

Laudon, K. C. & Laudon, J. P., 2012. *Management Information System*. New York: Prentice Hall.

Lima, A. F., 2015. *Pesquisa de Opinião Pública*. [Online] Available at: <http://www.ebah.com.br/content/ABAAAqpmAA/apostila-pesquisa-opiniao-publica?part=3> [Acesso em 06 02 2015].

Mecatrônica Atual, 2013. *Mecatrônica Atual*. [Online] Available at: <http://www.mecatronicaatual.com.br/educacao/1460-mes-sistemas-de-execuo-da-manufatura-manufacturing-execution-system> [Acesso em 07 02 2015].

Presidência da República do Brasil, 2000. *Site da Presidência da República do Brasil*. [Online] Available at: http://www.planalto.gov.br/ccivil_03/leis/19986.htm [Acesso em 01 08 2013].

Presidência da República, 2000. *Lei 9986*. [Online] Available at: http://www.planalto.gov.br/ccivil_03/leis/19986.htm [Acesso em 10 01 2014].

Process Mining Group, TU/E, 2013. *Process Mining*. [Online] Available at: <http://www.promtools.org/prom6/prom63.html>

Process Mining Task Force, 2014. *Process mining task force*. [Online] Available at: [http://www.processmining.org/online/conformanceevaluation?s\[\]=l&s\[\]=checker](http://www.processmining.org/online/conformanceevaluation?s[]=l&s[]=checker) [Acesso em 21-10 2014 2014].

R.S. Mans, M. S. M. S. W. v. d. A., 2008. *Application of Process Mining in Healthcare – A Case Study in a Dutch Hospital*. Berlin Heidelberg, Springer-Verlag, p. 425 a 438.

Recker, J., 2013. *Scientific Research in Information Systems*. Queensland: Springer.

Reginato, R., 2007. *Sistemas SCADA e Sistemas Supervisórios*. [Online] Available at: http://www.foz.unioeste.br/~romeu/CIP/2_Aula_scada.pdf [Acesso em 6 2 2015].

Rozinat, A., 2010. *Process Mining Conformance and extention*. Eindhoven: TU/E.

Rozinat, A. & Aalst, W. M. v. d., 2008. Conformance checking of processes based on monitoring real behavior. pp. 64-95.

Santoro, F. M., 2014. *Pesquisa Método 2013*. Rio de Janeiro(RJ): s.n.

Santos, R. C. d., 2014. *UM MODELO DE APOIO À TOMADA DE DECISÃO PARA ANÁLISE DE SIMILARIDADE E UNIFICAÇÃO DE PROCESSOS DE NEGÓCIO*, Rio de Janeiro: Copper - UFRJ.

SAP, 2015. *SAP BRASIL*. [Online] Available at: <http://www.sap.com/brazil/solution/lob/finance.html> [Acesso em 12 2015].

Sarbanes, P. & Oxley, M., 2002. *SARBANES-OXLEY*. [Online] Available at: <http://www.soxlaw.com/> [Acesso em 01 08 2013].

Song, M., 2012. *IEEE CIS Task Force on Process Mining*. [Online] Available at: http://www.win.tue.nl/ieeetfpm/lib/exe/fetch.php?media=:casestudies:unist_samsung_case_study.pdf [Acesso em 06 10 2014].

UNISIST, 2012. *IEEE Task Force on Process Mining*. [Online] Available at: http://www.win.tue.nl/ieeetfpm/doku.php?id=casestudies:process_mining_at_samsung_electro-mechanics

UNIST - SNUBH, 2012. *Process Mining at Seoul National University Bundang Hospital*. [Online] Available at: http://www.win.tue.nl/ieeetfpm/doku.php?id=casestudies:process_mining_at_seoul_national_university_bundang_hospital

van der Aalst, W. A. A. d. M. A. K. A. A. F. B. T. B. T. .. & P. L. (. J., 2012. *Process mining manifesto*. s.l., Springer Berlin Heidelberg., pp. (pp. 169-194)..

Weigand, H. & Bukhsh, F. A., 2013. *Smart Auditing - Innovating Compliance Checking*. Viena, IEEE.

Wil M.P. van der Aalst, K. M. v. H. ., J. M. v. d. W. V., 2010. Auditing 2.0: Using Process Mining to Support Tomorrow's Auditor. *Computing Now*, pp. 90-93.

8. Apêndices

8.1. Pesquisa sobre qualidade na fabricação de medicamentos.

A pesquisa sobre a qualidade de fabricação de medicamentos, foi realizada utilizando a ferramenta de formulário do Google que registra a data e hora de cada resposta, a planilha resultante é protegida contra gravação. O formulário está publicado no endereço:

https://docs.google.com/forms/d/1huNuQ_WQaq6dPLryXRY7AjQMJtel_ME40ZKkLBSDGdA/viewform?usp=send_form.

Tabela 8-1 - Pesquisa sobre qualidade de fabricação de medicamentos

Pesquisa sobre a qualidade na fabricação de medicamentos

Página 1 de 2

Pesquisa sobre a qualidade na fabricação de medicamentos

Este formulário é parte da pesquisa de Mestrado de Guilherme Neves Lopes .

*Obrigatório

1. **A qualidade do produto fabricado está diretamente relacionada com a conformidade do processo de fabricação? ***

Considere que a matéria prima foi testada e aprovada.
Marcar apenas uma oval.

	1	2	3	4	5	
Discordo	<input type="radio"/>	Concordo Totalmente				

2. **Qual o seu nível funcional na empresa ? ***

Responda o que mais se aproxima.
Marcar apenas uma oval.

- Operador
 Encarregado
 Supervisor
 Gerente
 Diretor

3. **Qual a influência da qualidade da matéria prima na qualidade final do medicamento ?***

1 - nenhuma influência , 2 - pouca influência , 3 - média influência , 4 - grande influência , 5 - muito grande influência
Marcar apenas uma oval.

	1	2	3	4	5	
Nenhuma	<input type="radio"/>	Muito grande				

4. **Cadastre o seu email para receber o resultado da pesquisa.**

(opcional)

.....

https://docs.google.com/forms/d/1huNuQ_WQaq6dPLryXRY7AjQMJtel_ME40ZKk... 06/01/2015

5. **Qual a influência da conformidade do processo de fabricação na qualidade do produto ?***

Considere o processo de fabricação validado . 1 - nenhuma influência , 2 - pouca influência , 3 - média influência , 4 - grande influência , 5 - muito grande influência
 Marcar apenas uma oval.

1 2 3 4 5

Nenhuma Muito grande

6. **Há quantos anos você trabalha na fabricação de medicamentos ?***

Marcar apenas uma oval.

1 a 3 anos

3 a 5 anos

5 a 10 anos

mais de 10 anos

8.2. Exemplos de preenchimento do formulário de informações sobre o domínio

Tabela 8-2 Formulário de informações sobre o domínio equipamento xyz

Especialista do Domínio:	ANALISTA KXX
E-mail	xxxxx@xxxx
Telefone	21 2121 2121
Data e Hora:	21/11/2014 15:43
Cargo:	Supervisor de Produção
Função (descrição da função em relação ao domínio):	Gestor do departamento de produção, atender o plano de produção, manter o treinamento dos funcionários, reduzir, manter a área com os requerimentos de inspeção.
Empresa	Pharma y
Pesquisador (nome do responsável por realizar a entrevista.)	GUILHERME NEVES
Processo estudado (nome do processo)	Equipamento xyz
Objetivos do Processo:	Esterilização de materiais utilizados na produção, partes da máquina de enchimento e recravação.
Descrição do Processo:	O processo é dividido em dois tipos de ciclos hpv1 para materiais não porosos e hpv2 para materiais porosos.

	<p>No hpv1, o material é embalado, é verificado os parâmetros de carga padrão validada para cada tipo ou grupo de materiais, carregamento manual da autoclave e distribuição de acordo com a carga padrão, fechamento da autoclave via plc, seleciona carga, start. Emissão de 2 tipos de relatórios gerado pelo yok, report em forma de tabela.</p> <p>Obs. – sistema redundante de registro report e carta gráfica do yok.</p>
Qual o atributo define de forma unitária cada instância do processo.	Número da carga
Atributos (descreva a lista de atributos que devem ser considerados no log)	Número da carga, temperatura, pressão, TimeStamp, f0 (índice de letalidade da esterilização),
Formato do Log (texto, banco de dados, planilha, etc.)	Perguntar para ti
Qual atributo marca o início e fim do processo e como identifica-lo?	Botão de inicio Atividade de fim do processo.
Qual a norma ou regulamento se aplicam a este processo?	RDC 17
Quais são os parâmetros críticos do processo em relação à regulamentação?	Temperatura, pressão, F0, tempo do ciclo.
Quais são as atividades críticas em relação à regulamentação que não podem acontecer?	Temperatura abaixo ou acima e tempo mínimo e máximo.
Quais são as atividades críticas em relação à regulamentação que precisam acontecer?	<ol style="list-style-type: none"> 1. Embalagem de material 2. Carregamento manual 3. Arrumação do material dentro da autoclave. 4. Fechamento da porta. 5. Seleção da receita 6. Start do processo.
Critérios de decisão (conforme, desvio e regulatório)	F0, tempos de cada etapa e temperatura.

O objetivo desta fase é levantar as informações do domínio necessárias para a obtenção da instância a priori do processo, que vai ser usada como instância padrão para comparação com as outras instâncias.

As primeiras ações a serem executadas nesta fase é a entrevista com o especialista do domínio e a solicitação dos procedimentos operacionais, utilize as respostas da entrevista e os procedimentos para preencher formulário de informações sobre o domínio (Formulário 1) que tem seus campos listados abaixo:

Especialista do domínio: Neste campo preencha o nome do entrevistado.

Data e Hora: data e hora do início da entrevista.

Cargo: cargo do entrevistado.

Função: breve descrição das atividades que entrevistado realiza no domínio.

Empresa: nome da empresa do entrevistado.

Pesquisador: Nome do pesquisador responsável pelo preenchimento do formulário.

Processo estudado: o nome do processo estudado.

Objetivo do Processo: qual o resultado esperado deste processo.

Descrição do processo: descrição das etapas do processo.

Atributos: nome e exemplo dos atributos utilizados neste processo, exemplo: número do lote, temperatura, tamanho do frasco, ciclo do processo, pressão diferencial, fluxo laminar, etc.

Qual o atributo marca o início e fim do processo e qual o valor ele deve assumir? Nesta pergunta responda o nome do atributo e qual o valor que este atributo assume na atividade inicial do processo e qual o valor que o atributo assume na atividade final da instância do processo, e se são sempre as mesmas atividades e valores de início e fim de cada instância do processo.

Qual a norma ou regulamento se aplicam a este processo? Responda quais as regulamentações são aplicadas a este processo, exemplo RDC 17, RDC 58, CFR PART 11 e onde podem ser encontradas as regulamentações, site, livro, documentos internos.

Quais são os parâmetros críticos do processo em relação a regulamentação? Responda quais os valores críticos dos atributos das instâncias do processo. Escreva em forma de regras, por exemplo? Se o atributo atividade for igual

a erro 53 então valor proibido. Se o tempo da atividade aquecimento a 300 °c < 30 min então valor proibido, se a condutividade da água for > 3 então valor proibido, vá elaborando todas as regras que ferem os parâmetros exigidos pelas regulamentações.

Quais são as atividades críticas em relação à regulamentação que não podem acontecer? Responda quais as atividades ou sequência de atividades não podem acontecer, deve ser escrito como regras, exemplo:

- Se o atributo atividade receber o valor tampa do tanque aberta, acontecer depois da atividade início da fabricação então atividade proibida,

- Se o atributo atividade desliga o fluxo laminar acontecer depois de carregamento de frascos, então atividade proibida.

Escrever quantas regras forem necessárias até esgotar todas as restrições.

Quais são as atividades críticas em relação à regulamentação que precisam acontecer? Responda com a lista de atividades que precisam acontecer, caso exista uma sequência pré-determinada especificar a sequência, exemplo:

Ligar o equipamento.

Entrar com a senha do usuário.

Especificar o ciclo

Digitar o número do lote.

Ligar a esteira.

Ligar o aquecimento.

Ligar o resfriamento.

Ligar o fluxo laminar e depois ligar o aquecimento.

Escreva todas as atividades obrigatórias e as sequências obrigatórias

Critérios de decisão (conforme, desvio e regulatório)? Responda como classificar as instâncias, exemplo:

Instância conforme: toda a instância que contém as atividades da a priori e somente elas, e não atendem a nenhuma das regras de restrição das instâncias do processo.

Instância com desvio não regulatório: toda a instância que contém todas as atividades da instância a priori, mas contém outras atividades e não atende a nenhuma das regras de restrição do domínio.

Instância com desvio regulatório: qualquer instância que atenda uma das regras que são restrições regulatórias.

Preencha todas as definições e os critérios de classificação.

Tabela 8-3 Formulário de informações sobre o domínio sistema XYZ

Especialista do Domínio:	ANALISTA ZZZ
Data e Hora:	21/11/2014 14:42
Cargo:	Supervisor de Manutenção
Função (descrição da função em relação ao domínio):	Manutenção preventiva, corretiva, calibração, instalação de equipamentos, coordenar as atividades de manutenção do parque industrial e utilidade.
Empresa	PHARMA XPTO
Pesquisador (nome do responsável por realizar a entrevista.)	Guilherme neves
Processo estudado (nome do processo)	SISTEMA XYZ
Objetivos do Processo:	Fornecer vapor puro para a autoclave.
Descrição do Processo:	A partir do vapor saturado vindo da caldeira (produzido a partir de água potável) alimenta um trocador de calor, que aquece um tanque de vaporização da água purificada a fim de produzir vapor puro.
Atributos (descreva a lista de atributos que devem ser considerados no log)	(plc halen bradley, ihm pnelview 500.) Condutividade, temperatura, pressão, hora, nome do operador.
Formato do Log (texto, banco de dados, planilha, etc.)	Perguntar para ti.
Qual atributo marca o início e fim do processo e como identifica-lo?	Botão start para início e aborta para final.
Qual a norma ou regulamento se aplicam a este processo?	Verificar na documentação.
Quais são os parâmetros críticos do processo em relação à regulamentação?	Condutividade < 1, 3 Pressão 2 quilos. Máximo 2.5 quilos

Quais são as atividades críticas em relação à regulamentação que não podem acontecer?	Aumento da condutividade e baixo nível de água.
Quais são as atividades críticas em relação à regulamentação que precisam acontecer?	<ol style="list-style-type: none"> 1. Ligar caldeira de vapor saturado. 2. Pressão de vapor saturado. 3. Nível de água pura no sistema 4. Descarrega água parada 5. Inicia o processo automático. 6. Aquece até produzir vapor. 7. Controla a pressão ente 1,9 e 2.5 quilos de pressão. 8. Controla a válvula de vapor saturado. 9. Abort. 10. Restart.
Critérios de decisão (conforme, desvio e regulatório)	Valores de condutividade e pressão.

O objetivo desta fase é levantar as informações do domínio necessárias para a obtenção da instância a priori do processo, que vai ser usada como instância padrão para comparação com as outras instâncias.

As primeiras ações a serem executadas nesta fase é a entrevista com o especialista do domínio e a solicitação dos procedimentos operacionais, utilize as respostas da entrevista e os procedimentos para preencher formulário de informações sobre o domínio (Formulário 1) que tem seus campos listados abaixo:

Especialista do domínio: Neste campo preencha o nome do entrevistado.

Data e Hora: data e hora do início da entrevista.

Cargo: cargo do entrevistado.

Função: breve descrição das atividades que entrevistado realiza no domínio.

Empresa: nome da empresa do entrevistado.

Pesquisador: Nome do pesquisador responsável pelo preenchimento do formulário.

Processo estudado: o nome do processo estudado.

Objetivo do Processo: qual o resultado esperado deste processo.

Descrição do processo: descrição das etapas do processo.

Atributos: nome e exemplo dos atributos utilizados neste processo, e: número do lote, temperatura, tamanho do frasco, ciclo do processo, pressão diferencial, fluxo laminar, etc.

Qual o atributo marca o início e fim do processo e qual o valor ele deve assumir? Nesta pergunta responda o nome do atributo e qual o valor que este atributo assume na atividade inicial do processo e qual o valor que o atributo assume na atividade final da instância do processo, e se são sempre as mesmas atividades e valores de início e fim de cada instância do processo.

Qual a norma ou regulamento se aplicam a este processo? Responda quais as regulamentações são aplicadas a este processo, exemplo RDC 17, RDC 58, CFR PART 11 e onde podem ser encontradas as regulamentações, site, livro, documentos internos.

Quais são os parâmetros críticos do processo em relação à regulamentação? Responda quais os valores críticos dos atributos das instâncias do processo. Escreva em forma de regras por exemplo? Se o atributo atividade for igual a erro 53 então valor proibido. Se o tempo da atividade aquecimento a $300\text{ }^{\circ}\text{C} < 30\text{ min}$ então valor proibido, se a condutividade da água for > 3 então valor proibido, vá elaborando todas as regras que ferem os parâmetros exigidos pelas regulamentações.

Quais são as atividades críticas em relação à regulamentação que não podem acontecer? Responda quais as atividades ou sequência de atividades não podem acontecer, deve ser escrito como regras, exemplo:

- Se o atributo atividade receber o valor tampa do tanque aberta, acontecer depois da atividade início da fabricação então atividade proibida,

- Se o atributo atividade desliga o fluxo laminar acontecer depois de carregamento de frascos, então atividade proibida.

Escrever quantas regras forem necessárias até esgotar todas as restrições.

Quais são as atividades críticas em relação à regulamentação que precisam acontecer? Responda com a lista de atividades que precisam acontecer, caso exista uma sequência pré-determinada especificar a sequência, exemplo:

Ligar o equipamento.

Entrar com a senha do usuário.

Especificar o ciclo
Digitar o número do lote.
Ligar a esteira.
Ligar o aquecimento.
Ligar o resfriamento.
Ligar o fluxo laminar e depois ligar o aquecimento.

Escreva todas as atividades obrigatórias e as sequências obrigatórias

Critérios de decisão (conforme, desvio e regulatório)? Responda como classificar as instâncias, exemplo:

Instância conforme: toda a instância que contém as atividades da a priori e somente elas, e não atendem a nenhuma das regras de restrição das instâncias do processo.

Instância com desvio não regulatório: toda a instância que contém todas as atividades da instância a priori, mas contém outras atividades e não atende a nenhuma das regras de restrição do domínio.

Instância com desvio regulatório: qualquer instância que atenda uma das regras que são restrições regulatórias.

Preencha todas as definições e os critérios de classificação.

8.3. Exemplo de aplicação do filtro de atributos para descobrir instâncias completas

A partir de uma lista de valores obrigatórios, crie um filtro de atributo com a opção Mandatory, esta opção elimina as instâncias que não contém pelo menos um valor selecionado no filtro. Como na verificação de instâncias completos nós queremos que todas as atividades estejam presentes em todas as instâncias filtradas, não podemos criar um único filtro com todas as atividades senão quando uma condição for atendida o filtro vai considerar a instância e o que queremos é que todas as atividades estejam presentes em cada instância resultante.

Comece entrando na ferramenta Disco, siga o passo a passo abaixo:

Clique no botão Windows e depois em todos os programas.

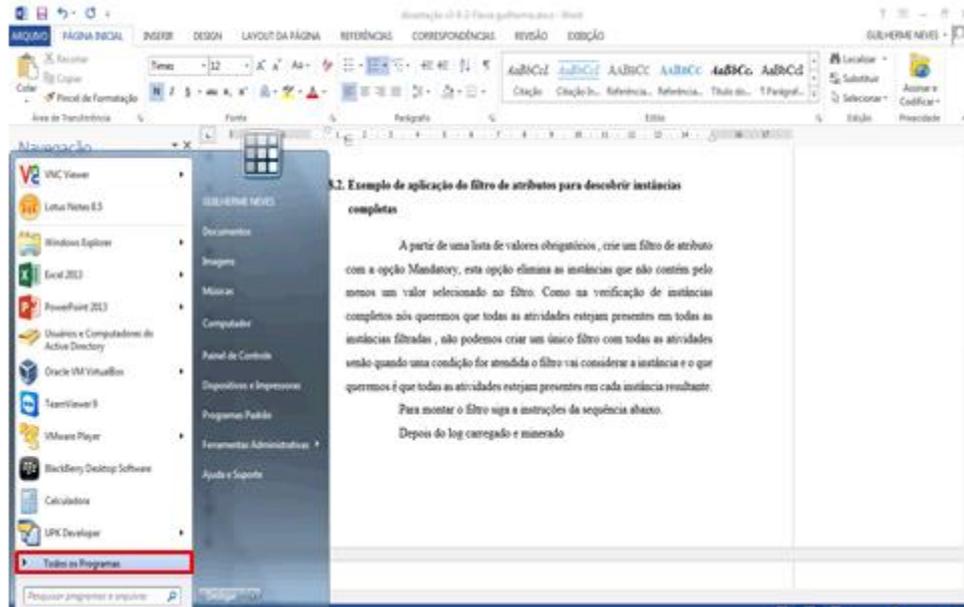


Figura 8-1 - Tela todos os programas

Clique na pasta Disco.



Figura 8-2 - Tela com a pasta de instalação do disco

Dê um duplo clique no aplicativo Disco.

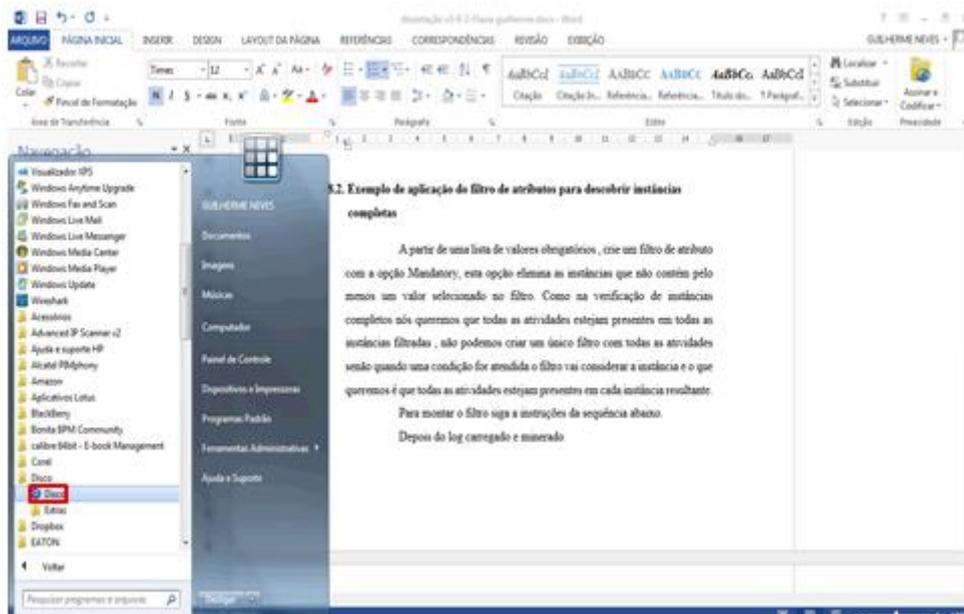


Figura 8-3 - Tela com o aplicativo Disco
 Clique na tecla abrir arquivo.

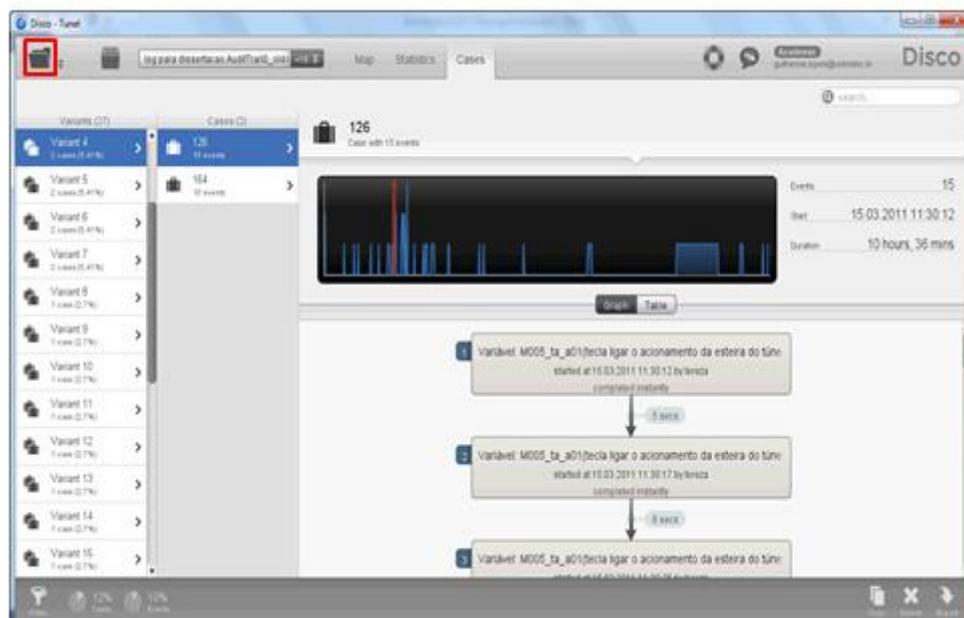


Figura 8-4 - Tecla abrir arquivo no Disco
 Selecione o arquivo do log e clique em abrir.

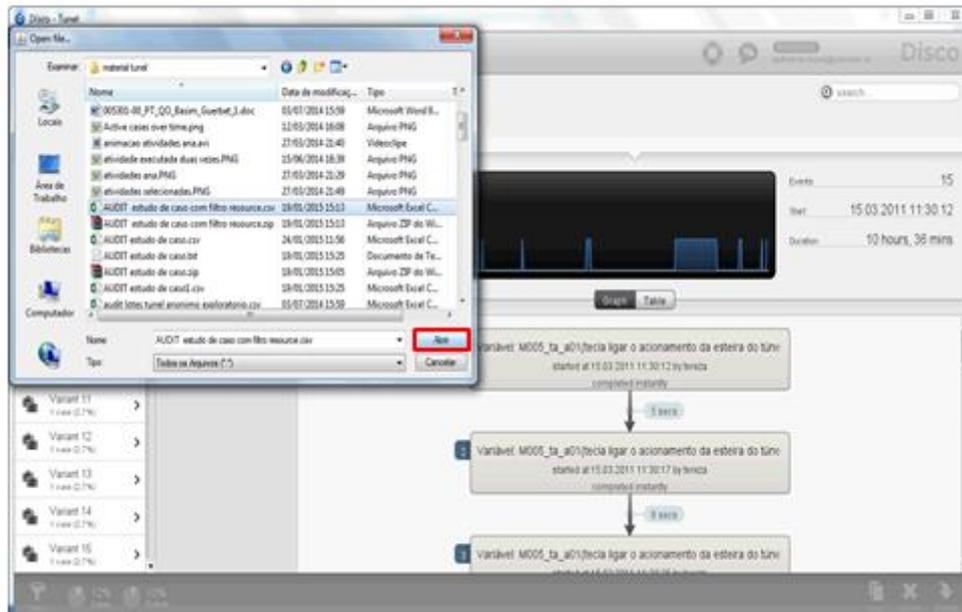


Figura 8-5 - Seleção do arquivo de log

Com o arquivo de log carregado, agora é necessário fazer o de-para das colunas do log com os atributos da ferramenta. O Case_id corresponde ao identificador único das instâncias do processo, então associe a coluna que contém o identificador da instância ao Case_id, clique no cabeçalho da coluna e depois no ícone ID.

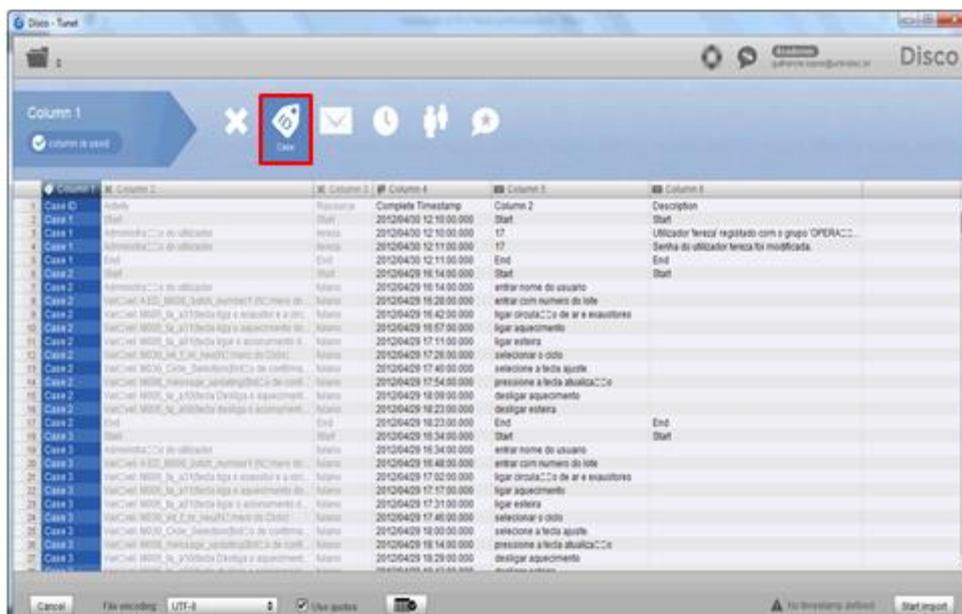


Figura 8-6 – De-para da coluna do log com o atributo case_id

Os eventos do log correspondem ao atributo Activity, faça o de-para clicando no cabeçalho da coluna onde estão os eventos do log e na tecla activity.

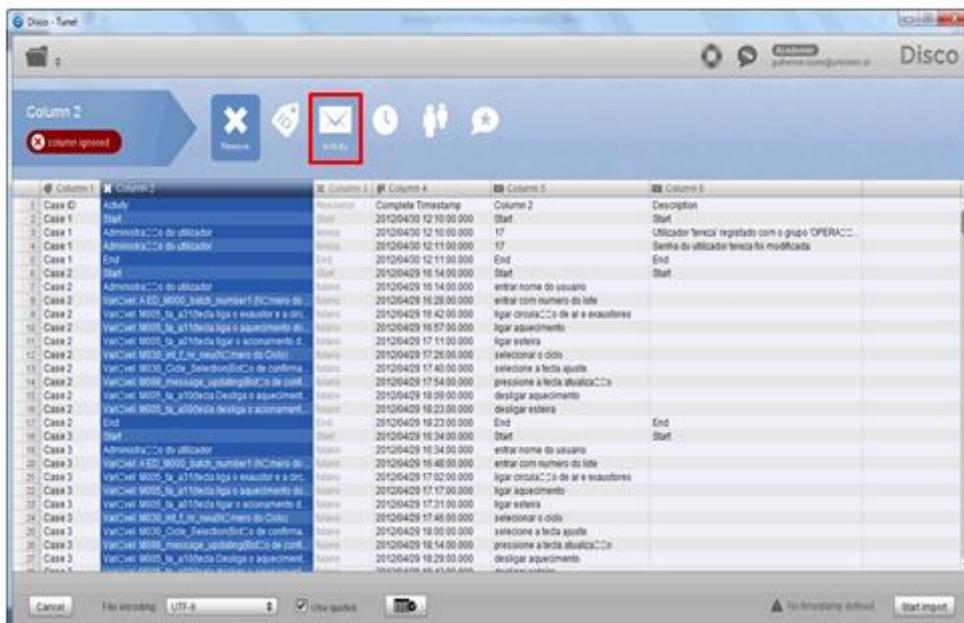


Figura 8-7 - De-para da coluna do log com Activity

O registro de data e hora que iniciou o evento equivale ao atributo Timestamp, faça o de-para clicando no cabeçalho da coluna onde se encontra o registro de tempo e no ícone Timestamp.

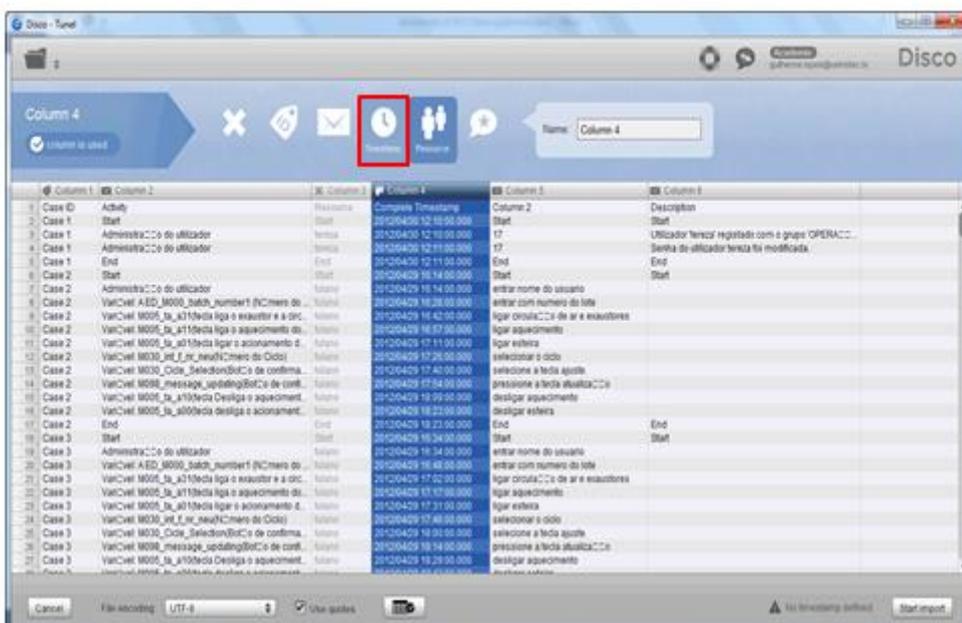


Figura 8-8 - De-para da coluna do log com o atributo TimeStamp

Como existe uma diversidade de formatos de data e hora é necessário configurar o padrão de data e hora. Clique no botão Pattern.

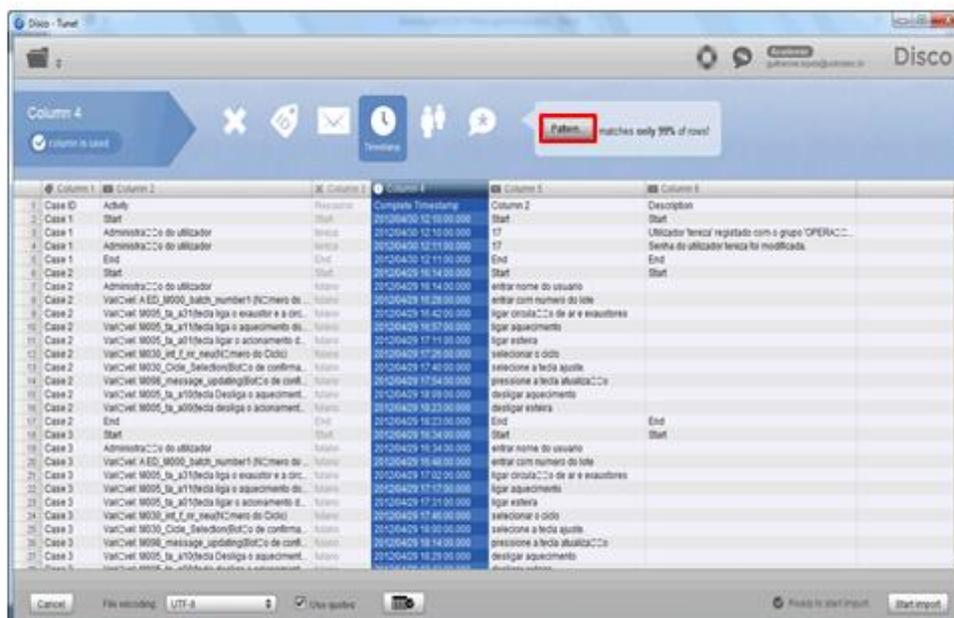


Figura 8-9 - Tecla de configuração de formato do timestamp

Selecione o formato de data e hora que corresponde ao padrão do log e aperte enter.

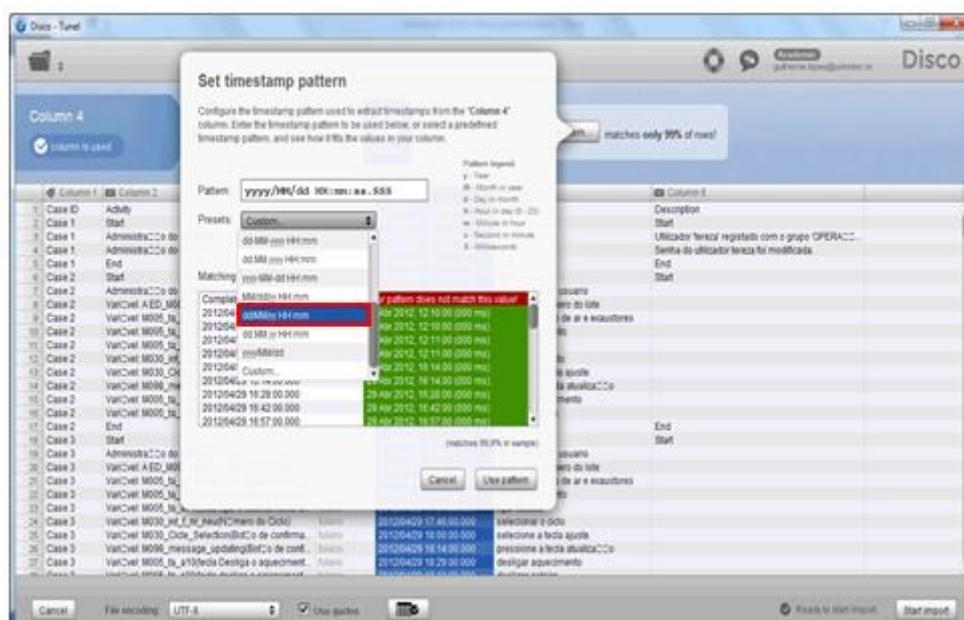


Figura 8-10 - Selecionando o padrão do TimeStamp

O campo do log que contém as informações de qual recurso executou um evento equivale ao atributo Resource no Disco, para fazer o de-para clique no

cabeçalho da coluna onde se encontra o recurso e no ícone Resource da Ferramenta Disco.

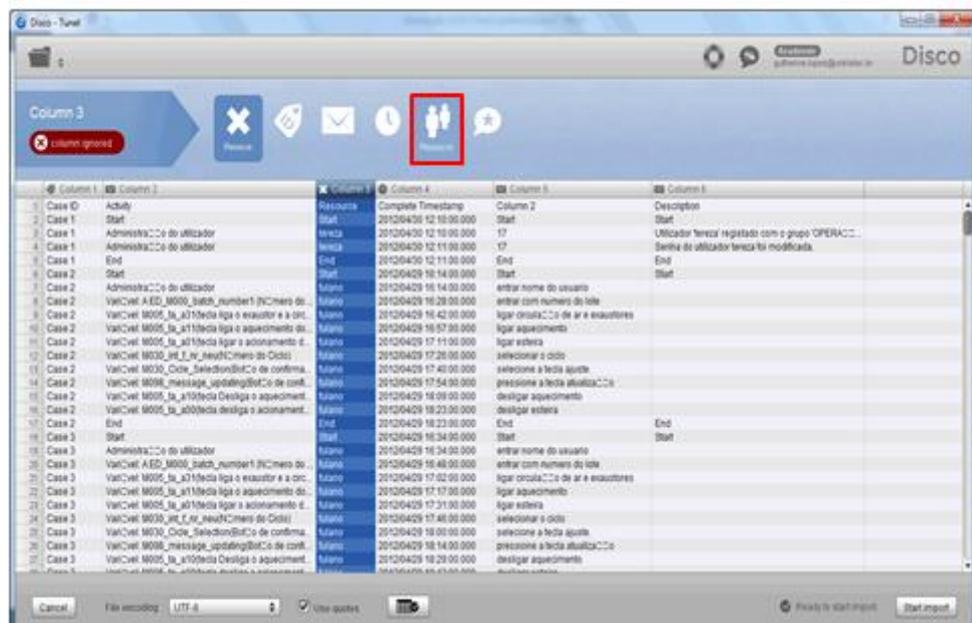


Figura 8-11 - De-para da coluna do log com o atributo Resource

As outras colunas do log que não correspondem a nenhum dos atributos listados acima devem ser marcados como outros, para isto clique no cabeçalho da coluna e no ícone * da ferramenta disco, faça isto para todas as colunas restantes.

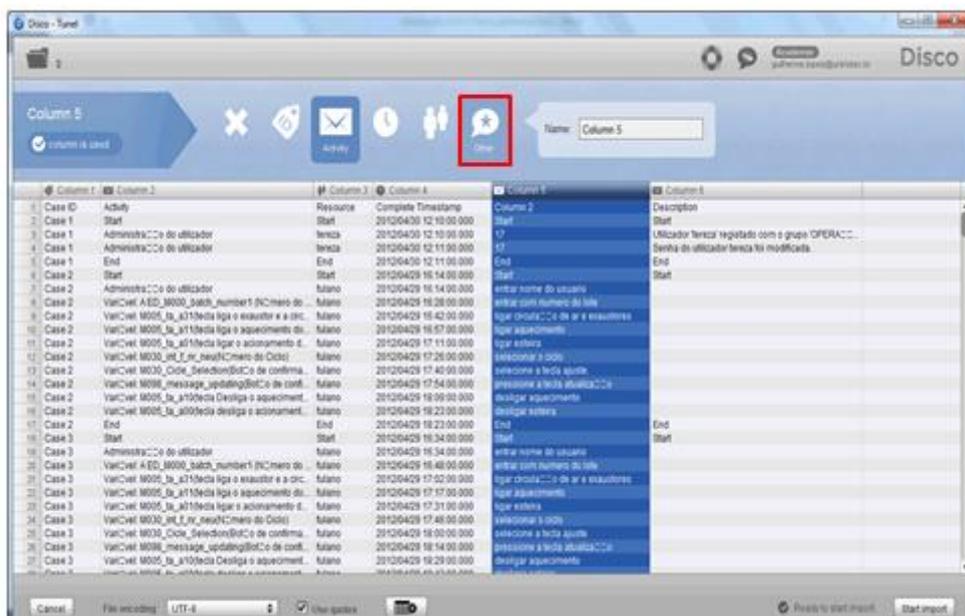


Figura 8-12 - Marcando as outras colunas do log como outros

Após todas as colunas estarem associadas a um atributo da ferramenta Disco clique em Start Import, este comando vai importar e minerar o log produzindo um modelo de processo.

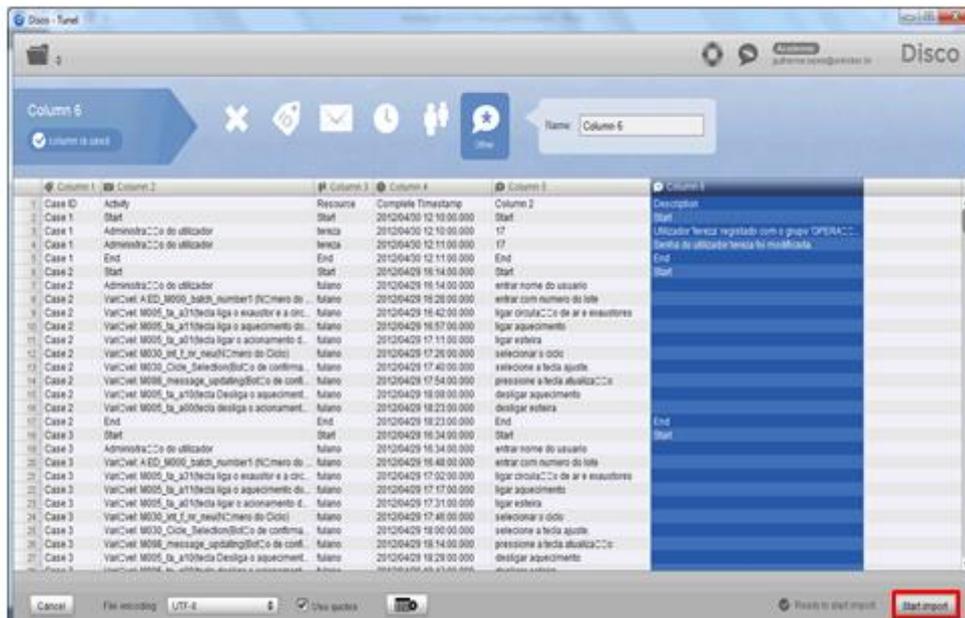


Figura 8-13 - Tecla de Start Import.

Com o modelo de processo na tela clique no botão filtrar como mostrado na Figura 8-14 - Modelo de processo Minerado e Tecla de filtro.

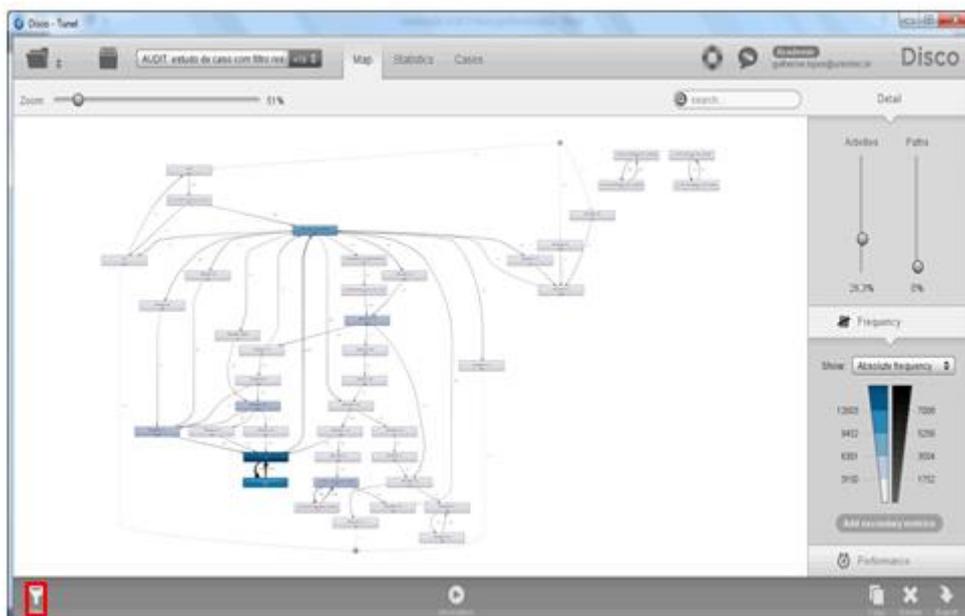


Figura 8-14 - Modelo de processo Minerado e Tecla de filtro

Para montar o filtro siga a instruções da sequência abaixo.

A partir da tela principal do filtro posicione o botão click to add new filter.

Clique na tecla conforme a Figura 8-15 - Tecla Adicionar filtro.

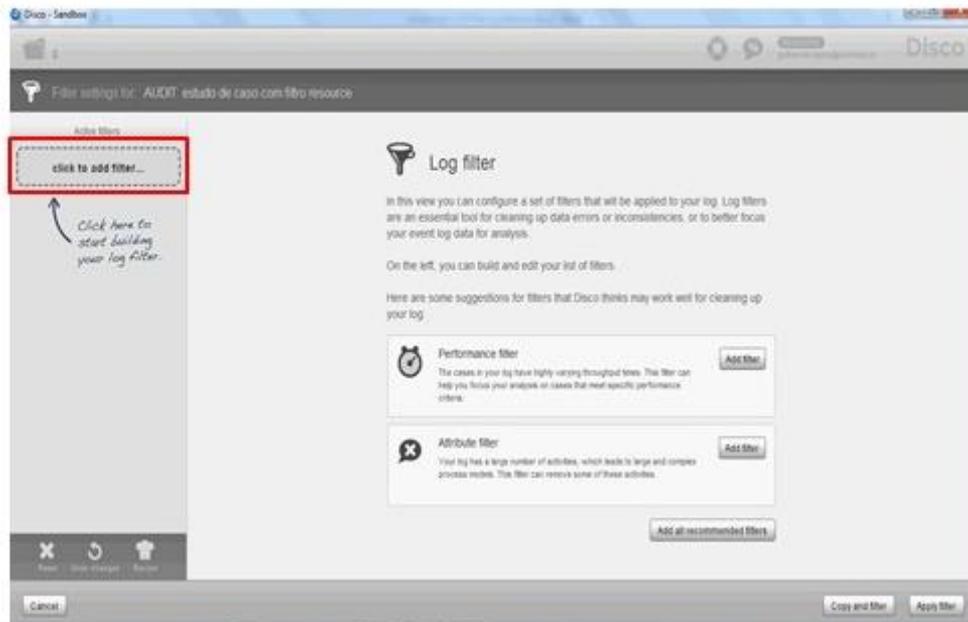


Figura 8-15 - Tecla Adicionar filtro

Selecione o filtro de atributo.

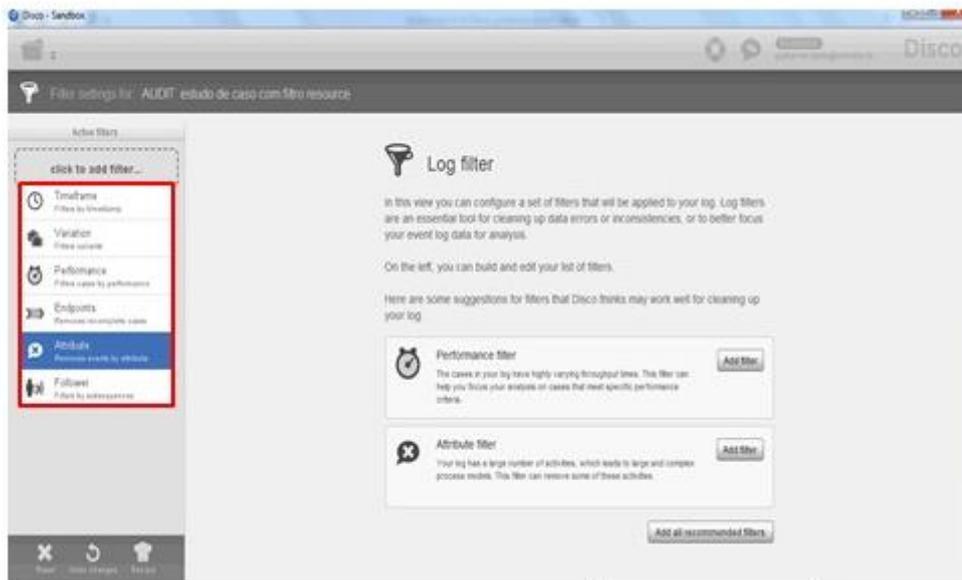


Figura 8-16 - Selecionando o tipo de filtro

Selecione o evento que deve ser filtrado e escolha o modo de filtro Mandatory conforme Figura 8-17 - Selecionando Evento a ser filtrado, este modo de filtragem vai eliminar todas as instâncias que não tiverem o evento selecionado. Como vamos fazer um filtro em cascata só vamos aplicar o filtro no final.

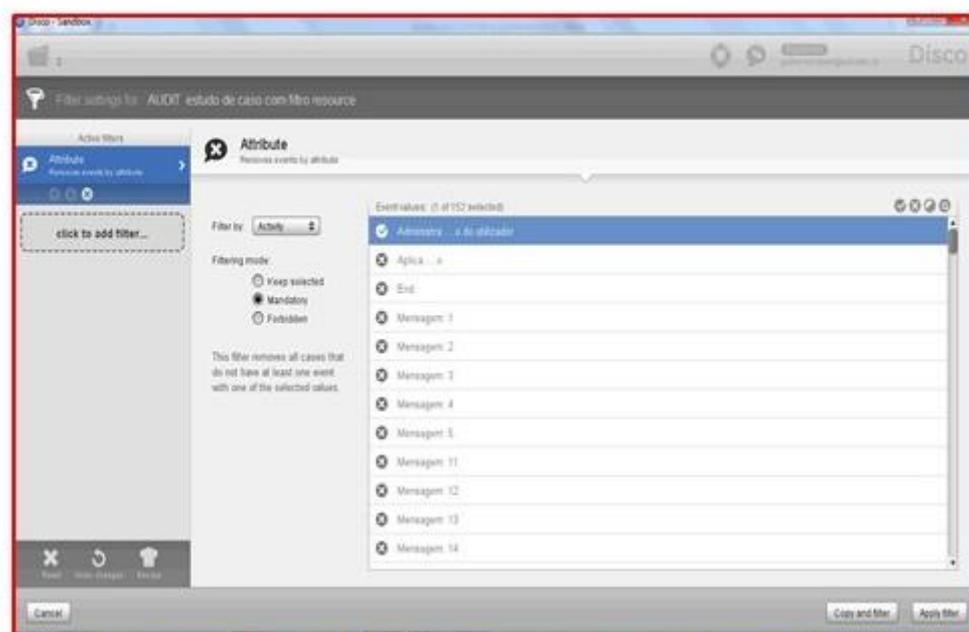


Figura 8-17 - Selecionando Evento a ser filtrado

Para selecionar o segundo evento, clique novamente no botão Click to Add Filter, selecione o novo evento, refaça este passo até o último evento que precisa ser filtrado e no final clique em Apply Filter conforme mostrado na Figura 8-18 - Filtro em cascata.

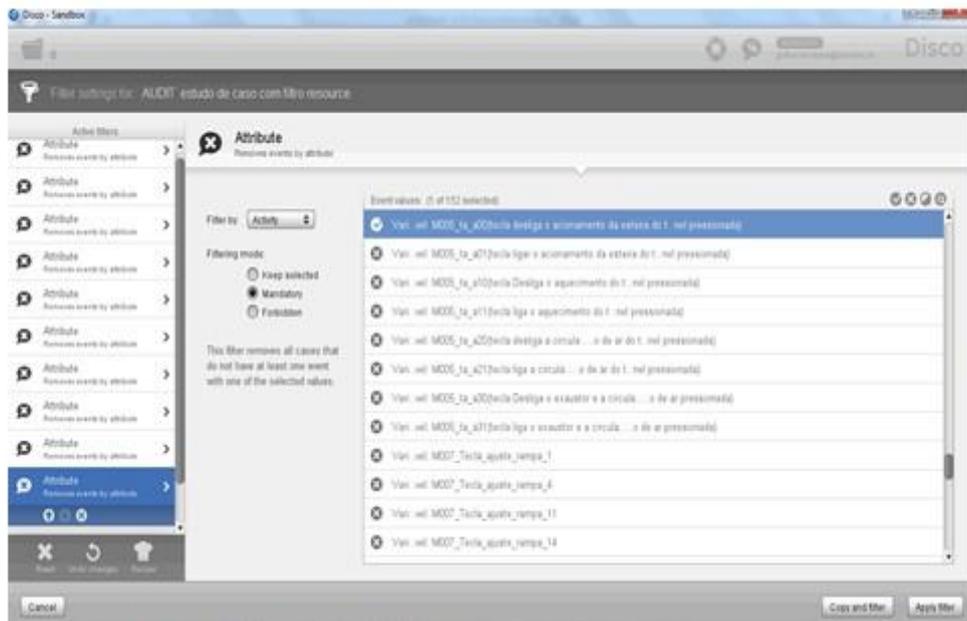


Figura 8-18 - Filtro em cascata

O resultado será o modelo de processo apenas com as instâncias que contém todos os eventos utilizados na construção do filtro conforme mostrado na Figura 8-19 - Modelo de processo após a aplicação do filtro.

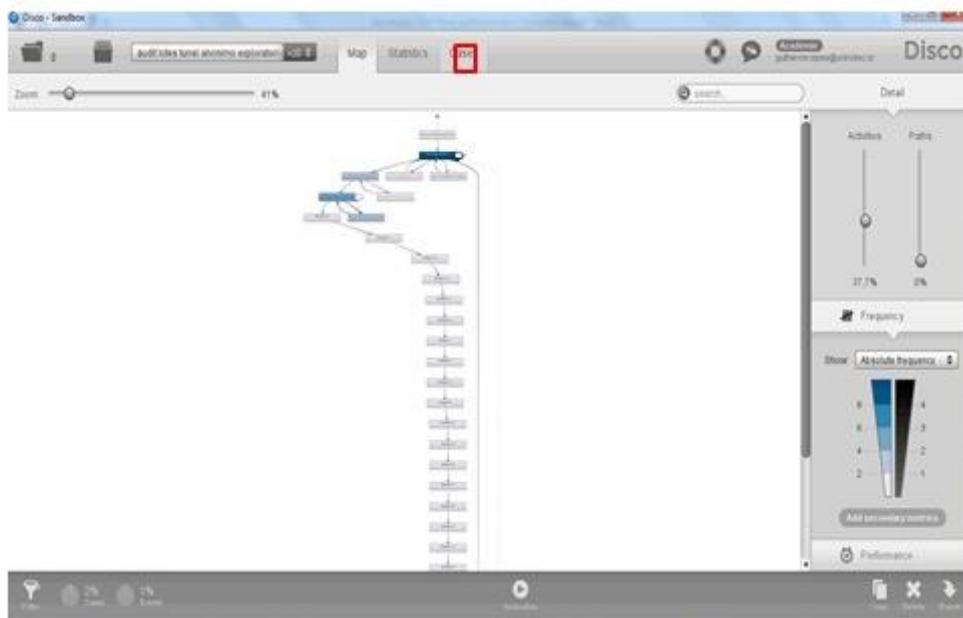


Figura 8-19 - Modelo de processo após a aplicação do filtro

8.4. Exemplo de aplicação de filtro para descobrir instâncias com desvio regulatório com atividades proibidas.

Para descobrir as instâncias que contêm atividades proibidas, é necessário aplicar o filtro de atributo, no exemplo abaixo foi importado um log de eventos e a atividade proibida é a mensagem 44. Veremos a seguir o passo a passo da aplicação do filtro.

O primeiro passo é a seleção do arquivo de log, e fazer o de-para entre os campos do arquivo e a ferramenta de mineração, cada campo do log deve estar associado a um atributo da ferramenta de mineração. O campo case_id, TimeStamp, resource e activity são obrigatórios. Esta associação é feita clicando na coluna e no atributo no menu superior como mostrado na Figura 8-20 - Seleccionando os campos e fazendo o relacionamento de-para.

Case_id	Received	Timestamp	DataFacts	UserID	ObjectID	Description	Comment	Checked
1	08012013 08:07	+2:00	operador1	Variable A ED_8000_batch_number? (Número de l...	Assinado: Alteração do valor da variável A ED_800	Zia-20		
2	08012013 08:09	+2:00	operador1	Variable M005_la_a110eda liga o aquecimento do ...	Alteração do valor da variável M005_la_a110eda k.	YAGZ		
3	08012013 08:09	+2:00	operador1	Variable M005_la_a100eda Desliga o aqueciment...	Alteração do valor da variável M005_la_a100eda D.	YAGZ		
4	08012013 08:08	+2:00	operador1	Variable M005_la_a110eda liga o aquecimento do ...	Alteração do valor da variável M005_la_a110eda k.	YAGZ		
5	08012013 08:10	+2:00	operador1	Mensagem: 141	Confirmação da mensagem 5178 Pressão alta no	YAGZ		
6	08012013 08:10	+2:00	operador1	Variable M005_la_a110eda liga o aquecimento do ...	Alteração do valor da variável M005_la_a110eda k.	YAGZ		
7	08012013 08:10	+2:00	operador1	Variable M005_la_a100eda Desliga o aqueciment...	Alteração do valor da variável M005_la_a100eda D.	Zia-20		
8	08012013 08:10	+2:00	operador1	Variable M005_la_a110eda liga o aquecimento do ...	Alteração do valor da variável M005_la_a110eda k.	Zia-20		
9	08012013 08:11	+2:00	operador1	Variable M030_mt_f_m_novoNúmero de Cidco)	Alteração do valor da variável M030_mt_f_m_novN.	YAGZ		
10	08012013 08:11	+2:00	operador1	Variable M030_mt_f_m_novoNúmero de Cidco)	Alteração do valor da variável M030_mt_f_m_novN.	YAGZ		
11	08012013 08:12	+2:00	operador1	Variable M030_mt_f_m_novoNúmero de Cidco)	Alteração do valor da variável M030_mt_f_m_novN.	YAGZ		
12	08012013 08:12	+2:00	operador1	Variable M030_mt_f_m_novoNúmero de Cidco)	Alteração do valor da variável M030_mt_f_m_novN.	YAGZ		
13	08012013 08:12	+2:00	operador1	Variable M030_mt_f_m_novoNúmero de Cidco)	Alteração do valor da variável M030_mt_f_m_novN.	Zia-20		
14	08012013 08:12	+2:00	operador1	Variable M030_mt_f_m_novoNúmero de Cidco)	Alteração do valor da variável M030_mt_f_m_novN.	YAGZ		
15	08012013 08:12	+2:00	operador1	Variable M030_mt_f_m_novoNúmero de Cidco)	Alteração do valor da variável M030_mt_f_m_novN.	YAGZ		
16	08012013 08:12	+2:00	operador1	Variable M030_mt_f_m_novoNúmero de Cidco)	Alteração do valor da variável M030_mt_f_m_novN.	YAGZ		
17	08012013 08:12	+2:00	operador1	Mensagem: 85	Confirmação da mensagem 3400 Temperatura m...	YAGZ		
18	08012013 08:13	+2:00	Sistema	Arguio	Dados de arguio fechados	YAGZ		
19	08012013 08:14	+2:00	Sistema	Arguio	Dados de arguio alertas de novo	Zia-20		
20	08012013 08:15	+2:00	operador1	Variable A ED_8000_batch_number? (Número de l...	Assinado: Alteração do valor da variável A ED_800	YAGZ		
21	08012013 08:20	+2:00	Sistema	Administração do utilizador	Utilizador já não log out	YAGZ		
22	08012013 08:22	+2:00	operador1	Administração do utilizador	Utilizador "operador1" registado com o grupo OPER.	YAGZ		
23	08012013 08:22	+2:00	operador1	Mensagem: 85	Confirmação da mensagem 3400 Temperatura m...	YAGZ		
24	08012013 08:22	+2:00	operador1	Mensagem: 85	Confirmação da mensagem 3400 Temperatura m...	YAGZ		
25	08012013 08:22	+2:00	operador1	Mensagem: 85	Confirmação da mensagem 3400 Temperatura m...	YAGZ		
26	08012013 08:22	+2:00	operador1	Mensagem: 85	Confirmação da mensagem 3400 Temperatura m...	YAGZ		
27	08012013 08:22	+2:00	operador1	Mensagem: 85	Confirmação da mensagem 3400 Temperatura m...	YAGZ		
28	08012013 08:22	+2:00	operador1	Mensagem: 85	Confirmação da mensagem 3400 Temperatura m...	YAGZ		

Figura 8-20 - Selecionando os campos e fazendo o relacionamento de-para
 Após a associação estar completa o log deve ser importado e minerado, para tal clique no botão start import.

Figura 8-21 – Importando o log de eventos
 O modelo de processo minerado será mostrado, e agora pode-se configurar o filtro desejado clicando no botão selecionado.

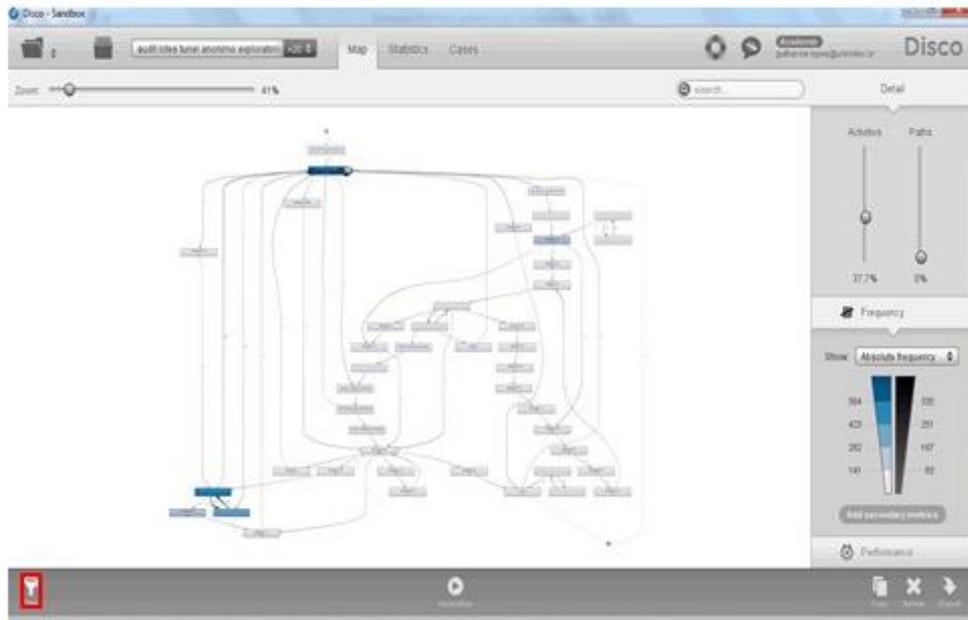


Figura 8-22 - Modelo de processo gerado

A próxima etapa é a adição de um filtro, clique no botão selecionado conforme a Figura 8-23 - Adicionando o filtro.

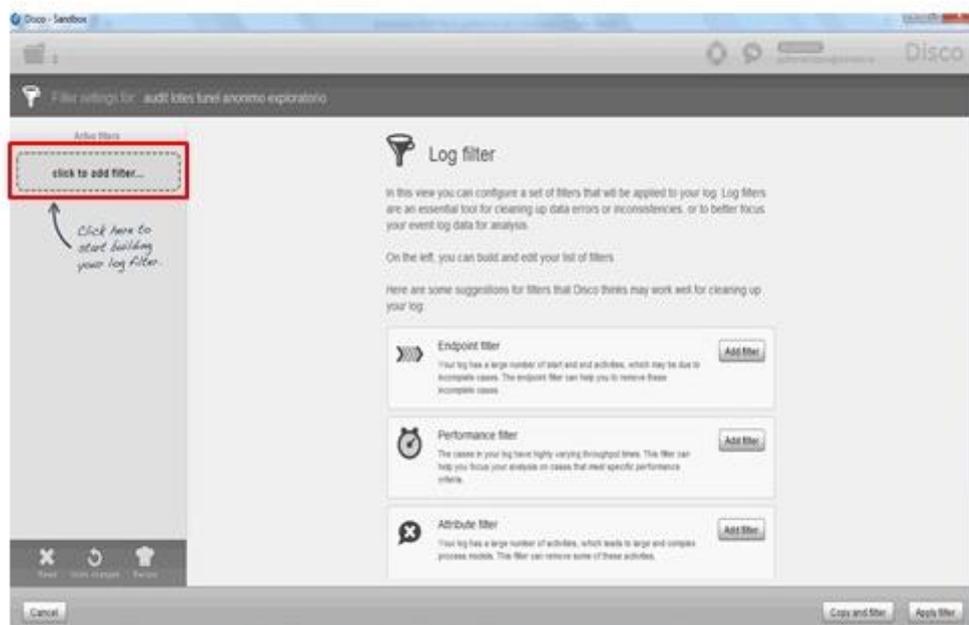


Figura 8-23 - Adicionando o filtro

A próxima etapa é a escolha do tipo de filtro, no caso estudado o filtro adequado é o de atributo, pois queremos filtrar o valor que um determinado atributo assume. Selecione a opção Attribute conforme mostrado na Figura 8-24 - Selecionando o filtro de atributo.

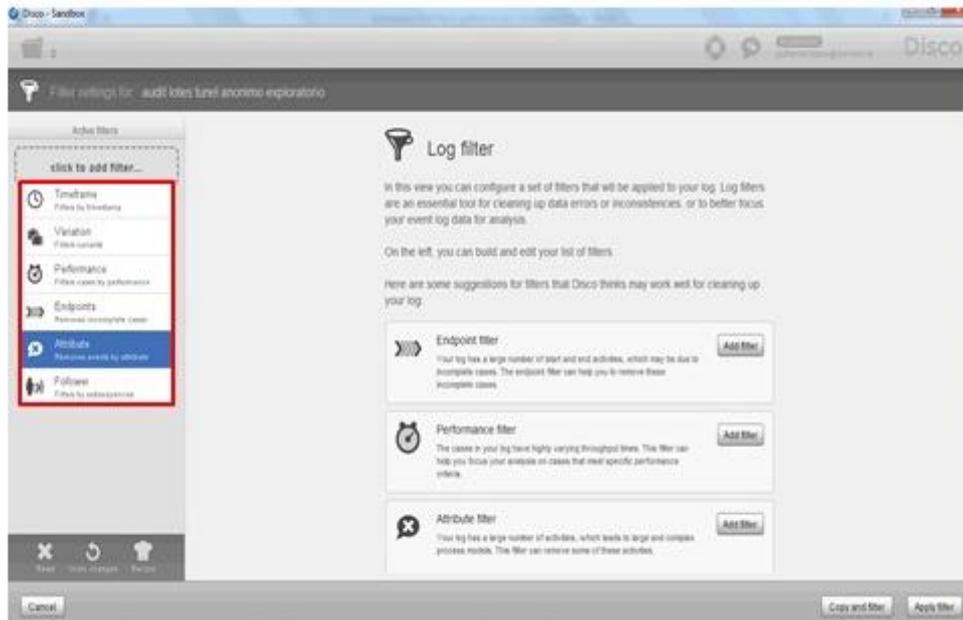


Figura 8-24 - Selecionando o filtro de atributo

Configure o filtro selecionando a atividade que deve ser filtrada e selecionado o Filtering Mode Mandatory, conforme a Figura 8-25 - Selecionando a atividade e o tipo de filtro, este modo de filtragem vai eliminar todas as instâncias que não possuam pelo menos um dos valores selecionados. Portanto o arquivo resultante serão todas as instâncias que contém valor selecionado. Clique no botão Apply Filter para aplicar o filtro.

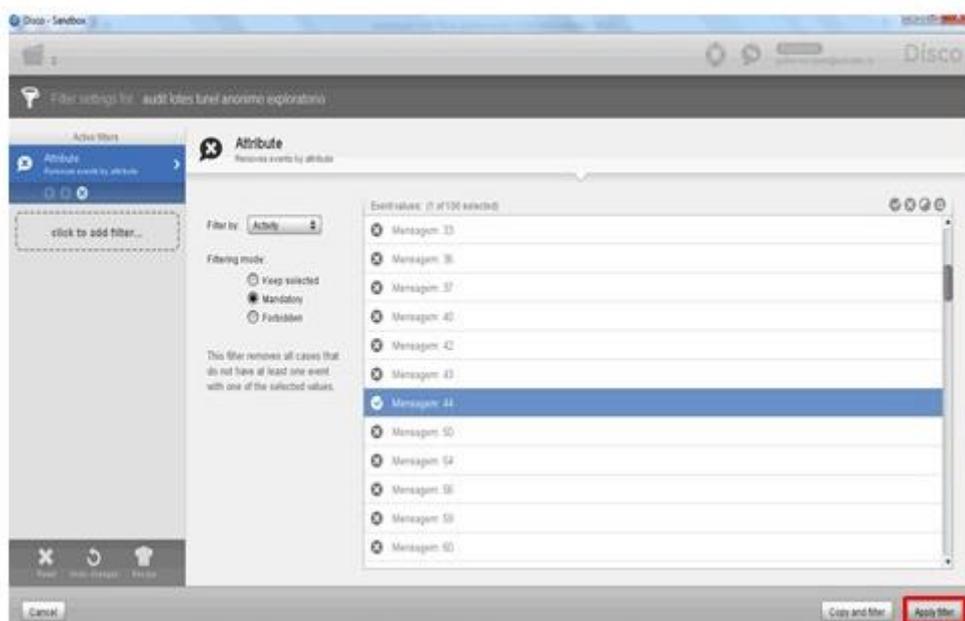


Figura 8-25 - Selecionando a atividade e o tipo de filtro

Após o filtro aplicado a ferramenta mostra o modelo de processos resultante conforme Figura 8-26 - Resultado da aplicação do filtro.

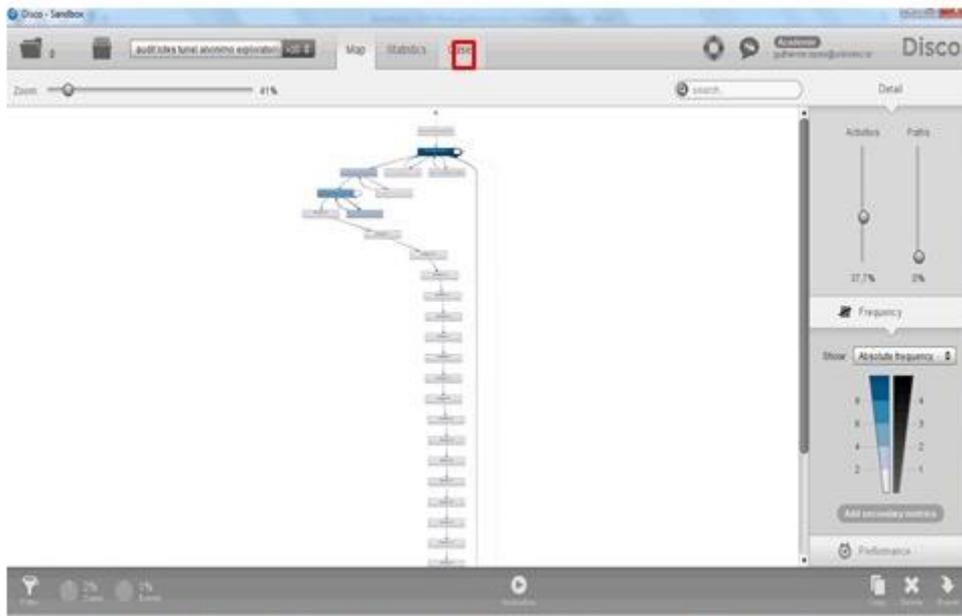


Figura 8-26 - Resultado da aplicação do filtro

Clicando em Cases, podemos verificar quais as instâncias resultaram, neste exemplo somente uma instâncias tem a atividade selecionada no filtro, podemos ver a instância resultante e as atividades executadas por ela na Figura 8-27 - Verificando as instâncias resultantes.

The screenshot shows the Disco software interface in the 'Cases' view. The main area displays a list of cases, with one case selected. The selected case shows a list of activities and their descriptions. The interface includes a search bar at the top right, a 'Cases' panel on the left, and a 'Detail' panel on the right showing the selected case's details.

Atividade	Recursos	Data	Tempo	Descrição
1 Variável A EO_8000_batida_number1 (Número de batida)	Sistema	12.01.2013	14:52:00	Assinador: Alteração de us
2 Administração do utilizador	Sistema	12.01.2013	14:57:00	Utilizador já filiado no
3 Administração do utilizador	Sistema	13.01.2013	18:30:00	Utilizador sem tempo 1 d
4 Administração do utilizador	Sistema	13.01.2013	18:30:00	Utilizador sem tempo 2 d
5 Administração do utilizador	Sistema	14.01.2013	06:15:00	Utilizador e-mail regis
6 Variável 9005_Su_a210fecta liga a circulação de ar do fone1 pressionado	evento	14.01.2013	06:15:00	Alteração do valor da vari
7 Variável 9005_Su_a210fecta desliga a circulação de ar do fone1 pressionado	evento	14.01.2013	06:15:00	Alteração do valor da vari
8 Variável 9005_Su_a210fecta liga a circulação de ar do fone1 pressionado	evento	14.01.2013	06:15:00	Alteração do valor da vari
9 Variável 9005_Su_a210fecta desliga a circulação de ar do fone1 pressionado	evento	14.01.2013	06:15:00	Alteração do valor da vari
10 Variável 9005_Su_a210fecta Desliga o exaustor e a circulação de ar pressionado	evento	14.01.2013	06:15:00	Alteração do valor da vari
11 Variável 9005_Su_a210fecta liga o exaustor e a circulação de ar pressionado	evento	14.01.2013	06:15:00	Alteração do valor da vari
12 Variável 9005_Su_a210fecta liga o exaustor e a circulação de ar pressionado	evento	14.01.2013	06:15:00	Alteração do valor da vari
13 Variável 9005_Su_a210fecta Desliga o exaustor e a circulação de ar pressionado	evento	14.01.2013	06:15:00	Alteração do valor da vari
14 Variável 9005_Su_a210fecta liga o exaustor e a circulação de ar pressionado	evento	14.01.2013	06:15:00	Alteração do valor da vari
15 Mensagem 218	evento	14.01.2013	06:15:00	Confirmação da mensag
16 Mensagem 217	evento	14.01.2013	06:15:00	Confirmação da mensag
17 Mensagem 216	evento	14.01.2013	06:15:00	Confirmação da mensag

Figura 8-27 - Verificando as instâncias resultantes

Este procedimento se aplica a qualquer atributo e qualquer valor existente no log, portanto pode-se utilizar o mesmo procedimento apenas alterando o valor e o atributo.

8.5. Exemplo de aplicação de filtro para descobrir instâncias com desvio regulatório devido à restrição de valor.

Uma restrição muito comum no domínio estudado é o tempo, determinados processos tem um tempo mínimo e máximo de duração, se uma instância estiver fora desta especificação indica que alguma atividade não foi executada por tempo suficiente ou pelo contrário o tempo de execução foi excessivo podendo comprometer o produto. Embora no estudo de caso não tenha aparecido esta restrição ela pode aparecer em outros processos e é interessante que o analista saiba como utilizá-la.

Após importar o arquivo de log adicione o filtro conforme a Figura 8-28 - Adicionando Filtro.

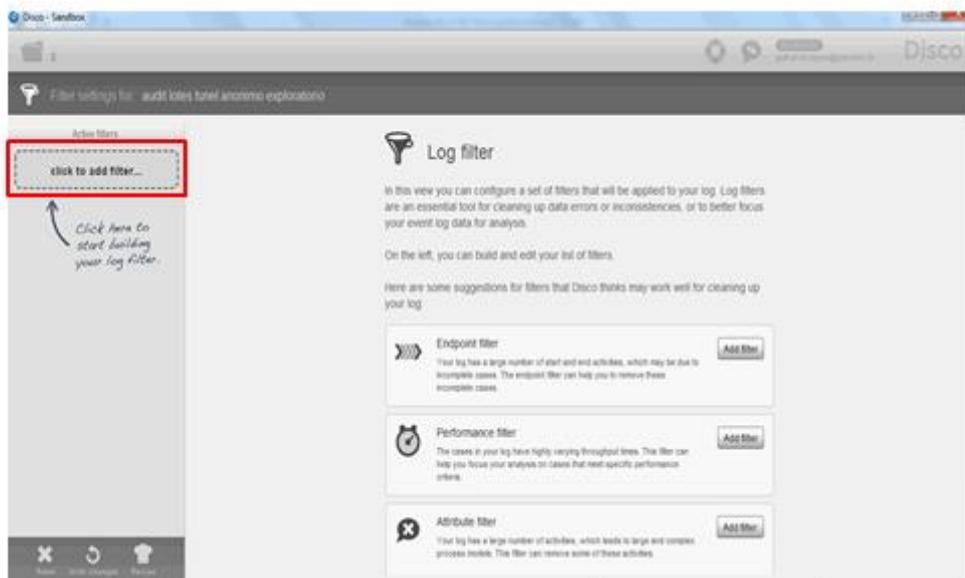


Figura 8-28 - Adicionando Filtro

Selecione o filtro de desempenho na lista conforme Figura 8-29 - Seleção do filtro .

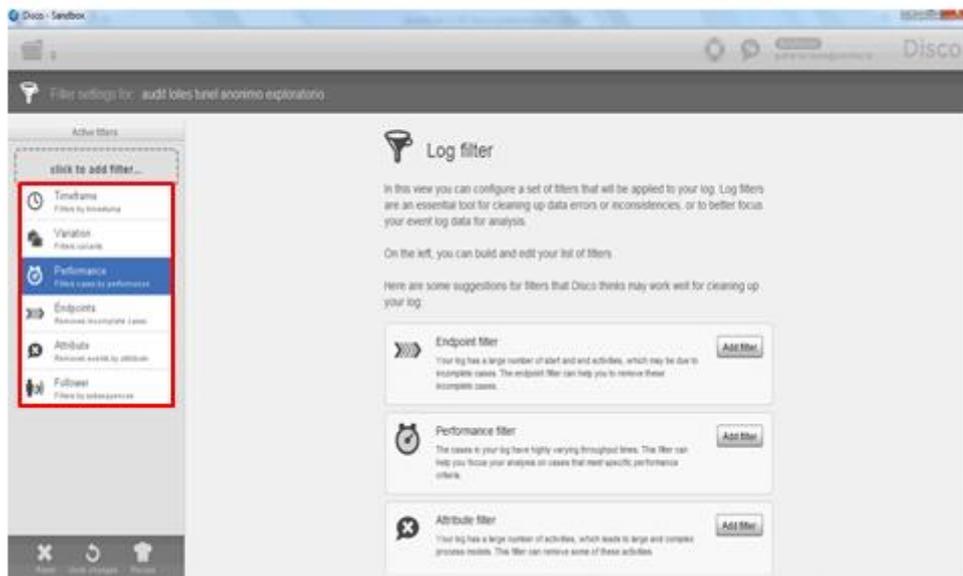


Figura 8-29 - Seleção do filtro Desempenho

Configure a duração mínima e máxima de cada instância conforme a Figura 8-30 - Configurando o intervalo de tempo.

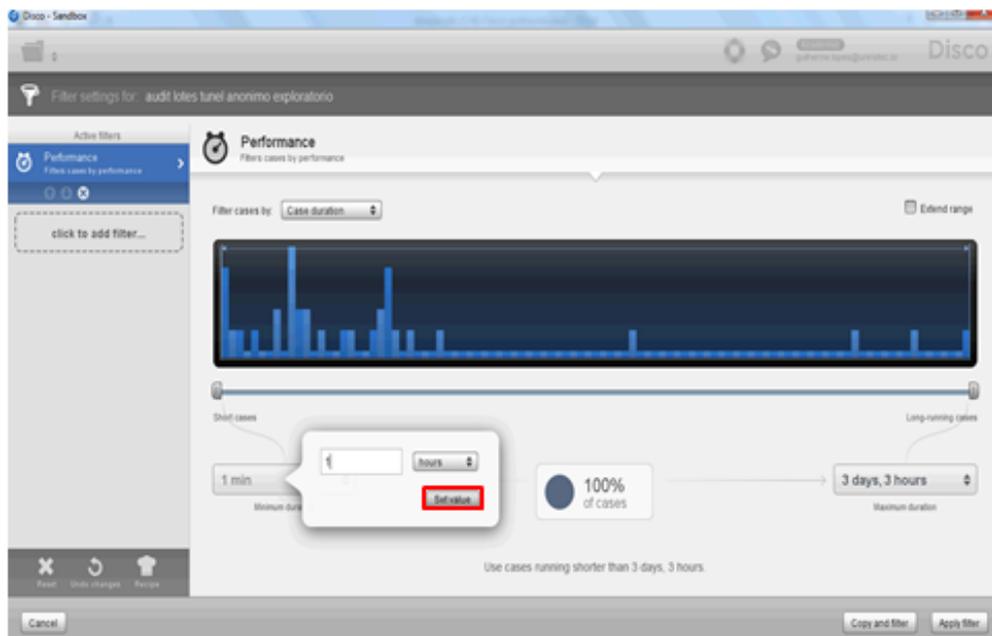


Figura 8-30 - Configurando o intervalo de tempo

Aplique o filtro pressionando o botão Apply Filter conforme a Figura 8-31 - Aplicando o filtro.

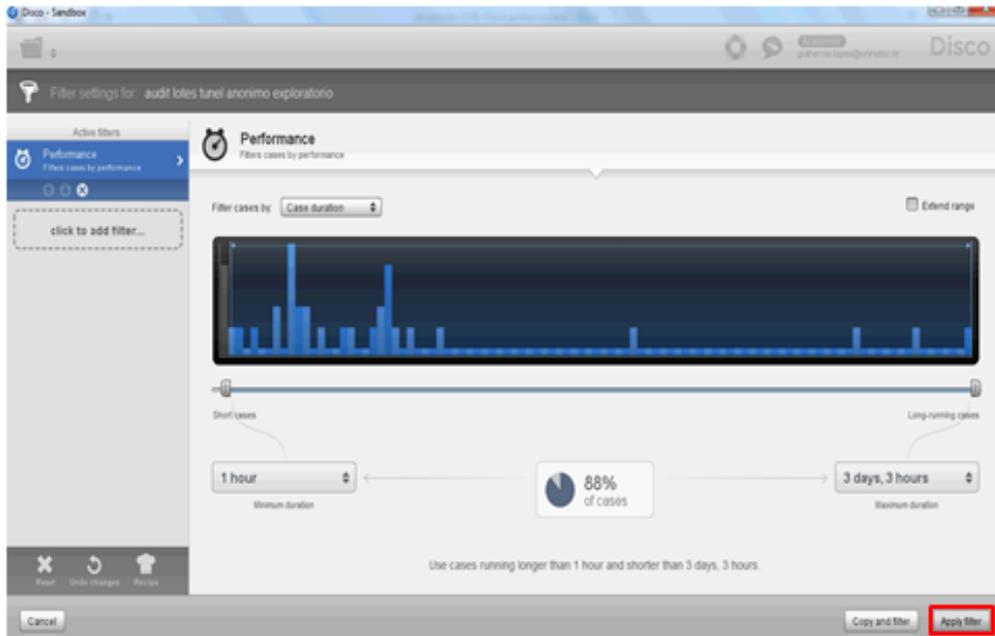


Figura 8-31 - Aplicando o filtro

Verifique o resultado da aplicação do filtro, as estatísticas mostram a quantidade de eventos, número de instâncias, duração e outros dados conforme a Figura 8-32 - Estatísticas do resultado da aplicação do filtro.

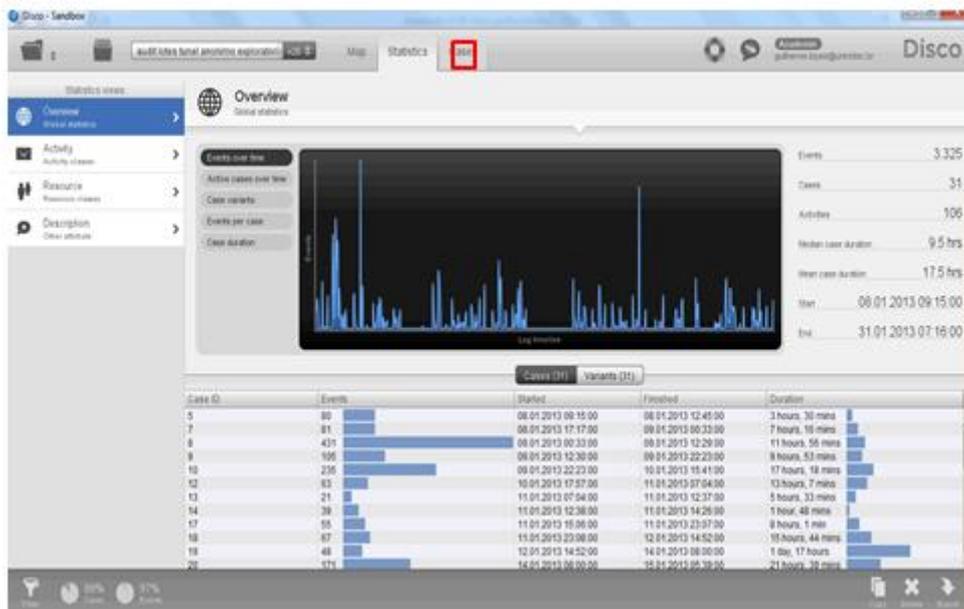


Figura 8-32 - Estatísticas do resultado da aplicação do filtro

Clicando na aba cases, podemos observar as instâncias resultantes e as atividades realizadas por elas, bem como as variações, conforme mostrado na Figura 8-33 - Instâncias resultantes após a aplicação do filtro.

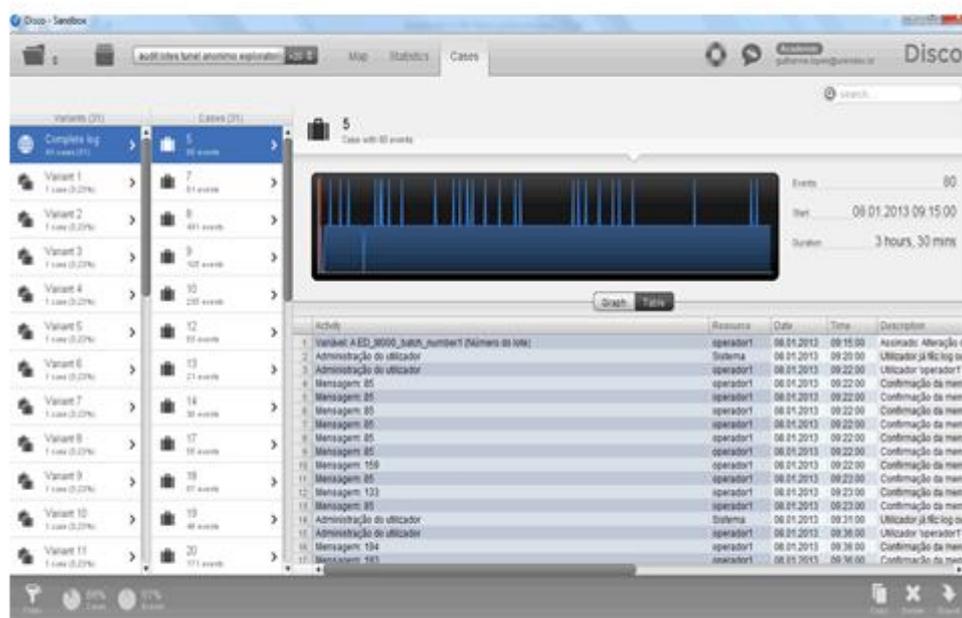


Figura 8-33 - Instâncias resultantes após a aplicação do filtro.

8.6. Estudo de caso, aplicação de filtros para descoberta de instâncias completas no log com filtro de Resources.

Esta parte do estudo de caso é ilustrada um problema muito comum no levantamento de informações sobre processos de negócios, a falta de informação ou informação errada do especialista do domínio. Na abordagem tradicional de modelagem de processos, normalmente verifica-se por amostragem o que vai possibilitar erros da verificação da conformidade do processo. Uma das grandes vantagens da mineração de processos é a possibilidade de verificação de 100% das instâncias do processo possibilitando a verificação realmente completa entre o processo modelado e o processo executado. Conforme visto em 5.4, o especialista do domínio informou que as atividades executadas por recursos específicos poderiam ser removidas, foi aplicado o filtro e removidas todas as atividades executadas pelos recursos, o log resultante foi utilizado para a verificação de instâncias completas. A construção do filtro é mostrada na Figura 8-34 - Aplicação correta do filtro para descoberta de instâncias completas, cada filtro é criado com apenas uma atividade e a cada atividade que precisa ser filtrada é gerado um filtro em cascata.

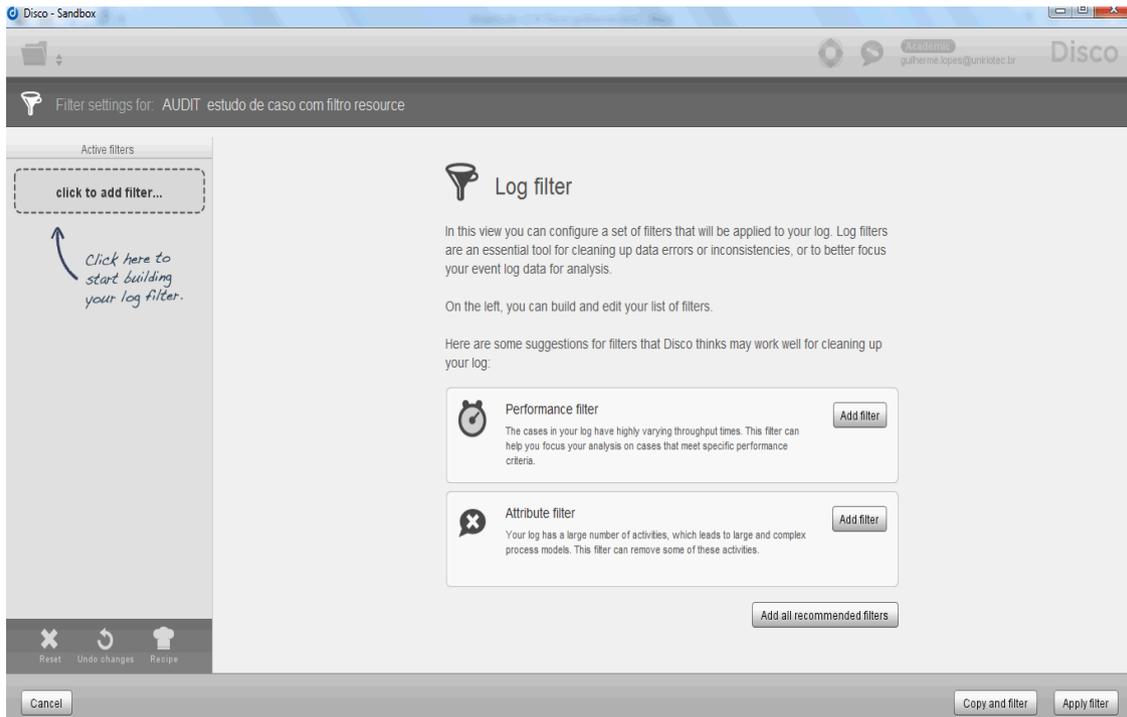


Figura 8-34 - Aplicação correta do filtro para descoberta de instâncias completas

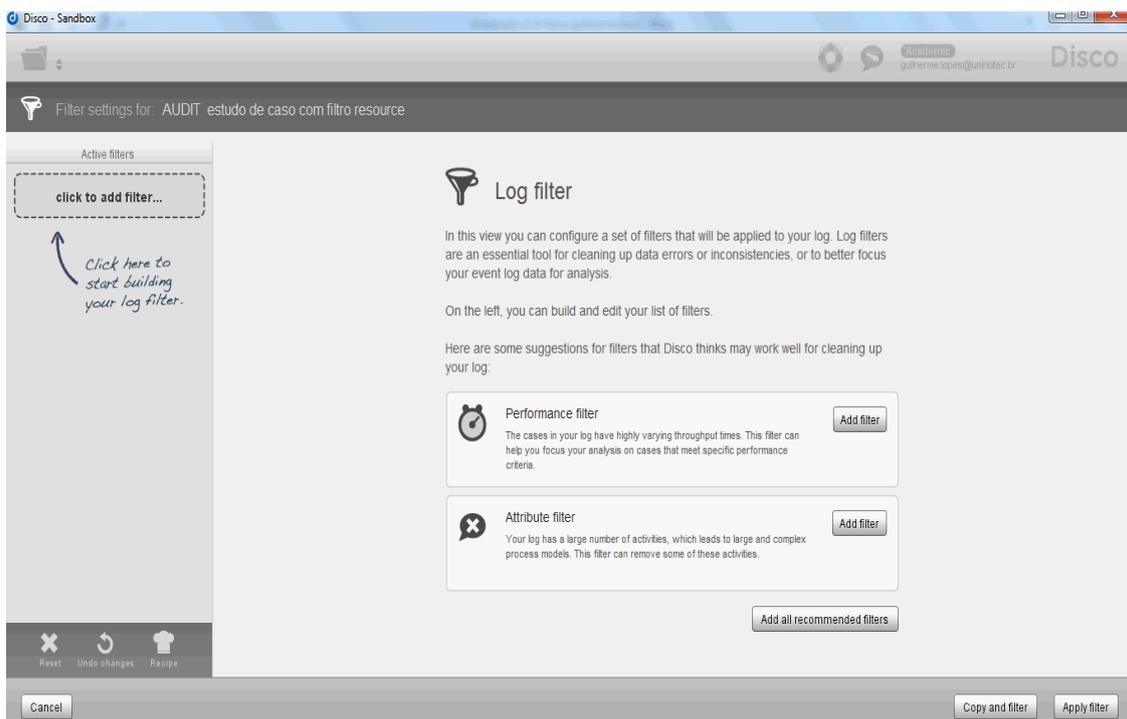


Figura 8-35 - Aplicação correta do filtro para descoberta de instâncias completas sequência 1

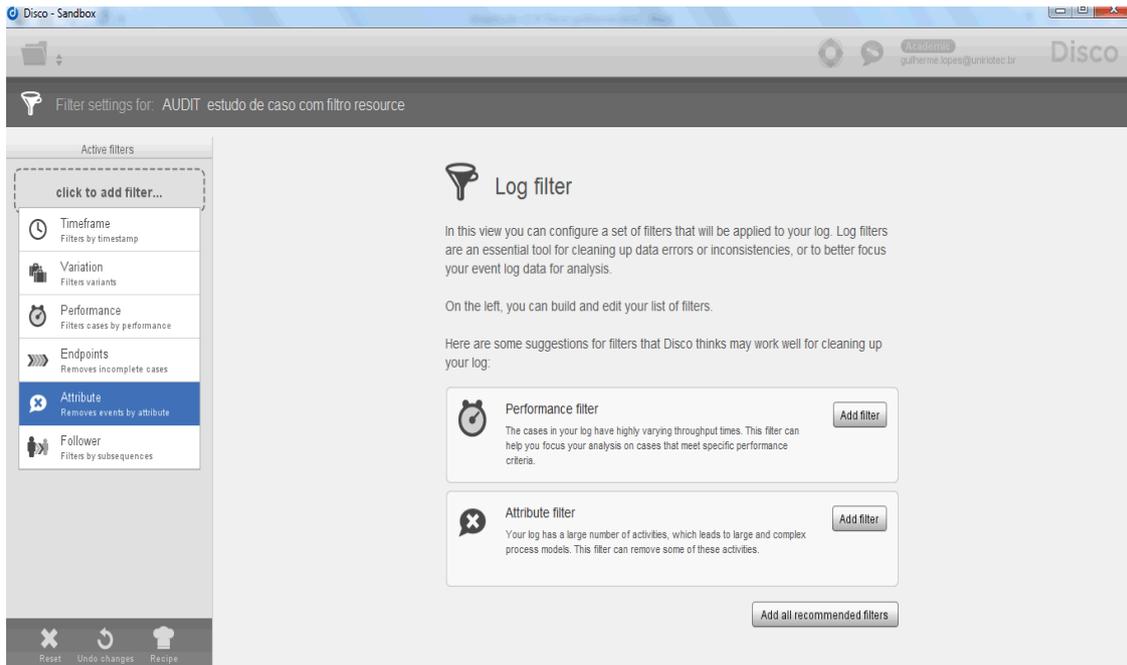


Figura 8-36 - Aplicação correta do filtro para descoberta de instâncias completas sequência 2

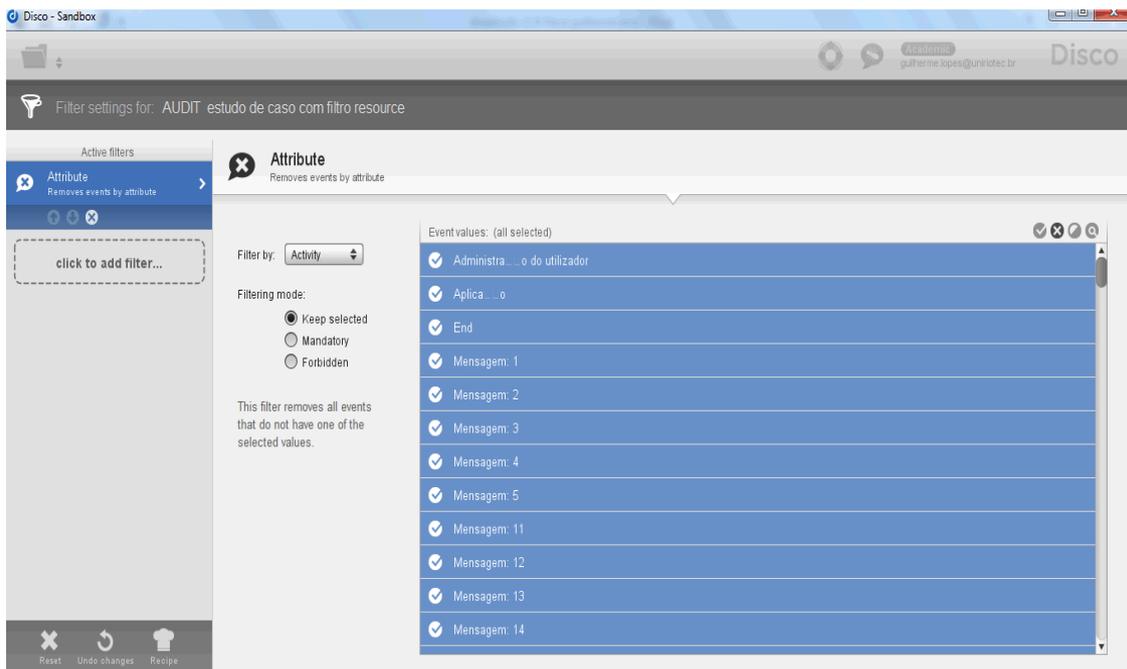


Figura 8-37 - Aplicação correta do filtro para descoberta de instâncias completas sequência 3

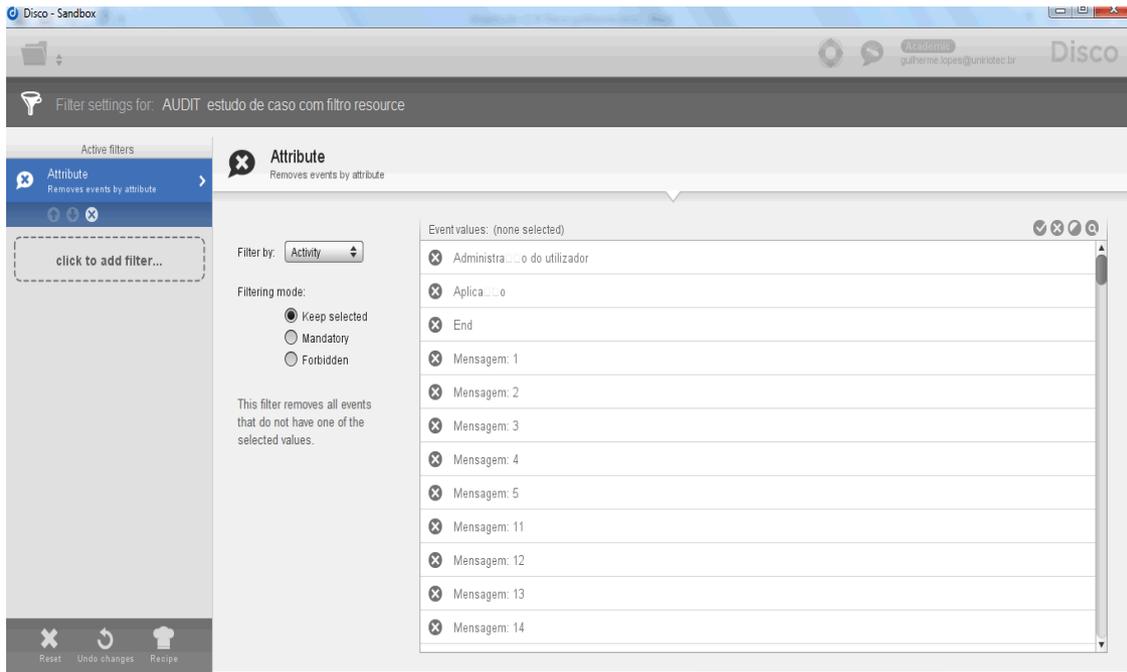


Figura 8-38 - Aplicação correta do filtro para descoberta de instâncias completas sequência 4

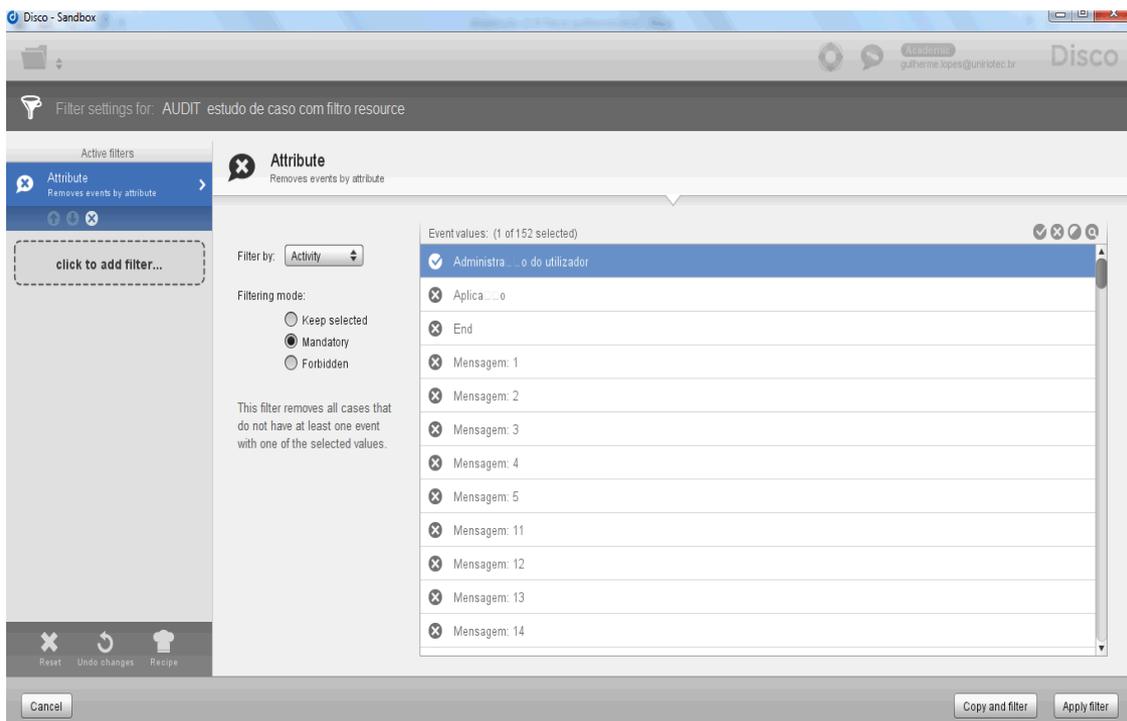


Figura 8-39 - Aplicação correta do filtro para descoberta de instâncias completas sequência 5

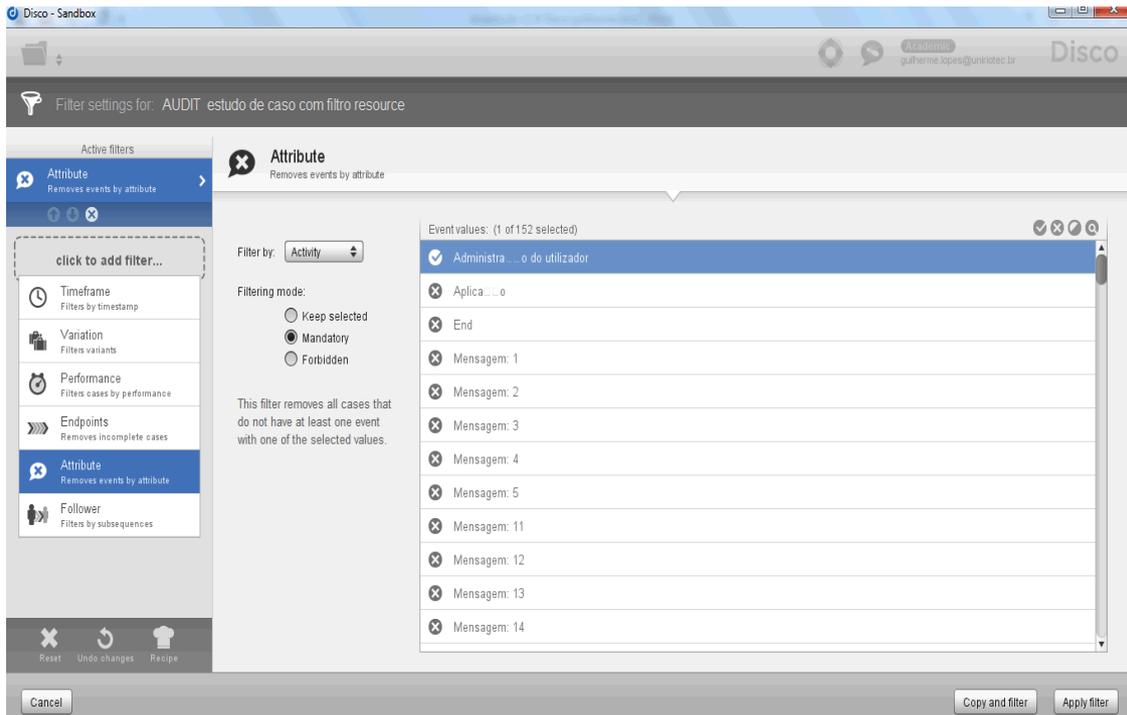


Figura 8-40 - Aplicação correta do filtro para descoberta de instâncias completas sequência 6

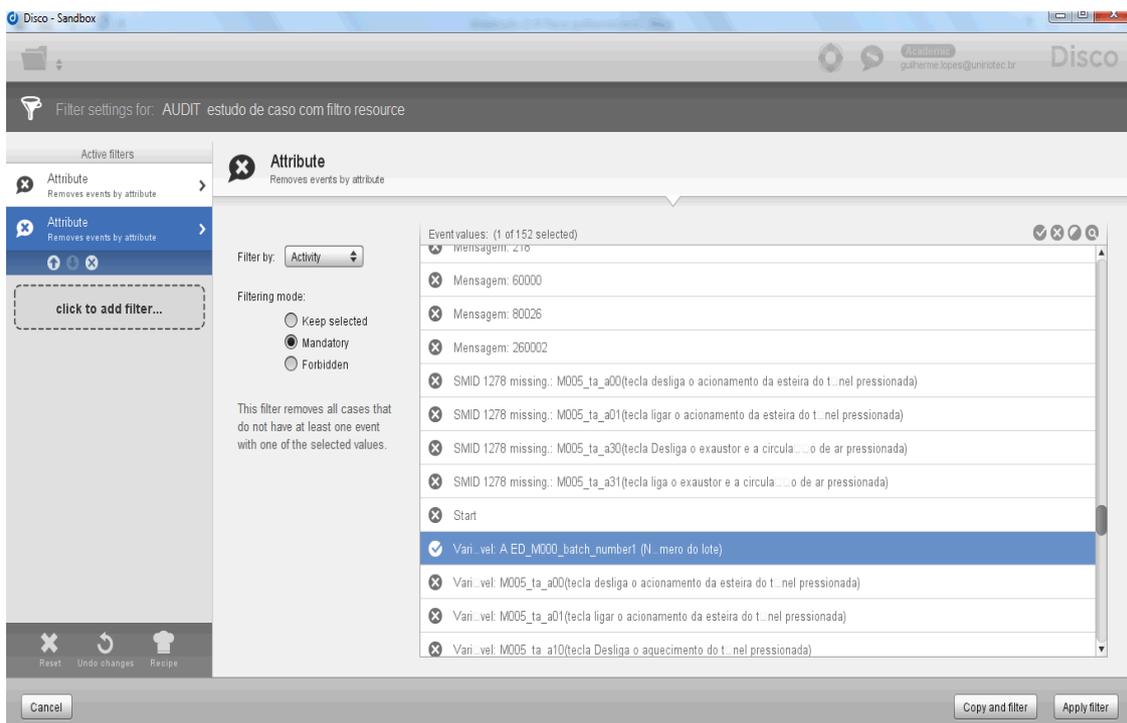


Figura 8-41 - Aplicação correta do filtro para descoberta de instâncias completas sequência 7

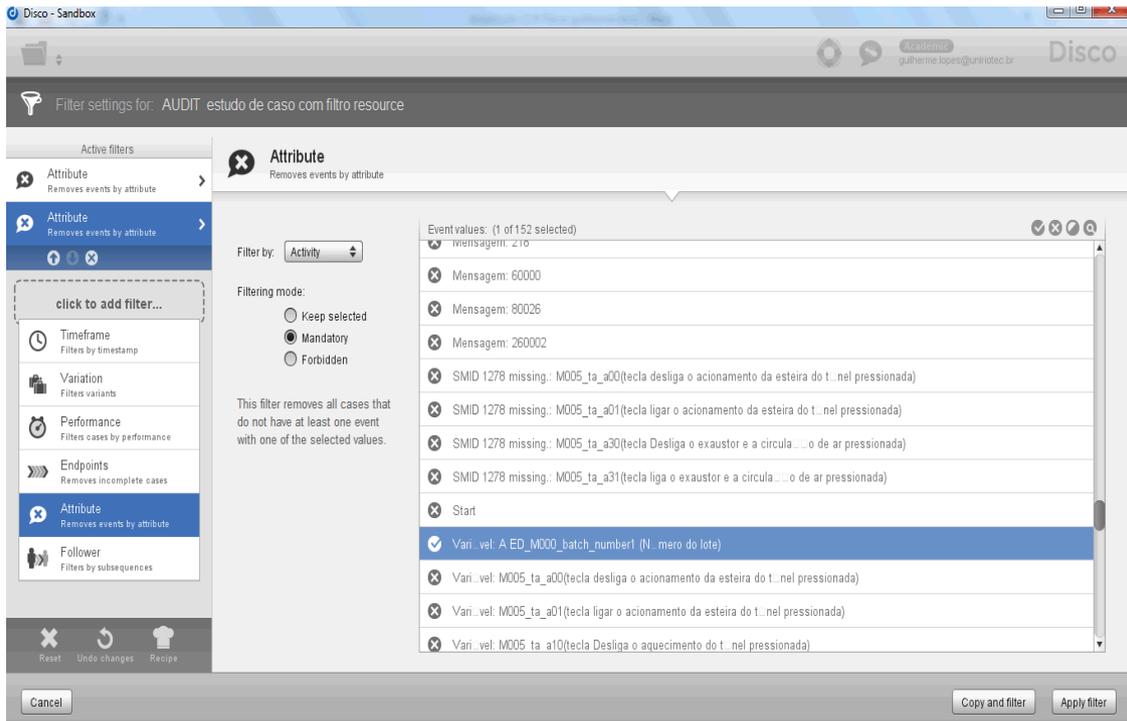


Figura 8-42 - Aplicação correta do filtro para descoberta de instâncias completas sequência 8

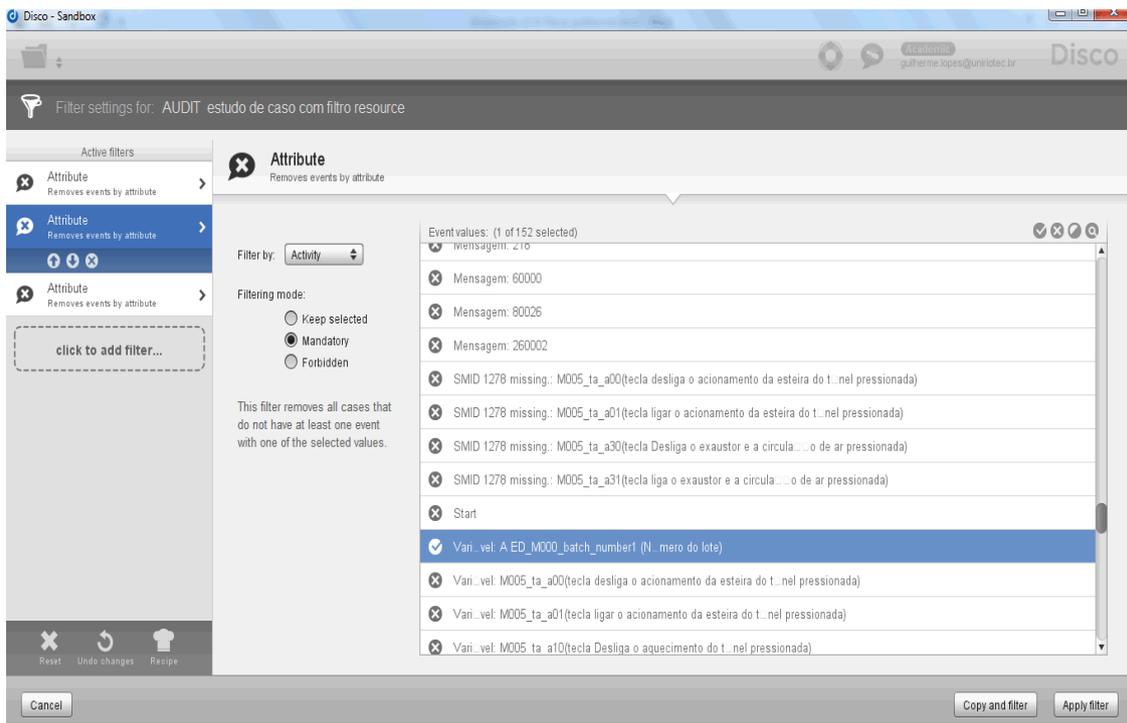


Figura 8-43 - Aplicação correta do filtro para descoberta de instâncias completas sequência 9

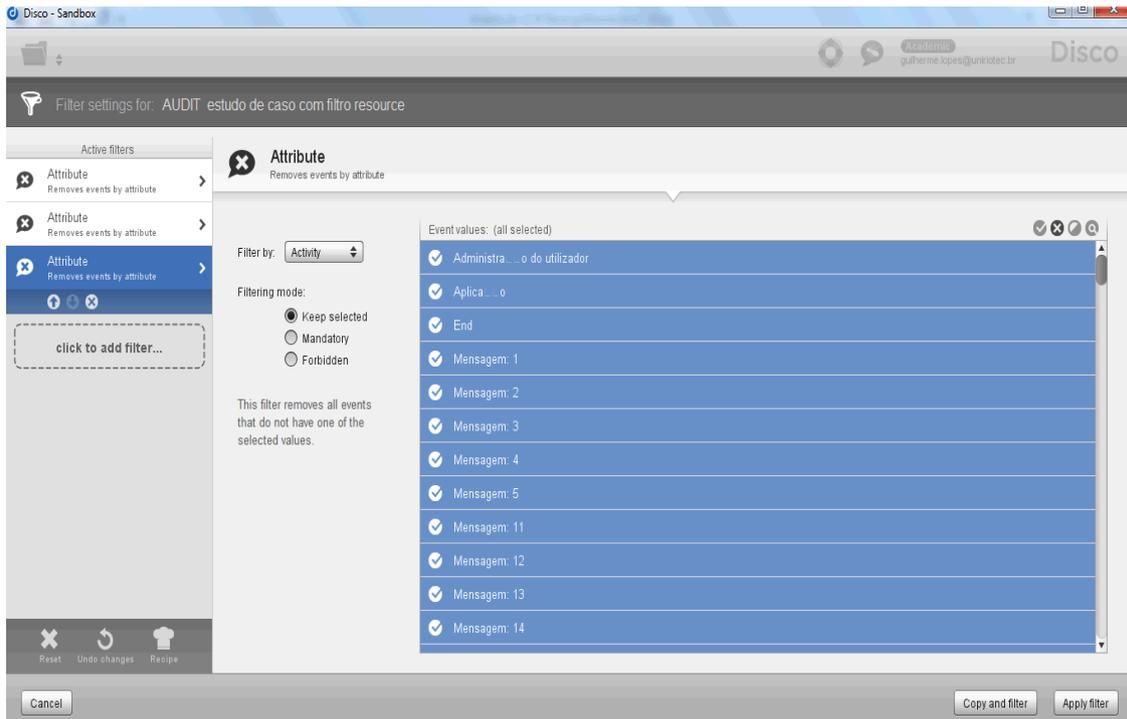


Figura 8-44 - Aplicação correta do filtro para descoberta de instâncias completas sequência 9

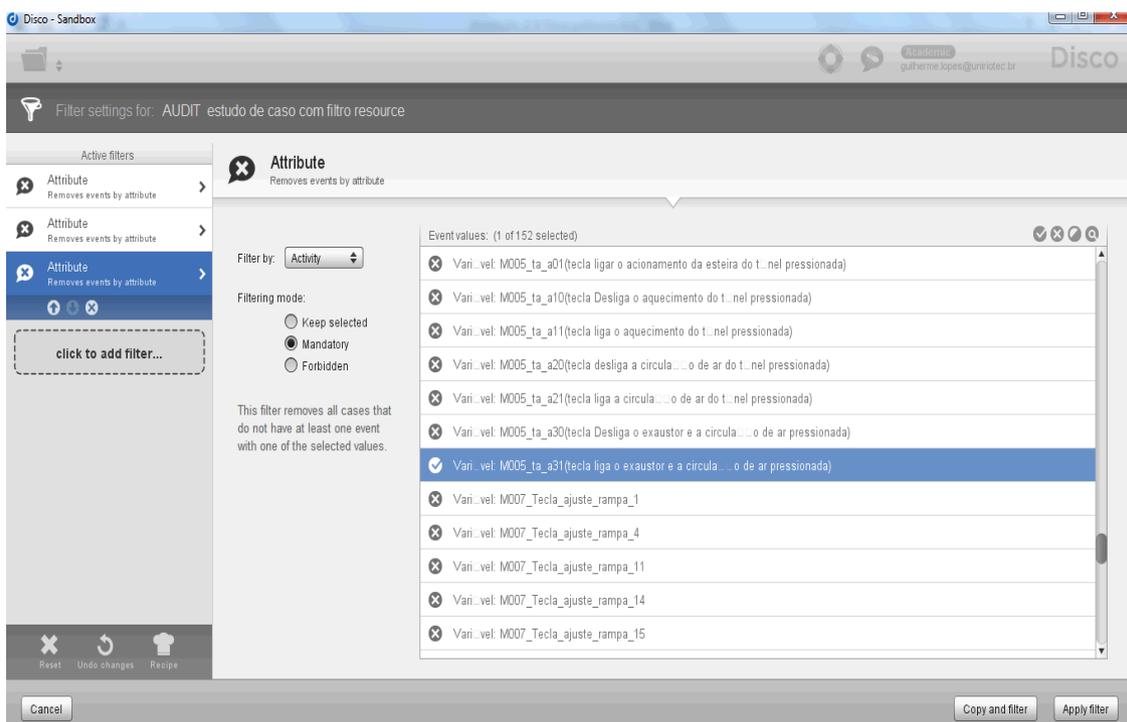


Figura 8-45 - Aplicação correta do filtro para descoberta de instâncias completas sequência 10

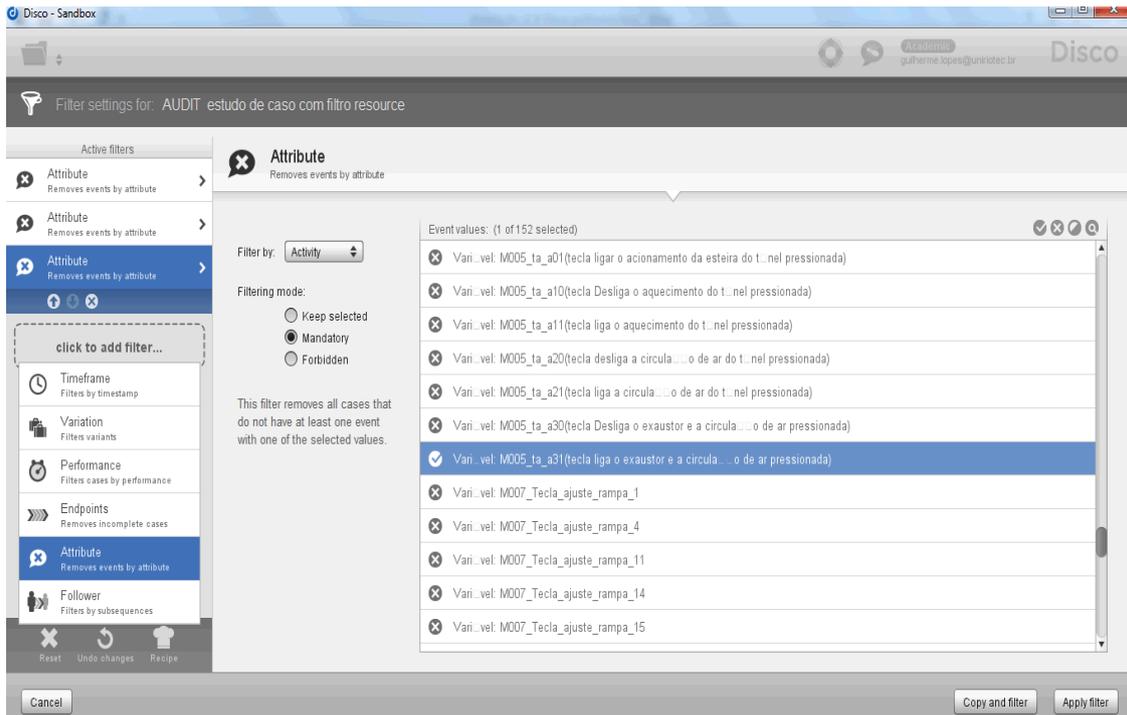


Figura 8-46 - Aplicação correta do filtro para descoberta de instâncias completas sequência 11

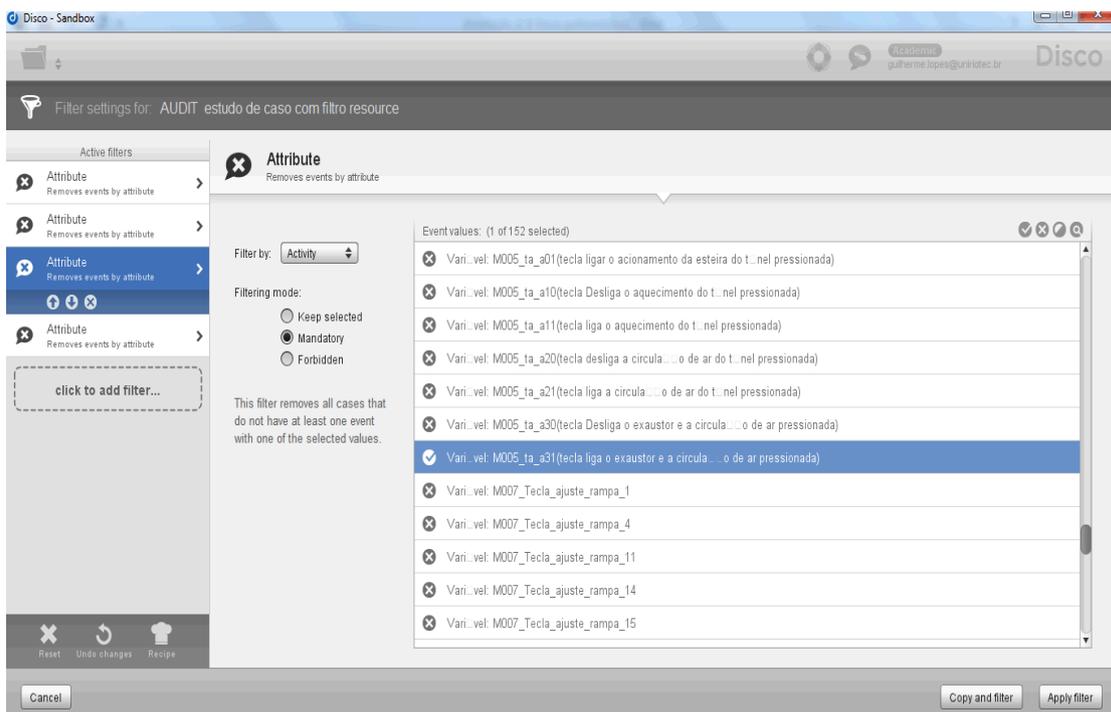


Figura 8-47 - Aplicação correta do filtro para descoberta de instâncias completas sequência 12

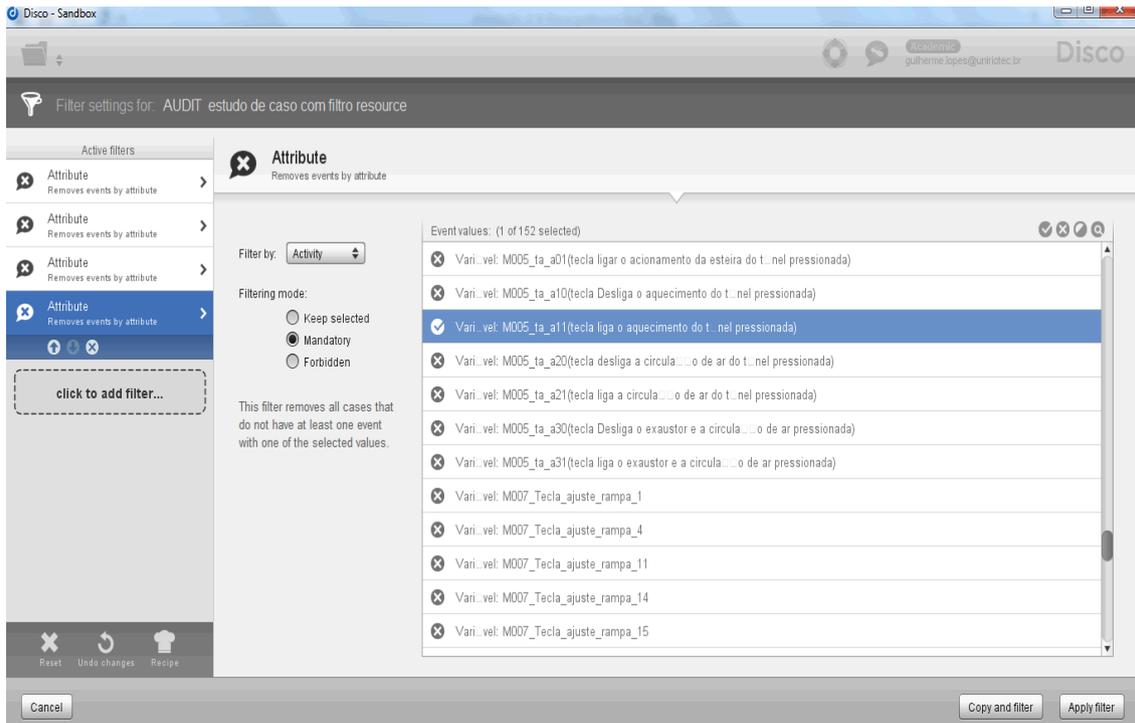


Figura 8-48 - Aplicação correta do filtro para descoberta de instâncias completas sequência 13

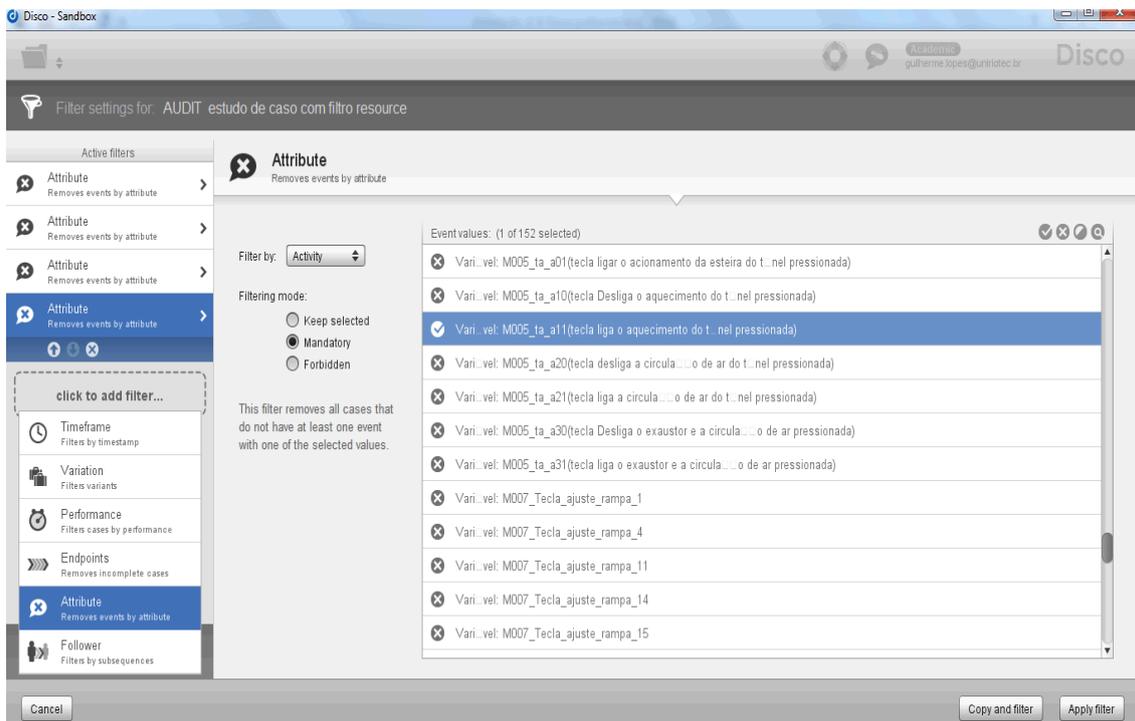


Figura 8-49 - Aplicação correta do filtro para descoberta de instâncias completas sequência 14

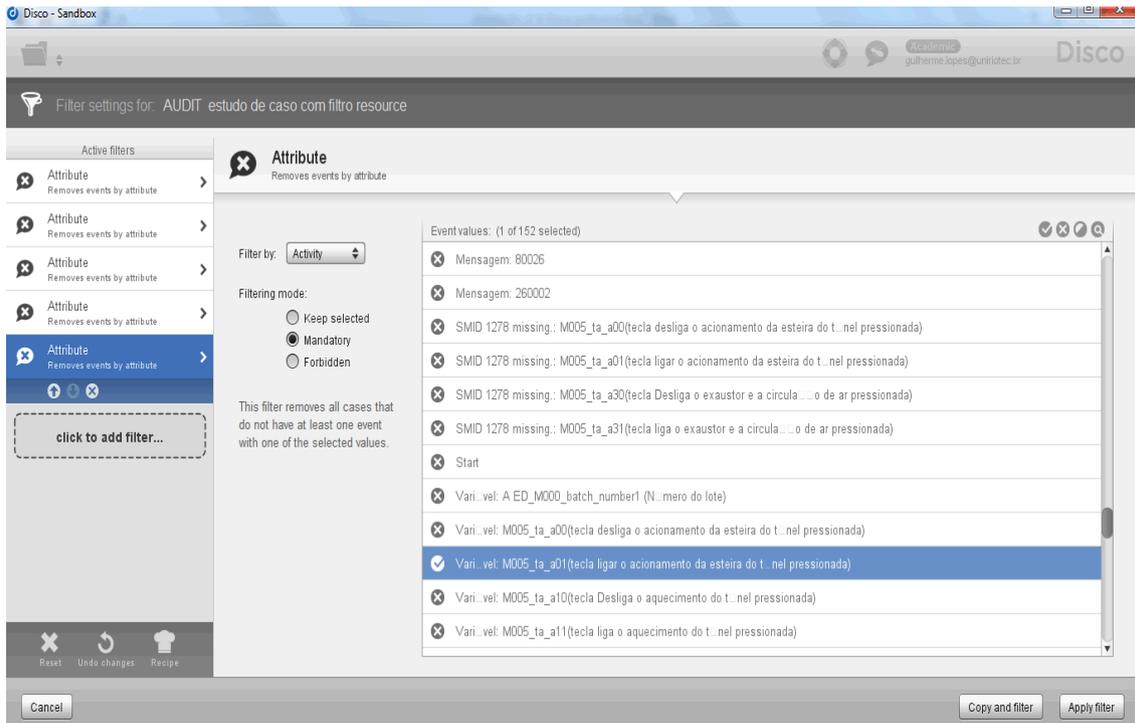


Figura 8-50 - Aplicação correta do filtro para descoberta de instâncias completas sequência 15

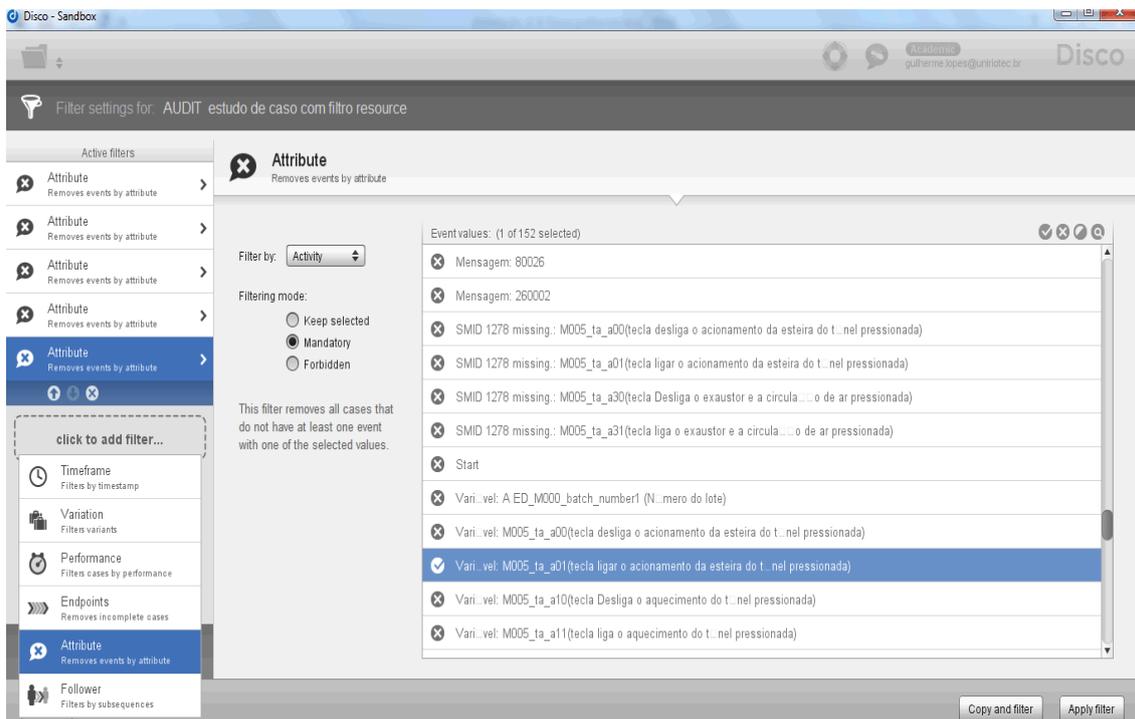


Figura 8-51 - Aplicação correta do filtro para descoberta de instâncias completas sequência 16

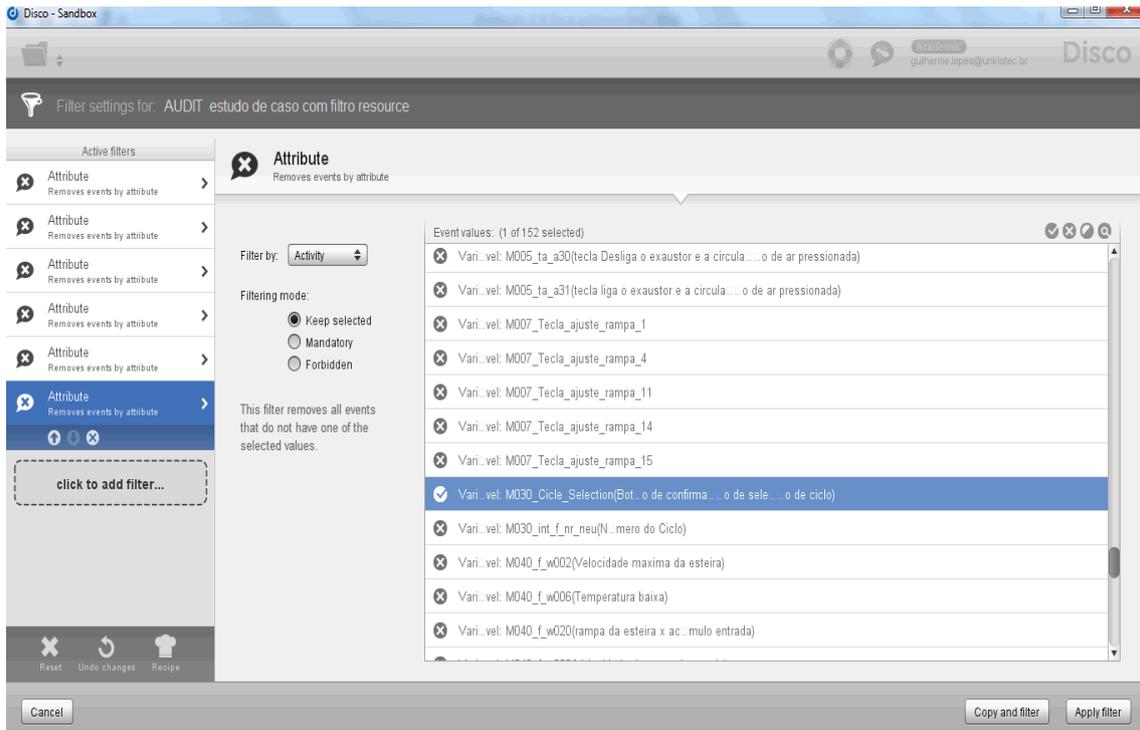


Figura 8-52 - Aplicação correta do filtro para descoberta de instâncias completas sequência 17

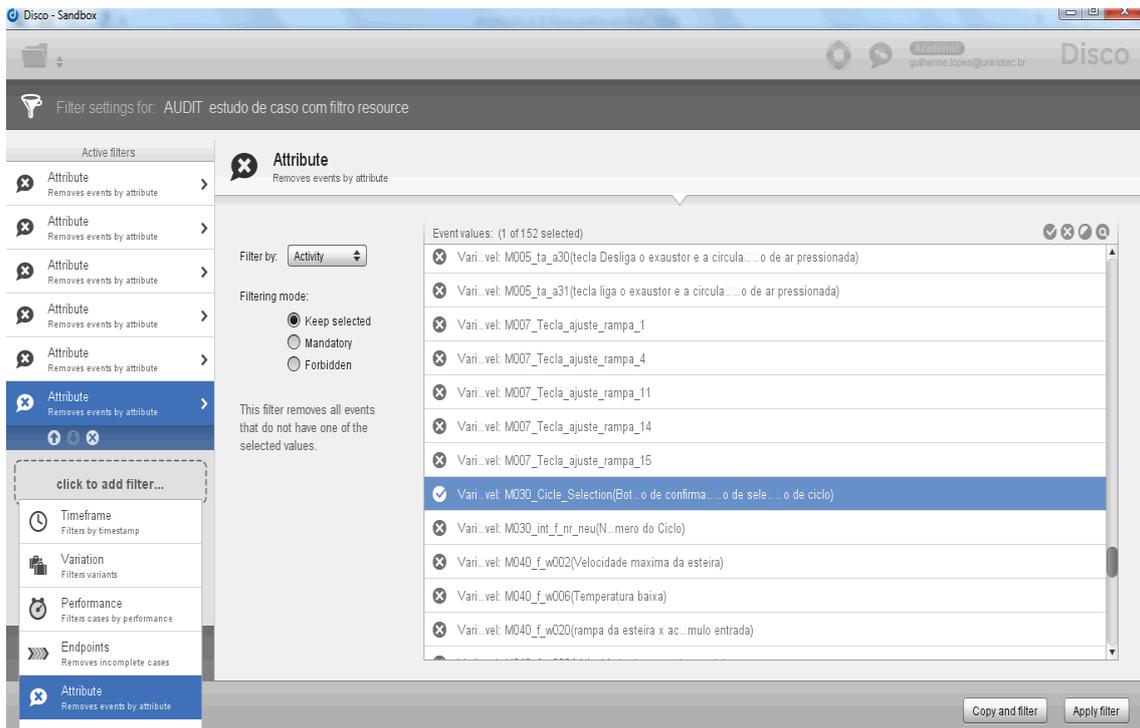


Figura 8-53 - Aplicação correta do filtro para descoberta de instâncias completas sequência 18

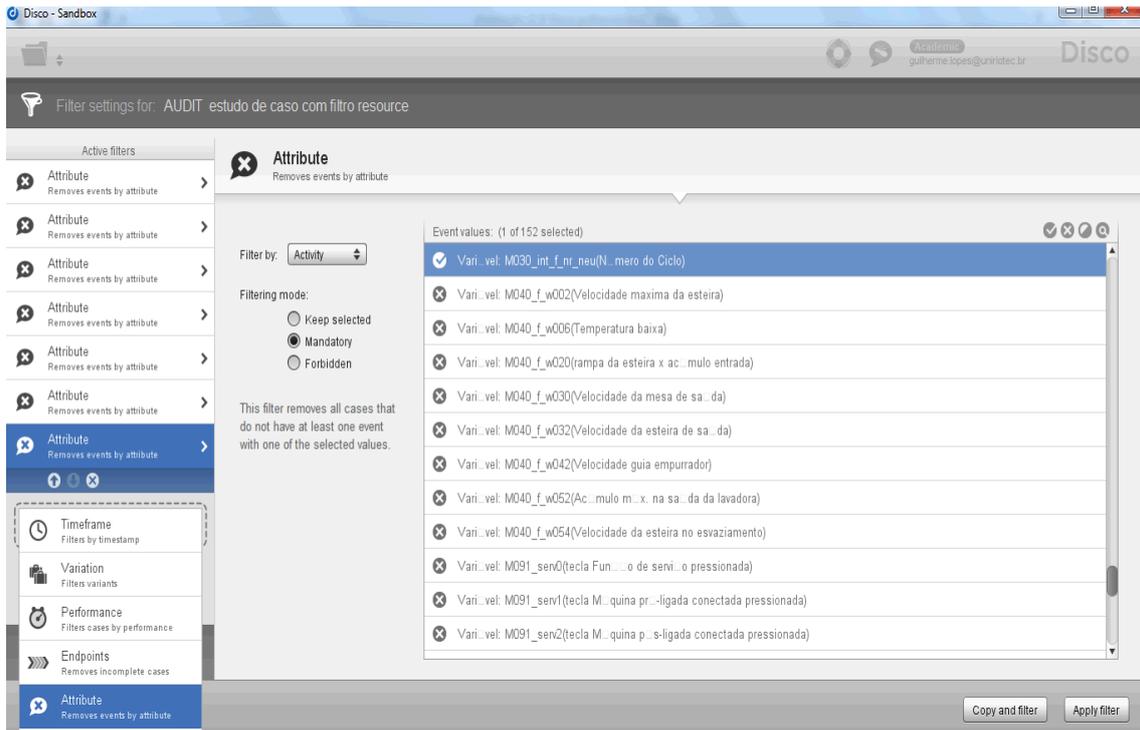


Figura 8-54 - Aplicação correta do filtro para descoberta de instâncias completas sequência 19

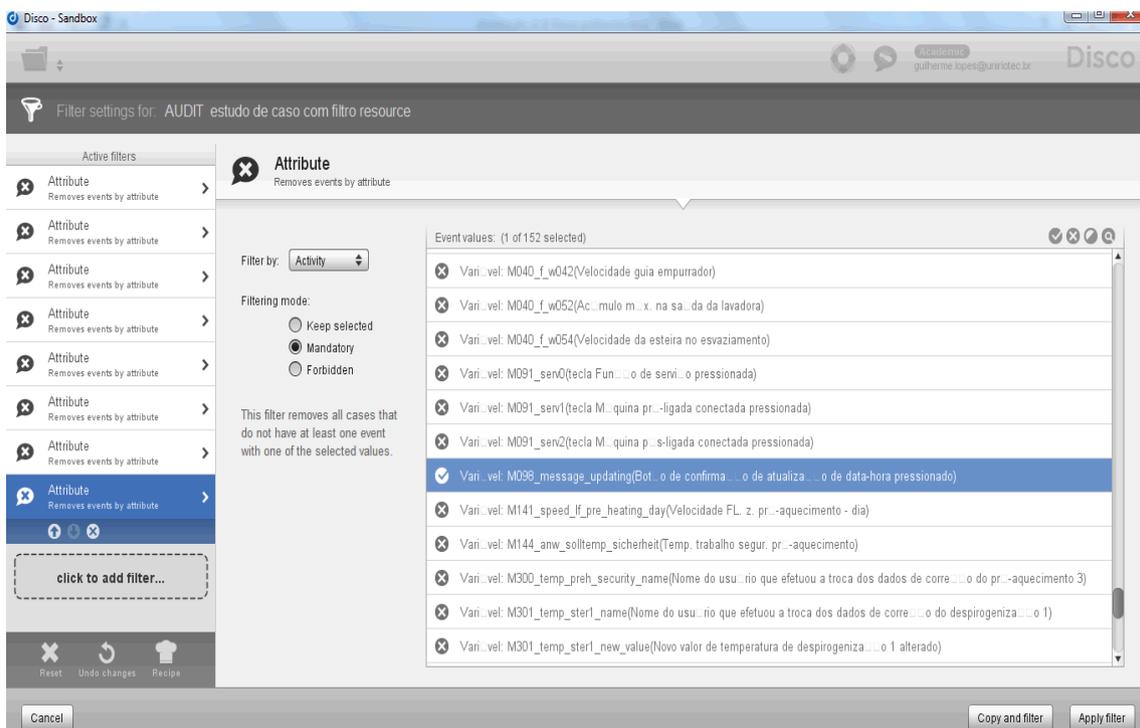


Figura 8-55 - Aplicação correta do filtro para descoberta de instâncias completas sequência 19

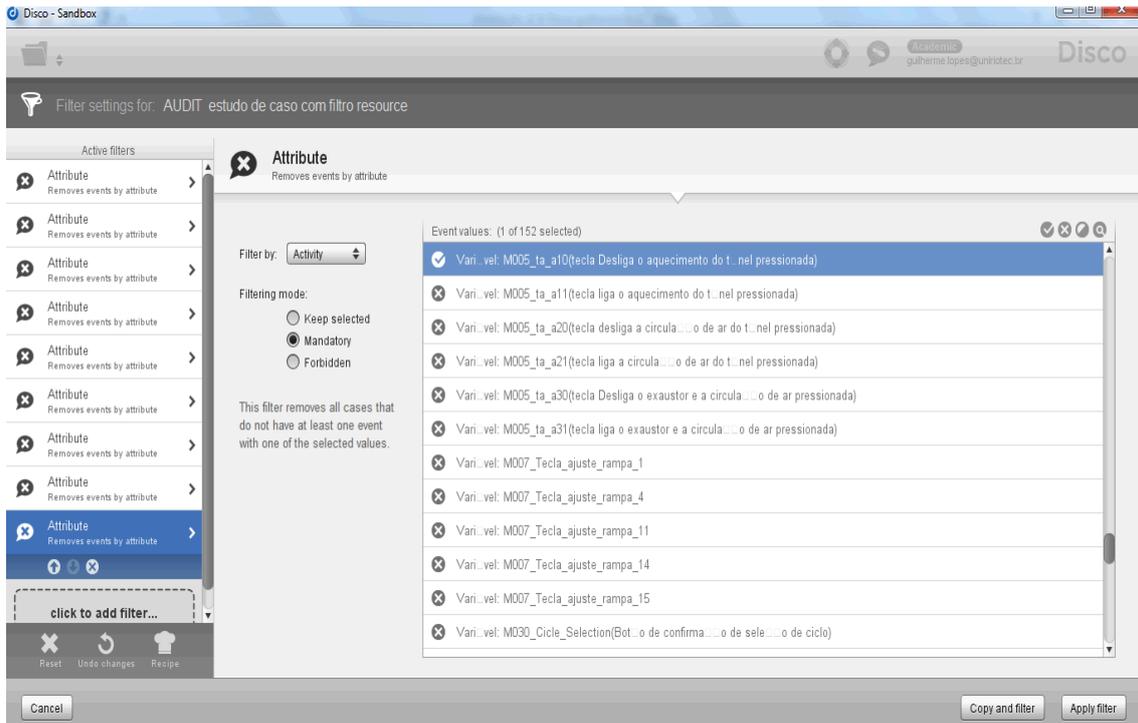


Figura 8-56 - Aplicação correta do filtro para descoberta de instâncias completas sequência 20

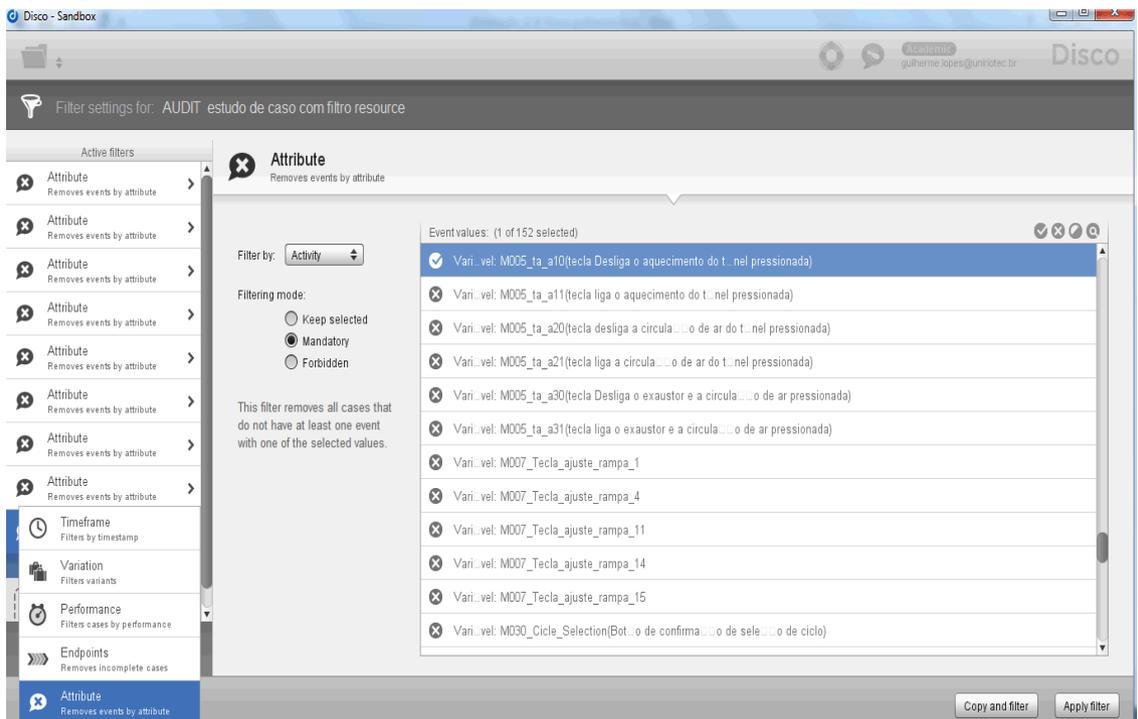


Figura 8-57 - Aplicação correta do filtro para descoberta de instâncias completas sequência 21

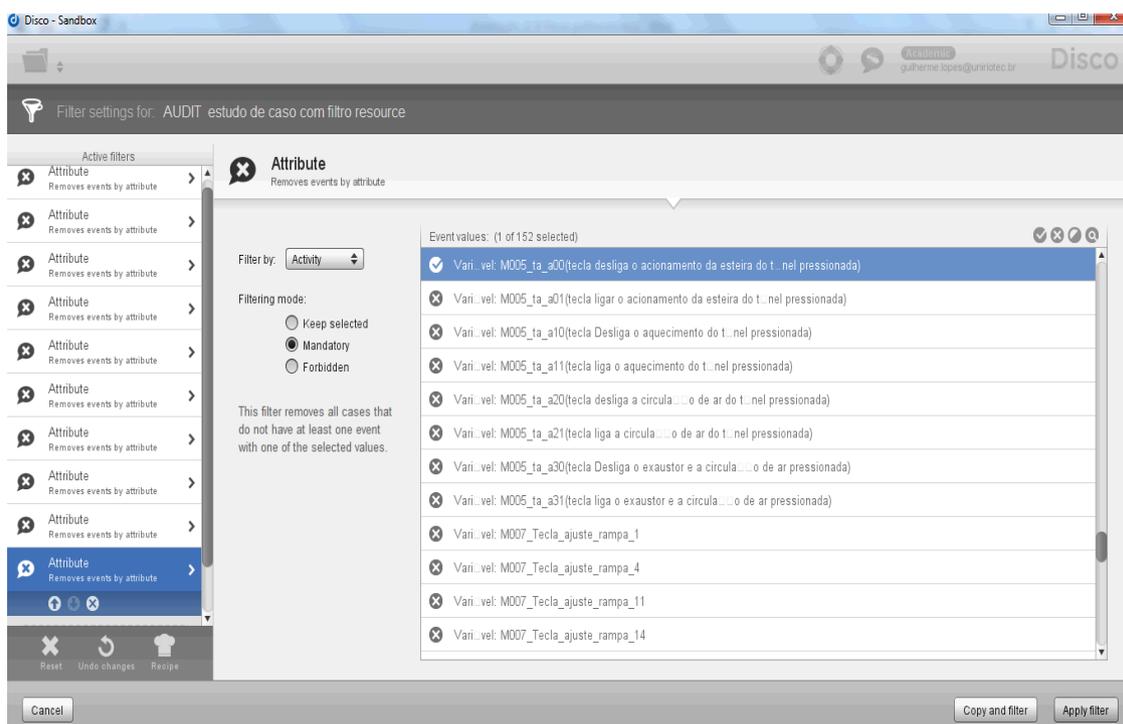


Figura 8-58 - Aplicação correta do filtro para descoberta de instâncias completas sequência 22

Após a seleção da última atividade aplica-se o filtro, neste caso recebemos uma mensagem de erro, indicando que a aplicação do filtro irá gerar um log em branco Figura 8-59 - Aplicação correta do filtro para descoberta de instâncias completas sequência 23. Isto significa que nenhuma instância executou todas as atividades obrigatórias. Isto ocorreu pois as atividades removidas do log fazem parte do processo, e não poderiam ser retiradas. Isto significa que o levantamento de informações sobre o domínio ficou incompleto e terá que ser refeito com nova entrevista ou com suporte de documentação como manuais, especificação funcional ou procedimentos operacionais.

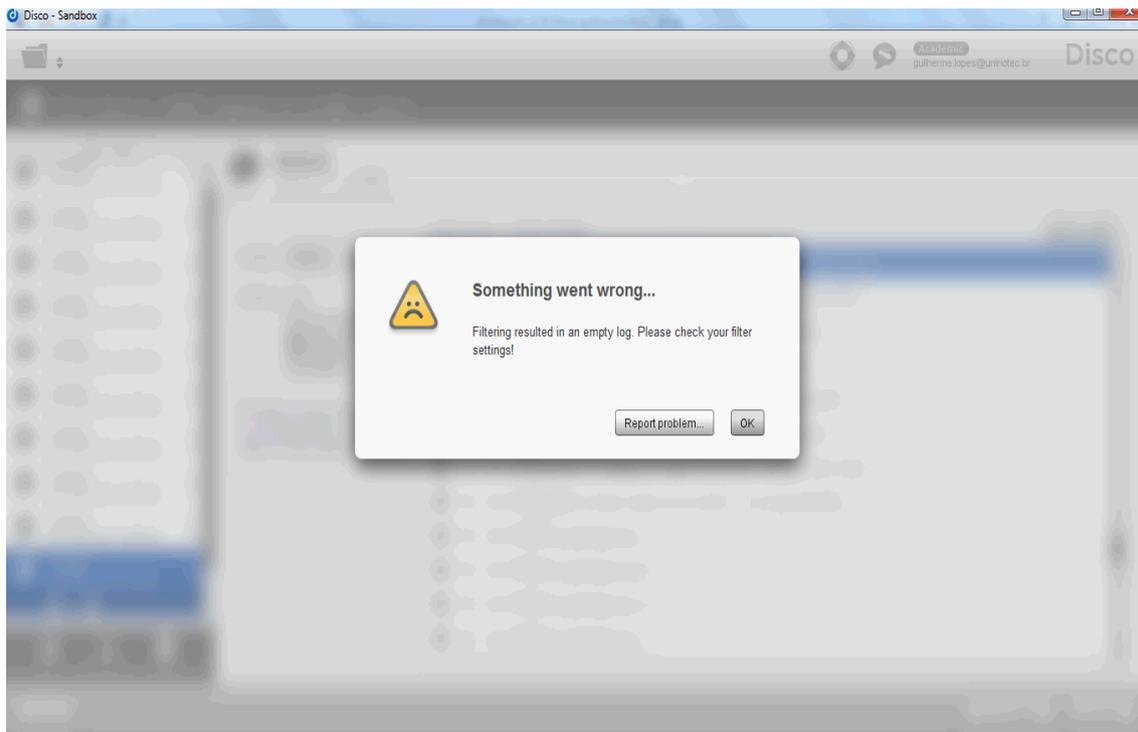


Figura 8-59 - Aplicação correta do filtro para descoberta de instâncias completas sequência 23

8.7. Estudo de caso, Aplicação do filtro de atributo para descoberta de instâncias completas no log original com todas as instâncias.

Para a aplicação do filtro de atributos que tem como objetivo eliminar do log as instâncias incompletas. Como nós já encontramos as instâncias com desvio regulatório, o correto é minerar o log que tenha apenas instâncias que sejam ou conformes ou com desvios não regulatórios, para isto geramos um novo log eliminando todas as instâncias com desvios regulatórios aplicando a regra :

Se valor do evento é Mensagem 141

Então elimina a instância do Log

A partir do log original com todas as instâncias aplicamos o filtro de atributo para eliminar as instâncias com atividades proibidas conforme mostrado na Figura 8-60 - Filtro para eliminar atividades proibidas. Foi utilizado o filtro de atributo no modo Forbidden, este modo elimina todas as instâncias que possuem a atividade selecionada.

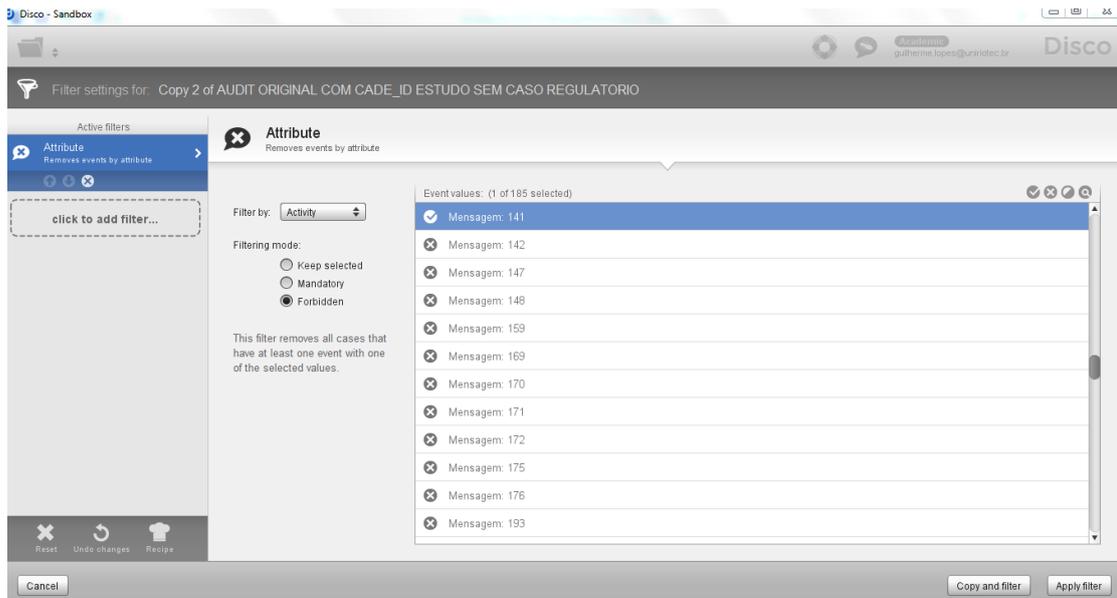


Figura 8-60 - Filtro para eliminar atividades proibidas

No menu estatístico observa-se que restaram 541 instâncias (cases), esta instância não contém nenhuma atividade proibida.

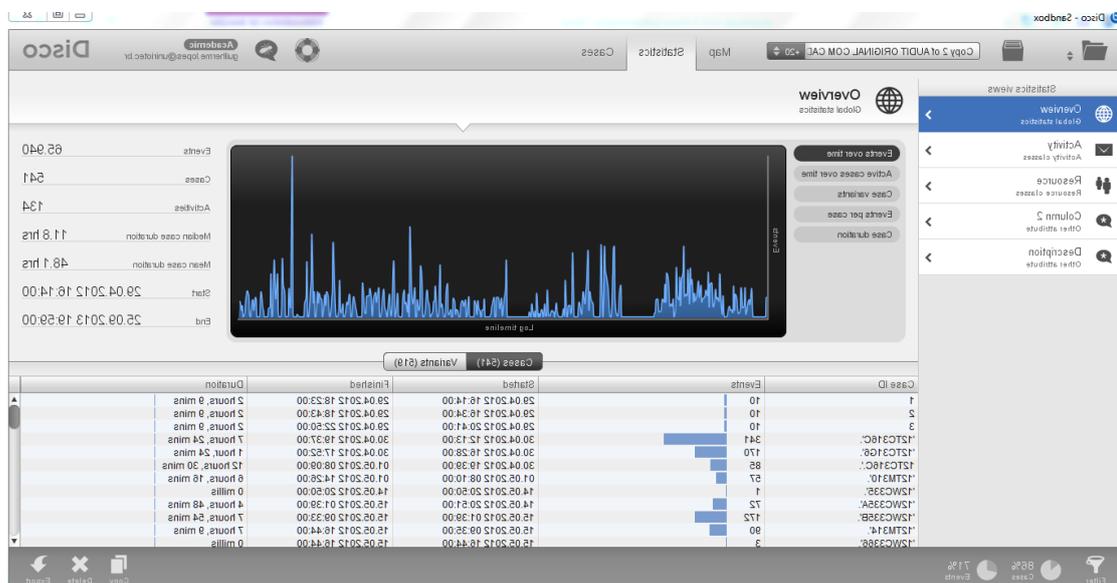


Figura 8-61 - Estatísticas após a aplicação do filtro

Este log foi exportado gerando um novo arquivo de log sem nenhuma instância com atividade proibida, ou seja sem desvio regulatório. O arquivo de log foi gerado para que não haja confusão na hora de aplicar os filtros para instâncias completa, facilitando o entendimento e diminuindo o processamento. O arquivo foi

gerado e recebeu o nome Copy 2 of AUDIT ORIGINAL COM CADE_ID ESTUDO SEM CASO REGULATORIO.csv.

Para a etapa de descoberta das instâncias completas foi importado o novo log com mostram as Figura 8-62 a Figura 8-69.

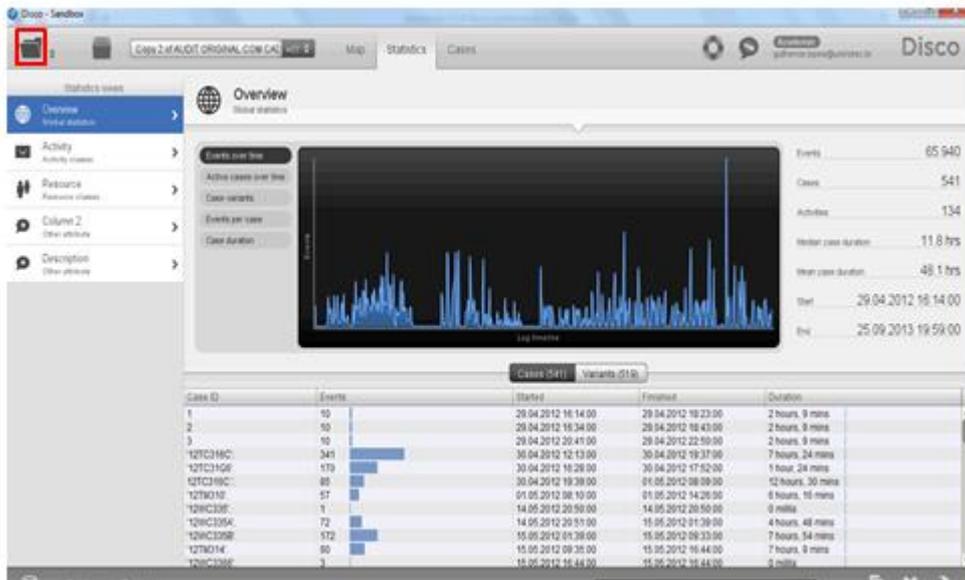


Figura 8-62 - Carregando um novo log

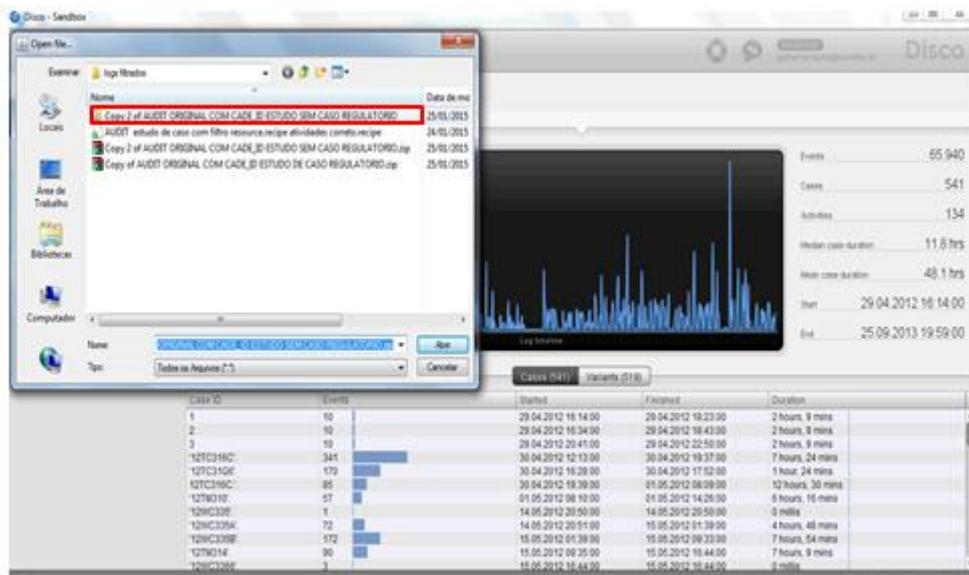


Figura 8-63 - Selecionando a pasta para importação

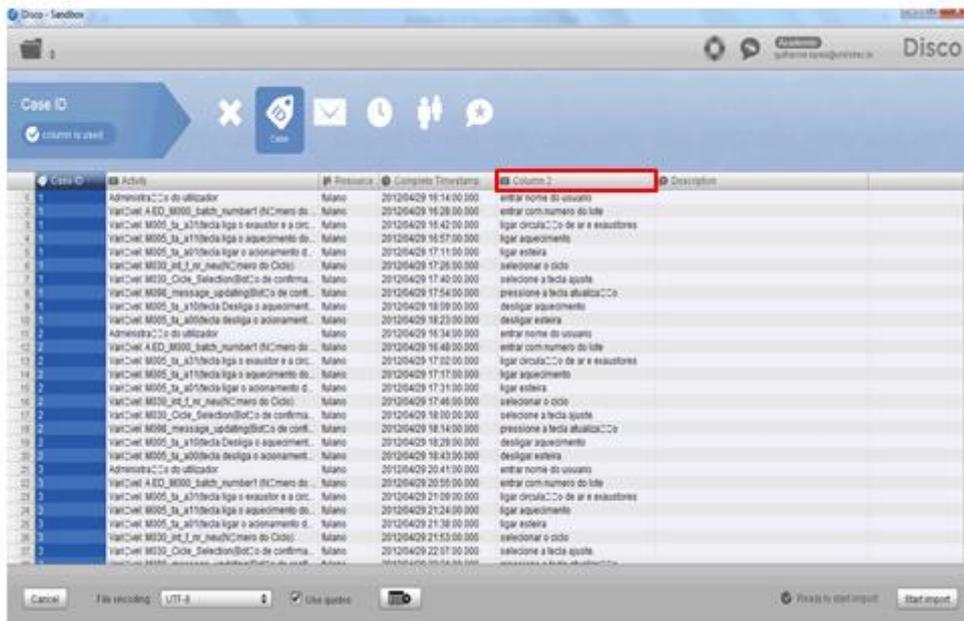


Figura 8-66 - De-para Case_id

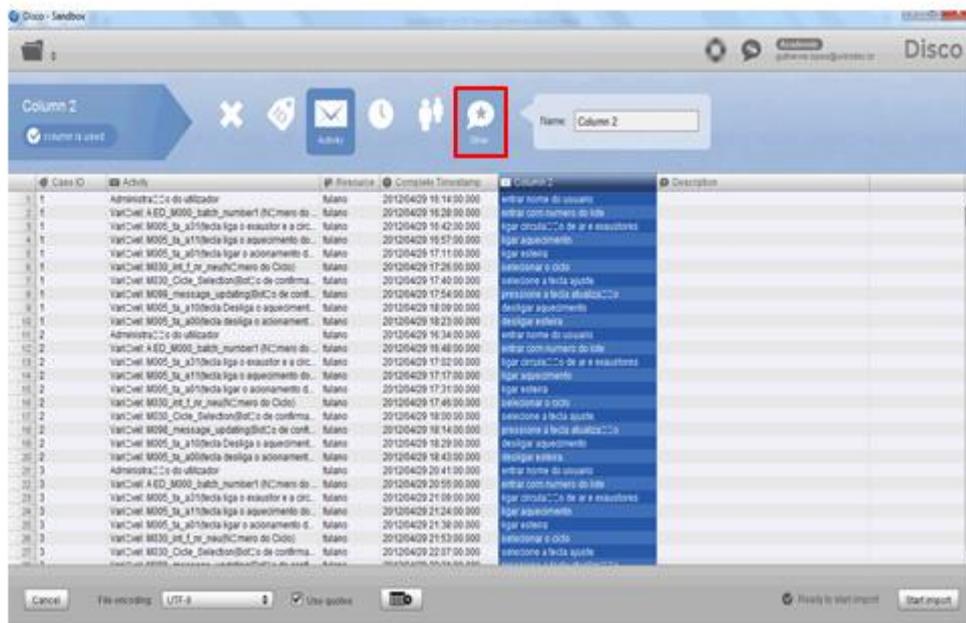


Figura 8-67 - De-para outros atributos

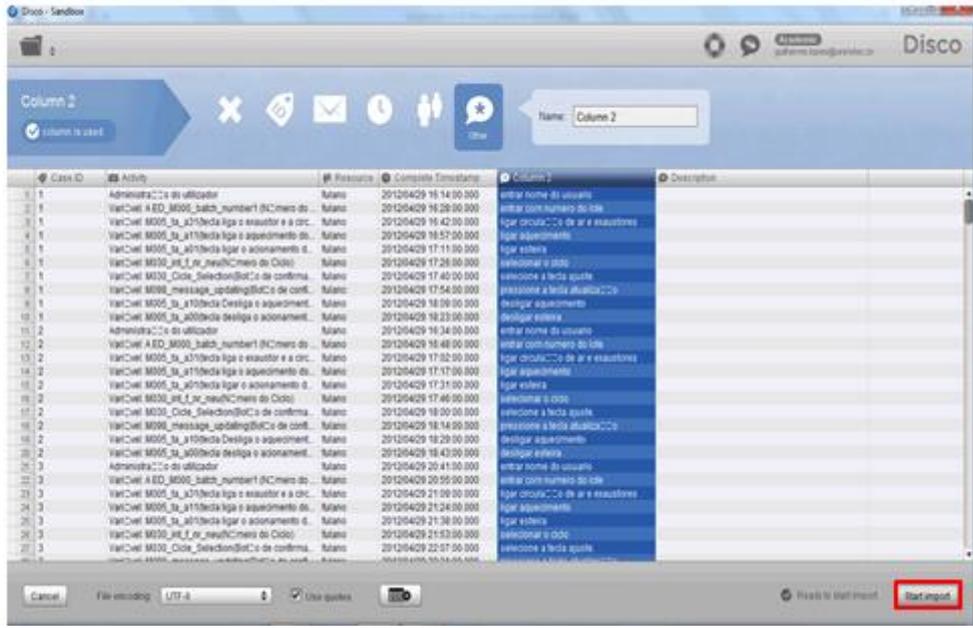


Figura 8-68 - Importando o log

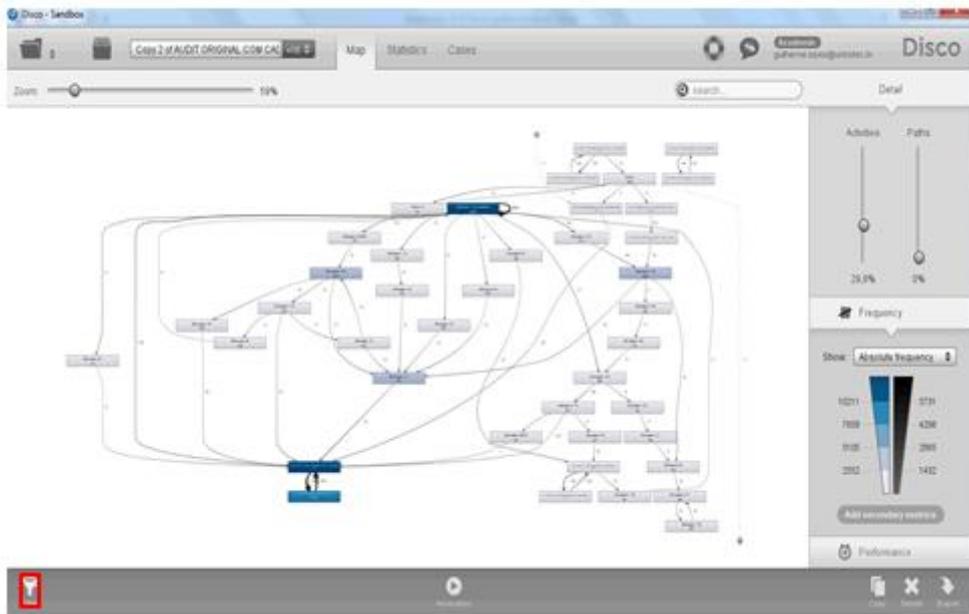


Figura 8-69 - Modelo de processo minerado

O filtro aplicado de forma encadeada conforme mostrado na sequencia abaixo atende ao seguinte critério:

Filtro = Verdadeiro **SE** a E b E c E d... z **SENÃO** falso. Onde a, b, c.. z são as atividades selecionadas.

Nas Figura 8-70 - Adicionando o filtro a Figura 8-83 - Aplicando o filtro são mostrados os passos da construção do filtro, a cada adição seleciona-se a atividade que vai ser filtrada, somente no final o filtro é aplicado.

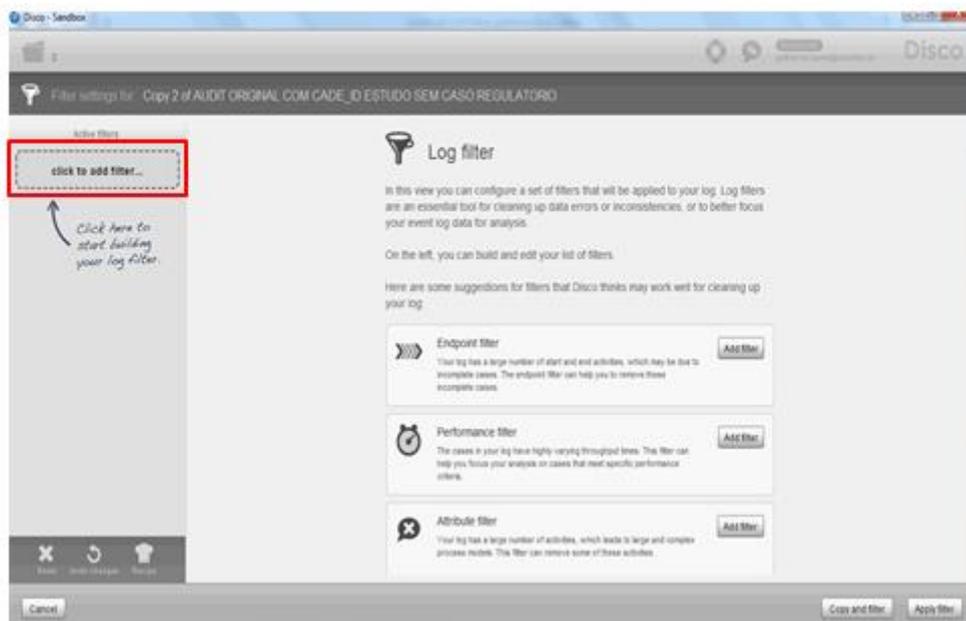


Figura 8-70 - Adicionando o filtro

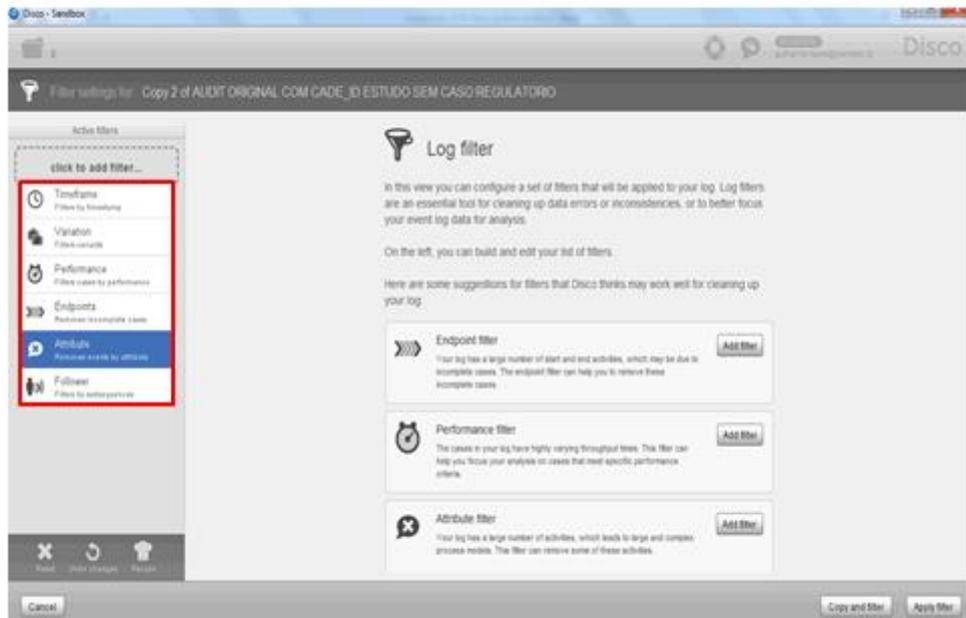


Figura 8-71 - Escolhendo o tipo de filtro

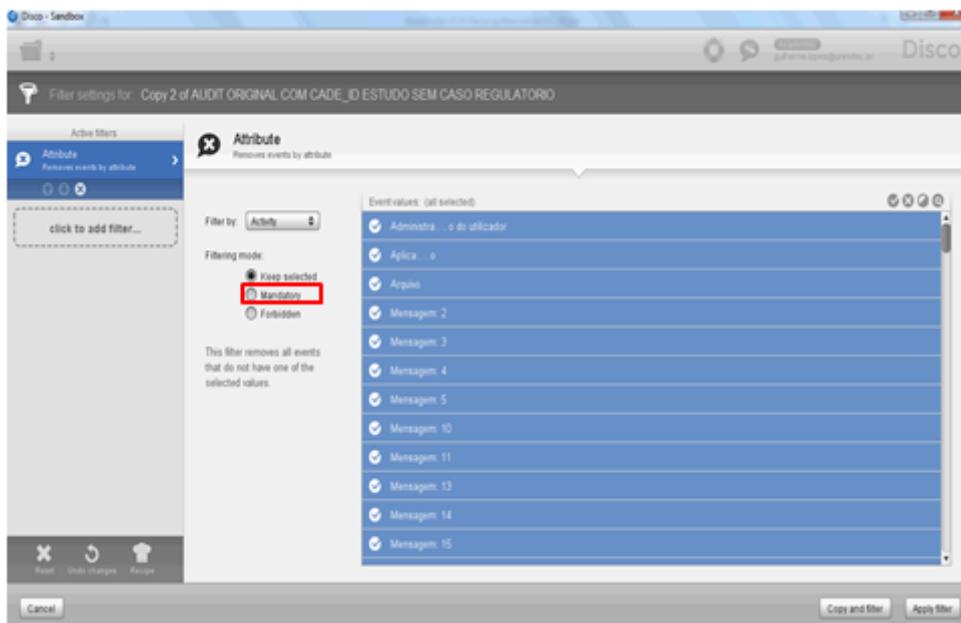


Figura 8-72 - Selecionando o modo do filtro

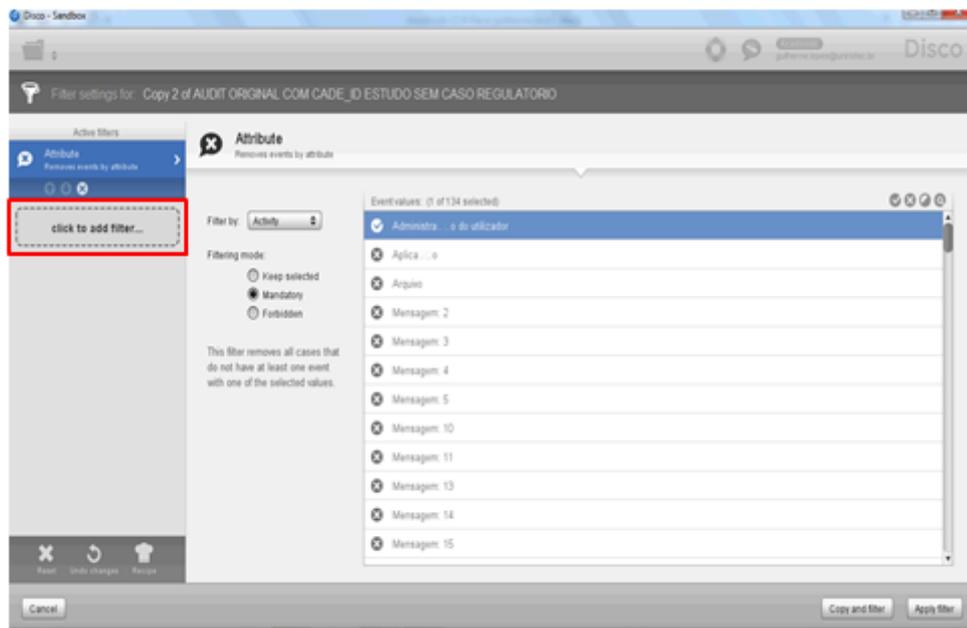


Figura 8-73 - Selecionando a primeira atividade
 Administrador do utilizador

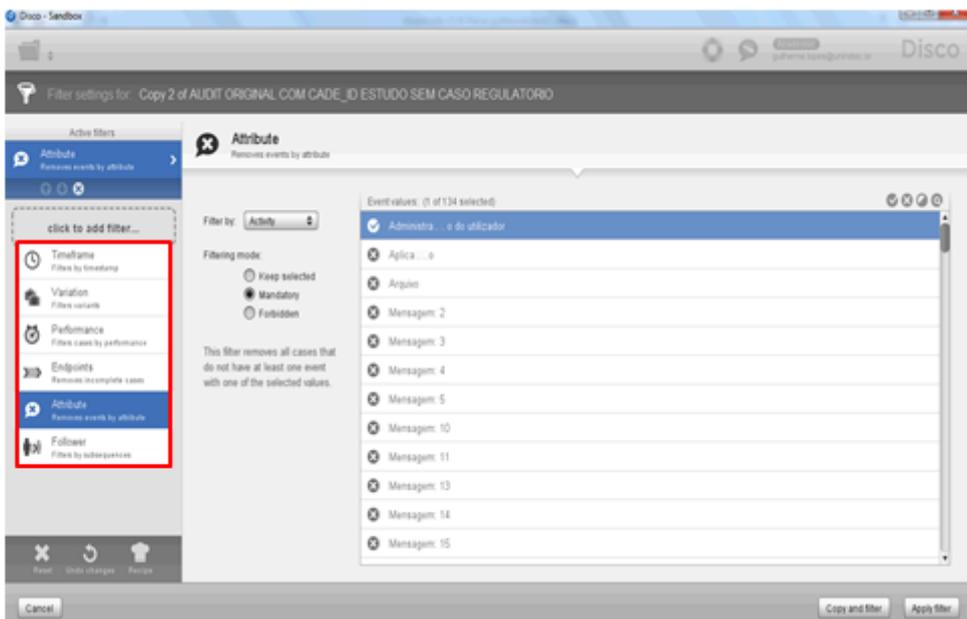


Figura 8-74 - Selecionando a segunda atividade
 Variável: A ED_M000_batch_number1 (Número do lote)

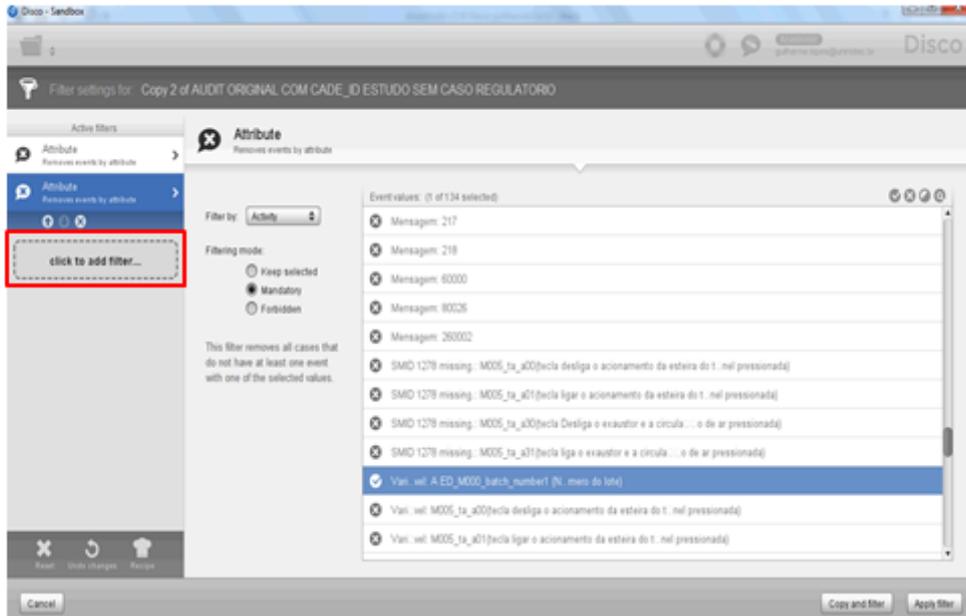


Figura 8-75 - Selecionando a terceira atividade

Variável: M005_ta_a31(tecla liga o exaustor e a circulação de ar pressionada)

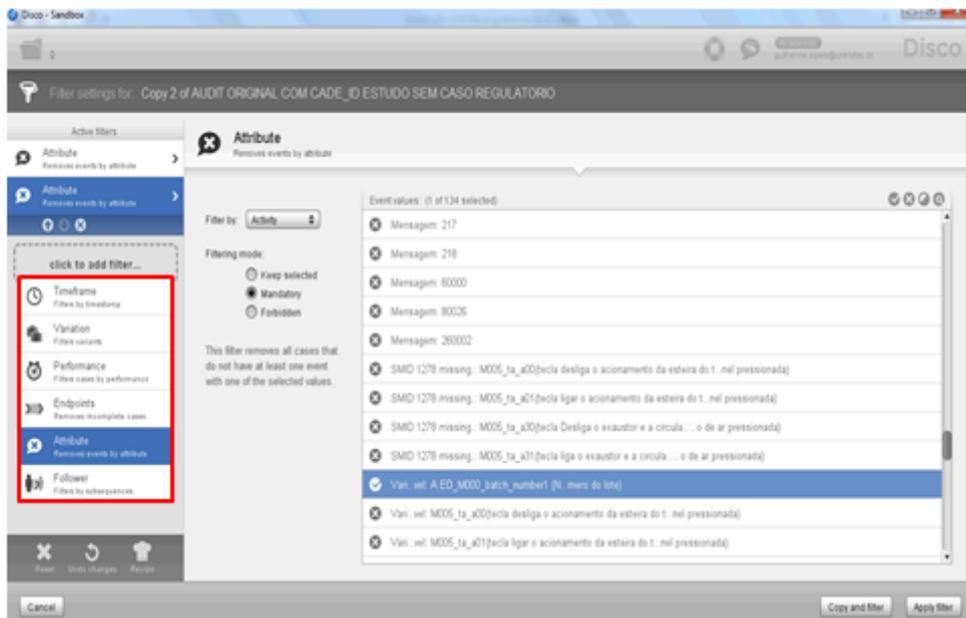


Figura 8-76 - Selecionando a quarta atividade

Variável: M005_ta_a11(tecla liga o aquecimento do túnel pressionada)

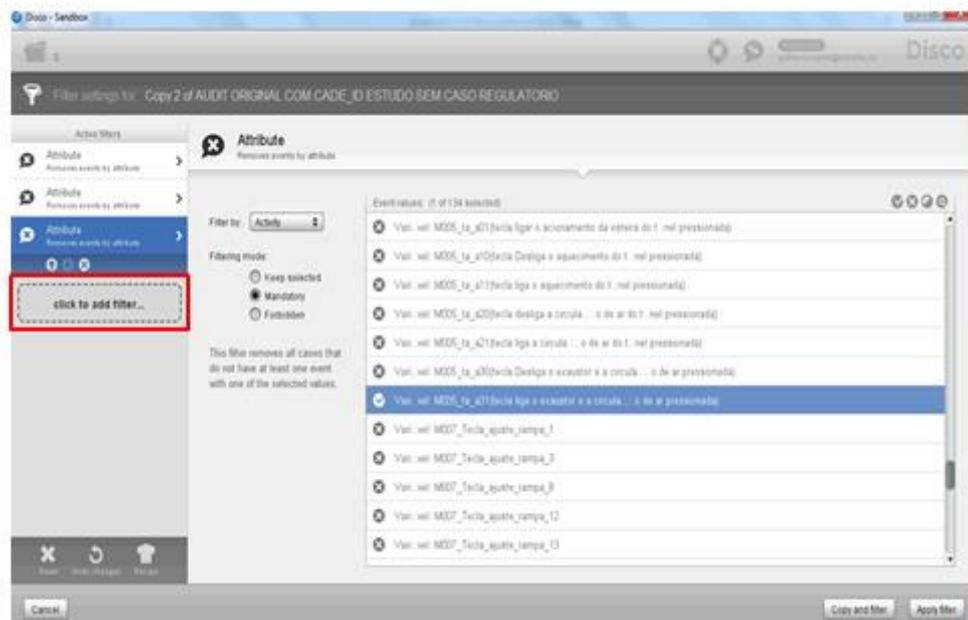


Figura 8-77 - Selecionando a quinta atividade

Variável: M005_ta_a01(tecla ligar o acionamento da esteira do túnel pressionada)

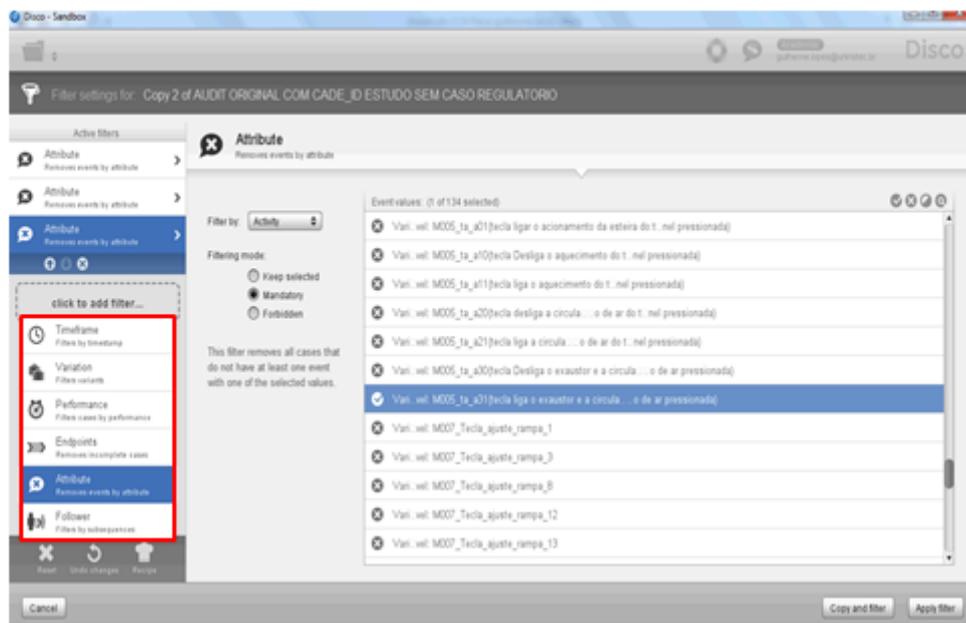


Figura 8-78 - Selecionando a sexta atividade

Variável: M030_int_f_nr_neu(Número do Ciclo)

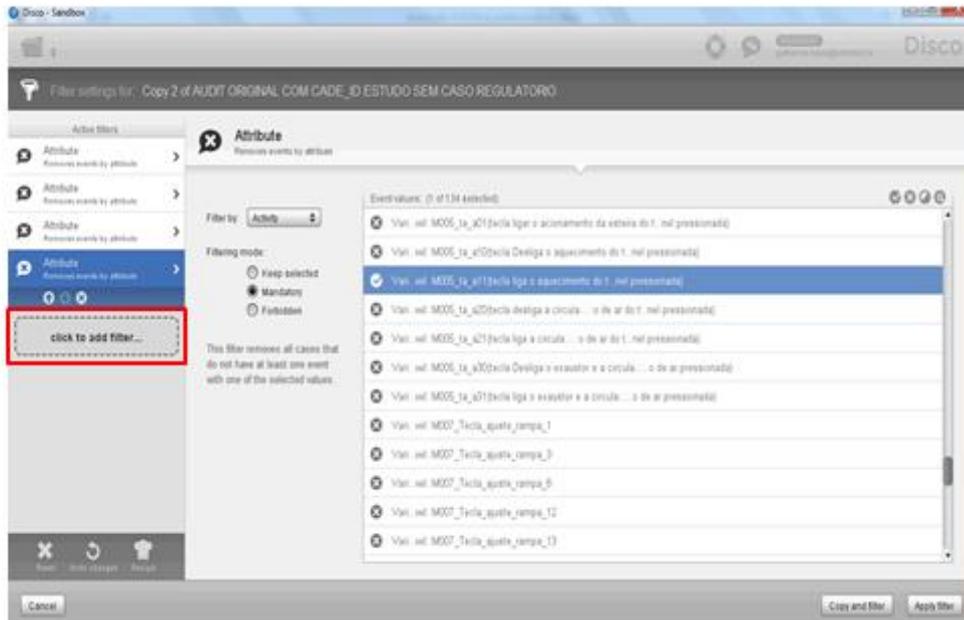


Figura 8-79 - Selecionando a sétima atividade

Variável: M030_Cicle_Selection(Botão de confirmação de seleção de ciclo)

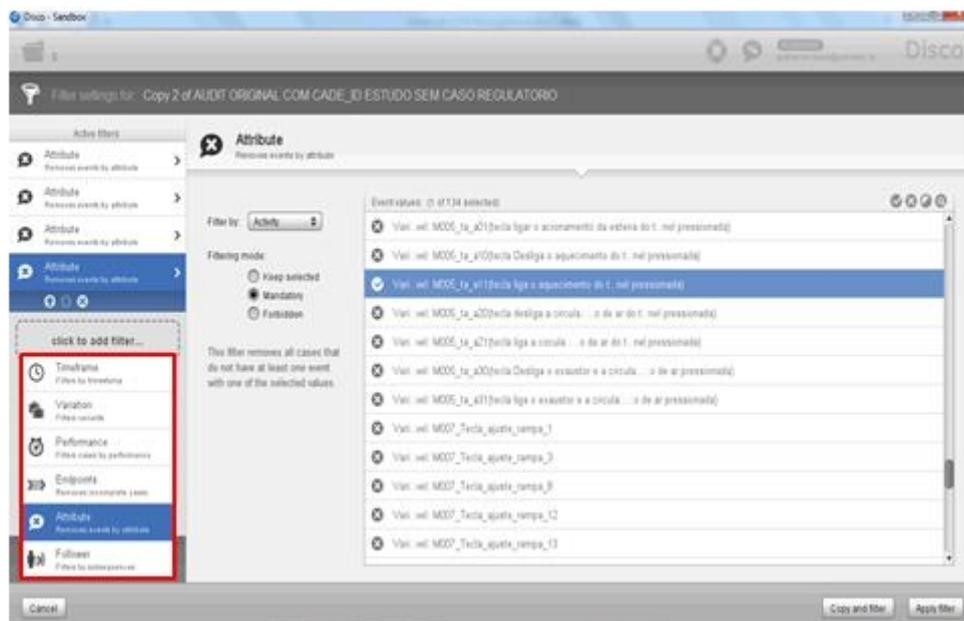


Figura 8-80 - Selecionando a oitava atividade

Variável: M098_message Updating(Botão de confirmação de atualização de data-hora pressionado)

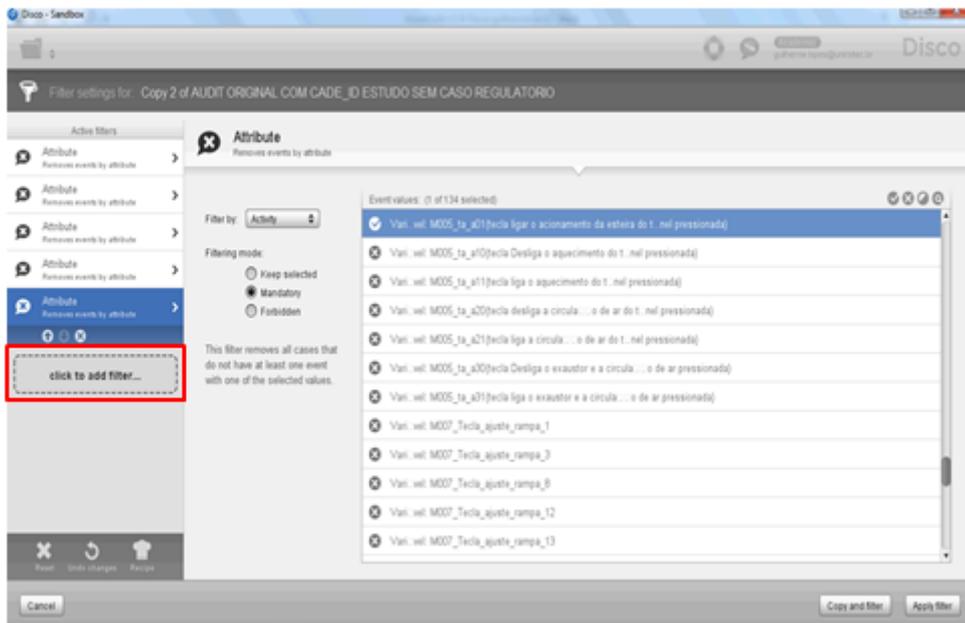


Figura 8-81 - Selecionando a nona atividade

Variável: M005_ta_a10(tecla Desliga o aquecimento do túnel pressionada)

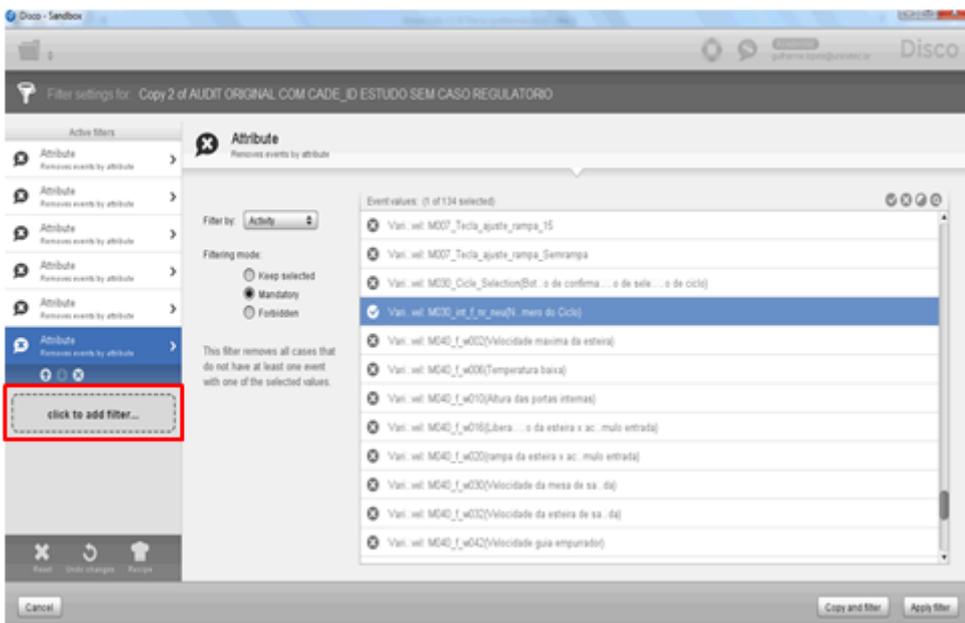


Figura 8-82 - Selecionando a última atividade

Variável: M005_ta_a00(tecla desliga o acionamento da esteira do túnel pressionada)

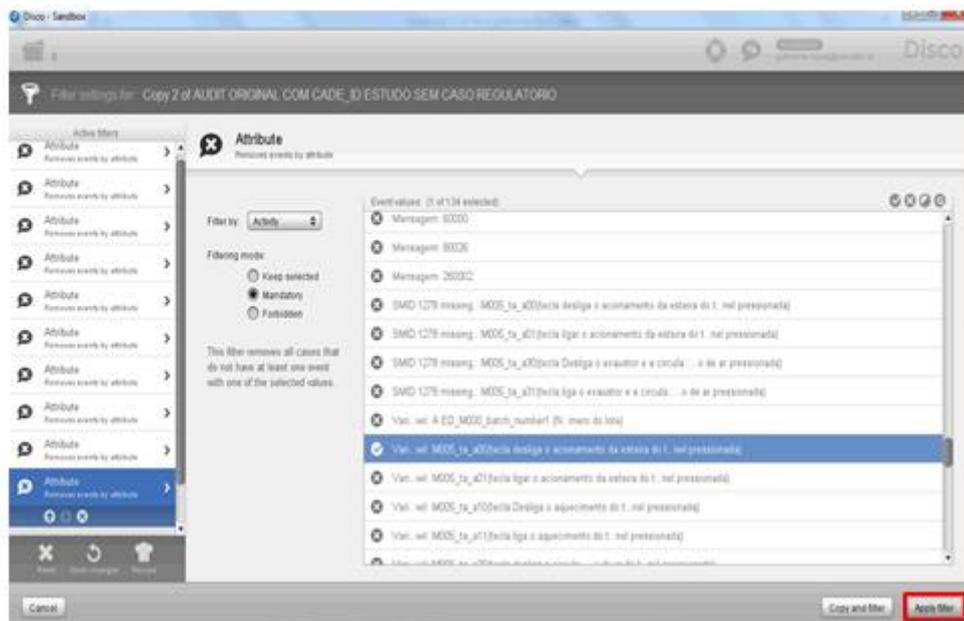


Figura 8-83 - Aplicando o filtro

A partir do modelo de processo gerado pode-se verificar que ele se está muito mais simples, e que não contém nenhuma atividade proibida. Isto significa que os produtos fabricados por estas instâncias tem mais chances de estarem próprios para o consumo, no entanto todas as variantes que fora da instância padrão executaram atividades desnecessárias, indicando falha no equipamento, falta de treinamento do operador ou problemas de manutenção. Estas informações podem contribuir para a melhoria do processo.

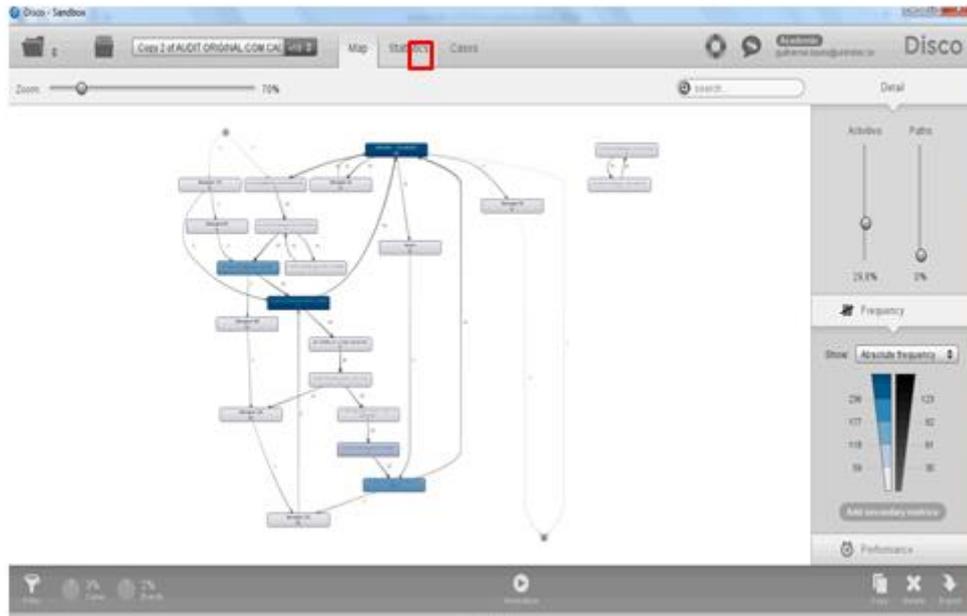


Figura 8-84 - Modelo de processo gerado após aplicação do filtro

Após a aplicação do filtro verifica-se que apenas 20 instâncias (cases) estão completos e não possuem atividades proibidas, o que está de acordo com as informações do especialista do domínio que informou que o log fornecido corresponde ao primeiro ano de implantação do sistema onde os ajustes e treinamento dos operadores estava sendo realizado.

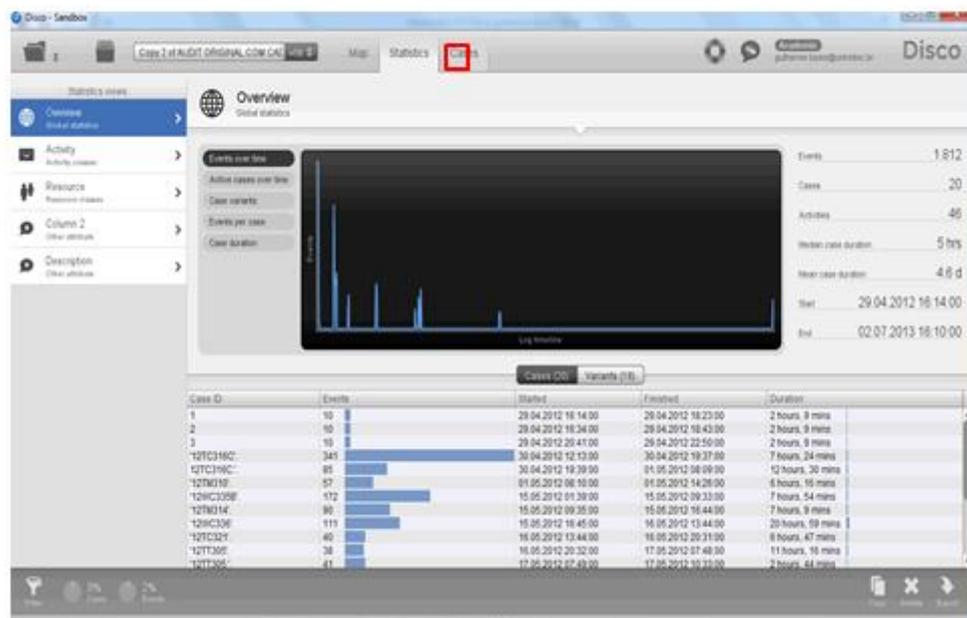


Figura 8-85 – Estatísticas após a aplicação do filtro

Pode-se verificar cada variante e os tipos de instâncias agrupadas com informações detalhadas de atividades, recursos (resources), hora de execução (timestamp), como esta variantes contém instâncias semelhantes, a análise de problema pode ser feito na variante diminuindo muito o trabalho do especialista do domínio. Como mostrado na Figura 8-86 - Instâncias Resultantes.

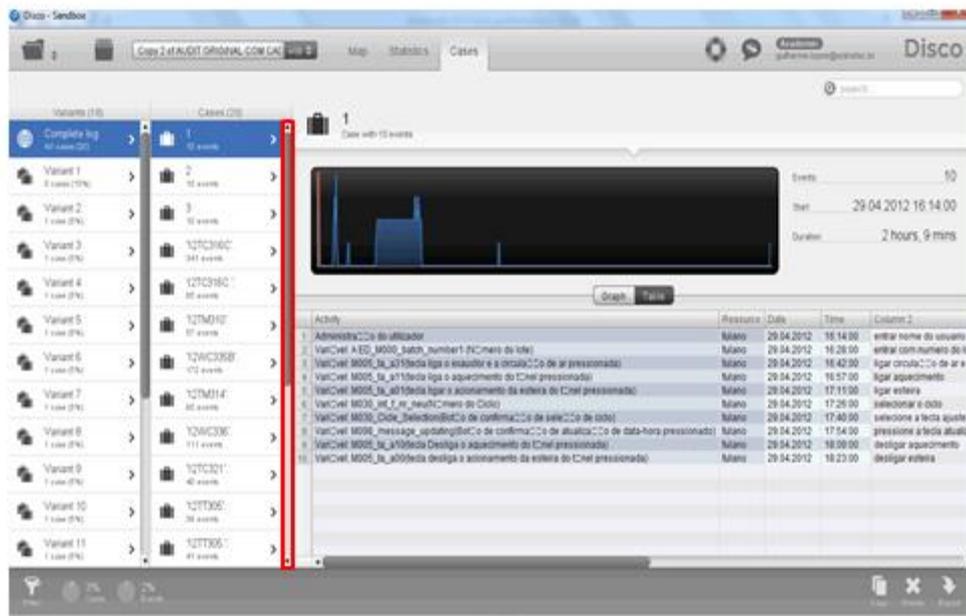


Figura 8-86 - Instâncias Resultantes

A instância Apriori foi nomeada com Instância número 1 para facilitar a identificação. Nota-se que na variante onde se encontra a Instância Apriori existem mais 2 instâncias, estão são as instâncias conformes, todas as instâncias que estiverem nesta variante estarão conformes pois executam as mesmas atividades da instância Apriori Instância. No estudo de caso não há outras restrições em relação à instâncias Apriori, mas poderíamos ter outras restrições como dependência de uma atividade com outra ou seja atividade A sempre vem antes de B ou por exemplo tempo mínimo entre A e B > 30 min. Nestes casos a variante continuaria valendo, mas a construção do filtro terá que levar em conta estas restrições.

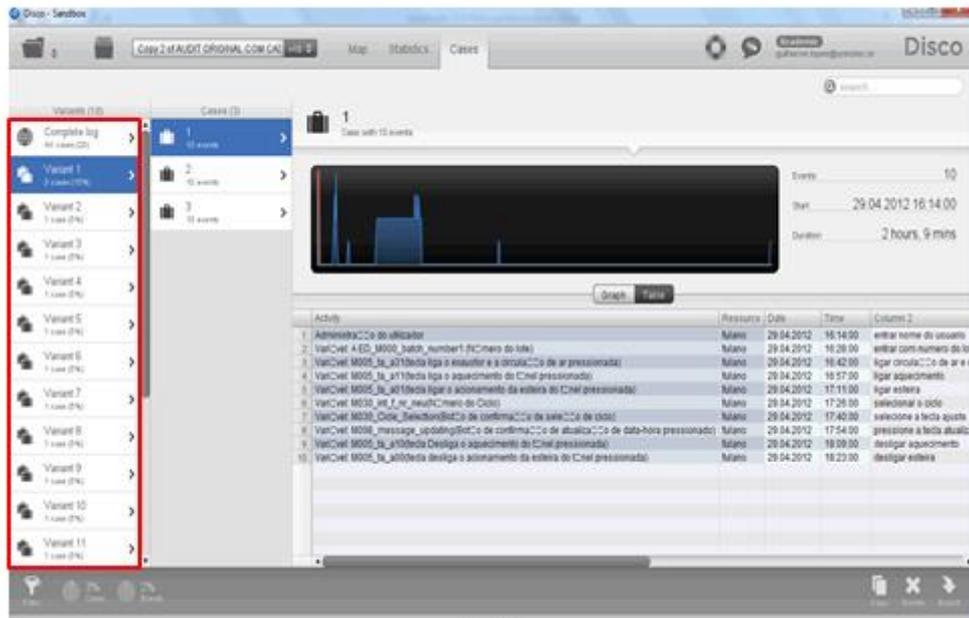


Figura 8-87 - Instância A priori

As demais instâncias do processo estão agrupadas em outras variantes e poderão ser visualizadas com o nível de abstração de variante ou instância a instância conforme Figura 8-88 - Demais instâncias completas com desvio não regulatório.

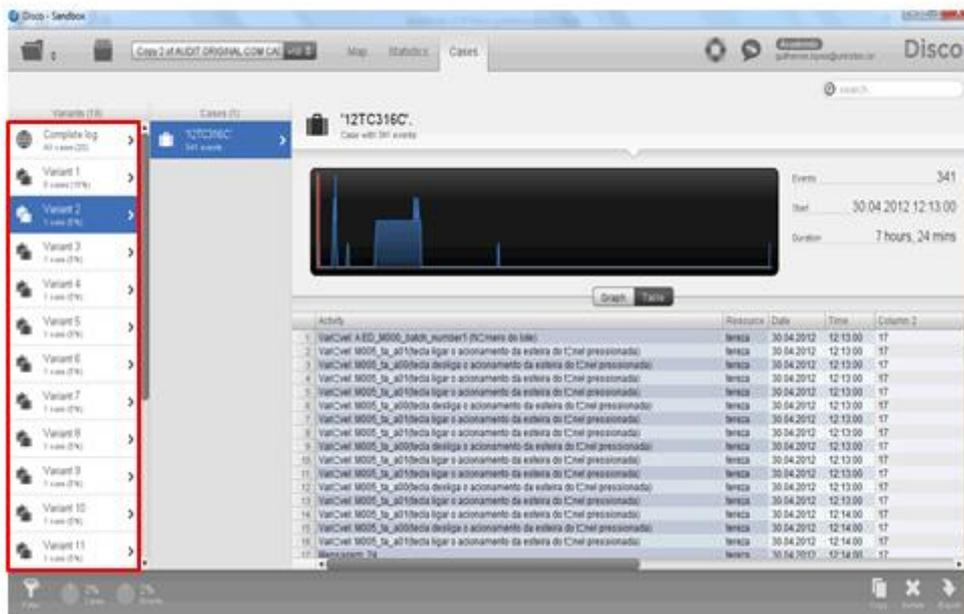


Figura 8-88 - Demais instâncias completas com desvio não regulatório