



DESCOBERTA E ANÁLISE DE PROCESSOS DE NEGÓCIO NÃO ESTRUTURADOS

Alexandre de Oliveira Ribeiro

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia de Sistemas e Computação, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia de Sistemas e Computação.

Orientador: Jano Moreira de Souza

Rio de Janeiro
Setembro de 2013

DESCOBERTA E ANÁLISE DE PROCESSOS DE NEGÓCIO NÃO
ESTRUTURADOS

Alexandre de Oliveira Ribeiro

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO
LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA
(COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE
DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE
EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

Examinada por:

Prof. Jano Moreira de Souza, D.Sc.

Prof. Geraldo Bonorino Xexéo, D.Sc.

Prof.^a Fernanda Araujo Baião Amorim, D.Sc.

RIO DE JANEIRO, RJ - BRASIL

SETEMBRO DE 2013

Ribeiro, Alexandre de Oliveira

Descoberta e Análise de Processos de Negócio Não Estruturados/ Alexandre de Oliveira Ribeiro. – Rio de Janeiro: UFRJ/COPPE, 2013.

IX, 130 p.: il.; 29,7 cm.

Orientador: Jano Moreira de Souza

Dissertação (mestrado) – UFRJ/ COPPE/ Programa de Engenharia de Sistemas e Computação, 2013.

Referências Bibliográficas: p. 126-130.

1. Gerenciamento de Processos de Negócio. 2. Mineração de Processos. 3. Processos Desestruturados. I. Souza, Jano Moreira de. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Sistemas e Computação. III Título.

À minha família.

AGRADECIMENTOS

Agradeço primeiramente a Deus, por ter criado a vida, o universo e tudo mais. Agradeço à minha mãe, Giselda Maria, ao meu pai, Antonio Carlos, à minha irmã, Gisele, por terem me apoiado sempre, desde que nasci, até este momento da minha vida. Não se constrói um mestre sem antes educar uma criança.

Agradeço à minha namorada Ana Luiza por todo o apoio, compreensão e motivação durante todo o tempo que cursei o mestrado. Ela esteve do meu lado em todos os momentos felizes e de dúvida que passei durante todo este tempo, garantindo o suporte para que eu continuasse seguindo em frente.

Agradeço aos professores Alexandre Assis, Marta Mattoso, Henrique Cukierman, Carlos Eduardo Mello, Geraldo Xexéo, Jonice Oliveira e Ricardo Marroquim por terem ministrado as matérias que cursei durante o mestrado. Agradeço em especial o professor Jano Moreira de Souza, pela orientação e todo o apoio para o desenvolvimento desta dissertação de mestrado.

Agradeço a todos meus amigos e companheiros de LabBD, Carlos, Pedro, Emerson, Júlia, Thaís, Wagner, Gustavo, Rafael, Renan e João Luiz, e aos meus amigos e companheiros que cursaram o mestrado comigo, Allan, Diego, Filipe e Bernardo.

Agradecimentos especiais ao Marco Vaz, por todas as oportunidades dadas e o incentivo para cursar o mestrado, por toda a paciência e apoio durante o processo de construção da minha dissertação, e por todos os conselhos dados desde que eu era aluno de graduação. Sem ele este trabalho não teria sido possível.

DESCOBERTA E ANÁLISE DE PROCESSOS DE NEGÓCIO NÃO ESTRUTURADOS

Alexandre de Oliveira Ribeiro

Setembro/2013

Orientador: Jano Moreira de Souza

Programa: Engenharia de Sistemas e Computação

Um dos objetivos em todas as organizações é a melhoria de seus processos para redução de custos, retrabalho e variabilidade dos procedimentos de trabalho. A mineração de processos é uma área do conhecimento humano que visa extrair conhecimento a partir de um conjunto de registros de atividades de processos, permitindo às organizações um entendimento de seus processos de negócio. Entretanto, a seleção de um conjunto de dados que representam um processo pode ser uma tarefa desafiadora, pois nem todos os processos possuem estrutura rígida em seus fluxos de execução, e nem todas as entidades relacionadas ao processo, como pessoas, atividades e departamentos, estão descritas de maneira uniforme nestes dados.

Esta dissertação apresenta uma abordagem de descoberta de processos não estruturados focada no entendimento dos dados provenientes de execuções de processos suportados por sistemas de informação. Uma ferramenta foi desenvolvida para permitir a curadoria de dados presentes em registros de execução de processos através de um dicionário capaz de unificar todos os valores referentes a uma entidade. Esta ferramenta também é capaz de realizar uma busca exploratória para apoiar a seleção de registros de execuções relacionados a processos não estruturados, propondo uma linguagem gráfica para retornar instâncias de processos através de seus encadeamentos de eventos.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

DISCOVERING AND ANALYSIS OF UNSTRUCTURED BUSINESS PROCESSES

Alexandre de Oliveira Ribeiro

September/2013

Advisor: Jano Moreira de Souza

Department: Computer Science Engineering

One of the main goals in every organization is to enhance their core processes to reduce cost, rework and the variability. Process mining is an emerging subject area which aims to extract knowledge from a set of process execution logs, allowing organizations to have a better understanding of their processes. However, selecting a representative set of execution logs which represent a business process can be a challenging task. Not all business processes have a well-defined structure defining their execution flow. Besides, entities regarding those processes, such as people, activities or departments, are not always described uniformly in the process execution logs.

This dissertation presents a process mining approach focused on the understanding of the data present in those logs. A system was developed to allow the curation of data present in process execution logs, supported by a data dictionary capable of unifying multiple values referring to an entity. This system also supports the explorative search to select process execution logs related to unstructured processes, proposing a graphic language to retrieve process instances based on their activities flow.

ÍNDICE

Capítulo 1 – Introdução	1
1.1. Problema	2
1.2. Objetivos	5
1.3. Organização	8
Capítulo 2 – Gerenciamento de Processos de Negócio	10
2.1. Processo de Negócio	10
2.2. Gerenciamento de Processos de Negócio	12
2.3. BPMN	15
2.4. Redes de Petri	20
Capítulo 3 – Mineração de Processos	22
3.1. Algoritmos de Descoberta	25
3.1.1. Algoritmo α	30
3.1.2. Minerador Heurístico	32
3.1.3. Minerador Genético	37
3.2. XES	40
3.3. Principais Desafios	43
3.4. Ferramentas Relacionadas	49
3.4.1. ProM	49
3.4.2. ProM Import	52
3.4.3. XESame	54
3.4.4. MANA	56
Capítulo 4 – Linguagem Gráfica de Busca	59
4.1. Elementos Gráficos da Linguagem	62
4.2. Algoritmo de Busca	65
Capítulo 5 – Ferramenta	69
5.1. Fluxo de Atividades	72
5.2. Entidades e Modelagem de Dados	75
5.2.1. Entidades dos Registros de Processos	75
5.2.2. Entidades do Dicionário de Dados	79
5.3. Módulos da Ferramenta	81
5.3.1. Dicionário de Dados	82
5.3.2. Cadastro e Seleção de Processos	86

5.3.3.	Filtragem de Instâncias	87
5.3.4.	Mineração	93
5.3.5.	Visualização e Edição de Modelos	95
5.3.6.	Animação	98
5.4.	Características Diferenciais	99
Capítulo 6 – Estudo de Caso		104
6.1.	Importação dos Dados	105
6.2.	Preenchimento do Dicionário	107
6.3.	Primeiro Estudo de Caso	111
6.4.	Segundo Estudo de Caso	117
Capítulo 7 – Conclusões		122
7.1.	Resultados e Contribuições	122
7.2.	Limitações e Trabalhos Futuros	123
Referências		126

Capítulo 1 – Introdução

O gerenciamento de processos de negócio é uma área que vem recebendo bastante atenção de gestores interessados em melhorar a competitividade de suas empresas no mercado. Entretanto, técnicas tradicionais de gerenciamento de processos de negócio costumam lidar com opiniões e impressões das pessoas sobre a realidade, em vez de serem baseadas em fatos. A mineração de processos é uma disciplina que provê um conjunto de técnicas e ferramentas capaz de extrair conhecimento a partir de dados sobre ocorrências reais de processos com objetivo de apoiar o entendimento e aperfeiçoamento dos mesmos, eliminando subjetividades (van der Aalst 2011).

Muitos sistemas de informação na atualidade são construídos para dar suporte a processos de negócio. Nestes sistemas é possível fazer o registro de eventos de processos, que permitem o armazenamento de informações sobre como transcorreram suas atividades, quais tarefas foram executadas, quando, onde, por quem, etc. Estes registros servem como base para as técnicas de mineração de processos, que são capazes de gerar modelos que explicam o comportamento de um processo extraindo o fluxo de atividades a partir de seus dados.

A mineração de processos possui objetivos de descoberta, verificação e extensão de processos. As atividades de descoberta fazem a utilização de algoritmos para gerar modelos de processos a partir de seus registros de execução. A verificação compara modelos de processos a registros dos processos correspondentes a estes modelos, inferindo se a realidade evidenciada pelos registros ocorre de fato como os modelos indicam. A extensão de processos visa melhorar modelos de processos já existentes a partir de um conjunto de dados sobre como estes foram executados, sendo uma destas

melhoras a correção das áreas do modelo que não refletem a realidade (van der Aalst 2011).

Este trabalho possui foco na descoberta de processos, em especial processos que não possuem estrutura rígida em seus fluxos de execução, que são chamados de processos não estruturados. O interesse por este tipo de processo vem da necessidade de permitir que as instituições entendam como funcionam seus processos. Este conhecimento pode ser obtido a partir de dados presentes nos sistemas de registro de protocolos destas instituições, que em geral fazem o acompanhamento de processos de forma passiva, sem impor um fluxo de execução, gerando processos não formalmente estruturados.

O restante deste capítulo contextualiza o estudo realizado junto à área de mineração de processos, apresentando os principais problemas a serem observados e tratados, os objetivos deste trabalho, uma breve descrição sobre a ferramenta desenvolvida para suporte a atividades de mineração de processos alinhada com estes objetivos e uma introdução aos tópicos dos demais capítulos, expondo o formato em que esta dissertação está organizada.

1.1. Problema

Algoritmos de descoberta de processos são capazes de gerar modelos a partir de registros de atividades, embora nem sempre os resultados sejam satisfatórios do ponto de vista de extração do conhecimento por pessoas. A presença de ruídos nestes registros pode causar representações não condizentes com a realidade de um processo, enquanto processos que não seguem uma estrutura rígida podem resultar em modelos com aparência de espaguete, de alta complexidade e difícil entendimento (Weijters *et al.* 2006). Um modelo com aparência de espaguete é ilustrado na Figura 1.

haja uma investigação cuidadosa dos dados. Também há a questão de registros onde valores adquiridos através de formulários que permitem a entrada de textos livres são utilizados para caracterizar os atributos de um evento ocorrido, possibilitando inúmeras formas de representação para o nome de uma atividade ou um departamento de uma organização, por exemplo. Esta gama de valores, apesar de geralmente ser de fácil entendimento por humanos, é de difícil interpretação por máquinas, sendo necessária a criação de mecanismos para indicar a correspondência destes valores a entidades bem definidas, minimizando o problema de múltiplas representações.

As técnicas de mineração de processos necessitam um conjunto de registros de eventos como ponto de partida, porém, nas ferramentas existentes que dão suporte a estas técnicas, o enfoque para obtenção destes dados está no mapeamento destes registros para um padrão de armazenamento de dados que é utilizado pela ferramenta, sem que haja qualquer tipo de filtragem e tratamento destes dados (Esposito 2012).

Neste cenário, o responsável por realizar esta tarefa deve possuir o conhecimento sobre como estes dados são armazenados no sistema de onde serão extraídos e quais destes dados são referentes a instâncias do processo que será alvo de estudos, além de ter a proficiência necessária em técnicas de computação para realizar a extração destes dados e correta formatação para utilização em ferramentas de mineração de processos. Torna-se indispensável a construção de mecanismos para facilitar a tarefa de seleção de registros de eventos de processos capazes de mitigar os problemas apresentados nesta seção.

Para que as atividades de mineração de processos obtenham maior grau de sucesso, é necessário que as ferramentas utilizadas sejam exploratórias, isto é, admitindo a comparação do que se espera do comportamento perfeito de um processo e o que acontece na realidade através de várias iterações, permitindo que ajustes sejam

feitos e melhores resultados sejam obtidos a partir de conhecimentos incrementais adquiridos.

1.2. Objetivos

O objetivo desta pesquisa é desenvolver mecanismos que propiciem maior exploração e entendimento dos dados provenientes de execuções de processos para melhorar os resultados obtidos pelos principais algoritmos de descoberta de processos descritos na literatura, além de uma ferramenta de mineração de processos que dê suporte a estes mecanismos. Este objetivo visa aumentar as chances de êxito na extração de informações relevantes que evidenciam o comportamento de processos a partir de uma variedade maior de sistemas de informação que registram eventos de processo. Embora seja possível realizar estas análises em outras ferramentas, por vezes os resultados não são satisfatórios devido aos problemas citados na seção anterior.

Outro objetivo, alinhado ao primeiro, é a continuação dos estudos relacionados a mineração de processos desestruturados no cenário de organizações públicas brasileiras iniciados em (Esposito 2012). Novas técnicas são desenvolvidas neste trabalho visando superar as dificuldades encontradas na utilização da ferramenta MANA (Esposito 2012), sobretudo as dificuldades relacionadas à baixa qualidade de dados encontrados em sistemas que registram eventos de processos do governo federal brasileiro. Esta pesquisa apresenta dois estudos de caso para ilustrar o uso da ferramenta desenvolvida com dados extraídos a partir de um destes sistemas.

Este estudo apresenta uma linguagem gráfica de busca capaz de selecionar instâncias de processos com base em seus encadeamentos de eventos, possuindo o intuito de facilitar a exploração da base de dados de processos evitando a necessidade de conhecimento em técnicas de computação pelas pessoas diretamente envolvidas com o

processo a ser minerado. A criação de metáforas para simplificar o processo de busca e seleção de registros de eventos está alinhada com o desafio de melhorar a usabilidade dos sistemas de mineração de processos para facilitar seus entendimentos por não especialistas, conforme descrito no Manifesto de Mineração de Processos (van der Aalst *et al.* 2012), um documento redigido pelos principais pesquisadores da área que indica direcionamentos para futuras pesquisas na área de mineração de processos e os desafios existentes na atualidade. Também existe a possibilidade de fazer a exploração da base de dados selecionando instâncias de processos a partir de filtros textuais que atuam sobre seus atributos, ou ainda combinando estes filtros à linguagem gráfica de busca.

Este trabalho propõe ainda a criação de um dicionário de dados para dar suporte ao entendimento dos dados presentes nos registros de processos analisados. Este dicionário visa mitigar o problema de falta de padrão para as informações presentes nestes registros e permite que os algoritmos de descoberta de processos entendam grafias distintas que se referem a uma mesma entidade como sendo um termo único. Outra aplicação do dicionário é propiciar a criação de correspondências entre termos, para que modelos de processo sejam gerados utilizando um nível de granularidade uniforme em sua representação, evitando que partes de um modelo possuam quantidade diferente de detalhes que outras.

Outro mecanismo para melhorar o entendimento das atividades relacionadas a mineração de processos é o versionamento das iterações realizadas na ferramenta desde a criação de um modelo de processo até seu estado atual. Este versionamento permite ao usuário verificar a evolução das buscas e instâncias de processo retornadas, dos parâmetros utilizados nos algoritmos de descoberta de processos e dos resultados obtidos, evitando que informações sejam perdidas durante a exploração dos dados e a criação de modelos.

Para a utilização da ferramenta é necessário primeiro realizar a importação dos dados referentes a registros de execução de processos. Completado este procedimento, um grupo inicial de entidades representativas em relação aos processos que serão analisados devem ser cadastradas no dicionário de dados, com intuito de facilitar a identificação e permitir a padronização de quais valores textuais presentes nos registros de execução de processos correspondem às entidades que constituirão os modelos descobertos.

As etapas seguintes do fluxo de uso da ferramenta, denominadas exploração e mineração, formam um ciclo que deve ser repetido até que o modelo de processo construído seja satisfatório. A etapa de exploração consiste na utilização da linguagem gráfica de busca e dos filtros supracitados para percorrer os registros de execuções de processos presentes na base de dados e retornar somente as instâncias que correspondem ao processo que será analisado. A etapa de mineração utiliza um algoritmo de descoberta de processos para construir um modelo de processo com base nos registros selecionados na etapa de exploração, com parâmetros configuráveis pelo usuário. A cada nova iteração deste ciclo, existe a possibilidade da pessoa que utiliza a ferramenta adquirir conhecimentos incrementais e realizar os ajustes necessários na exploração e mineração para obter resultados melhores.

Por fim, a ferramenta apresenta um conjunto de funcionalidades para prover o usuário de informações relevantes a partir do modelo que foi construído, permitindo a identificação de comportamentos não desejados, gargalos na execução do processo, etc. Para facilitar o entendimento dos envolvidos, há como realizar alterações nestes modelos para evidenciar informações julgadas interessantes.

Maiores detalhes sobre esta ferramenta são expostos no Capítulo 5, onde são discutidos todos os módulos, o fluxo de trabalho e os principais diferenciais em relação a outras ferramentas capazes de realizar a descoberta de processos não estruturados.

1.3. Organização

Esta dissertação está dividida em sete capítulos, onde o primeiro apresenta uma introdução ao trabalho realizado, o segundo e terceiro expõe a literatura relacionada a gerenciamento de processos de negócio e mineração de processos, o quarto e quinto apresentam propostas para apoiar a descoberta de processos não estruturados, o sexto discute os resultados obtidos pelo trabalho e o sétimo apresenta suas conclusões. Os tópicos abordados por estes capítulos são os seguintes:

O Capítulo 2 discorre sobre o que é gerenciamento de processos de negócio. Neste capítulo são expostas definições sobre o que são processos de negócio e apresentadas as notações para modelagem de processos que serão utilizadas no decorrer deste trabalho.

O Capítulo 3 discute o que é mineração de processos, apresentando algoritmos capazes de realizar descoberta de processos. São introduzidos os principais desafios da área e ferramentas existentes relacionadas a descoberta de processos não estruturados.

O Capítulo 4 apresenta uma proposta de linguagem de busca gráfica para selecionar registros de eventos de processos através de seus encadeamentos de eventos. São reveladas as abstrações suportadas por esta linguagem e é apresentado um algoritmo para realizar a busca destes registros.

O Capítulo 5 expõe a ferramenta construída para dar suporte a atividades de descoberta de processos não estruturados, discutindo em minúcias o fluxo de trabalho proposto, a arquitetura da ferramenta e seus principais diferenciais.

O Capítulo 6 mostra dois estudos de caso realizados utilizando a ferramenta com registros de eventos de processos extraídos de uma aplicação real que dá suporte a processos não estruturados.

O Capítulo 7 encerra este trabalho, apresentando suas conclusões, discorrendo sobre os resultados obtidos, indicando as dificuldades encontradas e sugerindo propostas para a realização de estudos futuros.

Capítulo 2 – Gerenciamento de Processos de Negócio

Este capítulo apresenta os principais conceitos relacionados a gerenciamento de processos de negócio considerados relevantes para o entendimento do objeto da pesquisa, introduzindo definições sobre o que são processos de negócio e expondo o ciclo de vida de gerenciamento de processos de negócio. As representações de modelos de processos que serão utilizadas neste trabalho são apresentadas e seus conceitos e o significado de suas notações gráficas são explicitados.

2.1. Processo de Negócio

O advento da Revolução Industrial trouxe uma mudança no paradigma do trabalho na sociedade. Antes deste período, um artesão era responsável por todas as etapas de produção. Com o surgimento das indústrias modernas, as pessoas passaram a ser treinadas para executar tarefas especializadas em uma parcela menor de etapas de produção. O aspecto industrial de organização do trabalho acarretou em grande aumento de produtividade, agrupando pessoas com especialidades similares e dispendo-as de acordo com o processo de produção. Deste então há um grande interesse na indústria em organizar e otimizar seus processos.

Um processo de negócio pode ser definido como um conjunto de atividades logicamente relacionadas e executadas com fins de obter um produto relevante para o negócio (Davenport e Short 1990). Um processo também pode ser definido como uma forma de converter insumos em produtos realizada de uma maneira confiável, reproduzível e consistente (Zairi 1997). Entre os elementos que compõem um processo, estão insumos que obedecem a requisitos bem especificados, um conjunto de tarefas ou atividades bem definidas, uma sequência lógica entre as atividades que transformarão os insumos em produtos e um padrão esperado para o produto ou resultado (Bulletpoint

1996 apud Zairi 1997). Jacobson *et al.* (1994) definem ainda um processo de negócio como um conjunto de atividades realizadas para atender um cliente.

Segundo Melão e Pidd (2000), processos de negócio podem ser considerados como um conjunto de partes compostas por pessoas, tarefas, estruturas e tecnologias que interagem entre si e com fatores externos para alcançar algum objetivo, possuindo insumos, produtos, transformações e delimitações para separar o que faz parte do processo e o ambiente onde este se insere.

A *Workflow Management Coalition* (1999) define um processo de negócio como um conjunto de atividades que satisfazem um objetivo de negócio, definindo ainda o termo atividade como um conjunto de trabalhos que compõe um segmento ou passo lógico a ser executado em um processo. Este grupo também propõe a classificação de uma atividade como manual ou automatizada, distinguindo-a com base na utilização ou não de suporte por técnicas de computação.

Existem dois aspectos importantes para um processo de negócio: sua execução correta e sua eficiência (Reijers 2003). A execução de forma correta trata em garantir que as atividades de um processo ocorram sem falhas ou desvios de algum padrão, caso este exista, e a eficiência busca minimizar o tempo, recursos e esforços empreendidos no processo.

De acordo com van der Aalst e van Hee (2004), os processos podem ser categorizados conforme seus níveis de padronização e frequência de execução. Os processos *ad hoc* e projetos não seguem um padrão definido e são criados e planejados para serem executados somente uma vez, de acordo com as demandas de seu problema. Os processos de produção em massa customizáveis são planejados para que múltiplas execuções sigam algum padrão, incorporando possíveis variações em suas tarefas que se adequam com as necessidades específicas de um cliente. Por fim, os processos de

produção em massa são planejados para execução em alta frequência e devem seguir um padrão rígido, sem margens para desvios.

Outro tipo de categorização é quanto ao foco do resultado a ser obtido pelo processo (Reijers 2003). Os processos de produção estão diretamente alinhados com as atividades fim das instituições, produzindo bens de consumo ou provendo serviços a entidades externas. Os processos de suporte servem como apoio aos processos de produção, garantindo que as atividades fim sejam abastecidas de todos os recursos necessários para que possam ser executadas. Os processos de gerenciamento possuem o objetivo de coordenar os processos de uma organização, sendo responsáveis pelas decisões táticas e estratégicas para o bom funcionamento dos mesmos.

2.2. Gerenciamento de Processos de Negócio

O termo gerenciamento de processos de negócio, ou *BPM* (sigla da expressão na língua inglesa *Business Process Management*) não possui um significado bem definido, mas geralmente está atrelado ao planejamento, análise, modelagem, implantação e controle de processos de negócio (Reijters 2003). Uma definição abrangente sobre o termo inclui suportar processos de negócio empregando métodos, técnicas e software para planejar, executar, controlar e analisar processos operacionais envolvendo humanos, organizações, aplicações, documentos e outras fontes de informação (van der Aalst *et al.* 2003). Gerenciamento de Processos de Negócio também pode ser definido como uma evolução de gerenciamento de *workflows*, onde um *workflow* é a automação de um processo de negócio, total ou parcial, em que documentos, informação ou tarefas são passadas de um ator para outro para realizar ações, de acordo com um conjunto de regras (Weske *et al.* 2004). As ferramentas disponíveis para gerência de processos de

negócio geralmente enfatizam a representação gráfica dos processos, enriquecidas com suporte a workflows e integração com outros sistemas (Hepp *et al.* 2005).

Entre as principais atividades de *BPM*, o planejamento de um processo de negócio está diretamente ligado à modelagem de processos, pois em geral o produto das tarefas de planejamento é um modelo de processo descrito de acordo com alguma linguagem de representação de processos (Weske *et al.* 2004).

Neste capítulo são descritas duas destas linguagens, que são utilizadas para representar modelos de processo ao longo deste trabalho: o BPMN e as Redes de Petri. A primeira notação vem se tornando a linguagem padrão para representação de processos de negócio (zur Muehlen e Recker 2008), enquanto a segunda possui relevância histórica para a literatura da área, sendo utilizada nas descrições de algoritmos de mineração de processos presentes nesta dissertação.

Outro aspecto importante de *BPM* é a análise de processos de negócio, que visa investigar propriedades não triviais a partir da observação de processos. As análises abrangem uma vasta gama de atividades, como simulação de casos e medição de desempenho (Weske *et al.* 2004). Uma vez que um processo está descrito em alguma linguagem com semântica bem definida, é possível fazer a verificação de várias propriedades, incluindo partes do modelo que nunca são executadas, gargalos em atividades, entre outros.

Estas atividades descritas acima estão presentes no ciclo de vida *BPM*, uma abstração criada para representar as atividades mais comuns desempenhadas na gerência de processos de negócio. Segundo Weske (2004), este ciclo abrange quatro etapas que dão suporte a operacionalização de processos de negócio: projeto, implantação, execução e análise. A Figura 2 ilustra o ciclo de vida de gerência de processos de

negócio. As etapas do ciclo devem ser executadas continuamente, não havendo um ponto onde o ciclo esteja finalizado.



Figura 2 – Ciclo de vida de gerência de processos de negócio. Adaptado de (Weske 2004).

A etapa de projeto consiste em fazer uma investigação de como o processo de negócio é ou deveria ser executado em uma instituição, e a partir destas informações se constrói um modelo que representa este processo. A etapa de implantação faz a implantação dos projetos gerados na etapa anterior em sistemas que dão suporte a processos, como sistemas de *workflow* (Weske *et al.* 2004). A etapa de execução é quando de fato as atividades dos processos são realizadas, pondo em prática o que foi planejado. A etapa de análise faz a verificação e validação dos processos de negócio, identificando erros de projeto ou pontos de melhoria para os processos que estão em execução.

A ferramenta desenvolvida neste trabalho está contida na etapa de análise, não sendo necessária a realização das etapas anteriores para a utilização da mesma. Em ambientes onde os processos não são formalmente estruturados esta fase tem importância fundamental para a descoberta de como estes processos são realmente executados. A utilização desta ferramenta dá suporte à investigação de características destes processos, provendo informações como quais são as suas atividades, locais onde

estas atividades acontecem, seus responsáveis, possíveis gargalos na execução do processo, etc.

2.3. BPMN

A *Business Process Model and Notation* (notação e modelagem de processo de negócio, em tradução livre), ou *BPMN*, é uma linguagem capaz de representar modelos de processos de negócio proposta pelo OMG (acrônimo do inglês *Object Management Group*), uma instituição composta pelos principais desenvolvedores de ferramentas de gestão de processos de negócio, como Oracle e IBM (Recker 2010). O BPMN tem recebido grande atenção tanto pela academia quanto pelos práticos da indústria e vem se tornando a linguagem padrão para captura de processos de negócio, especialmente em análises de domínio e planejamento em alto nível de sistemas (zur Muehlen e Recker 2008). Parte do sucesso de sua adoção está no fato que o BPMN oferece um padrão de notações gráficas simples e de fácil utilização, permitindo que analistas de negócio documentem seus processos de negócio e comuniquem suas ideias tanto para parceiros de negócio quanto para pessoas suas próprias organizações (Völzer 2011). Outra característica importante do BPMN é a possibilidade de converter modelos para a sintaxe BPEL (*Business Process Execution Language*), o padrão de facto para implantação de processos de negócio a partir de serviços *web*, que é suportado por grande número de plataformas (Ouyang 2006).

O BPMN define um diagrama de processo de negócio, que é um tipo de gráfico representando fluxo de atividades que incorpora objetos adaptados para modelagem de processos, como junções *AND*, disjunções *AND*, junções *XOR*, disjunções *XOR*, et cetera (Ouyang 2006).

Esta foi a notação escolhida para representar modelos de processos de negócio na ferramenta desenvolvida neste trabalho. Os motivos incluem a utilização de uma notação que é amplamente utilizada e sua simplicidade e eficiência em construir e comunicar ideias sobre como funcionam os processos a partir de seus modelos.

A definição completa de *BPMN* inclui pelo menos 50 entidades diferentes, que podem ser agrupadas em quatro categorias: objetos de fluxograma, conectores, raias e artefatos (zur Muehlen e Recker 2008). Os objetos de fluxograma constituem os elementos básicos do modelo, e entre eles estão os eventos, atividades e *gateways*. Os conectores fazem ligações entre os objetos de fluxograma através da utilização de arestas com diferentes tipos de representação. As raias são utilizadas para separar os objetos de fluxograma em domínios distintos, como grupos de atividades executadas em um departamento. Por fim, artefatos podem ser adicionados ao modelo para enriquecê-lo de informações onde apropriado, como anotações em texto que adicionam comentários a atividades ou eventos. A Figura 3 ilustra um modelo de processo de venda descrito utilizando *BPMN*. No exemplo existem duas raias, indicando em qual setor a atividade decorre, além de um artefato do tipo comentário, indicando que são utilizadas embalagens ecológicas na atividade de embalar o produto.

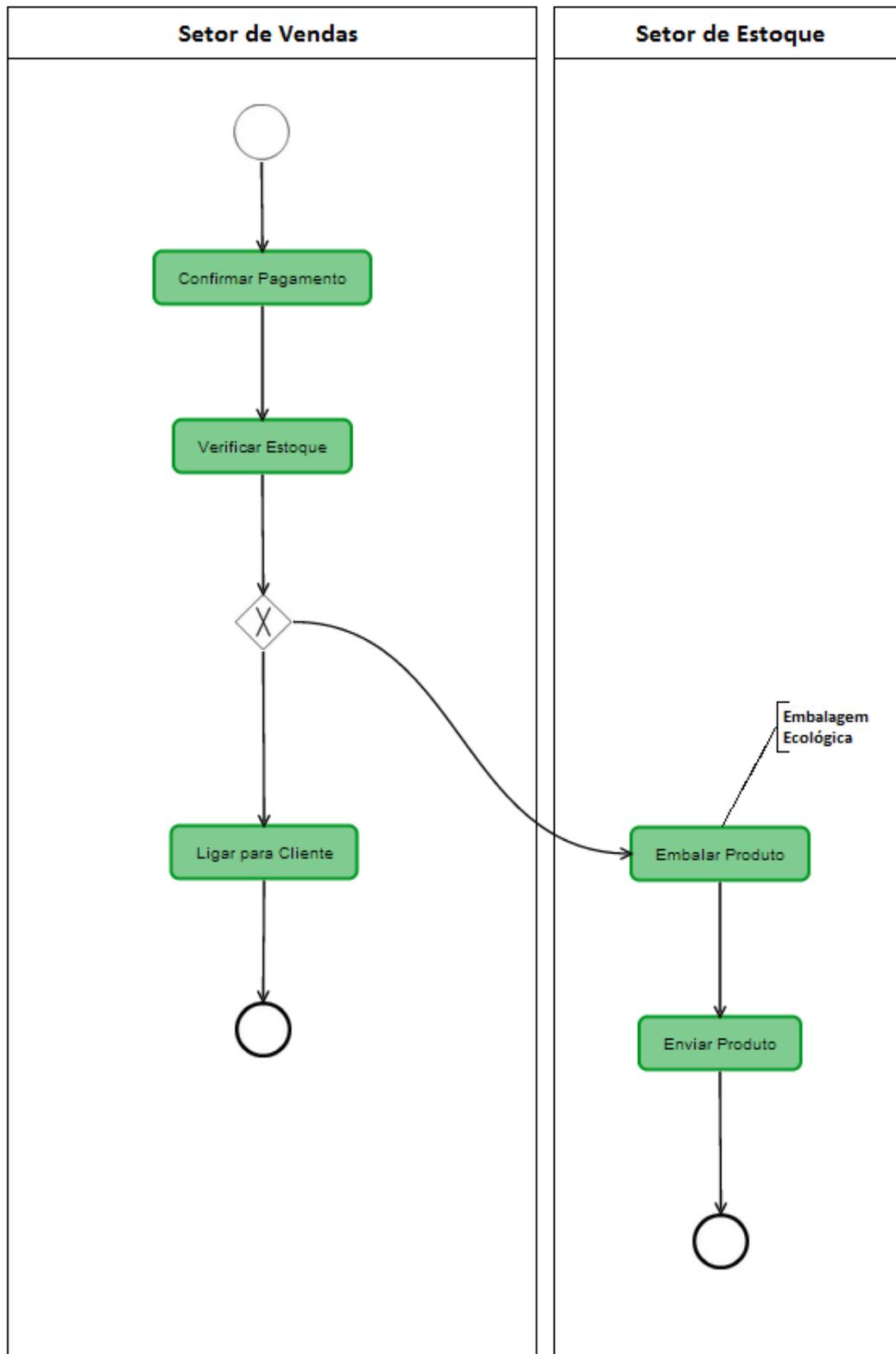


Figura 3 – Modelo de processo de negócio descrito em BPMN.

A definição completa e atual do padrão pode ser encontrada em (OMG 2011). Abaixo segue uma adaptação da descrição das principais entidades incluídas no padrão de acordo com a definição atual, junto às representações gráficas correspondentes.

Evento. É algo que acontece durante o curso de um processo, geralmente possuem uma causa que serve como condição de disparo. São representados por círculos, e existem três tipos: início, que indica quando um processo deve ser iniciado, fim, que indica quando um processo é finalizado, e intermediário, que é qualquer evento que ocorre entre um evento início e evento fim. A Figura 4 exemplifica um evento início à esquerda, um evento intermediário no centro e um evento fim à direita.



Figura 4 – Representação de Eventos em BPMN. Adaptado de (OMG 2011).

Atividade. É um termo genérico para um trabalho realizado em um processo. Existem dois tipos de atividades: subprocesso e tarefa. São representadas por retângulos com bordas arredondadas. Incluem o nome da atividade dentro do retângulo. A Figura 5 exemplifica uma atividade do tipo tarefa.



Figura 5 – Representação de uma Atividade em BPMN. Adaptado de (OMG 2011).

Fluxo de Sequência. É utilizado para indicar a ordem em que as atividades são executadas em um processo. É representado por uma linha contínua com uma seta na extremidade indicando qual é a próxima atividade. A Figura 6 ilustra duas atividades conectadas por um fluxo de sequência.



Figura 6 – Representação de um Fluxo de Sequência em BPMN. Adaptado de (OMG 2011).

Gateway. É utilizado para controlar a convergência e divergência de fluxos de sequência de um processo. São representados por um losango com um símbolo em seu interior, que indica o tipo do *gateway*. A Figura 7 exemplifica três anotações distintas de *gateways*. O elemento central representa um *gateway AND*, indicando paralelismo entre as atividades: no caso de divergência de fluxo, todos os fluxos de sequência que saem deste *gateway* devem ser executados, e no caso de convergência de fluxo, todos os fluxos de sequência que entram devem estar finalizados para que o elemento conectado à saída deste *gateway* possa ser executado. Os elementos na extremidade representam *gateways XOR*, indicando decisão exclusiva: somente um dos fluxos de sequência deve ser seguido ou finalizado para que o próximo elemento conectado à saída deste *gateway* possa ser executado. Nota-se que no caso de não haver um símbolo indicando o tipo do *gateway*, o mesmo é considerado um *gateway XOR*.



Figura 7 – Representação de Gateways em BPMN. Adaptado de (OMG 2011).

2.4. Redes de Petri

As Redes de Petri são uma notação para descrever e estudar informações de processos concorrentes, assíncronos, distribuídos, paralelos, não determinísticos e/ou estocásticos (Murata 1989). Um modelo em forma de rede de Petri captura tanto a estrutura estática de processos quanto os aspectos dinâmicos de seus comportamentos (Desel e Reisig 1998). As redes de Petri são de grande importância para a área de mineração de processos, pois são a forma de representação escolhida para apresentar os processos descobertos através de diversos algoritmos, incluindo o Algoritmo α , um dos mais estudados na área (van der Aalst, 2011).

Uma rede de Petri é composta por uma tupla $T(L,T,A)$, onde L representa os lugares da rede, T suas transições e A os arcos que conectam os lugares e transições; além de uma marcação M , que indica o estado atual da rede através da quantidade de fichas existentes em cada lugar pertencente ao conjunto L (Goltz e Reisig 1983). Uma transição pode ser disparada em uma determinada marcação caso todos os arcos desta transição estejam ligados a lugares que possuem pelo menos uma ficha. O comportamento da rede é definido pelas mudanças de marcação causadas pelo disparo de transições (Desel e Reisig 1998). A Figura 8 representa uma rede de Petri. Os lugares da rede são representados por círculos, enquanto as transições são representadas por barras. Existem quatro fichas de marcação, uma em cada um dos lugares P_0 , P_1 , P_2 e P_3 . É possível disparar as transições T_0 , T_1 e T_3 . Nota-se que não é possível disparar a transição T_2 pois não há fichas em P_4 .

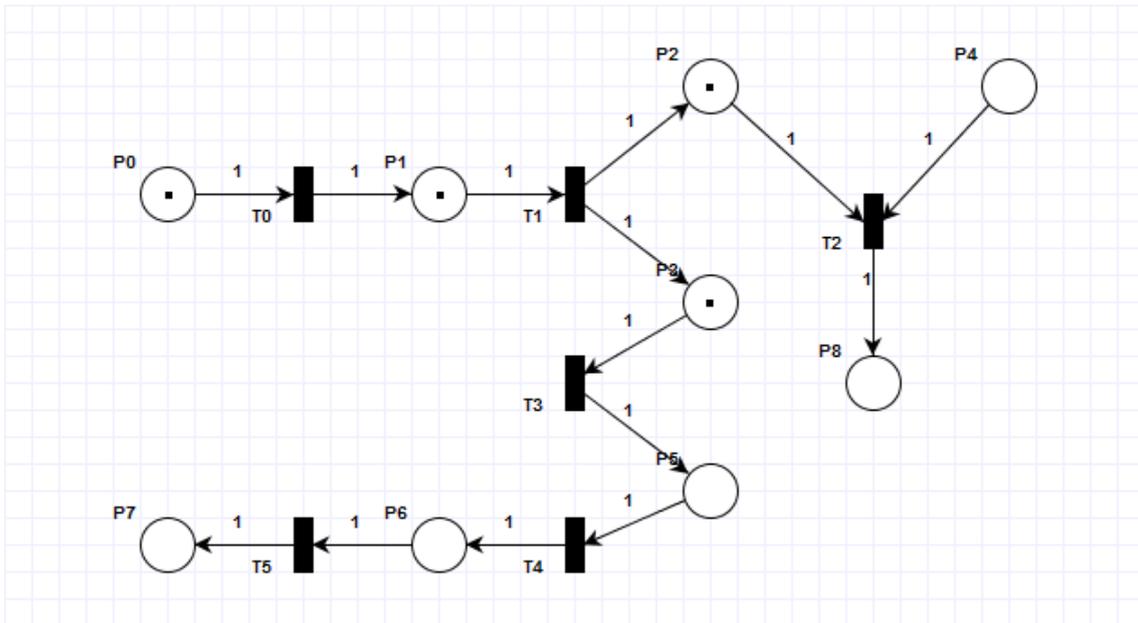


Figura 8 – Exemplo de uma rede de Petri.

Existe a possibilidade de modelar o controle de fluxo de atividades de um workflow em uma rede de Petri. As atividades são representadas por transições, enquanto relações de dependência causal entre atividades são representadas por lugares e arcos. Este caso especial de rede de Petri é chamado de *WF-net* (van der Aalst *et al.* 2004).

Capítulo 3 – Mineração de Processos

A ideia geral de mineração de processos é extrair conhecimento a partir de registros de eventos capturados por sistemas de informação (Weijters *et al.* 2006), complementando abordagens existentes de gerenciamento de processos de negócio (van der Aalst 2011). Enquanto técnicas tradicionais de gerenciamento de processos de negócio são baseadas em modelos construídos de forma manual (Weske 2007), a partir de registros documentais e opiniões pessoais, a mineração de processos possibilita a construção de modelos de processo baseada em evidências de atividades que foram registradas por sistemas de informação, eliminando a subjetividade das ideias e opiniões de indivíduos.

O artefato básico para a mineração de processos é o registro de evento. Cada evento deve corresponder a uma atividade bem definida de um processo e deve ser relacionado a outros eventos que fazem parte de uma mesma instância de processo, que é o conjunto de todas as atividades que fizeram parte de uma única execução de um processo. Estes registros podem ser extraídos de qualquer sistema de informação que armazene este tipo de dados, entre estes os sistemas de *CRM*, *ERPs*, sistemas de *workflow* e sistemas de protocolo, não se limitando a estes.

Para que seja possível realizar atividades de mineração de processos é necessário que cada registro de evento contenha informações sobre qual instância este evento fez parte, à qual atividade ele corresponde, além de sua ordem sequencial em relação ao conjunto de eventos da respectiva instância, para que possa ser estabelecido um encadeamento entre as atividades executadas durante o curso desta instância de processo, ou então a informação de quando este evento ocorreu, de forma que seja possível estabelecer uma ordem cronológica entre as atividades executadas. Informações como o responsável pela execução da atividade, o departamento ou

unidade organizacional onde esta foi realizada, entre outras, também podem fazer parte dos registros de eventos, permitindo complementar o conhecimento extraído através da mineração de processos. A Tabela 1 representa um conjunto de registros de eventos criados a partir de execuções de processos.

Tabela 1 – Exemplo de Conjunto de Registros de Eventos.

Instância	Atividade	Responsável	Departamento	Data e Hora
1	Atividade X	Pedro	RH	01/07/2000 08:40
1	Atividade A	Paulo	RH	01/07/2000 12:10
2	Atividade X	José	RH	01/07/2000 16:20
3	Atividade K	Ana	Vendas	02/07/2000 09:50
1	Atividade Z	Pedro	RH	02/07/2000 10:20
3	Atividade Y	Maria	Vendas	02/07/2000 15:10
2	Atividade A	Pedro	RH	02/07/2000 18:40
2	Atividade Y	Pedro	RH	03/07/2000 14:40
2	Atividade Z	João	RH	03/07/2000 16:30
3	Atividade Z	João	Vendas	03/07/2000 17:20
1	Atividade L	Pedro	RH	03/07/2000 18:45

Para o sucesso das atividades de mineração de processos, há duas premissas que devem ser observadas: os registros de eventos devem ser confiáveis, e o processo manifestado nos registros existe de fato. Estas premissas implicam em que todos os eventos presentes nos registros devem possuir uma atividade lógica correspondente no processo, e que existe uma forma perfeita de representação para um determinado processo (van der Aalst e Günther 2007). É importante que os eventos sejam registrados sob um nível de detalhamento em comum, para que não haja a necessidade de combinar múltiplos registros de uma atividade para que esta esteja no mesmo nível de abstração que outra a ser representada no modelo.

Os modelos de processos podem ser representados em diferentes perspectivas, podendo mais de uma estar presente em um modelo, devido à interseção de informações (van der Aalst 2011). Entre elas, podemos destacar:

- Fluxo de Atividades: esta perspectiva é focada no encadeamento e ordenação de atividades realizadas durante um processo. O mapeamento deve caracterizar os possíveis caminhos que um processo deve seguir a partir de seu início.
- Organizacional: representação do relacionamento entre os atores dentro de uma instituição e os eventos. Estes atores podem ser pessoas, departamentos, cargos. Tem como objetivo explicitar os responsáveis por cada atividade dentro do fluxograma de eventos de um processo.
- Temporal: leva em consideração a duração e frequência dos eventos, permitindo a descoberta e visualização de gargalos na execução de processos, assim como assegurar a qualidade em relação à duração das atividades e a verificação de atrasos.

A área de mineração de processos pode ser dividida em três objetivos (van der Aalst e Günther 2007): descoberta de processos, verificação de conformidade e extensão de processos.

A descoberta de processos visa construir um modelo através de algoritmos que utilizam registros de eventos de um determinado processo (van der Aalst 2012). O modelo de processo neste caso é descritivo, ou seja, é a representação de como os processos são realmente executados. Diversos algoritmos podem ser utilizados para realizar a descoberta, na Seção 4.1 deste trabalho serão apresentados e discutidos alguns deles.

A verificação de conformidade tem como objetivo avaliar a adequação de registros de eventos a um determinado modelo de processo pré-existente. Este modelo pode ter sido obtido a partir de atividades de descoberta em mineração de processos ou modelado por uma pessoa que detém conhecimento sobre como este processo é ou deveria ser executado. Neste caso, o modelo de processos é prescritivo, ou seja, é um guia que representa o fluxo de atividades de um determinado processo indicando seu

comportamento padrão para ser posteriormente comparado ao comportamento deste processo observado a partir de seus registros de eventos.

A extensão de processos tem o propósito de expandir, incrementar ou enriquecer modelos de processo existentes a partir de dados provenientes de execuções de processos, além de permitir que modelos não adequados ao que é observado nos registros de eventos possam ser corrigidos (van der Aalst 2012).

3.1. Algoritmos de Descoberta

Esta seção descreve os algoritmos de descoberta de processos, que possuem como objetivo criar modelos que representam processos com base em dados provenientes de registros de execução dos mesmos. A pesquisa desta classe de algoritmos teve início na década de 1990 a partir da necessidade de modelar processos suportados por sistemas de gerenciamento de *workflows*, paralelamente com a necessidade de descoberta de processos de negócio (van der Aalst 2012).

Os resultados apresentados pelos algoritmos de descoberta de processos podem não ser satisfatórios do ponto de vista de entendimento dos usuários, principalmente em casos onde estes algoritmos são empregados para investigar processos não estruturados, gerando modelos descritos pela literatura como em formato de espaguete. Esta denominação é devida à grande quantidade de elementos e ligações presentes no modelo. Embora complexos, estes resultados estão corretos do ponto de vista de representar o processo, descrevendo com fidelidade o que acontece no conjunto de registros de eventos utilizado como base para a construção do modelo (Günther e van der Aalst, 2007). A Figura 9 ilustra um caso de processo com aspecto de espaguete.

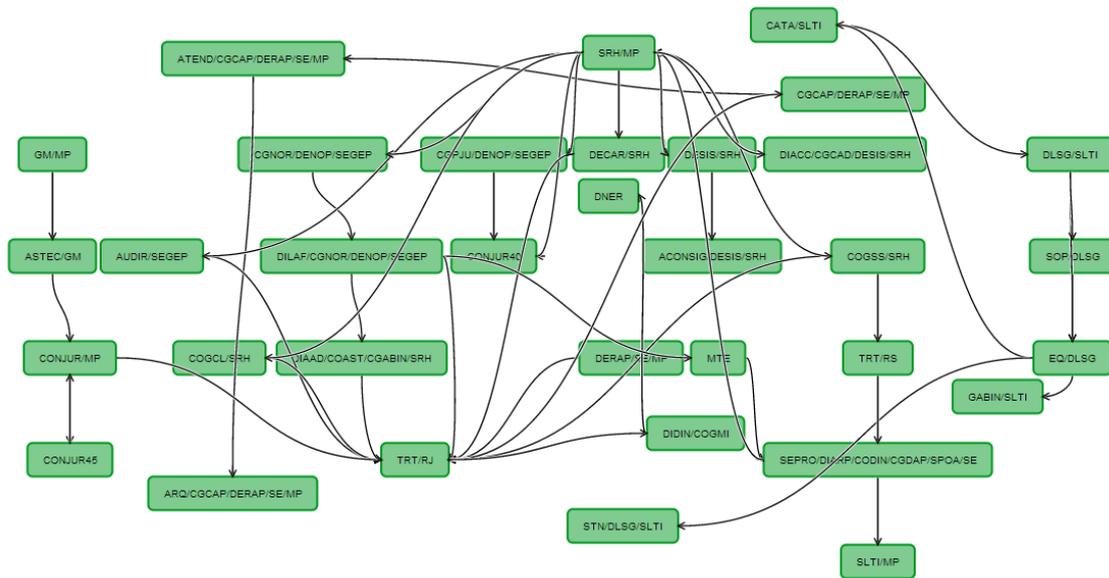


Figura 9 – Exemplo de modelo de processo com aspecto de espaguete.

Para garantir melhor qualidade e entendimento dos modelos gerados a partir de técnicas de descoberta de processos, podem ser utilizados quatro critérios: aptidão para reprodução dos registros, simplicidade, precisão e generalização. Estes critérios possuem conceitos conflitantes entre si e precisam ser balanceados de acordo com o tipo de resultado desejado (Buijs *et al.* 2012), conforme ilustrado pela Figura 10.

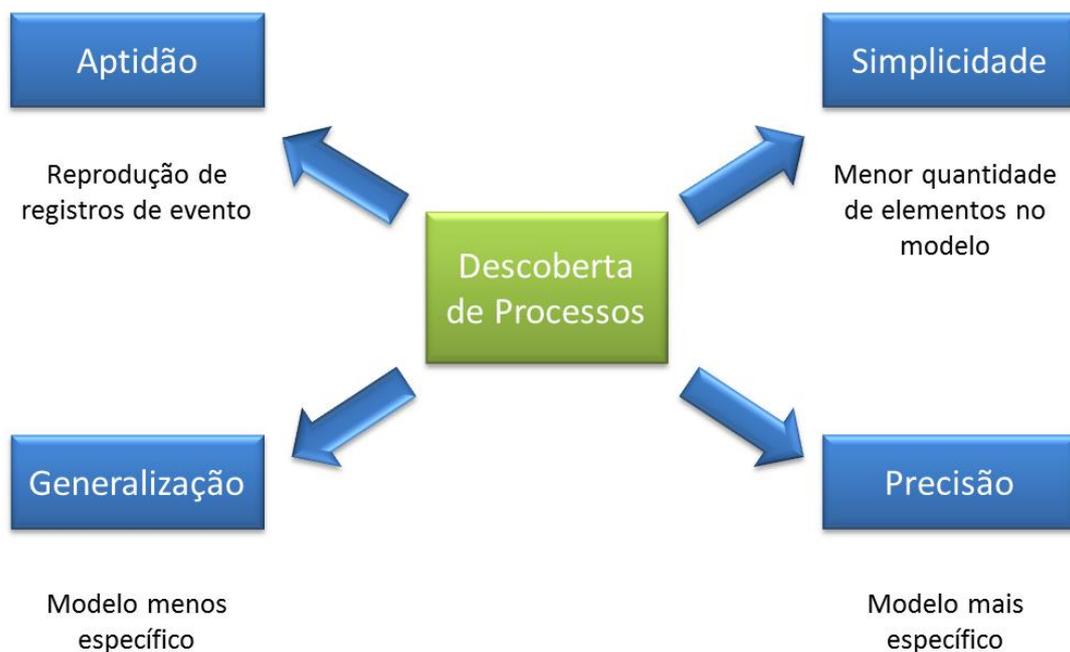


Figura 10 – Balanceamento entre critérios de qualidade. Adaptado de (Buijs *et al.* 2012).

A aptidão para reprodução de registros indica qual a capacidade de um modelo em reproduzir com fidelidade o fluxo de atividades presente nos registros de eventos de um processo. Algoritmos como o Minerador Heurístico, apresentado neste capítulo, usam este critério para basear seus resultados (Buijs *et al.* 2012).

O critério de simplicidade indica o grau de complexidade de um modelo. O modelo mais simples capaz de explicar o comportamento do processo como observado nos registros de eventos é considerado o melhor modelo. Métricas de simplicidade podem envolver a quantidade de arcos e nós presentes em um modelo. Optar por simplicidade geralmente resulta abrir mão de capacidade de reprodução de registros e da precisão.

A partir da aptidão de reprodução de registros e da simplicidade é possível criar um tipo de representação capaz de garantir avaliações próximas ao máximo de otimização para estes critérios. Uma WF-Net contendo um lugar de espera, onde todas as transições possíveis estão ligadas a este lugar, é capaz de reproduzir o comportamento de todas as observações presentes em um conjunto de registros, além de ser considerado simples por ter poucos elementos em sua representação. Este modelo, descrito como tendo a aparência de uma flor, é ilustrado pela Figura 11.

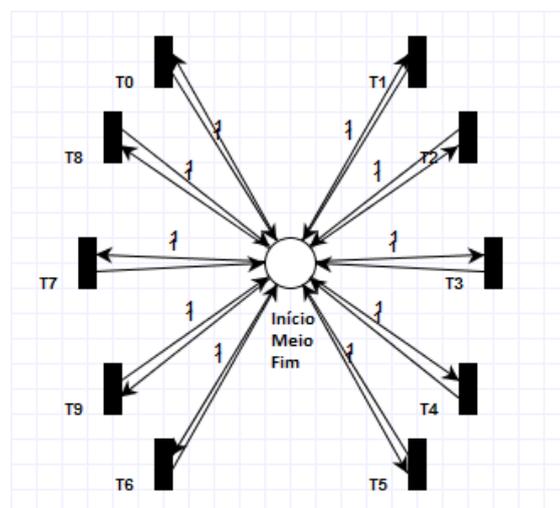


Figura 11 – WF-Net com aparência de flor.

Além deste modelo abrangente que permite todas as combinações possíveis de sequências de atividades de um processo, é possível criar um modelo restritivo que enumera as sequências de atividades presentes nos registros de eventos, também garantindo perfeita aptidão para reprodução de registros. As duas métricas restantes, a precisão e a generalização, visam equilibrar a construção de um modelo para que este não fique em um destes dois extremos, representados pelos modelos em formato de flor e enumerador descritos acima.

A precisão visa diminuir a quantidade de comportamento não observado nos registros que pode ser reproduzido pelo modelo, enquanto a generalização visa maximizar esta métrica. O tratamento dado ao ruído presente nos registros pode influenciar a criação de modelos mais precisos ou mais generalistas (van der Aalst 2011).

No restante desta seção será apresentado o Algoritmo α (van der Aalst e Günther 2007), devido à sua importância acadêmica e capacidade de resolver problemas de concorrência entre atividades de um processo. Várias ideias deste algoritmos são incorporadas em diversos estudos sobre descoberta de processos (van der Aalst 2012), sendo o próprio Algoritmo α objeto de extensões que visam solucionar suas deficiências, como o algoritmo proposto em (de Medeiros *et al.* 2003), que tem como objetivo tornar possível a detecção de ciclos no fluxo de atividades dos processos analisados.

Outro algoritmo a ser apresentado é o Minerador Heurístico (Weijters *et al.* 2006), que leva em consideração a frequência observada das sequências entre as atividades de um processo para delimitar e eliminar possíveis ruídos presentes no conjunto de registros utilizado como entrada do algoritmo. Por fim, será exposto o emprego de algoritmos genéticos na descoberta de processos. Devido à simplicidade do

algoritmo e a sua robustez em relação aos ruídos presentes nos registros de eventos provenientes de casos reais de uso, além de sua capacidade de obter resultados de forma determinística, a ferramenta desenvolvida neste trabalho faz uso somente do Minerador Heurístico para atividades de descoberta de processos.

Para a apresentação dos algoritmos definidos nesta seção serão utilizadas seguintes as notações, definidas em (Weijters *et al.* 2006):

- T é um conjunto de atividades, contendo o universo de todas as atividades existentes nos registros de eventos.
- T^* é o conjunto de todas as sequências de atividades possíveis combinando quaisquer números de atividades presentes em T .
- $\sigma \in T^*$ é uma sequência arbitrária de atividades representando uma instância de processo.
- $W \subseteq T^*$ é um multiconjunto de instâncias de processos, permitindo que instâncias de processos com a mesma sequência de atividades se repitam.

Também se faz necessário definir notações para dependências causais entre atividades. Sejam A e B duas atividades distintas que pertencem a T :

- $A >_w B$ se e somente se existe uma instância de processo σ pertencente a W onde a atividade B segue imediatamente a atividade A .
- $A \not>_w B$ se e somente se não existe uma instância de processo σ pertencente a W onde a atividade B segue imediatamente a atividade A .
- $A \rightarrow_w B$ se e somente se $A >_w B$ e $B \not>_w A$.
- $A \#_w B$ se e somente se $A \not>_w B$ e $B \not>_w A$.
- $A \parallel_w B$ se e somente se $A >_w B$ e $B >_w A$.

3.1.1. Algoritmo α

O Algoritmo α é um dos algoritmos mais frequentemente citados em pesquisas de mineração de processos, por ser um dos primeiros a resolver satisfatoriamente a questão de concorrência entre as atividades dos processos, onde uma atividade possui duas ou mais atividades sucessoras com possibilidade de serem executadas ao mesmo tempo (van der Aalst 2011).

O Algoritmo α geralmente não é utilizado para realizar descoberta de processos a partir de registros de eventos de aplicações reais devido à sua sensibilidade a ruído, que por muitas vezes estão presentes nos registros de eventos (van der Aalst 2011). Outros problemas, como a detecção de ciclos no fluxo de atividades do processo e detecção de atividades repetidas, são discutidos em estudos posteriores, não fazendo parte do escopo original do algoritmo (de Medeiros *et al.* 2003).

Este algoritmo tem como objetivo construir uma variante de redes de Petri, chamada WF-net, capaz de simular a reexecução das atividades de um processo (van der Aalst *et al.* 2004). Seu funcionamento básico é varrer os registros de eventos procurando por padrões nas sequências de atividades, assumindo que há uma relação de dependência entre duas tarefas quando ocorrem casos do tipo $A \rightarrow_w B$, onde a atividade A precede a atividade B e a atividade B nunca precede a atividade A. O Algoritmo α (van der Aalst *et al.* 2004) pode ser definido como:

Passo 1. É criado o conjunto T_w , correspondente a todas as atividades presentes no conjunto de registros W . Os elementos deste conjunto servirão como base para construir as transições na rede de Petri que será gerada por este algoritmo.

Passo 2. É criado o conjunto T_i , correspondente a todas as atividades que iniciam alguma instância de processo σ presentes no conjunto de registros W .

Passo 3. É criado o conjunto T_f , correspondente a todas as atividades que finalizam alguma instância de processo σ presente no conjunto de registros W .

Passo 4. É criado um conjunto de tuplas (A, B) , visando achar o conjunto $p(a, b)$ que representa as posições da rede, onde A corresponde às transições presentes em T_w que precedem $p(a, b)$, e B corresponde às transições em T_w que sucedem $p(a, b)$, observando que $a \in A$ e $b \in B$, e que a relação de dependência $a \rightarrow_w b$ deve ser verdadeira. Para todas as possíveis combinações entre elementos presentes em A , ilustradas por a_i e a_j , a relação $a_i \#_w a_j$ é verdadeira, e para todas as possíveis combinações entre elementos presentes em B , ilustradas por b_i e b_j , a relação $b_i \#_w b_j$ é verdadeira.

Passo 5. São eliminadas do conjunto de tuplas (A, B) todas as tuplas (A', B') que representam um subconjunto de elementos de tuplas contidas em (A, B) , restando apenas as tuplas que englobam maior quantidade de elementos.

Passo 6. É construído o conjunto de posições da rede P_w , a partir das posições restantes $p(A, B)$ adicionadas às posições $p(\text{início}, T_i)$ e $p(T_f, \text{fim})$, que representam as posições iniciais e finais da rede, respectivamente.

Passo 7. É construído o conjunto de transições da rede F_w , retornando as transições presentes nos conjuntos A, B, T_i e T_f .

Passo 8. É definido que o resultado do algoritmo α considerando o conjunto de registros de eventos W é a rede de Petri definida por (P_w, T_w, F_w) .

Nota-se que a variante de redes de Petri resultante, chamada de WF-net, contém somente posições, representando o estado atual do processo; transições, representando as atividades do processo; e arcos direcionados, que fazem a ligação entre as posições e

transições. A Figura 12 exemplifica uma WF-net que possui cinco atividades. Os círculos representam as posições e as barras representam as transições.

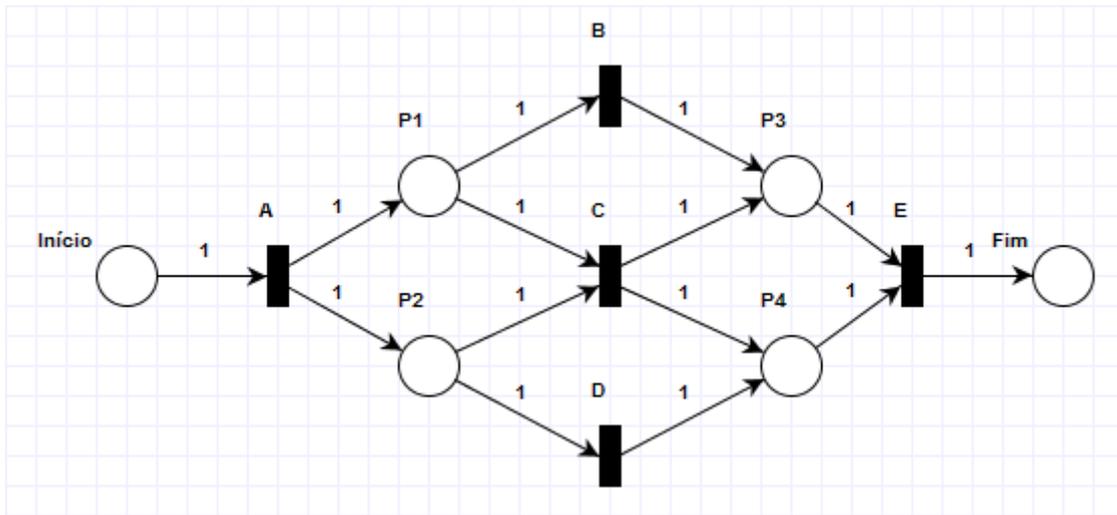


Figura 12 – Exemplo de WF-net criada a partir do Algoritmo α .

3.1.2. Minerador Heurístico

O Minerador Heurístico (Weijters *et al.* 2006) é um algoritmo de mineração de processos capaz de gerar modelos com possibilidade de diminuir a sensibilidade a possíveis ruídos presentes nos registros de eventos, apresentando robustez para utilização em situações reais que frequentemente lidam com ruído ou incompletude nos registros coletados para análise. Em vez considerar todas as relações de dependência $A \rightarrow_w B$ como no Algoritmo α , sua premissa é que somente as dependências que aparecem com maior frequência nos registros serão utilizadas na construção do modelo de processo, enquanto as dependências de menor frequência serão consideradas ruído e conseqüentemente descartadas.

O ponto de partida do algoritmo é a construção de uma matriz de dependência entre as atividades, com todas as combinações possíveis entre as atividades em T_w , a partir da métrica $A \Rightarrow_w B$ descrita na Equação 1. A notação $|A \rightarrow_w B|$ representa a

quantidade de vezes que a atividade A precede a atividade B no conjunto de registros de eventos W.

$$A \Rightarrow w B = \left(\frac{|A > w B| - |B > w A|}{|A > w B| + |B > w A| + 1} \right)$$

Equação 1 – Cálculo de dependência entre atividades.

É possível observar a partir desta equação que seu limite inferior tende a -1, quando $|B > w A|$ é um número bem maior que $|A > w B|$, e seu limite superior tende a 1, quando $|A > w B|$ é um número bem maior que $|B > w A|$. Para ilustrar o raciocínio por trás desta métrica, serão utilizados os resultados presentes na **Tabela 2**:

Tabela 2 – Variação dos valores de $A \Rightarrow w B$.

$ A > w B $	$ B > w A $	$A \Rightarrow w B$
1	0	0,5
2	0	0,66
3	0	0,75
5	0	0,83
10	0	0,90
25	0	0,96
50	0	0,98
100	0	0,99
100	1	0,97
100	5	0,89
100	10	0,81
1000	10	0,97

Nota-se que existe um limiar onde a quantidade de vezes que a atividade A precede a atividade B se torna pouco influente sobre o resultado de $A \Rightarrow w B$. Por outro lado, $A \Rightarrow w B$ começa a diminuir com rapidez conforme vão aparecendo casos $B > w A$. Assumindo que a atividade B é dependente de A e que existe um número suficientemente grande de registros em W, então o valor de $A \Rightarrow w B$ deveria ser

próximo a 1, considerando que o número $|B > w A|$ seria zero em um cenário ideal, e ordens de grandeza menor que $|A > w B|$ em caso de haver ruído.

A partir deste raciocínio, é possível definir arbitrariamente um limiar ao valor de $A \Rightarrow w B$ para considerar que existe uma dependência entre as atividades A e B. A definição do algoritmo (Weijters *et al.* 2006) prevê ainda dois outros limiares com base nas métricas explicadas anteriormente, sendo o primeiro a quantidade mínima de observações de um padrão (exemplo: $|A > w B|$ deve ser maior que 10), e o segundo a maior diferença aceitável entre o maior valor de $A \Rightarrow w B$ encontrado para qualquer outro valor de $A \Rightarrow w B$ (exemplo: 0,025).

Para ilustrar o algoritmo, supomos o conjunto de registros de eventos W a seguir, e partir deste, podemos inferir sua matriz de dependência entre atividades conforme a **Tabela 3**:

$$W = \{ 20x(ABE), 25x(ABD), 1x(BAE), 30x(ACD), 9x(ADE) \}$$

Tabela 3 – Matrix de dependência para o conjunto de registros W.

-	A	B	C	D	E
A	-	0,93	0,96	0,9	0,5
B	-0,93	-	0	0,96	0,95
C	-0,96	0	-	0,96	0
D	-0,9	-0,96	-0,96	-	0,9
E	-0,5	-0,95	0	-0,9	-

A partir da matriz de dependência entre atividades é possível criar um grafo de dependência, ilustrado pela Figura 13 e pela Figura 14, que servirá de base para o modelo de processo a ser construído. Entretanto, estes passos ainda não são suficientes para resolver questões como detecção de ciclo entre tarefas e dependências de maior complexidade entre tarefas, como relações AND e XOR.

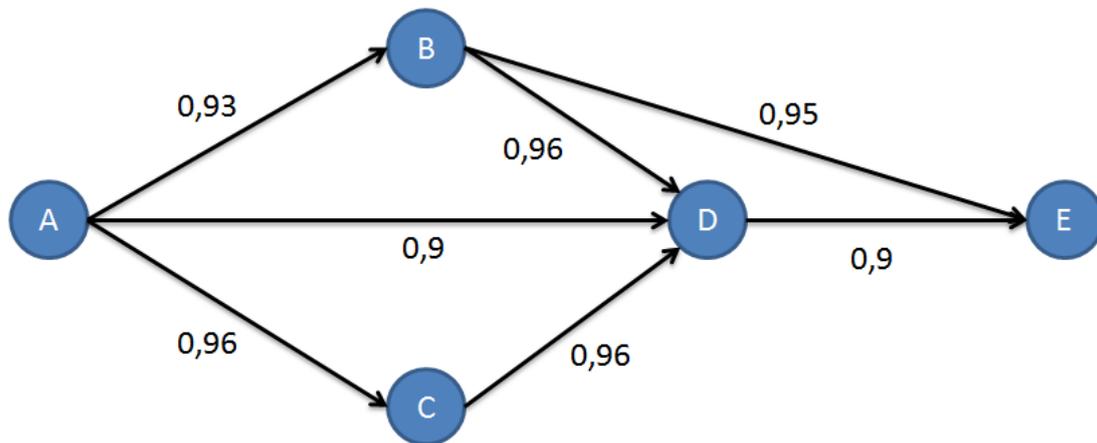


Figura 13 – Grafo de Dependência.

A Figura 13 representa um grafo de dependência usando mínimo de 0,9 para a dependência calculada pela Equação 1, com dependência mínima de 0,1 subtraídos do valor máximo obtido por esta equação, além da necessidade de cada dependência aparecer 5 vezes nos registros de eventos.

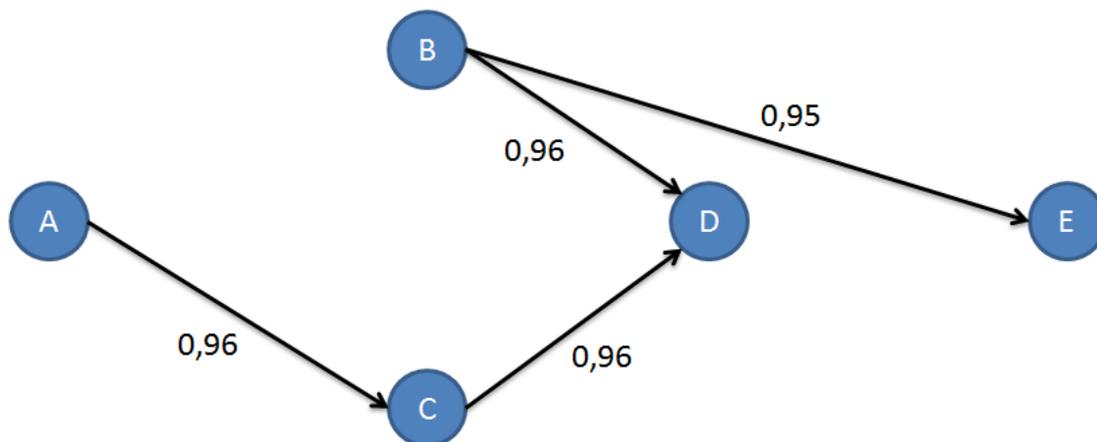


Figura 14 – Grafo de Dependência.

A Figura 14 representa um grafo de dependência usando mínimo de 0,95 para a dependência calculada pela Equação 1, com dependência mínima de 0,05 subtraídos do valor máximo obtido por esta equação, além da necessidade de cada dependência aparecer 10 vezes nos registros de eventos. Nota-se que antes era possível executar todas as atividades a partir de A, e neste grafo quase metade das arestas foram

consideradas como ruído, tornando o grafo menos conectado. Não é mais possível alcançar as atividades B e E partindo de A, e a atividade E só é possível ser alcançada partindo da atividade B.

O algoritmo propõe soluções para ciclos com um passo, como nos registros {AAD, AAAD, AAAAD}, e ciclos com dois passos, como os registros {BCBCD, BCBCBCD, BCBCBCBCD}. Para buscar ciclos de um passo, é necessário calcular um valor $A \Rightarrow w A$ de dependência para cada atividade em Tw, a partir da Equação 2:

$$A \Rightarrow w A = \left(\frac{|A \succ w A|}{|A \succ w A| + 1} \right)$$

Equação 2 – Cálculo para verificar ciclos com um passo.

Ciclos com dois passos podem ser encontrados a partir do valor de $A \Rightarrow^2 w B$, como descrito na Equação 3. A notação $|A \gg w B|$ indica a quantidade de vezes que um ciclo $A - B - A$ ocorre no conjunto de registros W.

$$A \Rightarrow^2 w B = \left(\frac{|A \gg w B| + |B \gg w A|}{|A \gg w B| + |B \gg w A| + 1} \right)$$

Equação 3 – Cálculo para verificar ciclos com dois passos.

O resultado destas novas matrizes de dependência podem ser incorporados ao grafo de dependência observando os limiares definidos como parâmetros de entrada do algoritmo.

Por fim, o Minerador Heurístico também é capaz de detectar relações AND, onde após uma atividade A ser executada, duas atividades B e C deveriam ser executadas, paralelamente ou em qualquer ordem, gerando registros como {ABC, ACB}; assim como relações XOR, onde após uma atividade A ser executada, somente uma operação B ou uma operação C poderiam ser executadas, gerando registros como

{AB, AC}, não sendo possíveis registros onde acontecem os padrões {ABC, ACB}. Para verificar ambos os tipos de relações, são calculadas as dependências entre atividades conforme a Equação 4:

$$A \Rightarrow w B \wedge C = \left(\frac{|B > w C| + |C > w B|}{|A > w B| + |A > w C| + 1} \right)$$

Equação 4 – Cálculo de dependências do tipo AND e XOR.

Para ilustrar o raciocínio por trás desta fórmula, utilizamos um conjunto de registros $W = \{ 10x(XYZK), 10x(XZKY), 10x(XOK) \}$. Calculando $X \Rightarrow w Y \wedge Z$ e $X \Rightarrow w Z \wedge Y$ obtemos o valor 0,95; enquanto os valores para ambos $X \Rightarrow w Y \wedge O$ e $X \Rightarrow w Z \wedge O$ são 0,00. Nota-se que quando existe uma relação AND entre duas atividades, o valor para a equação é alto quando comparado ao caso de existir uma relação XOR entre as atividades, quando o valor equação se aproxima de 0. Torna-se necessário então utilizar mais dois limiares como parâmetros de entrada do algoritmo, para definir o valor de corte para a Equação 4, por exemplo acima 0,8 para que uma relação de dependência seja considerada do tipo AND e abaixo de 0,1 para que uma relação de dependência seja considerada do tipo XOR, definindo o nível de ruído aceitável nos registros.

3.1.3. Minerador Genético

Outra abordagem proposta na literatura que visa lidar com os problemas de ruídos nos registros de eventos é a utilização de algoritmos genéticos em mineração de processos (de Medeiros *et al.* 2006). Algoritmos genéticos representam uma classe de algoritmos de aprendizado por máquina que se baseia no processo de evolução natural (Goldberg e Holland 1988). Um algoritmo genético é um procedimento iterativo que mantém uma população de tamanho constante de possíveis soluções para um problema.

A cada iteração do algoritmo, chamada de geração, características da população corrente são avaliadas, e uma nova população de possíveis soluções é formada a partir destas avaliações através de técnicas como combinação de características de seus indivíduos ou mutação de características (Grefenstette 1986).

Experimentos utilizando registros de eventos sintéticos e reais indicam que é possível obter modelos de processo que são completos, ou seja, reproduzem todo o comportamento presente nos registros, e também são precisos, ou seja, não permitem comportamento extra ausente dos registros (de Medeiros *et al.* 2006). Ao usar algoritmos genéticos para descobrir modelos de processos, devem-se levar em conta três aspectos: como os processos serão representados, a construção de uma métrica para avaliar a aptidão do modelo para reproduzir registros de eventos e a criação de operadores genéticos, como mutação e combinação, para garantir que todos os pontos no espaço de busca sejam alcançáveis (Medeiros *et al.* 2007).

A implantação de um algoritmo genético para mineração de processos foi proposta por de Medeiros *et al.* (2006) e disponibilizada como uma extensão para a ferramenta ProM. A representação dos processos nesta abordagem é feita através de matrizes de causalidade, que posteriormente podem ser convertidas em redes de Petri. Uma matriz de causalidade é uma tupla (A, C, I, O) , onde:

- A é um conjunto finito de atividades.
- C é uma matriz de relação de causalidades entre os elementos presentes em A.
- I são as possíveis atividades que podem acontecer antes de uma determinada atividade. A presença de atividades em um mesmo grupo indicam uma conjunção, enquanto em grupos distintos indica um disjunção. Exemplo: $I(H) = \{\{B, C\}, \{D, E\}\}$. Para que a atividade H possa acontecer, as atividades B e C devem ser realizadas, ou então as atividades D e E.

- O são as possíveis atividades que podem acontecer após uma determinada atividade.

A presença de atividades em um mesmo grupo indicam uma conjunção, enquanto em grupos distintos indica um disjunção, analogamente como no conjunto I.

A métrica de aptidão escolhida nesta abordagem está descrita nas Equações 5 e 6. Esta equação foi construída para que em um ambiente livre de ruídos, um indivíduo perfeito tivesse o resultado igual a 1, valor de seu limite superior.

$$\frac{\textit{atividades corretamente executadas} - \textit{punição}}{\textit{total de atividades}}$$

Equação 5 – Cálculo da métrica de aptidão.

O termo punição da Equação 5 equivale a:

$$\frac{\textit{fichas ausentes}}{\textit{registros} - \textit{registros com fichas ausentes} + 1} + \frac{\textit{fichas não disparadas}}{\textit{registros} - \textit{registros com fichas não disparadas} + 1}$$

Equação 6 – Cálculo do termo de punição da métrica de aptidão.

Por fim, os operadores genéticos utilizados nesta abordagem aplicam o elitismo, combinação e mutação para a construção de indivíduos para a próxima geração. Uma percentagem dos melhores indivíduos de acordo com a métrica de aptidão, chamados indivíduos de elite, são passados para a próxima geração sem nenhuma alteração. O

restante dos indivíduos são criados através de combinação, onde um par de indivíduos pais produzem um par de indivíduos filhos, combinando seus conjuntos C, I e O a partir de uma atividade escolhida aleatoriamente. Por fim, a mutação ocorre com probabilidade definida por uma taxa de mutação, adicionando ou removendo atividades do conjunto A e fazendo alterações no conjunto I e O.

A condição de parada do algoritmo é alcançada quando: é encontrado um indivíduo com valor para métrica de aptidão igual a 1; após N gerações; ou após o valor máximo da métrica de aptidão entre todos os seus indivíduos não tenha se alterado por K gerações consecutivas.

Embora seja possível encontrar um modelo com o valor da métrica de aptidão igual a 1 utilizando esta abordagem, não necessariamente este modelo encontrado possui precisão o suficiente para se aproximar do comportamento correto do processo que gerou os registros de eventos, permitindo mais casos de execução que possível no processo.

3.2. XES

O XES (*eXtensible Event Stream*) é um padrão de armazenamento de registros de eventos de processos construído pela Força Tarefa em Mineração de Processos do IEEE e publicado em 2010. Este padrão foi criado com o intuito de solucionar as limitações e substituir o padrão MXML, o principal formato de armazenamento de registros de eventos de processos utilizado por ferramentas de mineração de processos até então (Verbeek *et al.* 2011).

O padrão XES não prevê um conjunto fixo de atributos obrigatórios para os eventos ou instâncias de processo, podendo cada um destes elementos ter qualquer

número de atributos (van der Aalst 2011). A Figura 15 contém um diagrama que representa o modelo de armazenamento do padrão XES, adaptado de (Günther 2009).

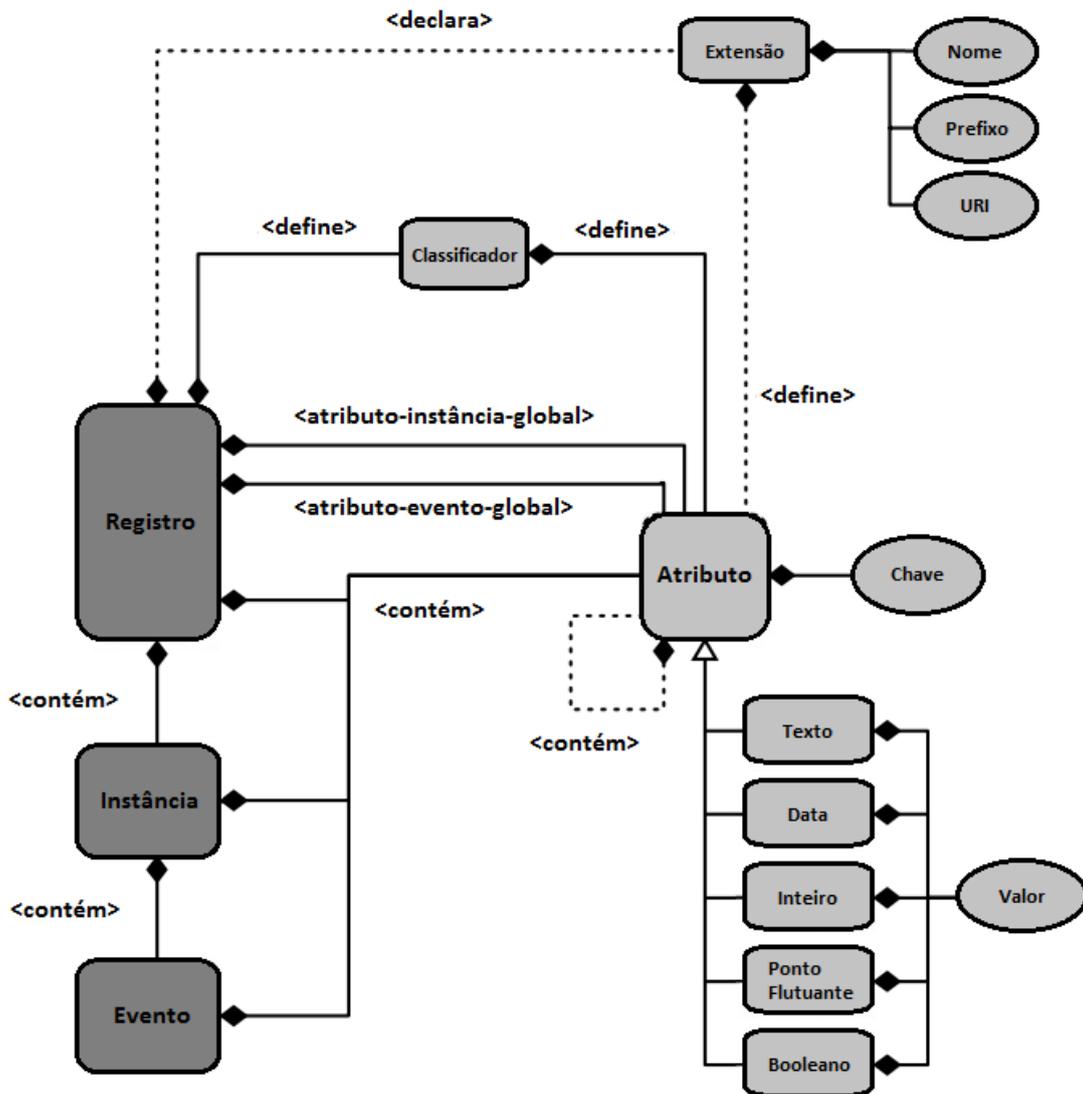


Figura 15 – Meta Modelo do padrão XES, adaptado de (Günther 2009).

Um documento XES contém um registro que consiste em muitas instâncias de processo. Cada instância de processo define uma lista sequencial de eventos correspondentes a uma execução deste processo. Cada registro, instância ou evento pode conter qualquer número de atributos, que podem ser aninhados (van der Aalst 2011). Pode-se declarar que algum atributo em particular é obrigatório, para tanto, deve-

se especificar uma lista de atributos que são válidos globalmente para todos os elementos daquele registro.

A semântica de um atributo é definida por sua extensão, que pode ser padrão ou definida pelo usuário. Existem cinco extensões que foram definidas como padrão (Verbeek *et al.* 2011):

- **Conceito.** Define um atributo com um nome claramente entendido para as instâncias e eventos presentes em um registro.
- **Ciclo de vida.** Define um atributo com a transição associada a um evento. Um exemplo de transição pode ser o início de uma atividade, um agendamento para esta ser executada, ou um evento indicando que a atividade foi finalizada.
- **Organização.** Define três atributos para um evento: recurso, papel e grupo. Um recurso é algo ou alguém que executa o evento, o papel caracteriza o tipo do recurso que foi empregado, como um cargo, e um grupo determina uma posição do recurso na organização, como um departamento.
- **Tempo.** Define um atributo com a data e hora em que um evento foi executado.
- **Semântica.** Define um atributo com uma referência a um modelo de conceito de uma ontologia, válido para cada elemento presente em um registro.

Os classificadores indicam um conjunto de atributos que, para quaisquer dois eventos cujos valores para os atributos neste conjunto são idênticos, estes dois eventos são considerados iguais para este classificador (van der Aalst 2011). Estes atributos devem ser obrigatórios. Classificadores são utilizados para mapear diversos eventos em uma mesma classe de evento durante a execução de atividades de mineração de processos.

3.3. Princípios e Principais Desafios

Em 2009 foi criada a Força Tarefa em Mineração de Processos no Instituto de Engenheiros Eletricistas e Eletrônicos (IEEE), com a finalidade de promover pesquisa, desenvolvimento, educação, implementação, evolução e entendimento na área de mineração de processos (van der Aalst *et al.* 2012). Esta força tarefa, cujos membros incluem diversas empresas e universidades que lidam com mineração de processos, foi responsável por redigir e publicar no ano de 2011 o Manifesto de Mineração de Processos (van der Aalst *et al.* 2012), um documento com declarações sobre os princípios e intenções a serem seguidas pelo grupo e por pessoas que realizam pesquisa ou atuam na área.

Os princípios apresentados no manifesto tem como finalidade a prevenção de erros que poderiam ser facilmente evitados durante a execução de atividades de mineração de processos. Estes princípios são:

Princípio 1 – Registros de eventos devem ser tratados com prioridade. Dado que a mineração de processos depende fortemente na qualidade dos registros de eventos que são processados por seus algoritmos, é importante garantir que os registros são confiáveis e estão completos de todas as suas atividades, que por sua vez são bem definidas.

Princípio 2 – A extração de registros deve ser direcionada por perguntas. Deve-se conhecer *a priori* as necessidades de conhecimento que serão objetos de estudo, para as quais as técnicas de mineração de processos buscarão a resposta. Somente devem ser extraídos para análise os registros de eventos relacionados a estas necessidades de conhecimento.

Princípio 3 – Estruturas básicas de controle de fluxo devem ser suportadas. A técnica de mineração de processos empregada deve lidar com concorrência (relação AND), ciclos e escolhas (relação XOR) no fluxo de atividades do modelo de processos que é estudado.

Princípio 4 – Os eventos registrados devem estar relacionados a elementos do modelo de processos. Deve ser fácil identificar qual elemento no registro de evento corresponde a qual elemento no modelo de processo, sem possibilidade para ambiguidade, não se limitando à perspectiva de fluxo de atividades. Elementos como unidade organizacional ou data da execução do processo devem ser representados também caso estejam dentro do escopo da mineração sendo realizada.

Princípio 5 – Modelos devem ser tratados como abstrações da realidade que este se propõe a representar. Um processo pode ser representado de diversas formas, de acordo com a necessidade de informação de cada envolvido em seu estudo. A granularidade das informações representadas e as entidades presentes no modelo devem seguir alguma abstração de realidade que seja fácil de ser entendida e que esteja de acordo com os objetivos da análise do processo sendo realizada.

Princípio 6 – A mineração de processos deve ser contínua. Devido à natureza dinâmica dos processos, que podem mudar tanto sua estrutura de fluxo de atividades quanto parâmetros como duração ou localidade em suas execuções, técnicas de mineração de processos devem ser empregadas constantemente para descobrir estas mudanças, evitando que o resultado obtido esteja defasado em relação à realidade.

O manifesto também lista quais são os principais desafios que deveriam ser abordados na área de mineração de processos na atualidade, assim como um conjunto de ações para mitigar estes desafios, deixando em aberto a adição de novos desafios em

versões atualizadas do manifesto, conforme novas necessidades surgirem a partir de avanços da pesquisa na área. Estes desafios, adaptados de (van der Aalst *et al.* 2012), são expostos a seguir, sucedidos por uma descrição sobre suas relações com este trabalho:

Desafio 1 – Identificação e curadoria de dados relativos a eventos. Os registros de eventos muitas vezes apresentam ruído, são incompletos, apresentam diferentes níveis de granularidade de informação ou não possuem dados necessários que melhorariam o entendimento do processo. Metodologias e técnicas devem ser criadas para evitar estes problemas e garantir que o **Princípio 1** seja seguido.

Desafio 2 – Suporte a registros de eventos com níveis de granularidade distintos. Algumas aplicações geram um conjunto de registros de eventos com alta granularidade, tornando difícil a aplicação de técnicas de mineração de processos devido ao grande volume dos dados sendo tratados. Outras aplicações seguem um caminho contrário, registrando menos informação que necessário nos registros de eventos para que o resultado final seja satisfatório. Ferramentas devem ser criadas para identificar quando um destes dois extremos está presente no conjunto de dados a ser tratado.

Desafio 3 – Criação de *benchmarks* representativos. Ainda não existe consenso sobre métricas de qualidade na área de mineração de processos. Critérios objetivos devem ser desenvolvidos para avaliar os resultados obtidos e o desempenho de técnicas de mineração de processo, assim como a criação de conjunto de registros que capturam características diversas importantes para os critérios de avaliação para servirem como padrão para os *benchmarks*.

Desafio 4 – Suporte a processos que possuem variação em sua execução. Alguns processos possuem comportamentos distintos variando de acordo com fatores externos,

como sazonalidade, quantidade de recursos disponíveis, etc. Mais pesquisa deve ser direcionada a esta área para entender a variação de conceito por trás dos processos.

Desafio 5 – Diminuição do viés de representação ao utilizar técnicas de descoberta de processos. Algumas técnicas de descoberta de processos estão fortemente acopladas com a representação de seus resultados. A escolha de uma linguagem ou padrão para descrever graficamente os processos pode limitar o algoritmo que o utiliza, ao propositalmente não buscar algum tipo de informação que não pode ser representada. Esforços devem ser feitos para desacoplar os algoritmos de suas representações.

Desafio 6 – Melhor balanceamento entre critérios de qualidade. A presença de ruído, assim como a não observação de todos os casos possíveis de execução de um determinado processo em seus registros de eventos, tornam difícil optar pela simplificação, generalização ou precisão na hora de construir um modelo que melhor representa o processo. Algoritmos melhores devem ser desenvolvidos visando o balanceamento entre critérios de qualidade, e todos os parâmetros utilizados por estes algoritmos devem ser facilmente entendidos para quem os utiliza.

Desafio 7 – Mineração de processos transorganizacional. Para permitir que um processo que apresenta continuidade em mais de uma organização seja minerado, é necessário que seus registros de eventos com distintas estruturas devido à forma que cada organização o armazena sejam combinados, mantendo um mesmo padrão de granularidade e abstração, o que pode ser uma tarefa com grau de complexidade elevado. Outra questão relacionada a este desafio é o segredo de informações que tornam as organizações competitivas no mercado, sendo necessária a criação de algoritmos capazes de preservar estes segredos.

Desafio 8 – Suporte operacional. Entre as principais atividades de suporte operacional estão a detecção, predição e recomendação relacionadas à execução de processos. Novas técnicas devem ser desenvolvidas para suportar estas atividades.

Desafio 9 – Combinação entre mineração de processos e outros tipos de análise de dados. Existe uma vasta gama de técnicas capazes de extrair informações a partir de dados, como as técnicas de classificação, regressão, agrupamento, descoberta de padrões, etc. Por fim, ainda existem relatórios que permitem a visualização destes dados, permitindo que humanos aproveitem sua capacidade de reconhecer padrões visualmente que técnicas automatizadas de computação não conseguem. Deve-se procurar integrar estas técnicas junto a atividades de mineração de processos.

Desafio 10 – Melhorar a usabilidade para não especialistas. Algoritmos e atividades sofisticadas de mineração de processos devem ser ocultadas atrás de interfaces amigáveis para o usuário, além de tornar possível sugerir o que pode ser feito, com intuito de aumentar as chances de sucesso na utilização das ferramentas.

Desafio 11 – Melhorar o entendimento dos resultados para não especialistas. Os resultados obtidos pelas técnicas de mineração de processos não necessariamente são corretamente interpretados, tornando-se necessário diminuir chances de não especialistas inferirem informações erradas a partir das análises geradas. As técnicas de mineração de processos também devem ser capazes de verificar possíveis problemas nos resultados obtidos, fazendo a verificação da qualidade dos mesmos, e indicando quando não há evidência o suficiente para suportar conclusões confiáveis em relação ao que é apresentado ao usuário.

Este estudo se propõe a endereçar três destes desafios. A curadoria de dados relativos a eventos de processos e o suporte a registros de eventos com níveis de

granularidade distintos são tratados pelo dicionário de dados presentes na ferramenta desenvolvida neste trabalho. Este dicionário tem como objetivo fazer a correspondência de múltiplos valores textuais a um termo compreensível pelo usuário, padronizando múltiplas grafias referentes a uma mesma entidade lógica para que estes textos sejam entendidos pelos algoritmos de descoberta de processos como sendo um valor único, permitindo a curadoria dos dados referentes a eventos ocorridos em processos. O suporte a níveis distintos de granularidade é feito a partir de relações de hierarquia entre os termos do dicionário, que possibilita realizar a inferência se o valor a ser utilizado está correto ou se um termo substituto correspondente será utilizado na criação de um modelo de processo onde todos os seus elementos pertençam ao mesmo nível de uma hierarquia.

Por fim, a ferramenta foi desenvolvida com foco no desafio de melhorar a usabilidade de sistemas capazes de realizar atividades de mineração de processos por não especialistas. Esta ferramenta implanta um ambiente onde é possível fazer uma busca exploratória em uma base de dados de registros de processos, empregando entre seus mecanismos uma linguagem gráfica de busca que permite a seleção destes registros utilizando as mesmas abstrações em que os modelos de processos são representados, evitando que o usuário necessite conhecimentos de tecnologia da informação relacionados a extração de dados dos sistemas que suportam processos de negócio para realizar esta tarefa a cada vez que um conjunto diferente de dados tenha que ser selecionado para análise, uma das barreiras de dificuldade que impede não especialistas de utilizarem ferramentas como o ProM (Esposito 2012).

Outros mecanismos que lidam com estes desafios estão descritos com mais detalhes no Capítulo 5, que expõe a ferramenta desenvolvida neste trabalho.

3.4. Ferramentas Relacionadas

Esta seção expõe as ferramentas existentes que são capazes de realizar atividades de descoberta de processos. É apresentada a ferramenta ProM (van Dongen *et al.* 2005), onde é possível trabalhar com o estado da arte em técnicas de mineração de processos através de suas extensões, e a ferramenta MANA (Esposito 2012), que possui uma abordagem diferente do ProM, permitindo a filtragem exploratória de registros de eventos de processos e apresentando um fluxo de atividades para a realização da descoberta de processos.

As ferramentas ProM Import (Günther e van der Aalst 2006) e XESame também são expostas nesta seção. Embora não estejam diretamente ligadas com a descoberta de processos em si, estas são importantes para permitir a exportação de dados referentes a processos de diversos sistemas e uma posterior adequação a formatos compatíveis com a ferramenta ProM.

Ferramentas que realizam atividades de Business Process Management e Business Process Intelligence (Grigori *et al.* 2004), cujos objetivos não estão diretamente associados à descoberta de processos, não fazem parte do escopo desta exposição.

3.4.1. ProM

O ProM (van Dongen *et al.* 2005) é uma ferramenta genérica e de código aberto que permite a criação de extensões relacionadas a mineração de processos para que sejam utilizadas em conjunto com outras extensões dentro de um mesmo ambiente padronizado. A ferramenta é capaz de utilizar registros de eventos no formato MXML, e a partir de sua sexta versão, no formato XES (Günther 2009).

A ferramenta ProM possui mais de 200 extensões (Buijs 2010) distribuídas em cinco categorias distintas (van der Aalst *et al.* 2005):

- **Mineração** – faz a implantação de algum algoritmo de mineração de processos, como a construção de redes de Petri e o Minerador Heurístico.
- **Exportação** – permite a exportação de dados e outros objetos da ferramenta, como modelos gráficos gerados durante atividades de mineração.
- **Importação** – permite abrir dados que foram exportados de outras ferramentas, como fórmulas, modelos e instâncias de processos armazenadas em outros formatos.
- **Análise** – realiza a análise de propriedades de instâncias de processos e modelos gerados em atividades de mineração de processos, como a verificação se uma rede de Petri é viva ou a avaliação do grau de conformidade de uma determinada instância de processo a um modelo.
- **Conversão** – permitem a conversão entre diferentes formatos de dados, como a transformação de um modelo para sua versão em formato de rede de Petri.

Entre suas extensões, podemos citar a implantação dos algoritmos de descoberta de processos presentes em seção anterior deste capítulo, além de extensões resultantes de outros trabalhos, expostas a seguir:

- **Minerador de Redes Sociais** – utiliza registros de eventos para gerar um mapa da rede social das pessoas envolvidas nos processos (van der Aalst e Song 2004).
- **Minerador de Decisões** – a partir de um modelo de processos e um conjunto de registros de eventos, busca descobrir qual a influência que os atributos dos registros de eventos possuem na escolha da rota a ser percorrida no fluxograma de atividades de um processo (Rozinat e van der Aalst 2006).

- **Verificador de Lógica Temporal Linear** – extensão para verificação de conformidade, faz a averiguação se os registros de eventos em estudo obedecem a regras de um comportamento desejado que é modelado em lógica temporal linear (van der Aalst *et al.* 2005).

A ferramenta ProM não propõe um fluxo de trabalho, sendo na prática uma coleção flexível de técnicas implantadas a partir de extensões que podem ser executadas a critério do utilizador do sistema. As técnicas e extensões presentes na ferramenta exigem entendimento prévio, por vezes sendo necessária a leitura de artigos científicos relacionados, o que dificulta a utilização por usuários não especialistas (Esposito 2012).

O ProM conta ainda com pelo menos onze extensões que empregam o uso de ontologias junto a técnicas de mineração de processos (de Medeiros *et al.* 2008b). Estas extensões possuem o objetivo de gerenciar as ontologias referentes a uma organização e permitir que registros de processos sejam complementados com estes dados, que passam a ser utilizáveis por outras extensões capazes de interpretar ontologias. As diferenças desta abordagem e do dicionário de dados implantado na ferramenta desenvolvida neste trabalho são discutidas no Capítulo 5.

A **Figura 16** mostra o gerenciador de extensões da ferramenta ProM. Como a ferramenta possui um grande número de extensões e sua interface depende de cada extensão, além de cada extensão poder modelar os processos estudados empregando padrões distintos, como redes de Petri ou a utilização de notação BPMN, é difícil caracterizar a interface mostrada ao usuário (van Dongen *et al.* 2005).



Figura 16 – Captura de tela do gerenciador de extensões da ferramenta ProM.

3.4.2. ProM Import

O ProM Import (Günther e van der Aalst 2006) é uma ferramenta que possui o objetivo de realizar a extração de registros de eventos de sistemas de informação que suportam processos e fazer a adequação destes registros ao padrão MXML.

Assim como o ProM, esta ferramenta permite o desenvolvimento de extensões, as quais são construídas para facilitar a extração de dados de algum sistema específico. Entre a gama de sistemas suportados pelas extensões estão versionadores de arquivos, como o Subversion e CVS, ferramentas de Business Process Intelligence, como o Aris PPM, o SGBD MS Access, além de ser possível obter a exportação a partir de arquivos com valores separados por um delimitador. Uma outra característica da ferramenta é a possibilidade de realizar a sanitização e anonimização de informações confidenciais.

O funcionamento do ProM Import se dá em duas etapas: a primeira é o mapeamento das entidades presentes nos registros de eventos em relação às entidades

correspondentes do padrão MXML, como instância de processo, eventos, tipos de evento, data de realização do evento, etc. Esta primeira etapa deve ser realizada por uma pessoa especializada que detém conhecimento sobre o sistema onde as informações são extraídas e sobre o padrão MXML. A segunda etapa é a extração dos dados em si, que pode ser realizada por qualquer pessoa quando necessário, uma vez que o especialista já configurou corretamente o mapeamento das entidades.

É importante frisar que o ProM Import não tem como objetivo realizar a exploração da base de dados, como na ferramenta desenvolvida neste trabalho e a ferramenta MANA (Esposito 2012). Neste caso, a pessoa que realizará os estudos, para cada processo, deve conhecer a priori todas as possíveis formas em que este processo pode ser armazenado no sistema que gera os registros de eventos, além de possuir conhecimento em TI para realizar o mapeamento entre seus dados e o padrão MXML, em oposto das ferramentas exploratórias, que permitem o usuário buscar quais registros de eventos serão utilizados nos estudos sem a necessidade de conhecimento em técnicas de tecnologia da informação.

A **Figura 17** é uma captura de tela do ProM Import utilizando a extensão para importação de dados a partir de arquivos com valores separados por um delimitador.

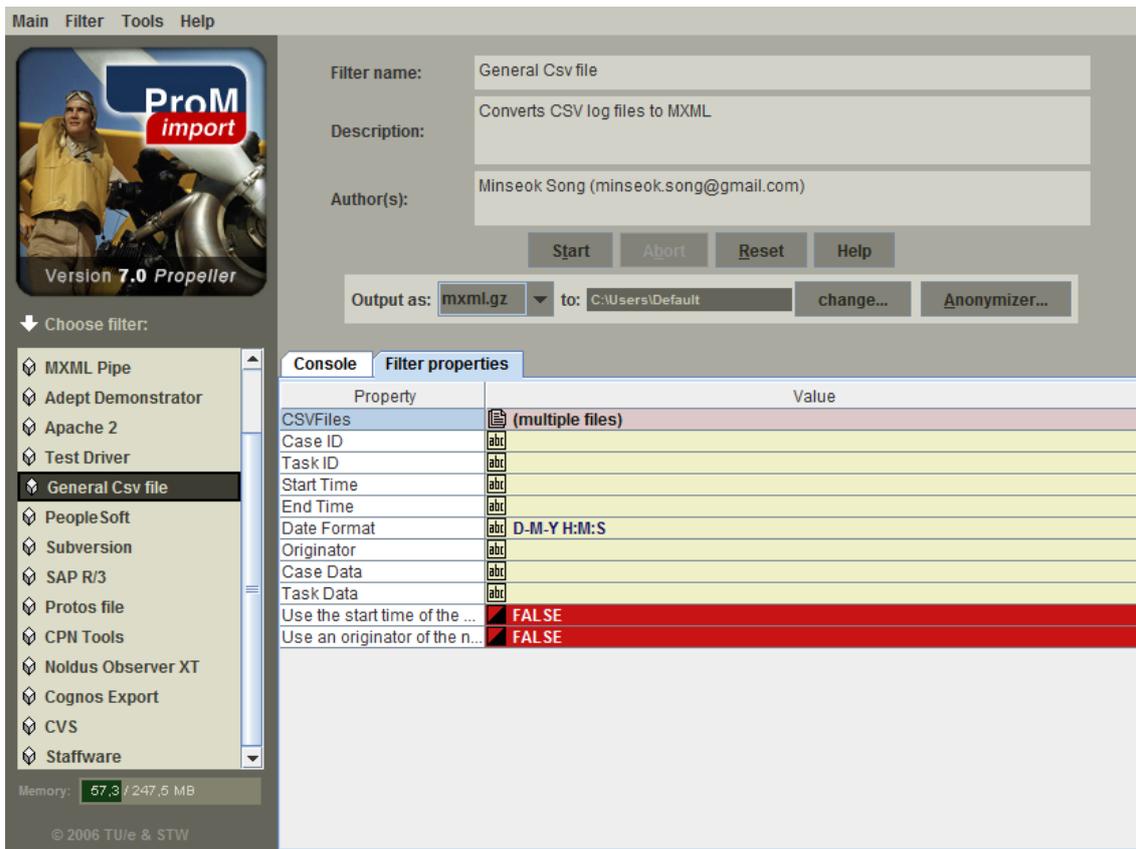


Figura 17 – Captura de tela da ferramenta ProM Import.

3.4.3. XESame

O XESame (Buijs 2010) é uma ferramenta que tem como finalidade realizar a extração de registros de eventos de sistemas de informação que suportam processos e fazer a adequação destes registros ao padrão XES (Günther 2009). O XESame vem incluído na versão corrente da ferramenta ProM na data em que esta dissertação foi escrita.

O XESame não funciona à base de extensões, sendo que a única forma de obter acesso aos dados que serão exportados é estabelecer conexão a uma base de dados através do padrão ODBC. Caso os registros de eventos estejam armazenados em algum formato que não permite acesso através do padrão ODBC, é necessária mais uma etapa

na exportação dos dados, que consiste em realizar a conversão dos dados para garantir o acesso pela ferramenta XESame.

Assim como o ProM Import, é necessário fazer o mapeamento das entidades dos registros de eventos em suas entidades respectivas que compõe o padrão XES. Este mapeamento é realizado utilizando comandos de busca de dados na linguagem SQL. O responsável por realizar a exportação dos dados deve ser um especialista em tecnologia da informação detentor de conhecimentos sobre a forma em que o processo específico alvo de estudos está armazenado no banco de dados e sobre o padrão XES, além de capaz de ter sucesso neste procedimento, limitando a quantidade de pessoas que podem utilizar o XESame. Esta ferramenta também não permite o processo de busca exploratória de processos na base de dados.

A Figura 18 ilustra uma captura de tela onde o mapeamento de atributos de uma instância de processos é realizado na ferramenta XESame.

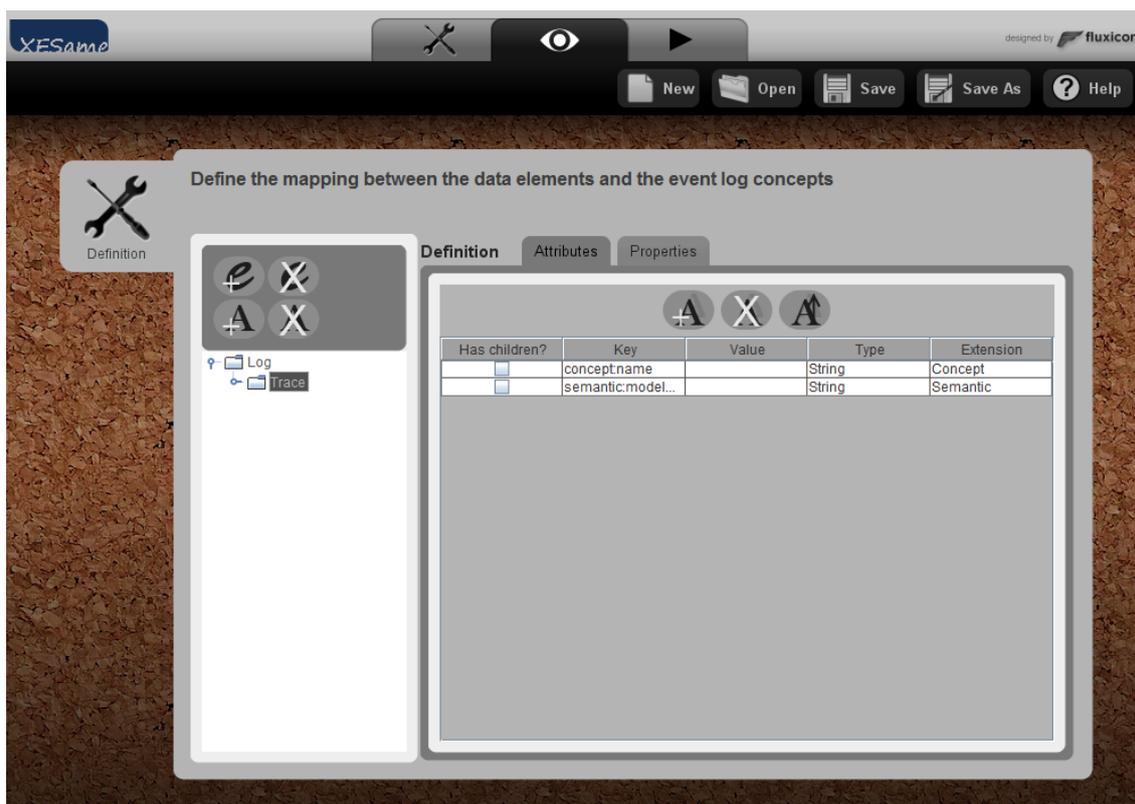


Figura 18 – Captura de tela da ferramenta XESame.

3.4.4. MANA

A ferramenta MANA (Esposito 2012) é uma ferramenta capaz de realizar atividades de mineração de processos, com foco na descoberta de processos não estruturados. Possui uma abordagem distinta ao ProM, apresentando um fluxo de atividades para seu uso e uma metodologia para ser aplicada com fins de mapeamento e reengenharia de processos.

Diferentemente do ProM, onde a mineração de processos ocorre a partir de dados presentes em arquivos no formato XML, a ferramenta MANA dá a possibilidade de realizar buscas exploratórias na base de dados de processos cadastrada no sistema, eliminando a necessidade de um especialista em tecnologia da informação construir um destes arquivos para cada processo a ser analisado, permitindo que mais pessoas possam utilizar a ferramenta. Esta busca exploratória é realizada através de filtros textuais, que são capazes de selecionar instâncias diretamente através de seus atributos, ou indiretamente através da seleção de atributos de eventos. A captura de tela representada pela Figura 19 ilustra a funcionalidade de filtro de instâncias.

MANA

Consulta em Análise: contendo 411 Instâncias de Processo - Nova Consulta

Filtros Instâncias Clustering Mineração Modelos Animação Desempenho Ajuda

Defina filtros para selecionar as instâncias que serão trabalhadas.
 Ao concluir a seleção, utilize o módulo de mineração para gerar modelos de processo.

Buscar Novos Filtros

Atributo Valor

Filtros Atuais

Nome	Comparação	Valor	Excluir
Assunto	=	SOLICITA SENHA DE ACESSO AO SIASG.	<input type="button" value="Excluir"/>
Unidade	=	MF	<input type="button" value="Excluir"/>
Unidade	=	MEC	<input type="button" value="Excluir"/>
Unidade	=	SPOA/SE	<input type="button" value="Excluir"/>
Unidade	=	STN/DLSG/SLTI	<input type="button" value="Excluir"/>
Unidade	=	EQ/DLSG	<input type="button" value="Excluir"/>
Unidade	=	MS	<input type="button" value="Excluir"/>
Unidade	=	MJ	<input type="button" value="Excluir"/>
Unidade	=	SRF	<input type="button" value="Excluir"/>
Unidade	=	MMA	<input type="button" value="Excluir"/>
Unidade	=	GABI/DLSG/SLTI	<input type="button" value="Excluir"/>
Unidade	=	GABIN/SLTI	<input type="button" value="Excluir"/>
Número	=	25000079751200411	<input type="button" value="Excluir"/>
Número	=	25002004101200915	<input type="button" value="Excluir"/>
Número	=	25003001829200543	<input type="button" value="Excluir"/>
Número	=	25008003071200457	<input type="button" value="Excluir"/>

Figura 19 – Captura de tela da ferramenta MANA.

A ferramenta MANA possui um padrão de armazenamento próprio para os dados provenientes de execuções de processos, que tem a visão de ser uma base canônica capaz de representar qualquer tipo de informação relacionada a processos. Não existe funcionalidade ou ferramenta auxiliar para efetuar a carga de dados nesta base canônica, sendo necessário haver uma pessoa especialista em tecnologia da informação para realizar esta tarefa, porém todos os dados de um sistema podem ser carregados de uma vez e filtrados posteriormente, em contrapartida do ProM, que necessita de uma carga para cada processo a ser estudado.

A descoberta de processos nesta ferramenta é realizada através do Minerador Heurístico (Weijters *et al.* 2006) e os mapas dos processos descobertos são construídos utilizando a notação BPMN. Não existe a possibilidade de criação de extensões na ferramenta para utilização de outros algoritmos de descoberta ou outras formas de

representação de processos como no ProM, sendo necessária a alteração do código fonte da própria ferramenta para que novas funcionalidades sejam inseridas.

A metodologia indicada pela ferramenta consiste em um ciclo de cinco etapas, que busca a melhoria contínua dos processos de uma organização. As etapas deste ciclo podem ser executadas mais de uma vez, de modo a que o conhecimento incremental adquirido na execução anterior ajude a produzir melhores resultados. Estas etapas estão descritas a seguir:

- **Carga de Instâncias** – através de atividades de extração, transformação e carregamento, os registros de eventos são extraídos de algum sistema e carregados na base de dados da ferramenta MANA.
- **Identificação** – através dos filtros textuais, é selecionado um grupo de instâncias que representam o processo que será alvo de estudos.
- **Mineração** – etapa onde é feita a descoberta de um modelo que representa os registros de eventos das instâncias selecionadas.
- **Análise e Visualização** – são apresentados dados sobre o processo minerado, como atraso médio de cada atividade, e é possível fazer a visualização do modelo gerado na etapa de mineração, junto com uma animação que representa a execução dos eventos, permitindo a identificação de possíveis gargalos no processo.
- **Reengenharia** – o modelo do processo que foi descoberto deve ser alterado conforme necessário a partir das análises realizadas na etapa anterior. Esta alteração então deve ser posta em prática, fazendo também uma avaliação se é preciso realizar alterações no sistema que dá suporte ao processo que é remodelado.

Capítulo 4 – Linguagem Gráfica de Busca

Este capítulo expõe uma proposta para criação de uma linguagem visual de busca capaz de selecionar registros de eventos de processos com base em seus encadeamentos de eventos. O problema que esta linguagem se propõe a resolver é categorizado, e um algoritmo para realizar operações de filtragem nestes registros de eventos de processos a partir de um modelo construído utilizando esta linguagem é apresentado.

A utilização de elementos gráficos para filtrar instâncias de processos baseadas em seus encadeamentos de eventos é importante para a realização da busca exploratória, pois permite ao usuário uma rápida visualização sobre a especificação de todos os casos de execução de um processo que serão retornados pela busca, especialmente em casos onde ocorrem paralelismos entre atividades ou desvios de fluxo. Enquanto representações textuais são limitadas a uma ordenação linear, notações gráficas podem ser mais eficientes na comunicação de ideias devido à arrumação espacial de seus elementos em um modelo (Larkin e Simon 1987).

Esta linguagem foi desenvolvida com intuito de facilitar a filtragem de instâncias em grandes conjuntos de dados relativos a processos não estruturados, onde não existe uma definição clara de como deve se comportar o processo que é analisado, e tampouco existem limites claros sugerindo quais as fronteiras para indicar que uma instância pertence a um processo. Os dados presentes em registros de eventos nem sempre apontam qual o tipo de processo referente a uma determinada instância, e quando esta informação está presente, não necessariamente ela possui relevância suficiente para ser utilizada. Este problema é ilustrado pela Figura 20.

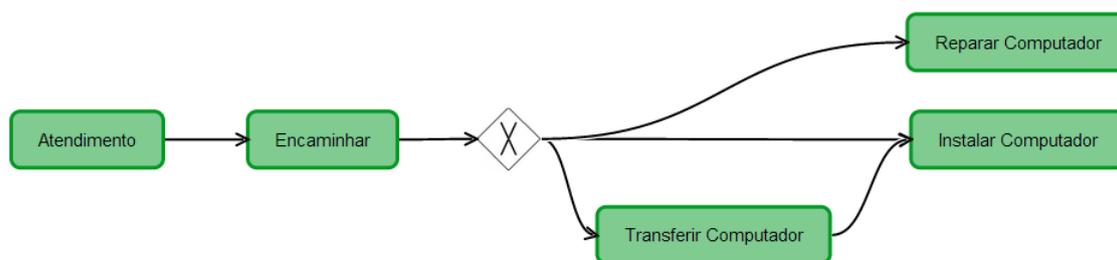


Figura 20 – Processo de Suporte de TI

A Figura 20 exemplifica três processos distintos referentes a suporte de TI, porém no conjunto de registros de eventos de processos a única informação presente sobre qual processo estas instâncias são provenientes é a que pertencem a processos cujos tipos são iguais a “suporte”. Quando é necessário descobrir informações sobre o processo de reparar computador, que é distinto dos processos de instalar e transferir de computadores, a informação que a instância pertence a um processo de suporte de TI não é suficiente para categorizá-la em um destes três tipos de processos, sendo necessário efetuar outras filtragens para se chegar ao resultado desejado.

Outro problema é a possibilidade da existência de casos onde uma instância é incorretamente classificada quanto ao seu tipo por quem cadastrou os dados no sistema. Embora algoritmos de descoberta de processos possuem um certo grau de robustez a ruídos (Weijters *et al.* 2006), torna-se necessário desenvolver novos filtros capazes de identificar ruídos para assegurar melhores resultados em atividades de mineração de processos.

A filtragem de instâncias de processos a partir de regras de encadeamento entre seus eventos permite realizar buscas em um espaço maior de informações, utilizando operadores de busca com características temporais, enquanto os mecanismos tradicionais de filtragem de registros de eventos são baseados somente em operadores de comparação aos atributos de instâncias e eventos. A Figura 21 ilustra um caso onde

somente é possível extrair corretamente as informações sobre as execuções de um processo utilizando características temporais dos eventos.

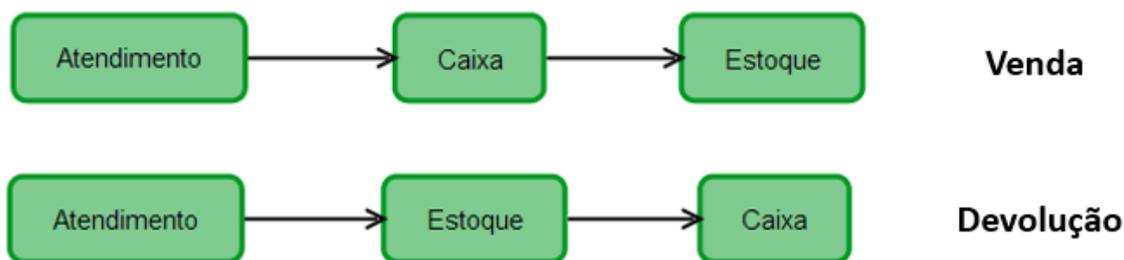


Figura 21 – Processos com mesmas atividades organizadas em ordens distintas.

Observa-se que ambos os processos passam pelos mesmos departamentos de uma organização, porém em ordens distintas. Utilizando filtros que buscam pelos setores de atendimento ao cliente, estoque e caixa, não é possível separar os processos de compra e devolução de um produto. Ao fazer a verificação temporal de que durante o processo de compra o cliente deve passar no caixa em um tempo anterior a sua ida ao setor de estoque, torna-se possível apontar diferenças entre os dois processos.

A opção por criar uma linguagem gráfica em contrapartida de uma textual se deu pela possibilidade da utilização dos mesmos elementos de representação que são empregados na representação do modelo de processos a ser descoberto ou enriquecido, com intuito de facilitar o entendimento de quem utiliza o sistema, observando o desafio de melhorar a usabilidade por não especialistas apresentado no manifesto de mineração de processos (van der Aalst *et al.* 2012). É possível fazer uma conversão direta para filtros textuais, embora verificou-se que a representação gráfica dos possíveis fluxos de atividades de um processo utiliza menos elementos que a representação textual, especialmente em casos onde existem vários desvios de fluxo, conforme explicitado no Capítulo 6.

4.1. Elementos Gráficos da Linguagem

Os elementos gráficos para representar o fluxo de atividades dos processos escolhido são baseados na notação BPMN, a mesma empregada na representação de modelos de processos que são alvos de estudo na ferramenta desenvolvida por este trabalho. Foi necessária a criação de novas abstrações para apontar possíveis fluxos entre atividades: a negação de um fluxo e a indicação que um fluxo passará por uma determinada atividade que não necessariamente será a próxima. Abaixo são expostos os possíveis elementos de representação da linguagem visual de busca e seus significados.

Filtro de Evento. Filtro que busca eventos de instâncias de processos baseado em atributos. O filtro é representado pela notação de atividade do BPMN. Ao selecionar um filtro é necessário indicar valores aos quais os atributos de um evento devem obedecer para serem retornados. O filtro deve possuir um nome significativo indicando as operações que este realiza, e por motivos de simplificação e legibilidade do modelo, estas operações devem ser explicitadas somente quando o filtro é selecionado. A Figura 22 ilustra um filtro.



Figura 22 – Representação de um Filtro.

Evento de Início e Evento de Fim. São utilizadas as mesmas notações de eventos do BPMN indicando o início e o fim de um processo. Um filtro de evento ligado a um evento de início deve indicar instâncias de processos que iniciam por eventos que possuem atributos em acordo com as operações presentes neste filtro, enquanto um evento de fim possui funcionamento análogo, porém com os eventos de instâncias sendo os últimos eventos observados na execução das respectivas instâncias.

A Figura 23 indica um filtro ligado a um evento de início à esquerda e um filtro ligado a um evento de fim à direita.



Figura 23 – Representação de Eventos de Início e Fim.

Fluxo Direto. Dois filtros ligados através de um fluxo direto, representado por uma linha contínua de cor preta, indicam uma ordem de procedência imediata entre eventos de um processo. Os eventos que obedecem ao filtro indicado pela aresta incidente devem ser observados imediatamente após os eventos que obedecem ao filtro de onde a aresta se inicia, sem nenhum outro evento entre estes dois, para que uma instância de processo obedeça a esta regra. A Figura 24 ilustra filtros conectados através de fluxos diretos.

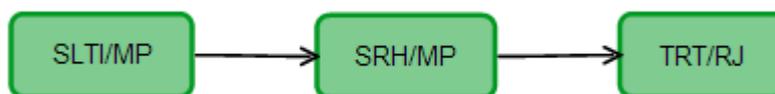


Figura 24 – Representação de fluxos diretos.

Fluxo Indireto. Dois filtros ligados através de um fluxo indireto, representado por uma linha pontilhada de cor preta, indicam uma ordem de procedência entre eventos de um processo. Os eventos que obedecem ao filtro indicado pela aresta incidente devem ser observados após os eventos que obedecem ao filtro de onde a aresta se inicia, não necessariamente de forma imediata e com possibilidade de haver um ou mais outros eventos entre os que obedecem aos filtros, para que uma instância de processo obedeça a esta regra. A Figura 25 ilustra filtros conectados através de fluxos indiretos.



Figura 25 – Representação de fluxos indiretos.

Negação de Fluxo. Representada por uma linha pontilhada de cor vermelha, os eventos que obedecem ao filtro da aresta incidente não podem ser observados após o fluxo de atividades que acaba na origem da aresta para que uma instância de processo obedeça a esta regra. A Figura 26 ilustra uma negação de fluxo.



Figura 26 – Representação de negação de fluxo.

Gateway XOR. Representado por um losango com um sinal de X em seu centro, indica que os fluxos que incidem ou são originados neste artefato devem ser exclusivos, somente um deve ser executado. Para que uma instância de processo obedeça a esta regra, deve-se verificar se o caminho observado é válido para algum destes fluxos, caso positivo, ele não pode se encaixar em nenhum outro. A Figura 27 ilustra o uso de um *Gateway XOR*.



Figura 27 – Representação de Gateways XOR.

Gateway AND. Representado por um losango com um sinal de positivo em seu centro, indica que os fluxos que incidem ou são originados neste artefato devem ser ambos executados. A Figura 28 ilustra o uso de um *Gateway AND*.



Figura 28 – Representação de Gateways AND.

4.2. Algoritmo de Busca de Instâncias de Processo

Nesta seção é apresentado um algoritmo para verificar quais instâncias obedecem ao que foi modelado na linguagem de busca visual. Embora procurou-se fazer otimizações para diminuir o tempo de busca e o algoritmo apresentado é o resultado de várias iterações com intuito de investigar o impacto no tempo total de execução, não há garantias que este é um algoritmo ótimo sob a ótica de algum critério pré-estabelecido. O algoritmo foi construído visando minimizar o tempo de busca em modelos onde há mais de um possível fluxo de atividades. Os passos e suas explicações são apresentados a seguir:

Passo 1. Todos os possíveis fluxos presentes nos critérios de busca especificados pelo usuário são identificados e separados, criando uma tabela para indicar se uma instância obedece aos critérios de busca caso obedeça a um ou mais fluxos simultaneamente. Este passo pode ser considerado como uma tradução da linguagem visual para uma linguagem que pode ser armazenada em forma textual. A Figura 29 exemplifica o processo que será buscado, enquanto a

Tabela 4 indica seus possíveis fluxos de eventos e a Tabela 5 aponta a combinação entre os fluxos para considerar se uma instância que os obedece é válida ou não. Caso existam dois ou mais fluxos desconexos na busca, ambos devem ser considerados.

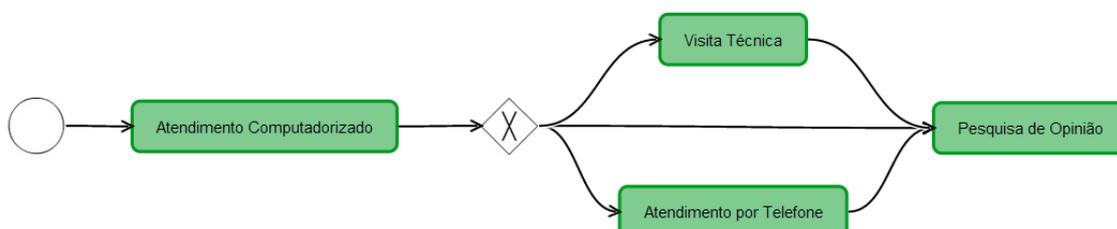


Figura 29 – Busca por processo referente a um serviço de atendimento ao cliente.

Tabela 4 – Possíveis fluxos de eventos.

Fluxo 1	
Inicia em	Atendimento Computadorizado
Continua em	Pesquisa de Opinião
Fluxo 2	
Inicia em	Atendimento Computadorizado
Continua em	Visita Técnica
Continua em	Pesquisa de Opinião
Fluxo 3	
Inicia em	Atendimento Computadorizado
Continua em	Atendimento por Telefone
Continua em	Pesquisa de Opinião

Tabela 5 – Tabela verdade indicando se uma instância que obedece aos fluxos deve ser retornada.

Fluxo 1	Fluxo 2	Fluxo 3	Resultado
Verdadeiro	Verdadeiro	Verdadeiro	Falso
Verdadeiro	Verdadeiro	Falso	Falso
Verdadeiro	Falso	Verdadeiro	Falso
Verdadeiro	Falso	Falso	Verdadeiro
Falso	Verdadeiro	Verdadeiro	Falso

Falso	Verdadeiro	Falso	Verdadeiro
Falso	Falso	Verdadeiro	Verdadeiro
Falso	Falso	Falso	Falso

Observa-se que existe uma bifurcação implicando em três fluxos possíveis no processo buscado devido a um *Gateway XOR*. Por ser uma operação de exclusividade, a instância só deve ser retornada caso obedeça a somente um destes fluxos.

Passo 2. Para cada filtro proveniente de fluxos diretos ou indiretos é feita uma verificação de quais eventos obedecem aos critérios estabelecidos no filtro, e então a instância de cada um destes eventos é retornada. Para cada possível fluxo é guardada uma lista com as instâncias que obedecem a todos os seus filtros, eliminando as instâncias que falham em algum critério em um dos filtros. Para cada filtro é adicionada uma lista de eventos que o obedece, para evitar que seja necessário realizar uma operação para verificar a conformidade do evento com o filtro novamente.

Passo 3. Para cada instância resultante do Passo 2, são retornados seus eventos em uma lista ordenada temporalmente. É feita uma verificação desta lista com o possível fluxo de atividades ao qual esta instância pode corresponder. Para cada filtro deste fluxo disposto em ordem sequencial, é feita a verificação se o evento menos recente da lista está presente na lista de eventos que obedece ao filtro em questão. Caso negativo, é feita a verificação se o próximo evento menos recente obedece a este filtro, e caso positivo, este passo é realizado para o próximo filtro da sequência. São observados casos especiais onde não é possível haver desacordo entre o evento e o filtro em questão, como no acontecimento de existir um fluxo direto ou de haver artefatos indicando início ou fim de uma execução de processo. Caso a lista sequencial de filtros seja percorrida por inteiro e não ocorra nenhuma violação de regra, a instância é

adicionada a uma lista indicando que esta obedece ao possível fluxo de atividades da busca.

Passo 4. É feita verificação para cada instância presente nas listas resultantes do Passo 3 para quais possíveis fluxos a instância é válida. Esta informação é comparada à tabela de combinação de fluxos criada no Passo 1, e caso a combinação seja válida, a instância é adicionada ao resultado final da busca.

Capítulo 5 – Ferramenta

A compreensão das atividades que são executadas em um processo dá possibilidades aos gestores de melhorar o trabalho realizado em uma organização, seja ela pública ou privada. Porém, no contexto de organizações públicas, características como a rigidez na estrutura organizacional, resistência a mudanças e descontinuidade administrativa se opõem à agilidade para desenvolvimento e implantação de soluções que atendem novas demandas da população.

A ferramenta desenvolvida neste estudo surgiu pela necessidade da realização de atividades de mineração de processos em organizações públicas brasileiras a partir de dados extraídos de seus sistemas de protocolos. As características destes sistemas, mais precisamente o fato de apresentarem um conjunto limitado de dados, por vezes havendo registros de diversos processos em um mesmo repositório onde a classificação do tipo de processo ao qual as instâncias se referem não é clara, tornam necessária a ajuda constante de pessoas com conhecimento sobre os diversos processos executados na organização para interpretação dos dados presentes nos registros de atividades. Estas características elevam o grau de complexidade das tarefas de buscar e extrair registros de eventos relacionados para que sejam utilizados nos algoritmos de descobertas de processo apresentados no Capítulo 3.

Para amenizar os problemas apresentados, a busca e seleção dos registros de eventos a serem estudados deve ser realizada de forma exploratória, onde existem várias iterações de consultas e em cada uma delas o usuário é capaz de melhorar os resultados baseado no conhecimento obtido ao realizar a busca anterior (White *et al.* 2006). A ferramenta apresentada neste capítulo prevê dois mecanismos para a realização destas buscas exploratórias. O primeiro é um mecanismo de busca textual, onde o usuário é capaz de entrar com valores para os atributos das abstrações suportadas pelo sistema

relativas aos processos de negócio, como o período de quando este foi executado, atividades realizadas, a unidade organizacional onde este ocorreu, etc. O segundo mecanismo é a implementação da linguagem de busca visual descrita no Capítulo 4 deste trabalho. Para apoiar esta busca exploratória, a ferramenta ainda prevê a possibilidade de rever os dados de todas as iterações na ferramenta que o usuário teve até chegar ao estado atual do modelo, incluindo a evolução dos filtros, os ajustes dos parâmetros dos algoritmos de descoberta de processos e os modelos resultantes que foram gerados.

A criação desta ferramenta foi motivada também pelas dificuldades observadas a partir da utilização da ferramenta MANA (Esposito 2012) com estes dados extraídos de sistemas de protocolos, representando uma evolução desta ferramenta para incluir, além de outras formas para realizar a busca exploratória, a criação de mecanismos para o tratamento destes dados. Entre as dificuldades encontradas está a não uniformidade dos dados presentes nos registros de eventos provenientes destes sistemas, resultantes da possibilidade de entrada de informações através de campos de texto livre onde o usuário pode inserir qualquer valor (Esposito *et al.* 2013). Deste modo, uma entidade como um departamento pode ocorrer nos dados sendo descrito em diversas grafias diferentes, além de não haver padronização hierárquica nos dados, podendo um mesmo departamento ser registrado sob a forma de um subdepartamento ou alguma outra unidade organizacional em nível hierárquico distinto, o que dificulta as atividades de mineração de processos e o entendimento das pessoas interessadas.

Para mitigar este problema, a ferramenta desenvolvida neste trabalho propõe a criação de um dicionário de dados, o qual possui capacidade de especificar que um determinado dado textual corresponde a uma entidade que faz sentido para o usuário, além de permitir a criação de associações e hierarquização entre as entidades presentes

no dicionário. Desta forma torna-se possível enumerar as distintas formas em que uma determinada atividade, departamento ou pessoa está representada nos dados e pode-se considerar este conjunto de grafias distintas como um elemento único durante a etapa de mineração e descoberta de processos, resultando em modelos mais simples e com menos ruídos.

Os usuários que possuem pouco conhecimento acerca de uma atividade, unidade organizacional, cargo ou qualquer outro tipo de termo pode realizar uma consulta no dicionário para buscar mais informações sobre o que se trata, sendo mais um tipo de auxílio que uma pessoa tem ao realizar atividades de mineração de processos. A partir da página que contém a definição de um termo também é possível fazer verificação de quais instâncias de processos estão relacionadas ao mesmo.

Outro benefício deste dicionário é a possibilidade da normalização do nível de detalhamento de um modelo de processo durante sua construção, utilizando elementos em um mesmo nível de granularidade de informação ou hierarquia sem que haja discrepância com uma determinada seção de um modelo com maior nível de detalhes que outra, comportamento que pode acontecer na construção de modelos de processo a partir de registros de eventos sem este tipo de tratamento.

Por fim, a criação de um dicionário de dados permite a inserção de informações na ferramenta a partir de outras fontes além dos registros de eventos de processo, ampliando a gama de informações disponíveis para a descoberta e construção de modelos de processo. Pode-se por exemplo preencher qual cargo executa uma determinada tarefa ou então carregar o organograma de uma instituição, tornando admissível que um modelo de processo seja elaborado do ponto de vista de cargos e atividades e aumentando o leque de possíveis níveis de detalhamento em modelos que levam em conta os departamentos de uma instituição.

Embora a implantação destas soluções fosse possível na ferramenta MANA (Esposito 2012), sua estrutura interna para representação de registros de execução de processos necessitava alterações para que novos tipos de dados fossem inseridos, limitando a quantidade de dados possíveis de serem utilizados na ferramenta sem que mudanças fossem realizadas. Optou-se por mudar a representação destes dados para uma estrutura que admite novos tipos de informação e compatível com o padrão XES apresentado no Capítulo 3. Esta mudança verificou-se bastante trabalhosa para ser aplicada à ferramenta MANA e resultou no desenvolvimento de uma nova ferramenta de mineração de processos, com enfoque maior em atividades de busca e descoberta de processos.

5.1. Fluxo de Atividades

O objetivo da ferramenta desenvolvida neste trabalho é apoiar a busca e descoberta de processos de negócios com base em registros de execuções de processos e dar suporte à análise destes processos. O fluxo operacional desta ferramenta, a ser apresentado com maiores detalhes ao longo deste capítulo, consiste nas etapas de importar, mapear, explorar, minerar e analisar. A ordem destas etapas indica apenas a necessidade de criação de um conjunto inicial de insumos para que a próxima etapa possa ser realizada. Uma vez que todos os insumos sejam criados, é possível executar as etapas em qualquer ordem, funcionando como módulos independentes utilizando os dados já existentes: o usuário pode sair do resultado do algoritmo de mineração, voltar à fase de mapeamento para indicar que dois atributos de eventos como dois departamentos de uma organização estão relacionados, e depois ir à fase de mineração novamente, sem necessidade de passar pela etapa de exploração, já que os insumos da etapa de mineração já estavam salvos anteriormente.

A etapa de importação consiste em extrair de um conjunto de dados referentes a execuções de processos as informações necessárias para realizar as atividades de mineração de processos, formatando o que foi extraído para inserir nas tabelas da base de dados desta ferramenta, que serão expostas na Seção 5.1 deste capítulo. Esta etapa requer o auxílio de uma pessoa que conheça técnicas de exportação e importação de dados de sistemas de informação. Devido à baixa frequência com que esta etapa é realizada e seu elevado grau de complexidade não foi planejada funcionalidade neste sistema para auxiliar a etapa de importação, deixando em aberto a possibilidade de criação de uma ferramenta auxiliar nos moldes do ProM Import (Günther e van der Aalst 2006) e XESame (Buijs 2010) apresentados na Seção 3.4 deste trabalho.

A etapa de mapeamento consiste em fazer a correspondência entre os dados que foram importados e o dicionário de dados presente na ferramenta. O objetivo é fazer o tratamento para que dados com grafias distintas sejam interpretados como fazendo parte de um único elemento e também cadastrar novos termos que correspondem a entidades que ainda não foram contempladas no dicionário de dados junto a suas associações com outras entidades. Embora seja possível seguir o fluxo de atividades da ferramenta e realizar a mineração de processos sem executar esta etapa ou até mesmo executá-la em um momento posterior, idealmente este mapeamento deve ser feito logo após a importação dos dados para que as etapas seguintes possam usufruir das informações presentes no dicionário de dados.

A exploração da base de dados de processos é realizada através da linguagem gráfica de busca descrita no Capítulo 4, que possibilita desenhar abstrações referentes a encadeamentos entre atividades, unida a filtros textuais, que permitem retornar um grupo de dados que obedecem a estes critérios. Esta exploração é realizada através de várias iterações, permitindo que refinamentos nos termos de busca sejam feitos para

retornar o conjunto de dados desejado. É possível criar filtros que agem diretamente sobre as instâncias, selecionando-as através de seus atributos, além filtros que buscam por instâncias através de seus encadeamentos de eventos.

A etapa de mineração aplica algoritmos de descoberta de processos nos registros de eventos identificados na etapa de exploração, criando um modelo estrutural descritivo correspondente ao fluxo de atividades do determinado processo, possibilitando a visualização de como o processo é executado. Caso o modelo criado não seja satisfatório, é possível refinar os parâmetros dos algoritmos e realizar a construção de um novo modelo.

Por último, a etapa de análise consiste em fazer a visualização dos processos descobertos através de suas representações gráficas e relatórios. Nesta etapa é possível ver uma animação que contém o comportamento da execução das instâncias de processos através do modelo, permitindo a identificação dos comportamentos mais frequentes e possíveis problemas de desempenho no processo, como a existência de gargalos. Também é possível fazer alterações no modelo gerado para que este se torne mais inteligível para os envolvidos em seu estudo, como também criar novos modelos de forma manual com base nos modelos de processo descobertos.

5.2. Entidades e Modelagem de Dados

Esta seção apresenta as entidades que compõem a ferramenta desenvolvida neste trabalho pertencentes a dois grupos: o armazenamento de dados provenientes de execuções de processos e a composição do dicionário de dados. A divisão é planejada de modo que cada grupo isolado possa servir como um repositório de dados para aplicações externas a partir da criação de mecanismos de acesso a estes dados, porém não foram desenvolvidas soluções para facilitar este intercâmbio de informações, ficando como uma sugestão de trabalho futuro.

5.2.1. Entidades dos Registros de Processos

O grupo de entidades apresentado nesta subseção tem como objetivo armazenar os dados provenientes de execuções de processos e foi criado com intuito de maximizar a compatibilidade com a representação de registros de eventos no padrão XES (Günther 2009), utilizado em ferramentas que dão suporte a atividades de mineração de processos. São utilizados os mesmos princípios que o padrão XES se baseia: simplicidade, onde a informação deve ser armazenada em sua forma mais simples de representação, sendo fácil de ler e escrever os dados neste formato; flexibilidade, possibilitando que os registros de eventos presentes em diversos sistemas possam ser representados nas entidades propostas; extensibilidade, permitindo que futuras alterações nas entidades não as tornem incompatíveis com suas versões anteriores; e expressividade, fazendo com que as informações representadas percam o mínimo de informação possível a partir das informações originais a partir da utilização de semânticas interpretáveis por pessoas em seus metadados.

Os conceitos principais representados por estas entidades são os de instância de processo e evento de processo. Uma instância de processo é um conjunto de um ou mais

eventos de processo que correspondem a uma única execução de um processo. Um evento de processo corresponde a uma atividade realizada durante a execução de um processo. Para garantir a flexibilidade, expressividade e possibilidade de representação no padrão XES, faz-se necessário a criação de entidades que representam os atributos e tipos de atributos que enriquecem de informação e caracterizam as instâncias e eventos de processos. A Figura 30 apresenta o diagrama de entidade-relacionamento criado para suportar estas entidades, suas descrições estão presentes na Tabela 6.

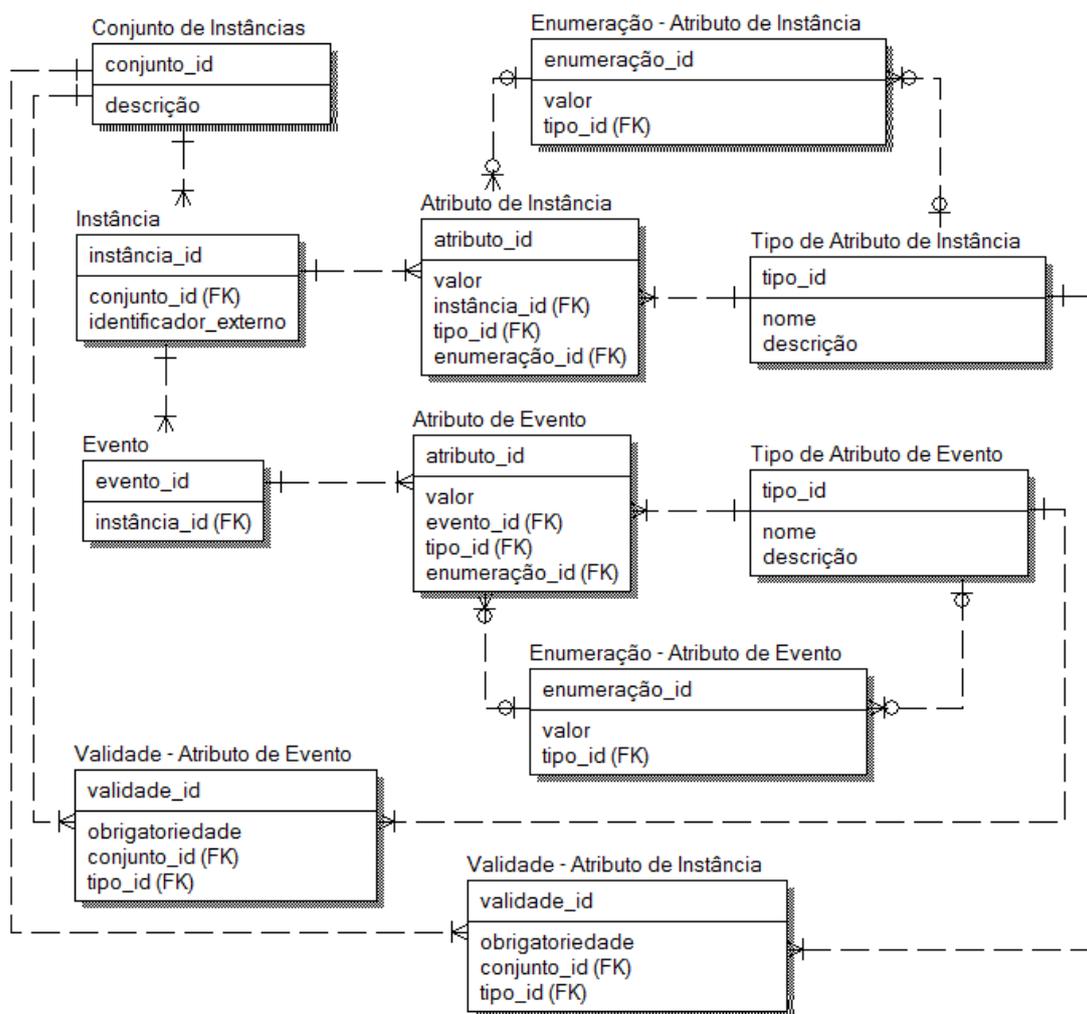


Figura 30 – Diagrama Relacional dos Registros de Processos.

Tabela 6 – Descrição das entidades relacionadas a registros de processos.

Tabela	Descrição
Conjunto de Instâncias de Processo	Indica um grupo de registros de eventos de processos oriundas de um mesmo sistema de informação que possuem informações relacionadas entre si, como atributos de eventos e atributos de instâncias em comum.
Instância de Processo	Corresponde a uma execução de um processo. Apesar de haver uma entidade específica para guardar os atributos referentes a instâncias de processos, é assumido que todas as instâncias de processos devem possuir algum identificador para que possam ser relacionadas aos dados provenientes dos sistemas onde os eventos de processos são registrados. Uma instância de processo deve estar associada a um conjunto de instâncias de processos.
Tipo de Atributo de Instância de Processo	Indica todos os possíveis tipos de atributos para as instâncias de processo. Esta lista é expansível, podendo comportar quantos tipos de atributos quanto forem necessários, sem que haja necessidade de conhecimento a priori de quais dados devem estar presentes para representar uma instância de modelo de processo. Cada tipo de atributo deve indicar como deve ser a representação dos dados que são armazenados, por exemplo: data, texto, número.
Atributo de Instância de Processo	Deve especificar um valor para um tipo de atributo a uma determinada instância de processo.
Evento de Processo	Corresponde a uma atividade executada durante o curso de um processo. Deve indicar a instância de processo à qual fez parte. Optou-se por utilizar data de execução para caracterizar a ordem em que as atividades são executadas em uma instância de processo pois durante a realização da pesquisa não foram encontrados casos de registros de eventos sem datas nos sistemas presentes nos casos de estudo e em outros sistemas que foram sondados através de entrevistas na tentativa de se obter a base de dados.
Tipo de Atributo de Evento de Processo	Similar ao tipo de atributo de instância de processo, é uma lista expansível que indica todos os possíveis tipos de atributos para eventos de processo.
Atributo de Evento de Processo	Atribui um valor para um tipo de atributo a um determinado evento de processo.
Validade de Atributos de Instâncias	Indica quais os atributos de instâncias de processo são obrigatórios e quais outros podem estar presentes para um determinado conjunto de instâncias de processo.
Validade de Atributo de Eventos	Indica quais os atributos de eventos de processo são obrigatórios e quais outros podem estar presentes para os eventos de processos presentes em um determinado conjunto de instâncias de processo.
Enumeração de Valores de um Tipo de Atributo de Evento	Contém uma lista expansível de todos os possíveis valores para um determinado tipo de atributo de evento. Esta entidade foi criada por razões de otimização das operações de busca no banco de dados e para diminuir o espaço necessário para armazenamento de dados, substituindo os N valores em texto iguais presentes na tabela de atributo de instância por somente o campo de identificação do registro correspondente a este texto nesta tabela de enumeração de valores.
Enumeração de Valores de um Tipo de Atributo de Instância	Análogo à entidade anterior, porém armazenando possíveis valores de tipos de atributo de instância.

A Tabela 7 faz uma comparação entre as entidades aqui descritas e as entidades que compõe o padrão XES descrito no Capítulo 3, indicando as correspondências. A principal diferença entre os dois tipos de armazenamento de dados é que o padrão XES permite atributos que possuem um subconjunto de outros atributos, como o caso de um atributo endereço poder possuir internamente os atributos de logradouro, número, complemento, etc. O classificador, que efetua a comparação entre dois eventos para averiguar se são iguais (van der Aalst 2011), possui uma entidade análoga nesta ferramenta que indica quais tipos de atributo de evento serão conferidos pelo algoritmo de descoberta de processos durante sua execução, que pode indicar também tipos de termo no dicionário para realizar esta comparação. A extensão, que define semântica para os atributos (Verbeek *et al.* 2011), não possui correspondência na ferramenta desenvolvida neste trabalho, enquanto as entidades de enumeração de valores de tipos de atributo foram criadas por razões de otimização, sem equivalência no padrão XES.

Tabela 7 – Correspondência entre as entidades do padrão XES e da ferramenta desenvolvida.

Padrão XES	Ferramenta Desenvolvida
Registro	Conjunto de Instâncias de Processo
Instância	Instância de Processo
Evento	Evento de Processo
Atributo	Atributo de Instância de Processo Tipo de Atributo de Instância de Processo Atributo de Evento de Processo Tipo de Atributo de Evento de Processo
Atributos de Instância Globais	Validade de Atributos de Instância
Atributos de Evento Globais	Validade de Atributos de Evento
Classificador	-
Extensão	-
-	Enumeração de Valores de um Tipo de Atributo de Evento
-	Enumeração de Valores de um Tipo de Atributo de Instância

5.2.2. Entidades do Dicionário de Dados

A seguir são apresentadas as entidades que compõem o dicionário de dados presente na ferramenta. O objetivo destas entidades é apontar entre os valores de atributos presentes nos registros de processos quais são termos correspondentes em um dicionário de dados que possui explicações ao usuário do sistema, propiciando maior compreensão do mesmo acerca destes dados. É possível criar associações entre os termos presentes no dicionário para indicar como os termos se relacionam, permitindo a construção de estruturas hierárquicas como um organograma de uma instituição, ou a construção de associações simples, como entre pessoas, atividades e cargos. Os dados presentes nos registros de eventos de processo podem ocorrer com múltiplas grafias, como abreviados em parte, escritos por extenso ou com acentuação suprimida, e o dicionário deve ser capaz de promover a identificação destes dados permitindo indicar a qual termo do dicionário o dado corresponde. A Figura 31 apresenta o diagrama de entidade-relacionamento criado para suportar estas entidades, suas descrições estão presentes na Tabela 8.

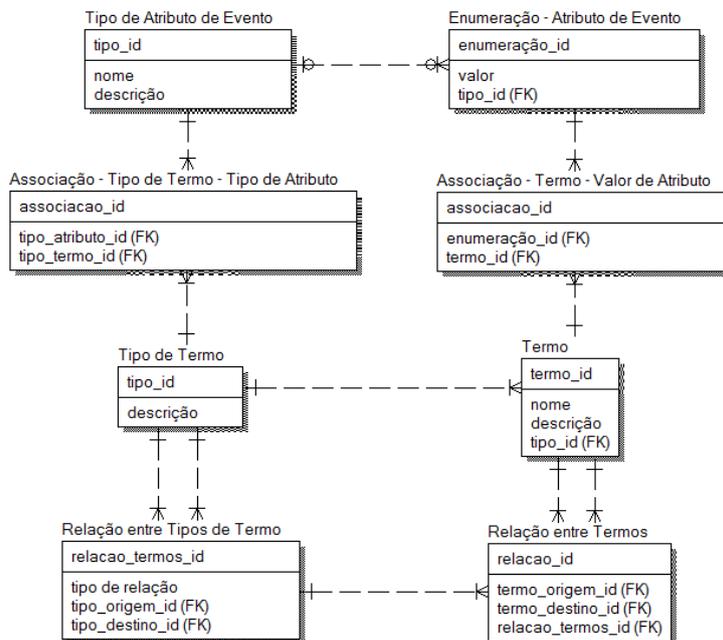


Figura 31 – Diagrama Relacional do Dicionário de Dados.

Tabela 8 – Descrição das entidades relacionadas ao dicionário de dados.

Tabela	Descrição
Tipo de Termo	Classifica um termo, devendo possuir uma descrição indicando o que correspondem os termos associados. Exemplo: <i>Departamento</i> .
Associação entre Tipos de Atributo de Evento e Tipos de Termo	Indica que os termos de um determinado tipo devem possuir correspondência com atributos de evento de um determinado tipo. Exemplo: o tipo de atributo de evento <i>Unidade Organizacional</i> corresponde ao tipo de termo <i>Departamento</i> .
Relação entre Tipos de Termo	Indica que dois tipos de termo possuem uma regra de associação. A esta regra devem ser dados um nome e uma descrição explicativas. A relação deve se enquadrar nas seguintes categorias: <i>faz parte de</i> , <i>é um</i> . Exemplo: o tipo de termo <i>Departamento</i> faz parte do tipo de termo <i>Organização</i> .
Termo	Uma entrada no dicionário que define um termo contendo um texto explicativo. Exemplo: <i>Secretaria de RH é o local da organização onde os recursos humanos são gerenciados</i> .
Associação entre Atributos e Termos	Indica que valores de atributos de evento correspondem a um termo. Exemplo: o valor <i>RH/DP</i> para o tipo de atributo <i>Unidade Organizacional</i> corresponde ao termo <i>Secretaria de RH</i> para o tipo de termo <i>Departamento</i> .
Relação entre Termos	Indica que um termo está associado a outro de acordo com uma regra de associação entre seus tipos de termos. Exemplo: o termo <i>Secretaria de RH</i> do tipo de termo <i>Departamento</i> faz parte de um termo <i>Universidade</i> do tipo de termo <i>Organização</i> .

5.3. Módulos da Ferramenta

Esta seção expõe os detalhes dos módulos da ferramenta desenvolvida neste trabalho explicando suas principais funcionalidades. A Tabela 9 faz menção aos módulos e indica onde estes estão situados no fluxo de atividades da ferramenta introduzido na Seção 5.1. As funcionalidades de gerenciamento e edição do *dicionário de dados* estão situadas na etapa de mapeamento, enquanto a consulta ao *dicionário de dados*, a *seleção de processos* e a *filtragem de instâncias* fazem parte da etapa de exploração. Há um módulo homônimo correspondente à etapa de mineração. Por fim, a *visualização e edição de modelos* e a *animação* compõe a etapa de análise. As ilustrações de uso contidas nesta seção são meramente figurativas e possuem o objetivo de servir como base para explicações, casos reais de uso da ferramenta podem ser observados no Capítulo 6.

Tabela 9 – Correspondência entre as entidades do padrão XES e da ferramenta desenvolvida.

Etapa	Módulo
Mapeamento	Dicionário de Dados
Exploração	Dicionário de Dados Seleção de Processos Filtragem de Instâncias
Mineração	Mineração
Análise	Visualização e Edição de Modelos Animação

5.3.1. Dicionário de Dados

Este módulo contém as funcionalidades de gerenciamento e consulta ao dicionário de dados presente na ferramenta. A Figura 32 ilustra a página central de onde é possível acessar o cadastro de todas as entidades relacionadas ao dicionário de dados. Nesta tela é exibida uma lista com todos os tipos de termo, suas associações com outros tipos de termo e tipos de atributo de evento e instância, quantos valores destes tipos atributos já foram associados e quantos ainda não possuem associação. Idealmente o preenchimento destas informações deve ser realizado logo após a importação dos dados na ferramenta, porém o volume de dados que necessita passar por este tratamento por vezes torna esta tarefa pouco eficiente.

Tipo de Termo	Descrição	Tipos de Atributo	Relações	Entradas	Atributos sem Associação
Órgão	Unidade Organizacional que está logo abaixo da Presidência da República na estrutura hierárquica do Governo Federal.	2	1	53	29258
Sub-Órgão	Unidade Organizacional que está logo abaixo de um Órgão na estrutura hierárquica do Governo Federal.	2	2	201	29343
Departamento	Unidade Organizacional que está logo abaixo de um Sub-Órgão na estrutura hierárquica do Governo Federal.	2	2	948	29429
Setor	Unidade Organizacional que está logo abaixo de um Departamento na estrutura hierárquica do Governo Federal.	2	2	1035	29447
Divisão	Unidade Organizacional que está logo abaixo de um Setor na estrutura hierárquica do Governo Federal.	2	2	598	29449
Sub-Divisão	Unidade Organizacional que está logo abaixo de uma Divisão na estrutura hierárquica do Governo Federal.	2	1	83	29449

Figura 32 – Página central de gerenciamento do Dicionário de Dados.

Para amenizar este problema foi criada uma funcionalidade capaz de listar os valores de atributos de evento das instâncias de processo selecionadas a partir dos filtros disponíveis na ferramenta para que estes sejam cadastrados no dicionário. É exibida uma tabela com a associação destes valores de atributo com os termos do dicionário,

permitindo que o usuário rapidamente possa averiguar quais atributos ainda não possuem associação, focando somente nos dados que de fato serão utilizados pelo usuário durante a descoberta de um processo específico. É possível realizar a associação entre termos e valores de atributo de evento a partir desta lista e efetuar a criação de termos caso não exista esta correspondência. A menor quantidade de dados retornada através de filtros criados pelo usuário facilita que este tenha maiores informações sobre o contexto em que os dados estão inseridos, permitindo verificar se as associações exibidas estão corretas e a retificação dos casos onde há erro. É necessário especificar antes da exibição da lista qual o tipo de atributo de evento será exibido para que a tabela seja criada exibindo somente os tipos de termo possíveis de serem associados. A Figura 33 ilustra esta funcionalidade, chamada de *revisar dicionário*. Neste exemplo o usuário deseja associar um termo do tipo *setor* para um atributo de evento do tipo *unidade organizacional* com valor *GABIN/SLTI*.

Revisar Dicionário

🔍 Busque ou crie um termo para associá-lo aos atributos dos eventos.

Atributo	Órgão	Sub-Órgão	Departamento	Setor
STN/DLSG/...	Ministério do Planejamento, Orçamento e Gestão	Secretaria de Logística e Tecnologia da Informação	Departamento de Logística e Serviços Gerais	
GABIN/SLTI	Ministério do Planejamento, Orçamento e Gestão	Secretaria de Logística e Tecnologia da Informação	Gabinete	

Figura 33 – Funcionalidade de revisão do Dicionário de Dados.

Uma vez preenchidos, estes dados estão disponíveis para consulta no sistema a qualquer momento e podem ser utilizados nas etapas seguintes no fluxo de atividades da

ferramenta. Outros módulos da ferramenta também fazem o emprego destes dados, seus detalhes estão presentes nas descrições dos módulos seguintes a este. A Figura 34 ilustra o atalho existente no cabeçalho de todas as telas da ferramenta que permite o acesso rápido às informações do dicionário. No exemplo exibido é realizada uma consulta para todos os termos que contém em qualquer posição o texto “*Ministério do Pla*” e são retornadas três ocorrências.



Figura 34 – Atalho para consulta aos dados presentes no dicionário.

Ao selecionar um termo do dicionário de dados é possível acessar uma tela do sistema que apresenta todas as informações relacionadas ao mesmo, ilustrada pela Figura 35. O título desta tela contém o nome do termo e logo abaixo está uma descrição entrada pelos usuários do sistema que deve possuir o objetivo de ajudar no entendimento de seu significado. Abaixo da descrição estão listados todos os termos associados com suas respectivas classes de associação e tipos de termo sendo indicados. O exemplo a seguir mostra um termo do tipo *sub-órgão* chamado *Secretaria-Executiva*, que faz parte de um *órgão* chamado *Ministério do Planejamento, Orçamento e Gestão* e que contém sete *departamentos*. Por fim, são exibidos os atributos de eventos com seus respectivos tipos que estão associados ao termo.

Secretaria-Executiva

Editar

Sub-Órgão do Ministério do Planejamento responsável por orquestrar o trabalho das demais unidades organizacionais deste Órgão.

- [Ministério do Planejamento, Orçamento e Gestão](#) (Órgão)
 - Secretaria-Executiva (Sub-Órgão)
 - [Assessoria Técnica e Administrativa](#) (Departamento)
 - [Corregedoria](#) (Departamento)
 - [Departamento de Coordenação e Governança das Empresas Estatais](#) (Departamento)
 - [Departamento de Órgãos Extintos](#) (Departamento)
 - [Diretoria de Tecnologia da Informação](#) (Departamento)
 - [Gabinete](#) (Departamento)
 - [Subsecretaria de Planejamento, Orçamento e Administração](#) (Departamento)

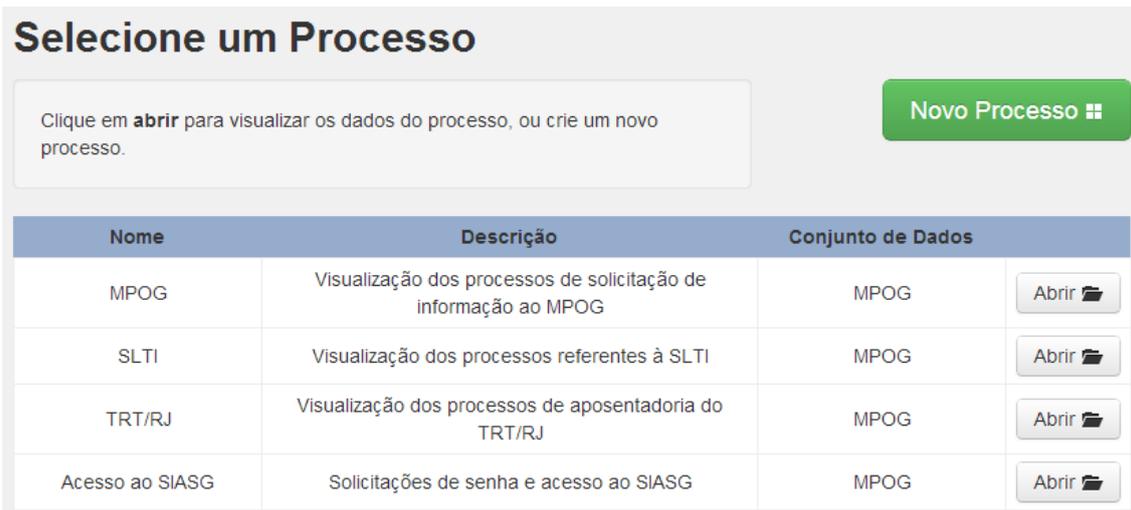
Atributos Associados		
SE/MP Unidade Organizacional (Extenso)	SE/MP Unidade Organizacional	DERAP/SE/MP Unidade Organizacional
DERAP/SE/MP Unidade Organizacional (Extenso)	DIALE/GERAP/DF/DERAP/SE/MP Unidade Organizacional (Extenso)	DIAPJ/GERAP/DERAP/SE/MP Unidade Organizacional (Extenso)
SECAP/CGCAP/DERAP/SE/MP Unidade Organizacional (Extenso)	COSIS/GERAP/DERAP/SE/MP Unidade Organizacional (Extenso)	DIEST/CGCAP/DERAP/SE/MP Unidade Organizacional
CGCAP/DERAP/SE/MP Unidade Organizacional	COFER/CGCAP/DERAP/SE/MP Unidade Organizacional (Extenso)	GEPRO/DERAP/SE/MP Unidade Organizacional
GABIN/DERAP/SE/MP Unidade Organizacional	DICLP/GERAP/DERAP/SE/MP Unidade Organizacional (Extenso)	GAB/GERAP/DF/DERAP/SE/MP Unidade Organizacional
GAB/CGCAP/DERAP/SE/MP Unidade Organizacional (Extenso)	SEDOC/DILOG/CGEAD/DERAP/SE/MP Unidade Organizacional (Extenso)	SEREC/DIVAP/GERAP/DERAP/SE/MP Unidade Organizacional

Figura 35 – Tela de informações sobre um termo do dicionário de dados.

Caso alguma informação do termo esteja incorreta é possível acessar uma tela contendo um formulário com todos os seus dados, onde pode-se realizar a associação ou dissociação com outros termos e atributos de eventos, ou ainda alterar o tipo, descrição ou nome do termo.

5.3.2. Cadastro e Seleção de Processos

A seleção do processo a ser estudado é a primeira ação que o usuário deve realizar na etapa de exploração do fluxo de atividades da ferramenta. É exibida uma tela com todos os processos cadastrados no sistema, ilustrada pela Figura 36. É possível iniciar as atividades de exploração para um novo processo acessando o formulário de cadastro de processos onde novos processos são criados.



Nome	Descrição	Conjunto de Dados	
MPOG	Visualização dos processos de solicitação de informação ao MPOG	MPOG	Abrir 
SLTI	Visualização dos processos referentes à SLTI	MPOG	Abrir 
TRT/RJ	Visualização dos processos de aposentadoria do TRT/RJ	MPOG	Abrir 
Acesso ao SIASG	Solicitações de senha e acesso ao SIASG	MPOG	Abrir 

Figura 36 – Lista de processos cadastrados no sistema.

Ao selecionar um processo é exibida uma tela contendo informações sobre todas as iterações no fluxo de atividades da ferramenta já realizadas para o processo. Uma das premissas deste trabalho é que é necessário guardar informações sobre iterações anteriores no fluxo de atividades da ferramenta de modo a aceitar que a pessoa que a utiliza possa revisitar resultados anteriores e verificar a evolução do modelo de processo que é estudado. Para cada iteração, a ferramenta desenvolvida permite a visualização dos filtros que foram utilizados na busca junto a suas instâncias retornadas, além dos parâmetros utilizados no algoritmo de descoberta de processos e seus modelos resultantes. A Figura 37 ilustra a tela da ferramenta que contém esta lista de iterações.

Histórico de Versões

TRT/RJ
Visualização dos processos de aposentadoria do TRT/RJ

Versão	Data de Criação	Qtd. de Instâncias	Instâncias	Parâmetros	Modelo
1	09/04/2013 16:00:34	28			
2	09/04/2013 16:23:12	2			
3	09/04/2013 16:27:30	2			

Figura 37 – Lista de iterações do fluxo de atividades da ferramenta de um processo.

5.3.3. Filtragem de Instâncias

Este módulo tem o objetivo de realizar a exploração dos registros de instâncias de processo, combinando filtros de busca para que um conjunto destes registros seja selecionado e utilizado como base para o modelo do processo a ser descoberto. É possível fazer esta seleção retornando diretamente as instâncias de processo com base em seus atributos, indiretamente a partir de seus encadeamentos de eventos, ou ainda combinando estas duas formas de busca.

Para o gerenciamento de filtros que agem sobre os atributos de instância existe um formulário onde é permitido escolher um tipo de atributo, um comparador e um valor e adicioná-los à lista de filtros ativos. Este formulário é ilustrado pela Figura 38. Como estes filtros são capazes de verificar se uma instância é válida a partir de um conjunto de dados menor que o utilizado para verificação de encadeamento de eventos, tornando a busca menos custosa quanto mais particulares forem estes filtros, pois o encadeamento de eventos somente é checado para as instâncias retornadas após esta filtragem inicial.

Filtros de Instâncias			
Filtros de Eventos		Filtros de Eventos (Texto)	
Instâncias Filtradas			
Atributo	Comparação	Valor	Remover
Selecione Atri	igual a	valor do atributo	Adicionar
Selecione Atributo	contém	desfazimento	Remover
Descrição	contém	equipamento	Remover
Descrição	contém	informática	Remover

Figura 38 – Formulário contendo filtros de instâncias.

Existem duas formas de especificar filtros que verificam encadeamento de eventos de processos: uma forma textual, realizada através de um formulário, onde é possível verificar todos os eventos com seus filtros em uma tela, e uma forma gráfica, que faz a implantação da linguagem gráfica de busca proposta no Capítulo 4. As duas funcionalidades são complementares, as informações preenchidas em uma seção estão disponíveis também no outro formato, possibilitando a troca do estilo de realização da busca a qualquer momento.

Ao adicionar um evento utilizando o método textual é obrigatório o preenchimento de um valor para algum de seus atributos para que este atue como filtro. Também deve ser especificado o fluxo de sequência deste evento em relação aos demais. Existem cinco opções de fluxo de sequência: iniciar, continuar, passar por, continuar e terminar, passar por e terminar. Caso seja selecionada a opção de iniciar, um evento de início é adicionado antes do evento em questão. As opções de continuar e passar por indicam fluxos diretos e indiretos ligando o novo evento ao anterior. As opções complementares que indicam o término de uma instância de processo acrescenta um evento de fim após o evento adicionado. O formulário de criação de um filtro de evento é ilustrado pela Figura 39.

Adicionar Filtro x

Fluxo	Atributo	Comparação	Valor
Selecionar Fluxo Selecionar Fluxo Começa em Continua em Passa por Continua em e Termina em Passa por e Termina em	Pessoa	igual a	José

Figura 39 – Formulário para criação de filtro de evento.

Os possíveis casos de execução das instâncias são calculados cada vez que um novo evento é adicionado ou removido. Cada caso exibe uma sequência linear com o encadeamento de seus eventos junto a seus filtros. Caso todos os filtros de atributos de um evento sejam removidos, este evento também é removido.

Esta representação textual sobre o encadeamento de eventos pode tornar-se complexa conforme a quantidade de possíveis casos de execução cresce. A Figura 40 ilustra um exemplo que possui quatro eventos e três casos de execução. Nota-se que o evento que deve possuir o atributo do tipo *Unidade Organizacional* igual ao valor *Secretaria Executiva* é repetido em todos os casos de execução possíveis para a instância. Esta forma de representação pode não ser intuitiva em episódios como a adição de um novo filtro de evento após o filtro com denominação *Unidade Organizacional*, onde um novo caso de execução seria computado e apareceria nesta lista.

Filtros de Instâncias Filtros de Eventos Filtros de Eventos (Texto) Instâncias Filtradas

▼ Adicionar Filtro

Fluxo de Eventos 1

Filtro	Atributos	Opções
Começa em Secretaria Executiva	Unidade Organizacional igual a Secretaria Executiva	
Passa por Secretaria de TI	Unidade Organizacional igual a Secretaria de TI	

Fluxo de Eventos 2

Filtro	Atributos	Opções
Começa em Secretaria Executiva	Unidade Organizacional igual a Secretaria Executiva	
Passa por Dept. de Logística	Unidade Organizacional igual a Dept. de Logística	

Fluxo de Eventos 3

Filtro	Atributos	Opções
Começa em Secretaria Executiva	Unidade Organizacional igual a Secretaria Executiva	
Passa por Dept. de Serviços Gerais	Unidade Organizacional igual a Dept. de Serviços Gerais	

Figura 40 – Forma textual de busca por encadeamento de eventos.

O mesmo exemplo representado usando a linguagem gráfica de busca é ilustrado pela Figura 41, onde cada evento é exibido uma vez. Nota-se que é intuitivo fazer a verificação dos possíveis casos de execução presentes no diagrama, que possui uma quantidade menor de elementos para representar uma mesma ideia. Alguns elementos textuais são substituídos por símbolos, como a adição de um evento de início antes de filtros que possuem o fluxo de sequência que indica início da instância de processo, e fica claro que um novo caso de execução será criado ao ser adicionado um filtro após o filtro denominado *Secretaria Executiva* devido à presença de um *gateway XOR*.

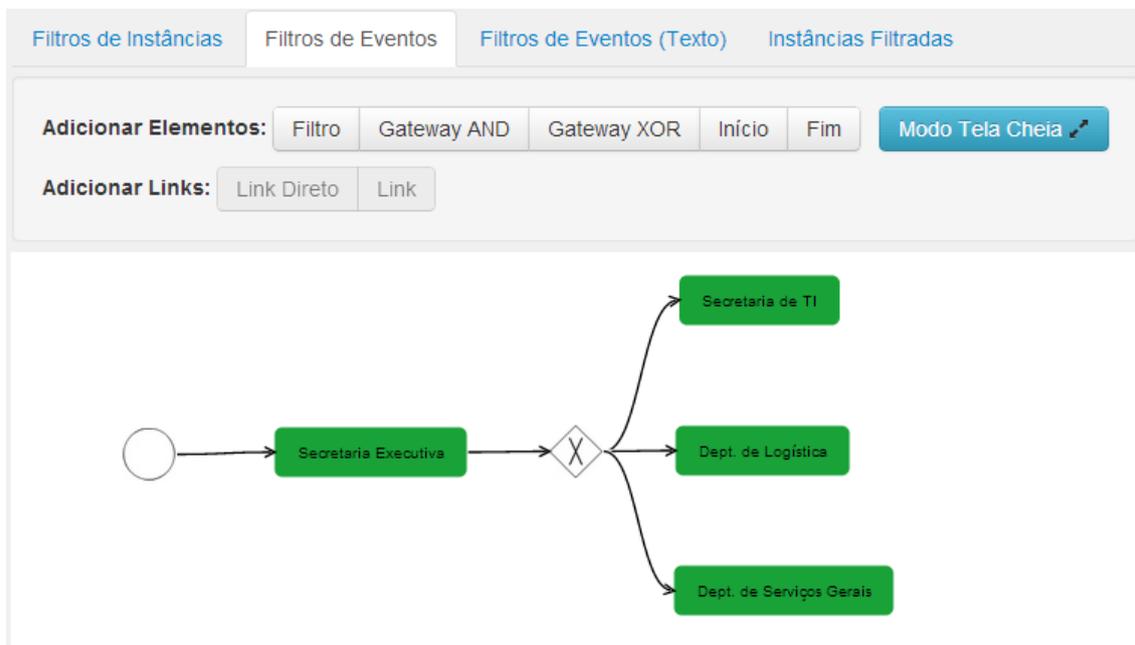


Figura 41 – Utilização da linguagem gráfica de busca para adição de filtros de eventos.

Acessando a funcionalidade que faz a implantação da linguagem de busca gráfica é exibida uma área para criação de filtros que representam o encadeamento de eventos das instâncias de processo a serem retornadas. Há como adicionar um Filtro de Evento, *Gateway AND*, *Gateway XOR*, Evento de Início e Evento de Fim nesta área, que podem ser conectados por fluxos diretos, indiretos e negações de fluxo. Ao selecionar um filtro de evento, as informações sobre os tipos, comparadores e valores de atributos que este evento deve obedecer são exibidas, onde é possível alterar o nome deste filtro de evento no diagrama e adicionar ou remover atributos a serem comparados.

Ao realizar a busca, é retornado um conjunto inicial com as instâncias que obedecem aos filtros de atributos de instâncias, e depois são realizados os passos do algoritmo descrito no Capítulo 4, ignorando as instâncias que não estão presentes neste conjunto inicial. O resultado da busca é salvo pela ferramenta e a tela exhibe o último resultado encontrado, evitando que todas as operações de busca sejam calculadas a cada vez que este conjunto de dados seja requisitado. A Figura 42 ilustra a tela do sistema

contendo as instâncias filtradas. Nesta tela é exibida uma lista contendo as instâncias juntamente a seus atributos e eventos ordenados cronologicamente.

Filtros de Instâncias Filtros de Eventos Filtros de Eventos (Texto) Instâncias Filtradas

388 Instâncias selecionadas.

Agrupamento	Quantidade de Instâncias	
Todas as Instâncias do Processo	388	Visualizar Instâncias

Atualizar Instâncias

Mostrando instâncias de 1 a 10.

Instância	Quantidade de Eventos	Início	Fim	Visualizar
2238759	8 eventos	30/07/2007	12/11/2007	[Expandir]
2377275	7 eventos	04/04/2008	09/09/2008	[Expandir]
2375896	7 eventos	02/04/2008	09/09/2008	[Expandir]
2370742	7 eventos	25/03/2008	08/09/2008	[Expandir]
2368184	7 eventos	19/03/2008	19/06/2008	[Expandir]
2375627	7 eventos	02/04/2008	21/05/2008	[Expandir]
2369901	6 eventos	24/03/2008	08/09/2008	[Expandir]
2368180	7 eventos	19/03/2008	19/06/2008	[Expandir]
2354266	7 eventos	25/02/2008	11/03/2008	[Expandir]
2377689	6 eventos	04/04/2008	07/04/2008	[Expandir]

Início 1 2 3 4 5 Fim

Figura 42 – Lista contendo as instâncias retornadas pela busca.

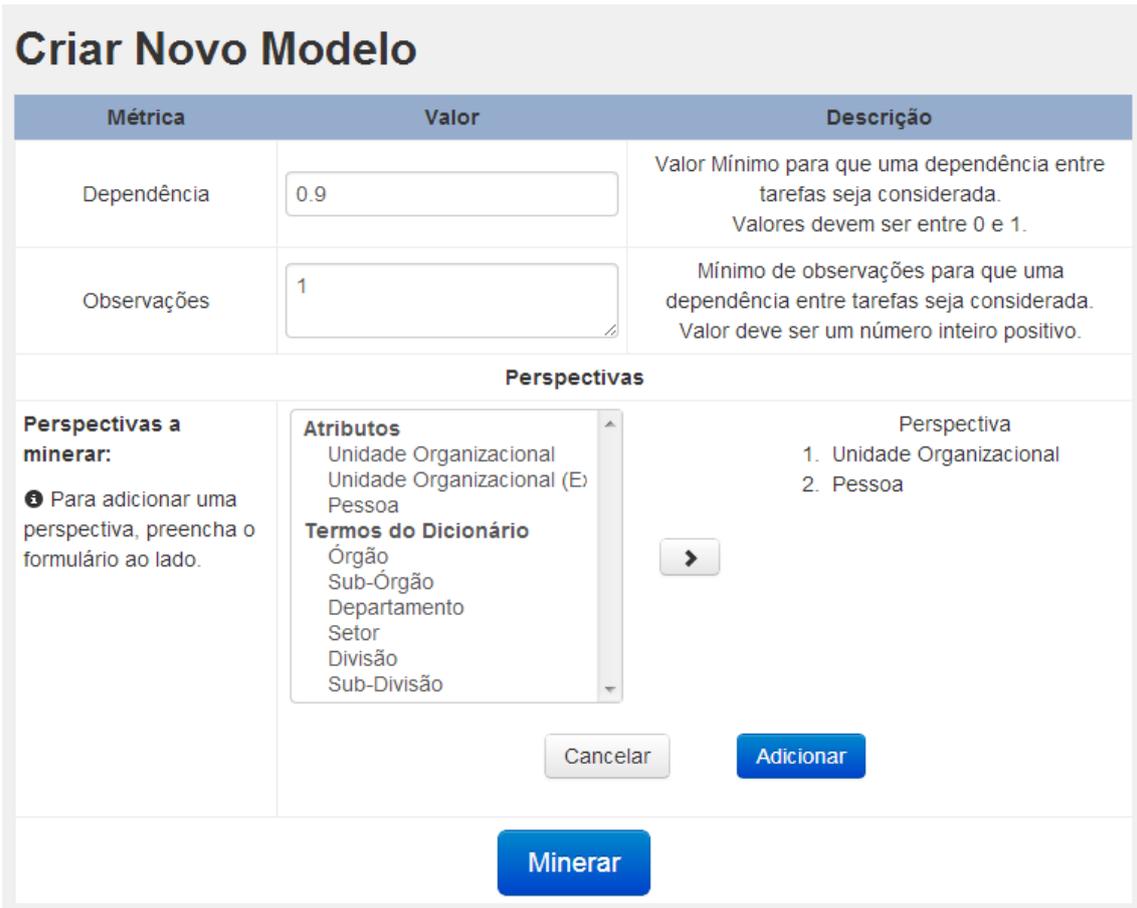
5.3.4. Mineração

O módulo de mineração é responsável por realizar as atividades de descoberta de processos na ferramenta. As instâncias de processo selecionadas na etapa de exploração são utilizadas pelo algoritmo de descoberta em conjunto com seus parâmetros de configuração para gerar modelos de processos. O Minerador Heurístico foi escolhido o algoritmo para a realização destas tarefas, decisão motivada por sua capacidade em lidar com registros de processos reais onde é possível a existência de ruído (Weijters *et al.* 2006) e pela reprodutibilidade de seus resultados a partir de um mesmo conjunto de dados e parâmetros. A ferramenta admite a implantação de outros algoritmos de descoberta de processos sem que seja necessário alterar as funcionalidades presentes em outros módulos.

O usuário pode escolher sob qual perspectiva os dados serão analisados durante a construção do modelo, resultando em esquemas que podem representar as tarefas realizadas no processo, as pessoas envolvidas, os cargos que realizaram as tarefas, os departamentos onde as tarefas foram executadas, etc. É possível gerar modelos sob a perspectiva de qualquer combinação entre tipos de atributo de evento e tipos de termo do dicionário. Cada evento no modelo descoberto será representado pela combinação de dados escolhida como perspectiva.

Além de escolher as perspectivas, o usuário deve entrar com os valores dos parâmetros do algoritmo que são utilizados para construir o modelo. Devido à complexidade de explicar de forma clara e satisfatória os conceitos do que cada parâmetro representa e seus efeitos no modelo resultante às pessoas que utilizam a ferramenta, optou-se por deixar que o usuário escolha os valores para somente dois parâmetros, fixando valores para os demais. Estes dois parâmetros são a quantidade mínima de ocorrências que uma atividade necessita para aparecer no modelo e o valor

mínimo para a métrica que calcula a dependência entre duas atividades para que seja adicionado um fluxo de sequência entre estas duas atividades, ambos explicados com detalhes no Capítulo 3. É possível alterar o formulário que registra os parâmetros a serem utilizados no algoritmo para que valores de outros parâmetros possam ser atribuídos livremente. A Figura 43 ilustra a tela do sistema onde pode-se escolher quais perspectivas serão utilizadas para a criação de um modelo de processos e o formulário contendo os parâmetros supracitados.



Criar Novo Modelo

Métrica	Valor	Descrição
Dependência	0.9	Valor Mínimo para que uma dependência entre tarefas seja considerada. Valores devem ser entre 0 e 1.
Observações	1	Mínimo de observações para que uma dependência entre tarefas seja considerada. Valor deve ser um número inteiro positivo.

Perspectivas

Perspectivas a minerar:
 ⓘ Para adicionar uma perspectiva, preencha o formulário ao lado.

Atributos

- Unidade Organizacional
- Unidade Organizacional (E)
- Pessoa

Termos do Dicionário

- Órgão
- Sub-Órgão
- Departamento
- Setor
- Divisão
- Sub-Divisão

Perspectiva

1. Unidade Organizacional
2. Pessoa

Figura 43 – Formulário com as informações utilizadas no algoritmo de descoberta de processos.

É importante salientar que a utilização do dicionário de dados durante a descoberta de processos permite que eventos pouco frequentes, que ocasionalmente ficariam de fora do modelo descoberto por não possuírem valor para alguma métrica acima do limiar desejável, sejam agrupados por decorrência de serem compatíveis com

um mesmo termo, gerando valores acima deste limiar. Um exemplo deste caso é quando dois setores não possuem frequência mínima de ocorrência nos registros de processos selecionados, porém quando são agrupados em uma única gerência, a quantidade de vezes que este novo elemento aparece é superior ao mínimo necessário.

Ao iniciar o processo de mineração, a ferramenta faz a transformação dos dados referentes às instâncias de processos selecionadas na etapa de exploração para o formato XES para que estes possam ser utilizados pela biblioteca que calcula todas as métricas referentes ao Minerador Heurístico. O modelo de processo é construído fazendo a comparação entre os parâmetros definidos pelo usuário e o resultado do cálculo das métricas, adicionando, por exemplo, as atividades que possuem mais ocorrências que o limiar indicado pelo parâmetro presentes no vetor que informa quantas vezes uma atividade ocorreu. Por fim, os elementos do modelo descoberto são passados a uma biblioteca de visualização de grafos para que estes sejam ordenados de forma inteligível em um diagrama e as informações sobre a posição de cada elemento neste modelo são salvas para que sejam utilizadas pelas funcionalidades de visualização de processos da ferramenta.

5.3.5. Visualização e Edição de Modelos

Este módulo é responsável por mostrar os resultados das atividades de descoberta de processos ao usuário e permitir que os modelos gerados possam ser alterados para facilitar seu entendimento por todos os interessados. É possível criar um novo modelo com base em algum modelo descoberto para que este seja modificado livremente, permitindo, entre outras coisas, que o modelo criado manualmente sirva como base para explicitar mudanças necessárias para otimização do processo em questão.

Para a representação dos modelos de dados foi escolhida a notação BPMN (OMG 2011), descrita no Capítulo 2, devido à sua simplicidade para construção de modelos e eficiência em comunicar ideias sobre o funcionamento dos processos descritos utilizando a notação. Dentre o conjunto de elementos de representação do BPMN são utilizados somente os fluxos de sequência, atividades, *gateways AND* e *XOR* e eventos de início e fim. A Figura 44 ilustra a visualização de modelos de processo da ferramenta.

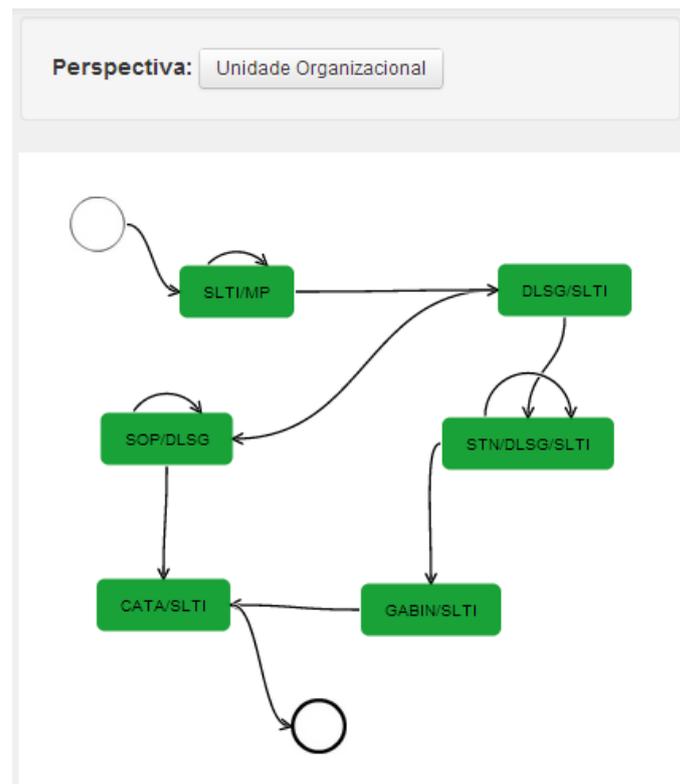


Figura 44 – Modelo de processos descoberto pela ferramenta.

A ferramenta propicia a análise de um processo sob perspectivas diferentes, escolhidas durante a etapa de mineração. Para facilitar o entendimento da correspondência entre elementos de modelo nas diversas perspectivas em que estão presentes, pode-se mudar a cor de um elemento e verificar estas alterações refletidas nos elementos relacionados que fazem parte de outras perspectivas. Esta associação entre

elementos de modelos de diferentes perspectivas é inferida através do dicionário de dados.

A partir da visualização de modelos de processo também é possível realizar atividades de exploração de registros de eventos de processos. Ao escolher um ou mais elementos do modelo de processo existe a opção de verificar quais instâncias de processo percorrem aquele caminho, exibindo uma lista com os atributos destas instâncias agrupados por valor e ordenados por frequência. Após a análise dos atributos, é possível fazer a seleção de parte destes e acionar novamente o algoritmo de descoberta de processos, utilizando somente o subconjunto de instâncias selecionadas através dos atributos e os mesmos parâmetros de entrada. Também é possível retornar ao formulário presente no módulo de mineração e alterar as perspectivas e parâmetros do algoritmo. A Figura 45 ilustra uma tela do sistema durante a realização deste procedimento.

Atributo	Quantidade
<input type="checkbox"/> DESFAZIMENTO DE EQUIPAMENTO DE INFORMATICA.	388
<input type="checkbox"/> ENCAMINHA A RELACAO DE EQUIPAMENTOS DE INFORMATICA DISPONIVEIS PARA O DESFAZIMENTO.	6
<input type="checkbox"/> ENCAMINHA RELACAO DE EQUIPAMENTOS DE INFORMATICA PARA FINS DE DESFAZIMENTO.	2
<input type="checkbox"/> ENCAMINHA PLANILHA DE EQUIPAMENTOS DE INFORMATICA PARA DESFAZIMENTO.	2
<input type="checkbox"/> INFORMA QUE ESTA SECAO JUDICIARIA DISPOE DE EQUIPAMENTOS DE INFORMATICA EM PROCESSO DE DESFAZIMENTO DE BENS MOVEIS OCIOSOS, ANTIECONOMICOS E IRRECUPERAVEIS.	1
<input type="checkbox"/> RETIFICA O OFICIO GAB/472/2007, DATADO DE 10/10/2007, RETIRA DA RELACAO DE MATERIAIS DE INFORMATICA PARA FINS DE DESFAZIMENTO OS EQUIPAMENTOS DA RELACAO QUE SEGUE.	1
<input type="checkbox"/> ENCAMINHA RELACAO DE EQUIPAMENTOS DE INFORMATICA DISPONIVEIS PARA DESFAZIMENTO.	1
<input type="checkbox"/> RELACAO DE EQUIPAMENTOS DE INFORMATICA E COMPONENTES PARA DESFAZIMENTO.	1

Figura 45 – Seleção de instâncias através de seus atributos e atalho para mineração.

5.3.6. Animação

O módulo de animação é responsável pela encenação da execução das instâncias de processo selecionadas na etapa de exploração a partir dos modelos gerados na etapa de mineração. Ao selecionar um destes modelos é possível especificar um intervalo de tempo para que seja realizada uma análise de como o processo foi executado durante este período. A partir desta animação podem-se observar quais são os comportamentos mais frequentes e verificar a existência de problemas de desempenho e gargalos em um processo.

A animação se dá adicionando círculos que percorrem os fluxos de sequência entre as atividades de um modelo. Não havendo informações sobre a duração de um evento, este intervalo é calculado a partir da diferença entre o tempo de ocorrência do evento e do seu próximo. Ao clicar em um destes círculos, o sistema exibe uma lista contendo todos os fluxos de sequência a serem percorridos por aquela instância durante sua execução.

A Figura 46 ilustra um instante da animação de um modelo de processo. É possível pausar ou mudar a aceleração da animação em qualquer momento, permitindo que o usuário observe os acontecimentos no ritmo desejado.

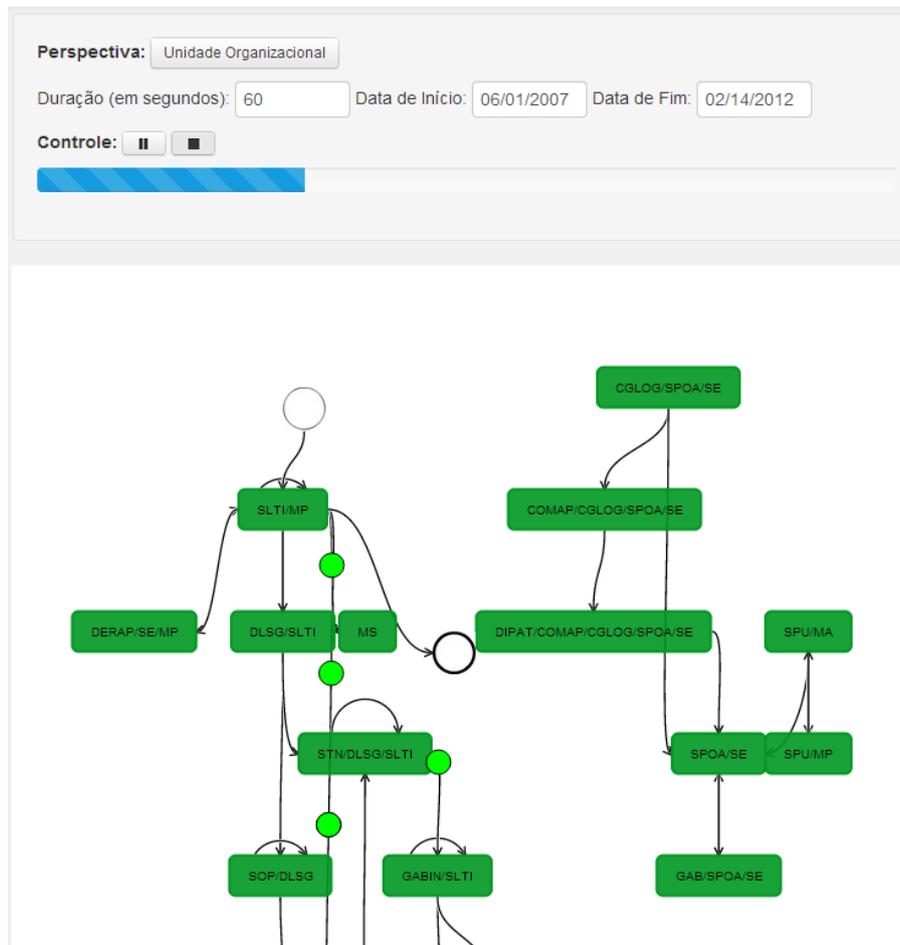


Figura 46 – Animação de um modelo de processo.

5.4. Características Diferenciais

Esta seção resume os principais diferenciais desta ferramenta em relação às abordagens descritas no Capítulo 3, escolhidas por possuírem capacidade de realizar atividades de descoberta de modelos de processos desestruturados. A Tabela 10 apresenta uma comparação entre a ferramenta desenvolvida neste trabalho e as ferramentas ProM e MANA, explicitando estes diferenciais. Uma explicação detalhada sobre cada um destes tópicos é exposta em seguida.

Tabela 10 – Características diferenciais da ferramenta desenvolvida.

Característica	Ferramenta Desenvolvida	MANA	ProM
Acesso a dados de iterações anteriores	Sim	Não	Não
Filtragem de instâncias de processos por encadeamento de eventos	Sim	Não	Não
Linguagem gráfica para auxílio na busca de instâncias de processos	Sim	Não	Não
Múltiplas perspectivas	Sim	Parcialmente	Sim
Dicionário de dados	Sim	Não	Sim
Fluxo de trabalho para analistas não especializados	Sim	Sim	Não
Suporte ao padrão XES	Parcialmente	Parcialmente	Sim

Acesso a Dados de Iterações Anteriores. A ferramenta desenvolvida neste trabalho é a única que provê mecanismos para guardar e manter acessíveis dados referentes a filtros, instâncias de processos recuperadas pela busca, parâmetros utilizados no algoritmo de descoberta de processos e os modelos gerados por este algoritmo. Esta característica permite que o usuário possa examinar o impacto de cada uma das mudanças realizadas durante uma iteração no fluxo de uso da ferramenta nos resultados obtidos, possibilitando a verificação da evolução incremental do conhecimento adquirido através das iterações.

A ferramenta MANA guarda somente os filtros e instâncias retornadas pela busca da iteração corrente, entretanto todos os modelos de processo gerados estão disponíveis para consulta. O ProM não trabalha com o conceito de busca exploratória, portanto não guarda estas informações, porém o usuário pode manter um histórico por conta própria sobre todos arquivos contendo os registros de processos selecionados para análise, como também seus modelos de processo resultantes.

Filtragem de Instâncias de Processo por Encadeamento de Eventos. As ferramentas ProM e MANA não permitem este tipo de filtro, embora o MANA permita selecionar instâncias a partir de atributos de um de seus eventos. A ferramenta proposta é a única das comparadas que suporta filtros por encadeamento de eventos, dando ao analista mais possibilidades para definir o comportamento procurado de um processo e refinar sua pesquisa.

Linguagem Gráfica para Auxílio na Busca de Instâncias de Processos. A ferramenta desenvolvida neste trabalho, por ser a única que permite buscar instâncias de processos através de seus encadeamentos de eventos, também é a única que implantou um mecanismo gráfico para o auxílio desta tarefa. Uma vantagem desta característica é permitir ao analista realizar a busca de instâncias de processos cujo comportamento pode ser definido pelo usuário em um diagrama que utiliza as mesmas metáforas de representação dos modelos de processos na ferramenta. Outra vantagem importante é que representações gráficas podem superar as limitações da natureza linear de representações em forma de texto para este tipo de filtragem, que permite a construção de estruturas de desvio de fluxo e paralelismo.

Múltiplas Perspectivas. A ferramenta proposta neste estudo, assim como o ProM, permite a construção de modelos de processos baseados em quaisquer combinação de seus atributos. Deste modo é possível visualizar como acontece um processo de negócio em diversas perspectivas, como o fluxo de atividades, rede social construída durante a execução de um processo, etc. A ferramenta MANA permite um número limitado de perspectivas, sendo necessário alterar a ferramenta para aceitar a inclusão de novos tipos de dados e definir como estes serão utilizados na construção de uma nova perspectiva.

Dicionário de Dados. Esta ferramenta propõe a criação de um dicionário de dados com o objetivo de nivelar o nível de granularidade de detalhes na construção de modelos e ajudar a realização da curadoria dos registros de processos, permitindo que grafias distintas referentes a uma mesma entidade sejam interpretadas como um termo único. Uma vez que os valores são mapeados, todas as funcionalidades disponíveis na ferramenta podem fazer o uso destas informações, incluindo o algoritmo de descoberta de processos. A ferramenta ProM possui um grupo de extensões que implantam funcionalidades similares através da criação de ontologias. Entretanto, a utilização das informações complementares presentes em registros de processos anotados de ontologias não é universal, sendo necessária a criação de extensões customizadas para a interpretação de cada ontologia específica. Durante a realização deste estudo foram encontradas somente onze extensões que dão suporte a ontologias dentro de um universo de mais de duas centenas disponíveis para o ProM (Buijs 2010).

Fluxo de trabalho para analistas não especializados. A ferramenta desenvolvida, assim como o MANA, implanta um fluxo de trabalho para que não especialistas em mineração de processos possam fazer uso da ferramenta, criando mecanismos para superar as dificuldades encontradas no uso da ferramenta ProM. Entre estas dificuldades está justamente a falta de uma proposta linear de trabalho no ProM, que funciona como um *framework* composto por diversas extensões capazes de realizar atividades de mineração de processos, onde não há uma indicação clara de quais passos devem ser realizados pelo usuário para que este obtenha sucesso em sua análise.

Entre os mecanismos utilizados para facilitar o fluxo de trabalho estão a busca exploratória de dados, a linguagem gráfica para facilitar a criação de filtros de encadeamento de eventos e a integração dos registros de processos com um dicionário de dados para facilitar o entendimento dos mesmos.

Suporte ao padrão XES. Somente o ProM dá suporte abrangente ao padrão XES de armazenamento de registros de processos, utilizando-o diretamente para realização das análises de processos na ferramenta. O ProM possui também uma ferramenta auxiliar que apoia a extração de dados de sistemas que registram eventos de processos e a formatação para o padrão XES.

Entretanto, para realizar a busca exploratória em grande volume de dados de registros de execuções de processos é necessário importá-los para sistemas de gerenciamento de base de dados, sistemas estes que armazenam dados utilizando menos espaço e facilitam a recuperação das informações, se comparados ao armazenamento de dados em arquivos XML utilizado pelo padrão XES.

O principal diferencial da ferramenta desenvolvida em relação a este tópico é a utilização de banco de dados com entidades que tentam simular o padrão XES, mesclando suas características de simplicidade, flexibilidade e extensibilidade à capacidade de lidar com grande volume de dados de SGBDs.

Por fim, tanto a ferramenta proposta neste trabalho e a ferramenta MANA dão suporte parcial ao padrão XES, fazendo utilizando uma biblioteca para a conversão interna dos dados para este formato de modo que sejam utilizados junto à biblioteca que calcula as métricas do Minerador Heurístico.

Capítulo 6 – Estudo de Caso

Esta seção apresenta dois estudos de caso que utilizam a ferramenta desenvolvida neste trabalho em atividades de descoberta de processos não estruturados com a finalidade de exemplificar seu uso. Os dados utilizados neste estudo de caso foram extraídos de uma aplicação real, o Sistema de Controle de Processos e Documentos (CPROD) do Ministério do Planejamento, Orçamento e Gestão (MPOG), órgão do governo brasileiro. O sistema analisado faz o acompanhamento dos processos de negócio sem impor estrutura rígida aos processos, que por consequência não possuem fluxos de trabalho bem definidos.

Entre as dificuldades encontradas na análise dos registros de processos estão a pobreza semântica dos dados encontrados, implicando na necessidade de recorrer a ambientes externos visando obter informações para identificar algum termo presente em algum registro de execução de processo, e a possibilidade de múltiplos valores presentes nos dados corresponderem a somente uma unidade organizacional, pois o sistema permite o registro de dados através de campos textuais livres sem padronização.

O primeiro estudo de caso investiga os processos registrados na Secretaria de Patrimônio da União que remetem a doação de bens. Os benefícios do emprego do dicionário de dados durante as atividades de descoberta deste processo são evidenciados, comparando os modelos baseados somente nos dados presentes nos registros de processos e os modelos gerados com o auxílio de termos do dicionário.

O segundo estudo de caso investiga a relação entre a Secretaria de Patrimônio da União e a Subsecretaria de Planejamento, Orçamento e Administração através dos processos que percorrem este caminho. Este estudo de caso evidencia o uso dos

mecanismos gráficos criados na ferramenta para suportar atividades de descoberta de processos.

As duas primeiras seções deste capítulo explicitam como foram realizadas as etapas de importação e mapeamento do fluxo de atividades proposto da ferramenta, enquanto as seções seguintes mostram como ocorreram as atividades de exploração, mineração e análise para seus respectivos casos de uso.

6.1. Importação dos Dados

O primeiro passo para ambos os estudos de caso consiste em extrair os dados do Sistema de Controle de Processos e Documentos e importar estes dados para a ferramenta desenvolvida neste trabalho. O CPROD é responsável por efetuar o recebimento, a distribuição, o registro, a tramitação, a formalização, a expedição e o arquivamento de processos do Ministério do Planejamento, Orçamento e Gestão (Brasil 2002).

A base de dados do CPROD apresenta 119 tabelas, porém somente uma pequena parte possui informações referentes aos processos que tramitam pelo MPOG. Os conceitos mais importantes para a mineração de processos, o de instância e evento, estão armazenados nas tabelas *Protocolo* e *Trâmite*, respectivamente. O conjunto de dados obtido para a realização dos experimentos dispunha da quantidade de 3.440.676 instâncias de processos e 17.443.323 eventos compondo estas instâncias.

Um protocolo possui mais de vinte atributos, entre eles alguns que poderiam ser calculados através da informação presente em seus trâmites, como a data de abertura, data do último trâmite, município, unidade federativa e país; e outros que não possuem relevância para a descoberta de processos, como a quantidade de cópias, quantidade de páginas, código de barras, entre outros. A partir de associações com a tabela Protocolo,

foi possível obter uma informação que indica a classe para os possíveis documentos e processos do sistema, porém esta informação está ausente na maioria dos protocolos. Outra associação indica os interessados nos processos, geralmente uma ou mais pessoas físicas, com pouca relevância para a análise realizada neste estudo de caso. Decidiu-se por extrair as informações sobre a classe e o atributo de assunto presente na tabela de protocolo, que contém uma breve descrição em texto livre sobre o que o processo se trata, para utilização como atributos de instância de processo no sistema desenvolvido neste trabalho.

Os trâmites destes protocolos indicam em seus dados a sua situação, unidade organizacional de origem, unidade organizacional de destino, data de criação, data de recebimento na unidade organizacional de destino, criador e responsável pelo recebimento. Nota-se que o CPROD não possui informações sobre as atividades realizadas quando alguém cadastra um trâmite no sistema, sendo registrados somente os atores e unidades organizacionais envolvidos no trâmite. A partir do manual do sistema (Brasil 2002) é possível inferir que as atividades são realizadas até ato de cadastro do trâmite, e que seus dados de recebimento são úteis somente para informar que a parte que o receberá tomou ciência do mesmo. Deste modo, optou-se por utilizar a data de criação do trâmite como a indicação de temporalidade do evento realizado. Foram adicionadas como atributos de evento as informações de nome e sigla da unidade organizacional de origem, como também o usuário criador do trâmite, indicando os responsáveis pelo mesmo. Entretanto, nem todos os valores de nome das unidades organizacionais estavam disponíveis, e em alguns casos este valor era igual à sigla.

A Tabela 11 indica os dados que foram exportados do CPROD e suas respectivas correspondências nas entidades do banco de dados desta ferramenta.

Tabela 11 – Correspondência entre dados do CPROD e entidades da ferramenta.

CPROD	Entidades dos Registros de Processos da Ferramenta
Protocolo	Instância de Processo
Classe	Atributo de Processo – Tipo “ <i>descrição</i> ”
Assunto	Atributo de Processo – Tipo “ <i>tipo</i> ”
Trâmite	Evento de Processo
Unidade	Atributo de Evento – Tipo “ <i>unidade organizacional</i> ” Atributo de Evento – Tipo “ <i>unidade organizacional (extenso)</i> ”
Responsável pelo Trâmite	Atributo de Evento – Tipo “ <i>pessoa</i> ”

6.2. Preenchimento do Dicionário

A etapa de mapeamento, responsável pelo preenchimento do dicionário de dados com informações para auxiliar o entendimento dos dados presentes nos registros de processos importados na ferramenta, foi dividida em dois passos. O primeiro passo consistia em cadastrar termos correspondentes às entidades presentes nos registros de processos e foi realizado com o auxílio de mecanismos para automatizar a importação dos dados a partir de um sistema externo. O segundo passo, realizado manualmente, fazia o relacionamento dos termos cadastrados aos possíveis valores presentes nos atributos de eventos. Havia somente dois tipos destes atributos: unidades organizacionais e pessoas.

As entidades correspondentes às unidades organizacionais foram carregadas a partir do SIORG - Sistema de Organização e Inovação Institucional do Governo (Brasil 2009), mantido pelo MPOG. Durante a realização deste procedimento estavam disponíveis para consulta dados sobre mais de 53.000 órgãos em pelo menos 1.400 municípios brasileiros, todas as informações eram oficiais e retiradas do Diário Oficial (Brasil 2009). As informações eram disponibilizadas em uma página da *web* onde era necessária a interação com o usuário para carregar as unidades organizacionais subordinadas de cada órgão. Esta característica aliada ao grande volume de dados

motivou a construção de um *crawler* para criação automática de termos no dicionário a partir destes órgãos, onde cada termo possuía também a informação sobre onde estava subordinado, gerando um relacionamento com outro termo. A Figura 47 ilustra a página web de onde foram carregadas as informações.



Figura 47 – Tela do SIORG contendo parte da hierarquia de órgãos governamentais.

Não havia informação sobre o nível de hierarquia de cada unidade organizacional, estando disponível para consulta somente relações de subordinação

entre órgãos. Como era necessário atribuir um tipo para cada termo cadastrado, foram escolhidos arbitrariamente seis tipos para atribuição aos termos de acordo com sua distância à Presidência da República, órgão raiz do organograma apresentado pelo SIORG. Estes tipos são, ordenados a partir da raiz: *Órgão*, *Sub-órgão*, *Departamento*, *Setor*, *Divisão* e *Sub-divisão*.

Optou-se por não cadastrar no dicionário termos referentes ao outro tipo de atributo de evento, as pessoas, pois não foi encontrada uma fonte de dados confiável e com quantidade representativa de informações. Era possível inferir o relacionamento entre pessoas e as unidades organizacionais através do conjunto de atributos de um determinado evento, porém, devido a acontecimentos como realocações em outros órgãos ou promoções de cargo, este relacionamento só poderia ser considerado válido para este evento.

Por fim, o segundo passo, que consistiu em efetuar as associações entre os possíveis valores de unidade organizacional aos termos cadastrados no dicionário, foi realizado através do módulo de dicionário de dados da ferramenta. Como os dados são provenientes de um sistema de protocolos do MPOG, com a maioria de seus eventos possuindo como atributo uma unidade organizacional deste órgão, e o estudo de caso se referia a processos que acontecem dentro do MPOG, então este cadastro de associações foi limitado somente a unidades organizacionais do Ministério do Planejamento, Orçamento e Gestão. Os demais valores foram cadastrados durante o estudo de caso conforme necessidade. Para cada valor de atributo associado a um termo, também eram adicionadas relações com os termos cujo o qual era subordinado. A Figura 48 ilustra o resultado do preenchimento do dicionário para o termo “*Secretaria de Logística e Tecnologia da Informação*”.

Secretaria de Logística e Tecnologia da Informação

 [Ministério do Planejamento, Orçamento e Gestão](#) (Órgão) Editar 

-  Secretaria de Logística e Tecnologia da Informação (Sub-Órgão)
 -  [Comissão de Coordenação do SISP](#) (Departamento)
 -  [Comissão Interministerial de Sustentabilidade na Administração Pública](#) (Departamento)
 -  [Coordenação 1](#) (Departamento)
 -  [Coordenação 2](#) (Departamento)
 -  [Departamento de Governo Eletrônico](#) (Departamento)
 -  [Departamento de Logística e Serviços Gerais](#) (Departamento)
 -  [Departamento de Serviços de Rede](#) (Departamento)
 -  [Departamento de Sistemas de Informação](#) (Departamento)
 -  [Departamento de Suporte à Gestão do Sistema de Transferências Voluntárias da União](#) (Departamento)
 -  [Divisão](#) (Departamento)
 -  [Gabinete](#) (Departamento)

Atributos Associados		
SLTI/MP Unidade Organizacional	SLTI/MP Unidade Organizacional (Extenso)	CATA/SLTI Unidade Organizacional
CATA/SLTI Unidade Organizacional (Extenso)	GABIN/SLTI Unidade Organizacional	GABIN/SLTI Unidade Organizacional (Extenso)
STN/DLSG/SLTI Unidade Organizacional	TERC/DLSG Unidade Organizacional	ASS/DLSG Unidade Organizacional (Extenso)
DLSG/SLTI Unidade Organizacional (Extenso)	SOP/DLSG Unidade Organizacional	EQ/DLSG Unidade Organizacional (Extenso)

Figura 48 – Informações sobre o termo *Secretaria de Logística e Tecnologia da Informação*.

Nota-se que as onze unidades organizacionais que aparecem na Figura 48 como subordinadas à *Secretaria de Logística e Tecnologia da Informação* estão representadas por termos do tipo *Departamento*. A lista de valores de atributos associados foi suprimida neste exemplo devido a sua extensão. A correspondência destes valores foi aferida a partir de suas siglas ou abreviações, com *GABIN/SLTI* estando associado diretamente ao termo *Gabinete* e conseqüentemente aos termos *Secretaria de Logística e Tecnologia da Informação* e *Ministério do Planejamento, Orçamento e Gestão*, por exemplo.

6.3. Primeiro Estudo de Caso

O objetivo deste estudo de caso é explicitar o uso do dicionário de dados para realizar a curadoria dos dados presentes nos registros de processos e demonstrar um exemplo em que este mecanismo pode ser utilizado para apoiar a descoberta de um processo não estruturado. As instâncias referentes à doação de bens foram escolhidas para análise neste estudo de caso por terem uma quantidade razoável de ocorrências e por não haver um padrão no fluxo de atividades entre as instâncias, resultando em um processo não estruturado.

Os procedimentos realizados para este estudo de caso são iniciados ao ser cadastrado um novo processo para análise na ferramenta. Neste cadastro, o processo é batizado de *Doação de Bens* e é informado que será utilizado o conjunto de registros de processos com origem no sistema CPROD. Após o cadastro do processo, a tela de filtros por atributos de instâncias é acessada, onde são estabelecidos filtros para atributos do tipo *descrição* contendo as palavras *doação* e *bens*. Estes filtros são ilustrados pela Figura 49.

Atributo	Comparação	Valor	Remover
Selecionar Atr ▾	igual a ▾	valor do atributo	Adicionar
Descrição	contém	doação	Remover
Descrição	contém	bens	Remover

Figura 49 – Filtros de atributos de instâncias buscando as palavras *descrição* e *bens*.

A busca utilizando estes filtros retorna 1870 instâncias. A partir de uma rápida análise dos dados presentes nestes registros, verifica-se que eles estão divididos em dois principais grupos em relação ao local onde são executados: o primeiro é composto por

unidades organizacionais da Secretaria de Patrimônio da União, enquanto o segundo é composto por departamentos da Secretaria de Logística e Tecnologia da Informação. Estes dois grupos de unidades organizacionais pertencem ao Ministério do Planejamento, Orçamento e Gestão.

O Minerador Heurístico foi utilizado para construir um modelo de processo que faz a descrição das instâncias retornadas pela busca. Foi escolhido o valor de 19 ocorrências mínimas para que cada elemento esteja representado no modelo, o que é aproximadamente 1% da quantidade de instâncias sendo analisadas, e o valor de 0,8 para o parâmetro de dependência entre os atributos de evento. O modelo foi gerado utilizando somente as informações sobre unidade organizacional presentes nos registros. Uma visão geral sobre a topologia do modelo minerado é mostrada na Figura 50.

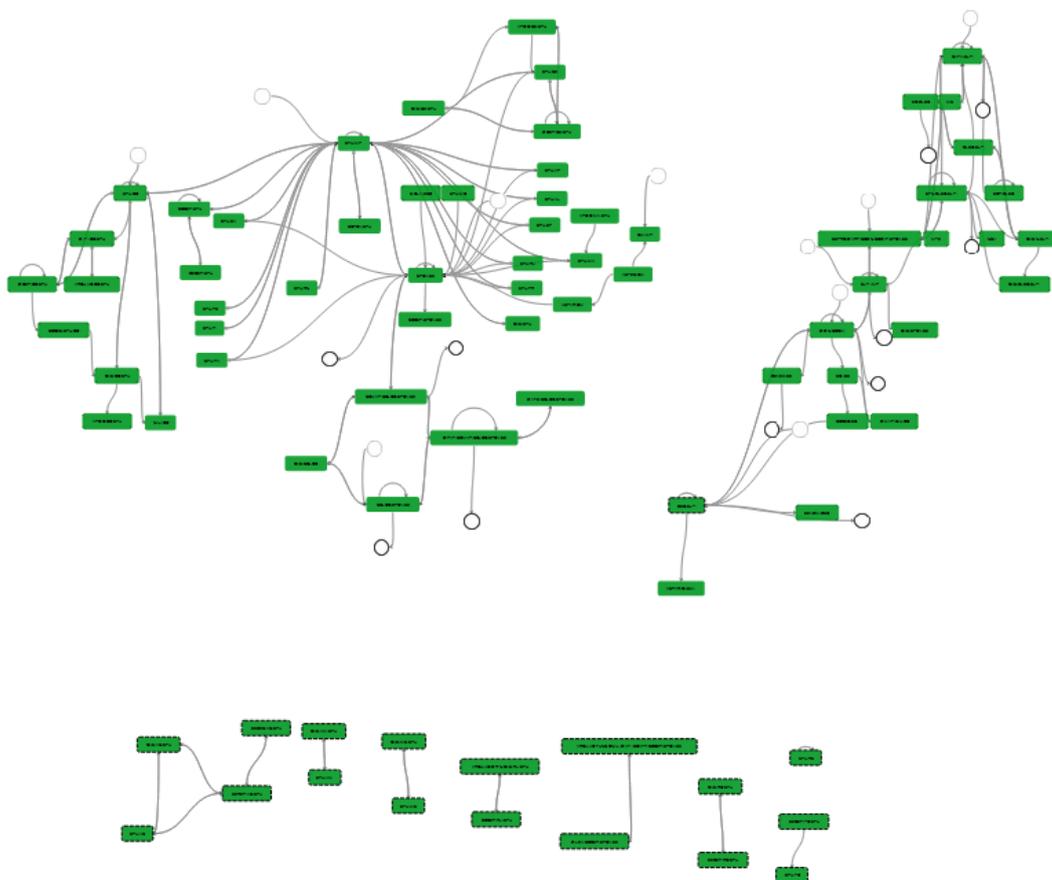


Figura 50 – Modelo de processo descoberto pelo Minerador Heurístico.

Observa-se que o modelo é composto por vários grupos não conectados entre si, e existem dois principais grupos, corroborando com a análise realizada na tela que exibe os detalhes das instâncias de processos utilizadas na construção deste modelo.

Este modelo é composto por 79 elementos entrelaçados por 181 conexões, esta grande quantidade de elementos torna difícil seu entendimento. Outro problema encontrado é a variação do nível de abstração dos elementos presentes neste modelo, alguns fazem menção direta a um Ministério, enquanto outros representam unidades organizacionais cinco níveis abaixo da raiz apresentada pelo SIORG.

Optou-se por restringir a investigação deste processo somente para o conjunto de registros relacionados à Secretaria do Patrimônio da União. Para isto, foi criado um filtro de encadeamento de eventos utilizando a linguagem gráfica proposta neste trabalho. Em geral, as atividades deste grupo de instâncias começavam em departamentos da Secretaria do Patrimônio da União e terminavam na Subsecretaria de Planejamento, Orçamento e Administração. Foi desenhado um encadeamento de eventos para refletir este comportamento, representado pela Figura 51.

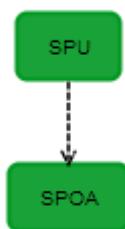


Figura 51 – Encadeamento de eventos utilizado como filtro.

Para uniformizar o nível de abstração na representação destes dados em um modelo e unificar múltiplos valores pertencentes a uma mesma entidade, o dicionário de dados foi utilizado. Foram escolhidas para isso as abstrações de *sub-órgão* e *departamento* presentes no dicionário, representando unidades organizacionais até dois níveis de subordinação a partir do Ministério do Planejamento, Orçamento e Gestão.

Este procedimento também permite que valores menos frequentes que porventura ficaram de fora do modelo e que representam um mesmo sub-órgão ou departamento sejam agrupados e alcancem valores superiores ao limiar de ocorrência mínima.

Nem todos os valores de atributos já possuíam termos do dicionário associados durante a etapa de mapeamento. Esta questão foi solucionada acessando a funcionalidade da ferramenta que lista somente os valores presentes nos atributos de instâncias selecionadas para análise, onde era possível verificar as associações e atribuir termos a estes valores. A Figura 52 ilustra esta funcionalidade. Neste exemplo figuram somente 5 dos 267 valores de unidade organizacional presentes nas instâncias selecionadas. Algumas entidades possuíam múltiplas grafias, sobretudo apresentando textos terminados com “MP”, “MPOG” ou sem sufixo indicando que pertencem ao Ministério do Planejamento, Orçamento e Gestão.

Revisar Dicionário

🔍 Busque ou crie um termo para associá-lo aos atributos dos eventos.

Atributo	Órgão	Sub-Órgão	Departamento
SPU/GO	Ministério do Planejamento, Orçamento e Gestão	Secretaria do Patrimônio da União	Superintendência do Patrimônio da União em Goiás
DIIFI/GO/SPU	Ministério do Planejamento, Orçamento e Gestão	Secretaria do Patrimônio da União	Superintendência do Patrimônio da União em Goiás
CGADL/SPU/MP	Ministério do Planejamento, Orçamento e Gestão	Secretaria do Patrimônio da União	Coordenação-Geral da Amazônia Legal
GAB/PA/SPU	Ministério do Planejamento, Orçamento e Gestão	Secretaria do Patrimônio da União	Superintendência do Patrimônio da União no Pará
GAB/MA/SPU	Ministério do Planejamento, Orçamento e Gestão	Secretaria do Patrimônio da União	Superintendência do Patrimônio da União no Maranhão

Figura 52 – Tela do sistema que permite revisar o dicionário.

Uma vez que as associações entre termos do dicionário e os atributos presentes nos registros de execução foram preenchidas, tornou-se possível utilizar estes dados

junto ao Minerador Heurístico para que um modelo com nível de abstração uniforme fosse construído. Foram utilizados os mesmos parâmetros da iteração anterior na etapa de mineração. O modelo descoberto, representado pela Figura 53, faz o agrupamento dos 79 elementos presentes no primeiro modelo em somente 8. Departamentos da Superintendência do Patrimônio da União em diversos estados foram representados como um só e alguns valores foram abreviados para permitir a impressão do modelo nesta dissertação.

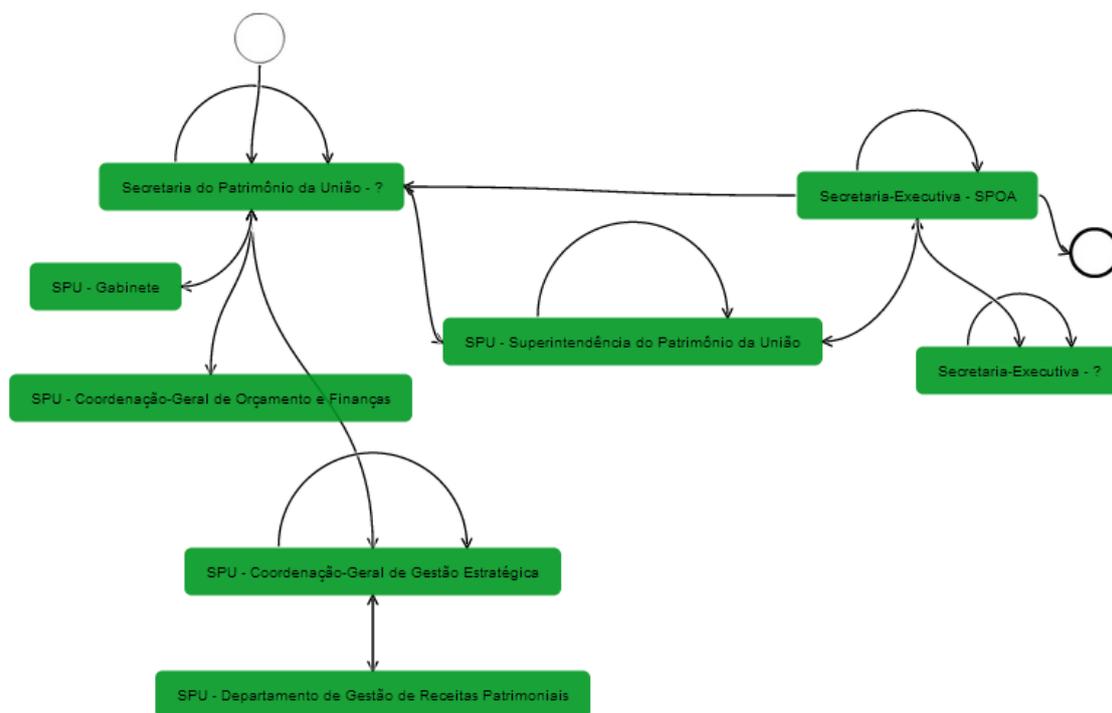


Figura 53 – Modelo de processo construído utilizando dados do dicionário.

Nota-se que a quantidade reduzida de elementos no modelo utilizando dados do dicionário tornou mais fácil seu entendimento se comparado ao modelo inicial gerado pelo Minerador Heurístico.

Os dados analisados neste estudo de caso não possuem informações sobre as tarefas executadas em cada unidade organizacional, por isto não foi possível gerar um modelo de processo baseado na perspectiva de fluxo de atividades. Entretanto, obteve-

se um modelo que descreve o caminho que o processo percorre dentro do Ministério do Planejamento, Orçamento e Gestão. A partir deste modelo existe a possibilidade de realização de diagnósticos preliminares sobre o processo, que podem servir de base para outras tarefas relacionadas ao gerenciamento de processos de negócio da organização e motivar o surgimento de ideias sobre atividades de reengenharia e implantação de melhorias.

A análise sobre o processo presente neste estudo de caso seria um desafio se realizada nas ferramentas ProM e MANA. Apesar de simples, os passos executados pela busca exploratória não são totalmente suportados em ambas estas ferramentas. Mesmo selecionando um conjunto inicial de registros com base em um critério trivial, seria necessário que o responsável pela análise possuísse conhecimento sobre como os dados estão armazenados no CPROD e como importa-los à ferramenta ProM. A segunda restrição na busca seria ainda mais desafiadora, pois envolve a análise do encadeamento de eventos das instâncias de processo para determinar quais instâncias devem prosseguir na análise, o que pode não ser praticável em um conjunto grande de dados sem o conhecimento de técnicas de programação.

Embora não suporte a busca por encadeamento de eventos de processos, a ferramenta MANA poderia se basear em uma das unidades organizacionais presentes no filtro para realizar a seleção de instâncias de processo. Não há garantias que o conjunto retornado pela busca seria exatamente igual, porém os registros que seguem o caminho pela Secretaria de Logística e Tecnologia da Informação poderiam ser deixados de fora da análise como neste estudo de caso.

Entretanto, a variabilidade nas nomenclaturas utilizadas para representar as unidades organizacionais e a natureza não estruturada do processo analisado dificultariam a obtenção de bons resultados utilizando a ferramenta MANA, pois não há

tratamento dos dados para evitar que o modelo gerado possua uma grande quantidade de elementos, como o modelo ilustrado pela Figura 50. Este problema poderia ser solucionado caso a ferramenta ProM fosse utilizada, porém as extensões que criam as ontologias para representar as entidades presentes nos dados são de difícil entendimento e possuem pouca documentação disponível.

6.4. Segundo Estudo de Caso

O objetivo deste estudo de caso é explicitar o uso dos mecanismos gráficos criados na ferramenta para apoiar atividades de descoberta de processos, não se limitando à linguagem gráfica de busca de registros de processos. O objeto de análise é a relação entre as duas unidades organizacionais apresentadas no estudo de caso anterior, a Secretaria do Patrimônio da União e a Subsecretaria de Planejamento, Orçamento e Administração.

As atividades são iniciadas repetindo o procedimento de criação de um novo processo visto no primeiro estudo de caso. O segundo passo, em vez de filtrar instâncias diretamente através de seus atributos limitando a busca somente a um tipo de processo, é criar um encadeamento de eventos indicando que o processo deve ser iniciado na SPU e em algum momento posterior deve passar pela SPOA. A Figura 54 ilustra o encadeamento de eventos de processo criado para agir como filtro.



Figura 54 – Encadeamento de eventos de processo utilizado como filtro.

Este filtro foi mais específico que o praticado no primeiro estudo visando minimizar a quantidade de unidades organizacionais distintas que aparecem no modelo descoberto. A busca retornou 707 instâncias de processos de diversos tipos.

A animação permitiu identificar um comportamento frequente neste conjunto de instâncias. As quatro unidade organizacionais presentes neste caminho, desde *SPU/MP* até *SPOA/SE*, foram selecionadas para que as informações sobre as instâncias de processo que fazem este percurso fossem retornadas. Foi retornada uma lista contendo os atributos de descrição e tipo mais frequentes. Ao selecionar o atributo de tipo de instância, só havia um valor, remetendo a auxílio moradia. A Figura 56 mostra o formulário contendo os atributos de instância e as opções de realizar uma nova etapa de mineração. Foram selecionadas 75 instâncias a partir deste formulário cujo atributo tipo era igual a *AUXILIO MORADIA*, e o Minerador Heurístico foi acionado utilizando os mesmos parâmetros de mineração da iteração anterior.



Figura 56 – Formulário que permite o filtro de instâncias com base em seus atributos.

O resultado deste procedimento contendo o novo modelo construído é ilustrado pela Figura 57. Nota-se que o modelo representado na Figura 55 descreve uma coleção de instâncias de processos referentes a diversos processos, enquanto o modelo representado na Figura 57 corresponde somente ao processo cujo tipo é Auxílio Moradia.

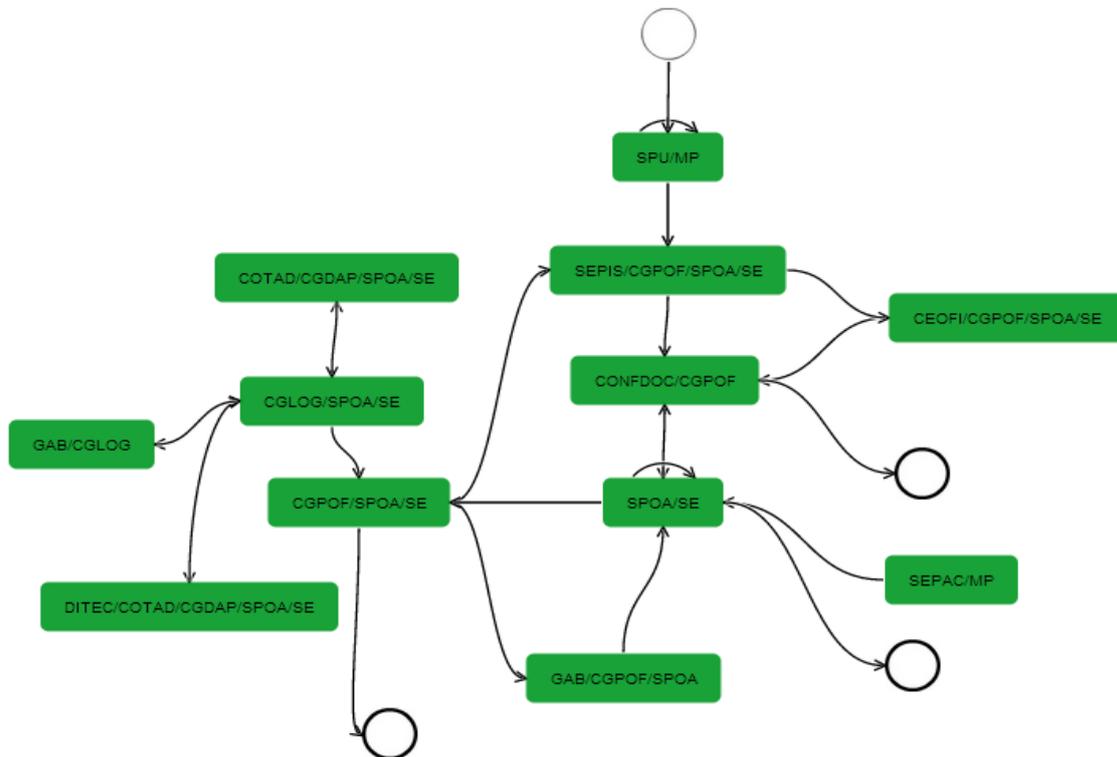


Figura 57 – Modelo de processo gerado a partir de um novo conjunto de instâncias.

Entretanto, a partir da verificação dos atributos do tipo *descrição* destas instâncias de processo foi identificado que existem dois grupos majoritários: pedido de reembolso de aluguel e pedido de imóvel funcional. As instâncias do grupo relacionado a reembolso de aluguel foram selecionadas por apresentarem maior quantidade. O processo de mineração foi realizado novamente, gerando o modelo representado pela Figura 58.

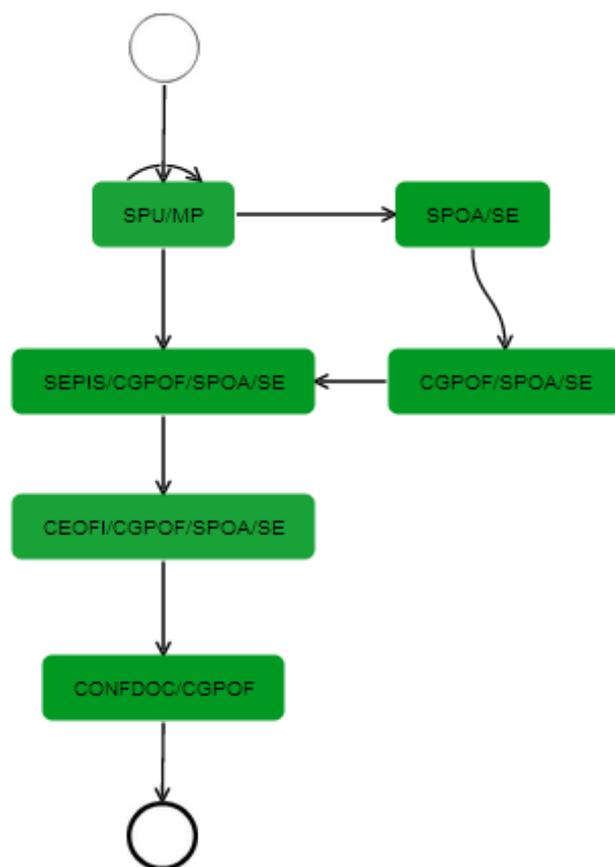


Figura 58 – Processo descoberto referente a reembolso de aluguel.

Este estudo de caso apresentou uma nova abordagem para atividades de descoberta de processos que permite investigar dados referentes a execuções de processos de uma forma exploratória a partir de mecanismos gráficos. Esta abordagem permite selecionar instâncias de processo para uma nova análise diretamente a partir de um modelo descoberto, sem a necessidade de voltar à fase de criação de filtros. Este mecanismo possibilita que o usuário elimine do modelo comportamentos julgados como não pertencentes a um processo, permitindo a construção de modelos de processos não estruturados mais simples a partir da poda de um modelo com maior quantidade de elementos em sua representação.

Capítulo 7 – Conclusões

Este capítulo encerra este trabalho apresentando suas conclusões. Os principais resultados e contribuições obtidos por este estudo são expostos, em seguida são apresentados trabalhos futuros que possuem o objetivo de alcançar melhorias em atividades de descoberta de processos não estruturados.

7.1. Resultados e Contribuições

Como resultado deste trabalho foi desenvolvida uma ferramenta capaz de realizar atividades de mineração de processos, com foco em ambientes que produzem processos desestruturados e possuem registros de eventos pobres em informação, de acordo com os objetivos propostos no Capítulo 1. Entre as principais características desta ferramenta está o armazenamento de registros de execuções de processos através de uma base de dados flexível e expansível, capaz de representar diferentes tipos de atributos relacionados a estes registros. Esta característica foi proposta na seção de trabalhos futuros da dissertação de mestrado que deu origem à ferramenta MANA (Esposito 2012), e embora a evolução da ferramenta MANA foi planejada como caminho inicial para atingir os objetivos deste trabalho, a alteração da maneira em que os registros de eventos de processo são armazenados tornou esta tarefa muito custosa, motivando a criação de uma nova ferramenta.

Uma das contribuições deste trabalho é a proposta de uma linguagem gráfica de busca que permite a filtragem de instâncias de processos a partir de seus encadeamentos de eventos. Esta linguagem permite refinar mais especificamente critérios de busca definindo um comportamento desejado para o processo, e propicia que o analisa realize buscas utilizando a mesma metáfora na qual os modelos de processo são representados.

Outra contribuição é o estudo de possibilidades para enriquecer modelos de processos descobertos a partir de registros de eventos pobres em informação. Foi construído um dicionário de dados para permitir que informações coletadas a partir de outras fontes pudessem ser utilizadas pelos algoritmos de descoberta de processos em conjunto com os registros de eventos. Este dicionário de dados também é capaz de oferecer mais informação acerca dos modelos construídos ao permitir que uma pessoa faça uma consulta sobre alguma atividade, departamento, cargo, pessoa ou qualquer outro elemento presente no modelo.

Por fim, foram apresentados dois estudos de caso que mostram o uso da ferramenta a partir de casos reais de processos não estruturados. Os dados para os experimentos foram extraídos de um sistema que faz o acompanhamento de processos do Ministério do Planejamento, Orçamento e Gestão, um órgão do governo federal do Brasil. O objetivo destes estudos de caso foi evidenciar os benefícios das contribuições supracitadas em atividades de mineração de processos não estruturados.

7.2. Limitações e Trabalhos Futuros

A ferramenta desenvolvida neste trabalho pode ser melhorada em diversos aspectos. Uma das possibilidades consiste em alterar a arquitetura da ferramenta para permitir o desenvolvimento de relatórios e indicadores de desempenho de processos de negócio através de extensões. Através destes relatórios é possível obter informações que ajudam o entendimento de como estes processos de fato ocorrem nas organizações. Entretanto, não é possível definir *a priori* as informações necessárias que causariam maior impacto positivo neste entendimento, tornando-se necessário facilitar a expansão de possibilidades de estudos realizados nesta ferramenta.

Uma das frentes de pesquisa não explorada devido à falta de dados confiáveis para testes foi a modificação de algoritmos de descoberta de processos para adicionar a capacidade de inferir regras de associação no dicionário de dados durante a descoberta dos modelos. Alguns dados quando inseridos em um contexto podem indicar a presença de outros, que porventura ocorreram porém não estão presentes nos registros de processos analisados. O objetivo é permitir que o modelo descoberto seja enriquecido com estas informações. Esta inferência pode ser realizada de formas distintas, como a partir de dados estatísticos ou regras de associação.

A criação de funcionalidades que ajudem a importação de dados provenientes de outras ferramentas é essencial. Embora não elimine a necessidade de uma pessoa especialista em tecnologia da informação para auxiliar a realização desta tarefa, o esforço dispendido na importação pode ser amenizado ao tornar possível o mapeamento direto entre as bases de dados do sistema que registra os eventos de processos e a ferramenta desenvolvida neste trabalho através de visões de banco de dados. Este procedimento também permitiria a realização das atividades de mineração de processos em tempo real logo após os eventos de processos serem registrados, sem a necessidade de uma nova carga de dados para atualização. Outra possibilidade relacionada à importação de dados é a criação de uma ferramentas análogas ao XESame ou ProM Import, apresentados no Capítulo 3.

Apesar da ferramenta possuir versionamento para cada iteração realizada na tentativa de se construir um modelo de processos que mais se adequa à realidade, não existem métricas definidas para comparar a evolução entre uma iteração ou outra, ficando a cargo do usuário fazer esta avaliação através de seus conhecimentos tácitos. É necessário estudar quais métricas se adequam melhor na medição da evolução incremental dos modelos de processos descobertos.

Por fim, também é interessante criar mecanismos que permitam a exportação dos dados gerados na ferramenta para diversos formatos. Dentre os exemplos de exportação está a disponibilização nos formatos dos padrões padrão XES e MXML dos registros de eventos selecionados durante a etapa de busca exploratória, permitindo que o trabalho de um usuário possa ser continuado em outros sistemas que possuem soluções não contempladas na ferramenta desenvolvida.

Referências

- Brasil, *Manual de Gerenciamento de Documentos e Utilização do CPROD.NET do Ministério do Planejamento, Orçamento e Gestão*, Brasília, 2002. Disponível em: http://www.planejamento.gov.br/secretarias/upload/Arquivos/publicacao/cprod/cprod_manual_do_usuario.pdf. Acesso em: 10 de Setembro de 2013.
- Brasil, Decreto Número 6944, Diário Oficial da União, Brasília, Agosto de 2009.
- Buijs, J., *Mapping Data Sources to XES in a Generic Way*. Dissertação de Mestrado, Technische Universiteit Eindhoven, Eindhoven, 2010.
- Buijs, J., van Dongen, B. F., van der Aalst, W. M., “On the Role of Fitness, Precision, Generalization and Simplicity in Process Discovery”. In: *On the Move to Meaningful Internet Systems: OTM 2012*, Springer Berlin Heidelberg, pp. 305-322, 2012.
- Davenport, H., Short, J., “The New Industrial Engineering: Information Technology and Business Process Redesign”. *Sloan Management Review*, v. 31, n. 4, pp. 11-27, 1990.
- Desel, J., Reisig, W., “Place/transition Petri Nets”. In: Lectures on Petri Nets I: Basic Models, v. 1491, *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 122-173, 1998.
- Esposito, P. M., *MANA: Identificação, Mineração, Análise e Reengenharia de Processos de Negócio*. Dissertação de Mestrado, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2012.
- Esposito, P. M., *et al.*, “MANA: Identifying and Mining Unstructured Business Processes”. In: *Business Process Management Workshops*, Springer Berlin Heidelberg, pp. 199-204, 2013.
- Goldberg, D., Holland, J., “Genetic Algorithms and Machine Learning”. *Machine Learning*, v. 3, n. 2, pp. 95-99, 1988.
- Goltz, U., Reisig, W., “The Non-Sequential Behaviour of Petri Nets”. *Information and Control*, v. 57, n. 2, pp. 125-147, 1983.

- Grefenstette, J., "Optimization of Control Parameters for Genetic Algorithms". *IEEE Transactions on Systems, Man and Cybernetics*, v. 16, n. 1, pp. 122-128, 1986.
- Grigori, D., Casati, F., Castellanos, M., *et al.*, "Business Process Intelligence". *Computers in Industry*, v. 53, n. 3, pp. 321-343, 2004.
- Günther, C., van der Aalst, W. P., "Fuzzy Mining – Adaptive Process Simplification Based on Multi-Perspective Metrics". In: *Business Process Management*, Springer Berlin Heidelberg, pp. 328–343, 2007.
- Günther, C., van der Aalst, W. P., "A Generic Import Framework for Process Event Logs". In: *Business Process Management Workshops*, Springer Berlin Heidelberg, pp. 81-92, 2006.
- Günther, C., XES Standard Definition. Fluxicon Process Laboratories, 2009.
- Hepp, M., *et al.*, "Semantic Business Process Management: A Vision Towards Using Semantic Web Services for Business Process Management". *IEEE International Conference on e-Business Engineering*, pp. 535-540, 2005.
- Jacobson, I., Ericsson, M., Jacobson, A., *The Object Advantage: Business Process Reengineering With Object Technology*. 1^a ed., Estados Unidos da América, Addison-Wesley Professional, 1994.
- Larkin, J., Simon, H., "Why a Diagram is (Sometimes) Worth Ten Thousand Words", *Cognitive Science*, v. 11, pp. 65-99, 1987.
- de Medeiros, A. K. A., van der Aalst, W. M., Weijters, A., "Workflow Mining: Current Status and Future Directions". In: *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*, v. 2888, Springer Berlin Heidelberg, pp. 389–406, 2003.
- de Medeiros, A. K. A., Weijters, A., van der Aalst, W. P., "Genetic Process Mining: A Basic Approach and Its Challenges". In: *Business Process Management Workshops*, Springer Berlin Heidelberg, pp. 203-215, 2006.

- de Medeiros, A. K. A., Weijters, A., van der Aalst, W. P., “Genetic Process Mining: An Experimental Evaluation”. *Data Mining and Knowledge Discovery*, v. 14, n. 2, pp. 245-304, 2007.
- de Medeiros, A. K. A., Guzzo, A., Greco, G., van der Aalst, W. M., Weijters, A., van Dongen, B. F., Saccà, D., “Process Mining Based on Clustering: A Quest for Precision”. In: *Business Process Management Workshops*, Springer Berlin Heidelberg, pp. 17-29, 2008.
- de Medeiros, A. K. A., van der Aalst, W. P., Weijters, A., *Semantics Utilized for Process Management Within and Between Enterprises*, 2008b.
- Melão, N., Pidd, M., “A Conceptual Framework for Understanding Business Processes and Business Processes Modelling”. *Informational Systems Journal*, v. 10, n. 2, pp. 105-129, 2000.
- Murata, T., “Petri Nets: Properties, Analysis and Applications”. *Proceedings of the IEEE*, v. 77, n. 4, pp. 541-580, 1989.
- OMG, *Business Process Model and Notation (BPMN) Version 2.0*, Janeiro de 2011.
- Ouyang, C., “From BPMN Process Models to BPEL Web Services”. *International Conference on Web Services*, pp. 285-292, 2006.
- Recker, J., “Opportunities and Constraints: The Current Struggle with BPMN”. *Business Process Management Journal*, v. 16, n. 1, pp. 181-201, 2010.
- Reijers, H., *Design and Control of Workflow Processes*. 1^a ed., Alemanha, Springer-Verlag, 2003.
- Rozinat, A., van der Aalst, W. P., “Decision Mining in ProM”. In: *Business Process Management*, Springer Berlin Heidelberg, pp. 420-425, 2006.
- Rodríguez, A., Medina, E., Piattini, M., “A BPMN Extension for the Modeling of Security Requirements in Business Processes”. *IEICE Transactions on Information and Systems*, v. E90-D, n. 4, pp. 745-752, 2007.

- van der Aalst, W. M., *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. 1^a ed., Alemanha, Springer-Verlag, 2011.
- van der Aalst, W. M., Hofstede, A., Weske, M., “Business Process Management: A Survey”. In: *Business Process Management*, v. 2678, Springer Berlin Heidelberg, pp. 1-12, 2003.
- van der Aalst, W. M., “Process Mining: Overview and Opportunities”, *ACM Transactions on Management Information Systems*, v. 3, n. 2, artigo 7, Julho de 2012.
- van der Aalst, W. M., Günther, C. W., “Finding structure in unstructured processes: The case for process mining”. In: *Seventh International Conference on Application of Concurrency to System Design*, pp. 3-12, Bratislava, Julho de 2007.
- van der Aalst, W. M., Weijters, T., Maruster, L., “Workflow Mining: Discovering Process Models from Event Logs”. *IEEE Transactions on Knowledge and Data Engineering*, v. 16, n. 9, pp. 1128–1142, Setembro de 2004.
- van der Aalst, W.M., Andriansyah, A., de Medeiros, A. K. A., Arcieri, F., Baier, T., Blickle, T., Bose, J.C., van den Brand, P., Brandtjen, R., Buijs, J., “Process Mining Manifesto”. In: *BPM 2011 Workshops Proceedings*, pp. 169–194, Clermont-Ferrand, França, Agosto de 2011.
- van der Aalst, W. P., De Beer, H., van Dongen, B., “Process Mining and Verification of Properties: An Approach Based on Temporal Logic”. In: *On the Move to Meaningful Internet Systems: CoopIS, DOA and ODBASE*, Springer Berlin Heidelberg, pp. 130-147, 2005.
- van der Aalst, W. P., van Hee, K., *Workflow Management: Models, Methods and Systems*. 1^a ed., Estados Unidos da América, The MIT Press, 2004.
- van der Aalst, W. P., Song, M., “Mining Social Networks: Uncovering Interaction Patterns in Business Processes”. In: *Business Process Management*, Springer Berlin Heidelberg, pp. 244-260, 2004.

- van Dongen, B., de Medeiros, A. K. A., Verbeek, H., *et al.*, “The ProM Framework: A New Era in Process Mining Tool Support”. In: *Applications and Theory of Petri*, Springer Berlin Heidelberg, pp. 444-454, 2005.
- Verbeek, H., *et al.*, “XES, XESame and ProM 6”. In: *Information Systems Evolution*, Springer Berlin Heidelberg, pp. 60-75, 2011.
- Völzer, H., “An Overview of BPMN 2.0 and Its Potential Use”. *Lectures Notes in Business Information Processing*, v. 67, Springer Berlin Heidelberg, pp. 14-15, 2011.
- Weijters, A., van der Aalst, W. M., de Medeiros, A. K. A., *Process mining with the heuristics miner-algorithm*. In: Tech. Rep. WP 166, Technische Universiteit Eindhoven, 2006.
- Weske, M., *Business Process Management: Concepts, Languages, Architectures*. 1^a ed., Alemanha, Springer-Verlag, 2007.
- Weske, M., van der Aalst, W. P., Verbeek, H., “Advances in Business Process Management”. *Data & Knowledge Engineering*, v. 50, n. 1, pp. 1-8, 2004.
- White, R., Kules, B., Drucker, S., “Supporting Exploratory Search: Introduction”. *Communications of the ACM*, v. 9, n. 4, pp. 36-39, 2006.
- Workflow Management Coalition, *Terminology & Glossary: A Workflow Management Coalition Specification*. Reino Unido, 1999.
- Zairi, M., “Business Process Management: A Boundaryless Approach to Modern Competitiveness”. *Business Process Management Journal*, v. 3, n. 1, pp. 64–80, 1997.
- zur Muehlen, M., Recker, J., “How Much Language is Enough? Theoretical and Practical Use of the Business Process Modeling Notation”. In: *Advanced Information Systems Engineering*, Springer Berlin Heidelberg, pp. 465-479, 2008.